**Brain Structural Organization Revealed by Unbiased Cell-Type Distribution Clustering**

Rodrigo Muñoz-Castañeda[1,2,4], Ramesh Palaniswamy[1], Jason Palmer[1], Rhonda Drewes[1], Corey Elowsky[1], Karla E. Hirokawa[3], Nicholas Cain[2], Kannan Umadevi Venkataraju[1], Julie A. Harris[3] and Pavel Osten[1]

[1] previous affiliation: Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724.

[2] Appel Alzheimer's Disease Research Institute, Feil Family Brain and Mind Research Institute, Weill Cornell Medicine, New York, NY, 10021, USA

[3]Allen Institute for Brain Science

[4]corresponding author: rom4023@med.cornell.edu

## ABSTRACT

Cell type composition across brain regions is a crucial anatomical factor shaping both local and long-range brain circuits. In this study, we employed single-cell resolution imaging of the mouse brain, combined with automated computational analyses, to map the distribution of thirty cell types defined by unique gene marker expressions across the entire brain. Leveraging Cre recombinase mouse models for cell type labeling, we generated a comprehensive atlas of cell-type-specific distributions throughout the male and female brain. This analysis revealed that major anatomical brain areas can be characterized by distinct cell-type composition signatures. Furthermore, these distributions provide a powerful framework for identifying and refining the boundaries of different brain regions through unbiased clustering of whole-brain cell density data. This includes mapping of cortical areas and layers, hippocampal subregions, and thalamic and hypothalamic nuclei. Our findings suggest that cell type composition is intricately linked to

the structural organization of the brain, with even subtle variations potentially contributing to distinct, brain region-specific brain functions.

## INTRODUCTION

The mammalian brain is composed of numerous distinct cell types, each distributed in specific densities and organized into cell type-specific circuits within anatomically and functionally defined regions. Given that cell type composition establishes the structural foundation for the formation of function-specific neural circuits, it plays a pivotal role in supporting all brain functions, ranging from information processing and integration to the regulation of behavior and emotional responses[1,2].

The classical anatomical division of the mammalian brain includes seven major structures: the cortex, cerebral nuclei, thalamus, hypothalamus, midbrain, hindbrain, and cerebellum. The spatial organization and subregions of these structures have been characterized in brain atlases using various histological techniques, such as Nissl staining, which visualizes brain regions based on differences in cell density and size, and acetylcholinesterase (AChE) immunolabeling, which highlights anatomical structures based on cholinergic innervation. However, while cytological stains and immunolabeling as well as projection tracing (both anterograde and retrograde) have been invaluable for mapping brain regions, traditional anatomical approaches have been limited by the use of few brain replicates and dependence on the expertise of a small number of anatomical experts. As a result, discrepancies exist between anatomical atlases, including between mouse brain atlases based on Nissl and AChE staining, Nissl alone, and more recent approaches using highly averaged tissue autofluorescence [3–5].

Efforts to reconcile these differences have included consortium-based studies that integrate multimodal data from various laboratories to provide a detailed description of specific anatomical areas, such as the mouse motor cortex [4,5]. Additionally, digital integration of multiple atlases with the Common Coordinate Framework (CCF) of the mouse brain has been implemented to help to standardize anatomical definitions [4,5]. Given the critical role of the mouse as a model organism in neuroscience—particularly due to the genetic accessibility it offers for tracing and manipulating specific cell types using Cre recombinase mouse lines [3]—it remains essential to further refine anatomical boundaries as well as characterize the cell type populations defined by these cell type-specific Cre lines.

Here we apply thirty cell type-specific Cre mouse lines to quantitatively map the respective cell type populations in the adult brain and to use these patterns to derive combinatorial cell type signatures for major brain regions and subregions. These Cre mice include 27 of the most commonly used cell type-specific knock-in Cre lines, in which the Cre recombinase is expressed from an IRES cassette inserted after the stop codon in the 3' of the cell type marker gene, thus achieving faithful Cre expression in conjunction with the native gene product [6]. In addition, we include 3 bacterial artificial chromosome (BAC) transgenic lines in which Cre was inserted after a promoter of the Ntsr1, Rbp4 and Tlx3 genes[7,8].

## RESULTS

### Whole brain visualization and quantification of gene expression-defined brain cell-types

The whole brain mapping approach applied in the current study included the following: First, we used genetically engineered "Cre-driver" mouse lines expressing the Cre recombinase from cell type-specific promoters to have a genetic access for fluorescent cell type labeling. Second, thirty

Cre driver mice were crossed with “Cre-reporter” mice which express nuclear H2B-GFP protein (or tdTomato protein) upon Cre recombination (Figure 1A, Table 1). This resulted in nuclear fluorescent labeling of the respective gene expression-defined cell types in the off-spring double transgenic mice (to simplify and maintain consistency, we will refer to gene expression-defined cell types as cell types throughout the rest of the paper). Third, both male and female double transgenic whole brains were imaged by serial two-photon tomography (STPT) and the resulting serial section datasets were registered to a Reference Mouse Brain and computationally analyzed using a modification of the previously published quantitative brain-wide (qBrain) mapping platform [3,4] (qBrain2.0) (see Figure 1). Specifically, whole brain imaging of the Cre driver-reporter mouse brains was performed using STPT at XY resolution of 1 um and Z-spacing of 50 um, obtaining a total of 280 coronal sections. 3D registration was performed using Elastix Toolbox[9] and both the reference and the experimental brains were pre-processed to enhance their intrinsic anatomical features (see Methods). The resulting transformations were then used to warp the original resolution images (1x1x50 um) onto the Common Coordinate Framework (CCF) of the mouse brain[3] (Figure 1B). After data registration onto CCF the data were anatomically segmented by the previously modified version of the Allan Reference Atlas (ARA) for the mouse brain[3]. The cell type-specific fluorescent nuclei were detected by convolutional neural networks (CNNs) trained on ground truths datasets representing a range of brain regions with varying density of cellular labelling (Figure 1C, Methods). Once completed, all brain images and processed datasets were transferred to the Brain Image Library (BIL) for public sharing. Finally, we implemented a density-based DBSCAN clustering for each cell type to test the use of density-based cell type distribution for unbiased segmentation of anatomical brain areas.

**Cell type distribution can define whole brain anatomical organization**

The Cre–lox system identifies cell populations expressing the tagged gene at any point during development and/or in the adulthood[10]. Therefore, we first determined which labeling among the thirty Cre-lox driver mice represent faithfully adult expression of the marker gene and which includes also development-based expression. This was done by a comparison of the cell type fluorescent labeling in our datasets to *in-situ* hybridization data derived from the Allen Institute for Brain Science data portal[11], the Gene Expression Nervous System Atlas (GENSAT) portal[12], and previously published works[13,14]. This revealed that 26 Cre-lox mice largely correctly label adult gene expression, whereas the following comprise a strong developmental labeling in addition to the adult expression: nNOS (neuronal nitric oxide synthase) marks a sparse population of inhibitory neurons in the cortex and other areas in the adult brain, whereas nNOS-Cre-based labeling marks high density of cells, reflecting a strong nNOS developmental expression perhaps in all neurons (Supp. Figure 1)[15]; SERT labels serotoninergic neurons in the adult midbrain raphe nuclei[16,17]. In contrast, SERT Cre-based labeling marks dense neuronal populations in the cortex and the thalamus, reflecting known SERT expression in the development which plays a critical role in the formation of cortico-thalamic connection patterns[18,19]; VGlut2 labels adult glutamatergic neurons primarily outside the cortex[20,21], but VGlut2-cre labeling includes all cortical areas. This cortical developmental VGlut2 expression has been previously documented [21,22] and is proposed to participate in the refinement of pyramidal neuron dendritic arbors [23]. Finally, VGlut3 is expressed primarily in cholinergic neurons of the striatum in the adult brain [24], in the dorsal raphe by subpopulation of

serotoninergic neurons [25], and at low density in all major areas [26,27]. However, VGlut3-Cre labeling identified additional cell populations in the cortex and the thalamus.

To begin the analysis of our data, all datasets were visualized in three-dimensional (3D) point-cloud views, revealing distinct anatomical patterns for the fluorescently labeled cells types, such as broad and fairly uniform patterns for neuron-specific SNAP25 or glia cell-specific Olig2, broad but varied patterns for inhibitory neuron-specific GAD2 or Slc32 and inhibitory subtype-specific PV, Calb, Calb2, SST, VIP and Pdyn, and excitatory neuron-specific VGluT1, VGluT2, VGluT3 and Emx1, as well as more restricted patterns highlighted by the hypothalamic Avptm marking vasopressinergic cells and Oxt marking oxytocinergic cells, or the cholinergic neurons marker ChAT expressed in the basal forebrain and the striatum (Figure 2A and Supplementary Figure 2). These distinct patterns suggest that gross anatomical areas may be defined by their shared composition of specific cell types and cell type combinations. To test this, we clustered 208 anatomical areas based on the similarity of their cell type expression profile, using a normalized expression of each cell type across all brain areas (Figure 2B, Methods). This analysis revealed four major clusters: three clusters comprised mostly homogenous brain areas: cluster 1 (green, Figure 2B) comprised mainly cortical areas, cluster 2 (orange) mainly thalamic nuclei and cluster 4 (red) mainly hypothalamic nuclei. The last cluster 3 (purple) comprised the remaining brain areas, including the striatum, midbrain, hindbrain, cerebellum and olfactory areas (Figure 2B). These data suggest that major brain structures indeed share common cell type distribution signatures. For example, the cortical cluster (cluster 1, green) is defined by a relative high density of most glutamatergic markers, including the broad glutamatergic markers VGlut1, VGlut2 and Emx, and the layer-specific markers Tlx3, Rbp4, Ctgf, Ntsr1, Cux2 and Rorb (Figure 2C). This highlights the significance of glutamatergic projection neurons in the cortical

signaling and organization. Also predominant in this cluster are the general neuronal marker Snap25, the neuropeptide NPY, the hormone-releasing factor Crh and tyrosine hydroxylase (TH). Next, the thalamic cluster (cluster 2, orange) is defined by the relative high density of the glutamatergic marker VGlut3 and the serotoninergic marker SERT, reflecting the developmental roles of these genes in the thalamus. Finally, the most relevant cell types in the hypothalamic cluster (cluster 4, red) are the neuropeptides Oxt and Avptm, and the dopamine marker DAT.

