

# Single cell transcriptomics of primate sensory neurons identifies cell types associated with chronic pain

Jussi Kupari <sup>1,7</sup>, Dmitry Usoskin<sup>1,7</sup>, Marc Parisien <sup>2</sup>, Daohua Lou<sup>1</sup>, Yizhou Hu <sup>1</sup>, Michael Fatt<sup>1</sup>, Peter Lönnerberg<sup>1</sup>, Mats Spångberg<sup>3</sup>, Bengt Eriksson<sup>3</sup>, Nikolaos Barkas <sup>4</sup>, Peter V. Kharchenko<sup>4</sup>, Karin Loré <sup>5,6</sup>, Samar Khoury<sup>2</sup>, Luda Diatchenko <sup>2</sup>✉ & Patrik Ernfors <sup>1</sup>✉

Distinct types of dorsal root ganglion sensory neurons may have unique contributions to chronic pain. Identification of primate sensory neuron types is critical for understanding the cellular origin and heritability of chronic pain. However, molecular insights into the primate sensory neurons are missing. Here we classify non-human primate dorsal root ganglion sensory neurons based on their transcriptome and map human pain heritability to neuronal types. First, we identified cell correlates between two major datasets for mouse sensory neuron types. Machine learning exposes an overall cross-species conservation of somatosensory neurons between primate and mouse, although with differences at individual gene level, highlighting the importance of primate data for clinical translation. We map genomic loci associated with chronic pain in human onto primate sensory neuron types to identify the cellular origin of chronic pain. Genome-wide associations for chronic pain converge on two different neuronal types distributed between pain disorders that display different genetic susceptibilities, suggesting both unique and shared mechanisms between different pain conditions.

<sup>1</sup>Division of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden. <sup>2</sup>Alan Edwards Centre for Research on Pain, Department of Anesthesia, School of Medicine, School of Dentistry, McGill University, Montréal, QC, Canada. <sup>3</sup>Astrid Fagraeus Laboratory, Comparative Medicine, Karolinska Institutet, Stockholm, Sweden. <sup>4</sup>Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA. <sup>5</sup>Division of Immunology and Allergy, Karolinska Institutet, Stockholm, Sweden. <sup>6</sup>Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden. <sup>7</sup>These authors contributed equally: Jussi Kupari, Dmitry Usoskin. ✉email: [luda.diatchenko@mcgill.ca](mailto:luda.diatchenko@mcgill.ca); [patrik.ernfors@ki.se](mailto:patrik.ernfors@ki.se)

The dorsal root ganglion (DRG) consists of a variety of neuron types, each tuned to detect and transduce different physical stimuli. These neuron types can broadly be divided into low-threshold mechanosensitive neurons responsible for sensing touch and high-threshold nociceptors, which are involved in pain, temperature, and itch<sup>1–4</sup>. However, a comprehensive classification of DRG neurons is critical for understanding exactly how somatosensation works and for providing insights into the cellular basis for acute and chronic pain. Rodents represent the main species for studies on the cellular and molecular basis of nociception and the greatest insights with respect to molecular classification of neuronal types have been obtained from mouse, where single-cell RNA-sequencing (scRNA-seq) has led to a molecular taxonomy of existing types of sensory neurons<sup>5–9</sup>.

This has enabled the identification of molecular types representing richly myelinated A-fiber low-threshold mechanoreceptors (LTMRs) and limb proprioceptors. The remaining neuronal types in the scRNA-seq are assigned as weakly myelinated or unmyelinated neurons. One of these is a C-fiber LTMR (C-LTMR) neuron type that expresses *Vglut3* (*Slc17a8*) and tyrosine hydroxylase (*Th*) that likely is not involved in pain sensation<sup>1–5</sup>. Nociception is largely conferred through unmyelinated peptidergic C-fiber neuron types and a few lightly myelinated A $\delta$ -nociceptors, a *Trpm8* expressing cluster of neurons, as well as cell types marked by expression of *Mrgprd*, *Mrgpra3*, or *Sst* (named NP1, NP2, and NP3 types of neurons, respectively<sup>8</sup>). This molecular classification agrees remarkably well with previous studies based on myelination and conduction velocity, neurochemical features and termination patterns peripherally in the skin and centrally in the spinal cord and is also consistent with the known ontogeny of DRG neuron types<sup>5</sup>. As a result, there have been significant advances in understanding the cellular and molecular characteristics of sensory neurons found in mouse DRG.

