Comprehensive monitoring of tissue composition using in vivo imaging of cell nuclei and deep learning

Amrita Das Gupta[a], Jennifer John [a], Livia Asan[a,c], Carlo Beretta [a], Thomas Kuner [a], Johannes Knabbe [a,b]

a Department of Functional Neuroanatomy, Institute for Anatomy and Cell Biology, Heidelberg University, Germany.

b Department of General Psychiatry, Centre for Psychosocial Medicine, Heidelberg University, Germany.

c Current address: Department of Neurology and Center for Translational Neuro- and Behavioral Sciences (C-TNBS), University Hospital Essen, Germany.

Correspondence: johannes.knabbe@uni-heidelberg.de, thomas.kuner@uni-heidelberg.de

Abstract

Comprehensive analysis of tissue composition has so far been limited to ex-vivo approaches. Here, we introduce NuCLear (Nucleus-instructed tissue composition using deep learning), an approach combining in vivo two-photon imaging of histone 2B-eGFP-labeled cell nuclei with subsequent deep learning-based identification of cell types from structural features of the respective cell nuclei. Using NuCLear, we were able to classify almost all cells per imaging volume in the secondary motor cortex of the mouse brain (0.25 mm^3 containing ~25000 cells) and to identify their position in 3D space in a non-invasive manner using only a single label throughout multiple imaging sessions. Twelve weeks after baseline, cell numbers did not change yet astrocytic nuclei significantly decreased in size. NuCLear opens a window to study changes in relative abundance and location of different cell types in the brains of individual mice over extended time periods, enabling comprehensive studies of changes in cellular composition in physiological and pathophysiological conditions.

Introduction

Understanding the plasticity and interaction of different brain cell types in vivo has been a limiting factor for the investigation of large-scale structural brain alterations associated with diverse physiological and pathophysiological states. Until now, to investigate the effect of an experimental intervention on multiple cortical cell types and their respective spatial interactions, each would require an individual labelling strategy using a set of specific promoters and fluorophores. This would require prioritizing cell types from the outset and thereby limit the scope to a small subset of the population. With multiple cell types to consider, the number of experiments and animals

investigated may become intractable. To date, studies assessing whole tissue composition quantitatively have primarily employed ex vivo approaches using manual or automated analysis of antibody staining in tissue sections (Ero, Gewaltig et al. 2018), isotropic fractionation (Herculano-Houzel and Lent 2005), in situ hybridization (Nagendran, Riordan et al. 2018) or single cell sequencing (Zeisel, Hochgerner et al. 2018). Other studies used manual (Schmitz and Hof 2007) or automated stereological approaches (Zhang, Yan et al. 2017, Pakkenberg, Olesen et al. 2019). All these studies were conducted ex vivo and a subset required the dissolution of the studied organ's cellular architecture. The ability to identify multiple cell types and their location in 3D space within a single subject, here referred to as 'composition', in the integrity of the living brain to quantify and observe changes over time delivers new opportunities.

Here, we propose an experimental approach to study neurons and glial cells, especially, astroglia, microglia and oligodendroglia, as well as endothelial cells in the same mouse in vivo, using a single genetically modified mouse line expressing a fusion protein of histone 2B and enhanced green fluorescent protein (eGFP) in all cell nuclei. Our approach uses a deep learning method, implementing artificial neural networks to classify each nucleus belonging to a specific cell type, making it possible to perform a nucleus-instructed tissue composition analysis using deep learning (NuCLear). Additionally, determining the precise coordinates of every nucleus within the imaged space allows one to assess spatial relations of cells, e.g. the degree of cell clustering vs. even distribution of cells. This can be utilized as an indirect marker of glial territory size, which has been proven to be relevant in pathogenesis of disease such as Alzheimer's (Bouvier, Jones et al. 2016), or give information about neuron-glia-vasculature proximity (Zisis, Keller et al. 2021). Apart from the application in longitudinal in vivo imaging of the mouse brain, the concept of this

technique could be applicable to many other cellular imaging techniques such as confocal and widefield imaging of different organs, when ground truth data for classifier training is available. This will make NuCLear a powerful approach for future analyses of large-scale automated tissue composition.

Results

For the comprehensive identification and tracking of cell type composition in the mouse brain over the course of weeks and even months, a mouse line constitutively expressing a human histone 2B-eGFP (H2B-eGFP) fusion protein in all cell nuclei was imaged using 2-photon microscopy after implantation of a chronic cranial window (Figure 1A, Supplementary Figure 1A). Volumetric images taken from this mouse line show a nuclei distribution resembling the well-known DAPI staining (Supplementary Figure 1B, C). The idea behind the approach was to use nuclei as proxy for cells and train a neuronal network with ground truth data to classify nuclei belonging to distinct cell types. In brief, the proposed method consists of three main steps: (1) Automated nuclei segmentation in the raw data (detection and labelling of nuclei), (2) feature extraction and (3) classification of individual nuclei using pretrained classifiers for each desired cell type (Figure 1B, C, C(i)). In our study we selected five different cell types for classification: neurons, astroglia, microglia, oligodendroglia, and endothelial cells. Neurons were further classified into inhibitory or excitatory subtypes.

To be able to train a classifier for each cell type, ground truth data had to be generated. In addition to the H2B-eGFP-labeled nuclei, a red fluorescent marker protein was introduced (tdTomato, DyLight 549 or mCherry) to mark a specific cell type. The colocalization of both the red and green

83 label allowed us to assign each nucleus to a specific cell type and to extract ground truth data for
84 nucleus classification in the green channel. Nuclei belonging to a specific cell type were manually
85 selected and their features were used to train a neuronal network classifier in a supervised way
86 (Figure 2A). To avoid a bleed through effect of the red fluorescent protein tdTomato into the eGFP
87 channel, which could influence the quality of the classification, we induced the Cre-ERT2-
88 dependent expression of tdTomato via intraperitoneal injection of tamoxifen (Supplementary
89 Figure 1D, right) for cell type identification *after* the first imaging of the nuclei (Figure 2B,
90 overlapping emission spectra for eGFP and tdTomato are shown in Supplementary Figure 1E,
91 bleed through of the tdTomato signal into the eGFP channel is demonstrated in Supplementary
92 Figure 1F). As microglia were the only cells to change their positions after induction with
93 tamoxifen, we chose acute injection of tomato-lectin coupled to DyLight 594 to label these cells,
94 which did not produce any signal alterations in the eGFP-channel (Supplementary Figure 1G).
95 Neurons were labelled using a cortically injected AAV expressing the mCherry red fluorescent
96 protein under the neuron-specific synapsin promotor (syn1-mCherry), which did not affect the
97 eGFP-signal due to its specific fluorescence properties with a more right-shifted emission
98 spectrum compared to tdTomato (Supplementary Figure 1E, G). To further distinguish between
99 different neuronal subtypes, excitatory and inhibitory neurons were labelled by AAV-mediated
100 expression of CamKIIα-mCherry and mDLX-mRuby2 respectively (Supplementary Figure 2A, B). To
101 prepare the ground truth datasets of tdTomato expressing mice, nuclei were manually selected
102 from the pre-induction timepoint, wherever the green and red fluorescence signals overlapped
103 in the post-induction timepoint images (Figure 2C, Supplementary Figure 4B). For the microglial

104 and neuronal datasets, nuclei that overlapped in the green and red channel were manually
105 selected (Figure 2C, Supplementary Figure 2B).
106 Automated segmentation of nuclei in the raw data was achieved using the StarDist neural
107 network (Schmidt, Weigert et al. 2018), trained on manually traced and labelled ground truth
108 datasets of the H2B-eGFP mouse line (Supplementary Figure 1H(i-ii),I). The trained network
109 showed a high nucleus detection accuracy of 94 % as well as a good shape segmentation of nuclei
110 (Supplementary Figure 1H (iii-iv), I). From the binary mask of individual nuclei and their respective
111 pixel intensities in the raw image, in total 107 features were extracted using the PyRadiomics
112 package (van Griethuysen, Fedorov et al. 2017), including 3D-diameter, flatness, grey level non-
113 uniformity, entropy, first order minimum or coarseness (Figure 2D) (for a full list and short
114 description of the features see Supplementary Tables 1 and 2). To reduce possible overfitting of
115 the classification algorithm, a subset of 12 features were automatically selected using a
116 sequential feature selection algorithm (Supplementary Table 2). Radar plots with these nuclear
117 features visualize the differences between the cell types, for example neuronal nuclei having a
118 larger 3D diameter and minor axis length in comparison to microglial and astroglial nuclei that
119 exhibit a larger pixel entropy (Figure 2E). For certain features, significant differences between the
120 cell types can be shown as well, for example flatness being significantly higher in nuclei of
121 microglia than in nuclei of neurons (Figure 2F). Furthermore, combinations of two or three
122 distinctive features allow for the visual separation of nuclei of different cell types in the 2D and
123 3D space (Supplementary Figure 3). When analyzing inhibitory and excitatory neurons, nuclei
124 showed significant differences in shape and texture features (Supplementary Figure 2C (i, ii, iii,
125 iv)).

126 Having demonstrated the ability to distinguish between cell types using only nuclear features, we
127 created a neural network model for classifying cell types based on their features (Supplementary
128 Figure 4A). Utilizing the dataset obtained from the five reporter mouse models, a single neural
129 network classifier was trained for each cell type. The purpose of each classifier was to distinctly
130 differentiate its corresponding cell type from the diverse array of other cell types (Figure 1C(i),
131 3A). To increase the amount of training data and equalize the nuclei counts for each cell type,
132 thus reducing training bias, synthetic data was generated from the features of the original dataset
133 (Supplementary Figure 4C). Synthetic data distribution fit well to the distribution of the original
134 data for each cell type (Supplementary Figure 4C, D). After each nucleus was classified by all
135 trained classifiers, it was either assigned to a single class (neurons, astrocytes, microglia,
136 oligodendroglia, endothelial cells), two or more classes (undecided) or to none (unclassified).
137 Precision and recall rates for the model were high for neurons and endothelial cells (Figure 3B).
138 Due to their relative similarity, glial cells exhibited lower precision and recall rates. The
139 classification accuracy for the entire training dataset was highest for neurons (98 %) and
140 endothelial cells (99 %), whereas the classifier showed slightly lower accuracies for all glial cell
141 types, especially for microglia (96 %) (Figure 3C, D, for the respective confusion plots see
142 Supplementary Figure 5). To be able to distinguish between inhibitory and excitatory neurons,
143 another classifier was trained on the ground truth data from excitatory and inhibitory cells. The
144 classifier had a 93% overall accuracy (Supplementary Figure 2D), enabling a good distinction
145 between inhibitory and excitatory neuronal cells. The 3D classification results for one example
146 cortical volume containing 20123 nuclei (Supplementary Figure 6A, B) visualizes the relative
147 abundance and positioning of cell types within the volume and their mutual spatial relationship

148 (see also Supplementary Movie 1 for an animated version of the segmentation and classification
149 process).
150
151 To apply NuClear in a longitudinal way in mice, we studied 16 imaging volumes from 8 mice (male,
152 12-14 weeks at baseline) in layer 2/3 of the secondary motor cortex 12 weeks apart. When the
153 trained model was applied, two features not chosen by the automated feature selection
154 algorithm differed significantly between the classes: neuronal nucleus volume (Figure 3E) and
155 endothelial mean surface to volume ratio, a measure for the sphericity of the nucleus (Figure 3F).
156 Both findings are expected for each of the cell types and hence demonstrate that even features
157 that were not used for training the neural network model are able to differentiate between the
158 classes. At baseline a density of 43000 neurons per mm^3 and of 57000 glial cells per mm^3
159 (including unclassified cells, that were counted as glial cells) could be calculated, resulting in a glia
160 to neuron ratio of 1.3 (Figure 3G). The density of excitatory neurons was 37000 cells/mm^3,
161 compared to 6000 cells/mm^3 for inhibitory neurons (Supplementary Figure 2E). The density of
162 cells 12 weeks after baseline imaging remained unchanged after correcting for multiple
163 comparisons (Figure 3H). When normalizing these counts to baseline, only the number of
164 astrocytes shows a trend to increase (Figure 3I), which could be either due to astrogliosis caused
165 by aging or by continued reaction to the chronic window implantation. Interestingly, astroglial
166 nuclear volume significantly decreased over time (Figure 3J), which might be due to altered
167 transcriptional activity known to occur in astrocytes with aging (Boisvert, Erikson et al. 2018,
168 Clarke, Liddelow et al. 2018), whereas the nuclei volume in inhibitory and excitatory neurons did
169 not change over time (Supplementary Figure 2F). This latter result illustrates that our method can

extract even subtle and unexpected changes from of a vast parameter space of different aspects of tissue composition.