**Glutamatergic and GABAergic organization of the brain.**

The distribution of the glutamatergic and GABAergic cell types is one of the core principles of brain cell type organization and has been studied by different approaches ranging from in situ hybridization, immunolabeling to MERFISH [28,29]. Here we quantified the two cell populations at a single cell resolution across the whole brain based on excitatory neuron expression of VGlut1 or VGlut2 and inhibitory neuron Gad2 expression (Figure 3A). The VGlut1 expression was used as a glutamatergic marker in the cortex and VGlut2 as a glutamatergic marker in subcortical brain areas. To visualize the relative glutamatergic to GABAergic cell expression across the brain, the individual brain areas were organized in the descending order of glutamatergic to GABAergic ratios (Figure 3B). This revealed that areas with the highest glutamatergic cell type content belong predominantly to the thalamus and the cortex, while areas with the highest GABAergic cell content are derived from the striatum, pallidum and hypothalamus. However, there were several distinctions from to this overall organization. For example, cortical and thalamic areas comprising higher proportion of GABAergic cells included the accessory olfactory bulb (AOB), main olfactory bulb (MOB), nucleus of the lateral olfactory tract (NLOT), basomedial amygdalar nucleus (BMA) and endopiriform nucleus, ventral part (EPv) in the

cortex and the intergeniculate leaflet of the lateral geniculate complex (IGL), peripeduncular nucleus (PP), ventral part of the lateral geniculate complex (LGv) and reticular thalamus (RT) in the thalamus (Figure 3B-C). This reveals a distinct excitatory to inhibitory neuron composition for the olfactory system, especially the olfactory bulb, compared to other cortical regions, in agreement with a recent study using MERPHIS [28]. This is further visualized in a voxel-wise distribution selectively for the olfactory system (Figure 3C). The thalamic nuclei with higher GABAergic population correspond to the most lateral thalamic areas, showing a pronounced spatial organization. This also visualized in Figure 3C.

In contrast, the brain areas with the highest proportion of GABAergic cells include the striatum, pallidum, hypothalamic, midbrain, hindbrain and cerebellum. Among these there were also areas that comprised higher proportion of glutamatergic cells including the mammillary nuclei (MM, PMd and PMv), paraventrivular hypothalamic nucleus (PVHdp, PVHm, PVHp and PVHmpv), ventromedial hypothalamic nucleus (VMHvl, VMHa, VMHc and VMHdm), subthalamic (STN) and parasubthalamic (PSTN) nuclei, and supraoptic nucleus (SO) in the thalamus; inferior colliculus (IC), midbrain trigeminal nucleus (MEV), periaqueductal grey (PAG, pedunculopontine nucleus (PPN), nucleus sagulum (SAG), cuneiform nucleus and ventral tegmental area (VTA) in the midbrain; and the deb nucleus of the anterior commissure (BAC) and triangular nucleus of the septum (TRS) in the pallidum.

**Cell type-based cortical organization**

The whole-brain analysis revealed that all cortical areas, with the exception of the frontal pole and olfactory MOB, AOB areas, share a cell type-based signature (Figure 2B). However, our earlier study of selected inhibitory cell types revealed that cell type distribution can also identify

distinct signatures for cortical areas and layers[3]. In the current study we extended on the earlier work by selecting 24 cell types with cortical expression which were analyzed across cortical areas and layers. These data are visualized by hierarchical clustering and flat-map organization in Figure 4. Additionally, examples of raw data highlighting the most significant features in the overall cell type-based cortical expression and organization are shown in Figure 5.

We first clustered all cortical areas based on the similarity of their cell type profile across all layers (see Methods)[30]. This revealed four major clusters and 8 subclusters (Figure 4A), which were visualized in 3D renderizations (Figure 4B). The first cortical cluster (C1; light orange) corresponds to frontal-medial areas and it is divided into 4 subclusters, orbitolateral (C1.1; orange subcluster), rostral (C1.2; green subcluster), frontal pole (C1.3; blue cluster) and medial (C1.4; red cluster; Figure 4A-B). This cortical cluster is characterized by relatively higher density of the following markers: broad GABAergic markers Gad2, Slc32, and Calb2, the serotoninergic marker SERT, the glutamatergic markers Rbp4 and Ctgf, the hormone releasing neurons Tac1, and the oligodendrocytes marker Olig2 (Figure 4C). The second cluster (C2; light purple) is composed by somatosensory, visual and auditory areas, indicating an overall homogeneous cellular organization among sensory processing regions (Figure 4A-B). Furthermore, this cluster also includes the posterior parietal association cortex (PTLp), which has been described as a connection hub between sensory areas and other cortical association regions [31–33]. This cluster is characterized by the relative higher density of the inhibitory markers Gad2, NPY, Pv, Sst and Calb, the layer-specific glutamatergic markers Cux2 and Ntsr1, and the hormone releasing neurons marker Crh (Figure 4C). The third cluster (C3; light brown) is composed by lateral association areas. Compared with other clusters, the most characteristic cell types of this cluster are the layer-specific glutamatergic markers Rorb and Ctgf (Figure 4A, right

panel). The fourth cortical cluster (C4; light pink) is composed by motor and medial areas, both association- and sensory-related. It can be divided into two subclusters, a visual-association cluster (C4.1; pink) and a somatomotor medial cluster (C4.2; grey). The visual-association subcluster (C4.1) includes the anteromedial visual cortex (VISam) and the agranular retrosplenial cortex (RSPagl). Notably, VISam has been shown to be the primary source of inputs to RSPagl [34,35]. The most relevant cell types of this major cluster are the glutamatergic markers Emx, VGlut1, VGlut2, VGlut3 and Tlx3, the inhibitory markers Pdyn and VIP, the hormone releasing marker Crh, the cholinergic marker ChAT, and the oligodendrocytes marker Olig2 (Figure 4C). Overall, these results underscore that distinct cell type composition can characterize anatomically and functionally related cortical areas.

Next, to understand the organization of the cortex based on layer-specific cell type distribution, we analyzed each cortical layer and defined their cell type signatures (Figure 4D-E). The analysis of layer 2/3 revealed four clusters and seven subclusters. The first cluster (C1, light blue) corresponds to frontolateral areas and is divided into two spatially overlapping subclusters. The first subcluster, C1.1 (orange), is more medio-lateral, while the second subcluster, C1.2 (green), is more fronto-lateral (Figure 4D). The most relevant cell types characterizing this cluster are the glutamatergic markers Ntsr1, Rorb and Rbp4, and the hormone-releasing marker Tac1 (Figure 4E). The second cluster (C2, light red) corresponds to sensory areas, including somatosensory, visual, and auditory regions. This is similar to the sensory whole cortical cluster (C2, light purple; Figure 4A). Additionally, the layer 2/3 sensory cluster also contains the association area PTLp. This cluster is characterized by the expression of the glutamatergic markers VGlut1 and Cux2, and the inhibitory markers NPY, Pv, Calb, and Sst (Figure 4E). The third cluster (C3, light brown) is composed of motor and association areas, both medial and

lateral, and is subdivided into two different subclusters: lateral and medial association areas (C3.1, purple) and motor areas (C3.2, brown). The most prominent cell types defining this cluster are the glutamatergic neurons VGlut3, Tlx3 and Ctgf. The fourth cluster (C4, light pink) consists of visual-association areas and is subdivided into two subclusters: a visual and dorsolateral association cluster (C4.1; pink) and a medial dorsoventral association cluster (C4.2, grey), which includes the caudal retrosplenial cortex (RSPd and RSPv). The cellular composition of this cluster is characterized by the expression of the glutamatergic markers Emx and VGlut2, the serotoninergic marker SERT, the cholinergic neuron marker ChAT, the GABAergic cell markers Gad2, Slc32, Calb2, Pdyn, and VIP, hormone-releasing neurons Crh, and the oligodendrocyte marker Olig2. Overall, relative predominance of glutamatergic cells is distributed among all clusters of layer 2/3. However, GABAergic cells show relative predominance in the two sensory clusters (C2 and C4), highlighting the importance of layer 2/3 interneurons in sensory processing.

Layer 4 is exclusively present in the granular cortex, thus excluding the more medial and lateral areas. Our analysis of layer 4 organization identified four major clusters, which can be further subdivided into five subclusters (Figure 4D). The first cluster (C1, light orange) is primarily composed of auditory areas, as well as lateral visual areas. The single most relevant cell type expressed in this cluster is the GABAergic marker Gad2 (Figure 4E). The second major cluster (C2, green) corresponds to somatosensory areas, which also includes the visceral cortex (VISC). Analysis of cell type distribution revealed that the most relevant cell types in this cluster are the GABAergic markers NPY, Sst, and Pv, the cholinergic marker ChAT, and the oligodendrocyte marker Olig2 (Figure 4E). The third cluster (C3, blue) is composed solely of the gustatory cortex, which has been previously identified as a multisensory integration area [36]. This

cluster shows relative high density of cells labeled by glutamatergic markers, such as Emx, VGlut1, VGlut2, Cux2, and Tlx3, as well as inhibitory neuronal markers including Slc32, Pdyn, Calb, Calb2, and VIP, and the hormone-releasing factor Crh. The fourth cluster (C4, light red) corresponds to visual and association areas and subdivided into two different subclusters: a visual subcluster (C4.1, red) and an association subcluster (C4.2, purple). This cluster is characterized by the relative high density of glutamatergic neuron markers Ntsr1, Rorb, Rbp4, Ctgf, and VGlut3, the serotoninergic marker SERT, and the hormone-releasing factor Tac1. Overall, gustatory and visual cluster areas (C3 and C4) have the most dense distribution of glutamatergic neurons among all layer 4 regions. On the other hand, somatosensory and gustatory cluster areas (C2 and C4) are distinguished by relatively the most-dense GABAergic cell distribution.

Layer 5 analysis revealed three major clusters and seven subclusters. The first cluster (C1, light blue) corresponds to frontal-medial areas and is subdivided into four subclusters. The first subcluster, C1.1 (blue), includes only the posteromedial visual cortex (VISpm). The second subcluster, C1.2 (orange), encompasses the primary and posterolateral visual cortex. The third subcluster, C1.3 (green), covers the frontomedial cortical areas, and the fourth subcluster, C1.4 (red), consists of the retrosplenial cortex (Figure 4D). The C1 cluster is characterized by relative higher densities of the glutamatergic cell markers Emx, VGlut1, VGlut2, Cux2, Ntsr1, Rbp4 and Ctgf and the serotoninergic marker SERT. Additionally, the GABAergic markers Gad2, Slc32, Pdyn, Calb, Calb2 and Sst, hormone-releasing factor neurons Tac1 and Crh, and the oligodendrocyte marker Olig2 are also predominant in this cluster. A more detailed analysis shows that the VISpm subcluster (C1.1) has the relative highest density of glutamatergic markers, notable for its predominant expression of broad glutamatergic markers such as Emx,

VGlut1, VGlut2, and Cux2. However, GABAergic cells Gad2, Slc32, Sst and Pv show relative higher density in the primary and posterolateral visual cortices subcluster (C1.2; Supp. Figure 5B). The second major cluster (C2, light purple) is composed by motor and sensory areas. It is characterized by relative high densities of the glutamatergic markers VGlut3 and Tlx3, and the GABAergic markers NPY, Sst and Pv. The cluster 3 (C3, light brown) is composed by fronto-lateral association areas and can be subdivided into two subclusters: C3.1 (brown) composed by more lateral areas and C3.2 (pink) composed by more frontal areas. The cell types with higher relative density in this cluster are the glutamatergic markers Ctgf and Rorb.

