Cellular and Molecular Organization of the Rhesus Macaque Dorsal Horn with Comparison to Mouse

Cynthia M. Arokiaraj[1], Michael Kleyman[2], Alexander Chamessian[3,4], Stephanie Shiers[5], Byungsoo Kang[2], Meaghan M Kennedy[2], Yitao Chen[2], Richard P. Dum[6], Ryan Patterson[7], David A. Lewis[8], Yawar Qadri[3], Ariel J. Levine[7], Theodore J. Price[5], Andreas R. Pfenning[2,9] and Rebecca P. Seal[1].

1. Department of Neurobiology and Pittsburgh Center for Pain Research, University of Pittsburgh School of Medicine, Pittsburgh, PA, 15213
2. Department of Computational Biology, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, 15213
3. Department of Anesthesiology, Duke University Medical Center, Durham, NC, 27708
4. Department of Anesthesiology, Washington University Pain Center, Washington University School of Medicine, St Louis, MO 63110, USA.
5. Department of Neuroscience and Center for Advanced Pain Studies, University of Texas at Dallas, Richardson, TX, 75080
6. Department of Neurobiology, Systems Neuroscience Institute, University of Pittsburgh School of Medicine, Pittsburgh, PA, 15213
7. Institute of Neurological Disorders and Stroke, NIH, Bethesda, MD, 20892
8. Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA 15261.
9. Neuroscience Institute, Carnegie Mellon University, Pittsburgh PA, 15213.

Corresponding Author: rpseal@pitt.edu

Summary

The spinal dorsal horn transforms incoming somatosensory information and transmits it supraspinally to generate modality-specific sensory percepts. The lack of an established framework for the molecular and cellular organization of the dorsal horn across species has greatly hampered delineating precisely how this region processes somatosensory information, including pain. Furthermore, the translation potential of the rodent work is unclear without data from higher order species. To fill these gaps, we performed single nucleus RNA-sequencing of Rhesus macaque dorsal horn and compared the results to a recently reported meta-analysis of mouse. Because dorsal horn laminae serve as a key organizing principle for function, we also determined the laminar location of each identified cell type. The work provides a comprehensive cross-species cellular and molecular database that will be critical for decoding the logic of dorsal horn somatosensory circuits and validating preclinical targets.

Keywords: Rhesus macaque, human, mouse, spinal cord, single nuclear RNA-sequencing, cell types, classification, neuropeptides

Introduction

The spinal cord dorsal horn is a principal site for somatosensory integration and transmission not only for normal sensory processing but also for maladaptive states such as chronic pain[1,2]. Classification schemes to describe neurons in the dorsal horn were initially based on major groups that were organized in a dorsoventral orientation starting with the most dorsal edge, also known as the marginal layer, followed by the substantia gelatinosa, the nucleus proprious and the Dorsal Nucleus of Clarke. Bror Rexed in the early 1950's, took into account the lamination pattern of the neuronal cytoarchitecture, and redefined the classification as layers or laminae that likewise extended dorsoventrally[3].

Consistent with the concept that laminar organization underlies functional organization, central terminals of functionally distinct classes of primary afferent fibers terminate within specific laminae. Unmyelinated or lightly myelinated primary sensory neurons that sense temperature or noxious stimuli terminate primarily in laminae I-II but some also send fibers to lamina V[4-6]. Unmyelinated primary sensory neurons that transmit information related to itch or innocuous mechanical stimuli terminate in lamina II, while myelinated fibers that carry innocuous, tactile information terminate in inner lamina II to V, also known as the low threshold mechanoreceptor-recipient zone[1,7-9]. Lastly, myelinated afferent fibers that are primarily involved in proprioceptive pathways terminate in the deeper dorsal horn which includes lamina VI in cervical and lumbosacral enlargements, and in the ventral horn[10,11]. Thus, each lamina of the dorsal horn receives distinct modalities of sensory input.

Within the Rexed laminae are diverse types of interneurons and projection neurons that make up the spinal microcircuits. The precise classification of individual cell types is critical for delineating how the dorsal horn processes sensory information. Historically, interneurons have been classified by their location, morphology, neuronal connectivity, neurotransmitter type and action potential firing patterns[7,12-15]. However, even taking into consideration these features, the neurons remain highly heterogeneous. Recent single cell transcriptomic studies of embryonic and adult mouse spinal cord have added significant molecular detail and a subsequent meta-analysis further aligned the cell type classifications[16-21]. Nevertheless, despite its importance to fundamental studies of spinal cord function and the validation of targeted therapeutics, whether this classification scheme holds true across species, particularly with respect to primates, is not known.

To address this, we generated a high-resolution single-nucleus transcriptomic map from Rhesus macaque and then performed computational analyses to compare gene expression profiles to the mouse meta-analysis dataset. Together, the work produced a classification scheme of the dorsal horn cell types that translates well across species from rodent to primate. To further solidify the classification scheme, we also compared the laminar distributions of the identified cell types from each species. Lastly, we examined species-specific differences in gene expression. These differences may reflect evolutionary changes that could impact function and translatability.

Results

Major Cell Types of the Macaque Dorsal Horn

To determine the cellular and molecular organization of the macaque dorsal horn, we performed single-nuclear RNA sequencing (snRNA-seq) on lumbar tissue (L4-5) harvested from young adult (3-year-old) male Rhesus macaques (n=3) using the 10X Genomics Chromium platform. For the isolation of the dorsal horn nuclei, the spinal grey matter of each animal was dissected, transected at the central canal and mechanically dissociated (**Figure 1A**). Deep sequencing of the single nucleus libraries produced an average of ~50M reads per animal. To optimize alignment of the reads to the macaque genome (Mmul10), we employed a custom annotation method (**Methods**) that mapped 91% of the reads to the genome[22]. After quality control filtering, 12,243 nuclei were retained for further analyses. Importantly, we observed that the number of UMIs highly correlated to the number of genes detected, indicating a high-quality sample (**Figure S1A**). The three biological replicates were then co-normalized to remove cell-specific read sampling biases using Scran and subsequently integrated and batch-corrected with Scanorama (**Figure S1B**)[23,24]. Nuclei were assigned to their cell type based on the expression of well-established marker genes. An average of 1,720 genes and 3,700 UMIs were detected per cell type (**Figures 1B-C** and **S1C**)[16]. The neuronal marker, RNA Binding Fox-1 Homolog 3 (*RBFOX3)*, identified 2,698 nuclei, which comprised ~22% of the total (**Figure 1B-C**). The largest cell class was formed by oligodendrocytes (39%), which were identified by the gene encoding Myelin Basic Protein (*MBP*) (**Figure 1B-C**). These cells could be further sub-clustered into two groups based on the expression of quinoid dihydropteridine reductase (*QDPR*) in one case or of Dpy-19 like C-mannosyltransferase 1 (*DPY19L1*) together with S100 calcium binding protein B (*S100B*) in the other (**Figure 1B-C**). Astrocytes were present in about the same numbers as neurons and were also sub-clustered into two groups based on high (Astrocyte 1) or low (Astrocyte 2) expression of Glial fibrillary acidic protein (*GFAP*). Oligodendrocyte precursor cells, microglia, meninges, ependymal cells and Schwann cells all made up smaller individual clusters (**Figure 1B-C**).

Transcriptomically Defined Neuronal Clusters

We next sub-clustered the neuronal nuclei into 16 distinct clusters that could be broadly characterized as either excitatory or inhibitory using the known gene markers: vesicular glutamate transporter 2 (*SLC17A6*) for excitatory and vesicular inhibitory transporter (*SLC32A1*), as well as GABA synthesizing genes i.e. glutamate decarboxylase 1 and 2 (*GAD1 and GAD2*) for inhibitory (**Figures 2 and S2**)[25]. Two additional clusters were unusual in that they each contained both excitatory nuclei and inhibitory nuclei, but otherwise had computationally indistinguishable gene expression profiles (**Figure S2A).** This is characteristic of clusters previously identified to be a part of the intermediate and ventral regions of the spinal cord (Laminae VI-VII)[16]. Because of the challenge of assigning these nuclei to either excitatory or inhibitory clusters, we combined the nuclei into one midVentral domain (midVen) cluster (**Figure S2B**).

Our final dataset, excluding the midVentral cluster, consisted of 1,954 nuclei, with 64% classified as excitatory and 36% as inhibitory (**Figure 2A-B**). From these nuclei, we identified 11 distinct excitatory clusters (GLUTs 1-11) and 5 distinct inhibitory clusters (GABAs 1-5). Each cluster had an average of 139 nuclei (range is from 49 to 222 nuclei per cluster), 4,000 genes and 10,000 UMIs (**Figure. 2B-D**). Relationships between the 16 individual clusters were visualized using the Ward's hierarchical clustering method together with the correlation matrix (**Figures 2E and S2C**). As expected, the greatest distance in the resulting dendrogram spanned from the excitatory to the inhibitory clusters, indicating these two groups share the least similarity of gene expression profiles. We next identified marker genes for each cluster, which were chosen by being well-represented among the nuclei of the cluster as well as highly enriched in that cluster (**Figure 2E-F).** For some clusters, a combination of marker genes was used to obtain higher representation of nuclei within the cluster.

Laminar Distribution of Marker Genes

The laminar location of a cell type provides some indication of its potential functional role. Many of the clusters identified by snRNA-seq in mice show enrichment within specific laminae. To determine how the transcriptomically defined macaque clusters are distributed within the Rexed laminae of the lumbar dorsal horn, we performed RNAscope® fluorescence *in situ* hybridization in L4-5 spinal cord slices taken from both male and female macaques using the selected marker genes (**Figures 3 and 4**)[26]. Nearly all cluster markers showed expression patterns that were highly or significantly enriched in either the superficial (I-II) or deep

dorsal horn (III-V) laminae. For example, of the total neurons in GLUT1, identified using probes to either Corticotropin-releasing hormone receptor 2 (*CRHR2*) or Collagen, type XIII, alpha 1 (*COL13A1*), 86% were located within laminae I-II (**Figure 3**). Similarly, 88% of neurons in the GLUT2 cluster (identified using Neuromedin U receptor 2 (*NMUR2))*, 69% of neurons in the GLUT5 cluster (identified by the overlap of Collagen Type XXIV Alpha 1 Chain (*COL24A1*) and Neuromedin B receptor (*NMBR*)), and 81% of neurons in the GLUT6 cluster (identified by the overlap of Tachykinin precursor 1 (*TAC1)* and Anosmin 1 (*ANOS1*)) were located in the superficial laminae. In contrast*,* GLUT3 (identified by the overlap of *TAC3* and Neuromedin U (*NMU*) probes), GLUT4 (identified as positive for COL5A2 and negative for the inhibitory neuron marker Paired box 2 (PAX2)), GLUT7 (identified by expression of either Collagen Type XXI Alpha 1 Chain (*COL21A1*) or *MAF*), GLUT9 (detected by expression of both Tolloid like 2 (*TLL2*) and ADAM Metallopeptidase With Thrombospondin Type 1 Motif 16 (*ADAMSTS16*)) and GLUT10 (identified with a probe to Otogelin-like (*OTOGL*)) clusters were enriched in the deep dorsal horn. Only two clusters, GLUT8 and GLUT11, showed no laminar enrichment. GLUT8 was identified by the overlap of MAF BZIP Transcription Factor A (*MAFA)* and *RBFOX3,* and GLUT11 by prodynorphin (*PDYN*) that did not overlap with *PAX2*.

The inhibitory clusters were similarly analyzed (**Figure 4**). Neurons in the GABA5 cluster, identified by the co-expression of *PDYN* and *PAX2*, are located primarily within laminae I-II (**Figure 4**). GABA4 inhibitory neurons marked by tyrosine-protein kinase *MET* and pleiotrophin (*PTN)* expression are also enriched in the superficial laminae. Neurons belonging to GABA1 were visualized by the overlap of MBL Associated Serine Protease 1 (*MASP1*) and *PAX2* and to GABA2 by expression of both GDNF antisense RNA 1 (*GDNF-AS1*) and ras responsive element binding protein 1 (*RREB1).* Because neuropeptide Y (*NPY*) expressing neurons are found in GABA2 as well as GABA3, we defined the latter population as NPY neurons that did not also express *RREB1.* Markers for these three clusters (GABA 1-3) were enriched in the deep dorsal horn. Lastly, all cluster markers showed a similar distribution in males and females.

Cross-species Comparison of Mouse and Macaque Neuronal Dorsal Horn Clusters

A meta-analysis that unified six single cell sequencing datasets from the mouse, put forward a harmonized atlas of defined neuronal clusters for the dorsal horn that were also organized into distinct families[19]. To determine the extent of molecular and cellular conservation between macaque and mouse dorsal horn neurons, we therefore compared our macaque dataset to this ~10,000 neuron mouse dorsal horn atlas. To test whether populations were conserved across the mouse and macaque, we first examined sets of mouse marker genes (**Figure S3A**). Specifically, each neuronal nucleus from macaque was compared to each major mouse class (excitatory, inhibitory or midVentral) based on the top 100 marker genes from the mouse meta-analysis. From this analysis, a high correlation was observed across species for both excitatory and inhibitory clusters (**Table S1 and Figure S3A**). Additionally, nuclei assigned to the midVentral cluster in macaque show the highest correlation to the midVentral clusters in mouse (**Figure S3A**). The macaque nuclei were then assigned a score depending on how similar their gene expression profiles were to the individual mouse clusters. We used these scores to annotate each of the macaque clusters with the most likely corresponding cell types in mouse[27]. Results from this type of analysis also showed a good correspondence between the nuclei of individual macaque and mouse clusters (**Figure S3B)**. There were however a few exceptions to this correspondence. The macaque GABA2 and GABA4 clusters did not correspond well with any of the mouse clusters (described in more detail below), and the mouse clusters Excit-07 and Excit-17 showed no correspondence to any of the macaque clusters (**Figure S3B**). A lack of orthologous populations for Excit-07 and Excit-17 is likely because these two mouse clusters contain neurons derived only from the Rosenberg dataset, which examined spinal cords taken from mice at early postnatal ages (P2 and P11)[19,20].

