ANALYSIS

New insights into the classification and nomenclature of cortical GABAergic interneurons

Javier DeFelipe[1,2], Pedro L. López-Cruz[3], Ruth Benavides-Piccione[1,2], Concha Bielza[3], Pedro Larrañaga[3], Stewart Anderson[5], Andreas Burkhalter[6], Bruno Cauli[7], Alfonso Fairén[8], Dirk Feldmeyer[9,10], Gord Fishell[11], David Fitzpatrick[12], Tamás F. Freund[13], Guillermo González-Burgos[14], Shaul Hestrin[15], Sean Hill[16], Patrick R. Hof[17], Josh Huang[18], Edward G. Jones[†], Yasuo Kawaguchi[19], Zoltán Kisvárday[20], Yoshiyuki Kubota[19], David A. Lewis[14], Oscar Marín[8], Henry Markram[16], Chris J. McBain[21], Hanno S. Meyer[22], Hannah Monyer[23], Sacha B. Nelson[24], Kathleen Rockland[25], Jean Rossier[26], John L. R. Rubenstein[27], Bernardo Rudy[28], Massimo Scanziani[29], Gordon M. Shepherd[30], Chet C. Sherwood[31], Jochen F. Staiger[32], Gábor Tamás[33], Alex Thomson[34], Yun Wang[35,36], Rafael Yuste[37] and Giorgio A. Ascoli[4]

Abstract | A systematic classification and accepted nomenclature of neuron types is much needed but is currently lacking. This article describes a possible taxonomical solution for classifying GABAergic interneurons of the cerebral cortex based on a novel, web-based interactive system that allows experts to classify neurons with pre-determined criteria. Using Bayesian analysis and clustering algorithms on the resulting data, we investigated the suitability of several anatomical terms and neuron names for cortical GABAergic interneurons. Moreover, we show that supervised classification models could automatically categorize interneurons in agreement with experts' assignments. These results demonstrate a practical and objective approach to the naming, characterization and classification of neurons based on community consensus.

[†]*In memory of Ted Jones*

Correspondence to J.DF., P.L. and G.A.A.
e-mails:
defelipe@cajal.csic.es;
pedro.larranaga@fi.upm.es;
ascoli@gmu.edu

The problem of classifying and naming neurons has been a topic of debate for over 100 years. Nevertheless, a satisfactory consensus remains to be reached, even for restricted neuronal populations such as the GABAergic interneurons of the cerebral cortex. Over the past two decades, the amount of morphological, molecular, physiological and developmental data has grown rapidly, making classification harder rather than easier. A consistent neuronal classification and terminology will help researchers to manage this multidisciplinary knowledge, and is needed for specialists in neuroscience subfields to establish and maintain effective communication and data sharing[1]. As in other domains of science, taxonomies can be empirical or scientific. This distinction was well described by John Hughlings Jackson[2] in 1874: "There are two ways of investigating diseases, and two kinds of classification corresponding thereto, the empirical and the scientific. The former is to be illustrated by the way in which a gardener classifies plants, the latter by the way in which a botanist classifies them. The former is, strictly speaking, only an arrangement. The gardener arranges his plants as they are fit for food, for ornament, etc. One of his classifications of ornamental plants is into trees, shrubs, and flowers. His object is the direct application of knowledge to utilitarian purposes. It is, so to speak, practical. The other kind of classification (the classification properly so-called) is rather for the better organization of existing knowledge, and for discovering the relations of new facts; its principles are methodical guides to further investigation. It is of great utilitarian value, but not directly."

In spite of the many studies performed since the original findings of Santiago Ramón y Cajal, it is surprising that we still lack a catalogue of neuron types and names that is accepted by the general scientific community. Recognizing this problem, the International Neuroinformatics Coordinating Facility (INCF) has recently established a Neuron Registry within the Program on Ontologies of Neural Structures (PONS), with the aim to identify known neuron types on the basis of their defining properties[3] (see the INCF Program on Ontologies of Neural Structures website). A collation of terms referring to neuron types is available as part of the Neuroscience Information Framework (NIF) from NeuroLex[4] (see the NeuroLex website).

A milestone towards a future classification of GABAergic interneurons in the cerebral cortex was the standardization of the nomenclature of their properties[1]. However, at that time it was not possible to identify a set of anatomical traits that unambiguously define an interneuron class. In this Analysis article, we describe a new, community-based strategy for defining a morphological taxonomy. Our goal was to establish a list of terms that could be used by all researchers in the field to distinguish neuronal morphologies. Because the developmental and evolutionary processes that gave rise to these morphologies are incompletely understood, we sought a practical rather than a scientific classification: a 'gardener's approach'. To this end, we selected a limited number of neuron types and morphological properties based on studies performed over the years in many laboratories. These neuron types and morphological properties are not meant to be imposed but rather are proposed, with the goals of incorporating community feedback and reaching consensus.

In this article, we first provide an overview of historical and current issues involved in classifying cortical neurons and, in particular, interneurons. We then

Author addresses

[1]Laboratorio Cajal de Circuitos Corticales, Centro de Tecnología Biomédica (CTB), Universidad Politécnica de Madrid, Campus Montegancedo S/N, Pozuelo de Alarcón, 28223 Madrid, Spain.

[2]Instituto Cajal (CSIC), Avenida Doctor Arce 37, 28002 Madrid, Spain.

[3]Departamento de Inteligencia Artificial, Universidad Politécnica de Madrid, Campus de Montegancedo S/N, 28660, Boadilla del Monte, Madrid, Spain.

[4]Center for Neural Informatics, Structure, and Plasticity Molecular Neuroscience Department, Krasnow Institute for Advanced Study MS2A1, George Mason University, Fairfax, Virginia 22030-4444, USA.

[5]Children's Hospital of Philadelphia, University of Pennsylvania School of Medicine, ARC 517, 3615 Civic Center Boulevard, Philadelphia 19104-5127, USA.

[6]Department of Anatomy and Neurobiology, Washington University School of Medicine, St. Louis, Missouri 63110, USA .

[7]Réseau Cortical et Couplage Neurovasculaire, Laboratoire de Neurobiologie des Processus Adaptatifs UMR 7102 CNRS-UPMC, 9 Quai Saint-Bernard, 75 005 Paris, France.

[8]Instituto de Neurociencias, CSIC & Universidad Miguel Hernández, Sant Joan d'Alacant 03550, Alicante, Spain.

[9]Research Centre Juelich, Institute of Neuroscience and Medicine, INM-2, Function of Neuronal Microcircuits Group, Leo-Brandt-Strasse, D-52425 Juelich, Germany.

[10]Rheinisch-Westfaelische Technische Hochschule, Aachen University, Department of Psychiatry and Psychotherapy, Pauwelsstr. 30, D-52074 Aachen, Germany.

[11]NYU Neuroscience Institute Department of Physiology and Neuroscience, New York University Langone Medical Center, New York 10016, USA.

[12]Max Planck Florida Institute for Neuroscience, 1 Max Planck Way, Jupiter, Florida 33468, USA.

[13]Hungarian Academy of Sciences, Institute of Experimental Medicine, H-1083 Budapest, Szigony u. 43, Hungary.

[14]Department of Psychiatry, University of Pittsburgh School of Medicine, Biomedical Science Tower, 200 Lothrop Street, Pittsburgh, Philadelphia 15261, USA.

[15]Department of Comparative Medicine, Stanford University School of Medicine, 300 Pasteur Drive, Edwards 312/314 Stanford, California 94305, USA.

[16]Brain Mind Institute Ecole Polytechnique, Fédérale de Lausanne, CH-1015 Lausanne, Switzerland.

[17]Fishberg Department of Neuroscience and Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA.

[18]Beckman Building, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA.

[19]Division of Cerebral Circuitry, National Institute for Physiological Sciences, Myodaiji, Okazaki 444-8787, Japan.

[20]Department of Anatomy, Histology and Embryology, University of Debrecen, Debrecen, Hungary.

[21]Eunice Kennedy Shriver National Institute of Child Health and Human Development, 35 Lincoln Drive, Bethesda, Maryland 20892-3715, USA.

[22]Department of Neurosurgery, Technical University of Munich, Ismaninger Str. 22, 81675 Munich, Germany.

[23]Department of Clinical Neurobiology University Hospital and German Cancer Research Center (DKFZ) Heidelberg, 69120 Heidelberg, Germany.

[24]Department of Biology and Center for Behavioral Genomics, Brandeis University, Waltham, Massachusetts 02454, USA.

[25]Department of Anatomy and Neurobiology, Boston University School of Medicine, 72 East Concord Street, Boston, Massachusetts 02118, USA.

[26]CPN INSERM U894, Hôpital Sainte Anne, 2 ter rue d'Alésia, 75014 Paris, France.

[27]Genetics, Development and Behavioral Sciences Building, 1550 4th Street, University of California, San Francisco, California 94143-2611, USA.

[28]New York University, 522 First Avenue, Smilow Research Building, New York, New York 10016, USA.

[29]Howard Hughes Medical Institute, University of California, San Diego, Neurobiology Section 0634, Division of Biology, 9500 Gilman Drive, La Jolla, California 92093-0634, USA.

[30]Department of Neuroscience, Yale University School of Medicine, New Haven, Connecticut 06520-8001 USA.

[31]Department of Anthropology, Center for the Advanced Study of Hominid Paleobiology and GW Institute for Neuroscience, The George Washington University, Washington, DC 20052, USA.

[32]Georg-August-Universität Göttingen – Zentrum Anatomie, Abteilung Neuroanatomie, Kreuzbergring 36, D-37075 Göttingen, Germany.

[33]Department of Physiology, Anatomy and Neuroscience. University of Szeged, Szeged, H-6726, Hungary.

[34]Department of Pharmacology, University College of London School of Pharmacy, London, WC1N 1AX, England.

[35]School of Optometry and Ophthalmology, Wenzhou Medical College, Wenzhou, Zhejiang 325027, China.

[36]Steward St. Elizabeth's Medical Center. Tufts Medical School, Boston, Massachusetts 02135, USA.

[37]Howard Hughes Medical Institute, Columbia University, Biological Sciences, 550 West 120th Street, New York, New York 10027, USA.
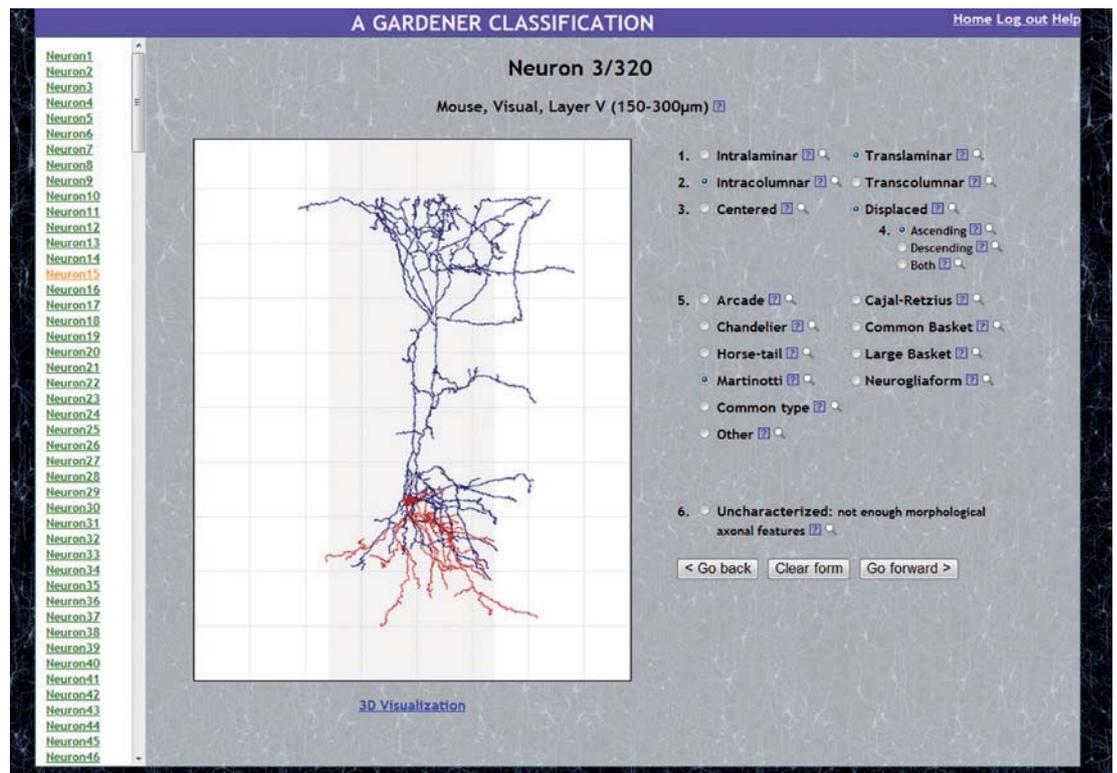
Figure 1 | **The web-based interactive system.** Screenshot of one of the 320 neurons included in the web-based interactive system. Also shown are the six axonal features and their categories (with possible values for each feature) displayed for the experts so they can select, for each feature, the category that is the most appropriate to describe the morphology of the neuron.

describe a novel, web-based interactive system (FIG. 1) that collected data about the terminological choices for a set of 320 cortical interneurons by 42 experts in the field. We used several analysis methods to empirically test the consistency, clarity and any emerging agreement on these terminological choices. This article deals primarily with neocortical GABAergic aspiny or sparsely spiny non-pyramidal neurons with non-projecting axons. Unless otherwise specified, we refer to these neurons, for simplicity, as 'cortical interneurons'.

Historical overview

Two major classes of cortical neurons: principal cells and interneurons. Before the discovery of the Golgi method, the existence of different morphological types of cortical neurons was already recognized[5]. Since then, researchers have tried to deduce the functional role of neurons from their morphological characteristics. Observations of nerve tissue preparations stained with a carmine dye, a technique that was introduced by Joseph von Gerlach (1820–1896) and Rudolf Berlin (1833–1897), led to the suggestion that neurons could be classified into three main cell types[6] (quoted in REF. 7) based on the shape of their somata: pyramidal cells (triangular somata), granular cells (small and irregular somata) and spindle-shaped cells (fusiform somata). This was the beginning of cytoarchitectural studies that were based mainly on the density and laminar distribution of different neuronal

shapes. However, carmine staining and other methods available at that time only allowed visualization of neuronal cell bodies and a small portion of their proximal processes, making further characterization of cortical neurons difficult. By contrast, Golgi-stained preparations allowed much more complete staining of the neuron, including most of its parts (soma, dendrites and the unmyelinated axon), enabling visualization of their finer morphological details in young animals[8]. This led to a fuller characterization of neurons, allowing for the first time the exploration of their possible interconnections.

According to Cajal[9], Golgi suggested that, in general, there were two morphologically and physiologically different types of neurons: motor (type I) neurons and sensory (type II) neurons. Motor neurons had long axons that not only gave rise to collaterals but also projected outside the grey matter. Sensory neurons had short axons that arborized near the parent cell and did not leave the grey matter. The former cells (with long axons) were thought to have a motor function because their axons were considered to be continuous with the motor roots, whereas the second type were thought to be sensory because their axonal branches were linked with afferent fibres. Cajal argued that it was not possible to maintain such a physiological distinction and designated Golgi's two types as cells with a long axon (projection neurons) and as cells with a short axon (intrinsic neurons or interneurons), avoiding any consideration

of their possible physiological roles. Since then, the term 'interneuron' has commonly been used as synonymous with short-axon cell[10,11]. Notably, some neurons are axonless, such as retina amacrine cells and olfactory granule cells.

Researchers soon realized that, in the cerebral cortex, most neurons were pyramidal cells with axons that were seen to enter or be directed towards the white matter (for example, see REF. 12). Therefore, pyramidal cells started being generally considered as both 'principal cells' and projection neurons (that is, cells with long axons).

Furthermore, from observations using the Golgi method, it was obvious that neurons showed a great diversity of morphologies. Thus, in addition to the terms based on the shape of the soma, neuroanatomists described neurons with names that were somewhat descriptive of their dendritic morphology and axonal arborization, alone or in combination. However, with few exceptions, no general consensus has emerged for naming cortical neurons. For example, at present most neuroscientists agree on the usage of terms such as pyramidal neuron, non-pyramidal neuron, interneuron and chandelier (or axo-axonic) cell. These cell types are readily distinguished by their clear morphological attributes. However, other common names, such as double bouquet cell, Martinotti cell, neurogliaform cell and basket cell, seem to lack a consensual definition. In these cases, the same name is often assigned to neurons of varying morphologies by different authors, and various terms are inconsistently adopted in different laboratories to represent the same cell classification. As a consequence, virtually every author has his or her own classification scheme and neuron terms, making the comparison and exchange of information among laboratories rather difficult, if not impossible.

What is a cortical interneuron? By our definition, a cortical interneuron is a short-axon cell — that is, a neuron with an axon that does not leave the neocortex — and has a soma that is located in the cerebral cortex. Most cortical interneurons lack the typical somatodendritic morphological characteristics that are used to identify projection neurons, namely a pyramidal-shaped cell body and an apical dendritic tree that is distinct from and lies opposite to the basal dendritic arbor. However, the absence of these features should not be used to define interneurons, as they are neither necessary nor sufficient for distinguishing interneurons from projection neurons. Indeed, there are interneurons that have a somatodendritic morphology resembling that of pyramidal cells (for example, the so-called 'pyramidal basket cells'[13], and projection neurons that have a non-pyramidal appearance in their somata and dendrites[14].

Traditionally, interneurons have been subdivided into two main groups[14]: spiny non-pyramidal cells and aspiny or sparsely spiny non-pyramidal cells. Spiny non-pyramidal cells are located in the middle cortical layers, especially in layer IV of primary sensory cortices. They comprise a morphologically heterogeneous group of interneurons with ovoid, fusiform and triangular somata. Most spiny non-pyramidal cells

are excitatory (specifically, glutamatergic[15]), and their axons are distributed within layer IV or in the adjacent layers above or below the somatic location[16]. Aspiny or sparsely spiny non-pyramidal cells usually have axons that remain near the parent cell, although some run prominent collaterals in the horizontal dimension (that is, parallel to the cortical surface) or vertical dimension (that is, ascending and/or descending to reach other cortical layers). These interneurons appear to be mostly GABAergic and constitute ~10–30% of the total neuron population, with the percentage varying substantially between cortical layers, areas and species[17,18]. They are the main component of inhibitory cortical circuits.

Following the approach of the Petilla terminology[1], we concentrate our effort on GABAergic cortical interneurons, thus excluding the majority of spiny non-pyramidal cells from the classification attempt. This choice is motivated by functional considerations, in the sense that the neurotransmitter released by a neuron is intimately linked to the role of this neuron in the circuitry. Moreover, restricting the scope of this classification to GABAergic interneurons has a practical reason, given the availability of reliable methods to identify GABA and related chemicals, such as its synthesizing enzymes (glutamate decarboxylase 65 (GAD65) and GAD67). Despite this relatively narrow definition, GABAergic cortical interneurons are located in all cortical layers and show a great variety of morphological, biochemical and physiological characteristics. Thus, rather than attempting a comprehensive classification of cortical interneurons, we focus on a group of less controversial cell types for which relatively more abundant experimental evidence converges on a limited number of defining properties within the anatomical domain.

Clarifications and remarks. In light of the above definitions, and before classifying specific interneuron types, it is useful to consider a number of points regarding the morphology and naming of cortical neurons raised by the collective work of many investigators.

First, over the years, the term interneuron has been most commonly used when referring to aspiny or sparsely spiny GABAergic non-pyramidal cells. These cells constitute the majority of interneurons and have come to epitomize the 'typical' cortical interneuron. As noted above, however, a minority of GABAergic interneurons are spiny[19]. Moreover, many interneurons that will become aspiny as they develop are spiny in the neonate[14]. For clarity, we propose to add the term 'spiny' to their name.

Second, some GABAergic non-pyramidal cortical cells (spiny and aspiny alike) project to other cortical areas[20,21] and might not therefore be strictly considered as interneurons. We propose to add the term 'projecting' to their name.

Third, there are glutamatergic spiny non-pyramidal and pyramidal cells (mostly in layer IV of sensory cortices) with locally confined axons that are distributed near the parent cell soma and do not leave the cortical grey matter. Therefore, they might be considered to

be short-axon neurons. However, because these cells are both morphologically and neurochemically rather distinct, we prefer to avoid the term 'interneuron' for glutamatergic spiny cells and propose to call them 'intrinsic (or local) glutamatergic spiny cells' instead.

Fourth, although most GABAergic interneurons have a non-pyramidal somatodendritic phenotype, some display a pyramidal (triangular) somatic shape. To minimize confusion, we propose to use the term 'triangular' to describe the somatic morphology of these interneurons.

