

# Cell-type-specific microarray data and the Allen atlas: quantitative analysis of brain-wide patterns of correlation and density

Pascal Grange<sup>1,\*</sup>, Michael Hawrylycz<sup>2</sup>, Partha P. Mitra<sup>1</sup>

<sup>1</sup> Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, United States

<sup>2</sup> Allen Institute for Brain Science, Seattle, Washington 98103, United States

\*E-mail: [pascal.grange@polytechnique.org](mailto:pascal.grange@polytechnique.org)

## Abstract

The Allen Atlas of the adult mouse brain is used to estimate the region-specificity of 64 cell types whose transcriptional profile in the mouse brain has been measured in microarray experiments. We systematically analyze the preliminary results presented in [arXiv:1111.6217], using the techniques implemented in the Brain Gene Expression Analysis toolbox. In particular, for each cell-type-specific sample in the study, we compute a brain-wide correlation profile to the Allen Atlas, and estimate a brain-wide density profile by solving a quadratic optimization problem at each voxel in the mouse brain. We characterize the neuroanatomical properties of the correlation and density profiles by ranking the regions of the left hemisphere delineated in the Allen Reference Atlas. We compare these rankings to prior biological knowledge of the brain region from which the cell-type-specific sample was extracted.

# Contents

<b>1</b>	<b>Introduction and background</b>	<b>3</b>
<b>2</b>	<b>Methods</b>	<b>4</b>
2.1	Correlations between cell-type-specific microarray data and the Allen Atlas . . .	4
2.2	A linear model relating cell-type-specific microarray data to the Allen Atlas . . .	5
<b>3</b>	<b>Results</b>	<b>6</b>
3.1	Rankings of brain regions induced by correlation and density profiles . . . . .	6
3.1.1	Ranking brain regions by correlation profile . . . . .	8
3.1.2	Ranking brain regions by density profile . . . . .	8
3.2	Anatomical data for cell-type-specific samples . . . . .	10
3.3	Neuroanatomical patterns of results, grouped by the regions in the 'big12' annotation of the left hemisphere . . . . .	12
3.3.1	Basic cell groups and regions . . . . .	12
3.3.2	Cerebral cortex . . . . .	12
3.3.3	Hippocampal region . . . . .	24
3.3.4	Striatum . . . . .	29
3.3.5	Pallidum . . . . .	35
3.3.6	Thalamus . . . . .	37
3.3.7	Midbrain . . . . .	40
3.3.8	Pons . . . . .	46
3.3.9	Medulla . . . . .	49
3.3.10	Cerebellum . . . . .	52
<b>4</b>	<b>Discussion of the linear model</b>	<b>55</b>
4.1	Competition between pairs of similar cell types . . . . .	56
4.1.1	Methods . . . . .	56
4.1.2	Results . . . . .	58
4.2	The set of pyramidal neurons . . . . .	69
4.3	Background value in microarray data . . . . .	73
<b>5</b>	<b>Tables of plots of estimated densities, original and corrected for cross-hybridization</b>	<b>81</b>
<b>6</b>	<b>Conclusions</b>	<b>104</b>
<b>7</b>	<b>Acknowledgments</b>	<b>104</b>
<b>8</b>	<b>Tables of brain-wide correlations</b>	<b>105</b>
<b>9</b>	<b>Tables of estimated brain-wide densities</b>	<b>117</b>
<b>10</b>	<b>Tables of description of cell-type specific samples</b>	<b>129</b>

## 1 Introduction and background

Neuroanatomy is experiencing a renaissance thanks to molecular biology and computational imaging [1]. The Allen Brain Atlas (ABA), the first Web-based, genome-wide atlas of gene expression in the adult mouse brain (eight-week old C57BL/6J male mouse brain), was obtained using an unified automated experimental pipeline [2, 3, 4, 5, 6, 7, 8]. The resulting data set consists of *in situ* hybridization (ISH) digitized image series for thousands of genes. These image series are co-registered to the Allen Reference Atlas (ARA) [9], which allows to compare ISH data to classical neuroanatomy.

The gene expression data were aggregated into a volumetric grid: The reference mouse brain is partitioned into  $V = 49,742$  cubic voxels of side 200 microns. For a voxel  $v$ , the *expression energy* of the gene labeled  $g$  is defined [10] as a weighted sum of the greyscale-value intensities of pixels  $p$  intersecting the voxel:

$$E(v, g) := \frac{\sum_{p \in v} M(p)I(p)}{\sum_{p \in v} 1}, \quad (1)$$

where  $I(p)$  is the intensity image and  $M(p)$  is a Boolean mask, worked out in the image-processing pipeline, with value 1 if the pixel is expressing the gene of interest. The ABA has led to the development of software for data exploration and analysis such as the Web-based Neuroblast [11] and Anatomic Gene Expression Atlas (AGEA) [10]. NeuroBlast allows users to explore the correlation structure between genes, while the AGEA is based on spatial correlation between voxels.

More recently, we developed the Brain Gene Expression Analysis (BGEA) MATLAB toolbox [12, 13], which allows to manipulate the gene-expression energies of the brain-wide ABA on the desktop as matrices [14, 15]. In particular, data corresponding to different ISH experiments (i.e. different genes) can be combined for computational analysis, and used to study other data sets.

A complementary (cell-based) approach to the study of gene-expression energy in the brain uses microarray experiments to study co-expression patterns in a small set of brain cells of the same type. We studied cell-type-specific microarray gathered from different studies [16, 17, 18, 19, 20, 21, 22, 23], already analyzed in [24], for  $T = 64$  cell types<sup>1</sup>. Each of these cell types is characterized by the expression of  $G_T = 14,580$  genes.

The region specificity of the transcriptional profiles of cell types is an open problem (preliminary results were presented in [25], without a systematic anatomical analysis). Given a cell

---

<sup>1</sup>The studies differ in the way cells are visually-identified are separated from their environments. The methods are laser-capture microdissection (LCM) [20, 17], translating ribosome affinity purification (TRAP) [19, 22], fluorescence-activated cell-sorting (FACS) [21], immuno panning (PAN) [18], and manual sorting [16, 23]. See Tables 62 and 63 for references to the study from which each of the cell-type-specific samples was drawn.

type extracted from a given brain region, it is hard to know where else in the brain cells with a similar transcriptional profile can be found. We computed brain-wide correlation profiles between each of the cell-type-specific samples and the Allen Atlas. Moreover, we propose a linear model decomposing the gene-expression energy of the collection of all genes in the Allen Atlas over the set of cell-type-specific data. The model provides an estimate of the brain-wide density of the cell types. For each cell type, brain regions are then ranked according to the average value of correlation, and to their contribution to the total density of the cell type.

## 2 Methods

We determined the set of genes that are represented in *both* data sets (there are  $G = 2,131$  such genes). We extracted the columns of the matrices  $E$  and  $C$  corresponding to these genes, and rearranged them to reflect the same order in each data set.  $E$  is assumed to be a voxel-by-gene matrix denoted by  $E$  (with  $V$  rows and  $G$  columns), and type-by-gene matrix denoted by  $C$  (with  $T$  rows and  $G$  columns). The columns of both matrices corresponding to the same set of genes, ordered in the same way:

$$E(v, g) = \text{expression of gene labeled } g \text{ in voxel labeled } v, \quad (2)$$

$$C(t, g) = \text{expression of gene labeled } g \text{ in cell type labeled } t. \quad (3)$$

The computational analysis of these two data sets was undertaken using the Brain Gene Expression Atlas MATLAB toolbox (BGEA) we developed in MATLAB [13].

The present analysis is focused on 4,104 genes for which experiments conducted using both sagittal and coronal sections were available at the time of analysis. To minimize reproducibility issues, we computed the correlation coefficients between sagittal and coronal data volumes (see [14] for a plot of values) and selected the genes in the top-three quartiles of correlation (3,041 genes) for further analysis.

### 2.1 Correlations between cell-type-specific microarray data and the Allen Atlas

For a fixed cell type labeled  $t$ , the  $t$ -th row of the type-by-gene matrix  $C$  defined in equation 3 corresponds to a vector in a  $G$ -dimensional gene space. On the other hand, for a fixed voxel labeled  $v$  in the mouse brain, the  $v$ -th row of the the voxel-by-gene matrix  $E$  defined in equation 2 gives rise to another vector in the same gene space. We computed the correlation  $\text{Corr}(v, t)$  between voxel labeled  $v$  and cell type labeled  $t$ :

$$\text{Corr}(v, t) = \frac{\sum_{g=1}^G (C(t, g) - \bar{C}(g))(E(v, g) - \bar{E}(g))}{\sqrt{\sum_{g=1}^G (C(t, g) - \bar{C}(g))^2} \sqrt{\sum_{g=1}^G (E(v, g) - \bar{E}(g))^2}}, \quad (4)$$

$$\bar{C}(g) = \frac{1}{T} \sum_{t=1}^T C(t, g), \quad (5)$$

$$\bar{E}(g) = \frac{1}{V} \sum_{v=1}^V E(v, g). \quad (6)$$

For a cell type labeled  $t$ , the correlation values defined across all voxels in the ABA, as defined in Equations 4, 5 and 6 give rise to a brain-wide *correlation profile* between this cell type and the Allen Atlas, with all genes taken into account, since the genes are summed over in Equation 4. The correlation at each voxel reflects a measure of similarity between the expression profile of the cell type  $t$  and the gene expression profile defined within a local cube of side 200 microns in the ABA. See Tables 40–50 for maximum-intensity projection images and images of individual sections of these volumetric profiles for each of the  $T = 64$  cell types in this study. The next section contains an analysis of the brain-wide correlation profiles, based on the average correlation across brain regions defined by the ARA.

## 2.2 A linear model relating cell-type-specific microarray data to the Allen Atlas

The above-defined analysis of correlations between the ABA and cell-type-specific data is not a decomposition of the signal into cell-type-specific components. In this section, we propose a linear model to attempt such a decomposition, using the cell-type-specific samples as a base of gene space.

Let us denote by  $C_t$  the vector in gene space obtained by taking the  $t$ -th row of the type-by-gene matrix  $C$  defined in equation 3:

$$C_t(g) = C(t, g), \quad 1 \leq g \leq G. \quad (7)$$

To decompose the gene expression at a voxel of the mouse brain into its cell-type-specific components, let us introduce the positive quantity  $\rho_t(v)$  denoting the contribution of cell-type  $t$  at voxel  $v$ , and propose the following linear model:

$$E(v, g) = \sum_{t=1}^T \rho_t(v) C_t(g) + \text{Residual}(v, g). \quad (8)$$

Both sides are estimators of the amount of mRNA for gene  $g$  at voxel  $v$ . The residual term in Equation 8 reflects the fact that  $T = 64$  cell types are not sufficient to sample the whole diversity of cell types in the mouse brain, as well as noise in the measurements, reproducibility issues, and non-linearities in the relations between numbers of mRNAs, expression energies and microarray data [26].

To find the parameters  $\rho$  that provide the best fit of the model 8, we have to minimize the residual term by solving the following problem:

$$(\rho_t(v))_{1 \leq t \leq T, 1 \leq v \leq V} = \operatorname{argmin}_{\phi \in \mathbf{R}_+(T, V)} \mathcal{E}_{E, C}(\phi), \quad (9)$$

where

$$\mathcal{E}_{E,C}(\phi) = \sum_{v=1}^V \left( \sum_{g=1}^G \left( E(v, g) - \sum_{t=1}^T \phi(t, v) C_t(g) \right)^2 \right). \quad (10)$$

The right-hand side of Equation 10 is a sum of positive quadratic functions of  $\phi(\cdot, v)$ , one per voxel  $v$ . The minimization problem 9 can therefore be solved voxel by voxel. For each voxel  $v$  we have to minimize a quadratic function of a vector with  $T$  positive components:

$$\forall v \in [1..V], \quad (\rho_t(v))_{1 \leq t \leq T} = \operatorname{argmin}_{\nu \in \mathbf{R}_+^T} \sum_{g=1}^G \left( E(v, g) - \sum_{t=1}^T \nu(t) C_t(g) \right)^2. \quad (11)$$

We solved these quadratic optimization problems (one per voxel), using the CVX toolbox for convex optimization [27, 28].

For each cell type  $t$ , the coefficients  $(\rho_t(v))_{1 \leq v \leq V}$  yield an estimated brain-wide density profile for this cell type. See Tables 51–61 for maximum-intensity projection images and individual image sections of these density profiles for all the cell types in this study. The next section contains an analysis of results related to the anatomical brain regions defined by the ARA.

## 3 Results

### 3.1 Rankings of brain regions induced by correlation and density profiles

To analyze the neuroanatomical properties of the results, we made use of two non-hierarchical systems of annotation, available for the left hemisphere at a resolution of 200 microns. We will refer to them as:

- the 'big12' annotation, consisting of 12 regions of the left hemisphere (together with a more patchy group of voxels termed 'basic cell groups of regions') whose names, sizes and shapes are shown in Table 1.
- the 'fine' annotation, a refinement of 'big12' into 94 regions, reflecting structures further down the ARA hierarchy.

For each cell-type-specific sample in this study, anatomical metadata specify the brain region from which the sample was extracted (see Tables 64 and 65). For each brain-wide quantity computed (i.e., correlation and density profiles, defined in Equations 4 and 8), we can determine how these quantities vary across brain regions. Specifically, these profiles can be used to produce *rankings* of brain regions in the Allen Reference Atlas.


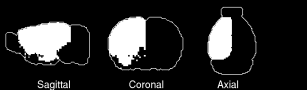

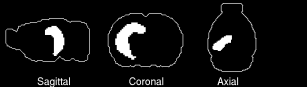



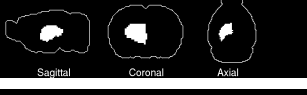
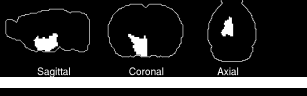
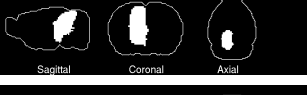
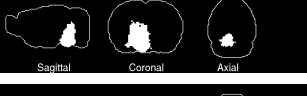

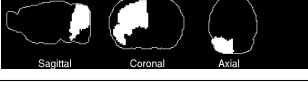
Brain region	Symbol	Percentage of hemisphere	Profile of region (maximal-intensity projection)
Basic cell groups and regions	Brain	4.6	
Cerebral cortex	CTX	29.5	
Olfactory areas	OLF	9.2	
Hippocampal region	HIP	4.3	
Retrohippocampal region	RHP	4	
Striatum	STR	8.6	
Pallidum	PAL	1.9	
Thalamus	TH	4.3	
Hypothalamus	HY	3.5	
Midbrain	MB	7.8	
Pons	P	4.6	
Medulla	MY	6.2	
Cerebellum	CB	11.5	

Table 1: Brain regions in the coarsest annotation of the left hemisphere in the ARA (referred to as the 'big12' annotation).

We have to choose a non-hierarchical partition of the brain according to the ARA, with  $R$  regions, available in a digitized form, co-registered with the voxel-based gene-expression energies (for definiteness and ease of presentation we consider the 'big12' annotation of the ARA, Table 1). Let  $V_r$  denote the set of voxels belonging to region labeled  $r$  in this partition.

### 3.1.1 Ranking brain regions by correlation profile

For a cell type labeled  $t$ , we can compute the average correlation with the Allen Atlas in each region (labeled  $r$ ) of the ARA:

$$\overline{\text{Corr}}(r, t) = \frac{1}{|V_r|} \sum_{v \in V_r} \text{Corr}(v, t). \quad (12)$$

For each cell type, the brain regions in the ARA can be ranked by decreasing values of the average correlation  $\overline{\text{Corr}}(\cdot, t)$  defined in Equation 12. The region with highest average correlation is called the *top region by correlation* for the cell type labeled  $t$ . A bar diagram of the average correlations between granule cells (index  $t=20$ , illustrated in Figure 2(a) of the main text), and the regions of the 'big12' annotation, is shown on Figure 1 (the symbols of the brain regions can be found in Table 1). The top region by correlation for granule cells is the cerebellum.

In the figures derived from the brain-wide correlation profiles (Tables 40–50), the maximal-intensity projections of each correlation profile are supplemented by a section through the top region by correlation.

### 3.1.2 Ranking brain regions by density profile

For a cell type labeled  $t$ , we can compute the fraction of the total brain-wide density density contributed by voxels of each region labeled  $r$  in the ARA:

$$\bar{\rho}(r, t) = \frac{1}{|\sum_{v \in \text{Brain Annotation}} \rho_t(v)|} \sum_{v \in V_r} \rho_t(v), \quad (13)$$

where **Brain Annotation** is the set of voxels included in the annotation (for the 'big12' annotation this set consists of the left hemisphere, as can be seen from the projections in Table 1). The brain regions in the ARA are ranked according to the fractions of density defined in Equation 18. For each cell type  $t$ , the fractional densities supported by the brain regions sum to 1:

$$\forall t \in [1..T], \quad \sum_{r=1}^R \bar{\rho}(r, t) = 1. \quad (14)$$

The region with highest fraction of density for a given cell type is called the *top region by density* for this type. A bar diagram of the fractions of the density profile for granule cells (cell type



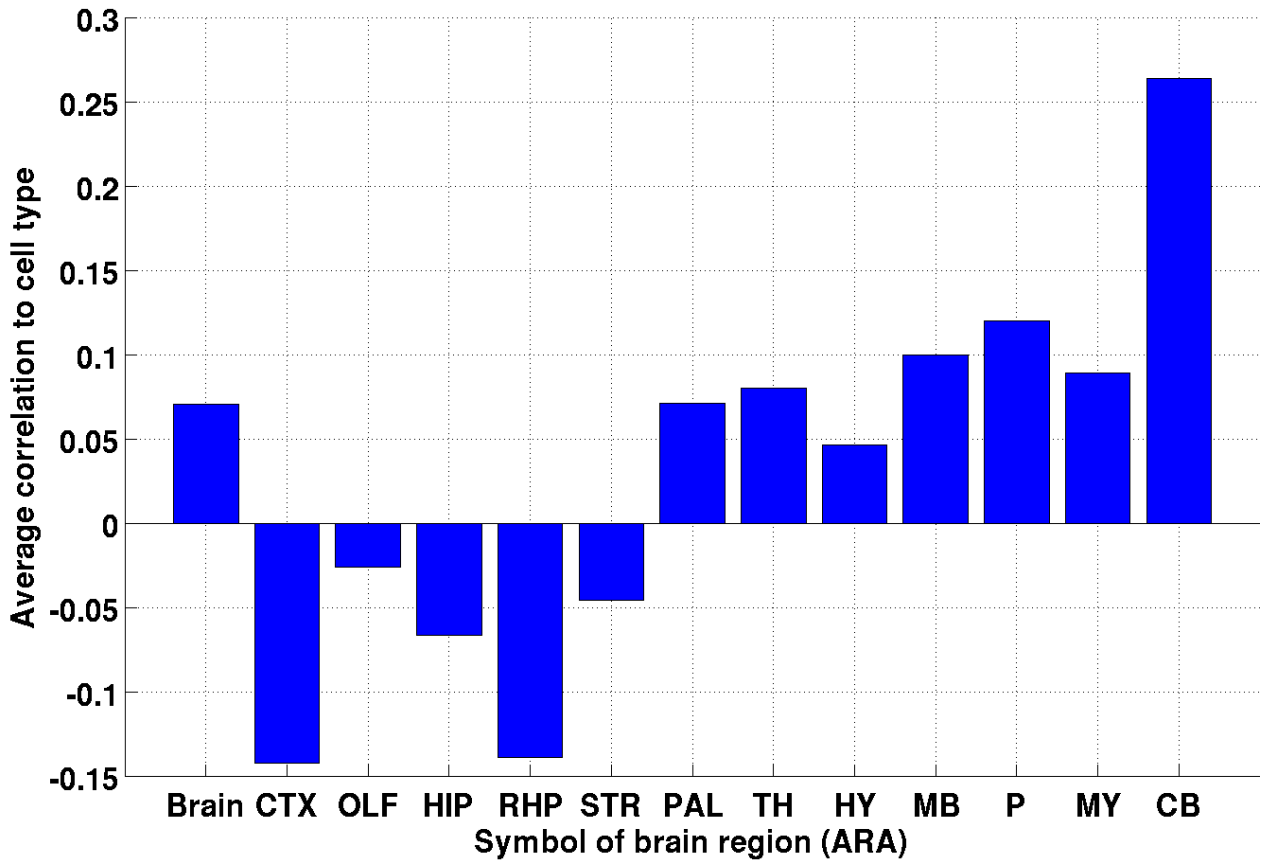


Figure 1: Bar diagram of the average correlations between granule cells and the voxels in the regions of the 'big12' annotation of the ARA, as defined in Equation 12 (granule cells correspond to the value  $t = 20$  for the cell-type index). The symbols of the regions in the 'big12' annotation read as in Table 1: Basic cell groups and regions = Brain, Cerebral cortex = CTX, Olfactory areas = OLF, Hippocampal region = HIP, Retrohippocampal region = RHP, Striatum = STR, Pallidum = PAL, Thalamus = TH, Hypothalamus = HY, Midbrain = MB, Pons = P, Medulla = MY, Cerebellum = CB.

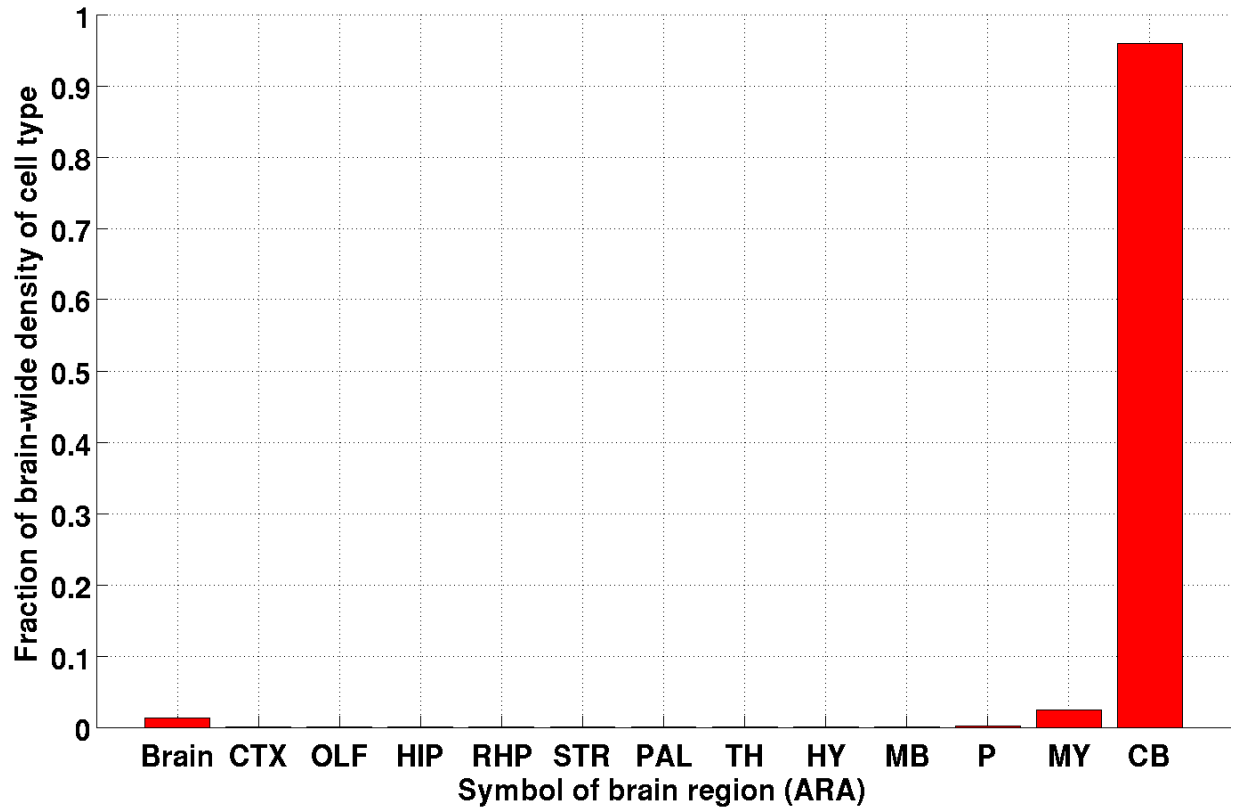


Figure 2: Fractions of density of granule cells in the regions of the 'big12' annotation of the ARA, as defined in Equation , for  $t = 20$ .

index  $t=20$ ), is shown on Figure 2. The top region by density for granule cells is the cerebellum.

In the figures derived from the brain-wide correlation density profiles (Tables 51–61), the maximal-intensity projections of each density profile are supplemented by a section through the top region by density.

### 3.2 Anatomical data for cell-type-specific samples

For each of the  $T = 64$  cell-type-specific samples in this study, anatomical metadata indicate the brain region from which the sample was extracted (see Tables 64 and 65). Eight of the twelve regions of the coarsest version of the Allen Reference Atlas (the 'big12' annotation) are represented in this data set (see Table 2), with an over-representation of the cerebral cortex (41 of the 64 samples come from the cerebral cortex, whereas the region occupies  $\sim 29.5$  percent of the volume of the brain, see Table 1). The cerebellum is the only other brain region in the 'big12' annotation to be over-represented by cell-type-specific samples compared to the

Brain region	Number of cell-type-specific samples	List of sample indices in the list of 64 samples in this study
Basic cell groups and regions	0	$\emptyset$
Cerebral cortex	41	$\{2, 3, [6, \dots, 10], 14, 22, 24, 26, [29, \dots, 48], 50, [53, \dots, 56], 58, [60, \dots, 64]\}$
Olfactory areas	0	$\emptyset$
Hippocampal region	2	$\{49, 57\}$
Retrohippocampal region	0	$\emptyset$
Striatum	3	$\{13, 15, 16\}$
Pallidum	1	$\{11\}$
Thalamus	1	$\{59\}$
Hypothalamus	0	$\emptyset$
Midbrain	3	$\{4, 5, 10\}$
Pons	1	$\{51\}$
Medulla	1	$\{12\}$
Cerebellum	11	$\{1, [17, \dots, 21], 23, 25, 27, 28, 52\}$

Table 2: **Anatomical data for the cell-type-specific samples.** Eight of the regions defined by the coarsest version of the Allen Reference Atlas are represented in our data set. See Tables 64 and 65 for a more detailed account of the anatomy of the of the cell-type-specific samples.

volume of the brain it occupies. The regions that are not represented are the olfactory areas, the retrohippocampal region and the hypothalamus. The group of voxels labeled 'Basic cell groups of regions', which includes the white matter, is also unrepresented.

For each of the cell-type-specific samples, we computed the ranks of the brain regions in the ARA according to correlation and density profiles, as defined in Equations 12 and 18. It is interesting to compare the computed *top region by density* and the *top region by correlation* to the brain region from which the cell-type-specific sample was extracted (listed in Tables in S9). In the rest of this section, we group the cell-type-specific samples that were extracted from a given brain region, and compare this region to the top region by correlation and to the top region by density (except for the set of voxels called 'Basic cell groups', which is discussed first as it appears as the top region in the results for a number of cell types).

Description (index)	Origin of sample	Fraction of density in 'Basic cell groups and regions' (%)
Mature Oligodendrocytes (22)	Cerebral cortex	53.6
Astroglia (29)	Cerebral cortex	56.0
Astrocytes (30)	Cerebral cortex	19.7
Astrocytes (31)	Cerebral cortex	69.9
Astrocytes (32)	Cerebral cortex	51.4
Oligodendrocytes (36)	Cerebral cortex	56.8
Astroglia (28)	Cerebellum	74.3

Table 3: Cell-type-specific samples that have 'Basic cell groups and regions' as their top region by density.

### 3.3 Neuroanatomical patterns of results, grouped by the regions in the 'big12' annotation of the left hemisphere

#### 3.3.1 Basic cell groups and regions

The patchy group of voxels assigned the label 'Basic cell groups and regions' in the digitized version of the Allen Atlas at a resolution of 200 microns (see Table 1) is not found in the list of brain regions from which the cell-type-specific samples analyzed here were extracted (see Table 2). However, this set of voxels is the top region by correlation and/or density for several cell types, and an inspection of the maximum-intensity projection of the correlation and density profiles for these cell types reveals an anatomical pattern that resembles the brain's white matter structures, including the *arbor vitae*. As can be seen in Table 3, most of these cell types were extracted from the cerebral cortex, except the astroglia (sample index  $t=28$ , [19]), which were extracted from the cerebellum). The sum of the density profiles of these cell types is illustrated on Figure 3. All the cell types whose top region by density is 'Basic cell groups and regions' are oligodendrocytes, astroglia or astrocytes.

Figures 4 and 5 show results for a class of astrocytes [18] (cell-type index 31) extracted from the cerebral cortex. The brain-wide correlation and density profiles exhibit a pattern resembling white-matter structures, with the most caudal component corresponding to the *arbor vitae*.

#### 3.3.2 Cerebral cortex

Some of the cell-type-specific samples extracted from the cerebral cortex have a correlation and/or density profile resembling white matter (which is included within the 'Basic cell groups and regions' in the ARA). See the previous subsection, Table 3 and Figure 3 for a separate analysis of these cell types.

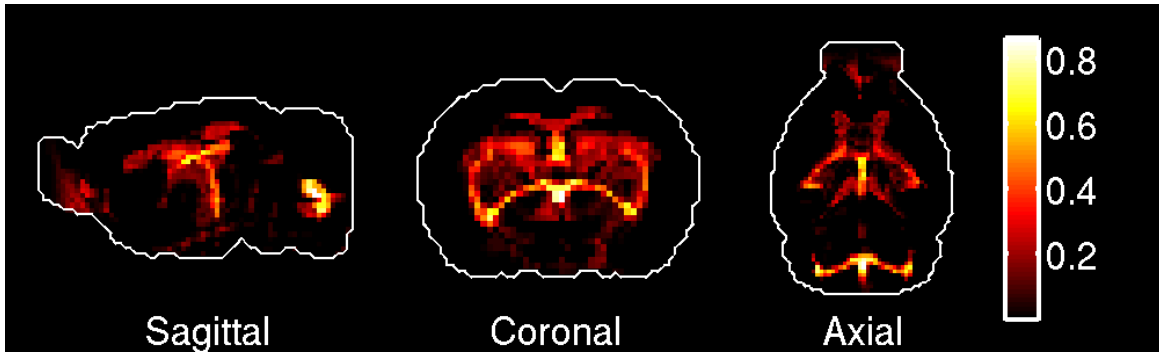
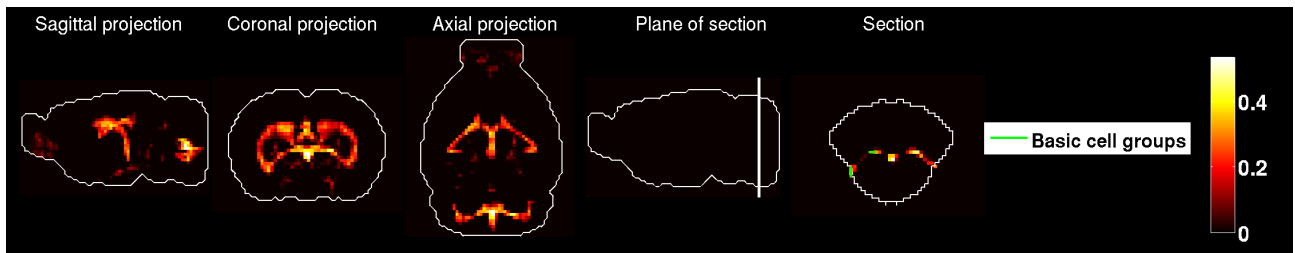


Figure 3: Maximum-intensity projection of the sum of density profiles of cell-type-specific samples that have 'Basic cell groups and regions' as their top region by density, listed in Table 3.

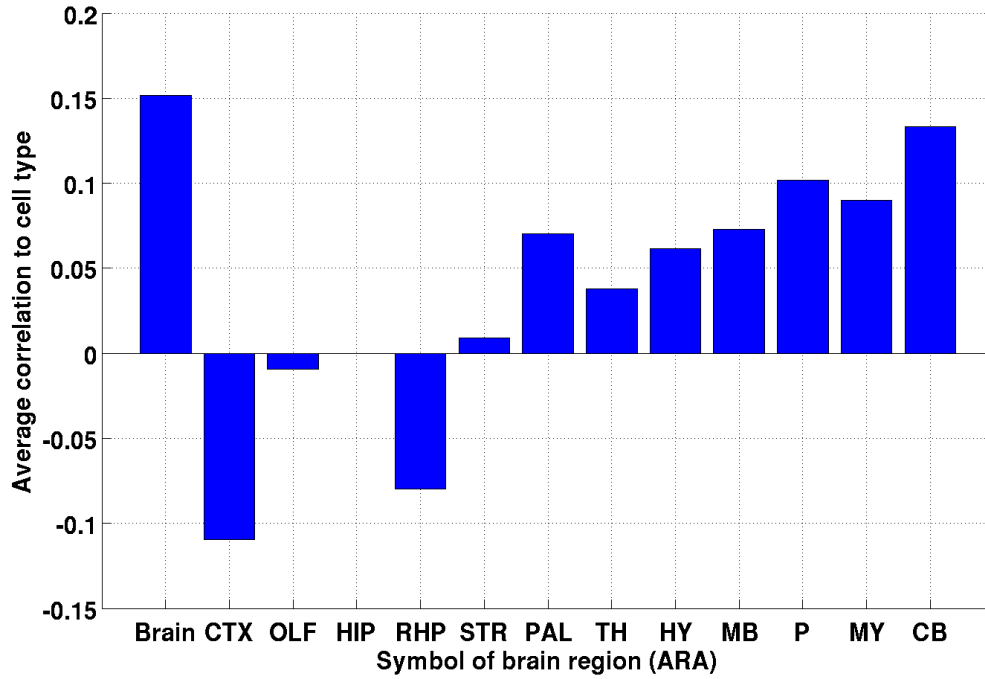


(a)

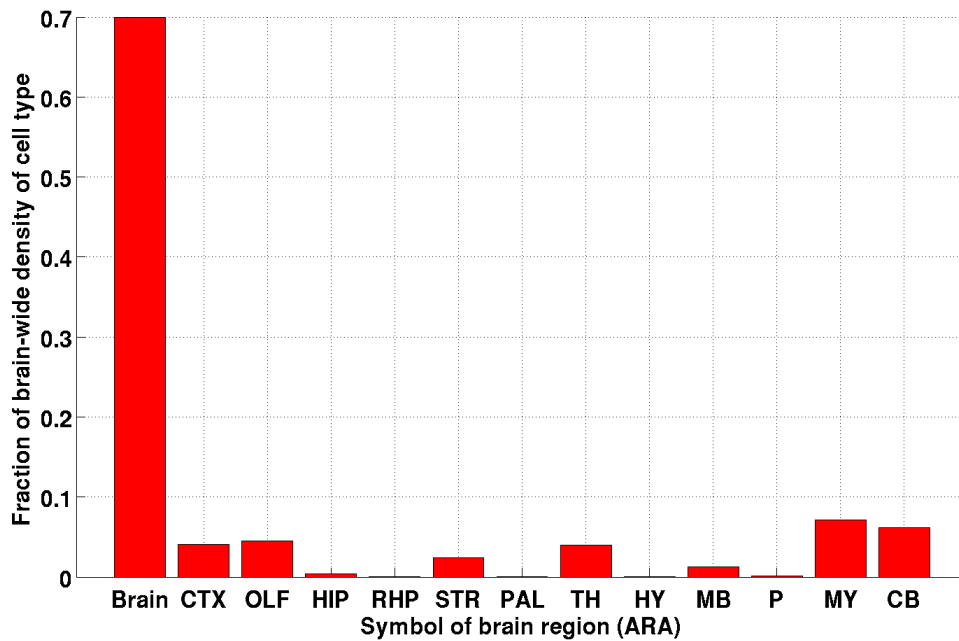


(b)

Figure 4: **Astrocytes (cell-type index 31)**. (a) Heat map of the brain-wide correlation profile. (b) Heat map of the estimated brain-wide density profile. 'Basic cell groups and regions' is the top region by correlation and by density, hence the choice of section and its legend.

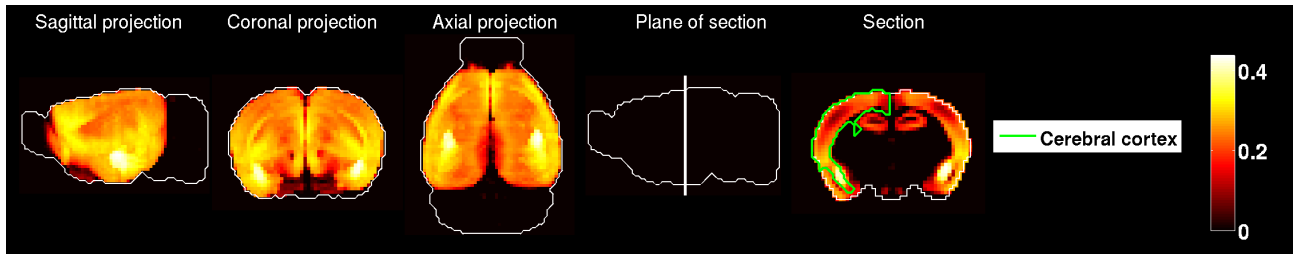


(a)

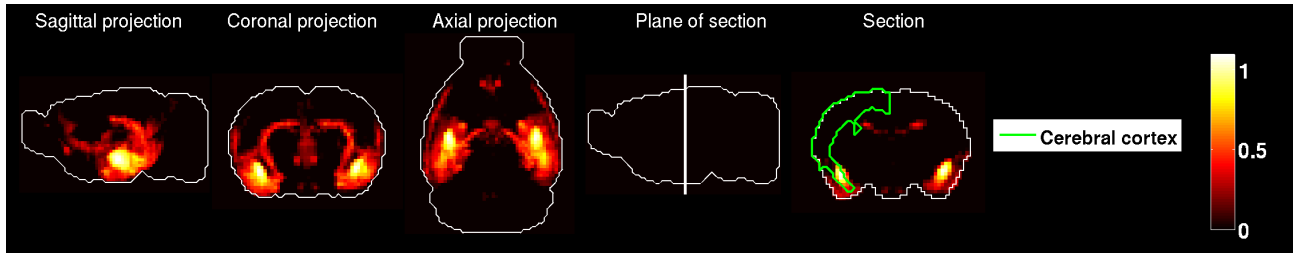


(b)

Figure 5: **Astrocytes (cell-type index 31)**. Average correlation between the cell type and the Allen Atlas, in the regions of the 'big12' annotation of the ARA. Even though the astrocytes were extracted from the cerebral cortex, their expression profile is negatively correlated on average with the voxels of the cerebral cortex.



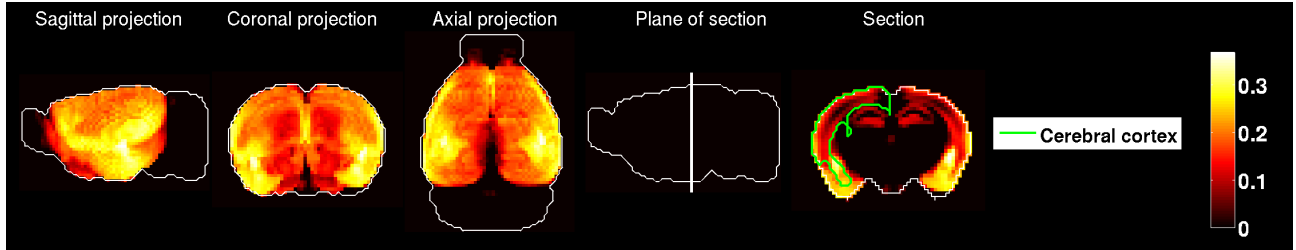
(a)



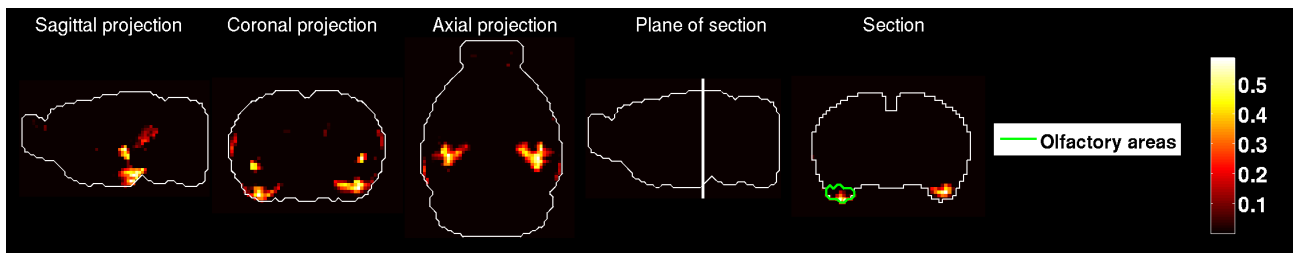
(b)

Figure 6: **Pyramidal neurons (cell-type index 48, studied in [23])**. (a) Heat map of the brain-wide correlation profile. (b) Heat map of the estimated brain-wide density profile.

neurons, index 48, and glutamatergic neurons, index 53, both studied in [23]) were extracted from the amygdala (see Tables 64–65). The amygdala is not one of the brain regions in the 'big12' annotation of the ARA. The amygdala is split between the subcortical plate (which is included in the cerebral cortex in the numerical version of the ARA at a resolution of 200 microns) and the olfactory areas [9]. A visual inspection of the correlation and density profiles for both these cell types (Figures 6 and 7) allows to detect a pattern resembling the amygdala, and indeed the cerebral cortex and the olfactory areas rank first and second by the fraction of the density profile they support (see Table 4). It is interesting to examine the fraction of the density profile supported by the various subdivisions of the olfactory areas (according to the 'fine' annotation), especially for glutamatergic neurons, for which olfactory areas support more than 64 % of the total density profile (see Table 5). The Cortical amygdalar area and Piriform-amygdalar area are among the main subregions of the olfactory areas contributing to the density profile of both cell types, which confirms the visual impression of an amygdalar pattern.



(a)



(b)

Figure 7: **Glutamatergic neurons (cell-type index 53, studied in [23])**. (a) Heat map of the brain-wide correlation profile. (b) Heat map of the estimated brain-wide density profile.

Description (index)	Fraction of density in cerebral cortex (%)	Fraction of density in olfactory areas (%)	Rank of Cerebral cortex in 'big12'	Rank of Olfactory areas in 'big12'
Pyramidal neurons (48)	38.5	29.3	1	2
Glutamatergic neurons (53)	30.2	64.9	2	1

Table 4: Cell-type-specific samples extracted from the amygdala, with fractions of their density profiles supported by the cerebral cortex and the olfactory areas, which are the two brain regions in the ARA that intersect the amygdala.



Description (index)	Subregion of Olfactory areas in the ARA ('fine' annotation)	Fraction of density in the region(%)	Fraction of Olfactory areas occupied by the region
Pyramidal neurons (48)	Piriform area	40.3	28.4
	Cortical amygdalar area	29.3	5.5
	Postpiriform transition area	15.4	2.2
	Piriform-amygdalar area	9.1	1.9
	Nucleus of the lateral olfactory tract	4.0	1.0
	Anterior olfactory nucleus	1.4	9.5
	Taenia tecta	0.5	3.2
	Main olfactory bulb	0	41.4
	Accessory olfactory bulb	0	1.6
Glutamatergic neurons (53)	Cortical amygdalar area	59.9	5.5
	Piriform-amygdalar area	19.1	1.9
	Postpiriform transition area	13.2	2.2
	Piriform area	7.4	28.4
	Main olfactory bulb	0.4	41.4
	Accessory olfactory bulb	0	1.6
	Anterior olfactory nucleus	0	9.5
	Taenia tecta	0	3.2
Nucleus of the lateral olfactory tract	0	1.0	

Table 5: Subregions of the olfactory areas ranked by the fraction of the density profile they support, for the two cell-type-specific samples extracted from the amygdala. The amygdalar regions are over-represented for both cell types, compared to the fraction of the olfactory areas they occupy.

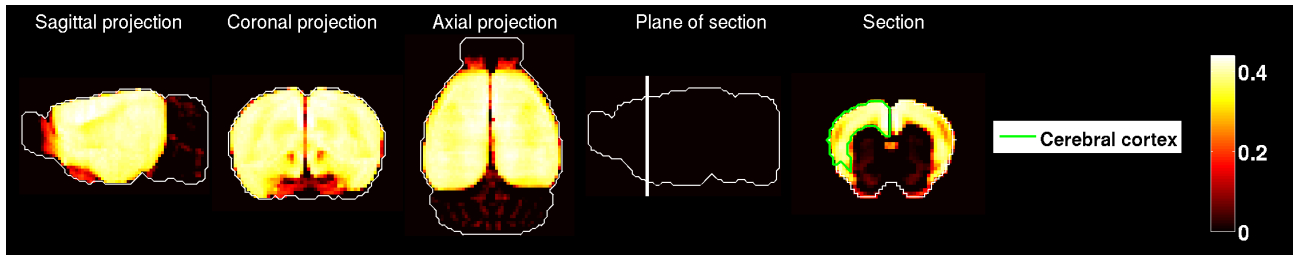
Description (index)	Fraction of density supported in cerebral cortex (%)	Next region in 'big12'	Fraction of density supported in next region (%)
Pyramidal neurons (7)	25.4	Olfactory areas	24.9
Pyramidal neurons, callosally projecting, P14 (40)	76.6	Olfactory areas	15.9
Pyramidal Neurons (45)	71.8	Retrohippocampal region	14.8
Pyramidal Neurons (46)	93.6	Retrohippocampal region	1.8
Pyramidal Neurons (47)	84.5	Retrohippocampal region	7.2

Table 6: Cell-type-specific samples extracted from the cerebral cortex, for which the cerebral cortex is ranked first both by average correlation and fraction of density profile supported.

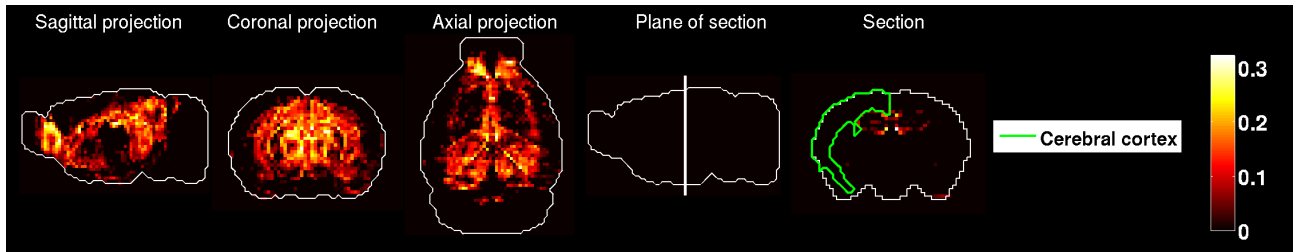
Having treated separately the cell types that correlate best or have highest density fits to the white matter (see Table 3), as well as the two amygdalar cell types, we are left with 29 cortical cell types. For 5 of these – all pyramidal neurons – the cerebral cortex is the top region in the 'big12' annotation of the ARA, both by average correlation and by fraction of the density profiles. These pyramidal neurons were all extracted from adult animals, except for the cell type labeled 40 (P14); see Tables 62,63 for age data. See Figure 8–10 for heat maps of the brain-wide correlation and density profiles for these cell types.

Some cell-type specific samples extracted from the cortex as per the anatomical data of Table 2 have the cerebral cortex as top region by correlation, but not as top region by density. There are 8 such cell types (see Table 7), of which 4 have zero density profiles in the left hemisphere. The other four cell types have Olfactory areas or Retrohippocampal region as their top region by density.

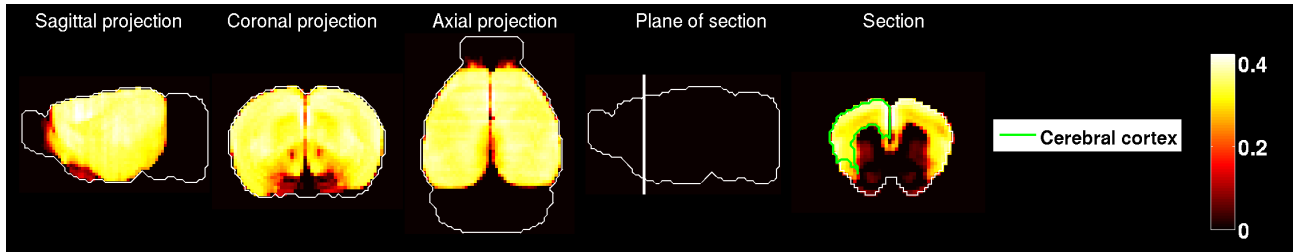
Some cortical cell-type-specific samples do not have the cerebral cortex as their top region by correlation or by density. There are 15 such cell types, 7 of which have zero density in the left hemisphere (they are not detected by the linear model, see Table 8). Six cell types have the cortex as their second ranked region by average correlation, while the top region is either the hippocampal region or the retrohippocampal region.



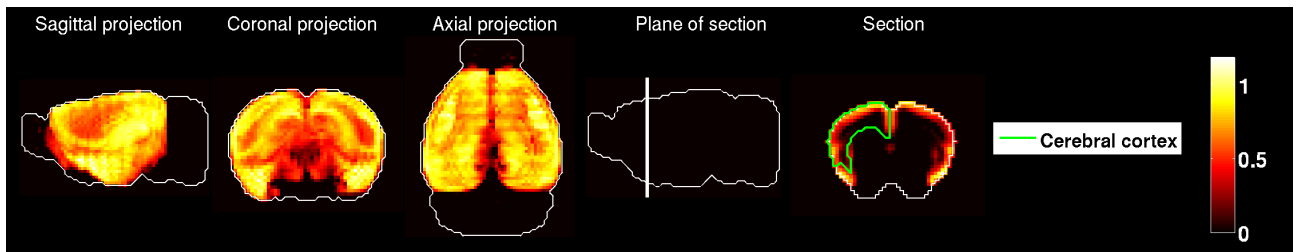
(a)



(b)

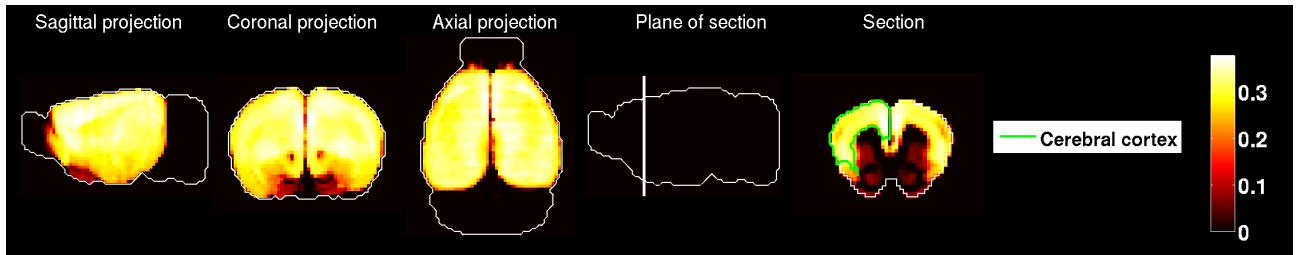


(c)

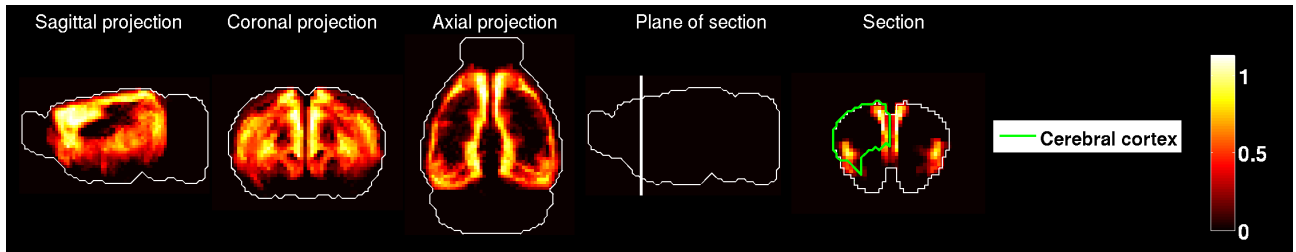


(d)

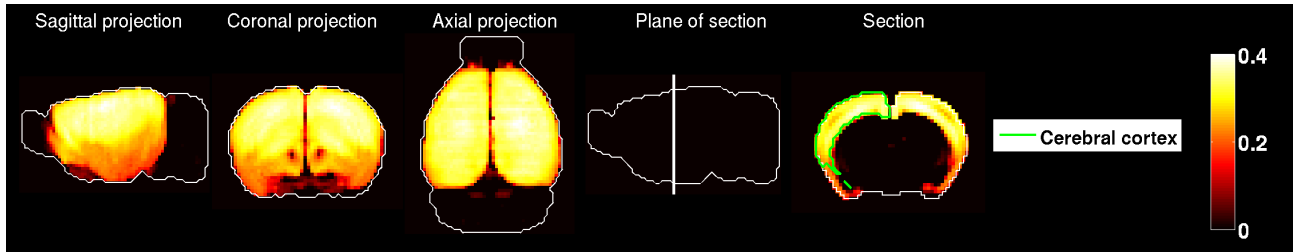
Figure 8: **Pyramidal neurons for which the cerebral cortex is the top region both by correlation and by density (I).** (a) Heat map of the brain-wide correlation profile (cell-type index 7). (b) Heat map of the estimated brain-wide density profile (cell-type index 7). (c) Heat map of the brain-wide correlation profile (cell-type index 40). (d) Heat map of the estimated brain-wide density profile (cell-type index 40).



