nature neuroscience

Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing

Dmitry Usoskin[1], Alessandro Furlan[1], Saiful Islam[1], Hind Abdo[1], Peter Lönnerberg[1], Daohua Lou[1], Jens Hjerling-Leffler[1], Jesper Haeggström[2], Olga Kharchenko[1], Peter V Kharchenko[3,4], Sten Linnarsson[1,5] & Patrik Ernfors[1,5]

The primary sensory system requires the integrated function of multiple cell types, although its full complexity remains unclear. We used comprehensive transcriptome analysis of 622 single mouse neurons to classify them in an unbiased manner, independent of any a priori knowledge of sensory subtypes. Our results reveal eleven types: three distinct low-threshold mechanoreceptive neurons, two proprioceptive, and six principal types of thermosensitive, itch sensitive, type C low-threshold mechanosensitive and nociceptive neurons with markedly different molecular and operational properties. Confirming previously anticipated major neuronal types, our results also classify and provide markers for new, functionally distinct subtypes. For example, our results suggest that itching during inflammatory skin diseases such as atopic dermatitis is linked to a distinct itch-generating type. We demonstrate single-cell RNA-seq as an effective strategy for dissecting sensory responsive cells into distinct neuronal types. The resulting catalog illustrates the diversity of sensory types and the cellular complexity underlying somatic sensation.

The ability to perceive and discriminate diverse types of sensation is reflected by the existence of specialized dorsal root ganglion (DRG) sensory neurons tuned to respond to specific stimuli. Currently, neuronal types are classified based on how a handful of established markers relate to functional aspects rather than by an unbiased comprehensive strategy that would allow an objective classification founded on unique transcriptional states. The relation of neuronal types to the physical and chemical stimuli underlying touch, itch, pain, temperature and proprioception remains largely elusive because of lack of an unbiased classification of types and absence of markers for their identification. Furthermore, as discrimination between different stimuli could arise both by the dedication of neuronal subtypes for specific modalities of sensation and/or by qualitative and quantitative differences in the integrated summation of activities from ensembles of broadly responsive neurons of different types[1,2], subtype assignment of neuronal types could lend new insight into sensory physiology.

Recent advances in single-cell transcriptomics have opened an unprecedented opportunity for studying gene regulation at high resolution. In particular, we and others have suggested a strategy for cell type discovery and classification[3–5] based on the idea that functional cell-type identity will be reflected in the gene expression profile of individual cells. Hence, *de novo* cell type discovery is possible by unbiased single-cell sampling, RNA-seq of each cell and unsupervised grouping of similar expression profiles to reveal cell populations[6,7]. Here we have used single-cell RNA-based dissection of the DRG for classification of neuronal types, which reveals the diversity and complexity of primary sensory system underlying somatic sensation. Thus, the establishment of an unbiased classification of sensory types, supplemented with full information on specific gene expression, will serve as a catalog of the cellular and molecular basis for somatic sensation. The utility of the data is exemplified by a more in-depth analysis of a distinct class of neurons involved in transduction of inflammatory-related itch.

RESULTS

Identification of discrete sensory neuron populations

We performed RNA-seq on 799 single cells from the mouse lumbar DRGs. A total of 2.76 billion reads were generated, of which 1.3 billion were properly barcoded and aligned unambiguously to known exons using RefSeq gene models. On average, 1.14 million reads were mapped to 3,574 ± 2,010 distinct genes in each cell (**Supplementary Fig. 1**). Principal component analysis (PCA) of expression magnitudes across all cells and genes revealed five distinct main clusters (**Fig. 1** and **Supplementary Video 1**). One cluster lacked expression of pan-neuronal markers and expressed the non-neuronal markers β2 microglobulin (*B2m*), vimentin (*Vim*) and several collagens (for example *Col6a2*), on the basis of which we concluded that these cells were non-neuronal. Examining the expression of known markers in the four neuronal clusters (**Fig. 1b** and **Supplementary Video 2**), we identified the first, the NF cluster, as expressing neurofilament heavy chain (*Nefh*) and parvalbumin (*Pvalb*), previously associated

[1]Division of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden. [2]Division of Physiological Chemistry, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden. [3]Center for Biomedical Informatics, Harvard Medical School, Boston, Massachusetts, USA. [4]Division of Hematology/Oncology, Children's Hospital, Boston, Massachusetts, USA. [5]These authors jointly directed this work. Correspondence should be addressed to P.E. (patrik.ernfors@ki.se) or S.L. (sten.linnarsson@ki.se).

Received 6 June; accepted 22 October; published online 24 November 2014; doi:10.1038/nn.3881

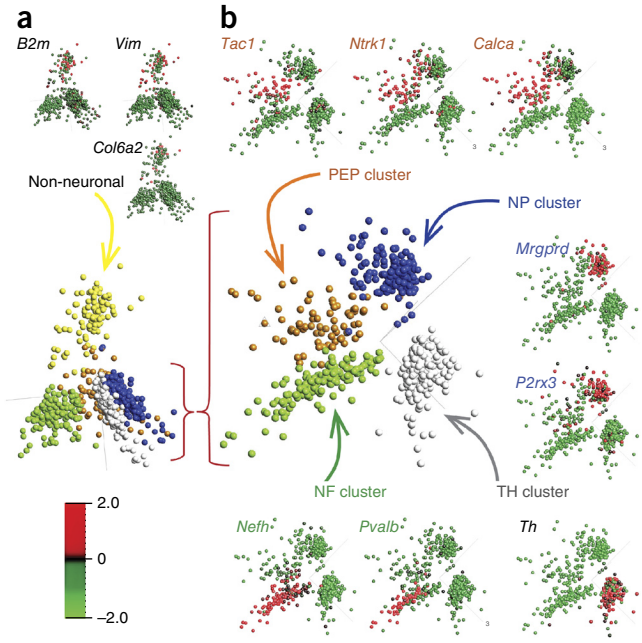


Figure 1 Single-cell RNA-seq of DRG cells reveals cellular heterogeneity. (**a**) Bottom: PCA of single-cell gene expression patterns for 799 individual cells assigned 731 of them to five clusters, one of which contained non-neuronal cells. Insets (top) show the same PCA plot with cells colored according to relative expression of representative genes mapped back to non-neuronal cells. (**b**) PCA of 622 neuronal cells comprising the four principal neuronal types. Insets show the same PCA plot with cells colored according to relative expression of representative genes mapped back to the different populations. Group identity was defined as peptidergic nociceptors (PEP), non-peptidergic nociceptors (NP), neurofilament containing (NF) and tyrosine hydroxylase containing (TH). *Tac1* encodes substance P; *Ntrk1*, TRKA; *Calca*, CGRP; *Mrgprd*, MGRPRD; *P2rx3*, P2X3; *Nefh*, neurofilament heavy chain; *Pvalb*, parvalbumin; *Th*, tyrosine hydroxylase. Color key represents normalized gene expression with mean 0 and variance 1.

with myelinated DRG neurons. The second, the PEP cluster, showed expression of substance P (*Tac1*), TRKA (*Ntrk1*) and calcitonin gene-related peptide (CGRP, also known as *Calca*), previously associated with peptidergic nociceptors. The third, the NP cluster, showed expression of *Mrgprd* and *P2rx3*, previously associated with nonpeptidergic nociceptors. The fourth, the TH cluster, showed distinct expression of tyrosine hydroxylase (*Th*), which has been described in a distinct subclass of unmyelinated neurons[8,9]. 622 cells were classified as neurons, 68 cells had ambiguous assignment and 109 cells were non-neuronal (**Fig. 2a**).

To discover further subtypes of neurons within the main classes, each of the four principal neuronal types was subjected to PCA including only those genes that were differentially expressed when compared against another subtype (the 'outgroup'; see Online Methods). Three new clusters, NF1, NF2/3 and NF4/5, appeared from the NF cluster when the TH cluster was used as an outgroup (**Fig. 2a,b**). Supporting the robustness of this finding, the same three groups appeared when either the PEP or NP clusters were used as the outgroup. The NP neurons split into two groups, NP1 and NP2/3, when the TH cluster was used as outgroup, and the PEP cluster split into two groups when NP and TH clusters were used as outgroups (**Fig. 2a,b**). In all cases, the same subclusters were identified with the different outgroups. We were unable to find any heterogeneity in the TH cluster of neurons. These results show that the genes differentially expressed between groups of cells also represent those that distinguish subtypes within the groups. Notably, in every instance, comparing groups in this manner led to subdivisions within only one of the groups, resulting in a hierarchical relationship among the cell types. Further iterative analysis (**Fig. 2a,b**, PCAs 4–6) led to the identification of a total of eleven neuronal classes in this data set, which we designate NF1 to NF5, NP1 to NP3, PEP1, PEP2 and TH. Notably, the identified basic hierarchical relationship of the eleven different neuronal types is largely consistent with the known developmental origin of sensory neuron types[8,10].

Sensory neuron cell types are historically characterized by distinct soma sizes. Two main categories that were identified initially, large light and small dark cells[11,12], were later extended by introduction of molecular markers. For example, somata of TRKC$^+$ muscle afferents were shown to be large, TRKB$^+$ mechanoreceptors of medium size and TRKA$^+$ nociceptors small[8,13]. Our results agree with these studies and furthermore show that the identified neuronal classes consist of neurons with distinct and characteristic soma sizes (**Fig. 2c**), lending further support to the classification.

We next determined the patterns of expression distinguishing different neuronal clusters and examined whether they were consistent with prior knowledge of sensory neuron function. To identify genes uniquely characterizing each of the eleven groups, we compared the expression within each neuronal type to all other identified types using a recently developed Bayesian approach to single-cell differential expression (SCDE)[14] (**Supplementary Table 1**). The differentially expressed ($P < 0.05$) genes in each comparison were ranked by the fold change (maximal likelihood estimate), and the top fifty genes from each list were pooled into a common set of 452 unique genes. The relative expression levels of these population-specific genes displayed a clear separation between neuronal types (**Fig. 2d**). Gene ontology analysis predicted distinct functional characteristics of different types (**Fig. 2e** and **Supplementary Table 2**), as well as confirming the non-neuronal population (**Supplementary Table 3**).

