

## **Molecular development of muscle spindle and Golgi tendon organ sensory afferents revealed by single proprioceptor transcriptome analysis**

Katherine M. Oliver<sup>1,3</sup>, Danny M. Florez-Paz<sup>2,3</sup>, Tudor C. Badea<sup>4</sup>, George Z. Mentis<sup>1-3</sup>, Vilas Menon<sup>1</sup>,  
and Joriene C. de Nooij<sup>1,3,5,\*</sup>

<sup>1</sup>Dept. of Neurology, Vagelos College of Physicians and Surgeons, Columbia University Medical Center, 630 West 168<sup>th</sup> Street, New York, NY 10032.

<sup>2</sup>Dept. of Pathology, Columbia University Medical Center, 650 West 168<sup>th</sup> Street, New York, NY 10032.

<sup>3</sup>Columbia University Motor Neuron Center, Columbia University Medical Center, 650 West 168<sup>th</sup> Street, New York, NY 10032.

<sup>4</sup>Retinal Circuit Development & Genetics Unit, National Eye Institute, Bethesda, MD 20892, USA.

<sup>5</sup>Lead contact

\*Corresponding author: [sd382@cumc.columbia.edu](mailto:sd382@cumc.columbia.edu)

## Abstract

Anatomical and physiological analyses have long revealed differences between proprioceptive groups Ia, II, and Ib sensory neurons, yet the molecular correlates of these three muscle afferent subtypes remain unknown. We performed single cell RNA sequencing of genetically identified adult proprioceptors and, using unbiased bioinformatics approaches, detected five molecularly distinct neuronal clusters. Validation of cluster-specific transcripts in dorsal root ganglia (DRG) and skeletal muscle provides evidence these clusters correspond to functionally distinct muscle spindle (MS) or Golgi tendon organ (GTO) afferent proprioceptors. Remarkably, while we uncovered just one type of GTO afferents, four of the five clusters represent MS afferents, thus demonstrating a previously unappreciated diversity among these muscle proprioceptors. *In vitro* electrophysiological recordings reveal just two broadly distinct proprioceptor types, and suggest that the refinement of functional subtype diversity may occur along multiple axes of maturation. Lineage analysis between proprioceptor transcriptomes at different developmental stages show little or no correlation for transcripts that define adult MS or GTO afferents, supporting the idea that proprioceptor subtype identity emerges late in development. Together, our data provide the first comprehensive molecular signature for groups Ia and II MS afferents and group Ib GTO afferents, and offer new strategies for genetic interrogation of the role of these individual proprioceptor subtypes in regulating voluntary motor behavior.

# Introduction

Sensory feedback from muscle, skin, and joints is critical for the normal execution of voluntary motor tasks. Collectively referred to as the proprioceptive sense, this afferent information informs the central nervous system (CNS) on the position of the body and limbs in space (Sherrington, 1907; Proske and Gandevia, 2012). Many decades of study of the intramuscular sense organs and their afferent innervation have led to predictions on how proprioceptive feedback integrates with other sensory modalities to influence central motor circuits and calibrate motor output, yet most of these inferences await further corroboration in behaving animals (Pearson, 2004; Windhorst, 2007; Scott, 2016). At present, such studies remain challenging due to the lack of genetic access to individual proprioceptor subtypes.

Proprioceptive feedback derives in large part from specialized mechanoreceptive organs in skeletal muscle, with additional contributions from skin and joint receptors. Extensive anatomical and physiological analysis have revealed two types of muscle receptors: muscle spindles and Golgi tendon organs (Matthews, 1972). Muscle spindle (MS) mechanoreceptors are considered the main drivers for the sense of limb position and movement (kinesthetic sense) (Bastian, 1887; Goodwin et al., 1972, Proske and Gandevia, 2018). They are embedded within the belly of skeletal muscles and consist of encapsulated intrafusal muscle fibers that are typically innervated by one primary (group Ia) and several secondary (group II) proprioceptive sensory neurons (pSNs) (Figure 1A) (Matthews, 1972; Banks et al., 1982). Both types of afferents are responsive to stretch of the intrafusal fibers, such that voluntary or passive changes in limb position (i.e. muscle length) result in increased or decreased firing rates (Hunt, 1990). Unique among sensory organs, MSs are subject to (CNS-directed) efferent motor control through dynamic and static gamma motor neurons ( $\gamma$ MN) which innervate the contractile polar ends of the intrafusal muscle fibers and effectively set the gain for group Ia/II afferent discharge frequency (Crowe and Matthews, 1964; Macefield and Knellwolf, 2018). Despite supplying the same sensory end organ, group Ia and group II afferents exhibit distinct intra-spindle innervation patterns, activation thresholds, and conduction velocities, features that appear to render them biased to qualitatively different information of muscle stretch. For instance, while both group Ia and II afferents relay changes in static muscle length, the dynamic sensitivity of group Ia afferents enables them to also signal the rate of change in muscle length (Brown et al., 1967; Vincent et al., 2017). Presently, to what extent these properties are intrinsic to the neurons or to their association with different intrafusal muscle fibers, remains poorly understood.

In contrast to MSs, Golgi tendon organs (GTOs) are mechanoreceptive organs that are concerned with the sense of effort and respond to changes in muscle load (Jami, 1992). GTO receptors are located at the myotendinous junction of skeletal muscles, and are innervated by a single group Ib afferent (Barker, 1974; Jami, 1992). Group Ib afferents have similar activation thresholds and conduction velocities as group Ia MS afferents, but respond to different stimuli; they show little activation to muscle stretch but are highly sensitive to contractions of the motor units that are in series with the tendon organ they innervate (Houk and Henneman, 1967) (Figure 1A). Group Ib afferents also exhibit dynamic sensitivity and can signal rapid changes in contractile force (Davies et al., 1995). Due to their physiological similarities to Ia MS afferents, selective activation of group Ib GTO afferents remains challenging, leaving a number of uncertainties regarding their central pathways and physiological role in motor control.

Underscoring their functional differences, MS and GTO afferents engage with different local reflex circuits and projection neurons (Brown, 1981; Pierrot-Deseilligny and Burke, 2012; Shresta et al., 2012; Vincent et al., 2017). Most notably, MS afferents (and in particular group Ia afferents), but not GTO afferents, establish monosynaptic contacts onto homonymous  $\alpha$ MNs that innervate the same muscle target, while GTO afferents mostly inhibit homonymous MNs through di/tri-synaptic connections (Brown & Fyffe, 1979; 1981; Eccles et al., 1957, Laporte & Lloyd 1952). Most other spinal excitatory or inhibitory reflex circuits receive overlapping input from groups Ia, II, and Ib afferents yet are biased towards a given proprioceptor subtype or combination of these subtypes (Jankowska, 1992; Jankowska and Edgley, 2010; Pierrot-Deseilligny and Burke, 2012). Similarly, information of all three classes of afferents is relayed to higher brain circuits through projection neurons of the DSCT and VSCT but the influence of group Ia, II or Ib feedback appears to be weighted differentially across the different tracts (Shresta et al., 2012). Thus, the three proprioceptive muscle afferents subtypes (groups Ia, II, and Ib) innervate one of two anatomically and functionally distinct mechanoreceptive organs, exhibit diverse physiological properties and engage with different central targets. Undoubtedly, these proprioceptor subtype-specific features originate from the expression of distinct repertoires of transcription factors, receptor molecules, or ion channels. Yet, over a century and a half after these afferents were first described (Kölliker, 1826; Kühne, 1863; as referenced in Matthews, 1972), a definitive molecular signature for functionally distinct proprioceptive muscle afferents remains to be established.

Based on assumptions that MS and GTO afferent identities would differ at the stage of muscle target innervation (~e14.5), prior attempts to identify subtype-selective molecules for these proprioceptors typically involved molecular screens that focused on early developmental (embryonic/neonatal) stages (Lee et al., 2012; de Nooij et al., 2015). However, our recent studies indicate that molecular differences between proprioceptive muscle afferents may emerge only at later stages - well after MS and GTO afferents innervate their nascent receptor organs - and possibly through the influence of extrinsic signaling factors (Wu et al., 2019). Advances in next generation RNA sequencing (RNAseq) technologies have vastly expanded our general insight into the molecular diversity of adult somatosensory neuron classes, including proprioceptors (Usoskin et al., 2015; Li et al., 2016; Zeisel et al., 2018; Sharma et al., 2020). Nevertheless, since proprioceptors make up only ~10% of the total DRG population, under-sampling of this neuronal class has thus far precluded a molecular classification of MS or GTO afferent subtypes.

Here we devised an intersectional genetic strategy to selectively label and isolate muscle proprioceptors, thus enabling us to maximize the sampling density for single cell transcriptome analysis. Single cell RNAseq of adult proprioceptors and subsequent bioinformatics analysis identified five molecularly distinct neuronal clusters – two more than expected based on previous anatomical and functional characterization (e.g. groups Ia, II, and Ib). Validation of these neuronal clusters indicates that they translate into distinct proprioceptor subtypes observed *in vivo*. Specifically, proprioceptors marked by expression of *Calbindin 1*, *Calbindin 2 (Calretinin)*, and *vGlut2* correspond to group Ia MS afferents, while proprioceptors marked by expression of *Pou4f3* associate with group Ib GTO afferents. Based on the observation that *Pou4f3*<sup>+</sup> proprioceptors account for most (if not all) group Ib afferents, we conclude that the three remaining neuronal clusters comprise different types of group II MS afferents. This suggests that, unlike muscle force, kinesthetic and positional information is provided through four molecularly distinct MS afferent subtypes. Interestingly, *in vitro* electrophysiological recordings broadly distinguish between rapidly adapting (RA) and slowly adapting (SA) afferents and suggest that molecularly distinct MS and GTO proprioceptor subtypes may be classified differently when just considering their intrinsic physiological properties. To examine the molecular pathways through which distinct proprioceptor types emerge, we also performed single cell RNAseq analysis of proprioceptors at earlier developmental stages. Lineage analysis of transcripts that identify MS or GTO proprioceptor subtypes in the adult show weak (for p12) or no associations (p0, e14.5) with these transcripts at earlier developmental time points. These developmental analyses support the idea that proprioceptor subtype

identity is established at a late developmental stage, after the afferents innervate their nascent receptor targets in muscle.

Together, these data offer valuable new insights into the diversity of proprioceptive MS and GTO afferents and the molecular underpinnings of their physiological properties. Moreover, the identification of molecules that unambiguously delineate MS and GTO afferent subtypes offers a means to genetically dissect the relative contributions of these afferents in the execution of coordinated motor behavior.

## Results

### Intersectional genetic labeling of proprioceptive muscle afferents.

Proprioceptive sensory neurons in DRG can be identified by the co-expression of Parvalbumin (PV) and the Runt-domain transcription factor Runx3 (Rx3) (de Nooij et al., 2013; Wu et al., 2019). In contrast, the singular expression of these molecules marks subsets of low threshold cutaneous mechanoreceptive afferents with rapidly or slowly adapting response properties, respectively (de Nooij et al., 2013; Abaira and Ginty, 2013). To permit the selective isolation of PV<sup>+</sup>Rx3<sup>+</sup> proprioceptors from these other types of mechanoreceptors, we generated a *Rx3:FlpO* allele, thus enabling an intersectional genetic strategy in conjunction with a previously generated *PV:Cre* allele (Figure 1A, Figure S1A)(Hippenmeyer et al., 2005). In assessing the accuracy and efficacy of *Rx3:FlpO*-mediated reporter expression in the context of a FlpO-dependent GFP reporter mice (*RCE:FRT*; Sousa et al., 2009), we find that on average  $91.2 \pm 1.9$  % of Rx3<sup>+</sup> DRG neurons express GFP following FlpO-mediated excision of the transcriptional stop (n=6 L2 DRG) (Figure S1C-G). Consistent with the notion that Rx3 is initially more widely expressed in embryonic (e) sensory progenitors (e11-e13.5; Figure S1B), at postnatal day (p) 0 the number of GFP<sup>+</sup> neurons exceeded the number of Rx3<sup>+</sup> neurons ( $30.9 \pm 2.1$ % of Islet1<sup>+</sup>GFP<sup>+</sup> neurons lack expression of Rx3; n=6 DRG) (Figure S1G). We find, however, that these Rx3<sup>off</sup>GFP<sup>+</sup> neurons (in which GFP expression is activated by the transient developmental expression of Rx3) do not express PV, suggesting that this population does not include PV<sup>+</sup> cutaneous mechanoreceptors (Figure S1E). In addition, GFP reporter expression is not detected in spinal interneurons or motor neurons (Figure S1J). Likewise, aside from the expected proprioceptor sensory endings, within muscle only a few GFP<sup>+</sup> satellite cells were detected (Figure S1K). These data indicate that, at least within DRG, spinal cord, and muscle, the *Rx3:FlpO*

reporter faithfully reflects endogenous Rx3 expression, and presents a useful new tool that enables genetic access to Rx3<sup>+</sup> DRG sensory neurons.

We next examined the *Rx3:FlpO* allele in the context of a *PV:Cre* driver and the *Ai65:double stop:tdTomato*<sup>+</sup> reporter (hereafter *PVRx3:tdT*; Hippenmeyer et al., 2005; Madisen et al., 2015). In *PVRx3:tdT* mice, expression of tdTomato is expected to be limited to DRG neurons that express *both* PV and Rx3 at any one stage of their development (Figure 1A, Figure S2A-E). Indeed, we find that tdT invariably coincides with PV<sup>+</sup>Rx3<sup>+</sup> neurons at all segmental levels analyzed (T10, L2, L5), and is not observed in neurons that express only Rx3 or PV (Figure 1B). At p2, an average of 95.7 ± 1.3 % of PV<sup>+</sup>Rx3<sup>+</sup> neurons co-express tdT (n=18 DRG) (Figure 1B, 1C). Consistent with these observations, we find that tdT<sup>+</sup> afferents associate with MS and GTO sensory end organs in muscle (Figure 1D). Expression of tdT is, however, also observed in a few Rx3<sup>+</sup> neurons with very low levels of PV expression (PV<sup>low</sup>Rx3<sup>+</sup>; Figure S2C-F), possibly representing a different neuronal subset that either transiently expresses PV, or that express PV at levels much lower than typically observed in proprioceptive muscle afferents. Indeed, while expression of tdTomato in spinal cord is strictly limited to sensory afferent projections (i.e. no expression is observed in spinal neurons or microglia) we noted a few afferents that appear to project to deep dorsal horn targets, suggesting these could represent cutaneous low-threshold mechanoreceptive afferents (Figure S2G). To examine this in more detail, we investigated innervation of tdT<sup>+</sup> axons in the glabrous and hairy skin of the mouse forepaw. These analyses revealed that, in addition to muscle afferents, two types of skin projecting afferents appear to express both PV and Rx3 at some point during their development. The first of these correspond to Merkel cell afferents (characterized by their association with Troma1<sup>+</sup> Merkel cells) (Figure 1E, Figure S3B), which were previously known to express Rx3, but not PV (de Nooij et al., 2013). Intriguingly, we only observed tdT<sup>+</sup> Merkel cell afferents in forelimb skin and not in the skin of the back, possibly suggesting that their (transient) expression of PV may depend on regional restricted regulatory control. The second class of cutaneous tdT<sup>+</sup> afferents we detected appears to innervate the nail (Figure 1F, Figure S3C). The terminals of these neurons resemble slowly-adapting (SA) Ruffini endings that have previously been observed in the nailbed in human (Pare et al., 2003; Biszerks et al., 2009). In contrast, tdT expression was not observed in longitudinal lanceolate endings, or in afferents that associate with Meissner or Pacinian corpuscles (Figure S3D-I). Together these data demonstrate that use of the intersectional *PVRx3:tdT* reporter provides an efficient and nearly exclusive means to manipulate or isolate proprioceptive muscle afferents.

# Single cell RNA sequencing of proprioceptors reveals molecularly distinct subtypes

The ability to label pSNs in DRG, provides a means to significantly enrich these neurons for single cell transcriptome analysis without losing sequencing power to unrelated cell types. Considering the dynamic nature of proprioceptor subtype development (Wu et al., 2019), we chose to perform our initial analysis at an adult stage when the sensory system is believed to be fully mature. In addition, given that we expected the transcriptional differences between proprioceptors to be fairly limited (compared to differences between proprioceptors and nociceptors), we favored a high-depth method to detect mid- and low-expressed genes, as opposed to profiling more cells at lower depth (Menon, 2018). For this reason, we selected a plate-based sequencing platform for all our transcriptome analyses - a choice we believe was justified based on a post-hoc analysis of sequencing data obtained from adult proprioceptors sequenced through plate- and droplet-based sequencing platforms (Figure S4A, S4B). To enable an unbiased molecular analysis of proprioceptor transcriptomes, DRG from all segmental levels were obtained from adult ( $\geq$ p56) *PVRx3:tdT* mice. DRG were dissociated and single tdT<sup>+</sup> neurons were purified through fluorescent activated cell sorting (FACS), followed by plate sequencing (Figure 2A). Neurons (480 in total) were sampled from three different experiments and derived from animals of either sex (totaling four males and two females). Cells with low gene detection (<2,000 genes; 30 cells in total) or with significant contamination from attached satellite cells were eliminated by filtering for the satellite/Schwann cell markers *Apoe* and *Mpz* (cells with >10% of the *Apoe*/*Mpz* mean transcript level were removed from downstream analysis; 242 cells in total) (Figure S4C). After filtering out these contaminated cells, we detected a mean of 6,688 genes and 44,511 Unique Molecular Identifiers (a proxy for transcripts) per neuron for the remaining 208 neurons (Figure S4D, S4E). With the exception of differences in global reads, we uncovered no qualitative differences between plates, and no batch correction was necessary.