Neurons in the mouse meta-analysis were sub-clustered to a higher degree than in the macaque dataset, resulting in a greater number of smaller clusters. For this reason, we also compared the macaque clusters with the higher-order clusters of the mouse, termed families. In these analyses, the orthologous genes of key markers from each of the mouse families were plotted onto the newly generated macaque UMAP (**Figure 5A**). From this, we observed that GLUT1 and GLUT2 clusters correspond to the Reln family; GLUT3, GLUT4, GLUT5 and GLUT6 to the Sox5 family; GLUT7 to the Maf family; GLUT8 to the Cpne4 family; GLUT9 and GLUT10 to Prkcg family; GLUT11 to the Rreb1 family; GABA1 to the Cdh3 family; GABA3 to the Npy family and GABA5 to the Pdyn family (**Figures S2B and S3B**).

As previously mentioned for the mouse cluster comparison, neither GABA2 nor GABA4 showed a direct correspondence to the mouse families. Upon closer examination, we observed that macaque UMAPs of orthologous genes of two of the mouse family markers, *RORB* and *ADAMTS5*, show an interesting separation within both the GABA2 and GABA4 clusters (**Figure 5A-B**). Indeed, further sub-clustering of GABA2 and GABA4 revealed a correlation of GABA2_1 and GABA4_1 with the mouse Rorb family and of GABA2_2 and GABA4_2 with the Adamts5 family (**Figure 5B and Figure S3B**). To reveal how the gene markers chosen for macaque clusters are distributed within the mouse clusters (and mouse families) as well as how the mouse family markers are distributed within the macaque clusters, we generated dotplots for the cluster distribution of both sets of markers (**Figure S3C**).

Based on the strong correspondence of top marker genes between mouse and macaque cell populations, we tested whether the datasets could be integrated beyond the cluster level, in order to visualize the similarity of single nuclei. We normalized the raw count data across three macaque samples and the adult mouse sample using Seurat and then integrated these datasets using LIGER[28]. There is very little separation according to sample or species, indicating that the datasets were indeed integrated (**Figure 5C, top**). Regardless of species, cells belonging to the same class and family had a strong tendency to be near each other within the UMAP plot, suggesting that the datasets were well-integrated and cell type signatures drive the visualization and clustering (**Figure 5C, middle and bottom**). These results further support the presence of consistent cell populations across macaque and mouse.

Laminar distributions of neuropeptides in macaque and human

In a mouse ontology study, genes related to neuropeptide signaling were found to be overrepresented in excitatory neurons of the dorsal horn[29]. Further, the laminar distribution of the neuropeptide genes such as cholecystokinin (*CCK*), neurotensin (*NTS*), gastrin releasing peptide (*GRP*), neuropeptide FF-amide peptide precursor (*NPFF*), thyrotropin releasing hormone (*TRH), TAC1*, *TAC2* and *NMU* show expression patterns restricted to one or two dorsal laminae. Furthermore, several of these neuropeptides have been targeted for study in the rodent dorsal horn and have been shown to be important for pain or itch[30-32]. We were therefore interested to see if the laminar distributions of these genes are conserved across species including in human (**Figure 6A and Table S2**). TAC1 and GRP expressing neurons, implicated in pain and

itch, respectively, are located mainly within lamina II of mouse and we find that these peptides have a similar distribution in macaque and human. Similarly, the laminar location of the peptides TAC3 (*Tac2* homolog in primate) and NMU located in laminae II and III of mouse are conserved in macaque and human. NTS neurons, which have been implicated in central sensitization phenomenon of windup and largely express PKCγ in mouse, are similarly enriched in lamina III of macaque and human. Interestingly, while *CCK* expressing neurons are abundant in laminae III and IV of mouse and macaque dorsal horn, they are equally abundant in the superficial laminae in human.

We also examined the distribution of two other neuropeptide genes which are markers of inhibitory clusters in both macaque and mouse, *NPY* and *PDYN.* Both peptides have been implicated in pain and itch in mice, though NPY for mechanical itch and PDYN for chemical itch (**Figure 6B**). The *NPY* gene is exclusively expressed by inhibitory neurons distributed across all dorsal horn laminae of mouse and we see a similar distribution in macaque and human[17,33,34]. With respect to the *PDYN* gene, the neurons are primarily inhibitory and reside within laminae I-II in all three species. Interestingly in mice, the number of excitatory dynorphin neurons present in the medial part of the L4 segment, which receives glabrous skin input, is significantly greater compared to the L2 segment, which only receives hairy skin input[35,36]. Such a restricted distribution along the medio-lateral axis of the lumbar enlargement is not observed for excitatory PDYN neurons in macaque or human. Thus, with a few potentially biologically interesting and relevant exceptions, the laminar distributions of these neuropeptide genes are conserved suggesting similar roles across species.

Comparisons of select functionally relevant genes between mouse and macaque

Cells belonging to 11 principal progenitor domains in the developing mouse spinal cord are defined by specific combinations of transcription factors, which govern neuronal cell type specification along the dorso-ventral axis of the mature spinal cord. Six of these domains (dI1-6), as well as two late-born domains (dIL_A and dIL_B) give rise to post-mitotic dorsal horn interneurons[37,38]. Given that some of these embryonic transcription factors continue to express in major cell classes (i.e. excitatory or inhibitory groups) in adult, we examined whether cell types and transcription factors show a similar correspondence in the macaque spinal cord[39]. Lim homeobox transcription factor 1-beta (LMX1B) defines the dI5 and late born excitatory

neurons, while gastrulation brain homeobox 1 (GBX1) is required for the development of a subset of PAX2 inhibitory neurons[39,40]. We observed a similar separation in the macaque where *LMX1B* is widely expressed among the excitatory clusters and *GBX1* is found specifically in *PAX2* expressing inhibitory clusters (**Figure S4B**). Interestingly, the Brn3a transcription factor (also known as POU class 4 homeobox 1 - *POU4F1*), which is a marker of excitatory populations- dI1-3 and dI5 during mouse and presumably macaque development, is present in the GLUT4 and GLUT11 clusters of these 3-year-old macaques[41]. The Maf group of transcription factors (MafA, MafB and c-Maf) are largely expressed by neurons in laminae II-IV of the mouse dorsal horn at postnatal ages as well as in adult[39]. In the macaque, *MAFB* is not detected in the dorsal horn, but *MAF* and *MAFA* serve as marker genes of GLUT7 and GLUT8, respectively (**Figure S4B**). Moreover, as in mouse, *MAF* expression is restricted to laminae III-IV while *MAFA* expression was visualized in lamina II as well as in deeper laminae (**Figure 3**). Lastly, the Zinc finger homeobox 3 (*Zfhx3*) transcription factor was recently shown to be expressed ventrolaterally by excitatory and inhibitory interneurons in the mouse spinal cord from embryonic stages into adulthood[42]. Additionally, different types of long-range projection neurons were found to express *Zfhx3*. The orthologous gene in macaque, *ZFHX3*, is present in several excitatory and inhibitory clusters, with high expression in GLUT1, GLUTs7-10 and GABA3 (**Figure S4B**). Thus, projection neurons, in addition to local interneurons, may belong to these five excitatory clusters.

We also examined the distribution of ion channels and receptors among the macaque clusters. Overall, their expression patterns correspond well with what has been reported for the mouse. With respect to the voltage-gated sodium channels, the most commonly studied include Nav1.7 (*SCN9A*), Nav1.8 (*SCN10A*) and Nav1.9 (*SCN11A*) due to their role in pain in peripheral sensory neurons of mice and humans[43,44]. *SCN9A* is distributed across several clusters as in mice while *SCN10A* and *SCN11A* are lowly expressed in both species[16,17]. The TTX-resistant Nav1.5 channel, encoded by *SCN5A* is highly specific to the macaque GLUT8, similar to its expression in the corresponding mouse Cpne4 cluster (Excit -01) (**Figure S4C**). As in mice, we noticed that most of the calcium and transient receptor potential (TRP) channels of the subfamily C and M are present in a number of clusters, while the TRP channels of the V subfamily are sparsely expressed (**Figure S4D-E**)[17,34].

Similarly, with respect to neurotransmitter receptors, we find that the Gamma-aminobutyric acid reporter subunit alpha-1 (*GABRA1*) is highly expressed in GLUT7 and GLUT9 clusters and to a lower extent in GABA4 (**Figure S5D**). This correlates well with its expression in the corresponding clusters of the mouse. Most nicotinic acetylcholine receptors are not expressed or sparsely expressed in the macaque and mouse dorsal horn. However, we do detect high levels of *CHRNA7* across many of the macaque clusters (**Figure S5F**). The serotonin receptor encoding gene 5-Hydroxytryptamine Receptor 2A (*HTR2A*) is highly expressed in GLUT10, which corresponds well with its high expression in the mouse Prkcg family of clusters (**Figure S5G**). Polymorphisms in the *HTR2A* gene have been associated with chronic low back pain and fibromyalgia in human patients[45,46].

Genes differentially expressed between macaque and mouse transcriptomes

Although the expression profiles of genes within cell populations are conserved enough to identify strong cluster relationships between the mouse and macaque, we did find species-specific differences. To identify genes that show species-specific cluster distributions, we first identified genes that were differentially expressed across each corresponding macaque and mouse cell type and family using a Wilcoxon Rank Sum test[34]. To avoid confounds from development or the isolation of whole cells versus nuclei, we only used the gene profiles of nuclei from the meta-analysis that were isolated from adult mice[27]. To identify a conservative set of differences, we first chose highly significant genes of each cluster as those that are $p < 10^{-5}$ and have a log-fold-change greater than 2 relative to their expression in the other excitatory or inhibitory families. Second, the expression of the gene in the corresponding family of the other species had a log-fold-change less than 0 relative to the other clusters of that species (**Figures 7, S7-8 and Tables S3-4**). We identified fifteen genes that are enriched in the macaque excitatory clusters compared to the mouse (**Figure 7A)**. Conversely, only carbonic anhydrase 12 (*CA12)* is enriched in the mouse excitatory clusters in comparison to the macaque (**Figure 7B)**. A number of genes (12) are enriched in the macaque inhibitory clusters compared to the mouse, among these include *MET* and *PTN,* which are marker genes for GABA4 (**Figure S7 and Table S3**). Surprisingly, *CPNE4* is not enriched in the cluster corresponding to the mouse Cpne4 family i.e., GLUT8 (**Figure 7C-D**). However, one of the genes enriched in GLUT8 is the Piezo- type mechanosensitive ion Channel component 2 (*PIEZO2*). Consistent with this result, the number of *PIEZO2*

positive cells detected in the macaque dorsal horn is significantly higher than in the mouse, as visualized by *in situ* hybridization (**Figure 7E-F**). *PIEZO2* expressing primary sensory neurons have been studied for the critical role they play in mechanotransduction so their significantly larger presence in the macaque dorsal horn poses a question of how they may influence or be critical to the function of these neuronal populations[47,48]. Interestingly, different collagen genes are specifically enriched in several of the macaque excitatory clusters, thereby serving as useful marker genes (**Table S3**). The finding raises the question of what role(s) collagen proteins and specifically each type plays in dorsal horn function.

Discussion

Understanding the cellular and molecular organization of the dorsal horn is essential for determining how the cell populations form functional circuits to process somatosensory information. Our work provides a high-quality snRNA-seq dataset from the macaque dorsal horn. We have created an interactive database https://seallab.shinyapps.io/dorsalhornseq/ with which gene expression profiles can be examined within the 16 distinct neuronal cell types and referenced to the 11 mouse families. In addition, our analysis of the laminar organization of these neuronal cell types in macaque shows that most are enriched in the either the superficial or deep dorsal horn laminae, consistent with their corresponding mouse cell types and consistent with the concept that the laminar organization plays a role in the functional organization[32]. Clusters with similar gene expression profiles and laminar distributions may arise from similar developmental programs, which could also provide potential clues to their functional roles.

Our comparative analyses between macaque and mouse clusters show that the underlying molecular organization is also indeed largely conserved. Our analysis of features that have been classically used to describe cell types, such as their neurotransmitter phenotype or their expression of transcription factors or ion channels, also shows a high congruence with what has been reported in the rodent dorsal horn. Converging cross species evidence supporting the identification of conserved cell types and molecular signatures (as well as pointing out the differences) provides a critical resource for functional and translational studies, especially considering the past challenges of harmonizing transcriptomic datasets across studies even within the same species[34].