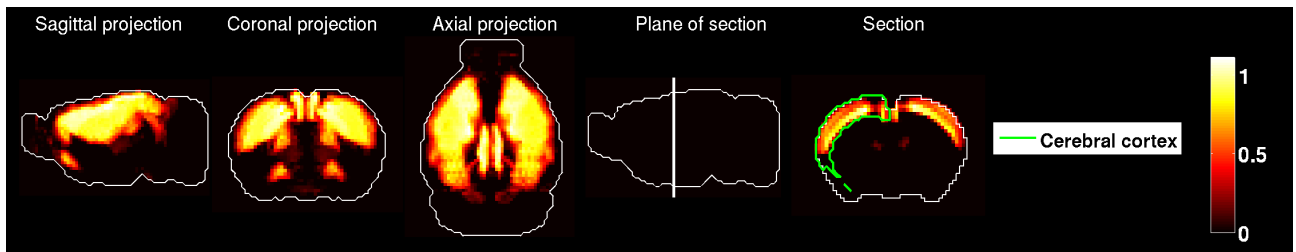
(a)



(b)



(c)



(d)

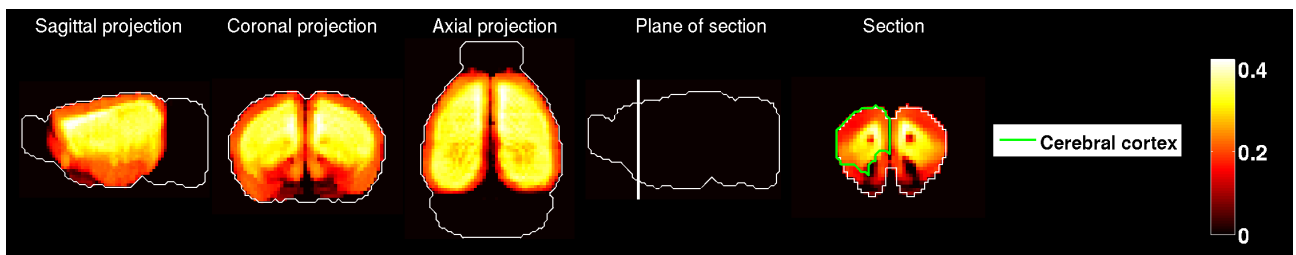
Figure 9: **Pyramidal neurons for which the cerebral cortex is the top region both by correlation and by density (II).** (a) Heat map of the brain-wide correlation profile (cell-type index 45). (b) Heat map of the estimated brain-wide density profile (cell-type index 45). (c) Heat map of the brain-wide correlation profile (cell-type index 46). (d) Heat map of the estimated brain-wide density profile (cell-type index 46).

Description (index)	Top region by density (percentage of density supported)	Rank of the cerebral cortex out of 13 regions	Fraction of density in cerebral cortex (%)
Pyramidal Neurons (2)	Olfactory areas (100)	2 (and last)	0
Pyramidal Neurons (8)	Olfactory areas (100)	2 (and last)	0
Mixed Neurons (9)	Cerebellum (100)	2 (and last)	0
Interneurons (14)	Olfactory areas (100)	2 (and last)	0
Neurons (26)	Olfactory areas (100)	2 (and last)	0
Pyramidal Neurons, Corticospinal, P14 (43)	Olfactory areas (72.4)	2	11.5
Pyramidal Neurons, Corticotectal, P14 (44)	Olfactory areas (33.0)	4	9.4
Pyramidal Neurons (50)	Retrohippocampal region (57.6)	3	10.7
GABAergic Interneurons, VIP+ (55)	Olfactory areas (96.7)	2	1.78

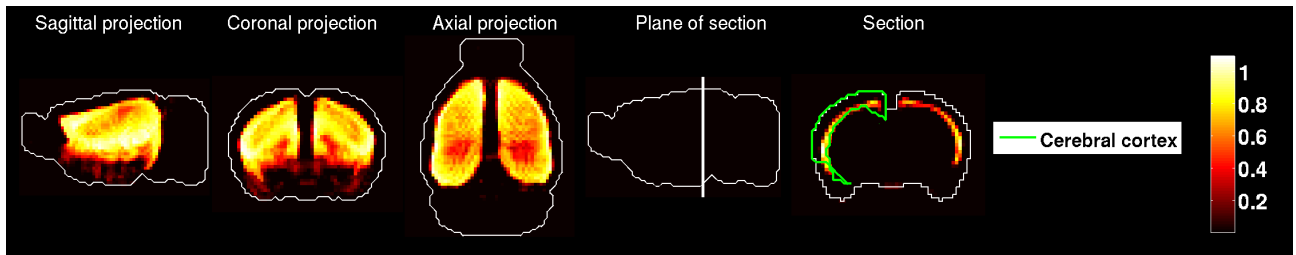
Table 7: Cell-type-specific samples extracted from the cerebral cortex, for which the cerebral cortex is ranked first by average correlation but not by the density.

Description (index)	Top region by correlation	Rank of cerebral cortex by correlation	Top region by density (percentage of density supported)	Rank of the cerebral cortex by density out of 13 regions	Fraction of density in cerebral cortex (%)
Pyramidal Neurons (3)	Hippocampal region	2	Olfactory areas (100)	2 (and last)	0
Mixed Neurons (33)	Retrohippocampal region	2	Olfactory areas (58.6)	3	4.8
Oligodendrocyte Precursors (37)	Hypothalamus	13	Hypothalamus (33.8)	7	2.6
Pyramidal Neurons, Callosally projecting, P3 (38)	Olfactory areas	6	Olfactory areas (100)	2 (and last)	0
Pyramidal Neurons, Callosally projecting, P6 (39)	Retrohippocampal region	2	Olfactory areas (99)	4 (and last)	0
Pyramidal Neurons, Corticospinal, P3 (41)	Olfactory areas	5	Olfactory areas (100)	2 (and last)	0
Pyramidal Neurons, Corticospinal, P6 (42)	Retrohippocampal region	2	Olfactory areas (86.6)	2	7.8
GABAergic Interneurons, VIP+ (54)	Retrohippocampal region	2	Striatum (36.6)	5	6.5
GABAergic Interneurons, SST+ (55)	Hypothalamus	10	Striatum (30.1)	8	2.4
GABAergic Interneurons, PV+ (58)	Medulla	7	Olfactory areas (93.4)	4 (and last)	0
GABAergic Interneurons, PV+, P7 (60)	Pallidum	11	Olfactory areas (99.2)	2	0.7
GABAergic Interneurons, PV+, P10 (61)	Pallidum	8	Olfactory areas (100)	2 (and last)	0
GABAergic Interneurons, PV+, P13-P15 (62)	Retrohippocampal region	2	Olfactory areas (100)	2 (and last)	0
GABAergic Interneurons, PV+, P25 (63)	Medulla	5	Olfactory areas (56.5)	6	2.1
GABAergic Interneurons, PV+ (64)	Medulla	5	Midbrain (28.1)	10	1.5

Table 8: Cell-type-specific samples extracted from the cerebral cortex, for which the cerebral cortex is ranked first neither by average correlation, nor by the fraction of density profile it supports.



(a)



(b)

Figure 10: **Pyramidal neurons for which the cerebral cortex is the top region both by correlation and by density (III).** (a) Heat map of the brain-wide correlation profile (cell-type index 47). (b) Heat map of the estimated brain-wide density profile (cell-type index 47).

### 3.3.3 Hippocampal region

Two cell-type-specific samples that were extracted from the hippocampus (see Table 2). For one of them pyramidal neurons (index 49, studied in [23]), the hippocampal region is the top region both by correlation and by density. We ranked the regions of the 'fine' annotation in the ARA by their contribution to the density profile of this sample (this ranking corresponds to Equation 18, with the region label  $r$  running over the  $R = 94$  regions in the 'fine' annotation). The first two regions are Ammon's horn (which contributes 48.8 percent of the total density of this cell type), followed by the dentate gyrus (which contributes 25.4 percent of the total density of this cell type). These two regions are the two subregions of the hippocampal region in the 'fine' annotation. Moreover, the data of [23] specify that the sample labeled 49 was taken from Ammon's horn, which indicates that the brain-wide correlation and density profiles are both consistent with prior biological knowledge for this cell type.

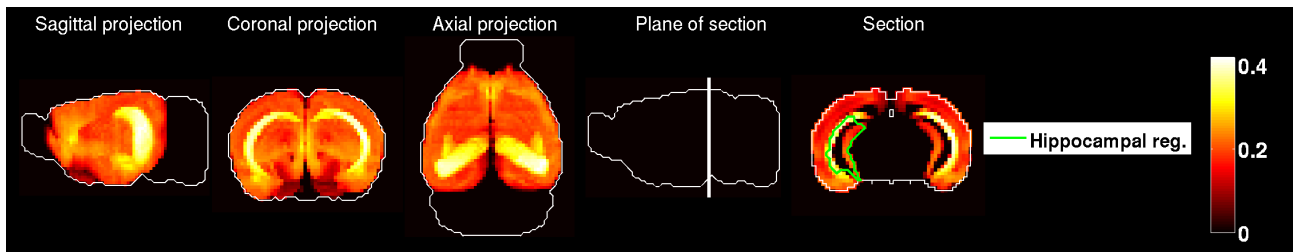
This sample (index 49) is the only one for which the hippocampus is ranked first by density. Another sample (index 3), is ranked first by correlation. This sample was not extracted from the hippocampus, but rather from the cerebral cortex (primary somatosensory area, layer 5). On the other hand, the cerebral cortex is ranked second by its fraction of density defined in Equation 18. This sample has an estimated density profile with values close to zero, except in a few voxels belonging to the olfactory bulb.

The second sample extracted from the hippocampus consists of somatostatin-positive GABAergic interneurons (index 57). The hippocampal region is ranked last by correlations and next-to-last by densities, whereas the first region by average correlation is hypothalamus, and the first region by density is midbrain (see Figure 12 and 14). Visual inspection of the Tables 49,50,60,61 shows that there is a lot of solidarity between the correlation profiles of the cell types labeled with indices between 54 and 64, which are all GABAergic interneurons. For these cell types, with the exception of the cell type labeled 55, whose top region by correlation is the cerebral cortex, the correlations are higher in regions of the brain that are more ventral than the region from which the samples where extracted.

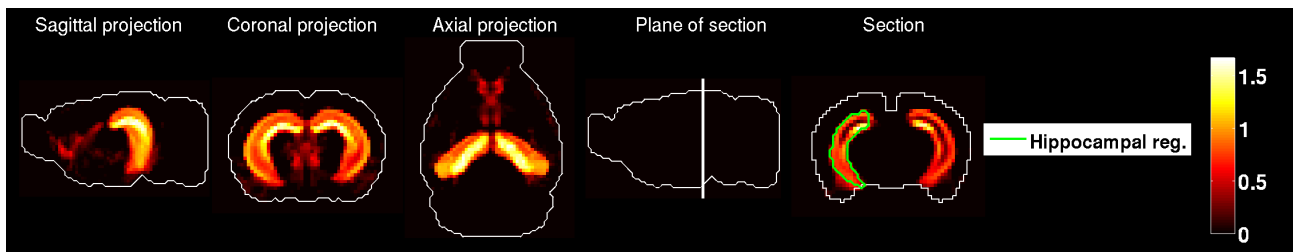


Description (index)	Origin of sample	Rank of region (out of 94) in the 'fine' annotation (by correlation)	Rank of region (out of 94) in the 'fine' annotation (by density)	Fraction of density in the region	Fraction of density supported in the hippocampal region
Pyramidal neurons (49)	Ammon's horn	3	1	48.7%	71.5%
GABAergic interneurons, SST+ (57)	Ammon's horn	91	55	0.1%	0.1%

Table 9: Anatomical analysis for the cell-type-specific samples extracted from the hippocampal region.

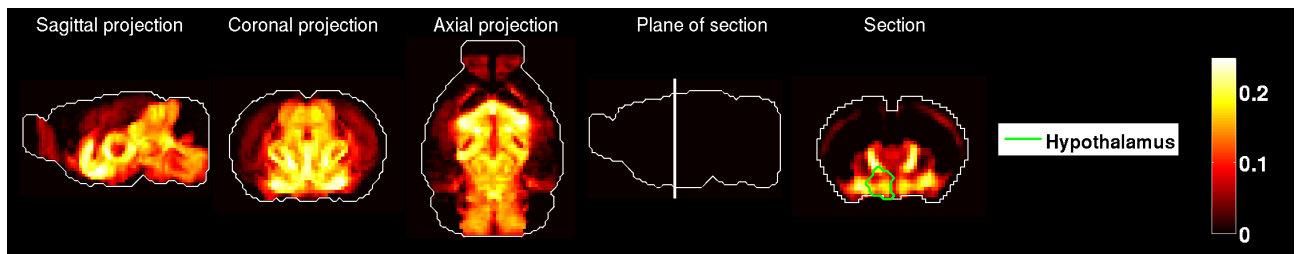


(a)

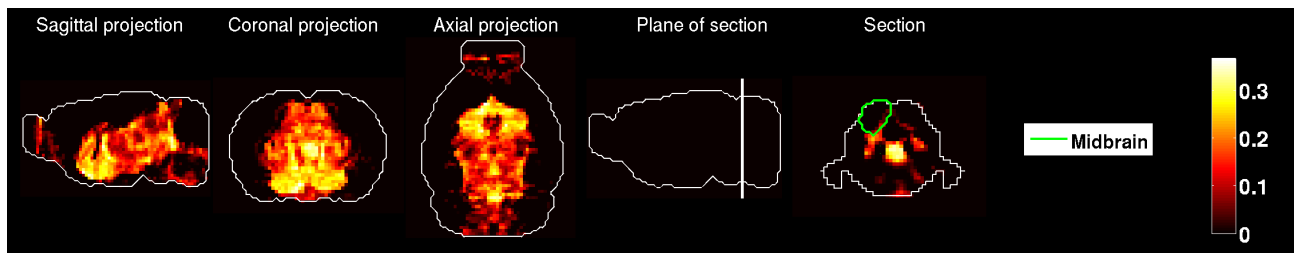


(b)

Figure 11: **Pyramidal neurons (cell-type index 49)**. (a) Heat map of the brain-wide correlation profile. (b) Heat map of the estimated brain-wide density profile.

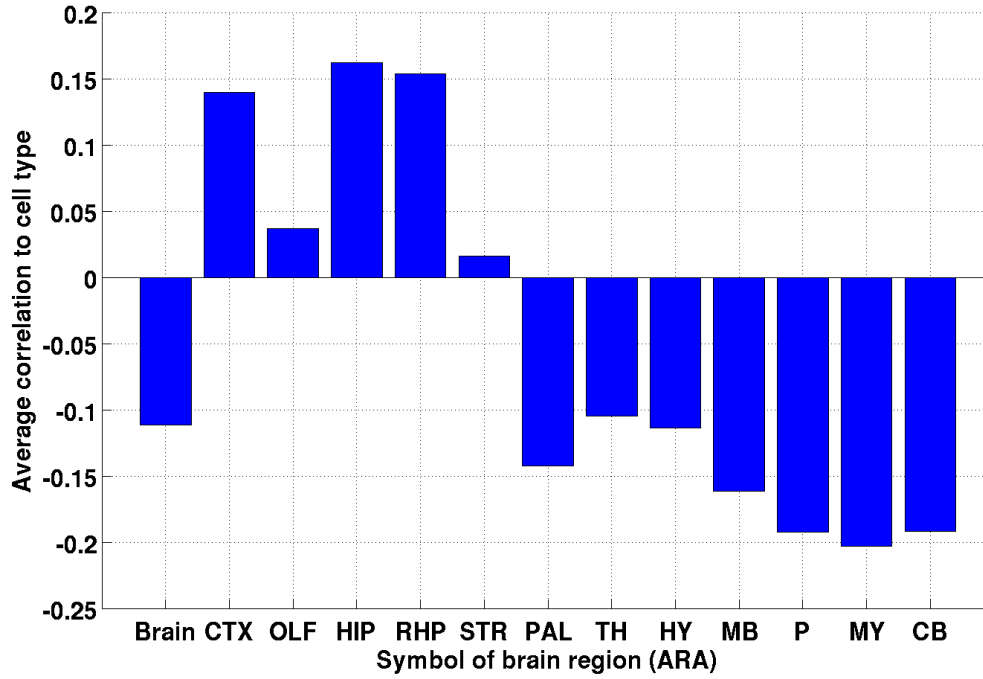


(a)

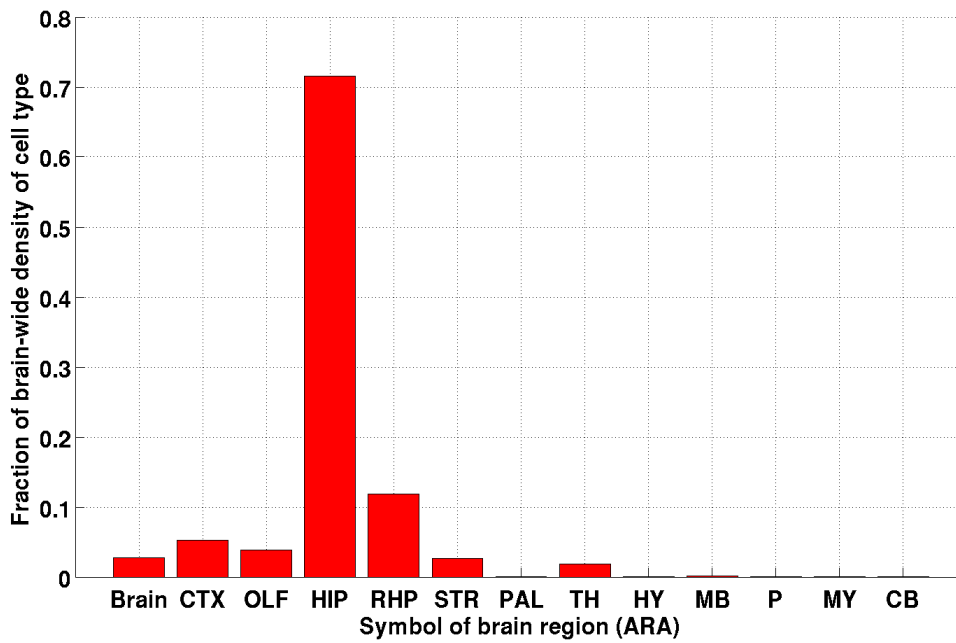


(b)

Figure 12: **GABAergic interneurons, SST+ (cell-type index 57)**. (a) Heat map of the brain-wide correlation profile. (b) Heat map of the estimated brain-wide density profile.

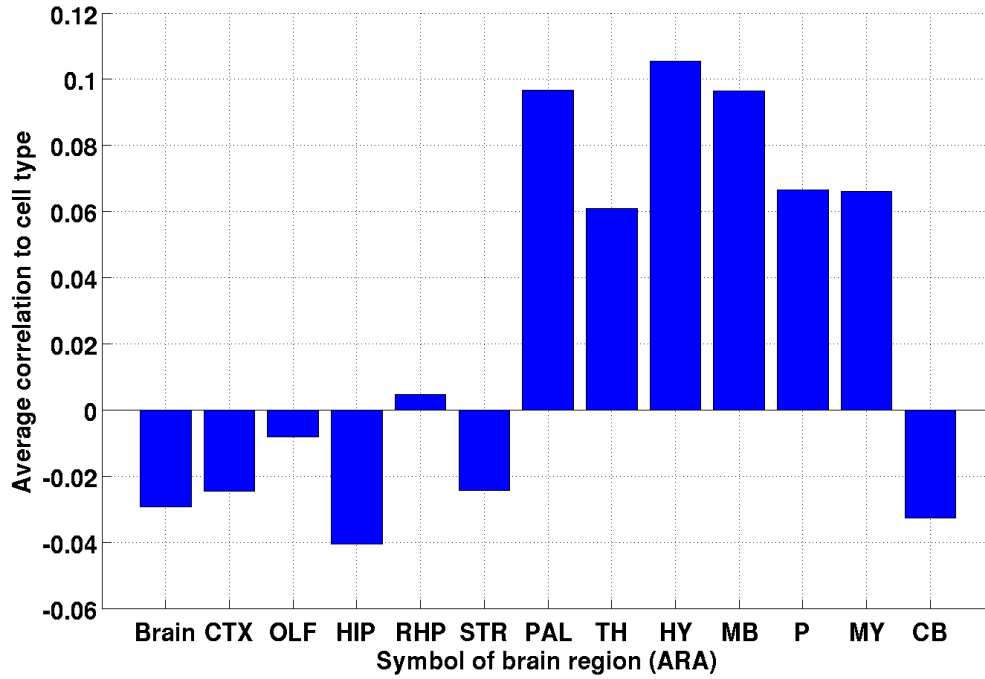


(a)

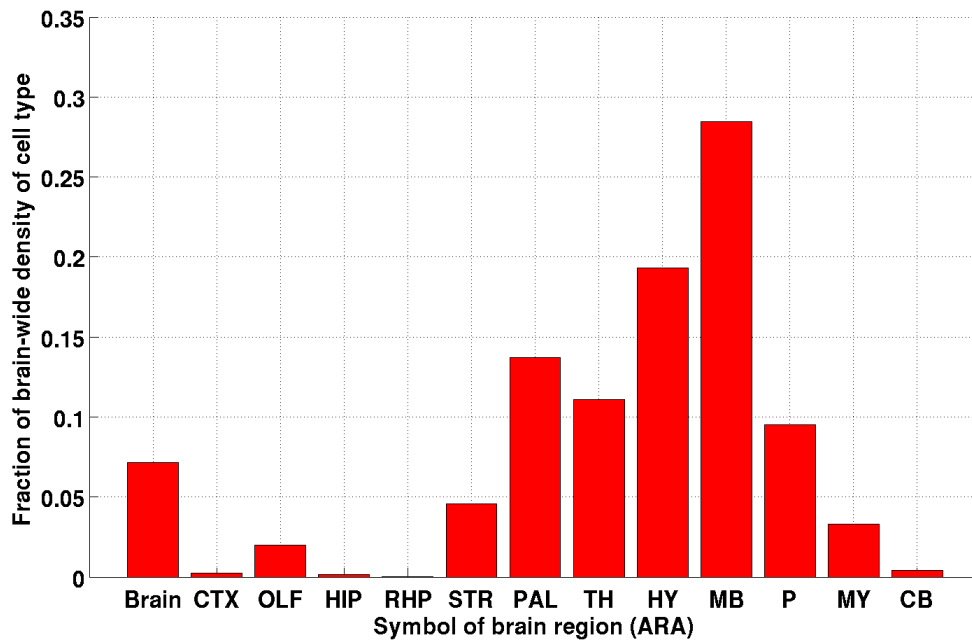


(b)

Figure 13: **Pyramidal neurons (cell-type index 49)**. Average correlation between the cell type and the Allen Atlas, in the regions of the 'big12' annotation of the ARA. The estimated density of this cell type is zero in the left hemisphere.



(a)



(b)

Figure 14: **GABAergic interneurons, SST+ (cell-type index 57)**. Average correlation between the cell type and the Allen Atlas, in the regions of the 'big12' annotation of the ARA.

Description (index)	Origin of sample	Rank of region (out of 12) in the 'big12' annotation (by correlation)	Rank of region (out of 12) in the 'big12' annotation (by density)	Fraction of density supported in the striatum
Cholinergic neurons (13)	Striatum	7	6	3.3%
<i>Drd1</i> + medium spiny neurons (15)	Striatum	1	2	30%
<i>Drd2</i> + medium spiny neurons (16)	Striatum	1	1	91.4%

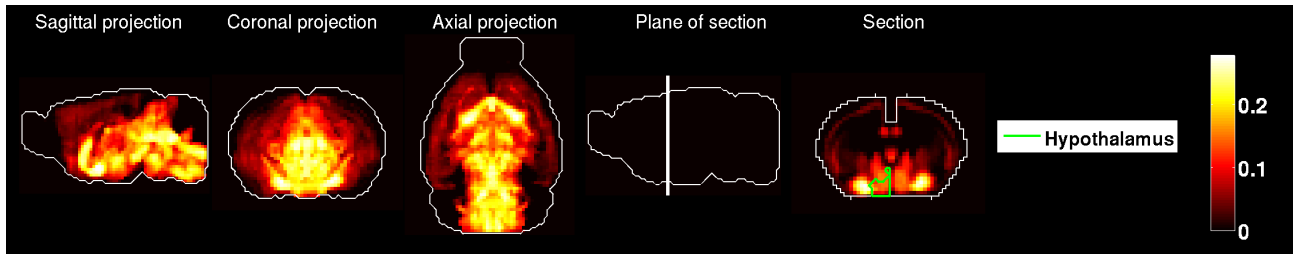
Table 10: Anatomical analysis for the cell-type-specific samples extracted from the striatum.

### 3.3.4 Striatum

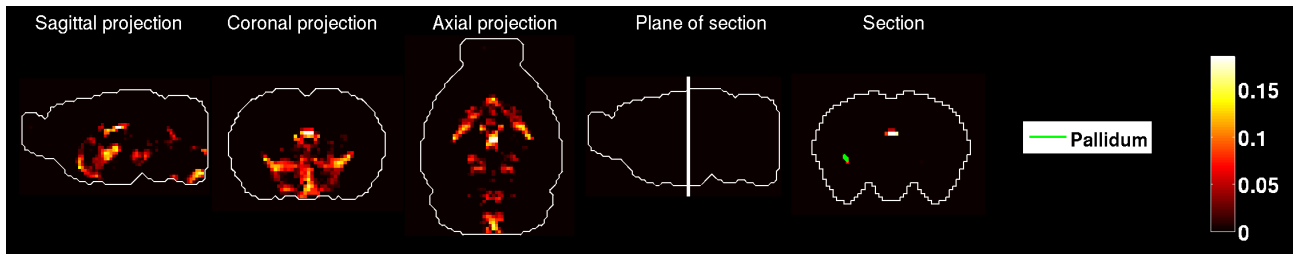
Three cell-type-specific samples (cholinergic neurons – index 13, *Drd1*+ medium spiny neurons – index 15, and *Drd2*+ medium spiny neurons – index 16) in the data set were extracted from the striatum.

The pallidum is the region in the 'big12' annotation that supports the highest fraction (43%) of the density of cholinergic neurons (index 13), and the striatum ranks only 7th by correlation and 6th by density (see Table 10 Figures 15 and 18).

On the other hand, medium spiny neurons [19] expressing both dopamine receptor types have striatum as their top region by correlations (see Figures 16, 17, 19, and 20). The striatum is the top region by density for *Drd2*+ medium spiny neurons, and the second region by density for *Drd1*+ medium spiny neurons, after the cerebral cortex.

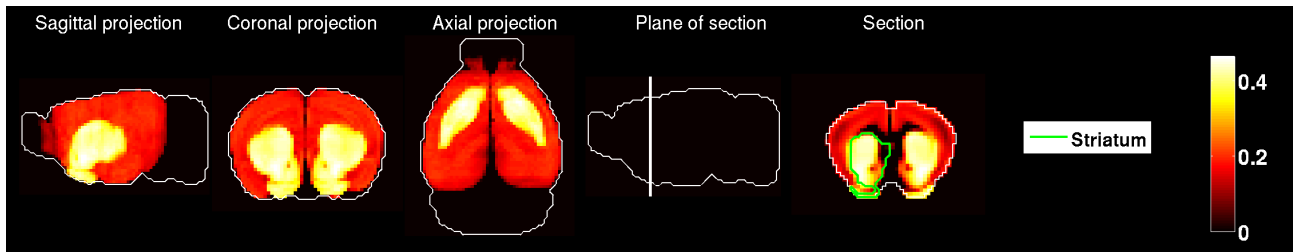


(a)

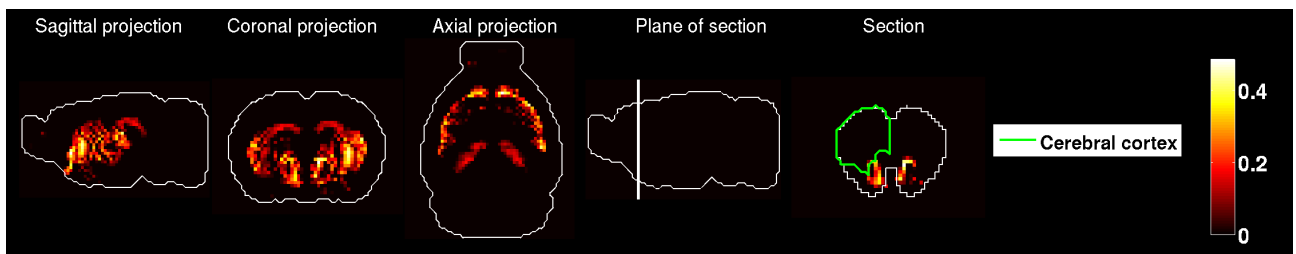


(b)

Figure 15: **Cholinergic neurons (cell-type index 13)**. (a) Heat map of the brain-wide correlation profile. (b) Heat map of the estimated brain-wide density profile.

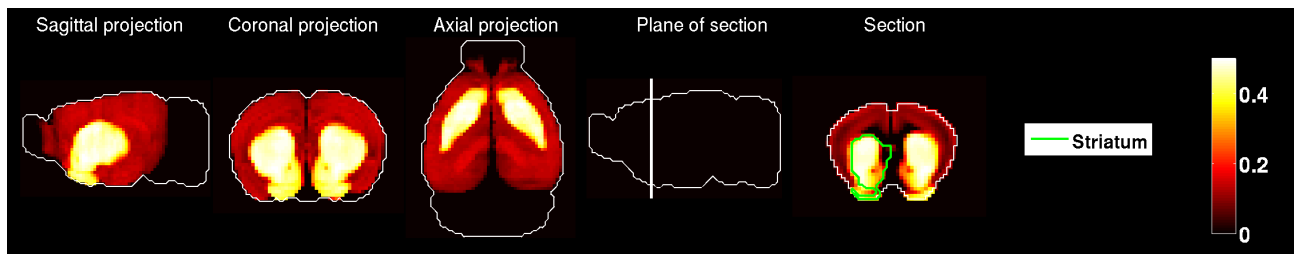


(a)

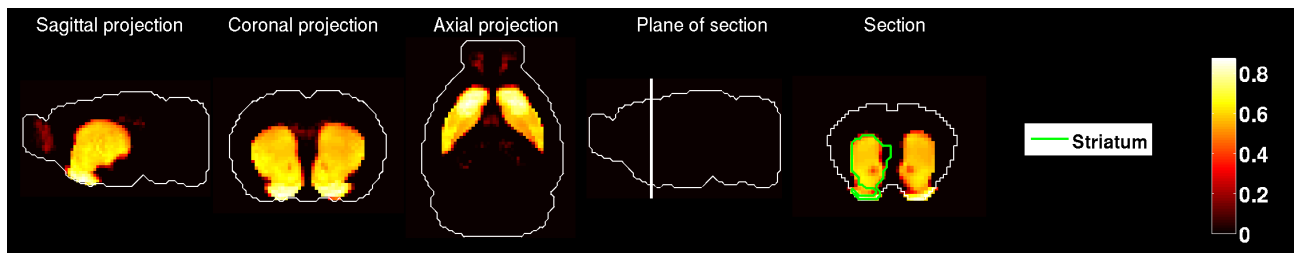


(b)

Figure 16: ***Drd1*+ medium spiny neurons (cell-type index 15)**. (a) Heat map of the brain-wide correlation profile. (b) Heat map of the estimated brain-wide density profile.

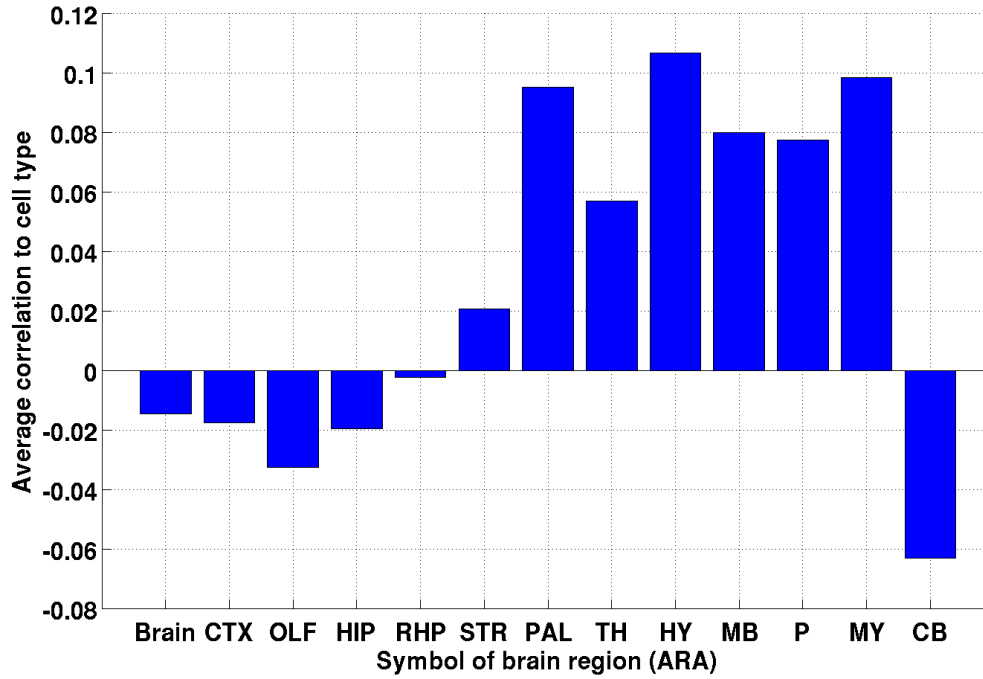


(a)

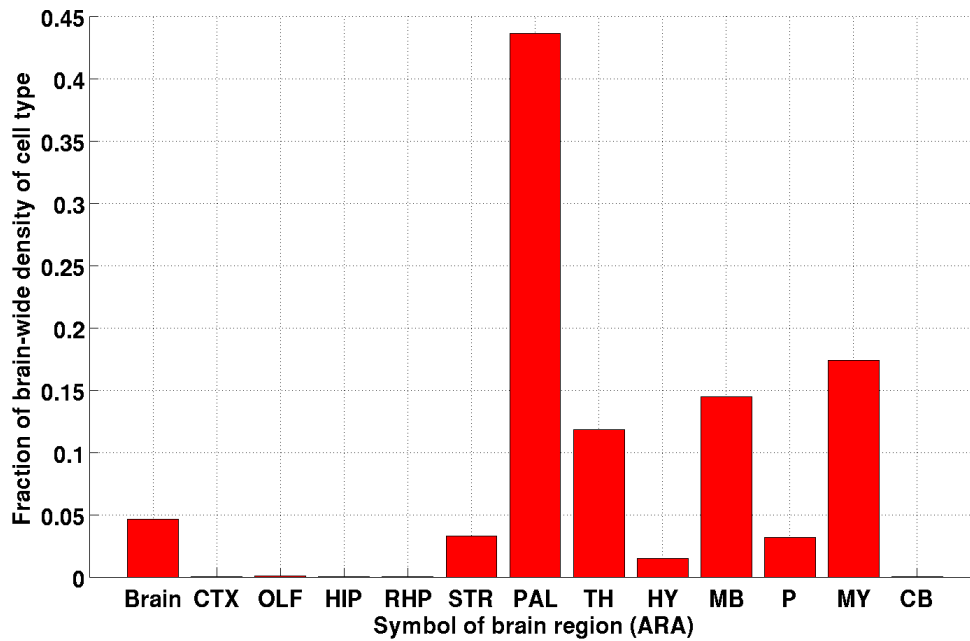


(b)

Figure 17: *Drd2*+ medium spiny neurons (cell-type index 16). (a) Heat map of the brain-wide correlation profile. (b) Heat map of the estimated brain-wide density profile.



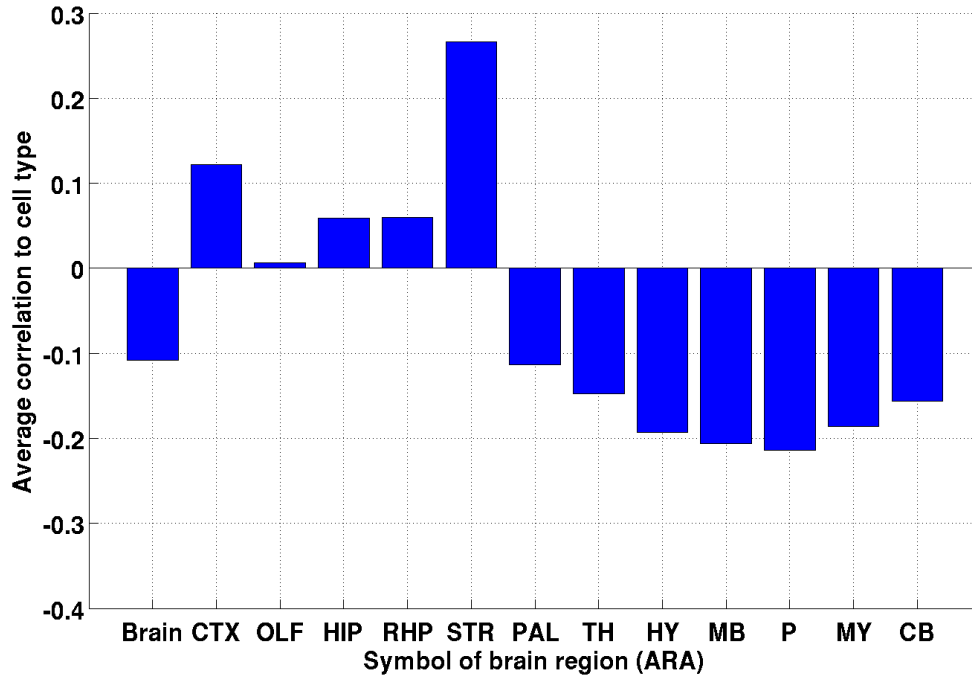
(a)



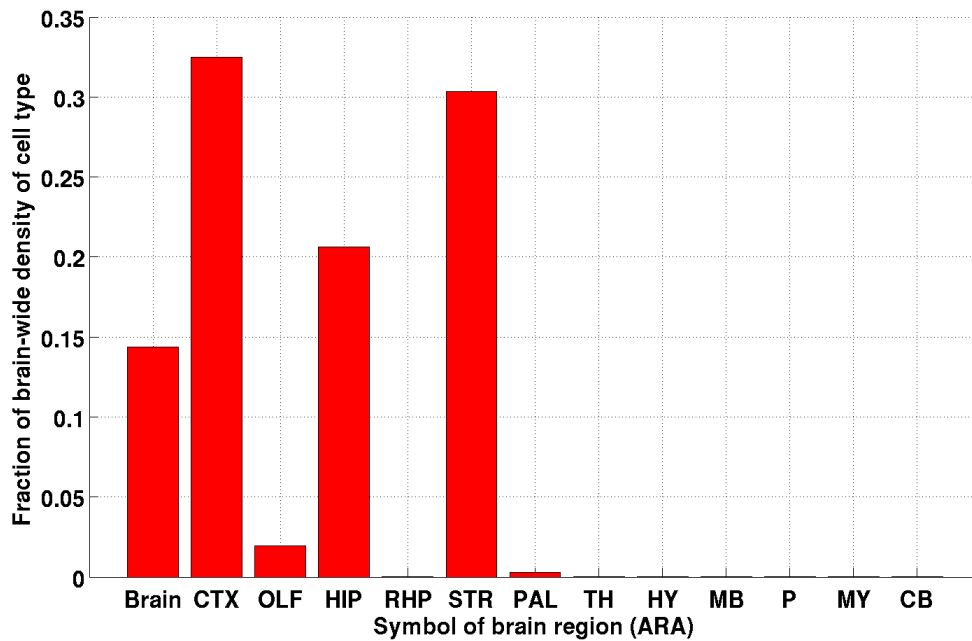
(b)

Figure 18: **Cholinergic neurons (cell-type index 13)**. (a) Average correlation between the cell type and the Allen Atlas, in the regions of the 'big12' annotation of the ARA. (b) Fractions of density of cell type in the regions of the 'big12' annotation of the ARA.



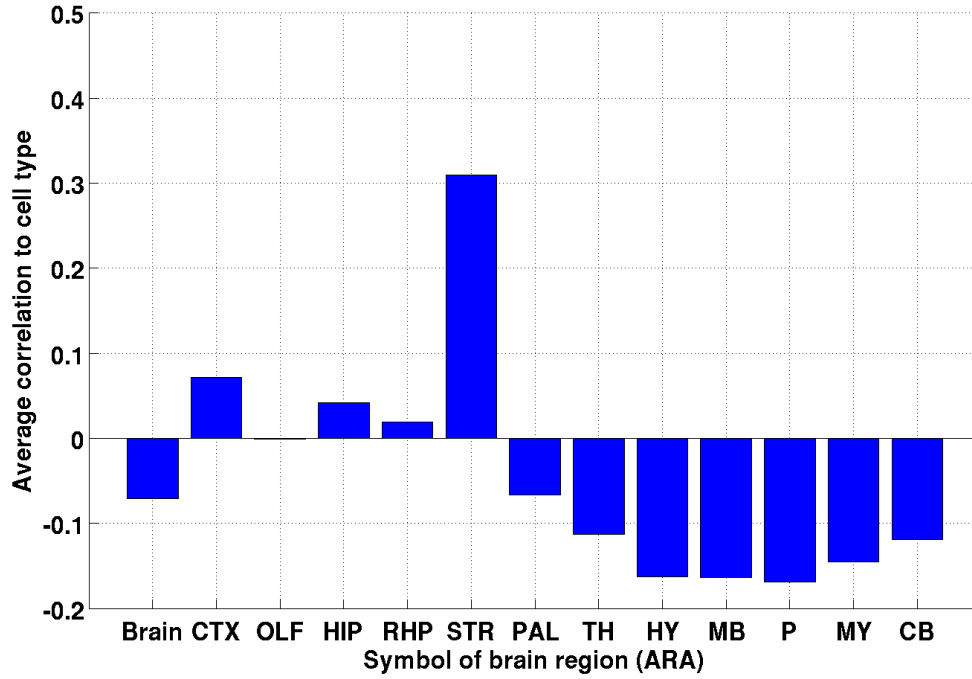


(a)

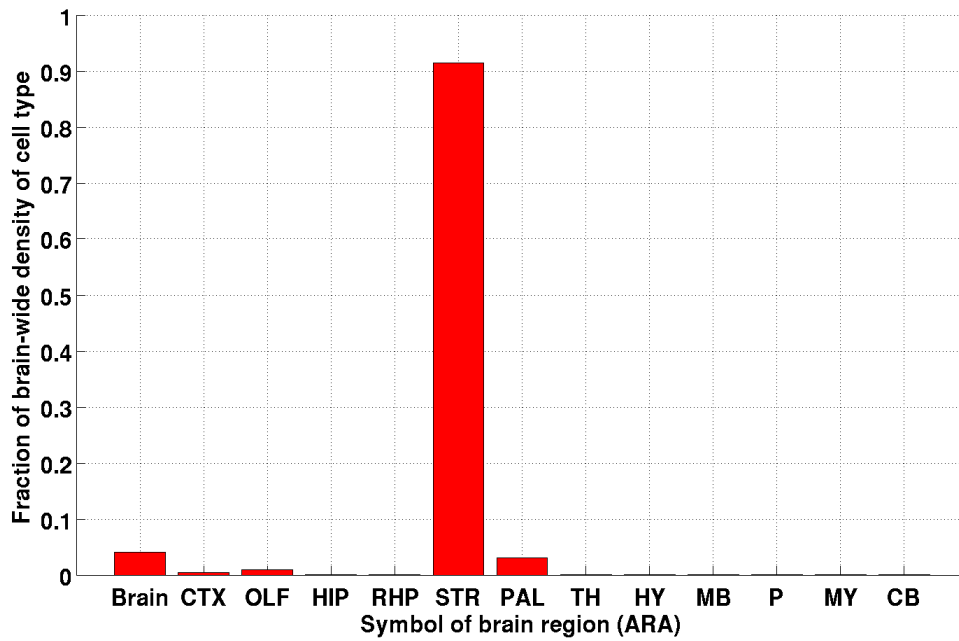


(b)

Figure 19: *Drd1*+ medium spiny neurons (cell-type index 15). (a) Average correlation between the cell type and the Allen Atlas, in the regions of the 'big12' annotation of the ARA. (b) Fractions of density of cell type in the regions of the 'big12' annotation of the ARA.



(a)



(b)

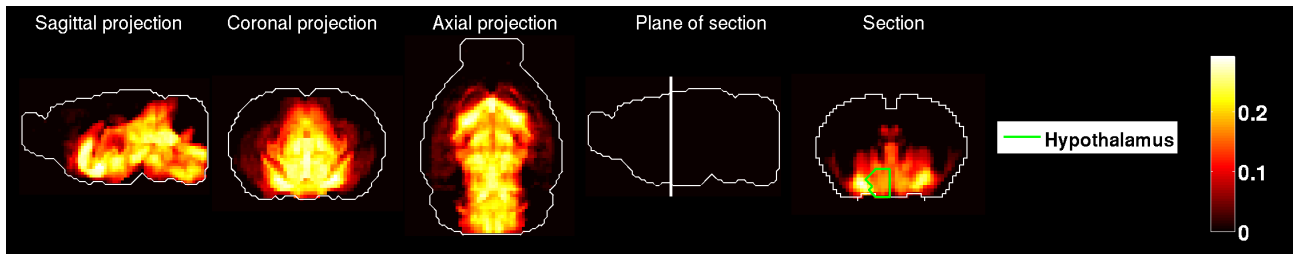
Figure 20: *Drd2*+ medium spiny neurons (cell-type index 16). (a) Average correlation between the cell type and the Allen Atlas, in the regions of the 'big12' annotation of the ARA. (b) Fractions of density of cell type in the regions of the 'big12' annotation of the ARA.

Description (index)	Origin of sample	Rank of region (out of 94) in the 'fine' annotation (by correlation)	Rank of region (out of 94) in the 'fine' annotation (by density)	Fraction of density in the region (%)	Fraction of density supported in the pallidum
Cholinergic projection neurons (11)	Pallidum ventral region	30	N/A	N/A	N/A%

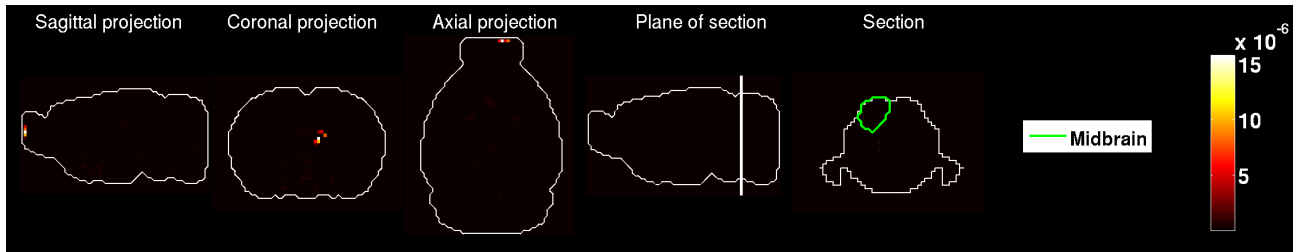
Table 11: Anatomical analysis for the cell-type-specific sample extracted from the pallidum.

### 3.3.5 Pallidum

The cell-type-specific sample with index 11 (Cholinergic projection neurons), which is the only sample obtained from the pallidum, has the hypothalamus as its top region by correlation, followed by pallidum (see Figures 21 and 22). Pallidum ventral region, which is the region in the 'fine' annotation from which the cell-type-specific sample was extracted, ranks 30th out of 94 regions by average correlation (see Table 11). This suggests that cholinergic projection neurons are not very region specific. Moreover, the estimated density of this cell type is zero in the left hemisphere.



(a)



(b)

Figure 21: **Cholinergic projection neurons (cell-type index 11)**. (a) Heat map of the brain-wide correlation profile. (b) Heat map of the estimated brain-wide density profile.

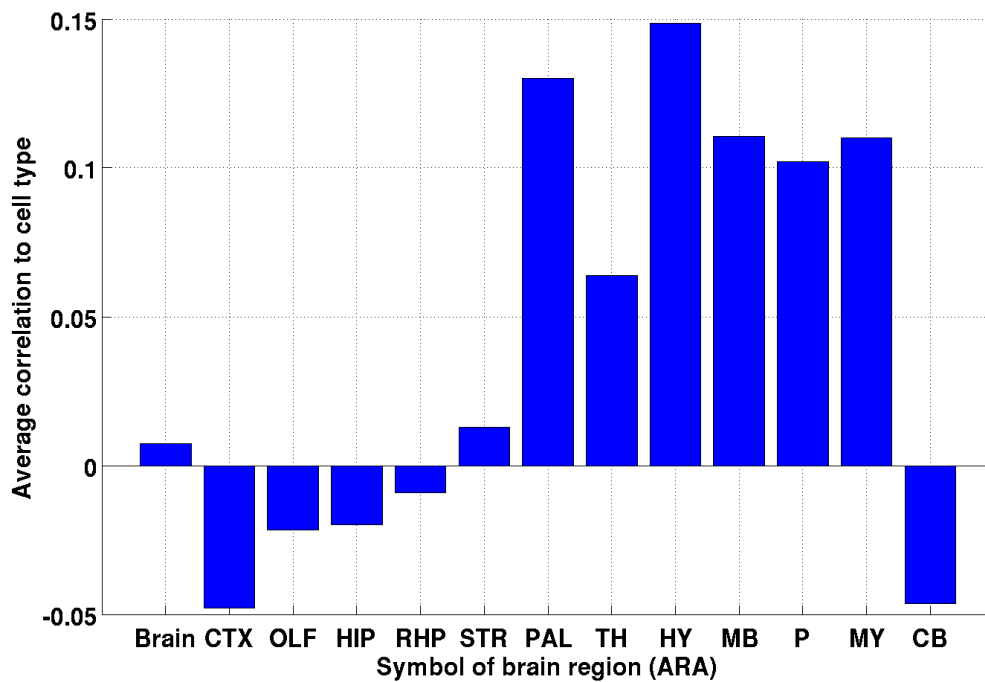


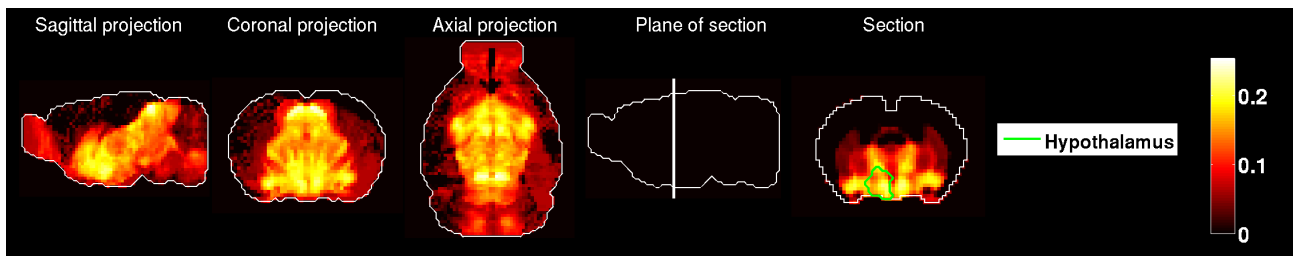
Figure 22: **Cholinergic projection neurons (cell-type index 11)**. Average correlation between the cell type and the Allen Atlas, in the regions of the 'big12' annotation of the ARA. The estimated density of this cell type is zero in the left hemisphere.

Description (index)	Origin of sample	Rank of region (out of 94) in the 'fine' annotation (by correlation)	Rank of region (out of 94) in the 'fine' annotation (by density)	Fraction of density in the region	Fraction of density supported in the thalamus
GABAergic Interneurons, PV+ (59)	Dorsal part of the lateral geniculate complex	63	31	0.2%	21.3%

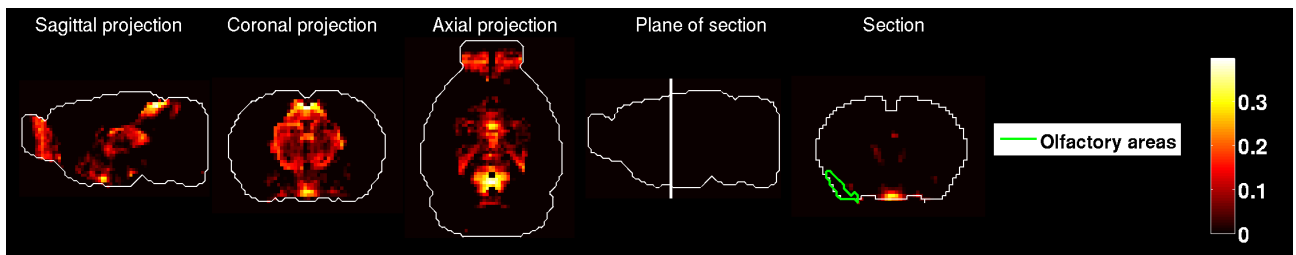
Table 12: Anatomical analysis for the cell-type-specific sample extracted from the thalamus.