The analysis of Layer 6 identified three main clusters and seven subclusters. Cluster 1 (C1, light orange) consists of frontal-medial areas and is divided into two subclusters: C1.1 (blue), which includes only the ventrolateral orbital area (ORBvl), and C1.2 (orange), which includes medial areas. This cluster is characterized by relative higher density of excitatory markers, such the glutamatergic markers Rbp4 and VGlut3, as well as the serotoninergic marker SERT and the cholinergic marker ChAT. Additionally, this cluster comprises relative higher density of the GABAergic markers Gad2, Slc32, Pdyn, Sst, Pv and Calb2, the hormone-releasing factor Tac1, and the oligodendrocyte marker Olig2. Cluster 2 (C2, light purple) is composed by sensory areas and it is divided into three subclusters: C2.1 (green) with auditory and lateral visual areas, C2.2 (red) with visual-association areas including the posteromedial visual cortex (VISpm) and RSPagl, and C2.3 (purple) corresponding to visuo-sensory areas. The cell types with relative high density in this cluster include the glutamatergic neurons Emx, VGlut1, VGlut2, Cux2, Ntsr1 and Tlx3, the GABAergic neurons Calb, NPY and VIP, and the hormone-releasing factor Crh. Cluster 3 (C3, light brown) includes motor and lateral association areas and is divided into two subclusters: C3.1 (brown) with prefrontal-motor areas, and C3.2 (pink) with

lateral association and sensory areas. The most relevant cell types in this cluster are glutamatergic neurons Rorb and Ctgf. Thus, sensory cluster (C2) is characterized by the relative highest density of glutamatergic markers among all layer 6 clusters. On the other hand, the frontomedial cluster (C1) is characterized by the relative highest density of GABAergic, serotoninergic and cholinergic markers.

These results emphasize that distinct patterns of cell type distributions can be used for unbiased identification and characterization of cortical areas, revealing that similar functional cortical regions have a similar cell type organization, both across and within cortical layers. For example, our analysis show that sensory areas cluster together based on their whole brain cell organization (cluster C2; Figure 4A-B), and this same clustering pattern is seen across cortical layers. Similarly, frontal-medial areas have a similar cellular organization both at the whole brain level and at layer-specific levels, with the exception of layer 2/3. Furthermore, the brain cellular organization seems to reflect previously known functional and connectivity differences among cortical areas.

**Density-based cluster analysis of the gene expression-defined cell types**

The above analyses used anatomical atlas to investigate the distribution of various cell types, demonstrating a distinct cell type patterns for different anatomical areas. This suggests that brain spatial organization and borders of brain regions may be defined in an unbiased way using the distribution patterns and local densities of the gene marker-defined cell types. To test this hypothesis, we next partitioned the spatial distribution of each cell type using spatial clustering of its cell densities calculated across the entire brain. The goal of this approach was to identify and visualize 3D spatial areas that can be defined solely based on a cell type and its local cell

density. This was performed using a spatial clustering algorithm Density-Based Spatial Clustering of Applications with Noise (DBSCAN)[37,38]. First, we chose ε = 150 um to define the radius around each cell within which the local cell densities were calculated. We then calculated the number of cells ("neighbors") within this radius for all detected cells and subsequently plotted each cell type's cell density histogram. This histogram represents the frequency distribution of cell densities for each cell type dataset. We then divided the frequency distribution into 9 percentiles, from the $10^{th}$ to the $90^{th}$ percentile, where the nth percentile represents the value where n percent of the cell densities lie to the left of the histogram. For example, the $10^{th}$ percentile cell density frequency identifies broad spatial areas defined by all cell densities higher than the 10% of density frequency. Conversely, the $90^{th}$ percentile identifies discreet spatial areas defined by cell densities higher than the 90% of density frequency. This resulted in 9 different density clusters for each cell type, partitioning the brain into progressively smaller areas comprising progressively greater densities at each percentile increment. Finally, we generated an overlay of all clusters generated for each cell type (see Figure 6A; see Methods).

The current analysis of the cell types most relevant for whole-brain and cortical organization included VGlut1, VGlut2, Rbp4, SERT, Npy and Pv (Figures 2-6). In order to have the general overview of their distribution and parcellation, we first visualized the $10^{th}$ percentile comprising broad areas for each cell type (Figure 6B). We also included one additional percentile for each cell type which highlighted examples of regional and subregional brain area specificity (see Figure 6C; see Supp. Figure 3 for all percentiles).

Starting with cortical cell type distribution, the analyses of the broad glutamatergic markers VGlut1 and VGlut2 revealed distinct density patterns, with a higher density in medial motor-association areas and sensory areas in a layer-specific manner, as seen in the 70th

percentile for VGlut1 and the 80th percentile for VGlut2 (Figure 6B-C). These density distributions are consistent with our earlier results indicating the relative high density of VGlut1 and VGlut2 in motor and medial areas, both association- and sensory-related (Figure 4A-C). On the other hand, Rbp4 distribution clearly highlighted the layer 5 across all cortical areas (blue cluster, Figure 6B). Furthemore, this density distribution shows a gradient along the different percentiles, with the medial and motor areas characterized by the highest densities (Supp. Figure 3). The serotoninergic marker SERT distribution highlights the layer 6 motor-medial and lateral areas (orange cluster; Figure 6B-C), with the motor-medial areas being characterized by the highest cell densities (Figure 6C). These motor-medial areas also show a thin but dense cluster of SERT cell density within layer 2 (Figure 6C and Supp. Figure 3). The interneuron subpopulation marker NPY shows a broad distribution in the cortex (blue cluster, Figure 6B-C), with higher densities in deep cortical layers of sensory areas (Supp. Figure 3). Finally, the interneurons subpopulation marker Pv also shows broad cortical expression (blue cluster; Figure 6C), although at the 20th percentile density revealed a a reduction in Pv distribution along both upper and lower cortical layers as well as in lateral association areas (Figure 6C).

Another area with clear subregional borders highlighted by the cell density analyses included the hippocampus. VGlut1 and VGlut2 displayed distinct patterns across the hippocampal formation (Figure 6B). The dentate gyrus (DG) has the highest density of both cell types (Figure 6C), followed by CA1 (Supp. Figure 3). This higher density of glutamatergic neurons in both DG and CA1 is consistent with previous works [39]. On the other hand, VGlut1 and VGlut2 cell densities delineate the fascicula cinerea (FC) at $10^{th}$ and $30^{th}$ percentile respectively (Figure 6B-C and Supp. Figure 3), an area that receives projections from lateral entorhinal and perirhinal cortices and projects exclusively to the dentate gyrus (DG) [40].

Interestingly, the FC area is also delineated in the serotoninergic marker SERT density distribution at a density that also highlights the CA3 hippocampus and the thalamus (red cluster; Figure 6B-C). Rbp4 cell density highlights the DG area (Figure 6B-C and Supp. Figure 3), which is also outlined by higher densities of the cells marked by the expression of the neuropeptide NPY (Figure 6B-C, Supp. Figure 3). Finally, Pv cell densities characterize the subiculum (SUB), medial CA1-so and caudal CA1-sp, with the SUB comprising the highest Pv cell density in the hippocampus (Figure 6C).

The majority of the thalamus, except the reticular thalamus, is highlighted by the cell density distribution of VGlut2 (Figure 6B-C), with the high 80$^{th}$ percentile density further delineating the parafascicular nucleus (PF; Figure 6C). Rbp4 expression highlights the mediodorsal (MD), the suprafascicular (SPF), parafascicular (PF) and the posterior complex (PO) nuclei of the thalamus (Figure 6C), with the PF and the medial PO being further highlighted by higher Rbp4 cell densities (Figure 6C). The serotoninergic marker SERT highlights the majority of thalamic nuclei, with the exception of the reticular thalamus (Figure 6B-C), with higher densities further highlighting the ventro-lateral thalamic nuclei (Supp. Figure 3). The neuropeptide NPY cell densities highlight the ventral posteromedial nucleus of the thalamus (VPM), medial thalamic nuclei Figure 6C) and the RT, the RT being further highlighted by the highest NPY cell densities (Figure 6C and Supp. Figure 3). Finally, Pv cell density highlights the paracentral nucleus of the thalamus (PCN) and the reticular thalamus (RT), the latter being the thalamic nucleus with the highest Pv cell density (Supp. Figure 3).

The selected cell types further highlight multiple hypothalamic regions. Similarly to the thalamus, VGlut2 highlights most of the hypothalamus (Figure 6C and Supp. Figure 3), with the high 70$^{th}$ (Supp Figure 7) and 80$^{th}$ (Figure 6C) percentiles densities further highlighting the

posterior (PH), dorsal premammilary (PMd), dorsomedial (DMH) and arcuate (ARH) hypothalamic nuclei. On the other hand, Rbp4 highlights hypothalamic areas at low densities, including the PMd and the PH (Figure 6C). NPY cell densities also highlight multiple hypothalamic regions: the tuberal nucleus ($10^{th}$ percentile), the paraventricular hypothalamic nucleus ($20^{th}$ percentile), DMH ($40^{th}$ percentile), and the median eminence (ME, $80^{th}$ percentile; Figure 6C and Supp. Figure 3).

The midbrain and the hindbrain have also multiple structures highlighted by the selected cell types. VGlut2 highlights areas in both structures at all percentile densities (Figure 6C). At $10^{th}$ percentile, it highlights the pontine grey (PG), superior colliculus (SC), and the periaqueductal grey (PAG; Figure 6C). At higher density, $80^{th}$ percentile, it highlights the precommisural nucleus (PRC; Figure 6C). Rbp4 highlights the nucleus prepositus (PRP), the vestibular nuclei (VNC) and the intermediate reticular nucleus at the low $10^{th}$ percentile density. At a higher density, $30^{th}$ percentile, PRP continues to be highlighted. Similar to the other brain regions, NPY marked cells are broadly distributed among the midbrain and the hindbrain areas at the $10^{th}$ percentile low density (Figure 6C). At a higher density, $40^{th}$ percentile, it highlights the spinal nucleus of the trigeminal, interpolar part (SPVI; Figure 6C). Finally, Pv marker also highlights multiple areas in these regions at low, $10^{th}$ percentile, density such as the ventral part of the lateral geniculate complex (LGv), the ventral tegmental area (VTA), the red nucleus (RN), the substantia nigra pars reticula (SNr), the oculomotor nucleus (III) and the PG (Figure 6C).

Taken together, these results show that cell type density can be used to define brain anatomical areas through 3D spatial analysis. This new method can allow unbiased identification of borders of anatomical brain regions and subregions in whole brain 3 dimensional datasets based solely on distinct local densities of specific cell types in different brain areas. This

approach thus provides a novel cell type-based approach to brain atlasing that can complement and refine previous efforts based on Nissl and immunolabeling or averaged tissue autofluorescence.

## DISCUSSION

In the current study we have characterized the 3D spatial distribution of thirty different cell populations defined by cell type-specific Cre mouse lines across the entire mouse brain at single cell resolution. This work was accomplished using a new whole-brain analysis pipeline, extending on our previously published quantitative brain-wide (qBrain) atlasing pipeline[3]. The current pipeline adds a more automated machine learning-based single-cell segmentation, a registration platform using intrinsic anatomical features (IAFs), and an unbiased cell density-based spatial distribution analysis and 3D spatial area parcellation (Figure 1, Figure 6 and Methods).