Interrogation of identified populations by gene expression

Identifying transcripts uniquely expressed in the different cell types lends insight into function and also provides tools for identifying, manipulating and studying the neuronal types. To evaluate the extent to which expression of a gene is restricted to specific populations, we used a 'fraction of positive cells' measure based on a thresholding procedure (see Online Methods). When applied to the merged list of 452 population-specific genes, this measure confirmed that most such genes show high fidelity of expression to one of the identified neuronal subtypes (list of merged top 50 genes, **Supplementary Table 4**; list of all genes, see http://linnarssonlab.org/drg/). These were further validated by scatter plot analyses of absolute gene expression level per cell for all cells across all neuronal types (**Fig. 3a**; a scatter plot for any gene can be visualized at http://linnarssonlab.org/drg/).

To validate *in vivo* relevance of the identified population-specific genes, we chose some genes from the top 50 lists with abundant and selective expression, as well as a high fraction of positive cells, for combinatorial immunohistochemical analyses with available antibodies (**Supplementary Table 5**). The selection is by no means comprehensive in this large data set and does not exclude the possibility that other and/or better markers can be identified. Nevertheless, with the selected markers, we were able to confirm the existence of the discrete neuronal types by double and triple immunohistochemical staining showing mutually exclusive expression patterns (**Fig. 3b–z** and **Supplementary Fig. 2**), in complete agreement with the RNA-seq data. LDHB (lactate dehydrogenase B) marked all NF classes. CACNA1H (calcium channel, voltage-dependent, T type, alpha 1H)

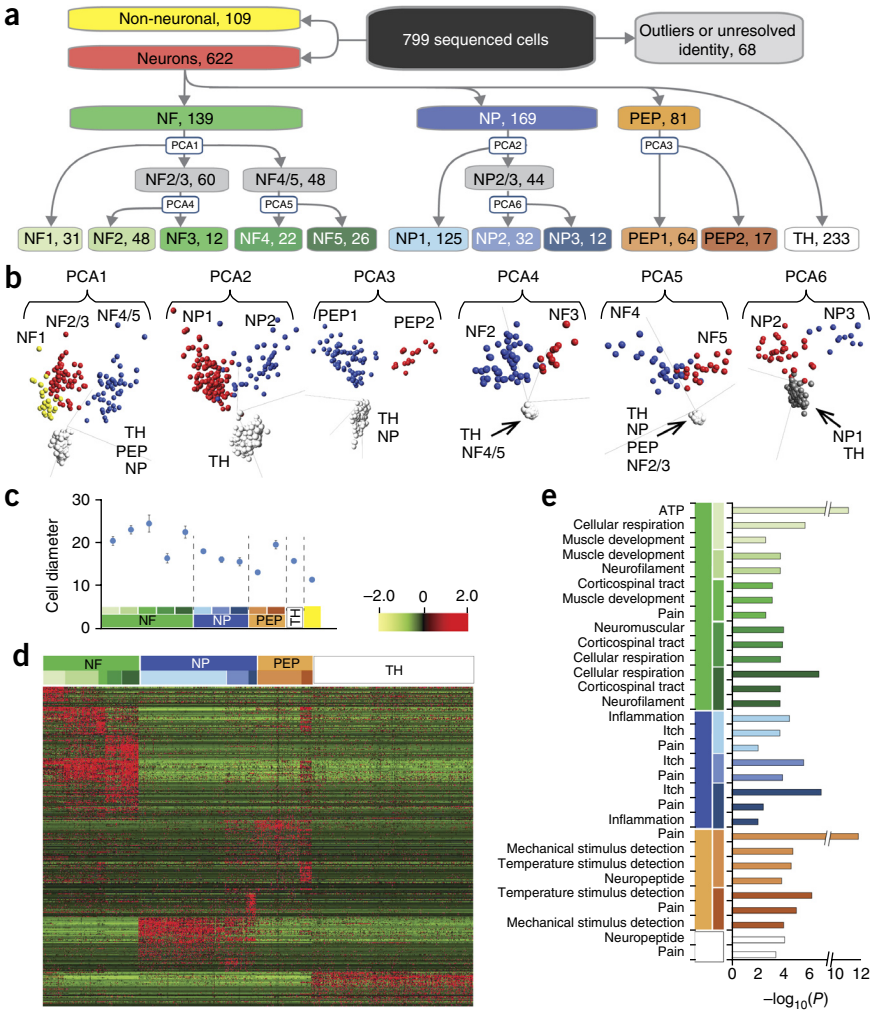


Figure 2 Identification of eleven principal types of sensory neurons by iterative unbiased PCA. (**a**) Outline of iterative analyses defining sensory neuron types. All cell types identified were cross-examined with all other cell types as outgroups. Cell type definitions with numbers of cells are indicated. In some graph nodes, we provide references to plots of representative PCAs, which were used for corresponding population splitting and are shown in **b** . (**b**) PCA plots showing identified cell clusters with outgroups stably generating cell clusters indicated at bottom (arrows). Each PCA was based on genes differentially expressed when tested against the indicated outgroups. The false discovery rates were set as follows: PCA1, 0.0017; PCA2, 0.02; PCA3, 0.005; PCA4, 2.33×10^{-6}; PCA5, 9×10^{-4}; PCA6, 0.047. (**c**) Cell diameters of identified neuronal types (mean ± s.e.m.) as measured on dissociated cells during the picking procedure. (**d**) Heat map of expression of the 50 most significantly enriched genes for each population (452 genes in total, excluding redundancy) in all 622 neurons grouped according to assigned population. See **Supplementary Table 4** for summary of neuronal population expression profile for genes presented in this panel. (**e**) Gene ontology analysis of neuronal types. The graph shows *P* values (Benjamini-Hochberg corrected) for the most significant specific terms reflecting sensory properties or functions (see also **Supplementary Table 2**).

marked NF1 and NF2 classes, which in turn could uniquely be identified by NECAB2 (N-terminal EF-hand calcium binding protein 2) and CALB1 (calbindin 1), respectively. NF3 was uniquely identified by FAM19A1 (family with sequence similarity 19, member A1) and NF4 and NF5 by CNTNAP2 (contactin associated protein-like 2), SPP1 (secreted phosphoprotein 1) and PV (parvalbumin, encoded by *Pvalb*). Double staining for these markers and the receptors TRKB (encoded by *Ntrk2*, neurotrophic tyrosine kinase, receptor, type 2) and TRKC (encoded by *Ntrk3*, neurotrophic tyrosine kinase, receptor, type 3), which previously have been used to identify myelinated sensory neurons, revealed unique complements of receptor expression in these classes. NP1–NP3 classes were confirmed by unique patterns of expression of PLXNC1 (plexin C1), P2X3 (purinergic receptor P2X, ligand-gated ion channel, 3, encoded by *P2rx3*), TRKA (encoded by *Ntrk1*, neurotrophic tyrosine kinase, receptor, type 1), CGRP and SST (somatostatin); PEP neurons were confirmed by their unique expression of TAC1 and FAM19A1 while the TH class was identified by expression of TH. Hence, these results show that the neuronal types identified by our unbiased, whole-genome, single-cell RNA profiling also exist *in vivo,* representing distinct entities, and that they can be distinguished by the corresponding protein markers.

Operational components and functional assignment of neurons

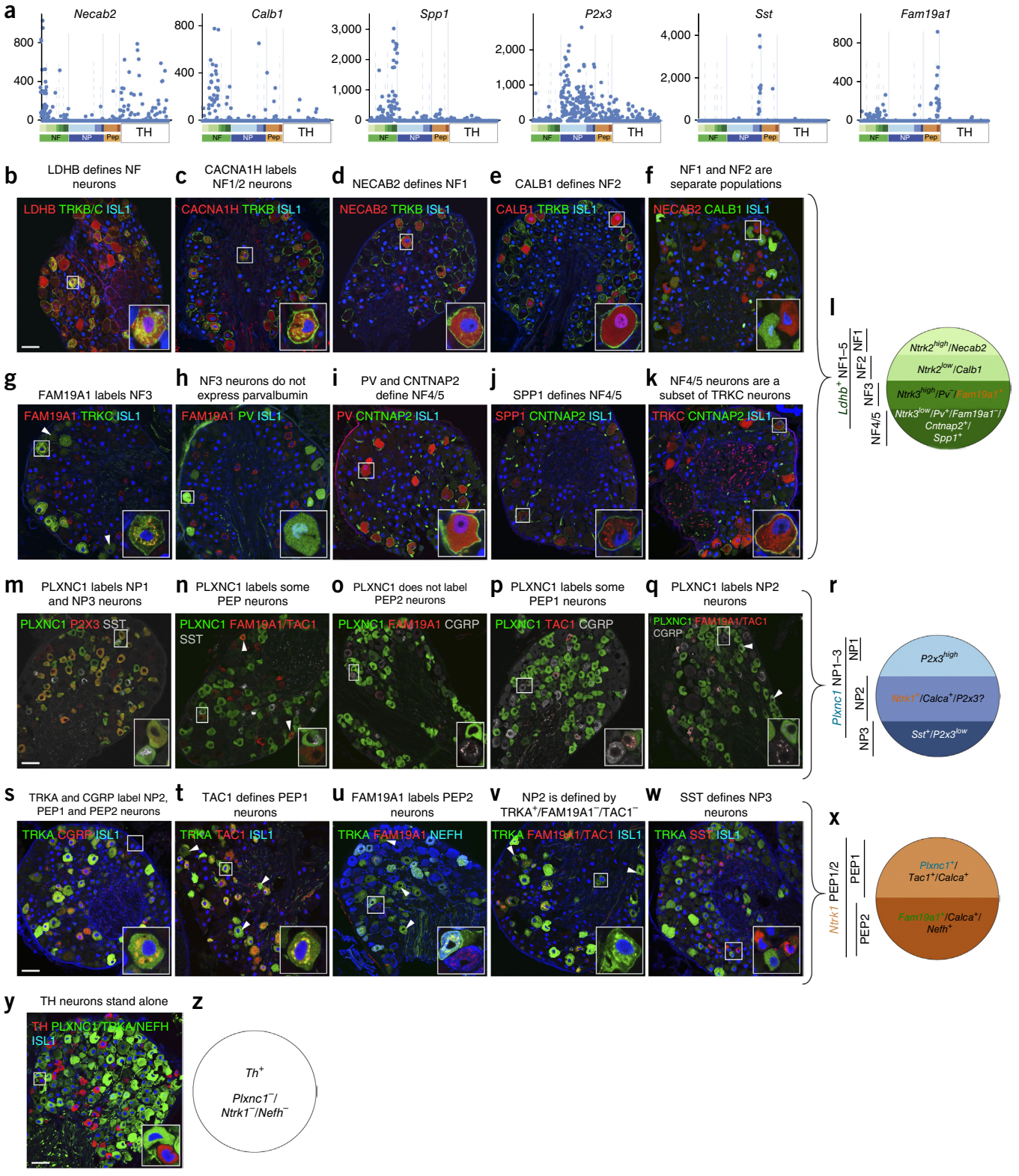
Transduction of sensation relies on the presence of many operational components, such as activation of sensory receptors, depolarization and action potential conduction by voltage-gated Ca^{2+}, Na^{+} and K^{+} channels (Ca_v, Na_v and K_v channels), and release of the neurotransmitter glutamate, neuropeptides and other signaling molecules onto the spinal cord second-order sensory neurons[15]. Sensory transduction can furthermore be influenced presynaptically by local excitatory and inhibitory circuits or descending fibers and by molecular changes associated with inflammatory and neuropathic pain. Analysis of expressed genes identified in the literature to be important in sensory transduction and pain revealed distinct complements of receptors, channels and cellular properties underlying stimuli reception, depolarization, action potential propagation and neurotransmitter release in the identified neuronal populations (**Fig. 4a** and **Supplementary Table 6**).