To determine any underlying structure in our proprioceptor population we employed an iterative clustering method to identify putative groupings of cells with similar transcriptional profile using an algorithm based on Tasic et al. 2018, and Baek et al. 2019 (see methods). These analyses resulted in the identification of five major clusters (Figure 2B, Figure S4F). Each of these clusters shows robust expression of canonical muscle proprioceptor markers, including *PV*, *Rx3*, *Ntrk3*, *Etv1* and *Whirlin* (Figure 2C; Figure S4F). We also identified four minor clusters (comprising fewer than ten cells), which we omitted from downstream analysis. One of these minor clusters does not express the pan proprioceptor marker *Whirlin*, suggesting that this cluster may correspond to some of the non-muscle



afferent contaminants such as the Merkel cell afferents or Ruffini-like endings (Figure 1E, 1F, Figure S3B, S3C, and data not shown) (de Nooij et al., 2015).

Considering that muscle proprioceptors are classified as either MS (group Ia, group II) or GTO afferents (group Ib), we anticipated that our transcriptome analysis would reveal two, or possibly three, molecularly distinct clusters. The discovery of five clusters led us to consider if other features of the proprioceptor phenotype (e.g. regional identity, flexor/extensor identity) could contribute to the clustering. To examine this, we determined the distribution of the expression of several known regional or muscle type molecules across the five clusters. With the exception of *Cadherin 13* (*Cdh13*) and to a lesser degree *Cortactin 1* (*Crtac1*), we find that many regional restricted molecules (e.g. Hox transcription factors) or transcripts associated with either dorsal or ventral muscle targets (e.g. *Sema5a*, *Vstm2b*, respectively), are randomly dispersed throughout our five main clusters (Catela et al., 2016; Poliak et al., 2016) (Figure S4F). Expression of *Cdh13* and *Crtac1* is biased to neonate proprioceptors that innervate dorsal (flexor) or ventral (extensor) limb muscle targets, respectively, but it remains unknown if this patterns is preserved in the adult (Poliak et al., 2016). These data suggest that the clustering revealed through our bioinformatics analysis is not driven by regional proprioceptor identity, but cannot exclude the possibility that some clusters may reflect some regional bias.

### Validation of identified pSN clusters

To investigate the biological relevance of the distinct pSN clusters, we first examined the features underlying the cluster distinctions. To do so, we performed an unbiased differential expression (DE) analysis between the neurons that comprise individual clusters (Robinson et al., 2010; McCarthy et al., 2012). These analyses revealed molecules that are present in either single or subsets of pSN clusters (Figure 2G). Interestingly, while many of our previously identified neonate proprioceptor subset markers show widespread expression across all five pSN clusters, the expression of three of these molecules appear biased towards certain clusters (Figure 2C-G, Figure S4G-J) (Wu et al., 2019). For example, expression of *Heart development protein with EGF like domains 1* (*Heg1*) is mainly associated with cluster 1, 2, and 3 neurons, while *Protocadherin 8* (*Pcdh8*) is prominently expressed in cluster 5 neurons, and is observed at lower level and frequency in clusters 2, 3, and 4 (Figure 2D, 2F, 2G). In addition, *Neurexophilin* (*Nxph1*) shows expression in clusters 2, 3 and 4, with little or no expression in clusters 1 and 5 (Figure 2E, 2G). Moreover, we find that the pattern of *Heg1*, *Pcdh8*, and *Nxph1* expression within the five bioinformatics-defined pSN clusters mirrors the combinatorial profile of these markers in adult

DRG (Figure 2G-J, Figure S4G). These data provide a first indication that our pSN clusters represent distinct proprioceptor types observed *in vivo*.

We next used RNA *in situ* hybridization (RNA scope) to assess the DRG expression pattern for several newly identified cluster-restricted molecules (Figure 3A). Specifically, we examined *i*) what proportion of tdT<sup>+</sup> pSNs express these markers, and *ii*) the extent of overlap with molecules that define other pSN classes. For these experiments, we first examined the expression of the cluster 1 markers *Hpse*, *Colq*, and *Agpat4* in adult *PVRx3:tdT* DRG (Figure 3B-G). Consistent with the proportion of cluster 1 neurons identified using bioinformatics approaches (51 out of 166), we find that these markers localize to ~20-40% of all tdT<sup>+</sup> neurons (Figure 3C, 3E, 3G). We also observed that *Hpse*, *Colq*, and *Agpat4* are generally co-expressed, but excluded from neurons with high levels of *Pcdh8* expression - a marker of cluster 5 pSNs (Figure 3B, 3D, 3F; Figure S5A). In contrast, expression of the cluster 5 markers *Itga2*, *Chad*, and *Pcdh17* generally coincides with high levels of *Pcdh8* expression (as well as with each other) (Figure 3H-K, Figure S5B, S5C). Expression of *Itga2*, *Chad*, and *Pcdh17* is observed in ~10-20% of all tdT<sup>+</sup> neurons, which is also in agreement with the proportion of cluster 5 neurons identified through our analysis (31 out of 166) (Figure 3I, 3K, Figure S5D). We next considered if any of these cluster-selective transcripts correspond to afferents that have different intra limb distributions. To do so, we examined the relative proportion of “marker<sup>+</sup>” pSNs in rostral lumbar DRG (L1-3; supplying body wall and proximal limb muscle targets) and caudal lumbar DRG (L4-5; supplying distal limb muscle targets) (S. Poliak, J. de Nooij, and T.M. Jessell, unpublished result). For each of the cluster 1 or 5 transcripts we examined, we find that their distribution is equivalent in rostral and caudal lumbar ganglia (Figure 3C, 3E, 3G, 3I, 3K, Figure S5D). We also examined the expression of cluster 2-4 markers *Cdh13* (cluster 4) and *Fam196a* (cluster 2 and 4); similar to cluster 1 and 5 transcripts, we find that both markers are expressed in subsets of pSNs (Figure 3L-O). In addition, within tdT<sup>+</sup> neurons, *Cdh13* often overlaps with expression of the pan-cluster 2-4 marker *Nxph1*<sup>+</sup> (Figure 3L). Together these data establish that our molecularly defined pSN clusters largely consist of non-overlapping proprioceptor subsets.

### **Molecularly defined pSN classes correspond to functionally distinct MS and GTO afferents**

The observation that our molecularly defined pSN clusters correspond to different proprioceptor subsets, prompted us to explore the biological relevance of these molecular distinctions. Specifically, we sought to determine if the neurons that comprise our different clusters align with any of the three known functional PSN subtypes: groups Ia and II MS afferents, or group Ib GTO afferents. We took

advantage of the observation that among the various class-specific transcripts we identified, we found a few markers for which validated immunological or genetic reagents are available. For class 1 neurons these transcripts are *Calbindin 2* (*Calb2*; encoding Calretinin), *Calbindin 1* (*Calb1*; encoding Calbindin), and the *vesicular glutamate transporter 2* (*Slc16a7*; encoding the vesicular glutamate transporter vGlut2), for class 4 neurons, *Tachykinin 1* (*Tac1*; encoding Neurokinin A and Substance P), and for class 5 neurons, the transcription factor *Pou4f3* (encoding Brn3c) (Figure 4A-F). To test if the expression of these molecules is restricted to either MS or GTO afferent proprioceptors, we examined their expression in adult muscle tissue, either directly (using specific antibodies), or indirectly (using genetic reporters) (Figure 4G-K).

We first focused our analysis on the class 1 molecules, Calbindin 1 (CB), vGlut2, and Calretinin (CR) (Figure 4G, Figure S6). While expression of all three markers is also observed in other DRG sensory neurons, we find that within muscle proprioceptors, expression of CB, CR, and vGlut2 is almost exclusively restricted to afferents that innervate muscle spindles (Figure 4G, Figure S6). Only occasionally did we observe expression of CB or vGlut2, but not CR, in GTO afferents (Figure S6, and data not shown). Moreover, these markers typically localize to afferents that innervate the equatorial region of the spindle, while afferents that innervate the polar contractile ends of the spindle are devoid of CR, CB, or vGlut2 expression (Figure 4G). Thus, within muscle proprioceptors, expression of CB, vGlut2, but especially CR, selectively mark group Ia MS afferents. Similarly, when using a *Calb2:Cre* allele in the context of the Cre-dependent *Mapt:loxP-STOP-loxP:GFP* reporter (hereafter *Calb2:GFP*) (Taniguchi et al., 2011; Hippenmeyer et al., 2005), we observed that GFP<sup>+</sup> terminals are almost invariably positioned in the equatorial region of the spindle (Figure 4H, Figure S6). Interestingly, some spindles lack a CR<sup>+</sup> or *Calb2:GFP*<sup>+</sup> Ia afferent. It is possible that this mosaic CR expression relates to technical issues given that muscle size, the density of connective tissue, or the spindle capsule may prevent adequate antibody penetration. Yet, even in extensor digitorum longus (EDL), a thin muscle in which most spindles are supplied by CR<sup>+</sup> Ia afferents, some spindles are devoid of a CR<sup>+</sup> Ia ending. The pattern of CR<sup>+</sup> Ia afferents in adult muscle raises the possibility that not all muscle spindles are innervated by a group Ia afferent, or alternatively, that there is further diversity among group Ia MS afferents. Consistent with the possibility of some degree of muscle bias, CR<sup>+</sup> pSNs are more prevalent in caudal than in rostral lumbar DRG (mean percentage  $\pm$  S.E.M. CR[L1-3]:  $11.6 \pm 2.3\%$ , n=24 sections, CR[L4-5]:  $23.9 \pm 1.9\%$ , n=40 sections) (Figure 4A, 4B).

We next assessed the afferent terminals of cluster 5 *Pou4f3*<sup>+</sup> neurons. In DRG, expression of *Pou4f3* is observed in a relatively large proportion of neurons, but labels just up to 25% of *PvRx3:tdT*<sup>+</sup> neurons (mean percentage  $\pm$  S.E.M. *Pou4f3*<sup>+</sup>,*TdT*<sup>+</sup> in T10:  $13.9 \pm 3.6\%$ , n=17 sections; in L5:  $25 \pm 3.2\%$ , n=78 sections) (Figure 4C, 4D). As for CR, the median number of *Pou4f3*<sup>+</sup>*tdT*<sup>+</sup> neurons per section is slightly higher at caudal (L4-5) than at rostral lumbar levels (L1-3), but for *Pou4f3* this difference does not reach statistical significance. To test if *Pou4f3*<sup>+</sup>*tdT*<sup>+</sup> neurons constitute a specific functional subtype of proprioceptors, we used a Cre-conditional *Pou4f3* mutant and chromogenic reporter (*Pou4f3*KOAP) in the context of a *PV:Cre* allele (Badea et al, 2011; Hippenmeyer et al., 2005). When heterozygous for both alleles, *PV:Cre* mediates excision of one *Pou4f3* allele and simultaneously activates expression of Alkaline Phosphatase (AP) (which remains under the control of endogenous *Pou4f3* promoter elements) (*PVPou4f3:AP*<sup>+</sup>). Consistent with previous observations that most DRG neurons express *Pou4f3* at early embryonic stages (Sharma et al., 2020), initially all pSNs (i.e. both MS and GTO afferents) are labelled by AP (Figure S7A-C). However, by p2, expression of AP is greatly diminished or absent in MS-innervating afferents but remains strongly expressed in GTO innervating afferents (Figure 4I, Figure S7A-C). We observed *PVPou4f3:AP*<sup>+</sup> afferents in all muscles we analyzed (including gluteus, soleus, EDL, plantaris, gastrocnemius, axial muscle) in a pattern that is consistent with GTO afferents in these muscle targets (Figure S7B, and data not shown). These data imply that within pSNs, expression of *Pou4f3* is selective for GTO afferents, and consequently, that cluster 5 neurons correspond to group Ib proprioceptors (Figure 4L). Consistent with these findings, we also observed *Pcdh8* expression in GTO afferents but not in MS-innervating afferents (Figure 4J, data not shown).

The observation that cluster 5 neurons appear to account for all group Ib GTO afferents suggests that the remaining cluster 2-4 neurons, collectively marked by expression of *Nxph1*, correspond to group II muscle afferents. Several lines of evidence support this idea. First, we and others previously showed that *Nxph1* expression in DRG is largely confined to proprioceptive sensory neurons (Wu et al., 2019; Sharma et al., 2020). Second, we showed that the cell bodies of *Nxph1*<sup>+</sup> neurons are smaller than the average *PV*<sup>+</sup> neuron cell body size (mean cell body diameter  $24.5 \pm 0.4 \mu\text{m}$  for *Nxph1*, n= 88 neurons;  $27.6 \pm 0.4 \mu\text{m}$  for *PV*, n= 183 neurons; p<0.001, Student's t-test), consistent with the notion that type II afferent neurons are smaller in caliber than groups Ia or Ib neurons, (Matthews, 1972; Wu et al., 2019). Third, based on morphological observations that MS are typically innervated by one group Ia and one to two group II afferents, the ratio of cluster 1 (group Ia) to cluster 2-4 (putative group IIs) we observed (1:1.53) aligns with the expected ratio for group Ia:group II afferents (1:1.5) (Banks et al., 1982; data not

shown; but see Sonner et al., 2017). To test if class 2-4 neurons represent group II afferents, we examined the expression of the class 4 marker *Tachykinin 1* (*Tac1*) in DRG and muscle. *Tac1* encodes Substance P and is typically associated with small diameter nociceptive neurons – not proprioceptors (Cao et al., 1998). When examining Substance P expression in DRG of *PVRX3:tdT* mice, we confirmed that ~10-15% of *tdT*<sup>+</sup> neurons co-label with Substance P (mean percentage  $\pm$  S.E.M. SubP at L1-3:  $13.8 \pm 2.5\%$ , n=33 sections, and at L4-5:  $9.0 \pm 1.6\%$ , n=11 sections) (Figure 4E, 4F). To assess the peripheral sensory endings of *Tac1*<sup>+</sup> pSNs in muscle, we used a *Tac1:Cre* allele crossed to the Cre-dependent *Ai14:tdTomato* reporter (hereafter: *Tac1:tdT*) (Harris et al., 2014; Madison et al., 2010). These experiments confirmed that *Tac1:tdT*<sup>+</sup> afferents can innervate the polar ends of MSs, suggesting they correspond to group II afferents (Figure 4K, Figure S8C). However, we also observed some *Tac1:tdT*<sup>+</sup> afferents that occupied the equatorial region of the spindle (indicating group Ia afferent identity), and noted numerous SubP<sup>off</sup>*tdT*<sup>+</sup> neurons in DRG (Figure S8B). These data suggest that some of the spindle innervating afferents (including group Ia afferents) may be labeled by the transient activation of *Tac1* expression in pSNs at earlier developmental stages. Nevertheless, these data show that transcriptionally diverse proprioceptor muscle afferents correspond to functionally distinct MS and GTO afferent subtypes, and reveal a previously unappreciated diversity among spindle afferents.

### **Intrinsic and circuit properties of MS and GTO afferents revealed by differential expression patterns of ion channels and neurotransmitter receptors**

The ability to correlate our molecularly defined pSN classes with functionally distinct groups Ia, II, and Ib muscle afferents offers an opportunity to examine these pSN subtypes with respect to their subtype-selective expression of transcription factors, receptor molecules, and ion channels (Figure S9). (An interactive web application to search for individual genes in proprioceptor subtypes will be made available at xxxx). GO analysis demonstrates that the gene expression differences between the pSN clusters relate mostly to categories involving ion channel activity or synaptic membranes (Figure S9A). However, a differential analysis of all voltage-gated sodium (Nav) or calcium (Cav) channels across the different proprioceptor classes revealed little variation in the expression of these channel subunits (Figure 5A, Figure S10A). This finding may reflect the fact that these channels are a common feature of the general proprioceptor identity. One notable exception is *Scn7a*, which encodes the atypical Na<sub>x</sub> channel (also known as Nav2.3, NaG), and appears selectively enriched in group Ib (class 5) and presumptive group II (class 4) afferents (Figure 5A) (Felts et al., 1997; Watanabe et al., 2000). In contrast to the relative uniform expression of Nav and Cav channels, we noted a much larger diversity with

respect to voltage-gated potassium (Kv) channels and their auxiliary  $\beta$  subunits (Figure 5B, Figure S10B). This diversity was apparent for individual pSN classes but also between group I afferents (Ia and Ib) and group II afferents. For instance, the A-type current modulator Kchip1 (*Kcnp1*) is upregulated in group I afferents relative to putative group IIs, while Kchip2 (*Kcnp2*) is more prevalent in group II compared to group I afferents (Jerng et al., 2004) (Figure 5B). Similarly, Kv7.3 (*Kcnq3*) is upregulated in group I afferents but shows little expression in group II afferents. Group I afferents also express higher levels of the Kv1.1 (*Kcna1*) and Kv1.2 (*Kcna2*) channels than group II afferents (Figure 5B).

Considering that potassium channels feature prominently in regulating neuronal firing properties (Jan and Jan, 2012; Tian et al., 2014), we wondered if the differences in their expression patterns would be reflected in the electrical responses of pSNs to maintained or repeated stimuli. To address this question, we performed whole-cell patch-clamp recordings on isolated adult pSNs (identified using the *PVRx3:tdT* reporter) (Figure 5C). Indeed, sustained current injections revealed two major firing patterns (Figure 5D-H). Nearly half of the neurons exhibited rapid adaptation and fired a single action potential (AP), even at large current injections (45%; 10/22 neurons) (Figure 5D-F). Most of the remaining neurons exhibited a tonic firing pattern (41%; 9/22 neurons), with firing frequencies up to  $53.95 \pm 7.7$  Hz (Figure 5D, 5E, 5H). A few neurons (14%; 3/22 neurons) displayed a small burst of APs with increasing currents (Figure 5D, 5G), suggesting that they may present a slightly less adapting variant of the single AP group. We note that these physiological analyses appear at odds with previous whole cell *in vitro* recordings of isolated PV<sup>+</sup> neurons (the majority of which are presumed to be pSNs). These prior studies detected either rapidly adapting (RA) (Woo et al., 2015) or slowly adapting (SA) neurons (Zheng et al., 2019) but not both. Possible explanations for these discrepancies include differences in the temperature at which the recordings were performed, the duration of culture time following dissociation, and/or differences in the addition of growth factors during culture (Woo et al., 2015; Zheng et al., 2019). Indeed, in our hands, rapidly adapting (RA) and tonic firing neurons were only observed when recordings were conducted at 37°C, after 24hrs of recovery at 37°C, and without addition of neurotrophins to the culture media, while recordings at room temperature only revealed rapidly adapting (74% of neurons) and burst responses (26% of neurons) (data not shown).