This study illustrates and can be mined for data that points out relationships between molecular and functional conservation across species. Glycinergic neurons, which through pharmacogenetic activation in mice have been shown to alleviate neuropathic injury-induced hyperalgesia and to block chloroquine- and histamine-induced itch, are prevalent in deeper laminae of the mouse dorsal horn[49]. We similarly find that the GABA1-3 clusters of macaques, which show the highest expression of the glycine transporter 2 (*SLC6A5*), are also enriched in the deeper laminae (III-V). Similarly, the hyperpolarization-activated cyclic nucleotide-gated cation channel *HCN4*, which has been implicated in mechanical allodynia (touch-evoked pain) under neuropathic conditions due to its presence on PKCγ neurons, is enriched in macaque clusters GLUT9 and GLUT10 which correspond to the mouse Prkcg family[50].

As described previously, the GABA2 and GABA4 clusters did not show a strong correspondence to any of the mouse families. However, sub-clustering them produced a strong correspondence to the mouse Rorb (GLUT2_1 and GLUT4_1) and Adamts5 (GLUT2_2 and GLUT4_2) families. We suspect that this may be due to a higher degree of similarity in the molecular profiles of the Rorb and Adamts5 families during embryonic/developmental ages than in the adult. Alternatively, it is possible that evolutionary differences in gene expression resulted in GABA2 and GABA4 clusters becoming more molecularly distinct in the macaque. Interestingly, both the GABA4_1 sub-cluster, and its corresponding subset from the mouse Rorb family (Inhibitory 03-05), show high expression of calretinin (*Calb2*) (https://seqseek.ninds.nih.gov/#/genes). Calretinin inhibitory neurons are largely concentrated in lamina II of the mouse dorsal horn and make up ~25% of the inhibitory neurons in this area[51]. Thus, the GABA4_1 cluster may represent a more superficially located population of the *RORβ* neurons than GABA2_1. A study involving the neuronal activity marker, phosphorylated extracellular signal related kinase (pERK), reported a role for inhibitory calretinin neurons in processing noxious mechanical information and to a lesser extent, noxious chemical stimulation[52].

Other inhibitory dorsal horn populations in mice that have been described by histochemical methods as not overlapping, apart from calretinin, include galanin (*Gal*) which co-expresses dynorphin, neuronal nitric oxide synthase (*Nos1*), parvalbumin (*Pvalb*) and *Npy*[35]. The expression of two of these genes is also largely segregated among the 5 macaque GABA clusters: *PVALB* is expressed in nuclei of the GABA2_1 cluster and *PDYN* in nuclei of the GABA5 cluster. In contrast, *NPY* is present at high levels

in many nuclei of the GABA3 cluster, but also is found in GABA2_1 and GABA5 clusters. *NOS1* is highly expressed by GABA1 cluster, but is also present in all other GABA clusters. *GAL* was detected at low levels in the macaque dorsal horn and thus overlap with *PDYN* also appeared minimal. Interestingly, despite the widespread expression of *NPY* and *NOS1* across macaque inhibitory clusters, these genes show minimal overlap. *PVALB* is not only expressed by nuclei of inhibitory clusters but is also expressed by a specific excitatory cluster- GLUT7, which corresponds to the Maf mouse family[35,53]. On the other hand, *NOS1* is highly expressed in several excitatory clusters with patterns similar to the mouse, while *PDYN* expression appears to be specific to GLUT11 [15,35,54].

Indeed, a lot of work has been done in mice to understand the role of neuropeptide-expressing excitatory dorsal horn neurons and their receptors. Several of these neuropeptides show modality-specific functions with respect to pain or itch processing and restricted laminar expression patterns[29]. *GRP* neurons in the macaque belong to GLUT1, GLUT3 and GLUT5 clusters which correspond to their inclusion in the mouse meta-analysis families - Reln (Excit 08-10) and Sox (Excit15, 18-19). In the rodent dorsal horn, Grp neurons present in lamina II receive primary afferent input from MrgprA3 pruritoceptors and are involved in spinal itch mechanisms[31,55]. The GRP receptor GRPR, found in laminae I-II of the rodent dorsal horn also transmits itch information[55]. Furthermore, GRPR neurons reportedly form synaptic contacts with projection neurons in lamina I[56]. The macaque GLUT2 cluster contains neurons expressing *GRPR* as well as a subset of the somatostatin (*SST*) neurons- a population that has also been implicated in the itch circuitry in mouse[36]. Interestingly, we find that the neuropeptide thyrotropin-releasing hormone (*TRH*) – expressing neurons belong to both GLUT1 and GLUTs 9-10 which correspond to the Reln and Prckg mouse families, respectively. Computationally, these neurons do not appear to overlap with the *GRP* population which is also present in GLUT1. TRH neurons in the mouse dorsal horn have not been functionally investigated in the context of pain or itch, but as a subset of the larger PKCγ neuron population, they may be involved in the transmission of neuropathic pain[30,57].

Both GLUT3 and GLUT4 are characterized by *TAC3* expression and are located in laminae II-III similar to the mouse meta-analysis clusters Excit-18-19, which are marked by *Tac2* expression. *TAC3* is also minimally expressed in GLUT11. Ablation of TAC2 neurons did not affect acute somatosensation or mechanical thresholds after nerve injury in mice[58]. However, there is evidence for a role for these neurons

in mechanical itch[59]. Additionally, 84% of TAC2 neurons express calretinin, which has been shown to be important for the transmission of mechanical allodynia under inflammatory conditions[30,51]. Both GLUT5 and GLUT6 show high expression of *TAC1* similar to their mouse meta-analysis counterparts Excit- 15-16 and in both species, *TAC1* is distributed in laminae I-II as well as in deeper laminae IV-V. We also find lower levels of *TAC1* expression in the macaque GLUT4 and GLUT11 clusters. Ablation of TAC1 dorsal horn neurons in mice did not alter reflexive-defensive behaviors but rather caused a deficit in nociceptive coping behaviors- suggestive of a role in the affective component of sustained pain[60].

The neuropeptide *CCK* is detected in several clusters. It is expressed in GLUT7 and at lower levels in GLUT9 which correlates with *Cck* expression in the mouse meta-analysis clusters Excit-05 (Part of Maf family) and Excit-03 (Part of Prkcg family), respectively. We also observe high expression of *CCK* in GLUT5 which expresses *TAC1.* One study reported that ~20% of CCK neurons in laminae I-III overlap with *Tac1* in mouse using histological methods[57]. Interestingly, we detected high expression of *CCK* in GLUT11 and this is not seen in the mouse meta—analysis counterpart. In mice, laminae III-IV CCK neurons are involved in the transmission of mechanical allodynia and heat hypersensitivity following inflammatory injury and transmission of mechanical allodynia following neuropathic injury[30]. It is notable that the distribution of *CCK* was different in the human spinal cord, where it is more highly expressed in the outer laminae. Given the differences in nociceptor primary afferent termination in humans, these neurons may play a specialized role in pain transmission in clinical pain disorders[61,62]. GLUT7 and GLUT9 express the neuropeptide-*NTS* which corresponds with what is seen in the mouse meta-analysis clusters Excit-03-05. Activation of NTS- lineage neurons in laminae II_i-III of the mouse dorsal horn resulted in an increase in withdrawal response to brush stimulation[32]. Dorsal root stimulation of NTS-lineage neurons was sufficient to evoke wind-up in lamina I spinoparabrachial neurons[63]. Overall, the expression of neuropeptides in distinct clusters shows a strong correlation between the mouse and macaque (**Table S5**)[29]. Furthermore, we show a striking conservation in their laminar distribution patterns across species including in human, suggestive of conserved somatosensory roles for these neuropeptides in the primate dorsal horn.

Projection neurons located in lamina I and in the deeper laminae III-VI transmit modality-specific somatosensory information to multiple brain regions[64]. As is the case for spinal interneurons, projection neurons are also molecularly diverse, however, molecular identification of the projection neurons is more

challenging in part because they are relatively small populations and because it requires back-labeling the neurons from distinct brain regions coupled with single cell sequencing or histology using cell type markers. Such studies have been performed to an extent in rodents, but also would be a major undertaking in macaque. Nevertheless, it is an important question that should be addressed in future studies. Extrapolating what is known in rodents to our new macaque database, we can speculate on which GLUT clusters contain projection neurons. Studies in rodents have focused primarily on spinoparabrachial projection neurons. Those expressing neurokinin 1-receptor (NK1R) are thought to account for a majority of lamina I projection neurons[64,65]. NK1R$^+$ projection neurons are activated by noxious stimuli and innocuous cooling and have been implicated in hyperalgesia under inflammatory or neuropathic pain conditions[66,67]. A subset of the NK1R neurons may also be involved in spinal itch mechanisms[68]. The *TACR1* gene which encodes the NK1 receptor is expressed in macaque GLUTs 4, 6, 7 and 11 clusters. Häring et al. identified an excitatory cluster – GLUT15 that consisted of spinoparabrachial neurons marked by the expression of *Lypd1* and *Elavl4*[17]. We noted that orthologs of these two mouse genes, *LYPD1* and *ELAVL4* are expressed together with *TACR1* in GLUT7 and GLUT11, but also in GABA1 and in deeper midVen clusters. A more recent study by Wercberger et al. analyzed the molecular signatures of spinoparabrachial projection neurons that are not NK1R$^+$, and identified a small proportion of projection neurons that expressed either *Cck,* neuronal pentraxin 2 *(Nptx2)* or neuromedin B (*Nmb)* many of which reside outside lamina I[69]. While *CCK* is expressed in several macaque excitatory clusters including GLUT7 and GLUT11, as is seen with *TACR1*, *NPTX2* is enriched in GLUTs 1, 2 and 8. Thus, the various, diverse populations of spinal projection neurons are likely spread across most GLUT populations.

While there is an overall conservation of gene expression within dorsal horn neurons from mouse and macaque, we also note species-specific differences. Some differences were initially identified when assigning marker genes to the macaque clusters. While marker genes such as *NMUR2*, are unique and highly expressed in correlating clusters of both mouse and the macaque, this was not observed with all clusters. For example, we now provide novel molecular markers such as *CRHR2* and *COL13A1*, which are not present in the mouse dorsal horn, as better candidates for the targeted study of their respective cell types in the macaque. Another notable difference in gene expression was observed with *PDYN* in excitatory neurons which are more prominent in the primate dorsal horn compared to the mouse. Sathyamurthy et al

study assigned the excitatory dynorphin neurons to the DE-15 cluster whereas they were not reported in the Haring et al study or in the meta-analysis[16,17,34]. Immunohistochemical studies in mice suggest excitatory dynorphin neurons are primarily located within the medial portion of lumbar segment L4, which receive input from glabrous skin[35,36]. Discrepancies in the reporting of these cells therefore may reflect their low abundance outside the lumbar enlargement in mice. Interestingly, in the rat lumbar dorsal horn, approximately ~50% of the dynorphin neurons in lamina I and 20% of those in lamina II were reported to be excitatory and to span the mediolateral axis[70]. In the primate, we observe 60-70% of the dynorphin neurons in lamina I and 30-40% of those in lamina II to be excitatory. Here also, the expression was not restricted to the medial dorsal horn. The excitatory dynorphin neurons are of interest, as a transcriptomic study conducted in mice showed a unilateral upregulation of genes specifically in these neurons in the superficial laminae ipsilateral to the site of surgical incision or carrageenan inflammation[71]. Additionally, another study showed that excitatory dynorphin neurons are implicated in noxious heat-induced burn injury[72]. Taken together, the excitatory dynorphin neurons may potentially be involved in acute phase hyperalgesia.

The work reported here provides a much-needed framework for future studies that seek to transfer both descriptive and causal work in mice into a deeper understanding of the primate somatosensory system, such as the molecular identity of spinal projection neurons, the functional role of the defined macaque clusters in somatosensation, and the validation of molecular targets derived from mouse studies for human therapeutics.

Acknowledgements

The authors thank Ian Cumming at the Duke Human Vaccine Institute Flow Cytometry Core Facility, Nicolas Devos of the Duke Sequencing and Genomic Technologies Shared Resource and Dr. Priyabrata Halder for help with spinal cord harvests. Additionally, we thank Michael Pavel Vannice, Samyuktha Lokanandi and Andrea L. Goliaszewski for help with *in situ* hybridization quantification and Suh Jin Lee for the spinal cord diagram in Figure 1. Finally, we thank the organ donors and their families for the donation of spinal cord for research purposes and Dr. Jeffrey Reese and Anna Cervantes for spinal cord recovery. This research was supported in part by the University of Pittsburgh Center for Research Computing through the resources provided. Funding was provided by NIH extramural support to R.P.S. (NS104964 and NS109792), C.M.A. (NS111791), D.A.L (MH051234), and T.J.P. (NS111929) and NIH Intramural support to A.J.L. (1ZIANS003153).

Contributions

C.M.A and R.P.S. conceived of the study. M.K, B.K, M.M.K and A.R.P performed the bioinformatic analyses. A.C and Y.Q processed the nuclei for snRNA-seq. C.M.A performed the macaque spinal cord *in situ* hybridization and S.S performed the human spinal cord *in situ* hybridization. C.M.A created the R shiny macaque dorsal horn web browser. D.A.L provided the Rhesus macaques and R.D. harvested tissue. C.M.A. and R.P.S. analyzed the data. A.R.P., A.J.L. and T.J.P provided intellectual input. R.P.S. and C.M.A wrote the manuscript with significant contributions from M.K, A.R.P, and A.C.

Competing financial interests: The authors declare no competing financial interests.

Resource Availability

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Rebecca P. Seal (rpseal@pitt.edu).

Materials Availability

This study did not generate any new materials.