Discussion

Here, we present a novel method for intravital, longitudinal, comprehensive, large scale and 3D determination of tissue composition based on a deep learning algorithm trained on radiomics features of cell nuclei (NuCLear -- Nucleus-instructed tissue composition using deep learning). We demonstrate that cell types and even subtypes can be reliably distinguished via the respective properties of their cell nucleus and that an inventory of cells and their 3D position can be generated for a large imaging volume. To demonstrate the usability of NuCLear, we analyzed volumetric images of the cortex of H2B-eGFP mice acquired with in vivo two photon microscopy. To be able to image a large dataset in a comparatively short time, we chose an imaging resolution with a z-step size of 2 µm and a lateral resolution of 0.3 µm. We were able to image a whole 3D volume of 700 μm^3 consisting of more than 20000 cells in around 20 minutes, making it possible to perform large scale data acquisition in a repeated manner, enabling longitudinal intravital monitoring of tissue histology over time and in response to perturbations, treatments, or adaptive processes such as learning or exploring.

Neurons and endothelial cells showed the best classification results regarding total accuracy and precision as well as recall rates (Figure 3B), as their nucleus feature profile turned out to be markedly different between each other as well as between the individual glial cell types (Figure 2E, F). Differentiation between the three different glial cell types was more difficult to achieve, resulting in lower overall accuracy rates as well as lower precision and recall rates, due to the

192 higher similarity of their respective nucleus feature profiles. Nevertheless, with an accuracy above
193 90% our approach can reliably be used to subclassify glia. We expect that a higher imaging
194 resolution as well as more training data would further increase the classifier performance.
195 Another reason for the comparatively lower classification accuracy in microglia might be due to
196 acute injection of tomato-lectin to label these cells, which could have introduced a bias towards
197 activated microglia. Nevertheless, the overall accuracy of the microglia classifier for the whole
198 training dataset was around 96%, owing to the fact, that non-microglial nuclei could be identified
199 correctly. For astrocytes, we used GFAP-labelled cells for classification, a label that primarily
200 marks reactive astrocytes and has a higher expression in older animals. Future studies could
201 optimize this labelling strategy. Additional confirmation of the classifier's precision can be inferred
202 by examining characteristics of the nuclei, like average nuclear volume and the surface-to-volume
203 ratio, which were not utilized in the training phase (Figure 3E, F). These features showed distinct
204 clustering and low standard deviation in each class as well as significant differences between the
205 cell types.
206
207 When applying the NuCLear to imaging volumes from Layer 2/3 of the secondary motor area and
208 calculating the number of cells per mm^3, a marked similarity of the number of neurons with
209 published results is evident (43413 ± 1003 vs 45374±4894) (Ero, Gewaltig et al. 2018) (Figure 3G).
210 Astrocyte numbers are similar (18629 ± 1007 vs 13258± 1416), microglia numbers differ more
211 (12411 ± 745 vs 24584.6± 2687). Oligodendrocyte numbers seem to be vastly different (4317 ±
212 154 vs 51436.1± 488), which might stem from the fact that only Layer 2/3 was analyzed, excluding
213 deeper cortical layers that have been shown to contain most oligodendrocytes (Tan, Kalloniatis et

214 al. 2009). When all glial cells as well as undecided and unclassified cells are summed up, a
215 Glia/Neuron ratio of 1.3 can be calculated, a result that is in line with known results of the rodent
216 cortex (Herculano-Houzel 2014). Candidates for unclassified cells could be the relatively large
217 population of oligodendrocyte precursor cells (about 5% of all cells in the brain (Dawson, Polito
218 et al. 2003)). After classifying the neurons into subtypes, 86% of all neurons were classified as
219 excitatory and 14% as inhibitory (37199±1387 vs 6214±1001 respectively), which is in par with
220 previous studies that reported 10-20% of neurons in layer 2/3 of the cortex being inhibitory
221 interneurons (Kummer, Mitric et al. 2020, Algamal, Russ et al. 2022). In conclusion, cell type
222 counts are comparable to published results, supporting the validity of our approach.
223
224 A decisive advantage of NuCLear over existing ex-vivo methods is the ability to study cell type
225 changes over time in the same imaging volume and thus achieving a higher statistical power in
226 experiments with fewer animals, which is crucial for complying with the “3R” rules in animal
227 research (reduce, replace, refine). We succeeded to image selected locations up to 1 year after
228 the implantation of chronic cranial windows. When analyzing the same imaging volumes from the
229 secondary motor cortex 12 weeks apart, we were able to detect a trend towards astrogliosis,
230 which might be due to the aging process leading to a higher GFAP-reactivity (Palmer and Ousman
231 2018). The significant decrease in mean nucleus volume of astroglia over time could be attributed
232 to altered transcriptional activity associated with aging (Boisvert, Erikson et al. 2018, Clarke,
233 Liddelow et al. 2018) since astroglial reactivity changes in the cortex over time (Clarke, Liddelow
234 et al. 2018). Another factor underlying astrogliosis might arise from the continued presence of
235 the chronic cranial window (Guo, Zou et al. 2017). Our study illustrates only one possible

application of NuCLear and subsequent analysis of 3D tissue composition. Here we focused on the identification of cell types, but the 3D datasets generated allow for additional analyses such as statistical distribution of cell types relative to each other, their nearest neighbor distances or inferences on physical tissue volume (Asan, Falfan-Melgoza et al. 2021). Our classifiers mostly rely on geometrical parameters such as nucleus shape, but also consider texture information to a certain degree. Pathophysiological conditions that alter nuclear shape and texture might influence classification accuracy. An increased sampling resolution may further increase textural information and thereby yield even higher classification accuracies. For now, segmentation and classification depend on the type of microscope used, quality of the images, depth of imaging as well the region which was analyzed. To make the classification more robust and accountable for different qualities of fluorescence signal, augmentations to the training data could be added such as 3D blurring with PSF-shaped convolutions.

NuCLear will be applicable to different organs as techniques have emerged in the last years to perform in vivo imaging through chronic windows in other rodent organs such as skin, abdominal organs, the tongue or spinal cord (Choi, Kwok et al. 2015), depending on the image quality and the possibility to acquire images during multiple timepoints. Other organisms and in vitro preparations like cell cultures or organoids can also be utilized, provided that the respective ground truths can be generated. It will be usable with a variety of microscopy techniques such as confocal-, light-sheet- or three-photon microscopy, making large-scale tissue analysis much more accessible. We assume that it will be possible to detect more cell types by adding classifiers trained with appropriate ground truth data. In this way, we propose a readily usable method to

258 implement large-scale tissue analysis ex-vivo or in-vivo to study effects of interventions on cell
259 type compositions of different organs.
260
261
262 **Materials and Methods**
263
264 Ethical approval
265 The entire study was conducted in line with the European Communities Council Directive
266 (86/609/ECC) to reduce animal pain and/or discomfort. All experiments followed the German
267 animal welfare guidelines specified in the TierSchG. The study has been approved by the local
268 animal care and use council (Regierungspräsidium Karlsruhe, Baden-Wuerttemberg) under the
269 reference numbers G294/15 and G27/20. All experiments followed the 3R principle and complied
270 with the ARRIVE guidelines (Percie du Sert, Hurst et al. 2020).
271
272 Animals
273 Adult transgenic mice expressing the human histone 2B protein (HIST1H2B) fused with enhanced
274 green fluorescence protein (eGFP) under the control of the chicken beta actin promoter (CAG-
275 H2B-eGFP) (Hadjantonakis and Papaioannou 2004) (B6.Cg-Tg(HIST1H2BB/EGFP)1Pa/J, Jackson
276 Laboratory; # 006069) were used in all animal experiments. For the reporter mouse lines, H2B-
277 eGFP mice were bred to carry a floxed sequence of the red fluorescent protein tdTomato and a
278 tamoxifen inducible Cre-recombinase under the expression of different cell type specific
279 promoters: GFAP (Glial Fibrillary Acidic Protein) for astroglia (HIST1H2BB/EGFP-GFAP/ERT2CRE-

280 CAG/loxP/STOP/loxP/tdTomato), PLP (Proteolipid Protein) for oligodendroglia (HIST1H2BB/EGFP-
281 PLP/ERT2CRE- CAG/loxP/STOP/loxP/tdTomato) and Tie2 (Tyrosine Kinase) for endothelial cells
282 (HIST1H2BB/EGFP-Tie2/ERT2CRE- CAG/loxP/STOP/loxP/tdTomato). Expression of the Cre-
283 recombinase was achieved over a course of up to 5 days with daily intraperitoneal injection of
284 two doses of 1mg Tamoxifen (Sigma Life sciences; dissolved in 1 part ethanol (99.8% absolute)
285 and 10 parts sunflower oil) (Figure 2B, Supplementary Figure 1D). Microglia were visualized using
286 intracortical injection of Lycopersicon Esculentum (Tomato) lectin coupled with DyLight 594
287 during the cranial window surgery after dilution by 1:39 in 150mM NaCl, 2.5 mM KCl, 10mM
288 HEPES at pH 7.4 (Schwendele, Brawek et al. 2012) (Figure 2B, Supplementary Figure 1D). For
289 neuronal labelling, an in house produced adeno-associated virus expressing mCherry under the
290 Synapsin promotor was cortically injected during the cranial window surgery (Figure 2B,
291 Supplementary Figure 1D). To label excitatory and inhibitory neurons, a viral labelling strategy
292 was implemented via intracortical injections, using AAV5-CamKIIα-mCherry (Addgene plasmid
293 114469) and AAV1-mDLX-mRuby2 (Addgene plasmid 99130) (Supplementary Figure 2) (Chan,
294 Jang et al. 2017). All mice were between 10-18 weeks old during baseline imaging (5 female/10
295 male).
296
297 <u>Chronic cranial window implantation</u>
298 A craniectomy procedure was performed on each mouse to enable in vivo two-photon imaging
299 following a predefined protocol as described before (Holtmaat, Bonhoeffer et al. 2009)
300 (Supplementary Figure 1A). Briefly, the mice were anesthetized with an intraperitoneal injection
301 (i.p.) of a combination of 60 µl Medetomidine (1mg/ml), 160 µl Midazolam (5 mg/ml) and 40 µl

302 Fentanyl (0.05 mg/ml) at a dosage of 3 µl/g body weight. The head was shaved, and the mice
303 were fixed with ear bars in the stereotactic apparatus and eye ointment was applied. Xylocain®
304 1 % (100 µl, Lidocaine hydrochloride) was applied under the cranial skin and 250 µl Carprofen (0.5
305 mg/ml) was injected subcutaneously (s.c.). The skin was removed to expose the skull and the
306 surface of the skull was made rough to allow the cement to adhere better. A skull island (approx.
307 6mm ∅) was drilled centered at 1 mm rostral to bregma using a dental drill and removed with a
308 forceps (#2 FST by Dumont) making sure not to damage the cortical surface. For improved imaging
309 condition, the dura was carefully removed from both hemispheres using a fine forceps (#5 FST by
310 Dumont). Normal rat ringer solution (NRR) was applied on the exposed brain surface to keep it
311 moist, and a curved cover glass was placed on top to cover it (Asan, Falfan-Melgoza et al. 2021).
312 With dental acrylic cement (powder: Paladur, Kulzer; activator: Cyano FAST, Hager & Werken
313 GmbH) the cover glass was sealed, and excess cement was used to cover the exposed skull and
314 edges of the skin. A custom designed 3-D printed holder was placed on top of the window and
315 any gaps were filled with cement to ensure maximum adhesion with the skull. After the procedure,
316 the mice were injected (i.p. / s.c.) with a mix of 30 µl Atipamezole (5 mg/ml), 30 µl Flumazenil
317 (0.1 mg/m) and 180 µl Naloxon (0.4 mg/ml) at a dosage of 6 µl per gram body weight. To ensure
318 proper recovery of the mice, 3 more doses of Carprofen were given every 8-12 hours and the
319 mice were placed onto a heating plate and monitored.
320

321 <u>In vivo two photon imaging</u>

322 Imaging was carried out with a two-photon microscope (TriM Scope II, LaVision BioTec GmbH)
323 with a pulsed Titanium-Sapphire (Ti:Sa) laser (Chameleon Ultra 2; Coherent) at excitation

