Supplementary Online Information S2

Analysis of raw data

Forty-two out of the 48 experts finished the experiment, and only data from these 42 experts are considered in the remainder of the analysis. We compared the categories assigned by the individual experts for each one of the six features (**Figs. S1-S6**). Some experts were 'outliers' in terms of their selections; for example, one expert categorized all the neurons as Intralaminar in Feature 1 (last expert in **Fig. S1**) and the same rater categorized almost all the neurons as Centered in Feature 3 (last expert in **Fig. S3**). With regard to Feature 5, high bars indicate that a high number of experts selected a particular category for a particular neuron in **Fig. S5**. On the contrary, short bars for a particular category and a particular neuron indicate that the corresponding neurons received very few votes in that neuron type. For example, it is possible to distinguish seven high bars for the Chandelier category indicating that experts agreed when assigning this particular category for those specific neurons. With regard to Feature 6 (**Fig. S6**), the majority of the experts considered that most of the neurons could be characterized and tried to classify them. Indeed, 35 out of 42 experts (83.33%) characterized more than 280 neurons, whereas two experts characterized less than 200 neurons (first two experts in **Fig. S6**).

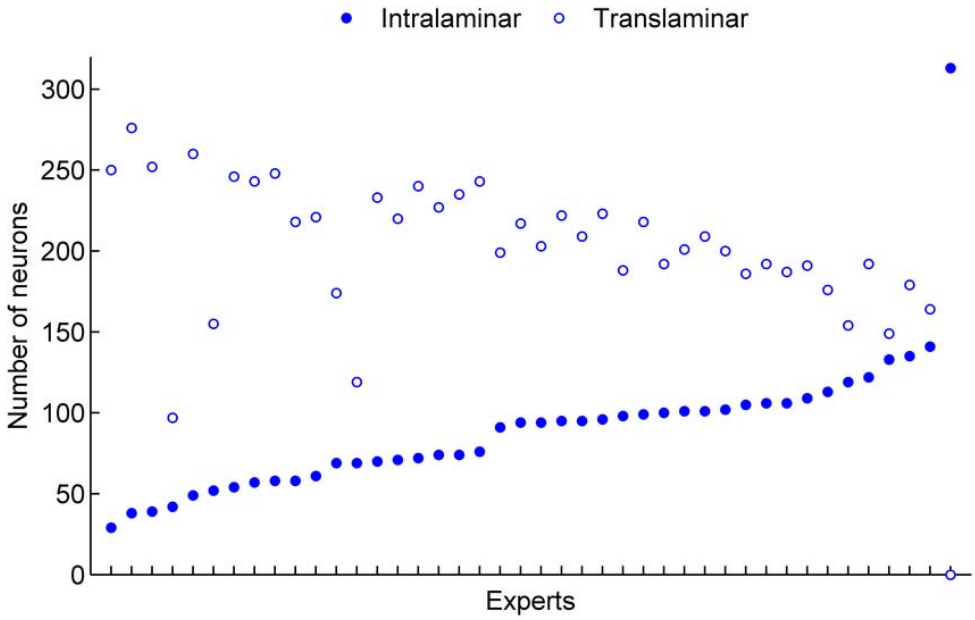

Fig. S1. Graphical representation of the ratings given to the different categories of Feature 1 by the 42 experts who completed the experiment. Experts are sorted in ascending order (in the x axis) based on the number of votes of the category Intralaminar.

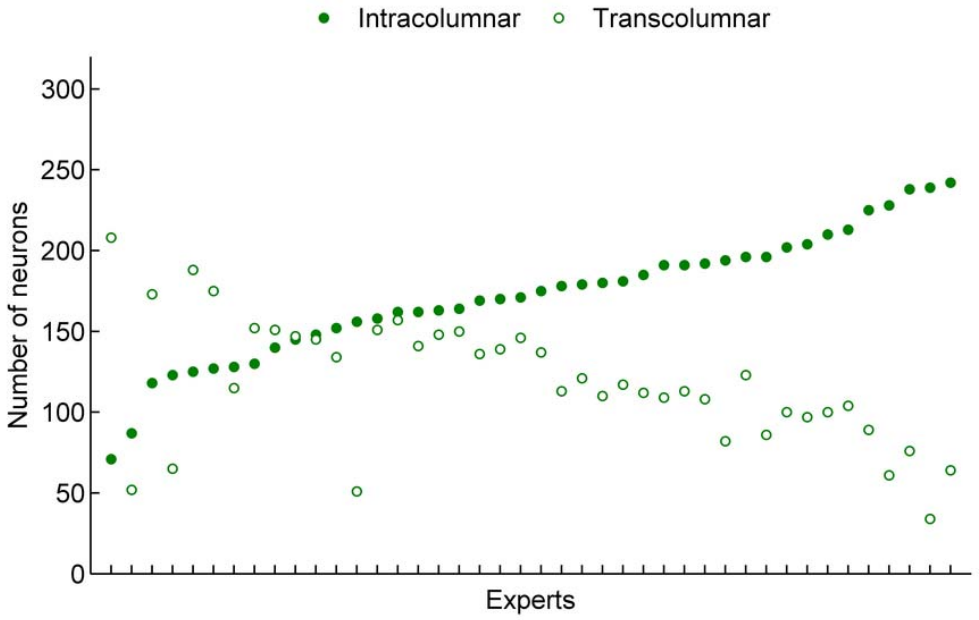

Fig. S2. Graphical representation of the ratings given to the different categories of Feature 2 by the 42 experts who completed the experiment. Experts are sorted in ascending order (in the x axis) based on the number of votes of the category Intracolumnar.

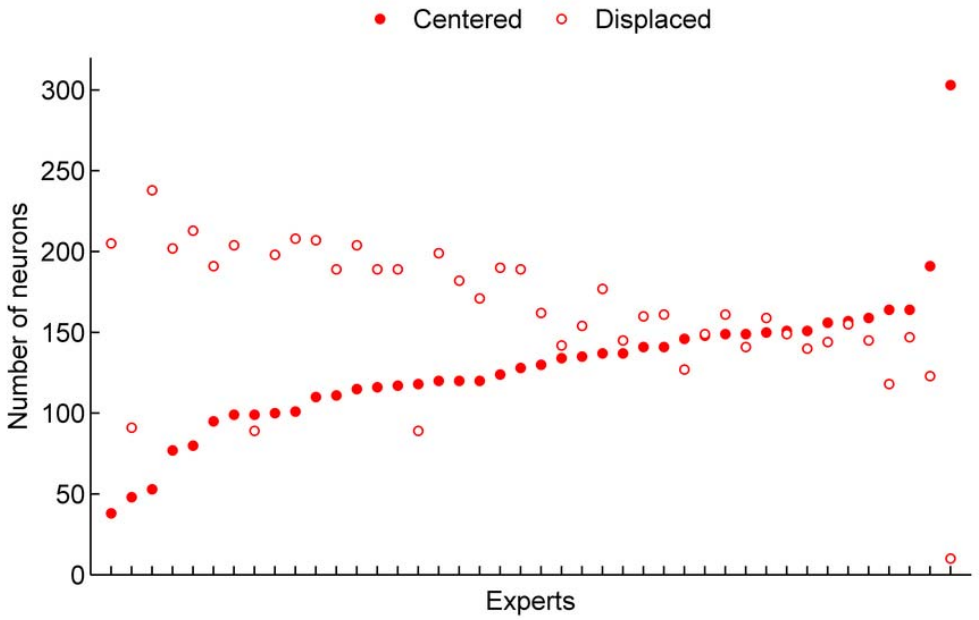

Fig. S3. Graphical representation of the ratings given to the different categories of Feature 3 by the 42 experts who completed the experiment. Experts are sorted in ascending order (in the x axis) based on the number of votes of the category Centered.

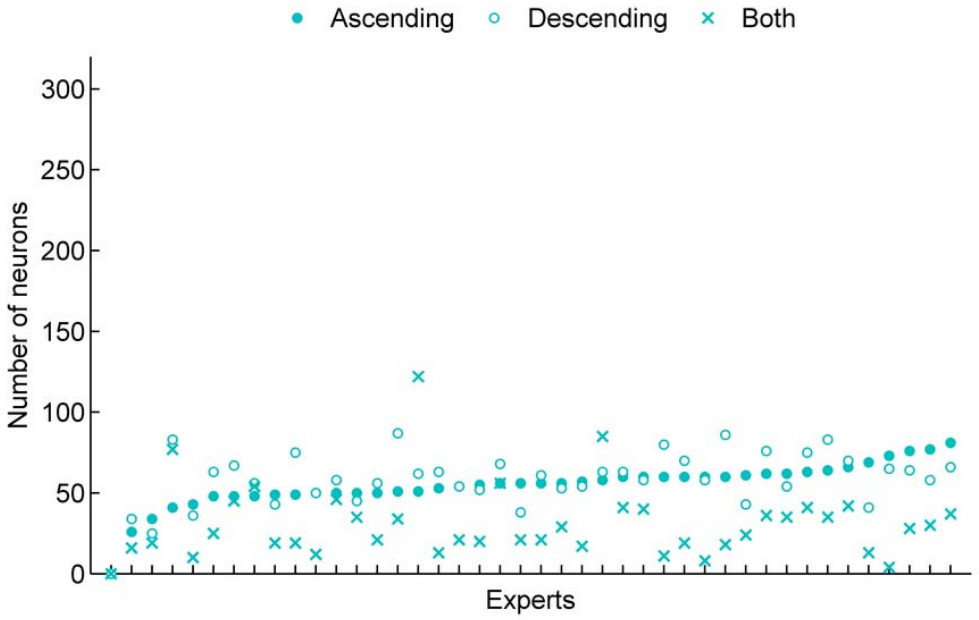

Fig. S4. Graphical representation of the ratings given to the different categories of Feature 4 by the 42 experts who completed the experiment. Experts are sorted in ascending order (in the x axis) based on the number of votes of the category Ascending.