Fifth, interneurons are highly diverse with regard to the morphology of their somata and of their dendritic and axonal arbors. For instance, interneurons displaying the same somatodendritic morphology may have different patterns of axonal arborization. Importantly, the axonal geometry is pivotal in establishing circuit connectivity. In several cases, axonal morphology is very distinct, facilitating comparisons of different interneurons. We therefore recommend, whenever historically tenable, using terms such as fusiform, stellate, multipolar, bitufted (neurons with two main dendrites running in opposite directions that, after a relatively short trajectory, resolve into two dendritic tufts) and bipolar (neurons with two principal long dendrites running in opposite directions and showing few dendritic collaterals) only to describe the somatic and/or dendritic morphology and not to name a particular interneuron type. Although these terms are useful descriptors of interneuron somatodendritic morphologies revealed by immunohistochemical staining against calcium-binding proteins and neuropeptides, such staining does not label the full extent of the axonal arbor and therefore does not allow one to unambiguously identify interneuron types. A good example is the double bouquet cell, a term adopted inconsistently in the literature. Some authors use this name for neurons with a bitufted dendritic morphology, regardless of the pattern of axonal arborization. Other authors use the term double bouquet cells for neurons with descending axons that form tightly intertwined bundles of long descending vertical collaterals resembling a horse tail[11]. Although these cells may have bitufted dendrites, interneurons with the same axonal patterns but with different somatodendritic morphologies also exist[14]. We propose that cortical interneurons identified by these characteristic axonal bundles be called 'horse-tail' cells.

Sixth, numerous neurons exist with axon collaterals that do not exhibit any orientation preferences. That is, they have more or less equal numbers of horizontal, oblique or vertical branches. In fact, most interneurons visualized in Golgi preparations or following intracellular labelling could match this description. We propose to introduce the term 'common type' to describe cells without any strikingly recognizable shape.

Seventh, an important morphological feature of cortical interneurons is the laminar and columnar reach of their axonal arbors. Following the Petilla terminology[1], we propose to describe neurons with an axonal arbor that is confined to a single layer as 'intralaminar', and neurons with an axonal arbor that is not confined to a single layer as 'translaminar'. Similarly, we refer to neurons with an axonal arbor that is confined to a single column as 'intracolumnar', whereas neurons with an axonal arbor that is not confined to a single column are referred to as 'transcolumnar' (FIG. 2).

Last, a relevant morphological feature of interneurons is the relative location of dendritic and axonal arbors. We propose to use the term 'centered' for neurons with dendritic and axonal arbors that are largely colocalized, and to use the term 'displaced' otherwise (FIG. 2). In the case of displaced neurons, axons of translaminar interneurons can be 'ascending' and/or 'descending' depending on whether, relative to the dendritic trees, they are distributed mostly towards the cortical surface, towards the white matter or approximately equally towards both.

Classification attempts

The Petilla terminology[1] considered the characteristics that are suitable for describing GABAergic cortical interneurons and organized them into morphological, physiological and molecular properties. Although the identity of a neuron is characterized by all of its properties, a typical experimental identification of a given neuron is commonly limited to a subset of properties. Indeed, most studies primarily (if not exclusively) rely on detailed anatomical, physiological or molecular evidence, and few studies use a balanced combination of these characteristics. Consequently, on the basis of existing data, neurons could in principle be classified using any of these groups of criteria. Several initial attempts at neuronal classification formulated from the Petilla terminology effort[1] are briefly summarized below.

Anatomical. The anatomical classification established in the Petilla terminology[1] divides GABAergic cortical interneurons into those targeting pyramidal cells or displaying no target specificity and, at least in the hippocampus, those specifically targeting other interneurons. Interneurons targeting pyramidal cells were further subdivided on the basis of the target location and included interneurons targeting the axonal initial segment (axo-axonic or chandelier cells), interneurons targeting the perisomatic region (basket cells) and interneurons targeting the dendrites. Basket cells were further distinguished, on the basis of their axonal morphology, into interneurons with tangential (horizontal) axons, interneurons with radial (vertical) axons, interneurons with both tangential and radial axons and interneurons with axons that are too local to discern a tangential or radial orientation. Dendrite-targeting interneurons were subclassified on an even finer scale as having either a shaft bias or a spine bias, with both of these categories finally separated on the basis of their axonal morphology. Shaft-biased interneurons have radial axons that either descend towards the white matter (these were termed willow cells) or ascend towards the pia (these were termed Martinotti cells). Spine-biased interneurons were further divided on the basis of their axonal patterns and include horse-tail and neurogliaform cells.
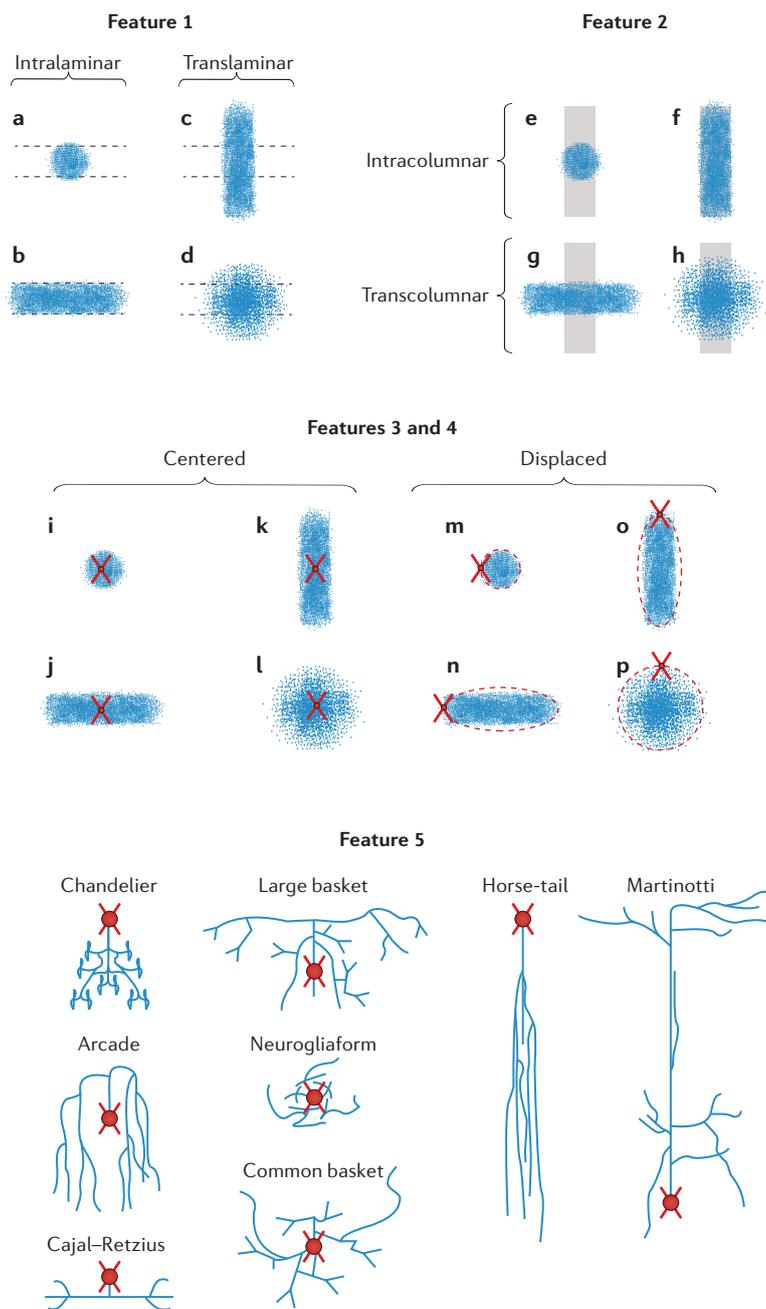
Feature 1

Intralaminar   Translaminar

a   c

b   d

Feature 2

Intracolumnar   e   f

Transcolumnar   g   h

Features 3 and 4

Centered   Displaced

i   k   m   o

j   l   n   p

Feature 5

Chandelier   Large basket   Horse-tail   Martinotti

Arcade   Neurogliaform

Cajal–Retzius   Common basket

Figure 2 | **Schematics of the morphological features.** For each feature, the experts had to select the category that best described the neuron on display. For feature 1, the categories were intralaminar (**a,b**) versus translaminar (**c,d**). For feature 2, they were intracolumnar (**e,f**) versus transcolumnar (**g,h**). For feature 3, the categories were centered (**i–l**) versus displaced (**m–p**). For feature 4, they were ascending, descending or both. (This feature applied only when neurons were translaminar and displaced; **o,p**.) For feature 5 (interneuron types), the categories were arcade, common basket, large basket, Cajal–Retzius, chandelier, horse-tail, Martinotti, neurogliaform, common type (not shown) or other (not shown). When an insufficient number of morphological axonal features are visualized for a given interneuron the cell is considered anatomically uncharacterized (feature 6; not shown). Dashed horizontal lines indicate the extent of the cortical layer. Vertical grey shadows indicate the extent of the cortical column. Axonal arborization is represented by blue dots. Soma and dendritic arborization are represented as red circles and crosses, respectively. Possible variations on the relative position of the somata with respect to the axonal arborization of displaced neurons are represented by red dashed ovals.

Molecular. The molecular classification of the Petilla terminology[1] divides cortical interneurons on the basis of the expression of specific molecular markers. In particular, five main groups of interneurons can be distinguished: those expressing parvalbumin (PV), including chandelier and basket cells; those expressing somatostatin (SOM), such as Martinotti cells; those expressing neuropeptide Y (NPY) but not SOM; those expressing vasoactive intestinal peptide (VIP); and those expressing cholecystokinin (CCK) but not SOM or VIP. These five groups can be further subdivided in multiple subtypes based on several molecular categories: transcription factors, neurotransmitters or their synthesizing enzymes, neuropeptides, calcium-binding proteins, neurotransmitter receptors, structural proteins, ion channels, connexins, pannexins and membrane transporters. For example, SOM-expressing interneurons can be subdivided depending on whether they also express NPY or calretinin (CR). Similarly, NPY-expressing interneurons and VIP-expressing interneurons can be subdivided depending on whether they also express CR[22,23]. A parallel effort to characterize interneurons based on transcription factors is also gaining traction[24]. This developmental classification separates cortical interneurons with an origin in the medial ganglionic eminence (MGE), lateral and dorsocaudal ganglionic eminence (CGE) and preoptic area (POA). The MGE group encompasses neocortical interneurons identified based on their molecular markers, including those expressing PV, SOM and, early in development, NPY. The CGE group includes the interneurons expressing both CR and VIP (horse-tail cells) and those expressing NPY later in development. The POA group expresses NPY. This mapping does not apply exactly to the hippocampus, as some expression differences have been reported in this area[25].

Physiological. The physiological classification of the Petilla terminology[1] identifies six main types of interneurons. Fast-spiking (FS) neurons show non-adapting spiking at steady-state, brief spikes and large fast after-hyperpolarizations and include continuous FS cells, delayed FS cells, stuttering FS cells and continuous stuttering FS cells. Non-adapting, non-fast spiking (NA-NFS) neurons display no apparent increase in the interspike interval at steady-state, and include continuous NA-NFS cells and burst-firing NA-NFS cells. Adapting (AD) neurons display a visually obvious increase in the interspike interval at steady-state and include continuous AD cells, bursting AD cells and delayed AD cells. Accelerating (AC) neurons display a decrease in the interspike interval at steady-state and include continuous AC cells and delayed AC cells. Irregular spiking (IS) neurons display an irregular interspike interval and include continuous IS cells and bursting IS cells. Lastly, intrinsic bursting (IB) neurons produce a stereotypical burst of two or more spikes riding on a depolarization envelope followed by a slow afterhyperpolarization potential and include rhythmic IB cells and initial IB cells.

Limitations of the Petilla terminology. Each of these classification schemes has limitations. For many cell types, the anatomical approach requires the identification of the subcellular postsynaptic target (or targets) in addition to the interneuron of interest. The molecular approach does not provide functional insight, as the functional roles of the most useful and commonly used markers are largely unknown. The physiological approach is greatly dependent on the experimental conditions and requires a complete specification and possibly standardization of experimental conditions to be widely acceptable. Thus, each of these complementary classifications provides only partial knowledge when taken individually, but a more comprehensive scheme involving multiple anatomical and functional criteria imposes considerable practical burdens.

Feature-based nomenclature proposal

As a pragmatic alternative and update to the anatomical characterization, we propose a taxonomic solution that is based mainly on axonal arborization patterns. We think that identification of these patterns may be among the most powerful tools available for the subclassification of interneurons.

Our classification design is based on six axonal features, numbered one to six (FIG. 2). These six features were selected as a representative subset of axonal morphological properties that may prove to be suitable for interneuron classification. After introducing all relevant definitions, we describe a web-based interactive system that is designed to evaluate this solution empirically, to test its potential for fostering consensus and to explore preliminary statistical patterns among the generated data (FIG. 1). Several statistical and pattern recognition techniques were used to achieve this goal, including the computation of agreement indices and the use of clustering and supervised classification algorithms.

First axonal feature. The first axonal feature refers to the distribution of the interneuron axonal arborization relative to cortical layers (FIG. 2). Within this feature, we propose two categories: intralaminar, which refers to interneurons with axonal arbors distributed predominantly in the layer of the parent soma; and translaminar, which refers to interneurons with axonal arbors distributed mainly above and/or below the cortical layer of the parent soma.

Second axonal feature. The second axonal feature refers to the distribution of the axonal arborization relative to the size of cortical 'columns' from a broad anatomical point of view. Certainly, the term column is vague[26,27], as it can refer to small-scale mini-columns (with a diameter of 50 μm), to larger-scale macro-columns (with a diameter of 300–500 μm) and to multiple different structures within both these categories (including barrel columns and ocular dominance columns, the extent of arborization of single thalamic afferent fibres, cytochrome oxidase blobs, individual dendritic arbors of pyramidal cells and tangential widths of axonal patches originated from pyramidal cells). Thus, we have arbitrarily set the size of a cortical column at a diameter of 300 μm — a value that is similar across several species and cortical areas for many

of these structures[28,29]. Within this feature, we propose two categories: intracolumnar, which refers to interneurons with axonal arbors primarily distributed at a distance from the parent soma that does not exceed 300 μm in the horizontal dimension (FIG. 2); and transcolumnar, which refers to interneurons with horizontal axonal collaterals exceeding a distance of 300 μm from the parent soma in the horizontal dimension.

Third axonal feature. The third axonal feature refers to the relative location of the axonal and dendritic arbors (FIG. 2). Within this feature, we propose the following categories: centered, which refers to interneurons with a dendritic arbor that is located mostly in the centre of the axonal arborization; and displaced, which refers to interneurons with a dendritic arbor that is shifted with respect to the axonal arborization (FIG. 2).

Fourth axonal feature. If a neuron is categorized as being both translaminar (for the first axonal feature) and displaced (for the third axonal feature), it can be further distinguished into the following categories[1]: ascending, which refers to interneurons with axonal arborization that is distributed mostly towards the cortical surface; descending, which refers to interneurons with axonal arborization that is distributed mostly towards the white matter; or 'both' (ascending and descending), which refers to interneurons with axonal arborization that is distributed towards both the cortical surface and the white matter (FIG. 2).

Fifth axonal feature: interneuron type. We defined a limited number of cell types for this classification step (see the Gardener Classification website) on the basis of recognizable morphological characteristics (FIG. 2) and the common usage of their name in the literature[14]. The first cell type — arcade or willow cells — denotes neurons with somata in layers II–VI, multipolar or bitufted dendrites and axons that give rise to axonal arcades, with predominantly vertical arbors and relatively long descending collaterals. The second cell type — common basket cells — denotes neurons with somata in layers II–VI, multipolar or bitufted dendritic arbors and axon collaterals that have numerous curved pre-terminal axon branches. The third cell type — large basket cells — denotes neurons with somata in layers II–VI, multipolar or bitufted dendrites and horizontally oriented axon collaterals that can reach a length of several hundred micrometres. These collaterals show numerous curved pre-terminal axon branches that innervate the somata and proximal dendrites of neurons. Frequently, these cells display one or several long descending axonal branches. The fourth cell type — Cajal–Retzius cells — denotes neurons with an axon plexus that is restricted to layer I and long dendrites with ascending branchlets to the pia. These neurons are not present in adult neocortex and in rodents persist only during the first two postnatal weeks[30] (but see REF. 31). Cajal–Retzius cells proper do not contain GABA or express GABA-synthesizing enzymes GAD65 and GAD67 (REFS 32,33). There are also GABAergic neurons with somata in layer I and prominent long horizontal

axon collaterals and/or dendrites[32], and these are often also named Cajal–Retzius neurons in the developing neocortex, despite their different molecular characteristics from Cajal–Retzius neurons proper[33]. Given the purely morphological nature of the present study, most of the authors practically considered any GABAergic neuron in layer I with horizontally oriented axonal arborization as a putative Cajal–Retzius cell. The fifth cell type — chandelier cells — denotes neurons with somata in layers II–VI, multipolar or bitufted dendritic arbors and terminal axon branches that form short vertical rows of boutons resembling candlesticks. These interneurons are also referred to as axo-axonic cells as they synapse on the axonal initial segment of their pyramidal targets. The sixth cell type — horse-tail cells — denotes neurons with somata mostly in layers II–III, multipolar, bitufted or bipolar dendrites and axons forming tightly intertwined bundles of long descending vertical collaterals. The seventh cell type — Martinotti cells — denotes neurons with somata in layers II–VI, multipolar, bitufted or bipolar dendrites and ascending axons that give rise to two axonal arbors, one near the soma and another at a variable distance above. This second plexus may be dense (axonal tuft) or diffuse and it can be either in the same layer as the soma of origin or in the layers above (ascending axons can travel from layer VI to layer I). The eighth cell type — neurogliaform cells — denotes neurons with somata in layers I-VI, multipolar dendritic arbors and that are characterized by very small and dense local axonal arborization around the parent cell body. Finally, we included the option common type to denote neurons with somata in layers I–VI, multipolar, bipolar or bitufted dendritic arbors and axon collaterals without any apparent target or orientation preference in the web-based interactive system (not shown in FIG. 2). Also, we added the option 'other' to label any given neuron with an alternative name in case the expert considered another term more appropriate.

Sixth axonal feature: uncharacterized versus characterized neurons. Interneurons that are uniquely characterized by peculiar morphological features can often be easily recognized, even when their axon is rather incompletely labelled. However, in many other cases, the axon needs to be fully labelled and reconstructed in order to distinguish the neuronal identity unequivocally. Thus, although it is not always necessary to visualize the full axonal and dendritic arborization to distinguish a given neuron, this is the preferred situation. Pragmatically, 'sufficiently complete' labelling simply means 'clear enough' to allow for the identification of a given morphological type. When an insufficient number of morphological axonal features are visualized for a given interneuron (because of incomplete staining, tissue slicing and so on), we propose that the cell should be deemed an anatomically 'uncharacterized' interneuron.

Study of inter-neuroscientist agreement
We designed and deployed an interactive web-based system (see the Gardener Classification website) to empirically test the level of agreement among 42 experts in the field in assigning the six features to individual cortical interneurons. The approach takes advantage of a common digital format to display, analyse and manipulate three-dimensional neuromorphological tracings reconstructed from light microscopy[34]. Images of the 320 interneurons included in the experiment were obtained either from the NeuroMorpho website[35] or by scanning two-dimensional drawings from previous publications. Altogether, this pool includes interneurons from different areas and layers of the cerebral cortex of the mouse, rat, rabbit, cat, monkey and human (Supplementary information S1). The database does not necessarily constitute a representative sample from the neuron population in different areas, layers and species. Furthermore, most of the anatomy recovered from electrophysiological work *in vitro* is conditioned by both slice thickness and plane of cut, which may vary across laboratories. Nonetheless, these conditions reflect the typical experimental variability that confronts researchers in the field.

Experienced neuroscientists who are knowledgeable in this field were asked to ascribe the categories they considered most appropriate to each neuron (there were six features and 21 categories in total; see FIG. 2). So, for feature 1 (F1) they would either ascribe a neuron the category intralaminar or the category translaminar. For feature 2 (F2), either intracolumnar or transcolumnar; for feature 3 (F3), either centered or displaced; for feature 4 (F4), ascending, descending or both; for feature 5 (F5), arcade common basket, large basket, Cajal–Retzius, chandelier, horse-tail, Martinotti, neurogliaform, common type or other; and for feature 6 (F6), either characterized or uncharacterized.

To study the agreement regarding the assignment of the features between neuroscientists, we computed typical statistical measures of inter-expert concordance for each feature and for each category (a possible value for a feature). We also identified sets of similar neurons using clustering algorithms. Furthermore, we induced from the data a Bayesian network model for each expert to enable us to analyse their choices by comparing the network structures of different neuroscientists (Supplementary information S1). With this approach, the possible reasoning of the experts can be inferred from the probabilistic models. Finally, we built automatic classifiers to assign each neuron to one category for each of the six features (Supplementary information S1).