### 3.3.6 Thalamus

Table 12 summarizes results for the cell-type-specific sample extracted from the thalamus (GABAergic Interneurons, PV+, index 59). Thalamus supports less of the density profile than Olfactory areas and Midbrain (it is ranked third by density and fourth by correlation). Moreover, there seems to be little solidarity between the ranking of regions by correlation and by density (see Figures 23 and 24).

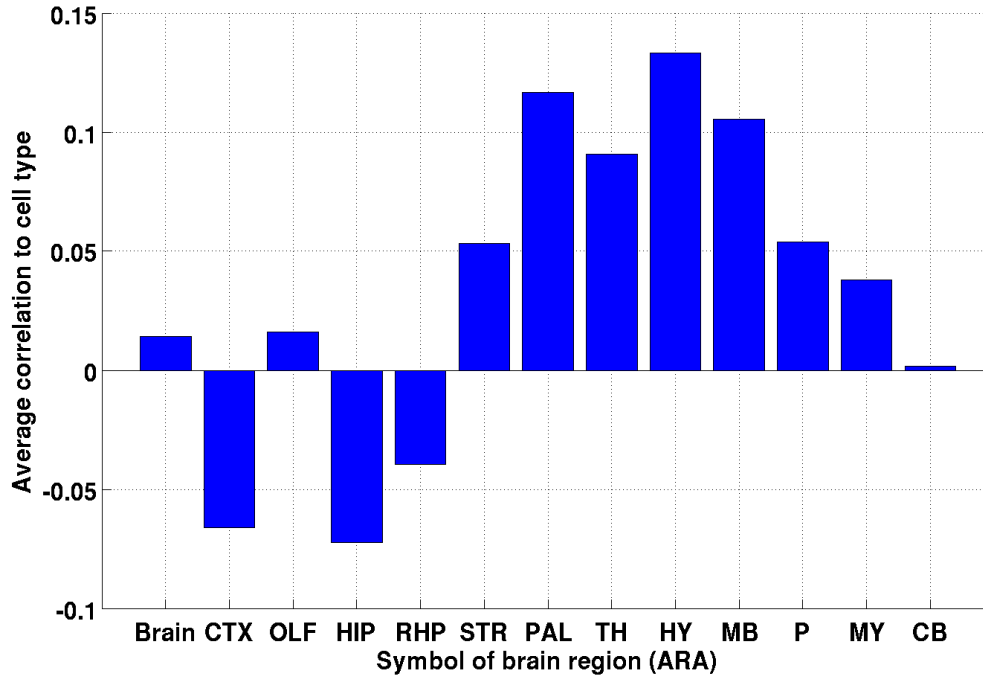


(a)

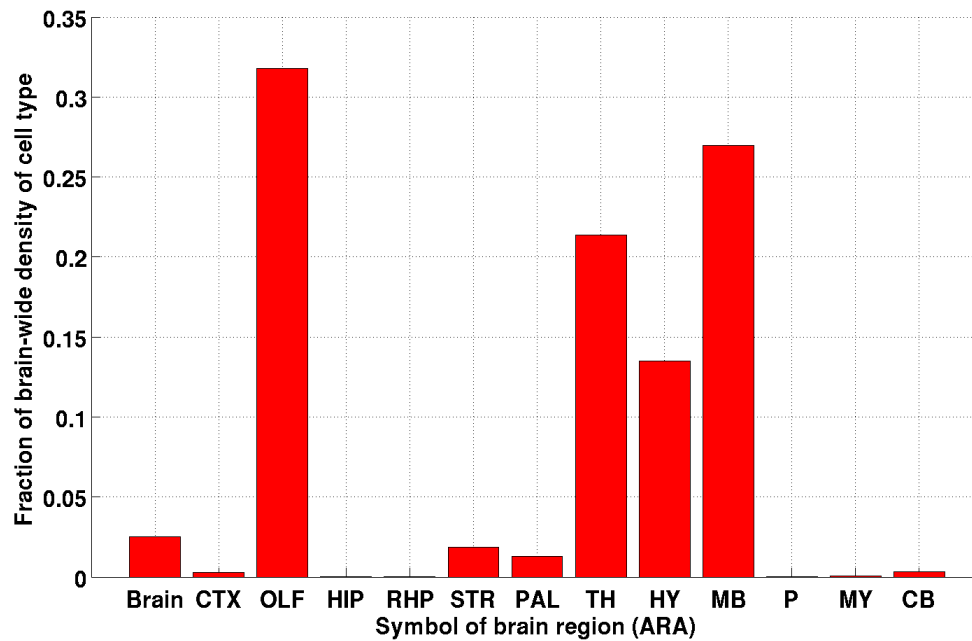


(b)

Figure 23: **GABAergic Interneurons, PV+ (cell-type index 59)**. (a) Heat map of the brain-wide correlation profile. (b) Heat map of the estimated brain-wide density profile.



(a)



(b)

Figure 24: **GABAergic Interneurons, PV+ (cell-type index 59)**. (a) Average correlation between the cell type and the Allen Atlas, in the regions of the 'big12' annotation of the ARA. (b) Fractions of density of cell type in the regions of the 'big12' annotation of the ARA.

### 3.3.7 Midbrain

Three cell-type-specific samples were extracted from the midbrain. Out of them, two (A9 dopaminergic neurons –index 4– and A10 dopaminergic neurons –index 5) have the midbrain as their top region by density (see Figures 25, 26, 28 and 29). For both of these samples, the midbrain ranks second by correlation, after the hypothalamus.

However, a visual inspection of the projections of correlation and density profiles of Figures 25 and 26 shows some heterogeneity across midbrain, with higher values in the ventral region of it. The 'big12' annotation is too coarse for this heterogeneity to be detected by our ranking procedures. We therefore ranked the regions of the 'fine' annotation are ranked by correlation and density according to Equations 12 and 18. The region that is ranked highest by density for A9 dopaminergic neurons is the 'Substantia nigra, compact part', which is a subregion of midbrain, and is indeed the finest anatomical label available (see Table 64). This region is ranked second by correlation, the first being 'Ventral tegmental area' (which is also a subregion of the midbrain).

For A10 dopaminergic neurons the region that is ranked first is the 'Hypothalamus' (which is the 'generic' subregion of hypothalamus in the 'fine' annotation, assigned to any voxels that are in hypothalamus but cannot be reliably assigned to a finer subdivision), but the second region is the 'ventral tegmental area, which is a subregion of midbrain, and is indeed the finest anatomical label available from Tables. It is ranked second by correlation, the first being 'Ventral tegmental area'.

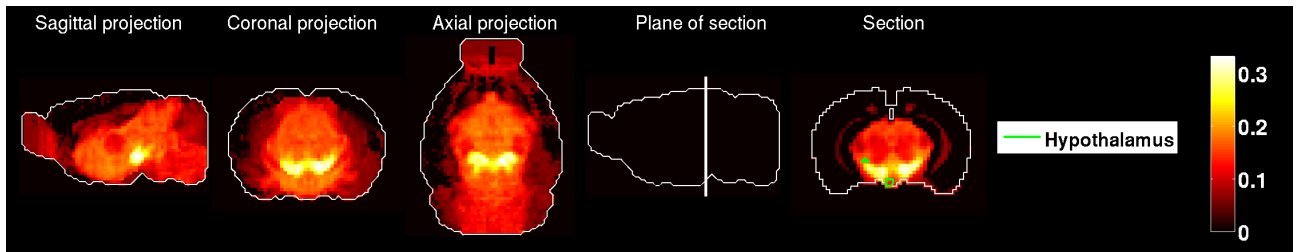
Whereas the midbrain supports a majority of the density for A9 dopaminergic neurons and A10 dopaminergic neurons, it supports only 9.8% of the density in Motor Neurons, Midbrain Cholinergic Neurons (whereas pons and medulla support 32.6% and 51.6% respectively, as can be see on Figure 30).

Whereas midbrain supports a majority of the density for A9 dopaminergic neurons and A10 dopaminergic neurons, it supports only 9.8% of the density in Motor Neurons, Midbrain Cholinergic Neurons (whereas medulla and pons support 32.6% and 51.6% respectively). However, it should be noted that very few cell types in our data come from medulla.

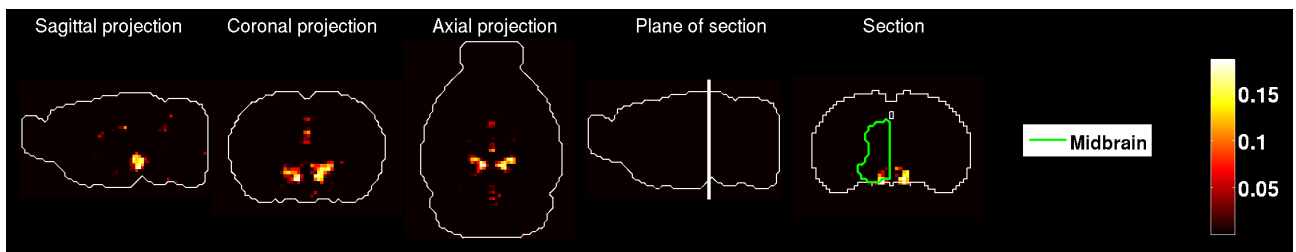


Description (index)	Origin of sample	Rank of region (out of 94) in the 'fine' annotation (by correlation)	Rank of region (out of 94) in the 'fine' annotation (by density)	Fraction of density in the region	Fraction of density supported in midbrain
A9 dopaminergic neurons (4)	Substantia nigra_ compact part	2	1	39 %	77 %
A10 dopaminergic neurons (5)	Ventral tegmental area	1	2	35 %	50 %
Motor Neurons, Midbrain Cholinergic Neurons (10)	Pedunculo-pontine nucleus	17	94 (zero density)	0	9.8 %

Table 13: Anatomical analysis for the cell-type-specific samples extracted from the midbrain.

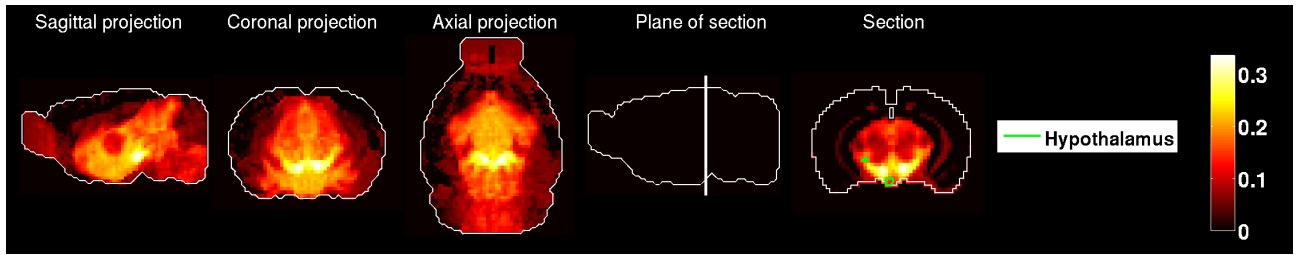


(a)

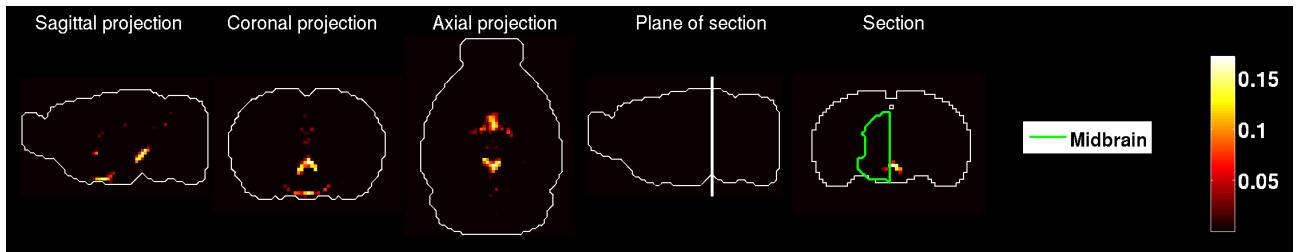


(b)

Figure 25: **A9 dopaminergic neurons (cell-type index 4)**. (a) Heat map of the brain-wide correlation profile. (b) Heat map of the estimated brain-wide density profile.

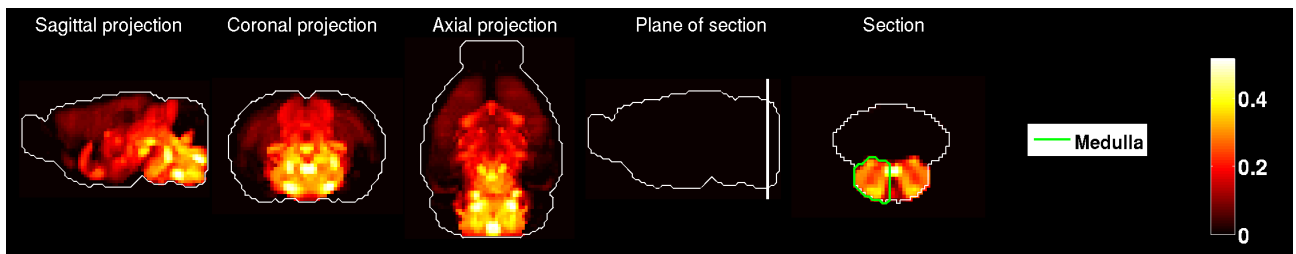


(a)

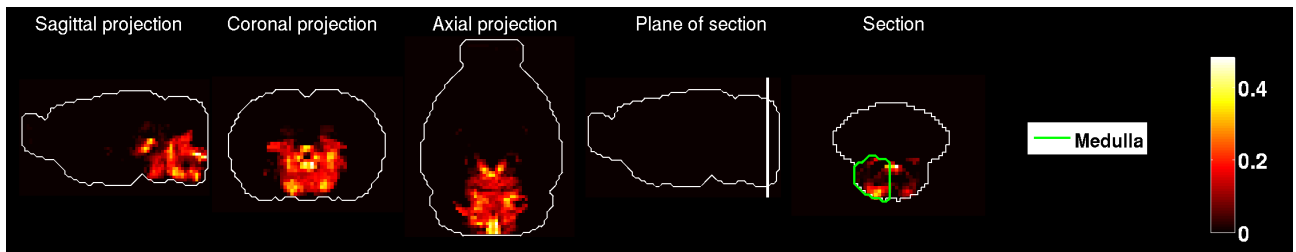


(b)

Figure 26: **A10 dopaminergic neurons (cell-type index 5)**. (a) Heat map of the brain-wide correlation profile. (b) Heat map of the estimated brain-wide density profile.

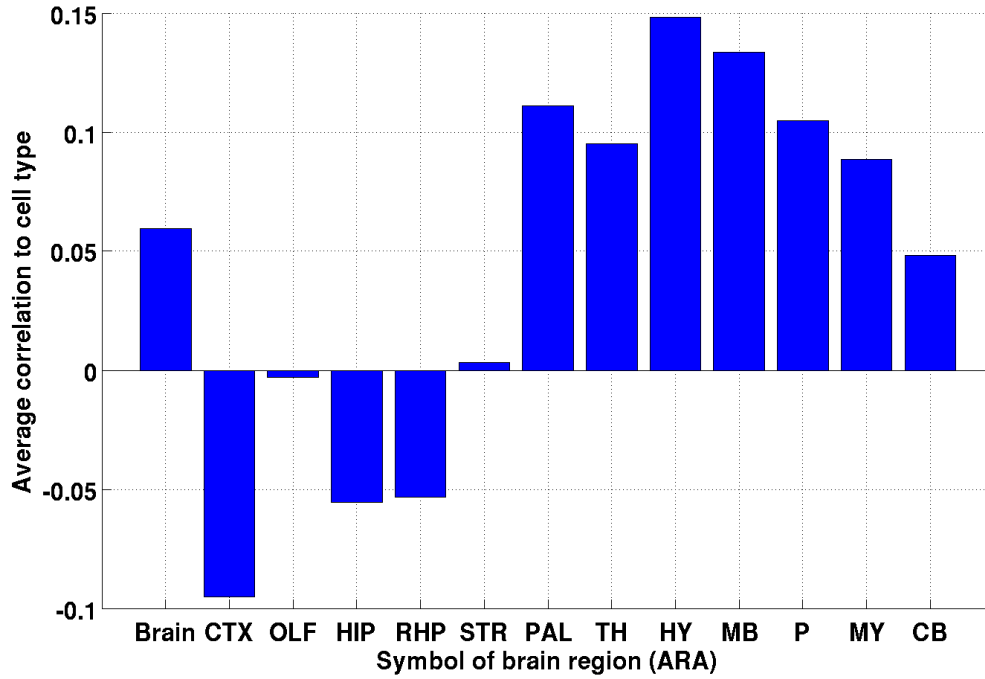


(a)

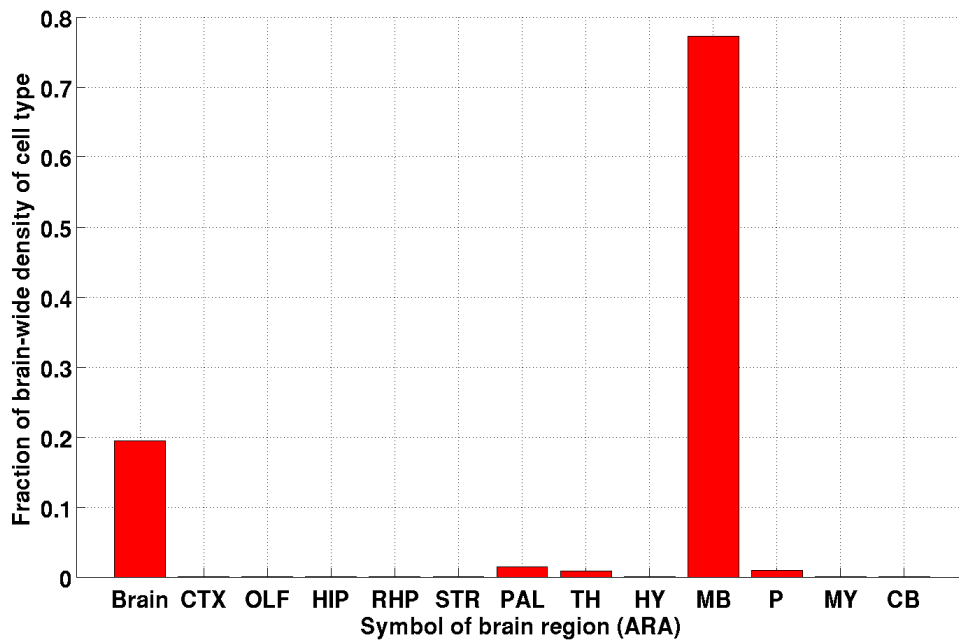


(b)

Figure 27: **Motor Neurons, Midbrain Cholinergic Neurons (cell-type index 10)**. (a) Heat map of the brain-wide correlation profile. (b) Heat map of the estimated brain-wide density profile.

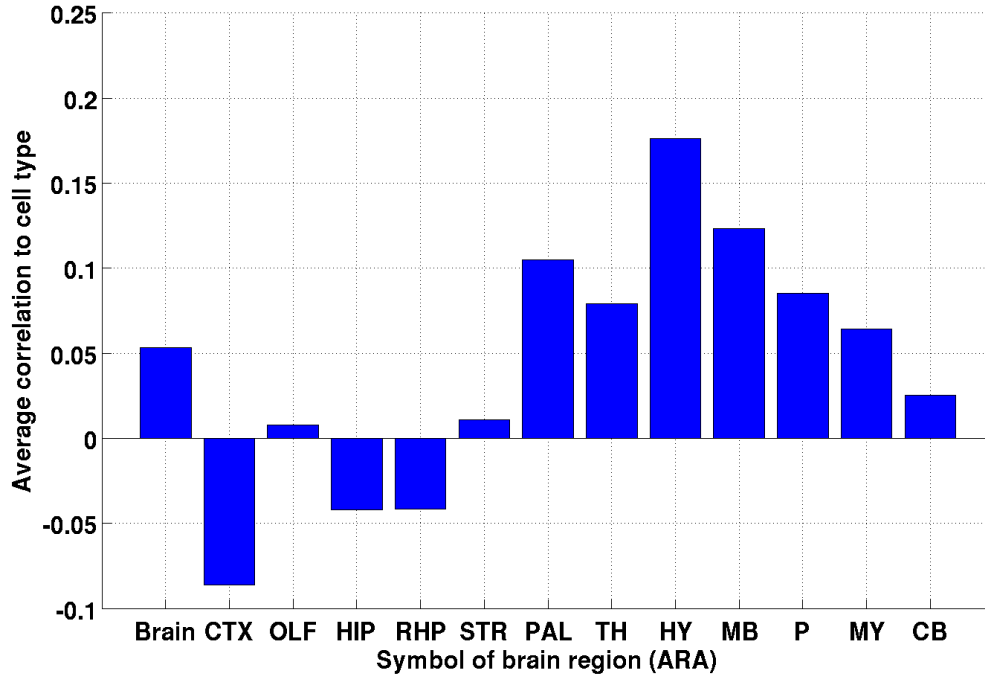


(a)

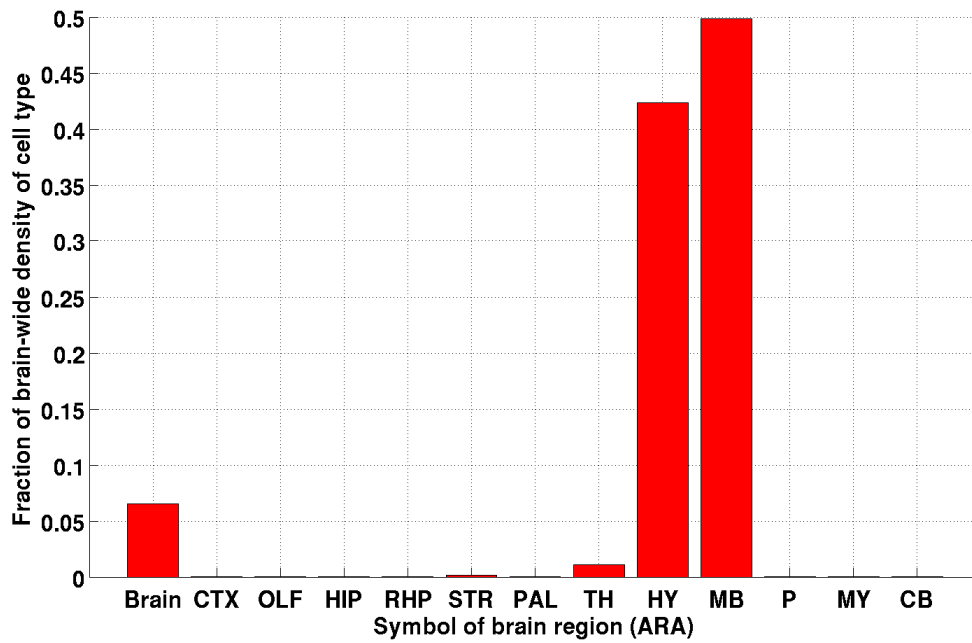


(b)

Figure 28: **A9 dopaminergic neurons (cell-type index 4)**. (a) Average correlation between the cell type and the Allen Atlas, in the regions of the 'big12' annotation of the ARA. (b) Fractions of density of cell type in the regions of the 'big12' annotation of the ARA.

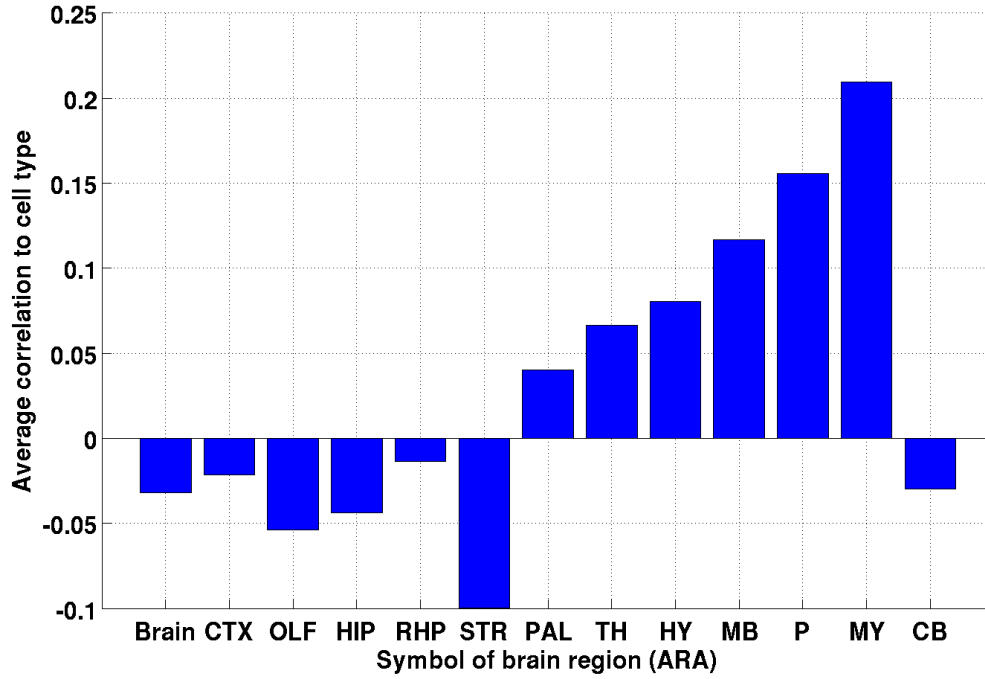


(a)

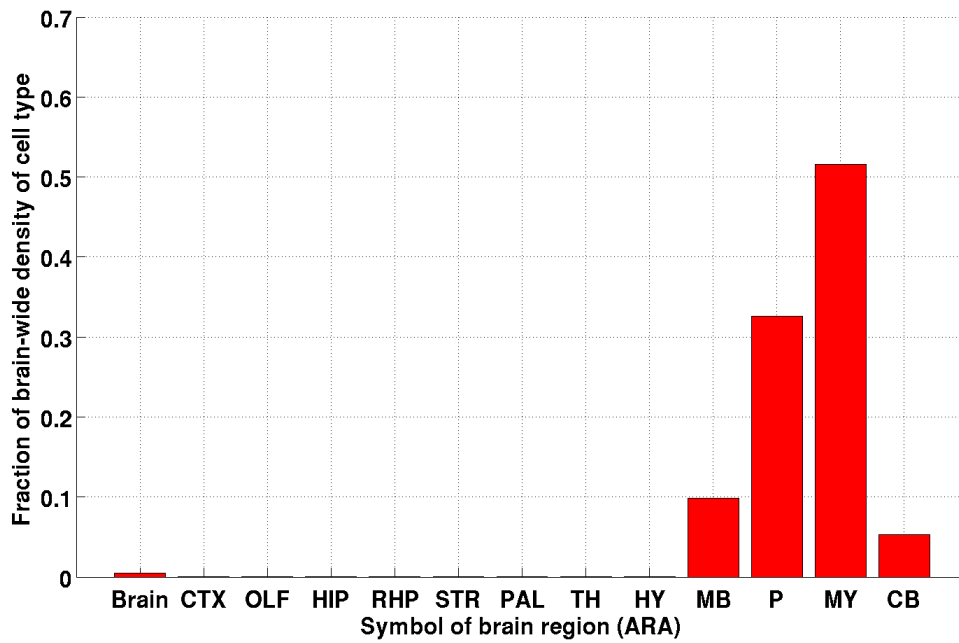


(b)

Figure 29: **A10 dopaminergic neurons (cell-type index 5)**. (a) Average correlation between the cell type and the Allen Atlas, in the regions of the 'big12' annotation of the ARA. (b) Fractions of density of cell type in the regions of the 'big12' annotation of the ARA.



(a)



(b)

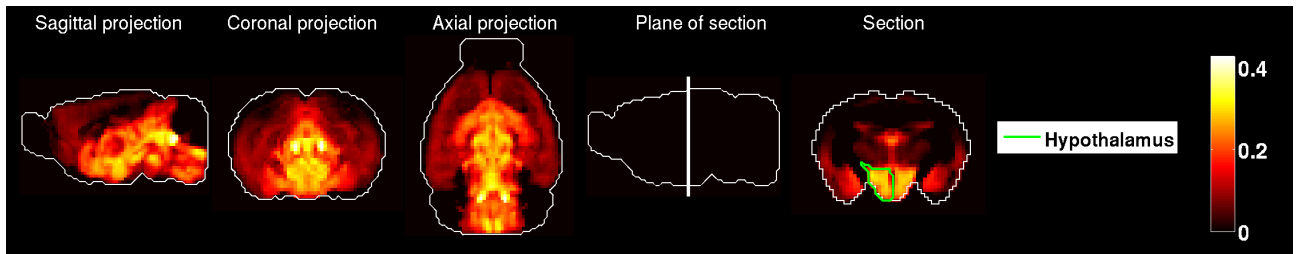
Figure 30: **Motor Neurons, Midbrain Cholinergic Neurons (cell-type index 10)**. (a) Average correlation between the cell type and the Allen Atlas, in the regions of the 'big12' annotation of the ARA. (b) Fractions of density of cell type in the regions of the 'big12' annotation of the ARA.

Description (index)	Origin of sample	Rank of region (out of 94) in the 'fine' annotation (by correlation)	Rank of region (out of 94) in the 'fine' annotation (by density)	Fraction of density in the region	Fraction of density supported in pons
Tyrosine Hydroxylase Expressing (51)	Pontine central gray	9	26	1.3 %	10.4 %

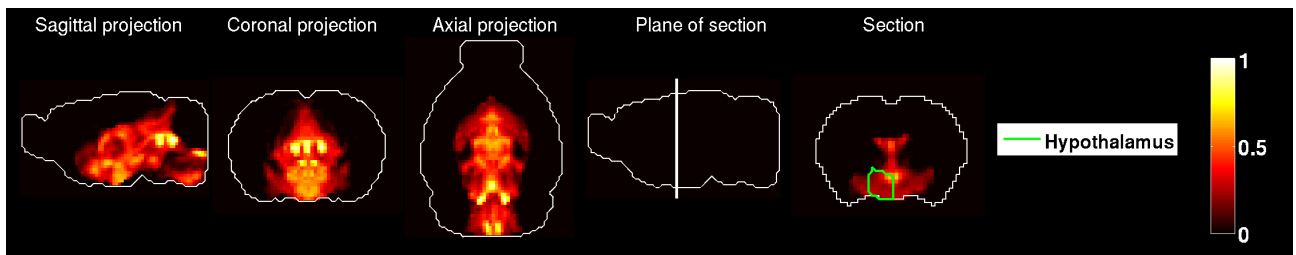
Table 14: Anatomical analysis for the cell-type-specific sample extracted from the pons.

### 3.3.8 Pons

One cell-type-specific sample in this study was extracted from the pons (index 51, unpublished data). According to the 'fine' annotation, it was extracted from the pontine central gray. Pons is ranked 4th in the 'big12' annotation both by density and correlation (after hypothalamus, midbrain and medulla, which support 28.9%, 26.7% and 12.5% of the density respectively, whereas pons supports 10.4% of the density for this cell type). The pontine central gray is ranked 26th out of 94 regions in the fine annotation. The fraction of the total estimated density of this cell type cumulated by the first 26 regions in the fine annotation is 81.9%. Conversely, this cell type is the second most important detected in the pons (after the GABAergic interneurons, PV+, index 64).

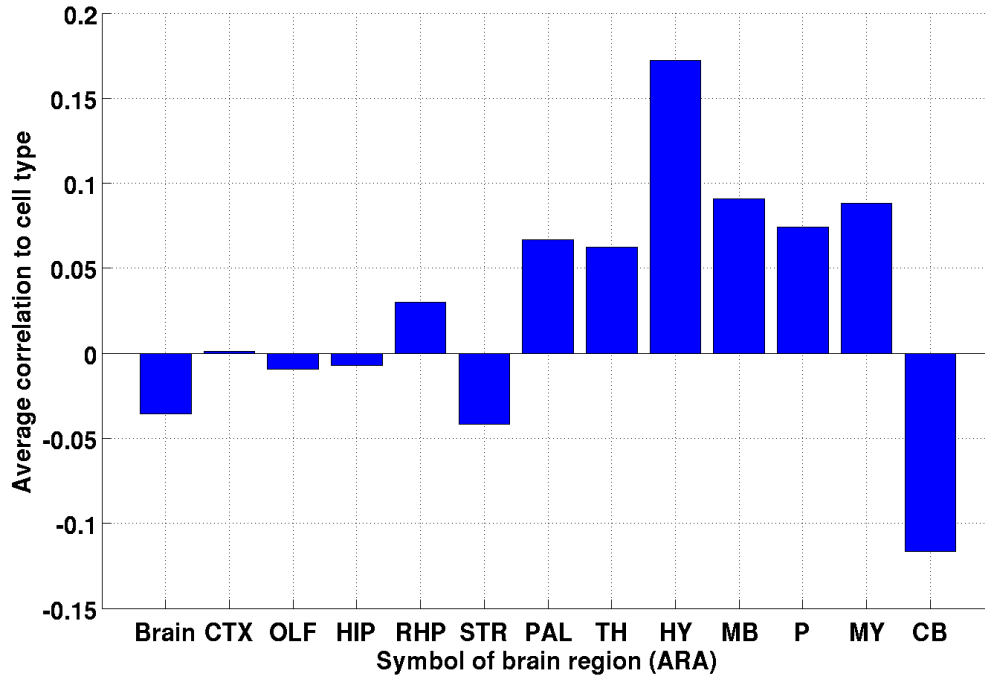


(a)

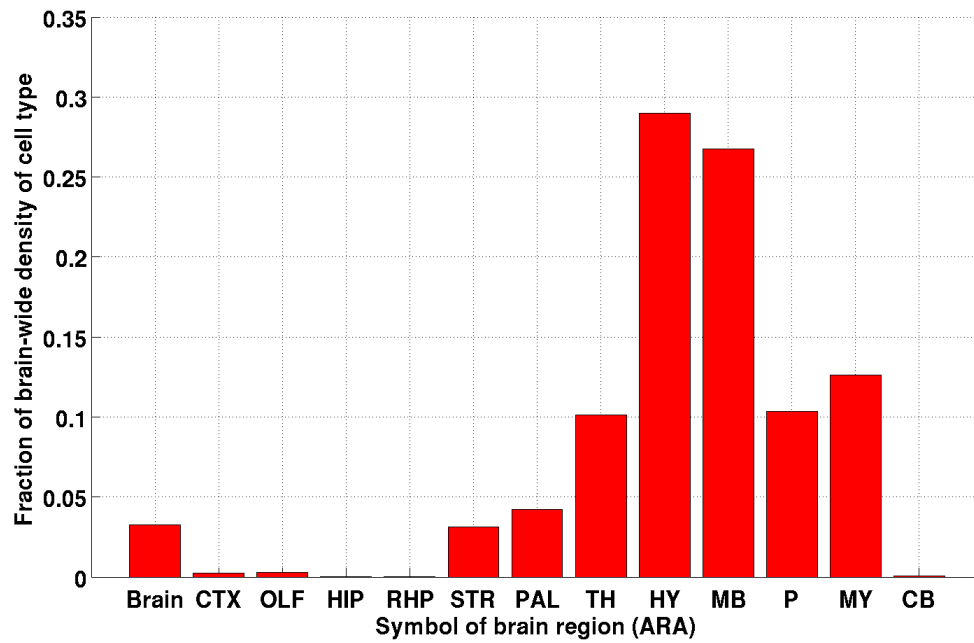


(b)

Figure 31: **Tyrosine Hydroxylase Expressing (cell-type index 51)**. (a) Heat map of the brain-wide correlation profile. (b) Heat map of the estimated brain-wide density profile.



(a)



(b)

Figure 32: **Tyrosine Hydroxylase Expressing (cell-type index 51)**. (a) Average correlation between the cell type and the Allen Atlas, in the regions of the 'big12' annotation of the ARA. (b) Fractions of density of cell type in the regions of the 'big12' annotation of the ARA.

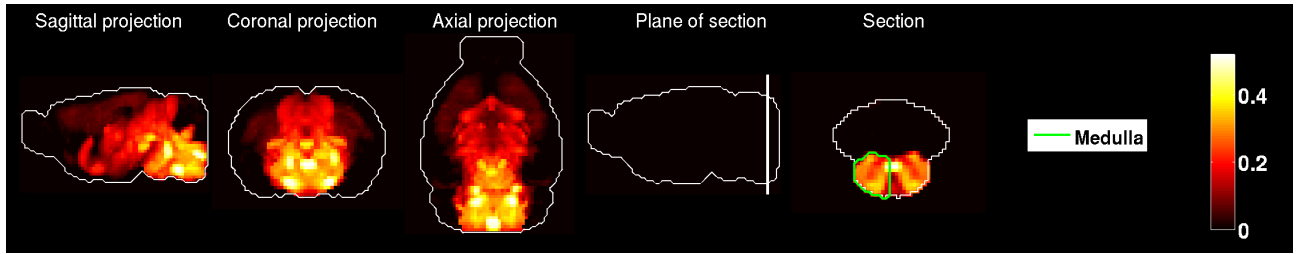


Description (index)	Origin of sample	Rank of region (out of 94) in the 'fine' annotation (by correlation)	Rank of region (out of 94) in the 'fine' annotation (by density)	Fraction of density in the region	Fraction of density supported in the medulla
Motor neurons, cholinergic interneurons (12)	Medulla	13	1	32.5 %	62.3%

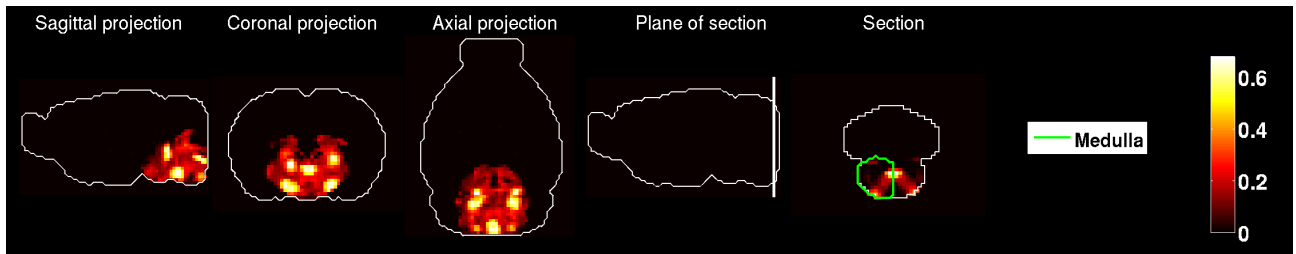
Table 15: Anatomical analysis for the cell-type-specific sample extracted from the medulla.

### 3.3.9 Medulla

The cell-type-specific sample with index 12 (motor neurons, cholinergic interneurons), was extracted from the spinal chord. The closest region in 'big12' is the medulla, which has a refinement into 15 different regions. Medulla is indeed the top region in the 'big12' annotation for this cell type, both by correlation and density (see Figures 33 and 34). The second region in 'big12' by density is the pons (33%), so the medulla and the pons support more than 95% of the estimated density of this cell type. The breakdown of the density among the regions of the medulla is shown in Table 16.



(a)

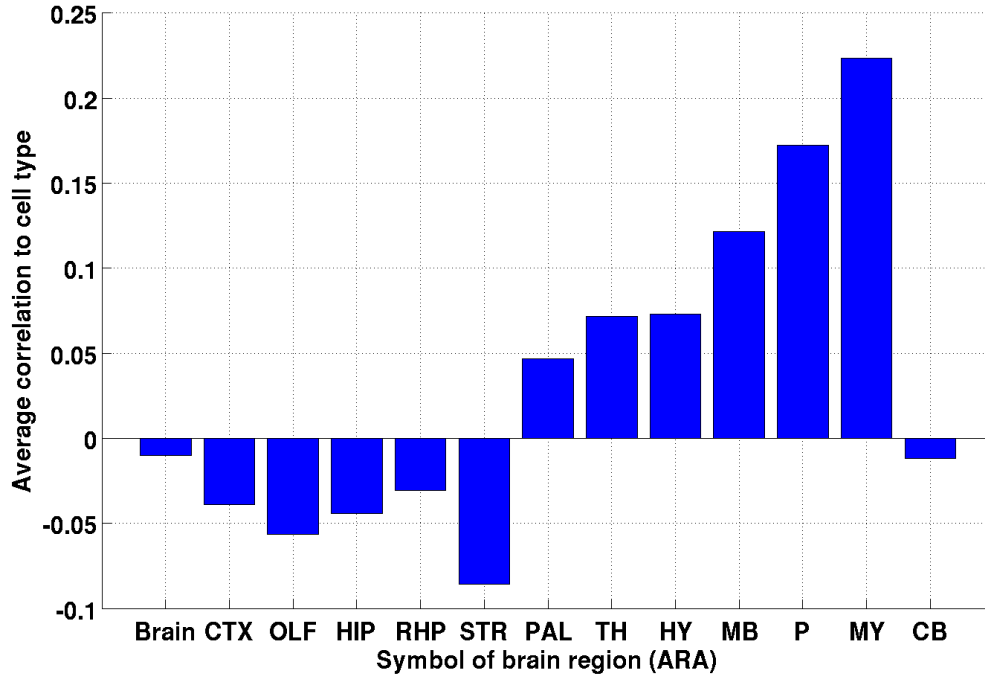


(b)

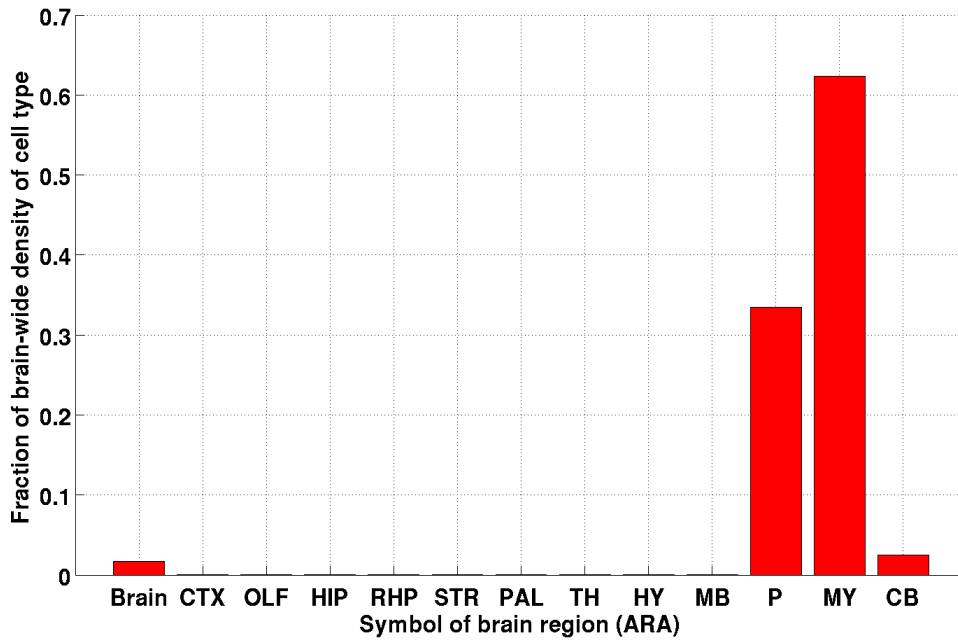
Figure 33: **Motor neurons, cholinergic interneurons (cell-type index 12)**. (a) Heat map of the brain-wide correlation profile. (b) Heat map of the estimated brain-wide density profile.

Brain region (in the 'fine' annotation)	Fraction of density supported in the region (%)
Medulla	32.5
Facial motor nucleus	6.9
Vestibular nuclei	5.5
Hypoglossal nucleus	3.6
Paragigantocellular reticular nucleus	3.3
Spinal nucleus of the trigeminal_ interpolar part	3.2
Lateral reticular nucleus	2.8
Spinal nucleus of the trigeminal_ oral part	1.8
Magnocellular reticular nucleus	1.3
Cochlear nuclei	1
Medulla_ behavioral state related	0.4
Spinal nucleus of the trigeminal_ caudal part	0.2
Dorsal column nuclei	0.2
Inferior olivary complex	0.2
Nucleus of the solitary tract	0

Table 16: Fractions of the density profile of cell-type index 12 supported by subsets of medulla in the ARA.



(a)



(b)

Figure 34: **Motor neurons, cholinergic interneurons (cell-type index 12)**. (a) Average correlation between the cell type and the Allen Atlas, in the regions of the 'big12' annotation of the ARA. (b) Fractions of density of cell type in the regions of the 'big12' annotation of the ARA.

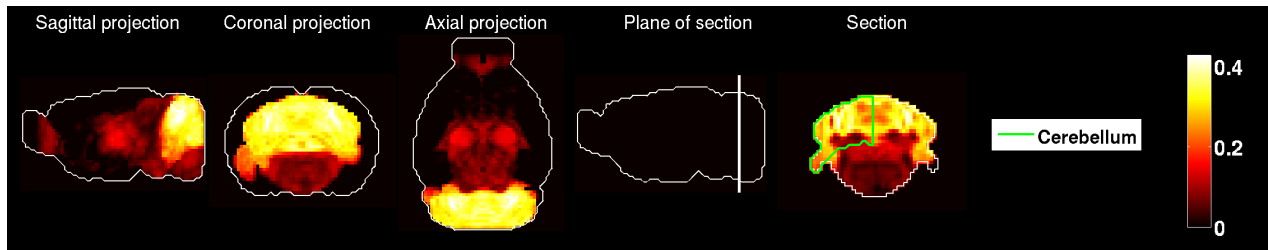
Description (index)	Top region by correlation	Top region by density (percentage of density supported)	Fraction of density supported in the cerebellum (%)
Purkinje Cells (1)	Cerebellum	Cerebellum	95.8
Golgi Cells (17)	Pons	N/A(*)	0
Unipolar Brush cells (some Bergman Glia) (18)	Cerebellum	Thalamus	0.2
Stellate Basket Cells (19)	Cerebellum	Medulla	18.8
Granule Cells (20)	Cerebellum	Cerebellum	96.0
Mature Oligodendrocytes (21)	Cerebellum	Cerebellum	39.9
Mixed Oligodendrocytes (23)	Pons	N/A(*)	0
Purkinje Cells (25)	Cerebellum	Olfactory areas	0
Bergman Glia (27)	Cerebellum	Olfactory area	5.5
Purkinje Cells (52)	Cerebellum	Thalamus	5.9
*Zero density in the left hemisphere.			

Table 17: Cell-type-specific samples extracted from the cerebellum.

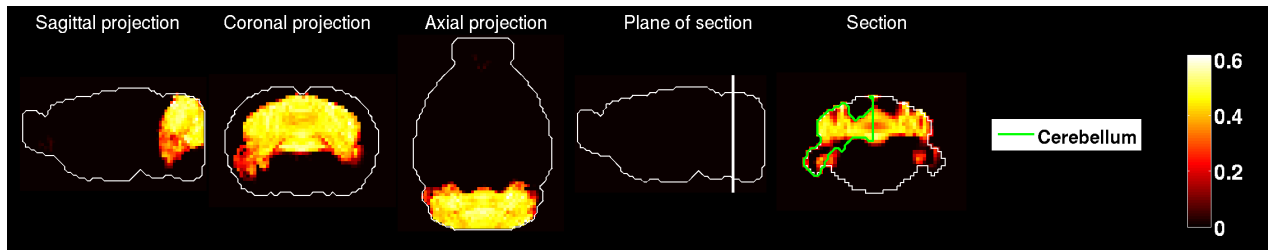
### 3.3.10 Cerebellum

Out of the 11 samples that were drawn from the cerebellum, 7 have the cerebellum as their top region by correlation, and 4 have the cerebellum as their top region by density (all of which also have the cerebellum as their top region by correlation, see 17). All the samples that have the cerebellum as their top region were indeed taken from the cerebellum. See Table 3 for the Mature oligodendrocytes (index 28), whose density profiles follows a white-matter pattern that includes the *arbor vitae*. See Figure 35 for a class of Purkinje cells [23] and Figure 36 for a class of mature oligodendrocytes [19], both extracted from the cerebellum. Their correlation and estimated density patterns are indeed mostly localized in cerebellum.

A remarkable class of Purkinje cells does not have the cerebellum as its top region by density. See Figure 37 for a class of Purkinje cells (index 52, unpublished) that correlates best with the Allen Atlas both in the thalamus and in the cerebellum, but that fits based on density coefficients only in the thalamus. This indicates that thalamus must contain cell types whose gene expression profile is closest to Purkinje cells in the present data set, but that thalamus is not sampled in enough detail by our microarray data set for these cell types to be distinguished from this class of Purkinje cells.

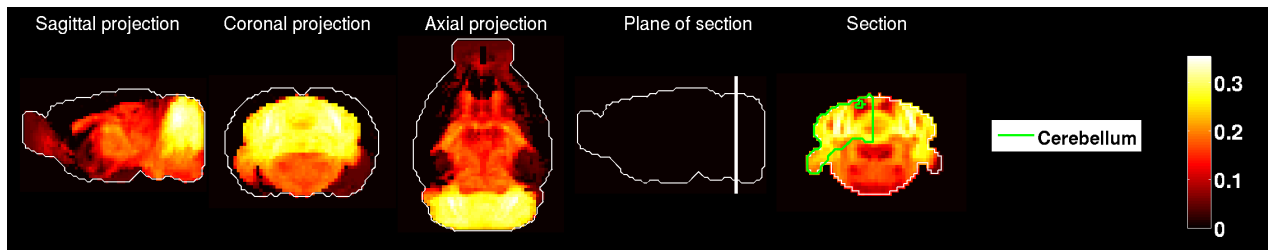


(a)

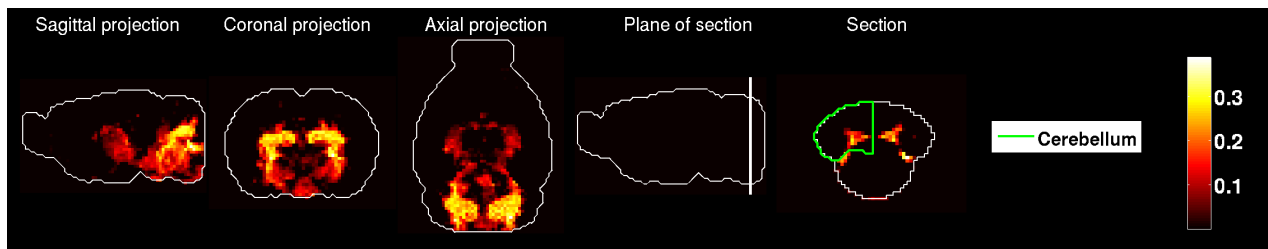


(b)

Figure 35: **Purkinje cells (cell-type index  $t = 1$ )**. (a) Average correlation between the cell type and the Allen Atlas, in the regions of the 'big12' annotation of the ARA. (b) Fractions of density of cell type in the regions of the 'big12' annotation of the ARA.