We next determined if the two main pSN firing patterns we observed correlate with other morphological or electrical features of the recorded neurons. While we did not observe significant differences between soma size across the two groups ( $851.4 \pm 3.2 \mu\text{m}^2$  for single AP,  $975.57 \pm 140.86$

$\mu\text{m}^2$  for burst firing, and  $817.9 \pm 6.4 \mu\text{m}^2$  for tonic firing), tonic neurons tended to be slightly smaller than RA neurons (Figure 5I). We find that the resting membrane potential for tonic neurons was significantly more depolarized when compared to RA neurons (Figure 5J). In addition, while capacitance, input resistance, and voltage threshold were not statistically different among the three groups, rheobase was statistically lower in tonic neurons (Figure 5K, 5L, 5M; Figure S11A). Since the shape of the action potential also results from the specific complement of ion channels, we also measured AP amplitude and half width. We find that AP amplitude was statistically larger in tonic neurons compared to single AP neurons, but we observed no difference in AP half width (Figure S11B-D). In addition, repetitive stimulation revealed no difference between the different neurons, and both RA and tonic neurons were able to sustain 1:1 firing up to nearly 140 Hz (Figure S11E). Based on previous observations that groups Ia and Ib afferents have dynamic (RA) response properties, we postulate that RA neurons correspond to groups Ia and Ib proprioceptors, while tonic neurons correspond to group II afferents (Hunt et al., 1990; Jami et al., 1992; Vincent et al., 2017). These observations support the idea that differential patterns of Kv ion channel subunit expression may underlie differences in firing properties between group I and group II proprioceptors. By extension, these data indicate that molecular distinct proprioceptor subtypes may be classified differently when considering their intrinsic physiological properties, with group Ia MS and group Ib GTO afferents demonstrating more resemblance to one another than groups Ia and II MS afferents.

Besides intrinsic group I and group II afferent-specific physiological features, our pSN transcriptome profiling offers insight into how MS and GTO afferent subtypes may operate at the circuit level. For example, we find that the five identified pSN classes exhibit different patterns of GABA<sub>A</sub> receptors: while all afferents express *Gabra2*, group Ia MS afferents co-express *Gabra3* and *Gabra5*, presumptive group II MS afferents co-express *Gabra3*, and group Ib GTO afferents co-express *Gabra1* and *Gabra5* (Figure 6A, 6C-H). Considering that proprioceptive muscle afferents are subject to central regulation through presynaptic inhibition (Rudomin, 2009), the differential expression of these *Gabra* subunits could offer a mechanism to regulate this central influence on MS and GTO afferents separately. In addition, we note that group Ib GTO afferents, but not groups Ia or II MS afferents, express appreciable mRNA levels of the G-protein coupled NPY receptor *Npy1r*, and the NMDA receptor *Grin3a* (Figure 6B, 6I-L). Similar to *Gabra* receptor units, the differential expression of these molecules in pSN subsets may offer new insight into proprioceptor engagement or inhibition within spinal circuits.

## Emergence of pSN subtype identity

To reveal when proprioceptive groups Ia, II, and Ib afferent identities first emerge, or the molecular mechanism that underlies their development, we also performed single cell RNAseq at key time-points during proprioceptor development. These stages included e14.5, when proprioceptors innervate their nascent receptor targets, p0, when animals are born (and first begin to apply goal-directed movement), and p12, when animals are able to rear themselves and walk. Similar as for adult proprioceptors, we used the intersectional *PVRx3:tdT* genetic reporter to isolate p0 and p12 pSNs (Figure 7A). Since PV is expressed at low levels at early embryonic stages, to efficiently isolate e14.5 pSNs we used a *TrkC-tdTomato* allele (Li, et al., 2011; hereafter *TrkC:tdT*), which labels all  $Rx3^+$  neurons at this stage (including  $PV^+Rx3^+$  pSNs and  $PV^{off}Rx3^+$  Merkel cells afferents) (Figure 7A). For all age groups, neurons were obtained from at least two different animals, and similar to the adult dataset, neurons were assessed for quality and were filtered for satellite cell contaminants. The total number of neurons with sufficient coverage (>2000 genes detected) and low contamination (<10% of mean Apoe, Mpz contaminant transcript counts) were 169 for e14.5, 55 for p0, and 154 for p12 (Figure S12A-E). Cluster analysis of all datasets combined demonstrated a nearly complete segregation by developmental stage, suggesting that maturation-related transcripts dominate the clustering over any other potential proprioceptor markers (Figure 7B). Unsupervised clustering of the individual e14.5, p0, p12 cellular transcriptomes was performed by applying the same iterative clustering method as described for adult neurons (see methods). These analyses revealed four clusters for the e14.5 neurons, three clusters for p0, and five clusters for the p12 neurons (similar as for adult) (Figure 7C-E). For the e14.5 data set however, two clusters (C3 and C4) show a near absence of expression of *ETV1* and *PV*, and an increased level of *Maf* expression (Figures S12F, S13A). These data suggest that these two e14.5 clusters represent subsets of non-proprioceptive *TrkC:tdT*<sup>+</sup> neurons, including Merkel cell afferents (de Nooij et al., 2013). Together these findings indicate that pSNs progressively become more diverse (from two to five clusters) during development.

To assess to what extent the genes that define pSN classes in the adult correlate with gene expression patterns at earlier developmental stages, we next performed a lineage analysis, focusing on transcripts that identified MS or GTO proprioceptor subtypes in the adult. Genes with strong co-expression patterns in adult clusters have weaker associations at p12, and essentially no associations at earlier time points (with exception of a few GTO afferent markers at p0) (Figure 7F). These data suggest that the expression patterns seen in adult pSN classes is not anticipated by the transcript patterns observed at earlier time points, but rather results from a dynamic interaction of lineage-defining genes.



Consistent with these observations, we find that expression of CR is not observed in Rx3<sup>+</sup>PV<sup>+</sup> pSNs at p0 (Figure 7G, 7H). Instead, CR only approaches its adult pattern of expression at p12 (Figure S13B). Similarly, Substance P-expressing pSNs are rare at p0 and only reach their adult expression pattern at p12 (Figure 7I, 7J, Figure S13C). In contrast, *Pou4f3* is expressed in many if not all DRG neurons at early stages and becomes confined to a subset of pSNs only later (Figure 7K, 7L, Figures S13A). These data indicate that the differential maintenance of *Pou4f3* expression is one of the first molecular distinctions between MS and GTO afferents.

The observation that expression of *Pou4f3* may predict pSN GTO subtype identity between e14.5 and p0 led us to consider if other transcription factors may anticipate the MS – GTO lineage selection at this developmental stage. To address this question, we performed a differential expression analysis at e14.5 and p0, focusing on the set of transcription factors that show elevated levels of expression in at least one class at either age group (Figure S14). These analyses again demonstrate elevated levels of *Pou4f3* expression in the p0-C2 cluster (suggesting they represent GTO afferents), but show little distinction in *Pou4f3* expression between the two pSN clusters (C1 and C2) observed at 14.5. Similarly, while this analysis revealed other transcription factors that are differentially expressed between MS and GTO afferents at p0 (e.g., *Id1*, *Id2*, *Id3*, *Tcf15*, *Tcf19*, *Crip1*, *Zfp235*, *Tcf7l2*), with the exception of *Zfp235*, none of these transcription factors are differentially expressed at e14.5 (Figure S14). Conversely, transcription factors that are differentially expressed between the C1 and C2 groups at e14.5 (*Cas21*, *Cux2*, *Pou4f1*, *Spry2*, *Spry4*), are equally distributed across all three p0 clusters. These data reinforce the idea that proprioceptive muscle afferents remain uncommitted during their early trajectory towards their peripheral target, and that transcriptional MS/GTO subtype identities emerge over a prolonged developmental period that extends after the afferents innervate their nascent receptors (Wu et al., 2019).

## Discussion

Proprioceptive muscle feedback is relayed through muscle spindle and GTO sensory afferents, yet little is known about the diversity among these afferents beyond the broad distinctions between their receptor and central targets. Here, we used single cell RNA sequencing and transcriptome analysis of genetically identified adult, adolescent, neonatal, and embryonic muscle proprioceptors to provide new and long overdue insights into the molecular differences between these afferent subtypes. While previous transcriptome studies have revealed general proprioceptor markers based on comparisons to other somatosensory modalities, our unbiased analyses selectively focused on proprioceptive muscle

afferents. This enabled us to identify proprioceptor subtype-selective transcripts that otherwise might exhibit more widespread expression in DRG. Apart from uncovering MS and GTO afferent-selective markers, our transcriptome analyses revealed a previously unappreciated level of MS afferent diversity, provides new insights into the intrinsic physiological properties of proprioceptors, and reinforces the idea that proprioceptor subtype identity may emerge through extrinsic signals.

### Functional organization of proprioceptive feedback

Using transcriptome analysis of single proprioceptors, we identified molecular markers for each of the main pSN subclasses, including groups Ia and II MS afferents, and group Ib GTO afferents. Specifically, among many other differentially expressed molecules, we demonstrate that expression of *Calretinin*, *Calbindin 1*, and *vGlut2* marks group Ia MS afferents, while expression of *Pcdh8* and *Pou4f3* is observed in group Ib GTO afferents. We also provide evidence that *Nxph1* is a general marker for group II MS afferents. Remarkably, instead of the three predicted clusters (groups Ia, II, and Ib), we identified two additional clusters of *Nxph1*<sup>+</sup> pSNs that we presume correspond to group II afferents. These results reveal a larger diversity in MS afferent subtypes than previously appreciated. The functional significance of this afferent diversity remains unclear, but may relate to the different requirements for afferents that relay positional information from axial, body wall or limb muscles. Indeed, we observed increased numbers of *Calretinin*<sup>+</sup> neurons in DRG that innervate limb muscle targets, while *Tac1*<sup>+</sup> neurons are more prevalent in ganglia that preferentially innervate proximal limb/body wall/axial muscle targets. It is also possible that some *Nxph1* neurons correspond to a separate set of group Ia afferents, or may constitute a hybrid class of group Ia/II afferents (given that some muscle spindles appear to lack innervation by *Calretinin*<sup>+</sup> afferents). In this context it is interesting to note that *in vivo* electrophysiological recordings have revealed classes of MS afferents that do not fully fit the firing behaviors expected for either group Ia or group II afferents, but instead exhibited mixed Ia/II response properties with some group Ia's lacking vibration sensitivity and some group II's showing dynamic sensitivity (Vincent et al., 2017). Our data indicate that the electrophysiological variability observed *in vivo* could in part have a molecular basis. We expect that future genetic strategies targeting individual MS afferent subtypes, beyond providing insight into their overall role in motor control, will reveal the biological relevance of this additional subtype diversity in the organization of body and limb position sense.

## Intrinsic properties of proprioceptor subtypes

A second surprising element of our studies concerns the identity of the transcripts that we find to define the different adult proprioceptor classes. Several of these molecules have not before been associated with a proprioceptor identity (*Hspe*, *Colq*, *Chad*, *Agpta4*), while other markers have typically been observed in sensory neurons other than proprioceptors (e.g. *vGlut2*, *Tac1*) (Zhang et al., 2018; Lallemend and Ernfors, 2012). Especially the latter observation suggests that identification of sensory neuron subtypes on the basis of one or two molecular markers is an unreliable strategy, and similar as observed for other neuronal classes, requires ‘multi-transcript authentication’ (Bikoff et al., 2106; Tasic et al., 2018, Shresta et al., 2018). Our data indicate that *vGlut2*, together with Calretinin and Calbindin, marks group Ia MS afferents. The expression of Calretinin and Calbindin group Ia MS afferents is not surprising given that both molecules are often observed to regulate  $\text{Ca}^{2+}$ -homeostasis in highly active neurons (Kovacks-Oller et al., 2019). However, the observation that adult group Ia MS afferents MS, in addition to *vGlut1*, also express *vGlut2* was unexpected and may have important implications for studies that assess inputs to spinal neurons on the basis of the presumed differential expression of these two glutamate transporters between proprioceptors and spinal neurons (Betley et al., 2009; Shakya-Shrestha et al., 2012). Why group Ia afferents would require both *vGlut1* and *vGlut2* remains as of yet unresolved. A similar intriguing observation is the expression of *Tac1* in a subclass of presumed group II afferents. *Tac1* encodes the neuropeptides Neurokinin A and Substance P, and represents one of the hallmarks of a nociceptive sensory identity (Cao et al., 1998). The expression of *Tac1* in group II MS afferents may hint at a direct role for muscle afferents in inflammatory muscle pain. Besides these MS afferent markers, the biological role for most newly identified MS or GTO afferent molecules remains largely unknown, let alone their relevance to pSN function. Thus, a better understanding of the biological activities of these molecules in general likely will offer new insights into proprioceptor development and/or function.

Aside from novel molecular markers, our transcriptome data offers valuable insights into the molecules that control the intrinsic electrical properties for individual proprioceptor subtypes. In contrast to voltage-gated sodium and calcium channel subunits, we detected a large degree of variability in the expression of voltage-gated potassium channels across the five pSN classes. These include differences between individual proprioceptor subtypes, as well as global differences between group I and group II afferents. Indeed, whole cell patch clamp recordings of pSNs revealed two main firing patterns upon current injection: rapidly adapting or slowly adapting. Based on prior observations,

we postulate that rapidly-adapting (RA) neurons may comprise group Ia or Ib afferents, while slowly-adapting (SA) tonic neurons may correspond to group II afferents (Hunt, 1990; Jamie, 1972; Vincent et al., 2017). While these afferent firing patterns may result from differences in several types of ion channels, we observed higher levels of Kv1 channel subunit expression in group I (class 1 and 5) afferents compared to group II afferents. This is of note because Kv1 subunits previously were implicated in regulating the firing properties of A $\beta$  SA- and A $\beta$ -RA afferents: afferents with lower Kv1 levels demonstrated tonic firing patterns, while those with higher levels adapted rapidly (Zheng et al., 2019). The tonic firing we observed for presumed group II afferents is consistent with observations that group II's generally provide information regarding static muscle length and have little dynamic sensitivity (Matthews, 1972; Hunt, 1990; Vincent et al., 2017). Most remaining pSNs (and presumed group I afferents) in our *in vitro* studies exhibited rapidly adapting responses, which seemingly is at odds with *in vivo* observations that group Ia and group Ib afferents relay both static and dynamic information about muscle length or contraction, respectively, but are reminiscent of the firing patterns described for Merkel cell afferents. (Hunt, 1990; Jamie, 1992). Merkel cell afferents exhibit rapidly and slowly adapting firing upon Merkel cell stimulation *in vivo*, but only show a rapidly adapting response to current injection in *in vitro* recordings (Woo et al., 2017; Maksimovic et al., 2017). Thus, the SA-adapting responses observed for groups Ia and Ib afferents *in vivo* may be due to receptor-based modulation (similar as proposed for Merkel cell afferents), possibly through sustained glutamatergic activation at the sensory terminal (Maksimovic et al., 2017; Bewick et al., 2005). Another notable observation in these *in vitro* recordings is that presumed groups Ia MS and Ib GTO afferents appear to have similar firing properties and vastly differ from group II MS afferents. This indicates that despite the molecular distinctions between these three proprioceptor subtypes, their classification follows different rules dependent on what neuronal feature is being assessed. Current injection experiments do not expose the full repertoire of electrical responses, however, and future studies may yet reveal physiological distinctions between groups Ia and Ib afferents that reflect differences in ion channel expression patterns.

### **Emergence of MS and GTO afferent identities**

Molecules that distinguish MS afferents from GTO afferents have thus far remained elusive. The inability to uncover such markers in previous studies may possibly be attributed to the long held conception that proprioceptor subtype identity is already established prior to pSN engagement with their peripheral receptors, thus targeting the search for such molecules to early developmental stages. Indeed,

expression of Calretinin (CR) was previously noted in group Ia muscle afferents in the chicken and rat (Duc et al., 1993; 1994). Yet our past studies consistently failed to observe CR in embryonic or neonatal mouse proprioceptors, letting us to conclude that CR presented a species-specific difference between chick, rat and mouse. We now confirm that CR is not expressed in embryonic or early postnatal proprioceptors and only reaches its mature pattern of expression by p12. Similarly, we find that expression of *Tac1*, a presumptive group II marker, only reaches its mature pattern of expression by p12. Both observations are supported by our longitudinal transcriptome analysis of e14.5, p0 and p12 proprioceptors and together imply that distinctions between groups Ia and II MS afferent identities only emerge during postnatal development. Interestingly, in previous studies we demonstrated that a disruption of normal muscle spindle development results in the failure to induce or maintain expression of *Heg1* (Tourtellotte and Milbrandt, 1998; Wu et al., 2019), a molecule we now find to be restricted to proprioceptors that innervate group Ia and II muscle spindles. *Nxph1* was unaffected in animals that lack spindles, indicating that spindle-derived signals may inform some but not all aspects of MS afferent identity. This suggests that other qualitative or quantitative differences in mesenchymal, muscle, or tendon signals, and/or differences in activity levels may determine the full complement of muscle spindle afferent phenotypes.