Data and Code Availability

The code and Jupyter notebooks used for the analysis of the macaque snRNA-seq has been made available at Github repository https://github.com/pfenninglab/dorsal_horn_snrnaseq. Additionally, the macaque dataset can be accessed through an R shiny application (https://seallab.shinyapps.io/dorsalhornseq/). This Shiny application was created using a Shiny Cell package[81].

The mouse meta-analysis pipeline is publicly available through the Github repository https://github.com/ArielLevineLabNINDS/SeqSeek_Classify_Full_Pipeline.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Macaque Spinal Cord Samples:

All lumbar spinal cord samples used in this study were obtained from *Macaca mulatta* provided by Dr. David Lewis at the University of Pittsburgh. Monkeys were housed in groups in the same social setting. All animals were deeply anesthetized with ketamine and pentobarbital and perfused transcardially with ice-cold artificial cerebrospinal fluid. All housing and experimental procedures were conducted in accordance with the guidelines of the US Department of Agriculture and the National Institutes of Health Guide for the Care and Use of Laboratory Animals and with the approval of the University of Pittsburgh Institutional Animal Care and Use Committee. No prior manipulations to the spinal cord were performed in these macaques. Three

male macaques (3-years old) were used for the single nuclear RNA-sequencing. Male and female macaque lumbar spinal cord samples (3-5 years old) were utilized for in situ hybridization studies. Though we did not observe sex differences for any of the genes using in situ hybridization, there is not enough statistical power to confidently assess if there are sex differences due to the limited availability of macaque tissue.

Human Spinal Cord Samples:

All human spinal cord procurement procedures were approved by the Institutional Review Boards at the University of Texas at Dallas. Donor information is provided in **Supplementary Table 5**. The human spinal cords were gradually embedded in OCT in a cryomold by adding small volumes of OCT over dry ice to avoid thawing. All tissues were sectioned at 20 µm onto SuperFrost Plus charged slides using a cryostat. Sections were only briefly thawed in order to adhere to the slide but were immediately returned to the -20°C cryostat chamber until completion of sectioning. Again, due to the limited availability of human donor tissue, not enough samples were assessed for sex differences at the gene level.

Nuclear Dissociation and Isolation

Snap-frozen lumbar spinal cord segments were removed from -80ºC storage and placed into separate Petri dishes containing a cold slurry of dissection buffer consisting of 1x Phosphate Buffered Saline (PBS, ThermoFisher Scientific; AM9625), 10% Dithiothreitol (DTT, Sigma; 43816-50ML).

From each animal, transverse sections (~50-75 mg) were made using a sterile razor blade. Three sections were selected for further dissection and the rest were discarded. The sections were cut in the coronal plane at the middle of the spinal cord to obtain dorsal and ventral halves. Any residual meningeal membranes were removed. The three dorsal halves were retained and the ventral halves discarded.

Nuclei were isolated according to Martelotto with some modifications[73]. The dorsal halves for each animal were transferred into separate Dounce homogenizers (Sigma) containing 1 mL of ice-cold homogenization buffer consisting of EZ Nuclei Lysis Buffer (Sigma; NUC101-1KT) with 0.5% RNasin Plus (Promega; N2615), 0.5% SUPERase-In (ThermoFisher; AM2696) and 1mM DTT. The samples were homogenized on ice using 20 strokes of Pestle A followed by 20 strokes of Pestle B. Any residual meningeal membrane was removed before switching pestles. The homogenate was filtered through a 50 µm filter (Sysmex; 04-004-

2327) into a 2 mL microcentrifuge tube (Eppendorf; 022431048). An additional 0.5 mL of homogenization buffer was used to wash the Dounce homogenizer and filter. The sample was then placed on ice while the remaining samples were processed. The sample was centrifuged at 500g at 4ºC for 5 min to obtain a crude pellet containing spinal nuclei. The supernatant was removed and discarded, being careful to not disturb the pellet. The pellet was resuspended in 1.5 mL of Homogenization Buffer and allowed to sit on ice for 5 mins. The samples were again centrifuged at 500x g, 4ºC for 5 min. The supernatant was removed and the pellet was resuspended in 1 mL of Nuclei Resuspension Buffer (NRB) consisting of 1x PBS, 1% Bovine Serum Albumin (BSA, Sigma; 2905-5GM) and 1% SUPERas-In followed by centrifugation at 500g, 4ºC for 5 mins. This wash step was repeated twice more for a total of 3 washes. The final pellet was resuspended in 0.5 mL of NRB containing 6 µM 4′,6-diamidino-2-phenylindole, dihydrochloride (DAPI, ThermoFisher; D1306). The suspension was filtered through a 20 µm filter (Sysmex; 04-004-2325) into a polypropylene tube and kept on ice.

Fluorescence Activated Nuclear Sorting (FANS) was performed to purify nuclei from debris on a FACSAria II (BD). Gates were set to isolate DAPI+ singlet nuclei based on forward scatter and side scatter as well as fluorescence intensity. The instrument was set to 40 pounds per square inch (psi) of pressure and a 70 µm nozzle was used, with sterile PBS sheath fluid. Nuclei were sorted into a 1.5 ml microcentrifuge tube containing 15 µl of NRB at 4ºC. For each sample, 18,000 events were sorted into the collection tube. The sorted nuclei and NRB total volume were approximately 45 µl, allowing for loading the entire suspension into the Chromium Single Cell 3' v3 solution (10x Genomics) without further manipulation. The 10x libraries were processed according to the manufacturer's instructions. Completed libraries were run on the Novaseq 6000 (Illumina).

Method Details

Alignment

Bcl files were converted to fastq format using the *cellranger mkfastq* command line tool. Fastq reads were then aligned to mmul10 genome and quantified into a raw UMI count matrix using the STAR aligner with the -solo option[74]. A custom transcriptome annotation that we previously developed was also given as input to the STAR aligner[22].

QC Filtering

Each biological replicate underwent a separate but standardized qc filtering, whereby empty droplets were removed using the *defaultDrops* function, which is a quantile filtering approach based on number of UMIs, with default parameters in the DropletUtils package[75]. Doublet filtering was also performed using the *cxds_bcds_hybrid* function, which uses a hybrid approach utilizing artificial generated doublets and unlikely coexpression pairs, with a cutoff of 1.0 in the SCDS package[76]. After both rounds of filtering 12,243 cells were left across all 3 replicates.

Normalization

The filtered digital gene expression (DGE) matrices were then loaded into Scanpy to perform the preclustering needed for Scran normalization[23,25]. Each biological replicate DGE was then normalized with log counts per million normalization, and denoised using iterative pca with 50 components. The actual preclustering was done via the Leiden community detection algorithm using Scanpy's Leiden function[77]. All 3 biological replicates were then co-normalized using the *computeSumFactors* function in Scran and *normalize* function in Scuttle with the preclusters as input[78]. The size factor normalization that was used was chosen for its ability to handle sparse single cell data which may have cell types with different expression values.

Integration of Neuronal cells and Clustering

Post normalization clustering was performed again using Scanpy's Leiden function[77]. The clusters were classified into oligodendrocytes 1 and oligodendrocytes 2, neurons, astrocyte 1 and astrocyte 2, oligodendrocyte precursor cells, microglia, meninges, ependymal cells and Schwann cells based on the expression of established marker genes. Neuronal clusters were then selected using the marker gene *RBFOX3*, which left a total of 2,698 cells across replicates. Data integration across biological replicates and neuron filtered gene expression matrices were performed using the integrate_scanpy function in Scanorama, which uses a panoramic stitching algorithm to integrated datasets by producing a batch corrected cell-cell distance graph. Leiden community detection was then performed with a resolution of 1.0

on the batch corrected cell-cell distance graph to get an initial integrated clustering[77]. It is important to note that while the Scanorama batch corrected data was used for clustering and integration, the 'un-corrected' data was used for marker gene and differential expression analysis to reduce possible bias that may have occurred during batch correction. Our initial clustering was organized into a dendrogram using ward hierarchical clustering. Further refinement of these clusters was performed by splitting clusters that had both excitatory and inhibitory markers (*SLC17A6* and *SLC32A1* respectively). We computed a midVentral and dorsal score for each cluster based on the average z-score scaled expression of genes that have a high discriminative ability (AUROC) between the spinal cord regions as computed by previous studies[16]. Two of the macaque clusters were thus grouped into a single cluster labeled midVentral. This cluster is excluded from our analysis of the dorsal horn populations as it lies outside our region of interest. The final list of dorsal horn clusters included 11 excitatory clusters and 5 inhibitory clusters.

Selecting Marker Genes

A combination of methods was used for the marker gene selection for the clusters. A binary matrix was constructed by thresholding the normalized gene expression values at a threshold of 0.2. The binary matrix was then used to compute the precision and sensitivity of each gene for each cluster. In addition, the top differentially expressed genes were computed using the *rank_gene_groups* function in Scanpy, and Wilcoxon rank-sum (Mann-Whitney-U) test. From this list, we removed genes that were co-expressed in non-neuronal cells if possible.

Comparison of Macaque and Mouse Expression Patterns

To identify orthologous genes across species, each mouse gene symbol was matched to the corresponding ENSEMBL ID using BioMart (v101). Then, BioMart was used to identify the orthologous human gene, filtering for one-to-one orthologs to avoid potential false positives. The macaque genome was annotated with orthologous human genes as previously described, motivated by previous successful approaches to annotate more complete gene structures using orthology to human[22,79]. The resulting datasets frame for macaque and mouse contained 15,157 one-to-one orthologs.

Orthologous genes of two of the mouse family markers, *RORB* and *ADAMTS5*, showed separation in the macaque UMAP, but were not independently clustered. GABA2 and GABA4 clusters were re-clustered using Leiden clustering (resolution 0.3)[77]. Three of the identified clusters represented *RORB* populations while the remaining two populations represented *ADAMTS5*. We split GABA2 and GABA4 clusters into two sub-clusters each, with GABA2_1 and GABA4_1 representing the RORB population and GABA2_2 and GABA4_2 representing ADAMTS5 populations.

A straightforward scoring approach was used for the cross-species comparisons to improve the interpretability. We identified the top 100 most enriched and most-depleted makers for each of the mouse cell classes (Excitatory, Inhibitory, and midVentral). Mouse and macaque markers for comparison were selected using the Wilcoxon rank sum approach in the "rank_genes_groups" function[25]. The relative levels of those markers, weighted by their cluster enrichment scores, created a score that we could use to annotate each cell in the macaque population for each of the major cell classes. These enrichment scores corresponded strongly with the individual marker-based annotation of the macaque cell classes. These scores were then used to identify significant shifts across the population using a t-test. The t-statistic confirms the broad differences in the distribution of the enrichment scores, with the exception of cluster GLUT11, which has features of midVentral neurons, but that we verify is highly similar to a specific sub-population of mouse excitatory neurons.

The same procedure was used to score each macaque cell for its cluster identify relative to its cell class (Excitatory or Inhibitory). We used the 50 most enriched and depleted markers, rather than 100, because the strength of enrichment declined more quickly down the ranked list for subtler differences that exist between subtypes of excitatory and inhibitory neurons in comparison to the strong differences that exist across those populations. Again, the annotation of cell clusters based on a t-test showed a strong match to annotation based on the most confident individual mouse markers.

To find genes specialized in a particular species, the mouse cells used were limited to the Sathyamurthy dataset which was obtained by snRNA-seq in adult mouse[16,34]. Rather than the processed data, where low

variance genes are removed, we used an unfiltered version of the mouse dataset. This allowed us to identify potential examples where the gene is specialized in macaque, but not mouse. Log fold differences were calculated using the procedure described above for both the mouse and macaque for each cell class relative to the others and for each family relative to the cell class. The family-level comparison was chosen to maximize the number of cells available for a rigorous identification in the differences in markers. We manually filtered out genes that exhibited a strong difference across species based on log2 fold change, but had very low abundance and were not significant. The orthologous marker genes of orthologous populations of cells were correlated with each other in their overall pattern (Pearson's R ranging from 0.058 to 0.21; p-value from $1*10^{-9}$ to $7*10^{-142}$). These correlations provide further support for our assignment of cell type families. The broad range reflected the relative abundances of the cell types rather than the lack of a strong match. To determine species-specific markers, we required a gene to have an adjusted p-value of $< 10^{-5}$ and a log2 fold difference of greater than 2. In addition, we required that the orthologous gene have a log2 fold difference of less than 0 in the other species. This highly stringent procedure may miss candidates, but the remaining ones are likely to be high quality.

Integration of Macaque and Mouse Single Nucleus RNA-Seq

Raw count matrices for macaque and adult mouse populations were read into R for analysis with the Seurat package[80]. We removed the nuclei annotated as mid-ventral to focus on the dorsal horn biology. As recommended in the LIGER package, the datasets from the mouse and three macaques were each scaled separated and integrated using the "RunOptimizeALS" function[28]. We performed a grid search across a small number of parameter values for lambda (5,7,10) and k (15,20,30,40). The parameters lamba=7, k=30 were chosen as it produced the fewest number of cells in mouse only or macaque only cells with default parameters. The dataset was further normalized according to species and sample and visualized using the UMAP visualization with default parameters.