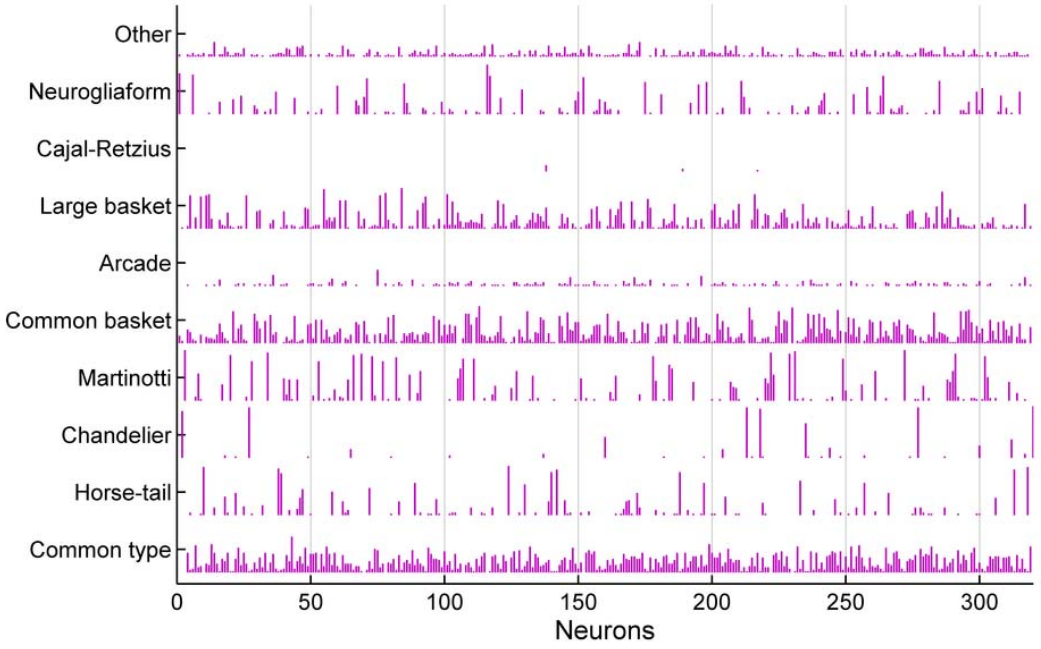

Fig. S5. Graphical representation of the number of experts who selected each category of Feature 5. A vertical bar is shown for each neuron and each category, representing the number of experts who selected that category for that neuron. High bars (e.g., for categories Chandelier, Horse-tail and Martinotti) show high agreement when classifying the neurons in these neuronal types. Contrarily, short bars (e.g., for categories Common type, Common basket, Large basket, Other, Arcade, etc.) represent low agreement.

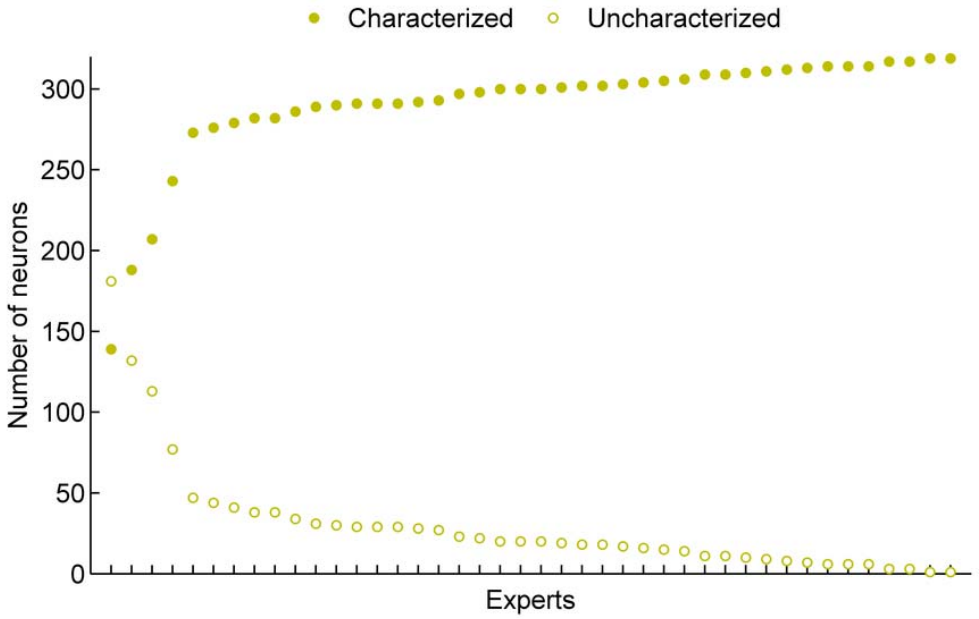

Fig. S6. Graphical representation of the ratings given to the different categories of Feature 6 by the 42 experts who completed the experiment. Experts are sorted in ascending order (in the x axis) based on the number of votes of the category Characterized.

Experts' agreement analysis

The difference between 'observed agreement' and chance-corrected Fleiss' pi index was particularly high for Feature 6 (F6), that is, for the decision on whether or not a neuron could be characterized; this feature had the highest observed agreement and the lowest Fleiss' pi value. We can detect frequent differences in the categories provided by some experts for this feature (see **Fig. S6**). In addition, more than 90% of the votes in Feature 6 were assigned to Characterized (see **Fig. S8A**), and such unbalanced prevalence tends to reduce the value of chance-corrected agreement indices. A permutation test reported statistically significant differences from chance agreement (uncorrected $p < 0.0001$) for all the features.

We also calculated the observed agreement (**Fig. S7A**) and Fleiss' pi index (**Fig. S7B**) for every category within each feature. The observed agreement for all categories in Feature 1, Feature 2 and Feature 3 was high, whereas Fleiss' pi values were lower. We also observed a high agreement for the categories Ascending and Descending in Feature 4, whereas agreement was lower for the category Both. For Feature 5, Chandelier, Horse-tail, and Martinotti were the most consensual interneuron types, with similar values for Fleiss' pi and the observed agreement. Experts' agreement values were low for the remaining categories, namely Arcade, Cajal-Retzius, and Other. With respect to Feature 6, the observed agreement was high for the category Characterized and relatively low for the category Uncharacterized. Some experts tried to characterize all the neurons whereas other experts frequently categorized them as Uncharacterized (**Fig. S6**). The differences in experts' biases and the unbalanced prevalence of the two categories explain the very low Fleiss' pi values for both Characterized and Uncharacterized categories. The values of the observed agreement and the category-

specific Fleiss' pi indices for all the categories of all the features significantly differed from chance agreement according to a permutation test (uncorrected $p < 0.0001$).

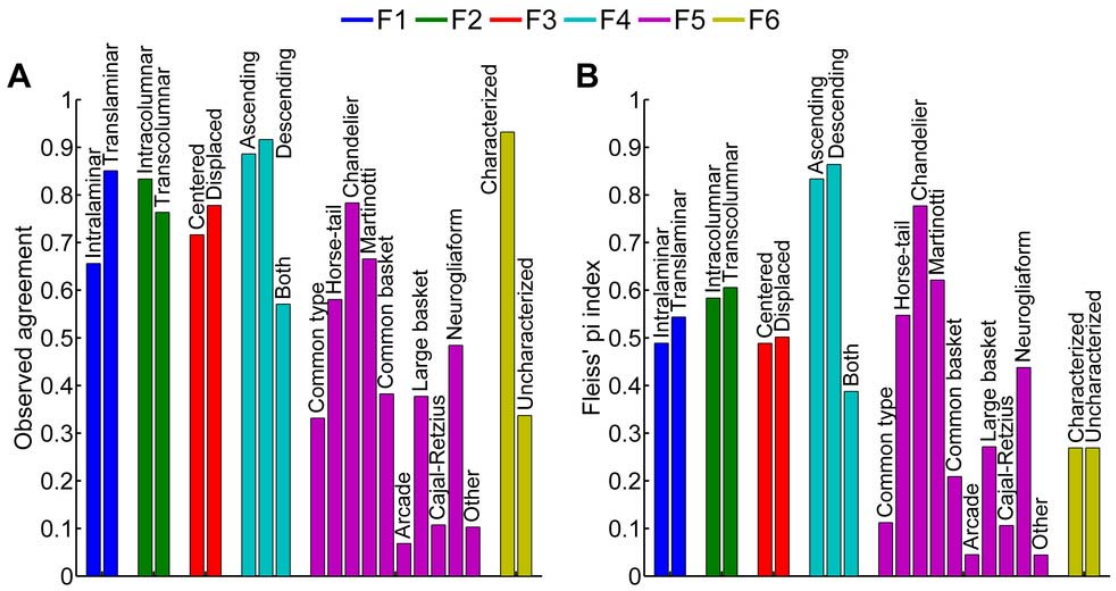

Fig. S7. (**A**) Observed agreement and (**B**) chance-corrected Fleiss' pi index for each category of every feature.

In a separate analysis, we tried to identify possible outliers in the group of experts by studying the influence of every one of the experts in the chance-corrected Fleiss' pi index computed for each feature (**Fig. S8A**). Additionally, we also removed groups of three experts in all possible combinations to further identify possible sets of experts contributing to low Fleiss' pi index values (**Fig. S8B**). Agreement increased for Features 1 and 3 when expert 33 was removed (as revealed by the small peak in the blue and red curves). This is consistent with the different selection of categories by this expert for this feature (**Fig. S1** and **S3**). Similarly, removing expert 23 increased the agreement for Feature 6, as shown by the peak in the ochre curve. The peaks in **Fig. S8B** corresponded to the subgroups of experts excluding expert 33 in Feature 1 and Feature 3, and expert 23 in Feature 6. For instance, this means that expert 33 selected categories for Feature 1 and 3 in a different way than the rest of the experts. The agreement for Feature 2, Feature 4 and Feature 5 did not vary when one or three experts were removed. The largest difference in Fleiss' pi index corresponded to the scenario where experts 23, 24 and 29 were removed in Feature 6. In this case, the agreement increased from 0.269 (when the 42 experts were considered) to 0.3628 (when 39 experts were considered). However, we did not remove any expert from the remainder of the analysis since there was not an expert (or a group of three experts) whose removal produced statistically significant Fleiss' pi index differences for all features.