The RNA profiles, gene ontology analysis and interrogation for the presence or absence of known markers previously used to identify sensory neuron types allows us to propose a relation of the identified sensory neuronal types to known modality-specific functions (**Fig. 4b**). The expression patterns of *Ntrk2*, *Ret*, *Calb1* and *Ntrk3* in NF1–NF3 identify these as low threshold mechanoreceptors[9,16–20]. The presence of TRKC and PV in NF4 and NF5 suggests these as limb proprioceptive neurons[21]. Consistently, all these are myelinated neurons as demonstrated by the expression of light, medium and heavy chain neurofilaments. The NP1 class appears to be involved in neuropathic pain[22–24]. NP1, NP2 and NP3 neuronal types are likely to participate in pruritus[25,26] and NP3 is likely to sense and transduce inflammatory itch. The PEP1 population, defined by expression of the marker *Tac1* (substance P) is consistent with thermosensitive (cold and heat) neurons[27–29] and PEP2 as defined by *Ntrk1* and *Nefh* expression as lightly myelinated Aδ nociceptors. Finally, the TH population,

containing high levels of *Piezo2* and exclusive *Slc17a8* (also known as *Vglut3*, vesicular glutamate transporter 3) expression, is consistent with a function as type C low-threshold mechanoreceptors (C-LTMRs) involved in mechanical pain and pleasant touch[9,30,31]. In addition, the selective expression of receptors for previously uncategorized pain and itch mediators and neuropeptide ligands predicted subtype-specific sensory responsiveness across the different neuronal types (**Table 1** and **Supplementary Tables 4** and **7**).

A small population tuned to respond to inflammatory itch

As a validation of the overall analysis, we examined predicted itch-transducing neurons in greater detail. Gene ontology analysis



supplemented with a custom category containing itch-related genes (see Online Methods), indicated responsiveness of NP1, NP2 and NP3 neurons to itch (**Fig. 2e**). Further analysis confirmed the presence of responsive receptors and predicted distinct response profiles to pruritic substances across these three classes of neurons (**Table 1**), with lysophosphatidic acid–responsive neurons (*Lpar3* and *Lpar5*) in NP1 class, chloroquine-responsive neurons (*Mrgpra3* and *Mrgprx1*) in NP2, interleukin (IL)-31 (*Il31ra* and *Osmr*)- and cysteine leukotriene (*Cysltr2*)-responsive in NP3. Histamine receptors (*Hth1*) were found in NP2 and NP3 and serotonin receptors (*Htr1f*, *Htr2a*) in NP3 and PEP2. Each class of unmyelinated neurons contained distinct complements of neuropeptides (**Table 1** and **Supplementary Table 7**).

In-depth immunohistochemical analyses of the NP3 population confirmed its unique properties. NP3 was distinct from NF1–NF5 and PEP2, as SST$^+$ neurons were NEFH (neurofilament heavy polypeptide) negative; distinct from NP2 and PEP1 and 2, as SST$^+$ neurons were separate from TRKA$^+$ neurons; distinct from TH and NP1, as SST did not colocalize with either TH or GFRA2 (**Fig. 5a–d**). Immunohistochemical identification of each class of unmyelinated neurons and quantification of neuronal size revealed NP3 neurons to be small in diameter (**Fig. 5e**). The NP3 class of neurons contained only low levels of RET and P2X3 and was negative for isolectin B4 (IB4) staining, markers previously used to define nonpeptidergic neurons (**Fig. 5f–k**). Consistent with our previous results, we confirmed high levels of the marker PLXNC1 in this population (**Fig. 5j–k**). In addition to receptors for cysteinyl leukotrienes (*Cysltr2*), IL-31 receptor (co-receptors *Il31ra* and *Osmr*) and the serotonin receptor (*Htr1f*), NP3 neurons were rich in other unique markers, including the neuropeptides natriuretic peptide, neurotensin and somatostatin (*Nppb*, *Nts* and *Sst*, respectively; **Table 1**), with confirmed full colocalization of SST, IL31RA and neurotensin (NTS) (**Fig. 5l,m**). *Il31ra* and cysteine leukotriene receptor expression in NP3 neurons indicate cellular responsiveness to the respective ligands and hence predict itching behavior in response to these inflammatory mediators. Although demonstrating causality will require conditional gene ablation in sensory neurons, we nevertheless recorded scratching bouts following subcutaneous injections of IL-31 and cysteine leukotriene D4 (LTD4). Both agents reliably triggered scratching (**Fig. 5n**). Furthermore, the SST marker for this class of neurons emerged unexpectedly late in development, with most neurons appearing after postnatal day 11 (**Fig. 5o–q**).

DISCUSSION

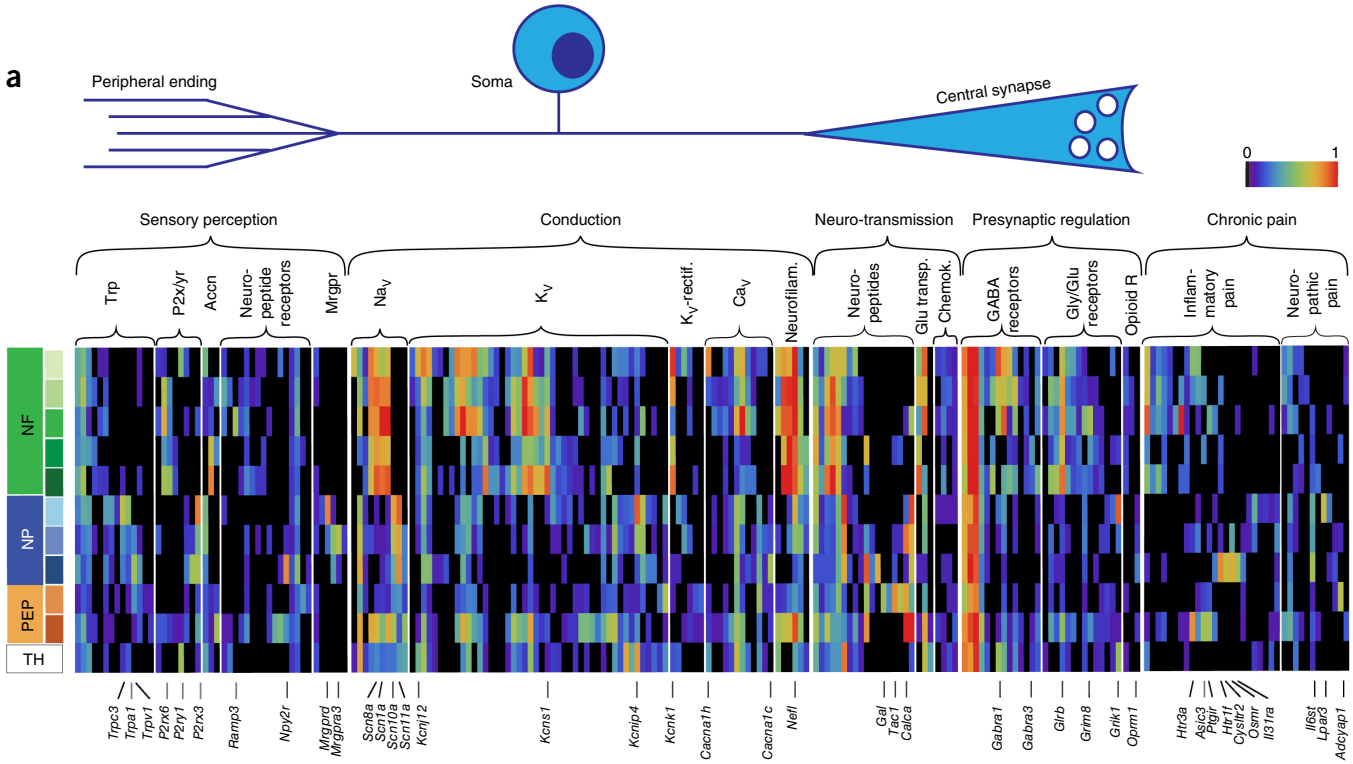
Despite intensive research using marker and physiological approaches, known functional heterogeneity among sensory neurons is not always consistent with marker-based subtype classification schemes. Using unbiased sampling and modeling of transcriptional states in single cells, we detected a high variance in gene expression among cells. We found that gene variance representing cell heterogeneity across analyzed sensory neurons clearly determined subtypes. We focused our analysis on sensory neurons from lumbar DRGs in order to have a full set of proprioceptors and include neurons innervating glabrous skin. Eleven fundamentally distinct types of sensory neurons were discovered and validated by immunohistochemical staining. Thus, we demonstrate the utility of single-cell RNA sequencing from a complex neural tissue for unbiased molecular discovery of previously unknown neuronal types. Our classification does not contradict a polymodal nature of C-fiber nociceptors, but it formally shows for the first time that discrete transcriptional states in different classes of unmyelinated polymodal neurons underlie distinct stimulus response.

The current scheme proposes that proprioception, C-LTMR, TrpM8-dependent cold detection[28,32], some forms of pruritus, and probably acute mechanical pain may be tuned to specific neuronal types relatively selectively, whereas other types of sensation are encoded by a broader integration of ensembles of neurons with different transcriptional states. Our results show, for example, that the well-studied heat sensitive channel TRPV1 (ref. 27) is present in the NP2, NP3 and PEP1 classes of neurons; the noxious heat channel TRPM3 (ref. 33) is present in NP1 and PEP1; and TRPA1, activated by pungent chemical such as mustard oil, ginger and clove and itch receptors, is present in all three NP classes as well as the TH class.