RNAscope® in situ hybridization

Fresh frozen lumbar spinal cord tissue samples were harvested from 3-year-old male and female macaques perfused with aCSF. The tissue was immediately placed in OCT and frozen on dry ice. L4-L6 lumbar spinal

cord was sectioned using a cryostat at 20 µm thickness on to Superfrost-charged slides and stored in -80ºC until the start of the assay. In situ hybridization was performed according to the Multiplex v2 Fluorescent (Advanced Cell Diagnostics) protocol for fresh frozen tissue after fixing the slides with cold 4% paraformaldehyde (PFA) for thirty minutes. The probes were designed and purchased from Advanced Cell Diagnostics. Signal amplification was carried out using the TSA Fluorescin, Cyanine 3 and Cyanine 5 reagents from Akoya Biosciences at 1:1500. All sections were co-stained for dapi. For in situ hybridization experiments conducted with human lumbar spinal cord, samples were fixed with cold 4% paraformaldehyde (PFA) for 15 minutes. The Multiplex v2 Fluorescent (Advanced Cell Diagnostics) protocol for fresh frozen tissue was followed with a 2-minute protease IV digestion. The Fluorescin, Cyanine 3 and Cyanine 5 reagents from Akoya Biosciences were used for probe visualization.

Combinations:

GLUT1	Crhr2; Col13a1
GLUT2	Nmur2
GLUT3	Tac3; Nmu
GLUT4	Col5a2; Pax2
GLUT5	Col24a1; Nmbr
GLUT6	Tac1; Anos1
GLUT7	Col21a1; Maf
GLUT8	Mafa; Rbfox3
GLUT9	Adamts16; Tll2
GLUT10	Otogl
GLUT11 and GABA5	Pdyn; Pax2
GABA1	Masp1; Pax2
GABA2	Gdnf-as1; Rreb1
GABA3	Npy; Rreb1
GABA4	Met; Ptn

Image Acquisition and Quantification

For macaque spinal cord experiments, representative images were acquired at 10X magnification using the Nikon A1R and Nikon's NIS-Elements imaging software and processed with ImageJ. Images taken for quantitative analysis were acquired using Nikon Eclipse 800. Laminar boundaries for the images were drawn with the Canvas X software using Atlas of the Spinal Cord: Mouse, Rat, Rhesus, Marmoset and

Human as a reference. The substantia gelatinosa of the primate dorsal horn is also easily recognizable due to its translucent nature and hence was used to demarcate the boundary between laminae II and III.

For human spinal cord experiments, multiple 10x images were acquired of the dorsal horn starting from the substantia gelatinosa to lamina 10. The acquisition parameters were set based on guidelines for the FV3000 provided by Olympus. The 10x images were stitched together manually using anatomical landmarks (particularly lipofuscin) that were common between images.

For the quantitative analysis, three or more closely spaced puncta were counted as a positive cell. To account for the presence of lipofuscin in the macaque/human spinal cord tissue, the 488 channel was left blank i.e., no probe or fluorophore was added, and this was used for background subtraction.

References

1 Koch, S. C., Acton, D. & Goulding, M. Spinal Circuits for Touch, Pain, and Itch. *Annu Rev Physiol* **80**, 189-217, doi:10.1146/annurev-physiol-022516-034303 (2018).

2 Peirs, C. & Seal, R. P. Neural circuits for pain: Recent advances and current views. *Science* **354**, 578-584 (2016).

3 Rexed, B. The cytoarchitectonic organization of the spinal cord in the cat. *J Comp Neurol* **96**, 414-495, doi:10.1002/cne.900960303 (1952).

4 Al-Khater, K. M., Kerr, R. & Todd, A. J. A quantitative study of spinothalamic neurons in laminae I, III, and IV in lumbar and cervical segments of the rat spinal cord. *J Comp Neurol* **511**, 1-18, doi:10.1002/cne.21811 (2008).

5 Al-Chaer, E. D., Feng, Y. & Willis, W. D. Comparative study of viscerosomatic input onto postsynaptic dorsal column and spinothalamic tract neurons in the primate. *J Neurophysiol* **82**, 1876-1882, doi:10.1152/jn.1999.82.4.1876 (1999).

6 Christensen, B. N. & Perl, E. R. Spinal neurons specifically excited by noxious or thermal stimuli: marginal zone of the dorsal horn. *J Neurophysiol* **33**, 293-307, doi:10.1152/jn.1970.33.2.293 (1970).

7 Abraira, V. E. *et al.* The Cellular and Synaptic Architecture of the Mechanosensory Dorsal Horn. *Cell* **168**, 295-310 e219, doi:10.1016/j.cell.2016.12.010 (2017).

8 Light, A. R., Trevino, D. L. & Perl, E. R. Morphological features of functionally defined neurons in the marginal zone and substantia gelatinosa of the spinal dorsal horn. *J Comp Neurol* **186**, 151-171, doi:10.1002/cne.901860204 (1979).

9 Bai, L. *et al.* Genetic Identification of an Expansive Mechanoreceptor Sensitive to Skin Stroking. *Cell* **163**, 1783-1795, doi:10.1016/j.cell.2015.11.060 (2015).

10 Maxwell, D. J. & Bannatyne, B. A. Ultrastructure of muscle spindle afferent terminations in lamina VI of the cat spinal cord. *Brain Res* **288**, 297-301, doi:10.1016/0006-8993(83)90106-3 (1983).

11 Mears, S. C. & Frank, E. Formation of specific monosynaptic connections between muscle spindle afferents and motoneurons in the mouse. *J Neurosci* **17**, 3128-3135 (1997).

12 Lima, D., Avelino, A. & Coimbra, A. Morphological characterization of marginal (lamina I) neurons immunoreactive for substance P, enkephalin, dynorphin and gamma-aminobutyric acid in the rat spinal cord. *J Chem Neuroanat* **6**, 43-52, doi:10.1016/0891-0618(93)90006-p (1993).

13 Lima, D. & Coimbra, A. A Golgi study of the neuronal population of the marginal zone (lamina I) of the rat spinal cord. *J Comp Neurol* **244**, 53-71, doi:10.1002/cne.902440105 (1986).

14 Grudt, T. J. & Perl, E. R. Correlations between neuronal morphology and electrophysiological features in the rodent superficial dorsal horn. *J Physiol* **540**, 189-207 (2002).

15 Graham, B. A. & Hughes, D. I. Defining populations of dorsal horn interneurons. *Pain* **161**, 2434-2436, doi:10.1097/j.pain.0000000000002067 (2020).

16 Sathyamurthy, A. *et al.* Massively Parallel Single Nucleus Transcriptional Profiling Defines Spinal Cord Neurons and Their Activity during Behavior. *Cell Rep* **22**, 2216-2225, doi:10.1016/j.celrep.2018.02.003 (2018).

17 Haring, M. *et al.* Neuronal atlas of the dorsal horn defines its architecture and links sensory input to transcriptional cell types. *Nat Neurosci* **21**, 869-880, doi:10.1038/s41593-018-0141-1 (2018).

18 Zeisel, A. *et al.* Molecular Architecture of the Mouse Nervous System. *Cell* **174**, 999-1014 e1022, doi:10.1016/j.cell.2018.06.021 (2018).

19 Russ, D. E. *et al.* A Harmonized Atlas of Spinal Cord Cell Types and Their Computational Classification. *bioRxiv*, 2020.2009.2003.241760, doi:10.1101/2020.09.03.241760 (2020).

20 Rosenberg, A. B. *et al.* Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360**, 176-182, doi:10.1126/science.aam8999 (2018).

21 Delile, J. *et al.* Single cell transcriptomics reveals spatial and temporal dynamics of gene expression in the developing mouse spinal cord. *Development* **146**, doi:10.1242/dev.173807 (2019).
22 He, J. *et al.* Transcriptional Diversity of Medium Spiny Neurons in the Primate Striatum. *bioRxiv*, 2020.2010.2025.354159, doi:10.1101/2020.10.25.354159 (2021).
23 Lun, A. T., McCarthy, D. J. & Marioni, J. C. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res* **5**, 2122, doi:10.12688/f1000research.9501.2 (2016).
24 Hie, B., Bryson, B. & Berger, B. Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. *Nat Biotechnol* **37**, 685-691, doi:10.1038/s41587-019-0113-3 (2019).
25 Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* **19**, 15, doi:10.1186/s13059-017-1382-0 (2018).
26 Wang, F. *et al.* RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *J Mol Diagn* **14**, 22-29, doi:10.1016/j.jmoldx.2011.08.002 (2012).
27 Bakken, T. E. *et al.* Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. *PLoS One* **13**, e0209648, doi:10.1371/journal.pone.0209648 (2018).
28 Liu, J. *et al.* Jointly defining cell types from multiple single-cell datasets using LIGER. *Nat Protoc* **15**, 3632-3662, doi:10.1038/s41596-020-0391-8 (2020).
29 Das Gupta, R. R., Scheurer, L., Pelczar, P., Wildner, H. & Zeilhofer, H. U. Neuron-specific spinal cord translatomes reveal a neuropeptide code for mouse dorsal horn excitatory neurons. *Sci Rep* **11**, 5232, doi:10.1038/s41598-021-84667-y (2021).
30 Peirs, C. *et al.* Mechanical Allodynia Circuitry in the Dorsal Horn Is Defined by the Nature of the Injury. *Neuron* **109**, 73-90 e77, doi:10.1016/j.neuron.2020.10.027 (2021).
31 Albisetti, G. W. *et al.* Dorsal Horn Gastrin-Releasing Peptide Expressing Neurons Transmit Spinal Itch But Not Pain Signals. *J Neurosci* **39**, 2238-2250, doi:10.1523/JNEUROSCI.2559-18.2019 (2019).
32 Gatto, G. *et al.* A Functional Topographic Map for Spinal Sensorimotor Reflexes. *Neuron* **109**, 91-104 e105, doi:10.1016/j.neuron.2020.10.003 (2021).
33 Iwagaki, N. *et al.* A combined electrophysiological and morphological study of neuropeptide Y-expressing inhibitory interneurons in the spinal dorsal horn of the mouse. *Pain* **157**, 598-612, doi:10.1097/j.pain.0000000000000407 (2016).
34 Russ, D. E. *et al.* A harmonized atlas of mouse spinal cord cell types and their spatial organization. *Nat Commun* **12**, 5722, doi:10.1038/s41467-021-25125-1 (2021).
35 Boyle, K. A. *et al.* A quantitative study of neurochemically defined populations of inhibitory interneurons in the superficial dorsal horn of the mouse spinal cord. *Neuroscience* **363**, 120-133, doi:10.1016/j.neuroscience.2017.08.044 (2017).
36 Huang, J. *et al.* Circuit dissection of the role of somatostatin in itch and pain. *Nat Neurosci* **21**, 707-716, doi:10.1038/s41593-018-0119-z (2018).
37 Alaynick, W. A., Jessell, T. M. & Pfaff, S. L. SnapShot: spinal cord development. *Cell* **146**, 178-178 e171, doi:10.1016/j.cell.2011.06.038 (2011).
38 Lai, H. C., Seal, R. P. & Johnson, J. E. Making sense out of spinal cord somatosensory development. *Development* **143**, 3434-3448, doi:10.1242/dev.139592 (2016).
39 Del Barrio, M. G. *et al.* A transcription factor code defines nine sensory interneuron subtypes in the mechanosensory area of the spinal cord. *PLoS One* **8**, e77928, doi:10.1371/journal.pone.0077928 (2013).
40 Buckley, D. M., Burroughs-Garcia, J., Kriks, S., Lewandoski, M. & Waters, S. T. Gbx1 and Gbx2 Are Essential for Normal Patterning and Development of Interneurons and Motor Neurons in the Embryonic Spinal Cord. *J Dev Biol* **8**, doi:10.3390/jdb8020009 (2020).