Analysis of the raw data. First, we performed a preliminary exploratory analysis of the raw data to study how the votes of the experts were distributed for the different features. We assessed the relative frequency of each category in the experiment. Less than 10% of neurons were rated as anatomically uncharacterized; as described above, this pertains to neurons with an insufficient number of morphological axonal features to allow classification. Thus, the vast majority of the neurons in the experiment were considered as characterized. The most frequently assigned categories of descriptive axonal features proposed in this study were translaminar, intracolumnar and displaced. The categories ascending and descending received a similar percentage of the ratings, whereas fewer neurons were assigned to the category both.
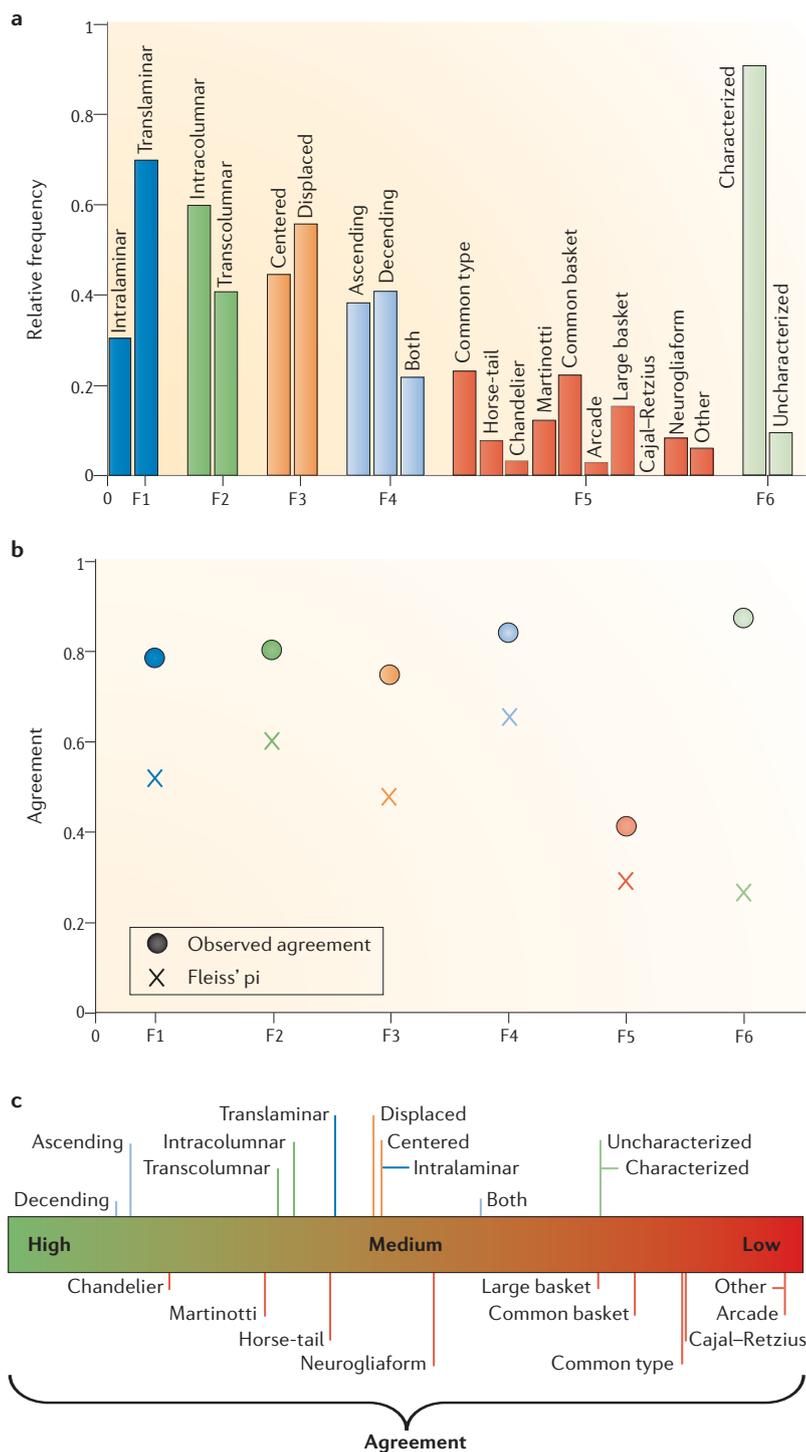
Figure 3 | **Agreement analysis. a** | Relative frequency of each category for each feature (F1 to F6): that is, the number of times a category was selected divided by the total number of ratings for the relevant feature. **b** | Overall observed agreement (circles) and chance-corrected Fleiss' pi index (crosses; see Supplementary information S1) for each feature, indicating the degree of concordance between the experts. **c** | Chance-corrected (Fleiss' pi index) agreement achieved in each category of each feature. Categories of the same feature are shown using lines with the same colour; for example, the categories intracolumnar and transcolumnar (which correspond to the second axonal feature) are shown with dark green bars. Interneuron types that were easily distinguished by the experts yielded high agreement (for example, the categories chandelier and Martinotti), whereas confusing categories, such as common type, common basket and large basket, yielded low chance-corrected agreement values.

We then assessed the frequency with which interneurons were assigned to specific interneuron types. The most commonly assigned interneuron types were common type, common basket and large basket. The interneuron types Martinotti, neurogliaform and horse-tail received an intermediate percentage of ratings, whereas chandelier and arcade received the lowest percentage of ratings. Only three cells were classified as Cajal–Retzius by six experts; the remaining experts classified these neurons as uncharacterized, common type, common basket, large basket, Martinotti or other.

Finally, we checked whether the names given to the 79 neurons that were scanned from original publications were maintained in the present experiment by the authors of those publications. Interestingly, the authors were frequently inconsistent in naming certain neurons. For example, some neurons named neurogliaform cells in the original publication were classified as uncharacterized in the current experiment by the same author.

Experts' agreement analysis. We computed statistical measures of inter-expert agreement to analyse the degree of concordance between the ratings given by the experts (Supplementary information S1). Here, the goal was to quantify the agreement among experts for each feature independently. We studied the agreement for both features and categories using the two most studied agreement indices: Fleiss' pi and Cohen's kappa indices.

We first analysed feature agreement. We found high levels of observed agreement between experts in the classification of neurons according to F1–F4 and F6 (observed agreement values exceeding 0.7; see FIG. 3b). The lowest level of inter-expert agreement (below 0.5) was found for F5.

The observed agreement values were then corrected for chance agreement. Thus, when the inter-expert coincidence was above random levels, the chance-corrected agreement indices yielded values above 0. After correcting for chance agreement (see Supplementary information S1), the highest chance-corrected Fleiss' pi inter-expert agreement was found for F4 (FIG. 3b). F1, F2 and F3 yielded intermediate chance-corrected agreement values, whereas F5 and F6 had low agreement. The difference between 'observed agreement' and Fleiss' pi index was particularly high for 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. This was due to the fact that the category prevalence of this feature was very unbalanced, such that characterized neurons were much more frequent than uncharacterized ones, reducing the values of the agreement measures (Supplementary information S2).

We then calculated the chance-corrected agreement achieved for each category of every feature (Supplementary information S2). Ascending and descending were the two categories with the highest inter-expert agreement, as indicated by the high values obtained for the chance-corrected Fleiss' pi index (see FIG. 3c and figure S7 in Supplementary information S2). Medium–high agreement levels were found for the categories intralaminar, translaminar, intracolumnar, transcolumnar, centered and displaced.
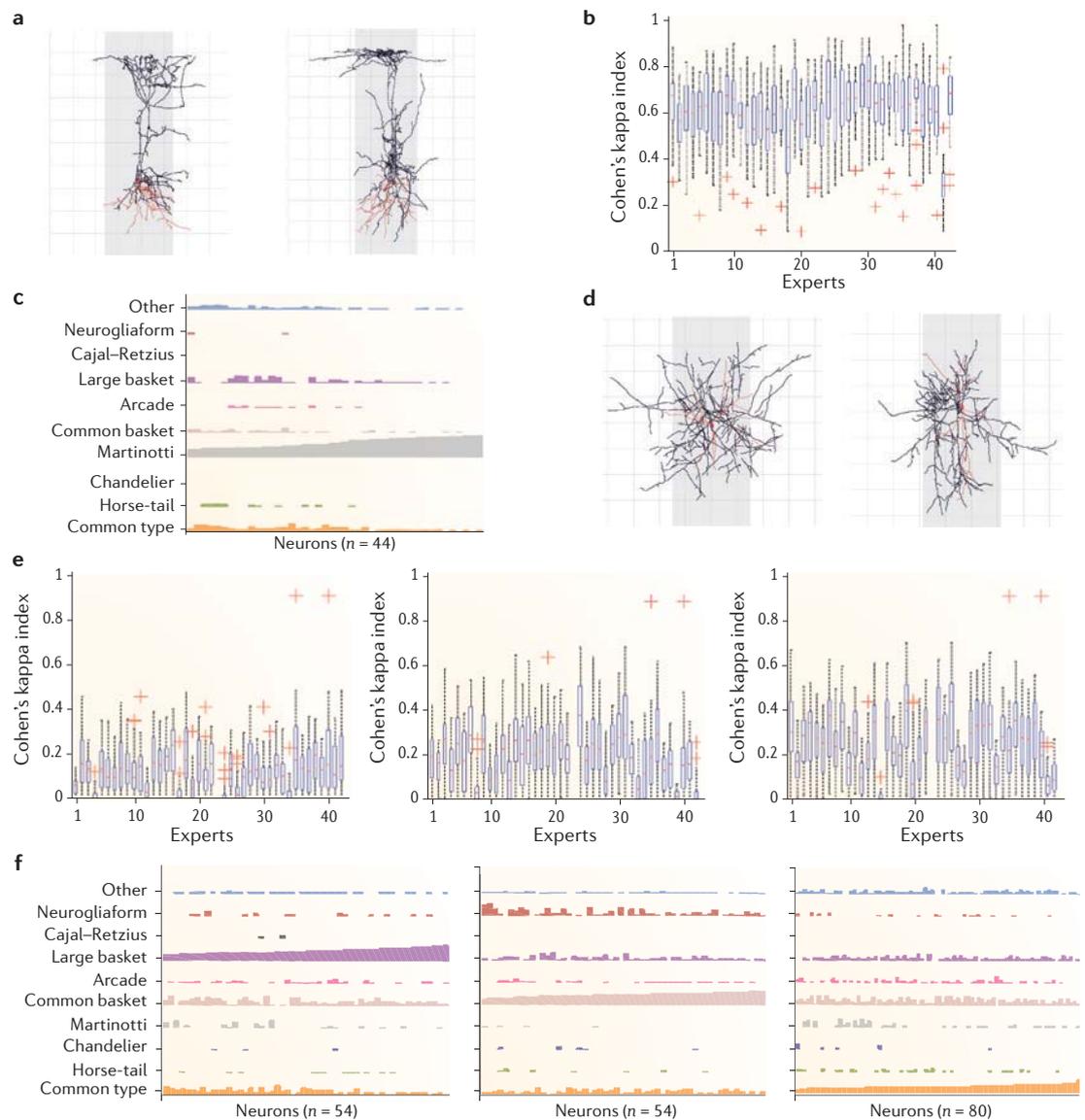
Figure 4 | **Examples of inter-expert agreement and disagreement.** **a** | Examples of neurons (neurons 3 and 272) categorized by 41 out of 42 experts as Martinotti. **b** | Box plots showing the agreement (quantified by Cohen's kappa index) between pairs of experts when comparing cells categorized as Martinotti against all the other interneuron types. For example, the first blue box shows the agreement values between the expert 1 and the other 41 experts when classifying interneurons as Martinotti cells. High values of Cohen's kappa index indicate high levels of inter-expert agreement. Apart from expert 41, the other experts yielded fairly high agreement when categorizing interneurons as Martinotti cells. The bottom and top of the blue boxes in the box plot are the lower and upper quartiles, respectively; the ends of the whiskers indicate the still considered typical values; the red crosses show outliers. **c** | A cluster of 44 neurons (shown from left to right) and the way they were assigned to one of the ten categories (each in different colour) of feature 5 by the experts. A vertical bar is shown for each category and each neuron, and the height of each bar indicates the number of experts who selected that category for that neuron. In the case of these 44 neurons, the neurons were classified as Martinotti by most of the experts. **d** | The left panel is an example of a neuron (neuron 31) that was categorized by 12 experts as common type, by 12 other experts as common basket, by 15 experts as large basket and by two experts as arcade. The right panel is another example of a neuron (neuron 274) that was categorized by 11 experts as common type, by 12 as common basket, by 14 as large basket, by one as chandelier, by one as arcade and by one as other. **e** | Low agreements between pairs of experts, as quantified by Cohen's kappa index, when categorizing interneurons as common type (left), common basket (middle) and large basket (right) against all the other interneuron types. **f** | Examples of clusters of neurons (54, 54 and 80 neurons, respectively) that show no unique category with high bars (compared with panel **c**). The graphs show that, in each cluster, the neurons received a high number of votes for common type, common basket and large basket rather than mainly for one category. Thus, the categories that were selected most often — large basket (left), common basket (middle) and common type (left) — were nevertheless selected less often (shorter bars) than the category Martinotti in panel **c** (longer bars). Note that a high number of experts also categorized the neurons as neurogliaform or common basket (high bars in middle and right panels, respectively).

Regarding F5, we found that the category chandelier yielded the highest agreement (that is, there was little disagreement between all experts over whether a given neuron should be classified as a chandelier cell). The level of agreement was high or medium for Martinotti, horse-tail and neurogliaform cells, whereas it was lower for the rest of the proposed interneuron types (large basket, common basket, common type, Cajal–Retzius, arcade and other). As in the above agreement analysis for F6, characterized and uncharacterized were the categories with the lowest level of chance-corrected inter-expert agreement (Supplementary information S2). Moreover, an analysis of chance-corrected Fleiss' pi index in which one or three experts were removed showed similar results, revealing those experts who contributed to the low agreement for some features (figure S8 in Supplementary information S2).

Additionally, we assessed whether Fleiss' pi values changed if two categories of F5 were merged into one category. The rationale for this was that certain pairs of categories seemed to overlap in terms of the neurons that were assigned to them. In fact, Fleiss' pi values increased when the categories common type, common basket and large basket were merged with each other (table S4 in Supplementary information S2); this reveals that these neuron types are ill-defined. By contrast, combining the Martinotti and/or chandelier categories with other categories yielded a lower chance-corrected agreement, suggesting that these neuron types are well defined. Furthermore, the above results were confirmed in a separate analysis using Cohen's kappa index (FIG. 4b,e). This index is defined for scenarios with two experts and two categories. Thus, we assessed the level of agreement between all possible pairs of experts, resulting in a comparison of each expert with all the other experts (figures S9–S11 in Supplementary information S2). For example, the first blue box in FIG. 4b summarizes the agreement between the first expert and the other 41 experts regarding the categorization of a neuron as Martinotti. Thus, this high-valued box means that this expert categorized the same neurons as Martinotti as the majority of the remaining experts. Also, we can conclude that there was a high agreement between experts for the category Martinotti, as all box plots (excluding expert 41) showed high Cohen's kappa index values (FIG. 4b). By contrast, there was a low level of agreement for common type, common basket and large basket cells, as reflected by the low values of the box plots (FIG. 4e). See Supplementary information S2 for further analyses regarding Cohen's kappa index.

Neuron clustering. We used clustering algorithms on the classification data from the experts to find groups of interneurons (clusters) with similar morphological properties. The rationale for this analysis was not to define interneuron types but to check whether the experts' votes for a given feature could separate neurons into clear groups. We performed the clustering analysis at two levels: neuron clustering for each feature and neuron clustering for all the features (Supplementary information S1).

First, we grouped the 320 neurons considering each feature independently. Thus, the clustering algorithm takes into account, for a given feature, which category was selected for each neuron by each individual expert. For F1–F3 and F6 (figures S12–S14 and S17 in Supplementary information S2), clear clusters of neurons could be identified for each category, whereas the clusters for F4 (figure S15 in Supplementary information S2) showed confusion about the category both. With regard to F5, we ran the algorithm to divide the set of 320 neurons into eight clusters (Supplementary information S1). FIG. 4c shows one cluster of neurons that clearly corresponds to Martinotti cells. By contrast, the panels in FIG. 4f show clusters that did not identify neurons corresponding to a single category; these clusters contained neurons that mainly corresponded to those categories for which no agreement was achieved by the experts. Results for the remaining clusters of F5 are reported in figure S16 in Supplementary information S2. These results indicate that although the scientific community was clear about some concepts (F1–F3 and F6), other categories (in F4 and F5) were controversial.

Next, we used another clustering algorithm to analyse the neurons, now taking into account all of the features at the same time. In this case, for a given neuron, the algorithm analyses the number of experts who selected each category of each feature without distinguishing between individual experts. This allowed us to study possible relationships between the features. FIG. 5 represents the clusters obtained in the analysis. We found some clusters containing neurons with clearly identified categories. For example, FIG. 5a shows a cluster of neurons that were clearly categorized as intralaminar, intracolumnar, centered and characterized. Furthermore, some of these neurons were mainly categorized as either common type, chandelier, common basket or neurogliaform. Similarly, FIG. 5b shows neurons that were mainly categorized as translaminar, transcolumnar, displaced. ascending, Martinotti and characterized. By contrast, FIG. 5c shows a cluster of neurons that were mainly categorized as translaminar and intracolumnar but that were not clearly categorized for the rest of the features. Finally, FIG. 5f shows a cluster of neurons showing no clearly identified categories, corresponding mainly to uncharacterized neurons.

Bayesian networks for modelling experts' opinions. Bayesian network models can capture the way by which an expert understands the (probabilistic) relationships among all the features (see REFS 36,37 for an application to neuroanatomy). As opposed to the analyses above, which focused on studying each feature independently, Bayesian networks enable us to analyse the associations between a set of features. The graphical representation of Bayesian networks allows one to visualize and inspect the relationships between the features. Here, we trained a probabilistic graphical model for each expert and used these models to analyse the experts' choice behaviours. In general, some Bayesian networks presented similar structures, whereas others showed different relationships between
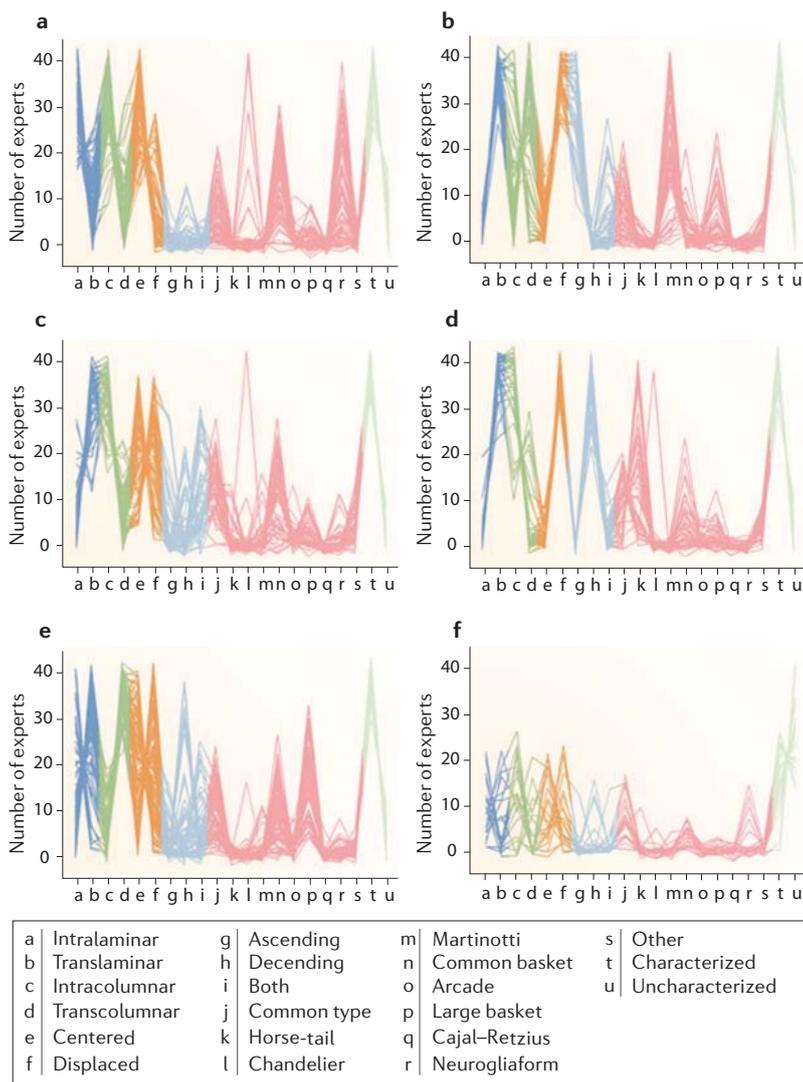
Figure 5 | **Clustering of neurons considering all features. a–f** | Parallel coordinate diagrams of clusters of neurons obtained with the *k*-means algorithm (*k* = 6) considering all of the features at the same time. Each line represents one neuron, showing the number of experts who selected each category of every feature when classifying that neuron. For example, panel **b** shows a cluster in which the majority of neurons were categorized by many experts as translaminar (dark blue), transcolumnar (dark green), displaced (orange), ascending (light blue), Martinotti (pink) and characterized (light green).

a	Intralaminar	g	Ascending	m	Martinotti	s	Other
b	Translaminar	h	Decending	n	Common basket	t	Characterized
c	Intracolumnar	i	Both	o	Arcade	u	Uncharacterized
d	Transcolumnar	j	Common type	p	Large basket		
e	Centered	k	Horse-tail	q	Cajal–Retzius		
f	Displaced	l	Chandelier	r	Neurogliaform		

the features. For example, FIG. 6 shows the Bayesian networks 'learned from' experts 16 and 27 when they selected Martinotti or common basket. The two models had a different structure, as shown by the variations in the relationships between the features (see FIG. 6 and figure S18 in Supplementary information S2). Additionally, Bayesian networks allow us to draw probabilistic conclusions about the categories. On the basis of Bayes' rule, we can also infer the likely reasoning of each expert and compare the behaviours of the different experts. Here, we selected some of the categories of F5 as evidence and inferred the most probable values for the rest of the features. This enabled us to identify the main properties for each interneuron type, allowing us to study the different conceptual thinking of the experts. In general, when

we studied categories with high levels of agreement, the propagated probabilities were similar in all the Bayesian networks. For example, when the category Martinotti was analysed, Bayesian networks yielded similar propagated probabilities (for example, FIG. 6a and figure S18 in Supplementary information S2). By contrast, when we analysed a category with a low level of agreement, the propagated probabilities were clearly different (for example, common basket in FIG. 6b and in figure S19 in Supplementary information S2). That is, experts had a similar concept for Martinotti cells, whereas, for common basket cells, they rather differed in their reasoning for assigning this interneuron type (see Supplementary information S2 for further details).