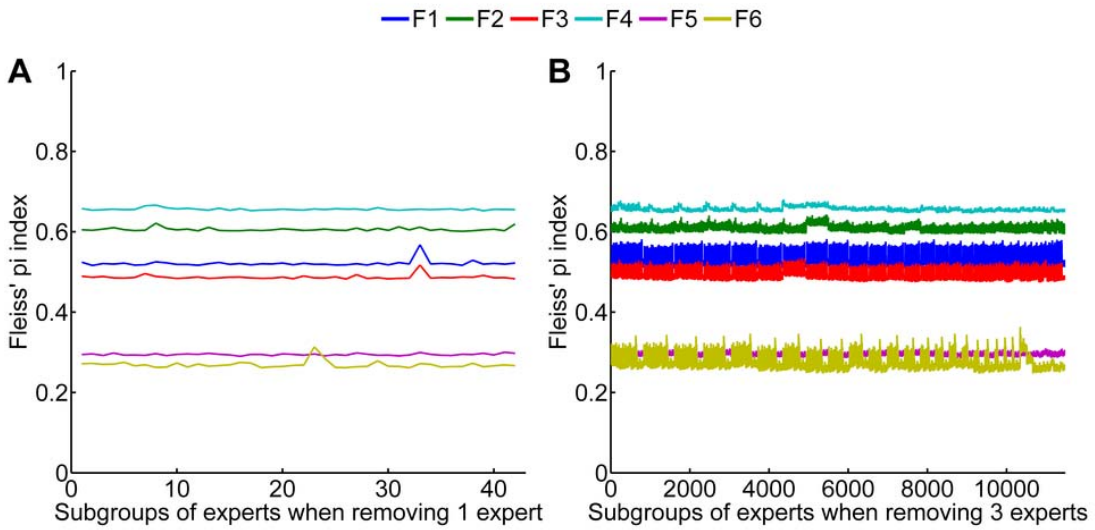

Fig. S8 (**A**) Fleiss' pi values for all the groups of experts obtained when removing one expert (42 possible subgroups) (**B**) and when three experts were removed (11,480 possible subgroups).

Next, we investigated whether the Fleiss' pi values increased or decreased when merging two categories of Feature 5. The rationale for this was to study possible overlapping between interneuron types. **Table S1** shows the values obtained in analyses in which a particular category (rows) was merged with another category (columns). The reference value obtained when all the interneuron types were considered as different categories was 0.2963 (**Fig. S8B**). Thus, Fleiss' pi values above this number will indicate categories that were confused with each other. Merging category Martinotti with any other category decreased Fleiss' pi value, with one exception, namely when categories Martinotti and Other were merged. The lowest Fleiss' pi value (0.2645) in **Table S1** was achieved when category Martinotti was merged with category Common basket. The Fleiss' pi value was also lower in all analyses in which category Chandelier was merged with any other category. Extending this analysis beyond pairwise merges, the Fleiss' pi value was lowest (0.2312) when categories Horse-tail, Martinotti, and Common basket were all merged together into a single category (not shown).

The highest Fleiss' pi value (0.3444) was achieved when categories Common type and Common basket were merged. Fleiss' pi value also increased when merging categories Common basket and Neurogliaform (0.3259), Common type and Large basket (0.3187), and Common basket and Large basket (0.3170). When we considered combinations of three neuronal types, the highest Fleiss' pi value (0.4110) was achieved when categories Common type, Common basket, and Large basket were merged into a single category (not shown).

Table S1. Fleiss' pi index values when a category of Feature 5 is merged with another category.

	Horse-tail	Chandelier	Martinotti	Common basket	Arcade	Large basket	Cajal-Retzius	Neurogliaform	Other
Common Type	0.2973	0.2891	0.2876	0.3444	0.3040	0.3187	0.2970	0.2836	0.3158
Horse-tail		0.2937	0.2854	0.2790	0.2969	0.2844	0.2962	0.2862	0.3102
Chandelier			0.2910	0.2922	0.2959	0.2909	0.2963	0.2944	0.2945
Martinotti				0.2645	0.2941	0.2916	0.2961	0.2803	0.2984

Common basket					0.3006	0.3170	0.2959	0.3259	0.2977
Arcade						0.3003	0.2963	0.2953	0.2982
Large basket							0.2973	0.2839	0.2961
Cajal-Retzius								0.2962	0.2965
Neuroglia form									0.2952

Additionally, Cohen's kappa index were computed for all the features along with two of its variants: the ratio between Cohen's kappa and its maximum value taking into account fixed marginals, and the Prevalence-Adjusted Bias-Adjusted kappa ($PABA_k$) index (see **Supplementary Online Information S1** for details). This analysis showed similar results to those obtained using previous agreement indices; thus, results are only shown for Feature 5 (**Figs. S9-S11**). The level of agreement between pairs of experts was highest for category Martinotti, followed by Chandelier, Horse-tail, and Neurogliaform categories. In contrast, there was low agreement for categories Common type (**Fig. S9A**), Common basket (**Fig. S9E**), and Large basket (**Fig. S9G**). The agreement values of the ratio k/k_{max} and $PABA_k$ were similarly low for these categories (**Figs. S10** and **S11**). However, Arcade and Other categories yielded low agreements for Cohen's kappa and the ratio between Cohen's kappa and its maximum value (**Fig. S9F-J** and **Fig. S10F-J**), whereas the agreement values of PABA were high (**Fig. S11F-J**). The low agreement found in these two categories is probably due to the low number of votes assigned by the experts to these categories. In fact, Arcade was the second category (after Cajal-Retzius) with fewest votes. Since PABA corrects for the differences in the number of votes, it yields much higher values (**Fig. S11**) than Cohen's kappa or the ratio between Cohen's kappa and its maximum value. Similar conclusions can be drawn for category Other.

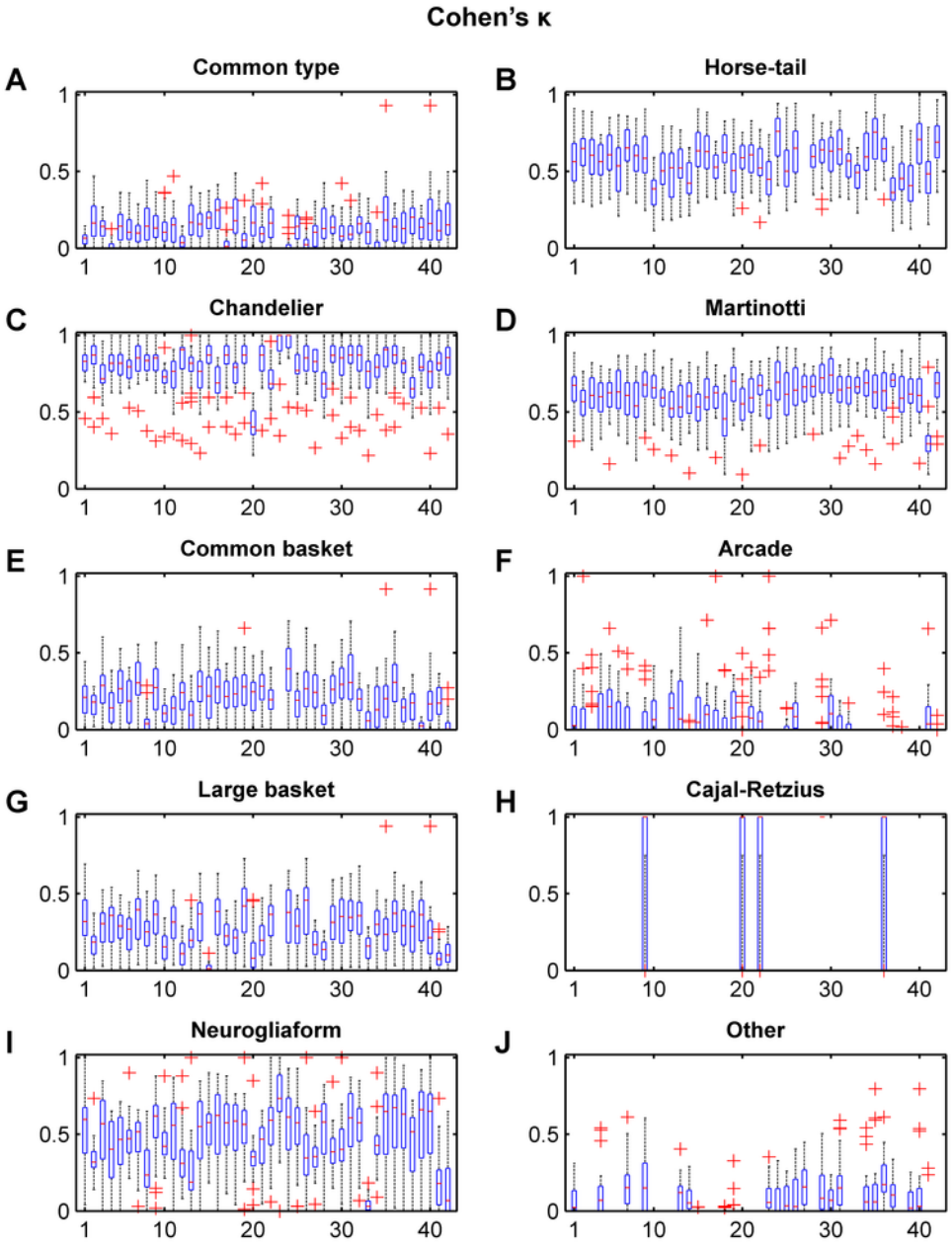

Fig. S9. Boxplots showing Cohen's kappa values for each pair of experts when comparing one category against all other categories in Feature 5. For example, the first box in panel A shows the agreement between the first expert (X-axis) and the rest of the experts for category Common type.