Transcriptional differences between MS and GTO afferents emerge at an earlier developmental stage, and can be observed as early as p0. At this stage, expression of the transcription factor *Pou4f3* is selectively maintained in prospective group Ib GTO afferents. However, we find no evidence that *Pou4f3*<sup>+</sup> neurons derive from a specific subset of e14.5 progenitors, indicating that the *Pou4f3*-group Ib lineage relationship first appears during the e14.5/p0 interval, after the afferents reach their nascent receptor targets. The absence of a clear lineage relationship between mature and early embryonic proprioceptors is observed for a broad panel of pSN subtype-selective molecules and parallel a recent study indicating that the transcriptomes that define adult somatosensory neuron subclasses bear little to no resemblance to the transcriptomes of early embryonic (e11.5) somatosensory neurons (Sharma et al., 2020). The notion that proprioceptor subtypes emerge relatively late in development reinforces the idea that MS or GTO afferent subtype selection is influenced by extrinsic signals (Wu et al., 2019; Sharma et al., 2020). The initial broad and overlapping expression patterns of many transcription factors within embryonic somatic sensory neurons, and their apparent regulation by extrinsic signals, may confer additional flexibility to respond to subtle changes in peripheral target tissues; differences in transcription factor expression levels can be rapidly amplified through a multitude of transcriptional

targets. The role of peripheral signals in the regulation of neuronal identities may be a general feature of sensory systems given that they interface between the external environment and the brain. Within the context of the proprioceptive sensory system such a mechanism may have evolved to facilitate adaptations to specific functional demands on the motor system. As such, a better understanding of the molecular mechanism by which Pou4f3 expression is selectively maintained in GTO afferents but not MS afferents should provide general insights into the developmental logic of sensory neuron subtype selection.

Taken altogether, these findings provide important new insights into the organization, molecular basis, and development of MS and GTO afferent proprioceptor subtype identities, and should form an important foundation for genetic studies aimed to further understanding of the role of these afferents in coordinated motor control.

**Acknowledgements:** We thank Barbara Han, Monica Mendelsohn, and Nataliya Zabello (Columbia Zukerman Mind and Brain Behavior Institute) for help with generating the *Runx3:FlpO* mice, David Ginty (Harvard) for providing the *TrkC:tdTomato* mice, Mike Kissner (Columbia Stem Cell Institute Flow core) for help with FACS plate sorting, Erin Bush (Columbia JP Sulzberger genome center) for help with single cell RNA sequencing, and Siddarth Bejugama for help with some of the initial RNA *in situ* analyses. We also greatly appreciate the time and effort of Niccolo Zampieri, Peter Sims, Eiman Azim, and Marcela Carmona, for providing comments on the manuscript. This research used the Genomics and High Throughput Screening Shared Resource, funded in part through the NCI Cancer Center Support Grant (NIH P30CA013696). In addition, G.Z.M. was supported by NINDS (NIH R01NS078375), The NIH Blueprint for Neuroscience Research, NIAA, NINDS (NIH R01AA027079), The SMA Foundation, and Target-ALS; T.C.B. was supported by NEI through the Intramural Research Program; V.M. and J.C.N. were supported by NINDS (NIH R01NS106715 to J.C.N.), and by the Thompson Family Foundation Initiative (J.C.N.).

**Author contributions:** J.C.N. conceived of the study, designed and performed all experiments with help from K.O., and analyzed data. D.F.P. and G.Z.M. performed electrophysiological experiments and analyzed data. T.C.B. provided valuable unpublished reagents. V.M. analyzed single cell RNA sequencing data. J.C.N. wrote the manuscript with input from all other authors.

**Declaration of Interests:** The authors declare no competing interests.

## Figure legends.

### Figure 1. An intersectional genetic labeling strategy for proprioceptive muscle afferents.

A. Schematic rendering of the groups Ia, II and Ib proprioceptor subtypes labeled in mice carrying a *PV:Cre*, *Rx3:FlpO*, and the double stop *Ai65:tdTomato* allele (*PVRx3:tdT*). In sensory neurons that co-express PV and Rx3, the dual activity of Cre and FlpO recombinases will allow expression of the tdTomato reporter.

B. Expression of Runx3, PV and tdTomato in DRG at p2 indicates efficient induction of the tdTomato reporter in  $PV^+Rx3^+$  sensory neurons, but not in  $PV^+Rx3^{off}$  (arrow head in *ii*), or  $PV^{off}Rx3^+$  (arrow in *ii*) neurons (*i* and *ii* are independent images).

C. Efficiency of tdTomato reporter induction in T10, L2, and L5 DRG in p2 *PVRx3:tdT* mice. Mean percentage  $\pm$  S.E.M. of  $tdT^+PV^+Rx3^+$  neurons (of all  $PV^+Rx3^+$ ) per ganglia for T10:  $98.0 \pm 1.3\%$ , for L2:  $94.4 \pm 2.5\%$ , for L5:  $94.8 \pm 1.9\%$ ; n=6 DRG per segmental level.

D. tdTomato expression in vGlut1<sup>+</sup> muscle spindle (MS) and Golgi tendon organ (GTO) afferent terminals in p10 EDL muscle of *PVRx3:tdT* animals. Diagram depicts muscle with red boxes showing areas in images below.

E, F. Expression of tdTomato in Merkel cell afferents present in fore paw glabrous skin (E), and in sensory neurons that appear to innervate the nail of fore paw digits (F) of adult ( $\geq p56$ ) *PVRx3:tdT* mice. Diagram depicts foot paw with red boxes showing areas in images below. In E, Merkel cells are indicated by expression of Troma1. (*i*, *ii*) panels show higher magnifications of boxed areas in E, F; (*ii*) panel shows tdTomato expression.

Scale 20 (B), or 50 (D-F)  $\mu m$ .

### Figure 2. Single cell transcriptome analysis of $PV^+Rx3^+$ neurons.

A. Schematic of single cell RNA sequencing strategy employed in experiments. DRG from adult ( $\geq p56$ ) *PVRx3:tdT* mice were collected and dissociated into single cell suspensions.  $tdT^+$  SNs were isolated using FACS. To ensure cell viability, neurons were also assessed for the presence of high Calcein Blue (CalB) and absence of Sytox red (SytxR) labeling while sorting individual neurons into a single well of a 96-well plate containing lysis buffer (see methods for details).

B. t-stochastic neighbor embedding (tSNE) plot of adult cells, colored by cluster membership. Clustering was done using a gene module and bootstrapping approach (with the hicat package – see methods).

C-F. tSNE plots of cells colored by expression of four genes. In each plot, the color represents no expression (grey) to maximum counts per million (CPM) value over all the cells (red).

G. Barplots showing a subset of differentially expressed genes across the five main clusters. For each gene, the maximum CPM value over all cells is shown on the right side of the corresponding bar plot. Genes were selected for differential expression across multiple pairs of clusters, with the added restriction of binary (on-off) expression, as far as possible, in at least two of the clusters.

H. Expression of *Heg1* and *Pcdh8* in tdT<sup>+</sup> pSNs in lumbar DRG of adult ( $\geq$  p56) *PVRx3:tdT* mice. Single neuron analysis (i-iii) of *Heg1* (h) and *Pcdh8* (p) expression exhibits three expression patterns across the five molecular identified clusters: cluster 1 tdT<sup>+</sup> neurons expressing high transcript levels of *Heg1* but little *Pcdh8* (i), cluster 2-4 tdT<sup>+</sup> neurons with comparable levels of *Heg1* and *Pcdh8* (ii), and cluster 5 tdT<sup>+</sup> neurons with high *Pcdh8* but no *Heg1* (iii). Some *Pcdh8* expressing neurons fall outside of the tdT<sup>+</sup> pSN population (marked by asterisk).

I. Expression of *Nxph1* and *Pcdh8* in tdT<sup>+</sup> pSNs in lumbar DRG of adult *PVRx3:tdT* mice. Single neuron analysis (i-iii) of *Nxph1* (n) and *Pcdh8* (p) expression shows presumed (i) cluster 1 tdT<sup>+</sup> neurons expressing very low transcript levels of *Pcdh8* but no *Nxph1*, (ii) cluster 2-4 tdT<sup>+</sup> neurons with comparable levels of *Nxph1* and *Pcdh8*, and (iii) cluster 5 tdT<sup>+</sup> neurons with high *Pcdh8* and no *Nxph1*. As for *Pcdh8*, *Nxph1* is also expressed in a few non-tdT<sup>+</sup> neurons (marked by asterisk).

J. Percentage of *Heg1*<sup>+</sup>, *Nxph1*<sup>+</sup>, and *Pcdh8*<sup>+</sup> tdT<sup>+</sup> SNs in  $\geq$  p56 lumbar DRG of *PVRx3:tdT* animals. Data points indicate counts for a single DRG section and include both low and high expressing neurons. Mean percentage  $\pm$  S.E.M. for *Heg1*: 63.5  $\pm$  3.2%, n=29 sections; for *Nxph1*: 57.8  $\pm$  2.7%, n=39 sections; for *Pcdh8*: 53.9  $\pm$  1.9%, n=94 sections.

Scale 10  $\mu$ m.

### Figure 3. Validation of molecular distinct proprioceptor subtypes.

A. Summary of transcripts examined in validation assays. Transcript expression levels are indicated by “++” (high expression levels in most neurons), “+” (intermediate expression level in most neurons), “+/-” (lower expression levels in all or a subset of neurons), or “-” (no or nearly no expression).

B-G. Expression of cluster I transcripts *Hpse*, *Colq*, and *Agpat4* in relation to expression of the cluster 5 transcript *Pcdh8* in  $\geq$  p56 lumbar DRG of *PVRx3:tdT* mice. In (B, D, F), images of individual neurons represent examples of observed transcript combinations other than those observed in the main image. High levels of *Hpse* (h) (in B), *Colq* (c) (in D), or *Agpat4* (a) (in F) transcript generally are mutually exclusive with high levels of *Pcdh8* (p) transcript (m indicates merged image). (C, E, G) Percentage of *Hpse*<sup>+</sup>tdT<sup>+</sup> neurons (C), *Colq*<sup>+</sup>tdT<sup>+</sup> neurons (E), or *Agpat4*<sup>+</sup>tdT<sup>+</sup> neurons (F) (of total tdT<sup>+</sup>) in  $\geq$  p56 rostral (L1-3) or caudal (L4-5) lumbar DRG of *PVRx3:tdT* animals. Mean percentage  $\pm$  S.E.M. *Hpse* (L1-3): 30.1  $\pm$  2.8%, n=28 sections; *Hpse* (L4-5): 25.4  $\pm$  2.2%, n=19 sections; *Colq* (L1-3): 36.7  $\pm$  3.5%, n=16 sections; *Colq* (L4-5): 36.1  $\pm$  3.6%, n=16 sections; *Agpat4* (L1-3): 23.2  $\pm$  2.8%, n=21 sections; *Agpat4* (L4-5): 24.7  $\pm$  2.8%, n=11 sections.

H-K. Expression of the cluster 5 transcripts *Itga2*, *Chad*, and *Pcdh8* in  $\geq$  p56 lumbar DRG of *PVRx3:tdT* mice. In (H, J), images of individual neurons represent examples of observed transcript combinations



other than those observed in the main image. High transcript levels of *Itga2* (i) (in H) or *Chad* (c) (in J) generally overlap with high levels of *Pcdh8* (p) transcript expression (m indicates merged image). (I,K) Percentage of *Itga2*<sup>+</sup>tdT<sup>+</sup> neurons (I), or *Chad*<sup>+</sup>tdT<sup>+</sup> neurons (J) (of total tdT<sup>+</sup>) in ≥ p56 rostral (L1-3) or caudal (L4-5) lumbar DRG of *PVRx3:tdT* animals. Mean percentage ± S.E.M. *Itga2*(L1-3): 11.5 ± 1.7%, n=34 sections; *Itga2*(L4-5): 12.4 ± 1.6%, n=21 sections; *Chad*(L1-3): 15.9 ± 2.6%, n=23 sections; *Chad*(L4-5): 19.5 ± 2.4%, n=9 sections.

L. Expression of cluster 2-4 transcript *Nxph1* and cluster 4 transcript *Cdh13* in ≥ p56 *PVRx3:tdT* DRG. Images of individual neurons represent examples of observed transcript combinations other than those observed in the main image. Single neuron analysis confirms that a subset of *Nxph1* (n) neurons co-express *Cdh13* (c; m indicates merged image).

M. Percentage of *Cdh13*<sup>+</sup>tdT<sup>+</sup> neurons (of total tdT<sup>+</sup>) in adult *PVRx3:tdT* DRG. Mean percentage ± S.E.M. *Cdh13*(L1-3): 39.6 ± 4.5%, n=22 sections, *Cdh13*(L4-5): 27.3 ± 5.3%, n=13 sections.

N. Expression analysis of *Hspc* (cluster 1), *Fam196a* (cluster 2 and 4), and *PV* in adult wild type DRG indicates the presence of cluster 1 *Hspc*<sup>+</sup> *Fam196a*<sup>off</sup>, and cluster 2 or 4 *Hspc*<sup>off</sup>, *Fam196a*<sup>+</sup> *PV* neurons. Images of individual neurons represent examples of observed transcript combinations other than those observed in the main image. A few cluster 1 neurons co-express *Hspc* and *Fam196a* (see also Figure 2G).

O. Percentage of *Fam196a*<sup>+</sup>tdT<sup>+</sup> neurons (of total tdT<sup>+</sup>) in adult *PVRx3:tdT* DRG. Mean percentage ± S.E.M. *Fam196a* (L1-3): 35.9 ± 5.7%, n=12 sections; *Fam196a*(L4-5): 31.4 ± 3.0%, n=12 sections.

Scale 10 μm.

#### Figure 4. Molecular distinct pSN clusters segregate with MS or GTO afferents.

A. Expression of Calretinin and tdT in ≥ p56 lumbar DRG of *PVRx3:tdT* mice.

B. Percentage of CR<sup>+</sup>tdT<sup>+</sup> neurons (of total tdT<sup>+</sup>) in adult *PVRx3:tdT* DRG. Mean percentage ± S.E.M. CR(L1-3): 11.6 ± 2.3%, n=24 sections, CR(L4-5): 23.9 ± 1.9%, n=40 sections, P<0.001 (Mann-Whitney U test).

C. Expression of Pou4f3, Runx3, and tdT in ≥ p56 lumbar DRG of *PVRx3:tdT* mice. Arrowheads indicate Pou4f3<sup>+</sup>Rx3<sup>+</sup>tdT<sup>+</sup> neurons.

D. Percentage of Pou4f3<sup>+</sup>tdT<sup>+</sup> neurons (of total tdT<sup>+</sup>) in adult *PVRx3:tdT* DRG. Mean percentage ± S.E.M. Pou4f3(T10): 13.9 ± 3.6%, n=17 sections, Pou4f3(L5): 25 ± 3.2%, n=78 sections.

E. Expression of Substance P and tdT in ≥ p56 lumbar DRG of *PVRx3:tdT* mice. Boxed area is also shown in top right corner to illustrate expression of Substance P.

F. Percentage of Sub. P<sup>+</sup>tdT<sup>+</sup> neurons (of total tdT<sup>+</sup>) in adult *PVRx3:tdT* DRG. Mean percentage ± S.E.M. SubP(L1-3): 13.8 ± 2.5%, n=33 sections, SubP(L4-5): 9.0 ± 1.6%, n=11 sections.

G. Expression of Calretinin (CR) (*i, i'*), Calbindin 1 (CB) (*ii*), and vGlut2 (*iii*) in GFP<sup>+</sup> muscle spindle (MS) afferents of p12 (*i, i'*) or adult ( $\geq$ p56) (*ii, iii*) muscle of *PV:Cre; Mapt:loxP-STOP-loxP:mGFP-iNLZ* (*PV:GFP*) mice. CR, CB, and vGlut2 expression is observed in GFP<sup>+</sup> MS afferents that innervate the equatorial region of the muscle spindle.

H. Expression of vGlut1 and GFP in a muscle spindle (MS) of p18 EDL muscle of *Calb2:Cre; Mapt:loxP-STOP-loxP:mGFP-iNLZ* (*Calb2:GFP*) mice. Similar as observed for Calretinin protein, in *Calb2:GFP* mice, expression of GFP labels MS afferents that innervate the equatorial region of the spindle.

I. Alkaline phosphatase (AP) labeling of GTO afferents in gluteus muscle of *PV:Cre; Pou4f3:loxP-STOP-loxP:PLAP* (*PV/Pou4f3:PLAP*) at p2 (*i, ii*), and at  $\geq$ p56 (*iii, iv*) mice. Dashed line in (*i*) indicates area of the myotendinous junction where GTOs reside. Arrow indicates muscle area where MSs are generally located.

J. *Pcdh8* expression localizes to GFP<sup>+</sup> GTO afferents in p12 gluteus muscle of *PV:GFP* mice.

K. tdT<sup>+</sup> MS afferents in EDL muscle of p20 *Tac1:Cre; Ai14:tdTomato* (*Tac1:tdT*) animals.

L. Summary of molecular distinct pSN subclasses and their association with either MS (group Ia, putative group II), or GTO (group Ib) muscle receptors.

Scale 10 (A, C, E), 20 (*Gii, Giii, H, J*), or 50 (*Gi*)  $\mu$ m. In (G-K), schematics indicate pSN subtype highlighted, and alleles used, in images below.

## Figure 5. Electrophysiological properties of muscle proprioceptors.

A, B. Barplots showing expression (in CPM) of annotated Na<sub>v</sub> (A) and K<sub>v</sub> (B) channels and auxiliary subunit genes across the five putative cell types identified from the clustering of single-cell RNAseq data.

(C-M) Physiological properties of adult (>p28) dissociated tdT<sup>+</sup> neurons as assessed through *in vitro* patch-clamp recordings.

C. Phase contrast images of adult DRG neurons in culture. A patch electrode is shown (left) on a PV<sup>+</sup> neuron expressing tdTomato (red; right).