Supervised classification of neurons: automatic classification. The ultimate goal of our experiment was to build a model that could automatically classify a neuron on the basis of its morphological characteristics and, more specifically, in terms of the six features defined in the present study. A supervised classifier is a model that can assign a category to a neuron based on its characteristics. Such a classifier must be trained with a dataset of neurons for which the true category is known. For this purpose, we used those neurons from the experiment (241) that had been reconstructed in three dimensions. We first measured 2,886 morphological parameters using Neurolucida Explorer. Then, we built mathematical models that could automatically classify these 241 neurons according to the values of their morphological parameters[38,39]. Because supervised classification tools require a single class value for each neuron, we used a naive approach of assigning to each neuron the category that received the highest number of votes[40] (Supplementary information S1). As a first approach, we built six classifiers — one per feature (F1–F6). Moreover, we tested several different supervised classifier algorithms (Supplementary information S1). The easiest problem was classifying the neurons as either characterized or uncharacterized. This problem was solved with the highest accuracy, with only two misclassified neurons for this feature; that is, for this feature, the result of the classifier matched that of the (majority of the) experts. For F1, F2 and F3, the classifiers also yielded fairly high accuracies (table S6 in Supplementary information S2). By contrast, the accuracy for Feature F4 was much lower. There could be two main reasons for this low accuracy. First, the category 'both' was confusing for the experts, so the assigned category using the majority of votes might not capture the true morphological properties of the neurons for this feature. Second, none of the used morphological variables (Supplementary information S1) might adequately capture the vertical orientation of the axon with respect to the soma. Considering additional variables that specifically refer to the orientation of the axon might help improve the accuracy of the classifiers for this feature. The classifiers also yielded low accuracies when distinguishing the categories in F5 (Supplementary information S2). These results were not surprising because distinguishing the nine proposed neuronal types proved to be difficult for the experts.

a   **Expert 16**

Expert 27
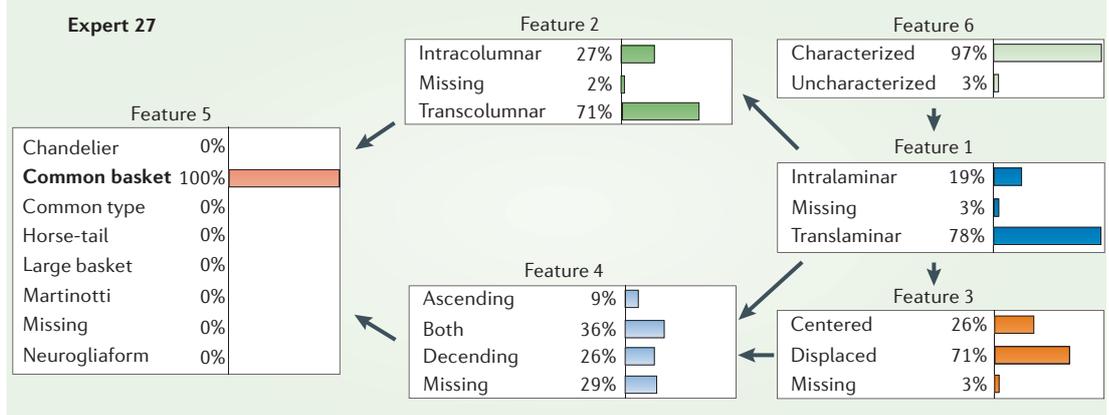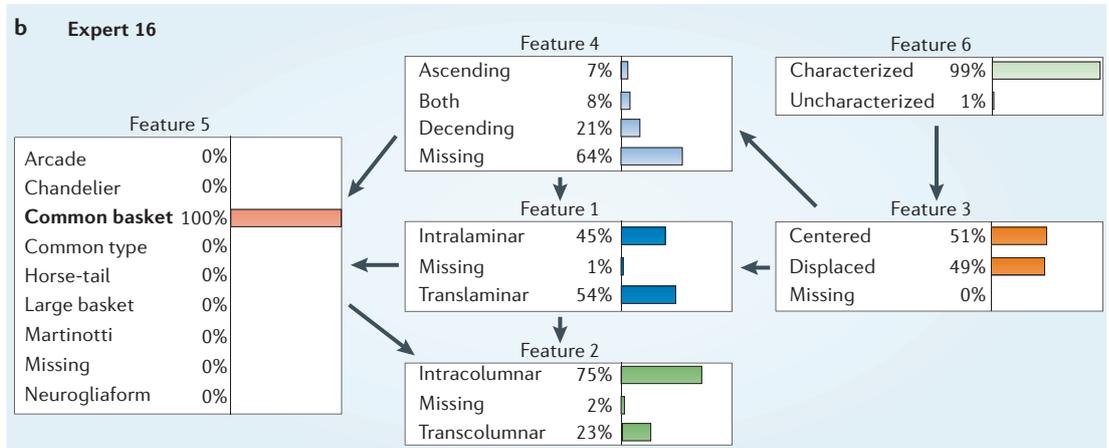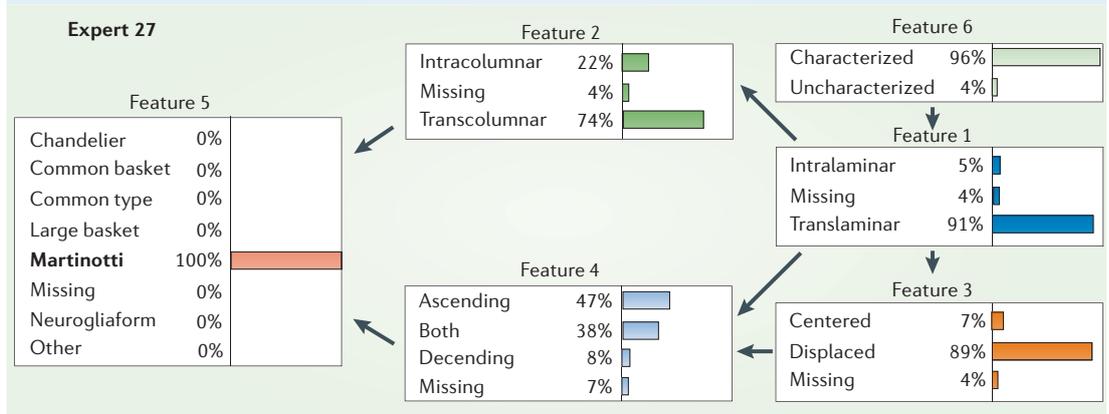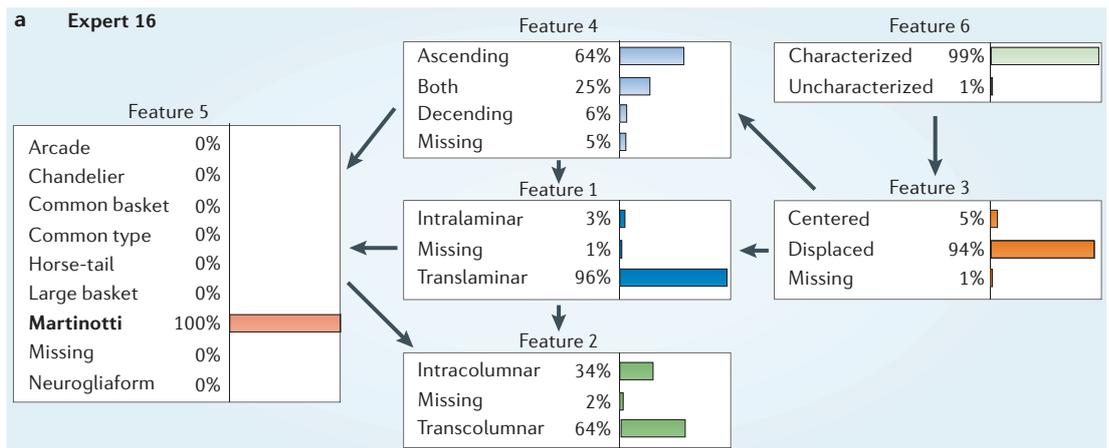
b   **Expert 16**

Expert 27

◄ Figure 6 | **Examples of Bayesian networks.** Examples of Bayesian network models of the choice behaviour of two experts (expert 16 and expert 27) when selecting the categories Martinotti (**a**) or common basket (**b**) in feature 5. In a Bayesian network structure, each feature is represented with a node (box) in the graph, and an arrow from one node X to another node Y in the graph represents the probabilistic dependence of Y on X (not shown here; see Supplementary information S1 for further details). Note that the direction of an arrow between two nodes does not necessarily reveal causality or hierarchy but merely shows a probabilistic relationship between the two corresponding features. When a category is selected (for example, Martinotti as neuron type in part **a**), probabilistic rules are used to propagate this information and to compute the conditional probability of any other node (for example, ascending as feature 4), shown by bar charts in this figure. Thus, the blue bar in feature 4 of part **a** means that if expert 16 called a neuron Martinotti, there was a 64% probability that he or she would consider it ascending. Similarities and differences between experts can be identified by comparing their Bayesian networks. For instance, arrows connecting feature 4 to feature 5 appear in both Bayesian networks, showing a common relationship for experts 16 and 27. Also, the propagated conditional probabilities can be used to compare experts' opinions. When selecting Martinotti, the propagated probabilities (shown by percentages and coloured bars) are similar in the two Bayesian networks; for example, translaminar in feature 1 has 96% probability for expert 16 and 91% probability in the panel for expert 27 (**a**). By contrast, the propagated probabilities when selecting common basket differ greatly; for example, there is 75% probability that expert 16 will select intracolumnar in feature 2 and 27% probability that expert 27 will select intracolumnar in feature 2 (**b**).

Additionally, we further analysed F5 by training binary classifiers that distinguished one category against all the other categories that were considered together. We drew similar conclusions as those obtained in previous analyses (Supplementary information S2). Finally, we observed a frequent disagreement between the categories common type, common basket and large basket throughout the analyses of the supervised classification experiment, and we therefore merged these three categories and repeated the automatic classification experiment. This increased the accuracy of the classification (Supplementary information S2).

Discussion and future directions

This study empirically and quantitatively demonstrates that the gardener's approach to neuron classification is untenable at this time and confirms the impression that different investigators use their own, mutually inconsistent schemes for classifying neurons based on morphological criteria. Many ambiguities are independent of the relative reconstruction quality and completeness of the tested neurons. A striking indication of the problem is that in several cases, experts assigned a different name to a neuron than the term they had chosen in their own original publication from which that same neuron was taken. This takes us back to the time of Cajal, who also inconsistently named various morphological types of interneurons. For example, Cajal termed neurons with different dendritic and axonal morphologies 'double bouquet cells' (*células bipenachadas* in Spanish; bitufted cells in English)[11]. In the present study, however, statistical

analyses of inter-expert agreement, application of Bayesian networks and different clustering and supervised classification algorithms clearly separated readily distinguishable interneuron types from apparently confusing interneuron names. High-consensus terms included chandelier and Martinotti cells, indicating that these are more easily identifiable interneuron types. Low-consensus terms included arcade, basket cells and Cajal–Retzius cells, suggesting that these are potentially less useful names. Researchers generally agreed on specific morphological features, such as ascending versus descending and intracolumnar versus transcolumnar axonal arbors.

A solution: the Neuroclassifier. How might the situation be improved? On the basis of the supervised classification models described here, we have started the development of a computer tool for automatic classification of neurons, a 'Neuroclassifier'. This machine will initially use probabilistic labels — based on the categories provided by experts — as neuron names and will evolve by combining supervised (known labels) and unsupervised (new labels) classification techniques. This may foster naming unification, robust classification and education of new students in the field through online learning techniques. As the scientific community uses the tool, more data will be incorporated into the Neuroclassifier, allowing model updates and increasing classification robustness and accuracy. Furthermore, other morphometric measurements encoding aspects of neuronal anatomy that are important for cortical circuit organization could be considered, including the percentage of axonal arbors that lie inside the cortical layer and column. Eventually, multiple correlative criteria — including molecular, physiological and synaptic connectivity attributes — would enable a more complete neuronal classification, which is a critical step towards better understanding of neuronal circuits.

Importantly, it should be kept in mind that the present analysis is limited to neurons from a small number of species, namely mammals commonly used in brain research. These include one lagomorph, two rodents, one felid and two primates. Although the results from our analysis may be consistent among these mammalian orders, the level of inter-expert agreement was not compared between species. Furthermore, the selection of interneurons from these species does not cover the probable variability of interneuronal morphologies among all mammalian families. In fact, except for the cat, the species in our study all belong to only one mammalian superorder — the Euarchontoglires. Although several 'canonical' neuronal morphologies are doubtlessly common to all placental mammals, some species (such as cetartiodactyls and xenarthrans) depart from the commonly observed neuron types[41,42]. Future inclusion of other species in the Neuroclassifier will allow detailed analysis of evolutionary conservation and species-specific neuron types.

1. Petilla Interneuron Nomenclature Group. Petilla Terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nature Rev. Neurosci.* **9**, 557–568 (2008). **A representative group of researchers proposed a standardized nomenclature of interneuron properties,** the Petilla terminology, which has been used as a recognized reference in the recent literature.

2. Jackson, J. H. On classification and on methods of investigation (1874) in *Selected Writings* (ed. Taylor, E., Hodder and Stoughton, 1931).

3. Hamilton, D., Shepherd, G. M., Martone, M. E. & Ascoli, G. A. An ontological approach to describing neurons and their relationships. *Front. Neuroinform.* **6**, 15 (2012). **This article provides a clear introduction to the neuroinformatics infrastructure requirements of**

neuronal classification, with a review of the outstanding technical, scientific and social challenges for the field.

4. Larson, S. D. & Martone, M. E. Ontologies for neuroscience: what are they and what are they good for? *Front. Neurosci.* **3**, 60–67 (2009).
5. von Kölliker, A. *Handbuch der Gewebelehre des Menschen.* (Engelmann, 1852) (in German).
6. Berlin, R. *Beitrag zur Strukturlehre der Grosshirnwindungen* (Junge, 1858) (in German).
7. Clarke, E. & O'Malley, C. D. *The Human Brain and Spinal Cord. A Historical Study Illustrated by Writings from Antiquity to the Twentieth Century.* (Univ. of California Press, 1968).
8. Golgi, C. Sulla struttura della sostanza grigia del cervello (Comunicazione preventiva). *Gazz. Med. Ital. Lombardia* **33**, 244–246 (1873) (in Italian).
9. Cajal, S. R. Nuevo concepto de la histología de los centros nerviosos. *Revista de Ciencias Médicas de Barcelona* **18**, 361–376 (1892); English translation available in DeFelipe, J. & Jones, E. G. *Cajal on the Cerebral Cortex* (Oxford Univ. Press, 1988).
10. Rakic, P. *Local Circuit Neurons* (The MIT Press, 1976).
11. DeFelipe, J. Cortical interneurons: from Cajal to 2001. *Prog. Brain Res.* **136**, 215–238 (2002).
12. Meynert, T. in *Handbuch der Lehre von den Geweben des Menschen und der Thiere* Vol. 1. (ed. Stricker, S.) 694–808 (Verlag von Wilhelm Engelmann, 1871) (in German).
13. Amaral, D. & Lavenex, P. in *The Hippocampus Book* (eds Andersen, P., Morris, R., Amaral, D., Bliss, T. & O'Keefe, J.) 37–114 (Oxford Univ. Press, 2006).
14. Jones, E. G. & Peters, A. (eds) *Cerebral Cortex: Volume 1: Cellular Components of the Cerebral Cortex* (Plenum Press, 1984).
    This volume represents a standard reference work on the histology of cortical neurons.
15. Feldmeyer, D., Egger, V., Lubke, J. & Sakmann, B. Reliable synaptic connections between pairs of excitatory layer 4 neurones within a single 'barrel' of developing rat somatosensory cortex. *J. Physiol.* **521**, 169–190 (1999).
16. Staiger, J. F. *et al.* Functional diversity of layer IV spiny neurons in rat somatosensory cortex: quantitative morphology of electrophysiologically characterized and biocytin labeled cells. *Cereb. Cortex* **14**, 690–701 (2004).
17. DeFelipe, J., Alonso-Nanclares, L. & Arellano, J. I. Microstructure of the neocortex: comparative aspects. *J. Neurocytol.* **31**, 299–316 (2002).
18. Meyer, H. S. *et al.* Inhibitory interneurons in a cortical column form hot zones of inhibition in layers 2 and 5A. *Proc. Natl Acad. Sci. USA* **108**, 16807–16812 (2011).
19. Kubota, Y. *et al.* Selective coexpression of multiple chemical markers defines discrete populations of neocortical GABAergic neurons. *Cereb. Cortex* **21**, 1803–1817 (2011).
20. Tomioka, R. & Rockland, K. S. Long-distance corticocortical GABAergic neurons in the adult monkey white and gray matter. *J. Comp. Neurol.* **505**, 526–538 (2007).
21. Melzer, S. *et al.* Long-range-projecting GABAergic neurons modulate inhibition in hippocampus and entorhinal cortex. *Science* **335**, 1506–1510 (2012).
22. Karagiannis, A. *et al.* Classification of NPY-expressing neocortical interneurons. *J. Neurosci.* **29**, 3642–3659 (2009).
23. Porter, J. T. *et al.* Properties of bipolar VIPergic interneurons and their excitation by pyramidal neurons in the rat neocortex. *Eur. J. Neurosci.* **10**, 3617–3628 (1998).
24. Welagen, J. & Anderson, S. Origins of neocortical interneurons in mice. *Dev. Neurobiol.* **71**, 10–17 (2011).
25. Tricoire, L. *et al.* A blueprint for the spatiotemporal origins of mouse hippocampal interneuron diversity. *J. Neurosci.* **31**, 10948–10970 (2011).
26. Rakic, P. Confusing cortical columns. *Proc. Natl Acad. Sci. USA* **105**, 12099–12100 (2008).
27. Rockland, K. S. Five points on columns. *Front. Neuroanat.* **4**, 22 (2010).
28. Malach, R. Cortical columns as devices for maximizing neuronal diversity. *Trends Neurosci.* **17**, 101–104 (1994).
29. Mountcastle, V. B. *Perceptual Neuroscience: The Cerebral Cortex.* (Harvard Univ. Press, 1998).
30. Chowdhury, T. G. *et al.* Fate of Cajal–Retzius neurons in the postnatal mouse neocortex. *Front. Neuroanat.* **4**,10 (2010).
31. Marín-Padilla, M. Cajal-Retzius cells and the development of the neocortex. *Trends Neurosci.* **21**, 64–71(1998).
32. Meyer, G., Goffinet, A. M. & Fairén, A. What is a Cajal-Retzius cell? A reassessment of a classical cell type based on recent observations in the developing neocortex. *Cereb. Cortex* **9**, 765–775 (1999).
33. Hevner, R. F., Neogi, T., Englund, C., Daza, R. A. & Fink, A. Cajal-Retzius cells in the mouse: transcription factors, neurotransmitters, and birthdays suggest a pallial origin. *Dev. Brain Res.* **141**, 39–53 (2003).
34. Ascoli, G. A. Mobilizing the base of neuroscience data: the case of neuronal morphologies. *Nature Rev. Neurosci.* **7**, 318–324 (2006).
    A broad perspective on the bioinformatics potential of digital neuronal morphology for the advancement of basic and computational neuroscience.
35. Ascoli, G. A., Donohue, D. E. & Halavi, M. NeuroMorpho.Org: a central resource for neuronal morphologies. *J. Neurosci.* **27**, 9247–9251 (2007).
36. Pearl, J. *Probabilistic Reasoning in Intelligent Systems* (Morgan Kaufmann, 1988).
37. López-Cruz, P. L., Bielza, C., Larrañaga, P., Benavides-Piccione, R. & DeFelipe, J. Models and simulation of 3D neuronal dendritic trees using Bayesian networks. *Neuroinformatics* **9**, 347–369 (2011).
38. Duda, R. O., Hart, P. E. & Stork, D. G. *Pattern Classification* 2nd edn (Wiley-Interscience, 2001).
    This textbook clearly describes the main supervised and unsupervised classification methods.
39. Jain, A. K., Duin, R. P. W. & Mao, J. Statistical pattern recognition: a review. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**, 4–37 (2000).
40. Raykar, V. C. *et al.* Learning from crowds. *J. Mach. Learn. Res.* **11**, 1297–1322 (2010).
41. Hof, P. R. *et al.* Cellular distribution of the calcium-binding proteins parvalbumin, calbindin, and calretinin in the neocortex of mammals: phylogenetic and developmental patterns. *J. Chem. Neuroanat.* **16**, 77–116 (1999).
42. Sherwood, C. C. *et al.* Neocortical neuron types in Xenarthra and Afrotheria: implications for brain evolution in mammals. *Brain Struct. Funct.* **213**, 301–328 (2009).