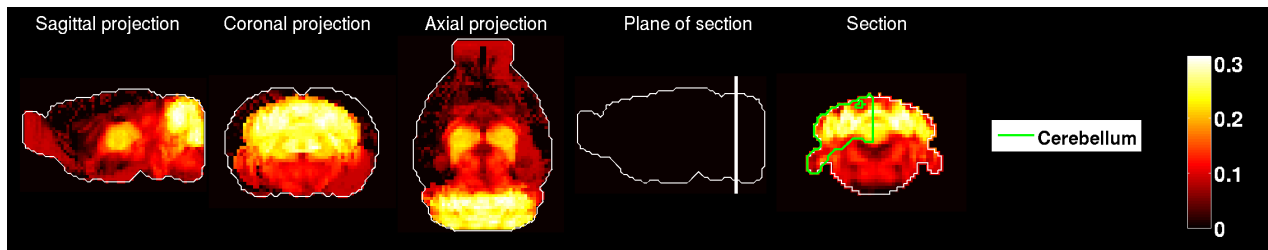


(a)

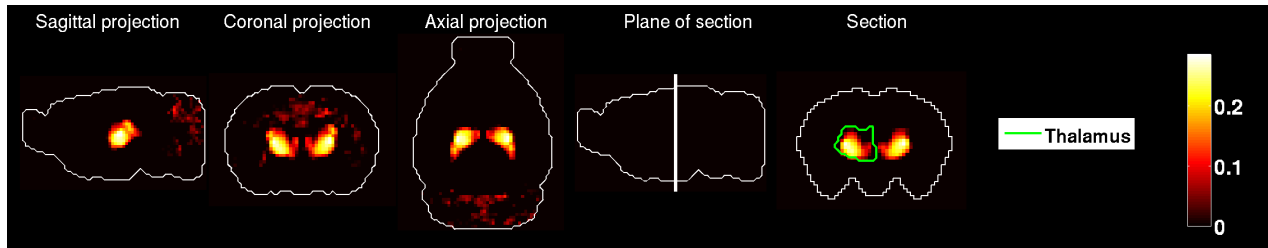


(b)

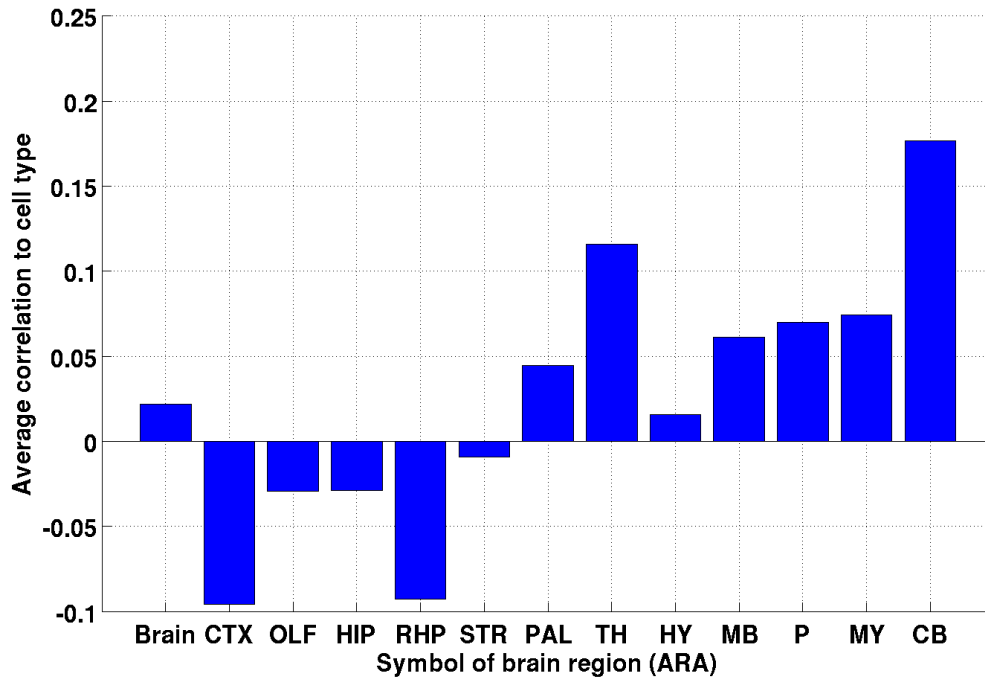
Figure 36: **Mature oligodendrocytes (cell-type index  $t = 21$ )**.(a) Average correlation between the cell type and the Allen Atlas, in the regions of the 'big12' annotation of the ARA. (b) Fractions of density of cell type in the regions of the 'big12' annotation of the ARA.



(a)



(b)



(c)

Figure 37: **Purkinje cells (cell-type index  $t = 52$ ).**(a) Average correlation between the cell type and the Allen Atlas, in the regions of the 'big12' annotation of the ARA. (b) Fractions of density of cell type in the regions of the 'big12' annotation of the ARA. Thalamus supports a majority of the estimated density. (c) Average correlations in the regions of the 'big12' annotation. Cerebellum is the top region, followed by thalamus.

## 4 Discussion of the linear model

In this section we discuss several limitations of the linear model. We address them by refitting the linear model to a modified by panel on cell-type-specific transcription profile:

- In the first subsection we replace pairs of highly similar cell types<sup>2</sup> by one of these cell types in the fitting panel, and verify that the estimated profile of the remaining cell type is close to the sum of the two estimated profiles for the pair of cell types in the original model (while the results for other cell types are stable).
- Moreover, we refit the model to a panel where the 18 transcription profiles of pyramidal neurons are replaced by one composite cell type equal to their average. The estimated density profile of the composite cell type is mostly cortical, and correlated to the sum of estimated profiles of pyramidal neurons (the rest of the results are stable).
- Finally, we study the response of the estimated densities to a correction in the cell-type-specific profiles consisting of a negative term (which offsets the cross-hybridization). The top regions of most cell types do not change, which confirms the anatomical conclusions of the linear model, but the residual terms go down, suggesting that the accuracy of the model is improved when cross-hybridization is taken into account.

Refitting of the model gives rise to a new set of estimated density profiles  $\rho_t^{\text{new}}$  of cell types, for  $t$  in a set of cell types, some of which have corresponding entries in the original model discussed above:

$$E(v, g) = \sum_{t=1}^{T'} \rho_t^{\text{new}}(v) C^{\text{new}}(t, g) + \text{Residual}^{\text{new}}(v, g). \quad (15)$$

If a cell type labeled  $t$  is present in both panels, we can compare the results of the two models quantitatively by computing the correlation coefficient between the two brain-wide density profiles labeled  $t$  and  $l$  in the respective models:

$$\mathcal{C}(\rho_t, \rho_l^{\text{new}}) = \frac{\sum_{v=1}^V (\rho_t(v) - \overline{\rho_t}) (\rho_l^{\text{new}}(v) - \overline{\rho_l^{\text{new}}})}{\sqrt{\left(\sum_{v=1}^V (\rho_t(v) - \overline{\rho_t})^2\right) \left(\sum_{v=1}^V (\rho_l^{\text{new}}(v) - \overline{\rho_l^{\text{new}}})^2\right)}}, \quad (16)$$

where the overlined quantities are cell-type-specific densities averaged over voxels:

$$\overline{\rho_t} = \frac{1}{V} \sum_{v=1}^V \rho_t(v), \quad \overline{\rho_l^{\text{new}}} = \frac{1}{V} \sum_{v=1}^V \rho_l^{\text{new}}(v). \quad (17)$$

---

<sup>2</sup>such as the two samples of medium spiny neurons, indices 15 and 16 in the original panel

To study the anatomical properties of the results in a refitted model, we can compute the average density in each brain region labeled  $r$ :

$$\bar{\rho}(r, t) = \frac{1}{|\sum_{v \in \text{Brain Annotation}} \rho_t(v)|} \sum_{v \in V_r} \rho_t(v), \quad (18)$$

and check for a cell type labeled  $t$  if the region with the maximum value of  $\bar{\rho}^{\text{new}}(., t)$  is identical to the one induced by  $\bar{\rho}(., t)$ . The top region is the one that is selected to choose the section when plotting the heat maps of the brain-wide density profile. It can therefore be read off when plotting the densities  $\rho_t$  and  $\rho_t^{\text{new}}$  side by side.

## 4.1 Competition between pairs of similar cell types

### 4.1.1 Methods

As can be expected from the names of the  $T = 64$  cell types in this study (see Tables 62, 63), some pairs of cell-type-specific transcription profiles can be very close to each other. This can lead to uniqueness problems in the predicted densities. Indeed, in a situation where two rows, labeled say  $l_1$  and  $l_2$  of the matrix  $C$  are identical, there is a one-parameter family of solutions to the optimization problem for the brain-wide densities of the two corresponding cell types:

$$\forall \phi \in [0, 1], \forall v \in [1..V] \quad \rho_{t_1}(v)C(t_1, \cdot) + \rho_{t_2}(v)C(t_2, \cdot) = (1 - \phi)\rho_{t_1}(v)C(t_1, \cdot) + (\rho_{t_2}(v) + \phi\rho_{t_1}(v))C(t_2, \cdot). \quad (19)$$

The transfer of any positive amount of signal between the two cell types at any voxel does not change the value of the sum (the density profiles of the two types are anti-correlated across the family of solutions).

We looked for families of cell-type-specific samples that are very similar to each other. We computed the pairwise type-by-type correlation matrix between centered cell-type specific transcription profiles:

$$\text{typeCorr}(t, t') = \frac{\sum_{g=1}^G (C(t, g) - \bar{C}(g)) (C(t', g) - \bar{C}(g))}{\sqrt{\sum_{g=1}^G (C(t, g) - \bar{C}(g))^2 \sum_{h=1}^G (C(t', h) - \bar{C}(h))^2}}, \quad (20)$$

The closer the entry  $\text{typeCorr}(t, t')$  is to 1, the more similar the two cell types  $t$  and  $t'$  are. We applied the thresholding procedure that was used for co-expression matrices of genes in the brain in [31]. For each value of  $\tau$  in the interval  $[-1, 1]$ , a threshold can be applied to the matrix  $\text{typeCorr}$  by putting to zero all entries lower than  $\tau$ :

$$\text{typeCorr}_\rho(t, t') = \mathbf{1}(\text{typeCorr}(t, t') \geq \rho) (1 + \text{typeCorr}(t, t')) \quad (21)$$

Applying Tarjan's algorithm [?] to the thresholded matrix  $\text{typeCorr}_\rho$  produces a partition of the  $T$  cell types into strongly connected components at this value of the threshold. At  $\tau = 1$ , all the cell types are disconnected (unless there are cell types with exactly the same value for



all genes, which is not the case in our data set). At  $\tau = -1$ , there is only one connected component, and all cell types are connected. When  $\tau$  decreases from 1 to  $-1$ , the more similar cell types group into connected components. We are interested in pairs of cell types that are connected at high values of  $\tau$  and do not connect with other cell types for the lowest possible value of  $\tau$ , when  $\tau$  is decreased (such cell types labeled  $t$  and  $t'$  are the closest pairs of cell types to the degenerate situation described by Equation ??).

Consider a pair of indices  $(t_1, t_2)$  singled out by the above analysis. Let us introduce the notation  $\hat{C}^i$  for the type-by-gene matrix obtained from  $C$  by leaving the  $i$ -th row out, we computed the corresponding density profiles denoted by  $\hat{\rho}^{t_1}$  and  $\hat{\rho}^{t_2}$ :

$$(\hat{\rho}_t^{t_1}(v))_{1 \leq t \leq T} = \operatorname{argmin}_{\nu \in \mathbf{R}_+^T} \sum_{g=1}^G \left( E(v, g) - \sum_{t=1}^T \hat{C}^{t_1}(t, g) \nu(t) \right)^2, \quad (22)$$

$$(\hat{\rho}_t^{t_2}(v))_{1 \leq t \leq T} = \operatorname{argmin}_{\nu \in \mathbf{R}_+^T} \sum_{g=1}^G \left( E(v, g) - \sum_{t=1}^T \hat{C}^{t_2}(t, g) \nu(t) \right)^2, \quad (23)$$

Intuitively, refitting the model with just one of the two similar types present ( $t_1$  or  $t_2$ ) should result in the remaining cell type inheriting the sum of the two density profiles estimated in the original model:

$$\hat{\rho}_{t_2}^{t_1} \stackrel{?}{\simeq} \rho_{t_1} + \rho_{t_2}, \quad (24)$$

$$\hat{\rho}_{t_1}^{t_2} \stackrel{?}{\simeq} \rho_{t_1} + \rho_{t_2}. \quad (25)$$

The profiles of the other cell types are expected to be stable:

$$\forall t \neq t_1, t_2, \hat{\rho}_t^{t_1} \stackrel{?}{\simeq} \rho_t, \quad (26)$$

$$\forall t \neq t_1, t_2, \hat{\rho}_t^{t_2} \stackrel{?}{\simeq} \rho_t. \quad (27)$$

To test this idea for each pair of cell types  $[t_1, t_2]$  singled out by the analysis of pairwise similarities, we computed the type-by-type matrices (of size  $T$  by  $T$ , so that they can be plotted as heat maps and compared more easily to the original  $T$  by  $T$  matrix of pairwise correlations between density profiles in the original model:

$$\hat{\Gamma}^{t_1}(t, u) = \mathcal{C}(\hat{\rho}_t^{t_1}, \rho_u) \quad \hat{\Gamma}^{t_1}(u, t) = \mathcal{C}(\rho_u, \hat{\rho}_t^{t_1}), \quad t, u \neq t_1, \quad t, u \neq t_1, \quad (28)$$

$$\hat{\Gamma}^{t_1}(t_2, t_2) = \mathcal{C}(\hat{\rho}_{t_2}^{t_1}, \rho_{t_1} + \rho_{t_2}). \quad (29)$$

$$\hat{\Gamma}^{t_1}(t_1, \cdot) = \hat{\Gamma}^{t_1}(\cdot, t_1) = 0. \quad (30)$$

$$\hat{\Gamma}^{t_2}(t, u) = \mathcal{C}(\hat{\rho}_t^{t_2}, \rho_u) \quad \hat{\Gamma}^{t_2}(u, t) = \mathcal{C}(\rho_u, \hat{\rho}_t^{t_2}), \quad t, u \neq t_2, \quad t, u \neq t_2, \quad (31)$$

$$\hat{\Gamma}^{t_2}(t_1, t_1) = \mathcal{C}(\hat{\rho}_{t_1}^{t_2}, \rho_{t_1} + \rho_{t_2}), \quad (32)$$

$$\hat{\Gamma}^{t_2}(t_2, \cdot) = \hat{\Gamma}^{t_2}(\cdot, t_2) = 0. \quad (33)$$

The matrices  $\hat{\Gamma}^{t_1}$  and  $\hat{\Gamma}^{t_2}$  should have (apart from a “cross” of zeroes at the row and column indices  $t_1$  and  $t_2$  respectively) omitted from the new fitting panel, a diagonal of high correlation coefficients, and off-diagonal terms that should be close to the off-diagonal terms of the type-by-type matrix  $\Gamma$  of correlations between cell types in the original model defined as follows:

$$\Gamma(t, u) = \mathcal{C}(\rho_t, \rho_u), \quad t, u \in [1..T]. \quad (34)$$

The matrix  $\Gamma$  is symmetric and has its diagonal entries all equal 1 by construction.

#### 4.1.2 Results

The pairs of cell types singled out by the above-described analysis are the following:

- (i) indices (15, 16), which are both medium spiny neurons (*Drd1+* and *Drd2+* respectively),
- (ii) indices (4, 5), which are both dopaminergic neurons (A9 and A10 respectively),
- (iii) indices (2, 3), which are both pyramidal neurons (but are not detected by the model except at a few voxels in the olfactory bulb).

- **The pair of medium spiny neurons (labeled by indices  $t_1 = 15$  and  $t_2 = 16$ ).** Having computed the refitted profiles  $\hat{\rho}^{15}$  and  $\hat{\rho}^{16}$ , which contain one medium spiny neurons instead of two, we can plot the density profiles of the remaining medium spiny neurons in each of the refitted models ( $\hat{\rho}_{16}^{15}$  and  $\hat{\rho}_{15}^{16}$ , Figures 38(b) and 38(c) respectively), as well as the sum of the densities of the two samples of medium spiny neurons in the original model (i.e.  $\rho_{15} + \rho_{16}$ , Figure 38(c)). The three plots are hard to distinguish from each other by eye (see Figure 38) and present the same characteristic concentration of signal in the striatum. This supports the conjectures 24 and 24, as each of the medium spiny neurons seems to inherit the signal of the pair labeled by indices 15 and 16 when the model is refitted, without picking up any significant other signal.

Computing the correlation matrices  $\hat{\Gamma}^{15}$  and  $\hat{\Gamma}^{16}$  and reading off the entries

$$\hat{\Gamma}_{16}^{15} = 0.9997, \quad \hat{\Gamma}_{15}^{16} = 0.9992 \quad (35)$$

confirms the visual impression of Figure 38, due to definitions 30 and 33.

Moreover, plotting the matrices  $\hat{\Gamma}^{15}$  and  $\hat{\Gamma}^{16}$  as heat maps (Figures 39b and Figures 39c) shows that the entries are roughly symmetric, and that the two matrices resemble the type-by-type matrix of  $\Gamma$  of pairwise correlations in the original model (Figure 39a), apart from the “cross” of zeroes at the omitted indices (row 15 and column 15 on Figure 39b, row 16 and column 16 on Figure 39c). Moreover, the diagonal coefficients of  $\hat{\Gamma}^{15}$  and  $\hat{\Gamma}^{16}$  are visibly close to 1 (their sorted values are plotted on Figure 40), with the exception of the cholinergic projection neurons (labeled  $t = 11$ ), whose density profile  $\rho_{11}^{15}$  has only a 0.133 correlation coefficient with  $\rho_{11}$ . However, this cell type has very little signal in any of the models ( $\rho_{11}$  has only 7 voxels with positive density,  $\hat{\rho}_{11}^{15}$  has 8 and  $\hat{\rho}_{11}^{16}$  has 7), and

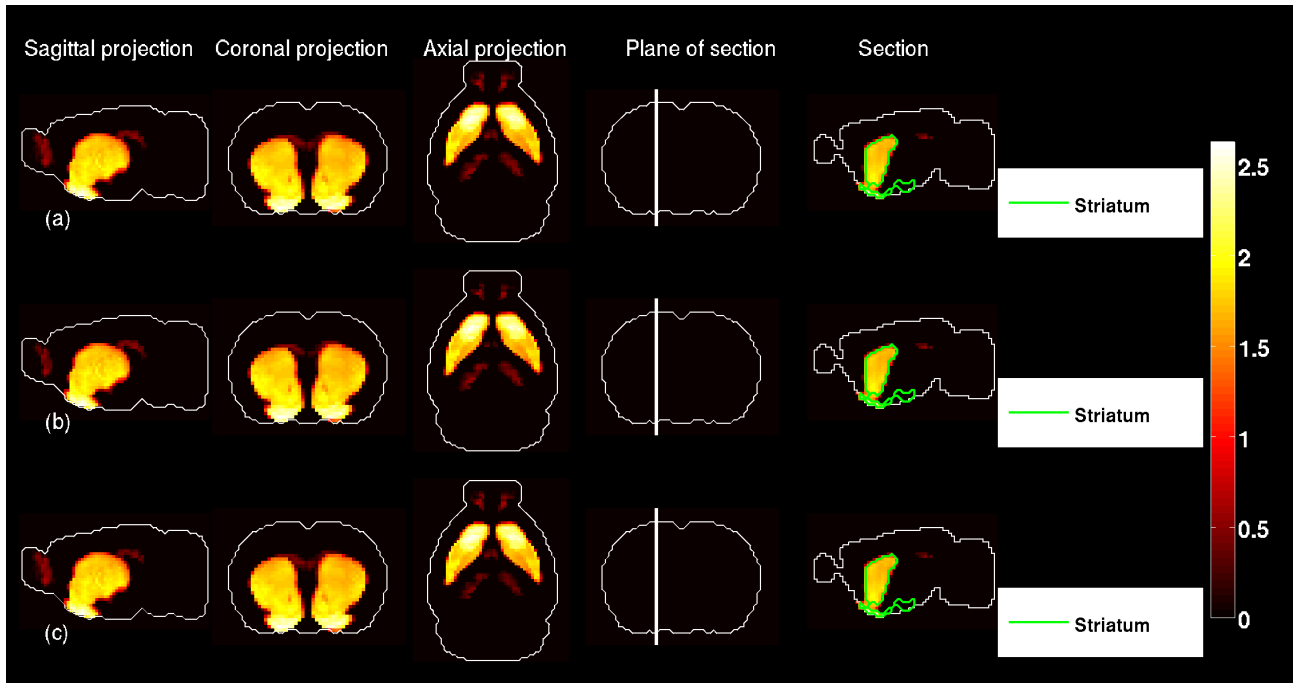


Figure 38: Medium spiny neurons are labeled by indices  $t_1 = 15$  and  $t_2 = 16$  in the original fitting panel. (a) Heat map of the sum of estimated brain-wide densities in the original model,  $\rho_{t_1} + \rho_{t_2}$ . (b) Heat map of  $\hat{\rho}_{t_1}^{t_2}$ , the density of cell type  $t_1 = 15$  when cell type  $t_2 = 16$  is omitted from the fitting panel; it is much closer to the sum  $\rho_{t_1} + \rho_{t_2}$  than to the rather sparse profile  $\rho_{t_1}$ . (c) Heat map of  $\hat{\rho}_{t_2}^{t_1}$ , the density of cell type  $t_2 = 16$  when cell type  $t_1 = 15$  is omitted from the fitting panel. The three profiles are very close to each other by eye, consistently with the conjectures of Equations 24 and 24.

these three profiles all represent less than  $10^{-8}$  times the sum of all the densities in the respective models:

$$\frac{\sum_v \rho_{11}(v)}{\sum_t \sum_v \rho_{11}(v)} = 1.53 \times 10^{-9}, \quad \frac{\sum_v \hat{\rho}_{11}^{15}(v)}{\sum_t \sum_v \hat{\rho}_{11}^{15}(v)} = 6.65 \times 10^{-9}, \quad \frac{\sum_v \hat{\rho}_{11}^{16}(v)}{\sum_t \sum_v \hat{\rho}_{11}^{16}(v)} = 1.53 \times 10^{-9}. \quad (36)$$

The low value of the correlation for cell-type index 11 is therefore compatible with the claim that density profiles are generally stable when replacing the pair of medium spiny neurons by just one of them, and that the remaining medium spiny neuron in the fitting panel inherits the sum of the density profiles predicted in the original model.

- **The pair (A9 dopaminergic neurons, A10 dopaminergic neurons) labeled by indices  $t_1 = 4$  and  $t_2 = 5$ .** We computed the refitted profiles  $\hat{\rho}^4$  and  $\hat{\rho}^5$ , which contain one sample of dopaminergic neurons instead of two. Again we can plot the density profiles of the remaining dopaminergic neuron in each of the refitted models ( $\hat{\rho}_5^4$  and  $\hat{\rho}_4^5$ , Figures 41(b) and 41(c) respectively), as well as the sum of the densities of A9 dopaminergic neurons and A10 dopaminergic neurons in the original model (i.e.  $\rho_4 + \rho_5$ , Figure 41(c)). The three plots show the highest fraction of their signal in the midbrain (hence the sections through midbrain in all three rows of Figure 41). Generally speaking they look quite similar. Again this supports the conjectures 24 and 24, in the case of  $t_1 = 4$  and  $t_2 = 5$ .

Computing the correlation matrices  $\hat{\Gamma}^4$  and  $\hat{\Gamma}^5$  and reading off the entries The correlations conjectured in Equations 30 and 33 are less close to 1 than in the case of medium spiny neurons, but still

$$\hat{\Gamma}_5^4 = 0.87, \quad \hat{\Gamma}_4^5 = 0.96, \quad (37)$$

and one can check that in each of the two refitted models, the sum of profiles  $\rho_4 + \rho_5$  has higher correlation with the remaining dopaminergic neurons than any cell type in the original fitting panel:

$$\hat{\Gamma}^{t_1}(t_2, t_2) = \max_{1 \leq t \leq T} \hat{\Gamma}^{t_1}(t_2, t), \quad \hat{\Gamma}^{t_2}(t_1, t_1) = \max_{1 \leq t \leq T} \hat{\Gamma}^{t_2}(t_1, t). \quad (38)$$

As in the case of medium spiny neurons, plotting the matrices  $\hat{\Gamma}^4$  and  $\hat{\Gamma}^5$  as heat maps (Figures 42b and Figures 42c) shows that the entries are roughly symmetric, and that the two matrices resemble the type-by-type matrix of  $\Gamma$  of pairwise correlations in the original model (Figure 42a), apart from the ‘‘crosses’’ of zeroes (at row 4 and column 15 on Figure 42b, row 16 and column 16 on Figure 42c). Moreover, the diagonal coefficients of  $\hat{\Gamma}^4$  and  $\hat{\Gamma}^5$  are all close to 1 (their sorted values are plotted on Figure 43). Again we conclude that refitting the model to a fitting panel containing one dopaminergic neuron instead of two leads to stable density profile, with the remaining dopaminergic neuron inheriting most of the sum of the two original density profiles  $\rho_4$  and  $\rho_5$ .

- **The pair of pyramidal neurons labeled by indices  $t_1 = 3$  and  $t_2 = 3$ .** We computed the refitted densities  $\hat{\rho}^2$  and  $\hat{\rho}^3$  and repeated the above analysis. Figures 45, 46 and 44 confirm conjectures 24 and 24. However, they are included mostly for completeness, as

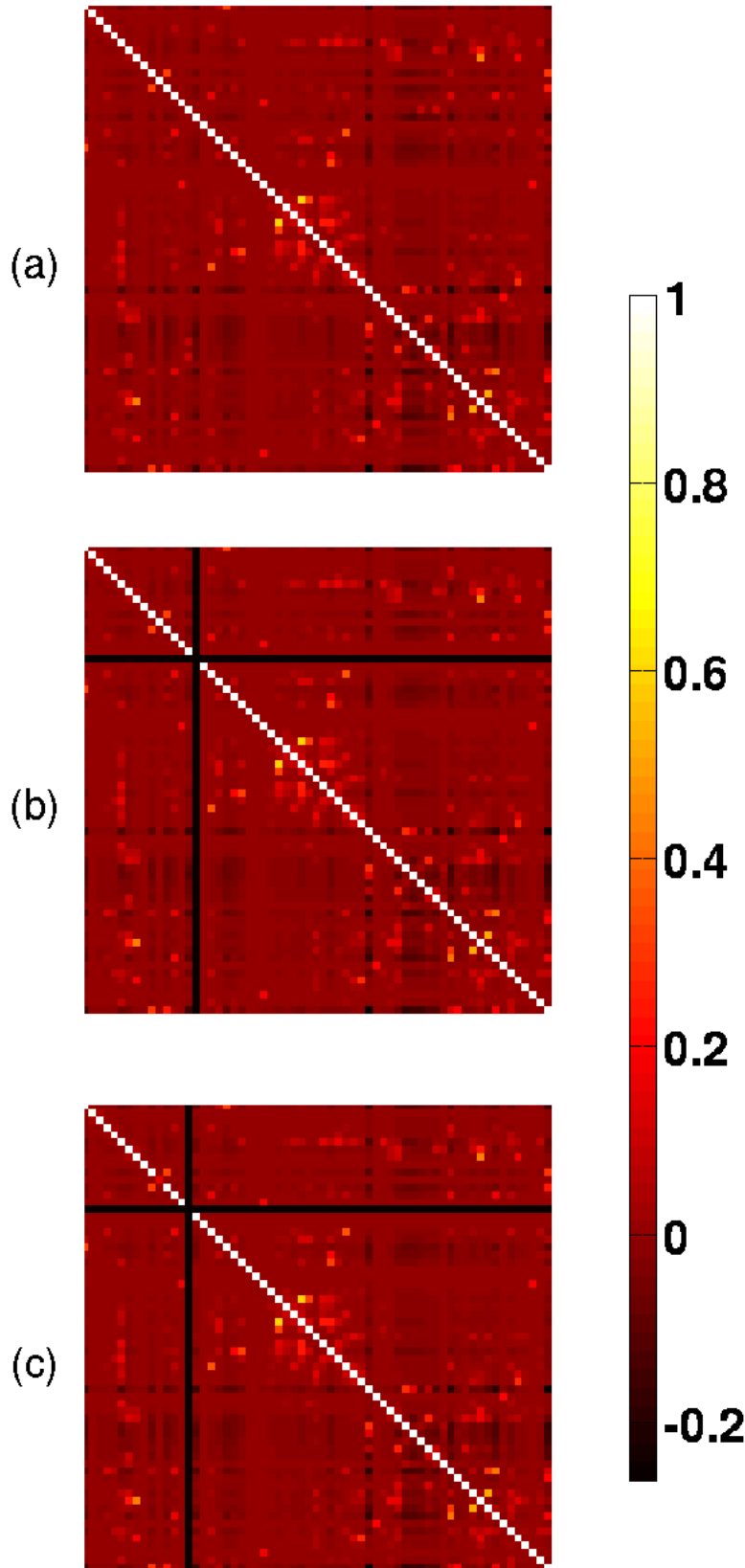


Figure 39: **Heat maps of the type-by-type matrices of pairwise correlations between density profiles.** (a) Using density profiles in the original model, or  $\Gamma$ , Equation 34. (b) Using density profiles  $\hat{\rho}^{15}$  from the model using only medium spiny neuron labeled 16, and densities from the original model, except at column 16 where the sum  $\rho^{15} + \rho^{16}$  is used (see Equation 30). (c) Using density profiles  $\hat{\rho}^{16}$  from the model using only medium spiny neuron labeled 15, and densities from the original model, except at column 15 where the sum  $\rho^{15} + \rho^{16}$  is used (see Equation 33). The three matrices look similar (apart from the zeroes at row and column indices 15 and 16 on (b) and (c) respectively). This suggests that the results of the model are stable upon using one sample of medium spiny neuron instead of two.

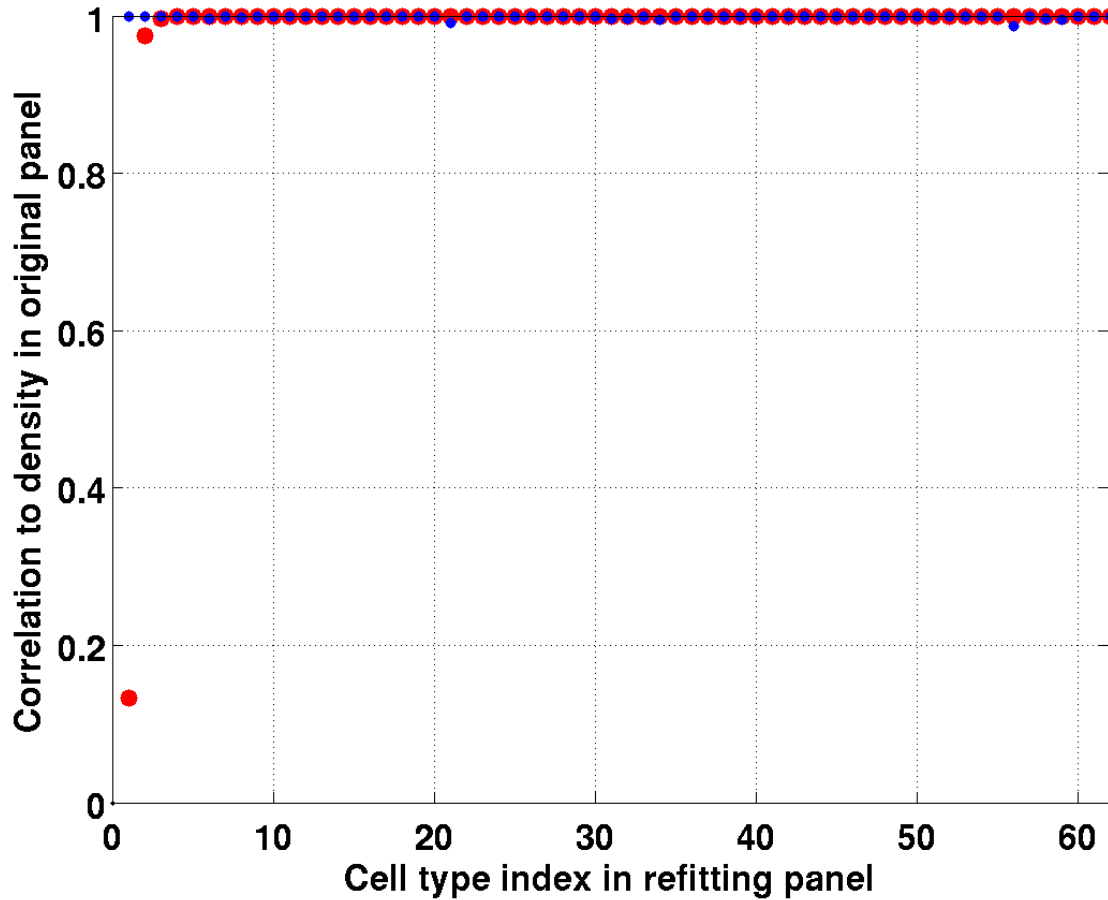


Figure 40: Sorted diagonal elements of the type-by-type correlation matrix  $\hat{\Gamma}^{15}$  (in red), and  $\hat{\Gamma}^{16}$  (in blue, plotted in the same order as the diagonal coefficients of  $\hat{\Gamma}^{15}$ ). All diagonal entries of  $\hat{\Gamma}^{15}$  and  $\hat{\Gamma}^{16}$  are above 0.97, except the entry of  $\hat{\Gamma}^{15}$  corresponding to index 11 (cholinergic projection neurons), but this cell type has very low density in all three models  $\rho$ ,  $r\hat{h}\rho^{15}$ ,  $\hat{\rho}^{16}$ .

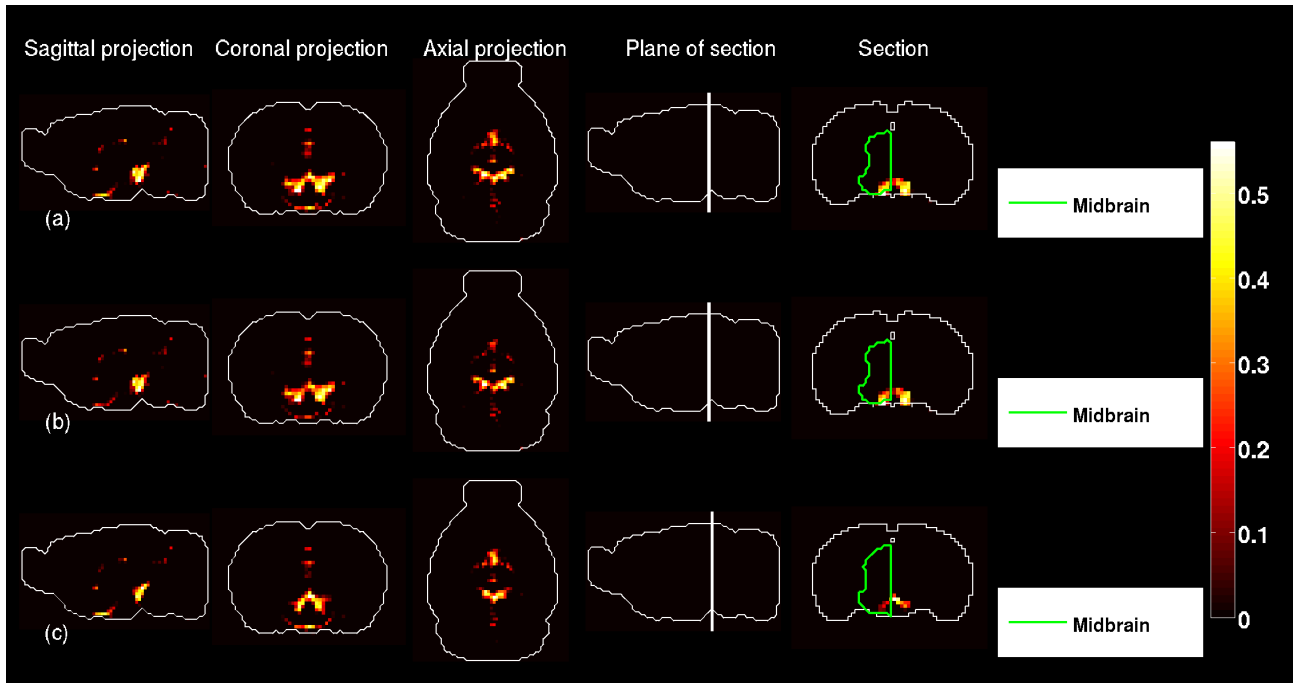


Figure 41: **A9 dopaminergic neurons and A10 dopaminergic neurons are labeled by indices  $t_1 = 4$  and  $t_2 = 5$  in the original fitting panel.** (a) Heat map of the sum of estimated brain-wide densities in the original model,  $\rho_{t_1} + \rho_{t_2}$ . (b) Heat map of  $\hat{\rho}_{t_1}^{t_2}$ , the density of cell type  $t_1 = 4$  when cell type  $t_2 = 5$  is omitted from the fitting panel. (c) Heat map of  $\hat{\rho}_{t_2}^{t_1}$ , the density of cell type  $t_2 = 5$  when cell type  $t_1 = 4$  is omitted from the fitting panel. The three profiles are very close to each other by eye, consistently with the conjectures of Equations 24 and 24.

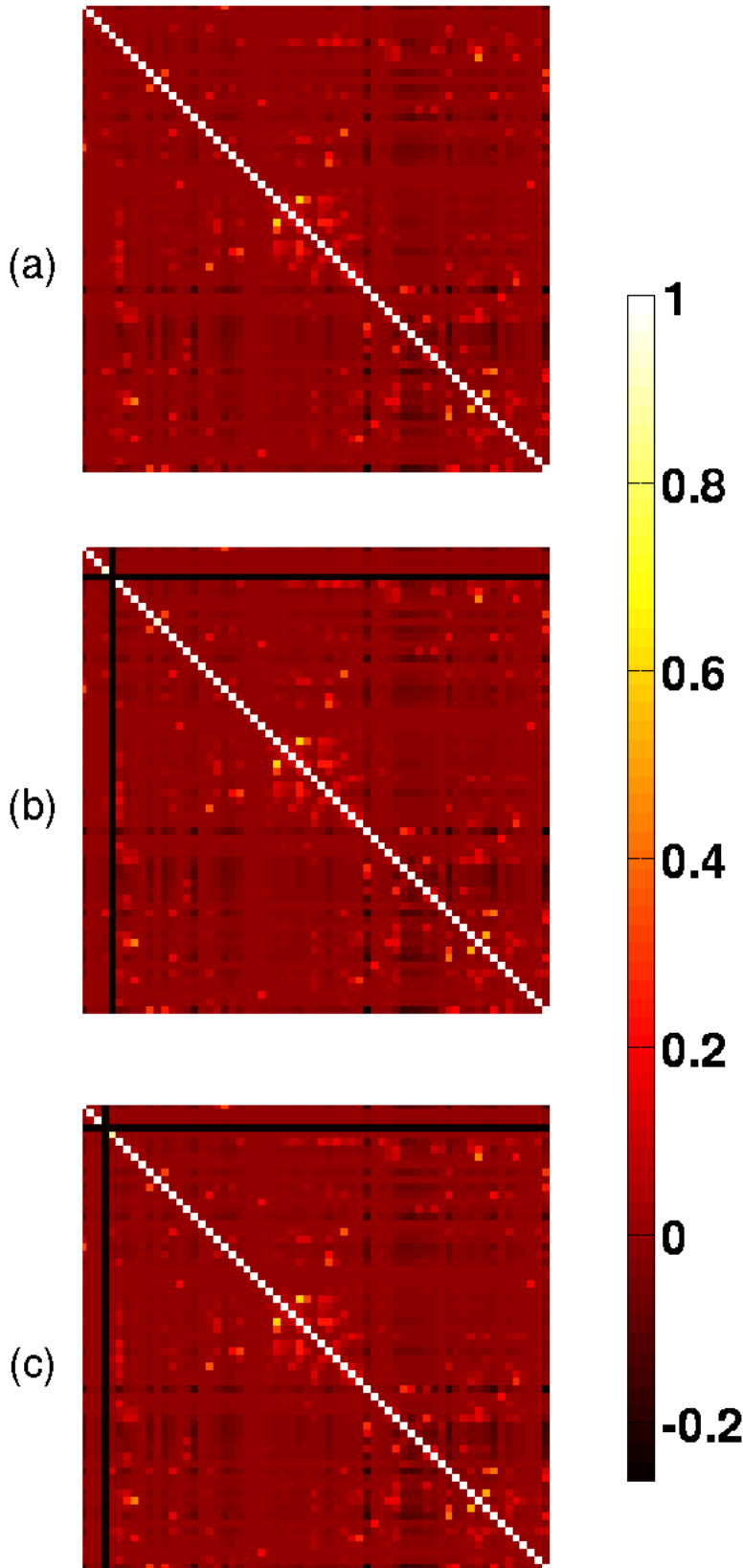


Figure 42: **Heat maps of the type-by-type matrices of pairwise correlations between density profiles, for  $t_1 = 4$  and  $t_2 = 5$ .** (a) The matrix  $\Gamma$  computed using density profiles in the original model, or  $\Gamma$ , Equation 34. (b) Using density profiles  $\hat{\rho}^4$  from the model using only A10 dopaminergic neurons, labeled  $t_2 = 5$ , and densities from the original model, except at column 5 where the sum  $\rho^4 + \rho^5$  is used (see Equation 30). (c) Using density profiles  $\hat{\rho}^5$  from the model using only A10 dopaminergic neurons, labeled  $t_1 = 4$ , and densities from the original model, except at column 4 where the sum  $\rho^4 + \rho^5$  is used (see Equation 33). The three matrices look similar (apart from the zeroes at row and column indices 4 and 5 on (b) and (c) respectively). This suggests that the results of the model are stable upon using either sample of dopaminergic neurons, instead of both of them.



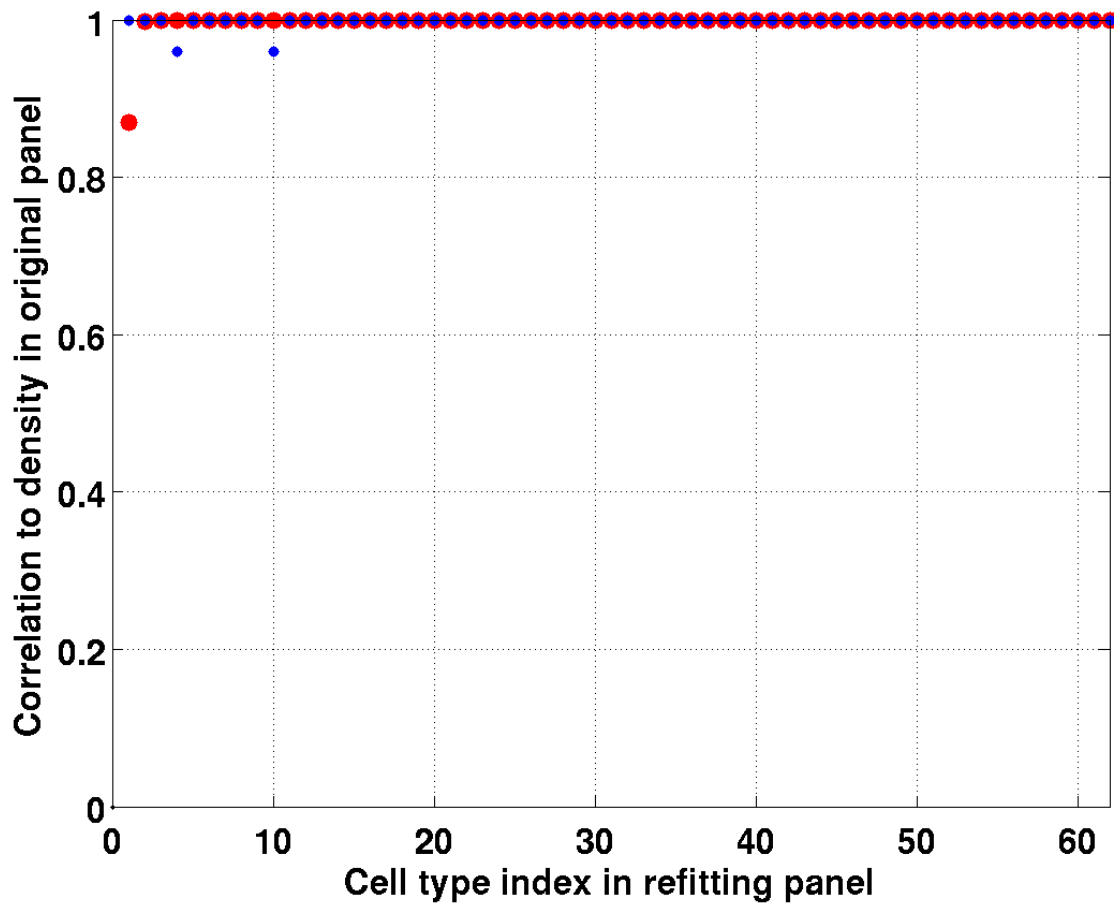


Figure 43: Sorted diagonal elements of the type-by-type correlation matrix  $\hat{\Gamma}^4$  (in red), and  $\hat{\Gamma}^5$  (in blue, plotted in the same order as the diagonal coefficients of  $\hat{\Gamma}^4$ ).

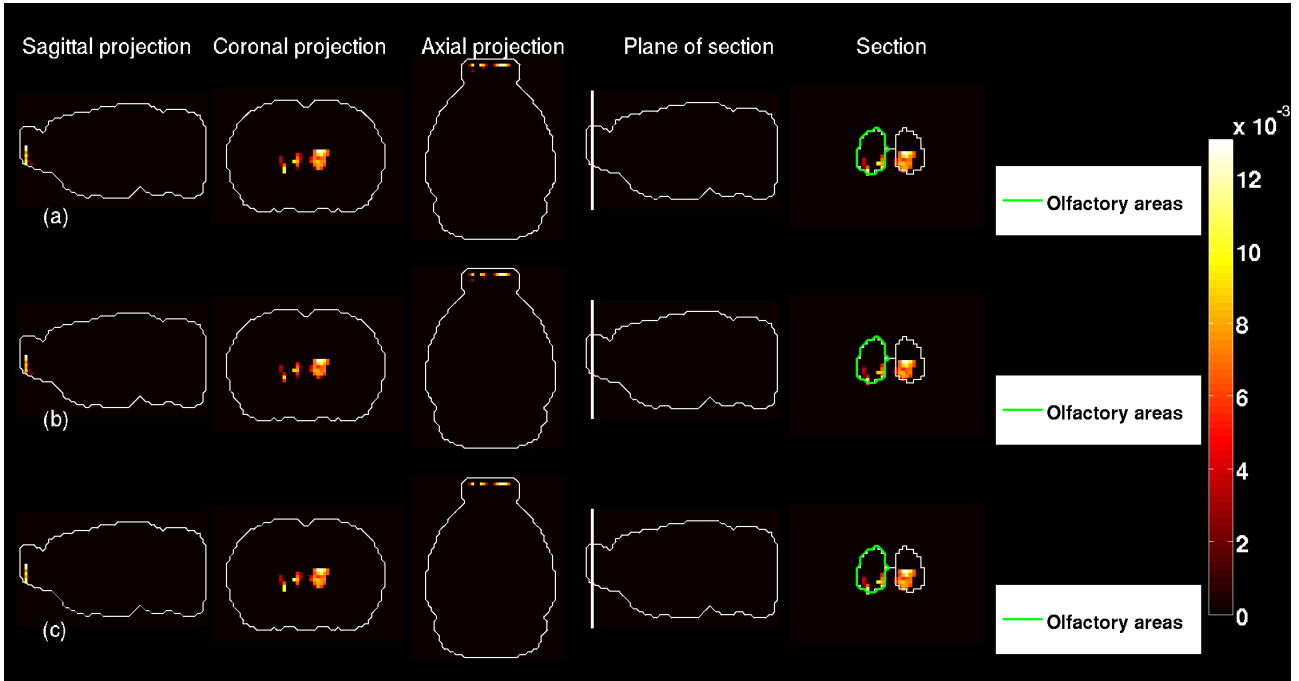


Figure 44: **The pair of pyramidal neurons labeled by indices  $t_1 = 2$  and  $t_2 = 3$  in the original fitting panel.** (a) Heat map of the sum of estimated brain-wide densities in the original model,  $\rho_{t_1} + \rho_{t_2}$ . (b) Heat map of  $\hat{\rho}_{t_1}^{t_2}$ , the density of cell type  $t_1 = 2$  when cell type  $t_2 = 3$  is omitted from the fitting panel; it is much closer to the sum  $\rho_{t_1} + \rho_{t_2}$  than to the rather sparse profile  $\rho_{t_1}$ . (c) Heat map of  $\hat{\rho}_{t_2}^{t_1}$ , the density of cell type  $t_2 = 3$  when cell type  $t_1 = 2$  is omitted from the fitting panel. The three profiles are very close to each other by eye, consistently with the conjectures of Equations 24 and 24.

the estimated densities of these two pyramidal neurons are very low in the original model (only 52 voxels, about 1 in 1000, have a positive density in the sum  $\rho_2 + \rho_3$ ). They stay low in the refitted model ( $\hat{\rho}_2^3$  has only 49 voxels with non-zero density and  $\hat{\rho}_3^2$  only 44). The high diagonal correlation coefficients of Figure 46 show that the rest of the estimated densities hardly change when refitting.

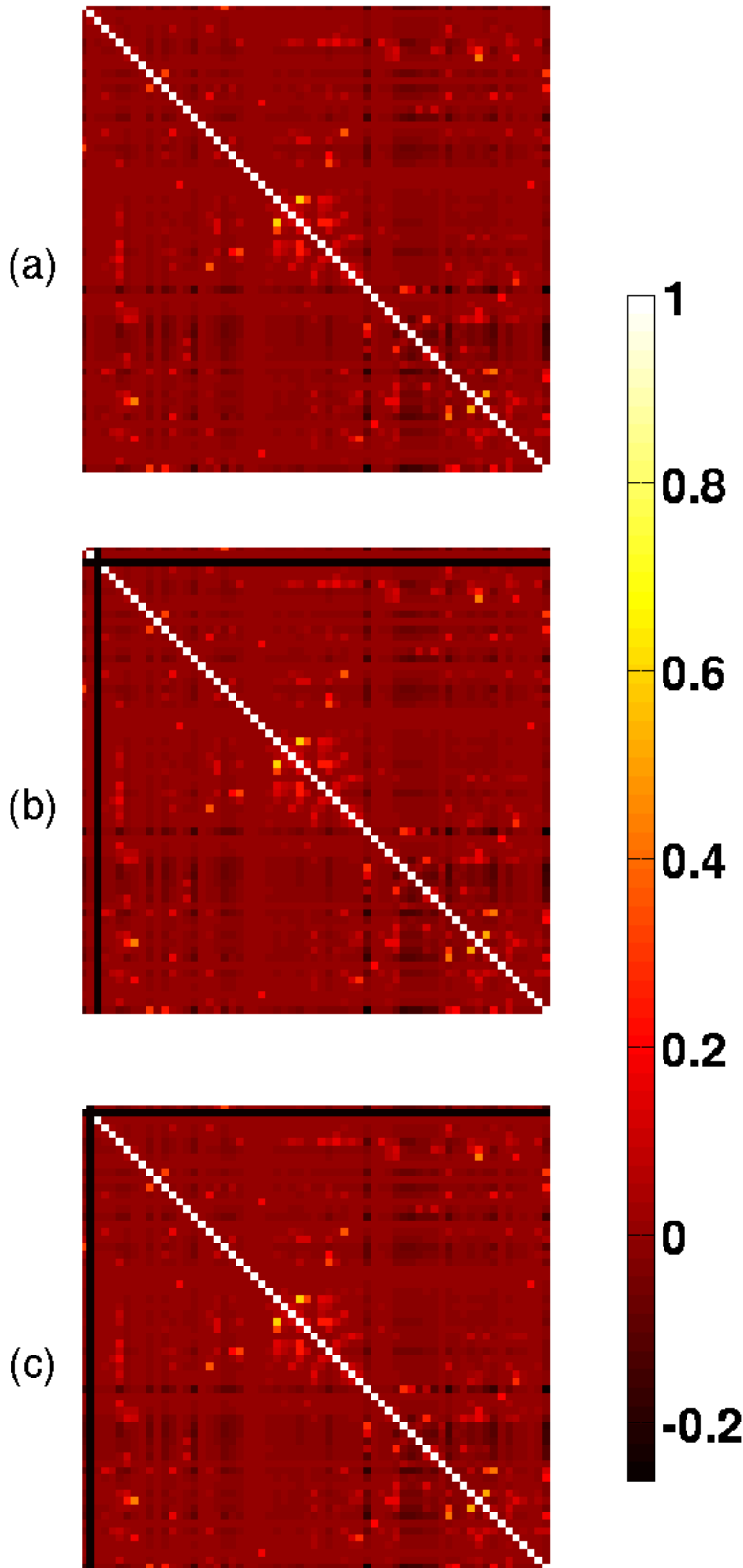


Figure 45: **Heat maps of the type-by-type matrices of pairwise correlations between density profiles, for  $t_1 = 2$  and  $t_2 = 3$ .** (a) The matrix  $\Gamma$  computed using density profiles in the original model, or  $\Gamma$ , Equation 34. (b) Using density profiles  $\hat{\rho}^2$  from the model using only the cell type labeled  $t_2$  (instead of  $t_1$  and  $t_2$ ), and densities from the original model, except at column 3 where the sum  $\rho^2 + \rho^3$  is used (see Equation 30). (c) Using density profiles  $\hat{\rho}^3$  from the model using only the cell type labeled  $t_1 = 2$  (instead of  $t_1$  and  $t_2$ ), and densities from the original model, except at column 2 where the sum  $\rho^2 + \rho^3$  is used (see Equation 33). The three matrices look similar (apart from the zeroes at row and column indices 2 and 3 on (b) and (c) respectively). This suggests that the results of the model are stable upon using either of the cell types labels (2, 3), instead of both of them.

