

Transcriptional profiling at whole population and single cell levels reveals
somatosensory neuron molecular diversity

Isaac M. Chiu^{1,2}, Lee B. Barrett¹, Erika Williams³, David E. Strochlic³, Seungkyu Lee¹, Andy D. Weyer⁴, Shan Lou⁵, Gregory Bryman¹, David P. Roberson¹, Nader Ghasemlou¹, Cara Piccoli¹, Ezgi Ahat¹, Victor Wang¹, Enrique J. Cobos^{1,6}, Cheryl L. Stucky⁴, Qiufu Ma⁵, Stephen D. Liberles³, and Clifford J. Woolf¹

¹F.M. Kirby Neurobiology Center, Boston Children's Hospital and Department of Neurobiology, Harvard Medical School, Boston, MA, USA.

²Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA, USA.

³Department of Cell Biology, Harvard Medical School, Boston, MA, USA.

⁴Department of Cell Biology, Neurobiology, and Anatomy, Medical College of Wisconsin, Milwaukee, WI, USA.

⁵Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA.

⁶Department of Pharmacology and Neurosciences Institute, University of Granada, Granada, Spain.

Corresponding Authors:

Isaac M. Chiu, email: Isaac_chiu@hms.harvard.edu

Clifford J. Woolf, e-mail: Clifford.woolf@childrens.harvard.edu

Abstract

The somatosensory nervous system is critical for the organism's ability to respond to mechanical, thermal, and nociceptive stimuli. Somatosensory neurons are functionally and anatomically diverse but their molecular profiles are not well-defined. Here, we used transcriptional profiling to analyze the detailed molecular signatures of dorsal root ganglion (DRG) sensory neurons. We used two mouse reporter lines and surface IB4 labeling to purify three major non-overlapping classes of neurons: 1)IB4⁺SNS-Cre/TdTomato⁺, 2)IB4⁻SNS-Cre/TdTomato⁺, and 3)Parv-Cre/TdTomato⁺ cells, encompassing the majority of nociceptive, pruriceptive, and proprioceptive neurons. These neurons displayed distinct expression patterns of ion channels, transcription factors, and GPCRs. Highly parallel qRT-PCR analysis of 334 single neurons selected by membership of the three populations demonstrated further diversity, with unbiased clustering analysis identifying six distinct subgroups. These data significantly increase our knowledge of the molecular identities of known DRG populations and uncover potentially novel subsets, revealing the complexity and diversity of those neurons underlying somatosensation.

Introduction

The somatosensory nervous system comprises diverse neuronal subsets with distinct conduction properties and peripheral and central innervation patterns, including small-diameter, unmyelinated C-fibers, thinly myelinated A δ -fibers, and large-diameter, thickly myelinated A α / β -fibers (Abraira and Ginty, 2013, Basbaum et al., 2009). Distinct sets of somatosensory neurons are thought to mediate different functional modalities, such as tactile sensation, proprioception, pruriception and nociception. During development, precise expression of neurotrophic receptors and transcription factors at different times controls the differentiation and connectivity of these diverse sensory afferent populations (Marmigere and Ernfors, 2007, Abraira and Ginty, 2013). Detection of thermal, mechanical, and chemical stimuli in the external or internal environment by the somatosensory neurons is mediated by expression of specific molecular transducers at their peripheral nerve terminals. For example, transient receptor potential (TRP) ion channels are activated in response to heat, cold, reactive chemicals, leading to cation influx and action potential generation (Basbaum et al., 2009, Dib-Hajj et al., 2010, Dubin and Patapoutian, 2010, Julius, 2013). Given the high degree of cellular diversity of the somatosensory system defined at developmental, anatomical, and functional levels, a classification scheme of different somatosensory neuron subtypes based on the comprehensive set of genes they express is so far lacking. Determining the detailed molecular organization of specific somatosensory neuron subtypes is however necessary for our understanding of their specification, normal function and contribution to disease.

Cell-type specific transcriptome analysis is increasingly recognized as important for the molecular classification of neuronal populations in the brain and spinal cord (Okaty et al., 2011). Fluorescence activated cell sorting (FACS) and other neuron purification strategies coupled with

transcriptional profiling by microarray analysis or RNA sequencing has allowed detailed molecular characterization of discrete populations of mouse forebrain neurons (Sugino et al., 2006), striatal projection neurons (Lobo et al., 2006), serotonergic neurons (Wylie et al., 2010), corticospinal motor neurons (Arlotta et al., 2005), callosal projection neurons (Molyneaux et al., 2009), proprioceptor lineage neurons (Lee et al., 2012), and electrophysiologically distinct neocortical populations (Okaty et al., 2009). These data have uncovered novel molecular insights into neuronal function. Transcriptional profiling technology at the single cell level is transforming our understanding of the organization of tumor cell populations and cellular responses in the immune system (Patel et al., 2014, Shalek et al., 2014), and has begun to be applied to neuronal populations (Citri et al., 2012, Mizeracka et al., 2013). This technology has been proposed as a useful approach to begin mapping cell diversity in the mammalian CNS (Wichterle et al., 2013).

To begin to define the molecular organization of the somatosensory system, we have performed cell-type specific transcriptional profiling of dorsal root ganglion (DRG) neurons at both whole population and single cell levels. Using two reporter mice, SNS-Cre/TdTomato and Parv-Cre/TdTomato, together with surface Isolectin B4-FITC staining, we identify three major, non-overlapping populations of DRG neurons encompassing almost all C-fibers and many A-fibers. SNS-Cre is a BAC transgenic mouse line expressing Cre under the *Scn10a* (Nav1.8) promoter (Agarwal et al., 2004) which has been shown to encompass DRG and trigeminal ganglia nociceptor lineage neurons, and in conditional gene ablation studies affects thermosensation, itch, and pain (Liu et al., 2010, Lopes et al., 2012, Lou et al., 2013). A widely used Nav1.8-Cre knock-in mouse line also exists (Abrahamsen et al., 2008, Stirling et al., 2005), but differs to some extent from the transgenic SNS-Cre mouse line. We find, for example, that

SNS-Cre/TdTomato reporter mice label 82% of total DRG neurons, which is slightly greater than Nav1.8-Cre/TdTomato reporter mice (75%) (Shields et al., 2012), implying capture of a larger neuronal population. Both the SNS-Cre lineage and Nav1.8-Cre lineage neurons include a large proportion of C-fibers and a smaller population of NF200⁺ A-fibers (Shields et al., 2012). As expected, the majority of TdTomato⁺ cells (90%) in the SNS-Cre/TdTomato line expressed Scn10a transcript encoding Nav1.8 when tested by RNA *in situ* hybridization (Liu et al., 2010). Our second reporter line used Parv-Cre, a knock-in strain expressing Ires-Cre under the control of the Parvalbumin promoter, which has been used in the study of proprioceptive-lineage (large NF200⁺ A-fiber) neuron function (de Nooij et al., 2013, Hippenmeyer et al., 2005, Niu et al., 2013). Finally we used IB4, which labels the surface of non-peptidergic nociceptive neurons (Basbaum et al., 2009, Stucky et al., 2002, Vulchanova et al., 1998).

Using these mice and the labeling strategies, we were able to FACS purify three major, non-overlapping populations of somatosensory neurons: (1) IB4⁺SNS-Cre/TdTomato⁺, (2) IB4⁻SNS-Cre/TdTomato⁺, (3) Parv-Cre/TdTomato⁺ neurons, and analyze their whole transcriptome molecular signatures. Differential expression analysis defined transcriptional hallmarks in each for ion channels, transcription factors and G-protein coupled receptors. Further analysis of hundreds of single DRG neurons identifies distinct somatosensory subsets within the originally purified populations, which were confirmed by RNA *in situ* hybridization. Our analysis illustrates the enormous heterogeneity and complexity of neurons that mediate peripheral somatosensation, as well as revealing the molecular basis for their functional specialization.

Results

Characterization of distinct DRG neuronal subsets for molecular profiling

To perform transcriptional profiling of the mouse somatosensory nervous system, we labeled distinct populations of DRG neurons. We bred SNS-Cre or Parv-Cre mice with the Cre-dependent Rosa26-TdTomato reporter line (Madisen et al., 2010). In SNS-Cre/TdTomato and Parv-Cre/TdTomato progeny, robust fluorescence was observed in particular subsets of neurons in lumbar DRG (Figure 1-figure supplement 1).

We next analyzed the identity of the SNS-Cre/TdTomato⁺ and Parv-Cre/TdTomato⁺ DRG populations by costaining with a set of widely used sensory neuron markers; Isolectin B4 (IB4) (for non-peptidergic nociceptors), Neurofilament-200 kDa (NF200) (for myelinated A-fibers) calcitonin-gene related peptide (CGRP) (for peptidergic nociceptors), and Parvalbumin (for proprioceptors) (Figure 1a). IB4 labeled a DRG subset that was completely included within the SNS-Cre/TdTomato population (Figure 1b, 98±0.87% IB4⁺ were SNS-Cre/TdT⁺; Figure 1c, 28.0±1.8% SNS-Cre/TdT⁺ neurons were IB4⁺). By contrast, IB4 staining was effectively absent in the Parv-Cre/TdTomato population (Figure 1b, 1.18±1.35% IB4⁺ were Parv-Cre/TdT⁺). CGRP also fell completely within a subset of the SNS-Cre/TdTomato population and also was absent in the Parv-Cre/TdTomato population (Figure 1b, 99.4±0.4% CGRP⁺ were SNS-Cre/TdT⁺; 1.5±2.05% CGRP⁺ were Parv-Cre/TdT⁺; Figure 1c, 45.1±3.9% SNS-Cre/TdT⁺ were CGRP⁺). Neurofilament heavy chain 200 kDa (NF200) was expressed by the majority of the Parv-Cre/TdT⁺ population (Figure 1b, 96.1±1.9%), but only a small proportion of the SNS-Cre/TdT⁺ population (16.9±1.9%). Parvalbumin protein was expressed by the majority of Parv-Cre/TdT⁺ neurons (Figure 1c, 81.4±3.4%), but was absent in the SNS-Cre/TdT⁺ population (Figure 1c, 0.8±0.2%). In the spinal cord, SNS-Cre/TdTomato fibers mostly overlapped with

CGRP and IB4 central terminal staining in superficial dorsal horn layers (Figure 1-figure supplement 1). By contrast, Parv-Cre/TdTomato fibers extended into deeper dorsal horn laminae, Clark's Nucleus, and the ventral horn (Figure 1-figure supplement 1). Taken together, these observations suggest that these two lineage reporter lines labeled two distinct populations of primary sensory afferents and the SNS-Cre/TdTomato population includes several subsets that can be partly delineated by IB4 staining (Venn diagram, Figure 1d). By NeuN staining, SNS-Cre/TdTomato labeled $82 \pm 3.0\%$ of all DRG neurons, while Parv-Cre/TdTomato labeled $12.5 \pm 1.7\%$ DRG neurons, indicating that the majority of primary afferents are included within these two populations. For transcriptome profiling analysis, we purified three non-overlapping sets of DRG neurons: 1) $IB4^+SNS\text{-}Cre/TdTomato^+$, 2) $IB4^-SNS\text{-}Cre/TdTomato^+$ and 3) $Parv\text{-}Cre/TdTomato^+$ neurons (Venn Diagram, Figure 1e).

Electrophysiology of somatosensory subsets

We analyzed the electrophysiological characteristics of the TdTomato-labeled populations using whole cell patch clamp recordings. Resting membrane potential was similar between SNS-Cre/TdT⁺ (-56 ± 5.2 mV) and Parv-Cre/TdT⁺ cells (-60 ± 5 mV). Analyzing firing characteristics, SNS-Cre/TdT⁺ neurons displayed broad, TTX-resistant action potentials, while Parv-Cre/TdT⁺ neurons all showed narrow, TTX-sensitive action potentials (Figure 2a). These differences were reflected in significant differences in action potential half-width ($p=0.0001$ by t-test, Figure 2b). Parv-Cre/TdT⁺ neurons also showed significantly larger capacitance than SNS-Cre/TdT⁺ neurons ($P=0.0017$ by t-test, Figure 2c). These differences in firing properties are likely due to distinct ion channel expression patterns. TTX-resistant action potentials are characteristics of Nav1.7 and Nav1.8 nociceptor-lineage neurons (Bean, 2007, Dib-Hajj et al., 2010, Shields et al., 2012).

Thus, both anatomically and by neurophysiology, these two lineage reporter mice labeled distinct DRG subsets.

FACS purification of DRG neuron populations

We performed FACS purification of distinct neuronal populations isolated from both adult (7-20 week old) male and female mice. To avoid multiple rounds of amplification of small quantities of RNA, which would arise from less-abundant neuronal populations such as Parv-cre/TdT⁺, we chose to pool DRGs from cervical to lumbar regions (C1-L6). DRG cells were enzymatically dissociated and subjected to flow cytometry following DAPI staining to exclude dead cells, and gating on TdTomato^{hi} populations (Figure 3). This allowed for purification of TdTomato⁺ neuronal somata with minimal contamination from fluorescent axonal debris and non-neuronal cells (Figure 3a). Analysis of our flow cytometry data showed SNS-Cre/TdT⁺ vs. Parv-Cre/TdT⁺ DRG cells matched the proportions ascertained by NeuN co-staining in DRG sections (Figure 3b). It also illustrates that a large percentage of DAPI⁺ live cells are non-neuronal. IB4-FITC surface staining allowed us to simultaneously purify the distinct IB4⁺ and IB4⁻ subsets within the SNS-Cre/TdT⁺ population (Figure 3c). Forward and side scatter light scattering properties reflect cell size and internal complexity, respectively. SNS-Cre/TdT⁺ neurons displayed significantly less forward scatter and side scatter than Parv-Cre/TdT⁺ neurons (Figure 3-figure supplement 1). For RNA extraction, DRG populations were sorted directly into Qiazol to preserve transcriptional profiles at the time of isolation.

Transcriptional profile comparisons of purified neurons vs. whole DRG

In total, 14 somatosensory neuron samples were FACS purified consisting of 3-4 biological replicates/neuron population (Table 1). We also analyzed RNA from whole DRG tissue for comparison with the purified neuron samples. Because of the small numbers of cells from

individual sensory ganglia and to eliminate the need for non-linear RNA amplification, total DRGs from three mice were pooled for each sample; following purification, RNA was hybridized to Affymetrix microarray genechips for transcriptome analysis.

Transcriptome comparisons showed few molecular profile differences between biological replicates, but very large inter-population differences (Figure 3-figure supplement 2). Importantly, whole DRG molecular profiles differed substantially from the FACS purified neurons. Myelin associated transcripts (Mpz, Mag, Mpz, Pmp2) that are expressed by Schwann cells, for example, showed significantly higher expression in whole DRG tissue than in purified subsets when expressed as absolute robust multi-array average normalized expression levels (Figure 3-figure supplement 2). Known nociceptor-associated transcripts (Trpv1, Trpa1, Scn10a, Scn11a) were enriched in SNS-Cre/TdT⁺ profiles, and known proprioceptor-associated transcripts (Pvalb, Runx3, Etv1, Ntrk3) were enriched in Parv-Cre/TdT⁺ profiles (Figure 3-figure supplement 2). Fold-change vs. Fold-change plots illustrated the transcriptional differences between purified neurons and whole DRG RNA (Figure 3-figure supplement), supporting the validity of FACS purification to analyze distinct somatosensory populations compared to whole tissue analysis, which includes mixtures of several neuron populations and many non-neuronal cells.

Hierarchical clustering and Principal components analysis

Hierarchical clustering of molecular profiles from IB4⁺SNS-Cre/TdT⁺, IB4⁻SNS-Cre/TdT⁺, and Parv-Cre/TdT⁺ neuron populations revealed a distinct segregation of these three DRG neuronal subsets, and large blocks of transcripts were enriched for each population (Heat-map, Figure 4a). Principal Components Analysis (PCA) showed clustering of samples into distinct groups. IB4⁻SNS-Cre/TdT⁺ neurons differed from Parv-Cre/TdT⁺ neurons along Principal Component 2

(14.49% variation, Figure 4b); IB4⁺ and IB4⁻SNS-Cre/TdT⁺ neurons differed along Principal Component 3 (2.58% variation, Figure 4b).

Somatosensory transcript expression across neuronal subsets

We next analyzed gene expression patterns for 36 key known functional mediators of somatosensation (Figure 5). The IB4⁺ and IB4⁻ SNS-Cre/TdTomato⁺ neuronal subsets were enriched for the TRP channels, neuropeptides, and G-protein coupled receptors (GPCRs) that are involved in thermosensation, nociception, and pruriception. B-type natriuretic polypeptide b (Nppb), recently identified to mediate itch signaling (Mishra and Hoon, 2013), was highly expressed by IB4⁻SNS-Cre/TdT⁺ neurons (>800 normalized expression), while gastrin-releasing peptide (GRP), also linked to pruriception (Sun and Chen, 2007), was not expressed at detectable levels in any of the purified subsets (<100 normalized expression). Piezo2 (Fam38b), a mechanosensory ion channel (Woo et al., 2014, Maksimovic et al., 2014, Coste et al., 2010), was highly expressed in all somatosensory subsets (>4000 normalized expression), with enrichment in SNS-Cre/TdT⁺ relative to Parv-Cre/TdT⁺ neurons. By contrast, Trpc1, a channel linked to cutaneous mechanosensation (Garrison et al., 2012) was enriched in Parv-Cre/TdT⁺ neurons, indicating a potential role in proprioception. C-tactile afferent markers Slc17a8 (Vglut3) and Th (Tyrosine hydroxylase) (Li et al., 2011, Seal et al., 2009) were enriched in IB4⁻SNS-Cre/TdT⁺ neurons, while Mrgprb4 (Vrontou et al., 2013) was enriched in IB4⁺SNS-Cre/TdT⁺ neurons. Mrgprd and Runx1 were enriched in IB4⁺SNS-Cre/TdT⁺ neurons, which are known markers of non-peptidergic nociceptors (Wang and Zylka, 2009, Chen et al., 2006). Expression of neurotrophic factor receptors (Ntrk1, Ntrk2, Ntrk3, Gfra2, Gfra3, Ret) also showed distinct segregation patterns among the IB4⁺SNS-Cre/TdT⁺, IB4⁻SNS-Cre/TdT⁺ and Parv-Cre/TdT⁺ populations. Pvalb, Cadherin 12 (Cdh12), Vglut1 (Slc17a7), and transcription factors (Runx3,

Etv1, Etv4) were highly enriched in Parv-Cre/TdT⁺ neurons relative to the other two subsets. The distribution of these known mediators or markers of somatosensory function reveals differences and similarities between the three populations that reflect their functional specialization and modality responsiveness.

Functional neuronal mediators segregate across somatosensory subsets

We next focused our analysis on the expression patterns of those families of genes that mediate different general neuronal functions. Neurons exhibit specific firing properties due to the coordinated activity of different voltage-gated ion channels (Dib-Hajj et al., 2010, Dubin and Patapoutian, 2010, Bean, 2007). We found that many voltage-gated sodium, calcium, potassium, and chloride channels were differentially expressed in the three purified DRG populations (Figure 6a-d). Focusing on sodium channels, Scn9a (Nav1.7), Scn10a (Nav1.8), and Scn11a (Nav1.9) were enriched both in the IB4⁺ and IB4⁻SNS-Cre/TdT⁺ populations (Figure 6a), agreeing with known roles in nociception (Dib-Hajj et al., 2010). Scn1a (Nav1.1), Scn8a (Nav1.6), and sodium channel beta subunits Scn1b, Scn4b were mainly expressed in Parv-Cre/TdT⁺ neurons (Figure 6a). Voltage-gated calcium channels, including L-type, N-type, and T-type channels, also showed differential expression (Figure 6b). SNS-Cre/TdT⁺ neurons were highly enriched for Cacna2d1 (α 2 δ 1) and for Cacna2d2 (α 2 δ 2), the pharmacological targets of gabapentin and pregabalin (Field et al., 2006, Wang et al., 1999, Patel et al., 2013); unexpectedly, Parv-Cre/TdT⁺ neurons were enriched for Cacna2d3 (α 2 δ 3) (Figure 6b), which contributes to heat nociception via supraspinal expression (Neely et al., 2010). Voltage-gated potassium channels showed perhaps the most striking expression patterns across somatosensory subsets (Top 60 most variably expressed shown in Figure 6c). Kcns1 (Kv9.1), where a common variant is associated with increased pain and whose down-regulation in large sensory neurons is

associated with increased neuropathic pain (Costigan et al., 2010, Tsantoulas et al., 2012), was expressed in Parv-Cre/TdT⁺ neurons (Figure 6c). The IB4⁺SNS-Cre/TdT⁺, IB4⁻SNS-Cre/TdT⁺, and Parv-Cre/TdT⁺ populations each showed distinct enrichment patterns for potassium channel genes, most of which have not yet been analyzed yet in terms of somatosensory function. Voltage-gated chloride channels also showed distinct expression patterns, with differential regulation of Clcn and Tweety family ion channel transcripts (Figure 6d). Surprisingly, the Ca²⁺ activated chloride channel Ano1 (Anoctamin 1), which has recently been linked to heat nociception (Cho et al., 2012), was absent in SNS-Cre/TdT⁺ populations but present in Parv-Cre/TdT⁺ neurons (Figure 6d).