FURTHER INFORMATION
Gardener Classification: http://cajalbbp.cesvima.upm.es/gardenerclassification
ICNF Program on Ontologies of Neural Structures: http://incf.org/core/programs/pons
NeuroLex: http://neurolex.org/wiki/Category:Neuron
NeuroMorpho: www. NeuroMorpho.Org

SUPPLEMENTARY INFORMATION
See online article: S1 | S2

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

Supplementary Online Information S1

A set of experts on cortical interneurons from different laboratories were asked to classify morphological reconstructions of interneurons (http://cajalbbp.cesvima.upm.es/gardenerclassification). The main characteristics of the experiment were:

- Number of experts who started the experiment: 48. The number of experts who completed the experiment was 42. We used only data from these completed experiments for our analyses.
- Number of neurons: 320.
    - 241 are 3D reconstructions from NeuroMorpho.Org (Ascoli et al., 2007).
    - 79 are 2D images scanned from older papers.
- Number of features for each neuron: 6. Total number of categories (each possible value of every feature): 21.
    1. Categories for Feature 1 (F1): Intralaminar vs Translaminar
    2. Categories for Feature 2 (F2): Intracolumnar vs Transcolumnar
    3. Categories for Feature 3 (F3): Centered vs Displaced
    4. Categories for Feature 4 (F4): Ascending vs Descending vs Both
    5. Categories for Feature 5 (F5) (interneuron type): Common type, Horse-tail, Chandelier, Martinotti, Common basket, Arcade, Large basket, Cajal-Retzius, Neurogliaform or Other
    6. Categories for Feature 6 (F6): Characterized vs Uncharacterized

Feature 6 models whether or not the labeled part of the neuron's morphology allows for its categorization in the rest of the features. When an expert selects Uncharacterized for a given neuron, s/he cannot provide values for any of the other features. Considering the opposite case, assigning a category to any feature from F1 to F5 implies that the neuron is Characterized, i.e., enough morphological detail is available to allow its categorization. Additionally, Feature 4 is only available when Translaminar is selected in Feature 1 and Displaced is selected in Feature 3. Therefore, the number of values available for each feature can differ between neurons, depending on the values selected by the experts in other features.

In the analysis, we first applied statistical techniques on the experts' selections. These analyses were used to study the agreement among the experts at both the feature and category levels. Different machine learning techniques were then used to confirm the agreement results and extract knowledge from the data. Clustering algorithms were applied to find groups of easily distinguishable neurons and to identify their defining properties. We used a Bayesian network approach to model the statistical relationships between the features and to study the underlying reasoning of each expert. Finally, supervised classification algorithms were run to induce models that were able to automatically classify the neurons taking into account an exhaustive set of morphological measurements of their three-dimensional reconstructions.

Experts' agreement analysis

We analyzed the agreement achieved by the 42 experts who completed the experiment involving the classification of the 320 neurons according to the six features. First, we computed the overall inter-expert agreement, as well as the chance-corrected Fleiss' pi agreement index (Fleiss, 1971) for each feature independently. Fleiss' pi index adjusts the observed agreement by subtracting the agreements between experts that are due to chance alone. Also, we computed the inter-expert agreement and Fleiss' pi index for each category of every feature, taking into account all the experts' ratings (Cicchetti and Feinstein, 1990).

We studied the sensitivity of the agreement to the set of experts, i.e. whether or not removing a small number of experts would significantly modify the agreement values. We wanted to identify whether any experts showed different choice behavior from the rest of the group (outliers). Therefore, we computed Fleiss' pi index when 1, 2 or 3 of the experts were removed from the analysis.

Also, we analyzed possible overlaps between categories of Feature 5. Pairs and triples of interneuron types were merged into one single category, and Fleiss' pi index was computed for all of these scenarios. We identified where combining interneuron types increased or decreased the agreement. If experts frequently confused two or more interneuron types, the agreement value increased when considering those interneurons as the same category. This was for example the case for categories Common type and Common basket (**Supplementary Online Information S2**). On the contrary, merging easily distinguishable pairs or combinations of interneuron types yielded decrements in the

agreement. This was for example the case for categories Martinotti and Common basket (**Supplementary Online Information S2**).

Additionally, we studied the agreement between each possible pair of experts for every feature using three different indices: Cohen's kappa (Cohen, 1960), Prevalence-Adjusted Bias-Adjusted (PABA) kappa (Byrt et al., 1993), and the ratio between Cohen's kappa and its maximum value given fixed marginals (Dunn, 1989). These indices can only be applied to binary features, while Features 4 and 5 are non-binary. Thus, for Feature 5, we used these indices to measure the agreement between each category versus all the other categories considered together. This allowed us to study each category independently for each pair of experts. We proceeded similarly with Feature 4.

Agreement indices
For each of the six features, we ran a classification experiment, in which a group of R experts classified a set of M items (neurons) into Q categories. The goal was to measure and analyze the degree of agreement between experts when categorizing the items. We denote r_{iq} as the number of experts who assigned the ith neuron to category q:

$$r_i = \sum_{q=1}^{Q} r_{iq}.$$

We denote n_{jq} as the number of items that expert j assigned to category q.

Overall observed agreement
The most straightforward way to assess consensus is by computing the observed agreement:

$$P_o = \frac{\sum_{q=1}^{Q} \sum_{i=1}^{M} r_{iq}(r_{iq} - 1)}{\sum_{i=1}^{M} r_i(r_i - 1)}.$$

This overall observed agreement has been widely criticized in the literature (e.g., Carletta, 1996). The observed agreement favors experiments with a low number of categories, Q. In addition, it does not take into account the different distributions of items among categories. Two solutions have been proposed to this problem: 1) adjusting the observed agreement for chance agreement; and 2) computing category-specific observed agreement.

Chance-corrected agreement coefficients
A solution to the problem of analyzing the agreement between experts in a classification study is correcting the observed value to erase the influence of chance agreements. Popping (1988) identified more than 40 different proposals of chance-corrected agreement indices. In general, most of the chance-corrected agreement indices have the following expression:

$$A = \frac{A_o - A_e}{1 - A_e}, \tag{1}$$

where A_o is the observed agreement and A_e is the expected agreement by chance. The numerator encodes the observed agreement beyond chance, and the denominator encodes the maximum agreement that can be achieved beyond chance. An index value of $A = 1$ means perfect agreement, whereas a value of $A = 0$ shows chance agreement. Negative values of A indicate agreement below chance. In this study we applied the two most studied agreement indices: Cohen's kappa (and some of its variants) and Fleiss' pi indices.

Cohen's kappa: Cohen's kappa index (Cohen, 1960) is defined for a two-expert ($R = 2$) and two-category ($Q = 2$) experiment. Only those items classified by both experts are considered. The results of the experiment can be reported as a cross-classification table (**Table S1**):

Table S1. Cross-classification table for an experiment with two experts and two categories (+ and –).

		Expert 2		
		+	-	Frequency
Expert 1	+	a	b	n_{1+}
	-	c	d	n_{1-}
Frequency		n_{2+}	n_{2-}	M

Cohen's kappa index has the structure of Equation (1), with

$$A_o^\kappa = \frac{a+d}{M}$$

and

$$A_e^\kappa = \frac{n_{1+}n_{2+} + n_{1-}n_{2-}}{M^2}.$$

Cohen's kappa index is developed under three assumptions: 1) the classified items are independent, 2) the categories are independent, exhaustive, and mutually exclusive, and 3) the experts operate independently and have different distributions for the categories. Cohen's kappa index is negatively affected by the different prevalence of the categories (prevalence problem) and by the degree of disagreement between the two experts (bias problem) (e.g., Feinstein and Cicchetti, 1990). Interpreting the magnitude of Cohen's kappa index is challenging because of these effects. Several standards have been proposed for interpreting the strength of agreement (e.g., Landis and Koch, 1977). These approaches are necessarily subjective and arbitrary, since the interpretation of Cohen's index depends on the field of science, the nature of the experiment, and the prevalence and bias effects in the data (Artstein and Poesio, 2008). Some variants of Cohen's kappa index follow.

- **Prevalence-Adjusted Bias-Adjusted kappa index:** Byrt et al. (1993) propose a Prevalence-Adjusted Bias-Adjusted kappa index (PABAκ) to minimize the effects of prevalence and bias. This value can be reported alongside Cohen's kappa to show the effects of prevalence and bias on the index value and to determine the sources of disagreement. To compute PABAκ, the cross-classification table is modified as in **Table S2**. The agreement cells *a* and *d* (main diagonal in **Table S1**) are changed to their mean *(a+d)/2*, removing the prevalence effect. The disagreement cells *c* and *b* (secondary diagonal in **Table S1**) are also changed to their mean value *(b+c)/2*, adjusting for the bias effect. PABAκ is Cohen's kappa index computed with the values in the modified **Table S2**.

Table S2. Modified cross-classification table for minimizing prevalence and bias effects and computing PABAκ.

		Expert 2		
		+	-	Frequency
Expert 1	+	*(a+d)/2*	*(b+c)/2*	n'_{1+}
	-	*(b+c)/2*	*(a+d)/2*	n'_{1-}
Frequency		n'_{2+}	n'_{2-}	M

- **Maximum kappa index:** Another approach for interpreting Cohen's kappa value is comparing it to the maximum value κ_{max} that can be achieved when the marginal frequencies of each expert are fixed (Sim and Wright, 2005). To compute κ_{max}, a modified cross-classification table is used, where the agreement cells (main diagonal) are set to the minimum of the marginal frequencies for their corresponding categories, as in

Table S3. The disagreement cells (secondary diagonal) are adjusted to maintain the marginal frequencies. The value of κ_{max} shows the maximum possible agreement taking into account the different prevalence and bias of the experts. The ratio κ/κ_{max} is usually used to measure the proportion of agreement that was achieved in the experiment taking into account the differences between experts.

Table S3. Modified cross-classification table for computing κ_{max}.

		Expert 2		Frequency
		+	-	
Expert 1	+	$min(n_{1+}, n_{2+})$	$n_{1+} - min(n_{1+},n_{2+}) ??n??1+??-??min??(??n??1+??,??n??2+??)??$	n_{1+}
	-	$n_{2+} - min(n_{1+}, n_{2+})$	$min(n_{1-}, n_{2-})$	n_{1-}
Frequency		n_{2+}	n_{2-}	M

Fleiss' pi index: When more than two experts join the experiment (*R>2*), Fleiss' (1971) generalization of Scott's (1955) pi index is the most commonly used chance-corrected agreement coefficient. When missing values are allowed (not all the experts have to classify all the items), Fleiss' pi can be adapted to give equal weight to each judgment or equal weight to each item (Artstein and Poesio, 2008). Giving an equal weight to each item, Fleiss' pi follows the structure in Equation (1) with

$$A_o^\pi = \frac{1}{M} \sum_{i=1}^{M} \sum_{q=1}^{Q} \frac{r_{iq}(r_{iq}-1)}{r_i(r_i-1)},$$

and

$$A_e^\pi = \sum_{q=1}^{Q} \left(\frac{1}{M} \sum_{i=1}^{M} \frac{r_{iq}}{r_i} \right)^2.$$

Fleiss' pi assumes that the marginal distribution of the categories is the same for each expert given the assumption that they are operating by chance. This is the main difference with Cohen's kappa index, where it is assumed that the marginal distributions of the categories for each expert are different.

Category-specific agreement indices
We can also study inter-expert agreement (observed and chance-corrected Fleiss' pi index) for each category in each feature individually.

Observed agreement: We can compute specific observed agreement values for each category *q=1,…,Q*, using a similar approach as with specificity and sensitivity:

$$P_{oq} = \frac{\sum_{i=1}^{M} r_{iq}(r_{iq}-1)}{\sum_{i=1}^{M} r_{iq}(r_i-1)}.$$

Several authors advocate the use of these category-specific indices (e.g., Cicchetti and Feinstein, 1990). Reporting these category-specific indices overcomes the problems of the overall observed agreement and avoids the need to correct for chance agreement.

Chance-corrected Fleiss' pi index: A different chance-corrected agreement index can be computed for each category using Fleiss' pi index. The chance-corrected agreement for a category $q = 1,…,Q$ is given by Equation (1) with

$$A_o^{\pi_q} = \frac{\sum_{i=1}^{M} r_{iq}(r_{iq} - 1)}{\sum_{i=1}^{M} r_{iq}(r_i - 1)},$$

and

$$A_e^{\pi_q} = \frac{1}{M}\sum_{i=1}^{M} \frac{r_{iq}}{r_i}.$$

Statistical tests for chance agreement

We performed a permutation test to check whether or not the values of the agreement indices explained above indicated an agreement above chance. A random experiment was generated by sampling categories for each feature maintaining the relative frequency of the categories in the complete experiment (shown in **Figure 3A** of the main text). For each expert and each neuron, we sampled a category for Feature 6. If the sampled category was *Characterized*, then categories for Features 1-3 and 5 were sampled. When *Translaminar* and *Displaced* categories were sampled for Features 1 and 3, then a category was randomly sampled for Feature 4. Ten thousand random experiments were generated and the observed agreement, Fleiss' pi and category-specific Fleiss' pi indices were computed. Then, the agreement value using the real classification data provided by the experts was compared to the cumulative distribution of the values of the agreement indices obtained with the randomly generated experiments. Statistical significance was established at a significance level $\alpha = 0.05$.

Neuron clustering

We applied unsupervised classification (clustering) algorithms to find groups of interneurons with similar characteristics according to the classifications provided by the experts.

Neuron clustering for each feature

First, we wanted to generate clusters for the set of M neurons considering each feature independently. For a given feature, we used the category assigned for each expert to each neuron (category value $q = 1,…,Q$) as information for the clustering. Therefore, the dataset used for the clustering algorithm had 320 instances (neurons), where each instance is an N-dimensional vector ($N = 42$ experts).
We applied the k-modes algorithm (Huang, 1998), an extension of the k-means algorithm (MacQueen, 1967) that manages categorical data. **Algorithm 1** sketches the main steps of the k-means algorithm, which are the same as in the k-modes algorithm. The goal of the k-means algorithm is to find the k cluster centers $C = \{c_1,…,c_k\}$ that minimize a measure of dissimilarity, where $k>1$ is a parameter of the algorithm indicating the number of clusters. For Features 1-3 and Feature 6 a number of clusters $k = 2$ was used. For Feature 4, three clusters ($k = 3$) were selected. Different numbers of clusters (six to ten) were analyzed for Feature 5. The clearest results were obtained with $k=8$. A neuron is assigned to the cluster with the closest center. Therefore, the fitness function to minimize is the sum of the distance of each item to the center of its cluster.
For categorical data, k-modes uses the Hamming distance to measure the distance between two items (neurons) or between an item and a cluster center. The set of cluster centers C is found by computing the modes of the items belonging to the cluster. Ties when computing the modes or when assigning items to clusters are broken randomly. In our implementation, the algorithm stopped when no change in the cluster centers occurred or when the fitness function had the same value for 100 consecutive iterations.

Algorithm 1. k-means clustering algorithm.
Input:
- k, number of clusters.
- Dataset of N-dimensional items x_i, $i = 1,…, M$.

Steps:

1. Initialize the *k* cluster centers *C* to *k* random items $x_{(1)}$, ..., $x_{(k)}$.
2. While cluster centers *C* change
   a. Assign each item x_i to the corresponding cluster with the closest center.
   b. Recompute *C* from the items in the cluster.
3. Return *C*.

The *k*-means algorithm (and also *k*-modes) can yield suboptimal solutions if it gets stuck in local minima. To avoid this situation, the algorithm was run 25 times with different initial values for the cluster centers in step 1 of **Algorithm 1**. The best result (minimum fitness function value) is reported in the Results section. The algorithms were implemented and run in Matlab.

Neuron clustering for all the features

We wanted to generate clusters of cells taking into account the agreement of the experts in all the features at the same time. For every neuron, we computed the number of experts that assigned the neuron to each category of every feature. Therefore, the dataset used in the clustering algorithm had 320 instances (neurons), and each instance was an *N*-dimensional vector ($N = 21$), corresponding to all the categories of the six features: Intralaminar, Translaminar, Intracolumnar, Transcolumnar, Centered, Displaced, Ascending, Descending, Both, Common type, Horse-tail, Chandelier, Martinotti, Common basket, Arcade, Large basket, Cajal-Retzius, Neurogliaform, Other, Characterized, and Uncharacterized.

We used the *k*-means algorithm to cluster cells according to the number of votes each neuron had in each category. Different numbers of clusters (six to ten) were analyzed. The clearest results were obtained with *k*=6. For continuous data, *k*-means uses Euclidean distance to compute the distance between every two items. Every time step 2b in **Algorithm 1** is performed, *k*-means computes the cluster centers as the centroid of the items in the cluster.

The algorithm was run 25 times with different initial values for the cluster centers to avoid local optima, similarly to *k*-modes, and the best result was shown. The clusters were illustrated using parallel coordinate diagrams (Wegman, 1990). Each line represents one neuron in the cluster and its height shows the number of experts who selected each category for that neuron. A small amount of noise drawn from a normal distribution (mean = 0, standard deviation = 0.75) was added to the values to ensure that all lines were visible.

Bayesian networks for modeling experts' opinions

We trained one Bayesian network (Pearl, 1988) on data from each expert, modeling the statistical relationships between the features. A Bayesian network is a kind of probabilistic graphical model that encodes a factorization of the joint probability distribution of the features (also called variables) in a given domain. Bayesian networks compactly represent the problem domain and can perform any kind of reasoning (causal, diagnostic, abductive, bidirectional, etc.) efficiently because of the local computations allowed by the probability factorization.

Formally, a Bayesian network can be defined as a pair $B = <G(X, A), P>$ with two main components:

- The graphical part $G(X, A)$ is a directed acyclic graph (DAG) used to capture the structure of the problem. The set of nodes (*X*) represents the variables, $X = (X_1, ..., X_n)$, included in the problem domain. The set *A* contains the directed edges (called arcs) connecting the nodes. In a DAG, the set of arcs cannot include a directed cycle. The probabilistic conditional (in)dependence relationships between the variables in the domain are codified in the set of arcs (*A*).
- The probabilistic component *P* includes the conditional probability distributions $P(X_i \mid Pa(X_i))$ associated with the variables X_i, $i=1,...,n$. For each variable X_i, we define the set of its parents as the set of variables with an arc going to X_i: $Pa(X_i) = \{Y \in X \mid (Y, X_i) \in A\}$.

A Bayesian network encodes a factorization of the joint probability distribution over all the variables in *X*:

$$P(X) = \prod_{i=1}^{n} P(X_i | Pa(X_i)).$$

Here, each feature in the experiment was modeled as a discrete variable in the Bayesian network, i.e., each Bayesian network contained six nodes. In the variables representing F1 to F5, we included one discrete value named "Missing". This value models the scenarios where a category was not provided,

either because Uncharacterized was selected, or because Translaminar and Displaced were not selected (for Feature 4).

We trained the Bayesian networks from the data using the GeNIe free modeling environment[1]. The greedy thick thinning algorithm (Dash and Cooper, 2004) with K2 scoring function (Cooper and Herskovits, 1992) was used to train the Bayesian network structure. K2 score function measures the joint probability of the Bayesian network G and a dataset D:

$$P(G, D) = P(G) \prod_{i=1}^{n} \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}!,$$

where $P(G)$ is the prior probability of the network G, r_i is the number of values of X_i, q_i is the number of possible configurations of $Pa(X_i)$, N_{ijk} is the number of instances in the dataset D where the variable X_i takes the k-th value x_{ik} and the set of parents $Pa(X_i)$ takes their j-th configuration, and *??*
$N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$.