D. The firing patterns for DRG PV<sup>+</sup> neuron expressing tdTomato displayed in percentages.

E. The frequency response for the three different firing patterns in proprioceptive neurons (n=22); color scheme as in (D). Tonic firing proprioceptors exhibited significantly higher frequencies compared to proprioceptors in which injected current (200-400 pA supra-threshold current steps for 2.5 s) evoked a single AP or burst firing (\*\*\*)  $p < 0.001$ ; One-way ANOVA; Tukey's *posthoc* test).

F. Superimposed traces of voltage (top) following current injection (bottom) for proprioceptors exhibiting a single action potential (AP) irrespective of the current injected.

G. Similar as in F, for a proprioceptor exhibiting a short duration burst of action potentials.

H. Traces from a tonic firing proprioceptor.

(I-M) Bargraphs showing differences in soma size (I), resting membrane potential (J), Rheobase (K), Capacitance (L), and voltage threshold (L) across proprioceptor neurons. Data expressed as mean  $\pm$  S.E.M.; \* $p < 0.05$ , \*\*\* $p < 0.001$  (One-way ANOVA; Tukey's *posthoc* test).

### Figure 6. Circuit features of MS and GTO afferent sensory subtypes.

A, B. Barplots showing expression (in CPM) of annotated GABA<sub>A</sub> (A) and NDMA (B) receptor subunit genes across the five putative cell types identified from the clustering of single-cell RNAseq data.

C-H. Distribution patterns of the GABA-A receptor units *Gabra1* (C), *Gabra3* (E), and *Gabra5* (G) in tdT<sup>+</sup> neurons in adult ( $\geq$ p56) *PVRx3:tdT* mice. Images of individual neurons represent examples of observed *Gabra*<sup>+</sup>tdT<sup>+</sup> neurons other than those observed in the main image. Percentage of *Gabra1*<sup>+</sup>tdT<sup>+</sup> (D), *Gabra3*<sup>+</sup>tdT<sup>+</sup> (F), and *Gabra5*<sup>+</sup>tdT<sup>+</sup> (H) (of all tdT<sup>+</sup> neurons) in p56 mice, at rostral (L1-3) and caudal (L4-5) lumbar levels. Mean percentage  $\pm$  S.E.M. *Gabra1*(L1-3): 30.6  $\pm$  3.8%, n=37 sections, *Gabra1*(L4-5): 34.2  $\pm$  4.8%, n=18 sections; *Gabra3*(L1-3): 73.1  $\pm$  3.7%, n=63 sections, *Gabra3*(L4-5): 61.4  $\pm$  4.9%, n=31 sections; *Gabra5*(L1-3): 19.3  $\pm$  2.4%, n=19 sections, *Gabra5*(L4-5): 20.5  $\pm$  2.7%, n=10 sections.

I. *Grin3a* receptor expression in lumbar DRG of  $\geq$ p56 *PVRx3:tdT* mice. *Grin3a* transcript is detected in a subset of tdT<sup>+</sup> neurons. Images of individual neurons represent examples of observed *Grin3a*<sup>+</sup>tdT<sup>+</sup> neurons other than those observed in the main image.

J. Percentage of *Grin3a*<sup>+</sup>tdT<sup>+</sup> neurons (of all tdT<sup>+</sup> neurons) in lumbar DRG of  $\geq$ p56 *PVRx3:tdT* mice. Mean percentage  $\pm$  S.E.M. *Grin3a*(L1-3): 45.9  $\pm$  8.4%, n=19 sections, *Grin3a*(L4-5): 33.3  $\pm$  6.4%, n=9 sections.

K. Expression of *Npy1r* transcript in tdT<sup>+</sup> neurons in lumbar DRG of  $\geq$ p56 *PVRx3:tdT* mice. Images of individual neurons represent examples of observed *Npy1r*<sup>+</sup>tdT<sup>+</sup> neurons other than those observed in the main image.

K. Percentage of *Npy1r*<sup>+</sup>tdT<sup>+</sup> neurons (of all tdT<sup>+</sup>) in lumbar ganglia. Mean percentage  $\pm$  S.E.M. *Npy1r*(L1-3): 49.9  $\pm$  5.4%, n=19 sections, *Npy1r*(L4-5): 49.3  $\pm$  6.7%, n=5 sections.

Scale 10  $\mu$ m

### Figure 7. Proprioceptor transcriptome profiling shows dynamic transcriptional changes across development.

A. Experimental paradigm for proprioceptor transcriptome analysis across different developmental stages. Developmental stages include e14.5, p0, p12, and adult ( $\geq$ p56). p0 and p12 neurons were isolated as described for the adult dataset. Due to the inefficient *PV:Cre*-mediated recombination at early developmental stages, e14.5 neurons were isolated using a *TrkC:tdTomato* reporter. As described for adult pSNs, viable tdT<sup>+</sup> pSNs neurons were isolated by FACS and directly deposited into 96-well plates.

B-E. tSNE plots at each time point profiled, with cells colored by cluster identity, using the same clustering technique (see Methods) used to identify putative cell types in the adult in Figure 2.

F. Coexpression patterns of genes with observed coexpression in the adult. For each time point, the heatmap represents pairwise gene-gene correlation values (Pearson's  $r$  using log-transformed CPM data). Because genes were selected based on co-expression in the adult (a subset of cluster-specific genes shown in Figure 2G), the correlations are strongest in the adult. At p12, only the putative group Ib genes show strong correlations, whereas by p0 and e14, the overall correlations are substantially weaker. This suggests that general cell type-specific gene coexpression patterns in the adult differ from those in development.

G-L. Developmental expression pattern of Calretinin (CR), Pou4f3 and Substance P (Sub.P). (G, I, K) Expression of CR (G), Pou4f3 (I), and substance P (K) in p0 lumbar DRG of *PVRx3:tdT* mice. (H, J, L) Percentage of CR<sup>+</sup> (H), Pou4f3<sup>+</sup> (J), or Sub. P<sup>+</sup> (L) pSNs (defined by *PVRx3:tdT*<sup>+</sup> or *PV<sup>+</sup>Rx3<sup>+</sup>* neurons) at different developmental stages. Mean percentage  $\pm$  S.E.M. for CR at p0:  $0 \pm 0\%$ , n=10 sections; at p6:  $4.8 \pm 1.6\%$ , n=26 sections; at p12:  $22.5 \pm 1.8\%$ , n=29 sections; at  $\geq$ p56:  $19.4 \pm 1.6\%$ , n=65 sections. Mean percentage  $\pm$  S.E.M. for Pou4f3 at e15.5:  $59.6 \pm 3.5\%$ , n=22 sections; at p0:  $13.5 \pm 3.0\%$ , n=17 sections; at p6:  $21.4 \pm 5.2\%$ , n=28 sections; at  $\geq$ p56:  $23.1 \pm 2.8\%$ , n=95 sections. Mean percentage  $\pm$  S.E.M. for Sub. P at p0:  $2.0 \pm 1.1\%$ , n=20 sections; at p6:  $4.5 \pm 1.2\%$ , n=25 sections; at p12:  $11.3 \pm 1.2\%$ , n=30 sections; at  $\geq$ p56:  $8.8 \pm 1.3\%$ , n=22 sections.

Scale 10  $\mu$ m

## KEY RESOURCES TABLE

(see Excel file)

## CONTACT FOR REAGENT AND RESOURCE SHARING.

Questions on scRNAseq analysis should be directed to Vilas Menon (vm2545@cumc.columbia.edu). All other requests for (information on) resources and reagents should be directed to Joriene de Nooij).

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

**Animal husbandry and mouse strains.** All experiments were performed according to National Institutes of Health guidelines and approved by the Institutional Animal Care and Use Committee of Columbia University. Mouse strains used for experiments are described in Star methods. *Runx3:flpO* animals were generated through homologous recombination of a *Runx3:FlpO* targeting vector in MM13 ES cells. Successful recombinants were identified through Southern blot analysis. *Rx3:FlpO* transgenic animals (heterozygous or homozygous) were identified through genotyping PCR analysis. Primers amplifying the *FlpO* allele are *FlpO-For*: 5'-GCATCTGGGAGATCACCGAG-3' and *FlpO-Rev*: 5'-GCCGTTCCAGGCGGGGTATCTG-3', which result in an 850 bp product. Primers used to distinguish homozygous from heterozygous or wild type animals: *Rx3ex6 (F)*: 5'-GCGCCCTACCACCTCTT-3' and *Rx3ex6(R2)*: 5'-TGGGAGCCACTGCCAGCTCTG-3'. These primers result in a 400 pb product in heterozygous and wild-type animals, but no product in homozygous mutant animals. For all experiments, animals of both sexes were used. Age of animals at time of analysis is indicated per experiment. Animals were given ad lib access to water and food.

## METHOD DETAILS

**Fluorescence activated cell sorting (FACS) of dissociated pSNs.** DRG from adult (p>56), adolescent (p12), neonatal (p0), and embryos (e14.5) of either sex were dissected (1 hour maximal) in ice-cold Hank's balanced salt solution (HBSS) and collected, on ice, in HBSS supplemented with 0.75% horse serum (HS). The total number of animals (of either sex) used for experiments was six for adult samples, four for p12, six for p0, and two for e14.5. Following dissection, DRGs were centrifuged at low speed and washed once with ice cold HBSS prior to dissociation through enzymatic digestion using Papain followed by Collagenase/Dispase (Malin et al., 2007; de Nooij et al., 2015). Papain digestion step consisted of 3 ml HBSS with ~16 units/ml Papain (Worthington), 0.83 mM L-Cysteine, 0.42 mM EDTA, and 20 units DNase I/ml (Roche). Collagenase/Dispase digestion was with 3 ml HBSS containing 1,066 units/ml Collagenase IV (Worthington), 4 units/ml Dispase (Worthington), and 20 units/ml DNase I (Roche). Duration of

digestion incubation times was 16 min for adult, 12 min for p12, 10 minutes for p0 and e14.5 (for both enzyme digestion steps). Solutions were exchanged by a low speed centrifuge step (4' at 800 rpm in table top Eppendorf centrifuge) and aspiration. After Collagenase/Dispase digestion, enzyme solution was replaced with 500  $\mu$ l HBSS supplemented with 20% HS and 20 units DNase I, and DRG/cell suspension was dissociated by slow mechanical trituration using a 200  $\mu$ l pipetman (~50 times). Following dissociation, cell suspension was incubated with 2  $\mu$ M Calcein Blue (CellTrace™ Calcein Blue, AM; Invitrogen) for 15 minutes at room temperature. After Calcein Blue labeling, cells were spun down (4' at 800 rpm) and resuspended in 500  $\mu$ l sorting solution (HBBS, 1% HS, 20 units/ml DNase I) supplemented with 0.01  $\mu$ M Sytox red (SYTOX™ Red Dead Cell Stain; Invitrogen). Just prior to FACS, dissociated cells were passed through 40-70  $\mu$ m gauze filters to clear remaining cellular aggregates. FACS of fluorescently-labelled (tdTomato<sup>+</sup>) neurons was performed at 12 psi, using a Becton Dickinson FACSaria (SORP model, 5-laser, 20 parameter), equipped with a 130  $\mu$ m nozzle, and using 586/15 (tdTomato), 450/50 (Calcein Blue), and 670/30 (Sytox red) filter sets. Fluorescent neurons were directly deposited in 96-well LoBind plates (Eppendorf) prefilled with 7.5  $\mu$ l lysis buffer/well. Lysis buffer contained 0.2% Triton X-100 (Sigma), SUPERaseIN (1 U/ $\mu$ l) (ThermoFisher), 2 mM deoxyribonucleotides (dNTPs) (ThermoFisher), and 2  $\mu$ M reverse transcriptase (RT) primer (Integrated DNA Technologies) (Snyder et al., 2019). Plates were stored at -80°C until processed for cDNA generation, library preparation and RNA sequencing.

**Single cell cDNA generation, library preparation, and RNA sequencing.** Single cell cDNA and library preparation was performed by the Columbia JP Sulzberger genome center as described previously (Snyder et al., 2019). In brief, after primer annealing (72°C for 3 min), reverse transcription (RT) was performed by adding 7.5  $\mu$ l of RT mix to each well. RT mix contained 2 M betaine (Affymetrix), 2 $\times$  Protoscript Buffer (New England Biolabs), 12 mM MgCl<sub>2</sub> (ThermoFisher), 10 mM dithiothreitol (ThermoFisher), 5.3 U of Protoscript II Reverse Transcriptase (New England Biolabs), 0.53 U of SUPERaseIN (ThermoFisher), and 2  $\mu$ M Template Switching Oligo (Integrated DNA Technologies). Reverse transcription was performed at 42°C for 90 min, followed by 10 cycles of 50°C for 2 min, 42°C for 2 min, 70°C for 10 min, followed by a 4°C hold. Excess primers were removed by adding 2  $\mu$ l of Exonuclease I (ThermoFisher) mix to each well (1.875 U of ExoI in water) and incubating at 37°C for 30 min, 85°C for 15 min, 75°C for 30 s, and 4°C hold. Following RT reactions, all wells were pooled into a single 15-ml falcon tubes, and complementary DNA (cDNA) was purified and concentrated using Dynabeads MyOne Silane beads (ThermoFisher) according to the manufacturer's instructions. The cDNA

was split into duplicate reactions containing 25  $\mu$ l of cDNA, 25  $\mu$ l of 2 $\times$  HIFI HotStart Ready Mix (Kapa Biosystems), and 0.2 M SMART PCR Primer and polymerase chain reaction (PCR) mix. cDNA was amplified as above, and duplicate reactions were combined and purified using 0.7 volume of AMPure XP beads (Beckman Coulter). The amplified cDNA was visualized on an Agilent TapeStation and quantified using a Qubit II fluorometer (ThermoFisher).

Sequencing libraries were constructed using Nextera XT (Illumina) as described in (Snyder et al., 2019). A custom i5 primer was used (Nextera PCR) with 0.6 ng of input cDNA, and 10 cycles of amplification were performed. Unique i7 indexes were used for each plate. After amplification, the library was purified with two rounds of AMPure XP beads, visualized on the TapeStation, and quantified using the Qubit II fluorometer. Libraries were sequenced on an Illumina NextSeq 500 using the 75-cycle High Output Kit [read lengths 26(R1)  $\times$  8(i)  $\times$  58(R2)]. Custom sequencing primers were used for Read 1 (SMART\_R1seq and ILMN\_R1seq; see Snyder et al., 2019). With each plate, we targeted ~70 million reads. Library pools were loaded at 1.8 pM with 30% PhiX (Illumina). All sequencing data will be deposited in the NCBI GEO database (accession # xxx).

**Single cell transcriptome analysis.** Analysis of single cell RNA sequencing datasets was performed essentially as described previously (Tasic et al., 2016).

**Alignment of sequencing reads.** Raw sequence reads were aligned to the mm10 mouse reference genome with transcriptome annotation derived from NCBI. Alignment was performed using the STAR aligner, and reads were demultiplexed by Plate-seq barcode and collapsed by Unique Molecular Identifiers to obtain final transcript counts for each cell. For visualization of gene expression and calculation of gene-gene correlations, we normalized UMI counts to CPM (counts per million). For each cell, the number of UMIs for each gene is divided by the total UMI count for the cell, and then multiplied by  $10^6$  to obtain CPM values. For differential gene expression among clusters, we used the raw UMI count values, since edgeR performs its own distribution-specific normalization.

**Filtering of satellite cell contaminants.** To avoid clustering cells with significant oligodendrocyte and glial gene contamination, we set the following thresholds, based on the distributions of transcripts over the entire data set: Mbp Transcripts per million (CPM) < 10,000; Apoe Transcripts per million (CPM) < 20.

**Cluster analysis.** After alignment and initial filtering of cells for non-neuronal signatures, we performed a bootstrapped, iterative clustering approach, using the hicat R package as outlined in Tasic et al. 2018. This clustering algorithm proceeds as follows: (1) Select all cells. (2) Select top 6,000 genes with the highest variance/mean (Fano Factor) ratio. (3) Perform a modified Weighted Gene Network Co-

expression Analysis to identify modules of genes, using a cutheight of 0.995 and soft-threshold power = 1. (4) When no gene modules with co-expression beyond chance are identified, terminate the iteration of clustering. Otherwise, cluster cells in the reduced dimensions corresponding to the eigengenes of the gene modules, using hierarchical clustering with Ward's Method, and a distance metric defined by cell-cell correlation in eigengene space. (5) Test robustness of clusters by summing the negative log<sub>10</sub> p-value of differentially expressed genes (defined by the R limma package) among each pair of clusters. (6) Merge clusters for pairs of clusters with a sum of negative log<sub>10</sub> p-values less than 100. We repeated steps 2-5 for each new cluster identified, until no further subclusters were identified. We next repeated steps 1-6 100 times, using randomly selected subsets comprising 80% of all cells, to generate a co-clustering matrix, where each entry represents the proportion of times a given pair of cells was found in the same cluster across 100 iterations. For each pair of clusters, we determined the mean co-clustering value over all cells belonging to both of the clusters. Clusters with mean co-clustering values >0.25 were merged.

*Differential gene expression and marker identification.* We used edgeR on the UMI count values to perform differential gene expression and marker identification. We calculated differential expression across all pairs of clusters. A subset of combinatorial marker genes was selected for display (in Figure 2) based on identification across multiple pairwise comparisons.

*t-SNE visualization.* Cells were visualized in t-SNE space using all differentially expressed genes across clusters (FDR-adjusted p-value < 0.05). After subsetting on these genes, we calculated the top 20 principal components on the log<sub>10</sub>(CPM+1) transformed data. These PC values were input into the Rtsne package to calculate t-SNE coordinates for visualization. Cells were not clustered in t-SNE space.