324 wavelengths of 860nm for DyLight 594 and 960nm for H2GFP, tdTomato and mCherry respectively.
325 A water immersion objective (16x; NA 0.8, Nikon) was used to obtain volumetric 3D stacks.
326 Individual frames consisted of 694 µm x 694 µm in XY with a resolution of 0.29 µm/pixel. Stacks
327 were obtained at varying depths up to 702 µm from the cortical surface with a step size of 2 µm
328 in Z. Prior to each imaging session, the laser power was adjusted to achieve the best signal to
329 noise ratio. Adaptations were made to minimize the effect of laser attenuation due to tissue
330 depth by creating a z-profile and setting the laser power for different imaging depths while
331 making sure to minimize oversaturation of pixels.
332 Mice were initially anesthetized using a vaporizer with 6% isoflurane, eye ointment was applied,
333 and mice fixed in a custom-built holder on the microscope stage. Isoflurane was adjusted
334 between 0.5 – 1.5% depending on the breathing rate of each mouse to achieve a stable breathing
335 rate of 55 – 65 breaths per minute with an oxygen flow rate of 0.9 – 1.1 l/min. A heating pad was
336 placed underneath the mouse to regulate body temperature. An infrared camera was used for
337 monitoring of the mice during the imaging.
338 For ground truth training data reporter mouse lines for neurons and neuronal subtypes, astroglia,
339 oligodendroglia and endothelial cells were imaged 2-4 weeks after the cranial window surgery.
340 Microglia reporter mice were imaged immediately after the cranial window surgery for a minimal
341 inflammatory reaction (for a full list of reporter mice numbers see Supplementary Figure 4G). For
342 the data obtained from adult H2B-eGFP mice used for classification, 3D volumetric stacks were
343 imaged in the secondary motor cortex at two different timepoints. This included baseline imaging
344 which was performed at 3-4 weeks after the chronic cranial window surgery and 12 weeks after

345 the baseline timepoint. To investigate a possible age effect, mice underwent a sham surgery at
346 the left hind paw 1 week after baseline surgery as part of a different study.
347
348 Automatic nuclei segmentation using StarDist
349 Automated nuclei segmentation was achieved using the StarDist algorithm, which implements a
350 neural network to detect star-convex polygons in 2D and 3D fluorescence imaging (Schmidt,
351 Weigert et al. 2018). The StarDist model was trained to segment nuclei in 3D using an in-house
352 developed Jupyter Notebook. A GUI based version of the software for training of the
353 segmentation (NucleusAI) can be found here: https://github.com/adgpta/NuCLear. Nuclei were
354 manually traced and segmented using the segmentation editor in Fiji (Schindelin, Arganda-
355 Carreras et al. 2012) (Supplementary movie 2). In total, 15 different crops from the two-photon
356 volumetric data with approximately 150 – 200 nuclei each were used for training the StarDist 3D
357 segmentation classifier (Supplementary Figure 1I).
358 Detection accuracy and precision were calculated after visualization of raw data and
359 segmentation in ImageJ/Fiji. True positive nuclei (TP), false positive nuclei (FP) and false negative
360 nuclei (FN) were manually counted. 4 arbitrary volumetric images were selected containing up to
361 80 nuclei per crop. Accuracy was calculated as TP / (TP + FP + FN). Precision was calculated as TP
362 / (TP + FP).
363
364 Feature extraction of segmented nuclei using PyRadiomics
365 Using the PyRadiomics python package (van Griethuysen, Fedorov et al. 2017), in total 107
366 radiomics features were extracted for each nucleus after segmentation using StarDist, including,

367 but not limited, to morphological features, intensity features and texture features (see
368 Supplementary table 1). Nuclei touching the borders of the imaging volume were excluded.
369

370 Extraction of data for training of the classifier

371 To perform supervised training of the deep neural network for cell type classification, a ground
372 truth dataset was created using the two-photon volumetric fluorescence data, automatically
373 segmented nuclei, which were assigned a unique label, and the radiomics features. Using the red
374 fluorescence channel, nuclei belonging to a specific cell type were manually identified after
375 creating a composite image of the green and red fluorescence images (Supplementary Figure 4B).
376 This allowed for identification of the label by synchronizing the composite image and the StarDist
377 output in ImageJ/Fiji and extraction of radiomics features for these individual nuclei.
378 Approximately 70-400 nuclei were identified for each cell type (Supplementary figure 4G).
379 To increase the number of nuclei for training and make it equal in size for all cell types and thus
380 avoid a bias in training, synthetic datasets were created from nuclei features using the synthetic
381 data vault (SDV) package (Patki, Wedge et al. 2016), which utilizes correlations between features
382 of the original dataset as well as mean, minimum or maximum values and standard deviations.
383 The synthetic data generated by the model fit to the original dataset with approximately 83%
384 accuracy (Supplementary Figure 4D).
385

386 Training of the classification model

387 All training and validation of the classifiers were performed with a custom MATLAB (R2021a,
388 Mathworks) script, visualization was performed in ImageJ/Fiji (Supplementary movie 3). Original

389 and synthesized data were merged into a single table and random indices were generated to
390 divide the dataset into a training, test, and validation dataset with 70%, 15% and 15% of the total
391 dataset, respectively. Datasets with different amounts of synthetic data combined with the
392 original dataset were created and training accuracy was later compared between them: orig =
393 original dataset, orig * 2.5 = dataset containing 2.5 times the amount of data as the original
394 dataset, orig * 2.5 (down sampled) = dataset down sampled to minimum sample count (after 2.5
395 fold increase) to equalize sample numbers for all cell types, orig * 9 = dataset containing 9 times
396 the amount of data as the original dataset, orig * 9 (down sampled) = dataset down sampled to
397 minimum sample count (after 9 fold increase) to equalize sample numbers for all cell types
398 (Supplementary Figure 4E, F). Down sampled datasets were created by selecting the class with
399 the minimum number of nuclei and removing random samples from the other classes to match
400 the nuclei count of the selected class. The same training, test, and validation datasets were used
401 for training five different classifiers, one for each cell type. In the training dataset, for each
402 classifier, the class of interest was assigned a unique identifier and all other cell classes were
403 denoted by another identifier, for example, when training the neuronal classifier, the neurons
404 were labelled as “Neuron” while all the other cell types in the dataset were labelled “Other”
405 (Figure 1C(i), Figure 3A, Supplementary Figure 4A). Each classifier had the same design. Initially,
406 a sequential feature selection algorithm “sequentialfs” was applied to the whole dataset to
407 extract 12 features with the highest variability to reduce data overfitting. These were fed into the
408 “feature input layer” of the classifier. The features were z-score normalized for better
409 compatibility between each class. A “fullyconnected” layer of 50 outputs was created and a mini-
410 batch size of 16 samples was set to reduce the time and memory consumption during training. A

411 batch normalization layer was added to reduce sensitivity to the initialization of the network. The
412 activation layer ReLU (Rectified Linear Unit) was added for thresholding operations. Another
413 "fullyconnected" layer was added, which sets the number of outputs identical to the number of
414 input classes and a second activation layer "softmax" was chosen for classifying the output into
415 the two separate groups. The "adam" adaptive moment estimation optimizer (Kingma and Ba
416 2017) was selected as the solver for the training network which updates the network weights
417 based on the training data in an iterative manner. The training data and the validation data were
418 shuffled every epoch, i.e. a complete pass or iteration through the entire training dataset during
419 the training process, before training and network validation, respectively. After each classifier
420 was trained, further automatic and visual validation was performed to check for accuracy.
421
422 Classification of the H2B-eGFP data
423 After acquiring volumetric images of H2B-eGFP nuclei (n=8 mice, male), automated segmentation
424 of nuclei was performed with the StarDist neuronal network (see above, Supplementary Figure
425 6A, C). Features for the segmented nuclei were extracted using PyRadiomics (see above) and then
426 classified with all the different classifiers trained on the orig * 9 dataset (see "Training of the
427 classification model" for a description of the dataset). Thus, a single nucleus would be either
428 labelled as belonging to one of the five classes or to the "Other" class. Nuclei that were assigned
429 to multiple classes, were labelled as "Undecided" (UD). Any nuclei that were identified as "Other"
430 by all the classifiers were labelled as "Unclassified" (UC). Nuclei of neurons were further
431 subdivided into excitatory and inhibitory subtypes.
432

433 Statistical analysis

434 All classified nuclei and their features were stored in a local MySQL database (MySQL Workbench
435 8.0). The data were imported into R (R. Core Team 2021) for statistical analysis. Cell count, mean
436 nuclei volume and first nearest neighbor distance were calculated for each cell type for different
437 timepoints. Since the distribution of the data was non-parametric, the Wilcoxon signed-rank test
438 was used for all statistical testing and p-values were corrected for multiple comparisons using
439 the "bonferroni"-method. Plots were created using the R package "ggplot2" (Wickham 2016).

440

441 **Contributions**

442 ADG: Data acquisition, implementation and development of segmentation and classification
443 pipeline and scripts, training of classifiers, conceptual inputs, generation of figures, writing of the
444 MS.

445 LA: Acquisition of experimental ground truth data with reporter mouse lines using 2PLSM imaging
446 and establishment of tomato-lectin labelling, proof-of-concept with initial development of
447 solutions for segmentation and classification.

448 CAB: Development of segmentation, application of StarDist, contributions to classification.

449 JJ: Data acquisition, training of StarDist for segmentation.

450 TK: Initial project idea and concept, writing of the MS, supervision.

451 JK: Conceptualization, implementation, experimentation, analysis, writing of the MS, project
452 management and supervision.

453

454 **Data and code availability**

All datasets used in the study are available on the heiData archive: https://heidata.uni-heidelberg.de/dataset.xhtml?persistentId=doi:10.11588/data/L3PITA

Code and exemplary data can be found on the github page: https://github.com/adgpta/NuCLear

Acknowledgements

We thank Michaela Kaiser for her excellent technical support during preparation and execution of experiments. We thank Sidney Cambridge for devising breeding schemes for mouse lines needed for ground truth generation and for help with mouse breeding. We gratefully acknowledge the support by the German Research Foundation (DFG) (SFB1158, project B08 awarded to TK), the data storage service SDS@hd, supported by the Ministry of Science, Research and the Arts Baden-Württemberg (MWK) and the DFG through grant INST 35/1314-1 FUGG, as well as the high-performance cluster bwForCluster MLS&WISO, supported by the MWK and the DFG through Grant INST 35/1134-1 FUGG.

References

Algamal, M., A. N. Russ, M. R. Miller, S. S. Hou, M. Maci, L. P. Munting, Q. Zhao, D. Gerashchenko, B. J. Bacskai and K. V. Kastanenka (2022). "Reduced excitatory neuron activity and interneuron-type-specific deficits in a mouse model of Alzheimer's disease." Commun Biol **5**(1): 1323.

Arganda-Carreras, I., C. O. S. Sorzano, J. Kybic and C. Ortiz-de-Solórzano (2008). bUnwarpJ : Consistent and Elastic Registration in ImageJ . Methods and Applications.

Asan, L., C. Falfan-Melgoza, C. A. Beretta, M. Sack, L. Zheng, W. Weber-Fahr, T. Kuner and J. Knabbe (2021). "Cellular correlates of gray matter volume changes in magnetic resonance morphometry identified by two-photon microscopy." Sci Rep **11**(1): 4234.

Baddeley, A. and R. Turner (2005). "spatstat: An R Package for Analyzing Spatial Point Patterns." Journal of Statistical Software **12**(6): 1 - 42.

Boisvert, M. M., G. A. Erikson, M. N. Shokhirev and N. J. Allen (2018). "The Aging Astrocyte Transcriptome from Multiple Regions of the Mouse Brain." Cell Rep **22**(1): 269-285.

Bouvier, D. S., E. V. Jones, G. Quesseveur, M. A. Davoli, A. F. T, R. Quirion, N. Mechawar and K. K. Murai (2016). "High Resolution Dissection of Reactive Glial Nets in Alzheimer's Disease." Sci Rep **6**: 24544.

Chan, K. Y., M. J. Jang, B. B. Yoo, A. Greenbaum, N. Ravi, W. L. Wu, L. Sanchez-Guardado, C. Lois, S. K. Mazmanian, B. E. Deverman and V. Gradinaru (2017). "Engineered AAVs for efficient noninvasive gene delivery to the central and peripheral nervous systems." Nat Neurosci **20**(8): 1172-1179.

Choi, M., S. J. Kwok and S. H. Yun (2015). "In vivo fluorescence microscopy: lessons from observing cell behavior in their native environment." Physiology (Bethesda) **30**(1): 40-49.

Clarke, L. E., S. A. Liddelow, C. Chakraborty, A. E. Munch, M. Heiman and B. A. Barres (2018). "Normal aging induces A1-like astrocyte reactivity." Proc Natl Acad Sci U S A **115**(8): E1896-E1905.

Dawson, M. R., A. Polito, J. M. Levine and R. Reynolds (2003). "NG2-expressing glial progenitor cells: an abundant and widespread population of cycling cells in the adult rat CNS." Mol Cell Neurosci **24**(2): 476-488.

Ero, C., M. O. Gewaltig, D. Keller and H. Markram (2018). "A Cell Atlas for the Mouse Brain." Front Neuroinform **12**: 84.

Guo, D., J. Zou, N. Rensing and M. Wong (2017). "In Vivo Two-Photon Imaging of Astrocytes in GFAP-GFP Transgenic Mice." PLoS One **12**(1): e0170005.

Hadjantonakis, A. K. and V. E. Papaioannou (2004). "Dynamic in vivo imaging and cell tracking using a histone fluorescent protein fusion in mice." BMC Biotechnol **4**: 33.

Herculano-Houzel, S. (2014). "The glia/neuron ratio: how it varies uniformly across brain structures and species and what that means for brain physiology and evolution." Glia **62**(9): 1377-1391.