Transient receptor potential (TRP) channels, ligand-gated ion channels, and G-protein coupled receptors (GPCRs) are integral in the detection of specific environmental stimuli. These different types of molecular transducers showed substantial differential expression across the three purified DRG populations (Figure 6e and Figure 7a-b). In our dataset, IB4⁻SNS-Cre/TdT⁺ neurons were enriched for specific TRP channels (Trpv1, Trpm8, Trpc7, Trpm6), while IB4⁺SNS-Cre/TdT⁺ neurons were enriched for others (Trpv2, Trpm4, Trpa1, Trpm3, Trpc6, Trpc5, Trpc3), and only a few TRP channels showed expression in Parv-Cre/TdT⁺ neurons (Trpm2, Trpc1) (Figure 6e). Ligand-gated ion channels also play key roles in nociception or other somatosensory functions. We found diverse expression patterns for HCN channels, P2X channels, 5-HT receptors (Htr3a, Htr3b) ionotropic glutamate receptors, GABA receptors, and Glycine receptors across the neuronal populations (Top 60 most variably expressed ligand-gated channels, Figure 7a). GPCRs, including Mas-related GPCRs, muscarinic glutamate receptors, neuropeptide receptors, as well as some orphan receptors showed significant expression in different somatosensory subsets (Top 60 most variably expressed GPCRs, Figure 7b). Taken

together, these data show complex patterns of ligand-gated molecular transducer expression that could play roles in functional specialization and signaling.

We also found that many transcription factors were differentially expressed across these three neuron populations (Top 60 most variably expressed TFs, Figure 7c). Many of these have not yet been explored in the somatosensory system, and could play roles in neuronal differentiation and maintenance of cell-type specification during adulthood. For example, *Klf7* and *Isl2* were expressed at high levels and enriched in SNS-Cre/TdT⁺ neurons (>1.5 fold, p<0.01, >5000 expression). Based on this analysis, these two transcription factors have now been used to facilitate the trans-differentiation of fibroblasts into nociceptor-like neurons in conjunction with other transcription factors (Wainger et al., 2014).

Pair-wise enrichment analysis of neuron populations

To obtain statistically significant and unbiased enrichment analysis, we next performed pair-wise comparisons of the three major neuronal subclasses. We first compared SNS-Cre/TdTomato and Parv-Cre/TdTomato neurons, yielding many differentially expressed (DE) genes in each neuron subset (SNS-Cre/TdT⁺: 907 genes, Parv-Cre/TdT⁺: 774 genes, p<0.05, 2-fold; Figure 8a, Supplementary File 1). Specific Gene Ontology (GO) categories and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were significantly enriched (Figure 8b). The most differentially expressed GO categories were GO:0006816~calcium ion transport and GO:0006813~potassium ion transport. Thus, we focused on calcium ion channels and potassium ion channels using volcano plot comparisons (Figure 8c). SNS-Cre/TdT⁺ vs. Parv-Cre/TdT⁺ neurons showed differential regulation of various calcium channels (SNS-Cre/TdT⁺: 9 genes, Parv-Cre/TdT⁺: 3 genes, 2-fold, p<0.01, Figure 8c-i) and potassium channels (SNS-Cre/TdT⁺: 15 genes, Parv-Cre/TdT⁺: 12 genes, 2-fold, p<0.01, Figure 8c-ii). Based on statistical criteria of

fold-change>2, $p<0.01$, all differentially expressed TRP channels were enriched only in SNS-Cre/TdT⁺ neurons, which may relate to their importance in thermosensation and nociception (8 genes, Figure 8c-iii).

In a second pairwise comparison, IB4⁺ were compared with IB4⁻ SNS-Cre lineage neurons (Figure 9). This analysis yielded 258 significantly enriched transcripts in IB4⁺ vs. 492 in IB4⁻ neurons (2-fold, $p<0.05$, Figure 9a, Supplementary File 2). GO categories differentially regulated between IB4⁺ and IB4⁻ subsets included those for ion transport, cell adhesion, and synaptic transmission (Figure 9b). Volcano plot analysis shows significant differential expression of ion channels between these two subsets (IB4⁻SNS-Cre/TdT⁺: 29 genes, IB4⁺SNS-Cre/TdT⁺: 16 genes, $p<0.05$; 2 Fold, Figure 9c-i). P2rx3 (P2X3) and Scn11a (Nav1.9), ion channels known to mark non-peptidergic nociceptors, were enriched 1.8-fold and 2.0-fold, respectively in IB4⁺SNS-Cre/TdT⁺ neurons. Interestingly, we found even greater enrichment for Trpc3 (7.98-fold) and Trpc6 (7.67-fold) in this subset. Focusing on cell adhesion, volcano plots showed differential enrichment for Nrnx3, Nrcam, and Ncam2 in IB4⁻SNS-Cre/TdT⁺ neurons, and Cdh1, Pvr11 in IB4⁺SNS-Cre/TdT⁺ neurons (Figure 9c-ii). Next, we focused on GPCR expression differences (Figure 9c-iii). Mrgprd, a widely used marker of non-peptidergic neurons (Wang and Zylka, 2009), was enriched 20.6-fold in IB4⁺ neurons. Interestingly, we found several GPCRs that were enriched as Mrgprd in IB4⁺SNS-Cre/TdT⁺ neurons but have not yet been characterized for function in this subset, including Agtr1a (20.4-fold), Gpr116 (15.7-fold), Lpar3 (11.8-fold), and Lpar5 (12.6-fold).

Single cell analysis reveals log-scale gene expression heterogeneity

We next performed single cell level transcriptional analysis of the three globally characterized DRG populations using Fluidigm parallel qRT-PCR gene expression technology. Distinct

transcriptional hallmarks for each FACS purified population were first defined by their differential expression in the microarray datasets (3-fold enrichment, Figure 10). Taqman assays were chosen corresponding to these enriched markers, and including 2 housekeeping genes (Gapdh and Actb), a complete group of 80 assays was used for single cell expression profiling (Table 2). We first used these assays to analyze 100-cell and 10-cell FACS sorted groups of each neuronal population (Figure 10-figure supplement 1), confirming the enrichment of various marker transcripts.

We next FACS sorted individual IB4⁺SNS-Cre/TdT⁺, IB4⁻SNS-Cre/TdT⁺, and Parv-Cre/TdT⁺ neurons into 96-well plates for Fluidigm analysis. A total of 334 individual neurons were purified and analyzed (IB4⁺SNS-Cre/TdT⁺ cells, n=132; IB4⁻SNS-Cre/TdT⁺ cells, n=110; and Parv-Cre/TdT⁺ cells, n=92, Table 1).

We found that the expression levels for specific transcripts across single cell datasets often displayed a log-scale continuum (Figure 11). Some transcripts were highly enriched in one subset of single cells (e.g. Mrgprd, Trpv1, P2rx3), but were often nonetheless expressed at detectable levels in other neuronal groups. This continuum of gene expression made it difficult to set “thresholds” for assigning the presence or absence of a particular transcript. Thus, we focused our definition of distinct subgroups not by absolute proportion of positive gene expression but by correlative and aggregate analysis. Other transcripts (e.g. Nppb, Runx3, Cdh12) showed expression patterns restricted in one population and were not present in other populations.

Hierarchical clustering of single cell data reveals distinct subgroups

Spearman-rank hierarchical clustering was performed on the Fluidigm expression data normalized to gapdh expression (columns represent single cells, Figure 12). This analysis

revealed a high degree of heterogeneity of transcriptional expression across the three DRG populations. The vast majority of single cells showed distinct patterns of expression of at least one neuronal transcript, including voltage-gated ion channels (Scn10a, Scn11a, Kcnc2, Kcnc1), ligand-gated channels (P2rx3, Trpv1, Trpa1), and Parvalbumin (Pvalb) indicating minimal amplification noise (Figure 12-figure supplement 1). Unbiased spearman rank analysis revealed seven distinct neuronal subgroups (Figure 12). Six out of seven groups had 24 or more individual cells (group I, 115 cells; group II, 50 cells; group III, 4 cells; group IV, 24 cells; group V, 24 cells; group VI, 24 cells; group VII, 93 cells). We chose one level of sample segregation to analyze, but other cellular subclasses are likely present at lower levels of clustering (Figure 12). Importantly, when hierarchical clustering was performed on data normalized to Actb, neuronal subgroups based on gapdh normalization segregated in a similar manner (data not shown). Principal components analysis showed distinct separation of the single cell subgroups along different principal components (Figure 13a), with Groups I and VII on disparate arms of PC2 (~5% variation), while Group V neurons segregated along PC3 (~1.88% variation). Parv-Cre/TdT⁺ neurons mainly fell within group VII (96.7% of the cells, Figure 13b). IB4⁺SNS-Cre/TdT⁺ and IB4⁻SNS-Cre/TdT⁺ neurons were distributed among subgroups II-VI (Figure 13b). Therefore, this analysis has uncovered potentially novel subgroups distributed across the SNS-Cre/TdT⁺ population that are not captured by the presence or absence of IB4 staining.

Major characteristics of distinct single cell subgroups

We next analyzed the major characteristics of each DRG single cell subgroup (Figure 12). Group I neurons were mostly IB4⁺ nociceptors enriched for Pr2x3, Scn11a, and Mrgprd, markers for non-peptidergic nociceptors. Our analysis found a large number transcriptional hallmarks for Group I neurons that were as well enriched as the known marker genes, including

Grik1, Agtr1a, Pde11a, Ggta1, Prkcq, A3galt2, Ptgdr, Lpar5, Mmp25, Lpar3, Casz1, Slc16a12, Lpyd1, Trpc3, Moxd1, Wnt2b (Figure 12, and Figure 12-supplement 2). Nearest neighbor analysis across all single cells found 13 transcripts with Pearson correlation>0.5 for Mrgprd, further showing a large cohort of genes that segregate in expression within group I neurons (Figure 14).

Group II neurons expressed high levels of Ntrk1 (Trka), Scn10a (Nav1.8), and Trpv1. We also found that they expressed significant levels of Aqp1 (Aquaporin 1), and a major proportion of Group II neurons also expressed Kcnv1 (Kv8.1). Group III consisted of only 4 cells and we thus did not consider it a true neuronal subclass.

Group IV neurons were characterized by the absence of Scn10a (Nav1.8) but the presence of Trpv1 expression (Figure 14-figure supplement 1). Although Group IV neurons were all labeled by SNS-Cre/TdTomato, they did not all show Scn10a gene expression, likely reflecting transient transcription of this transcript that is shutdown in some neurons during development (Liu et al., 2010).

Group V neurons were distinguished by Th (tyrosine hydroxylase) gene expression, a known marker for low-threshold C-mechanoreceptors (Li et al., 2011). Triple immunofluorescence with IB4 showed that TH fell mostly within the IB4⁻SNS-Cre/TdT⁺ subset (91.4±2.4% TH⁺ were IB4⁻SNS-Cre/TdT⁺, Figure 15-figure supplement 1). Th⁺ neurons also expressed high levels of Scn10a (Nav1.8) and Aqp1 (Aquaporin 1), but low/undetectable levels of Ntrk (Trka) and Trpv1 (Figure 14-figure supplement 1, 2).

Group VI neurons were a distinct population characterized by co-expression of Nppb and IL31ra (Figure 14). Nppb is a neuropeptide mediator of itch signaling from DRG neurons to spinal cord pruritic circuitry (Mishra and Hoon, 2013). IL31 is a T cell cytokine associated with

pruritus, and DRG neurons express the IL31 receptor (Sonkoly et al., 2006, Bando et al., 2006) Co-expression of IL31ra with Nppb suggests that these neurons may be specialized in mediating itch. Group VI neurons also showed high level expression of Scn10a, Scn11a, and Trpv1 relative to the other subsets (Figure 14-figure supplement 1, 2).

Group VII neurons consisted of 93 cells comprising the majority of Parv-Cre/TdT⁺ sorted single cells (Figure 12-13). These neurons showed high expression of Parvalbumin (Pvalb) and osteopontin (Spp1), Cadherin 12 (Cdh12) as well as proprioceptor associated transcription factors (Etv1, Runx3). Nearest neighbor analysis of Pvalb gene expression showed 13 transcripts with Pearson correlation>0.5 (Figure 14). These include a set of distinct ion channels (Kcnc1, Ano1), GPCRs (Pth1r, Gpcr5b), and other genes (Ndstr4, Car2, Wnt7a) that have not been functionally characterized in this subset.

Fluorescence *ISH* analysis of subgroup-specific characteristics

RNA *in situ* hybridization (ISH) was used to confirm specific localization of novel Group I, VI, and VII enriched transcripts (Table 3). The Group I marker Lysophosphatidic acid receptor 3 (Lpar3) labeled a subset of SNS-Cre/TdT⁺ neurons that did not overlap with Parv-Cre/TdT⁺ expression (Figure 15). We found similar results for Prkce (PKC θ), another Group I marker (Figure 15-figure supplement 2). The Group VI marker Il31ra also labeled a distinct subset of SNS-Cre/TdT⁺ neurons and did not colocalize with Parv-Cre/TdT⁺ neurons (Figure 15). By contrast, the group VII marker Gpcr5b did not stain SNS-Cre/TdT⁺ neurons but co-localized well with Parv-Cre/TdT⁺ proprioceptors (Figure 15). Double ISH found that itch-related Group VI marker IL31ra did not colocalize with group I markers Prkce or Lpar3, nor with group VII marker Gpcr5b (Figure 15). In confirmation of the Fluidigm data, double ISH found that IL31ra colocalized well with Nppb (Figure 15), thus confirming co-expression of two itch-related

425 markers in the same neuronal subset. Thus, expression profiling at single cell resolution reveals
426 an unsuspected degree of complexity of sensory neurons with elucidation of many new markers
427 and of different neuronal subtypes.

428

Discussion

Mapping neuronal circuitry and defining the molecular characteristics of specific neurons is critical to understanding the functional organization of the nervous system. The somatosensory system, all of whose primary sensory neurons are of neural crest origin, is highly complex, innervating diverse peripheral tissues and encoding thermal, mechanical, and chemical modalities across a broad range of sensitivities, from innocuous to noxious with different dynamic ranges (Li et al., 2011, Dubin and Patapoutian, 2010, Basbaum et al., 2009, Marmigere and Ernfor, 2007). Sensory neurons are currently classified based on myelination and conduction properties (i.e. C-, A α / β - or A δ -fibers) or their selective expression of ion channels (e.g. Trpv1, P2rx3, Nav1.8), neurotrophin receptors (e.g. TrkA, TrkB, TrkC, Ret), cytoskeletal proteins (e.g. NF200, Peripherin), and GPCRs (e.g. Mrgprd, Mrgpra3). However, combining these different classification criteria can result in complex degrees of overlaps, making a cohesive categorization of distinct somatosensory populations challenging. Transcriptome-based analysis has become recently a powerful tool to understand the organization of complex populations, including subpopulations of CNS and PNS neurons (Zhang et al., 2014, Mizeracka et al., 2013, Okaty et al., 2011, Okaty et al., 2009, Molyneaux et al., 2009, Sugino et al., 2006, Lobo et al., 2006, Lee et al., 2012). In this study, we performed cell-type specific transcriptional analysis to better understand the molecular organization of the mouse somatosensory system.

Our population level analysis revealed the molecular signatures of three major classes of somatosensory neurons. There were vast transcriptional differences between SNS-Cre/TdTomato⁺ and Parv-Cre/TdTomato⁺ neurons, potentially reflecting their developmental specification into neurons with quite different functional attributes and targets. As SNS-Cre is expressed mainly within TrkA-lineage neurons (Liu et al., 2010, Abdel Samad et al., 2010),

while Parv-Cre is expressed mainly in proprioceptor-lineage neurons (Hippenmeyer et al., 2005), these two populations reflect archetypical C- and A α / β -fibers, respectively. *Bourane et al* previously performed SAGE analysis of TrkA deficient compared to wild-type DRGs, which revealed 240 differentially expressed genes and enriching for nociceptor hallmarks (Bourane et al., 2007). Our FACS sorting and comparative population analysis identified 1681 differentially expressed transcripts (2-fold), many of which may reflect the early developmental divergence and vast functional differences between these lineages. While C-fibers mediate thermosensation, pruriception and nociception from skin and deeper tissues, Parv-Cre lineage neurons mediate proprioception, innervating muscle spindles and joints (Dubin and Patapoutian, 2010, Marmigere and Ernfors, 2007). Almost exclusive TRP channel expression in SNS-Cre/TdT⁺ neurons vs. Parv-Cre/TdT⁺ neurons may relate to their specific thermosensory and chemosensory roles. We also found significant molecular differences between the IB4⁺ and IB4⁻ subsets of SNS-Cre/TdT⁺ neuronal populations. Our analysis identified many molecular hallmarks for the IB4⁺ subset (e.g. *Agtr1a*, *Cas2l*, *Slc16a12*, *Moxd1*) that are as enriched as the currently used markers (P2rx3, *Mrgprd*), but whose expression and functional roles in these neurons have not yet been characterized. This analysis of somatosensory subsets covered the majority of DRG neurons (~95%), with the exception of TrkB⁺ A δ cutaneous low-threshold fibers (Li et al., 2011), which are NF200⁺ but we find are negative for SNS-Cre/TdTomato and Parv-Cre/TdTomato (Data not shown).

Single cell analysis by parallel quantitative PCR of hundreds of neurons demonstrated large heterogeneity of gene expression within the SNS-Cre/TdT⁺ neuron population, much greater than the current binary differentiation of peptidergic or non-peptidergic IB4⁺ subclasses. Interestingly, we found a log-scale continuum for many transcripts, including nociceptive genes

(e.g. Trpv1, Trpa1) showing high expression in IB4⁺ and IB4⁻ subsets and with lower but not absent levels in Parv-Cre/TdT⁺ cells. This may reflect transcriptional shut-down of genes during differentiation. Unbiased hierarchical clustering analysis of single cell data revealed at least six distinct neuronal subgroups. These findings reveal new molecular characteristics for known neuron populations and also uncover novel neuron subsets: Group I neurons consist of Mrgprd⁺Nav1.8⁺P2rx3⁺Nav1.9⁺ cells, which are polymodal non-peptidergic C-fibers, for which we identify a panoply of new molecular markers. Group II consists of Trka^{hi}Nav1.8⁺Trpv1⁺Aquaporin⁺ neurons, matching known characteristics of thermosensitive C-fibers; many of these expressed Kcnv1. Group V consists of Th⁺Nav1.8⁺Trka⁻Trpv1⁻ cells, matching characteristics of C-fiber low-threshold mechanoreceptors (C-LTMRs) (Li et al., 2011). Group VII consists of Pvalb⁺Runx3⁺Etv1⁺ neurons, which are mostly proprioceptor-lineage neurons for which we identified 12 molecular markers. *Lee et al* recently performed transcriptome analysis of purified TrkC-lineage proprioceptive neurons in the presence or absence of NT-3 signaling (Lee et al., 2012) and we note that Group VII neurons were similar to TrkC lineage cells in gene expression (Pth1r, Runx3, Pvalb). Group IV consists of Trpv1⁺Nav1.8⁻ neurons, which may represent a unique functional subgroup; *Wood et al* found that mice depleted for Nav1.8-lineage neurons retained a TRPV1 responsive subset (Abrahamsen et al., 2008). We uncover a new subset of neurons, Group VI, which appears to represent pruriceptive neurons based on their co-expression of IL31ra and Nppb.

While preparing this manuscript, several papers performing expression profiling of postnatal adult somatosensory neurons were published (Goswami et al., 2014, Thakur et al., 2014, Usoskin et al., 2014). We note that each study utilized distinct methodologies from our work: *Goswami et al* profiled Trpv1-Cre/TdTomato⁺ neurons compared to Trpv1-diphtheria toxin

498 depleted whole DRG tissue (Goswami et al., 2014). *Thakur et al* performed magnetic bead
499 selection to remove DRG non-neuronal cells, performing RNA-seq on residual cells enriched for
500 neurons (Thakur et al., 2014). *Usokin et al* performed an elegant single cell RNA-seq on
501 hundreds of DRG neurons that were picked in an unbiased fashion robotically (Usoskin et al.,
502 2014). We believe that our study possesses has unique features and certain advantages, as well
503 as limitations, in relation to these studies. In our study, we performed whole population analysis
504 of 3 major DRG subsets, which we followed by single cell granular profiling of hundreds of cells
505 from the same populations. We believe advantages of beginning with a differential analysis of
506 well-defined populations is that this facilitates correlation of the data back to function and
507 enables a highly specific comparative analysis to be performed between major neuronal
508 populations. Further definition of each population by shifting to a single cell strategy then
509 allows identification of functionally defined groups of cells. The same advantages of a
510 population based strategy is also a caveat, in that it could introduce pre-determined bias, which
511 *Usokin et al* purposely avoided by randomly picking single DRG neurons as a starting point. We
512 note that our analysis is the only one so far to utilize parallel qRT-PCR of single cells, which we
513 demonstrate is able to detect log-scale differences in expression (Figure 11), and may have better
514 detection sensitivities than single cell RNA-seq. In a comparison of the overall datasets, we
515 produce some similar findings with *Usokin et al*, including the finding of a distinct pruriceptive
516 population (IL31ra⁺ Group VI). However, our analysis showed higher definition of markers
517 present in Group I and Group VII neurons, as well as Group IV neurons (which was not
518 previously described), while *Usokin et al* detected TrkB⁺ neurons whereas we did not, as these
519 cells are not included in our sorted populations. We believe that our study and these recently

published papers will be useful foundation and resource for future analysis of the molecular determinants of sensory neuron phenotype.