The greedy thick thinning algorithm starts with an empty graph and iteratively adds the arc (without creating a cycle) that yields the maximum increase in the marginal likelihood. When no increment is possible, the algorithm iteratively removes arcs until no arc deletion yields a positive increase in the marginal likelihood. Then, the algorithm stops and returns the resulting Bayesian network structure. We did not allow any feature to be a parent of the variable corresponding to Feature 6. This restriction encodes the knowledge that categorizing a neuron as Uncharacterized disabled the rest of the features for categorization. Once the network structure was found, the maximum likelihood estimators of the parameters in the conditional probability tables of each node were computed from the counts in the data.

We analyzed the graphical structures and made inferences with the Bayesian networks to compare the underlying reasoning of different experts. We used the Bayesian networks to study differences in the experts' behavior. A preliminary analysis of the structures of the Bayesian networks consisted of counting the number of network structures where a given edge (arc connecting two nodes without considering its direction) appeared. Very frequent edges highlight common relationships and properties in a large number of experts.

Also, GeNIe software was used to perform inference on the Bayesian networks. We set some categories as selection of choice ("evidence") and used the exact inference algorithm based on message passing (Lauritzen and Spiegelhalter, 1988; Jensen et al., 1990) to update the probabilities of the variables in the Bayesian networks. We then analyzed different scenarios for a subset of Bayesian networks with either the same or different network structures. Similarities and differences between the experts' reasoning were identified by comparing the updated probabilities in the Bayesian networks.

Supervised classification of neurons
In order to build models that automatically identify each of the features studied in this work based on a set of quantitative morphological parameters, we selected the 241 neurons whose 3D reconstructions were available at NeuroMorpho.Org. The MicroBrightField Neurolucida package was used to perform the branched structure, convex hull, Sholl, fractal, fan-in diagram, vertex, and branch angle analyses. These analyses were conducted for the complete neuronal morphology as well as separately for the dendritic and axonal arbors. These analyses yielded a set of 2,886 morphological measures of each neuron, including:

- General information about the dendrites and the axons, e.g., the number of endings, the number of nodes (branching points), the total length and the mean length of each dendritic arbor.
- Morphometric measures of the soma such as the area, aspect ratio, compactness, convexity, contour size (maximum and minimum feret), form factor, perimeter, roundness and solidity.
- The total, mean, median and standard deviation of the length of the segments belonging to dendritic arbors and axons independently. Also, we performed these analyses dividing the segments by their centrifugal order from the soma.
- Number of nodes and segments of the complete dendrites and axons, and number of nodes and segments measured by centrifugal order.

---

[1] Developed by the Decision Systems Laboratory of the University of Pittsburgh: http://dsl.sis.pitt.edu.

- Convex hull analysis. We performed 2D and 3D convex hull analysis of the dendrites and the axon independently to obtain measures of the area, perimeter, volume and surface of the neuronal morphology.
- Sholl analysis. We computed the number of intersections in concentric spheres centered at the soma with increasing radii of 20 μm. We also used the number of endings, nodes and the total length of the segments included in those spheres.
- Fractal analysis. We computed the fractal dimension for the dendrites and the axon independently using the box-counting method (Mandelbrot, 1982). The fractal dimension is a quantity that indicates how completely the neuron fills space.
- Vertex analysis of the connectivity of the nodes in the branches to describe the topological and metric properties of the arbors. We used the number of nodes of each one of the three types: Va (branching points where the two child segments end), Vb (branching points where one of the child segments end) and Vc (branching points where the two child segments bifurcate). We also used the ratio Va/Vb and computed the number of nodes of each type by centrifugal order.
- Branch angle analysis. We used planar, local and spline angles that measure the direction of the branches at different levels. We computed the mean, standard deviation, and median of the three angles for dendrites and axon individually. Additionally, we computed the mean, standard deviation, and median of the angles of the segments grouped by centrifugal order.

Many variables were measured according to the centrifugal order of the segments they belonged to. Since neurons have different maximum centrifugal order, length, etc., each neuron had a different number of computable variables. For example, one neuron might have dendrites with a maximum centrifugal order of 9 and another neuron could have dendrites with a maximum centrifugal order of 5. Variables that measured neuron morphology at orders 6, 7, 8 and 9 were not computable in the second neuron, so we set those values to 0 to be manageable by the algorithms. Variables concerning the complete neuron morphology are not affected by this issue, since they were obtained from the data directly coming from the 3D reconstructions. For each one of the features in the experiment, we had to assign a single "true category" to each neuron. We used the most frequently occurring value in the 42 assignments made by the experts who completed the experiment, i.e., we applied a simple majority vote to assign a "true category" to each neuron for each feature. Using this approach, there were no neurons categorized as Arcade, Cajal-Retzius or Other by the majority of the experts.

The accuracy of the classifiers was estimated using the leave-one-out technique (Mosteller and Tukey, 1968). The following 10 different classification algorithms available in Weka software were applied using their default parameters (Witten and Frank, 2005).

- NB: Naïve Bayes classifier, where the conditional distributions of the continuous variables given the class values are modeled using Gaussian distributions (Pérez et al., 2006).
- NBdis: Discrete naïve Bayes classifier (Minsky, 1961). The continuous variables are discretized using a supervised discretization technique (Fayyad and Irani, 1993).
- RBFN: Neural network for classification tasks with one single hidden layer that uses Gaussian radial basis functions as activation functions (Bishop, 1995).
- SMO: Support vector machine with polynomial kernels implementing the sequential minimal optimization algorithm (Platt, 1998; Keerthi et al., 2001).
- IB1: Nearest neighbor classifier (Aha et al., 1991).
- IB3: Nearest neighbor classifier using 3 neighbors.
- JRip: Rule induction technique using RIPPER algorithm (Cohen, 1995).
- J48: Classification tree using C4.5 algorithm (Quinlan, 1993).
- RForest: Classification technique using a set of random tree classifiers (Breiman, 2001).
- RTree: Classification tree that chooses the variables at each node randomly.

Additionally, two variable selection methods were studied:

- Gain Ratio: A univariate filter algorithm that ranks the predictive variables according to their Gain Ratio with the class label and keeps the best 500 variables.
- CfsSubsetEvaluation: This algorithm tries to find a subset of predictive variables that is highly correlated with the class, but has low intercorrelation between the predictive variables. It starts with an empty subset and iteratively adds the variable that yields a subset with the highest correlation value. The correlation measures the symmetric uncertainty of each variable in the subset with the class (to maximize), and adjusts it to take into account the symmetric uncertainty between the predictive variables (to minimize). The symmetric uncertainty is a measure of correlation based on the marginal entropies and the joint entropies between pairs of variables (Hall, 1999).

These classification algorithms were applied in three different settings:

- **Classifiers for each feature independently:** Each one of the features in the experiment was considered independently. The number of class values was the same as the number of categories in the features, i.e., two class values for Feature 1, Feature 2, Feature 3, and Feature 6; and three class values for Feature 4. There were no neurons classified as Arcade, Cajal-Retzius or Other, so the classifiers for Feature 5 had 7 class values.

- **Binary classifiers for each category in Feature 5:** We induced a binary classifier (with two class values) to identify each category in Feature 5 versus all the other categories merged together. Neurons classified as Chandelier (3 neurons) or Neurogliaform (4 neurons) were very rare. Therefore, we did not induce binary classifiers for these two categories, because the class values would be too unbalanced for the classifiers to find the characterizing properties of these interneuron types.

- **Classifiers merging interneuron types:** Following the agreement results observed in the previous analyses, we decided to check whether the classification algorithms performed better when interneuron types that are difficult to distinguish were merged into one category. Therefore, we trained classifiers after having merged the categories corresponding to Common type, Common basket and Large basket into a single category, as these three interneuron types were frequently confused with each other. The rest of the categories were considered individually.

We performed an exact binomial test to test the hypothesis that the number of correctly classified neurons is greater than that expected with a base classifier always assigning the class with maximum prior probability. To estimate the number of correctly classified neurons, we multiplied the accuracy reported by the leave-one-out technique by 241. The null hypothesis is that the number of correctly classified neurons matches 241 times the maximum prior probability. The alternative hypothesis is that the number of correctly classified neurons is higher than 241 times the maximum prior probability. Statistical significance was established when the p-values were smaller than the significance level $\alpha = 0.05$.

References

Aha, D.W., Kibler, D., Albert, M.K. (1991) Instance-based learning algorithms. Machine Learning 6: 37—66.

Artstein, R. and Poesio, M. (2008) Inter-coder agreement for computational linguistics. Computational Linguistics 34(4): 555—596.

Ascoli, G.A., Donohue, D.E. and Halavi, M. (2007) Neuromorpho.org: A central resource for neuronal morphologies. Journal of Neuroscience 27(35): 9247—9251.

Bishop, C.M. (1995) Neural Networks for Pattern Recognition. New York: Oxford University Press.

Breiman, L. (2001) Random forests. Machine Learning 45(1): 5—32.

Byrt, T., Bishop, J., Carlin, J.B. (1993) Bias, prevalence and kappa. Journal of Clinical Epidemiology 46(5):423—429.

Carletta, J. (1996) Assessing agreement on classification tasks: The kappa statistic. Computational Linguistics 22(2): 249—254.

Cicchetti, D.V. and Feinstein, A.R. (1990) High agreement but low kappa: II. Resolving the paradoxes. Journal of Clinical Epidemiology 43(6): 551—558.

Cohen, J. (1960) A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20(1): 37—46.

Cohen, W.W. (1995) Fast effective rule induction. In: Prieditis, A., Russell, S.J. (Eds.) Proceedings of the Twelfth International Conference on Machine Learning, pp. 115—123. Morgan Kaufmann.

Cooper, G.F. and Herskovits, E. (1992) A Bayesian method for the induction of probabilistic networks from data. Machine Learning 9: 309—347.

Dash, D. and Cooper, G.F. (2004) Model averaging for prediction with discrete Bayesian networks. Journal of Machine Learning Research 5: 1177—1203.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). Journal of the Royal Statistical Society. Series B (Methodological) 39: 1—38.

Dunn, G. (1989) Design and Analysis of Reliability Studies: The Statistical Evaluation of Measurement Errors. Edward Arnold.

Fayyad, U.M. and Irani, K.B. (1993) Multi-interval discretization of continuous-valued attributes for classification learning. In: Bajcsy, R. (Ed.) Proceedings of the 13th International Joint Conference on Artificial Intelligence, pp. 1022—1027. Morgan Kaufmann.

Feinstein, A.R. and Cicchetti, D.V. (1990) High agreement but low kappa: I. The problems of two paradoxes. Journal of Clinical Epidemiology 43(6): 543—549.

Fleiss, J.L. (1971) Measuring nominal scale agreement among many experts. Psychological Bulletin 76(5): 378—382.

Hall, M. (1999) Correlation-based Feature Selection for Machine Learning. PhD Thesis, University of Waikato.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H. (2009) The WEKA data mining software: An update. SIGKDD Explorations 11(1).

Huang, Z. (1998) Extensions to the *k*-means algorithm for clustering large data sets with categorical values. Data Mining and Knowledge Discovery 2:283—304.

Jensen, F.V., Lauritzen, S.L. and Olesen, K.G. (1990) Bayesian updating in causal probabilistic networks by local computations. Computational Statistics Quarterly 4: 269–282.

Keerthi, S.S., Shevade, S.K., Bhattacharyya, C. and Murthy, K.R.K. (2001) Improvements to Platt's SMO algorithm for SVM classifier design. Neural Computation 13(3): 637—649.

Landis, J.R. and Koch, G.G. (1977) The measurement of observer agreement for categorical data. Biometrics 33(1): 159—174.

Lauritzen, S.L. and Spiegelhalter, D.J. (1988) Local computations with probabilities on graphical structures and their application to expert systems. Journal of the Royal Statistical Society, Series B, Statistical Methodology 50(2): 157–224.

MacQueen, J.B. (1967) Some methods for classification and analysis of multivariate observations. Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, pp. 281–297.

Mandelbrot, B. (1982) The Fractal Geometry of Nature. Freeman.

Minsky, M. (1961) Steps toward artificial intelligence. Proceedings of the Institute of Radio Engineers 49: 8—30.

Mosteller, F. and Tukey, J.W. (1968) Data analysis, including statistics. In: Lindzey, G. and Aronson, E. (Eds.) Handbook of Social Psychology, volume 2, pp. 80—203. Addison-Wesley.

Pearl, J. (1988) Probabilistic Reasoning in Intelligent Systems. Morgan Kaufmann.

Pérez, A., Larrañaga, P. and Inza, I. (2006) Supervised classification with conditional Gaussian networks: Increasing the structure complexity from naïve Bayes. International Journal of Approximate Reasoning 43: 1—25.

Platt, J.C. (1998) Fast training of support vector machines using sequential minimal optimization. In: Schölkopf, B., Burges, C. and Smola, A. (Eds.) Advances in Kernel Methods: Support Vector Learning, pp. 185—205. Cambridge: The MIT Press.

Popping, R. (1988) On agreement indices for nominal data. Sociometric Research: Volume 1, Data Collection and Scaling, pp. 90—105. St. Martin's Press.

Quinlan, R. (1993) C4.5: Programs for Machine Learning. San Mateo: Morgan Kaufmann.

Scott, W.A. (1955) Reliability of content analysis: The case of nominal scale coding. Public Opinion Quarterly 19(3): 321—325.

Sim, J. and Wright, C.C. (2005) The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. Physical Therapy 85(3): 257—268.

Wegman, E.J. (1990) Hyperdimensional data analysis using parallel coordinates. Journal of the American Statistical Association 85(411): 664—675.

Witten, I.H. and Frank, E. (2005) Data Mining. Practical Machine Learning Tools and Techniques, 2nd edition. Morgan Kaufmann.

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).
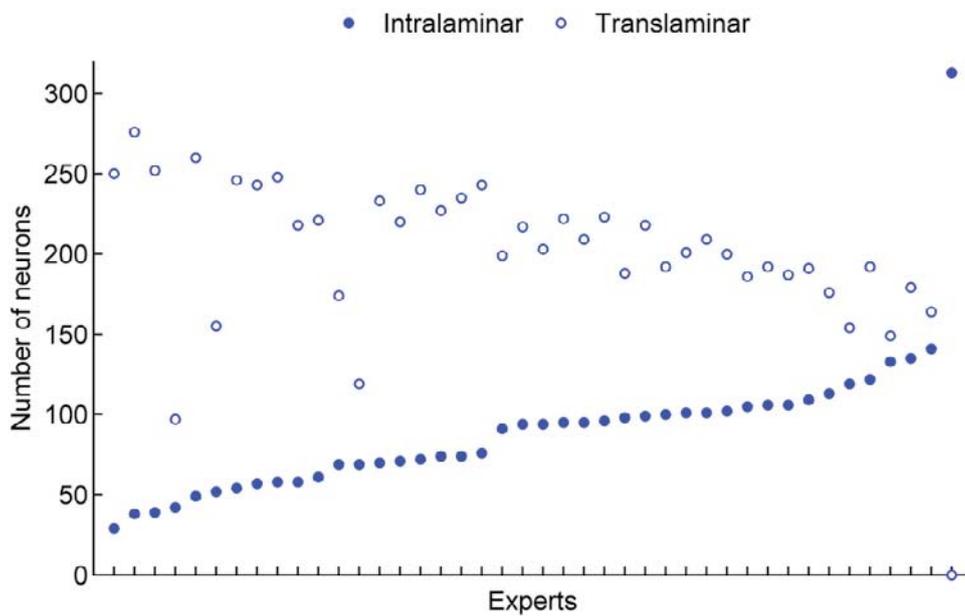

Fig. S1. Graphical representation of the ratings given to the different categories of Feature 1 by the 42 experts who completed the experiment. Experts are sorted in ascending order (in the x axis) based on the number of votes of the category Intralaminar.
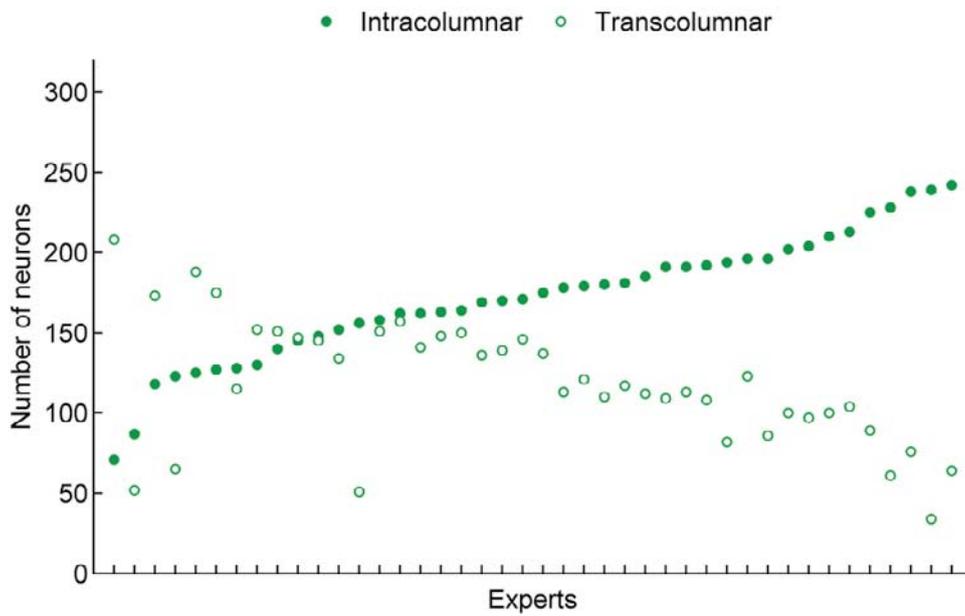
Fig. S2. Graphical representation of the ratings given to the different categories of Feature 2 by the 42 experts who completed the experiment. Experts are sorted in ascending order (in the x axis) based on the number of votes of the category Intracolumnar.
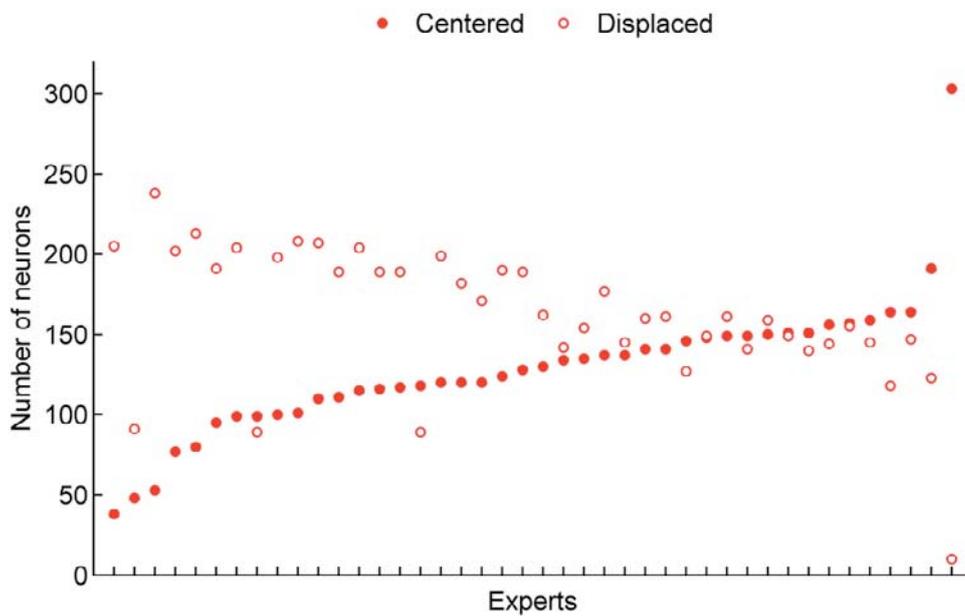