Figure 3 *In vivo* interrogation of identified neuronal types. (**a**) Representative graphs of single-cell absolute expression levels (reads per million, RPM) of some of the genes taken for *in vivo* validation. (**b–z**) Immunohistochemical identification of neuronal types. Validation of NF1–NF5 (**b–l**), of NP1–NP3, PEP1 and PEP2 (**m–x**) and of TH (**y**,**z**) classes of sensory neurons. (**b**) Triple immunohistochemistry for LDHB, combined TRKB/TRKC and ISLET1 (ISL1). Note LDHB$^+$/(TRKB$^+$ or TRKC$^+$) neurons (inset). (**c**) Triple immunohistochemistry for CACNA1H, TRKB and ISLET1. Note TRKB$^+$/CACNA1H$^+$ neurons (inset). (**d**) Triple immunohistochemistry for NECAB2, TRKB and ISLET1. Note NECAB2$^+$/TRKB^{high} neurons (inset). (**e**) Triple immunohistochemistry for CALB1, TRKB and ISLET1. Note CALB1$^+$/TRKB^{low} neurons (inset). (**f**) Triple immunohistochemistry for NECAB2, CALB1 and ISLET1. NECAB2 and CALB1 are mutually exclusive (inset). (**g**) Triple immunohistochemistry for FAM19A1, TRKC and ISLET1. Note FAM19A1$^+$/TRKC$^+$ neurons (inset) and FAM19A1$^-$/TRKC$^+$ neurons (arrowheads). (**h**) Triple immunohistochemistry for FAM19A1, PV and ISLET1. PV and FAM19A1 expression is mutually exclusive (inset). (**i**) Triple immunohistochemistry for PV, CNTNAP2 and ISLET1. Note PV$^+$/CNTNAP2$^+$ neurons (inset). (**j**) Triple immunohistochemistry for SPP1, CNTNAP2 and ISLET1. Note SPP1$^+$/CNTNAP2$^+$ neurons (inset). (**k**) Triple immunohistochemistry for CNTNAP2, TRKC and ISLET1. Note CNTNAP2$^+$/TRKC$^+$ neurons (inset). (**l**) Scheme showing the five subgroups of the major NF group: NF1 and NF2 classes are identified by expression of high *Ntrk2* (*Ntrk2^{high}*), *Necab2* and low *Ntrk2* (*Ntrk2^{low}*) and *Calb1*, respectively. *Cacna1h* is expressed by low- and high-*Ntrk2*-expressing neurons. NF3 is defined by expression of *Ntrk3* and *Fam19a1*, NF4 and NF5 by expression of *Ntrk3* and *Pvalb*, *Cntnap2* and *Spp1*. All NF neurons express *Ldhb*. (**m**) Triple immunohistochemistry for PLXNC1, P2X3 and SST. Note SST$^+$/PLXNC1$^+$/P2X3$^+$ neurons (inset). (**n**) Triple immunohistochemistry for PLXNC1, combined FAM19A1/TAC1 and SST. Note (FAM19A1 or TAC1)$^+$/PLXNC1$^+$ (arrowheads) and SST$^+$/FAM19A1$^-$/TAC1$^-$ neurons (inset). (**o**) Triple immunohistochemistry for PLXNC1, FAM19A1 and CGRP. FAM19A1 and PLXNC1 expression is mutually exclusive (inset). (**p**) Triple immunohistochemistry for PLXNC1, TAC1 and CGRP. Note PLXNC1$^+$/TAC1$^+$ neurons (inset). (**q**) Triple immunohistochemistry for PLXNC1, combined FAM19A1/TAC1 and CGRP. Virtually all CGRP$^+$/FAM19A1$^-$/TAC1$^-$ neurons are PLXNC1$^+$ (arrowheads and inset). (**r**) Scheme showing the three subgroups of the main NP group. All NP neurons express *Plxnc1*. NP subgroups are identified by combinatorial expression of *P2x3* (NP1), *Ntrk1* and *Calca* (NP2) and *Sst* and *P2x3* (NP3). (**s**) Triple immunohistochemistry for TRKA, CGRP and ISLET1. TRKA and CGRP show 1:1 colocalization (inset). (**t**) Triple immunohistochemistry for TRKA, TAC1 and ISLET1. Note TRKA$^+$/TAC1$^+$ (inset) and TRKA$^+$/TAC1$^-$ neurons (arrowheads). (**u**) Triple immunohistochemistry for TRKA, FAM19A1 and NEFH. Note TRKA$^+$/FAM19A1$^+$ and NEFH$^+$/FAM19A1$^+$ (PEP2 and NF3, respectively, inset) and TRKA$^+$/FAM19A1$^-$ neurons (arrowheads). (**v**) Triple immunohistochemistry for TRKA, combined FAM19A1/TAC1 and ISLET1. Note TRKA$^+$/FAM19A1$^-$/TAC1$^-$ neurons (inset and arrowheads). (**w**) Triple immunohistochemistry for TRKA, SST and ISLET1. SST and TRKA expression is mutually exclusive (inset). (**x**) Scheme showing the two subgroups of the main PEP group. PEP1 neurons are characterized by their unique expression of *Tac1*. Some of them express *Plxnc1*; PEP2 neurons are myelinated (*Nefh$^+$*) and express *Fam19a1*. All PEP neurons express *Ntrk1* and *Calca*. (**y**) Triple immunohistochemistry for TH, combined PLXNC1/TRKA/NEFH and ISLET1. Note TH$^+$/TRKA$^-$/PLXNC1$^-$/NEFH$^-$ neurons (inset). (**z**) Scheme showing the TH group. TH neurons are identified solely by expression of *Th*. Scale bars, 50 μm; scale bars are all of the same size and apply to all micrographs.

However, connectivity and signal integration from the same channel or receptor with regard to thresholds and signal quality may be different in each unique neuronal class. Our results thus support the notion that sensation can arise both through the selective activation of a dedicated neuronal type or as encoded by the integration of activity in ensembles of different neuronal classes.

Hierarchical clustering shows that a large number of genes display distinct patterns of expression across neuronal types, and the



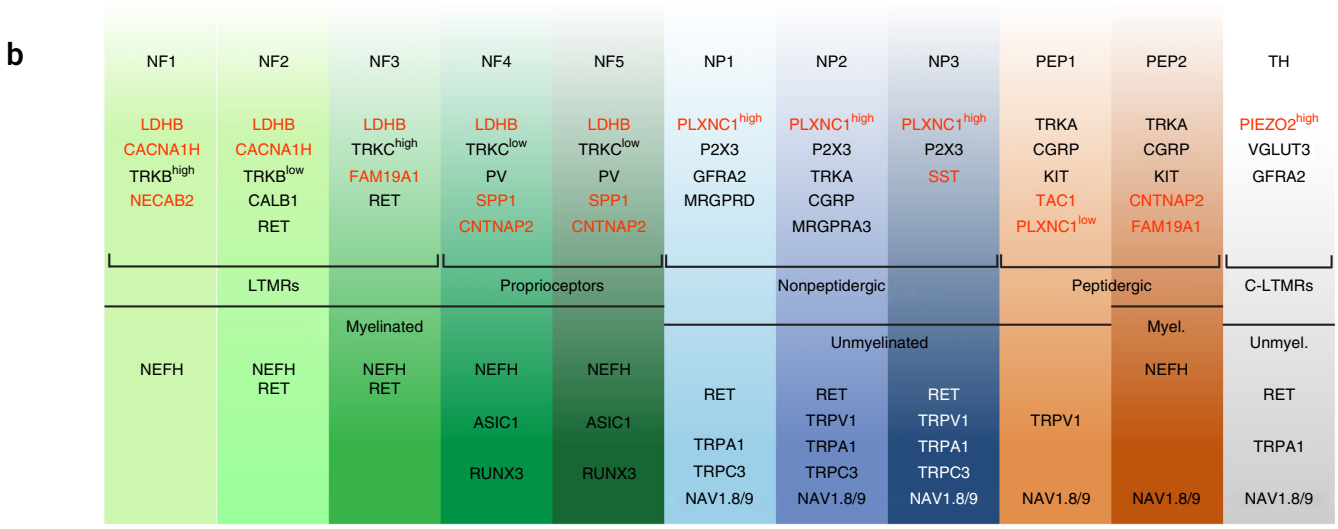


Figure 4 Unique patterns of expression of operational components in different sensory types. (**a**) Heat map of expression (fraction of positive cells) of the various operational components of sensory neurons in different classes of neurons. All neurons contain a peripheral terminal that is responsive to sensory stimuli, an axon that conducts action potentials from the periphery to the CNS and a central terminal where information is transferred to the CNS. Neuronal identity is reflected by the unique combination of activation properties encoded by the specificity of transducer ion channels at peripheral terminals and presence of voltage-gated ion channels eventually resulting in release of glutamate, neuropeptides and other signaling molecules in the spinal cord. Furthermore, the central terminal is presynaptically subject to both excitatory and inhibitory influences, and during inflammatory or neuropathic pain a number of gene products are known to participate in sensitization. Some selected genes are named below the heat map. See **Supplementary Table 6** for data presented in this panel. Trp, transient receptor potential cation channels; P2x/yr, purinergic receptor P2X or P2Y; Accn, amiloride-sensitive cation channels; K_V-rectif., open rectifier potassium channels (Kcnk), neurofilam., neurofilaments; Glu transp., glutamate transporters; chemok., chemokines; Opioid R., opioid receptors. (**b**) Proposed new sensory neuron classification based on unbiased full RNA transcriptome analyses. Gene products on top are suggested markers for identification of subtypes (red, new markers; black, selected previously used markers). Gene products at bottom are examples of a distribution of gene products commonly studied in the field. Proposed relation of sensory subtypes to modality-specific functions or previous functional classification is indicated below brackets. Myel., myelinated; Unmyel., unmyelinated.