Somatosensory lineage neurons subserve multiple functions: nociceptive, thermoceptive, pruriceptive, proprioceptive, and tactile. It is likely that additional granular analysis at the single cell level will further refine these subsets and uncover new molecular subclasses of neurons. As genomic technologies and single cell sorting methodologies evolve current limitations (e.g. RNA quantity) will be overcome and future analysis of thousands of single cells from distinct anatomical locations, developmental time-points, or following injury/inflammation will begin to reveal even more critical information about the somatosensory system.

This transcriptional analysis illustrates an unsuspected degree of molecular complexity of primary sensory neurons within the somatosensory nervous system. Functional studies are now needed to analyze the roles of the many newly identified sensory genes in neuronal specification and action. As we begin to explore the function, connectivity and plasticity of the nervous system we need to recognize this needs a much more granular analysis of molecular identity, since even the presumed functionally relatively simple primary sensory neuron, is extraordinarily complex and diverse.

Materials and Methods

Mice.

Parvalbumin-Cre (Hippenmeyer et al., 2005), ai14 Rosa26TdTomato mice (Madisen et al., 2010) were purchased from Jackson Labs (Bar Harbor, ME) and bred in the animal facility at Boston Children's Hospital. SNS-Cre transgenic mice (Agarwal et al., 2004) were from Rohini Kuner (University of Heidelberg, Germany). All animal experiments were conducted according to institutional animal care and safety guidelines at Boston Children's Hospital and at Harvard Medical School.

Ethics Statement.

All studies were conducted under strict review and guidelines according to the Institutional Animal Care and Use Committee (IACUC) at Boston Children's Hospital, which meets the veterinary standards set by the American Association for Laboratory Animal Science (AALAS). The experiments were reviewed and approved by the IACUC at Boston Children's Hospital under animal protocol number 13-01-2342R.

Immunostaining and Microscopy.

Mice were transcardially perfused with PBS followed by 4% Paraformaldehyde/PBS (Sigma-Aldrich). DRG, spinal cord, plantar tissue were dissected, post-fixed for 2 hr, cryoprotected in 30% sucrose/PBS, and frozen at -80°C in Optimal cutting temperature compound (OCT, Electron Microscopy Sciences). For plantar skin, 50 µm cryosections were cut onto Superfrost Plus slides (Thermo Fisher Scientific, USA) and imaged by confocal microscopy using 1 µm Z-stacks. For DRG, 14 µm cryosections were cut onto Superfrost Plus slides, stained with rabbit anti-CGRP (EMD Millipore, PC205L, 1:500), rabbit anti-tyrosine hydroxylase (Millipore, AB152, 1:1000), rabbit anti-NeuN (Millipore, A60, 1:1000), rabbit anti-Parvalbumin (Swant, Switzerland; PV25, 1:1000), followed by Alexa 488 or Alexa 647 goat anti-rabbit IgG (Life Technologies; 1:1000) or chicken anti-neurofilament 200 (Millipore, AB5539, 1:500), followed by Alexa 488 or Alexa 647 anti-chicken IgG (Life Technologies, 1:1000). For spinal cord, 20 µm cryosections were cut onto Superfrost Plus slides, stained with rabbit anti-CGRP or anti-PKCγ (1:1000), followed by Alexa 488 or Alexa 647 goat anti-rabbit IgG (Life Technologies, 1:1000). Isolectin B4-FITC (Vector Labs; 1:1000) or Isolectin B4-Alexa 647 (Life Technologies, 1:1000) were also used for staining. Sections were mounted in Prolong Gold antifade reagent (Life Technologies) prior to imaging using an Eclipse 50i epifluorescence microscope (Nikon, NY, USA). Fluorescent DRG images were thresholded and analyzed for cell size by NIH ImageJ software. For quantification, at least eight distinct 10X fields of lumbar DRG staining from n=3 animals were analyzed for co-localization and neuronal proportions. Statistical analysis and graphs were generated using Prism software (Graphpad).

For whole mount imaging, lumbar dorsal root ganglia, trigeminal ganglia, sciatic nerve, plantar skin, abdominal walls were dissected and mounted in PBS under glass coverslips. Confocal microscopy was conducted using a LSM700 laser-scanning confocal microscope (Carl Zeiss, Germany), using a 10X Zeiss EC plan-NEOFLUAR dry and a 40X Zeiss plan-APOCHROMAT oil objectives, with Z-stacks imaged at 1 µm steps, collected of up to 200 µm total. Three dimensional reconstructions were rendered as maximum projection images using Volocity software (Perkin Elmer).

RNA *in situ* hybridization

For *in situ* hybridization (ISH), mice were euthanized with CO₂. Lumbar L4-L6 DRGs were dissected and immediately frozen in OCT on dry ice. Tissue was cryosectioned (10-12 µm), mounted onto Superfrost Plus slides (VWR, USA), frozen at -80°C. Digoxigenin- and fluorescein-labeled anti-sense cRNA probes matching coding (Gprc5b, Lpar3, TdTomato, Ntrk2 (Trkb), Prkcg, Nppb, Il31ra) or untranslated regions were synthesized, hybridized to sections, and visualized as previously described (Liberles and Buck, 2006), with minor modifications in amplification strategy. Following overnight hybridization, slides were incubated with peroxidase conjugated anti-digoxigenin antibody (Roche Applied Sciences, IN, USA; 1:200) and alkaline phosphatase conjugated anti-fluorescein antibody (Roche Applied Sciences, 1:200) for 1 hour at room temperature. Tissues were washed and incubated in TSA-PLUS-Cy5 (Perkin Elmer) followed by HNPP (Roche Applied Sciences) according to manufacturer's instructions. Epifluorescence images were captured with a Leica TCS SP5 II microscope. Sequences of primers used for probe generation are listed in Table 3.

Neuronal Cultures and Electrophysiology.

For electrophysiological analysis of Parv-Cre/TdTomato and SNS-Cre/TdTomato neurons, DRGs were dissected, placed in HBSS, incubated for 90 min with 5 mg/ml collagenase, 1 mg/ml dispase II at 37°C. Cells were triturated in the presence of DNase I inhibitor, centrifuged through 10% BSA, resuspended in 1 ml of neurobasal medium, 10 µM Ara-C (Sigma), 50 ng/ml NGF, 2 ng/ml GDNF (Life Technologies), and plated onto 35-mm tissue culture dishes coated with 5 mg/ml laminin. Cultures were incubated at 37°C under 5% CO₂. Recordings were made at room temperature within 24 h of plating. Whole-cell recordings were made with an Axopatch 200A amplifier (Molecular Devices) and patch pipettes with resistances of 2-3 MΩ. The pipette capacitance was decreased by wrapping the shank with parafilm and compensated using the amplifier circuitry. Pipette solution was 5 mM NaCl, 140 mM KCl, 0.5 mM CaCl₂, 2 mM MgCl₂, 5 mM EGTA, 10 mM HEPES, and 3 mM Na₂ATP, pH 7.2, adjusted with NaOH. The external solution was 140 mM NaCl, 5 mM KCl, 2 mM CaCl₂, 2 mM MgCl₂, 10 mM HEPES, and 10 mM D-glucose, pH 7.4, adjusted with NaOH (Sigma). Current clamp recordings were made with the fast current-clamp mode. Command protocols were generated and data digitized with a Digidata 1440A A/D interface with pCLAMP10 software. Action potentials (AP) were evoked by 5 ms depolarizing current pulses. AP half width was measured at half-maximal amplitude. 500 nM Tetrodotoxin (TTX) were applied to block TTX-sensitive Na⁺ currents.

Flow Cytometry of Neurons.

DRGs from cervical (C1-C8), thoracic (T1-T13), and lumbar (L1-L6) different fluorescent strains were pooled from consisting of 7-20 week age-matched male and female adult mice (see Table 1). DRGs were dissected, digested in 1 mg/mL Collagenase A/2.4 U/mL Dispase II (enzymes from Roche), dissolved in HEPES buffered saline (Sigma) for 70 minutes at 37°C. Following digestion, cells were washed into HBSS containing 0.5% Bovine serum albumin (BSA, Sigma), filtered through a 70 µm strainer, resuspended in HBSS/0.5% BSA, and subjected to flow cytometry. Cells were run through a 100 µm nozzle at low pressure (20 p.s.i.) on a BD FACS Aria II machine (Becton Dickinson, USA). A neural density filter (2.0 setting) was used to allow visualization of large cells. Note: Initial trials using traditional gating strategies (e.g. cell size, doublet discrimination, and scatter properties²) did not eliminate non-neuronal cells. An important aspect of isolating pure neurons was based on the significantly higher fluorescence of the Rosa26-TdTomato reporter in somata compared to axonal debris, allowing accurate gating

for cell bodies and purer neuronal signatures. For microarrays, fluorescent neuronal subsets were FACS purified directly into Qiazol (Qiagen). To minimize technical variability, SNS-Cre/TdTomato (total, IB4⁺, IB4⁻) and Parv-Cre/TdTomato neurons were sorted on the same days. FACS data was analyzed using FlowJo software (TreeStar, OR, USA). For Fluidigm analysis, single cells or multiple cell groups from different neuronal populations were FACS sorted into individual wells of a 96-well PCR plate containing pre RNA-amplification mixtures. For microscopy, fluorescent neurons or axons were FACS purified into Neurobasal+B27 supplement+50ng/mL NGF, plated in poly-d-lysine/laminin-coated 8-well chamber slides (Life Technologies) and imaged immediately or 24 hr. later by Eclipse 50i microscope (Nikon). Flow cytometry was performed in the IDDRC Stem Cell Core Facility at Boston Children's Hospital.

Single Neuron Analysis.

Flow cytometry was used to purify 100 cell groups, 10 cell groups, or single cells into 96-well plates containing 9 μ l of a pre-amplification containing reaction mix from the CellsDirect One-Step qRT-PCR Kit (Life Technologies) mixture with pooled Taqman assays (purchased as optimized designs from Life Technologies). Superscript III RT Taq mix (Life Technologies) was used for 14 cycles to pre-amplify specific transcripts. We found that not every FACS sorted-well contained a cell; thus, a pre-screening method was utilized, where 2 μ l from each well was subjected to two-step quantitative PCR (qPCR) for Actb (β -Actin) using fast SYBR green master mix (Life Technologies) on an Applied Biosystems 7500 machine using the following primers: 5'-acactgtgccatctacgag-3' and 5'-gctgtggtggtgaagctga-3'. Wells showing Actb Ct values <20 were picked for subsequent analysis. Using the Fluidigm microfluidic multiplex qRT-PCR platform, pre-amplified well products were run on 96.96 dynamic arrays (Biomark) and assayed against 81 Taqman assays (Life Technologies). Specific assays were chosen based on differential expression by microarray analysis, functional category, and housekeeping genes (Table 2). Ct values were measured by Biomark software, relative transcript levels determined by 2^{- Δ Ct} normalization to Gapdh or Actb transcript levels. For each transcript, outliers of 5 standard deviations from the mean were excluded (set to 0) from our analysis. A total of 334 single cells were analyzed, consisting of IB4⁺SNS-Cre/TdT⁺ (n=132), IB4⁻SNS-Cre/TdT⁺ (n=110), Parv-Cre/TdT⁺ (n=92) neurons. Spearman rank average-linkage clustering was performed with the HierarchicalClustering module from the GenePattern genomic analysis platform and visualized using the HierarchicalClusteringViewer module of GenePattern (MIT Broad Institute). A specific level of hierarchical clustering was used to ascertain clustered neuron subgroups. The Population PCA tool was used for principal components analysis - <http://cbdm.hms.harvard.edu/LabMembersPges/SD.html>. Pearson correlation analysis of specific transcripts to all 80 probes across the single cell expression dataset was generated using nearest neighbor analysis by the GenePattern platform. Histogram plots of single cell data were generated in Excel (Microsoft, USA). Dot plots showing single cell transcript data across subgroups was generated in Prism software (Graphpad, USA).

Statistical analysis.

Sample sizes for experiments were chosen according to standard practice in the field. “n” represents the number of mice, samples, or cells used in each group. Bar and line graphs are plotted as mean \pm standard error of the mean (s.e.m.). Data meet the assumptions of specific statistical tests chosen, including normality for parametric or non-parametric tests. Statistical analysis of electrophysiology, neuronal cell counts, and flow cytometry were by One-way

ANOVA with Tukey's post-test or by unpaired, student's t-test. Data was plotted using Prism software (Graphpad, USA).

RNA Processing, Microarray Hybridization and Bioinformatics Analysis

RNA was extracted by sequential Qiazol extraction and purification through the RNeasy micro kit with on column genomic DNA digestion according to manufacturer's instructions (Qiagen, CA, USA). RNA quality was determined by Agilent 2100 Bioanalyzer using the Pico Chip (Agilent, CA, USA). Samples with RIN>7 were used for analysis. RNA was amplified into cDNA using the Ambion WT expression kit for Whole Transcript Expression Arrays, with Poly-A controls from the Affymetrix Genechip Eukaryotic Poly-A RNA control kit (Affymetrix, CA, USA). The Affymetrix Genechip WT Terminal labeling kit was used for fragmentation and biotin labeling. Affymetrix GeneChip Hybridization control kit and the Affymetrix GeneChip Hybridization, wash, stain kit was used to hybridize samples to Affymetrix Mouse Gene ST 1.0 GeneChips, fluidics performed on the Affymetrix Genechip Fluidics Station 450, and scanned using Affymetrix Genechip Scanner 7G (Affymetrix). Microarray work was conducted at the Boston Children's Hospital IDDR Molecular Genetics Core. For Bioinformatics analysis, Affymetrix CEL files were normalized using the Robust Multi-array Average (RMA) algorithm with quantile normalization, background correction, and median scaling. Hierarchical clustering and principal-component analysis (PCA) was conducted on datasets filtered for mean expression values greater than 100 in any population (Mingueneau et al., 2013), with elimination of noisy transcripts with an intra-population coefficient of variation (CoV) < 0.65. Spearman-rank average linkage analysis was conducted on the top 15% most variable probes across subsets (2735 transcripts) using the HierarchicalClustering module, and heat-maps generated using the HierarchicalClusteringViewer module of the GenePattern analysis platform (Broad Institute). The Population PCA tool was used (<http://cbdm.hms.harvard.edu/LabMembersPges/SD.html>). For pathway enrichment analysis, pairwise comparisons of specific neuronal datasets (e.g. Parv-Cre/TdTomato vs. SNS-Cre/TdTomato) were conducted. Differentially expressed transcripts (2-fold, $p < 0.05$) were analyzed using Database for Annotation, Visualization and Integrated Discovery (DAVID) (<http://david.abcc.ncifcrf.gov>). Pathway enrichment p-values for GO Terms (Biological Processes) or Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were plotted as heat-maps using the HeatmapViewer module of GenePattern. Differentially expressed transcripts were illustrated using volcano plots, generated by plotting fold-change differences against comparison p-values or $-\log(p\text{-values})$. Transcripts showing low intragroup variability ($\text{CoV} < 0.65$) were included in this differential expression analysis. Specific gene families, including ion channels (calcium, sodium, potassium, chloride, ligand-gated, TRP and HCN channels), GPCRs and transcription factors were highlighted on volcano plots.

Data Deposition

All microarray datasets are deposited at the NCBI GEO database (<http://www.ncbi.nlm.nih.gov/>) under accession number GSE55114. Data in Supplementary Files 1 and 2 are deposited at Dryad (<http://www.datadryad.org>).

Table 1. Transcriptional samples analyzed in this study.

Sample name	Sample Description	Type	n
SNS-Cre/TdT ⁺	SNS-Cre/TdTomato ⁺ FACS purified neurons	Neuron population	4
Parv-Cre/TdT ⁺	Parv-Cre/TdTomato ⁺ FACS purified neurons	Neuron population	4
IB4 ⁺ SNS-Cre/TdT ⁺	IB4 ⁺ SNS-Cre/TdT ⁺ FACS purified neurons	Neuron population	3
IB4 ⁺ SNS-Cre/TdT ⁺	IB4 ⁺ SNS-Cre/TdT ⁺ FACS purified neurons	Neuron population	3
Whole DRG	Homogenized DRG tissue	Whole tissue	3
IB4 ⁺ SNS-Cre/TdT ⁺ (individual neurons)	IB4 ⁺ SNS-Cre/TdT ⁺ FACS purified single cells	Single cells	132
IB4 ⁺ SNS-Cre/TdT ⁺ (individual neurons)	IB4 ⁺ SNS-Cre/TdT ⁺ FACS purified single cells	Single cells	110
Parv-Cre/TdT ⁺ (individual neurons)	Parv-Cre/TdT ⁺ FACS purified single cells	Single cells	92

Table 1. Transcriptional samples analyzed in this study.

In this study, we performed microarray profiling of FACS purified neuron populations, DRG tissue, and single neuron samples. This table summarizes the sample names, descriptions, types, and numbers of samples analyzed. For neuron populations and whole DRG tissue, each biological replicate consisted of pooled total DRG cells from n=3 animals.

Table 2. Taqman assays used for single cell transcriptional profiling

SNS-Cre/TdT ⁺ enriched (vs. Parv- Cre/TdT)	IB4 ⁺ SNS-Cre/TdT ⁺ enriched	IB4 ⁻ SNS-Cre/TdT ⁺ enriched	Parv-Cre/TdT ⁺ enriched
Trpv1	Mrgprd	Smr2	Pvalb
Trpa1	P2rx3	Npy2r	Runx3
Scn10a	Agtr1a	Nppb	Calb2
Scn11a	Prkeq	Kcnv1	Slit2
Isl2	Wnt2b	Prokr2	Spp1
Kcnc2	Slc16a12	Ptgir	Ano1
Galr1	Lpar3	Th	Stxbp6
Car8	Lpar5	Il31ra	St8sia5
Chrna3	Trpc3	Ntrk1	Ndst4
Atp2b4	Trpc6	Bves	Esrrb
Aqp1	Moxd1	Kcnq4	Esrrg
Chrna6	A3galt2	Htr3a	Gprc5b
Pde11a	St6gal2	S100a16	Car2
MrgprC11	Mrgprb4	Pou4f3	Pth1r
Syt5	Mrgprb5	Cgnl1	Wnt7b
Gfra3	Ptgdr		Kcnc1
Klf7	Ggta1		Etv1
Cysltr2	Grik1		Pln
Irf6	Mmp25		Cdh12
Prdm8	Casz1		
Etv5	Bnc2		
Stac	Klf5		
	Lypd1		
Housekeeping genes			
Gapdh	Actb		

Table 2. Taqman assays used for single cell transcriptional profiling.

To perform Fluidigm single cell analysis, Taqman assays were chosen to cover four categories of population-enriched transcripts first identified by microarray whole transcriptome analysis: 1) SNS-Cre/TdT⁺ (total population) enriched markers (vs. Parv-Cre/TdT⁺ neurons), 2) IB4⁺SNS-Cre/TdT⁺ enriched markers (vs. other 2 groups), 3) IB4⁻SNS-Cre/TdT⁺ markers (vs. other 2 groups), and 4) Parv-Cre/TdT⁺ markers (vs. other 2 groups). Taqman assays for housekeeping genes Gapdh and Actb were also included.

Table 3. RNA *in situ* hybridization probes.

Gene	Forward primer	Reverse Primer	Probe Length (bp)
Gpcr5b	5'-ATGTTCTGGT	5'-TCACCAATGGTG	1233
Lpar3	5'-TTGTGATCGTCCTGTGCGTG	5'-GCCTCTCGGTATTGCTGTCC	870
TdTomato	5'-ATCAAAGAGTTCATGCGCTTC	5'-GTTCCACGATGGTGTAGTCCTC	615
Prkcq	5'-TCTTGCTGGGTCAGAAGTACAA	5'-TCTGTGGTTGAGTGGAATTGAC	919
Nppb	5'-TGAAGGTGCTGTCCCAGATGATTC	5'-GTTGTGGCAAGTTTGTGCTCCAAG	545
Il31ra	5'-CTCCCCTGTGTTGTCCTGAT	5'-TTCATGCCATAGCAGCACTC	559

Table 3. RNA *in situ* hybridization probes.