Fig. S3. Graphical representation of the ratings given to the different categories of Feature 3 by the 42 experts who completed the experiment. Experts are sorted in ascending order (in the x axis) based on the number of votes of the category Centered.
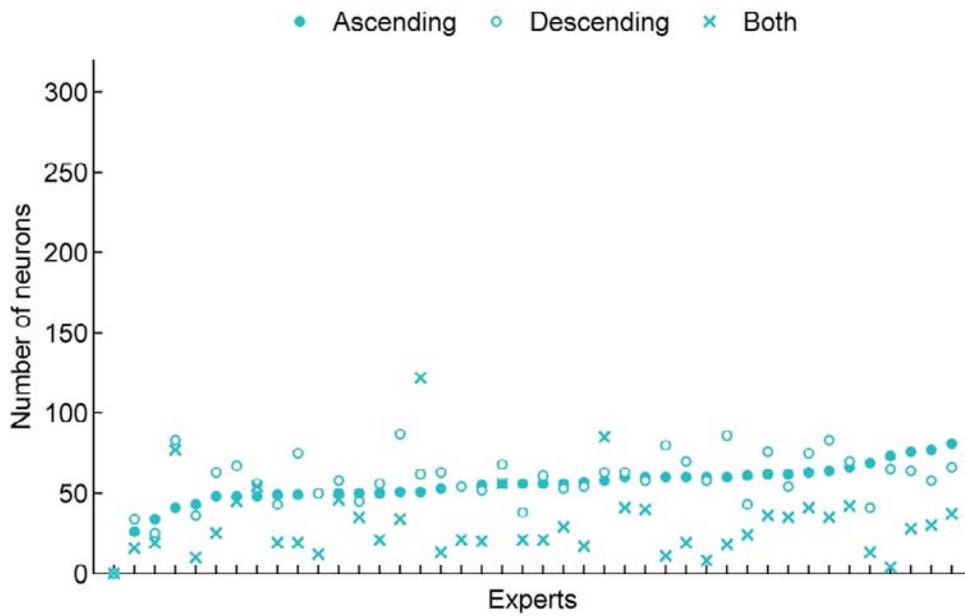
Fig. S4. Graphical representation of the ratings given to the different categories of Feature 4 by the 42 experts who completed the experiment. Experts are sorted in ascending order (in the x axis) based on the number of votes of the category Ascending.
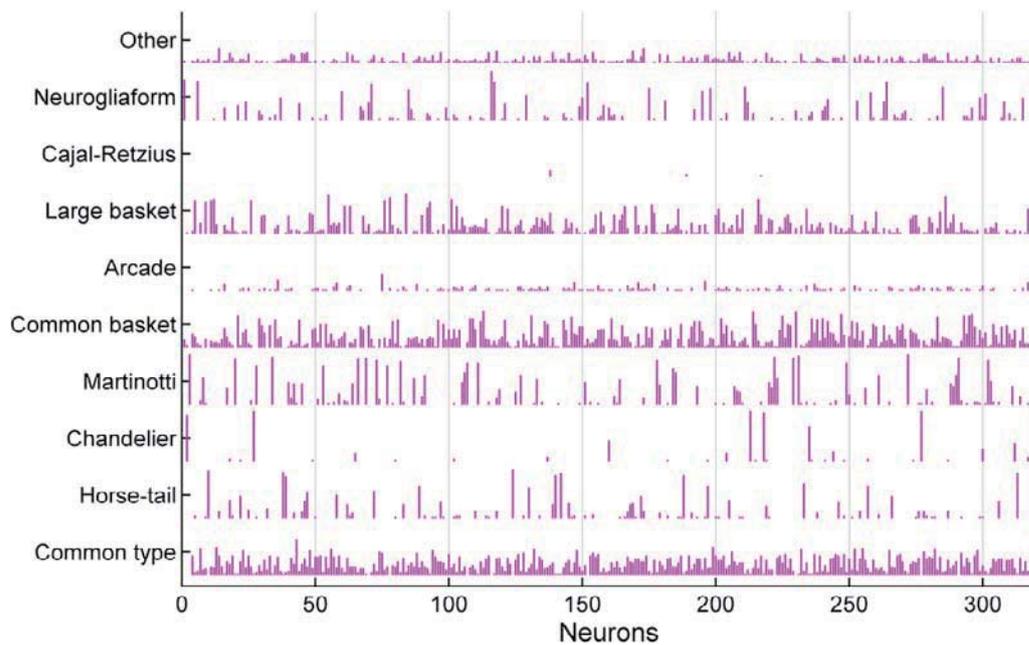

Fig. S5. Graphical representation of the number of experts who selected each category of Feature 5. A vertical bar is shown for each neuron and each category, representing the number of experts who selected that category for that neuron. High bars (e.g., for categories Chandelier, Horse-tail and Martinotti) 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.
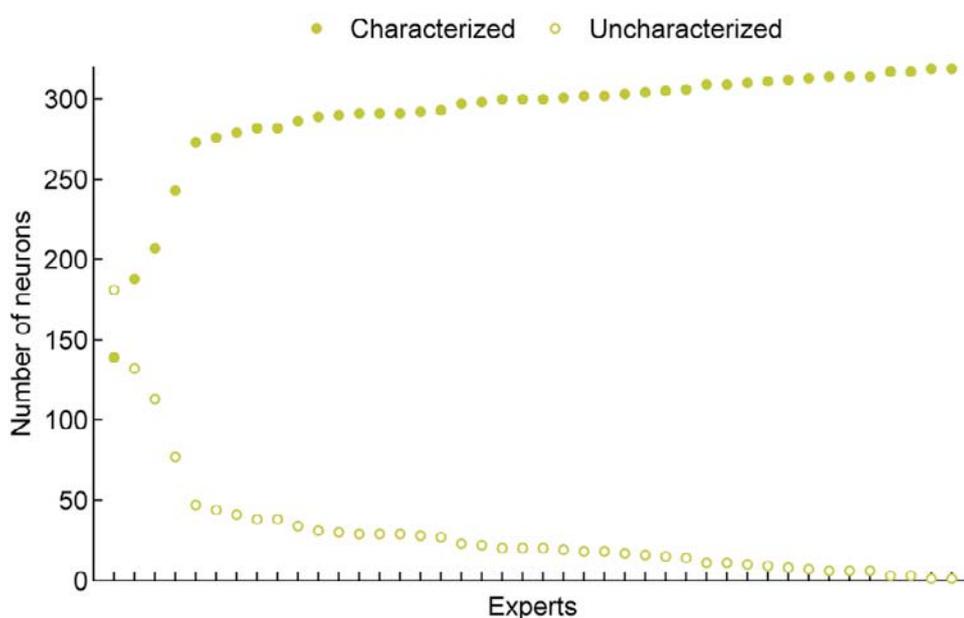
Fig. S6. Graphical representation of the ratings given to the different categories of Feature 6 by the 42 experts who completed the experiment. Experts are sorted in ascending order (in the x axis) based on the number of votes of the category Characterized.

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 $p < 0.0001$).
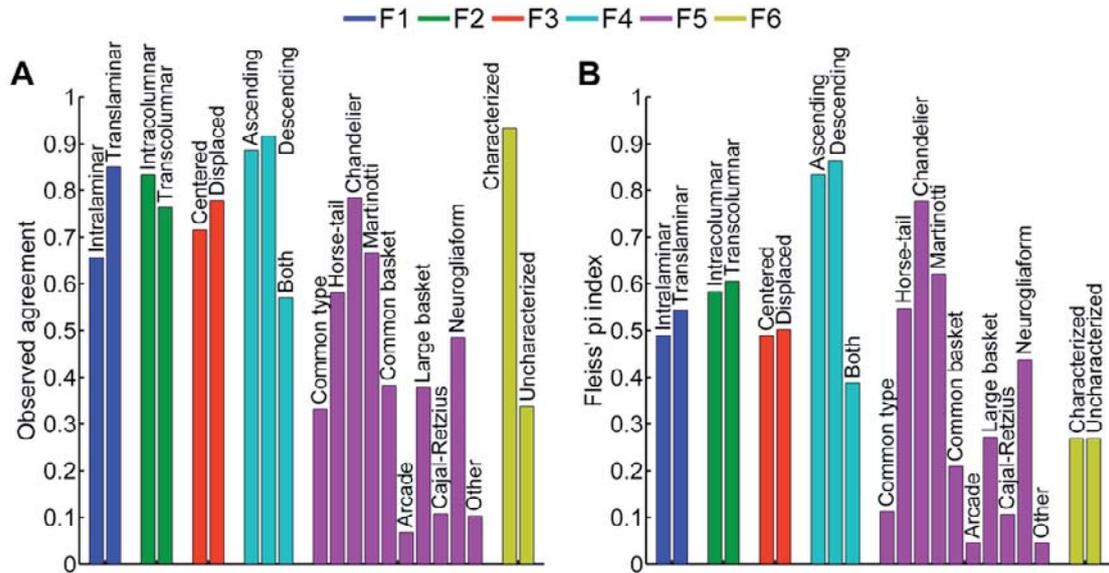
Fig. S7. (**A**) Observed agreement and (**B**) chance-corrected Fleiss' pi index for each category of every feature.

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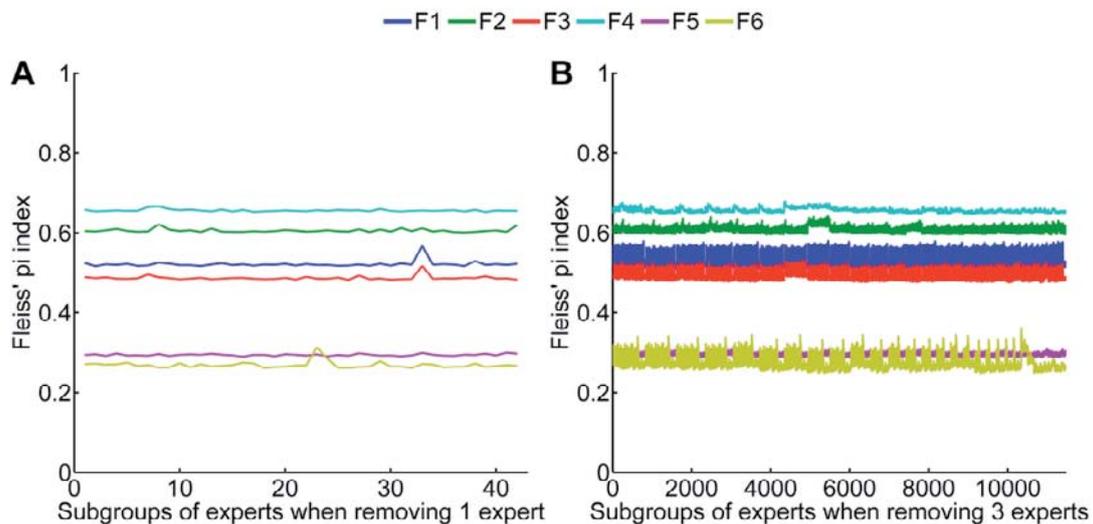
Fig. S8 (**A**) Fleiss' pi values for all the groups of experts obtained when removing one expert (42 possible subgroups) (**B**) and when three experts were removed (11,480 possible subgroups).

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 S4** 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 S4** 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 S4. 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
Neurogliaform									0.2952

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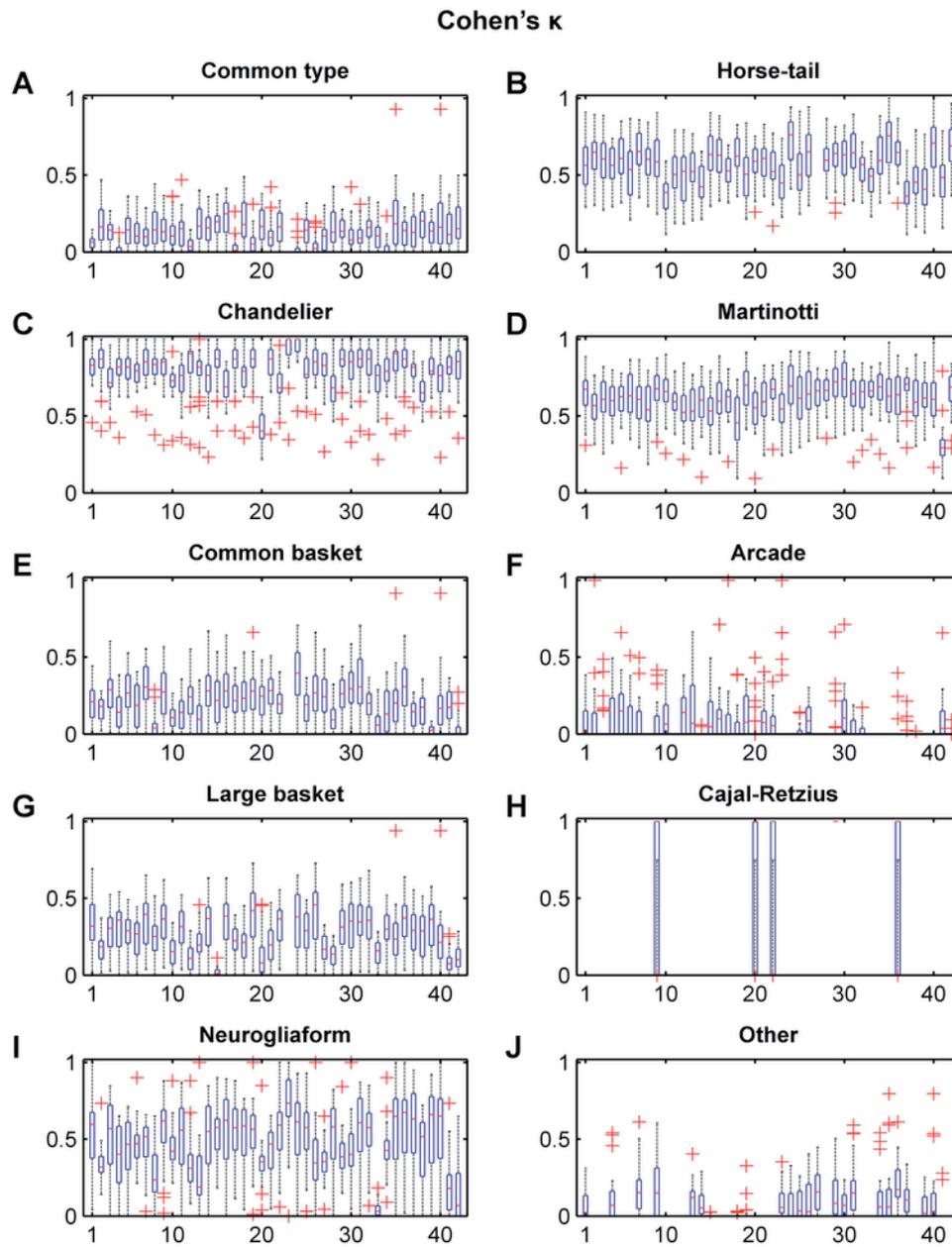
Fig. S9. Boxplots showing Cohen's kappa values for each pair of experts when comparing one category against all other categories in Feature 5. For example, the first box in panel A shows the agreement between the first expert (X-axis) and the rest of the experts for category Common type.

κ/κ_max ratio

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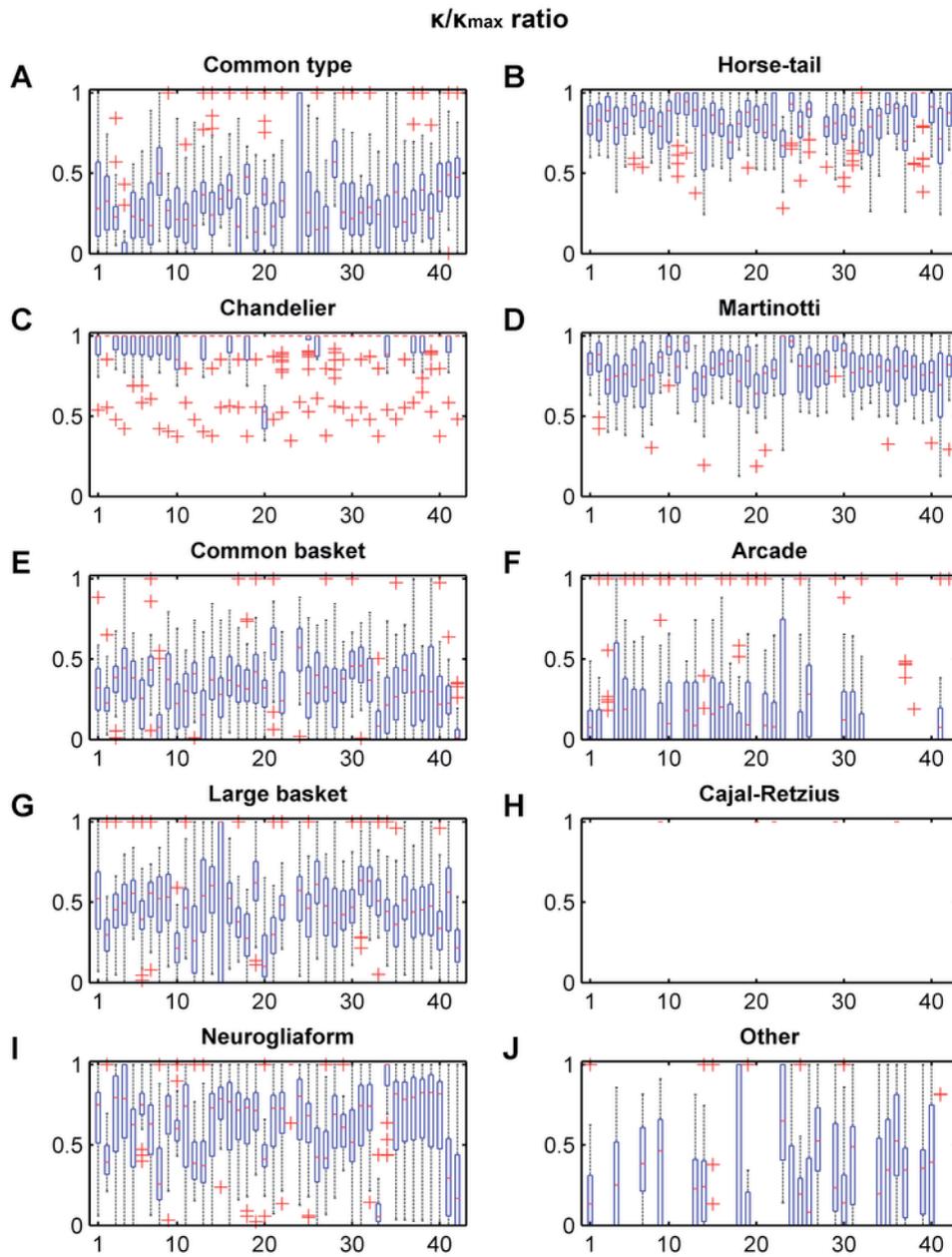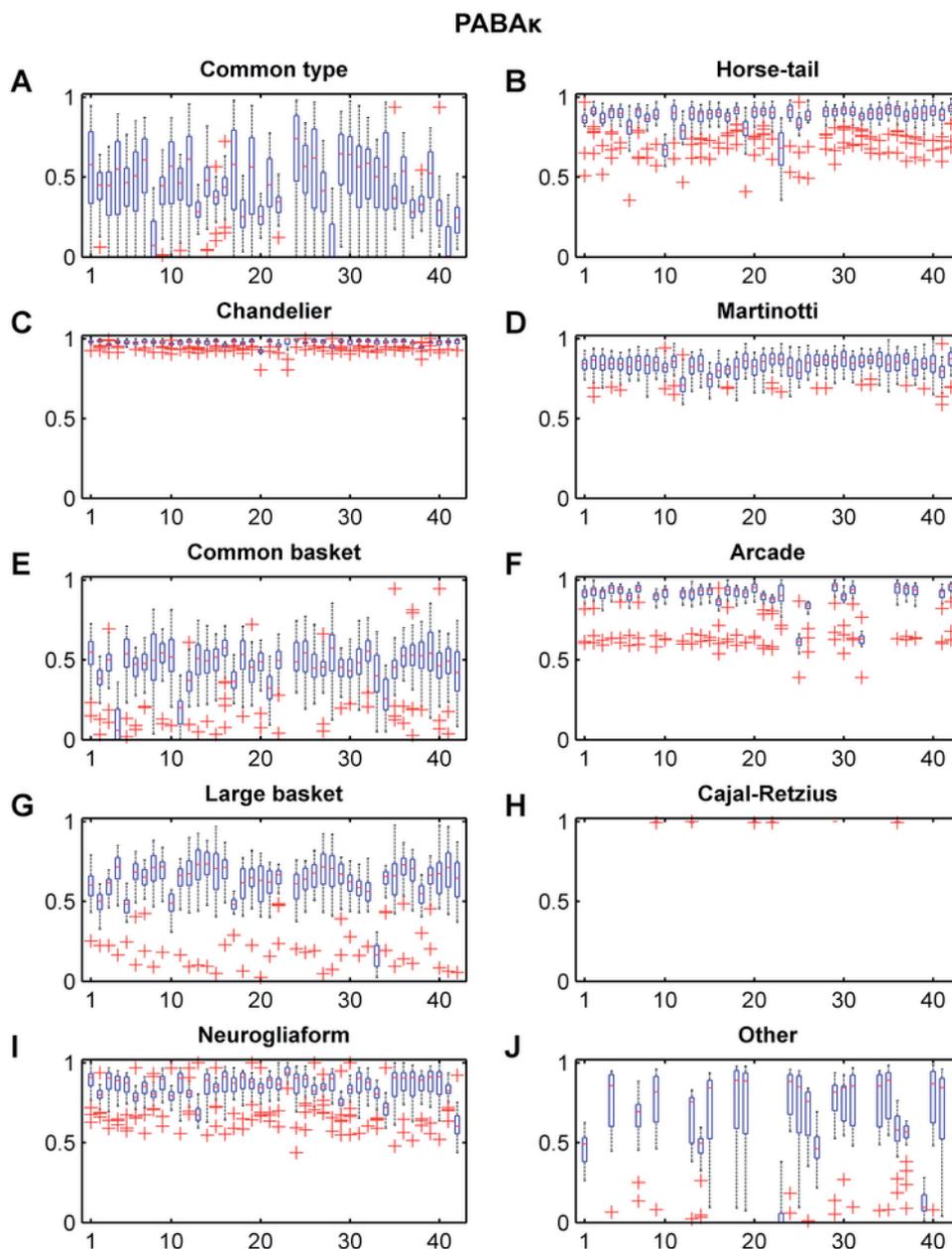
Fig. S11. Boxplots showing the Prevalence-Adjusted Bias-Adjusted kappa (PABA$_k$) values for each pair of experts when comparing one category against all the other categories in Feature 5.