We used these methods to study brain organization based on thirty cell type distribution in Cre recombinase mouse models using both males and females to control over possible sex differences. While most of these Cre lines correctly label the cell type populations expressing the respective cell-type marker gene in the adult brain, we also included four Cre lines driven by gene markers with strong developmental expression that is not sustain in the adult brain. This includes nNOS-Cre-based broad cell labeling across all brain regions[15]; SERT-Cre-based dense cell labeling in the cortex and the thalamus, reflecting the role of SERT in the formation of cortico-thalamic connections[18,19], VGlut2-Cre-based cortical cell labeling[21,22,23], and VGlut3-Cre-based cell labeling in distinct cortical regions and the thalamus. Since all these mice are knock-in Cre mice with Cre expression directly linked to the endogenous genes, the adult brain

labeling reflects the adult cell lineages marked by the transient developmental expression, rather than artificial Cre misexpression. We therefore included these mouse lines in the current analysis, with the stated caveat of developmental contribution to the adult cell counts and distributions.

Our results revealed that major brain anatomical areas can be clustered and defined by their distinct cell-type composition signatures (Figure 2). While differences in cell type distribution across brain regions are well known and have been most recently demonstrated by a whole mouse brain MERFISH study, which identified as many as 300 cell types based on an analysis of more than 5000 gene transcript [28], our study demonstrates that a relatively few major cell type-specific gene markers in the brain are sufficient to cluster functionally related brain areas of the cortex, hippocampus, thalamus and hypothalamus. This data thus emphasizes the importance of these major cell type populations in defining the functional architecture of the brain.

Finally, given the power of the selected cell type markers in characterizing functionally related brain areas, we asked whether their distribution alone may be sufficient to identify brain regions and define their spatial borders. We tested this hypothesis using unbiased spatial clustering of the cell type cell densities using DBSCAN and progressive segmentation of the whole-brain data from the $10^{th}$ to $90^{th}$ density percentile. The current analysis of these data for selected six, VGlut1, VGlut2, Rbp4, SERT, Npy and Pv Cre line datasets revealed that indeed this approach can identify spatial areas representing distinct brain regions and subregions across all cortical structures. This proof-of-concept data suggest that a number of anatomical regions can be spatially defined based solely on a major cell type and its cell density. A systematic application of this approach for all 30 cell types as well as for cell type combination is likely to

provide further delineations of many more brain areas, setting the stage for a complete mouse brain atlas defined entirely by cell type composition. Such atlas may complement current atlases and aid in refining brain borders in 3D at high spatial resolution.

## MATERIAL AND METHODS

### Cre-reported lines

Cre driver ‘‘knock-in’’ animals were crossed with reporter mice (CAG-LoxP-STOP-LoxP-H2B-GFP) as previously described [3]. Animal procedures were performed under the Cold Spring Harbor Laboratory Institutional Animal Care and Use Committee (IACUC) approval. All animals received food and water ad libitum and were housed under constant light and temperature conditions (12 hr cycle lights ON: 0600, lights OFF: 1800).

### Brain samples preparation and imaging of cell-type distributions

All animals were anesthetized with ketamine/xylazine and perfused transcardially with isotonic saline followed by 4% paraformaldehyde (PFA) in 0.1M phosphate buffer (PB, pH 7.4). Brains were post-fixed overnight at 4C after extraction in PFA and stored at 0.05M PB until imaging.

For imaging, brains were embedded and cross-linked with oxidized 4% agarose as previously described [3,41,42]. Whole brain imaging of Cre animals was performed using serial two-photon tomography (STPT). The entire brain was coronally imaged at an X,Y resolution of 1µm and Z-spacing of 50 µm [3,41,42].

### Whole mount NeuroTrace staining

Whole brain NeuroTrace staining was performed with a modification of iDISCO+ protocol[43,44]. In order to reduce methanol induce distortions in the tissue from the original iDISCO+ protocol, samples were incubated in a solution containing 3% Triton-X in PBS with heparin for 1 week. Thereafter, samples were gently washed in PTwH 4x 8h. Finally, brain samples were incubated in a solution containing PTwH and 20% NeuroTrace for 4 weeks at 37C. Samples were finally washed in PTwH 4x 8h. Before imaging in STPT, samples were post-fixed again in 4% PFA overnight at 4C to avoid sectioning artifacts.

**Anatomical features enhancing**

To improve whole brain registration, intrinsic anatomical features (IAFs) in the CCF and whole brain datasets were enhanced through preprocessing using custom MATLAB and Python scripts. First, the main IAFs in the reference brain were identified and highlighted on the reference brain, followed by the application of a Sobel operator to reduce noise and computational cost during image registration. For the imaged datasets, the stack images intensities were first homogenized to avoid signal interference during registration process using gamma compression, followed by the application of a Sobel operator to get the IAFs. This ensured that all images were consistent in format and that IAFs were enhanced identically. By enhancing IAFs in the datasets.

**Image registration**

3D whole brain datasets were registered to the CCF reference brain using a three-step process. First, images were preprocessed as previously described (see *Anatomical features enhancing* above for description). Second, a 3D affine transformation was calculated using the Elastix registration toolbox[45] with Advanced Mattes mutual information metric. This transformation was

then followed by a 3D B-spline transformation to further improve the registration alignment. Accuracy was checked both visually and using brain unique identified points as previously described[1].

**High resolution image registration transformation**

After image registration at low resolution, the output transformations were used to generate high-resolution registered datasets using custom MATLAB scripts. After registration it was generated a 3D displacement field that identifying the transformation of each pixel from the original image coordinate space onto the CCF reference space. Then, the displacement field of the initial registration was transformed onto the original high resolution coordinate system and used to compute the high-resolution registration transformations from the original images.

**STPT cell counting**

Automatic cell counting was performed using a convolutional neural network (CNN) trained on H2B-GFP nuclear signaling data [3,41]. First, an unsupervised detection algorithm based on structure tensor and connected components analyses was developed to identify and segment individual cells. The results were used to generate a dataset of 270 random segmented image tiles from three different datasets, containing approximately 1350 cells from multiple brain regions. This subset of data, both from raw and segmented with unsupervised algorithm, was then used to train the CNN for cellular detection.

Accuracy of the detection network was assessed using different datasets with sample images from multiple brain regions to account for different signal to noise ratios based on intrinsic autofluorescence. Cell detection on these datasets was done manually by 3 different experts and

final accuracy calculated based on the mean values comparing the resulting output from the trained network with the expert's manual detection.

**Hierarchical Cluster analysis**

To highlight the most relevant regions for each cell type, we applied maximum normalization based on their area densities, i.e. we normalized each cell type by the maximum brain region density. We then performed a hierarchical cluster analysis with brain region correlation based on cell types distribution as pairwise distance, and ward metric to calculate the distances. All analyses were performed using custom Python scripts.

**DBSCAN cluster analysis**

DBSCAN cell type density-based analysis was performed using custom Python scripts. First, we calculated, for each individual dataset, all pairs of points within a radius of 150 μm. We then calculated the $10^{th}$, $20^{th}$, $30^{th}$, $49^{th}$, $50^{th}$, $60^{th}$, $70^{th}$, $80^{th}$ and $90^{th}$ percentiles for each individual dataset. For each cell type we finally calculated then mean value for each percentile and then multiply by the number of datasets, obtaining the estimation of the percentiles for each cell type with all datasets. We then used the estimated percentile values to perform the DBSCAN algorithm at each percentile.

## REFERENCES


1. Swanson, L. W. & Lichtman, J. W. From Cajal to Connectome and Beyond. *Annu. Rev. Neurosci.* **39**, 197–216 (2016).
2. Swanson, L. W. & Bota, M. Foundational model of structural connectivity in the nervous system with a schema for wiring diagrams, connectome, and basic plan architecture. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 20610–20617 (2010).
3. Kim, Y. *et al.* Brain-wide Maps Reveal Stereotyped Cell-Type-Based Cortical Architecture and Subcortical Sexual Dimorphism. *Cell* **171**, 456-469.e22 (2017).
4. Muñoz-Castañeda, R. *et al.* Cellular anatomy of the mouse primary motor cortex. *Nature* **598**, 159–166 (2021).
5. BRAIN Initiative Cell Census Network (BICCN) *et al.* A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature* **598**, 86–102 (2021).
6. He, M. *et al.* Strategies and Tools for Combinatorial Targeting of GABAergic Neurons in Mouse Cerebral Cortex. *Neuron* **91**, 1228–1243 (2016).
7. Gong, S. *et al.* Targeting Cre Recombinase to Specific Neuron Populations with Bacterial Artificial Chromosome Constructs. *J. Neurosci.* **27**, 9817–9823 (2007).
8. Doyle, J. P. *et al.* Application of a Translational Profiling Approach for the Comparative Analysis of CNS Cell Types. *Cell* **135**, 749–762 (2008).
9. Klein, S., Staring, M., Murphy, K., Viergever, M. A. & Pluim, J. P. W. Elastix: A toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging* **29**, 196–205 (2010).

10. Song, A. J. & Palmiter, R. D. Detecting and Avoiding Problems When Using the Cre–lox System. *Trends in Genetics* **34**, 333–340 (2018).

11. Ng, L. *et al.* An anatomic gene expression atlas of the adult mouse brain. *Nature Neuroscience* **12**, 356–362 (2009).

12. Heintz, N. Gene Expression Nervous System Atlas (GENSAT). *Nat Neurosci* **7**, 483–483 (2004).

13. Daigle, T. L. *et al.* A Suite of Transgenic Driver and Reporter Mouse Lines with Enhanced Brain-Cell-Type Targeting and Functionality. *Cell* **174**, 465-480.e22 (2018).

14. Harris, J. A. *et al.* Anatomical characterization of Cre driver mice for neural circuit mapping and manipulation. *Front. Neural Circuits* **8**, (2014).

15. Bredt, D. S. & Snyder, S. H. Transient nitric oxide synthase neurons in embryonic cerebral cortical plate, sensory ganglia, and olfactory epithelium. *Neuron* **13**, 301–313 (1994).

16. Pollak Dorocic, I. *et al.* A Whole-Brain Atlas of Inputs to Serotonergic Neurons of the Dorsal and Median Raphe Nuclei. *Neuron* **83**, 663–678 (2014).

17. Ren, J. *et al.* Anatomically Defined and Functionally Distinct Dorsal Raphe Serotonin Sub-systems. *Cell* **175**, 472-487.e20 (2018).

18. Chen, X., Petit, E. I., Dobrenis, K. & Sze, J. Y. Spatiotemporal SERT expression in cortical map development. *Neurochemistry International* **98**, 129–137 (2016).

19. Chen, X. *et al.* Disruption of Transient Serotonin Accumulation by Non-Serotonin-Producing Neurons Impairs Cortical Map Development. *Cell Reports* (2015) doi:10.1016/j.celrep.2014.12.033.

20. Fremeau, R. T. *et al.* The Expression of Vesicular Glutamate Transporters Defines Two Classes of Excitatory Synapse. *Neuron* **31**, 247–260 (2001).

21. Nakamura, K. *et al*. Transiently increased colocalization of vesicular glutamate transporters 1 and 2 at single axon terminals during postnatal development of mouse neocortex: A quantitative analysis with correlation coefficient. *European Journal of Neuroscience* (2007) doi:10.1111/j.1460-9568.2007.05868.x.

22. Boulland, J. *et al.* Expression of the vesicular glutamate transporters during development indicates the widespread corelease of multiple neurotransmitters. *J of Comparative Neurology* **480**, 264–280 (2004).