Probesets used for RNA *in situ* hybridization analysis. Listed are gene symbols, sequences for forward and reverse primers, and resulting probe lengths.

Acknowledgements

We thank Victor Wang, Olesegun Babanyi, Ta-wei Lin, Catherine Ward, Richard Bennett, Kristen Cabal, and Noreen Francis for technical support; Sriya Muraldiharan and Amanda Strominger for immunostaining and quantification; Mark Hoon for Nppb probe information; Christian Von Hehn for helpful discussions on neuronal purification; Bruce P. Bean, Vijay Kuchroo for helpful advice. This work was supported by CJW NIH R37 NS039518; R01 NS038253; 1PO1 NS072040-01 AND Miriam and Sheldon Adelson Medical Foundation. Gene expression analysis were performed in the IDDRC Molecular Genetics Core facility at Boston Children's Hospital, supported by National Institutes of Health award NIH-P50-NS40828. Flow cytometry was performed in the IDDRC Stem Cell Core Facility at Boston Children's Hospital, supported by NIH-P30-HD18655. Microarray work was conducted at the Boston Children's Hospital IDDRC Molecular Genetics Core, supported by NIH-P30-HD 18655.

References

- Abdel Samad, O., Liu, Y., Yang, F. C., Kramer, I., Arber, S. & Ma, Q. 2010. Characterization of two Runx1-dependent nociceptor differentiation programs necessary for inflammatory versus neuropathic pain. *Mol Pain*, **6**: 45. doi: 10.1186/1744-8069-6-45.
- Abrahamsen, B., Zhao, J., Asante, C. O., Cendan, C. M., Marsh, S., Martinez-Barbera, J. P., Nassar, M. A., Dickenson, A. H. & Wood, J. N. 2008. The cell and molecular basis of mechanical, cold, and inflammatory pain. *Science*, **321**: 702-5. doi: 10.1126/science.1156916.
- Abraira, V. E. & Ginty, D. D. 2013. The sensory neurons of touch. *Neuron*, **79**: 618-39. doi: 10.1016/j.neuron.2013.07.051.
- Agarwal, N., Offermanns, S. & Kuner, R. 2004. Conditional gene deletion in primary nociceptive neurons of trigeminal ganglia and dorsal root ganglia. *Genesis*, **38**: 122-9. doi: 10.1002/gene.20010 [doi].
- Arlotta, P., Molyneaux, B. J., Chen, J., Inoue, J., Kominami, R. & Macklis, J. D. 2005. Neuronal subtype-specific genes that control corticospinal motor neuron development in vivo. *Neuron*, **45**: 207-21. doi: 10.1016/j.neuron.2004.12.036.
- Bando, T., Morikawa, Y., Komori, T. & Senba, E. 2006. Complete overlap of interleukin-31 receptor A and oncostatin M receptor beta in the adult dorsal root ganglia with distinct developmental expression patterns. *Neuroscience*, **142**: 1263-71. doi: 10.1016/j.neuroscience.2006.07.009.
- Basbaum, A. I., Bautista, D. M., Scherrer, G. & Julius, D. 2009. Cellular and molecular mechanisms of pain. *Cell*, **139**: 267-84. doi: 10.1016/j.cell.2009.09.028.
- Bean, B. P. 2007. The action potential in mammalian central neurons. *Nat Rev Neurosci*, **8**: 451-65. doi: 10.1038/nrn2148.
- Bourane, S., Mechaly, I., Venteo, S., Garces, A., Fichard, A., Valmier, J. & Carroll, P. 2007. A SAGE-based screen for genes expressed in sub-populations of neurons in the mouse dorsal root ganglion. *BMC Neurosci*, **8**: 97. doi: 10.1186/1471-2202-8-97.
- Chen, C. L., Broom, D. C., Liu, Y., De Nooij, J. C., Li, Z., Cen, C., Samad, O. A., Jessell, T. M., Woolf, C. J. & Ma, Q. 2006. Runx1 determines nociceptive sensory neuron phenotype and is required for thermal and neuropathic pain. *Neuron*, **49**: 365-77. doi: 10.1016/j.neuron.2005.10.036.
- Cho, H., Yang, Y. D., Lee, J., Lee, B., Kim, T., Jang, Y., Back, S. K., Na, H. S., Harfe, B. D., Wang, F., Raouf, R., Wood, J. N. & Oh, U. 2012. The calcium-activated chloride channel anoctamin 1 acts as a heat sensor in nociceptive neurons. *Nat Neurosci*, **15**: 1015-21. doi: 10.1038/nn.3111.
- Citri, A., Pang, Z. P., Sudhof, T. C., Wernig, M. & Malenka, R. C. 2012. Comprehensive qPCR profiling of gene expression in single neuronal cells. *Nat Protoc*, **7**: 118-27. doi: 10.1038/nprot.2011.430.
- Coste, B., Mathur, J., Schmidt, M., Earley, T. J., Ranade, S., Petrus, M. J., Dubin, A. E. & Patapoutian, A. 2010. Piezo1 and Piezo2 are essential components of distinct mechanically activated cation channels. *Science*, **330**: 55-60. doi: 10.1126/science.1193270.
- Costigan, M., Belfer, I., Griffin, R. S., Dai, F., Barrett, L. B., Coppola, G., Wu, T., Kiselycznyk, C., Poddar, M., Lu, Y., Diatchenko, L., Smith, S., Cobos, E. J., Zaykin, D., Allchorne, A.,

- Gershon, E., Livneh, J., Shen, P. H., Nikolajsen, L., Karppinen, J., Mannikko, M., Kelempisioti, A., Goldman, D., Maixner, W., Geschwind, D. H., Max, M. B., Seltzer, Z. & Woolf, C. J. 2010. Multiple chronic pain states are associated with a common amino acid-changing allele in KCNS1. *Brain*, **133**: 2519-27. doi: 10.1093/brain/awq195.
- De Nooij, 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**: 1055-68. doi: 10.1016/j.neuron.2013.01.015.
- Dib-Hajj, S. D., Cummins, T. R., Black, J. A. & Waxman, S. G. 2010. Sodium channels in normal and pathological pain. *Annu Rev Neurosci*, **33**: 325-47. doi: 10.1146/annurev-neuro-060909-153234.
- Dubin, A. E. & Patapoutian, A. 2010. Nociceptors: the sensors of the pain pathway. *J Clin Invest*, **120**: 3760-72. doi: 10.1172/jci42843.
- Field, M. J., Cox, P. J., Stott, E., Melrose, H., Offord, J., Su, T. Z., Bramwell, S., Corradini, L., England, S., Winks, J., Kinloch, R. A., Hendrich, J., Dolphin, A. C., Webb, T. & Williams, D. 2006. Identification of the alpha2-delta-1 subunit of voltage-dependent calcium channels as a molecular target for pain mediating the analgesic actions of pregabalin. *Proc Natl Acad Sci U S A*, **103**: 17537-42. doi: 10.1073/pnas.0409066103.
- Garrison, S. R., Dietrich, A. & Stucky, C. L. 2012. TRPC1 contributes to light-touch sensation and mechanical responses in low-threshold cutaneous sensory neurons. *J Neurophysiol*, **107**: 913-22. doi: 10.1152/jn.00658.2011.
- Goswami, S. C., Mishra, S. K., Maric, D., Kaszas, K., Gonnella, G. L., Clokie, S. J., Kominsky, H. D., Gross, J. R., Keller, J. M., Mannes, A. J., Hoon, M. A. & Iadarola, M. J. 2014. Molecular Signatures of Mouse TRPV1-Lineage Neurons Revealed by RNA-Seq Transcriptome Analysis. *J Pain*, **15**: 1338-59. doi: 10.1016/j.jpain.2014.09.010.
- 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 Biol*, **3**: e159. doi: 10.1371/journal.pbio.0030159.
- Julius, D. 2013. TRP channels and pain. *Annu Rev Cell Dev Biol*, **29**: 355-84. doi: 10.1146/annurev-cellbio-101011-155833.
- Lee, J., Friese, A., Mielich, M., Sigrist, M. & Arber, S. 2012. Scaling proprioceptor gene transcription by retrograde NT3 signaling. *PLoS One*, **7**: e45551. doi: 10.1371/journal.pone.0045551.
- 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. doi: 10.1016/j.cell.2011.11.027.
- Liberles, S. D. & Buck, L. B. 2006. A second class of chemosensory receptors in the olfactory epithelium. *Nature*, **442**: 645-50. doi: 10.1038/nature05066.
- Liu, Y., Abdel Samad, O., Zhang, L., Duan, B., Tong, Q., Lopes, C., Ji, R. R., Lowell, B. B. & Ma, Q. 2010. VGLUT2-dependent glutamate release from nociceptors is required to sense pain and suppress itch. *Neuron*, **68**: 543-56. doi: 10.1016/j.neuron.2010.09.008.
- Lobo, M. K., Karsten, S. L., Gray, M., Geschwind, D. H. & Yang, X. W. 2006. FACS-array profiling of striatal projection neuron subtypes in juvenile and adult mouse brains. *Nat Neurosci*, **9**: 443-52. doi: 10.1038/nn1654.

- Lopes, C., Liu, Z., Xu, Y. & Ma, Q. 2012. Tlx3 and Runx1 act in combination to coordinate the development of a cohort of nociceptors, thermoceptors, and pruriceptors. *J Neurosci*, **32**: 9706-15. doi: 10.1523/jneurosci.1109-12.2012.
- Lou, S., Duan, B., Vong, L., Lowell, B. B. & Ma, Q. 2013. Runx1 controls terminal morphology and mechanosensitivity of VGLUT3-expressing C-mechanoreceptors. *J Neurosci*, **33**: 870-82. doi: 10.1523/jneurosci.3942-12.2013.
- 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 Neurosci*, **13**: 133-40. doi: 10.1038/nn.2467.
- 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. doi: 10.1038/nature13250.
- Marmigere, F. & Ernfors, P. 2007. Specification and connectivity of neuronal subtypes in the sensory lineage. *Nat Rev Neurosci*, **8**: 114-27. doi: 10.1038/nrn2057.
- Mingueneau, M., Kreslavsky, T., Gray, D., Heng, T., Cruse, R., Ericson, J., Bendall, S., Spitzer, M. H., Nolan, G. P., Kobayashi, K., Von Boehmer, H., Mathis, D., Benoist, C., Best, A. J., Knell, J., Goldrath, A., Jojic, V., Koller, D., Shay, T., Regev, A., Cohen, N., Brennan, P., Brenner, M., Kim, F., Rao, T. N., Wagers, A., Heng, T., Ericson, J., Rothamel, K., Ortiz-Lopez, A., Mathis, D., Benoist, C., Bezman, N. A., Sun, J. C., Min-Oo, G., Kim, C. C., Lanier, L. L., Miller, J., Brown, B., Merad, M., Gautier, E. L., Jakubzick, C., Randolph, G. J., Monach, P., Blair, D. A., Dustin, M. L., Shinton, S. A., Hardy, R. R., Laidlaw, D., Collins, J., Gazit, R., Rossi, D. J., Malhotra, N., Sylvia, K., Kang, J., Kreslavsky, T., Fletcher, A., Elpek, K., Bellemare-Pelletier, A., Malhotra, D. & Turley, S. 2013. The transcriptional landscape of alphabeta T cell differentiation. *Nat Immunol*, **14**: 619-32. doi: 10.1038/ni.2590.
- Mishra, S. K. & Hoon, M. A. 2013. The cells and circuitry for itch responses in mice. *Science*, **340**: 968-71. doi: 10.1126/science.1233765.
- Mizeracka, K., Trimarchi, J. M., Stadler, M. B. & Cepko, C. L. 2013. Analysis of gene expression in wild-type and Notch1 mutant retinal cells by single cell profiling. *Dev Dyn*, **242**: 1147-59. doi: 10.1002/dvdy.24006.
- Molyneaux, B. J., Arlotta, P., Fame, R. M., Macdonald, J. L., Macquarrie, K. L. & Macklis, J. D. 2009. Novel subtype-specific genes identify distinct subpopulations of callosal projection neurons. *J Neurosci*, **29**: 12343-54. doi: 10.1523/JNEUROSCI.6108-08.2009.
- Neely, G. G., Hess, A., Costigan, M., Keene, A. C., Goulas, S., Langeslag, M., Griffin, R. S., Belfer, I., Dai, F., Smith, S. B., Diatchenko, L., Gupta, V., Xia, C. P., Amann, S., Kreitz, S., Heindl-Erdmann, C., Wolz, S., Ly, C. V., Arora, S., Sarangi, R., Dan, D., Novatchkova, M., Rosenzweig, M., Gibson, D. G., Truong, D., Schramek, D., Zoranovic, T., Cronin, S. J., Angjeli, B., Brune, K., Dietzl, G., Maixner, W., Meixner, A., Thomas, W., Pospisilik, J. A., Alenius, M., Kress, M., Subramaniam, S., Garrity, P. A., Bellen, H. J., Woolf, C. J. & Penninger, J. M. 2010. A genome-wide Drosophila screen for heat nociception identifies alpha2delta3 as an evolutionarily conserved pain gene. *Cell*, **143**: 628-38. doi: 10.1016/j.cell.2010.09.047.

- Niu, J., Ding, L., Li, J. J., Kim, H., Liu, J., Li, H., Moberly, A., Badea, T. C., Duncan, I. D., Son, Y. J., Scherer, S. S. & Luo, W. 2013. Modality-based organization of ascending somatosensory axons in the direct dorsal column pathway. *J Neurosci*, **33**: 17691-709. doi: 10.1523/JNEUROSCI.3429-13.2013.
- Okaty, B. W., Miller, M. N., Sugino, K., Hempel, C. M. & Nelson, S. B. 2009. Transcriptional and electrophysiological maturation of neocortical fast-spiking GABAergic interneurons. *J Neurosci*, **29**: 7040-52. doi: 10.1523/JNEUROSCI.0105-09.2009.
- Okaty, B. W., Sugino, K. & Nelson, S. B. 2011. Cell type-specific transcriptomics in the brain. *J Neurosci*, **31**: 6939-43. doi: 10.1523/JNEUROSCI.0626-11.2011.
- Patel, A. P., Tirosh, I., Trombetta, J. J., Shalek, A. K., Gillespie, S. M., Wakimoto, H., Cahill, D. P., Nahed, B. V., Curry, W. T., Martuza, R. L., Louis, D. N., Rozenblatt-Rosen, O., Suva, M. L., Regev, A. & Bernstein, B. E. 2014. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science*, **344**: 1396-401. doi: 10.1126/science.1254257.
- Patel, R., Bauer, C. S., Nieto-Rostro, M., Margas, W., Ferron, L., Chaggar, K., Crews, K., Ramirez, J. D., Bennett, D. L., Schwartz, A., Dickenson, A. H. & Dolphin, A. C. 2013. alpha2delta-1 gene deletion affects somatosensory neuron function and delays mechanical hypersensitivity in response to peripheral nerve damage. *J Neurosci*, **33**: 16412-26. doi: 10.1523/jneurosci.1026-13.2013.
- Seal, R. P., Wang, X., Guan, Y., Raja, S. N., Woodbury, C. J., Basbaum, A. I. & Edwards, R. H. 2009. Injury-induced mechanical hypersensitivity requires C-low threshold mechanoreceptors. *Nature*, **462**: 651-5. doi: 10.1038/nature08505.
- Shalek, A. K., Satija, R., Shuga, J., Trombetta, J. J., Gennert, D., Lu, D., Chen, P., Gertner, R. S., Gaublot, J. T., Yosef, N., Schwartz, S., Fowler, B., Weaver, S., Wang, J., Wang, X., Ding, R., Raychowdhury, R., Friedman, N., Hacohen, N., Park, H., May, A. P. & Regev, A. 2014. Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature*, **510**: 363-9. doi: 10.1038/nature13437.
- Shields, S. D., Ahn, H. S., Yang, Y., Han, C., Seal, R. P., Wood, J. N., Waxman, S. G. & Dib-Hajj, S. D. 2012. Nav1.8 expression is not restricted to nociceptors in mouse peripheral nervous system. *Pain*, **153**: 2017-30. doi: 10.1016/j.pain.2012.04.022.
- Sonkoly, E., Muller, A., Lauerma, A. I., Pivarcsi, A., Soto, H., Kemeny, L., Alenius, H., Dieu-Nosjean, M. C., Meller, S., Rieker, J., Steinhoff, M., Hoffmann, T. K., Ruzicka, T., Zlotnik, A. & Homey, B. 2006. IL-31: a new link between T cells and pruritus in atopic skin inflammation. *J Allergy Clin Immunol*, **117**: 411-7. doi: 10.1016/j.jaci.2005.10.033.
- Stirling, L. C., Forlani, G., Baker, M. D., Wood, J. N., Matthews, E. A., Dickenson, A. H. & Nassar, M. A. 2005. Nociceptor-specific gene deletion using heterozygous Nav1.8-Cre recombinase mice. *Pain*, **113**: 27-36. doi: 10.1016/j.pain.2004.08.015.
- Stucky, C. L., Rossi, J., Airaksinen, M. S. & Lewin, G. R. 2002. GFR alpha2/neurturin signalling regulates noxious heat transduction in isolectin B4-binding mouse sensory neurons. *J Physiol*, **545**: 43-50. doi: 10.1016/j.pain.2004.08.015.
- Sugino, K., Hempel, C. M., Miller, M. N., Hattox, A. M., Shapiro, P., Wu, C., Huang, Z. J. & Nelson, S. B. 2006. Molecular taxonomy of major neuronal classes in the adult mouse forebrain. *Nat Neurosci*, **9**: 99-107. doi: 10.1038/nn1618.
- Sun, Y. G. & Chen, Z. F. 2007. A gastrin-releasing peptide receptor mediates the itch sensation in the spinal cord. *Nature*, **448**: 700-3. doi: 10.1038/nature06029.

- Thakur, M., Crow, M., Richards, N., Davey, G. I., Levine, E., Kelleher, J. H., Agley, C. C., Denk, F., Harridge, S. D. & McMahon, S. B. 2014. Defining the nociceptor transcriptome. *Front Mol Neurosci*, **7**: 87. doi: 10.3389/fnmol.2014.00087.
- Tsantoulas, C., Zhu, L., Shaifta, Y., Grist, J., Ward, J. P., Raouf, R., Michael, G. J. & McMahon, S. B. 2012. Sensory neuron downregulation of the Kv9.1 potassium channel subunit mediates neuropathic pain following nerve injury. *J Neurosci*, **32**: 17502-13. doi: 10.1523/JNEUROSCI.3561-12.2012.
- Usoskin, D., Furlan, A., Islam, S., Abdo, H., Lonnerberg, P., Lou, D., Hjerling-Leffler, J., Haeggstrom, J., Kharchenko, O., Kharchenko, P. V., Linnarsson, S. & Ernfors, P. 2014. Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. *Nat Neurosci*, doi: 10.1038/nn.3881.
- Vrontou, S., Wong, A. M., Rau, K. K., Koerber, H. R. & Anderson, D. J. 2013. Genetic identification of C fibres that detect massage-like stroking of hairy skin in vivo. *Nature*, **493**: 669-73. doi: 10.1038/nature11810.
- Vulchanova, L., Riedl, M. S., Shuster, S. J., Stone, L. S., Hargreaves, K. M., Buell, G., Surprenant, A., North, R. A. & Elde, R. 1998. P2X3 is expressed by DRG neurons that terminate in inner lamina II. *Eur J Neurosci*, **10**: 3470-8. doi: 10.1523/jneurosci.3248-09.2009.
- Wainger, B. J., Buttermore, E. D., Oliveira, J. T., Mellin, C., Lee, S., Saber, W. A., Wang, A. J., Ichida, J. K., Chiu, I. M., Barrett, L., Huebner, E. A., Bilgin, C., Tsujimoto, N., Brenneis, C., Kapur, K., Rubin, L. L., Eggan, K. & Woolf, C. J. 2014. Modeling pain in vitro using nociceptor neurons reprogrammed from fibroblasts. *Nat Neurosci*, doi: 10.1038/nn.3886.
- Wang, H. & Zylka, M. J. 2009. Mrgprd-expressing polymodal nociceptive neurons innervate most known classes of substantia gelatinosa neurons. *J Neurosci*, **29**: 13202-9. doi: 10.1523/jneurosci.3248-09.2009.
- Wang, M., Offord, J., Oxender, D. L. & Su, T. Z. 1999. Structural requirement of the calcium-channel subunit alpha2delta for gabapentin binding. *Biochem J*, **342 (Pt 2)**: 313-20. doi: 10.1126/science.1235884.
- Wichterle, H., Gifford, D. & Mazzoni, E. 2013. Neuroscience. Mapping neuronal diversity one cell at a time. *Science*, **341**: 726-7. doi: 10.1126/science.1235884.
- Woo, S. H., Ranade, S., Weyer, A. D., Dubin, A. E., Baba, Y., Qiu, Z., Petrus, M., Miyamoto, T., Reddy, K., Lumpkin, E. A., Stucky, C. L. & Patapoutian, A. 2014. Piezo2 is required for Merkel-cell mechanotransduction. *Nature*, **509**: 622-6. doi: 10.1038/nature13251.
- Wylie, C. J., Hendricks, T. J., Zhang, B., Wang, L., Lu, P., Leahy, P., Fox, S., Maeno, H. & Deneris, E. S. 2010. Distinct transcriptomes define rostral and caudal serotonin neurons. *J Neurosci*, **30**: 670-84. doi: 10.1523/JNEUROSCI.4656-09.2010.
- Zhang, Y., Chen, K., Sloan, S. A., Bennett, M. L., Scholze, A. R., O'keeffe, S., Phatnani, H. P., Guarnieri, P., Caneda, C., Ruderisch, N., Deng, S., Liddelow, S. A., Zhang, C., Daneman, R., Maniatis, T., Barres, B. A. & Wu, J. Q. 2014. An RNA-Sequencing Transcriptome and Splicing Database of Glia, Neurons, and Vascular Cells of the Cerebral Cortex. *J Neurosci*, **34**: 11929-47. doi: 10.1523/JNEUROSCI.1860-14.2014.