Table 1 Expression profile (fraction of positive cells by thresholding method) of itch receptors and neuropeptides in small size neurons

	Population:	NP1	NP2	NP3	PEP1	PEP2	TH
	N:	125	32	12	64	17	233
	1 cell represents:	0.008	0.03	0.08	0.016	0.06	0.004
Symbol	Ratio to *Actb*						
Itch receptors							
Mrgprd	0.06	0.84	0.22	0	0.02	0	0.01
Trpa1	0.04	0.51	0.22	0.17	0.06	0	0.18
Lpar3	0.07	0.66	0.03	0	0	0	0.01
Lpar5	0.05	0.24	0.03	0.08	0	0	0.01
Mrgpra3	0.11	0.01	0.63	0.08	0	0	0
Mrgprx1	0.05	0.07	0.53	0.33	0	0	0
Cysltr2	0.12	0.03	0	0.67	0	0	0
Il31ra	0.06	0	0.03	0.58	0.02	0	0
Osmr	0.09	0.01	0.38	0.75	0.03	0	0.01
Trpv1	0.09	0.03	0.28	0.58	0.31	0	0
Hrh1	0.02	0	0.09	0.08	0	0	0
Htr2a	0.02	0.07	0.03	0.25	0.02	0.24	0
Htr1f	0.02	0	0.09	0.83	0	0	0
Htr3a	0.31	0	0.12	0	0.05	0.82	0
Neuropeptides							
Nmb	0.21	0.94	0.75	0.67	0.39	1.00	0.16
Calca	1.97	0.20	0.88	0.08	0.78	1.00	0.01
Nppb	0.30	0	0.03	0.83	0.03	0	0
Sst	0.39	0	0.03	0.83	0.02	0	0
Agrp	0.03	0.01	0.13	0.25	0.11	0	0
Nts	0.35	0.01	0.03	0.75	0.02	0	0
Tac1	11.94	0.03	0.09	0.17	0.83	0	0.04
Adcyap1	0.19	0.01	0.09	0	0.63	0	0.01

The population size and the fraction of the population that would correspond to one cell are shown at top. See **Supplementary Table 7** for full lists of itch-related and neuropeptide genes.

combination of patterned gene expression underlies cell type–selective functional properties. However, for each identified class, we also find sets of genes expressed exclusively within that class. Notably, nearly all those gene products are involved in ionic activity or modulation, cell adhesion, receptor signaling or signaling molecule release. Such genes likely participate in establishment of connectivity, induction of end organs, induction and regulation of excitation, and maintenance and modulation of cellular properties. Our data have extensive utility for explaining functional mechanisms in the somatosensory system. Despite the complexity of somatic sensation, our data predict that many acute stimuli with unknown cellular mediators are tuned to activate a dedicated neuronal subtype and thereby possibly distinct neuronal networks. For instance, our data predict that pro-nociception by the immunomodulatory lipid mediator sphingosine-1-phosphate via the receptor $S1P_1$ (*S1pr1*)[34–36] seems to occur almost exclusively by recruitment of NP3 neurons, whereas $S1P_3$ (*S1pr3*) mediates response in PEP1 and PEP2 neurons; the involvement of nerve injury–induced lysophosphatidic acid in the initiation and maintenance of neuropathic pain and cholestatic itch occurs via lysophosphatidic acid receptors 3 and 5 (*Lpar3* and *Lpar5*)[37,38], exclusively expressed in NP1 class of neurons; the pro-nociceptive effects of angiotensin II during neuropathic pain[39] is mediated via the angiotensin II type 1 receptor (*Agtr1a*) in NP1 neurons; inflammatory pain produced by the pro-nociceptive cytokine prokineticin 2 through prokineticin receptor 2 (*Prokr2*) is selective for PEP2 neurons; and neurotensin[40] could potentially mediate nociception directly through activation of the high-affinity neurotensin receptor (*Ntsr1*) present in some TH class neurons. Furthermore, VGLUT3 (coded by *Slc17a8*), involved in pleasant touch, is exclusive to the TH class of neurons.

Our data also allow us to predict multiple itch pathways (**Table 1**). Lysophosphatidic acid receptors (encoded by *Lpar3* and *Lpar5*), which can initiate itch triggered by their ligands, were unique to NP1 neurons. Chloroquine mediates its effects via interaction with MRGPRA3 receptors[26] (functionally coupled to the ionotropic channel TRPA1)[41], which is exclusively expressed in the NP2 class of neurons. Another chloroquine receptor, MRGPRX1 (also known as MRGPRC11), is also in the same class. Serotonin-induced itch can be assigned to NP3 owing to the specific presence of the serotonin receptors HTR1F and HTR2A. Some itch pathways, however, remain elusive because, in both NP2 and NP3, histamine receptor H1 (HRH1) was detected only at low levels (and only in NP2 was this statistically significant), and PAR2-dependent itch could not be assigned, probably owing to unreliable detection. Histamine-induced itch requires both PLCβ3 (phospholipase C, beta 3) and TRPV1 while serotonin requires PLCβ3 but not TRPV1 (refs. 42,43). We observed *Plcb3* in NP1, NP2, NP3 and to lesser degree in PEP1 neurons, while *Trpv1* is expressed in NP2, NP3 and PEP1, possibly delimiting histamine itch to classes expressing both these genes (NP2, NP3) and serotonin itch to the PLCβ3-containing class (NP3).

The prediction that IL-31 and cysteine leukotrienes can induce itch was experimentally addressed. As both molecules have been directly linked to inflammatory itch diseases, including atopic dermatitis[44,45], our results suggest that itch in such conditions could be tuned to the activation of the specific and relatively small population of neurons represented by the NP3 cluster. Neurons of this cluster selectively expressed somatostatin, natriuretic peptide and neurotensin. Expression of *Nppb* may represent transduction in the next stage of the neuronal pathway, as its receptor *Npra* is expressed in the spinal cord dorsal horn and blocking NPPB activity prevents itch induced by a variety of pruritogenic agents without affecting nociceptive behavior[46]. This itch pathway from primary NPPB-expressing sensory neurons involves secondary neurons in the dorsal horn of the spinal cord, from

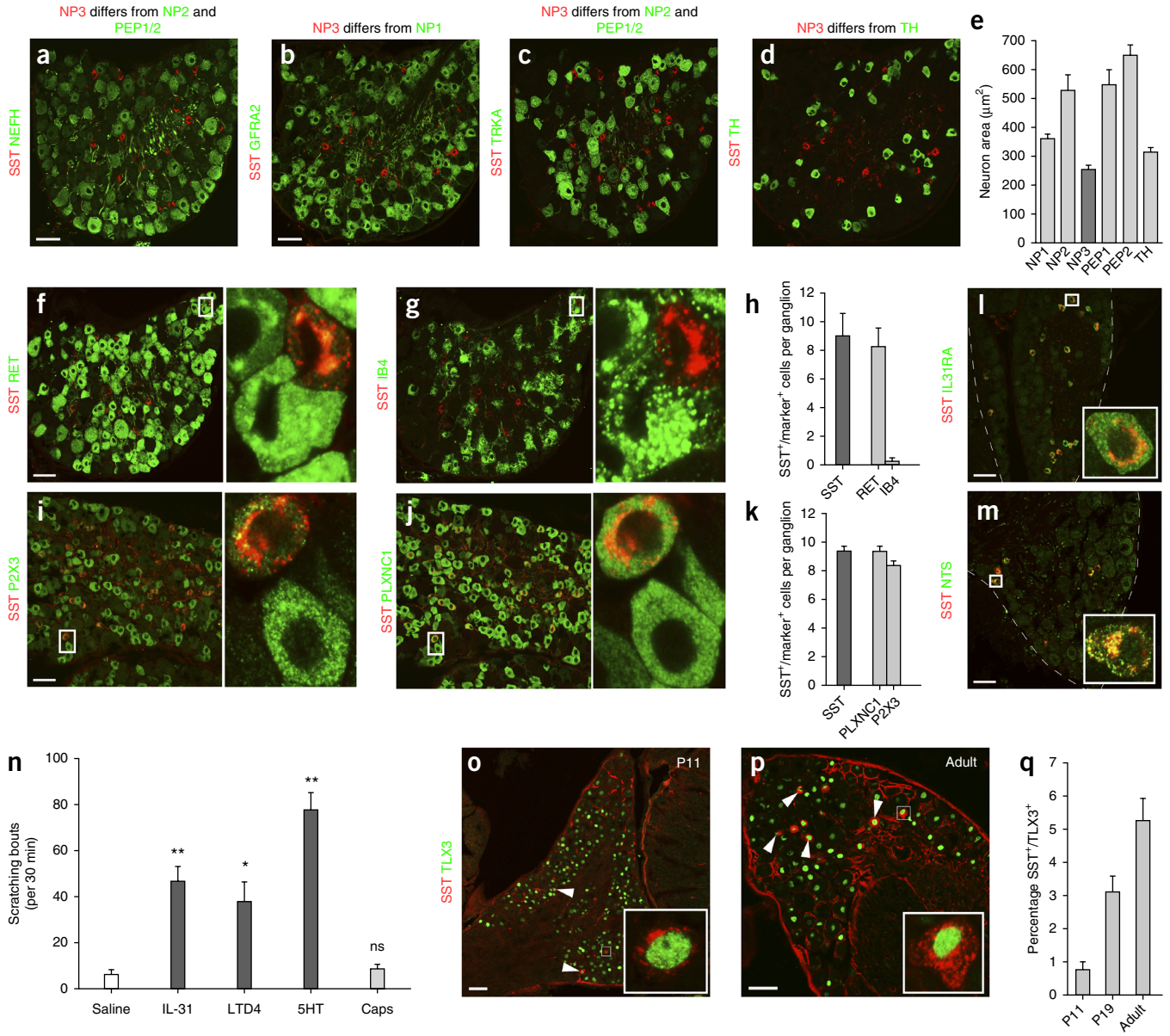


Figure 5 Diversity of itch responsive neuronal subtypes. (**a**–**d**) Double immunohistochemistry for SST and NEFH (**a**), GFRA2 (**b**), TRKA (**c**) or TH (**d**), representative markers for NF1–NF5, NP1, PEP1-PEP2-NP2 and TH neuronal classes, respectively, on adult lumbar DRG sections. (**e**) Quantification of the soma area for the different subclasses (mean ± s.e.m.). (**f**–**m**) Double immunohistochemistry for SST and RET (**f**) or IB4 (**g**), or P2X3 (**i**) or PLXNC1 (**j**) and for SST and IL31RA (**l**) or NTS (**m**). Quantification of SST⁺ neurons coexpressing RET, IB4, P2X3 or PLXNC1 is reported in **h**,**k** (mean ± s.e.m.). (**n**) Behavioral response (scratching bouts) following subcutaneous administration of compounds: saline, IL-31, LTD4, the serotonin analog 2-methyl-5-hydroxytryptamine (5HT) and capsaicin (Caps). Two-sided Mann-Whitney test. Significance compared to saline, $*P < 0.05$, $**P < 0.01$; ns, not significant; values: IL-31, $P = 0.0065$; LTD4, $P = 0.0105$; 5HT, $P = 0.008$; Caps, $P = 0.47$. Mean ± s.e.m. (**o**–**q**) Double immunohistochemistry for SST, to label neurons of the NP3 population, and TLX3 (T cell leukemia, homeobox 3), to quantify all neurons on wild-type postnatal day (P) 11, P19 and adult lumbar DRG. Arrowheads indicate SST⁺ neurons. Quantification in **q**, mean ± s.e.m. Immunohistochemistry in **a**,**c**, **b**,**d**, **f**,**g** and **i**,**j** was performed on the same sections. Scale bars are shared between these panel pairs. Specific signal emitted in far-red light from NEFH (**a**), TH (**d**), IB4 (**g**) and PLXNC1 (**j**) is shown as green. Scale bars, 50 µm.