Much less is known about characteristics of human DRG. Apart from information on size of the ganglia along the rostro-caudal axis, micro-anatomy including neuron size<sup>10–12</sup> and electrophysiological characteristics<sup>13–16</sup>, the molecular characterization of human DRG is still limited to bulk RNA-sequencing<sup>17–19</sup> and neurochemical analyses of gene products in a handful of studies<sup>20</sup>. Hence, the concordance of markers used in different studies and their relation to actual neuron types remain largely unknown. Nevertheless, by examining individual gene products, these studies suggest important species differences between human and mouse where, for example, *Nav1.8*, *Nav1.9*, *P2X3* receptor, and *TRPV1* are present in both small and large neurons in humans, but only small neurons in mouse, suggesting fundamental differences in molecular characteristics and principles of initiation and transduction of somatosensory stimuli between humans and rodent<sup>20</sup>.

In humans, rare and drastic mutations that explain different types of congenital insensitivity to pain and erythromelalgia have been identified, such as, for example, *SCN9A* (*Nav1.7*), *NTRK1* (*TRKA*), and *SCN11A* (*Nav1.9*)<sup>21–25</sup>. In addition to these rare causing mutations, it is known that the genetic risk for chronic pain is due to common variations with small effect size<sup>26</sup>. Close to half of the risk of developing chronic pain are attributable to genetic factors<sup>27–29</sup>, including musculoskeletal pain conditions<sup>28</sup>. For musculoskeletal pain there is statistical evidence for a diverse set of genes involved, with a marked overrepresentation of genes expressed in neurons and functionally associated with neurotransmission, indicating a strong heritable component caused by altered functions of neurons<sup>26</sup>. Pleiotropy of single-nucleotide polymorphisms (SNPs) among painful and non-painful conditions has also been shown<sup>30</sup>, even in human DRG<sup>31</sup>. It has recently become possible to connect genomic results to transcriptomics at the cellular level which allows for insights into the

cell types which are fundamental for disorders. Thus, taking advantage of scRNA-seq for mapping susceptibility genes to cell types, new insights have been made into the cell types involved, for example, in schizophrenia<sup>32,33</sup>, neuroticism<sup>34</sup>, intelligence<sup>35,36</sup>, and Alzheimer's disease<sup>37</sup>, but such analyses have not been attempted for chronic pain conditions.

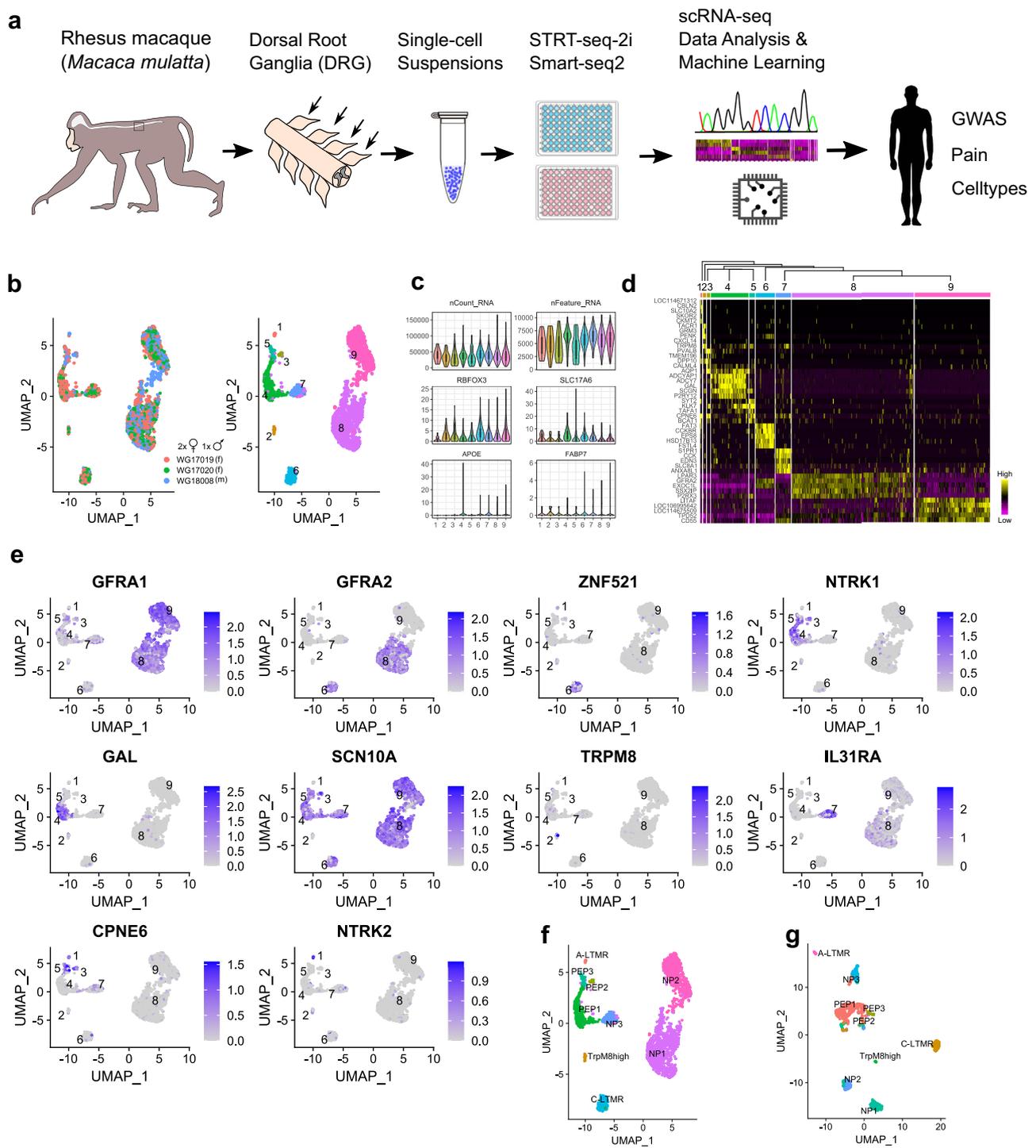
Knowledge on the molecular and cellular characteristics of primate DRG and their mouse correlates has critical implications for translating data from rodent models to human pain disorders<sup>38,39</sup> and allows mapping genomic loci implicated in chronic pain onto specific primate somatosensory neuron types. In this work, we explore the cellular basis of somatosensation in the non-human primate and identify sensory cell types linked to human chronic pain. Our results reveal an overall conserved cellular strategy for somatosensation between primate and mouse and identify that heritability for musculoskeletal pain converge on two discreet DRG sensory neuron types.