Figure Legends

Figure 1. Fluorescent characterization of SNS-Cre/TdTomato and Parv-Cre/TdTomato DRG populations. (a) SNS-Cre/TdTomato and Parv-Cre/TdTomato lumbar DRG sections imaged for TdTomato (red), IB4-FITC, anti-CGRP, or anti-Parvalbumin (green). Scale bars, 50 μ m. (b-c) Proportions of IB4⁺, CGRP⁺, NF200⁺, Parvalbumin⁺ populations expressing SNS-Cre/TdTomato or Parv-Cre/TdTomato, and converse TdTomato proportions expressing each co-stained marker (mean \pm s.e.m., n=8-20 fields from 3 animals) (c) Venn diagram depicting distinct DRG populations as labeled by Isolectin B4, NF200, and TdTomato populations. (b) For transcriptional profiling, three non-overlapping DRG populations were FACS purified: IB4⁺SNS-Cre/TdTomato⁺, IB4⁺SNS-Cre/TdTomato⁺, and Parv-Cre/TdTomato⁺ cells.

Figure 1-figure supplement 1. SNS-Cre/TdTomato and Parv-Cre/TdTomato DRG and spinal cord characterization.

(a) SNS-Cre and Parv-Cre mice were bred with Rosa26-TdTomato mice to generate lineage reporter progeny. (b) Confocal microscopy images of whole mount L4 DRG from TdTomato progeny. Scale bars, 50 μ m. (c) Lumbar spinal cord sections were stained with Isolectin B4-FITC (green) and anti-CGRP (blue). SNS-Cre/TdTomato fibers overlapped with CGRP and IB4 staining in dorsal horn laminae I-II. By contrast, Parv-Cre/TdTomato fibers extended to lamina III, Clark Nucleus (C.N.) and ventral horns. (d) Lumbar sections show SNS-Cre/TdTomato fibers in lamina II (colocalized with IB4), but not lamina III stained by anti-PKC- γ . Parv-Cre/TdTomato does not innervate superficial laminae. Scale bars, 100 μ m.

Figure 2. Electrophysiological properties of SNS-Cre/TdTomato and Parv-Cre/TdTomato neurons.

Whole cell current clamp recordings were conducted on SNS-Cre/TdTomato and Parv-Cre/TdTomato neurons in response to 200 pA injection. (a) Representative action potential waveforms before and after application of 500 nM TTX. (b) Statistical comparisons of action potential (AP) half-widths and capacitances between sensory populations (SNS-Cre/TdT⁺, n=13; Parv-Cre/TdT⁺, n=9; P-values by student's t-test).

Figure 3. FACS purification of distinct somatosensory neuron populations.

(a) Mouse DRG cells were stained with DAPI and subjected to flow cytometry. After gating on large cells by forward and side scatter (R1), dead cells were excluded by gating on the DAPI events; Next, TdTomato (hi) events were purified. Following purification, fluorescence and DIC microscopy show that the majority of sorted neurons are TdTomato⁺ (images on right) (b) Representative FACS plots of Parv-Cre/TdTomato⁺ and SNS-Cre/TdTomato⁺ DRG populations. Right, quantification of proportions of DAPI⁺ events in the DRG constituting each neuron population (n=5 SNS-Cre/TdTomato mice, n=4 Parv-Cre/TdTomato mice; p-values, student's t-test; Error bars, mean \pm s.e.m.) (c) Representative FACS plot shows relative percentages of IB4-FITC surface stained and IB4⁺ neuronal populations among the total SNS-Cre/TdTomato (hi) gate.

Figure 3-figure supplement 1. Flow cytometric sorting and analysis of TdTomato⁺ neurons.

(a) By FACS analysis, TdTomato labeled both "high" and "low" fluorescence populations (see gates). Purified high-expressing populations corresponded to neuronal cell bodies, while the

lower fluorescence consisted of fluorescent axonal debris, as shown by microscopy images post-sorting (right). **(b)** TdTomato neurons purified and plated onto glass slides. After 24 hr, post-sorted SNS-Cre/TdT⁺ neurons showed neurite outgrowth and relatively pure populations compared to unsorted SNS-Cre/TdT⁺ neurons. **(c)** Representative FACS plot overlay of light scattering properties for Parv-Cre/TdT⁺ vs. SNS-Cre/TdT⁺ populations. Comparison of forward and side scatter properties on left (SNS-Cre/TdT, n=4; Parv-Cre/TdT, n=4; error bars, s.e.m).

Figure 3 – figure supplement 2. Transcriptome analysis of purified neuronal samples relative to whole DRG tissues. (a) Individual expression profile comparisons of sorted neuron samples (Red and green show numbers of transcripts >2-fold differential expression). (b) Plots of absolute RMA normalized transcript levels for myelin associated, nociceptor associated, and proprioceptor associated genes in FACS purified SNS-Cre/TdT⁺ and Parv-Cre/TdT⁺ samples vs. whole DRG samples. P-values by One-way ANOVA: ***, p<0.001. (c) Fold-change vs. fold-change comparison of sorted neurons vs. whole DRG datasets (red transcripts are >2-fold enriched in whole DRG; blue transcripts are >2-fold enriched in both sorted subsets).

Figure 4. Hierarchical clustering and Principal components analysis of transcriptomes. (a) Hierarchical clustering of sorted neuron molecular profiles (top 15% probesets by coefficient of variation), showing distinct groups of transcripts enriched in IB4⁺SNS-Cre/TdT⁺, IB4⁻SNS-Cre/TdT⁺, and Parv-Cre/TdT⁺ neuron populations. **(b)** Principal component analysis shows distinct transcriptome segregation for the purified populations along 3 principal components axes.

Figure 5. Functional somatosensory mediators show clustered gene expression across purified DRG populations.

Heat-map showing relative transcript levels for known somatosensory mediators plotted across IB4⁺SNS-Cre/TdTomato⁺, IB4⁻SNS-Cre/TdTomato⁺, and Parv-Cre/TdTomato⁺ purified neuron transcriptomes (rows show individual samples; columns are specific transcripts). Genes were grouped based on known roles linked to thermosensation/nociception, pruriception, tactile function, neurotrophic receptors, and proprioception.

Figure 6. Heat-map distribution of voltage-gated and TRP channels across neuronal subsets.

Expression patterns of different sub-types of voltage-gated ion channels and transient receptor potential (TRP) channels were hierarchically clustered and analyzed across IB4⁺SNS-Cre/TdT⁺, IB4⁻SNS-Cre/TdT⁺ and Parv-Cre/TdT⁺ purified neuron samples (columns are individual samples, heat-maps). **(a)** Sodium channel levels **(b)** Calcium channel levels **(c)** Potassium channel levels (top 60 differentially expressed transcripts by CoV), **(d)** Chloride channel levels, and **(e)** TRP channel levels are plotted as heat-maps. For **a-e**, plotted transcripts show minimum expression>100 in at least one neuronal subgroup.

Figure 7. Heat-map distribution of ligand-gated ion channels, G-protein coupled receptors, and transcription factors across neuronal subsets. (a) Expression patterns of ligand-gated ion channels, including glutamatergic, chlorinergic, HCN, P2X channels, were analyzed by hierarchical clustering (columns are individual samples). **(b)** Differentially expressed G-protein coupled receptors (GPCRs) were clustered and plotted across sensory subsets (Top 60 by CoV are shown). **(c)** Differentially expressed transcription factors were clustered and plotted across

sensory subsets as a heat-map (Top 60 by CoV are shown). For **a-c**, plotted transcripts show minimum expression > 100 in at least one neuronal subgroup.

Figure 8. Differential volcano plot analysis of SNS-Cre/TdTomato vs. Parv-Cre/TdTomato transcriptomes. (a) Pairwise comparison of SNS-Cre/TdT⁺ vs. Parv-Cre/TdT⁺ profiles showing differentially expressed (DE) transcripts as a volcano plot (blue transcripts, Parv-Cre/TdT enriched; red, SNS-Cre/TdT enriched, 2-fold, p < 0.05). (b) Most enriched Gene ontology (GO) categories and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways in SNS-Cre/TdT vs. Parv-Cre/TdT enriched transcripts, plotted as heat-map of -log (p-value). (c) Volcano plots depicting (i) calcium channels, (ii) potassium channels, and (iii) TRP channels expression differences between populations. Individual transcripts highlighted (red, SNS-Cre/TdT⁺ enriched; green, Parv-Cre/TdT⁺ enriched; blue, not significantly different: 2-fold, p < 0.01).

Figure 9. Differential volcano plot analysis of IB4⁺ and IB4⁻ SNS-Cre/TdTomato subset transcriptomes. (a) Pairwise comparison of IB4⁺ SNS-Cre/TdT⁺ vs. IB4⁻ SNS-Cre/TdT⁺ neuronal profiles show differentially expressed (DE) genes by volcano plot (blue, IB4⁺ enriched; red, IB4⁻ enriched, 2-fold, p < 0.05). (b) Top Gene ontology (GO) categories of biological processes (BP) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways for IB4⁺ SNS-Cre/TdT⁺ and IB4⁻ SNS-Cre/TdT⁺ enriched transcripts, plotted as heat-maps of -log (p-value). (c) Volcano plots showing differential expression of (i) ion channels, (ii) cell adhesion molecules, and (iii) G-protein coupled receptors between neuronal populations (red, IB4⁺ enriched transcripts; green, IB4⁻ enriched; blue, not significantly different: 2-fold, p < 0.01).

Figure 10. Analysis of most enriched marker expression by IB4⁺, IB4⁻ SNS-Cre/TdTomato and Parv-Cre/TdTomato⁺ populations. (a-c) Fold-change/fold-change comparisons illustrate most differentially enriched genes in each subset (highlighted in color are 3-fold and 2-fold enriched numbers). (d) Heat-maps showing relative expression of the top 40 transcripts enriched in each of the three neuronal subsets (>3-fold), ranked by product of fold-change differences.

Figure 10-figure supplement 1. Fluidigm analysis of 100 and 10 cell-samples.

FACS sorted 100 cell or 10 cell samples consisting of IB4⁺ SNS-Cre/TdT⁺, IB4⁻ SNS-Cre/TdT⁺, and Parv-Cre/TdT⁺ neurons were analyzed by Fluidigm for 80 different transcript levels chosen based on microarray results, and normalized to Gapdh expression. Hierarchical clustering of transcript levels is shown for 100 cell and 10 cell groups as heat-maps.

Figure 11. Single cell transcript levels show log-scale distribution across neuronal populations.

Normalized transcript levels in single cells determined by parallel qRT-PCR are plotted on a log-scale comparing IB4⁺ SNS-Cre/TdT⁺, IB4⁻ SNS-Cre/TdT⁺, and Parv-Cre/TdT⁺ cells. (a) Nociceptor related transcript levels (Trpv1, Trpa1, Mrgprd, P2rx3, Nppb, Ptgir), (b) Proprioception related transcript levels (Pvalb, Runx3, Cdh12). Individual neurons are shown as dots in plots.

Figure 12. Hierarchical clustering analysis of single cell qRT-PCR data reveals distinct neuronal subgroups. Heat-map of 334 single neurons and 80 genes after spearman-rank hierarchical analysis of RT-PCR data (relative gene expression normalized to gapdh). Each

column represents a single sorted cell, and each transcript is shown per row. Clustering analysis finds seven distinct subgroups (I, II, III, IV, V, VI, VII). Characteristic transcript expression patterns that delineate each somatosensory subset are written below.

Figure 12-figure supplement 1. Expression of neuronal-associated transcripts across purified single cell samples by qRT-PCR. Heat-map showing expression levels of neuron-associated transcripts across single cells (from 1 to 334) by qRT-PCR.

Figure 12-figure supplement 2. Transcript expression levels for characteristic marker genes in single cell neuron Group I and Group VII. Plotted are normalized transcript levels of Group I and Group VII transcripts, ordered from highest to lowest expression (i.e. Grik1 to Wnt2b for Group I, Pvalb to Cdh12 for Group VII).

Figure 13. Single cell subgroups distribute differentially across originally purified populations.

(a) Principal Components Analysis of single cell transcriptional data shows distinct segregation of Groups I, V, and VII neurons. (b) Proportions of each neuronal subgroup relative to original labeled IB4⁺SNS-Cre/TdTomato⁺, IB4⁺SNS-Cre/TdTomato⁺, and Parv-Cre/TdTomato⁺ neurons.

Figure 14. Focused analysis of single cell heterogeneity and transcript enrichment in neuronal subgroups. (a) Relative expression levels of subgroup specific transcripts in single cells across each neuronal subgroup (each bar = 1 cell). Group I (Lpar3, Mrgprd), group VI (Il31ra, Nppb), and group VII markers (Gpcr5b) show subset enrichment and highly heterogeneous expression at the single cell level. (b-c) Nearest neighbor analysis by pearson correlation of Mrgprd and Pvalb transcript levels to all 80 probes across the single cell expression dataset was generated. Correlation levels go from left to right.

Figure 14-figure supplement 1. Defining the transcriptional characteristics of Group I, II, and IV neurons.

Transcript levels for selected genes that define the characteristics of specific neuronal subgroups Group I, II, and IV neurons were plotted across all 334 individual neurons. (a) Group I neurons were found with high levels of P2rx3, Lpar3. (b) Group II neurons show high levels of Ntrk1 and Kcnv1. (c) Group IV are characterized by Trpv1 expression but lack of Scn10a expression.

Figure 14-figure supplement 2. Expression plots of nociceptor-associated transcripts across single cell transcriptional data. Transcript levels for nociceptor associated genes (a) Trpa1, (b) Scn11a, and (c) Aqp1 were plotted across all individual neurons.

Figure 15. DRG subgroups I, VI, and VII characteristics defined by double RNA *in situ* hybridization. (a) Double RNA *in situ* hybridization in SNS-Cre/TdTomato and Parv-Cre/TdTomato lumbar DRG sections for TdTomato (red) with Lpar3, Il31ra, or Gpcr5b (green), which are Group I, VI, and VII markers respectively. Lpar3 and Il31ra expression colocalize with SNS-Cre/TdTomato but not Parv-TdTomato, while Gpcr5b colocalizes with Parv-Cre/TdTomato but not SNS-Cre/TdTomato. (b) Double *in situ* hybridization in lumbar DRG sections for group VI marker IL31ra vs. Group I marker Lpar3, Group VI marker Gpcr5b, or

Group VI marker Nppb. Il31ra and Nppb is shown in a distinct subset of DRG neurons. Scale bars, 100 μ m.

Figure 15-figure supplement 1. Immunofluorescence characteristics of DRG subgroup V.

(a) Expression plot shows enrichment of Th expression in Group V neurons. (b) SNS-Cre/TdTomato lumbar DRG sections were imaged for TdTomato (red), anti-TH (blue), and IB4-FITC (green). (c) Quantification of neuronal proportions TH⁺ neurons that are IB4⁻SNS-Cre/TdT⁺, IB4⁻SNS-Cre/TdT⁺, or Parv-Cre/TdT⁺ neurons expressing TH. Statistical analysis by student's t-test (n=8-10 fields from 3 mice each). Scale bars, 100 μ m.

Figure 15-figure supplement 2. Group I marker Prkcq is in a distinct subset of DRG neurons.

(a) Transcript levels for Prkcq plotted across all individual neuron subgroups. (b) Double *in situ* hybridization (ISH) of lumbar DRG sections for TdTomato (red) and for Lpar (green) shows that Prkcq⁺ neurons showed SNS-Cre/TdTomato expression whereas they did not express SNS-Cre/TdTomato. Scale bars, 100 μ m. (c) Double ISH of lumbar DRG sections shows that Prkcq does not colocalize with Group VI marker IL31ra.

Table Legends

Table 1. Transcriptional samples analyzed in this study.

In this study, we performed microarray profiling of FACS purified neuron populations, DRG tissue, and single neuron samples. This table summarizes the sample names, descriptions, types, and numbers of samples analyzed. For neuron populations and whole DRG tissue, each biological replicate consisted of pooled total DRG cells from n=3 animals.

Table 2. Taqman assays used for single cell transcriptional profiling.

To perform Fluidigm single cell analysis, Taqman assays were chosen to cover four categories of population-enriched transcripts first identified by microarray whole transcriptome analysis: 1) SNS-Cre/TdT⁺ (total population) enriched markers (vs. Parv-Cre/TdT⁺ neurons), 2) IB4⁺SNS-Cre/TdT⁺ enriched markers (vs. other 2 groups), 3) IB4⁻SNS-Cre/TdT⁺ markers (vs. other 2 groups), and 4) Parv-Cre/TdT⁺ markers (vs. other 2 groups). Taqman assays for housekeeping genes Gapdh and Actb were also included.