which gastrin-releasing peptide (GRP) is released to exert the NPPB itch-inducing activity[46]. In addition to NPPB as a primary itch neurotransmitter, GRP has been proposed as a transducer of itch from primary sensory neurons[47–49]. In accord with a recent study[50], *Grp* was not found to be expressed by primary sensory neurons in the present study. Instead, in addition to *Nppb*, expression of *Nts* and *Sst* was specific to the NP3 population, suggesting their participation in the functionality of these neurons as modulators or neurotransmitters in the periphery and/or centrally in the dorsal horn. Thus, our data support the existence of at least three classes of itch responsive neurons with unique response profiles: lysophosphatidic acid associated with cholestatic disorders would be tuned to NP1 neurons, chloroquine and histamine associated with acute itch would be tuned to NP2 neurons, and mediators such as IL-31 and cysteine leukotrienes, which are linked to chronic states of inflammatory itch, as well as histamine and serotonin, would engage NP3 neurons. As exemplified above, these results provide a comprehensive map of the cellular basis of somatic sensation, and the underlying

single-cell transcriptome data open up direct strategies for delineating the functional roles of the various neuronal types. More broadly, we demonstrate a strategy for unbiased cell type discovery and classification that should be applicable to other tissues and even organisms.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Gene Expression Omnibus: raw data for individual plates (libraries), GSE59739. Additional external resources are available at http://linnarssonlab.org/drg/ and include normalized data for all sequenced wells (expressed as reads per million, RPM), with full sample annotation; the expression profiles for all genes and neuronal populations (fraction of positive cells, with color coding), allowing visualization of scatter plots of expression (RPM) for any gene; visualizing tool for expression level posteriors; and online search engine for scatter plots of expression for any gene.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

The authors thank the CLICK Imaging Facility, supported by the Wallenberg Foundation. This work was supported by the Swedish Research Council for Medicine and Health, the Swedish Foundation for Strategic Research and Linné grants (DBRM grants), the Swedish Brain Foundation, Hållsten Foundation, Torsten Söderberg Foundation, Wallenberg Scholar and European Research Council advanced grant (232675) to P.E.; and by European Research Council starting grant (261063) to S.L.

AUTHOR CONTRIBUTIONS

D.U., S.L. and P.E. designed the study. D.U., A.F., D.L., O.K., H.A., J.H.-L., J.H. and S.I. carried out experiments. D.U., P.V.K., P.L., P.E. and S.L. performed data analysis, including statistical analyses. D.U., S.L. and P.E. wrote the manuscript in consultation with all authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Ma, Q. Population coding of somatic sensations. *Neurosci. Bull.* **28**, 91–99 (2012).
2. Ma, Q. Labeled lines meet and talk: population coding of somatic sensations. *J. Clin. Invest.* **120**, 3773–3778 (2010).
3. Shapiro, E., Biezuner, T. & Linnarsson, S. Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* **14**, 618–630 (2013).
4. Wichterle, H., Gifford, D. & Mazzoni, E. Neuroscience. Mapping neuronal diversity one cell at a time. *Science* **341**, 726–727 (2013).
5. Islam, S. *et al.* Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res.* **21**, 1160–1167 (2011).
6. Jaitin, D.A. *et al.* Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776–779 (2014).
7. Treutlein, B. *et al.* Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* **509**, 371–375 (2014).
8. Lallemend, F. & Ernfors, P. Molecular interactions underlying the specification of sensory neurons. *Trends Neurosci.* **35**, 373–381 (2012).
9. Li, L. *et al.* The functional organization of cutaneous low-threshold mechanosensory neurons. *Cell* **147**, 1615–1627 (2011).
10. Liu, Y. & Ma, Q. Generation of somatic sensory neuron diversity and implications on sensory coding. *Curr. Opin. Neurobiol.* **21**, 52–60 (2011).
11. Andres, K.H. *Z. Zellforsch. Mikrosk. Anat.* [Research on the fine-structure of spinal ganglia] **55**, 1–48 (1961).
12. Lieberman, A.R. Sensory ganglia. in *The Peripheral Nerve* (ed. Landon, D.N.) 188–278 (Chapman and Hall, London, 1976).
13. McMahon, S.B., Armanini, M.P., Ling, L.H. & Phillips, H.S. Expression and coexpression of Trk receptors in subpopulations of adult primary sensory neurons projecting to identified peripheral targets. *Neuron* **12**, 1161–1171 (1994).
14. Kharchenko, P.V., Silberstein, L. & Scadden, D.T. Bayesian approach to single-cell differential expression analysis. *Nat. Methods* **11**, 740–742 (2014).
15. Woolf, C.J. & Ma, Q. Nociceptors–noxious stimulus detectors. *Neuron* **55**, 353–364 (2007).
16. Abdo, H. *et al.* Dependence on the transcription factor Shox2 for specification of sensory neurons conveying discriminative touch. *Eur. J. Neurosci.* **34**, 1529–1541 (2011).
17. Bourane, S. *et al.* Low-threshold mechanoreceptor subtypes selectively express MafA and are specified by Ret signaling. *Neuron* **64**, 857–870 (2009).
18. Gascon, E. *et al.* Hepatocyte growth factor-Met signaling is required for Runx1 extinction and peptidergic differentiation in primary nociceptive neurons. *J. Neurosci.* **30**, 12414–12423 (2010).
19. Luo, W., Enomoto, H., Rice, F.L., Milbrandt, J. & Ginty, D.D. Molecular identification of rapidly adapting mechanoreceptors and their developmental dependence on ret signaling. *Neuron* **64**, 841–856 (2009).
20. Wende, H. *et al.* The transcription factor c-Maf controls touch receptor development and function. *Science* **335**, 1373–1376 (2012).
21. Ernfors, P., Lee, K.F., Kucera, J. & Jaenisch, R. Lack of neurotrophin-3 leads to deficiencies in the peripheral nervous system and loss of limb proprioceptive afferents. *Cell* **77**, 503–512 (1994).
22. Chen, C.L. *et al.* Runx1 determines nociceptive sensory neuron phenotype and is required for thermal and neuropathic pain. *Neuron* **49**, 365–377 (2006).
23. Luo, W. *et al.* A hierarchical NGF signaling cascade controls Ret-dependent and Ret-independent events during development of nonpeptidergic DRG neurons. *Neuron* **54**, 739–754 (2007).
24. Snider, W.D. & McMahon, S.B. Tackling pain at the source: new ideas about nociceptors. *Neuron* **20**, 629–632 (1998).
25. Han, L. *et al.* A subpopulation of nociceptors specifically linked to itch. *Nat. Neurosci.* **16**, 174–182 (2013).
26. Liu, Q. *et al.* Sensory neuron-specific GPCR Mrgprs are itch receptors mediating chloroquine-induced pruritus. *Cell* **139**, 1353–1365 (2009).
27. Caterina, M.J., Rosen, T.A., Tominaga, M., Brake, A.J. & Julius, D. A capsaicin-receptor homologue with a high threshold for noxious heat. *Nature* **398**, 436–441 (1999).
28. McKemy, D.D., Neuhausser, W.M. & Julius, D. Identification of a cold receptor reveals a general role for TRP channels in thermosensation. *Nature* **416**, 52–58 (2002).
29. Tominaga, M. *et al.* The cloned capsaicin receptor integrates multiple pain-producing stimuli. *Neuron* **21**, 531–543 (1998).
30. Coste, B. *et al.* Piezo1 and Piezo2 are essential components of distinct mechanically activated cation channels. *Science* **330**, 55–60 (2010).
31. Seal, R.P. *et al.* Injury-induced mechanical hypersensitivity requires C-low threshold mechanoreceptors. *Nature* **462**, 651–655 (2009).
32. Peier, A.M. *et al.* A TRP channel that senses cold stimuli and menthol. *Cell* **108**, 705–715 (2002).
33. Vriens, J. *et al.* TRPM3 is a nociceptor channel involved in the detection of noxious heat. *Neuron* **70**, 482–494 (2011).
34. Chi, X.X. & Nicol, G.D. The sphingosine 1-phosphate receptor, $S1PR_1$, plays a prominent but not exclusive role in enhancing the excitability of sensory neurons. *J. Neurophysiol.* **104**, 2741–2748 (2010).
35. Doyle, T., Finley, A., Chen, Z. & Salvemini, D. Role for peroxynitrite in sphingosine-1-phosphate-induced hyperalgesia in rats. *Pain* **152**, 643–648 (2011).
36. Welch, S.P., Sim-Selley, L.J. & Selley, D.E. Sphingosine-1-phosphate receptors as emerging targets for treatment of pain. *Biochem. Pharmacol.* **84**, 1551–1562 (2012).
37. Kremer, A.E. *et al.* Lysophosphatidic acid is a potential mediator of cholestatic pruritus. *Gastroenterology* **139**, 1008–1018 (2010).
38. Ueda, H., Matsunaga, H., Olaposi, O.I. & Nagai, J. Lysophosphatidic acid: chemical signature of neuropathic pain. *Biochim. Biophys. Acta* **1831**, 61–73 (2013).
39. Smith, M.T., Woodruff, T.M., Wyse, B.D., Muralidharan, A. & Walther, T. A small molecule angiotensin II type 2 receptor (AT_2R) antagonist produces analgesia in a rat model of neuropathic pain by inhibition of p38 mitogen-activated protein kinase (MAPK) and p44/p42 MAPK activation in the dorsal root ganglia. *Pain Med.* **14**, 1557–1568 (2013).
40. Dobner, P.R. Neurotensin and pain modulation. *Peptides* **27**, 2405–2414 (2006).
41. Wilson, S.R. *et al.* TRPA1 is required for histamine-independent, Mas-related G protein-coupled receptor-mediated itch. *Nat. Neurosci.* **14**, 595–602 (2011).
42. Han, S.K., Mancino, V. & Simon, M.I. Phospholipase Cβ 3 mediates the scratching response activated by the histamine H1 receptor on C-fiber nociceptive neurons. *Neuron* **52**, 691–703 (2006).
43. Imamachi, N. *et al.* TRPV1-expressing primary afferents generate behavioral responses to pruritogens via multiple mechanisms. *Proc. Natl. Acad. Sci. USA* **106**, 11330–11335 (2009).
44. Angelova-Fischer, I. & Tsankov, N. Successful treatment of severe atopic dermatitis with cysteinyl leukotriene receptor antagonist montelukast. *Acta Dermatovenerol. Alp. Pannonica Adriat.* **14**, 115–119 (2005).
45. Sonkoly, E. *et al.* IL-31: a new link between T cells and pruritus in atopic skin inflammation. *J. Allergy Clin. Immunol.* **117**, 411–417 (2006).
46. Mishra, S.K. & Hoon, M.A. The cells and circuitry for itch responses in mice. *Science* **340**, 968–971 (2013).
47. Sun, Y.G. *et al.* Cellular basis of itch sensation. *Science* **325**, 1531–1534 (2009).
48. Sun, Y.G. & Chen, Z.F. A gastrin-releasing peptide receptor mediates the itch sensation in the spinal cord. *Nature* **448**, 700–703 (2007).
49. Liu, X.Y. *et al.* B-type natriuretic peptide is neither itch-specific nor functions upstream of the GRP-GRPR signaling pathway. *Mol. Pain* **10**, 4 (2014).
50. Goswami, S.C. *et al.* Itch-associated peptides: RNA-Seq and bioinformatic analysis of natriuretic precursor peptide B and gastrin releasing peptide in dorsal root and trigeminal ganglia, and the spinal cord. *Mol. Pain* **10**, 44 (2014).