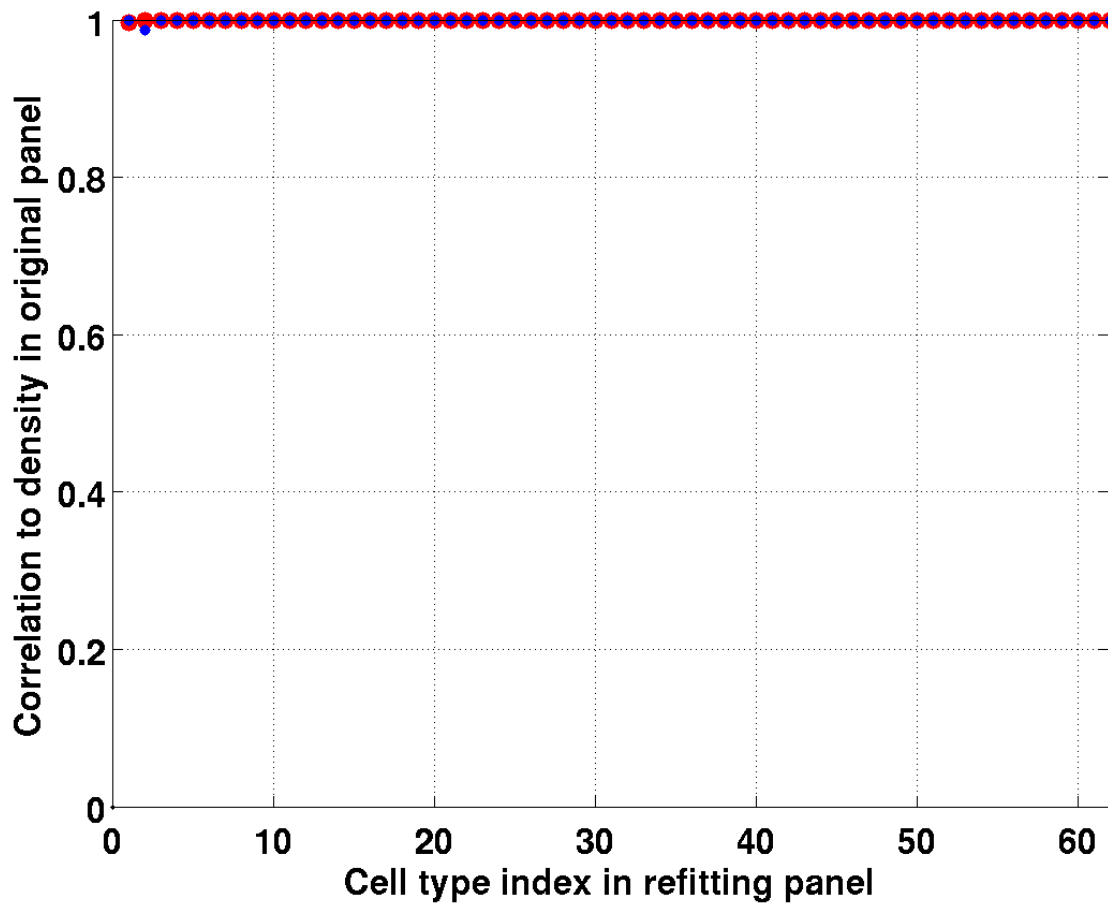


Figure 46: Sorted diagonal elements of the type-by-type correlation matrix  $\hat{\Gamma}^2$  (in red), and  $\hat{\Gamma}^3$  (in blue, plotted in the same order as the diagonal coefficients of  $\hat{\Gamma}^2$ ).

## 4.2 The set of pyramidal neurons

The previous subsection dealt with pairs of indices identified computationally for their strong similarity to each other, as well as dissimilarity with the rest of the fitting panel (they happened to have similar names in the taxonomy of cell types). In this section we refit the model based on a choice made using the names of the cell-type-specific samples: we group together the largest set of cell types with similar names (pyramidal neurons), and replace them with a composite cell type made of the average transcription profiles of all pyramidal neurons in the data set.

We combined all the  $T_{pyr} = 18$  transcriptomic profiles of pyramidal neurons into their average:

$$C^{pyr}(g) = \frac{1}{T_{pyr}} \sum_{i=1}^{T_{pyr}} C(t_i, g), \quad (39)$$

where  $t_i$  are the indices of the cell types labeled as pyramidal neurons. We refitted the model to a set of  $T' = 47$  cell types consisting of this composite pyramidal cell type with the data of the non-pyramidal cell types concatenated with the data in  $C^{pyr}$ .

More precisely, let us rewrite the new fitting panel as a matrix  $C^{new}$  of size  $T - T_{pyr} + 1$  by concatenating  $C^{pyr}$  and the rows of  $C$  corresponding to non-pyramidal indices. Let  $n_1, \dots, n_{T'}$  be the indices of the non-pyramidal cell types (row indices of  $C$ ), where  $T' = T - T_{pyr}$  is the number of non-pyramidal cell types (i.e. for all  $i$  in  $[1..T']$ ,  $C(n_i, \cdot)$  is the transcription profile of a non-pyramidal cell):

$$C^{new}(1, \cdot) = C^{pyr}(g) \quad (40)$$

$$\forall i \in [1..T'], C^{new}(i + 1, \cdot) = C(n_i, \cdot). \quad (41)$$

The brain-wide density profile is the refitted profile, denoted by  $\rho^{new}$ , are the solutions of the usual optimization problems at each voxel:

$$\forall v \in [1..V] \quad (\rho_t^{new}(v))_{1 \leq t \leq T'} = \operatorname{argmin}_{\nu \in \mathbf{R}_+^{T'+1}} \sum_{g=1}^G \left( E(v, g) - \sum_{t=1}^{T'+1} C^{new}(t, g) \nu(t) \right)^2. \quad (42)$$

There are fewer degrees of freedom in the optimization problem 42 than in the original model, so the fitting is expected to be less accurate, but one can ask how correlated the density  $\rho_1^{new}$  of the composite pyramidal neuron is to the sum of the density profiles of all the pyramidal neurons in the original model.

Let us denote by  $\rho^{pyr}$  the sum of the density profiles of all the pyramidal cells in the original model:

$$\rho^{pyr} = \sum_{i=1}^{T_{pyr}} \rho_{t_i}. \quad (43)$$

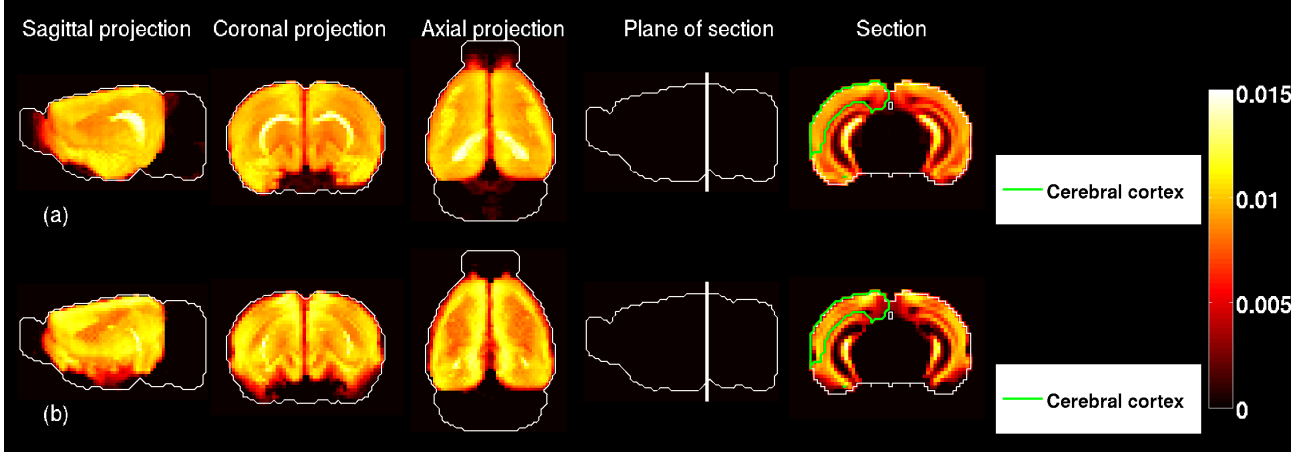


Figure 47: (a) Heat map of  $\rho^{\text{PYR}}$ , the sum of brain-wide densities of pyramidal neurons in the original model. (b) Heat map of the profile  $\rho_1^{\text{new}}$ , the estimated brain-wide density of the composite pyramidal cell defined in Equation 39. The correlation between these two brain-wide profiles is  $\Gamma^{\text{new}}(1, 1) = 0.8875$ .

The natural matrix of correlations  $\Gamma^{\text{new}}$  to compute is the following:

$$\Gamma^{\text{new}}(t, u) = \mathcal{C}(\rho_t^{\text{new}}, \rho_{n_u}) \quad \forall t > 1, \quad \forall u > 1, \quad (44)$$

$$\Gamma^{\text{new}}(1, 1) = \mathcal{C}(\rho_1^{\text{new}}, \rho^{\text{PYR}}), \quad (45)$$

$$\Gamma^{\text{new}}(1, u) = \mathcal{C}(\rho_1^{\text{new}}, \rho_{n_u}) \quad \forall u > 1, \quad (46)$$

$$\Gamma^{\text{new}}(u, 1) = \mathcal{C}(\rho_u^{\text{new}}, \rho^{\text{PYR}}) \quad \forall u > 1, \quad (47)$$

which is expected to be close to the following  $T' + 1$  by  $T' + 1$  matrix based on the profiles  $\rho$  of the original model only:

$$\Gamma^{\text{old}}(1, 1) = \mathcal{C}(\rho^{\text{PYR}}, \rho^{\text{PYR}}) \quad (48)$$

$$\forall i, j \in [1..T'] \quad \Gamma(i + 1, j + 1) = \mathcal{C}(\rho_{n_i}, \rho_{n_j}), \quad (49)$$

$$\Gamma^{\text{old}}(1, i + 1) = \Gamma^{\text{old}}(i + 1, 1) = \mathcal{C}(\rho^{\text{PYR}}, \rho_{n_i}). \quad (50)$$

The correlation  $\Gamma^{\text{new}}(1, 1) = 0.8875$  is indeed close to 1, moreover it is the maximum entry in the first row (the next-highest value in the first row is 0.3621) and in the first column of  $\Gamma^{\text{new}}$  (the next-highest value in the first column is -0.0033):

$$\operatorname{argmax}_{i \in [1..T'+1]} \Gamma^{\text{new}}(i, 1) = 1, \quad \operatorname{argmax}_{i \in [1..T'+1]} \Gamma^{\text{new}}(1, i) = 1. \quad (51)$$

The plots of  $\rho^{\text{PYR}}$  and  $\rho_1^{\text{new}}$  show that indeed the two profiles are mostly concentrated in the cerebral cortex. The most visible difference comes from the hippocampus, which is more highlighted in  $\rho^{\text{PYR}}$  than in the density  $\rho_1^{\text{new}}$  of the composite pyramidal cell type.

Figures 48 and 49 show some similarity between  $\Gamma^{\text{new}}$  and  $\Gamma^{\text{old}}$ , moreover the diagonal correlations in  $\Gamma^{\text{new}}$  are larger than 0.8 for 35 out of 48 cell types in the new fitting panel defined in Equation 41.

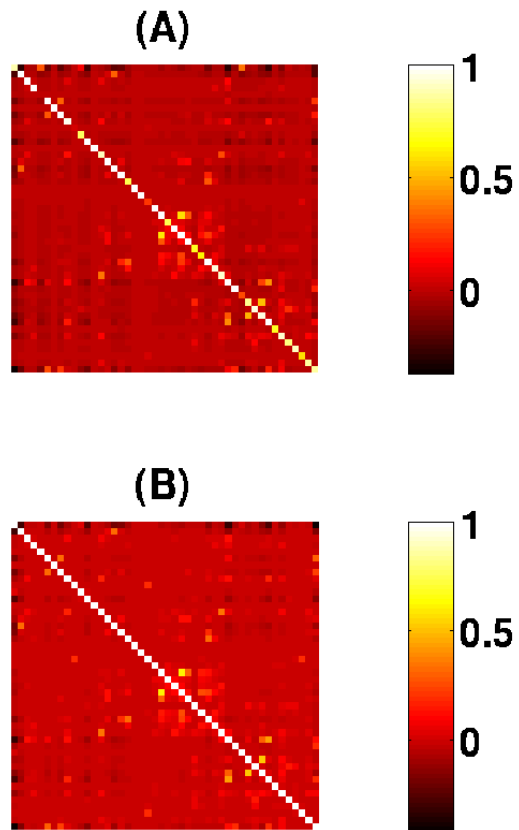


Figure 48: (a) Heat map of the type-by-type correlation matrix  $\Gamma^{\text{new}}$  defined in Equation 47.  
 (b) Heat map of the type-by-type correlation matrix  $\Gamma^{\text{old}}$  defined in Equation 50.

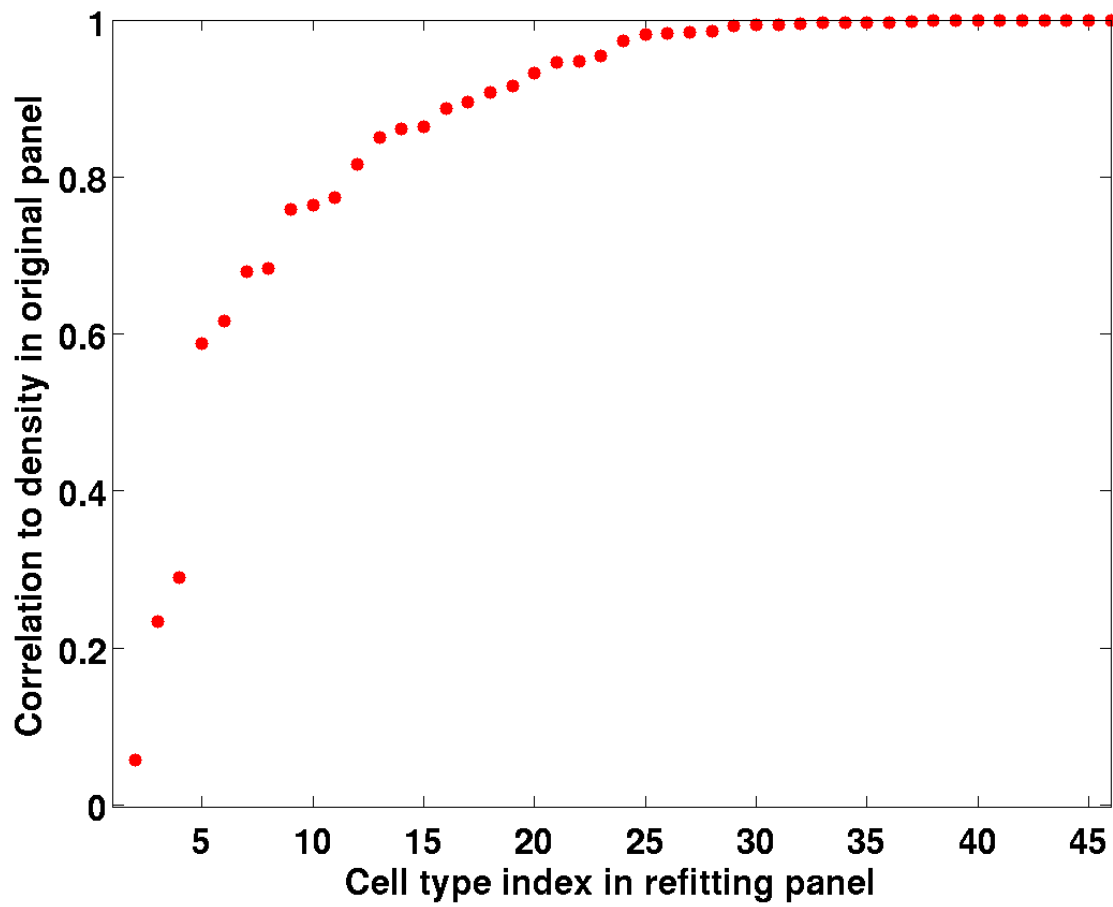


Figure 49: Sorted diagonal elements of the type-by-type correlation matrices  $\Gamma^{\text{new}}$  (a) and  $\Gamma^{\text{old}}$  (b).



### 4.3 Background value in microarray data

It can be deduced by examination of the ISH data set (the voxel-by-gene matrix  $E$  representing the Allen Atlas of the adult mouse brain) and the microarray data set (the type-by-gene matrix  $C$ ) that the background intensities are lower in the Allen Atlas than in the cell-type-specific microarray data. Out of the  $G = 2,131$  genes analyzed in the intersection of the two datasets: 1. no gene has zero signal in any of the cell types, moreover, the minimum entry of the type-by-gene matrix  $C$  is a relatively large fraction of the average value of the average of all the entries:

$$m_C := \min_{1 \leq t \leq T, 1 \leq g \leq G} C(t, g), \quad (52)$$

$$\frac{m_C}{\frac{1}{GT} \sum_{1 \leq t \leq T} \sum_{1 \leq g \leq G} C(t, g)} = 46.5\%. \quad (53)$$

2. Each gene has zero ISH signal in at least 0.38 percent of the voxels, and the average proportion of voxels with zero signal is larger than 4 percent

$$m_E := \min_{1 \leq v \leq V, 1 \leq g \leq G} E(v, g) = 0, \quad (54)$$

$$\frac{1}{VG} \sum_{1 \leq g \leq G} |\{v \in [1..V], E(v, g) = 0\}| = 4.16\%. \quad (55)$$

Let us denote the true expression signal of gene labeled  $g$  in cell type labeled  $t$  by  $C_{\text{true}}(t, g)$ . The cross-hybridization between genes  $g$  and  $g'$  can be modeled by a factor  $\mathcal{T}_{gg'}$  relating  $C$  to the true signal:

$$C(t, g) = \sum_{g'} C_{\text{true}}(t, g') \mathcal{T}_{g'g}. \quad (56)$$

A zero cross-hybridization would correspond to a matrix  $\mathcal{T}$  equal to the identity matrix in gene space. The simplest choice for a non-zero cross-hybridization assumes that a uniform fraction  $\alpha$  of the true signal of each gene is smeared across all other genes:

$$\mathcal{T}_{gg'} := (1 - \alpha) \delta_{gg'} + \alpha U_{gg'}, \quad (57)$$

where uniform cross-hybridization  $U$  is the rank-one matrix whose entries are all equal to one:

$$\forall g, g', U_{gg'} = 1. \quad (58)$$

Equation 56 can be used as follows to express  $C_{\text{true}}$  in terms of  $C$ :

$$C_{\text{true}} = \frac{1}{1 - \alpha} C \left( I_G + \frac{\alpha}{1 - \alpha} U \right)^{-1} = \frac{1}{1 - \alpha} C \sum_{k \geq 0} \left( -\frac{\alpha}{1 - \alpha} \right)^k U^k \quad (59)$$

$$= \frac{1}{1 - \alpha} C \left( I_G + U \sum_{k \geq 1} \left( -\frac{\alpha}{1 - \alpha} \right)^k \right) = \frac{1}{1 - \alpha} C (I_G - \alpha U), \quad (60)$$

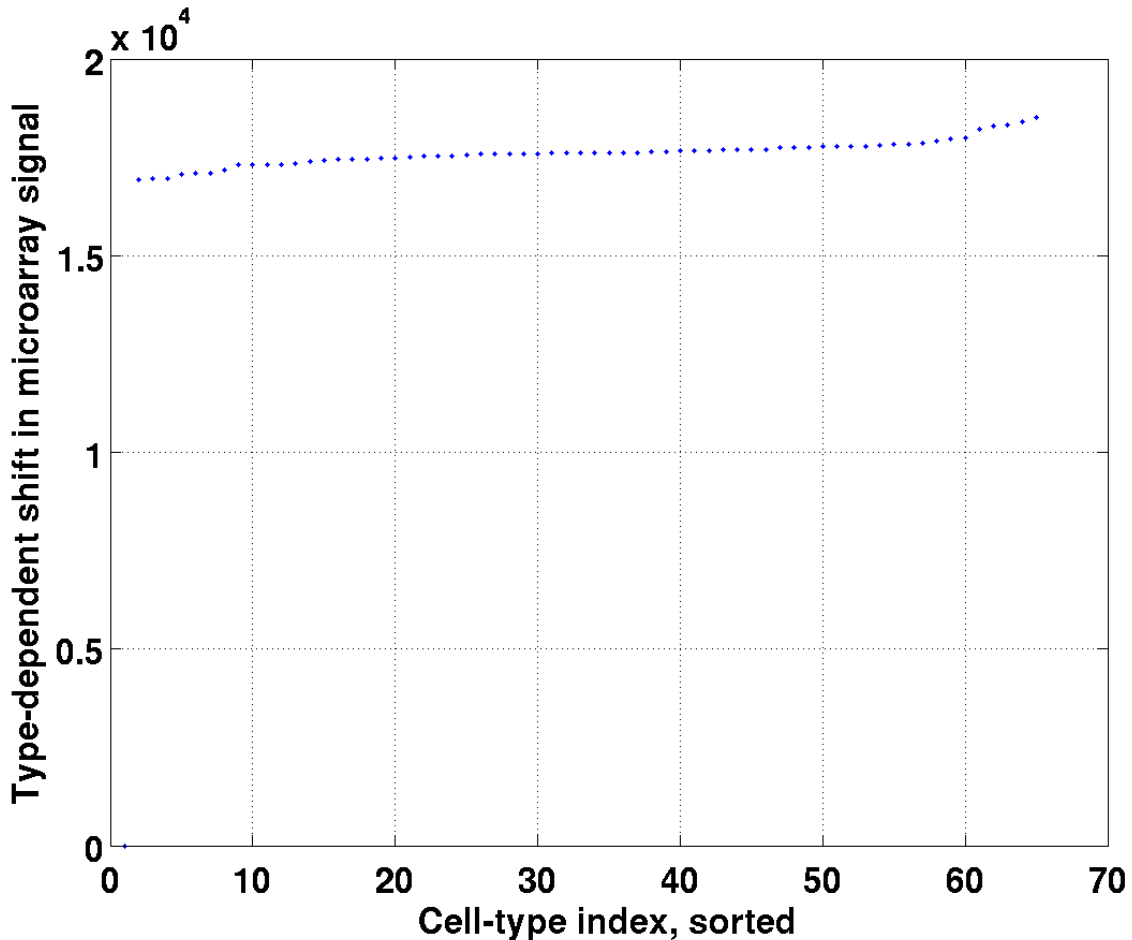


Figure 50: Sorted values of the type-dependent shift term  $\sum_{g'} C(t, g')$  in Equation 61. All the values are within 5.2 percent of the average value across cell types.

where we have used the fact that  $U$  is a projector, and as such has all its positive powers equal to  $U$ . We can therefore rewrite the entries of  $C_{\text{true}}$  (up to a collective multiplicative factor that does not modify the relative values of entries in  $C$ ) as the sum of the corresponding entry in  $C$  and a negative term that does only depend on the type index  $t$ , and not on the gene  $g$ :

$$C_{\text{true}}(t, g) = \frac{1}{1 - \alpha} \left( C(t, g) - \alpha \sum_{g'} C(t, g') \right). \quad (61)$$

Plotting the  $t$ -dependent shift-term in the above equation as a function of  $t$ , one observes (Figure 50) that it is roughly constant across cell types.

Consider the contribution of two cell types labeled  $t_1$  and  $t_2$ , and two genes labeled  $g_1$  and  $g_2$ , such that  $g_1$  is expressed only in  $t_1$  and  $g_2$  is expressed only in  $t_2$ . This situation corresponds to a diagonal submatrix matrix of size 2 in the true signal. Cross-hybridization between  $g_1$  and

$g_2$  adds non-diagonal elements to this matrix, and if they are large enough the determinant of the matrix can become zero, resulting in a degenerate solution to the system  $E(v, \cdot) = C x$ , even though the system  $E(v, \cdot) = C_{\text{true}} x$  is a linear system with positive solutions.

Moreover, suppose a perfect solution  $\rho^{\text{perfect}}$  exists at some voxel labeled  $v$  for the fitting to a true signal:

$$E(v, \cdot) = \sum_{t=1}^T \rho_t^{\text{perfect}}(v) C_{\text{true}}(t, \cdot) + 0. \quad (62)$$

If the model is fitted to a microarray signal that differs from  $C_{\text{true}}$  by a type-by-gene matrix with uniform entries all equal to some value  $c$ , one can write a test-vector  $\phi_t(v) \in \mathbf{R}_+^T$  as the sum of the vector  $\rho_t^{\text{perfect}}(v)$  and a term  $\delta\phi_t(v) \in \mathbf{R}^T$ , and optimize over the values of  $\delta\phi$  (whose optimal value goes to zero when  $c$  goes to 0):

$$\text{Err}_c(\phi) := \sum_{g=1}^G \left( E(v, g) - \sum_{t=1}^T (\phi_t(v) (C_{\text{true}}(t, g) + c)) \right)^2. \quad (63)$$

$$\phi_t(v) := \rho_t^{\text{perfect}}(v) \delta\phi_t(v) \quad (64)$$

$$\text{Err}_c(\phi) = \sum_{g=1}^G \left( \sum_{t=1}^T (C_{\text{true}}(t, g) + c) (\rho_t^{\text{perfect}}(v) + \delta\phi_t(v)) - \sum_{t=1}^T C_{\text{true}}(t, g) \rho_t^{\text{perfect}}(v) \right)^2 \quad (65)$$

$$= \sum_{g=1}^G \left( \sum_{t=1}^T (c \rho_t^{\text{perfect}}(v) + C_{\text{true}}(t, g) \delta\phi_t(v) + c \delta\phi_t(v)) \right)^2 \quad (66)$$

$$(67)$$

At  $c=0$ , each term in the sum over genes is identically zero, but one can see that the dependence on the gene index  $g$  cannot be maintained when  $c$  is increased, hence an increased value of the residual term.

The parameter  $\alpha$  of Equation 57 is not fixed by the previous analysis, and any value can be tried provided it does not give rise to any negative entry in  $C_{\text{true}}$  defined by 61. We can therefore simulate a worst-case scenario, by subtracting the lowest entry of  $C$  from all the other entries (which corresponds to the approximation where the  $t$ -dependence of the second term in 61 is neglected, i.e. where the dots on Figure 50 are considered to be on a horizontal line). We therefore took the following type-by-gene matrix  $C_{\text{offset}}$  as the new fitting panel:

$$C_{\text{offset}}(t, g) := C(t, g) - m_C. \quad (68)$$

We solved the following optimization problem at each voxel, giving rise to a new estimate called

$\rho_{\text{offset}}$  for the brain-wide density of each cell type:

$$(\rho_{\text{offset}_t}(v))_{1 \leq t \leq T} = \operatorname{argmin}_{\nu \in \mathbf{R}_+^T} \sum_{g=1}^G \left( E(v, g) - \sum_{t=1}^T C_{\text{offset}}(t, g) \nu(t) \right)^2. \quad (69)$$

The brain-wide correlation profiles between each cell types and the Allen Atlas do not change under the transformation  $C \rightarrow C_{\text{offset}}$  because the uniform shift in all the entries of the matrix  $C$  is compensated by the shift in the entries of the average row  $\bar{C}$ . For each cell type, labeled  $t$ , we can plot projections and sections of the new estimated profile derived  $\rho_{\text{offset}_t}$  from  $C_{\text{true}}$ , next to the original plot, for visual comparison. The top region by density is the one through which the section is taken, which shows that the top region is conserved **except for cell types labeled ??**.

Moreover, we can compute the following correlation coefficient between the new correlation profile and the one obtained in the original linear model for the same cell type, as defined in Equation 16 for any pair of brain-wide profiles:

$$\mathcal{C}(\rho_t, \rho_{\text{offset}_t}) = \frac{\sum_{v=1}^V (\rho_t(v) - \bar{\rho}_t) (\rho_{\text{offset}_t}(v) - \bar{\rho}_{\text{offset}_t})}{\sqrt{\left( \sum_{v=1}^V (\rho_t(v) - \bar{\rho}_t)^2 \right) \left( \sum_{v=1}^V (\rho_{\text{offset}_t}(v) - \bar{\rho}_{\text{offset}_t})^2 \right)}}, \quad (70)$$

It turns out that 35 of the  $T = 64$  cell types in the data set give rise so correlation values larger than 80 percent. Another correlation can be computed between the fractions of the density binned by the regions of the ARA (defined in Equation 18).

$$\mathcal{C}^{\text{ARA}}(\rho_t, \rho_{\text{offset}_t}) = \frac{\sum_{r=1}^R \left( \bar{\rho}(r, t) - \langle \bar{\rho}(\cdot, t) \rangle \right) \left( \rho_{\text{offset}}(r, t) - \langle \rho_{\text{offset}}(\cdot, t) \rangle \right)}{\sqrt{\left( \sum_{r=1}^R \left( \bar{\rho}(r, t) - \langle \bar{\rho}(\cdot, t) \rangle \right)^2 \right) \left( \sum_{r=1}^R \left( \rho_{\text{offset}}(r, t) - \langle \rho_{\text{offset}}(\cdot, t) \rangle \right)^2 \right)}}, \quad (71)$$

where  $R = 13$  is the number of regions in the coarsest version of the ARA, and the fractions of densities in regions are defined as in Equation 18:

$$\bar{\rho}(r, t) = \frac{1}{|\sum_{v \in \text{Brain Annotation}} \rho_t(v)|} \sum_{v \in V_r} \rho_t(v), \quad (72)$$

$$\bar{\rho}_{\text{offset}}(r, t) = \frac{1}{|\sum_{v \in \text{Brain Annotation}} \rho_{\text{offset}_t}(v)|} \sum_{v \in V_r} \rho_{\text{offset}_t}(v), \quad (73)$$

and used the values to compute the average of residual fraction across regions in the ARA:

$$\langle \bar{\rho}(\cdot, t) \rangle = \frac{1}{R} \sum_{r=1}^R \bar{\rho}(r, t), \quad \langle \rho_{\text{offset}}(\cdot, t) \rangle = \frac{1}{R} \sum_{r=1}^R \bar{\rho}_{\text{offset}}(r, t). \quad (74)$$

The values of  $\mathcal{C}^{\text{ARA}}(\rho_t, \rho_{\text{offset}_t})$  are plotted in Figure 51 (56 of them are larger than 80 percent).

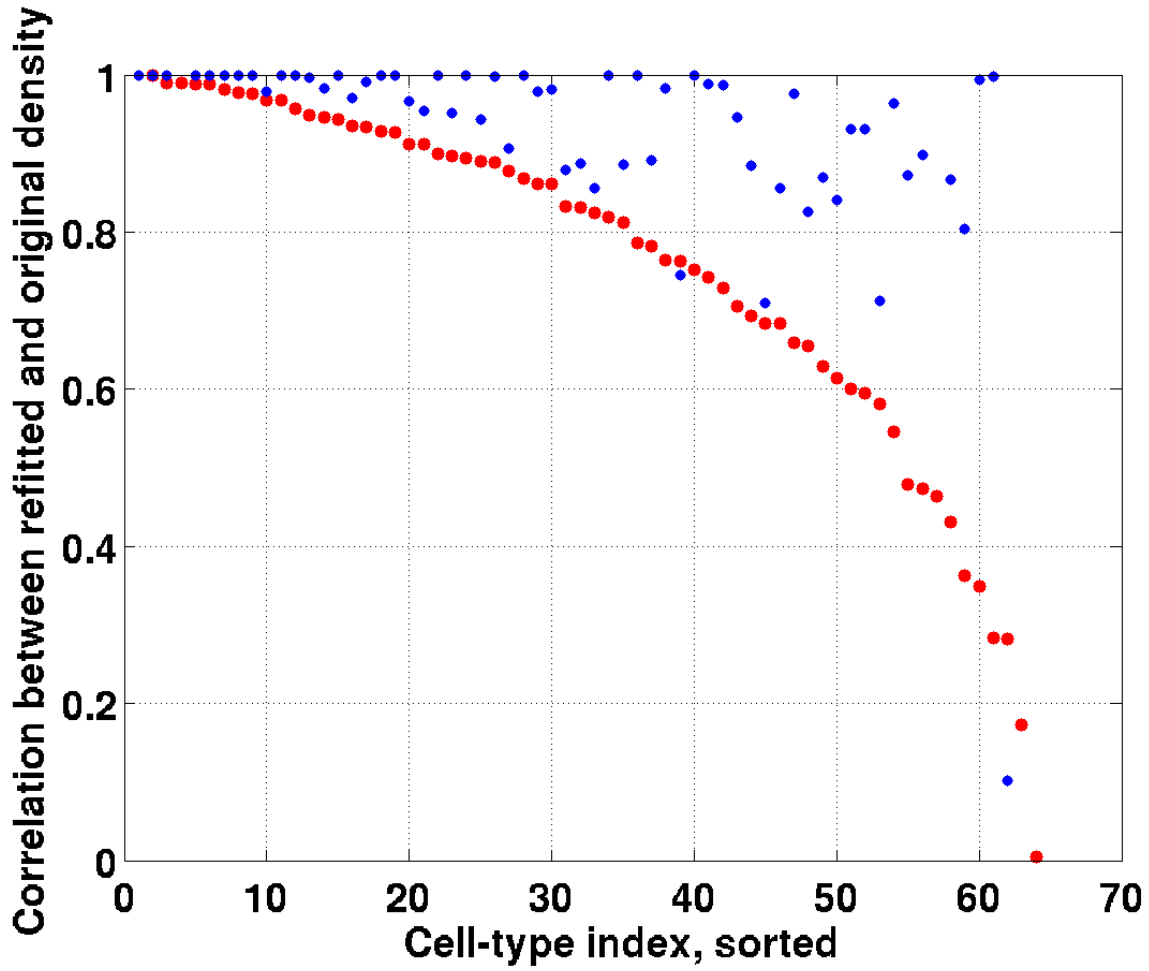


Figure 51: Sorted values of the correlation coefficient  $\mathcal{C}(\rho_t, \rho_{\text{offset}_t})$ , defined in Equation 70 for all values of the cell-type label  $t$  (in red) (see the tables for assignment of each value to a definite cell type, the average value is 75.3%). The correlations between anatomic fractions in the ARA (Equation 71) are plotted in blue (the average value is 92.8%, the cell types are ordered in the same way as for the red dots)

The residual terms can be compared in order to estimate whether the fit between the Allen Atlas and the cell types is improved by subtracting a cross-hybridization term to the microarray data, as it should from the heuristic Equation 67 (even though the offset matrix  $C_{\text{offset}}$  is not expected to give rise to a perfect fit due to the imperfections of our cross-hybridization model, it must be closer to the signal measured by ISH data as it brings some entries of the cell-type-specific matrix closer to zero). We computed the average fraction of the signal that is contained in the absolute value of the residual,

$$\text{ResFracLocal}(v) = \left| 1 - \frac{\sum_t \rho_t(v) C(t, g)}{E(v, g)} \right| \quad (75)$$

across each of the 13 regions in the coarsest partition of the ARA:

$$\text{ResFrac}(r) = \frac{1}{\sum_{v \in V_r} 1} \sum_{v \in V_r} \left| 1 - \frac{\sum_t \rho_t(v) C(t, g)}{E(v, g)} \right| \quad (76)$$

$$\text{ResFrac}_{\text{refit}}(r) = \frac{1}{\sum_{v \in V_r} 1} \sum_{v \in V_r} \left| 1 - \frac{\sum_t \rho_t(v) (C(t, g) - m_C)}{E(v, g)} \right| \quad (77)$$

where  $r$  labels regions in the ARA (see Figure 52).

Indeed the average residual fractions in the refitted model are consistently lower than in the original model fitted to the full microarray data (see Figure 52), which suggests a better correspondence to the ISH data. For a plot of the voxel-by-voxel values, see Figure 75. The average value across the brain goes down when going from the original model to the refitted model, and the highest values are localized in the same voxels in both models. The high values in the most rostral voxels can be attributed to the loss of sections from the olfactory bulbs in the ISH data, as these sections are the most fragile. The other cluster of voxels with high residual, which is much more caudal, could be related to registration errors.

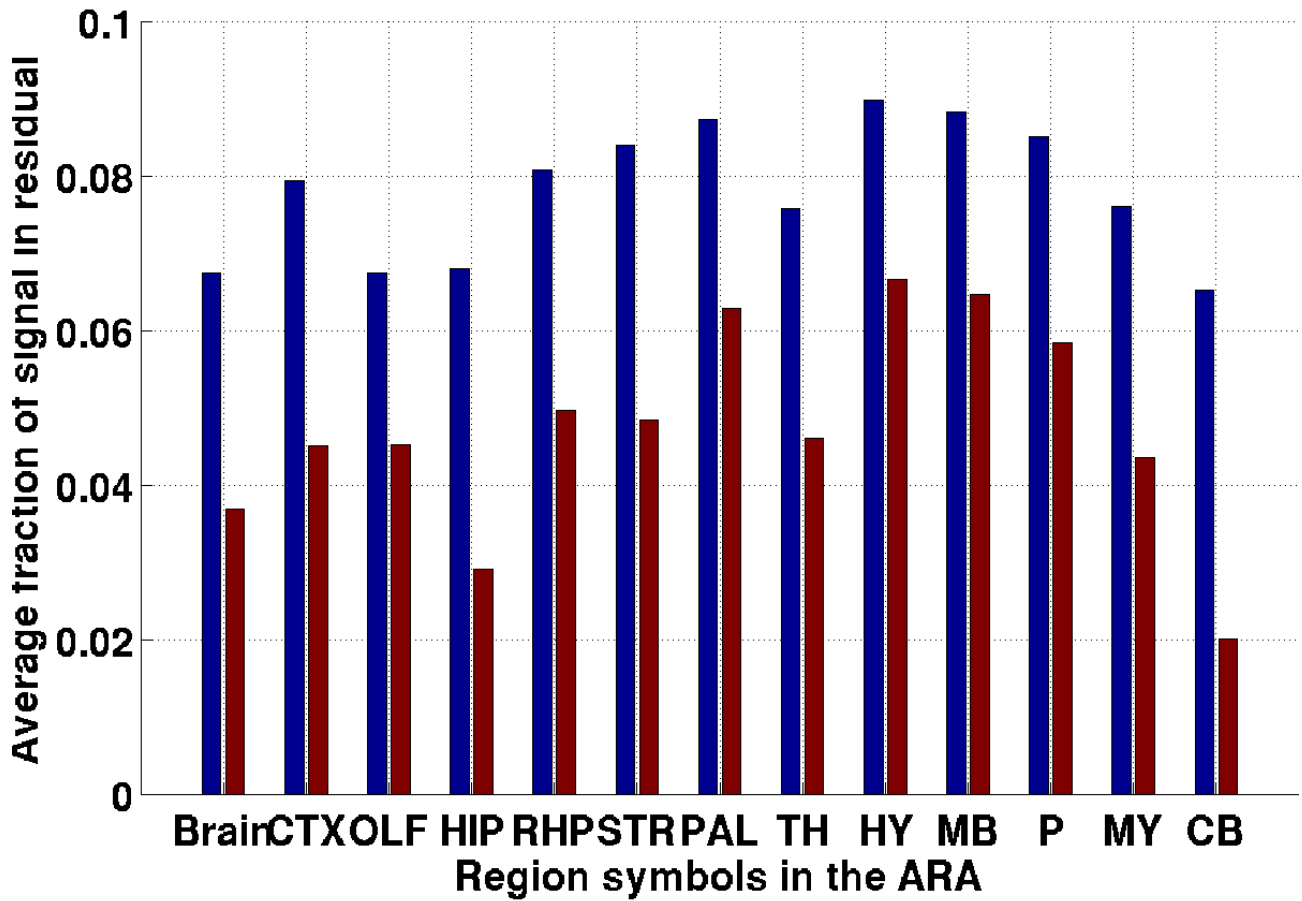


Figure 52: Fraction of absolute values of the residual averaged over the regions in the coarsest version of the ARA (for the original model in blue, and for the refitted model in red).

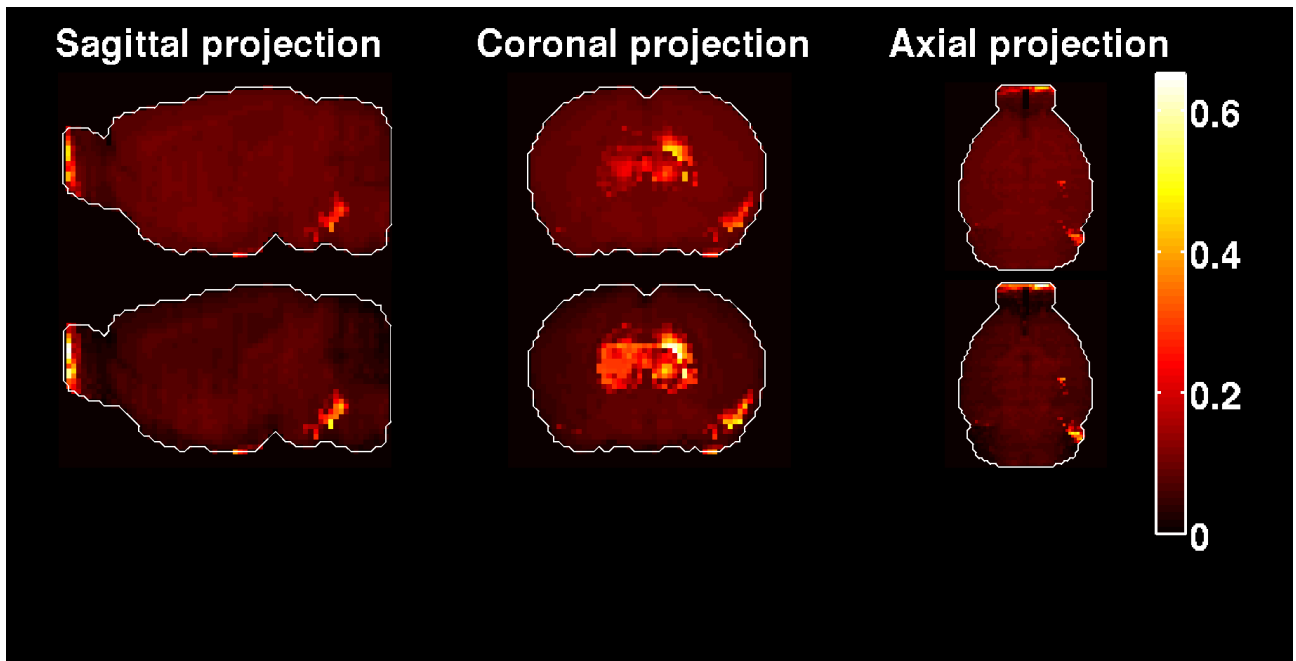


Figure 53: Maximal-intensity projections of the absolute values of the residual at each voxel (original model in the first row, refitted model in the second row). The maximum values are close to the boundary of the brain in both cases. The average value is (for the original model in the first row, average value 7.8%, for the refitted model in the second row, average value 4.4%).



## 5 Tables of plots of estimated densities, original and corrected for cross-hybridization

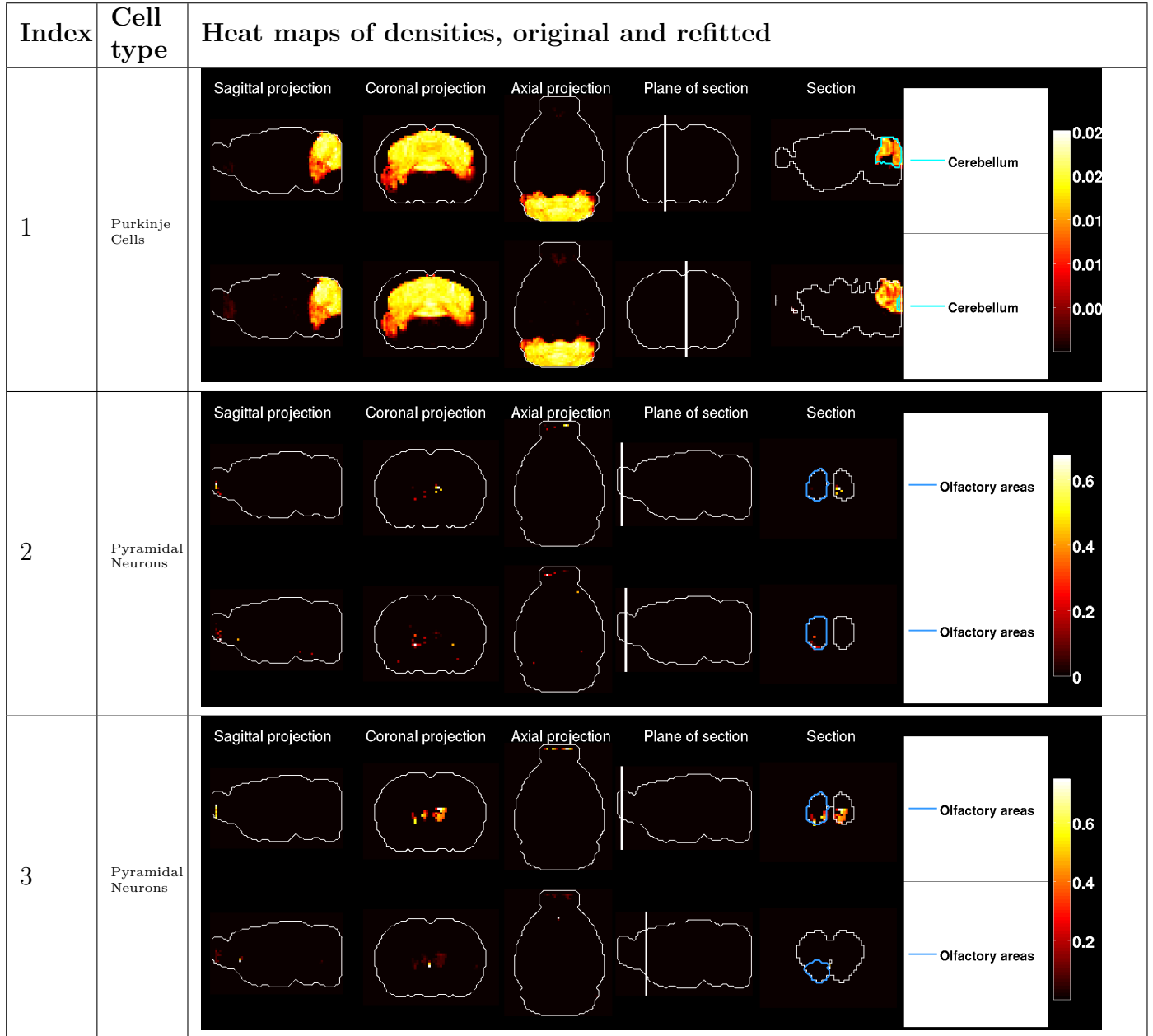


Table 18: Brain-wide density profiles of 3 cell types, in the original linear model (first row of each figure), and in the model fitted to microarray data incorporating the maximum uniform correction compatible with positive entries (second row of each figure).

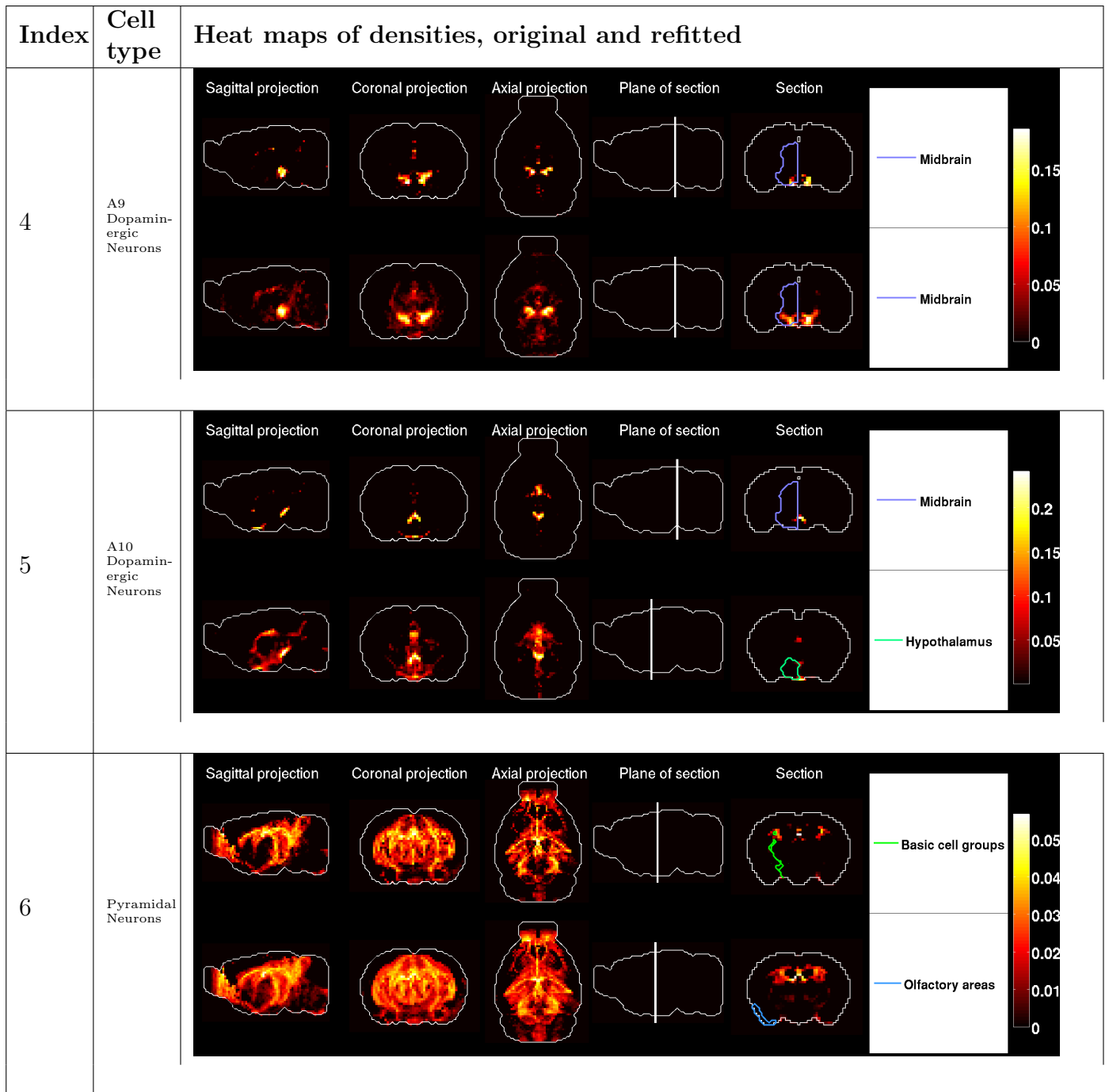


Table 19: Brain-wide density profiles of 3 cell types, in the original linear model (first row of each figure), and in the model fitted to microarray data incorporating the maximum uniform correction compatible with positive entries (second row of each figure).

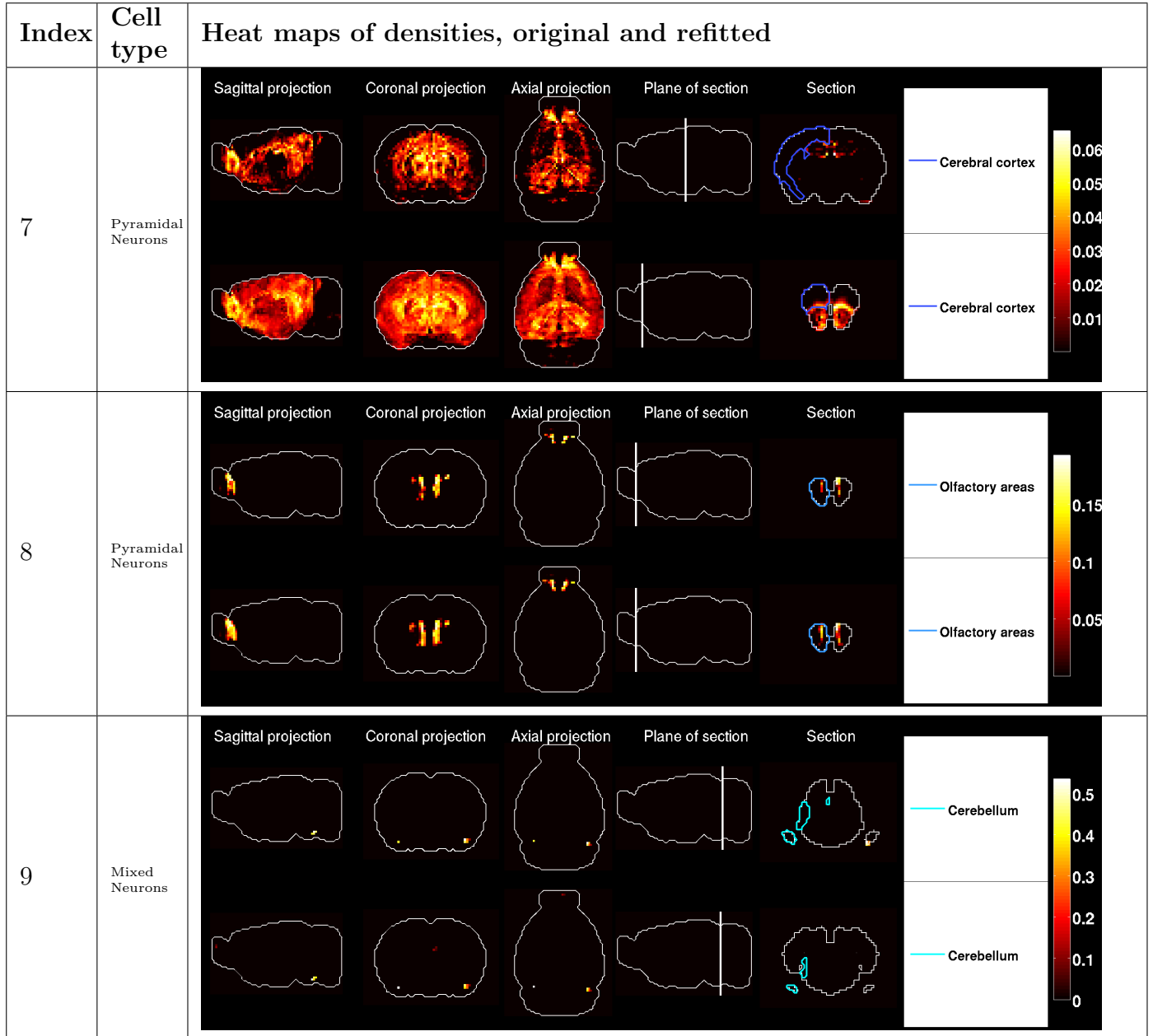


Table 20: Brain-wide density profiles of 3 cell types, in the original linear model (first row of each figure), and in the model fitted to microarray data incorporating the maximum uniform correction compatible with positive entries (second row of each figure).