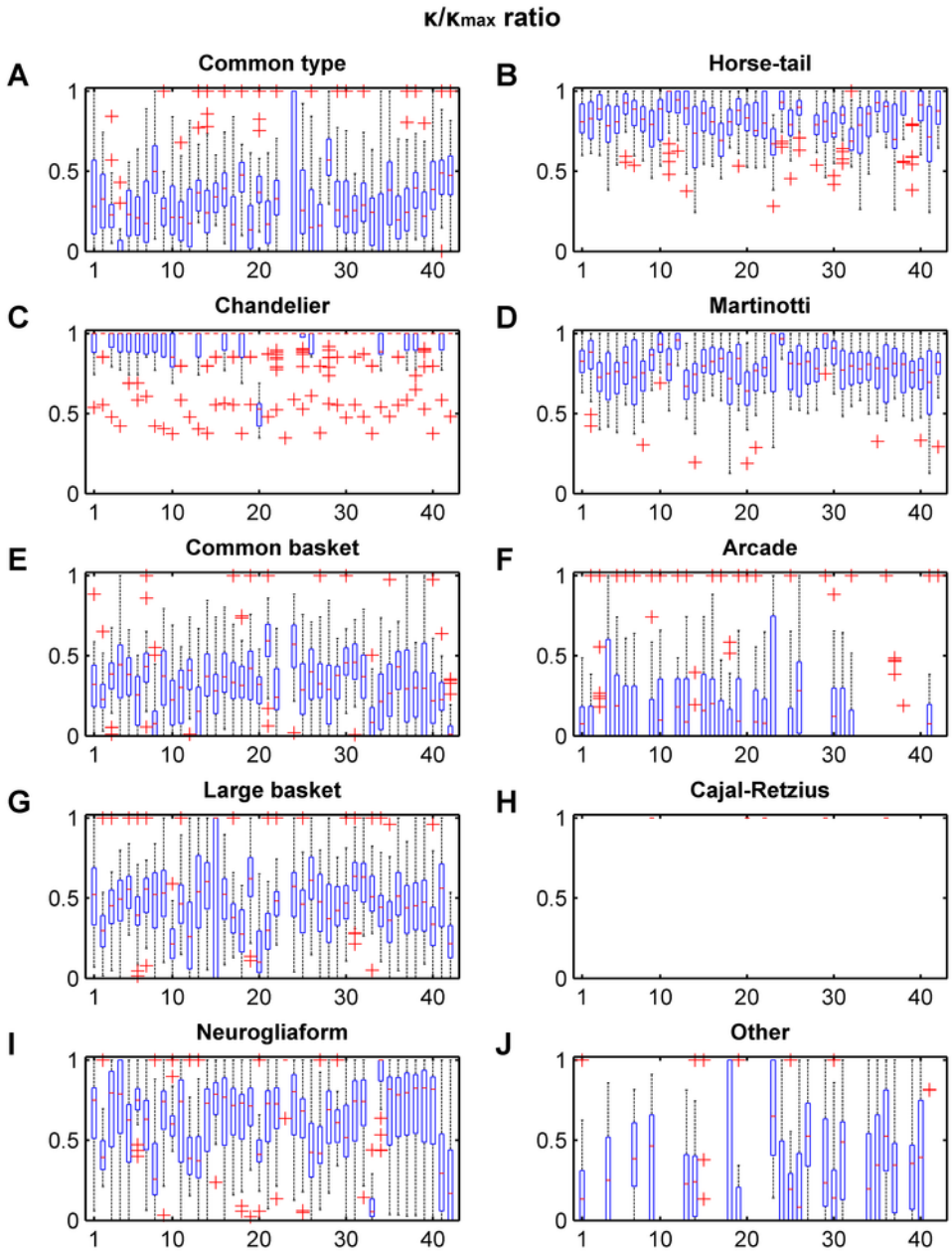

Fig. S10. Boxplots showing the ratios between Cohen's kappa and its maximum value given fixed marginal frequencies for the experts. The ratio values are computed for each pair of experts when comparing one category against all the other categories in Feature 5.

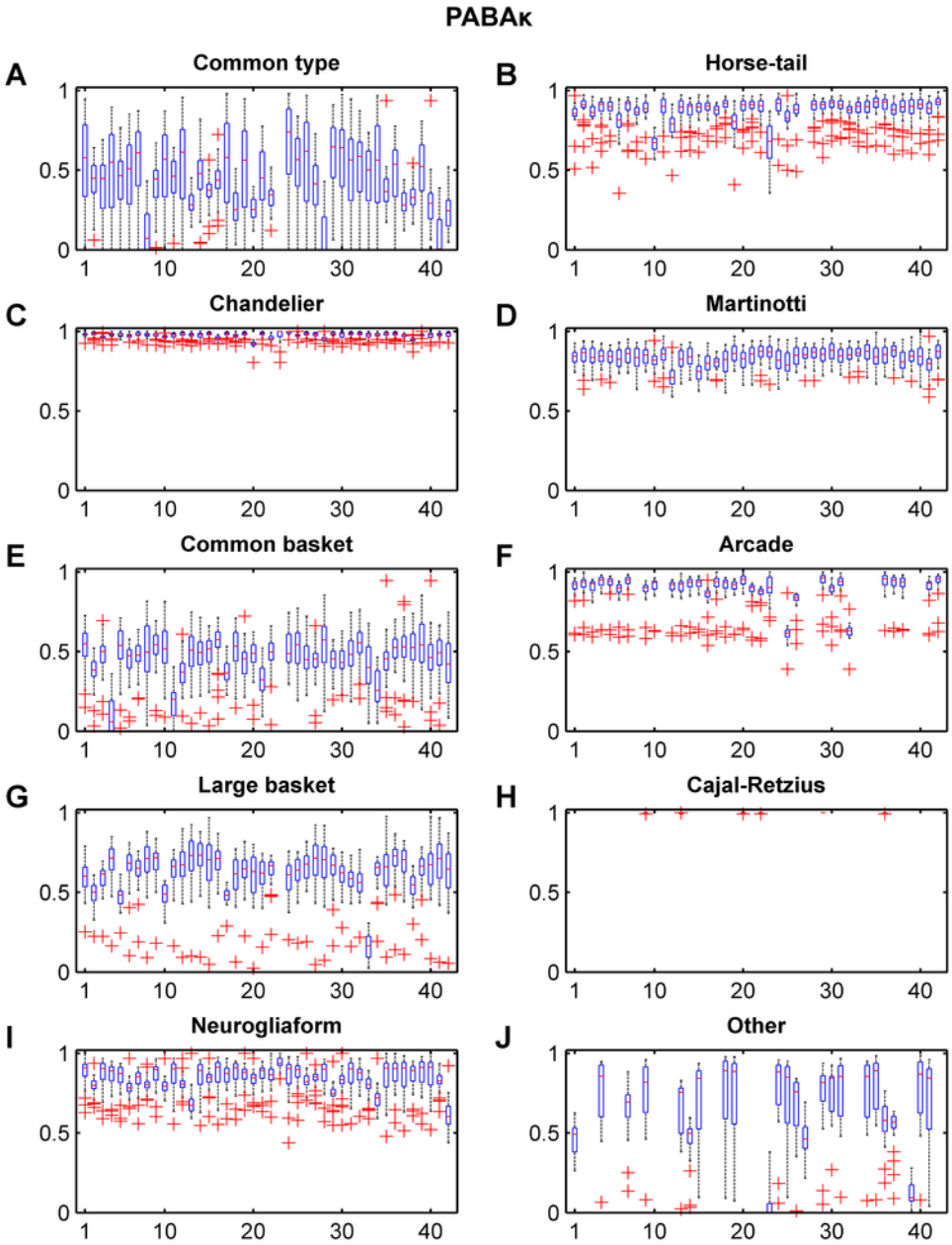

Fig. S11. Boxplots showing the Prevalence-Adjusted Bias-Adjusted kappa ($PABA_k$) values for each pair of experts when comparing one category against all the other categories in Feature 5.

Neuron clustering

We ran clustering algorithms to find groups of neurons at two levels: neuron clustering for each feature and neuron clustering for all features. These algorithms find clusters of neurons with similar properties. Then, we studied whether or not all the neurons in a cluster were assigned the same category within respective features by the experts.

First, we used the *k*-modes algorithm (**Supplementary Online Information S1**) to find clusters of neurons for each feature independently, based on the category selected by each expert for every neuron. For Feature 1 (**Fig. S12**), the *k*-modes algorithm (with k=2) separated neurons into one cluster of neurons mainly categorized by the experts as Translaminar (**Fig. S12A**), and another cluster of neurons mainly categorized as Intralaminar (**Fig. S12B**). The vertical bars in the graphs show the number of experts who selected each category for each neuron in the cluster. However,

note that the *k*-modes algorithm does not use this summarized information. Instead, it clusters the neurons using the category selected by each expert individually (see **Supplementary Online Information S1**). Similarly, single clusters were easily identified for every category of Feature 2 (**Fig. S13**) and 3 (**Fig. S14**). Regarding Feature 4 (**Fig. S15**), the *k*-modes algorithm (*k*=3) found two clusters of neurons mainly categorized by the experts as Ascending (**Fig. S15A**) and Descending (**Fig. S15B**), respectively. However, the third cluster (**Fig. S15C**) contains neurons categorized by the experts as Ascending, Descending or Both, showing confusion about the Both category. With respect to Feature 5, the *k*-modes algorithm (*k*=8) identified individual clusters containing neurons mainly categorized by the experts as Martinotti (**Fig. S16A**), Horse-tail (**Fig. S16B**), Chandelier (**Fig. S16F**) or Neurogliaform (**Fig. S16G**). However, category Neurogliaform was sometimes confused with categories Common type and Common basket (**Fig. S16C, E** and **G**). Other clusters included neurons that the experts categorized as Common type, Common basket, and Large basket (**Fig. S16D, E** and **H**). Thus, the *k*-modes clustering algorithm identified clusters where these three interneuron types were intermingled. The algorithm also showed that the Arcade category appeared distributed in all clusters (**Fig. S16**), although this category was more frequent in clusters in which Common type, Common basket, and Large basket categories were also frequent (**Fig. S16H**). As for Feature 6, the *k*-modes (*k* = 2) identified a cluster with neurons mainly categorized as Characterized (**Fig. S17A**), whereas **Fig. S17B** contains neurons categorized as either Characterized or Uncharacterized by different experts, showing disagreements for these neurons.