*GO Analysis.* Go analysis was performed for differentially expressed genes between each pair of clusters from the P56 cells. For each pair of clusters, we use edgeR to find differentially expressed genes, up in cluster 1 and down in cluster 2. The opposite comparison (down in cluster 1, up in cluster 2) was run as a separate gene list. Differentially expressed genes were filtered to those with log-Fold Change >1. The differential gene list was run through topGO, with the following parameters: nodeSize=5, annot=annFUN.org. Output of GO categories was filtered using the following criteria: (i) At least five overlapping genes between the differential gene list and the GO category, (ii) Fisher.elim and Fisher.classic adjusted p-values < 0.01, and (iii) Category is related to neurons and neuronal processes (manual filtering).

The final output consisted of a list of significant GO categories for each cluster (genes upregulated relative to each other cluster, in pairwise fashion).



*Longitudinal expression analysis.* For the cells obtained from the e14, p0, and p12 time points, the clustering was carried out as with the p56 data, and marker genes identified in the same way. For the gene-correlation analysis, we selected a subset of differential genes from p56, and calculated the Pearson correlation coefficient of the log(CPM+1) value for each pair of genes from this gene subset, over all the cells from a given time point. The plots in Figure 7 reflect these correlation values, arranged according to the hierarchical clustering of these genes (average linkage) in the p56 data.

**Immunohistochemistry.** Immunohistochemistry was performed on cryostat sections (15–30  $\mu$ m) or whole-mount muscle tissue as described previously (de Nooij et al., 2013). For p8 and older tissues, animals were anesthetized using Avertin (1.25%) and perfused by transcardial perfusion with PBS (~20 ml, adjusted based on age of animal) and 4% PFA (in PBS; ~20 ml). Tissues were dissected in PBS and post-fixed for two hrs (p12 and younger and muscle tissue) to o/n (p12 and older tissues) in 4% PFA in PBS at 4°C. (Please note, for staining with Pou4f3 antibody, tissues were perfused with 2% PFA in PBS and post-fixed for 30 minutes at RT). Following fixation, tissues were washed in PBS (3 exchanges of 10 minutes each) and, for cryostat sections, equilibrated in 30% sucrose (in 0.1 M PO<sub>4</sub> buffer), embedded in Tissue-Tek O.C.T. (Sakura Finetek), and stored at -80°C until use. Sections were cut on a Leica CM3050S cryostat and collected on Superfrost plus slides (Fisherbrand) and stored air tight at -80°C. When proceeding directly with immunostaining, slides/sections were air dried for ~15 minutes and rinsed once with PBS (5') before o/n incubation (at 4°C) with primary antibody in PBS, 1% BSA, 0.1% Triton X-100. Following primary antibody incubation, slides were washed twice in PBS (6-8 min/wash) and subsequently incubated with secondary antibody for 2 hrs at RT. Following incubation with secondary abs, slides were washed with two exchanges of PBS (6-8 min each) and cover slipped with fluoromount-G (Southern Biotech), or Vectashield (Vector labs). For whole mount muscle tissue, following PFA fixation and wash with PBS (in 24-well plates), muscles were dissected free (and teased into smaller segments depending on muscle size) and incubated in ice cold methanol (100%) for 5-10 minutes. After methanol permeation, muscle tissues were washes in PBS (4 exchanges of 5, 10, 15, and 15 minutes; at RT) before incubation in blocking solution (PBS, 1% BSA, 0.3% Triton X-100 ) for at least one hour under gentle agitation at RT or o/n at 4°C. Incubation of primary antibodies was performed in 24 well plates (250  $\mu$ l antibody soln./well) for 2-3 days at 4°C under gentle agitation. After primaries, muscles were washed at RT for 3 x 5 min and 5 x 1 hour in blocking solution, and subsequently incubated with secondary antibody (for 1-2 days at 4°C). After staining with secondary antibodies, muscle were washed as before with the final two washes performed in PBS without BSA/Triton X-100.

Muscles were kept in PBS at 4°C until mounted in Vectashied (Vector labs). Primary antibodies used in immunohistochemistry experiments are described in STAR methods. Fluorophore-conjugated secondary antibodies generated in donkey (Jackson Immuno Research Laboratories) were used at 1:1000 (FITC), or 1:500 (Cy3, Cy5) in PBS, 1% BSA, 0.1% Triton X-100. Images were acquired on LSM510 or LSM700 confocal microscopes (Carl Zeiss).

**Alkaline phosphatase labeling.** Alkaline phosphatase staining was essentially performed as described previously (Badea et al., 2003). In brief, tissues were fixed as described above and incubated at 68°C for 90 min. to inactivate endogenous alkaline phosphatase activity. Subsequently, tissues were washed twice (10 min each in AP pre-staining buffer (0.1 M Tris, 0.1 M NaCl, 50 mM MgCl<sub>2</sub>, pH 9.5) prior to incubation with AP staining buffer (prestaining buffer supplemented with 0.34 g/ml nitroblue tetrazolium (NBT), and 0.175 g/ml 5-bromo-4-chloro-3-indolyl-phosphate (BCIP) (Invitrogen). Alkaline phosphatase staining reaction was proceeded o/n at RT. After staining, tissues were washed three times for 20 min in PBS, 0.1% Tween 20, and postfixed in PBS with 4% paraformaldehyde o/n. After washes with PBS, samples were dehydrated through an ethanol series and then cleared with 2:1 benzyl benzoate (BB)/benzyl alcohol (BA). Images were acquired using a Zeiss Axioscope 2.

**RNAscope.** RNAscope analysis was performed using the ACDbio RNAscope kit for nonfixed tissue (ACD #320851) based on Griffith et al., 2019. In short, DRG extracted from adult mice were fixed for 15 minutes in 4% PFA, washed in PBS, and allowed to equilibrate in 30% sucrose for 2 hr. before embedding in OCT (Tissue Tek; Sakura Finetek). DRG were stored at 80°C until use. On day of experiment, DRG were sectioned at 20 µm, dried for 1 h at 30°C, washed in PBS, dehydrated in a series of 50%, 70%, and 100% EtOH steps, and pretreated with protease IV digestion for 30 minutes (at RT). Hybridization was performed in a humidified oven (ACDbio) at 40°C for 2 hr. Following hybridization, tissues were washed and processed for probe amplification and detection using Amp1 (30 min), Amp2 (15 min), Amp3 (30 min.), and Alt4 (A, B, or C; 15 min). All wash steps were performed at 40°C in the same humidified hybridization oven. Probes used in experiments are described in STAR methods. When combined with immunostaining, following Alt4 incubation step, slides were rinsed with PBS once and process as described above (*Immunohistological analysis*). Images were acquired on a Zeiss LSM700 confocal microscope.

**Electrophysiology.** Electrophysiological experiments were performed at ~37 °C, from dorsal root ganglia dissociated neuronal cultures performed 24 hours prior to the recording session from adult mice (age

range: 6 to 9 weeks old). Whole cell patch clamp recordings were performed using a Multiclamp 700B amplifier (Molecular Devices., Sunnyvale, CA, USA). Signals were acquired at frequency of 100 KHz and filtered at 10 KHz. Borosilicate electrodes were pulled using a puller P-1000 (Sutter instrument) with resistance ranging between 4 to 8 M $\Omega$  and filled with an intracellular solution containing (in mM): 130 K-gluconate, 10 NaCl, 10 HEPES, 1 EGTA, 1 MgCl<sub>2</sub>, 0.1 CaCl<sub>2</sub>, and 1 Na-GTP. pH was adjusted to 7.2 with KOH. Osmolarity was adjusted to 290 – 295 mOsmol Kg<sup>-1</sup> with sucrose. The extracellular solution contained (in mM): 145 NaCl, 3 KCl, 1.5 CaCl<sub>2</sub>, 1 MgCl<sub>2</sub>, 10 D-glucose and 10 HEPES. pH was adjusted to 7.4 with NaOH. Osmolarity ranged between 310 – 315 mOsmol Kg<sup>-1</sup>. The junction potential was corrected and taken into account for subsequent analysis. Only tdTomato positive cells were recorded, identified by epifluorescence using a Leica (DM 6000FS) upright fix-stage microscope. The diameter of the cells was measured with the Leica Application Suite X software (Leica Microsystems). After whole cell access was established, seal parameters were recorded and only cells with input resistance ( $R_{in}$ ) over 75 M $\Omega$  were analyzed, based on the value for  $R_{in}$  reported in other reports for large DRG neurons and specifically for Pavlb+ neurons (Woo et al, 2015; Liang et al, 2016). The input resistance for each cell was obtained from the slope of a steady-state (linear) current-voltage plot in response to a series of hyperpolarizing currents. After establishing the resting membrane potential in current clamp mode, cells were held at -60 mV and a series of current steps of 50 pA (duration of 250 ms) was injected to induce action potentials (APs), in order to reveal the firing pattern (assigned as: single AP, burst or phasic). Signals were acquired using a Digidata 1440A (Molecular Devices., Sunnyvale, CA, USA) controller by pClamp 10.3 software. To ensure supramaximal stimulation, the current injection was increased to 200 pA over the rheobase. The parameters measured were: i) capacitance, ii) resting membrane potential ( $V_m$ ), iii) rheobase, iv) voltage threshold for AP induction, v) AP amplitude and vi) AP half-width (see Figure S11). Additionally, we perform two protocols to identify the maximum firing frequency and to determine the ability of neurons to induce action potentials at different frequencies following a small current step injection. For the first protocol, cells were stimulated at supra-threshold values using 200 to 400 pA over rheobase with a sustained current for 2.5 s. The number of APs per second was calculated. For the latter, a series of current injections with a short duration (5 ms) at different frequencies were performed using currents over 200 to 300 pA of rheobase for a total duration of 2.5 s in each series. The frequencies analyzed ranged from 4 Hz to 200 Hz. The total number of APs was quantified. The recordings were analyzed off-line with the software Clampex 10.1 (Molecular Devices Corp., Sunnyvale, CA, USA).

**Quantification and statistical analysis.** For whole DRG neuronal counts, analyses were performed on serial cryostat sections (30  $\mu\text{m}$ ) of individual ganglia, with each section counted. Counted cell bodies and nuclei only included those with near full-size diameter to avoid double counting of cells or nuclei at the edge of sections. Except where stated otherwise, a minimum of three DRG, obtained from three experimental or WT animals were counted per segmental level, and per experimental condition. Average counts/DRG and SEM were calculated through Sigmaplot (Systat Software). For neuronal counts on tissue sections obtained at rostral or caudal lumbar DRG (e.g. RNAscope, immunohistological analysis), individual sections (20  $\mu\text{m}$ ) from cervical, thoracic, or lumbar DRG were analyzed. The average number of neurons/section were calculated by dividing the sum of all counted cells by the total number of counted sections. To estimate the proportion of pSNs expressing a given transcript, marker expressing neurons were counted across multiple tissue sections (at least 6 per probe combination) and gated to PVRx3 (tdT<sup>+</sup>) neurons. Percentage of PVRx3:tdT<sup>+</sup> neurons expressing a certain marker combination were calculated by dividing the number of marker neurons for each section with the number of PV<sup>+</sup> neurons for the same section. Average percentages and SEM were calculated through Sigmaplot (Systat Software). Statistical analysis (Student's t test or Mann-Whitney U test) on all counted neuronal populations or cell diameters was performed using Sigmaplot (Systat Software). Significance was accepted for  $p < 0.05$ .

Statistical analysis for electrophysiological recording experiments was performed using the software GraphPad Prism 6 (GraphPad, Inc.) and data are reported as mean  $\pm$  standard error of the mean (SEM). *Kolmogorov-Smirnov* test was used to test normal distribution. Significance was set as \* $p < 0.05$ , \*\*  $p < 0.01$  and \*\*\*  $p < 0.001$  as assessed by the One-way ANOVA; Tukey's *posthoc* test.

## References:

- Abraira, V.E., and Ginty, D.D. (2013). The sensory neurons of touch. *Neuron* 79, 618-639.
- Badea, T.C., & Nathans, J. (2011). Morphologies of mouse retinal ganglion cells expressing transcription factors Brn3a, Brn3b, and Brn3c: Analysis of wild type and mutant cells using genetically-directed sparse labeling. *Vision Research* 51, 269-279.
- Badea, T.C., Wang, Y., & Nathans, J. (2003). A noninvasive genetic/pharmacologic strategy for visualizing cell morphology and clonal relationships in the mouse. *J. Neurosci.* 23, 2314-2322.
- Baek, M., Menon, V., Jessell, T.M., Hantman, & A.W., Dasen, J.S. (2019). Molecular logic of spinocerebellar tract neuron diversity and connectivity. *Cell Reports* 27, 2620-2635.

- Banks, R.W., Barker, D., & Stacey, M.J. (1982). Form and distribution of sensory terminals in cat hindlimb muscle spindles. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 299, 329-364.
- Barker, D. (1974). The morphology of muscle receptors. In: *Handbook of Sensory Physiology* (pp. 1-190). Springer, New York.
- Bastian, H. (1888). The “muscular sense”; Its nature and localization. *Brain* 10, 1–36.
- Bewick, G.S., Reid, B., Richardson, C., & Banks, R.W. (2005). Autogenic modulation of mechanoreceptor excitability by glutamate release from synaptic-like vesicles: Evidence from the rat muscle spindle primary sensory ending. *J. Physiology* 562, 381–394.
- Bikoff, J.B., Gabitto, M.I., Rivard, A.F., Drobac, E., Machado, T.A., Miri, A., Brenner-Morton, S., Famojure, E., Diaz, C., Alvarez, F.J., Mentis, G.Z., & Jessell, T.M. (2016). Spinal inhibitory interneuron diversity delineates variant motor microcircuits. *Cell* 165, 207-219.
- Birznieks, I., Macefield, V.G., Westling, G., & Johansson, R.S. (2009). Slowly adapting mechanoreceptors in the borders of the human fingernail encode fingertip forces. *J. Neurosci.* 29, 9370-9379.
- Brown, A. G., and Fyffe, R. E. (1979). The morphology of group Ib afferent fibre collaterals in the spinal cord of the cat. *J. Physiology* 296, 215-226.
- Brown, A.G. (1981). *Organization in the Spinal Cord*. Springer, New York.
- Brown, A. G., and Fyffe, R. E. (1981). Direct observations on the contacts made between Ia afferent fibres and alpha-motoneurons in the cat's lumbosacral spinal cord. *J. Physiology* 313, 121-140.
- Brown, M.C., Engberg, I., & Matthews, P.B.C. (1967). Fusimotor stimulation and the dynamic sensitivity of the secondary ending of the muscle spindle. *J. Physiology* 189, 545–550.
- Cao, Y.Q., Mantyh, P.W., Carlson, E.J., Gillespie, A.M., Epstein, C.J., & Basbaum, A.I. (1998). Primary afferent tachykinins are required to experience moderate to intense pain. *Nature* 392, 390-394.
- Catela, C., Shin, M.M., Lee, D.H., Liu, J.P., & Dasen, J.S. (2016). Hox proteins coordinate motor neuron differentiation and connectivity programs through Ret/Gfra genes. *Cell Reports* 14, 1901-1915.
- Crowe, A., and Matthews, P.B.C. (1964). The effects of stimulation of static and dynamic fusimotor fibres on the response to stretching of the primary endings of muscle spindles. *J. Physiology* 174, 109–131.
- Davies, P., Petit, J., & Scott, J.J. (1995). The dynamic response of Golgi tendon organs to tetanic contraction of in-series motor units. *Brain Research* 690, 82–91.
- de Noij, J. C., Doobar, S., & Jessell, T.M. (2013). Etv1 inactivation reveals proprioceptor subclasses that reflect the level of NT3 expression in muscle targets. *Neuron* 77, 1055-1068.
- de Noij, J.C., Simon, C.M., Simon, A., Doobar, S., Steel, K.P., Banks, R.W., Mentis, G.Z., Bewick, G.S., & Jessell, T.M. (2015). The PDZ-domain protein Whirlin facilitates mechanosensory signaling in mammalian proprioceptors. *J. Neurosci.* 35, 3073-3084.
- Eccles, J.C., Eccles, R.M., & Lundberg, A. (1957). Synaptic actions on motoneurons caused by impulses in Golgi tendon organ afferents. *J. Physiology* 138, 227-252.

- Felts, P. A., Black, J.A., Dib-Hajj, S.D., & Waxman, S.G. (1997). NaG: A sodium channel-like mRNA shared by Schwann cells and other neural crest derivatives. *Glia* 21, 269-276.
- Goodwin, G.M., McCloskey, D.I., & Matthews, P.B. (1972). The contribution of muscle afferents to kinaesthesia shown by vibration induced illusions of movement and by the effects of paralysing joint afferents. *Brain* 95, 705–748.
- Goel, M., Li, T., and Badea, T.C. (2019). Differential expression and subcellular localization of Copines in mouse retina. *J Comp Neurol.* 527, 2245-2262.
- Harris, J.A., Hirokawa, K.E., Sorensen, S.A., Gu, H., Mills, M., Ng, L.L., Bohn, P., Mortrud, M., Ouellette, B., Kidney, J., Smith, K.A., Dang, C., Sunkin, S., Bernard, A., Oh, S.W., Madisen, L., & Zeng, H. (2014). Anatomical characterization of Cre driver mice for neural circuit mapping and manipulation. *Front Neural Circuits* 8, 76.
- Hippenmeyer, S., Vrieseling, E., Sigrist, M., Portmann, T., Laengle, C., Ladle, D.R., & Arber, S. (2005). A developmental switch in the response of DRG neurons to ETS transcription factor signaling. *PLoS Biology* 3, e159
- Houk, J., and Henneman, E. (1967). Responses of Golgi tendon organs to active contractions of the soleus muscle of the cat. *J. Neurophys.* 30, 466-481.
- Hunt, C. C. (1990). Mammalian muscle spindle: Peripheral mechanisms. *Physiology Reviews* 70, 643-663.
- Jami, L. (1992). Golgi tendon organs in mammalian skeletal muscle: Functional properties and central actions. *Physiology Reviews* 72, 623-666.
- Jan, L.Y., and Jan, Y.N. (2012) Voltage-gated potassium channels and the diversity of electrical signaling. *J. Physiology* 590, 2591-2599.
- Jankowska, E. (1992). Interneuronal relay in spinal pathways from proprioceptors. *Progr. in Neurobiol.* 38, 335-378.
- Jankowska, E., and Edgley, S.A. (2010). Functional subdivision of feline spinal interneurons in reflex pathways from group Ib and II muscle afferents; An update. *Eur. J. Neurosc.* 32, 881-893.
- Jerng, H.H., Pfaffinger, P.J., & Covarrubias, M. (2004). Molecular physiology and modulation of somatodendritic A-type potassium channels. *Mol. Cell. Neurosc.* 27, 343-369.
- Kölliker, A. (1826). On the terminations of nerves in muscles, as observed in the frog: And on the disposition of the nerves in the frog's heart. *Proc. of the Royal Society* 12, 65-84.
- Kovács-Öller, T., Szarka, G., Ganczer, A., Tengölics, Á., Balogh, B., & Völgyi, B. (2019). Expression of Ca<sup>2+</sup>-Binding buffer proteins in the Human and Mouse Retinal Neurons. *Int. J. Mol. Sciences* 20, e2229.
- Kozeka, K., and Ontell, M. (1981). The three-dimensional cytoarchitecture of developing murine muscle spindles. *Dev. Biol.* 87, 133–147.
- Kühne, W. (1863). Über die Endigung der Nerven in den Muskeln. *Virchows Archives of Pathological Anatomy and Physiology and of Clinical Medicine* 27, 508-523.
- Langfelder, P., and Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* 9, 559.