23. He, H. *et al.* Neurodevelopmental Role for VGLUT2 in Pyramidal Neuron Plasticity, Dendritic Refinement, and in Spatial Learning. *J. Neurosci.* **32**, 15886–15901 (2012).

24. Gras, C. *et al.* A Third Vesicular Glutamate Transporter Expressed by Cholinergic and Serotoninergic Neurons. *J. Neurosci*. **22**, 5442–5451 (2002).

25. Wang, H.-L. *et al*. Dorsal Raphe Dual Serotonin-Glutamate Neurons Drive Reward by Establishing Excitatory Synapses on VTA Mesoaccumbens Dopamine Neurons. *Cell Reports* **26**, 1128-1142.e7 (2019).

26. Fremeau, R. T., Voglmaier, S., Seal, R. P. & Edwards, R. H. VGLUTs define subsets of excitatory neurons and suggest novel roles for glutamate. *Trends in Neurosciences* **27**, 98–103 (2004).

27. Schäfer, M. K.-H., Varoqui, H., Defamie, N., Weihe, E. & Erickson, J. D. Molecular Cloning and Functional Identification of Mouse Vesicular Glutamate Transporter 3 and Its Expression in Subsets of Novel Excitatory Neurons. *Journal of Biological Chemistry* **277**, 50734–50748 (2002).

28. Zhang, M. *et al. A Molecularly Defined and Spatially Resolved Cell Atlas of the Whole Mouse Brain*. http://biorxiv.org/lookup/doi/10.1101/2023.03.06.531348 (2023) doi:10.1101/2023.03.06.531348.

29. Gonchar, Y. Multiple distinct subtypes of GABAergic neurons in mouse visual cortex identified by triple immunostaining. *Front. Neuroanat.* **1**, (2008).

30. Satopaa, V., Albrecht, J., Irwin, D. & Raghavan, B. Finding a 'Kneedle' in a Haystack: Detecting Knee Points in System Behavior. in *2011 31st International Conference on Distributed Computing Systems Workshops* 166–171 (IEEE, Minneapolis, MN, USA, 2011). doi:10.1109/ICDCSW.2011.20.

31. Zingg, B. *et al.* Neural networks of the mouse neocortex. *Cell* **156**, 1096–111 (2014).

32. Zhang, S. *et al.* Organization of long-range inputs and outputs of frontal cortex for top-down control. *Nat Neurosci* **19**, 1733–1742 (2016).

33. Lyamzin, D. & Benucci, A. The mouse posterior parietal cortex: Anatomy and functions. *Neuroscience Research* **140**, 14–22 (2019).

34. Fischer, L. F., Mojica Soto-Albors, R., Buck, F. & Harnett, M. T. Representation of visual landmarks in retrosplenial cortex. *eLife* **9**, e51458 (2020).

35. Vogt, B. A. & Miller, M. W. Cortical connections between rat cingulate cortex and visual, motor, and postsubicular cortices. *J of Comparative Neurology* **216**, 192–210 (1983).

36. De Araujo, I. E. & Simon, S. A. The gustatory cortex and multisensory integration. *Int J Obes* **33**, S34–S43 (2009).

37. Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. in *Proceedings of the Second International*

*Conference on Knowledge Discovery and Data Mining* 226–231 (AAAI Press, Portland, Oregon, 1996).

38. Schubert, E., Sander, J., Ester, M., Kriegel, H. P. & Xu, X. DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN. *ACM Trans. Database Syst.* **42**, 1–21 (2017).

39. Attili, S. M., Moradi, K., Wheeler, D. W. & Ascoli, G. A. Quantification of neuron types in the rodent hippocampal formation by data mining and numerical optimization. *Eur J of Neuroscience* **55**, 1724–1741 (2022).

40. Park, S.-B. *et al.* The fasciola cinereum subregion of the hippocampus is important for the acquisition of visual contextual memory. *Progress in Neurobiology* **210**, 102217 (2022).

41. Kim, Y. *et al.* Mapping social behavior-induced brain activation at cellular resolution in the mouse. *Cell reports* **10**, 292–305 (2015).

42. Ragan, T. *et al.* Serial two-photon tomography for automated ex vivo mouse brain imaging. *Nature methods* **9**, 255–8 (2012).

43. Renier, N. *et al.* IDISCO: A simple, rapid method to immunolabel large tissue samples for volume imaging. *Cell* **159**, 896–910 (2014).

44. Renier, N. *et al.* Mapping of Brain Activity by Automated Volume Analysis of Immediate Early Genes. *Cell* **165**, 1789–1802 (2016).

45. Klein, S., Staring, M., Murphy, K., Viergever, M. A. & Pluim, J. P. W. Elastix: A toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging* **29**, 196–205 (2010).

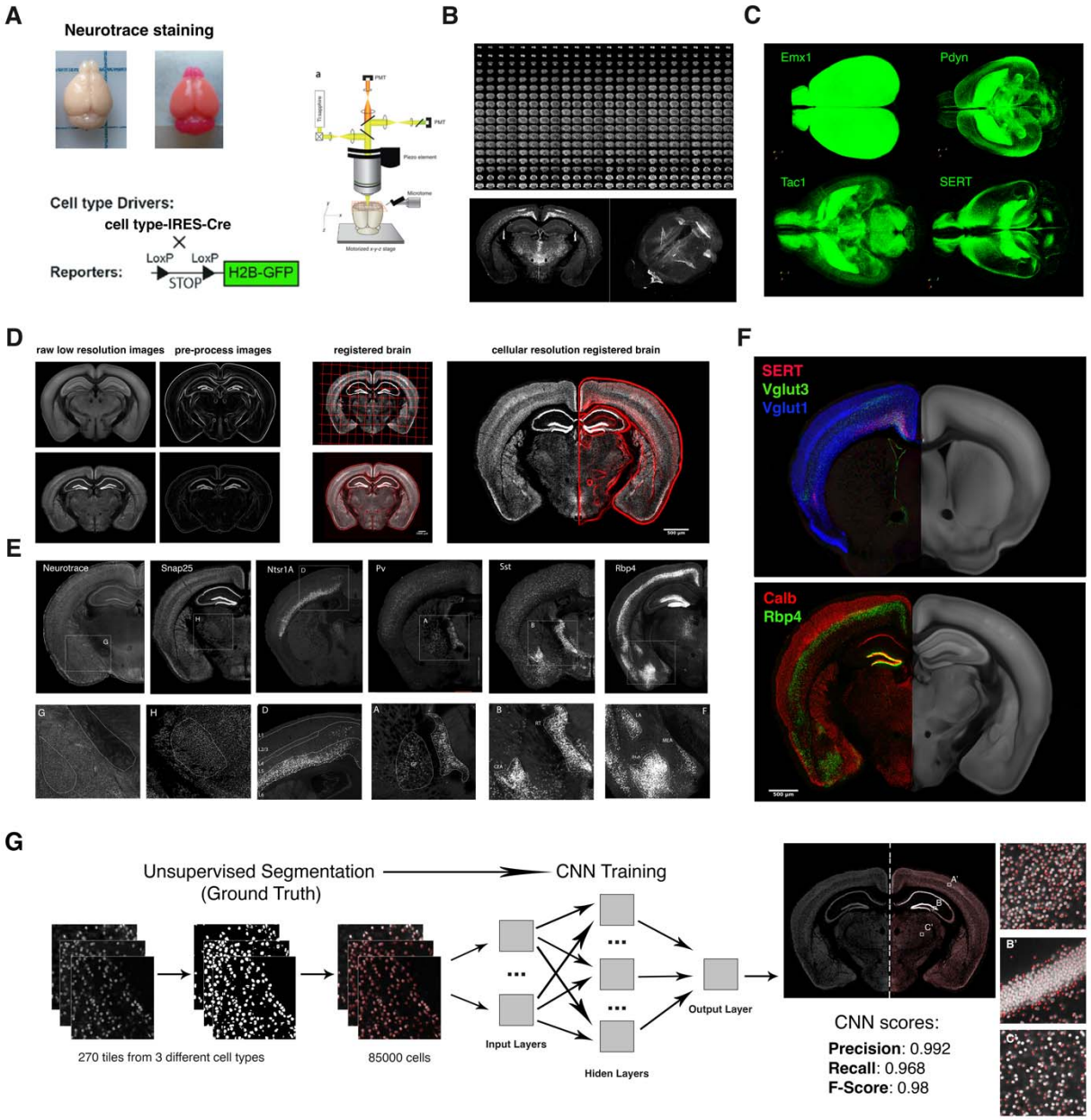


**Figure1: Whole brain analysis and visualization at cellular resolution**: **A)** Serial Two Photon Tomography (STPT) imaging of mouse brains after fluorescent Nissl staining (NeuroTrace) or from cell-type Cre-driver lines crossed with fluorescent protein Cre-reporter lines. (Left top) Mouse brain stained with NeuroTrace before and after. (Left) A diagram of Cre-LoxP labeling strategies. (Right) A schema of STPT instrument for automated whole mouse brain imaging. (Ab) (Top) Resulting 280 serial sections of whole mouse brain imaging at 1x1x50 um resolution. **B)** An example coronal section and (bottom right) a 3D renderization of the whole brain. **C)** Examples of 3D renderization of fluorescent H2BGFP cell distribution for Emx1, Pdyn, Tac1 and SERT Cre-driver lines. **D)** Intrinsic anatomical features (IAFs) based 3D registration of mouse brain at cellular resolution. (Left) Generation of anatomical features from downsampled 20x20x50 um whole brain images for both mean intensity reference brain(top) and an individual sample brain. (Middle) The resulting grid space deformation after registration (top) and the overlay of the STPT imaged brain onto the reference IAFs based brain (bottom). (Right) The resulting transformations obtained from the downsampled images were

applied to the original 1x1x50 um resolution brain images and warped onto a high-resolution reference space coordinate system. **E)** Examples of NeuroTrace and different Cre-driver lines crossed with the H2BGFP Cre report showing cell type-specific anatomical distribution before registration. **F)** Overlay of selected 4 Cre-driver lines SERT+VGlut3+VGlut1 (top) and Calb2+Rbp4 (bottom) after registration at the 1x1x50 um cellular resolution (left) onto the reference brain (right), showing their cellular distribution organization. **G)** (Left panel) Machine learning-based cell detection. Unsupervised segmentation strategy was used for the generation of a large dataset of marked cell nuclei from 3 different imaged brains used for training of a convolutional neural network (CNN). Resulting accuracy was validated using manually marked up ground truth data. (Right panel) Coronal plane of Snap-H2B-GFP with left hemisphere overlay onto the output segmentation, with details showing the segmentation results at the cortex (A'), hippocampal formation (B') and the thalamus (C').