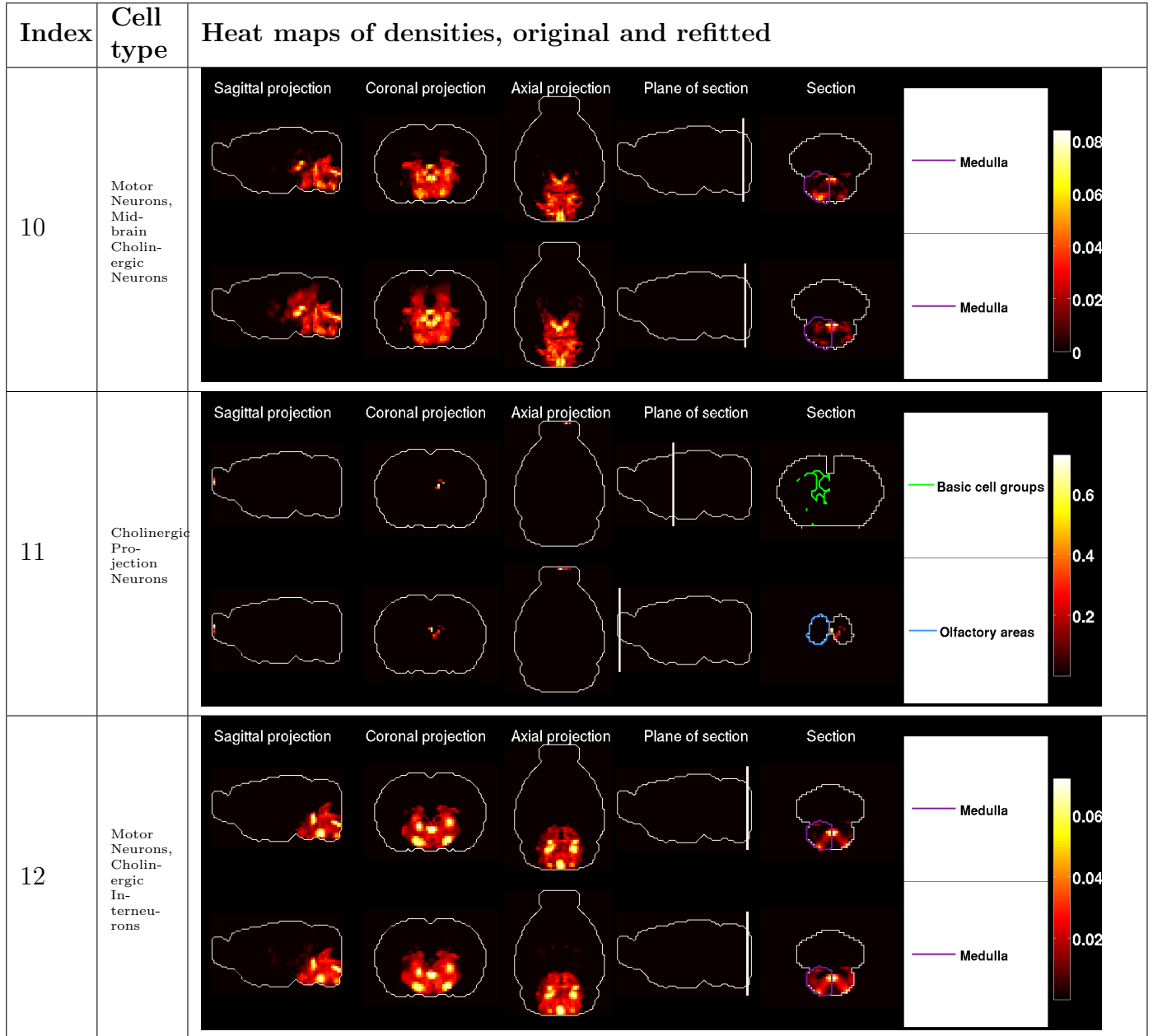


Table 21: Brain-wide density profiles of 3 cell types, in the original linear model (first row of each figure), and in the model fitted to microarray data incorporating the maximum uniform correction compatible with positive entries (second row of each figure).

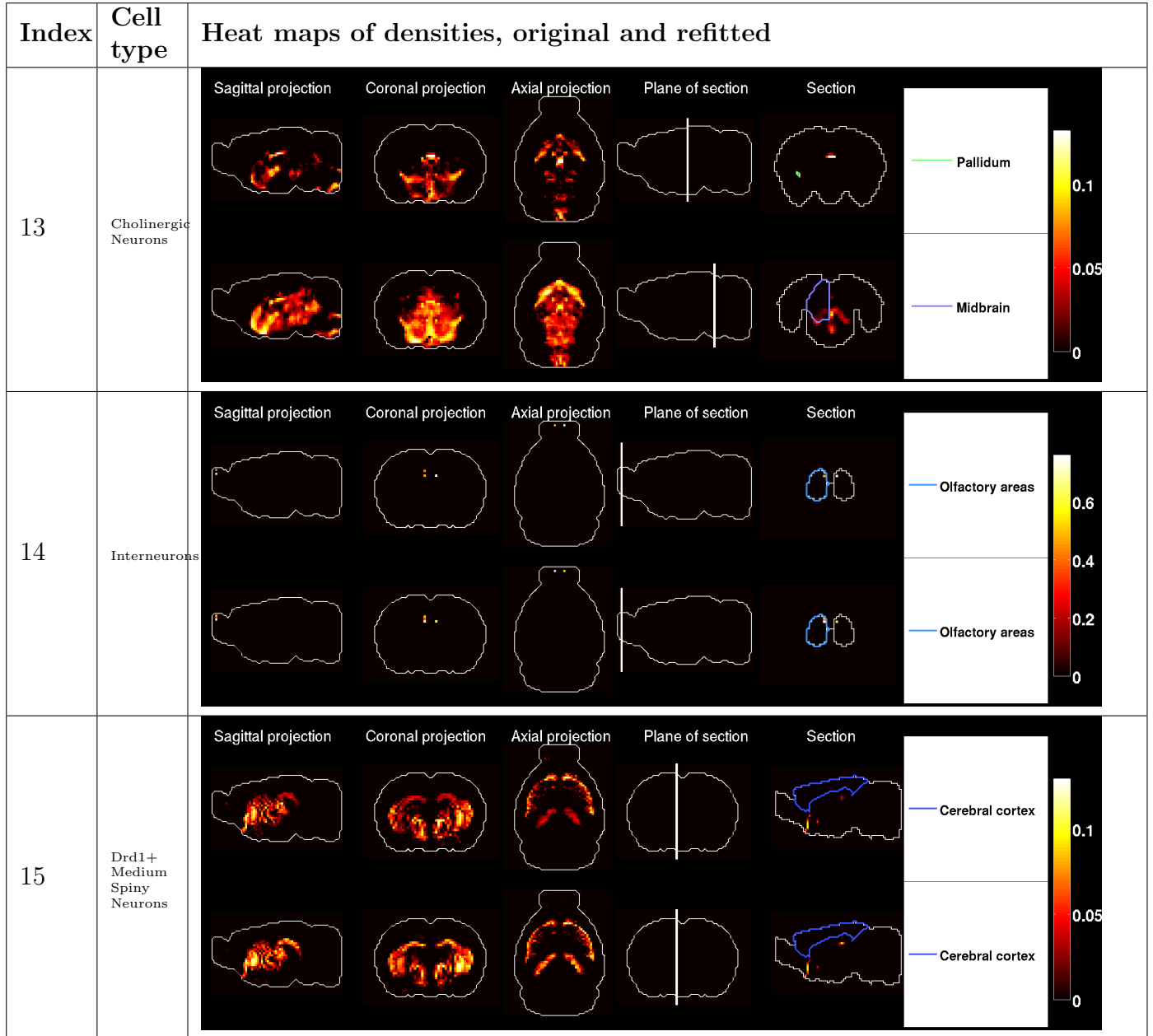


Table 22: Brain-wide density profiles of 3 cell types, in the original linear model (first row of each figure), and in the model fitted to microarray data incorporating the maximum uniform correction compatible with positive entries (second row of each figure).

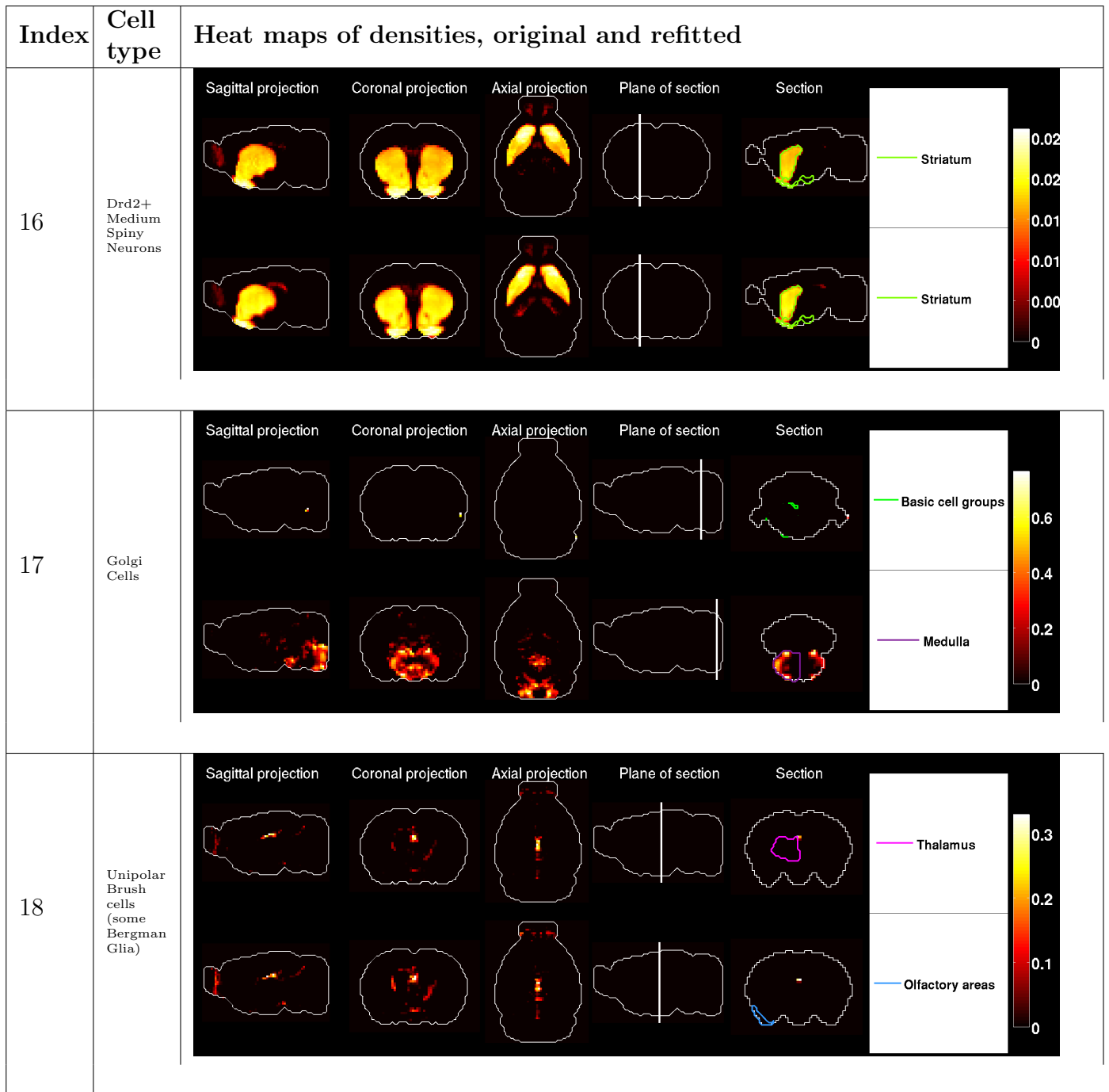


Table 23: Brain-wide density profiles of 3 cell types, in the original linear model (first row of each figure), and in the model fitted to microarray data incorporating the maximum uniform correction compatible with positive entries (second row of each figure).

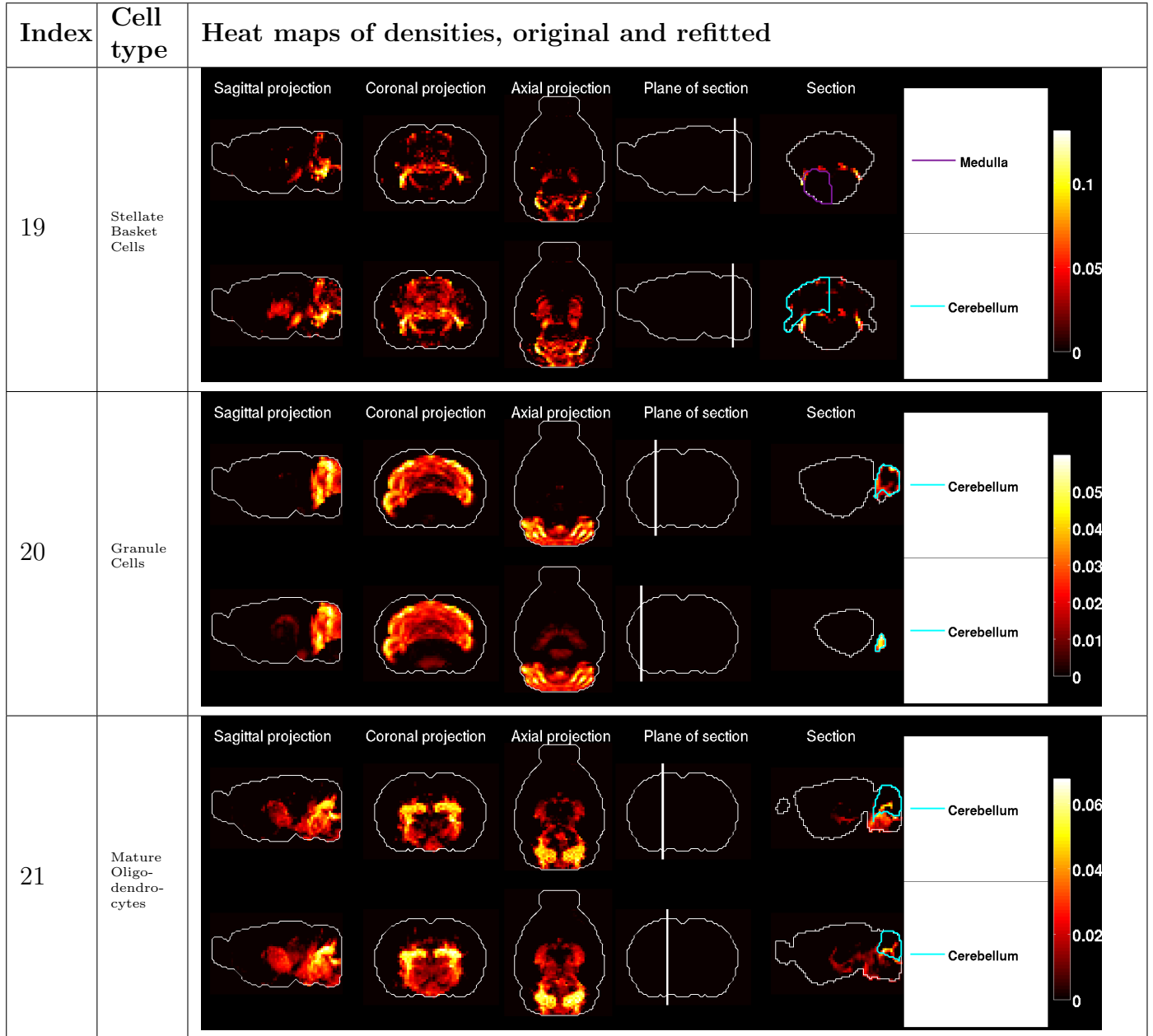


Table 24: Brain-wide density profiles of 3 cell types, in the original linear model (first row of each figure), and in the model fitted to microarray data incorporating the maximum uniform correction compatible with positive entries (second row of each figure).



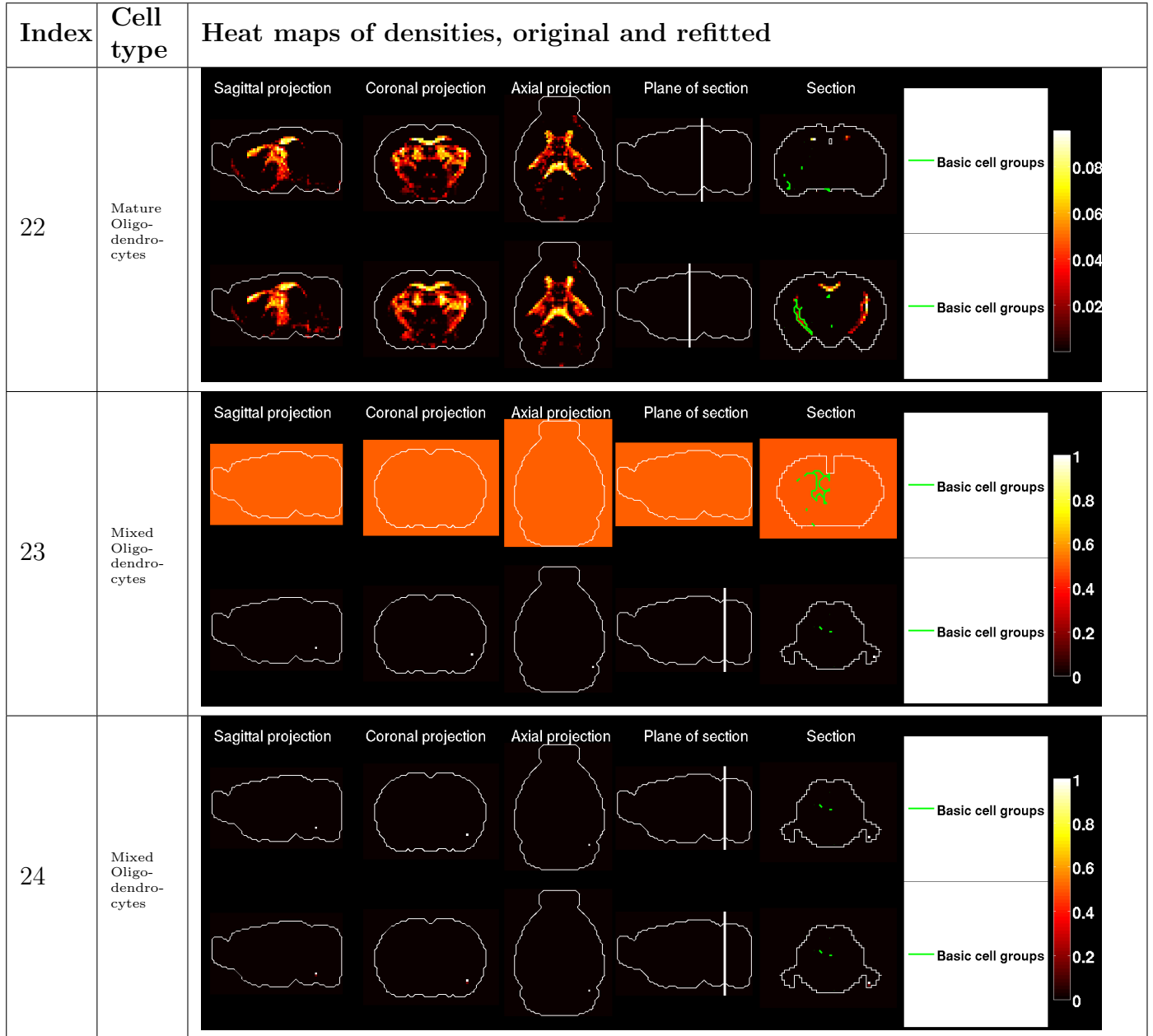


Table 25: Brain-wide density profiles of 3 cell types, in the original linear model (first row of each figure), and in the model fitted to microarray data incorporating the maximum uniform correction compatible with positive entries (second row of each figure).

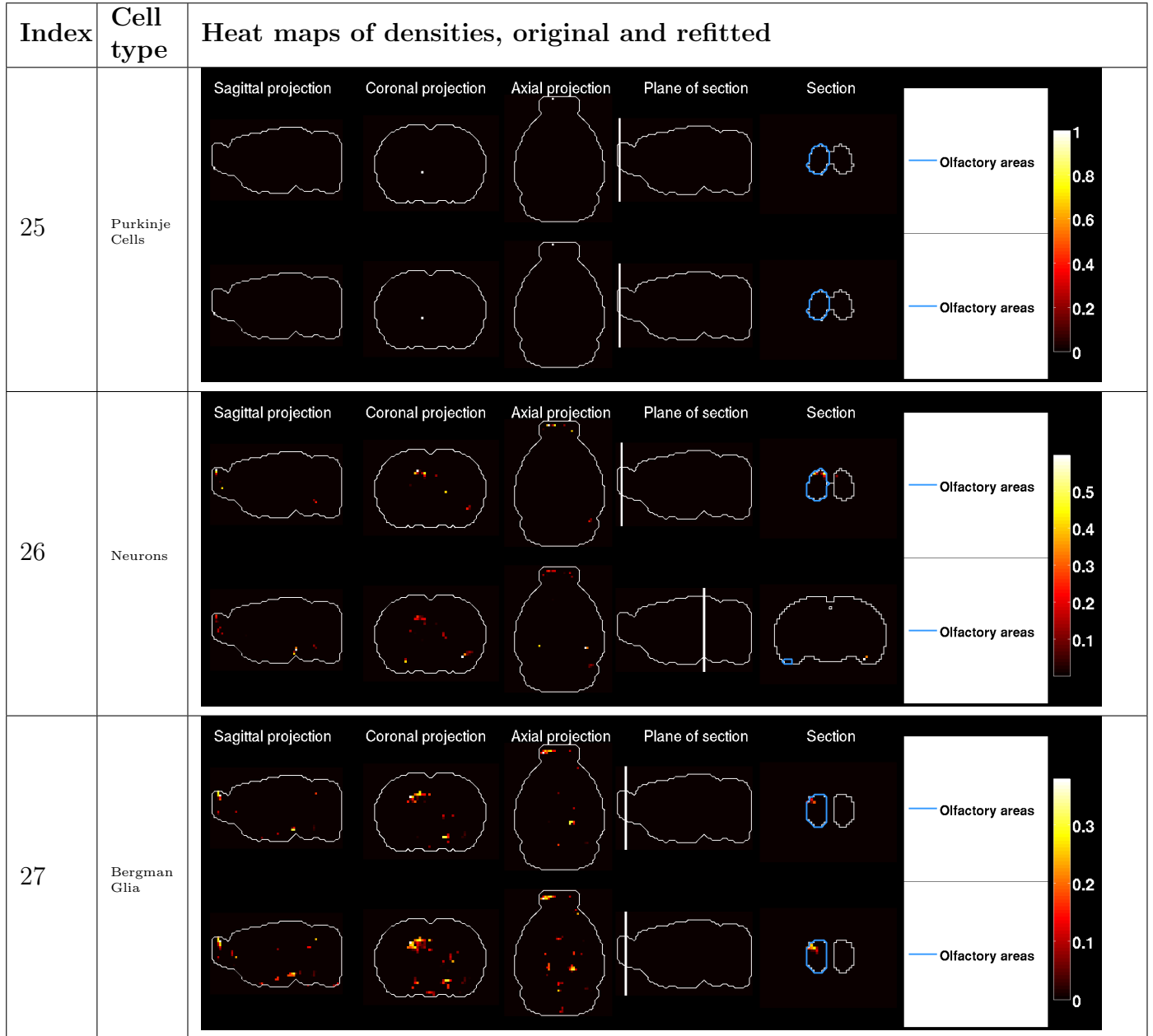


Table 26: Brain-wide density profiles of 3 cell types, in the original linear model (first row of each figure), and in the model fitted to microarray data incorporating the maximum uniform correction compatible with positive entries (second row of each figure).

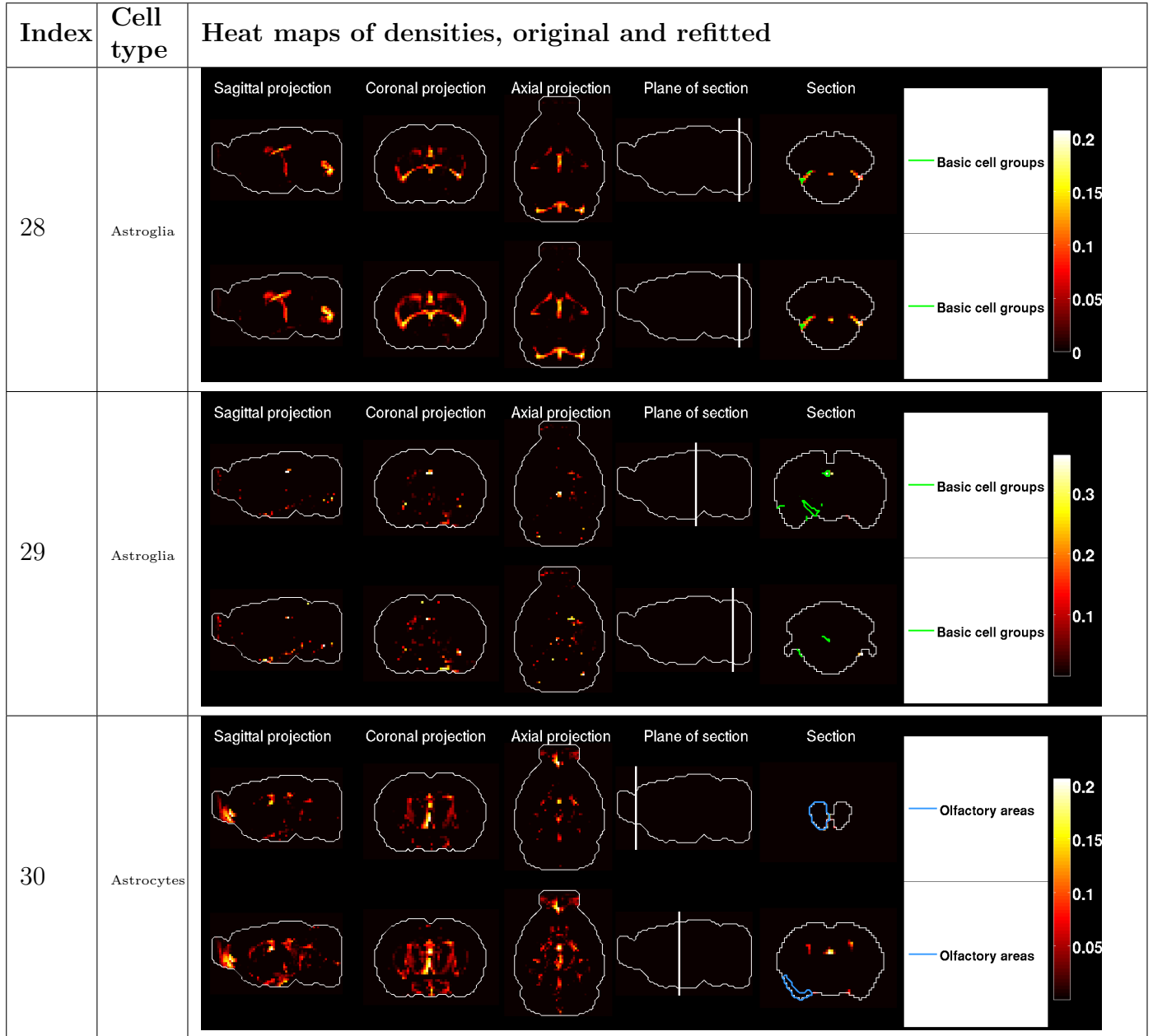


Table 27: Brain-wide density profiles of 3 cell types, in the original linear model (first row of each figure), and in the model fitted to microarray data incorporating the maximum uniform correction compatible with positive entries (second row of each figure).

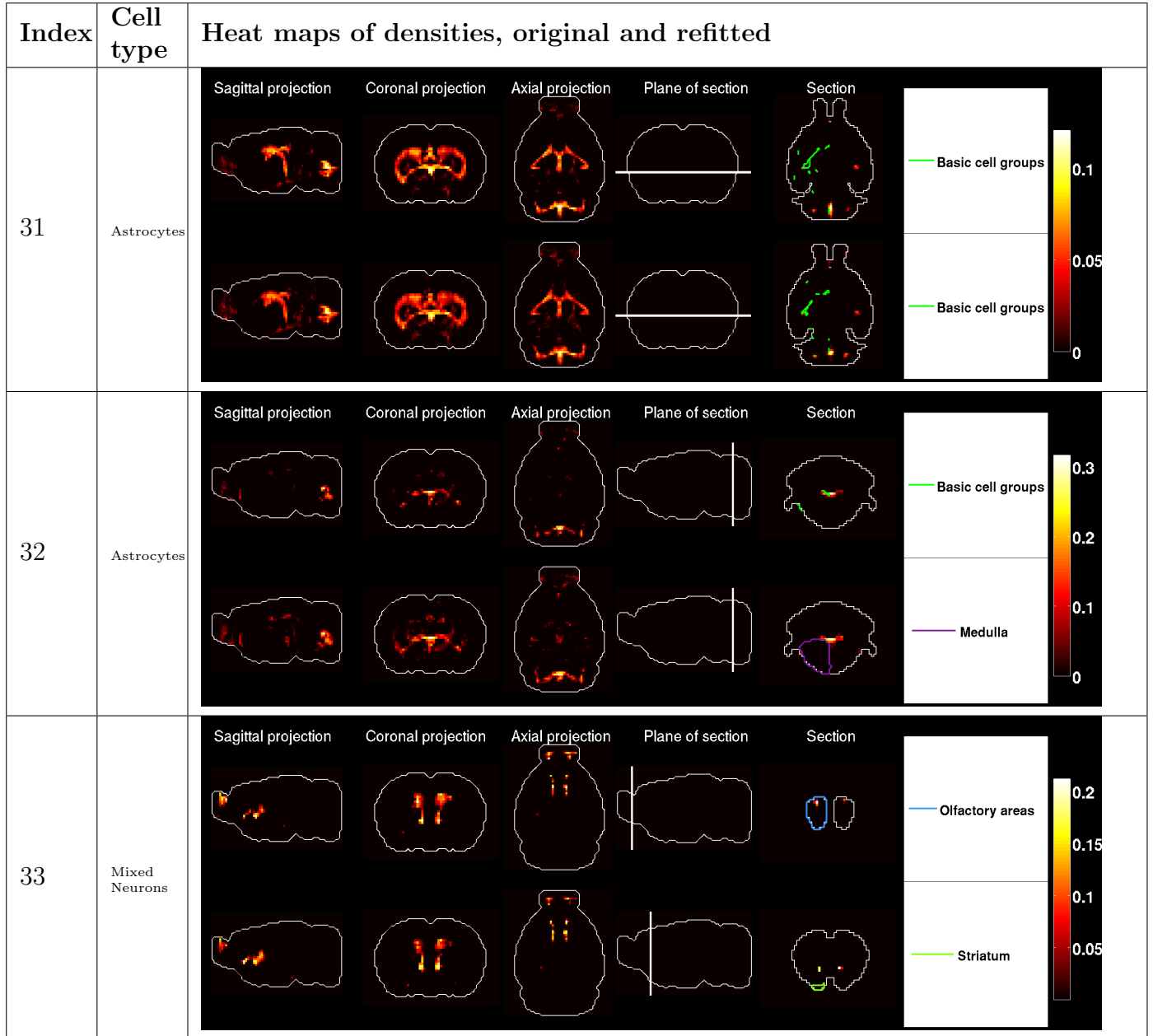


Table 28: Brain-wide density profiles of 3 cell types, in the original linear model (first row of each figure), and in the model fitted to microarray data incorporating the maximum uniform correction compatible with positive entries (second row of each figure).

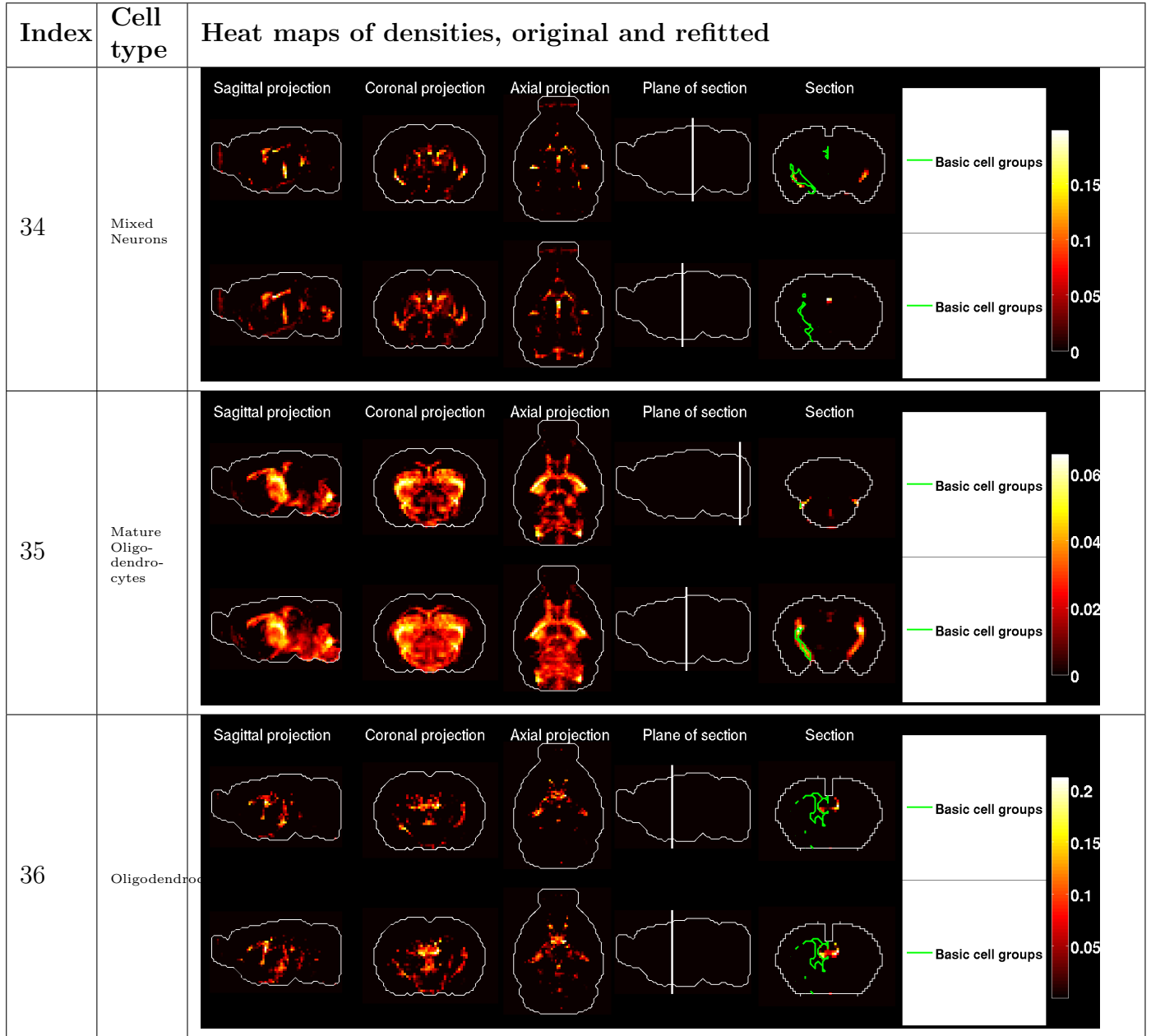


Table 29: Brain-wide density profiles of 3 cell types, in the original linear model (first row of each figure), and in the model fitted to microarray data incorporating the maximum uniform correction compatible with positive entries (second row of each figure).

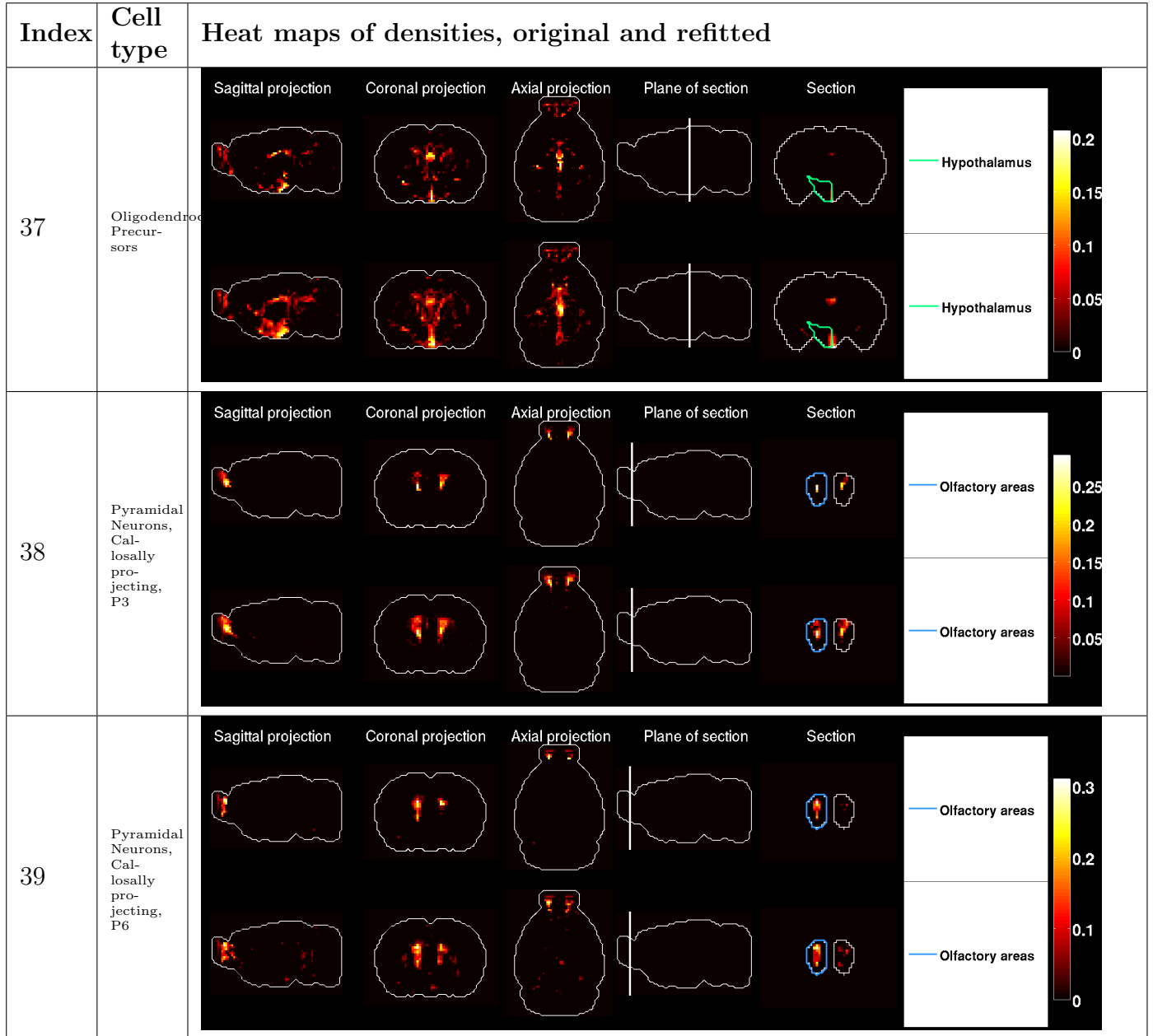


Table 30: Brain-wide density profiles of 3 cell types, in the original linear model (first row of each figure), and in the model fitted to microarray data incorporating the maximum uniform correction compatible with positive entries (second row of each figure).

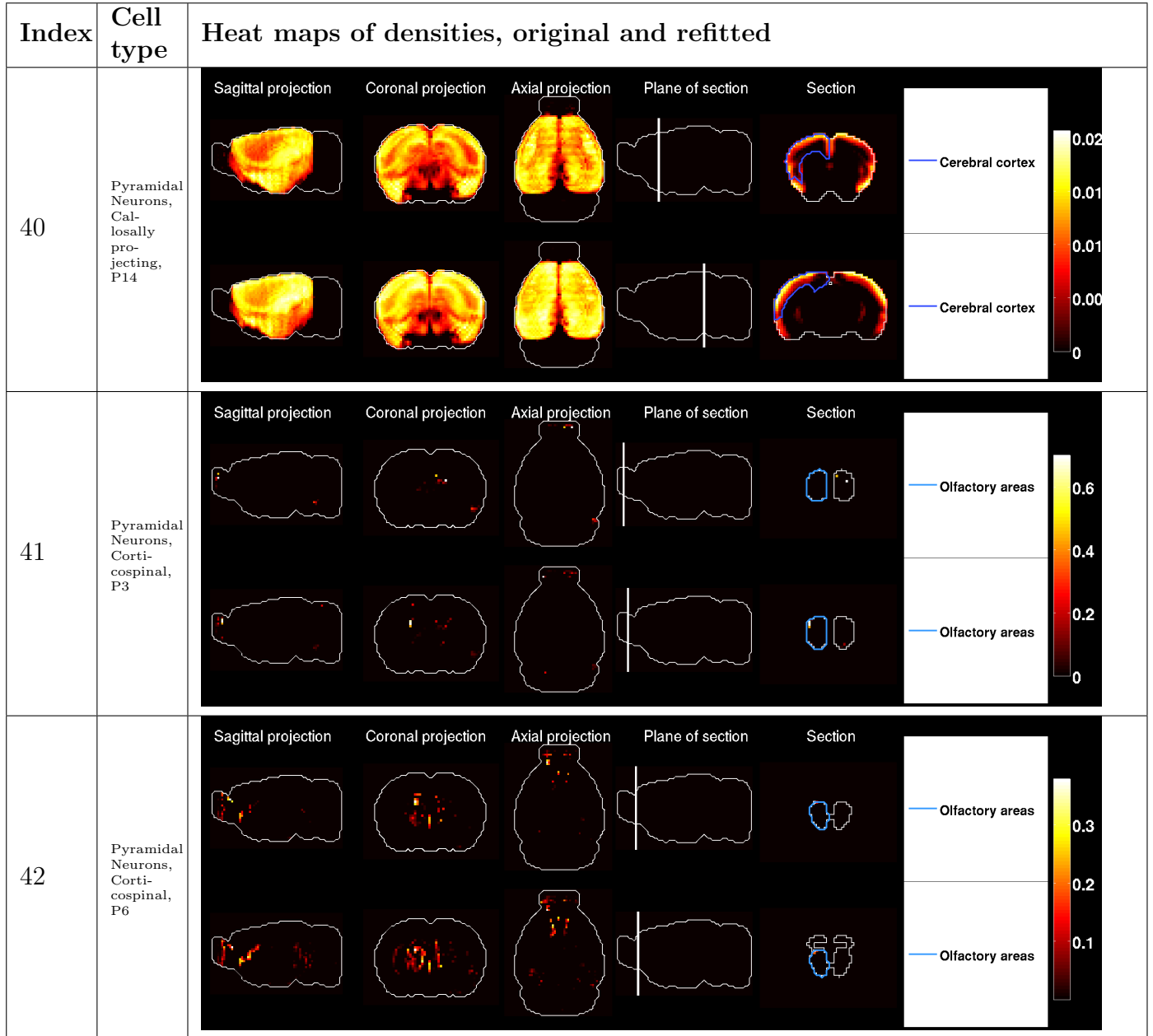


Table 31: Brain-wide density profiles of 3 cell types, in the original linear model (first row of each figure), and in the model fitted to microarray data incorporating the maximum uniform correction compatible with positive entries (second row of each figure).

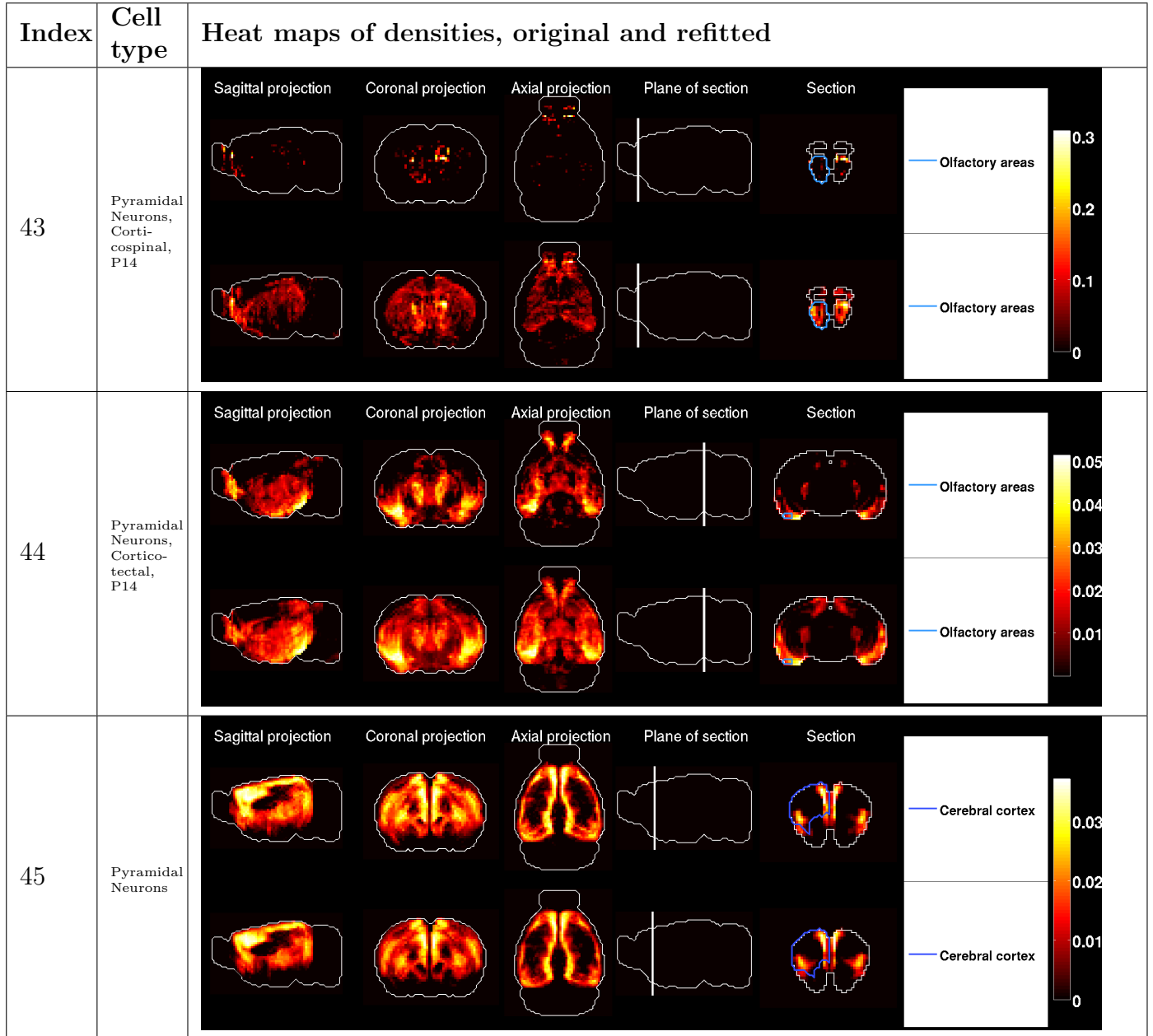


Table 32: Brain-wide density profiles of 3 cell types, in the original linear model (first row of each figure), and in the model fitted to microarray data incorporating the maximum uniform correction compatible with positive entries (second row of each figure).



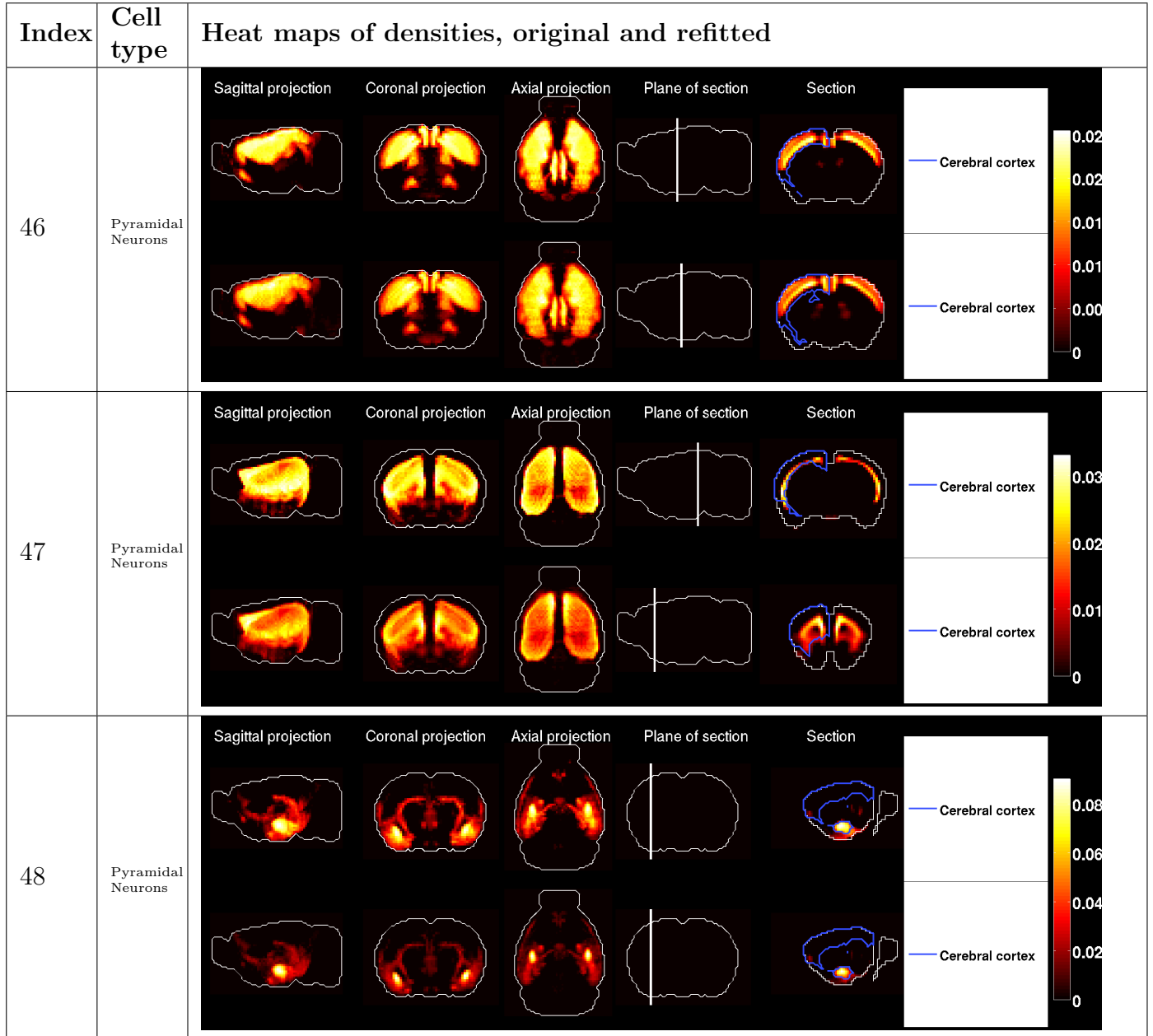


Table 33: Brain-wide density profiles of 3 cell types, in the original linear model (first row of each figure), and in the model fitted to microarray data incorporating the maximum uniform correction compatible with positive entries (second row of each figure).