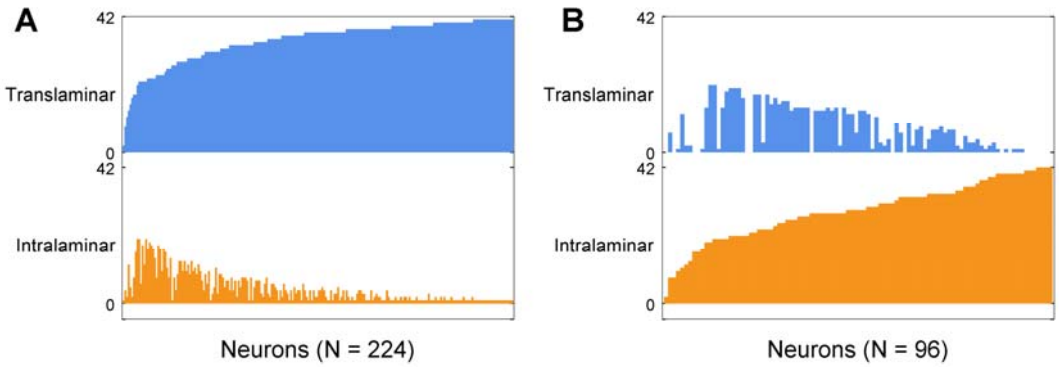

Fig. S12. ~~Graphs showing the c~~Clusters of neurons obtained with the *k*-modes algorithm (*k* = 2) for Feature 1. Vertical bars show the number of experts who selected each category for each neuron in the cluster. Neurons have been sorted in ascending order by the number of votes for clarity. Panels **A** and **B** clearly correspond to Translaminar and Intralaminar categories, respectively. The number of neurons (N) for each cluster is shown at the bottom of each panel.

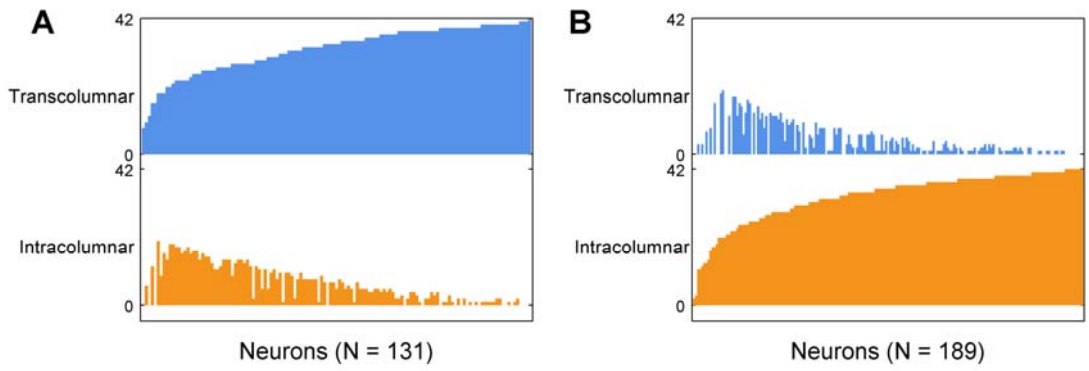

Fig. S13. Clusters of neurons obtained with the *k*-modes algorithm (k = 2) for Feature 2. Vertical bars show the number of experts who selected each category for each neuron in the cluster. Neurons have been sorted in ascending order by the number of votes for clarity. Panel **A** clearly corresponds to neurons mainly categorized as Transcolumnar, whereas panel **B** clearly corresponds to Intracolumnar.

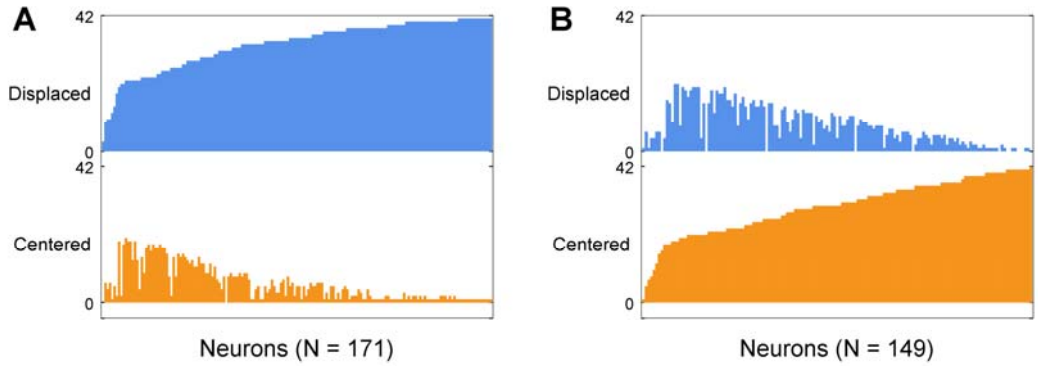

Fig. S14. Clusters of neurons obtained with the *k*-modes algorithm (k = 2) for Feature 3. Vertical bars show the number of experts who selected each category for each neuron in the cluster. Neurons have been sorted in ascending order by the number of votes for clarity. Panels **A** and **B** clearly correspond to Displaced and Centered categories, respectively.

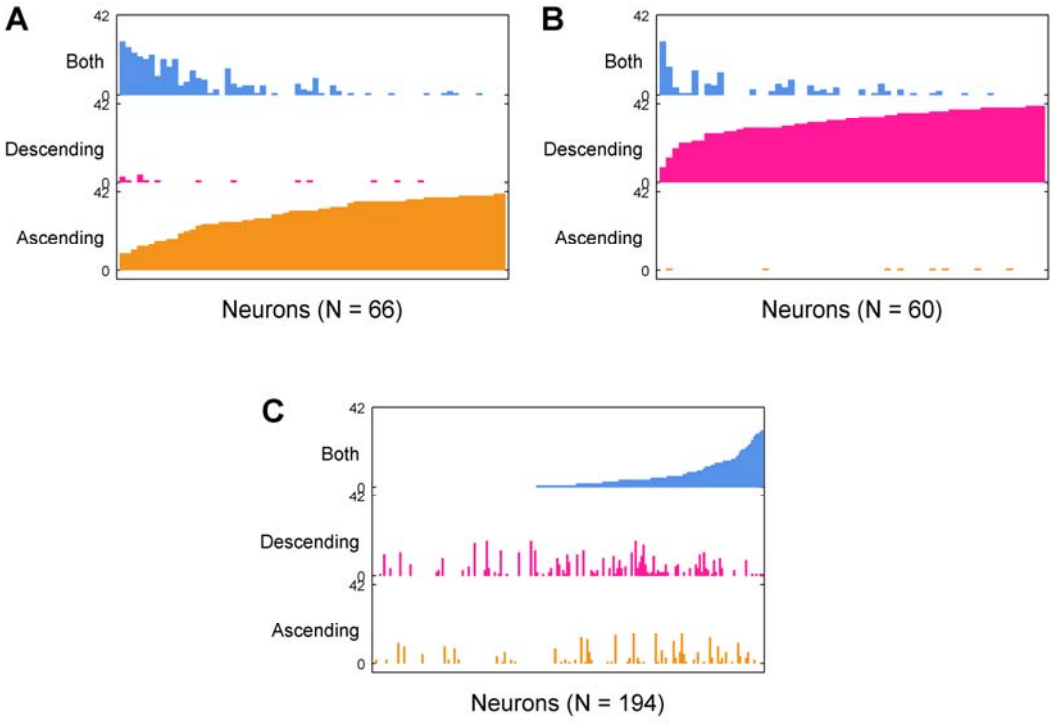

Fig. S15. Clusters of neurons obtained with the *k*-modes algorithm (k = 3) for Feature 4. Vertical bars show the number of experts who selected each category for each neuron in the cluster. Neurons have been sorted in ascending order by the number of votes for clarity. Panels **A** and **B** correspond to neurons mainly categorized as Ascending and Descending, respectively. Panel **C** shows neurons where different experts disagreed, categorizing them as Ascending, Descending or Both.

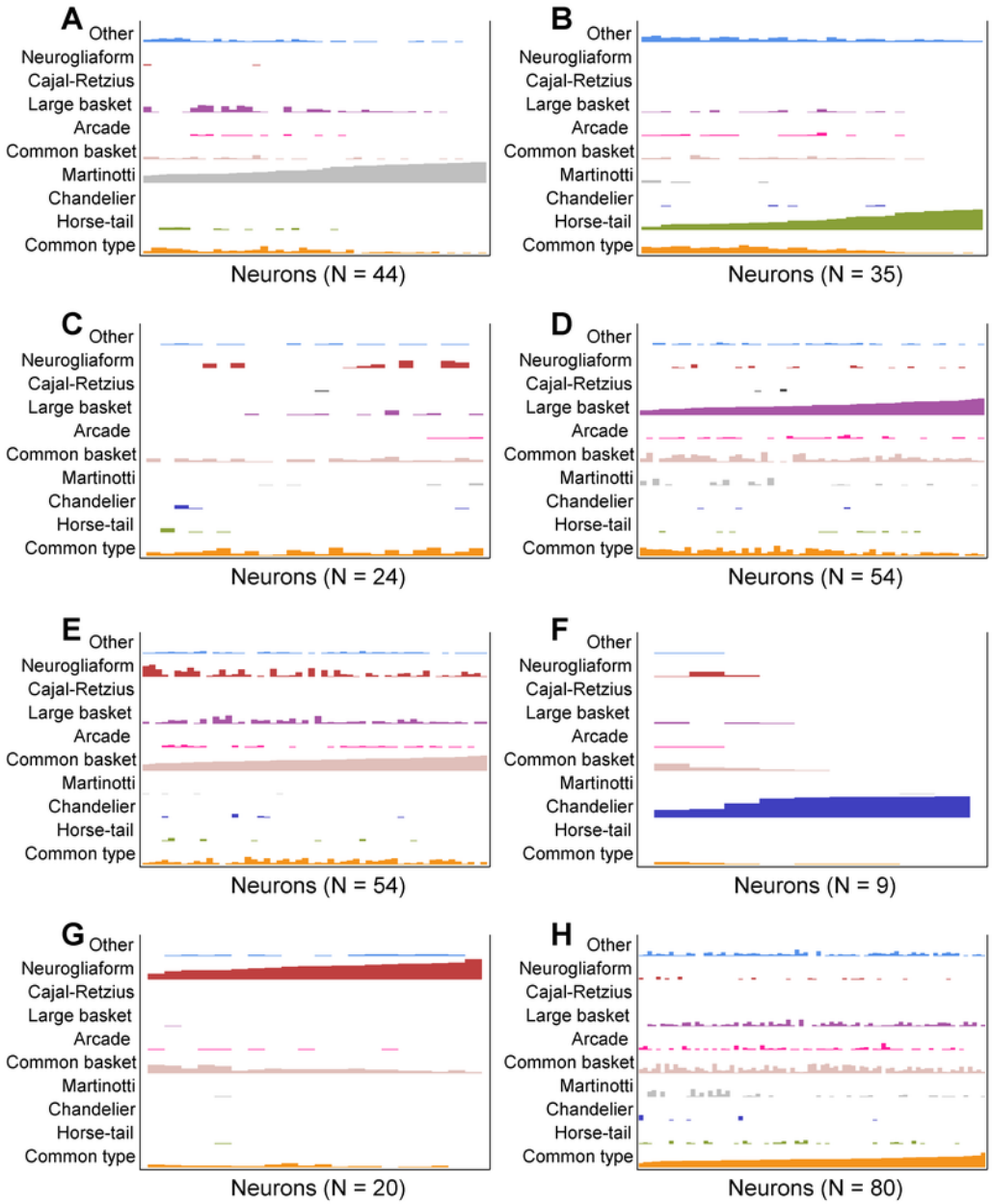

Fig. S16. Clusters of neurons obtained with the *k*-modes algorithm (k = 8) for Feature 5. Vertical bars show the number of experts who selected each category for each neuron in the cluster. Neurons have been sorted in ascending order by the number of votes for clarity. Panels **A** and **F** show clusters of neurons clearly corresponding to Martinotti and Chandelier cells, respectively. Other panels (e.g., **E**) show clusters of neurons that did not correspond to a single category.