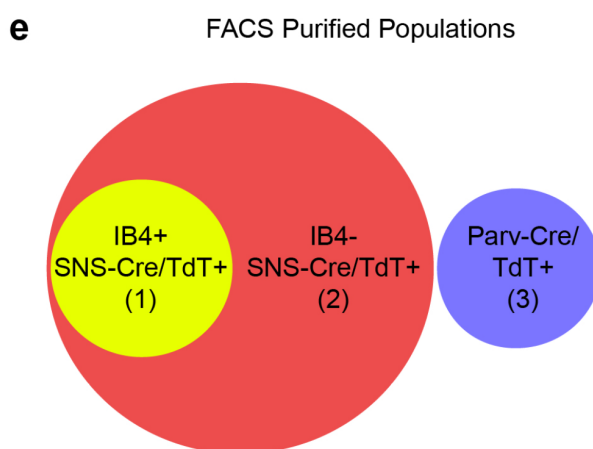
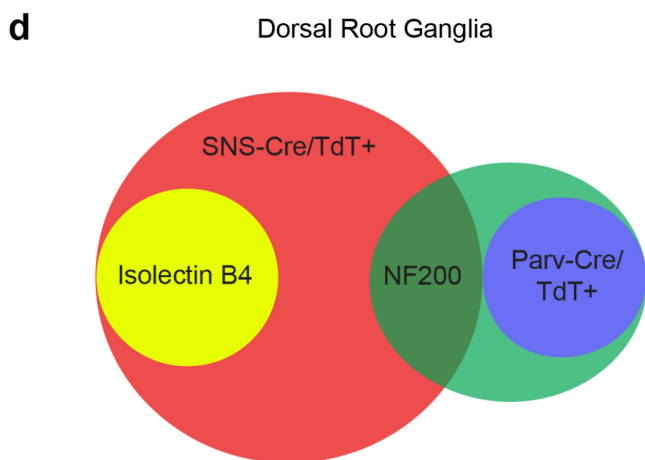
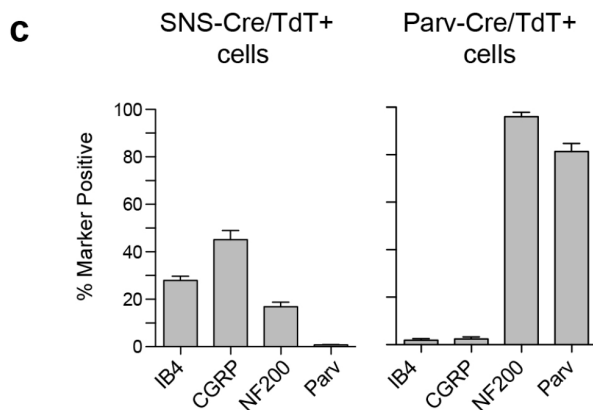
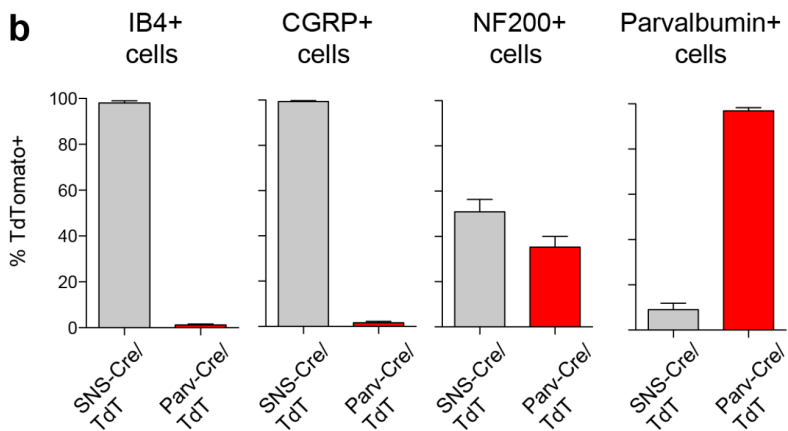
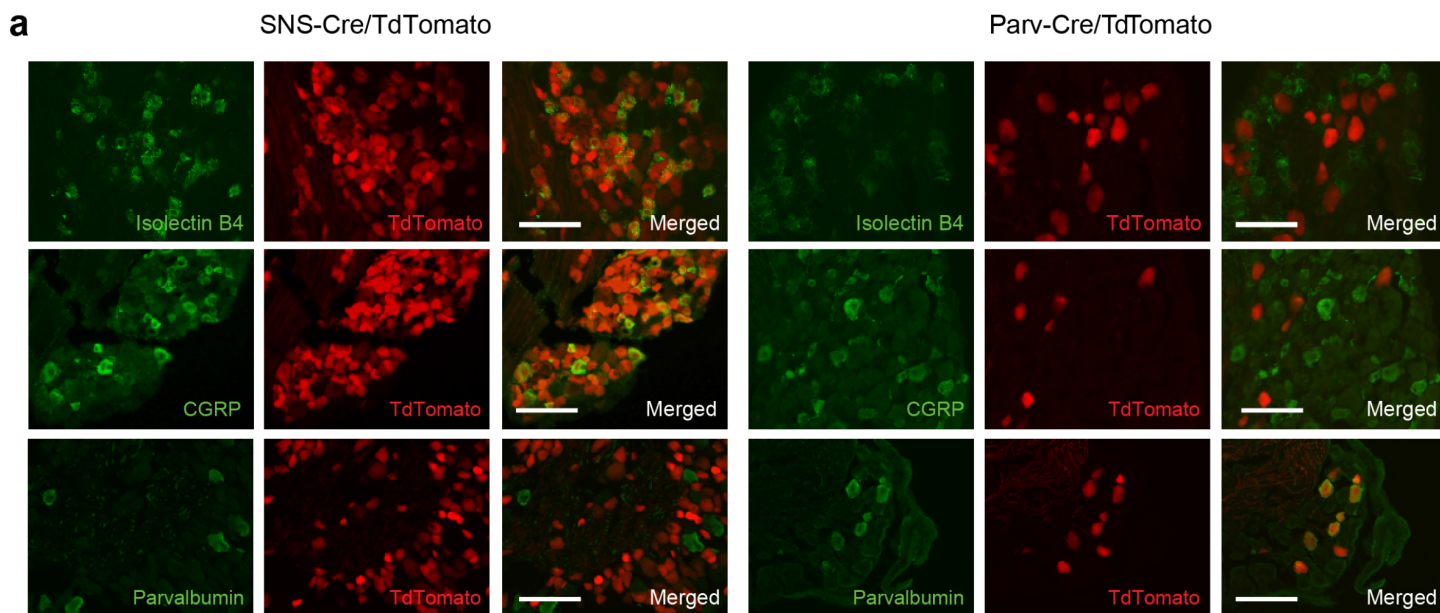
Table 3. RNA *in situ* hybridization probes.

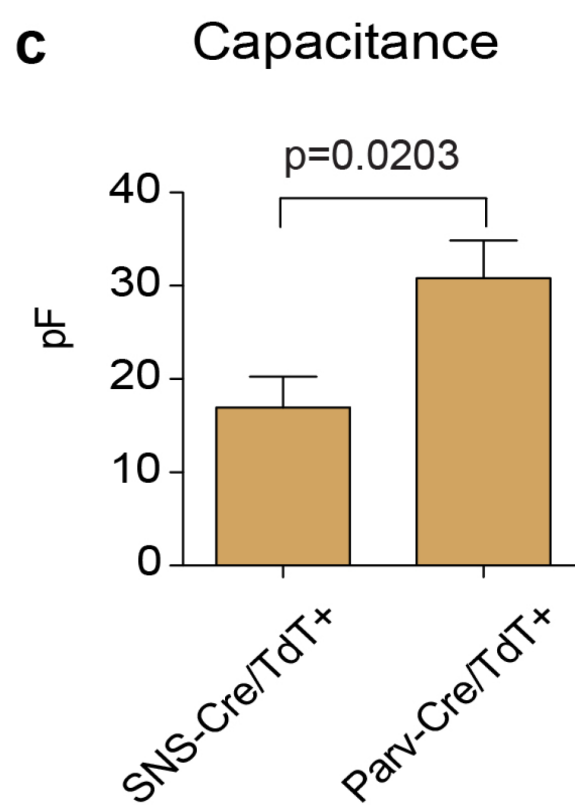
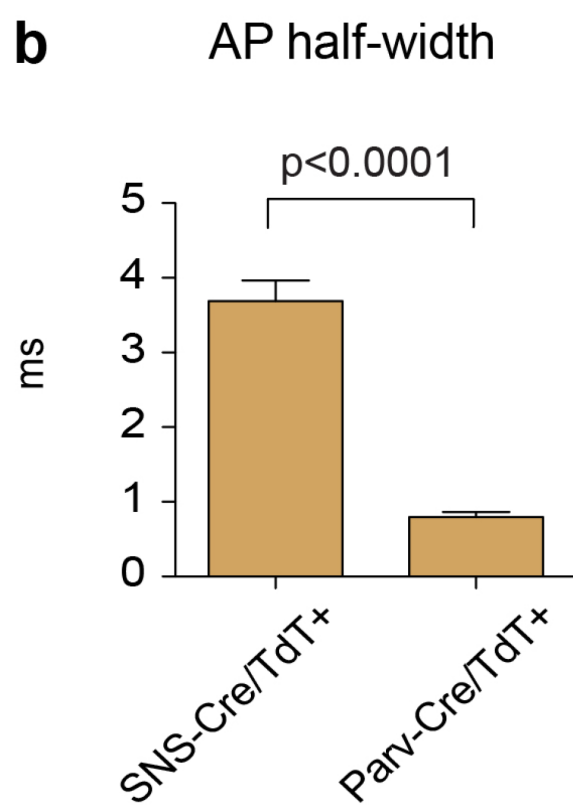
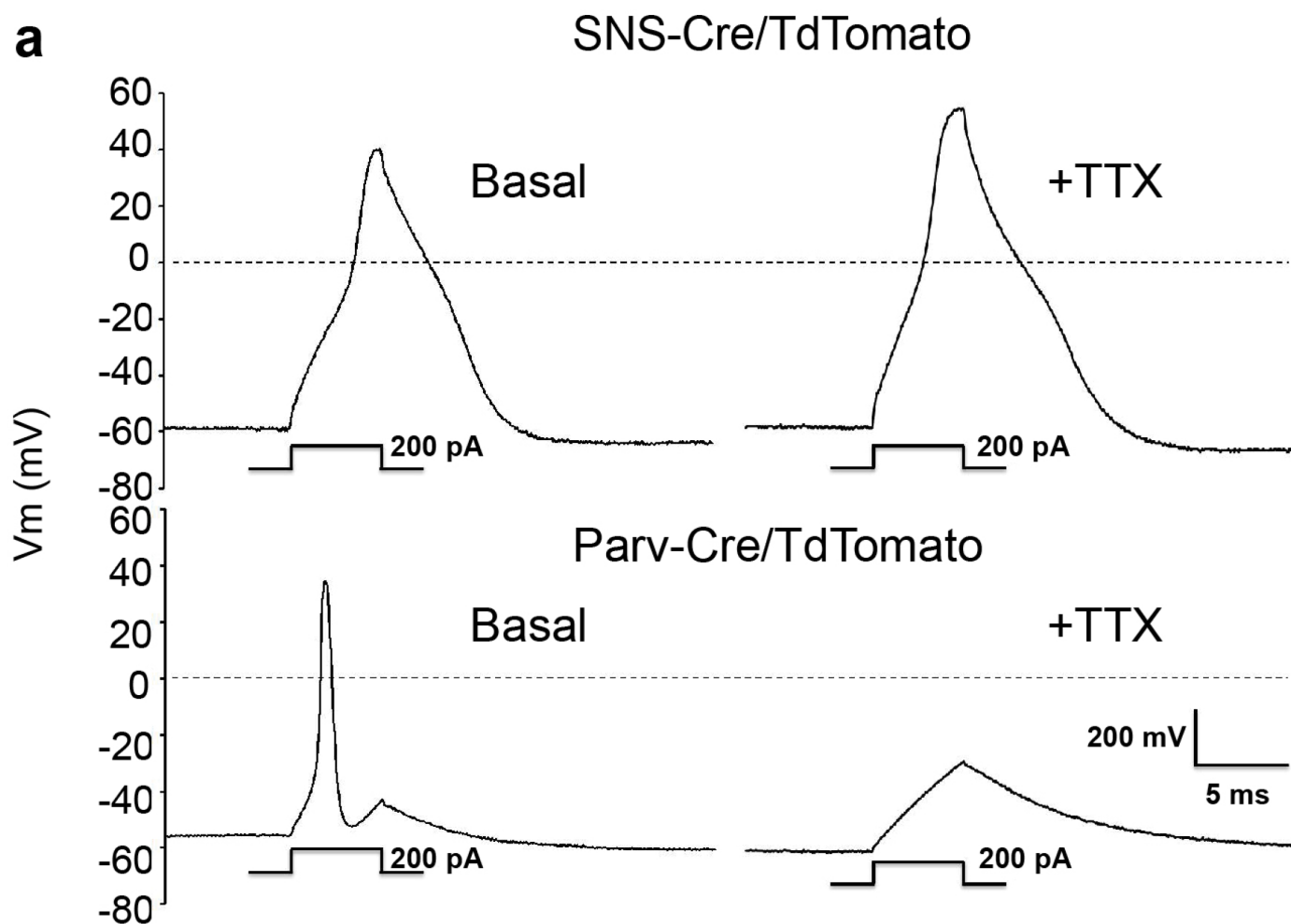
Probesets used for RNA *in situ* hybridization analysis. Listed are gene symbols, sequences for forward and reverse primers, and resulting probe lengths.

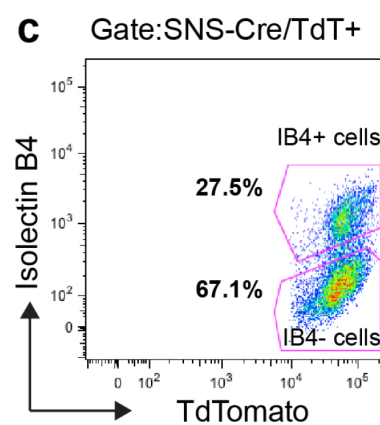
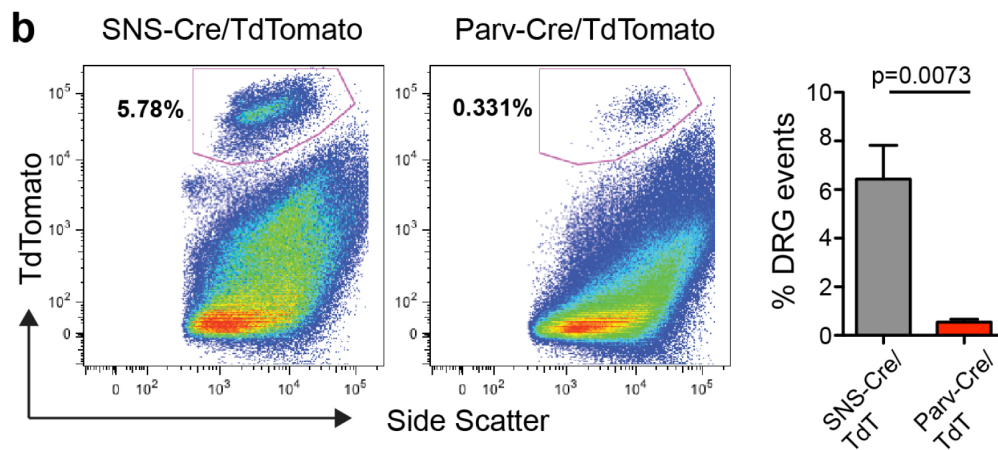
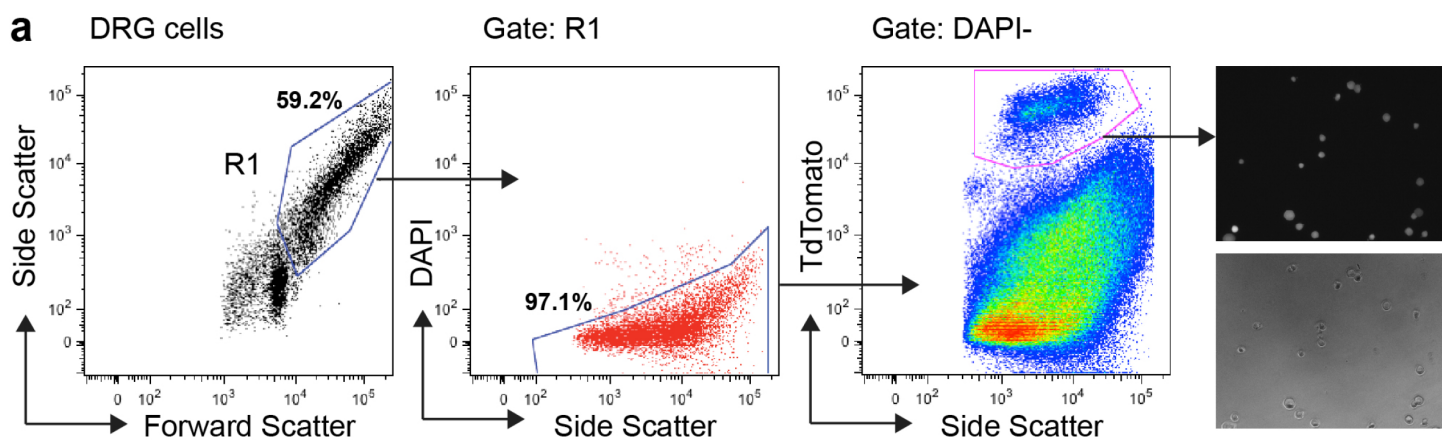
Supplementary File Legends

Supplementary File 1. Comparison of SNS-Cre/TdT vs. Parv-Cre/TdT neuron expression profiles. Differential expression analysis of microarray data from SNS-Cre/TdTomato⁺ neurons (n=4) vs. Parv-Cre/TdTomato⁺ neurons (n=4). Transcripts are ranked by fold-change, with the following information given: Affymetrix ID, genebank accession number, gene symbol, description, average RMA normalized levels, standard deviation, fold-change, p-value and FDR.

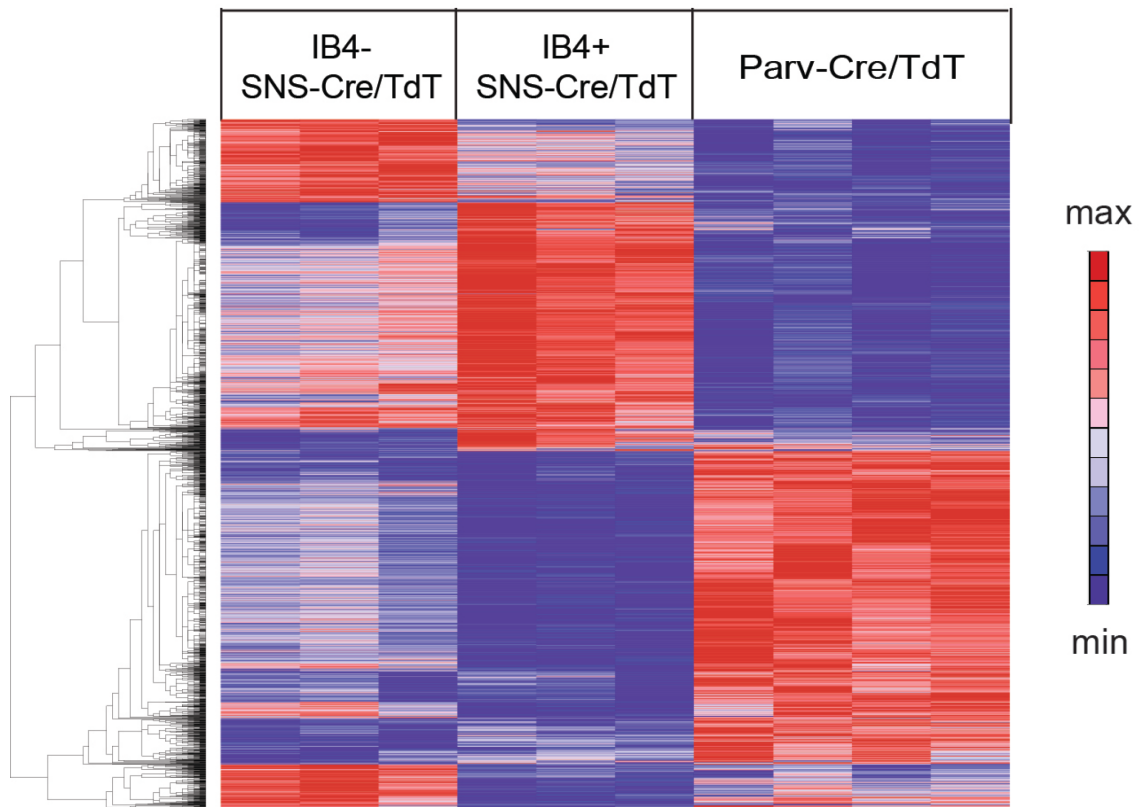
Supplementary File 2. Comparison of IB4 positive vs. IB4 negative SNS-Cre/TdT neuron profiles. Differential expression analysis of microarray data from IB4⁺SNS-Cre/TdTomato⁺ neurons (n=3) vs. IB4⁻SNS-Cre/TdTomato⁺ neurons (n=3). These cells were sorted from the same animals. Transcripts are ranked by fold-change, with the following information given: Affymetrix ID, genebank accession number, gene symbol, description, average RMA normalized levels, standard deviation, fold-change, p-value and FDR.