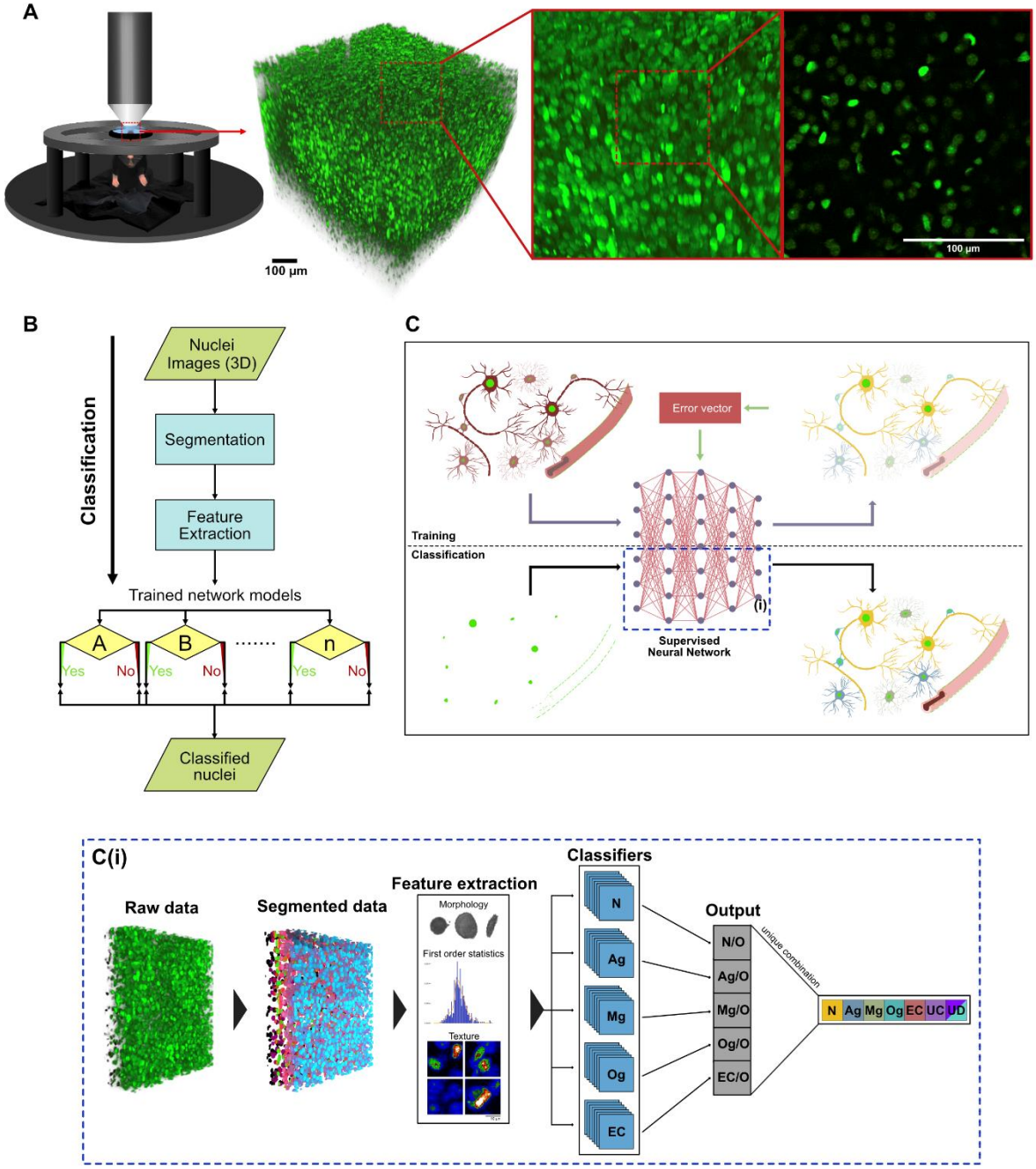
Herculano-Houzel, S. and R. Lent (2005). "Isotropic fractionator: a simple, rapid method for the quantification of total cell and neuron numbers in the brain." J Neurosci **25**(10): 2518-2521.

Holtmaat, A., T. Bonhoeffer, D. K. Chow, J. Chuckowree, V. De Paola, S. B. Hofer, M. Hubener, T. Keck, G. Knott, W. C. Lee, R. Mostany, T. D. Mrsic-Flogel, E. Nedivi, C. Portera-Cailliau, K. Svoboda, J. T. Trachtenberg and L. Wilbrecht (2009). "Long-term, high-resolution imaging in the mouse neocortex through a chronic cranial window." Nat Protoc **4**(8): 1128-1144.

Kingma, D. P. and J. Ba (2017). "Adam: A Method for Stochastic Optimization." arXiv preprint **arXiv:1412.6980**.

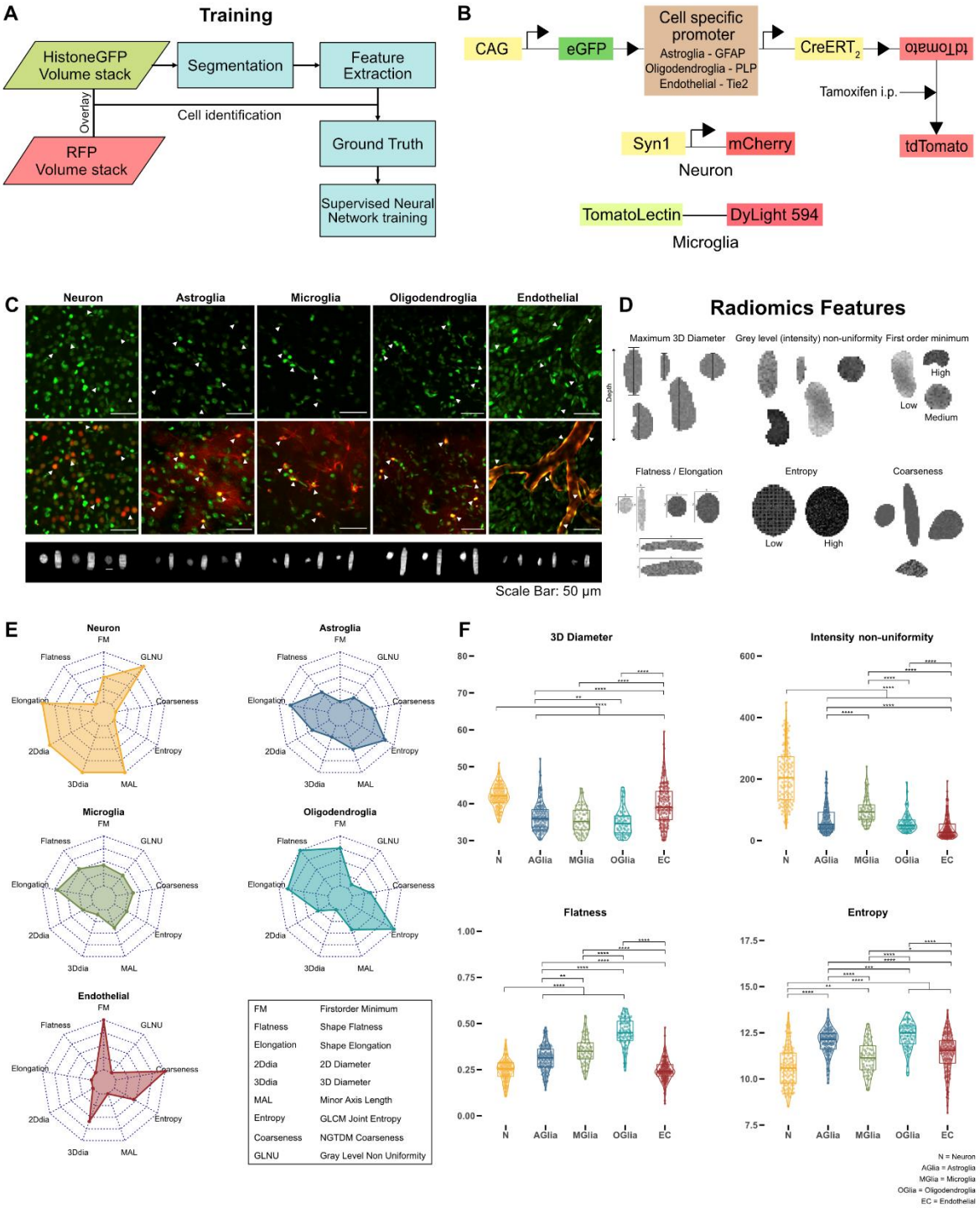
522 Kummer, K. K., M. Mitric, T. Kalpachidou and M. Kress (2020). "The Medial Prefrontal Cortex
523 as a Central Hub for Mental Comorbidities Associated with Chronic Pain." Int J Mol Sci **21**(10).
524 Nagendran, M., D. P. Riordan, P. B. Harbury and T. J. Desai (2018). "Automated cell-type
525 classification in intact tissues by single-cell molecular profiling." Elife **7**.
526 Pakkenberg, B., M. V. Olesen, S. S. Kaalund and K. A. Dorph-Petersen (2019). "Editorial:
527 Neurostereology." Front Neuroanat **13**: 42.
528 Palmer, A. L. and S. S. Ousman (2018). "Astrocytes and Aging." Front Aging Neurosci **10**: 337.
529 Patki, N., R. Wedge and K. Veeramachaneni (2016). The Synthetic Data Vault. 2016 IEEE
530 International Conference on Data Science and Advanced Analytics (DSAA).
531 Percie du Sert, N., V. Hurst, A. Ahluwalia, S. Alam, M. T. Avey, M. Baker, W. J. Browne, A.
532 Clark, I. C. Cuthill, U. Dirnagl, M. Emerson, P. Garner, S. T. Holgate, D. W. Howells, N. A.
533 Karp, S. E. Lazic, K. Lidster, C. J. MacCallum, M. Macleod, E. J. Pearl, O. H. Petersen, F.
534 Rawle, P. Reynolds, K. Rooney, E. S. Sena, S. D. Silberberg, T. Steckler and H. Wurbel (2020).
535 "The ARRIVE guidelines 2.0: Updated guidelines for reporting animal research." PLoS Biol
536 **18**(7): e3000410.
537 R. Core Team (2021). "R: A Language and Environment for Statistical Computing."
538 Schindelin, J., I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C.
539 Rueden, S. Saalfeld, B. Schmid, J. Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P.
540 Tomancak and A. Cardona (2012). "Fiji: an open-source platform for biological-image analysis."
541 Nat Methods **9**(7): 676-682.
542 Schmidt, U., M. Weigert, C. Broaddus and G. Myers (2018). Cell Detection with Star-Convex
543 Polygons, Cham, Springer International Publishing.
544 Schmitz, C. and P. R. Hof (2007). Design-Based Stereology in Brain Aging Research. Brain
545 Aging: Models, Methods, and Mechanisms. D. R. Riddle. Boca Raton (FL).
546 Schwendele, B., B. Brawek, M. Hermes and O. Garaschuk (2012). "High-resolution in vivo
547 imaging of microglia using a versatile nongenetically encoded marker." Eur J Immunol **42**(8):
548 2193-2196.
549 Tan, S. S., M. Kalloniatis, H. T. Truong, M. D. Binder, H. S. Cate, T. J. Kilpatrick and V. E.
550 Hammond (2009). "Oligodendrocyte positioning in cerebral cortex is independent of projection
551 neuron layering." Glia **57**(9): 1024-1030.
552 van Griethuysen, J. J. M., A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R. G. H.
553 Beets-Tan, J. C. Fillion-Robin, S. Pieper and H. Aerts (2017). "Computational Radiomics System
554 to Decode the Radiographic Phenotype." Cancer Res **77**(21): e104-e107.
555 Wickham, H. (2016). "ggplot2: Elegant Graphics for Data Analysis."
556 Zeisel, A., H. Hochgerner, P. Lonnerberg, A. Johnsson, F. Memic, J. van der Zwan, M. Haring, E.
557 Braun, L. E. Borm, G. La Manno, S. Codeluppi, A. Furlan, K. Lee, N. Skene, K. D. Harris, J.
558 Hjerling-Leffler, E. Arenas, P. Ernfors, U. Marklund and S. Linnarsson (2018). "Molecular
559 Architecture of the Mouse Nervous System." Cell **174**(4): 999-1014 e1022.
560 Zhang, C., C. Yan, M. Ren, A. Li, T. Quan, H. Gong and J. Yuan (2017). "A platform for
561 stereological quantitative analysis of the brain-wide distribution of type-specific neurons." Sci
562 Rep **7**(1): 14334.
563 Zisis, E., D. Keller, L. Kanari, A. Arnaudon, M. Gevaert, T. Delemontex, B. Coste, A. Foni, M.
564 Abdellah, C. Cali, K. Hess, P. J. Magistretti, F. Schurmann and H. Markram (2021). "Digital
565 Reconstruction of the Neuro-Glia-Vascular Architecture." Cereb Cortex **31**(12): 5686-5703.
566 **Figures**