ONLINE METHODS

Animals. C57Bl/6 mice were used for all experiments. Animals were kept in cages in groups, with food and water *ad libitum,* under 12-h light-dark cycle conditions. All animal work was performed in accordance with the national guidelines and local ethics committees of Stockholm.

Cell picking. Six hindlimb-innervating lumbar DRGs (L4–L6) from adult mice (6–8 week old, both males and females) were dissected and dissociated according to a previously published protocol[51] followed by Percoll gradient centrifugation[52]. L4–L6 neurons are largely limb innervating, and we note that composition and neuronal types could vary along the rostro-caudal axis. During the preparation procedure the sample was always kept on ice, apart from incubation in papain and collagenase/dispase, and in some preparations a cooled microscope stage was used. The final cell pellet was suspended in 300 µl of L15 medium supplemented with B27. Cells were picked with a robotic cell-picking setup, constructed in-house around an inverted Nikon microscope. Individual cells were sucked into a glass capillary attached to a CellTram syringe with visual control so that individual cells were captured. By an automated *x-y-z* axis drive, the capillary tip was positioned in a well of a 96-well plate containing 3 µl of lysis buffer, into which the cell was released. In total, 18 48-well plates were used (864 wells total) from 18 different animals in independent experiments. Picking of one plate containing 48 cells took from 40 to 60 min. Dissociated cells were photographed for cell size measurements before collection. Cells were picked randomly.

Single-cell RNA-seq. Cells were processed for RNA sequencing and sequenced on an Illumina Genome Analyzer IIx or HiSeq 2000, and reads were mapped to the mouse genome as previously described[5,53]. 16 out of 18 picked 48-well plates were pooled to generate libraries that included up to 96 cells for sequencing (L128–L228). Thus, each of these 8 libraries contained 96 sequenced wells. The last two libraries, L281 and L282, were processed separately and contained 48 wells per library. Out of a total of 864 processed wells, 799 contained a single cell as verified by visual inspection under microscope. The remaining wells were empty or contained more than one cell. The raw data for individual plates (libraries) are accessible at Gene Expression Omnibus (GEO) under accession code GSE59739. The expression level in RPM (reads per million) was taken as the number of reads mapped to all isoforms of a gene divided by all reads mapped to any annotated gene, using the RefSeq gene models mapped on the GRCh37/hg19 genome build. Our RNA-seq method reads only the 5′ end of transcripts[3], and hence there is no need to normalize for length, as in the commonly used RPKM (reads per kilobase per million) measure. These normalized data are provided at http://linnarssonlab.org/drg/ (together with annotation of individual wells, including both content and cell population assignment).

Principal component analysis (PCA). PCA of normalized data was performed using Qlucore Omics Explorer 2.3. PCA was performed using mean = 0, variance = 1 normalization. 2,858 genes contributing to differences between picking sessions ($q \leq 0.05$, ANOVA test) were excluded, and in total 22,478 genes were used for further analysis. PCA plot density pattern with one distant and four closely positioned groups was robust and observed with a wide range of settings for filtering by variance. **Figure 1a** (**Supplementary Video 1**) shows PCA based on 12,750 genes that passed filtering by variance at the level of 3.85×10^{-4}. The most distal cluster identified (yellow in **Fig. 1a**) represented a mixture of non-neuronal cells, while the remaining four clusters were neurons (see gene ontology analysis, **Supplementary Table 3**). Excluding the non-neuronal group, PCA led to a confirmation of the four clear distinct neuronal groups. Separation was robust and worked over a wide range of variance filters, the only filter applied. For the analysis in **Figure 1b** (**Supplementary Video 2**), the variance filter was set at 2.55×10^{-4}, which resulted in 11,658 genes contributing to the analysis.

Starting with identification of four neuronal groups, we assigned each of the 622 neurons to 1 of 11 neuronal types using an unbiased iterative PCA-based process. To do so, each of the initial groups was compared one by one against all other groups by applying parametric filtering of variables contributing to the difference between the two groups. In many cases this resulted in one of the clusters becoming compacted (this population is referred to as an outgroup population), while the other split into two or three clearly separated clusters. These newly identified subpopulations were annotated after verifying that (i) this splitting was stable at different levels of parametric variable filter (*q* values) and that (ii) alternative splitting was absent even when a different outgroup population is used. For subgroup splitting, representative PCA charts are shown in **Figure 2b** with appropriate PCA numbering referring to branching points in the hierarchical categorization in **Figure 2a**. Cells that could not be reliably assigned to a cluster were excluded from analysis and categorized as unclear identity.

When all cells that could be assigned to any of the lowest level clusters—that is, the eleven neuronal clusters—had been identified, the contents of the highest level groups were finalized. Hence, the four main groups were constructed only from cells that also were assigned to the final eleven clusters.

We also controlled for the hypothesis that 'orthogonal' sets of neuronal populations may exist—that is, that there may be 'superimposed' categories consisting of cells different from the main four categories. An example of this would be if cells in all four categories split between dorsal and ventral clusters, sharing common genes through all four groups. To test this, we performed PCA excluding the genes that contributed to the separation of the four clusters. This analysis did not reveal any noticeable density patterns, suggesting that the clusters are organized hierarchically, whereby each lower level cluster is a part of an upper level cluster and does not result from orthogonal superimposition.

Data handling and statistics. All assignment of cells to populations was performed on the basis of PCA. Afterwards, to identify genes categorizing individual neuronal classes, we used the recently published Bayesian approach to single-cell differential expression (SCDE) analysis method[14] to compare each cell class against all other clusters. Importantly, all genes, even those that were differentially expressed between picking sessions and that were excluded during neuronal type identification, were included in this analysis. Briefly, cell-specific error models were derived over the entire population of neuronal genes. As cells were collected in three separate picking sessions, the following batch-correction procedure was used in comparing expression of a given gene between any two groups: (i) fold expression difference posterior for a given gene was first estimated using standard procedure without any additional corrections; (ii) a separate fold expression difference posterior accounting only for potential batch effects was generated by randomly sampling sets of cells corresponding in size to the groups being compared, maintaining the original batch composition (that is, number of cells from different picking session) within each group; (iii) the final batch-adjusted fold expression difference posterior was calculated as a posterior distribution of the difference between the standard and batch-effect posteriors (that is, probability of observing a particular fold expression difference relative to batch-only posterior). This batch correction procedure is implemented in the version 1.1 of the SCDE package (http://pklab.med.harvard.edu/scde/index.html).

For behavioral experiment, no statistical methods were used to predetermine sample sizes, but our sample sizes are similar to those generally employed in the field. For analysis of behavioral experiment, a two-sided Mann-Whitney test was used, which requires no assumption apart from independence of measurements.

Gene ontology analysis. Lists of genes differentially expressed in each of the eleven groups (when compared to the other neuronal populations) were ranked by upregulation fold change (maximum likelihood estimate), and not more than 300 top genes for which the *P* value was 0.1 and higher were used for gene ontology (GO) analysis. Cytoscape 3.0.1 (ref. 54) with ClueGO[55] app was used. The murine GO (Biological Processes, version from 4 July 2013) was used with the following settings: type of analysis: single; GO terms level: 3–8; GO term restriction: 2 genes and 4%; evidence code: all. A significance threshold level of 0.05 was applied (**Supplementary Table 2**). After this, GO term groups reflecting similar sensory properties or functions were generated and custom terms were created generalizing such groups (**Supplementary Table 2**). Moreover, since we did not find any GO term representing itch among ontology terms, we used published data[56] to generate a list of 40 genes (**Supplementary Table 2**, last tab) expected to be expressed in sensory neurons and associated with itch. Using a calculation approach implemented in ClueGO, the custom 'itch' term (associated with this gene list) was introduced into the GO term statistics outcome table, generated by ClueGO (**Supplementary Table 2**). The final ClueGO results (**Fig. 2e**) were obtained by plotting the *P* value of the first (most significant) GO term representing each generalized term. To select gene lists for neuronal versus non-neuronal cells, we used SCDE (**Supplementary Table 3**), ranked genes first by fold change (maximum likelihood estimate) and then by *P* value, and selected the 300 top genes, the *P* values of which did not exceed 0.1. Then these two corresponding gene lists were analyzed by the ClueGO app in

doi:10.1038/nn.3881

the same way as described above for neuronal subpopulations. The resulting lists of population-specific GO terms were ranked according to *P* value, with a cutoff at $P < 0.05$. No further term selection was performed.