| Mouse Line - Abbreviated Nam | Mouse Line Full Name | Cell Type | Driver, Reporter | Generation Method | Type | Originating Lab (Donating Investigator | Reference(s) | Public Reposi | Stock # | Cortex | Pallidum | Thalamus | Hypothalamu | Midbrain | Hindbrain | Cerebellum |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Avpt4-Cre | B6.Cg-Avpt4m1.1(cre)HzeJ | Vasopressin neurons - GABAergic | Driver | Knock-in | IRES | Allen Institute for Brain Science | Harris et al. 2014 Front Neural Circuits (PMID: 25071457) | JAX | 23530 | | | | | | | |
| Calb-Cre | B6.129S-Calb1tm2.1(cre)HzeJ | Calbindin | Driver | Knock-in | IRES2 | Allen Institute for Brain Science | Daigle et al. 2018 Cell (PMID: 30007418) | JAX | 28532 | | | | | | | |
| Calb2-Cre | B6(Cg)-Calb2tm1(cre)Zjh/J | Calretinin | Driver | knock-in | IRES | Allen Institute for Brain Science | | JAX | 10774 | | | | | | | |
| ChAT-IRES-Cre | B6.129S-Chattm1(cre)Lowl/MwarJ | Cholinergic neurons | Driver | knock-in | IRES | Allen Institute for Brain Science | Daigle et al. 2018 Cell (PMID: 30007418) | JAX | 31661 | | | | | | | |
| Crh-Cre | B6(Cg)-Crhtm1(cre)Zjh/J | Corticotropin releasing hormone neurons | Driver | knock-in | IRES | Z. Josh Huang | Taniguchi et al. 2011 Neuron (PMID: 21943598) | JAX | 12704 | | | | | | | |
| Ctgf-T2A-dgCre | Ctgf-T2A-dgCre | L6b PNs | Driver | Knock-in | T2A | Allen Institute for Brain Science | Tasic et al. 2016 Nat Neurosci (PMID: 26727548) | JAX | 28535 | | | | | | | |
| Cux2-IRES-Cre | B6(Cg)-Cux2tm2.1(cre)Mull/Mmmh | L2/3/4 PNs | Driver | Knock-in | IRES | Ulrich Mueller | NIH Neuroscience Blueprint Cre Driver Network 2009 | MMRRC | 32778 | | | | | | | |
| DAT-Cre | B6.SJL-Slc6a3tm1.1(cre)Bkmn/J | Dopaminergic neurons | Driver | Knock-in | IRES | Cristina Backman | Backman CM , et al 2006 Genesis (PMID:16865686) | JAX | 6660 | | | | | | | |
| Emx1-IRES-Cre | B6.129S2-Emx1tm1(cre)Krj/J | Broad PNs | Driver | Knock-in | IRES | Kevin R. Jones | Gorski et al. 2002 JNeurosci (PMID: 12151506) | JAX | 5628 | | | | | | | |
| Gad2-Cre | Gad2tm2(cre)Zjh/J | Broad GABAergic | Driver | Knock-in | IRES | Z. Josh Huang | Z. Josh Huang | JAX | 10802 | | | | | | | |
| nNOS-Cre | B6.129-Nos1tm1(cre)Mgm/J | Nitric oxide neurons | Driver | knock-in | IRES | Martin G Myers | Leshan et al. 2012 Nat Med (PMID: 22522563) | JAX | 14541 | | | | | | | |
| NPY-Cre | B6.Cg-Npytm1(cre)Zman/J | Neuropeptide Y neurons | Driver | Knock-in | IRES | Boris Zemelman | Milstein et al. 2015 Neuron (PMID: 26402609) | JAX | 27851 | | | | | | | |
| Ntsr1-Cre; Ntsr1-Cre_GN220 | B6.FVB(Cg)-Tg(Ntsr1-cre)GN220Gsat/Mmucd | L6 PNs | Driver | Transgenic | BAC | Nathaniel Heintz and Charles Gerfen | Gong et al. 2007 J Neurosci (PMID: 17855595) | MMRRC | 30648 | | | | | | | |
| Olig2-Cre | B6.129-Olig2tm1.1(cre)Wdr/J | Oligodendrocytes | Driver | Knock-in | IRES | William Richardson | Zawadzka et al. 2010 Cell Stem Cell (PMID: 20569695) | JAX | 25567 | | | | | | | |
| Oxt-Cre | B6.129S-Oxttm1.1(cre)Dolsn/J | Oxytocin neurons | Driver | Knock-in | IRES | David Olson | Wu Z , et al. 2012 Plos One (PubMed:23028821) | JAX | 24234 | | | | | | | |
| Pdyn-IRES-Cre | B6.129S-Pdyntm1.1(cre)Mjkr/LowlJ | Prodynorphin neurons | Driver | knock-in | IRES | Bradford B. Lowell | Krashes et al. 2014 Nature (PMID: 24487620) | JAX | 27958 | | | | | | | |
| Pvalb-IRES-Cre | B6.129P2-Pvalbtm1(cre)Arbr/J | Parvalbumin | Driver | Knock-in | IRES | Silvia Arber | Hippenmeyer et al. 2005 PLoS Biol (PMID: 15836427) | JAX | 8069 | | | | | | | |
| Rbp4-Cre; Rbp4-Cre_KL100 | STOCK Tg(Rbp4-cre)KL100Gsat/Mmucd | L5 PNs | Driver | Transgenic | BAC | Nathaniel Heintz and Charles Gerfen | Gong et al. 2007 J Neurosci (PMID: 17855595) | MMRRC | 31125 | | | | | | | |
| Rorb-IRES2-Cre | Rorb-IRES2-Cre-D | L4/5 PNs | Driver | Knock-in | IRES2 | Allen Institute for Brain Science | Harris et al. 2014 Front Neural Circuits (PMID: 25071457) | JAX | 23526 | | | | | | | |
| SERT-Cre | B6.129(Cg)-Slc6a4tm1(cre)Xz/J | Serotoninergic neurons | Driver | knock-in | | Xiaoxi Zhuang | Zhuang et al 2005 J Neurosci Method (PMID: 15763133) | JAX | 14554 | | | | | | | |
| Slc32a1-IRES-Cre | Slc32a1-IRES-Cre | Broad GABAergic | Driver | Knock-in | IRES | Bradford B. Lowell | Vong et al. 2011 Neuron (PMID: 21745644) | JAX | 28862 | | | | | | | |
| Snap25-Cre | Snap25-IRES2-Cre-D | Pan neuronal | Driver | Knock-in | IRES2 | Allen Institute for Brain Science | Harris et al. 2014 Front Neural Circuits (PMID: 25071457) | JAX | 23525 | | | | | | | |
| Sst-IRES-Cre | Ssttm2.1(cre)Zjh/J | Somatostatin | Driver | Knock-in | IRES | Z. Josh Huang | Taniguchi et al. 2011 Neuron (PMID: 21943598) | JAX | 13044 | | | | | | | |
| Tac1-Cre | B6.129S-Tac1tm1.1(cre)Hze/J | Tacykinin1 neurons | Driver | Knock-in | IRES2 | Allen Institute for Brain Science | Harris et al. 2014 Front Neural Circuits (PMID: 25071457) | | | | | | | | | |
| TH-Cre | B6.Cg-7630403G23RikTg(Th-cre)1Tmd/J | Tyrosine hydroxylase neuons | Driver | knock-in | | Ted M Dawson | Savitt et al. 2005 J Neurosci (PMID: 16033881) | JAX | 8601 | | | | | | | |
| Tlx3-Cre; Tlx3-Cre_PL56 | STOCK Tg(Tlx3-cre)PL56Gsat/Mmucd | L5a neurons | Driver | Transgenic | BAC | Nathaniel Heintz and Charles Gerfen | Gong et al. 2007 J Neurosci (PMID: 17855595) | MMRRC | 36547 | | | | | | | |
| Vglut1-Cre | Slc17a7-IRES2-Cre-D | Glutamatergic transporter 1 neurons | Driver | Knock-in | IRES2 | Allen Institute for Brain Science | Harris et al. 2014 Front Neural Circuits (PMID: 25071457) | JAX | 23527 | | | | | | | |
| Vglut2-Cre | Slc17a6-IRES-Cre | Glutamatergic transporter 2 neurons | Driver | Knock-in | IRES | Bradford B. Lowell | Vong et al. 2011 Neuron (PMID: 21745644) | JAX | 28863 | | | | | | | |
| Vglut3-Cre | Slc17a8-IRES2-Cre-D | Glutamatergic transporter 3 neurons | Driver | Knock-in | IRES2 | Allen Institute for Brain Science | Harris et al. 2014 Front Neural Circuits (PMID: 25071457) | JAX | 28534 | | | | | | | |
| Vip-IRES-Cre | Viptm1(cre)Zjh/J | Vasoactive intestinal peptide neurons - GABAergic | Driver | Knock-in | IRES | Z. Josh Huang | Taniguchi et al. 2011 Neuron (PMID: 21943598) | JAX | 10908 | | | | | | | |