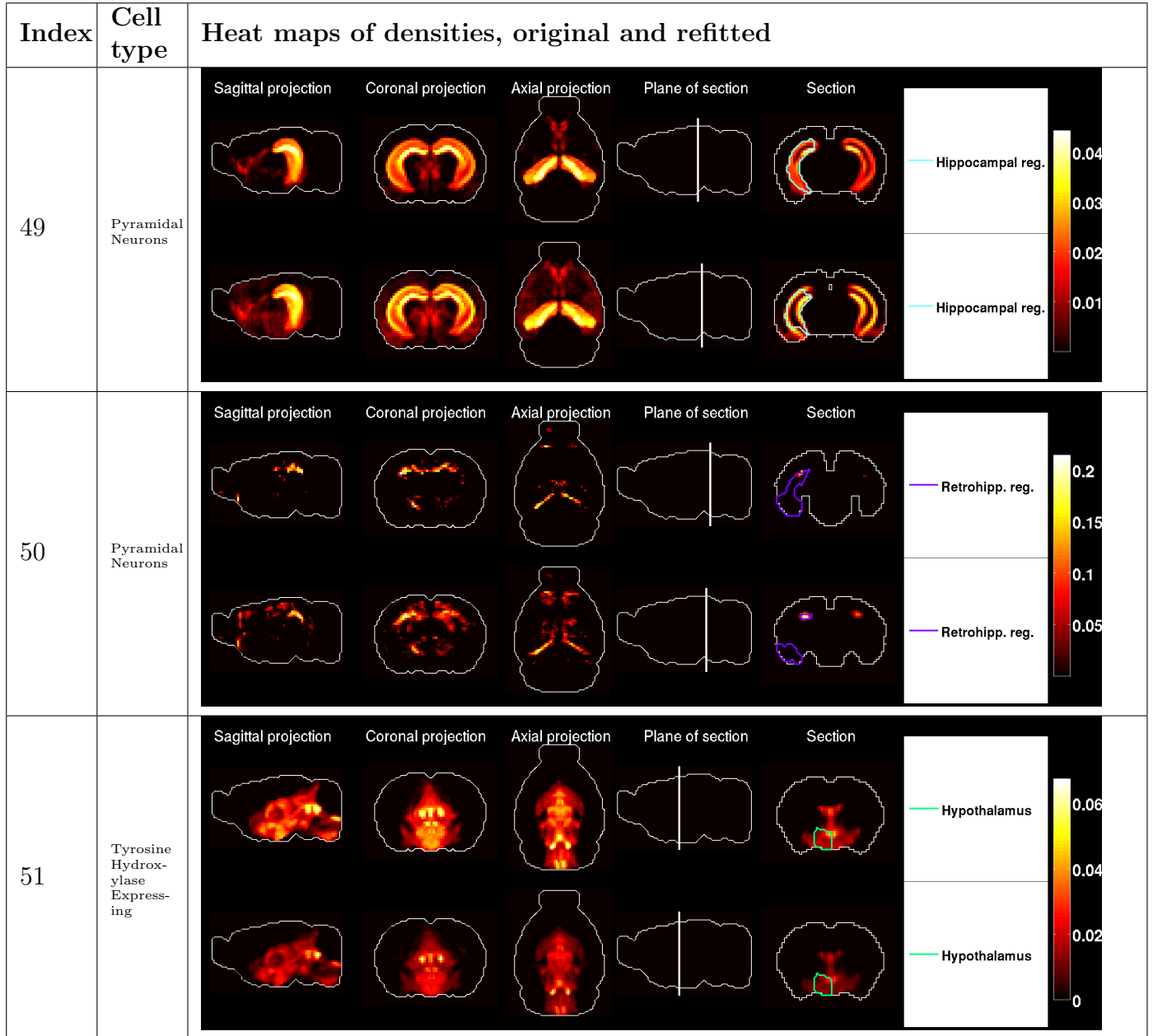


Table 34: Brain-wide density profiles of 3 cell types, in the original linear model (first row of each figure), and in the model fitted to microarray data incorporating the maximum uniform correction compatible with positive entries (second row of each figure).

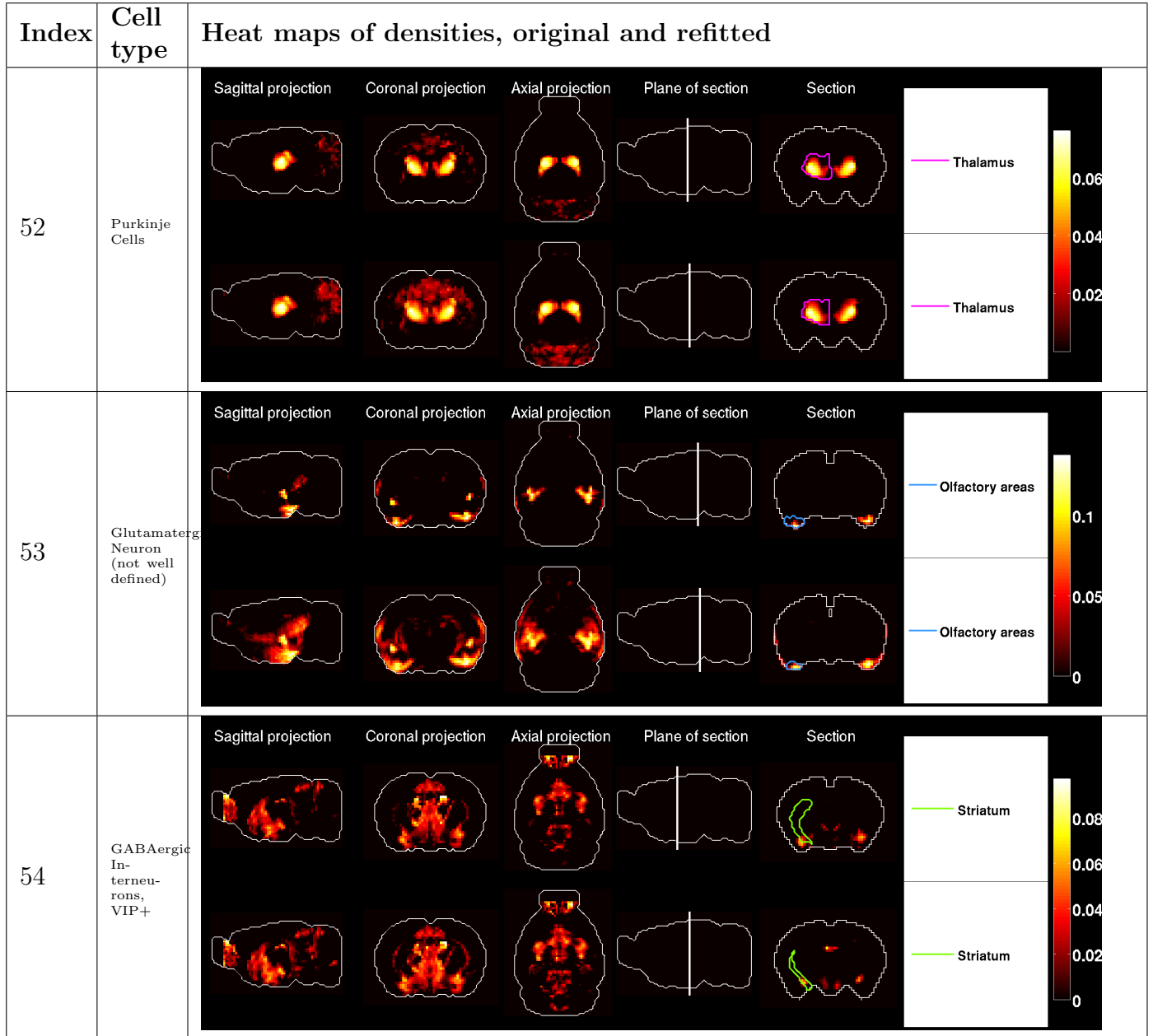


Table 35: Brain-wide density profiles of 3 cell types, in the original linear model (first row of each figure), and in the model fitted to microarray data incorporating the maximum uniform correction compatible with positive entries (second row of each figure).

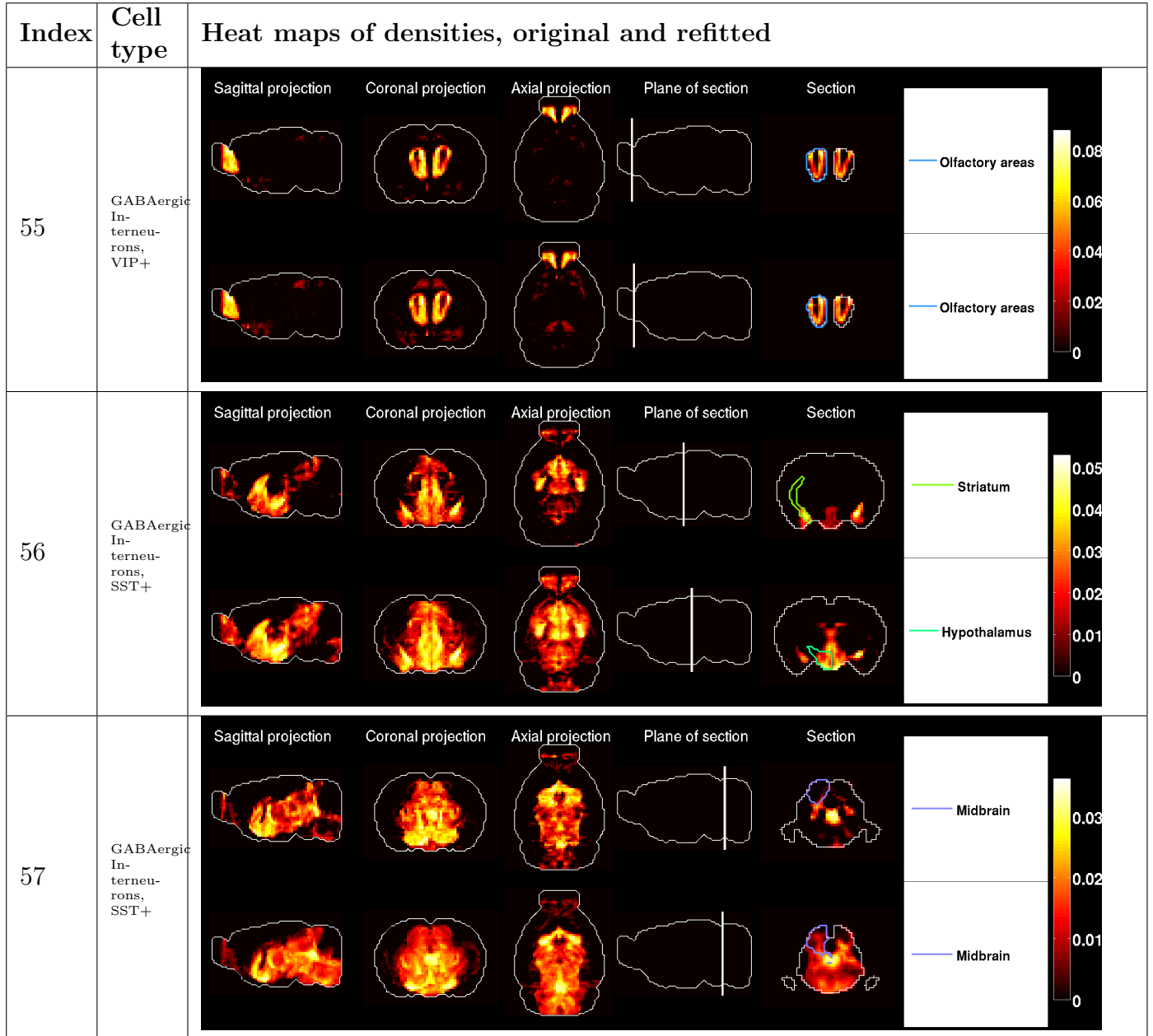


Table 36: Brain-wide density profiles of 3 cell types, in the original linear model (first row of each figure), and in the model fitted to microarray data incorporating the maximum uniform correction compatible with positive entries (second row of each figure).

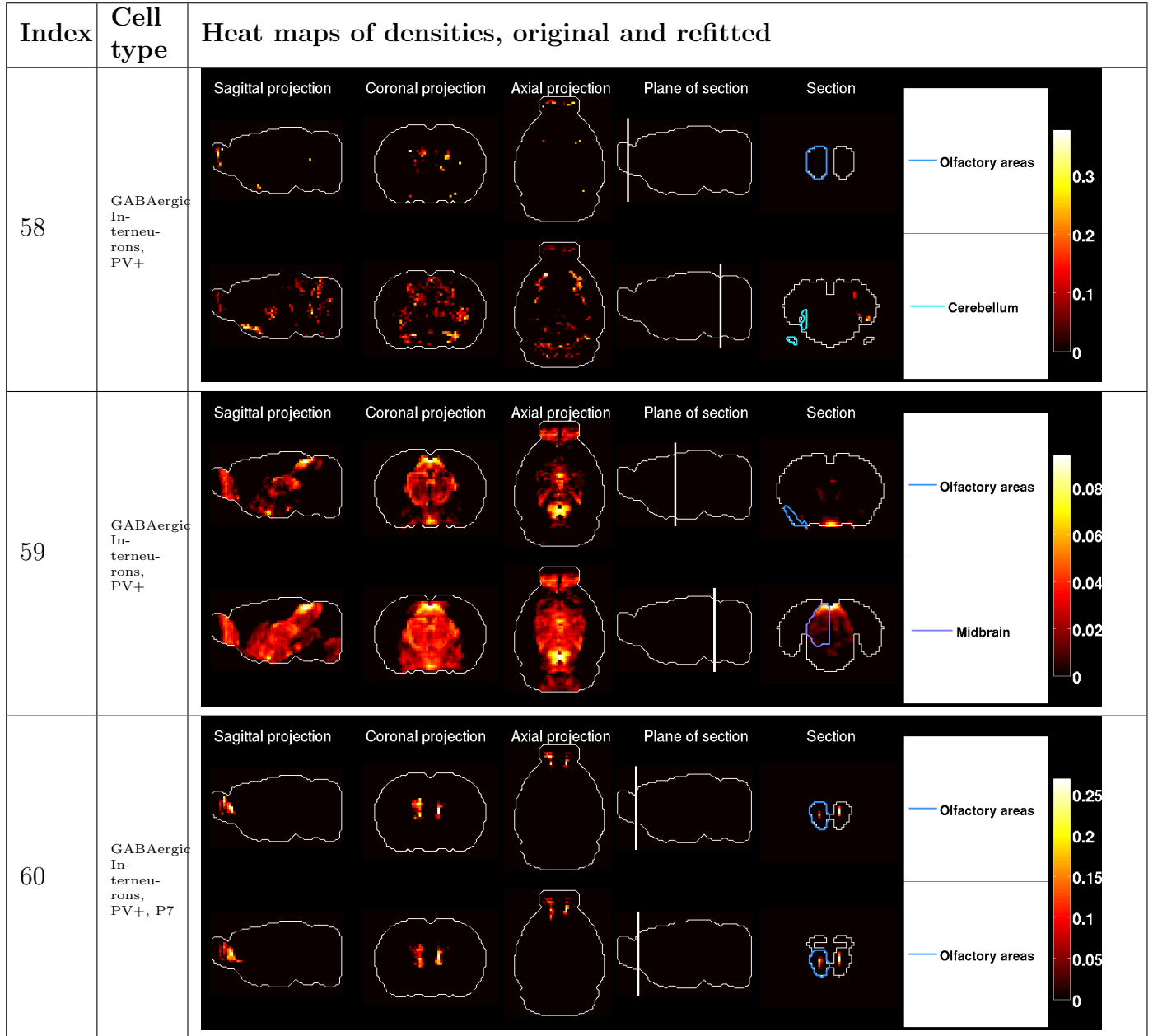


Table 37: Brain-wide density profiles of 3 cell types, in the original linear model (first row of each figure), and in the model fitted to microarray data incorporating the maximum uniform correction compatible with positive entries (second row of each figure).

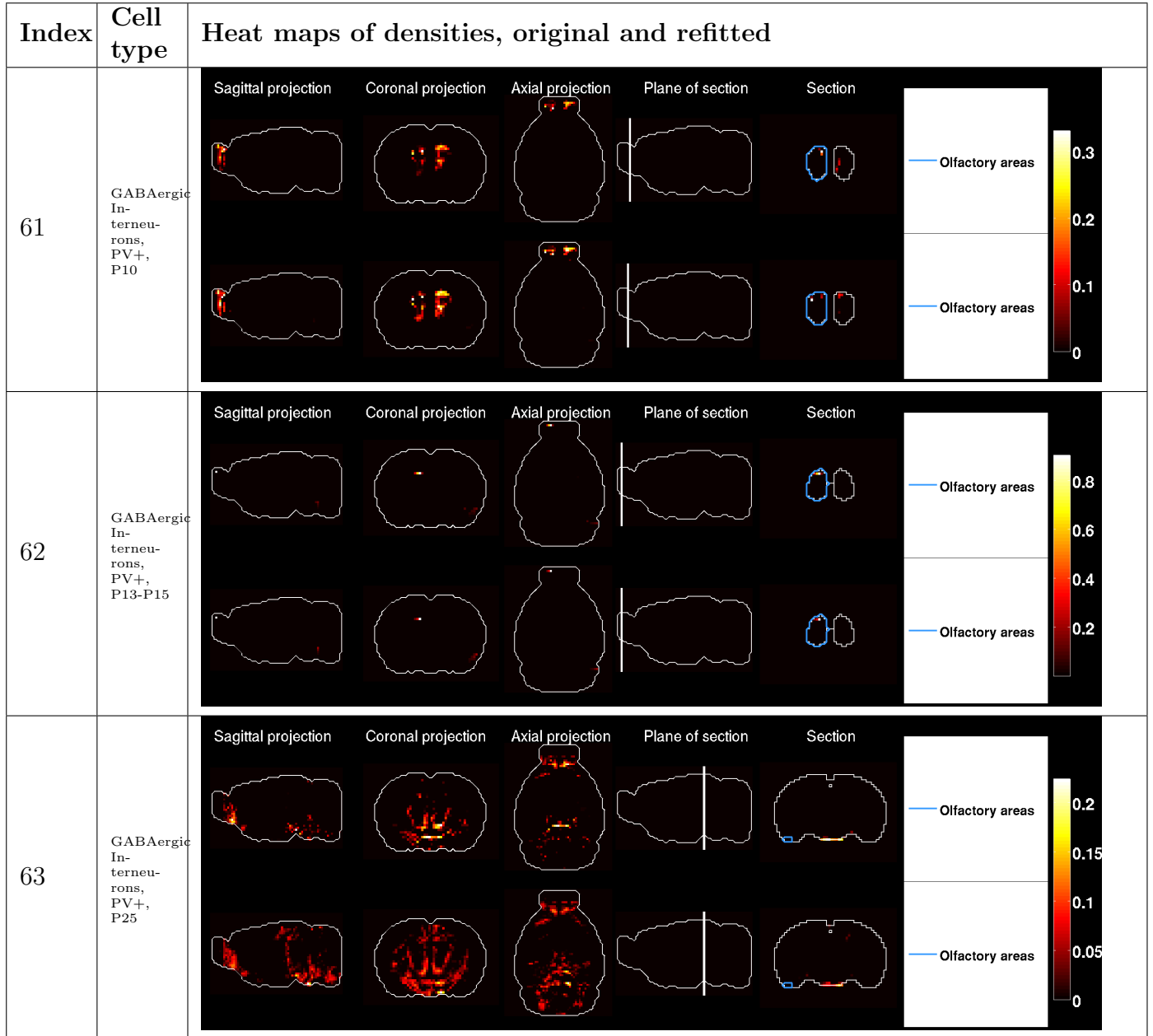


Table 38: Brain-wide density profiles of 3 cell types, in the original linear model (first row of each figure), and in the model fitted to microarray data incorporating the maximum uniform correction compatible with positive entries (second row of each figure).

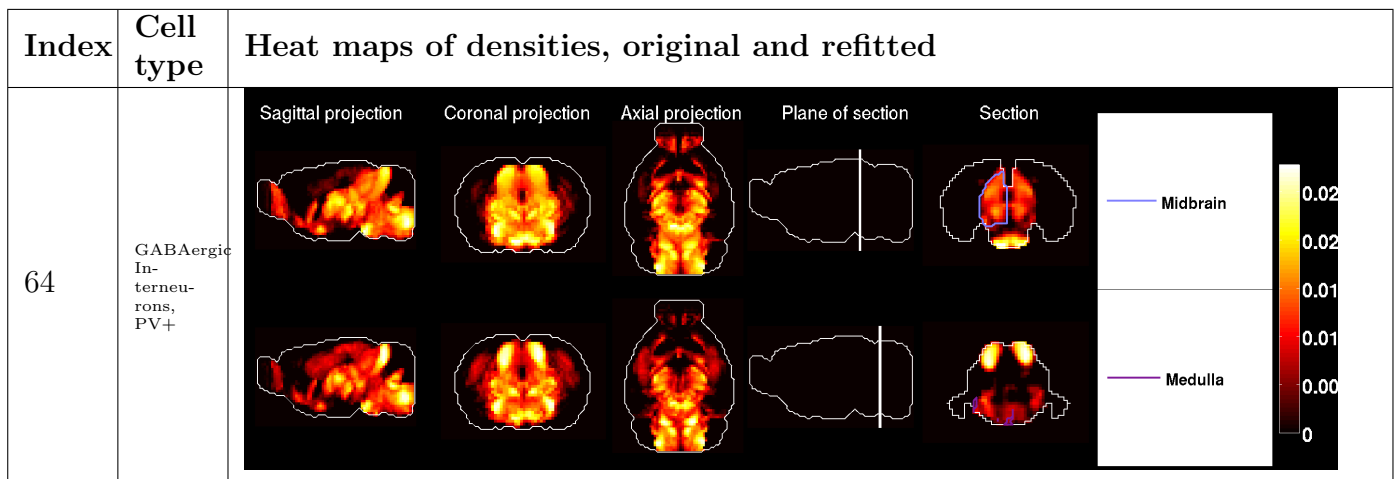


Table 39: Brain-wide density profiles of 3 cell types, in the original linear model (first row of each figure), and in the model fitted to microarray data incorporating the maximum uniform correction compatible with positive entries (second row of each figure).

## 6 Conclusions

Some of the neuroanatomical patterns discovered both in the correlation and density volumes confirm prior biological knowledge in the sense that the top region by correlation or density coincides with the region from which the corresponding cell-type-specific sample was extracted. This is the case in the cerebral cortex, the hippocampal region, the striatum, the ventral mid-brain, the medulla and the cerebellum.

The anatomical analysis conducted here is supervised in the sense that the brain regions are always taken to be one of the sets of voxels carrying the same label under some version of the voxelized Allen Reference Atlas. In the case of the two samples extracted from the amygdala (indices 48 and 53), our method cannot return the label 'amygdala' as a top region by correlation and density, since the amygdala is split between several regions. This kind of issue is bound to occur for any choice of atlas, since there is no universal agreement on the nomenclature of brain regions [29]. Clustering methods could be applied to the correlation and density profiles, following the approach taken in [30] for the analysis of the voxel-by-gene matrix of expression energies.

The rankings of brain regions yield some surprising results that can be linked to the relative paucity of cell types in the study, compared to the whole diversity of cells in the mouse brain, and to the fact that the cell samples are not distributed uniformly across the brain. It can be noted that there is more solidarity between the GABAergic interneurons (indices 54–64), than between their top regions (by correlation or density), and their anatomical origin.

A richer microarray data set with a more uniform sampling of brain regions will modify the numerical results. In particular, it will be interesting to see if positive densities in the hypothalamus and in the olfactory areas can be estimated for samples extracted from these regions.

## 7 Acknowledgments

We thank Benjamin Okaty, Sacha Nelson and Ken Sugino for collating and transferring the microarray data, and for help with the anatomical analysis of results. We thank Jason W. Bohland, Hemant Bokil, Lydia Ng and Domenico Orlando for discussions. This research is supported by the NIH-NIDA Grant 1R21DA027644-01, *Computational analysis of co-expression networks in the mouse and human brain*.



## 8 Tables of brain-wide correlations

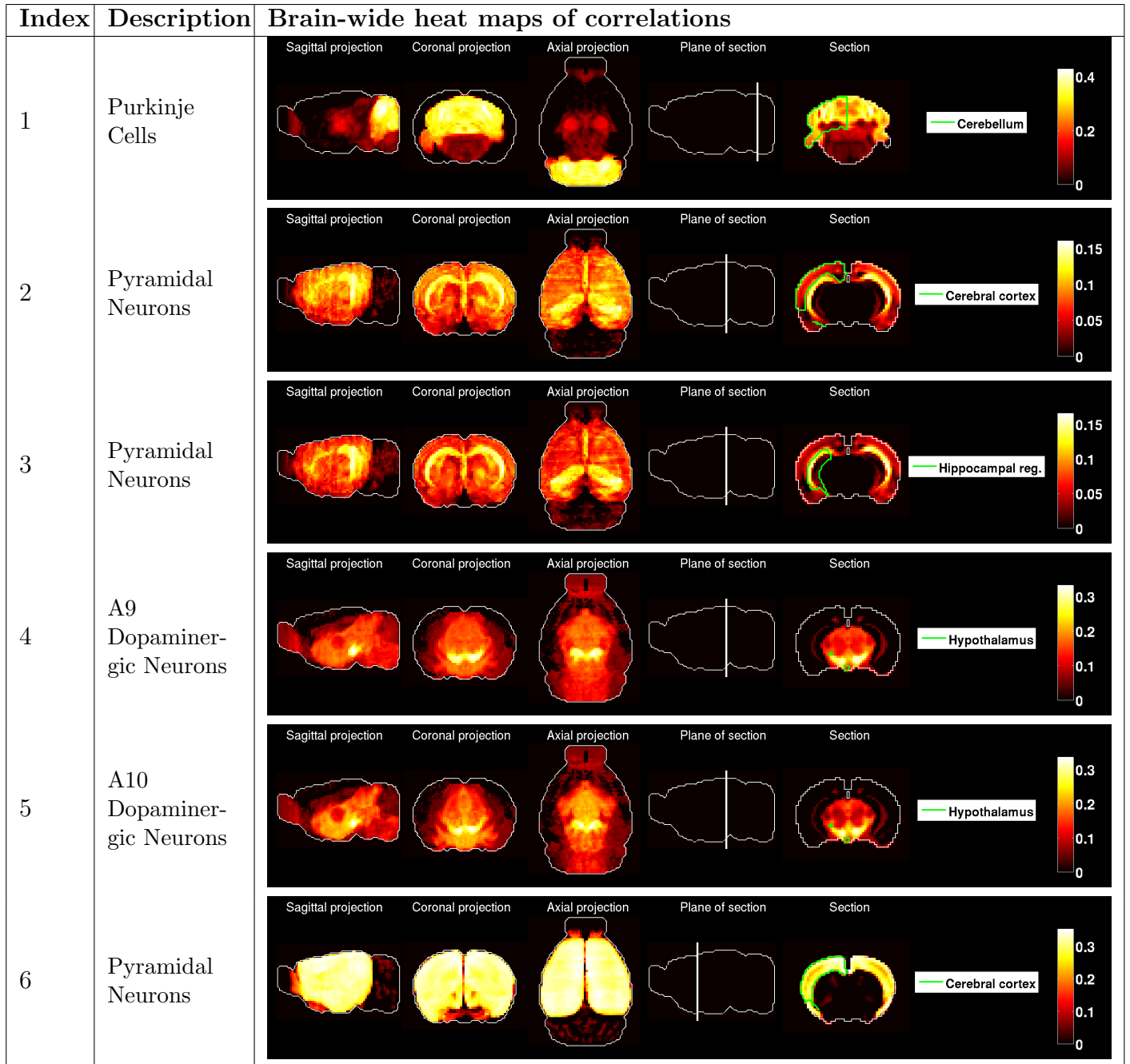


Table 40: Brain-wide correlation profiles between 6 cell types and the Allen Atlas.

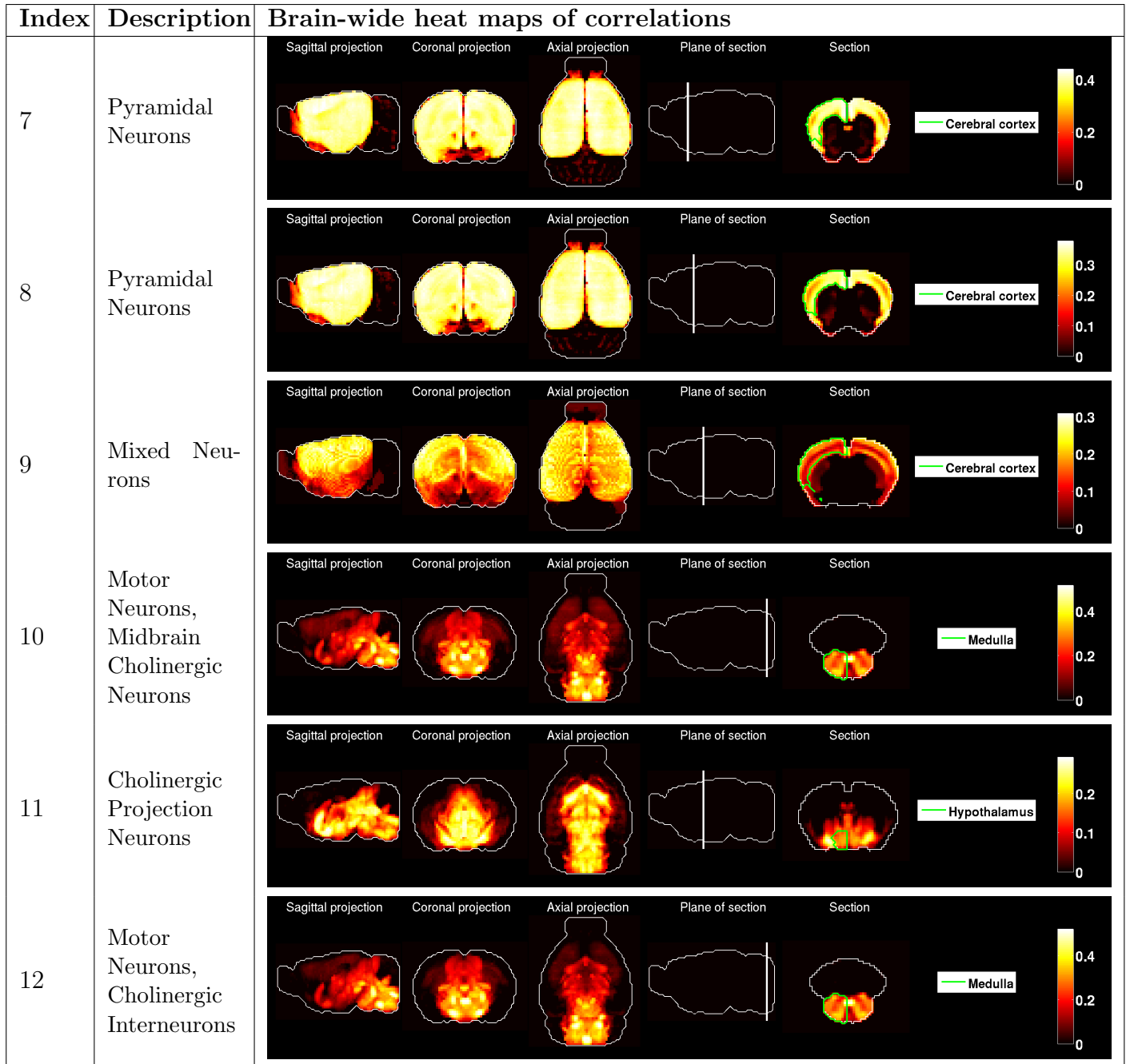


Table 41: Brain-wide correlation profiles between 6 cell types and the Allen Atlas.

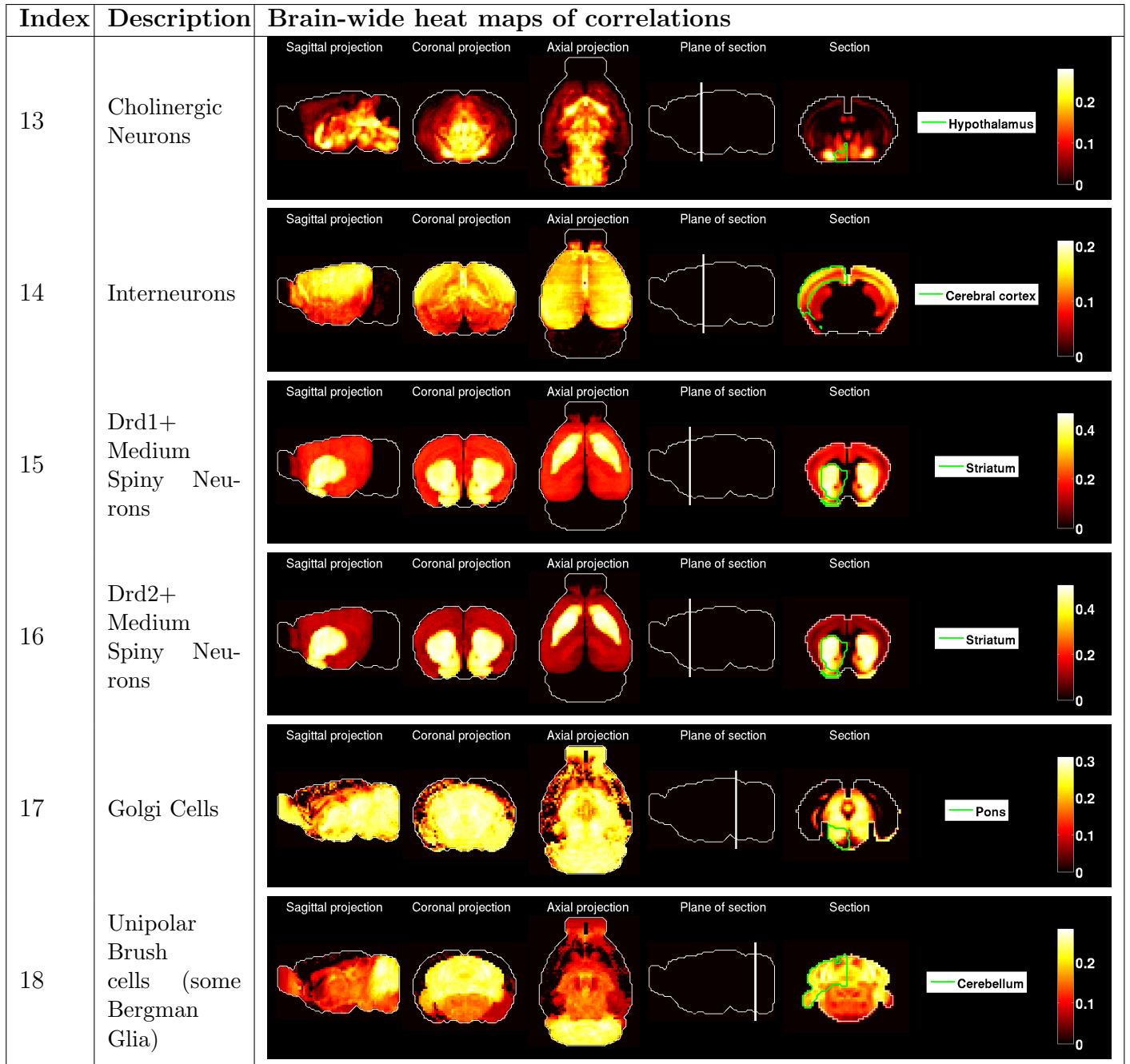


Table 42: Brain-wide correlation profiles between 6 cell types and the Allen Atlas.

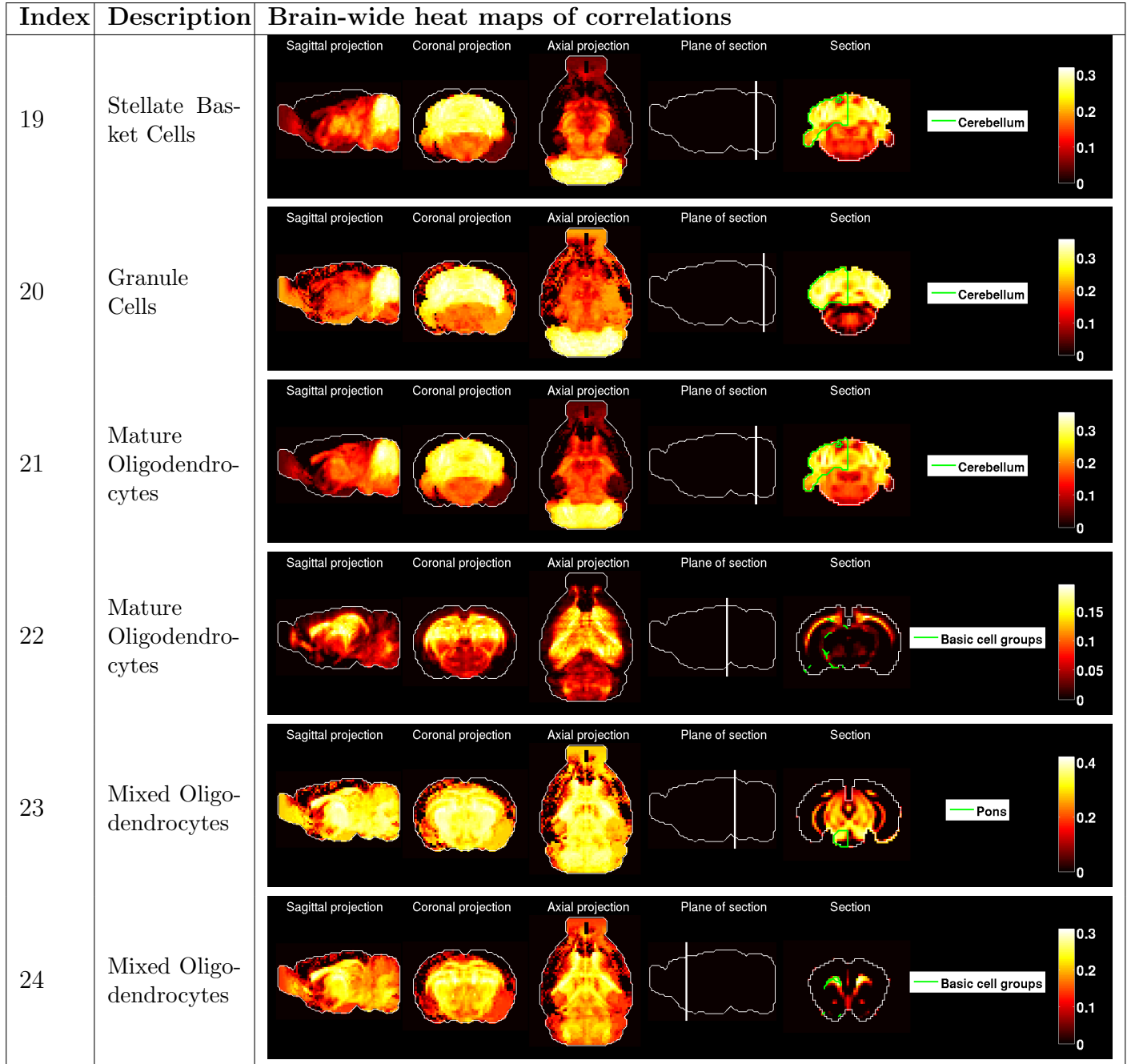


Table 43: Brain-wide correlation profiles between 6 cell types and the Allen Atlas.

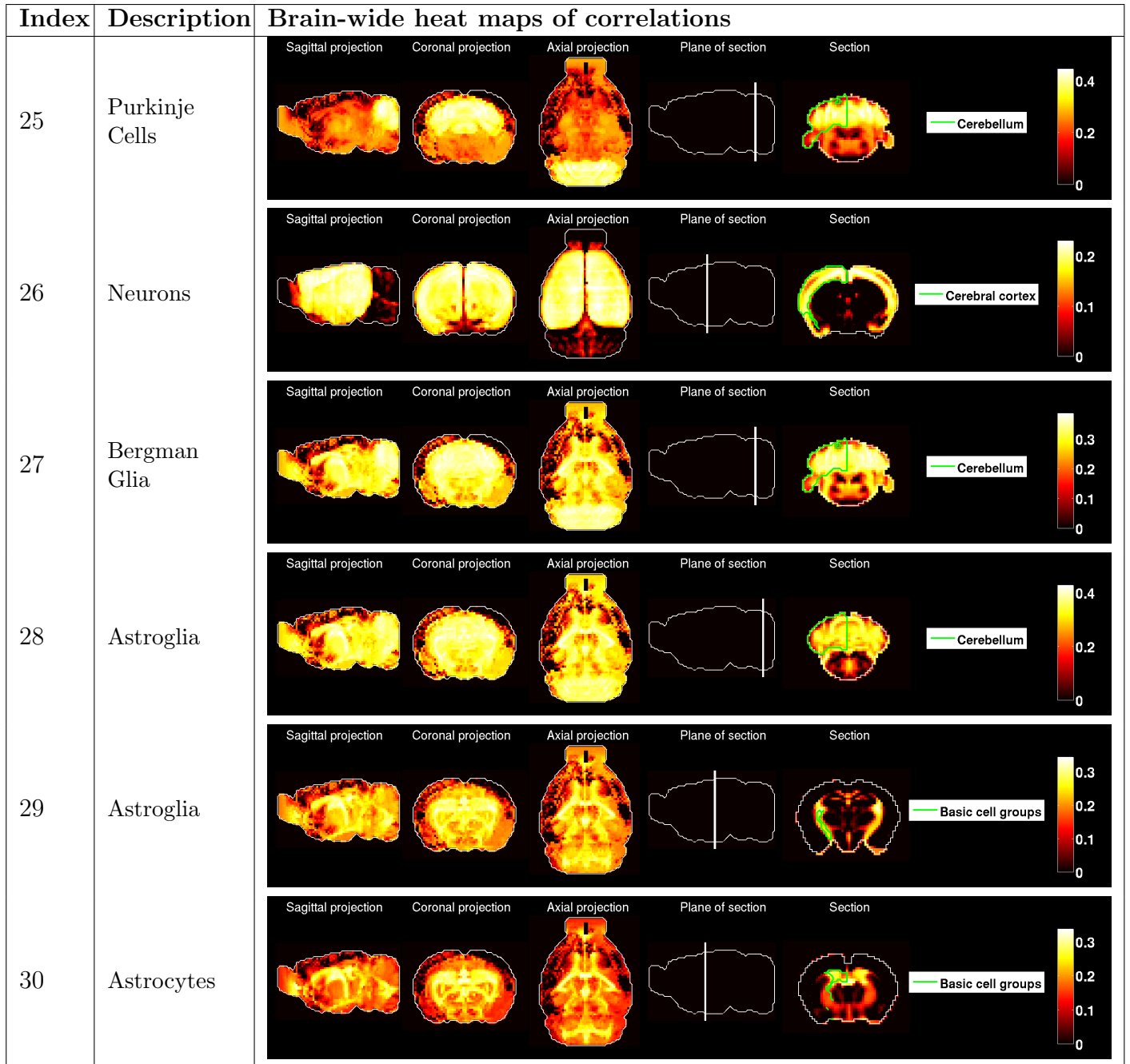


Table 44: Brain-wide correlation profiles between 6 cell types and the Allen Atlas.

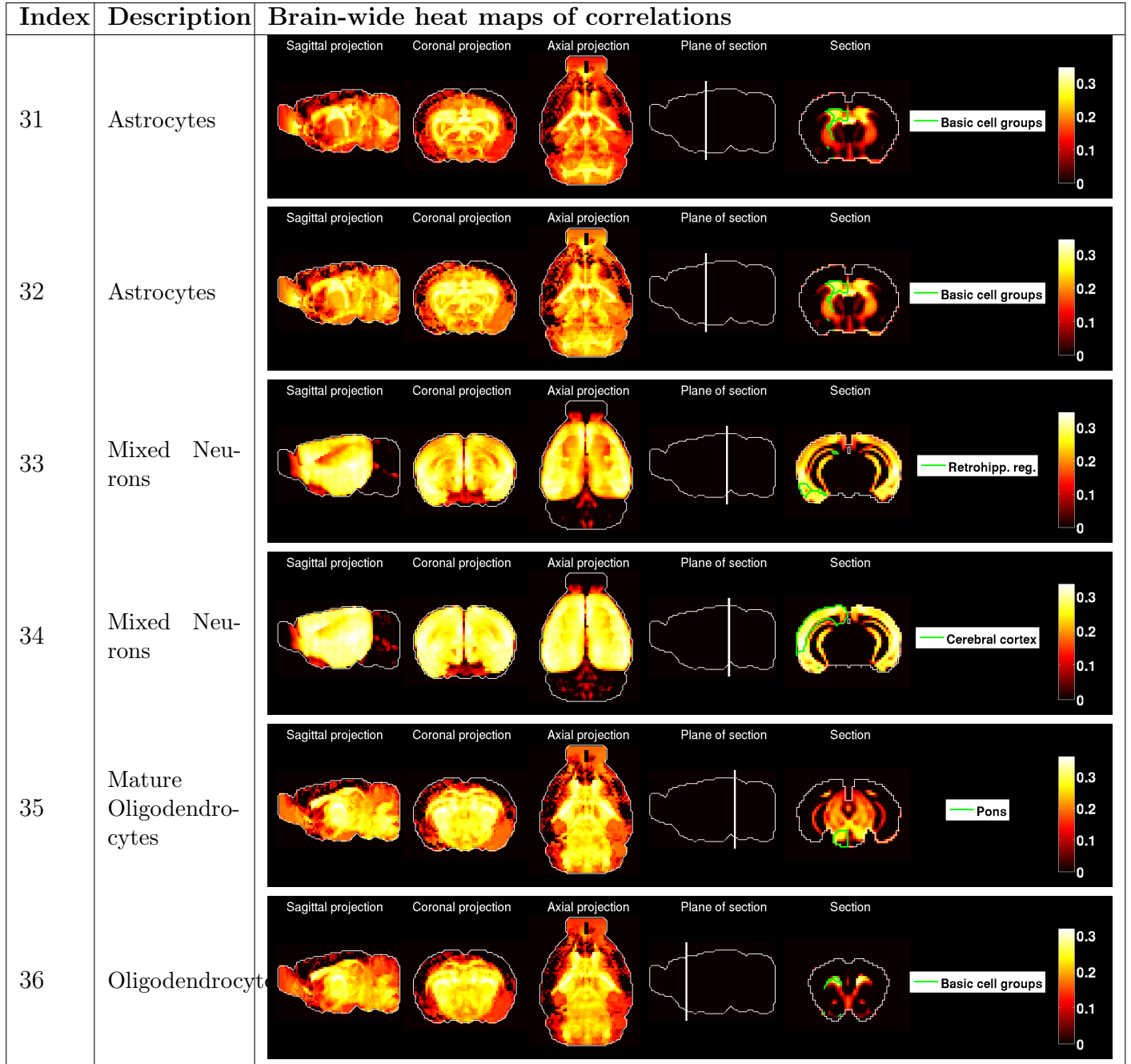


Table 45: Brain-wide correlation profiles between 6 cell types and the Allen Atlas.

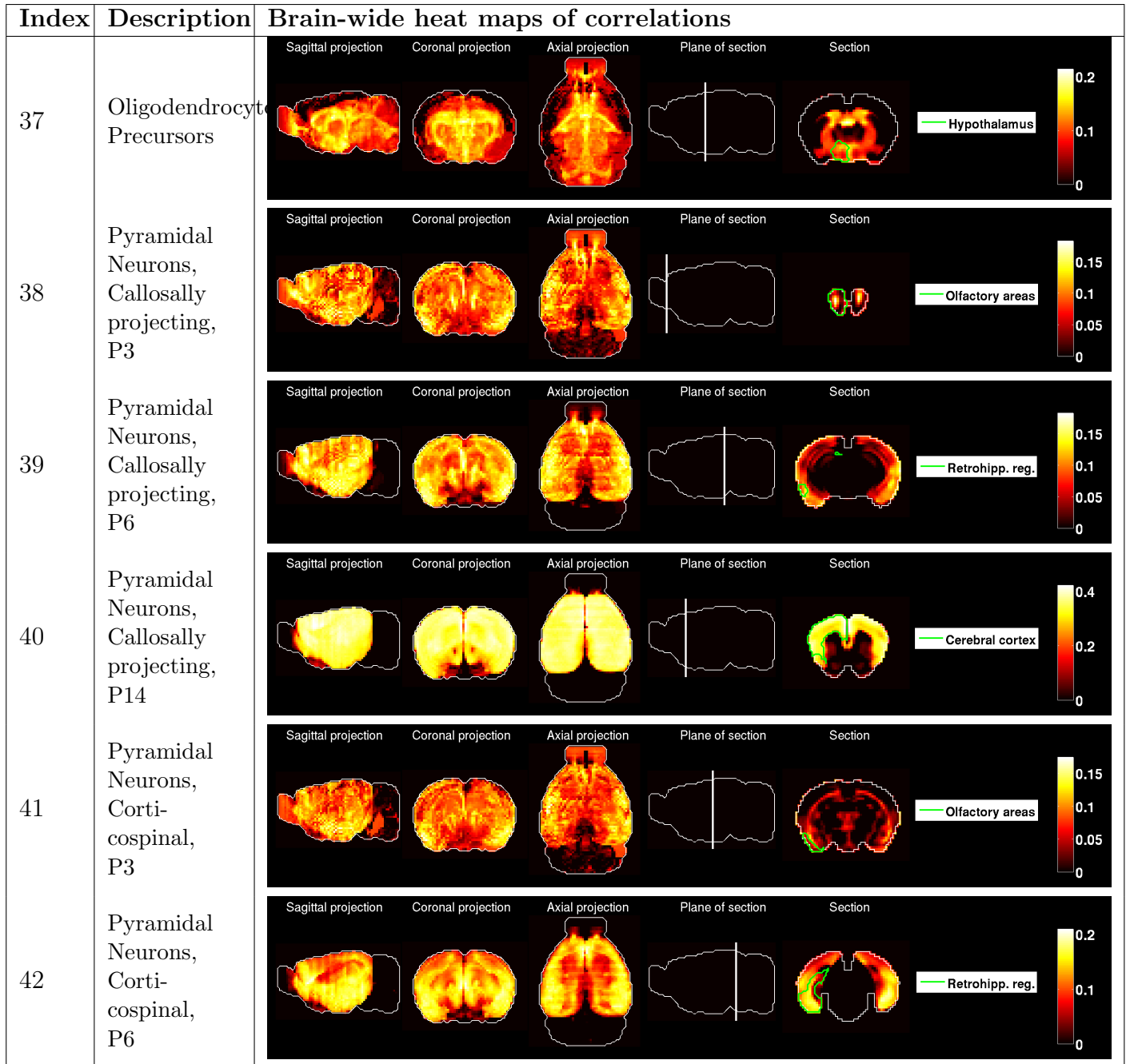


Table 46: Brain-wide correlation profiles between 6 cell types and the Allen Atlas.



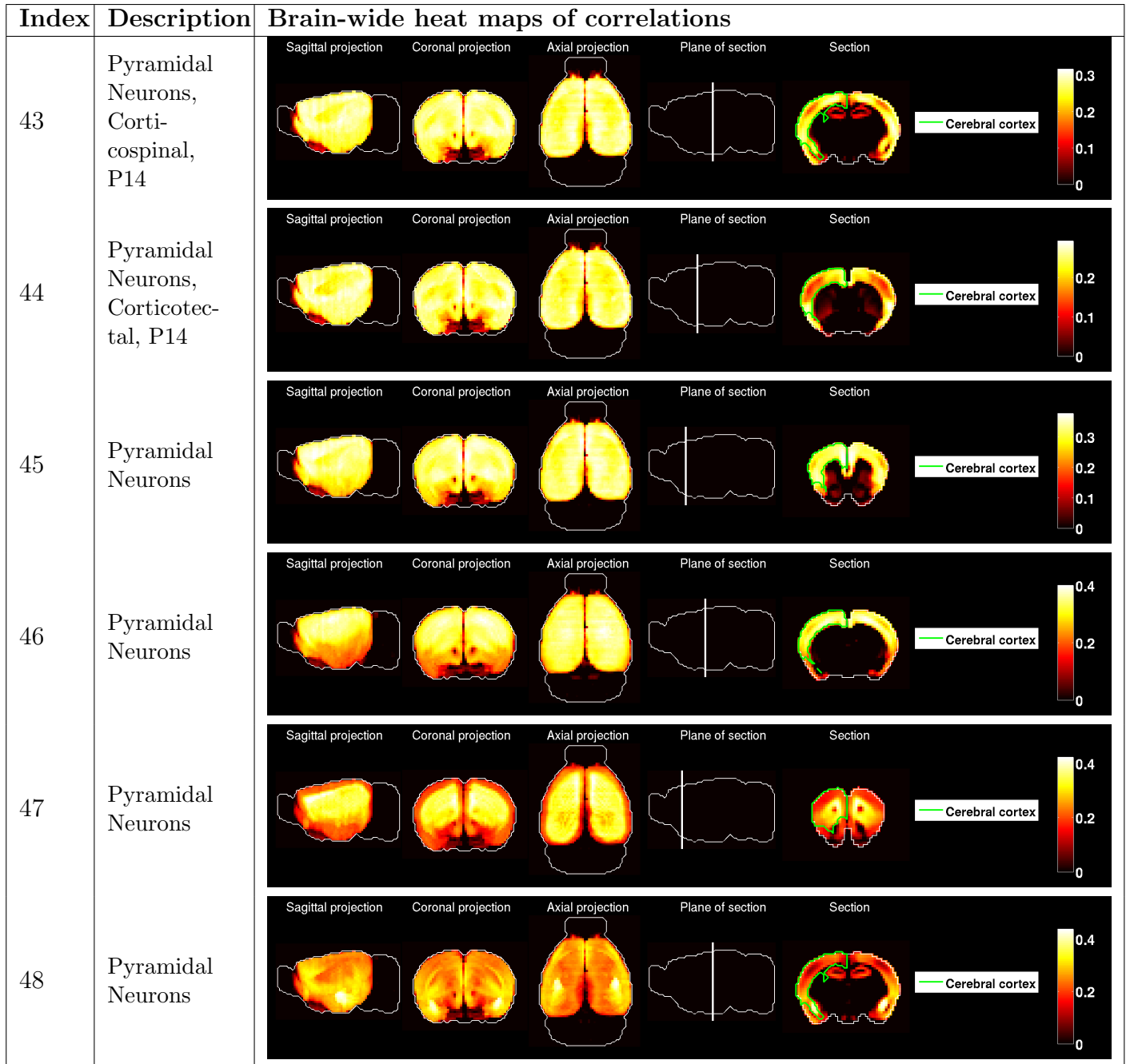


Table 47: Brain-wide correlation profiles between 6 cell types and the Allen Atlas.