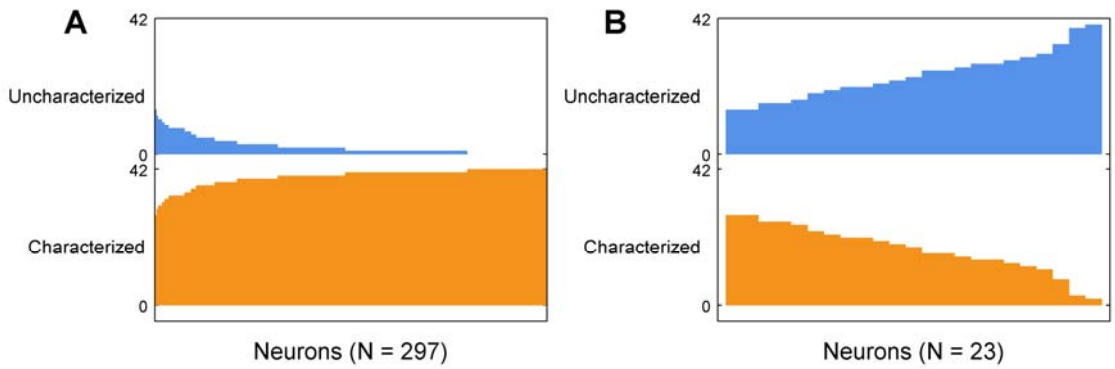

Fig. S17. Clusters of neurons obtained with the *k*-modes algorithm (k = 2) for Feature 6. Vertical bars show the number of experts who selected each category for each neuron in the cluster. Neurons have been sorted in ascending order by the number of votes for clarity. Panel **A** contains neurons mainly categorized as Characterized, whereas panel **B** contains neurons where different experts disagreed, categorizing them as either Characterized or Uncharacterized.

Bayesian networks for modeling experts' opinions

We trained a Bayesian network with the data provided by each expert. The 42 Bayesian network structures (one for each expert) were analyzed and probabilistic inferences were performed to reveal the underlying behaviors of the experts (**Supplementary Online Information S1**), i.e. how the experts made their choices about a neuron.

As an example, **Figs. S18** and **S19** show four Bayesian networks corresponding to four different experts (all figures for the remaining networks are available upon request). The Bayesian networks for experts 16 (**Fig. S18A**) and 17 (**Fig. S18B**) had the same structure and similar propagated probabilities when these experts assigned a neuron as a Martinotti cell in Feature 5. The greatest difference between the two networks occurred in Feature 2, where the probabilities of Intracolumnar and Transcolumnar were respectively 0.34 and 0.64 for expert 16, and 0.56 and 0.42 for expert 17. The Bayesian networks for experts 27 (**Fig. S18C**) and 32 (**Fig. S18D**) had a different structure from each other and from experts 16 and 17. However, the probabilistic reasoning on Feature 1 and on Feature 3 when the four experts considered a neuron as Martinotti cell was similar, e.g., these four experts agreed (assigning probabilities higher than 0.87) that Martinotti cells were Translaminar and Displaced. Differences between the experts could also be identified in the Bayesian networks, e.g., Feature 5 in **Fig. S18C** did not include as possible categories Arcade or Horse-tail cells, but included category Other. That means that expert 27 did not categorize any neuron as Arcade or Horse-tail.

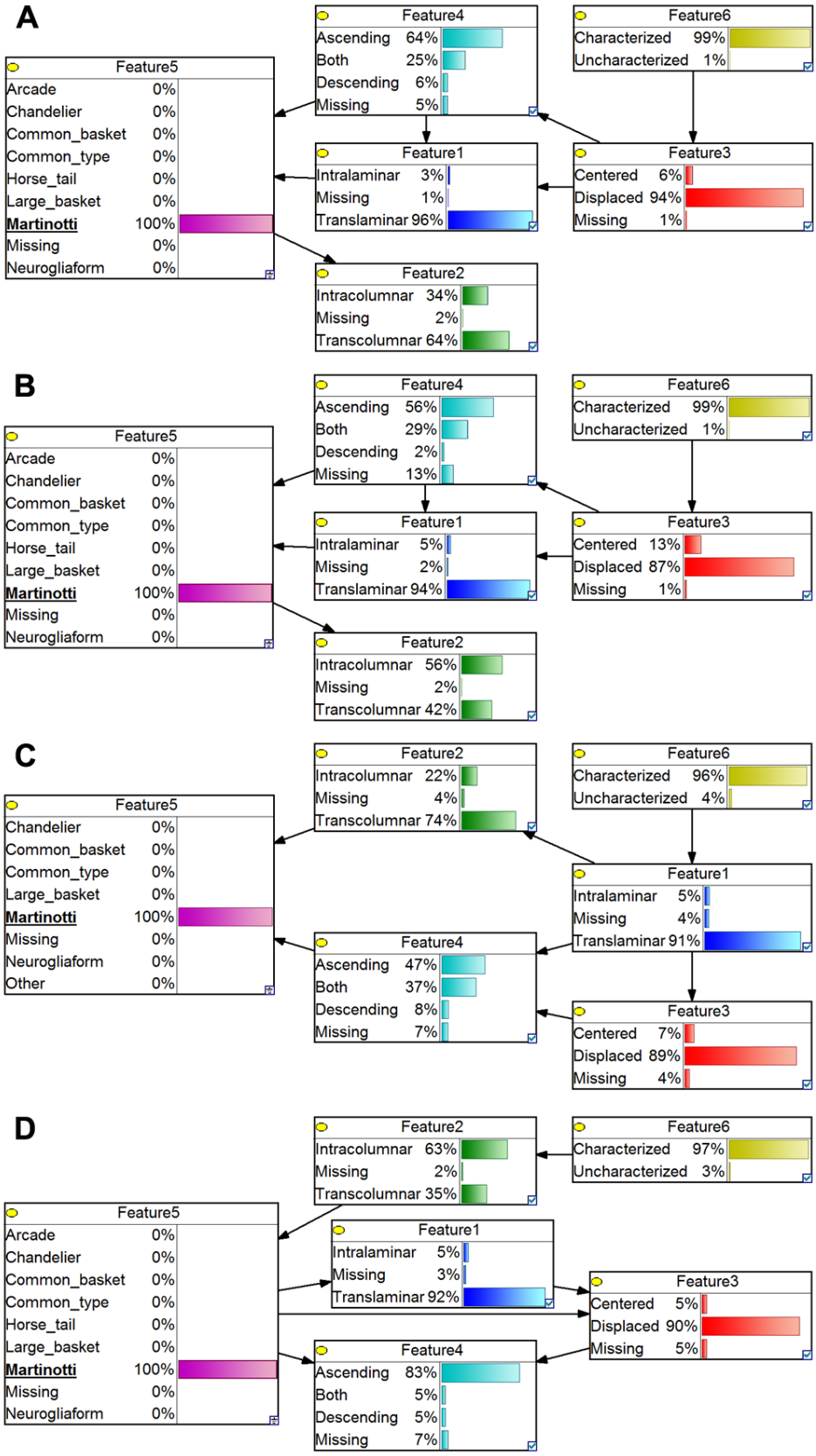

Fig. S18. Bayesian networks for experts 16 (**A**), 17 (**B**), 27 (**C**) and 32 (**D**). Martinotti was selected in Feature 5 and the probabilities were propagated through the Bayesian networks. Bar charts show the propagated probabilities of the remaining features conditioned on the Martinotti category.