Thresholding method to determine fraction of positive cells. Single-cell RNA-seq data are noisy and carry a high rate of false negatives, which means that summary statistics such as mean or even median can be highly misleading. To summarize the expression of genes within groups of cells (for example, 'cell types'), we calculated the fraction of expressing cells using the following thresholding procedure. The 'maximum' expression level for each gene was determined to be the average of the three cells with the highest expression. 5% of this 'maximum' level was designated as the threshold of expression. The fraction of cells expressing a given gene above this threshold was then determined (**Supplementary Fig. 3**). A list of all genes with an expression profile for neuronal populations calculated in this way, as well as a macro for performing such analysis, is presented at http://linnarssonlab.org/drg/. The same file also allows normalized data for each gene (RPM) to be easily visualized as a scatter plot of cells grouped by population.

Immunohistochemistry. Males 6–8 weeks old were perfused using 4% paraformaldehyde (PFA) in PBS (pH 7.4). DRGs were dissected out and postfixed in 4% PFA for 1 h at 4 °C. Samples were subsequently washed in PBS at 4 °C for 1 h and cryoprotected by incubating at 4 °C overnight in 20% sucrose followed by 30% sucrose in PBS. Tissue was then embedded in OCT and frozen at −20 °C. Samples were sectioned at 14 µm. The sections were used and processed as previously described[57]. At least two animals were used for each immunostaining, showing similar results.

The primary antibodies used were rabbit anti-NECAB2 (Atlas Antibodies, 1:1,000, HPA013998), goat anti-SPP1 (R&D Systems, 1:500, AF808), rabbit anti-LDHB (Abcam, 1:400, ab75167), rabbit anti-FAM19A1 (Atlas Antibodies, 1:100, HPA013407), rabbit anti-CNTNAP2 (Alomone Labs, 1:100, APZ-005), sheep anti-PV (R&D Systems, 1:500, AF5058), goat anti-CALB1 (R&D Systems, 1:1,000, AF3320), goat anti-TRKA (R&D Systems, 1:200, AF1056), goat anti-TRKB (R&D Systems, 1:500, AF1494), goat anti-TRKC (R&D Systems, 1:500, AF1404), sheep anti-PLXNC1 (R&D Systems, 1:200, AF5375), rabbit anti-P2X3 (Abcam, 1:1,000, ab10269), rat anti-SST (Millipore, 1:100, MAB354), goat anti-CGRP (Abcam, 1:1,000, ab36001), guinea pig anti-CGRP (Peninsula Laboratories, 1:1,000, T-5027), rabbit anti-TAC1 (Millipore, 1:1,000, AB1566), rabbit anti-TH (Pel-Freez, 1:1,000, P40101-0), sheep anti-TH (Novus Biologicals, 1:2,000, NB300-10), chicken anti-NF200 (Abcam, 1:10,000, ab4680), goat anti-GFRA2 (R&D Systems, 1:500, AF429), goat anti-RET (1:100, AF482), goat anti-IL31RA (R&D Systems, 1:500, AF2769), guinea pig anti-TLX3 (kind gift from T. Müller and C. Birchmeier, Max-Delbrück-Centrum for Molecular Medicine, Berlin, Germany, 1:10,000), Mouse anti-ISL1 (Developmental Studies Hybridoma Bank (DSHB), 1:100, 39.4D5-b) and rabbit anti-neurotensin (Abcam, 1:100, ab43833). For isolectin B4 staining, slides were incubated overnight in IB4-biotin (Invitrogen, 1:500) and subsequently incubated for 1–2 h in FITC-streptavidin (Jackson Laboratories 1:500). For detection, donkey 488-, 555- and 647-conjugated Alexa secondary antibodies against rabbit (Invitrogen A21206, A31572 and A31573, respectively), goat (Invitrogen A11055, A21432 and A21447, respectively), sheep (Invitrogen A11015, A21436 and A21448, respectively) and mouse (Invitrogen A21202, A31570 and A31571, respectively), donkey 488-, 549- and 647-conjugated Alexa secondary antibodies against chicken (Jackson ImmunoResearch 703-545-155, 703-505-155 and 703-605-155, respectively), donkey Cy3-conjugated secondary antibody against rat (Jackson ImmunoResearch 712-166-150) and donkey Cy3- and Dylight 647-conjugated secondary antibody against guinea pig (706-165-148 and 706-495-148, respectively) were used at a dilution of 1:1,000. Slices were subsequently mounted with glycerol mounting media (Merck).

Immunohistochemical staining was performed on the following numbers of animals and ganglia: **Figure 3** and **Supplementary Figure 2**, number of animals = 2, number of ganglia analyzed per condition = 4 or 5. **Figure 5a–d**, number of SST^+ neurons counted (co-stained for NEFH and TRKA) = 47; $SST^+/NEFH^+ = 0$; $SST^+/TRKA^+ = 0$; SST^+ (co-stained for GFRA2 and TH) = 44; $SST^+/GFRA2^+ = 0$; $SST^+/TH^+ = 0$. Number of animals = 2; number of ganglia analyzed per condition SST/NEFH/TRKA = 4; SST/GFRA2/TH = 5. **Figure 5f–k**: number of SST^+ neurons counted (co-stained for RET and IB4) = 36; $SST^+/RET^+ = 33$; $SST^+/IB4^+ = 1$; number of neurons SST^+ counted (co-stained for P2X3 and PLXNC1) = 28; $SST^+/P2X3^+ = 25$; $SST^+/PLXNC1^+ = 28$. Number of animals = 2; number of ganglia analyzed per condition SST/RET/IB4 = 5; SST/P2X3/PLXNC1 = 4. **Figure 5o–q**: number of ganglia analyzed per age: P11 = 9; P19 = 15; adult = 8 from one animal each.

Quantification of cell size. Soma area of neurons belonging to different subclasses were calculated on immunostained sections using ImageJ. For every neuronal subgroup, neurons expressing a distinctive, unique molecular marker for that subgroup were analyzed. Neurons belonging to the NP1 population were identified by double immunohistochemistry for PLXNC1 and combined antibodies against NEFH/TRKA/SST/TH. Neurons positive for PLXNC1 but negative for combined NEFH/TRKA/ SST/TH staining were selected for analysis (**Supplementary Fig. 2v**; neurons analyzed $n = 30$). Neurons belonging to the NP2 population were identified by double immunohistochemistry for TRKA and combined antibodies against TAC1 and FAM19A1. Neurons positive for TRKA but negative for combined TAC1/FAM19A1 staining were selected for analysis. Alternatively, triple immunohistochemistry for PLXNC1, CGRP and TAC1 was performed and $PLXNC1^+/CGRP^+/SP^-$ neurons were used for analysis. Note that CGRP is a common marker for NP2, PEP1 and PEP2, but only NP2 and PEP1 are $PLXNC1^+$ and only PEP1 is SP^+ (**Fig. 3** and **Supplementary Fig. 2**; neurons analyzed $n = 10$). Neurons belonging to the NP3 population were identified by immunohistochemistry for SST (**Figs. 3** and **5** and **Supplementary Fig. 2**; neurons analyzed $n = 27$). Neurons belonging to the PEP1 population were identified by double immunohistochemistry for TRKA and TAC1. Double-positive neurons were selected for analysis (**Fig. 3** and **Supplementary Fig. 2**; neurons analyzed $n = 15$). Neurons belonging to the PEP2 population were identified by double immunohistochemistry for TRKA and FAM19A1. Double-positive neurons were selected for analysis (**Fig. 3** and **Supplementary Fig. 2**; neurons analyzed $n = 32$). Neurons belonging to the TH population were identified by immunohistochemistry for TH (**Fig. 3** and **Supplementary Fig. 2**; neurons analyzed $n = 30$).

Behavioral analysis. Adult males (2–4 months old) were used. Experiments were performed under the light part of the day cycle. Animals were not subjected to any other prior experimentation (surgery, drug administration, behavioral). No animals were excluded from the analysis. Two days before the experimental day, the fur of the back was shaved and all mice were acclimated 24 h before and just before the experiment by placing them in transparent glass chambers for 30 min. After the habituation period, a 50-µl solution of each compound or vehicle was administered subcutaneously into the nape of the neck of the mouse. Behavior was then recorded for 30 min and scratching bouts were counted. The experimenter was not blinded to the identity of the compound used in each injection. Each scratching bout was defined as repeated scratching of the injecting site with the hindpaw followed by licking or biting the hindpaw and/or placing it on the floor. For Me-5HT, the number of scratching bouts induced by the vehicle was deduced from the total number of scratching bouts counted. The following drugs were used: IL-31 (Peprotech, 60 pmol in 0.9% NaCl), leukotriene D4 (LTD4, Cayman Chemical, 30 pmol in 0.9% NaCl), 2-methyl-5-hydroxytryptamine (Me-5HT, Sigma, 100 pmol in 0.9% NaCl, 1% ascorbic acid) and capsaicin (Sigma, 100 pmol in 0.9% NaCl, 7% Tween-80). Number of times experiments were repeated (where *n* is the number of animals in each group from *N* different litters): saline: $n = 6$, $N = 3$; IL-31: $n = 6$, $N = 4$; LTD4: $n = 5$, $N = 4$; Me-5HT: $n = 5$, $N = 3$; capsaicin: $n = 5$, $N = 3$.

A **Supplementary Methods** Checklist is available.

51. Usoskin, D. *et al.* En masse in vitro functional profiling of the axonal mechanosensitivity of sensory neurons. *Proc. Natl. Acad. Sci. USA* **107**, 16336–16341 (2010).
52. Hjerling-Leffler, J., Alqatari, M., Ernfors, P. & Koltzenburg, M. Emergence of functional sensory subtypes as defined by transient receptor potential channel expression. *J. Neurosci.* **27**, 2435–2443 (2007).
53. Islam, S. *et al.* Highly multiplexed and strand-specific single-cell RNA 5′ end sequencing. *Nat. Protoc.* **7**, 813–828 (2012).
54. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
55. Bindea, G. *et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091–1093 (2009).
56. Liu, T. & Ji, R.R. New insights into the mechanisms of itch: are pain and itch controlled by distinct mechanisms? *Pflugers Arch.* **465**, 1671–1685 (2013).
57. Furlan, A., Lubke, M., Adameyko, I., Lallemend, F. & Ernfors, P. The transcription factor Hmx1 and growth factor receptor activities control sympathetic neurons diversification. *EMBO J.* **32**, 1613–1625 (2013).