41 Zou, M., Li, S., Klein, W. H. & Xiang, M. Brn3a/Pou4f1 regulates dorsal root ganglion sensory neuron specification and axonal projection into the spinal cord. *Dev Biol* **364**, 114-127, doi:10.1016/j.ydbio.2012.01.021 (2012).
42 Osseward, P. J., 2nd *et al.* Conserved genetic signatures parcellate cardinal spinal neuron classes into local and projection subsets. *Science* **372**, 385-393, doi:10.1126/science.abe0690 (2021).
43 Emery, E. C., Luiz, A. P. & Wood, J. N. Nav1.7 and other voltage-gated sodium channels as drug targets for pain relief. *Expert Opin Ther Targets* **20**, 975-983, doi:10.1517/14728222.2016.1162295 (2016).
44 Akopian, A. N., Sivilotti, L. & Wood, J. N. A tetrodotoxin-resistant voltage-gated sodium channel expressed by sensory neurons. *Nature* **379**, 257-262, doi:10.1038/379257a0 (1996).
45 Yildiz, S. H. *et al.* Assessment of Pain Sensitivity in Patients With Chronic Low Back Pain and Association With HTR2A Gene Polymorphism. *Arch Rheumatol* **32**, 3-9, doi:10.5606/ArchRheumatol.2017.5846 (2017).
46 Nicholl, B. I. *et al.* Association of HTR2A polymorphisms with chronic widespread pain and the extent of musculoskeletal pain: results from two population-based cohorts. *Arthritis Rheum* **63**, 810-818, doi:10.1002/art.30185 (2011).
47 Woo, S. H. *et al.* Piezo2 is the principal mechanotransduction channel for proprioception. *Nat Neurosci* **18**, 1756-1762, doi:10.1038/nn.4162 (2015).
48 Ranade, S. S. *et al.* Piezo2 is the major transducer of mechanical forces for touch sensation in mice. *Nature* **516**, 121-125, doi:10.1038/nature13980 (2014).
49 Foster, E. *et al.* Targeted ablation, silencing, and activation establish glycinergic dorsal horn neurons as key components of a spinal gate for pain and itch. *Neuron* **85**, 1289-1304, doi:10.1016/j.neuron.2015.02.028 (2015).
50 Nakagawa, T. *et al.* Expression of the pacemaker channel HCN4 in excitatory interneurons in the dorsal horn of the murine spinal cord. *Mol Brain* **13**, 127, doi:10.1186/s13041-020-00666-6 (2020).
51 Gutierrez-Mecinas, M. *et al.* Expression of Calretinin Among Different Neurochemical Classes of Interneuron in the Superficial Dorsal Horn of the Mouse Spinal Cord. *Neuroscience* **398**, 171-181, doi:10.1016/j.neuroscience.2018.12.009 (2019).
52 Smith, K. M. *et al.* Functional heterogeneity of calretinin-expressing neurons in the mouse superficial dorsal horn: implications for spinal pain processing. *J Physiol* **593**, 4319-4339, doi:10.1113/JP270855 (2015).
53 Hughes, D. I. *et al.* Morphological, neurochemical and electrophysiological features of parvalbumin-expressing cells: a likely source of axo-axonic inputs in the mouse spinal dorsal horn. *J Physiol* **590**, 3927-3951, doi:10.1113/jphysiol.2012.235655 (2012).
54 Hughes, D. I. & Todd, A. J. Central Nervous System Targets: Inhibitory Interneurons in the Spinal Cord. *Neurotherapeutics* **17**, 874-885, doi:10.1007/s13311-020-00936-0 (2020).
55 Pagani, M. *et al.* How Gastrin-Releasing Peptide Opens the Spinal Gate for Itch. *Neuron* **103**, 102-117 e105, doi:10.1016/j.neuron.2019.04.022 (2019).
56 Bardoni, R. *et al.* Pain Inhibits GRPR Neurons via GABAergic Signaling in the Spinal Cord. *Sci Rep* **9**, 15804, doi:10.1038/s41598-019-52316-0 (2019).
57 Gutierrez-Mecinas, M. *et al.* Expression of cholecystokinin by neurons in mouse spinal dorsal horn. *J Comp Neurol* **527**, 1857-1871, doi:10.1002/cne.24657 (2019).
58 Duan, B. *et al.* Identification of spinal circuits transmitting and gating mechanical pain. *Cell* **159**, 1417-1432, doi:10.1016/j.cell.2014.11.003 (2014).
59 Chen, S. *et al.* A spinal neural circuitry for converting touch to itch sensation. *Nat Commun* **11**, 5074, doi:10.1038/s41467-020-18895-7 (2020).

60 Huang, T. *et al.* Identifying the pathways required for coping behaviours associated with sustained pain. *Nature* **565**, 86-90, doi:10.1038/s41586-018-0793-8 (2019).
61 Shiers, S. I. *et al.* Convergence of peptidergic and non-peptidergic protein markers in the human dorsal root ganglion and spinal dorsal horn. *J Comp Neurol*, doi:10.1002/cne.25122 (2021).
62 Shiers, S. I. *et al.* Convergence of peptidergic and non-peptidergic protein markers in the human dorsal root ganglion and spinal dorsal horn. *J Comp Neurol* **529**, 2771-2788, doi:10.1002/cne.25122 (2021).
63 Hachisuka, J. *et al.* Wind-up in lamina I spinoparabrachial neurons: a role for reverberatory circuits. *Pain* **159**, 1484-1493, doi:10.1097/j.pain.0000000000001229 (2018).
64 Cameron, D. *et al.* The organisation of spinoparabrachial neurons in the mouse. *Pain* **156**, 2061-2071, doi:10.1097/j.pain.0000000000000270 (2015).
65 Spike, R. C., Puskár, Z., Andrew, D. & Todd, A. J. A quantitative and morphological study of projection neurons in lamina I of the rat lumbar spinal cord. *European Journal of Neuroscience* **18**, 2433-2448, doi:https://doi.org/10.1046/j.1460-9568.2003.02981.x (2003).
66 Han, Z. S., Zhang, E. T. & Craig, A. D. Nociceptive and thermoreceptive lamina I neurons are anatomically distinct. *Nat Neurosci* **1**, 218-225, doi:10.1038/665 (1998).
67 Andrew, D. Sensitization of lamina I spinoparabrachial neurons parallels heat hyperalgesia in the chronic constriction injury model of neuropathic pain. *J Physiol* **587**, 2005-2017, doi:10.1113/jphysiol.2009.170290 (2009).
68 Carstens, E. E., Carstens, M. I., Simons, C. T. & Jinks, S. L. Dorsal horn neurons expressing NK-1 receptors mediate scratching in rats. *Neuroreport* **21**, 303-308, doi:10.1097/WNR.0b013e328337310a (2010).
69 Wercberger, R. & Basbaum, A. I. Spinal cord projection neurons: a superficial, and also deep, analysis. *Current opinion in physiology* **11**, 109-115, doi:10.1016/j.cophys.2019.10.002 (2019).
70 Sardella, T. C. *et al.* Dynorphin is expressed primarily by GABAergic neurons that contain galanin in the rat dorsal horn. *Mol Pain* **7**, 76, doi:10.1186/1744-8069-7-76 (2011).
71 Sapio, M. R. *et al.* The Persistent Pain Transcriptome: Identification of Cells and Molecules Activated by Hyperalgesia. *J Pain*, doi:10.1016/j.jpain.2021.03.155 (2021).
72 Varga, A. *et al.* Spinal Excitatory Dynorphinergic Interneurons Contribute to Burn Injury-Induced Nociception Mediated by Phosphorylated Histone 3 at Serine 10 in Rodents. *Int J Mol Sci* **22**, doi:10.3390/ijms22052297 (2021).
73 LG, M. *'Frankenstein' protocol for nuclei isolation from fresh and frozen tissue for snRNAseq.* , 2020).
74 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
75 Lun, A. T. L. *et al.* EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biol* **20**, 63, doi:10.1186/s13059-019-1662-y (2019).
76 Bais, A. S. & Kostka, D. scds: computational annotation of doublets in single-cell RNA sequencing data. *Bioinformatics* **36**, 1150-1158, doi:10.1093/bioinformatics/btz698 (2020).
77 Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep* **9**, 5233, doi:10.1038/s41598-019-41695-z (2019).
78 McCarthy, D. J., Campbell, K. R., Lun, A. T. & Wills, Q. F. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**, 1179-1186, doi:10.1093/bioinformatics/btw777 (2017).
79 Warren, W. C. *et al.* Sequence diversity analyses of an improved rhesus macaque genome enhance its biomedical utility. *Science* **370**, doi:10.1126/science.abc6617 (2020).
80 Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* **33**, 495-502, doi:10.1038/nbt.3192 (2015).

81 Ouyang, J. F., Kamaraj, U. S., Cao, E. Y. & Rackham, O. J. L. ShinyCell: Simple and sharable visualisation of single-cell gene expression data. *Bioinformatics*, doi:10.1093/bioinformatics/btab209 (2021).

Figure Legends

Figure 1. Major Cell Types of the Macaque Dorsal Horn
A. Schematic overview of the single-nuclear RNA-sequencing (snRNA-seq) experimental workflow of the macaque lumbar dorsal horn. **B.** Pie chart representation of the contribution of each major cell type present in the macaque dorsal horn which includes oligodendrocytes, neurons, astrocytes, oligodendrocyte precursor cells (OPC), microglia, meninges, ependymal cells and Schwann cells. **C.** Heatmap of the normalized mean expression of key marker genes used to identify each major cell type. The size of the circles depicts the percentage of cells within each cluster that express the gene.

Figure 2. Classification of Macaque Dorsal Horn Neurons
A. UMAP visualization plot of the neuronal clusters broadly classified as either excitatory (green) or inhibitory (red) based on their neurotransmitter identity. **B.** Bar graphs of (top) the total number of nuclei that are excitatory (1,355 nuclei) or inhibitory (870) and (bottom) the number of nuclei per cluster. **C.** UMAP visualization of the 1,954 neuronal nuclei that are colored based on their cluster identity from Leiden clustering. Excitatory clusters are prefixed with GLUT and inhibitory clusters with GABA. **D.** Violin plots depicting the number of genes (top panel) and UMIs (bottom panel) in each neuronal cluster. **E.** Neuronal clusters were hierarchically clustered as shown in the dendrogram. Marker genes for each major branch point of the dendrogram are listed. E.g., Neurons belonging to the GLUT1 cluster express *SLC17A6*, *GFRA1*, *CRHR2* and *COL13A1*. **F.** Heatmap of the normalized mean expression of marker genes (mentioned in the dendrogram) for the neuronal clusters.

Figure 3. RNAscope® In situ Validation of Excitatory Cluster Marker Genes
Panels depict the in situ hybridization marker gene combinations used to detect each excitatory cluster. Dorsal horn images are taken at 10x with the smaller insets showing a magnified image (20x) of the individual gene(s) as well as the merged image. Laminar boundaries(dashed lines) are drawn between II/III, III/IV and IV/V. UMAPs of the marker genes used in the in situ are shown. Pie charts show the percentage of cells expressing the cluster marker genes in superficial (I-II) and deep (III-V) dorsal horn. Histograms show the number of cells positive for the cluster marker gene(s) binned into superficial or deep dorsal horn for each spinal cord section. n=5 or 6 spinal cord sections from N=2-3 macaques. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001, not significant (ns). Scale bar = 100 μm

Figure 4. RNAscope® In situ Validation of Inhibitory Cluster Marker Genes
Panels depict the in situ hybridization marker gene combinations used to detect each inhibitory cluster. Dorsal horn images are taken at 10x with the smaller insets showing a magnified image (20x) of the individual gene(s) as well as the merged image. Laminar boundaries (dashed lines) are drawn between II/III, III/IV and IV/V. UMAPs of the marker genes used in the in situ are shown. Pie charts show the percentage of cells expressing the cluster marker genes in superficial (I-II) and deep (III-V) dorsal horn. Histograms show the number of cells positive for the cluster marker gene(s) binned into superficial or deep dorsal horn for each spinal cord section. n=5 or 6 spinal cord sections from N=2-3 macaques. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001, not significant (ns). Scale bar = 100 μm

Figure 5. Cross-species Comparison of Marker Genes Across Mouse and Macaque
A. UMAP visualization of the macaque dorsal horn neurons colored and annotated based on the mouse families from Russ et al., 2021. **B.** UMAP showing dorsal horn clusters along with GABA2 and GABA4 subclusters (top left). Leiden re-clustering of GABA2 and GABA4 (top right).

Visualization of *RORB* and *ADAMTS5* gene expression within the GABA2 and GABA4 clusters (bottom). **C.** UMPA visualization of the neuronal nuclei isolated from dorsal horn of macaque (left) and mouse (right). The datasets were integrated using LIGER and then visualized according to batch (top), cell type class (middle) and cell type family (bottom).

Figure 6. Cross-species Comparison of Neuropeptides Across Macaque and Human
A. Representative human dorsal horn in situ images of excitatory neuropeptides – *TAC1*, *GRP*, *CCK*, *NTS*, *TAC3*, and *NMU* after background subtraction to remove lipofuscin (left). All images were taken at 10x. Scale bar = 100 µm. Histograms show total in situ hybridization counts of neuropeptides from macaque and human lumbar dorsal horn. n=6 hemisections from N=2-3 animals/donors. **B**. Representative human dorsal horn in situ images of inhibitory neuropeptides– *NPY*, *PDYN* and *PAX2* after background subtraction to remove lipofuscin (left). All images were taken at 10x. Scale bar = 100 µm. Histograms show total in situ hybridization counts of neuropeptides from macaque and human lumbar dorsal horn. n=6 hemisections from N=2-3 animals/donors.

Figure 7. Macaque and Mouse-specific Gene Enrichment
A. Correlation plots of genes specialized in the macaque excitatory clusters compared to the adult mouse excitatory clusters. Significant genes (black circles) within each cluster show p < 10-3 and log-fold-change > 1 relative to other clusters of the macaque. Among these, genes are species-specific or specialized (yellow circles) if they show enrichment scores of p < 10-5 and log-fold-change (lfc) > 2 (compared across other clusters of that species) and the orthologous genes in the other species shows lfc < 0 (compared across other clusters of that species). All significant genes are plotted. Refer to Supplementary Tables 3-4. **B.** Specialized genes in mouse excitatory clusters are in yellow. Significant genes (black circles) within each cluster show p < 10-3 and log-fold-change > 1 relative to other clusters of the mouse. All significant genes are plotted. **C.** Correlation plot of genes specialized in the macaque cluster GLUT8 (*PIEZO2* is enriched in the GLUT8 but not in the corresponding clusters which form the Cpne4 family of mouse (mac lfc = 3.6774378, mouse lfc = -0.46002). **D.** Correlation plot of genes specialized in the Cpne4 mouse family of clusters. *Cpne4* is enriched in the mouse but not in the corresponding GLUT8 cluster of macaque (mac lfc = -0.1435162, mouse lfc = 9.116846). All significant genes are plotted (black circles). **E.** Macaque (left) and mouse (right) dorsal horn in situ hybridization of *PIEZO2* (green). Representative images are 10x. Scale bar = 100 µm. **F.** In situ hybridization counts of PIEZO2 gene show expression in a higher number of cells in the superficial and deep dorsal horn laminae of macaque compared to mouse.