- Lee, J., Friese, A., Mielich, M., Sigrist, M., Arber, S. (2012) Scaling proprioceptor gene transcription by retrograde NT3 signaling. *PLoS One* 7, e45551.
- Li, C.L., Li, K.C., Wu, D., Chen, Y., Luo, H., Zhao, J.R., Wang, S.S., Sun, M.M., Lu, Y.J., Zhong, Y.Q., Hu, X.Y., Hou, R., Zhou, B.B., Bao, L., Xiao, H.S., Zhang, X. (2016) Somatosensory neuron types identified by high-coverage single-cell RNA-sequencing and functional heterogeneity. *Cell Research* 26, 967.
- Li, L., Rutlin, M., Abaira, V.E., Cassidy, C., Kus, L., Gong, S., Jankowski, M.P., Luo, W., Heintz, N., Koerber, H.R., Woodbury, C.J., & Ginty, D.D. (2011). The functional organization of cutaneous low-threshold mechanosensory neurons. *Cell* 147, 1615-27.
- Macefield, V.G., and Knellwolf, T.P. (2018). Functional properties of human muscle spindles. *J. Neurophys.* 120, 452-467.
- Madisen, L., Zwingman, T.A., Sunkin, S.M., Oh, S.W., Zariwala, H.A., Gu, H., Ng, L.L., Palmiter, R.D., Hawrylycz, M.J., Jones, A.R., Lein, E.S., & Zeng, H. (2010). A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosc.* 13, 133-40.
- Madisen, L., Garner, A.R., Shimaoka, D., Chuong, A.S., Klapoetke, N.C., Li, L., van der Bourg, A., Niino, Y., Egolf, L., Monetti, C., Gu, H., Mills, M., Cheng, A., Tasic, B., Nguyen, T.N., Sunkin, S.M., Benucci, A., Nagy, A., Miyawaki, A., Helmchen, F., Empson, R.M., Knopfel, T., Boyden, E.S., Reid, R.C., Carandini, M., & Zeng, H. (2015). Transgenic mice for intersectional targeting of neural sensors and effectors with high specificity and performance. *Neuron* 85, 942-958.
- Maksimovic, S., Nakatani, M., Baba, Y., Nelson, A.M., Marshall, K.L., Wellnitz, S.A., Firozi, P., Woo, S.H., Ranade, S., Patapoutian, A., & Lumpkin, E.A. (2014). Epidermal Merkel cells are mechanosensory cells that tune mammalian touch receptors. *Nature* 509, 617-21.
- Matthews, P.B.C. (1972) *Mammalian muscle receptors and their central actions*. Edward Arnold Publishers, London.
- McCarthy, D.J., Chen, Y., & Smyth, G.K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research* 40, 4288-4297.
- Menon, V. (2018). Clustering single cells: A review of approaches on high-and low-depth single-cell RNA-seq data. *Briefings in Functional Genomics* 17, 240-245.
- Paré, M., Behets, C., & Cornu, O. (2003) Paucity of presumptive ruffini corpuscles in the index finger pad of humans. *J. Comp. Neurol.* 456, 260-266.
- Pearson, K.G. (2004). Generating the walking gait: Role of sensory feedback. *Progr. Brain Res.* 143, 123-129.
- Pierrot-Deseilligny, E., and Burke, D.J. (2012). *The circuitry of the human spinal cord: Spinal and corticospinal mechanisms of movement*. Cambridge University Press, New York.
- Pleil, K.E., Rinker, J.A., Lowery-Gionta, E.G., Mazzone, C.M., McCall, N.M., Kendra, A.M., Olson, D.P., Lowell, B.B., Grant, K.A., Thiele, T.E., & Kash, T.L. (2015). NPY signaling inhibits extended amygdala CRF neurons to suppress binge alcohol drinking. *Nat. Neurosc.* 18, 545-552.
- Poliak, S., Norovich, A.L., Yamagata, M., Sanes, J.R., & Jessell, T.M. (2016). Muscle-type identity of proprioceptors specified by spatially restricted signals from limb mesenchyme. *Cell* 164, 512-525.

- Proske, U., and Gandevia, S.C. (2012). The proprioceptive senses: Their roles in signaling body shape, body position and movement, and muscle force. *Physiological Reviews* 92, 1651-1697.
- Proske, U., and Gandevia, S.C. (2018). Kinesthetic Senses. *Comprehensive Physiology* 8, 1157-1183.
- Robinson, M.D., McCarthy, D.J., & Smyth, G.K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139-140.
- Rudomin, P. (2002). Selectivity of the central control of sensory information in the mammalian spinal cord. *Adv. Exp. Med. Biol.* 508, 157-170.
- Scott, J.J. (1991). Responses of Ia afferent axons from muscle spindles lacking a bag1 intrafusal muscle fibre. *Brain Res.* 543, 97-101.
- Scott, S. H. (2016). A functional taxonomy of bottom-up sensory feedback processing for motor actions. *Trends in Neurosc.* 39, 512-526.
- Shakya-Shrestha, S., Bannatyne, B.A., Jankowska, E., Hammar, I., Nilsson, E. & Maxwell, D.J. (2012). Excitatory inputs to four types of spinocerebellar tract neurons in the cat and the rat thoraco-lumbar spinal cord. *J. Physiology* 590, 1737–1755.
- Sharma, N., Flaherty, K., Lezgiyeva, K., Wagner, D.E., Klein, A.M., & Ginty, D.D. (2020). The emergence of transcriptional identity in somatosensory neurons. *Nature* 577, 392-398.
- Sherrington, C. (1907). On the proprio-ceptive system, especially in its reflex aspects. *Brain* 29, 467–482.
- Shrestha, B.R., Chia, C., Wu, L., Kujawa, S.G., Liberman, M.C., & Goodrich, L.V. (2018). Sensory neuron diversity in the inner ear is shaped by activity. *Cell* 174, 1229-1246.
- Snyder, M.E., Finlayson, M.O., Connors, T.J., Dogra, P., Senda, T., Bush, E., Carpenter, D., Marboe, C., Benvenuto, L., Shah, L., Robbins, H., Hook, J.L., Sykes, M., D'Ovidio, F., Bacchetta, M., Sonett, J.R., Lederer, D.J., Arcasoy, S., Sims, P.A., & Farber, D.L. (2019). Generation and persistence of human tissue-resident memory T cells in lung transplantation. *Science Immunology* 4, eaav5581.
- Sonner, M.J., Walters, M.C., & Ladle, D.R. (2017). Analysis of proprioceptive sensory innervation of the mouse soleus: A whole-mount muscle approach. *PLoS ONE* 12, e0170751.
- Sousa, V.H., Miyoshi, G., Hjerling-Leffler, J., Karayannis, T., & Fishell, G. (2009). Characterization of Nkx6-2-derived neocortical interneuron lineages. *Cerebral Cortex* 19, Suppl 1, i1-10.
- Taniguchi, H., He, M., Wu, P., Kim, S., Paik, R., Sugino, K., Kvitsani, D., Fu, Y., Lu, J., Lin, Y., Miyoshi, G., Shima, Y., Fishell, G., Nelson, S.B., & Huang, Z.J. (2011). A resource of cre driver lines for genetic targeting of GABAergic neurons in cerebral cortex. *Neuron* 71, 995-1013
- Tasic, B., Menon, V., Nguyen, T.N., Kim, T.K., Jarsky, T., Yao, Z., Levi, B., Gray, L.T., Sorensen, S.A., Dolbeare, T., Bertagnolli, D., Goldy, J., Shapovalova, N., Parry, S., Lee, C., Smith, K., Bernard, A., Madisen, L., Sunkin, S.M., Hawrylycz, M., Koch, C., & Zeng, H. (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosc.* 19, 335-346.
- Tasic, B., Yao, Z., Graybuck, L.T., Smith, K.A., Nguyen, T.N., Bertagnolli, D., Goldy, J., Garren, E., Economo, M.N., Viswanathan, S., Penn, O., Bakken, T., Menon, V., Miller, J., Fong, O., Hirokawa, K.E., Lathia, K., Rimorin, C., Tieu, M., Larsen, R., Casper, T., Barkan, E., Kroll, M., Parry, S., Shapovalova, N.V., Hirschstein, D., Pendergraft, J., Sullivan, H.A., Kim, T.K., Szafer, A., Dee, N., Groblewski, P., Wickersham, I., Cetin, A.,



- Harris, J.A., Levi, B.P., Sunkin, S.M., Madisen, L., Daigle, T.L., Looger, L., Bernard, A., Phillips, J., Lein, E., Hawrylycz, M., Svoboda, K., Jones, A.R., Koch, C., & Zeng, H. (2018). Shared and distinct transcriptomic cell types across neocortical areas. *Nature* 563, 72-78.
- Tian, C., Zhu, R., Zhu, L., Qiu, T., Cao, Z., & Kang, T. (2014). Potassium channels: Structures, diseases, and modulators. *Chemical Biology & Drug Design* 83, 1-26.
- Tourtellotte, W.G., and Milbrandt, J. (1998) Sensory ataxia and muscle spindle agenesis in mice lacking the transcription factor *Egr3*. *Nat. Genetics* 20, 87-91.
- Usoskin, D., Furlan, A., Islam, S., Abdo, H., Lönnerberg, P., Lou, D., Hjerling-Leffler, J., Haeggström, J., Kharchenko, P.V., Kharchenko, O., Linnarsson, S., & Ernfors, P. (2015). Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. *Nat. Neurosci.* 18, 145-153.
- Vincent, J.A., Gabriel, H.M., Deardorff, A.S., Nardelli, P., Fyffe, R.E.W., Burkholder, T., & Cope, T.C. (2017). Muscle proprioceptors in adult rat: Mechanosensory signaling and synapse distribution in spinal cord. *J. Neurophys.* 118, 2687-2701.
- Watanabe, E., Fujikawa, A., Matsunaga, H., Yasoshima, Y., Sako, N., Yamamoto, T., Saegusa, C., Noda, M. (2000). Nav2/NaG channel is involved in control of salt-intake behavior in the CNS. *J. Neurosci.* 20, 7743-7751.
- Windhorst, U. (2007) Muscle proprioceptive feedback and spinal networks. *Brain Research Bulletin* 73, 155-202.
- Woo S.H., Lukacs, V., de Nooij, J.C., Zaytseva, D., Criddle, C.R., Francisco, A., Jessell, T. M., Wilkinson, K. A., & Patapoutian, A. (2015). Piezo2 is the principal mechanotransduction channel for proprioception. *Nat. Neurosci.* 18, 1756-1762.
- Wu, D., Schieren, I., Qian, Y., Zhang, C., Jessell, T.M., & de Nooij, J.C. (2019). A role for sensory end organ-derived signals in regulating muscle spindle proprioceptor phenotype. *J. Neurosci.* 39, 4252-4267.
- Zeisel, A., Hochgerner, H., Lönnerberg, P., Johnsson, A., Memic, F., van der Zwan, J., Häring, M., Braun, E., Borm, L.E., La Manno, G., Codeluppi, S., Furlan, A., Lee, K., Skene, N., Harris, K.D., Hjerling-Leffler, J., Arenas, E., Ernfors, P., Marklund, U., & Linnarsson, S. (2018). Molecular architecture of the mouse nervous system. *Cell* 174, 999-1014.
- Zhang, F.X., Ge, S.N., Dong, Y.L., Shi, J., Feng, Y.P., Li, Y., Li, Y.Q., & Li, J.L. (2018). Vesicular glutamate transporter isoforms: The essential players in the somatosensory systems. *Progr. in Neurobiol.* 171, 72-89.
- Zheng, Y., Liu, P., Bai, L., Trimmer, J.S., Bean, B.P., & Ginty, D.D. (2019). Deep sequencing of somatosensory neurons reveals molecular determinants of intrinsic physiological properties. *Neuron* 103, 598-616.