Figure 1


A
100 µm
100 µm
B
Classification
Nuclei Images (3D)
Segmentation
Feature Extraction
Trained network models
A
B
n
Yes
No
Yes
No
Yes
No
Classified nuclei
C
Error vector
Training
Classification
Supervised Neural Network
(i)
C(i)
Raw data
Segmented data
Feature extraction
Morphology
First order statistics
Texture
Classifiers
N
Ag
Mg
Og
EC
Output
N/O
Ag/O
Mg/O
Og/O
EC/O
unique combination
N Ag Mg Og EC UC UD

568 **Figure 1. Imaging of cell nuclei and classification into different types. (A)** Illustration showing
569 acquisition of a 700 µm x 700 µm x 700 µm image volume of cell nuclei in the neocortex using in-
570 vivo 2-photon (2P) microscopy in a histone 2B-eGFP (H2B-eGFP) mouse. 3D reconstruction and
571 single imaging plane. **(B)** Flow diagram showing the classification process from acquisition of the
572 2P images until the final classification. **(C)** Illustration visualizing a simplified process of the
573 training and classification algorithms. Using an overlay of red (only one cell type labelled per
574 mouse) and green (nuclei with histone 2B-egfp expression: bright green) fluorescent images,
575 ground truth data were obtained to train a supervised neural network to classify nuclei into cell
576 types. Yellow: Neurons, blue: Astroglia, green: Microglia, turquoise: Oligodendroglia, red:
577 Endothelial cells **(C(i))** Visualization of the classification process described in (B) and the blue
578 dashed area in (C). Image volumes were obtained from H2B-eGFP mice, each nucleus was
579 automatically segmented. Features were extracted and each nucleus was classified using
580 pretrained classifiers. (N = Neuron, Ag = Astroglia, Mg = Microglia, Og = Oligodendroglia, EC =
581 Endothelial cells, UC = unclassified cells, UD: undecided cells (classified as belonging to multiple
582 classes)).

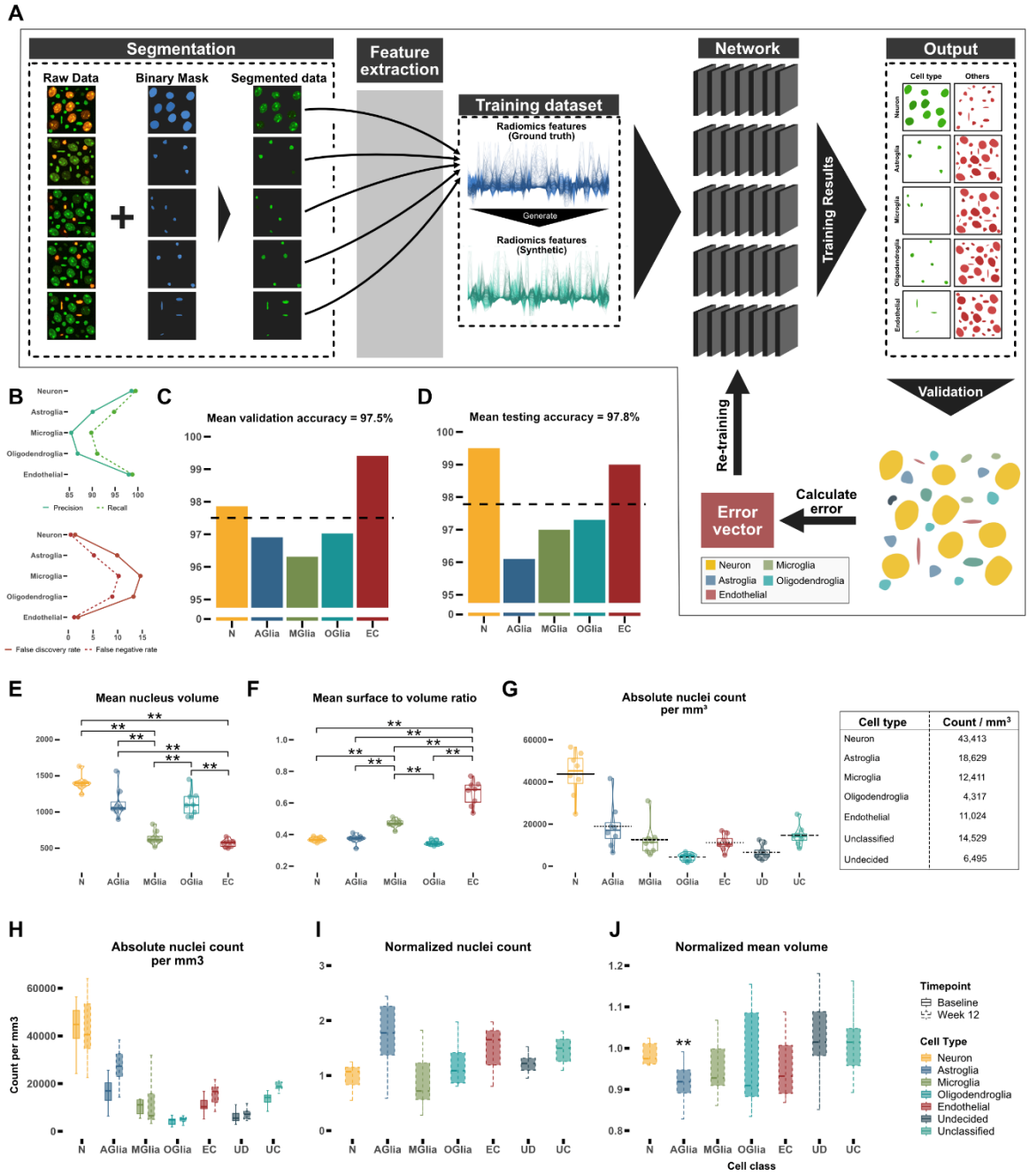
Figure 2


A
Training
HistoneGFP Volume stack
Segmentation
Feature Extraction
Overlay
Cell identification
RFP Volume stack
Ground Truth
Supervised Neural Network training
B
CAG
eGFP
Cell specific promoter
Astroglia - GFAP
Oligodendroglia - PLP
Endothelial - Tie2
CreERT2
tdTomato
Tamoxifen i.p.
tdTomato
Syn1
mCherry
Neuron
TomatoLectin
DyLight 594
Microglia
C
Neuron
Astroglia
Microglia
Oligodendroglia
Endothelial
Scale Bar: 50 µm
D
Radiomics Features
Maximum 3D Diameter
Grey level (intensity) non-uniformity
First order minimum
Depth
High
Low
Medium
Flatness / Elongation
Entropy
Coarseness
Low
High
E
Neuron
Astroglia
Microglia
Oligodendroglia
Endothelial
FM
GLNU
Coarseness
Entropy
MAL
3Ddia
2Ddia
Elongation
Flatness
FM Firstorder Minimum
Flatness Shape Flatness
Elongation Shape Elongation
2Ddia 2D Diameter
3Ddia 3D Diameter
MAL Minor Axis Length
Entropy GLCM Joint Entropy
Coarseness NGTDM Coarseness
GLNU Gray Level Non Uniformity
F
3D Diameter
Intensity non-uniformity
Flatness
Entropy
N
AGlia
MGlia
OGlia
EC
N = Neuron
AGlia = Astroglia
MGlia = Microglia
OGlia = Oligodendroglia
EC = Endothelial

583

584 **Figure 2. Strategies for training a neural network classifier using nucleus features. (A)** Flow
585 diagram showing classification training pipeline. Cell type specific red fluorescent proteins were
586 used to manually identify H2B-eGFP nuclei. This information was used to train a classifier for
587 each cell type. **(B)** Labelling strategies used for each cell type. Reporter mouse lines were
588 created by breeding H2B-eGFP mice to carry a floxed sequence of the red fluorescence protein
589 tdTomato and a tamoxifen inducible Cre-recombinase under the expression of different cell type
590 specific promoters which were used for identification of astroglia, oligodendroglia and
591 endothelial cells (see methods). Microglia and neurons were visualized using intracortical
592 injections of Lycopersicon Esculentum (Tomato) lectin and an AAV expressing mCherry under
593 the synapsin promoter respectively. **C)** Nuclei belonging to a specific cell type could be
594 identified by a red fluorescent marker (maximum intensity projections; z = 20 µm). 3D-
595 renderings of individual nuclei are shown in the lower panel. Scale bar: 50µm. **(D)** Illustrations
596 of a subset of radiomics features showing examples of different shape, intensity and texture
597 features. **(E)** Radar plots showing a subset of nuclear radiomics features for each cell type. **(F)**
598 Comparison of nuclear features (3D diameter in voxels, (voxel resolution in xyz: 0.29 µm x 0.29
599 µm x 2µm), intensity non uniformity, flatness, entropy) between cell types (Wilcoxon test,
600 Bonferroni correction for multiple comparisons, n of neurons: 135, n of astroglia: 137, n of
601 microglia: 62, n of oligodendroglia: 72, n of endothelial cells: 155, $p < 0.05$ *, $p < 0.01$ **, $p <$
602 0.001 ***), N = Neuron, AGlia = Astroglia, MGlia = Microglia, OGlia = Oligodendroglia, EC =
603 endothelial cells.
604

Figure 3

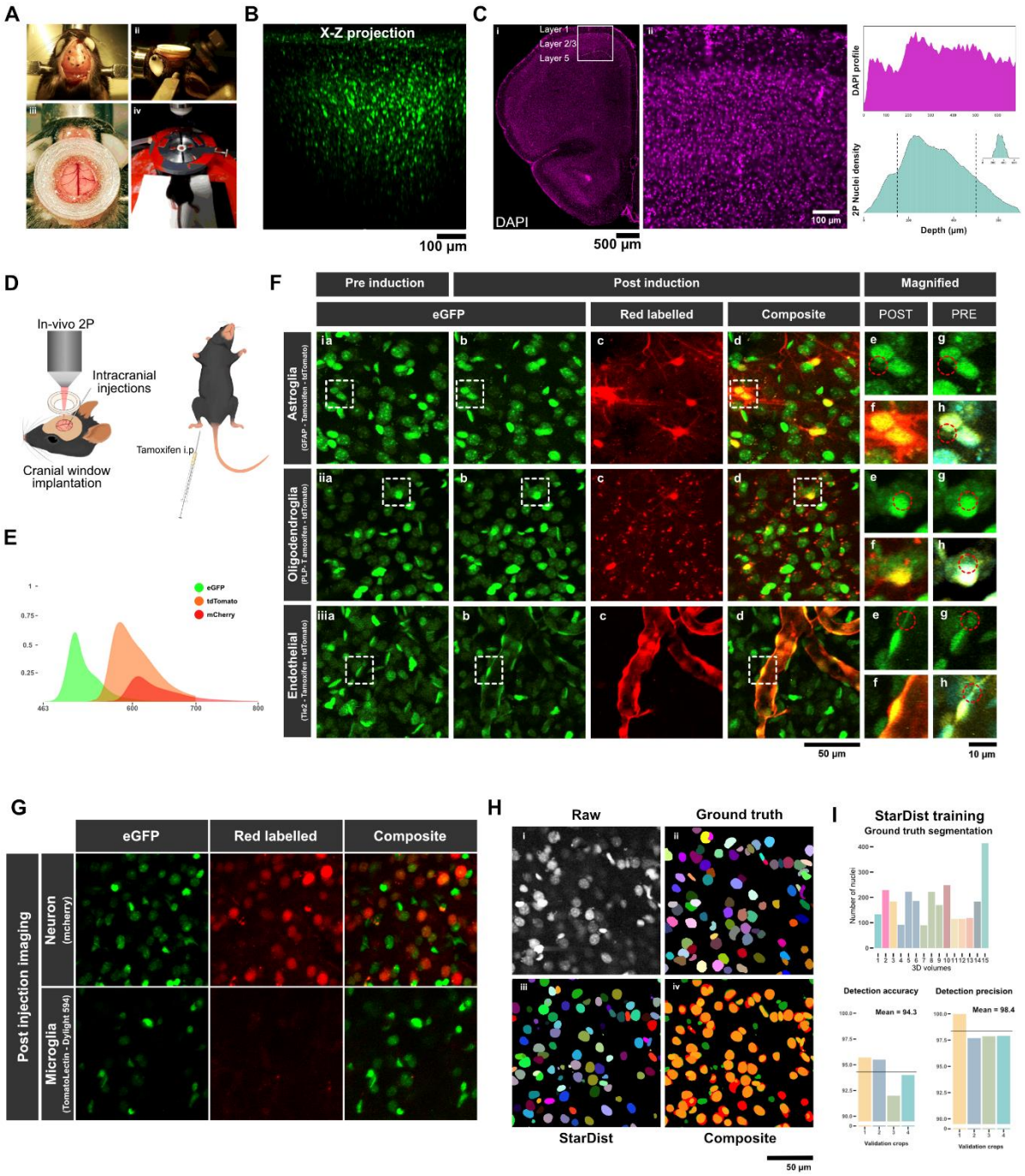


Cell type	Count / mm³
Neuron	43,413
Astroglia	18,629
Microglia	12,411
Oligodendroglia	4,317
Endothelial	11,024
Unclassified	14,529
Undecided	6,495