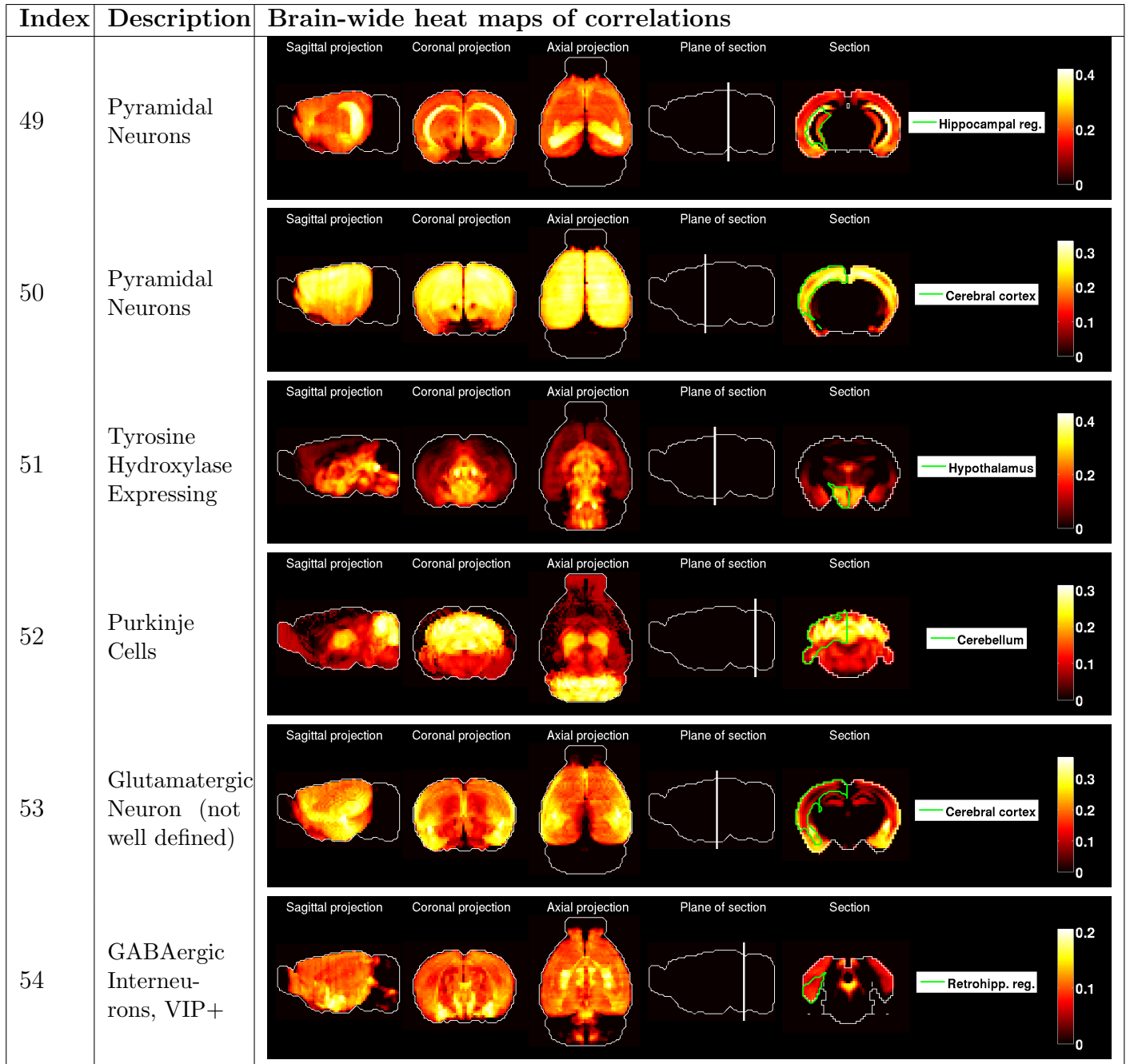


Table 48: Brain-wide correlation profiles between 6 cell types and the Allen Atlas.

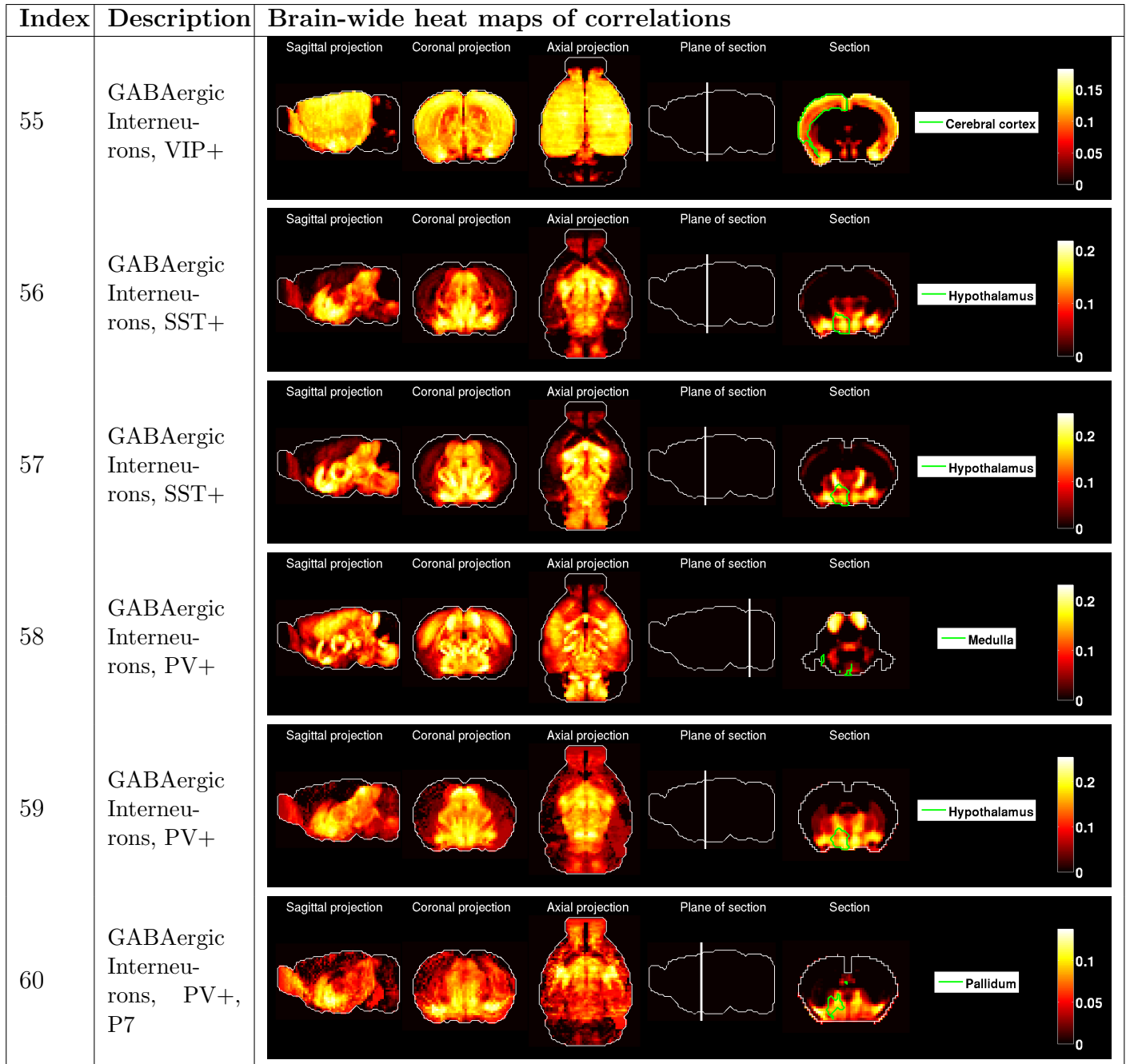


Table 49: Brain-wide correlation profiles between 6 cell types and the Allen Atlas.

Index	Description	Brain-wide heat maps of correlations
61	GABAergic Interneurons, PV+, P10	
62	GABAergic Interneurons, PV+, P13-P15	
63	GABAergic Interneurons, PV+, P25	
64	GABAergic Interneurons, PV+	

Table 50: Brain-wide correlation profiles between 4 cell types and the Allen Atlas.

## 9 Tables of estimated brain-wide densities

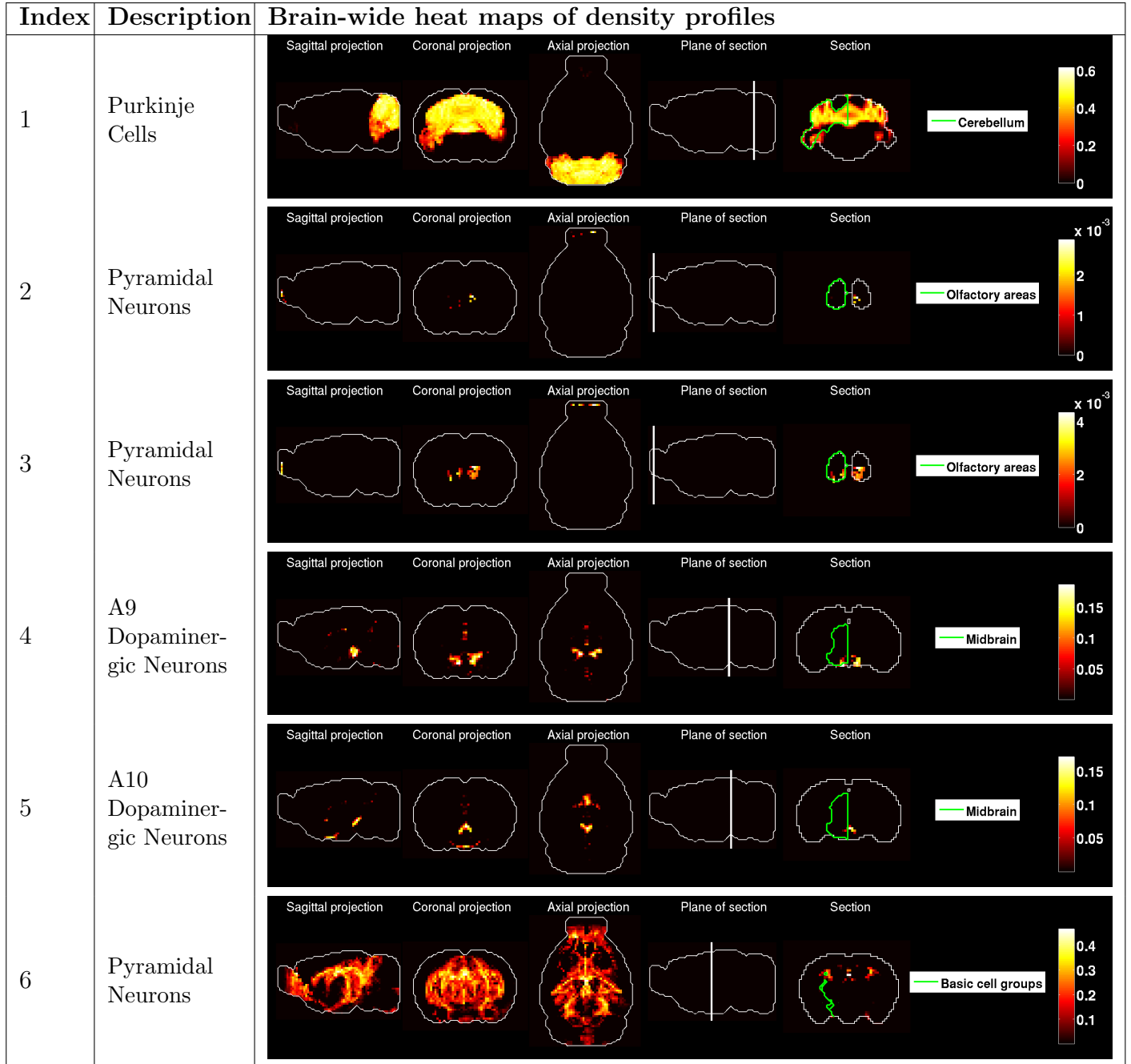


Table 51: Brain-wide density profiles of 6 cell types.

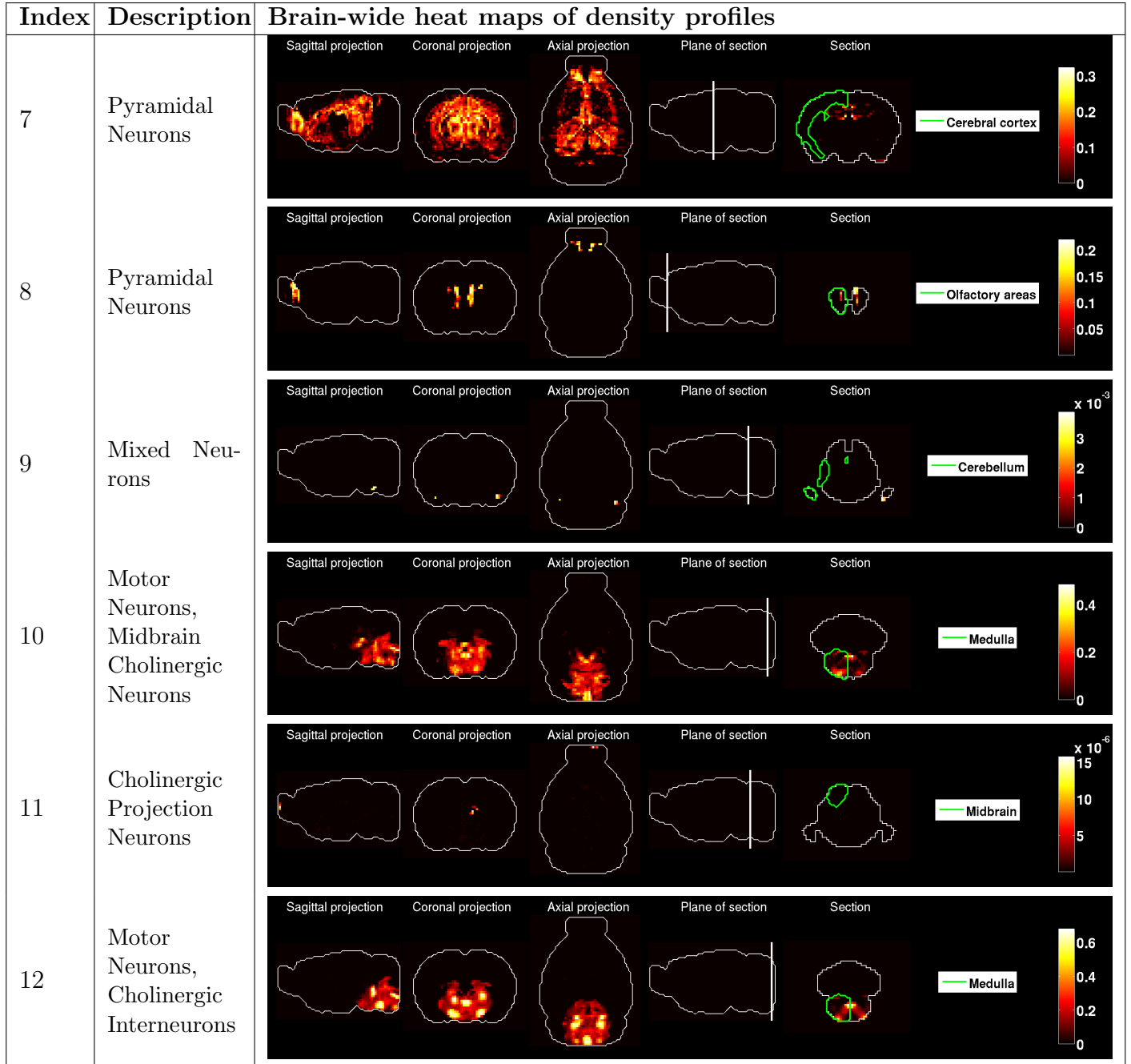


Table 52: Brain-wide density profiles of 6 cell types.

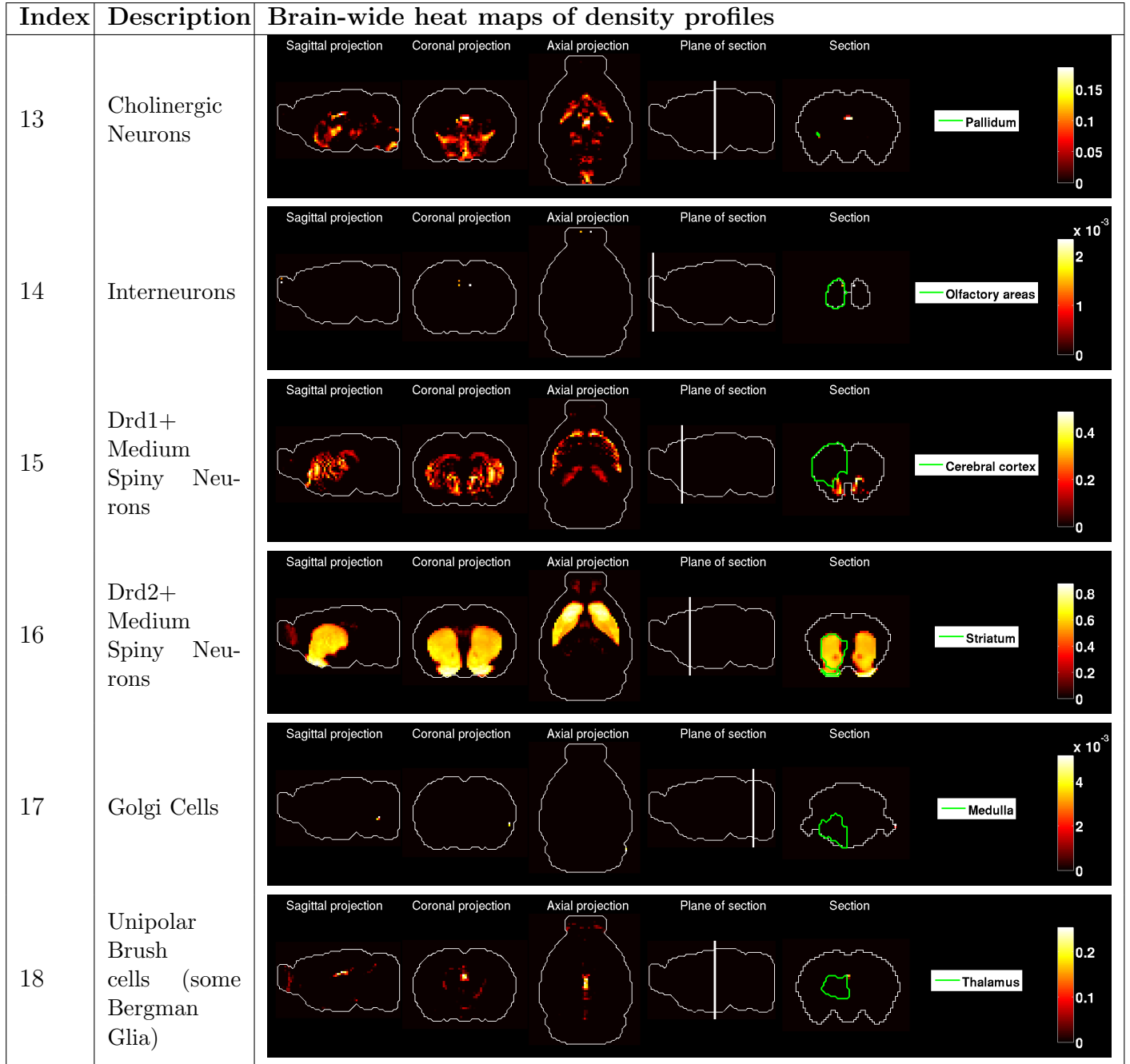


Table 53: Brain-wide density profiles of 6 cell types.



Index	Description	Brain-wide heat maps of density profiles
19	Stellate Basket Cells	
20	Granule Cells	
21	Mature Oligodendrocytes	
22	Mature Oligodendrocytes	
23	Mixed Oligodendrocytes	N/A
24	Mixed Oligodendrocytes	

Table 54: Brain-wide density profiles of 6 cell types.

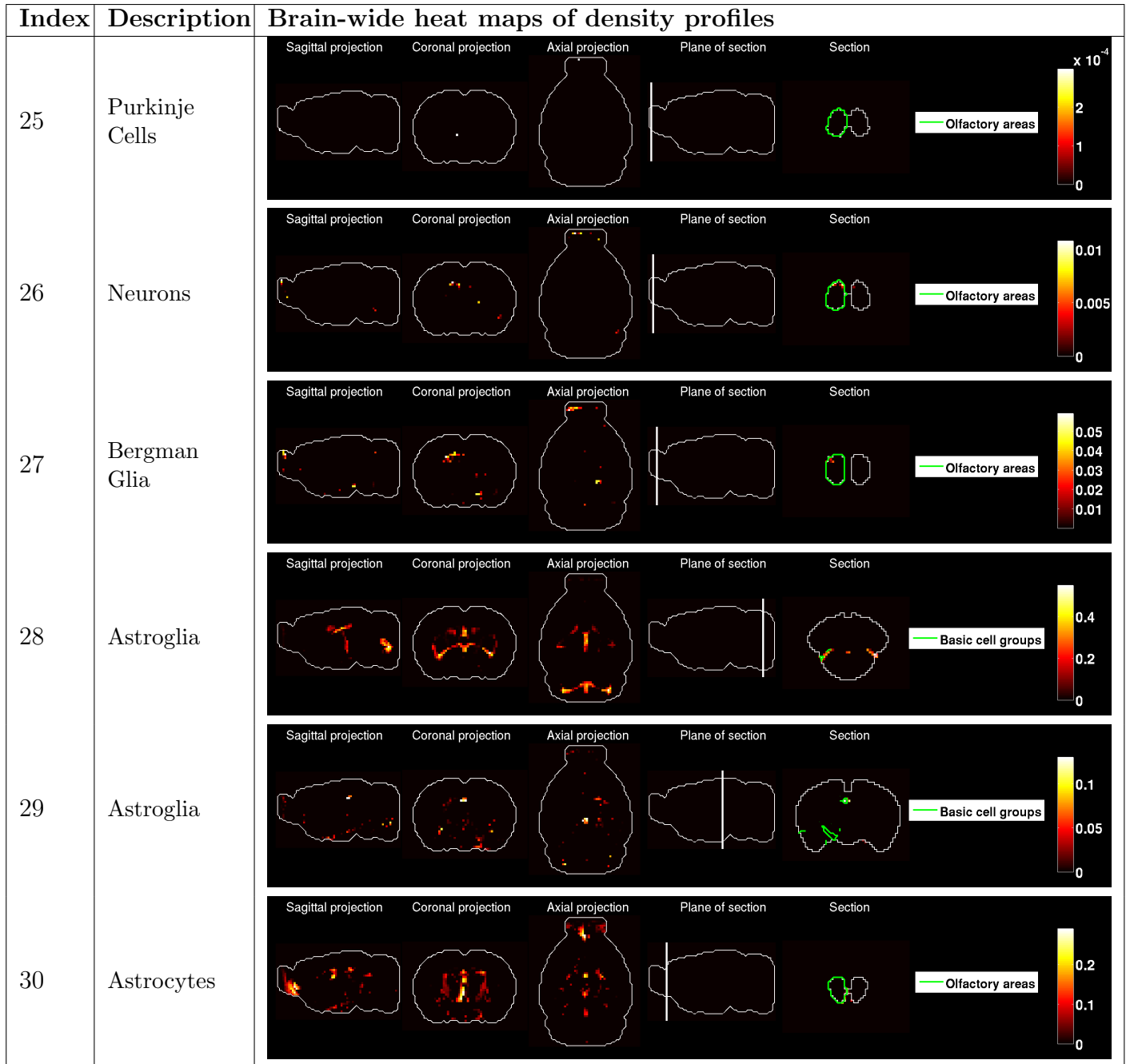


Table 55: Brain-wide density profiles of 6 cell types.

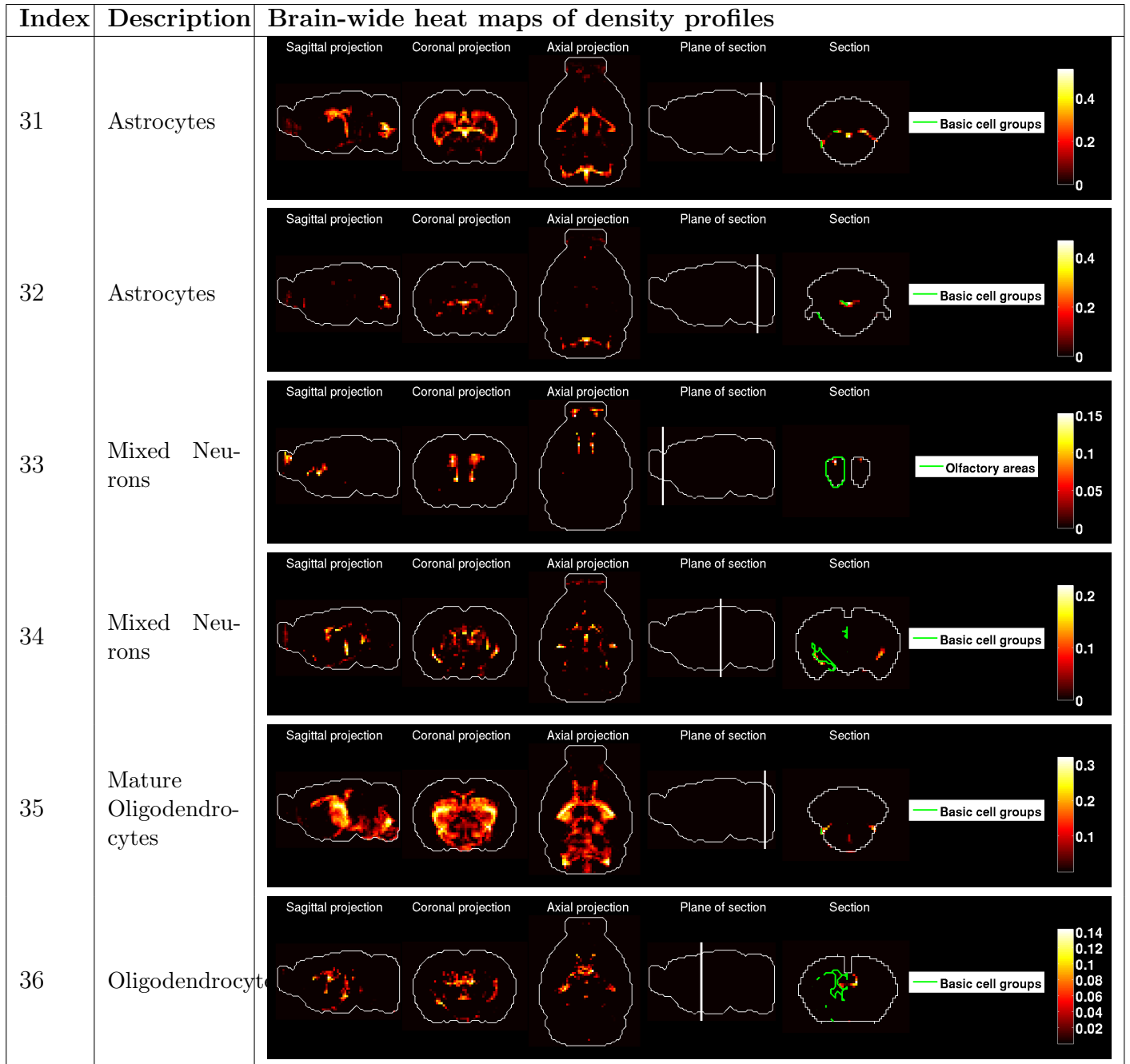


Table 56: Brain-wide density profiles of 6 cell types.

Index	Description	Brain-wide heat maps of density profiles
37	Oligodendrocyte Precursors	
38	Pyramidal Neurons, Callosally projecting, P3	
39	Pyramidal Neurons, Callosally projecting, P6	
40	Pyramidal Neurons, Callosally projecting, P14	
41	Pyramidal Neurons, Corti-cospinal, P3	
42	Pyramidal Neurons, Corti-cospinal, P6	

Table 57: Brain-wide density profiles of 6 cell types.

Index	Description	Brain-wide heat maps of density profiles
43	Pyramidal Neurons, Corticospinal, P14	
44	Pyramidal Neurons, Corticotectal, P14	
45	Pyramidal Neurons	
46	Pyramidal Neurons	
47	Pyramidal Neurons	
48	Pyramidal Neurons	

Table 58: Brain-wide density profiles of 6 cell types.

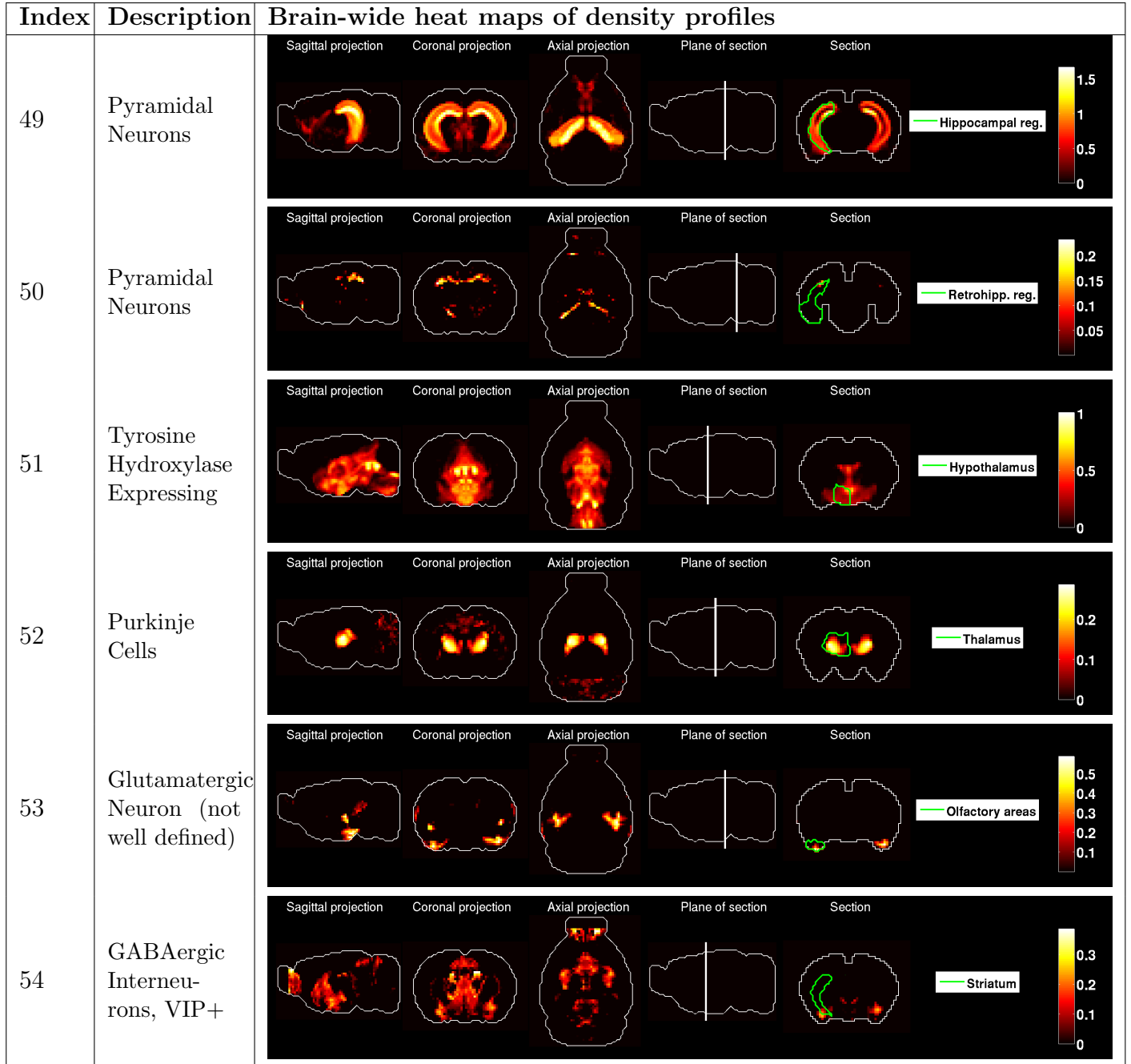


Table 59: Brain-wide density profiles of 6 cell types.

Index	Description	Brain-wide heat maps of density profiles
55	GABAergic Interneurons, VIP+	
56	GABAergic Interneurons, SST+	
57	GABAergic Interneurons, SST+	
58	GABAergic Interneurons, PV+	
59	GABAergic Interneurons, PV+	
60	GABAergic Interneurons, PV+, P7	

Table 60: Brain-wide density profiles of 6 cell types.

Index	Description	Brain-wide heat maps of density profiles
61	GABAergic Interneurons, PV+, P10	
62	GABAergic Interneurons, PV+, P13-P15	
63	GABAergic Interneurons, PV+, P25	
64	GABAergic Interneurons, PV+	

Table 61: Brain-wide fitting profiles between 4 cell-types and the Allen Atlas.



## 10 Tables of description of cell-type specific samples

Index	Description	Age of animal	Reference
1	Purkinje Cells	~P60	[17]
2	Pyramidal Neurons	~P60	[17]
3	Pyramidal Neurons	~P60	[17]
4	A9 Dopaminergic Neurons	Adult (precise age not given)	[20]
5	A10 Dopaminergic Neurons	Adult (precise age not given)	[20]
6	Pyramidal Neurons	Adult (precise age not given)	[19]
7	Pyramidal Neurons	Adult (precise age not given)	[19]
8	Pyramidal Neurons	Adult (precise age not given)	[19]
9	Mixed Neurons	Adult (precise age not given)	[19]
10	Motor Neurons, Midbrain Cholinergic Neurons	Adult (precise age not given)	[19]
11	Cholinergic Projection Neurons	Adult (precise age not given)	[19]
12	Motor Neurons, Cholinergic Interneurons	Adult (precise age not given)	[19]
13	Cholinergic Neurons	Adult (precise age not given)	[19]
14	Interneurons	Adult (precise age not given)	[19]
15	Drd1+ Medium Spiny Neurons	Adult (precise age not given)	[19]
16	Drd2+ Medium Spiny Neurons	Adult (precise age not given)	[19]
17	Golgi Cells	Adult (precise age not given)	[19]
18	Unipolar Brush cells (some Bergman Glia)	Adult (precise age not given)	[19]
19	Stellate Basket Cells	Adult (precise age not given)	[19]
20	Granule Cells	Adult (precise age not given)	[19]
21	Mature Oligodendrocytes	Adult (precise age not given)	[19]
22	Mature Oligodendrocytes	Adult (precise age not given)	[19]
23	Mixed Oligodendrocytes	Adult (precise age not given)	[19]
24	Mixed Oligodendrocytes	Adult (precise age not given)	[19]
25	Purkinje Cells	Adult (precise age not given)	[19]
26	Neurons	Adult (precise age not given)	[19]
27	Bergman Glia	Adult (precise age not given)	[19]
28	Astroglia	Adult (precise age not given)	[19]
29	Astroglia	Adult (precise age not given)	[19]
30	Astrocytes	P7-P8	[20]
31	Astrocytes	P17	[20]
32	Astrocytes	P17	[20]
33	Mixed Neurons	P7	[20]
34	Mixed Neurons	P16	[20]
35	Mature Oligodendrocytes	P17	[20]
36	Oligodendrocytes	P7	[20]
37	Oligodendrocyte Precursors	P7	[20]

Table 62: Description of cell-type-specific samples (I).

Index	Description	Age of animal	Reference
38	Pyramidal Neurons, Callosally projecting, P3	P3	[21]
39	Pyramidal Neurons, Callosally projecting, P6	P6	[21]
40	Pyramidal Neurons, Callosally projecting, P14	P14	[21]
41	Pyramidal Neurons, Corticospinal, P3	P3	[21]
42	Pyramidal Neurons, Corticospinal, P6	P6	[21]
43	Pyramidal Neurons, Corticospinal, P14	P14	[21]
44	Pyramidal Neurons, Corticotectal, P14	P14	[21]
45	Pyramidal Neurons	~P60	[23]
46	Pyramidal Neurons	~P60	[23]
47	Pyramidal Neurons	~P60	[23]
48	Pyramidal Neurons	~P60	[23]
49	Pyramidal Neurons	~P60	[23]
50	Pyramidal Neurons	~P60	Unpublished
51	Tyrosine Hydroxylase Expressing	~P45	Unpublished
52	Purkinje Cells	~P45	Unpublished
53	Glutamatergic Neuron (not well defined)	~P60	[23]
54	GABAergic Interneurons, VIP+	~P60	[23]
55	GABAergic Interneurons, VIP+	~P60	[23]
56	GABAergic Interneurons, SST+	~P60	[23]
57	GABAergic Interneurons, SST+	~P60	[23]
58	GABAergic Interneurons, PV+	~P60	[23]
59	GABAergic Interneurons, PV+	~P60	[23]
60	GABAergic Interneurons, PV+, P7	P7	[16]
61	GABAergic Interneurons, PV+, P10	P10	[16]
62	GABAergic Interneurons, PV+, P13-P15	P15	[16]
63	GABAergic Interneurons, PV+, P25	P25	[16]
64	GABAergic Interneurons, PV+	~P45	[16]

Table 63: Description of cell-type-specific samples (II).

## 11 Tables of anatomical origin of the cell-type-specific samples

For each of the cell-type-specific samples analyzed in this study, the following two tables give a brief description of the cell types, the region from which those cell types were extracted according to the coarsest (or 'big12') version of the Allen Reference Atlas, and the finest region to which it can be assigned according to the data provided in the studies [16, 17, 18, 19, 20, 21, 22, 23].

Index	Description	Region in the ARA ('big12')	Finest label in the ARA
1	Purkinje Cells	Cerebellum	Cerebellar cortex
2	Pyramidal Neurons	Cerebral cortex	Primary motor area; Layer 5
3	Pyramidal Neurons	Cerebral cortex	Primary somatosensory area; Layer 5
4	A9 Dopaminergic Neurons	Midbrain	Substantia nigra_ compact part
5	A10 Dopaminergic Neurons	Midbrain	Ventral tegmental area
6	Pyramidal Neurons	Cerebral cortex	Cerebral cortex; Layer 5
7	Pyramidal Neurons	Cerebral cortex	Cerebral cortex; Layer 5
8	Pyramidal Neurons	Cerebral cortex	Cerebral cortex; Layer 6
9	Mixed Neurons	Cerebral cortex	Cerebral cortex
10	Motor Neurons, Midbrain Cholinergic Neurons	Midbrain	Peduncolopontine nucleus
11	Cholinergic Projection Neurons	Pallidum	Pallidum_ ventral region
12	Motor Neurons, Cholinergic Interneurons	Medulla	Spinal cord
13	Cholinergic Neurons	Striatum	Striatum
14	Interneurons	Cerebral cortex	Cerebral cortex
15	Drd1+ Medium Spiny Neurons	Striatum	Striatum
16	Drd2+ Medium Spiny Neurons	Striatum	Striatum
17	Golgi Cells	Cerebellum	Cerebellar cortex
18	Unipolar Brush cells (some Bergman Glia)	Cerebellum	Cerebellar cortex
19	Stellate Basket Cells	Cerebellum	Cerebellar cortex
20	Granule Cells	Cerebellum	Cerebellar cortex
21	Mature Oligodendrocytes	Cerebellum	Cerebellar cortex
22	Mature Oligodendrocytes	Cerebral cortex	Cerebral cortex
23	Mixed Oligodendrocytes	Cerebellum	Cerebellar cortex
24	Mixed Oligodendrocytes	Cerebral cortex	Cerebral cortex
25	Purkinje Cells	Cerebellum	Cerebellar cortex
26	Neurons	Cerebral cortex	Cerebral cortex
27	Bergman Glia	Cerebellum	Cerebellar cortex
28	Astroglia	Cerebellum	Cerebellar cortex
29	Astroglia	Cerebral cortex	Cerebral cortex
30	Astrocytes	Cerebral cortex	Cerebral cortex
31	Astrocytes	Cerebral cortex	Cerebral cortex
32	Astrocytes	Cerebral cortex	Cerebral cortex
33	Mixed Neurons	Cerebral cortex	Cerebral cortex
34	Mixed Neurons	Cerebral cortex	Cerebral cortex
35	Mature Oligodendrocytes	Cerebral cortex	Cerebral cortex
36	Oligodendrocytes	Cerebral cortex	Cerebral cortex
37	Oligodendrocyte Precursors	Cerebral cortex	Cerebral cortex

Table 64: Anatomical origin of the cell-type-specific samples (I).

Index	Description	Region in the ARA ('big12')	Finest label in the ARA
38	Pyramidal Neurons, Callosally projecting, P3	Cerebral cortex	Cerebral cortex
39	Pyramidal Neurons, Callosally projecting, P6	Cerebral cortex	Cerebral cortex
40	Pyramidal Neurons, Callosally projecting, P14	Cerebral cortex	Cerebral cortex
41	Pyramidal Neurons, Corticospinal, P3	Cerebral cortex	Cerebral cortex
42	Pyramidal Neurons, Corticospinal, P6	Cerebral cortex	Cerebral cortex
43	Pyramidal Neurons, Corticospinal, P14	Cerebral cortex	Cerebral cortex
44	Pyramidal Neurons, Corticotectal, P14	Cerebral cortex	Cerebral cortex
45	Pyramidal Neurons	Cerebral cortex	Cerebral cortex, Layer 5
46	Pyramidal Neurons	Cerebral cortex	Cerebral cortex, Layer 5
47	Pyramidal Neurons	Cerebral cortex	Primary somatosensory area; Layer 5
48	Pyramidal Neurons	Cerebral cortex	Prelimbic area and Infralimbic area; Layer 5 (Amygdala)
49	Pyramidal Neurons	Hippocampal region	Ammon's Horn; Layer 6B
50	Pyramidal Neurons	Cerebral cortex	Primary motor area
51	Tyrosine Hydroxylase Expressing	Pons	Pontine central gray
52	Purkinje Cells	Cerebellum	Cerebellar cortex
53	Glutamatergic Neuron (not well defined)	Cerebral cortex	Cerebral cortex; Layer 6B (Amygdala)
54	GABAergic Interneurons, VIP+	Cerebral cortex	Prelimbic area and Infralimbic area
55	GABAergic Interneurons, VIP+	Cerebral cortex	Primary somatosensory area
56	GABAergic Interneurons, SST+	Cerebral cortex	Prelimbic area and Infralimbic area
57	GABAergic Interneurons, SST+	Hippocampal region	Ammon's Horn
58	GABAergic Interneurons, PV+	Cerebral cortex	Prelimbic area and Infralimbic area
59	GABAergic Interneurons, PV+	Thalamus	Dorsal part of the lateral geniculate complex
60	GABAergic Interneurons, PV+, P7	Cerebral cortex	Primary somatosensory area
61	GABAergic Interneurons, PV+, P10	Cerebral cortex	Primary somatosensory area
62	GABAergic Interneurons, PV+, P13-P15	Cerebral cortex	Primary somatosensory area
63	GABAergic Interneurons, PV+, P25	Cerebral cortex	Primary somatosensory area
64	GABAergic Interneurons, PV+	Cerebral cortex	Primary motor area

Table 65: Anatomical origin of the cell-type-specific samples (II).

## References

- [1] M. Bota, H.-W. Dong, L.W. Swanson, *From gene networks to brain networks*, Nature neuroscience (2003) **6** (8), 795–9.
- [2] E.S. Lein, M. Hawrylycz, N. Ao, M. Ayres, A. Bensinger, A. Bernard, A.F. Boe, M.S. Boguski, K.S. Brockway, E.J. Byrnes, L. Chen, L. Chen, T.M. Chen, M.C. Chin, J. Chong, B.E. Crook, A. Czaplinska, C.N. Dang, S. Datta, N.R. Dee, *et al.*, *Genome-wide atlas of gene expression in the adult mouse brain*. Nature **445**, 168–176 (2007).
- [3] L. Ng, M. Hawrylycz, D. Haynor, *Automated high-throughput registration for localizing 3D mouse brain gene expression using ITK*, Insight-Journal (2005).
- [4] S.M. Sunkin and J.G. Hohmann, *Insights from spatially mapped gene expression in the mouse brain*, Human Molecular Genetics, 2007, Vol. 16, Review Issue 2.
- [5] L. Ng, S.D. Pathak, C. Kuan, C. Lau, H. Dong, A. Sodt, C. Dang, B. Avants, P. Yushkevich, J.C. Gee, D. Haynor, E. Lein, A. Jones and M. Hawrylycz, *Neuroinformatics for genome-wide 3D gene expression mapping in the mouse brain*, IEEE/ACM Trans. Comput. Biol. Bioinform. (2007), Jul-Sep **4**(3) 382–93.
- [6] M. Hawrylycz, R.A. Baldock, A. Burger, T. Hashikawa, G.A. Johnson, M. Martone, L. Ng, C. Lau, S.D. Larsen, J. Nissanov, L. Puellas, S. Ruffins, F. Verbeek, I. Zaslavsky<sup>1</sup>, J. Boline, *Digital Atlasing and Standardization in the Mouse Brain*, PLoS Computational Biology **7** (2) (2011).
- [7] A.R. Jones, C.C. Overly and S.M. Sunkin, *The Allen Brain Atlas: 5 years and beyond*, Nature Reviews (Neuroscience), Volume **10** (November 2009), **1**.
- [8] M. Hawrylycz, L. Ng, D. Page, J. Morris, C. Lau, S. Faber, V. Faber, S. Sunkin, V. Menon, E.S. Lein, A. Jones, *Multi-scale correlation structure of gene expression in the brain*, Neural Networks **24** (2011) 933–942.
- [9] H.-W. Dong, *The Allen reference atlas: a digital brain atlas of the C57BL/6J male mouse*, Wiley, 2007.
- [10] L. Ng, A. Bernard, C. Lau, C.C. Overly, H.-W. Dong, C. Kuan, S. Pathak, S.M. Sunkin, C. Dang, J.W. Bohland, H. Bokil, P.P. Mitra, L. Puellas, J. Hohmann, D.J. Anderson, E.S. Lein, A.R. Jones, M. Hawrylycz, *An anatomic gene expression atlas of the adult mouse brain*, Nature Neuroscience **12**, 356–362 (2009).
- [11] L. Ng *et al.*, *NeuroBlast: a 3D spatial homology search tool for gene expression*, BMC Neuroscience 2007, **8**(Suppl 2):P11.
- [12] P. Grange, J.W. Bohland, M. Hawrylycz and P.P. Mitra, *Brain Gene Expression Analysis: a MATLAB toolbox for the analysis of brain-wide gene-expression data*, [arXiv:1211.6177 [q-bio.QM]].

- [13] The Brain Gene Expression Atlas MATLAB toolbox is downloadable from <http://brainarchitecture.org/allen-atlas-brain-toolbox>.
- [14] P. Grange, M. Hawrylycz, P.P. Mitra, *Computational neuroanatomy and co-expression of genes in the adult mouse brain, analysis tools for the Allen Brain Atlas*, Quantitative Biology (2013), in press, [arXiv:1301.1730 [q-bio.QM]].
- [15] P. Grange, P.P. Mitra, *Computational neuroanatomy and gene expression: optimal sets of marker genes for brain regions*, IEEE, in CISS 2012, 46th annual conference on Information Science and Systems (Princeton).
- [16] B.W. Okaty, M.N. Miller, K. Sugino, C.M. Hempel, S.B. Nelson, *Transcriptional and electrophysiological maturation of neocortical fast-spiking GABAergic interneurons*, J. Neurosci. (2009) **29(21)** 7040-52.
- [17] M.J. Rossner, J. Hirrlinger, S.P. Wichert, C. Boehm, D. Newrzella, H. Hiemisch, G. Eisenhardt, C. Stuenkel, O. von Ahsen, K.A. Nave, *Global transcriptome analysis of genetically identified neurons in the adult cortex*, J. Neurosci. 2006 **26(39)** 9956-66.
- [18] J.D. Cahoy, B. Emery, A. Kaushal, L.C. Foo, J.L. Zamanian, K.S. Christopherson, Y. Xing, J.L. Lubischer, P.A. Krieg, S.A. Krupenko, W.J. Thompson, B.A. Barres, *A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function*, J. Neurosci. 2008 **28(1)** 264-78.
- [19] J.P. Doyle, J.D. Dougherty, M. Heiman, E.F. Schmidt, T.R. Stevens, G. Ma, S. Bupp, P. Shrestha, R.D. Shah, M.L. Doughty, S. Gong, P. Greengard, N. Heintz, *Application of a translational profiling approach for the comparative analysis of CNS cell types*, Cell (2008) **135(4)** 749-62.
- [20] C.Y. Chung, H. Seo, K.C. Sonntag, A. Brooks, L. Lin, O. Isacson *Cell type-specific gene expression of midbrain dopaminergic neurons reveals molecules involved in their vulnerability and protection*. Hum. Mol. Genet. (2005) **14**: 1709–1725.
- [21] P. Arlotta, B.J. Molyneaux, J. Chen, J. Inoue, R. Kominami *et al.* (2005) *Neuronal subtype-specific genes that control corticospinal motor neuron development in vivo*, Neuron **45**: 207–221.
- [22] M. Heiman, A. Schaefer, S. Gong, Peterson JD, Day M, Ramsey KE, Surez-Farias M, Schwarz C, Stephan DA, Surmeier DJ, P. Greengard, N. Heintz, (2008) *A translational profiling approach for the molecular characterization of of CNS cell types*, Cell **135**: 738–748.
- [23] K. Sugino, C.M. Hempel, M.N. Miller, A.M. Hattox, P. Shapiro, C. Wu, Z.J. Huang, S.B. Nelson, *Molecular taxonomy of major neuronal classes in the adult mouse forebrain*, Nature Neuroscience **9**, 99-107 (2005).



- [24] B.W. Okaty, K. Sugino, S.B. Nelson, *A Quantitative Comparison of Cell-Type-Specific Microarray Gene Expression Profiling Methods in the Mouse Brain*, PLoS One (2011) **6(1)**.
- [25] P. Grange, J.W. Bohland, H. Bokil, S. Nelson, B. Okaty, K. Sugino, L. Ng, M. Hawrylycz and P.P. Mitra, *A cell-type based model explaining co-expression patterns of genes in the brain*, [[arXiv:1111.6217] [q-bio.QM]].
- [26] C.K. Lee, S.M. Sunkin, C.C. Kuan, C.L. Thompson, S. Pathak, L. Ng, C. Lau, S. Fischer, M. Mortrud, C. Slaughterbeck, A. Jones, E. Lein, M. Hawrylycz, *Quantitative methods for genome-scale analysis of in situ hybridization and correlation with microarray data*, Genome Biol. (2008); **9(1)**: R23.
- [27] Research, Inc. CVX: *Matlab software for disciplined convex programming*, version 2.0 beta. <http://cvxr.com/cvx>, September 2012.
- [28] M. Grant and S. Boyd, *Graph implementations for nonsmooth convex programs*, *Recent Advances in Learning and Control (a tribute to M. Vidyasagar)*, V. Blondel, S. Boyd, and H. Kimura, editors, pages 95-110, Lecture Notes in Control and Information Sciences, Springer, 2008. [http://stanford.edu/~boyd/graph\\_dcp.html](http://stanford.edu/~boyd/graph_dcp.html).
- [29] J.W. Bohland, H. Bokil, C.B. Allen, P.P. Mitra, *The Brain Atlas Concordance Problem: Quantitative Comparison of Anatomical Parcellations*, PLoS ONE (2009).
- [30] J.W. Bohland, H. Bokil, C.-K. Lee, L. Ng, C. Lau, C. Kuan, M. Hawrylycz, P.P. Mitra, *Clustering of spatial gene expression patterns in the mouse brain and comparison with classical neuroanatomy*, *Methods*, Volume **50**, Issue 2, February 2010, Pages 105-112.
- [31] I. Menashe, P. Grange, E.C. Larsen, S. Banerjee-Basu and P.P. Mitra, *Co-expression profiling of autism genes in the mouse brain*, PLoS Comput. Biol. 9(7): e1003128.