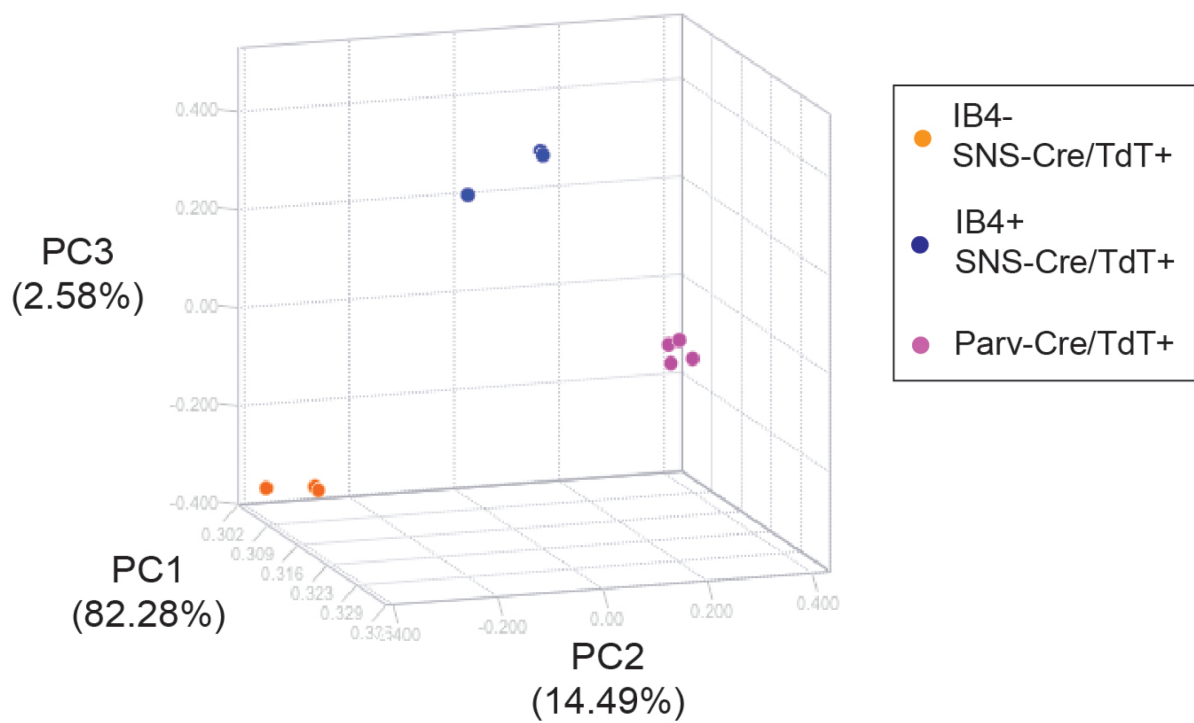




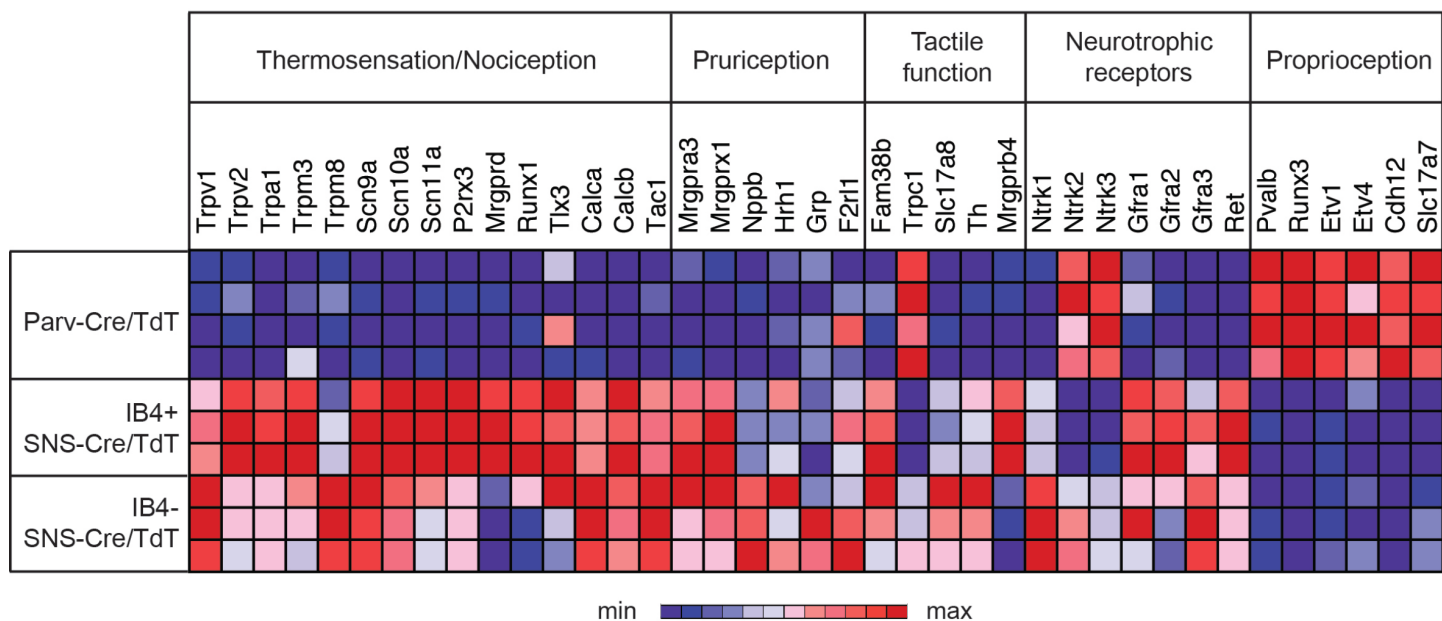
a Hierarchical Clustering



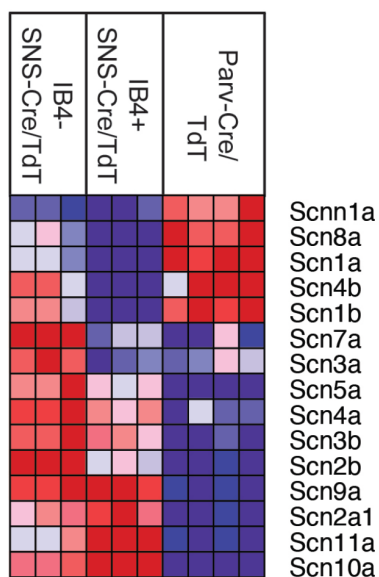
b Principal Component Analysis



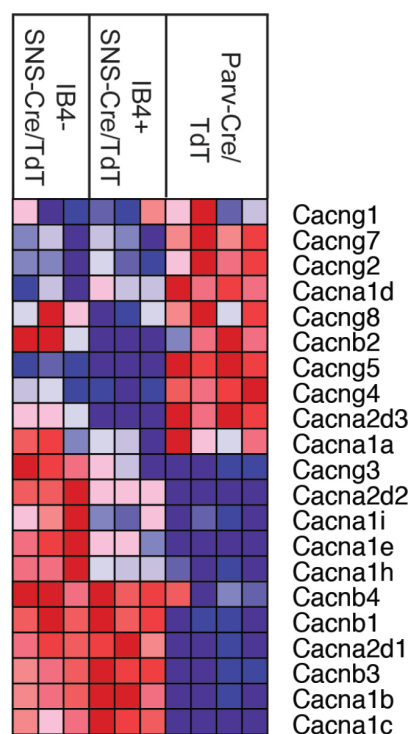
Somatosensory Mediators



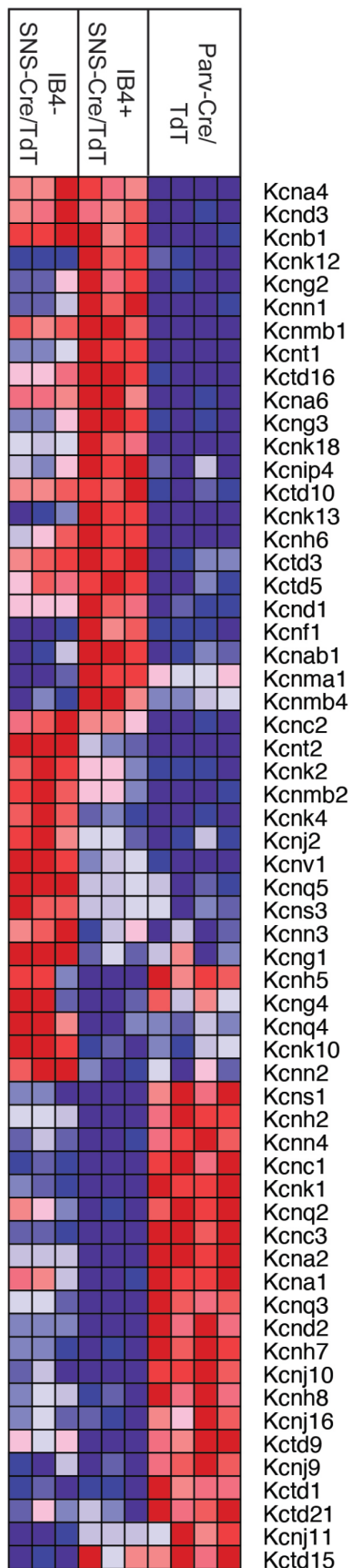
a Sodium Channels



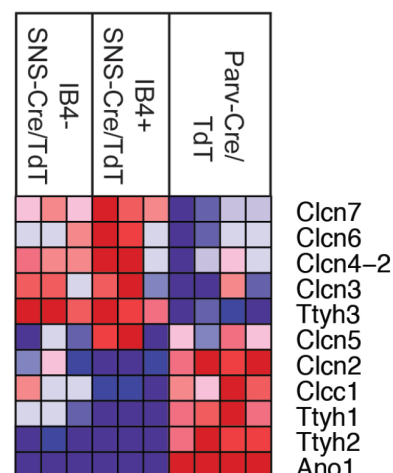
b Calcium Channels



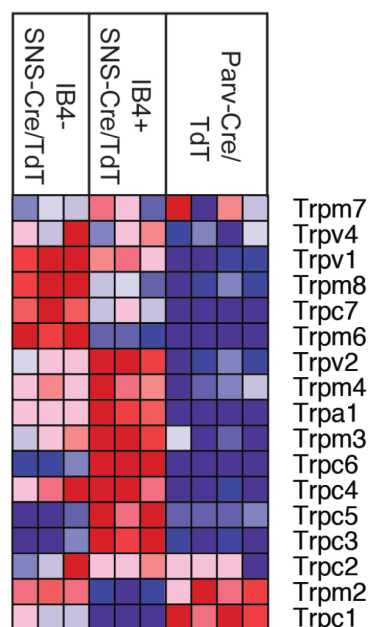
c Potassium Channels



d Chloride Channels

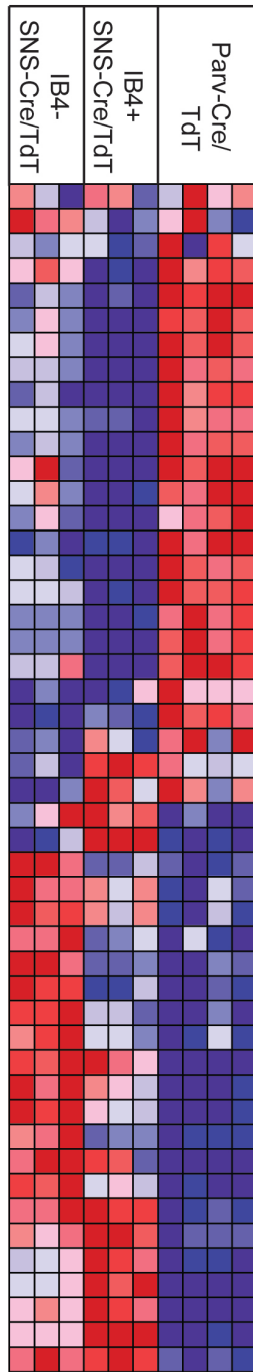


e TRP Channels



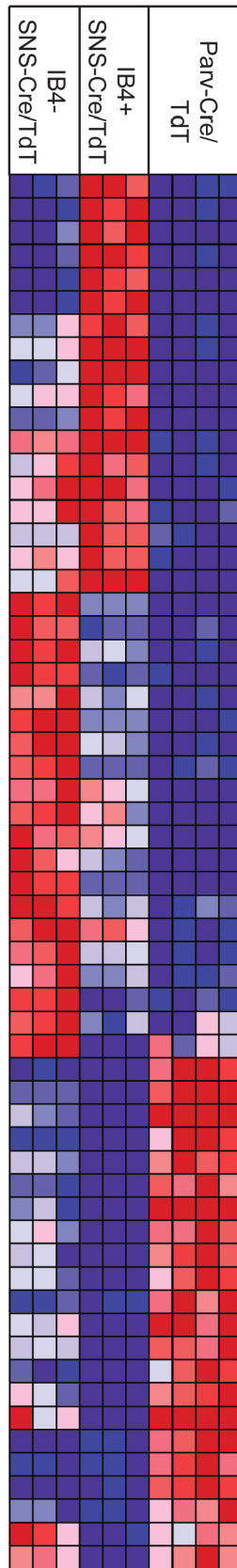
min max

a Ligand-gated
Ion Channels

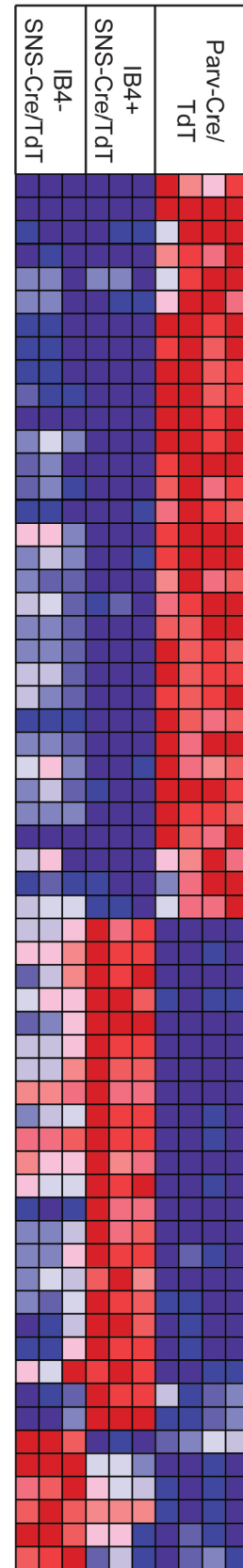


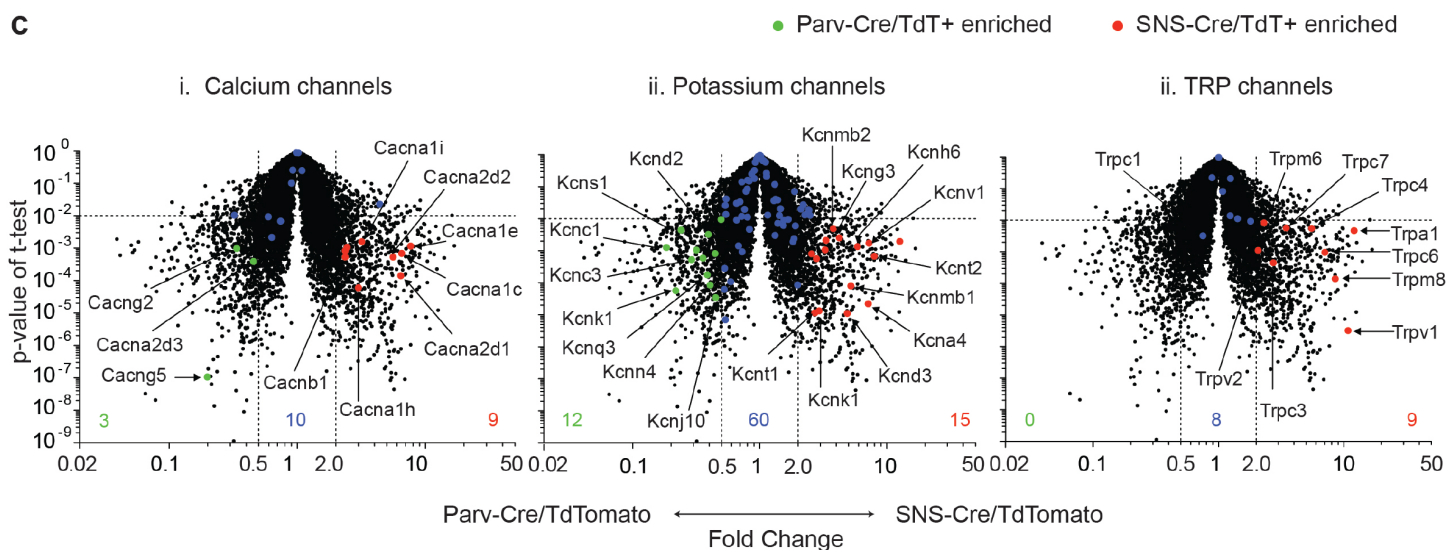
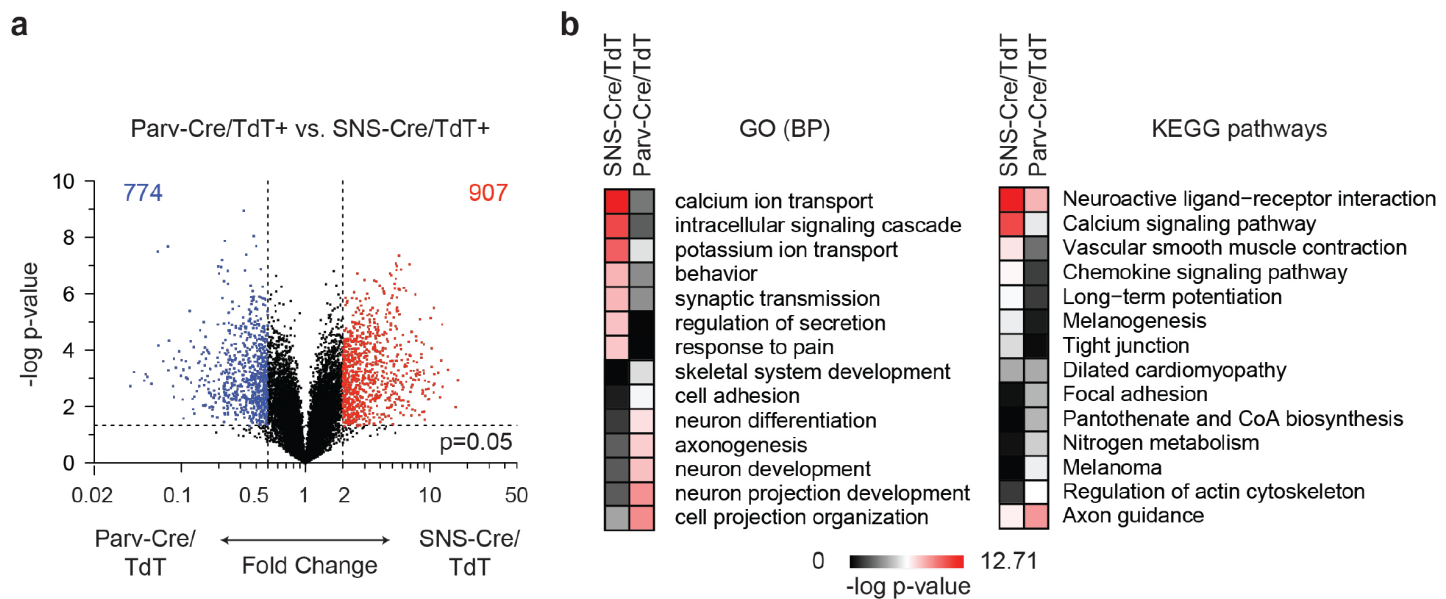
min max

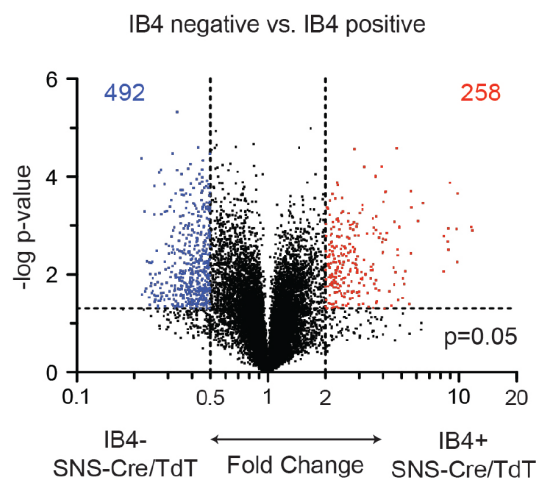
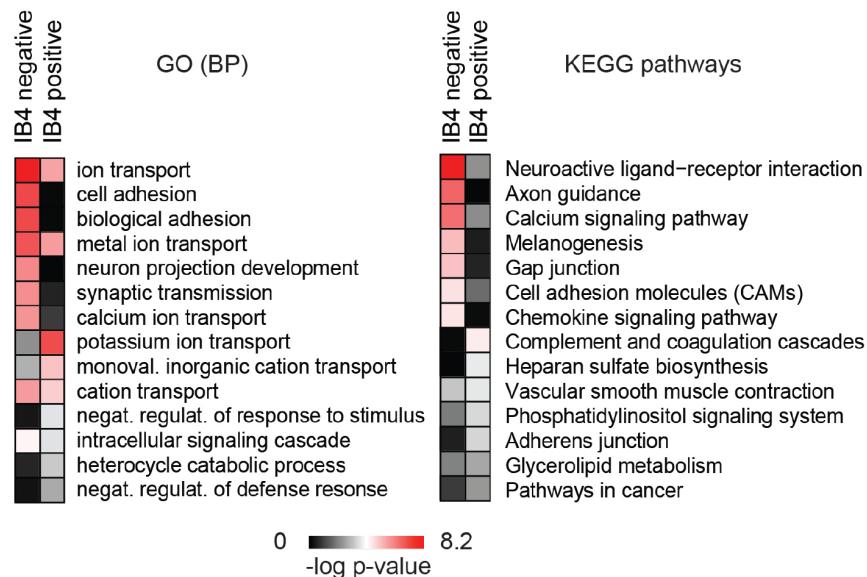
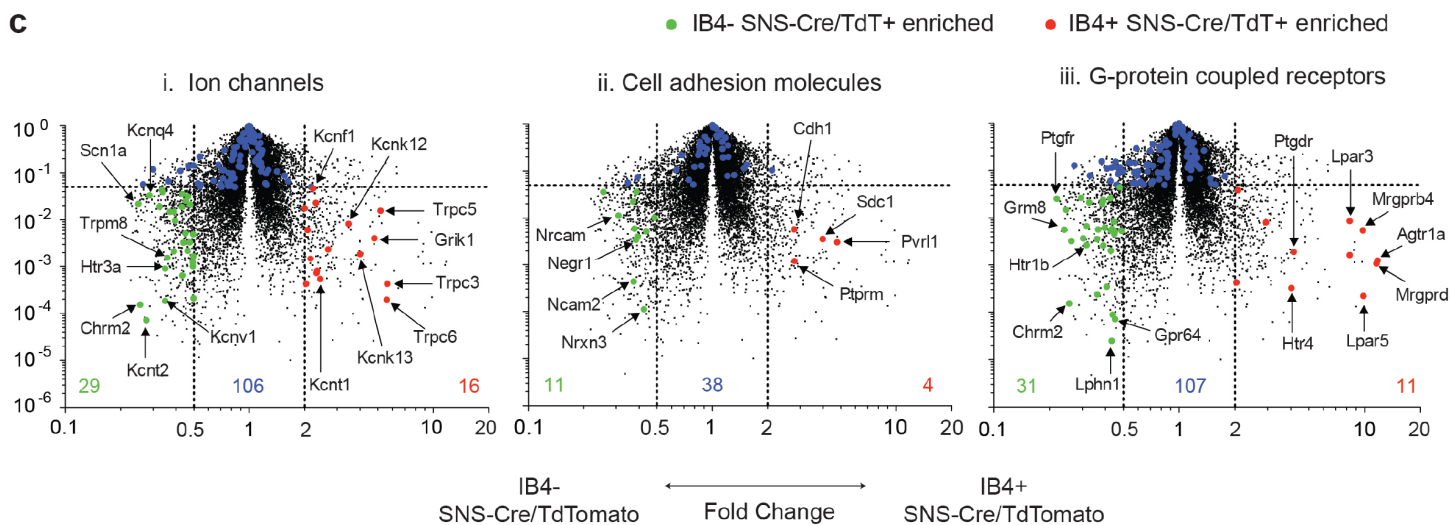
b G-protein
Coupled Receptors