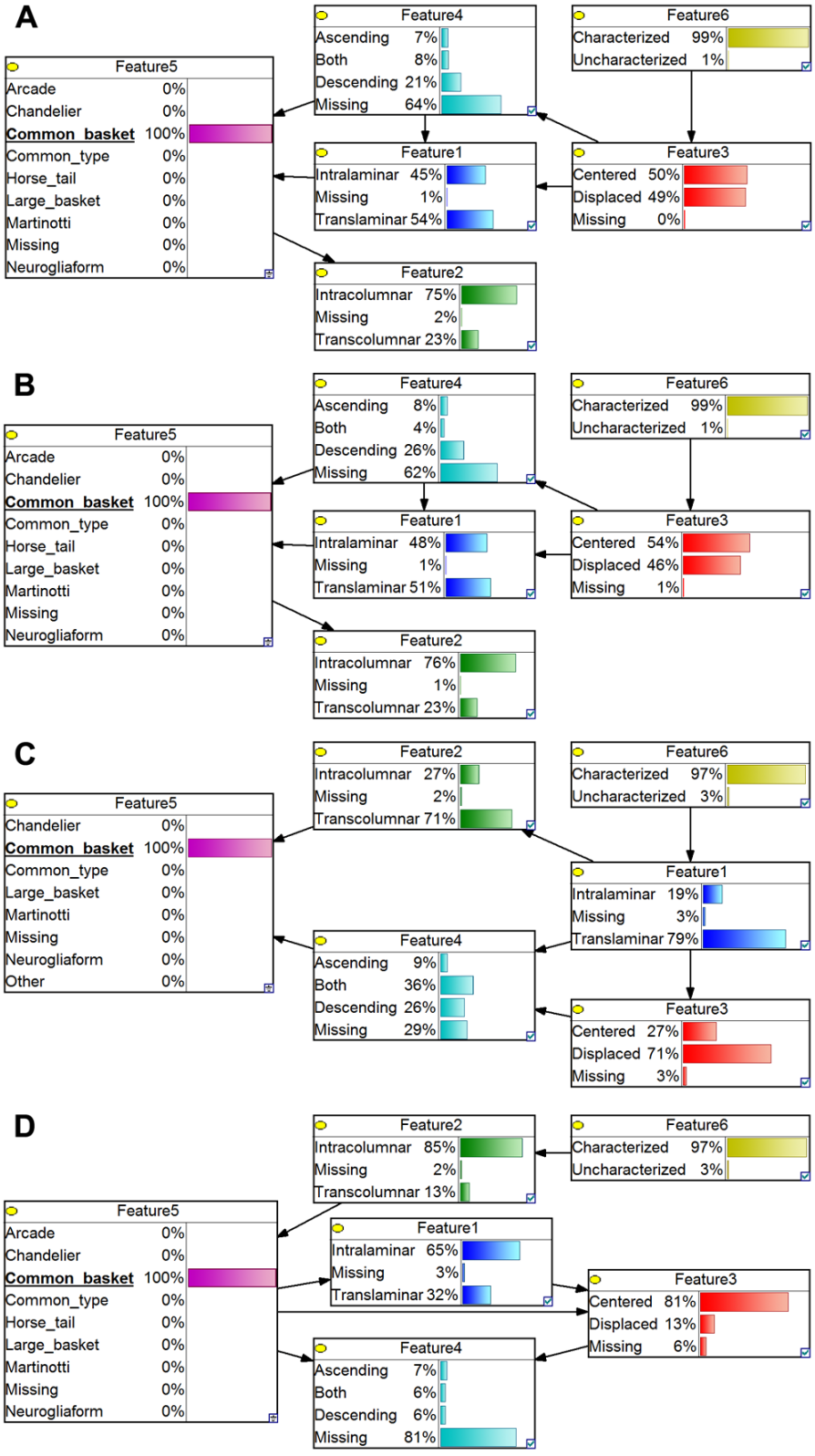

Fig. S19. Bayesian networks for experts 16 (**A**), 17 (**B**), 27 (**C**) and 32 (**D**). Common basket was selected in Feature 5 and the probabilities were propagated through the Bayesian networks. Bar charts show the propagated probabilities of the remaining features conditioned on the Common basket category.

We also used Bayesian networks to analyze the disagreements between experts about the classification of interneuron types. **Fig. S19** shows the Bayesian networks for the

same four experts when Common basket was selected as evidence in Feature 5 and the probabilities were propagated. The posterior probabilities for expert 16 (**Fig. S19A**) and expert 17 (**Fig. S19B**) were similar but they were different for expert 27 (**Fig. S19C**) and expert 32 (**Fig. S19D**). For example, regarding Feature 1, the probability of Translaminar was 0.79 in **Fig. S19C** and 0.32 in **Fig. S19D**. With respect to Feature 2, the probability of a Common basket being Transcolumnar was 0.71 in **Supplementary Fig. S19C**, whereas in the other three Bayesian networks the probability was below 0.23. For Feature 3, Centered was the most probable value in **Fig. S19C** and Displaced had the highest probability in **Fig. S19D**. Also, **Fig. S19C** shows a higher probability for the category Both in Feature 4 than the other Bayesian networks.

The analysis of the 42 Bayesian network structures is summarized in **Table S2**, including frequent relationships (high numbers) and rare relationships between features (**Supplementary Online Information S1**). Feature 1, Feature 3, and Feature 4 appeared frequently related; this could be explained by the fact that the categories Ascending, Descending, and Both are associated to the categories Translaminar and Displaced, describing the vertical orientation of the neuron. Feature 5 was frequently linked to Feature 1, Feature 2, and Feature 4 in more than half of the Bayesian network structures. Therefore, these three features (laminar, columnar, and ascending/descending) are identified in this analysis as relevant when describing morphological properties of interneuron types (Feature 5).

Table S2. Number of Bayesian networks out of 42 that include the possible (undirected) edge between the nodes in the corresponding row and column. Presence of an edge in the Bayesian network indicates that the choices of categories in those features by that expert are related. Frequency of relationships is highlighted with a gradient of color shades from red (most frequent) to white (non-existent or rare).

	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5	Feature 6
Feature 1		4	32	30	30	17
Feature 2			0	0	38	17
Feature 3				41	14	14
Feature 4					29	2
Feature 5						10
Feature 6						

Supervised classification of neurons: automatic classification

We aimed to build a model that could automatically classify the neurons in each of the six features on the basis of a set of 2,886 morphological measurements of the digital 3D reconstructions. From the total of 320 neurons, we used the 241 neurons for which digital 3D morphological reconstructions were available. We computed a set of 2,886 morphological variables for each neuron from the data provided by Neurolucida Explorer and included information about dendrites, axons, and soma, including length of dendritic and axonal arbor segments, convex hull, Sholl, fractal, vertex and branch-angle analyses (**Supplementary Online Information S1**). Six classification models or classifiers were built, each for predicting the value of a feature. A unique value (a class value in the supervised classification terminology) for each of the features was assigned to each neuron, and these values were based on the experts' 'majority votes' for that

neuron. We trained the classifiers using ten different supervised classification algorithms and two variable selection methods (**Supplementary Online Information S1**) for each feature. In particular, we used each of the ten different algorithms for building classifiers, with (1) all variables, (2) a subset of variables selected with the Gain Ratio method, and (3) another subset of variables selected with the CfsSubset method (**Supplementary Online Information S1**). This gives a total of 30 classifiers per feature. **Table S3** shows the accuracy of the classifiers, that is, the percentage of correct classifications by comparing, for each neuron, the outcome of the classifier with the 'majority vote' of the experts. The accuracy of the classifiers was estimated using the leave-one-out technique (Mosteller and Tukey, 1968). Thus, we trained the classifiers with all data except a single neuron, and used that neuron later for testing. This is repeated such that each neuron is left out once. The classifiers were able to distinguish whether or not a neuron was Characterized, as the best result in accuracy is 99.17% (2 neurons misclassified). The best performing classifiers for Feature 1 and Feature 2 yielded accuracy over 80%, whereas the best result for Feature 3 was 73.86%. The accuracy of the classifiers was below 70% for both Feature 4 and Feature 5. One explanation for the low accuracy for Feature 5 is that the class labels were not very reliable because the experts frequently disagreed when classifying the neurons in this feature. However, it is also possible that the interneuron classes could not be distinguished using the set of morphological measurements included in the study. Moreover, according to majority votes, the number of neurons assigned by the experts to the different interneuron types were unbalanced, with only three Chandelier cells and four Neurogliaform cells, but as many as 77 Common type cells and 68 Common basket cells. This makes it especially difficult for the classifiers to distinguish the least frequent neuronal types. Surprisingly, the classifiers achieved the lowest accuracy for Feature 4. This may be explained by the same two factors indicated above: the Both category was confusing to the experts, so the neurons might have been assigned to the wrong category. Also, there may be no morphological variables that capture the orientation of the axon. To test the significance of these results, we computed the category with maximum prior probability for the classifier induced for each feature independently:

- Feature 1: 0.7718 (achieved at Translaminar)
- Feature 2: 0.5187 (Transcolumnar)
- Feature 3: 0.5809 (Displaced)
- Feature 4: 0.3817 (Ascending)
- Feature 5: 0.3195 (Common type)
- Feature 6: 0.9544 (Characterized)

For every feature, the best classifier in **Table S3** outperformed a base classifier which always selects the class with maximum prior probability according to an exact binomial test (see asterisks in **Table S3** and **Supplementary Online Information S1**).

Table S3. Percent accuracy of the classifiers trained for each feature independently using ten different classification algorithms (in columns, see **Supplementary Online Information S1**) and three variable selection methods (in rows): NoFSS (no feature subset selection, i.e., all variables selected), Gain Ratio, and CfsSubset (**Supplementary Online Information S1**). The highest accuracy for each feature and variable selection method are highlighted in bold. Additionally, the overall highest accuracy for each feature is shaded in grey. A binomial test was used to check whether or not the classifiers outperformed a base classifier always selecting the category with maximum prior probability. Asterisks indicate a p-value < 0.05.

	NB	NBdisc	RBFN	SMO	IB1	IB3	JRip	J48	RForest	RTree
Feature 1: Intralaminar vs. Translaminar										
NoFSS	57.68	58.51	77.59	82.16*	72.2	73.44	82.57*	**85.48***	82.16*	75.93
Gain Ratio	64.73	54.36	79.67	82.99*	69.71	75.93	83.82*	**85.48***	84.23*	79.67
CfsSubset	75.93	75.1	81.33	84.23*	73.86	80.08	**84.65***	80.08	82.16*	80.08
Feature 2: Intracolumnar vs. Transcolumnar										
NoFSS	59.75*	62.66*	52.28	**75.52***	57.68*	65.56*	74.27*	68.46*	66.39*	58.09*
Gain Ratio	66.39*	63.07*	53.11	**76.35***	64.32*	65.98*	75.52*	68.88*	70.12*	65.98*
CfsSubset	72.61*	65.56*	76.76*	**81.33***	73.86*	73.03*	74.69*	70.54*	76.35*	69.29*
Feature 3: Centered vs. Displaced										
NoFSS	62.24	53.94	54.77	**68.88***	64.73*	68.05*	66.8*	67.63*	68.46*	62.24
Gain Ratio	64.73*	**73.03***	65.98*	70.54*	65.56*	71.37*	70.54*	66.39*	72.2*	68.46*
CfsSubset	68.88*	**73.86***	70.54*	73.03*	65.15*	68.05*	63.9*	71.78*	68.46*	65.15*
Feature 4: Ascending vs. Descending vs. Both										
NoFSS	34.44	27.8	44.4*	49.38*	41.91	38.59	33.61	**54.36***	40.25	37.76
Gain Ratio	43.57*	33.2	43.98*	**49.79***	41.91	42.32	43.57*	46.89*	45.64*	42.74
CfsSubset	47.3*	51.87*	47.3*	58.51*	47.3*	52.28*	48.13*	42.32	**60.17***	47.3*
Feature 5: Interneuron type (10 classes)										
NoFSS	56.02*	19.09	45.23*	**58.51***	50.62*	53.94*	50.62*	47.72*	52.28*	40.25*
Gain Ratio	60.17*	26.14	58.92*	**62.24***	49.79*	51.87*	48.55*	43.15*	58.09*	43.98*
CfsSubset	61*	43.57*	**61.41***	60.58*	58.09*	56.85*	53.94*	49.38*	56.85*	51.45*
Feature 6: Characterized vs. Uncharacterized										
NoFSS	77.18	88.38	95.85	**97.93***	97.51	97.51	**97.93***	97.51	96.27	95.85
Gain Ratio	**98.34***	73.86	97.51	96.68	97.1	97.51	97.93*	97.93*	97.51	**98.34***
CfsSubset	97.51	89.63	96.27	97.1	95.44	95.02	97.93*	96.27	97.51	**99.17***