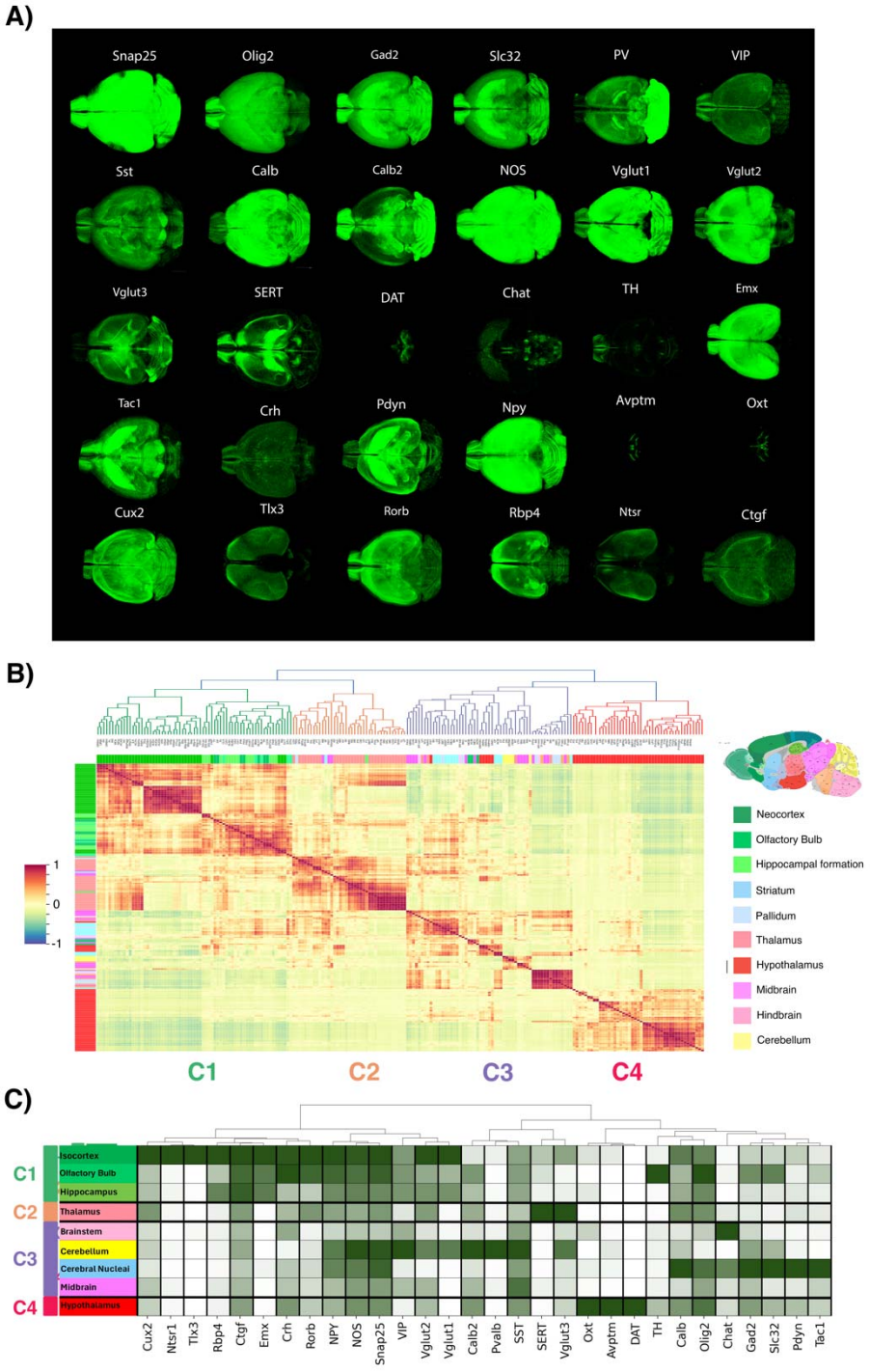


**Figure2: Whole brain 3D distributions visualization of cell types and analysis**: **A)** 3D point-cloud visualizations from 30 different cell types: Snap25 - Synaptosomal associated protein 25-; Olig2 - Oligodendrocyte Transcription Factor 2-; Gad2 - Glutamate Decarboxylase 2-; Slc32 - Vesicular GABA transporter-; Pv -Parvaobumin-; VIP - Vasoactive Intestinal Peptide-; Sst -Somatostatin-; Calb -Calbindin-; Calb2 -Calretinin-; NOS - Neuronal nitric oxide synthase 1-; VGlut1 - Vesicular glutamate transporter 1-; VGlu2 - Vesicular glutamate transporter 2-; VGlut3 - Vesicular glutamate transporter 3-; SERT -Serotonin Transporter-; Chat - Choline acetyltransferase -; DAT - Dopamine Transporter -; TH - Tyrosine Hydroxylase-;

Emx1 - Empty Spiracles Homeobox 1-; Tac1 - Tachykinin precursor 1-; Crh - Corticotropin-releasing hormone-; Pdyn - Prodynorphin-; Npy -Neuropeptide Y-; Avptm - arginine vasopressin-;-; Oxt - Oxytocine-; Cux2 - Cut Like Homeobox 2-; Tlx3 - T Cell Leukemia Homeobox 3; Rorb - RAR Related Orphan Receptor B-; Rbp4 - Retinol binding protein 4-; Ntsr1A - Neurotensin receptor type 1-; and Ctgf - Connective tissue growth factor-. **B)** Clustermap showing mouse brain organization based on the distribution of all the cells. Cluster analysis shows that brain areas organize in 4 major clusters based on their cell distribution: cortex (cluster C1, green dendrogram), thalamus (cluster C2, orange dendrogram); striatum, pallidum, midbrain, hindbrain and cerebellum (cluster C3, purple dendrogram) and hypothalamus (cluster C4, red dendrogram). Thus, major brain has a very distinctive cellular organization defining their internal structure. **C)** Normalized cell type distribution on major brain areas. X-axis show all cell types clustered by normalized expression, and Y-axis shows major brain areas organized by panel B) brain cluster organization.

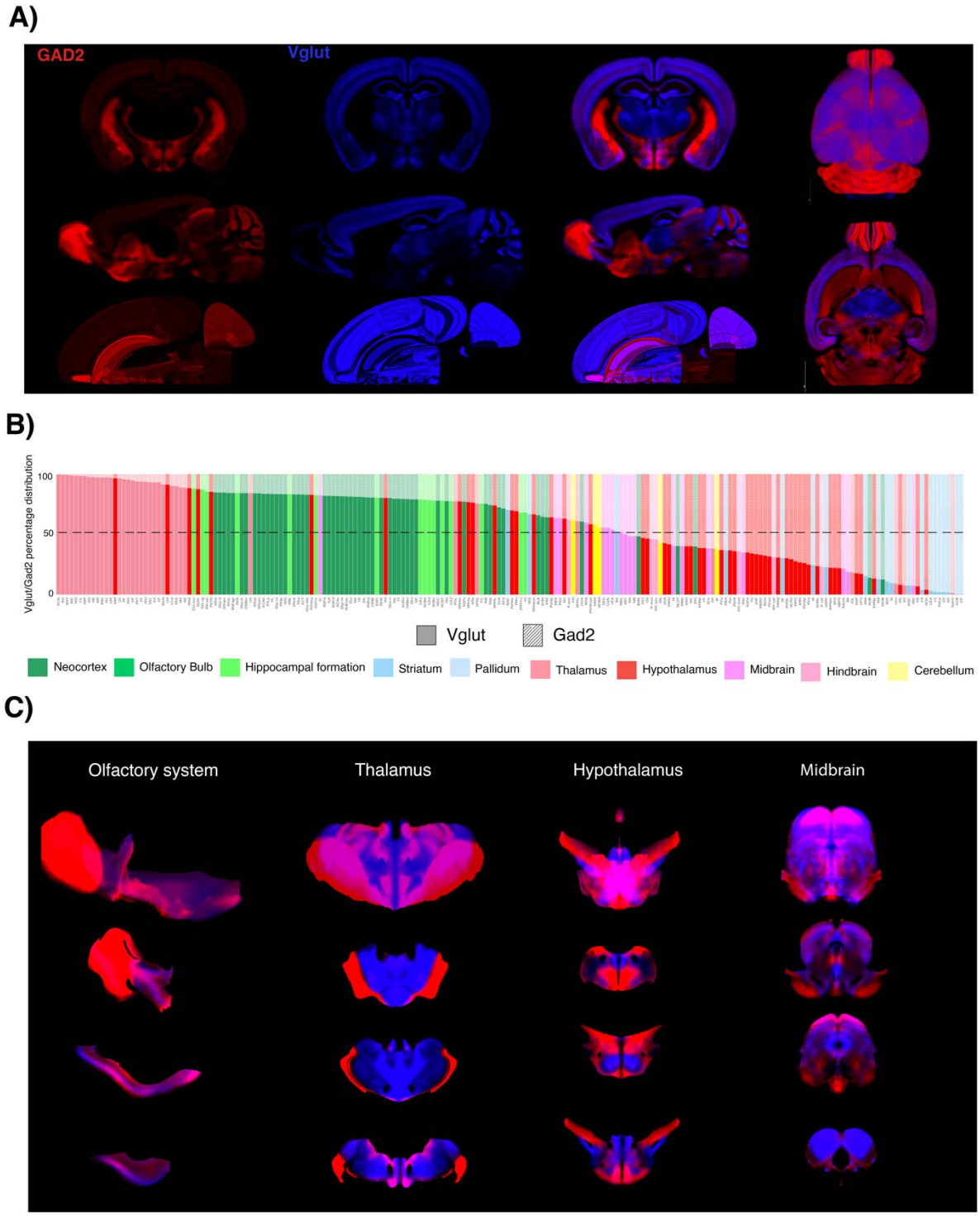


**Figure 3: Whole brain distribution of excitatory and inhibitory neurons. A)** Voxel-wise coronal (upper panel) and sagittal (middle panel) views showing wise distribution of Gad2 (red) and VGlut (blue) on the mouse brain, with a whole brain flatmap representation of their distribution among all brain areas (lower panel). Right panel shows overlay of Gad2 and VGlut point-clouds in 3D (upper panel) and horizontal view (lower panel). **B)** Ratio of VGlut:Gad2 cell distribution among all brain regions analyzed color-coded by major

regions. Neocortex, thalamus, midbrain and cerebellum show a predominance of excitatory neurons. In relation to the thalamic nuclei, GENv, RT and PP have a higher percentage of inhibitory neurons, contrary to all other thalamic nuclei. Similarly, in the midbrain SNr, RR, IV, VTA, AT and LT have a clear predominance of inhibitory neurons. **C)** Upper row shows 3D voxel-wise spatial distribution of VGlut and Gad2 in the olfactory system, the thalamus, the hypothalamus and the midbrain. Lower rows show three different planes (sagittal for the olfactory system and coronal for the other brain regions) displaying the VGlut and Gad2 spatial distribution.