605

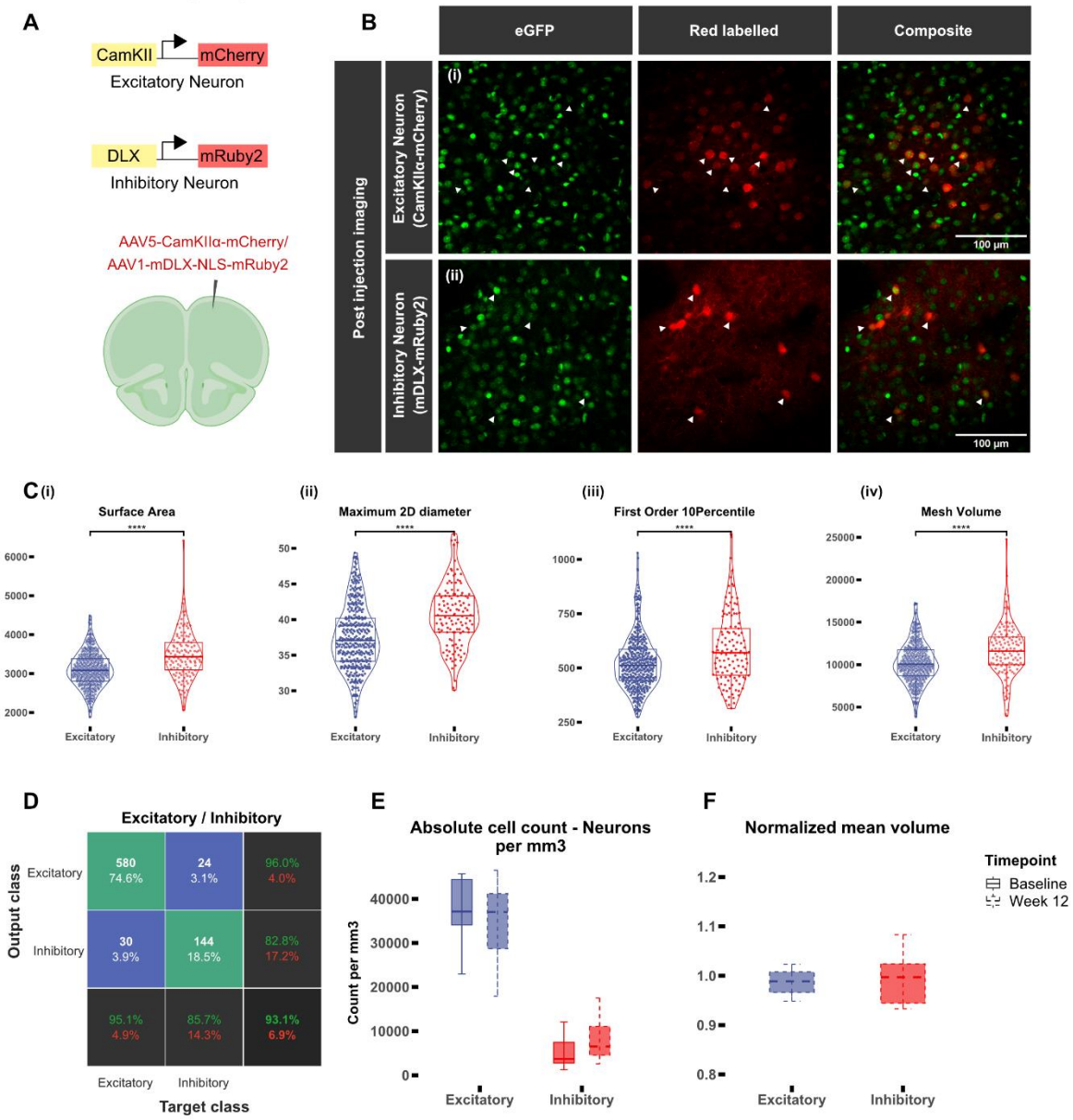
606 **Figure 3. Training a neural network for cell type classification. (A)** Schematic depiction of the
607 training process. After segmenting nuclei of the raw data with StarDist for every cell type,
608 radiomics features of nuclei were extracted from the ground truth data (blue) and synthetic data
609 (green) was generated. For each cell type, a single neuronal network was trained to distinguish
610 the corresponding nucleus type from the other nuclei. **(B)** Precision, recall, false discovery rate
611 and false negative rates for the test dataset. **(C, D)** Mean accuracy for all cell types in the validation
612 dataset (15 % of the ground truth dataset) and the test dataset (15 % of the ground truth dataset).
613 **(E, F)** Comparison of mean nucleus volume (unit: μm^3) and mean nuclear surface to volume ratio
614 (unit: μm^{-1}) of each cell type for classified volumes. **(G)** Number of nuclei per mm^3 in the
615 secondary motor cortex. **(H)** Number of nuclei per mm^3 in the secondary motor cortex 12 weeks
616 apart. **(I)** Comparison of normalized number of nuclei after 12 weeks. **(J)** Mean nucleus volume
617 after 12 weeks normalized to baseline. (Significance testing for C, D, E, F, G, H, I, J: Wilcoxon test,
618 p-values were corrected for multiple comparisons using the Bonferroni-method, n = 8 mice, $p <$
619 0.05*, $p < 0.01$ **, $p < 0.001$ ***), N = Neuron, AGlia = Astroglia, MGlia = Microglia, OGlia =
620 Oligodendroglia, EC = endothelial cells.
621
622 **Supplementary figures**

Supplementary Figure 1



624 **Supplementary figure 1 (A) (i,ii,iii)** Chronic cranial window implantation using a curved glass
625 cover slip and a custom 3D printed holder. **(iv)** Fixation of the anesthetized mouse in the custom
626 holder for imaging under the 2-photon microscope. **(B)** X-Z maximum intensity projection of a
627 two-photon volumetric stack (700µm x 700µm x 700µm) **(C)** Imaging position on a DAPI stained
628 brain slice. Slice thickness 50 µm. DAPI intensity profile showing similar distribution of intensity
629 as a two-photon image of nuclei. The decreasing nucleus density in the two-photon image stack
630 is a result of attenuation of fluorescence signal at higher depths. Inlay: Nucleus density
631 distribution of sub-volumes used for data analysis. **(D)** Scheme showing intracranial and
632 intraperitoneal injection strategies for labelling cell types. **(E)** Fluorescence emission spectra for
633 eGFP, tdTomato and mCherry (y axis: fluorophore emission normalized to quantum yield, source:
634 fpbase.org) **(F)** Labelling strategies for red fluorescence expression in reporter mouse lines.
635 Visualization of crosstalk between the eGFP and tdTomato signal after induction with tamoxifen.
636 Overlay of pre (cyan) and post (yellow) GFP after image alignment with the ImageJ plugin
637 bUnwarpJ (Arganda-Carreras, Sorzano et al. 2008). **(G)** No crosstalk is visible between GFP and
638 mCherry signals (upper panel) or eGFP and Tomato lectin-Dylight 594 signals (lower panel). **(H)**
639 (i) Raw data of H2B-eGFP signal (ii) manually labelled ground truth (iii) StarDist segmentation (iv)
640 composite image of ground truth (red) and StarDist segmentation (green). **(I)** Upper panel: Count
641 of manually segmented ground truth nuclei for StarDist training, each color depicts an individual
642 mouse. Lower panel: StarDist nucleus detection accuracy and precision, each bar represents an
643 individual imaging volume.

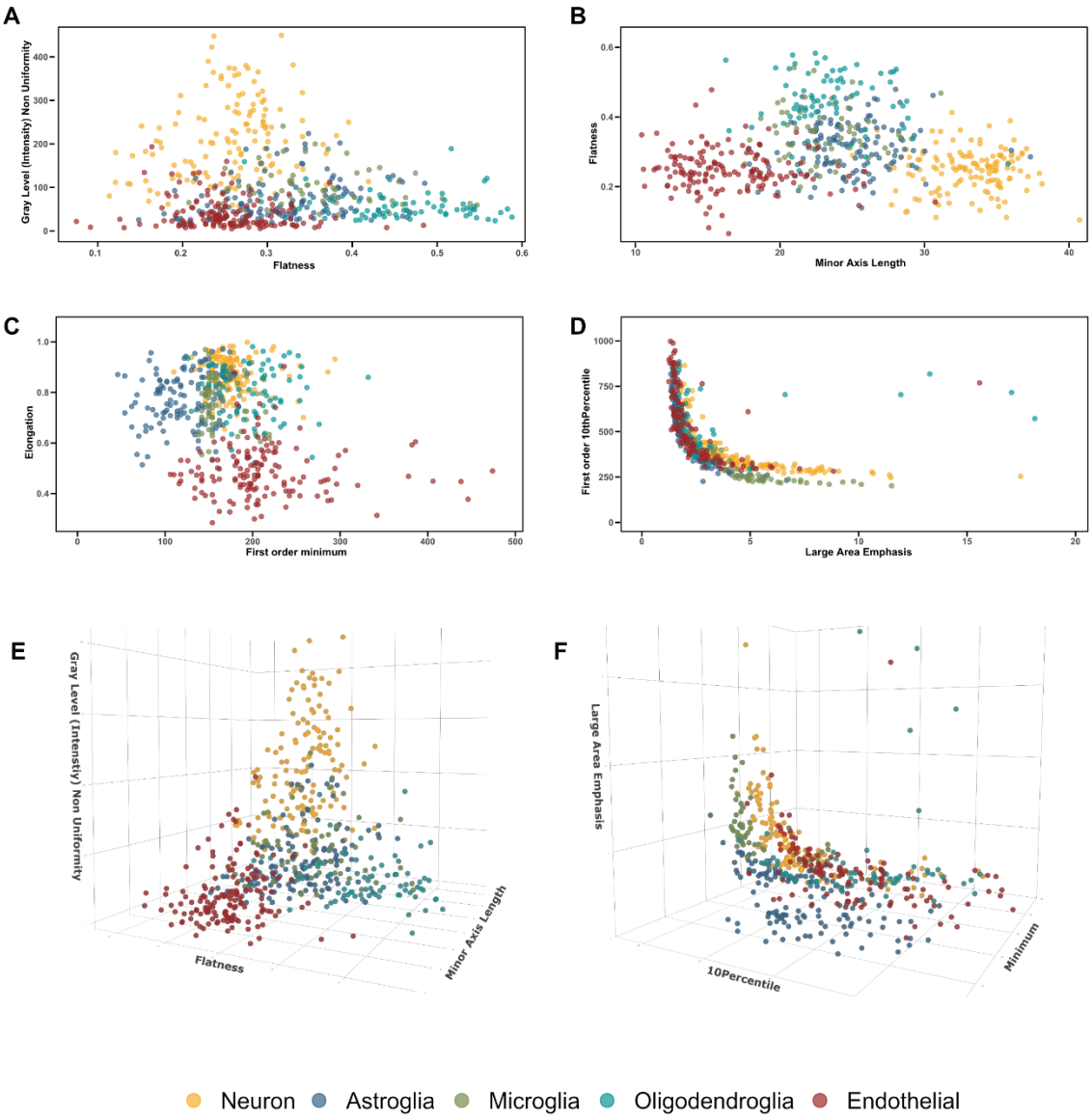
Supplementary Figure 2



644

645 **Supplementary Figure 2. Training a classifier to distinguish excitatory from inhibitory neurons.**
646 **(A)** Labelling strategy for excitatory and inhibitory neurons. Intracortical injections were
647 performed at ~400µm deep from the cortex of H2B-eGFP mice using AAV5-CamKIIα-mCherry and
648 AAV1-mDLX-NLS-mRuby2 to visualize excitatory and inhibitory neurons. **(B)** Post-injection images**.**
649 Inhibitory neurons were imaged 2-3 weeks after injection, excitatory neurons were imaged 3-4
650 weeks after injection. **(C)** Radiomics features showing differences between excitatory and
651 inhibitory neurons for (i) surface area (in pixel ^{2} [resolution x = 0.29 µm, y = 0.29 µm, z = 2 µm])
652 (ii) maximum 2D diameter (in pixel, same resolution as in (i)) (iii) first order 10th percentile (iv)
653 mesh volume in voxels (resolution as in (i)), (n of excitatory neurons: 396, n = 3 mice, n of
654 inhibitory neurons = 122, n = 2 mice). **(D)** Confusion plot for the classifier. Rows show the
655 predicted class (output class), and the columns show the true class (target class). Green fields
656 illustrate correct identification whereas blue fields illustrate erroneous identifications. The
657 number of observations and the percentage of observations compared to the total number of
658 observations are shown in each cell. Column on the far right shows the precision (or positive
659 predictive value) and false discovery rate in green and red respectively. Bottom row denotes recall
660 (or true positive rate) and false negative rate in green and red. Cell on the bottom right shows
661 overall accuracy of the classifier. **(E)** Number of nuclei per mm^3 in the secondary motor cortex at
662 baseline after 12 weeks (dashed line) (n = 8 mice). **(F)** Mean nucleus volume after 12 weeks
663 normalized to baseline (n = 8 mice). (Significance testing for C, E, F: Wilcoxon test, p-values were
664 corrected for multiple comparisons using the Bonferroni-method, $p < 0.05$*, $p < 0.01$ **, $p < 0.001$
665 ***).