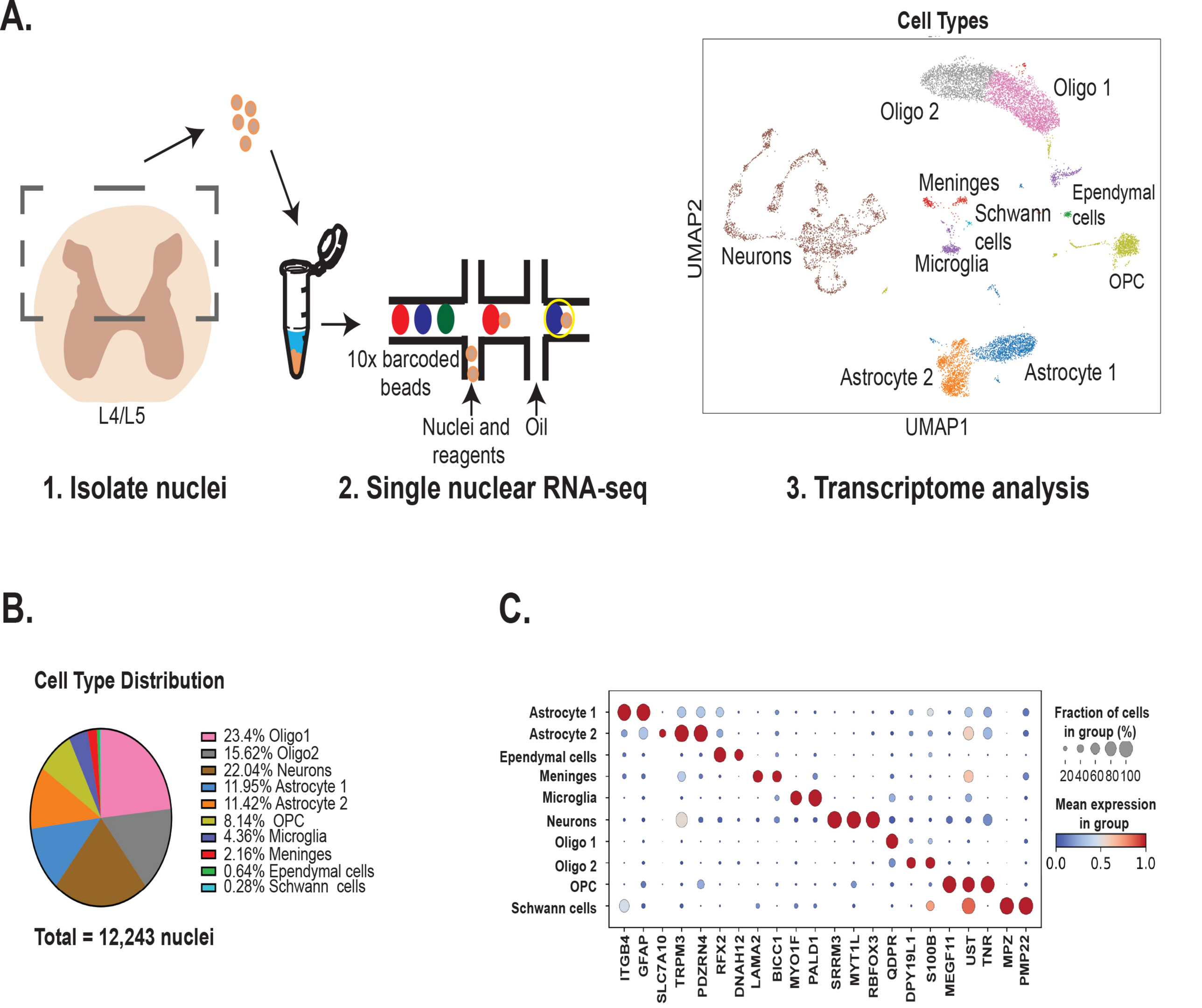

Figure 1. Arokiaraj et al

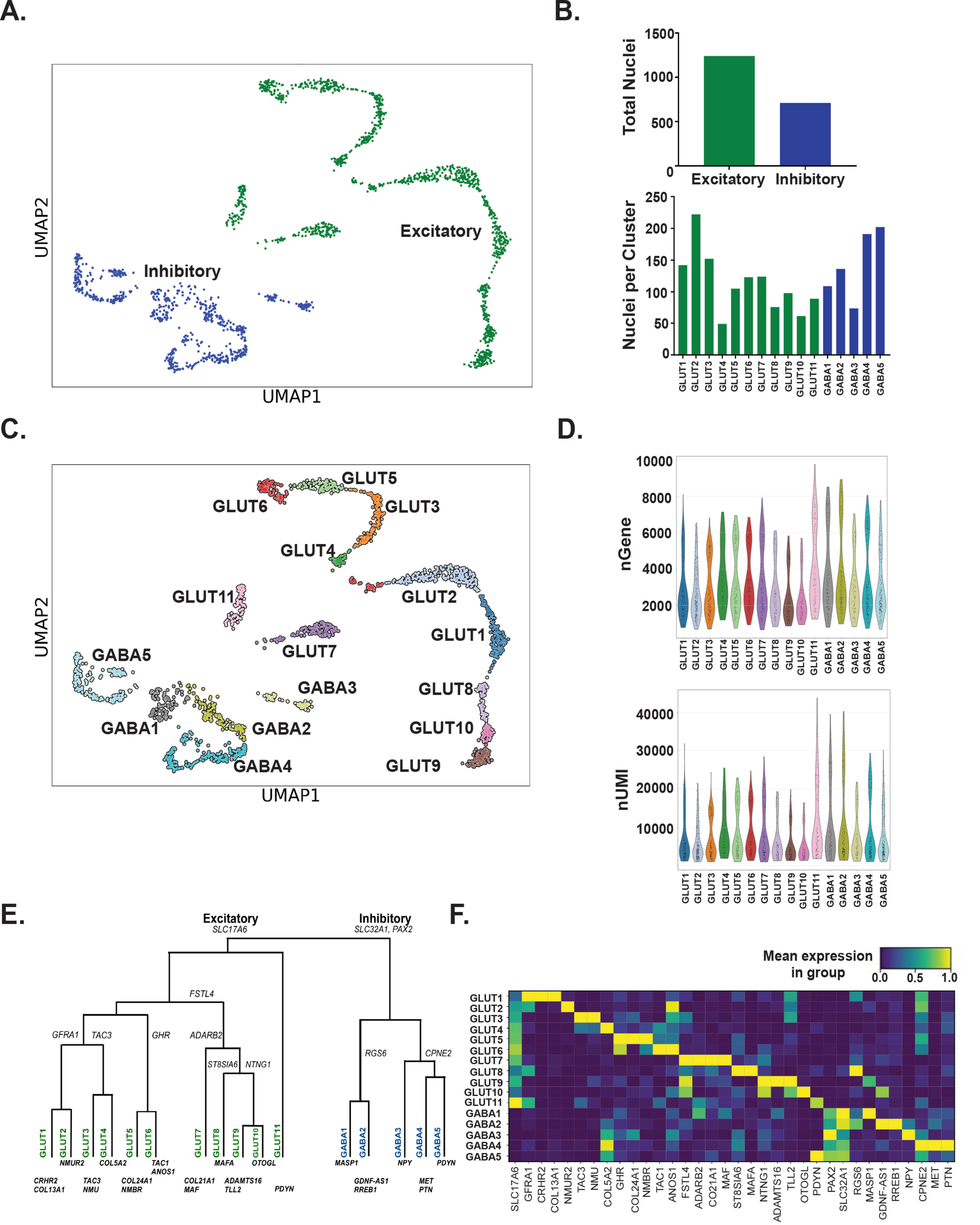

Figure 2. Arokiaraj et al

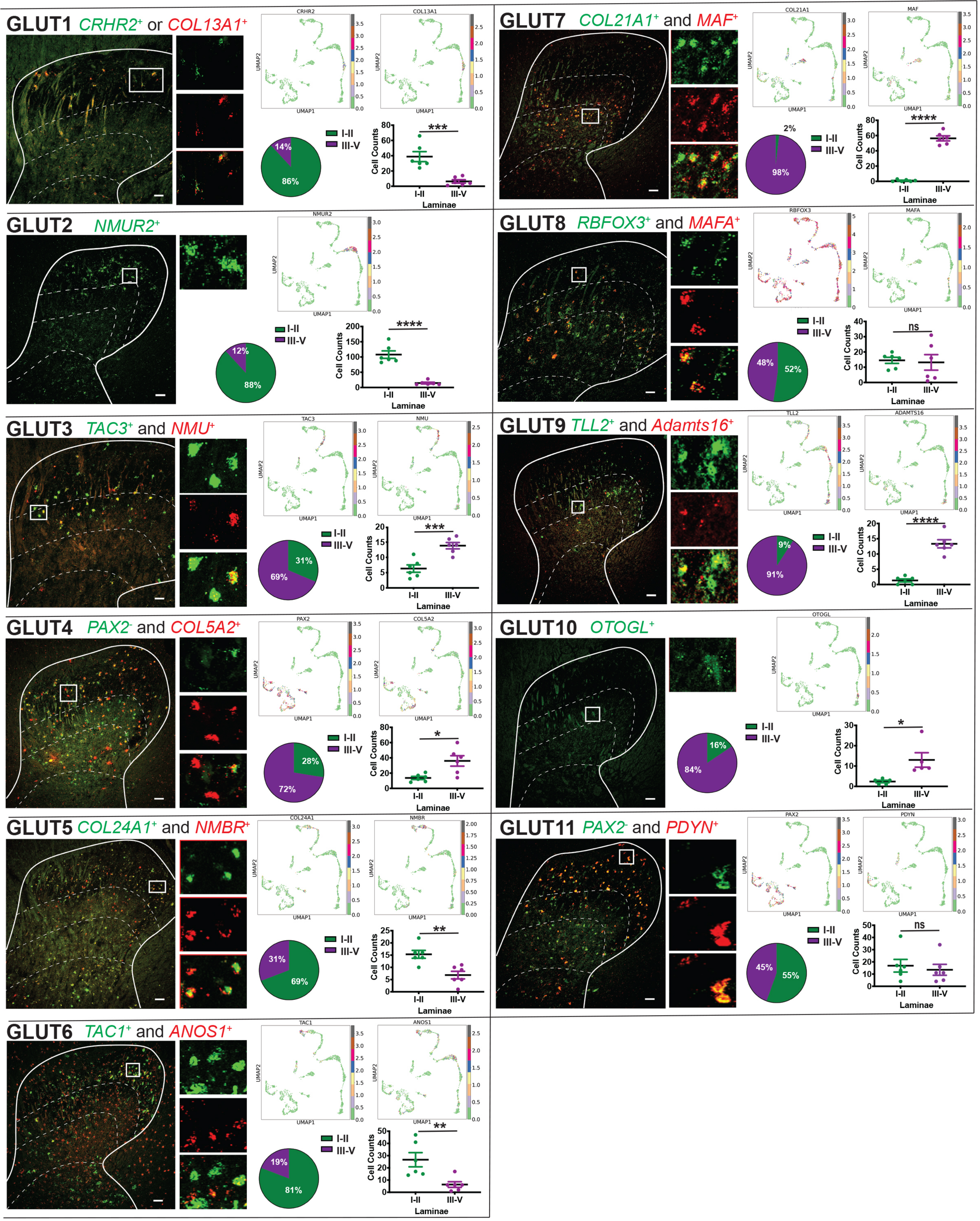

Figure 3. Arokiaraj et al

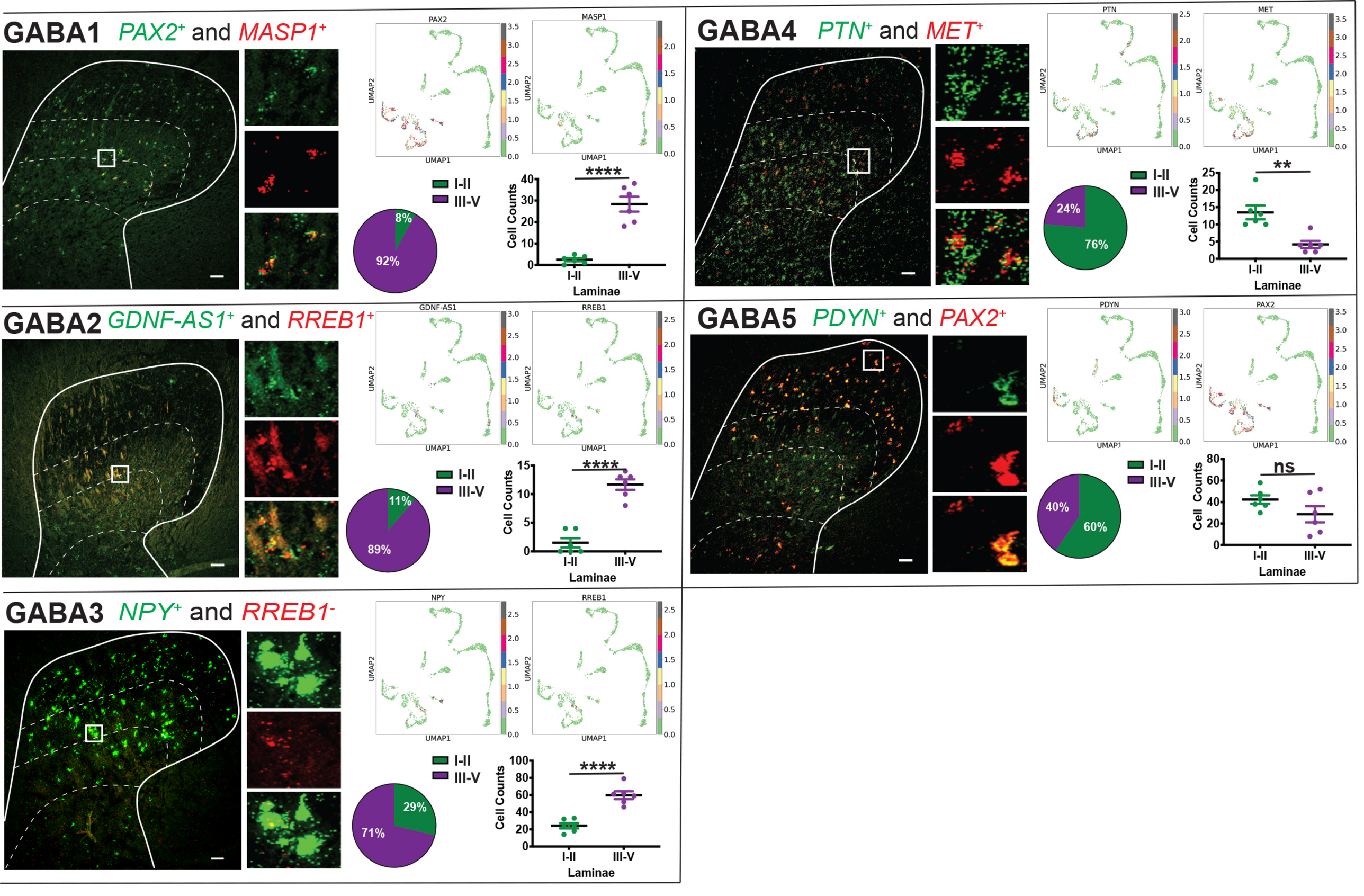