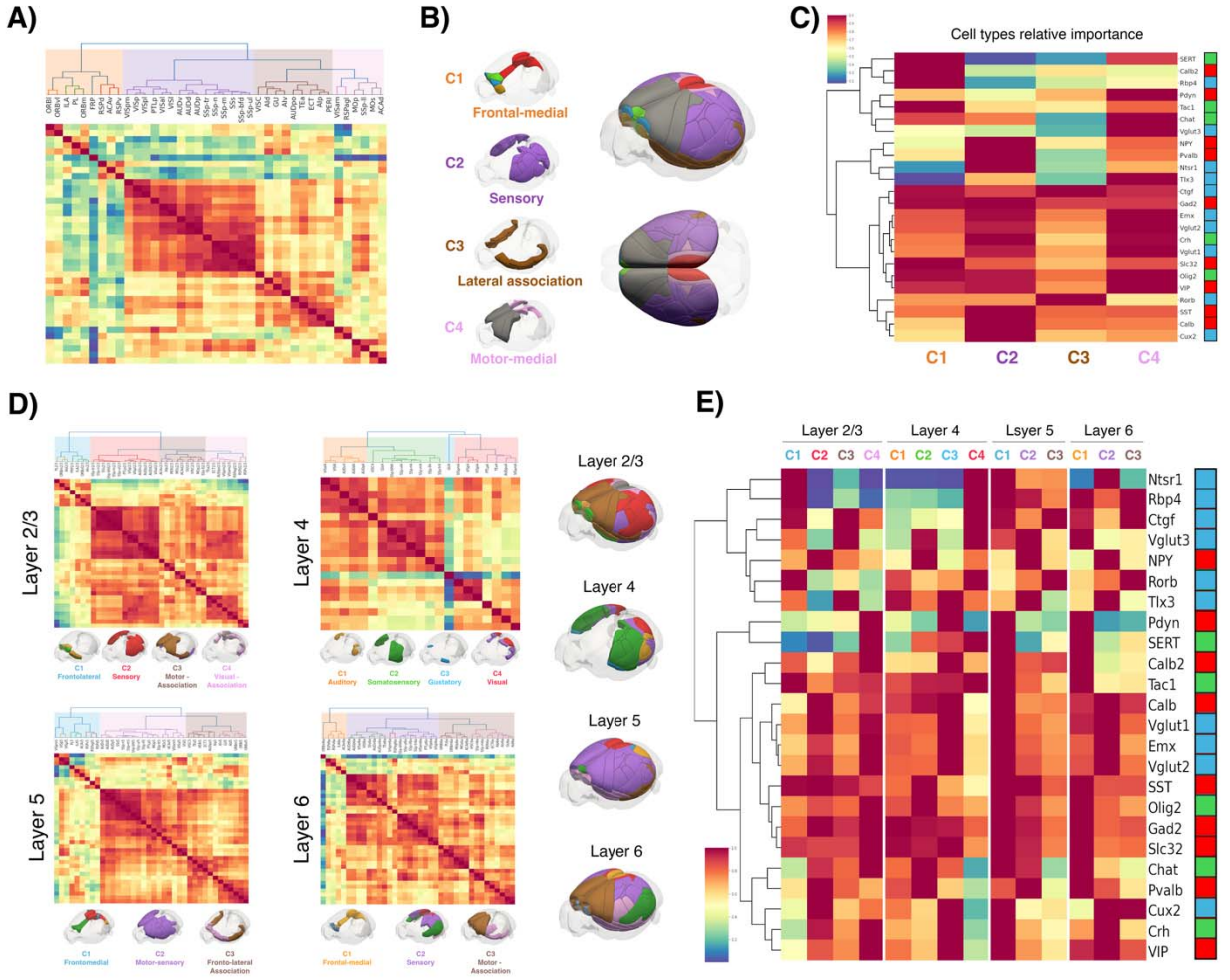


**Figure 4: Cell type-based cortical organization. A)** Cell type-based clustermap of the entire cortex revealing its organization into four major clusters, further subdivided into eight subclusters. **B)** Whole cortex 3D renderization showing the spatial distribution of all cortical areas according to the major cluster organization. Each individual area is color-coded based on the eight subclusters identified in the dendrogram (A). **C)** Normalized cell type distribution of major cortical clusters. X-axis shows the major cortical clusters defined in A); Y-axis shows all cell types clustered by normalized expression. Cell-type clustermap divides cells cortical distribution into three main clusters. **D)** Layer-based cell type-based clustermaps showing distinct cortical organizations for each four cortical layers and the spatial organizations. Layer 2/3 shows four major clusters further subdivided into seven cluster. Lower panel shows the spatial distribution of each major cluster clor-coded by on the seven subclusters identified in the dendrogram. Layer 4 shows four major clusters further subdivided into five cluster. Lower panel shows the spatial distribution of each major cluster clor-coded by on the five subclusters identified in the dendrogram. Layer 5 shows three major clusters further subdivided into seven cluster. Lower panel shows the spatial distribution of each major cluster clor-coded by on the seven subclusters identified in the dendrogram. Layer 6 shows 3 major clusters further subdivided into seven cluster. Lower panel shows the spatial distribution of each major cluster clor-coded by on the seven subclusters identified in the dendrogram. Right panel shows the 3D renderization of each layer cluster organization with all clusters together. **E)** Normalized cell type distribution per layer of all layer-based clusters from D). X-axis shows the major cortical clusters defined in D) for each layer; Y-axis shows all cell types clustered by normalized expression. Cell-type clustermap divides cells cortical distribution into four main clusters.

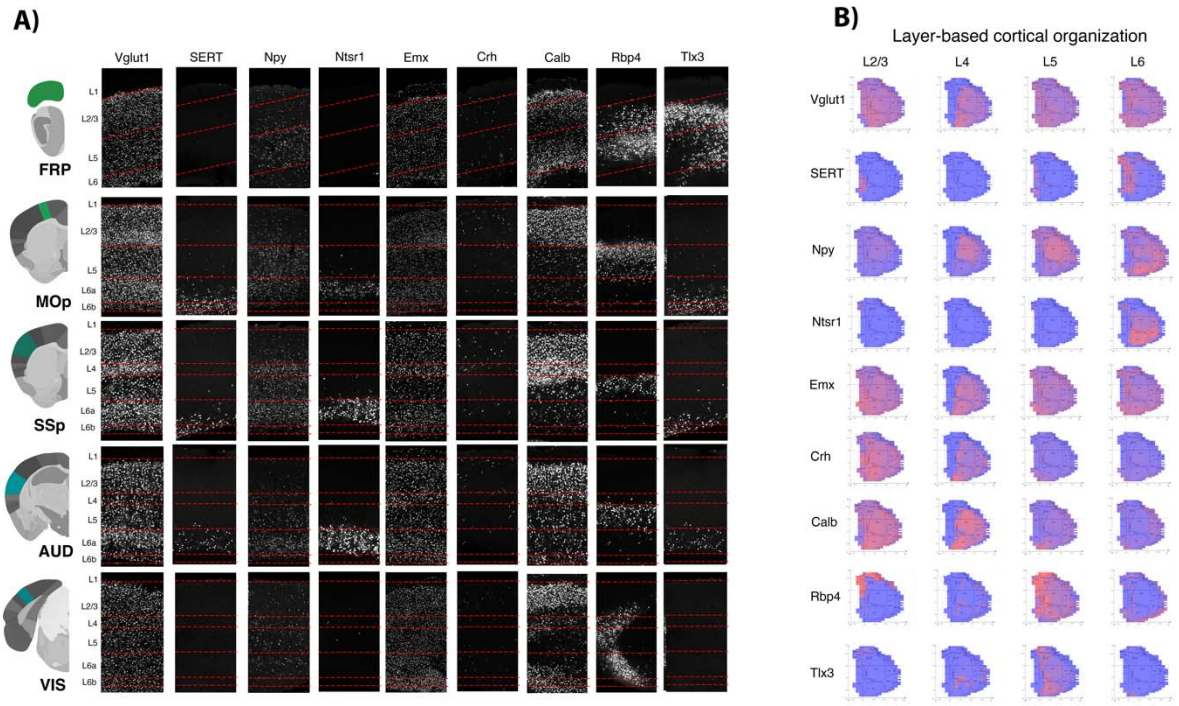


**Figure 5: Cortical density distribution: A)** STPT coronal sections showing cellular distribution of nine main cell types (VGlut1, SERT, NPY, Ntsr1, Emx, Crh, Calb, Rbp4 and Tlx3) in five major cortical areas: frontal pole (FRP); primary motor cortex (MOp); primary somatosensiry cortex (SSp); auditory cortex (AUD); and visual cortex (VIS). B) Layer-based cortical flatmaps of the same cell types seen on A).

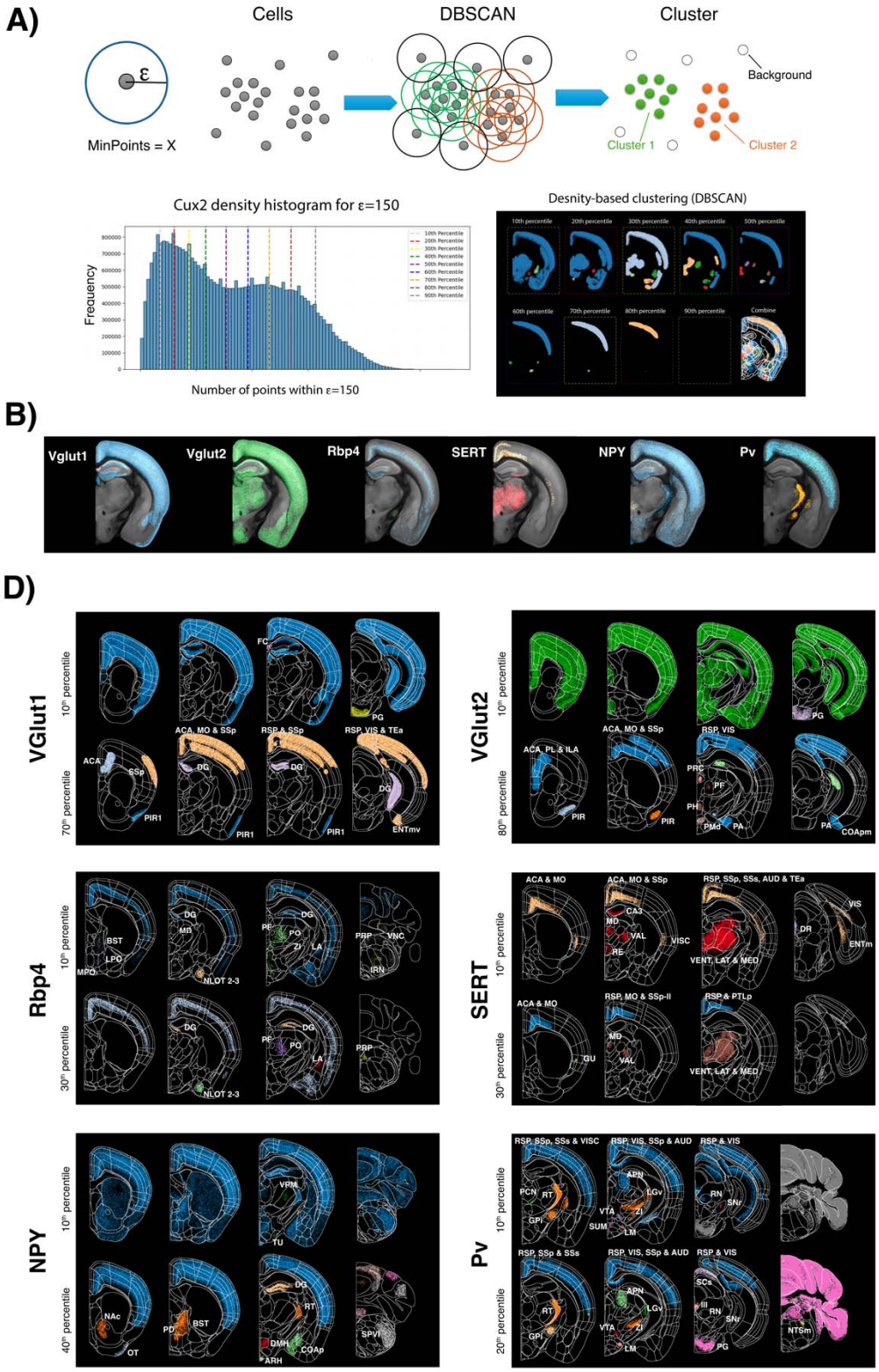


**Figure 6: Density-based whole brain spatial organization. A)** (Upper panle) Scheme showing density-based DBSCAN spatial clustering of brain cell distribution. e defines the maximum distance (radious) between two points for them to be considered neighbors (in our analysis ε=150 mm); MinPoints defines the minimum

number of points/neighbors within ε requiered to be considered a dense region, i.e. a cluster. (Lower panel) The definition of MinPoints is based on the histogram density frequency (left panel), thus taking into account the varying density of each individual cell type based on its intrinsic cell density distributions. MinPoints values are selected using 9 percentiles from the histogram density distribution (from $10^{th}$ to $90^{th}$ percentile), in this example for Cux2. Afterwards it is analyzed nine different DBSCAN clusterings, one for each percentile, and the resulting spatial density clusters are individually visualized in the whole brain (right panel), with each color representing the cells belonging to an individual cell density cluster. **B)** Coronal plane from six relevant cell types selected based fron the previous analyses, showing the spatial-density clustering at minimum density ($10^{th}$ percentile). **C)** Coronal planes showing the spatial-density distribution of the previous cell types at $10^{th}$ percentile and a second percentile contigent to each cell type: VGlut1 ($10^{th}$ and $70^{th}$ percentiles), VGlut2 ($10^{th}$ and $80^{th}$ percentiles), Rbp4 ($10^{th}$ and $30^{th}$ percentiles), SERT ($10^{th}$ and $30^{th}$ percentiles), NPY ($10^{th}$ and $40^{th}$ percentiles) and Pv ($10^{th}$ and $20^{th}$ percentiles). All coronal planes were overlay onto the CCFv3 to show the anatomical distributions of the density analysis.