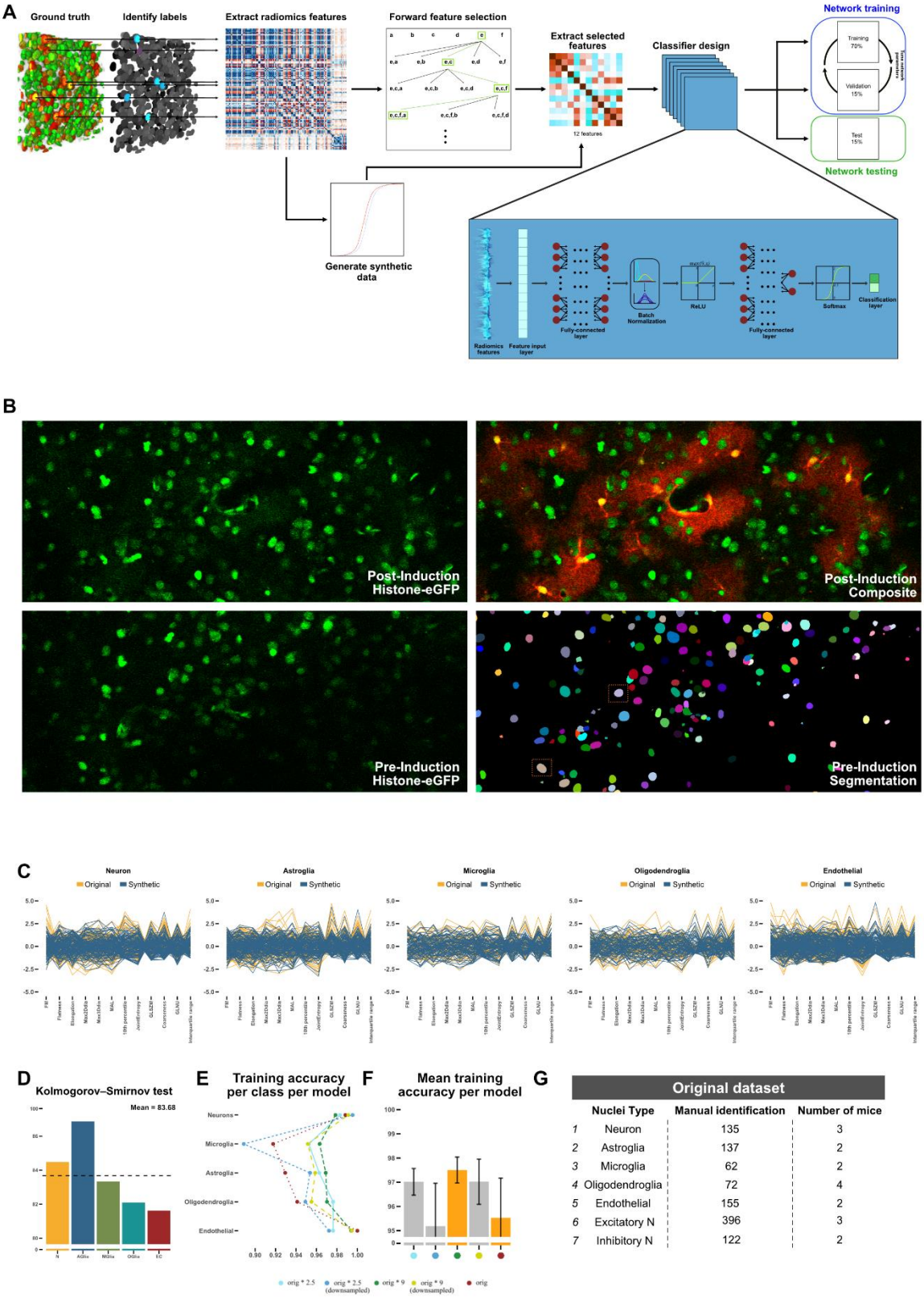
Supplementary Figure 3



666

667 **Supplementary Figure 3.** Combinations of nuclear morphology, intensity, and texture features
668 show clear distinction between cell types. **(A)** Flatness, gray level non-uniformity. **(B)** Minor axis
669 length, flatness. **(C)** First order minimum, elongation. **(D)** Large area emphasis, first order 10th
670 percentile. **(E, F)** Combinations of features in three dimensions.

Supplementary Figure 4



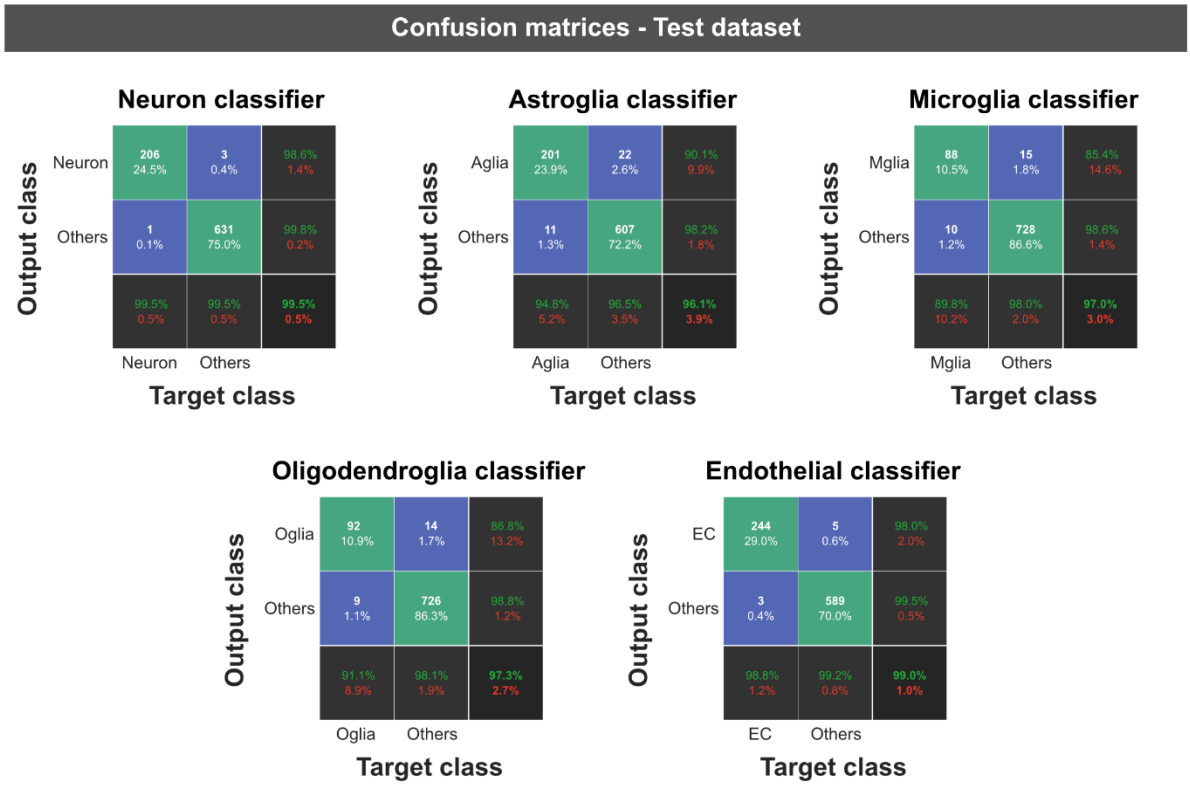
Original dataset			
	Nuclei Type	Manual identification	Number of mice
1	Neuron	135	3
2	Astroglia	137	2
3	Microglia	62	2
4	Oligodendroglia	72	4
5	Endothelial	155	2
6	Excitatory N	396	3
7	Inhibitory N	122	2

671

672 **Supplementary Figure 4 (A)** Visualization of the entire classification training process. After ground
673 truth data were selected, a sequential forward feature selection algorithm was applied to
674 extracted features from all nuclei of all cell types, which selected 12 features from the 107
675 radiomics features. Synthetic data was generated from all the radiomics features of all nuclei and
676 all cell types. The combined dataset was used to train a classifier with the training, validation and
677 test data comprising 70%, 15%, and 15% of the combined dataset. **(B)** Example showing manual
678 selection of ground truth data for supervised training. Post-induction green and red channels
679 were overlayed to create a composite. Any nuclei appearing yellow (possessing green and red
680 fluorescence) were identified in the pre-induction GFP channel and from its corresponding
681 segmentation, a label id was acquired and later used to identify the extracted features. **(C)**
682 Synthetic training data generated from the original dataset matched the features of the original
683 datasets. **(D)** Statistical analysis of similarity between the distribution of original data and
684 distribution of synthetic data (K-S test; mean = 83.68%). **(E)** Datasets with different amounts of
685 synthetic data were created and training accuracy was compared between them, orig = original
686 dataset, orig * 2.5 = dataset containing 2.5 times the amount of data as the original dataset, orig
687 * 2.5 (down sampled) = dataset down sampled to minimum sample count (after 2.5 fold increase)
688 to equalize sample numbers for all cell types, orig * 9 = dataset containing 9 times the amount
689 of data as the original dataset, orig * 9 (down sampled) = dataset down sampled to minimum
690 sample count (after 9 fold increase) to equalize sample numbers for all cell types. **(F)** Mean
691 accuracy of classifiers trained 5 times with different combinations of synthetic data; error bars
692 denote standard deviation (SD). light blue = orig * 2.5, dark blue = orig * 2.5 (downsampled),
693 green = orig * 9, yellow = orig * 9 (down sampled), red = orig. **(G)** Table listing the amount of

694 manually identified nuclei used in the training process (Excitatory N: excitatory neurons,
695 Inhibitory N: inhibitory neurons).