Figure 4. Arokiaraj et al

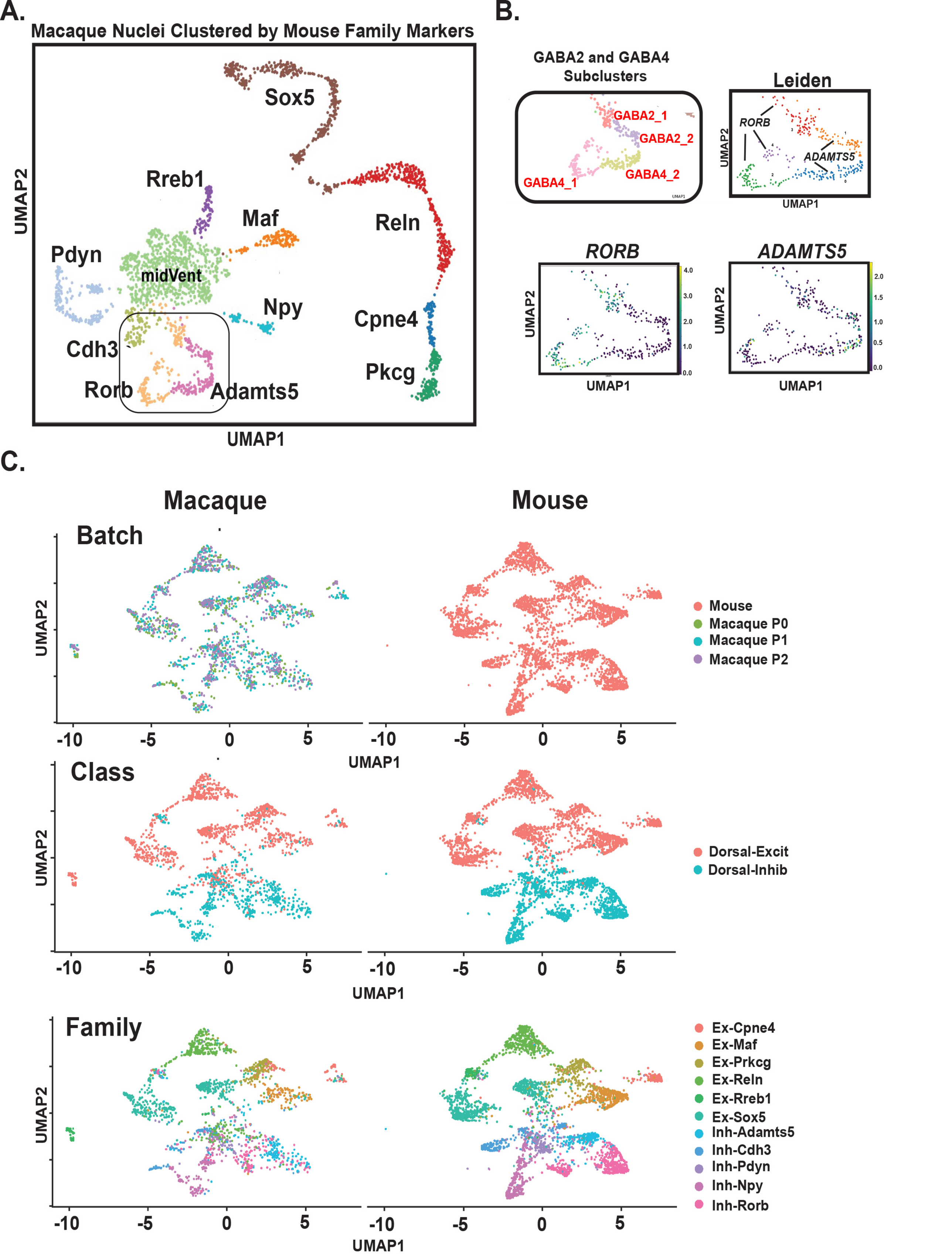

Figure 5. Arokiaraj et al

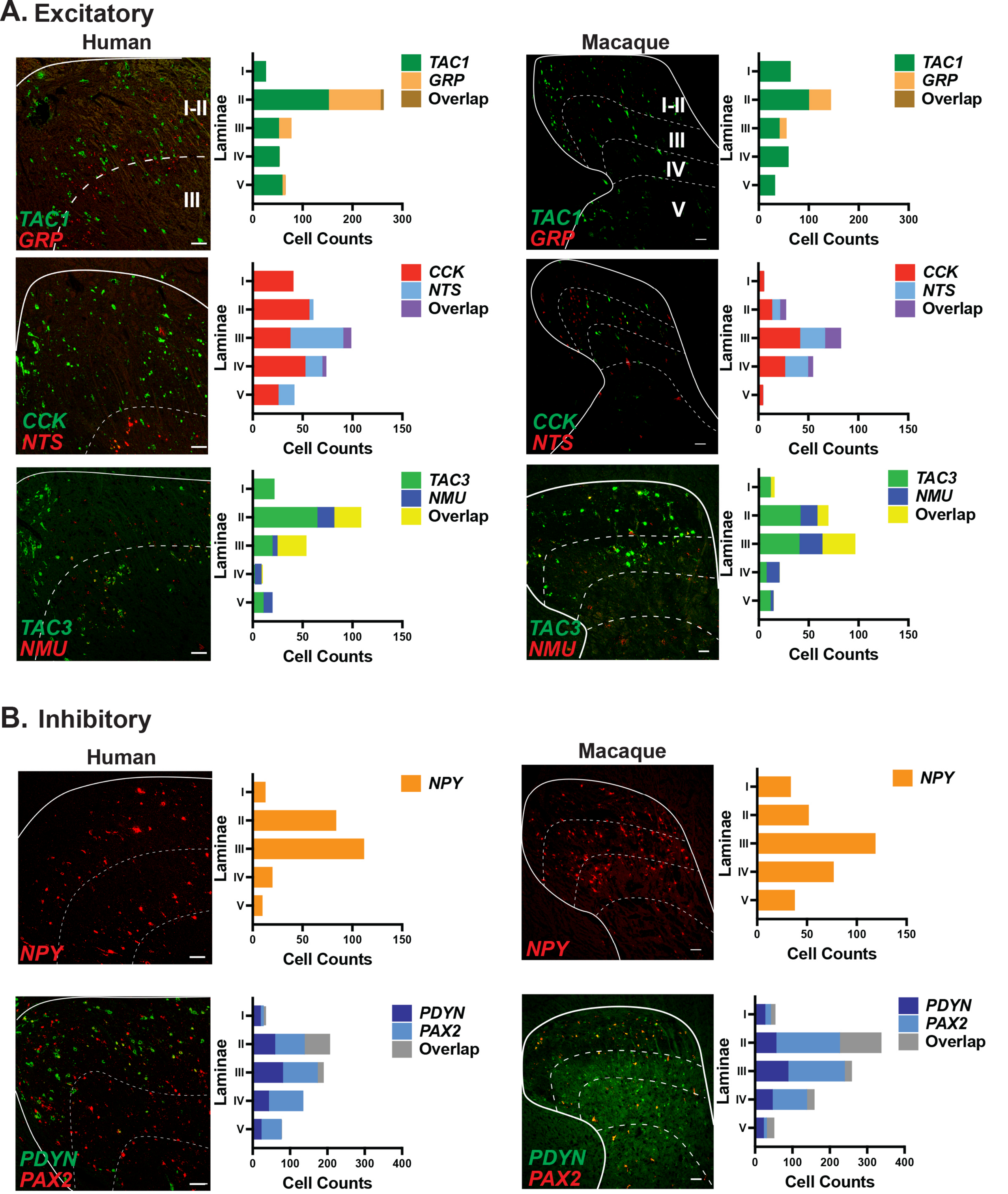

Figure 6. Arokiaraj et al

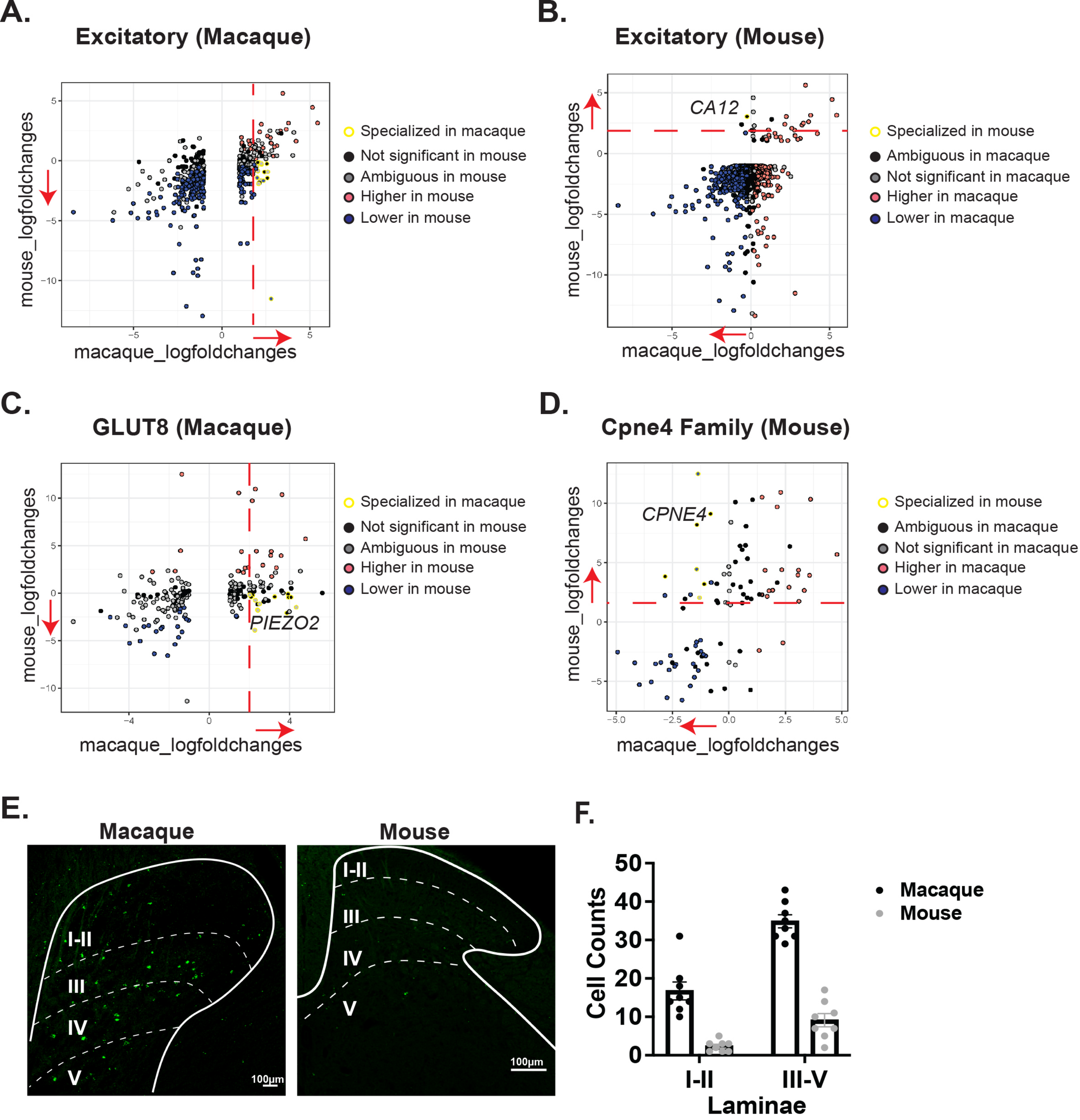

Figure 7. Arokiaraj et al