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. 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.
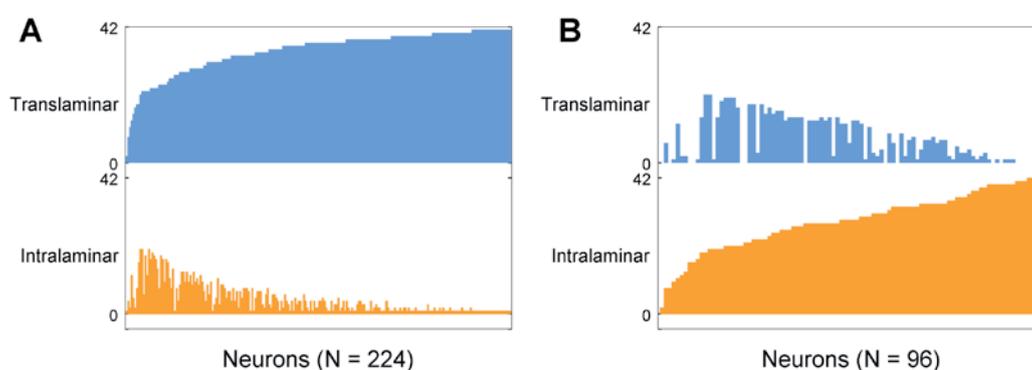

Fig. S15. Clusters of neurons obtained with the *k*-modes algorithm (*k* = 3) for Feature 4. Vertical bars show the number of experts who selected each category for each neuron in the cluster. Neurons have been sorted in ascending order by the number of votes for clarity. Panels **A** and **B** correspond to neurons mainly categorized as Ascending and Descending, respectively. Panel **C** shows neurons where different experts disagreed, categorizing them as Ascending, Descending or Both.
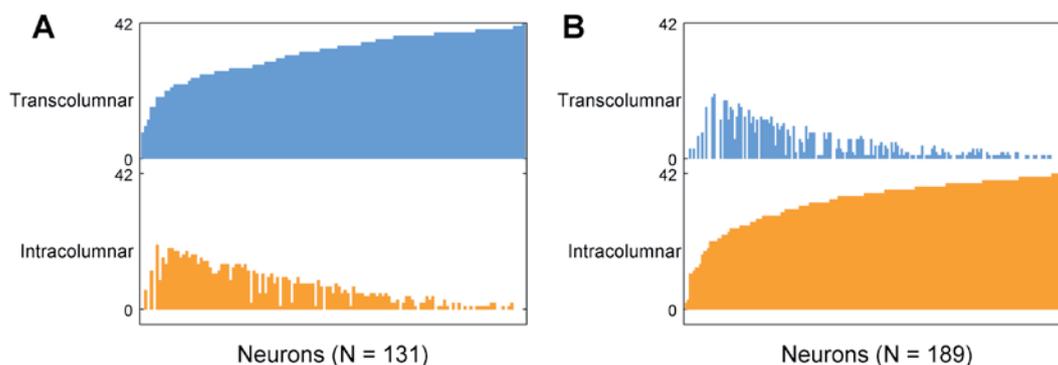
Fig. S16. Clusters of neurons obtained with the *k*-modes algorithm (*k* = 8) for Feature 5. Vertical bars show the number of experts who selected each category for each neuron in the cluster. Neurons have been sorted in ascending order by the number of votes for clarity. Panels **A** and **F** show 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.
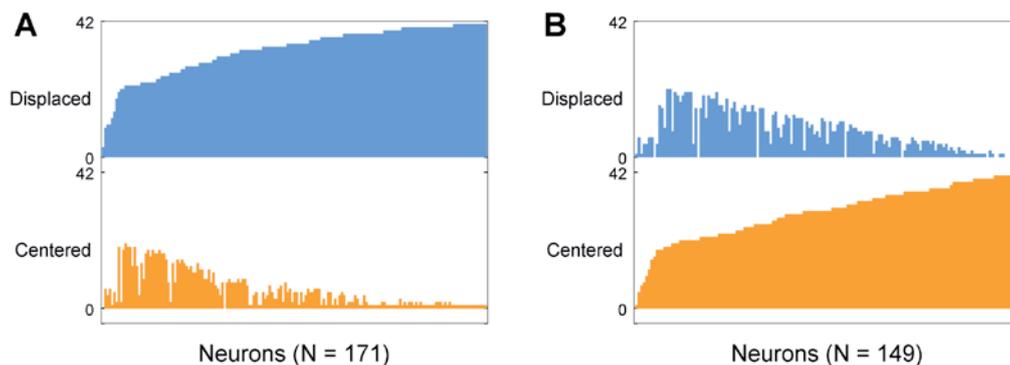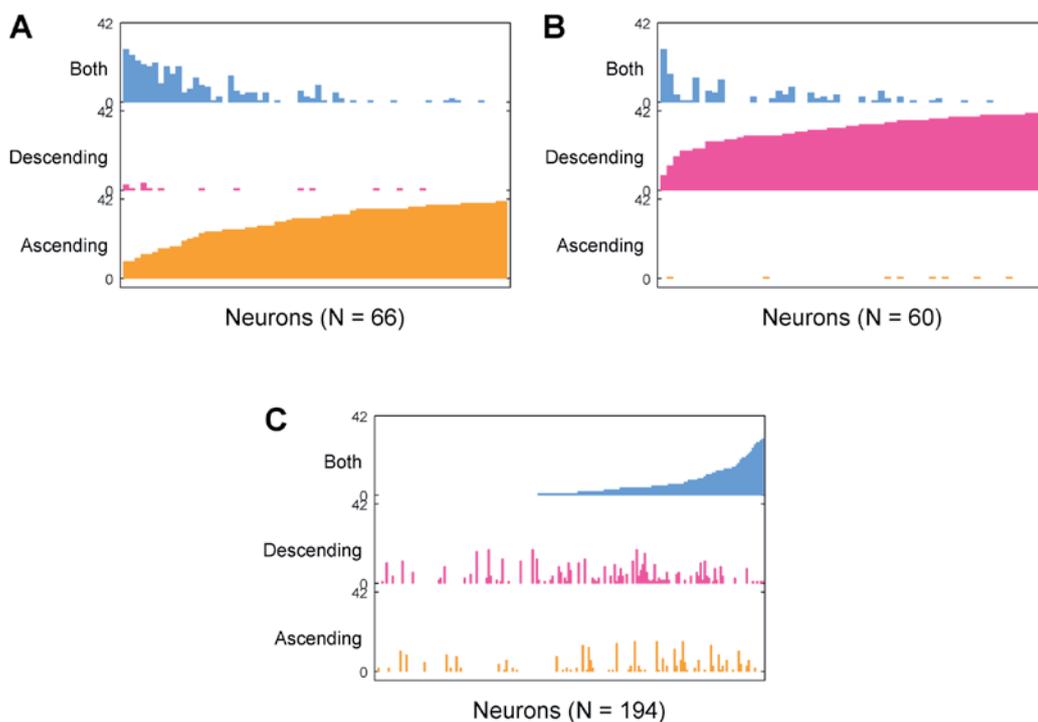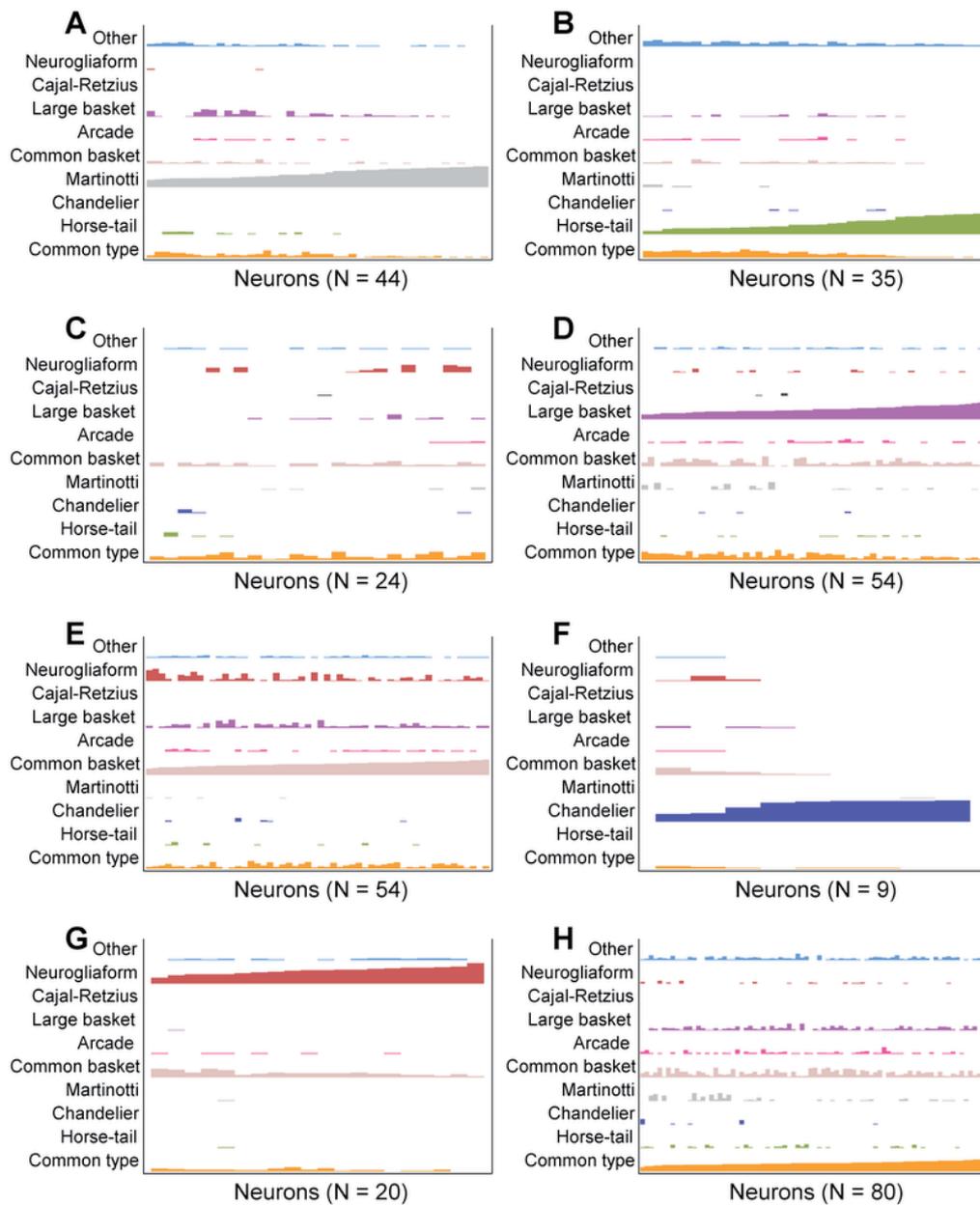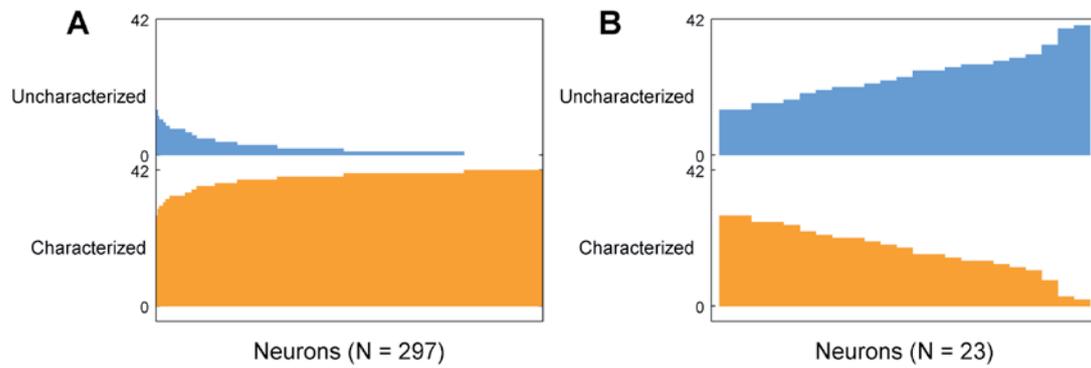
Fig. S17. Clusters of neurons obtained with the *k*-modes algorithm (*k* = 2) for Feature 6. Vertical bars show the number of experts who selected each category for each neuron in the cluster. Neurons have been sorted in ascending order by the number of votes for clarity. Panel **A** contains neurons 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.
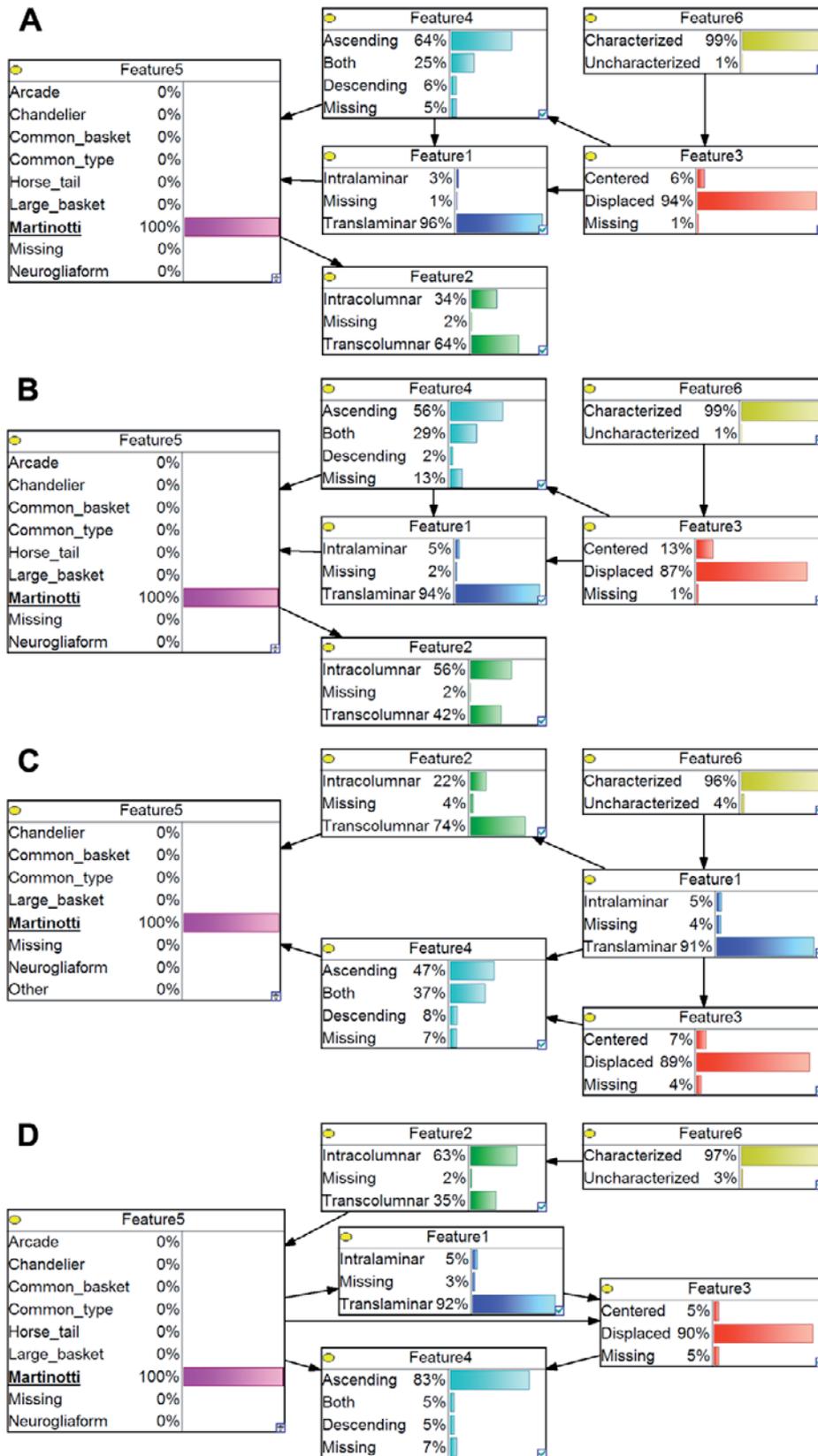
Fig. S18. Bayesian networks for experts 16 (**A**), 17 (**B**), 27 (**C**) and 32 (**D**). Martinotti was selected in Feature 5 and the probabilities were propagated through the Bayesian networks. Bar charts show the propagated probabilities of the remaining features conditioned on the Martinotti category.

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.

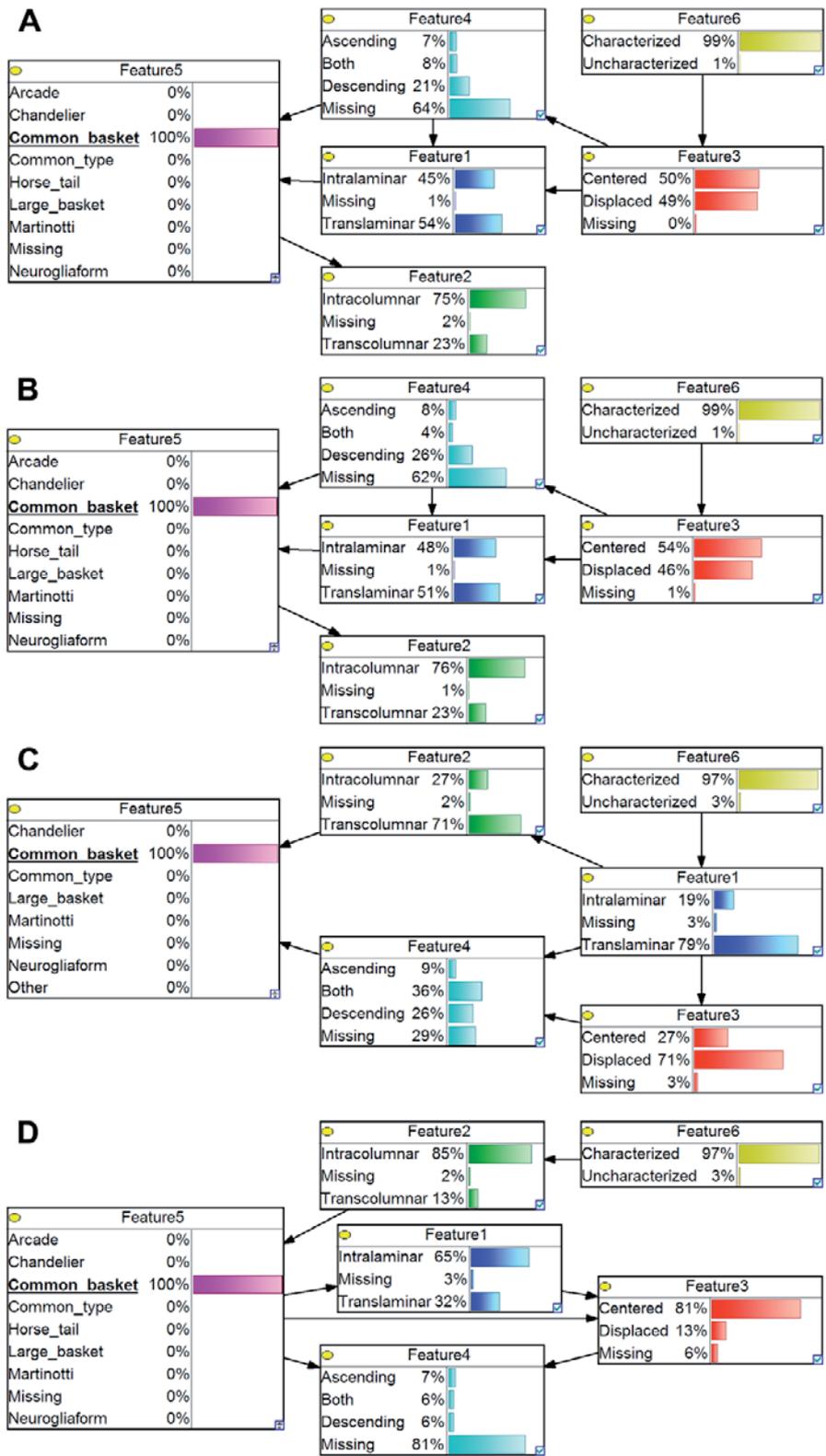We also used Bayesian networks to analyze the disagreements between experts about the classification of interneuron types. **Fig. S19** shows the Bayesian networks for the same four experts when Common

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 S5**, 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 S5. 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 S6** 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 S6** outperformed a base classifier which always selects the class with maximum prior probability according to an exact binomial test (see asterisks in **Table S6** and **Supplementary Online Information S1**).

Table S6. 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 S7** shows the confusion matrix of the best performing algorithm (SMO), which achieved an accuracy of 62.24% (**Table S6**). 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 S7**). 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 S7. 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 S8** 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 S8. 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 S9. 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*