To further analyze the results for Feature 5, **Table S4** shows the confusion matrix of the best performing algorithm (SMO), which achieved an accuracy of 62.24% (**Table S3**). The confusion matrix shows the performance of an algorithm by displaying the number of neurons of each true category (rows) matched to the categories predicted by the classifier (columns). Numbers in the main diagonal of the matrix (shaded) indicate the number of correctly classified neurons, i.e., those neurons whose true class was equal to the class predicted by the classifier. High values in the main diagonal of the matrix reflect very accurate classifiers. Contrarily, non-zero values outside the main diagonal of the confusion matrix show classification errors, i.e., cases in which the predicted class for the neurons did not match the true class. Some Martinotti cells were wrongly classified as Common type (9 cases), Large basket (4) and Chandelier (1). This was

similar to the results shown by the clustering algorithms. Horse-tail cells were wrongly classified as Common type cells. Also, the classifier often confused Common type, Common basket and Large basket neuron types (**Table S4**). The four Neurogliaform cells and two out of the three Chandelier cells were wrongly classified as Common basket. There were no neurons from classes Arcade, Cajal-Retzius, and Other, because none of the neurons were assigned to any of those categories by the majority of the experts.

Table S4. Confusion matrix for the SMO classifier and Gain Ratio for variable selection using Feature 5 data. The main diagonal of the matrix (shaded) indicate the number of correctly classified neurons, whereas non-zero values outside the main diagonal show the number of wrongly classified neurons.

		Predicted class									
		Common type	Horse-tail	Chandelier	Martinotti	Common basket	Arcade	Large basket	Cajal-Retzius	Neuroglia form	Other
True class	Common type	55	1	0	5	11	0	5	0	0	0
	Horse-tail	7	5	0	2	0	0	0	0	0	0
	Chandelier	0	0	1	0	2	0	0	0	0	0
	Martinotti	9	0	1	24	0	0	4	0	0	0
	Common basket	15	1	0	0	49	0	3	0	0	0
	Arcade	0	0	0	0	0	0	0	0	0	0
	Large basket	11	0	0	3	7	0	16	0	0	0
	Cajal-Retzius	0	0	0	0	0	0	0	0	0	0
	Neuroglia form	0	0	0	0	4	0	0	0	0	0
	Other	0	0	0	0	0	0	0	0	0	0

Then, we trained one binary classifier for categories in Feature 5 with more than 5 neurons, i.e. Common type, Horse-tail, Martinotti, Common basket, and Large basket. The goal was to check whether a particular category could be distinguished from all the other interneuron types (categories) considered together (**Supplementary Online Information S1**). **Table S5** shows the accuracies of the binary classifiers for each category. The classifiers for Horse-tail and Martinotti cells achieved high accuracies, whereas the classifiers for Common type, Common basket, and Large basket cells yielded lower accuracies. The maximum prior probabilities for these binary classifiers were:

- Common type vs. the rest: 0.6805 (achieved at the rest)
- Horse-tail vs. the rest: 0.9419 (the rest)
- Martinotti vs. the rest: 0.8423 (the rest)
- Common basket vs. the rest: 0.7178 (the rest)
- Large basket vs. the rest: 0.8465 (the rest)

The induced classifiers were not able to outperform the base classifier for Horse-tail and Large basket categories. There were few neurons categorized as Horse-tail by the majority of the experts, so it was difficult to induce classifiers able to distinguish this category, even though it was easily distinguishable for the experts. This limitation

should vanish when more data become available. Contrarily, neurons categorized as Large basket were difficult to distinguish for both experts and supervised classifiers.

Table S5. Percent accuracy of the binary classifiers (in columns, see **Supplementary Online Information S1**) induced for the categories in Feature 5 and two variable selection methods (in rows). Each classifier tried to identify whether a neuron belonged to a particular category vs. all other categories, and this was repeated for each category separately. The best results for each category and variable selection method are highlighted with bold face. The highest accuracy for a given category is shaded in grey. A binomial test was used to check whether or not the classifiers outperformed a base classifier always selecting the category with maximum prior probability. Asterisks indicate a p-value < 0.05.

	NB	NBdisc	RBFN	SMO	IB1	IB3	JRip	J48	RForest	RTree
Common type vs. the rest										
NoFSS	61.83	54.36	71.37	70.95	69.29	75.52*	77.59*	76.76*	**78.84***	68.88
Gain Ratio	67.22	63.49	75.10*	69.71	71.37	74.27*	75.10*	**77.18***	75.10*	68.88
CfsSubset	74.69*	64.32	75.10*	74.27*	76.76*	**78.84***	71.78	69.29	70.95	68.88
Horse-tail vs. the rest										
NoFSS	91.70	51.87	93.36	94.19	90.87	94.19	92.53	90.87	**94.61**	92.53
Gain Ratio	86.31	88.38	90.46	94.61	92.95	94.19	92.12	90.87	**95.02**	93.78
CfsSubset	92.53	72.61	93.36	**95.02**	94.61	93.36	92.12	93.78	93.78	94.19
Martinotti vs. the rest										
NoFSS	84.23	65.56	82.99	**88.80***	85.48	86.72	82.57	84.23	85.48	83.82
Gain Ratio	84.65	67.63	81.33	**88.38***	84.65	87.14	84.23	80.91	85.89	83.40
CfsSubset	85.89	77.18	86.31	87.97	87.97	**90.46***	84.65	84.23	87.55	85.48
Common basket vs. the rest										
NoFSS	68.46	54.77	71.78	**79.25***	77.18*	78.01*	78.84*	77.18*	78.42*	76.76*
Gain Ratio	72.61	51.87	75.52	79.25*	76.76*	77.59*	76.35	74.27	**83.40***	78.42*
CfsSubset	78.01*	78.84*	80.91*	**81.33***	77.59*	77.18*	79.25*	74.69	80.91*	79.25*
Large basket vs. the rest										
NoFSS	54.77	67.63	84.65	80.50	83.40	**85.06**	83.40	79.67	82.57	74.69
Gain Ratio	70.95	66.39	84.23	80.50	79.25	80.08	**84.65**	81.74	82.99	79.67
CfsSubset	81.74	59.34	81.33	82.99	80.91	82.57	84.23	82.57	**84.65**	80.08

Finally, when merging the three categories (Common type, Common basket, and Large basket) into one single category, the accuracy of the best classifier increased to 83.40%. When we merged only Common type and Common basket cells, the best classifier accuracy was 73.86%. When merging only Common type and Large basket, the best classifier accuracy was 69.29%. Lastly, merging only Common basket and Large basket cells resulted in the best accuracy among classifiers of 70.12%. These results suggest that Common type, Common basket, and Large basket are not well-defined categories. For all these experiments, the induced classifiers outperformed the base classifiers using the maximum prior probabilities:

- Common type + Common basket + Large basket vs. each neuron type: 0.7552 (achived at Common type + Common basket + Large basket)
- Common type + Common basket vs. each neuron type: 0.6017 (Common type + Common basket)

- Common type + Large basket vs. each neuron type: 0.4730 (Common type + Large basket)

Common basket + Large basket vs. each neuron type: 0.4357 (Common basket + Large basket)

Table S6. Percent accuracy of the classifiers (in columns, see **Supplementary Online Information S1**) for Feature 5 in an analysis in which Common type, Common basket, and Large basket or pairs among them were merged into one category. Two variable selection methods are used (in rows). The best results for each combination of categories and variable selection method are highlighted with bold face. The highest accuracy for a given combinations of categories is shaded in grey. A binomial test was used to check whether or not the classifiers outperformed a base classifier always selecting the category with maximum prior probability. Asterisks indicate a *p*-value < 0.05.

	NB	NBdisc	RBFN	SMO	IB1	IBk	JRip	J48	RForest	RTree
Common type + Common basket + Large basket vs. each neuron type										
NoFSS	79.25	20.75	78.42	**82.57***	74.69	79.67	77.18	69.29	79.25	70.12
Gain Ratio	77.18	32.37	73.86	80.91*	77.59	**82.16***	73.44	73.44	78.84	73.03
CfsSubset	80.91*	50.21	80.91*	80.91*	80.50*	**83.40***	74.27	75.10	**83.40***	74.69
Common type + Common basket vs. each neuron type										
NoFSS	**67.22***	21.99	58.51	66.80*	60.58	66.80*	58.51	53.53	62.66	59.75
Gain Ratio	64.32	27.39	63.49	**73.03***	63.07	65.56*	61.83	51.45	69.71*	61.83
CfsSubset	68.88*	39.83	68.88*	73.03*	70.12*	**73.86***	64.73	63.90	69.71*	63.07
Common type + Large basket vs. each neuron type										
NoFSS	60.17*	19.92	49.38	**64.73***	57.26*	59.75*	56.02*	51.45	59.34*	51.45
Gain Ratio	64.73*	27.80	59.75*	**68.46***	59.34*	63.07*	55.60	50.62	64.32*	54.36*
CfsSubset	65.56*	39.83	69.71*	64.73*	65.98*	**69.29***	57.68*	59.75*	64.73*	54.36*
Common basket + Large basket vs. each neuron type										
NoFSS	60.17*	16.60	54.36*	**65.56***	58.09*	61.00*	55.60*	52.70*	62.66*	54.36*
Gain Ratio	61.41*	51.45*	65.98*	64.32*	61.00*	65.56*	55.60*	49.79*	**67.63***	53.53*
CfsSubset	66.39*	41.08	68.46*	**70.12***	64.73*	66.80*	52.70*	58.09*	68.46*	55.19*