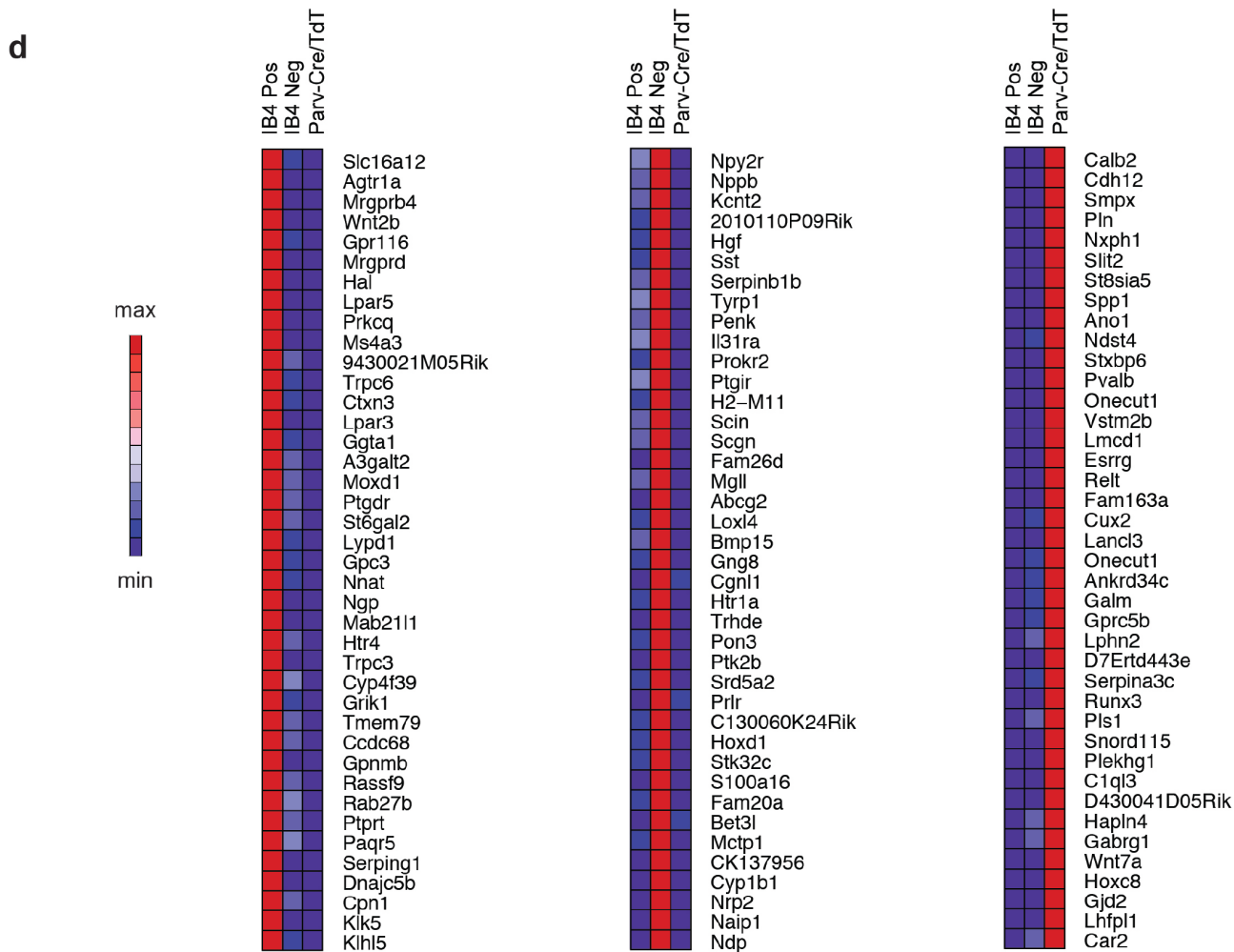
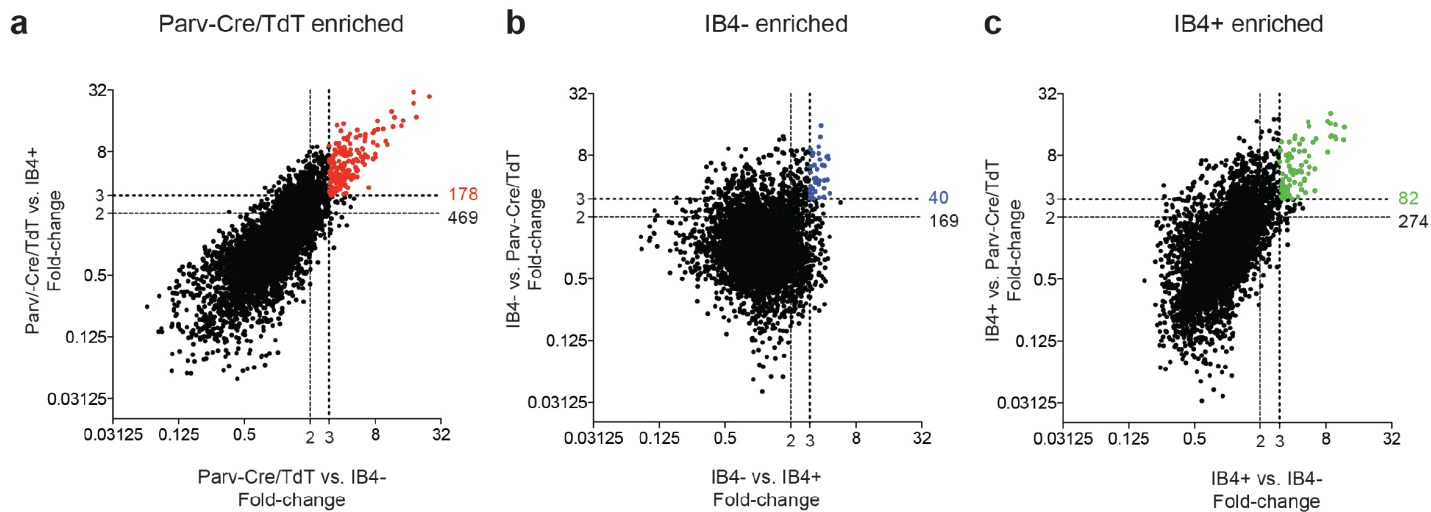


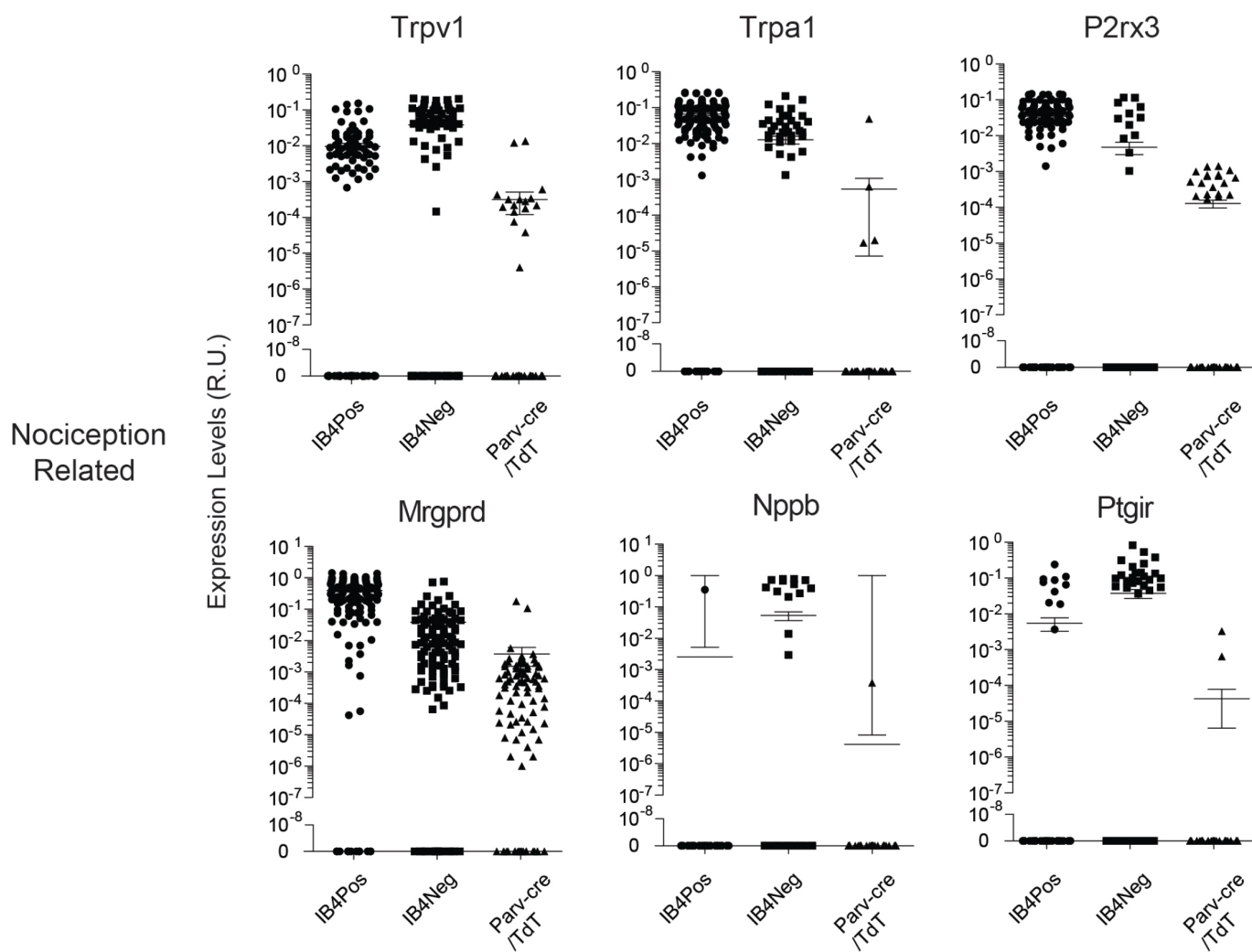
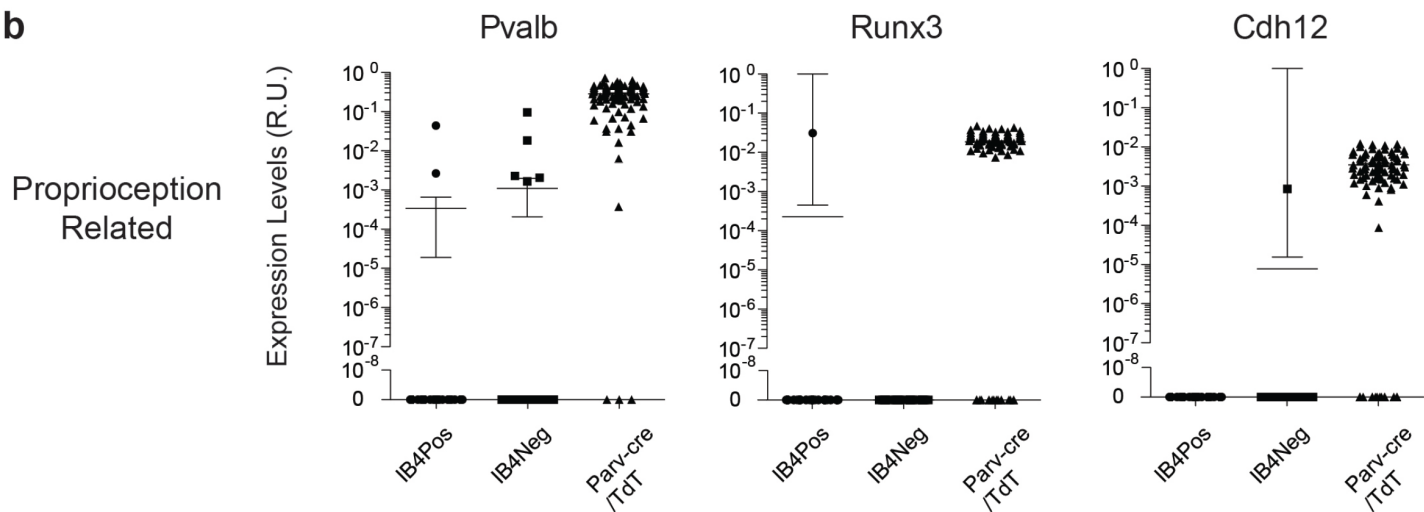
c Transcription
Factors



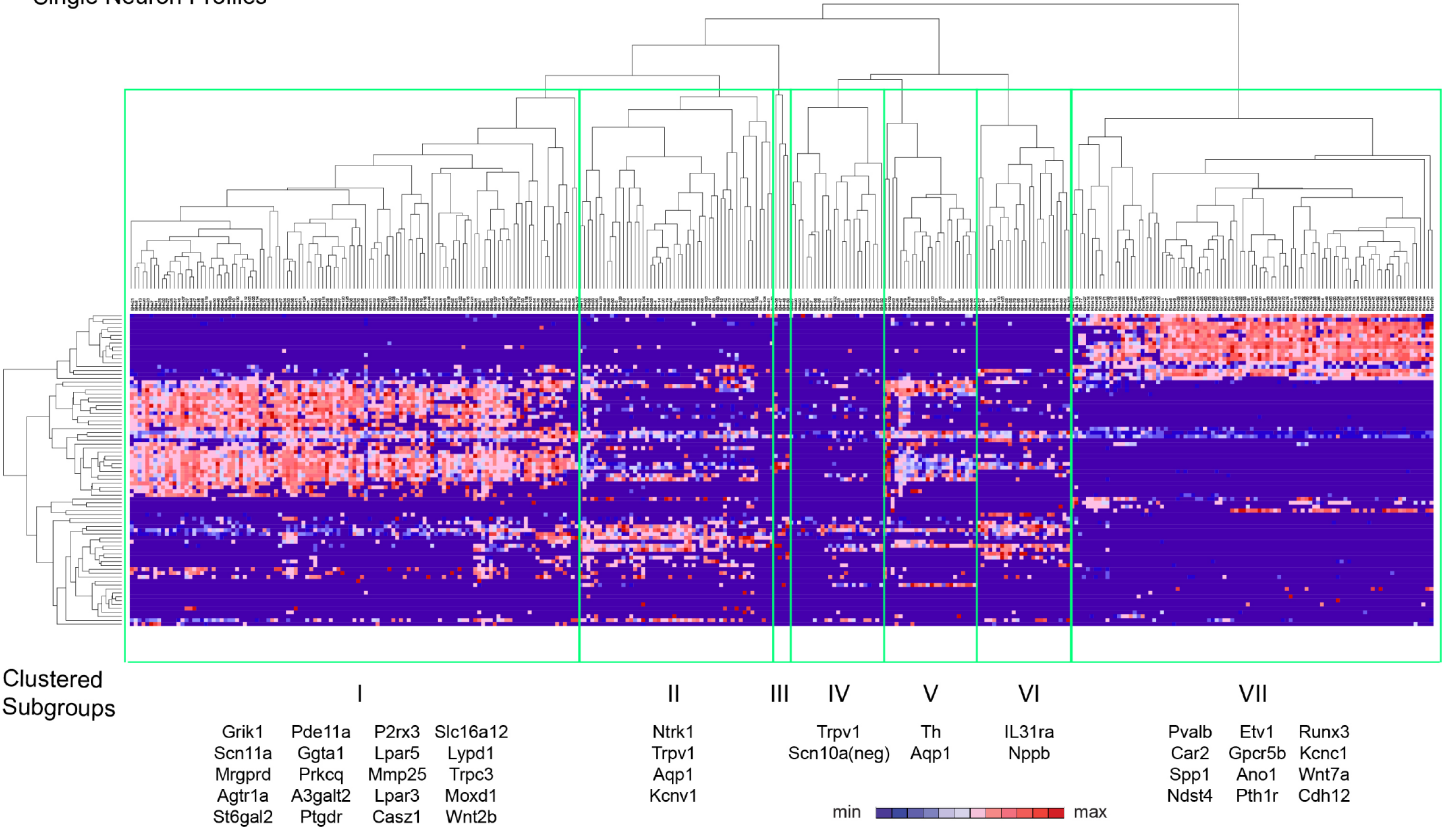


a**b****c**



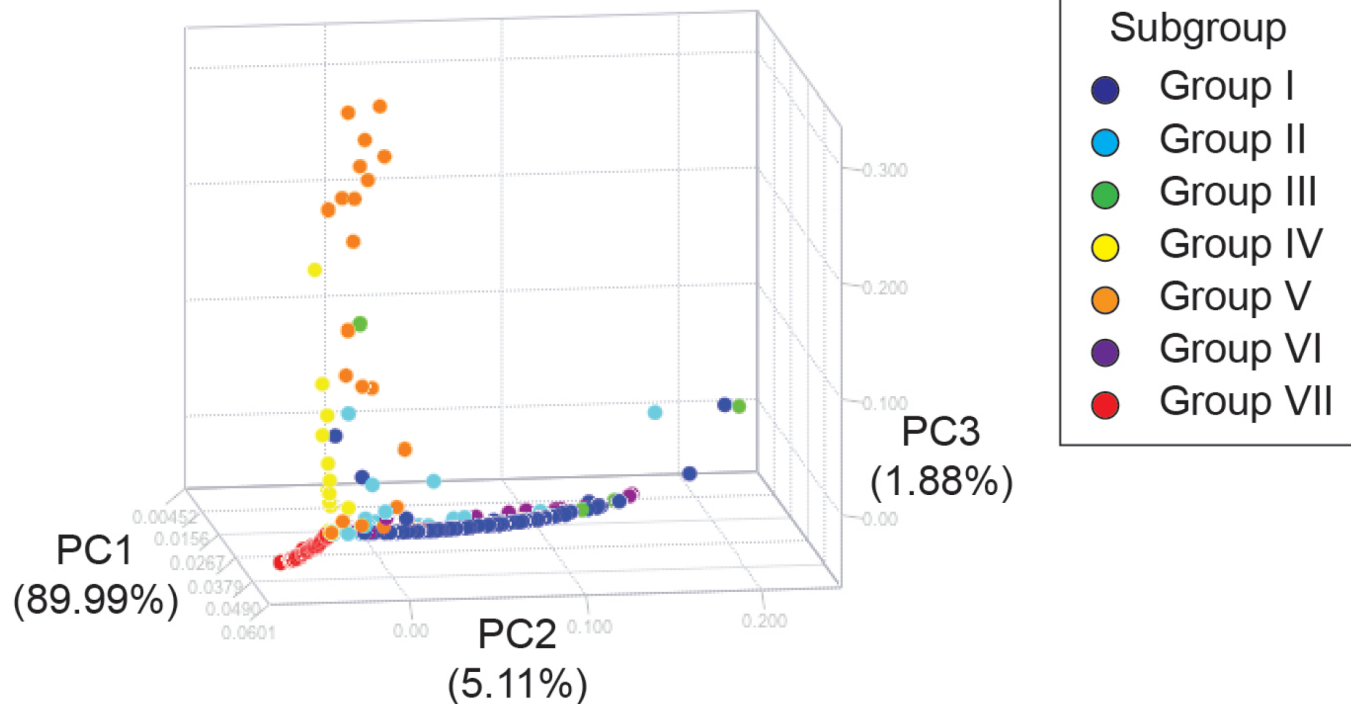
a**b**

Single Neuron Profiles



a

PCA Population Distances

**b**

Neuronal Subgroup Makeup