Figure 1

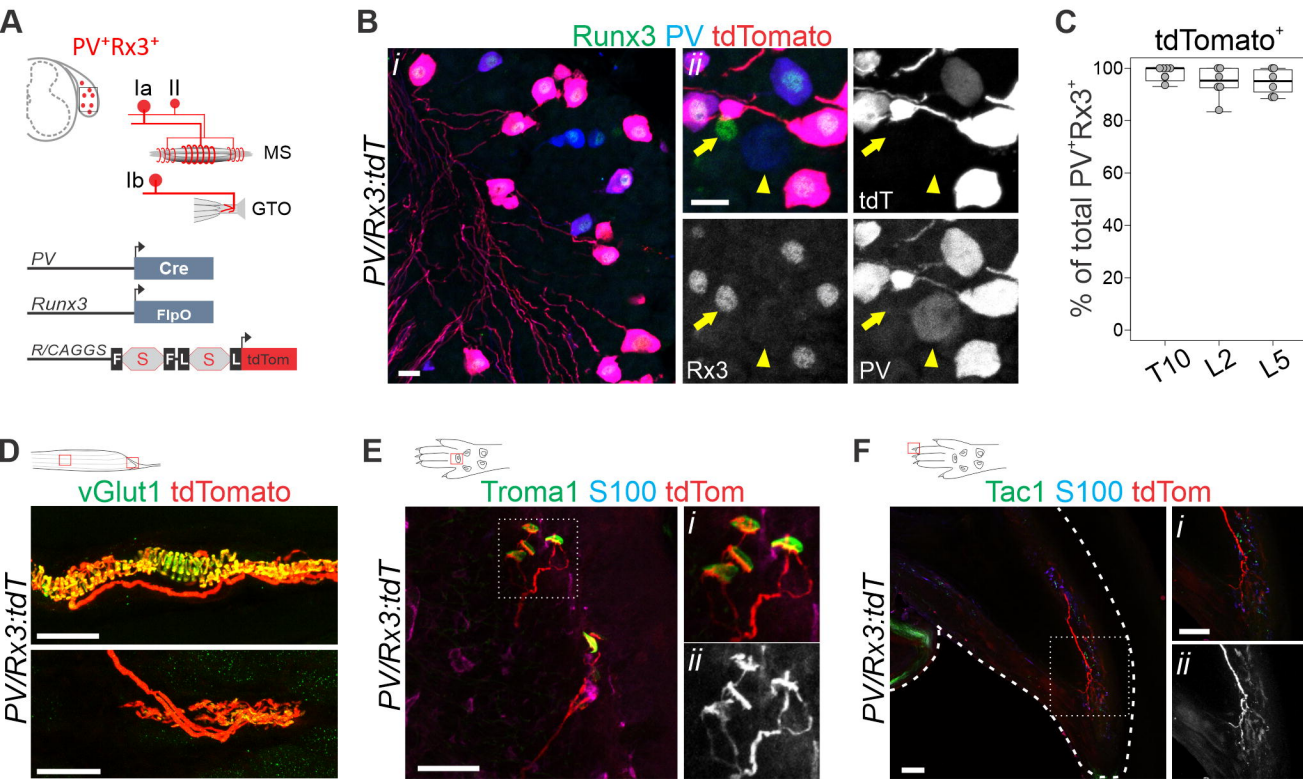


Figure 2

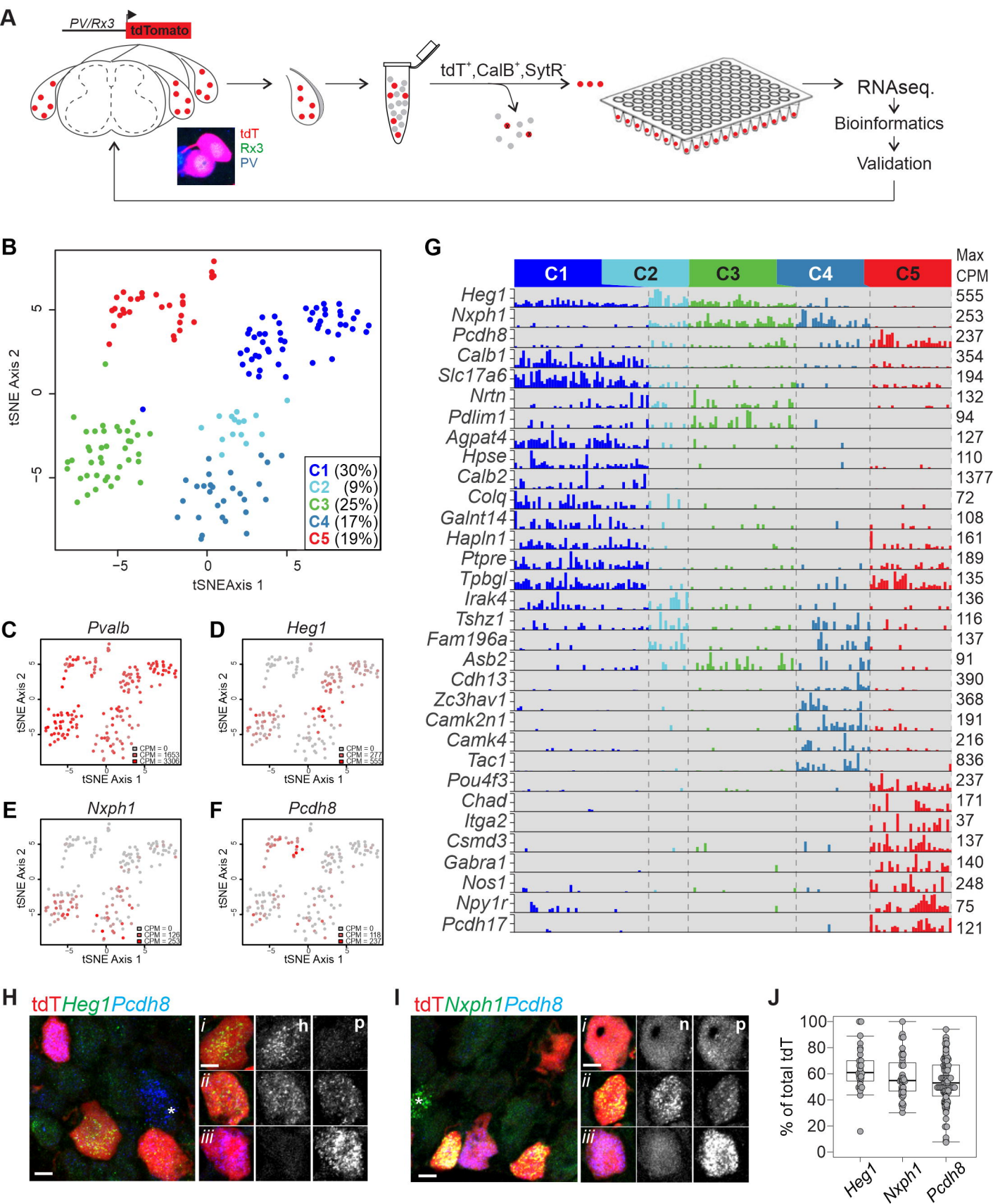


Figure 3

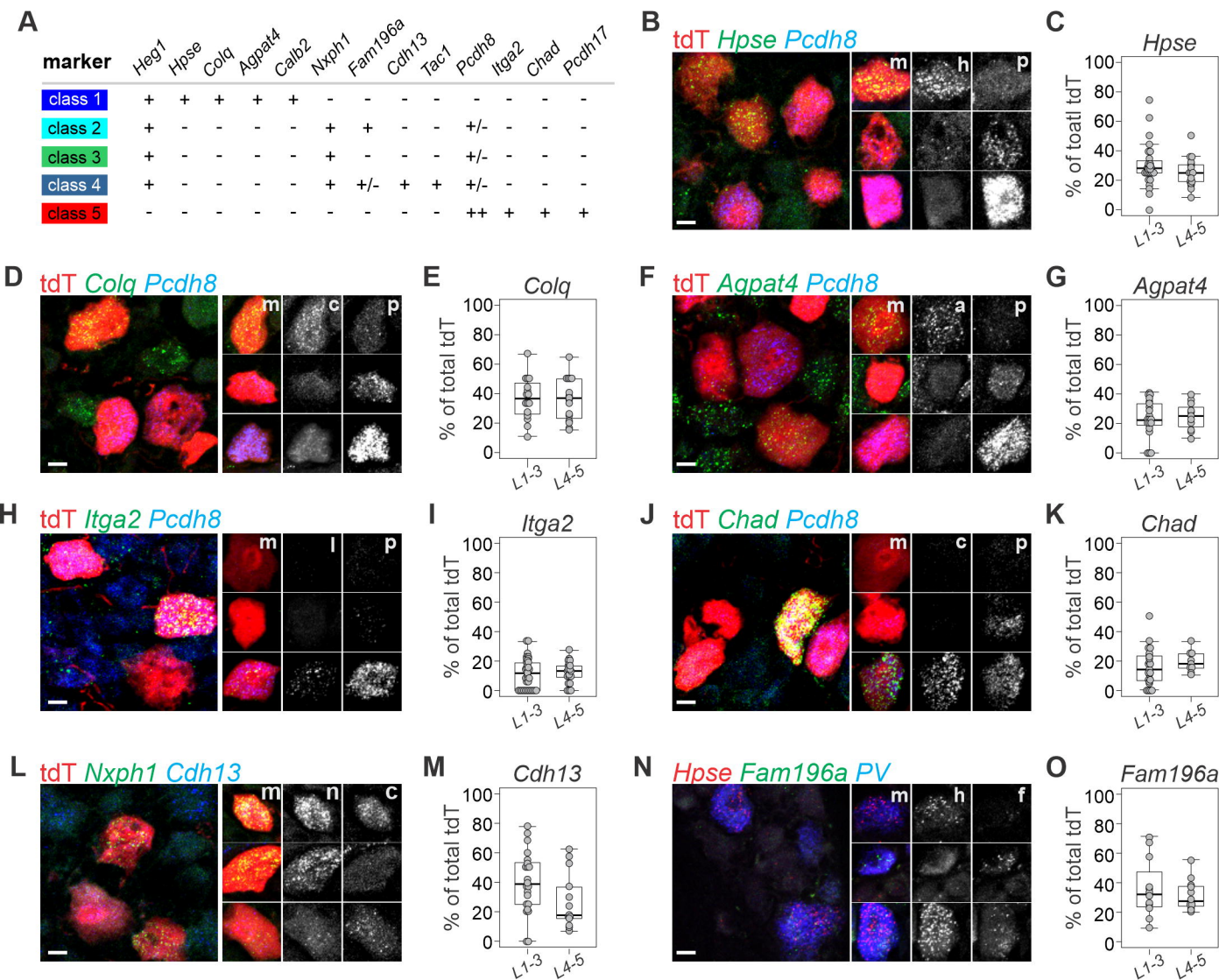




Figure 4

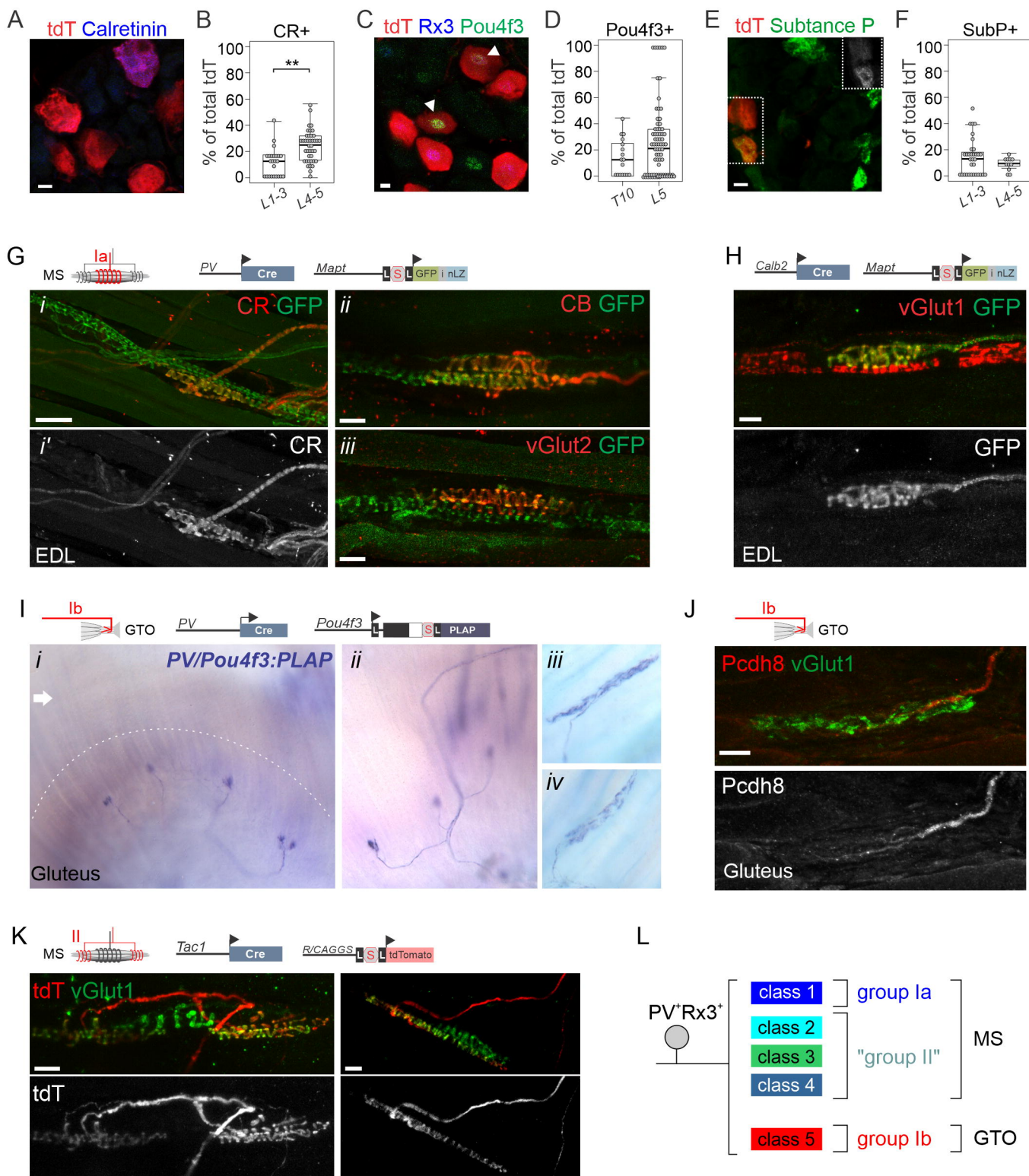


Figure 5

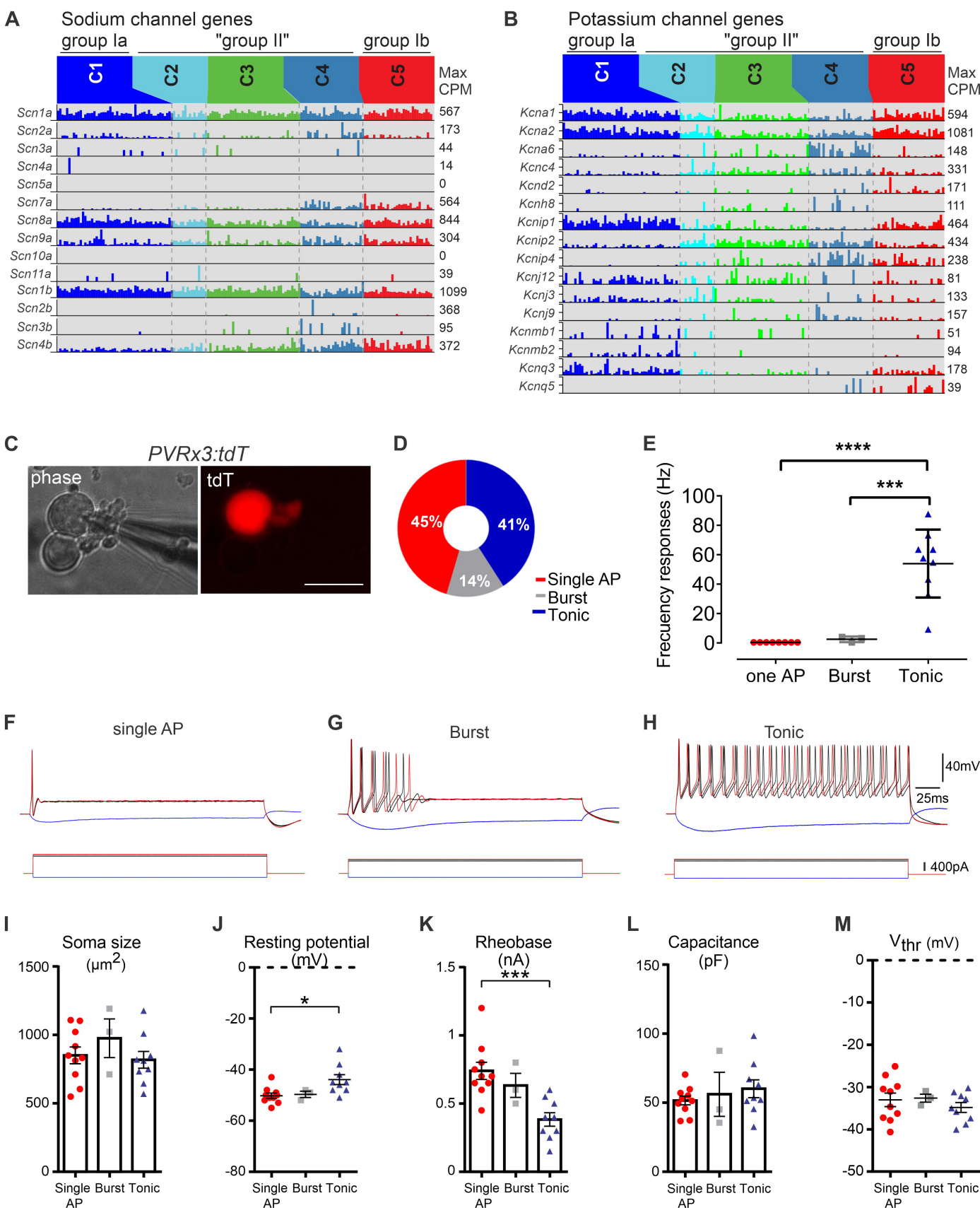


Figure 6

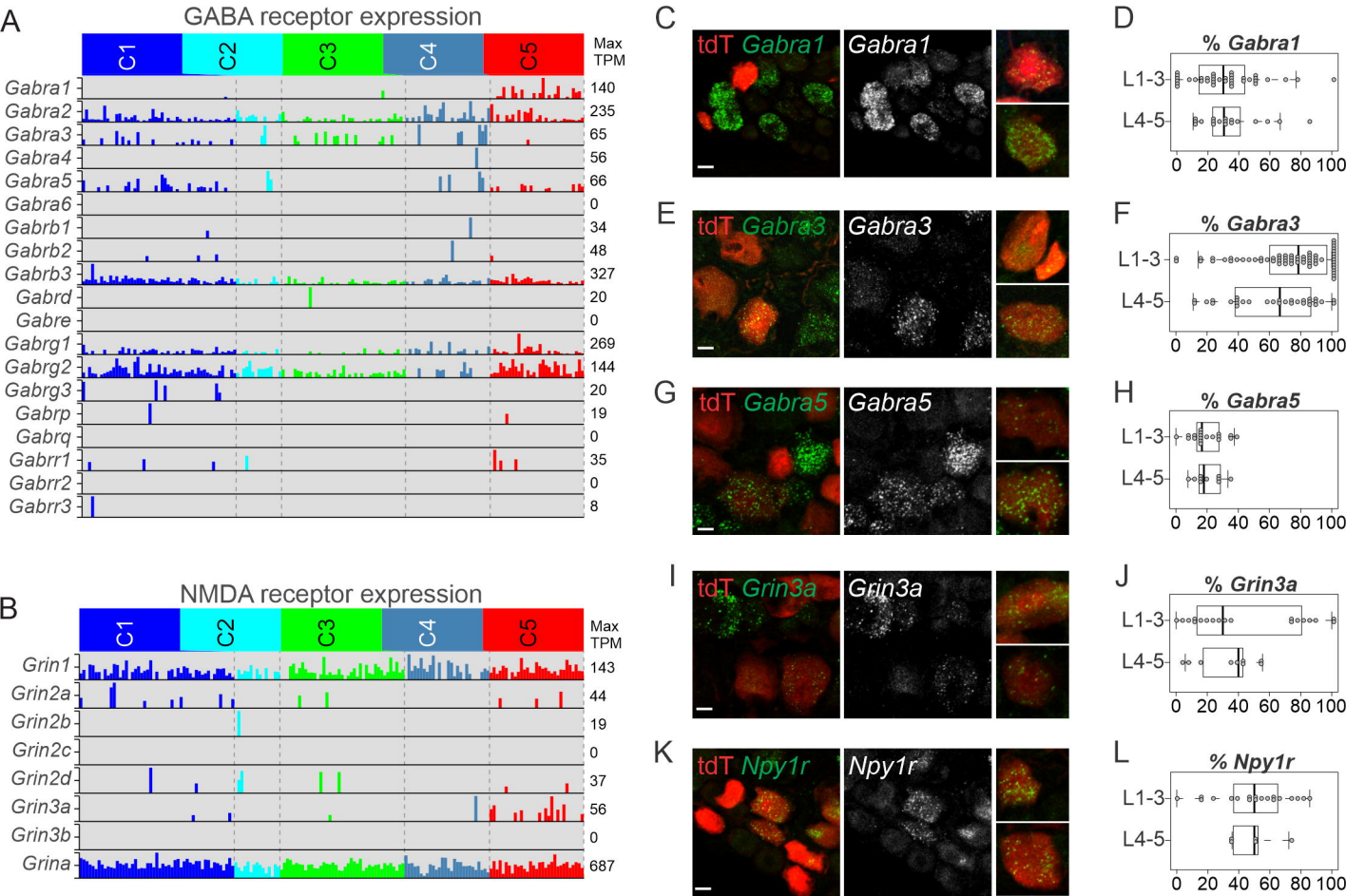




Figure 7