Supplementary Figure 5

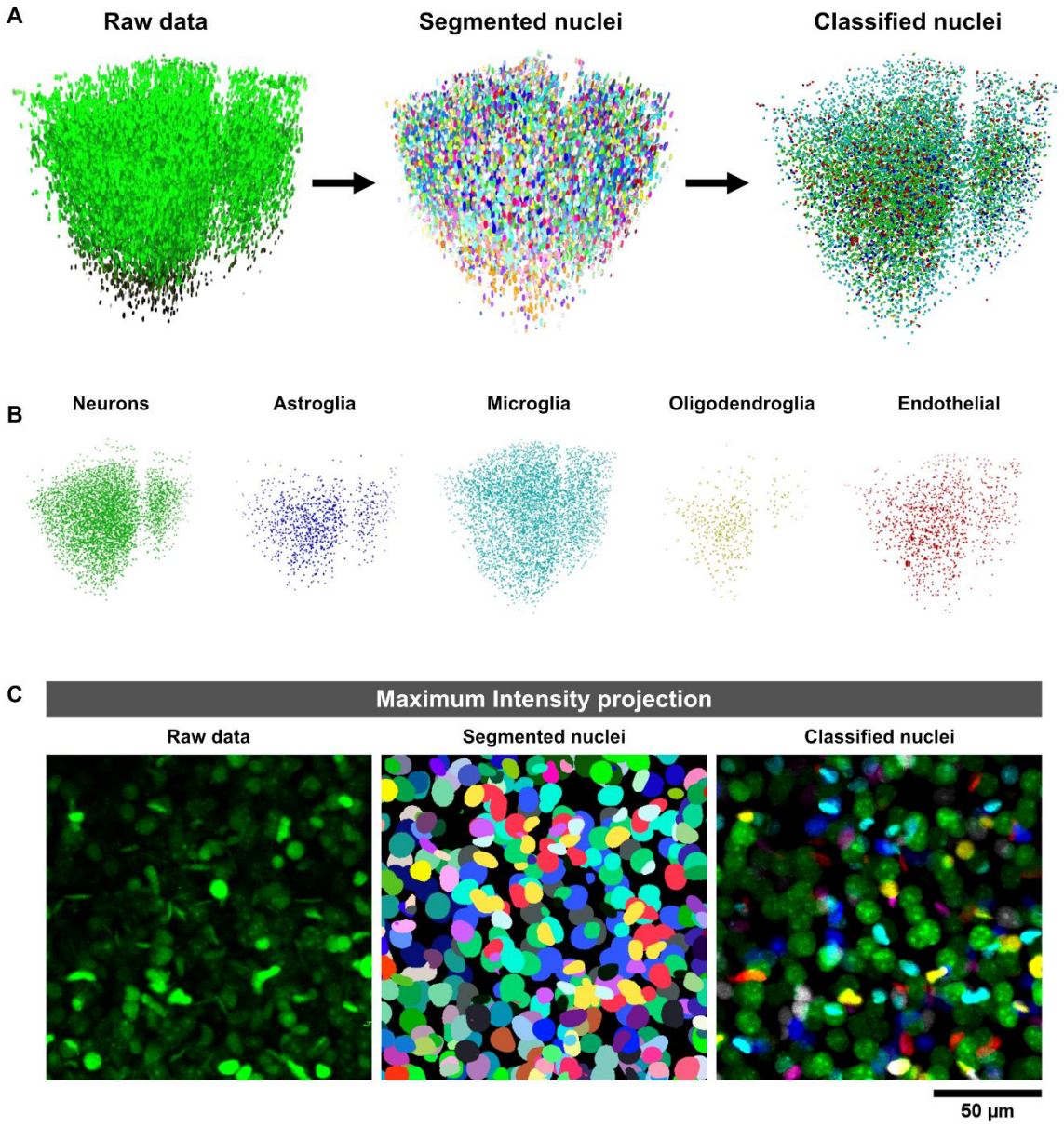


696

Supplementary Figure 5. Confusion matrices for each classifier for the test dataset. Since each classifier distinguishes between the desired class and every “other” class, the confusion matrix consists only of 4 fields. Rows show the predicted class (Output Class), and the columns show the true class (Target Class). Green fields illustrate correct identification of target and “other” class, blue fields illustrate erroneous identifications. The number of observations and the percentage of observations compared to the total number of observations are shown in each cell. Column on the far right shows the precision (or positive predictive value) and false discovery rate in green and red respectively. Bottom row denotes recall (or true positive rate) and false negative rate in

705 green and red. Cell on the bottom right shows overall accuracy of the classifier.

Supplementary Figure 6



Supplementary Figure 6. (A) 3-dimensional representation of raw, segmented and classified data (700 µm x 700 µm x 700 µm volume) visualizing comprehensive cell type composition. **(B)** Cell type distribution in a single volumetric stack. **(C)** Maximum intensity z-projection of a sub volume (150 µm x 150 µm x 100 µm) showing raw, segmented and classified nuclei in the X-Y axis.

Radiomics features							
Morphology		Intensity	Texture				
2-Dimensional	3-Dimensional	First Order Statistics	Gray Level Co-occurence Matrix (GLCM)	Gray Level Run Length Matrix (GLRLM)	Gray Level Size Zone Matrix (GLSZM)	Neighboring Gray Tone Different Matrix (NGTDM)	Gray Level Dependence Matrix (GLDM)
Mesh Surface	Mesh Volume	Energy	Autocorrelation	Short Run Emphasis (SRE)	Small Area Emphasis (SAE)	Coarseness	Small Dependence Emphasis (SDE)
Pixel Surface	Voxel Volume	Total Energy	Joint Average	Long Run Emphasis (LRE)	Large Area Emphasis (LAE)	Contrast	Large Dependence Emphasis (LDE)
Perimeter	Surface Area	Entropy	Cluster Prominence	Gray Level Non-Uniformity (GLN)	Gray Level Non-Uniformity (GLN)	Busyness	Gray Level Non-Uniformity (GLN)
Perimeter to Surface ratio	Surface Area to Volume ratio	Minimum	Cluster Shade	Gray Level Non-Uniformity Normalized (GLNN)	Gray Level Non-Uniformity Normalized (GLNN)	Complexity	Dependence Non-Uniformity (DN)
Sphericity	Sphericity	10th percentile	Cluster Tendency	Run Length Non-Uniformity (RLN)	Size-Zone Non-Uniformity (SZN)	Strength	Dependence Non-Uniformity Normalized (DNN)
Spherical Disproportion	Compactness 1	90th percentile	Contrast	Run Length Non-Uniformity Normalized (RLNN)	Size-Zone Non-Uniformity Normalized (SZNN)		Gray Level Variance (GLV)
Maximum 2D diameter	Compactness 2	Maximum	Correlation	Run Percentage (RP)	Zone Percentage (ZP)		Dependence Variance (DV)
Major Axis Length	Spherical Disproportion	Mean	Difference Average	Gray Level Variance (GLV)	Gray Level Variance (GLV)		Dependence Entropy (DE)
Minor Axis Length	Maximum 3D diameter	Median	Difference Entropy	Run Variance (RV)	Zone Variance (ZV)		Low Gray Level Emphasis (LGLE)
Elongation	Maximum 2D diameter (Slice)	Interquartile Range	Difference Variance	Run Entropy (RE)	Zone Entropy (ZE)		High Gray Level Emphasis (HGLE)
	Maximum 2D diameter (Column)	Range	Joint Energy	Low Gray Level Run Emphasis	Low Gray Level Zone Emphasis		Small Dependence Low Gray Level Emphasis
	Maximum 2D diameter (Row)	Mean Absolute Deviation (MAD)	Joint Entropy	High Gray Level Run Emphasis	High Gray Level Zone Emphasis		Small Dependence High Gray Level Emphasis
	Major Axis Length	Robust Mean Absolute Deviation (rMAD)	Informational Measure of Correlation (IMC) 1	Short Run Low Gray Level Emphasis	Small Area Low Gray Level Emphasis		Large Dependence Low Gray Level Emphasis
	Minor Axis Length	Root Mean Squared (RMS)	Informational Measure of Correlation (IMC) 2	Short Run High Gray Level Emphasis	Small Area High Gray Level Emphasis		Large Dependence High Gray Level Emphasis
	Least Axis Length	Standard Deviation	Inverse Difference Moment (IDM)	Long Run Low Gray Level Emphasis	Large Area Low Gray Level Emphasis		
	Elongation	Skewness	Maximal Correlation Coefficient (MCC)	Long Run High Gray Level Emphasis	Large Area High Gray Level Emphasis		
	Flatness	Kurtosis	Inverse Difference Moment Normalized (IDMN)				
		Variance	Inverse Difference (ID)				
		Uniformity	Inverse Difference Normalized (IDN)				
			Inverse Variance				
			Maximum Probability				
			Sum Average				
			Sum Entropy				
			Sum of Squares				

Supplementary Table 1: Radiomics features extracted from segmented nuclei. Features marked in green were used for cell type training and classification.

Selected Features			Definition
Morphology	3-Dimensional	Maximum 3D diameter	Largest pairwise Eucledian distance between two surface vertices
		Maximum 2D diameter (Row)	Largest pairwise Eucledian distance between two surface vertices in the sagittal plane
		Minor Axis Length	Second largest axis length of the ROI-enclosing ellipsoid
		Elongation	Relationship between two largest principal components in the ROI
		Flatness	Relationship between the largest and the smallest principal components in the ROI
Intensity	First order statistics	Minimum	Minimum gray level intensity within the ROI
		10th percentile	10th percentile of all the voxels in the ROI
		Interquartile Range	Difference between the 75th and 25th Percentile
Texture	Gray Level Co-occurence Matrix (GLCM)	Joint Entropy	Measure of randomness or variability in the neighborhood intensity values
	Gray Level Run Length Matrix (GLRLM)	Gray Level Non-Uniformity (GLN)	Similarity of gray-level intensity values in the image
	Gray Level Size Zone Matrix (GLSZM)	Large Area Emphasis (LAE)	Measures the distribution of large area size zones. Higher value indicates more coarseness
	Neighboring Gray Tone Difference Matrix (NGTDM)	Coarseness	Measures the average difference between the center voxel and its neighborhood

Supplementary Table 2: Morphology, intensity and texture features used for training and classification.

Supplementary Movie 1. Illustration of nucleus segmentation and classification using NuCLear. Raw data of H2B-eGFP-labeled nuclei (green), nucleus segmentation (randomly colored) and classification (apricot = neurons, red = microglia, blue = oligodendrocytes, dark green = astrocytes, magenta = endothelial cells)

Supplementary Movie 2. Visualization of the manual nucleus segmentation process.

Supplementary Movie 3. Visualization of a classified 3D volume

Supplementary Text 1: Explanations for technical terms used in the manuscript

Explanations for technical terms used in the manuscript:

Accuracy: Accuracy is a performance metric that quantifies the proportion of correctly classified data (both true positives and true negatives) out of all the data.

Augmentation: Augmentation involves applying various transformations, such as rotation, flipping, scaling, or adding noise, to the training data to increase its diversity. It helps improve the generalization and robustness of machine learning models.

Binary mask: A binary mask is a binary image where pixels are classified as either "foreground" (usually denoted as white) or "background" (usually denoted as black). Binary masks are commonly used to represent segmented regions, where each pixel is either part of the segmented structure (foreground) or not (background).

Coarseness: Coarseness is a texture feature that describes the scale of intensity variations within an image. It quantifies the contrast between regions of different intensities and their spatial frequency. A higher value indicates a lower spatial change rate and a locally more uniform texture.

Confusion plots: Confusion plots, also known as confusion matrices, are used to visualize the performance of a classification algorithm. They display the number of true positive, true negative, false positive, and false negative predictions made by the algorithm.

Crops: Crops refer to smaller regions or patches extracted from larger images.

Entropy: Entropy is a measure of the randomness or uncertainty in an image. It can be used to characterize the complexity of structures or textures within the image.

Epoch: An epoch is one complete cycle through the entire training dataset during the training process of a machine learning model in bioimage analysis.

False discovery rate: The false discovery rate (FDR) is a measure used to control the rate of false positive findings. It represents the proportion of incorrect positive predictions (false positives) among all positive predictions made by an algorithm.

False positive / false negative: False positive indicates an incorrect positive prediction, such as when the algorithm wrongly identifies background regions as part of the object of interest. False negative, on the other hand, refers to an incorrect negative prediction, where the algorithm fails to identify actual regions of interest.

First order statistics: First order statistics refers to the basic statistical measures computed directly from the pixel intensity values in an image, such as mean, median, standard deviation, etc.

Flatness: Flatness shows the relationship between the largest and smallest principal components in the region of interest (ROI) shape. The principal component analysis is performed using the physical coordinates of the voxel centers defining the ROI.

Gray level co-occurrence matrix: The gray level co-occurrence Matrix (GLCM) is a texture analysis method that quantifies the frequency of intensity value pairs at various spatial relationships within an image. It is commonly used to extract texture features.

Gray level dependence matrix: The gray level dependence matrix (GLDM) is a texture analysis technique that captures dependencies between pairs of pixels with specific intensity values in an image.

Gray level run length matrix: The gray level run length matrix (GLRLM) is a texture analysis technique that characterizes the lengths and frequencies of consecutive pixel runs with the same intensity value in an image.

Gray level size zone matrix: The gray level size zone matrix (GLSZM) is a texture analysis method that provides information about the distribution of connected regions of similar intensity values in an image.

Gray level non-uniformity: Gray level non-uniformity (GLNU) is a texture feature that quantifies the variance in pixel intensities within an image. It provides information about the spatial variations of grey levels in the image.

Ground truth: Ground truth refers to the manual or expert annotations of the correct regions of interest in an image. It serves as the gold standard for evaluating the performance of algorithms and machine learning models.

Neighboring gray tone difference Matrix: The neighboring gray tone difference matrix (NGTDM) is a texture analysis method that quantifies differences in pixel intensity values between neighboring pixels in an image.

Overfitting: Overfitting is a common problem in machine learning in general, where a model becomes too specialized on the training data and fails to generalize well to new, unseen data. It occurs when a model learns noise or irrelevant patterns from the training set.

Precision: Precision is a measure of the accuracy of positive predictions made by an algorithm. It calculates the proportion of true positive instances out of all positive predictions (true positives plus false positives).

Radiomics: Radiomics is a field in bioimage analysis that focuses on the extraction and analysis of quantitative features from images like texture, shape and morphology.

Recall: Recall, also known as sensitivity or true positive rate, is a measure of the algorithm's ability to correctly identify positive instances (true positives) out of all actual positive instances in the image.

Segmentation: Segmentation refers to the process of identifying and delineating regions of interest (ROI) within an image, separating them from the background or other structures. The goal is to partition the image into meaningful regions that correspond to specific biological structures or objects of interest.

Training: Training involves using a dataset with known ground truth annotations to teach a machine learning model or algorithm how to perform the task, such as segmentation or classification, on new, unseen images.

True positive / true negative: True positive refers to correctly identified positive instances (e.g., correctly identified regions of interest). True negative, on the other hand, refers to correctly identified negative instances (e.g., background regions correctly identified as not being part of the object of interest).

Validation: Validation is the process of assessing the performance and generalization capabilities of a trained algorithm on a separate dataset. It helps to ensure that the model's performance is consistent and reliable on new, unseen data.