## Results

**Molecular diversity of sensory neuron types in a non-human primate.** We prepared DRG cell suspensions for scRNA-seq from adult Rhesus macaques using two different platforms (Fig. 1a). First, cells from three macaques (two females and one male) were captured and sequenced using STRT-2i-seq. A total of 4742 cells were sequenced and the reads were aligned to the macaque genome *Mmu10* with gene names and annotations transferred from human (see “Methods”). The data were merged and then clustered using the anchoring-based integration and graph-based clustering approach implemented in Seurat<sup>40</sup>. Using iterative rounds of clustering and quality control, we identified and removed non-neuronal cells, injured neurons and ambiguous cells, and finally merged clusters with highly similar transcriptomic profiles (Supplementary Fig. 1a–g, see “Methods”). The remaining 2518 neurons formed nine separate clusters (Fig. 1b and Supplementary Fig. 1g). For validation, this cleaned dataset was also analyzed using Conos<sup>41</sup>, an approach that identifies multiple plausible inter-sample mappings and builds a joint graph of the datasets. This approach produced close to identical clusters to the original nine formed with Seurat (Supplementary Fig. 1h, i; see “Methods”). Gene expression patterns between the different animals showed near perfect positive correlation indicating high similarity of inter-individual transcriptome profiles (Supplementary Fig. 1j). The analyzed neurons contained 5687 genes and 38,624 unique transcripts per cell on average, expressed neuron and sensory neuron specific genes (*RBFOX3*, *SLC17A6*) throughout with limited expression of satellite-glia genes (*FABP7*, *APOE*), and showed unique gene expression profiles (Supplementary Fig. 1k, l and Fig. 1c, d).

After clustering, we used canonical mouse DRG neuron markers to assign tentative identities for the clusters based on their likely mouse counterparts (Fig. 1e, f): NP1 (cluster 8) and NP2 (cluster 9) were named based on the combination of *GFRA1* and *GFRA2* expression; C-LTMRs (6) were assigned using *GFRA2* and *ZNF521* (*Zpf521* in the mouse); *NTRK1* and *GAL* were used to identify PEP1 (4); *SCN10A* and *TRPM8* suggested the identity of the *Trpm8*<sup>high</sup> (2) cluster (negative for *SCN10A*); *IL31RA* expression was used for naming NP3 (7); *CPNE6* together with *NTRK2* was used to assign putative A-LTMRs (1), and *CPNE6* together with *NTRK1* and *SCN10A* were used to assign the PEP2 (3) cluster. The final cluster (5) also expressed *CPNE6*, *NTRK1*, and *SCN10A*, and was named PEP3.

A second dataset was prepared from five female macaques using Smart-seq2 technology. After clustering and cleaning steps (Supplementary Fig. 2a–d), these data included 1038 neurons showing >480,000 counts and >13,500 detected genes per cell on average



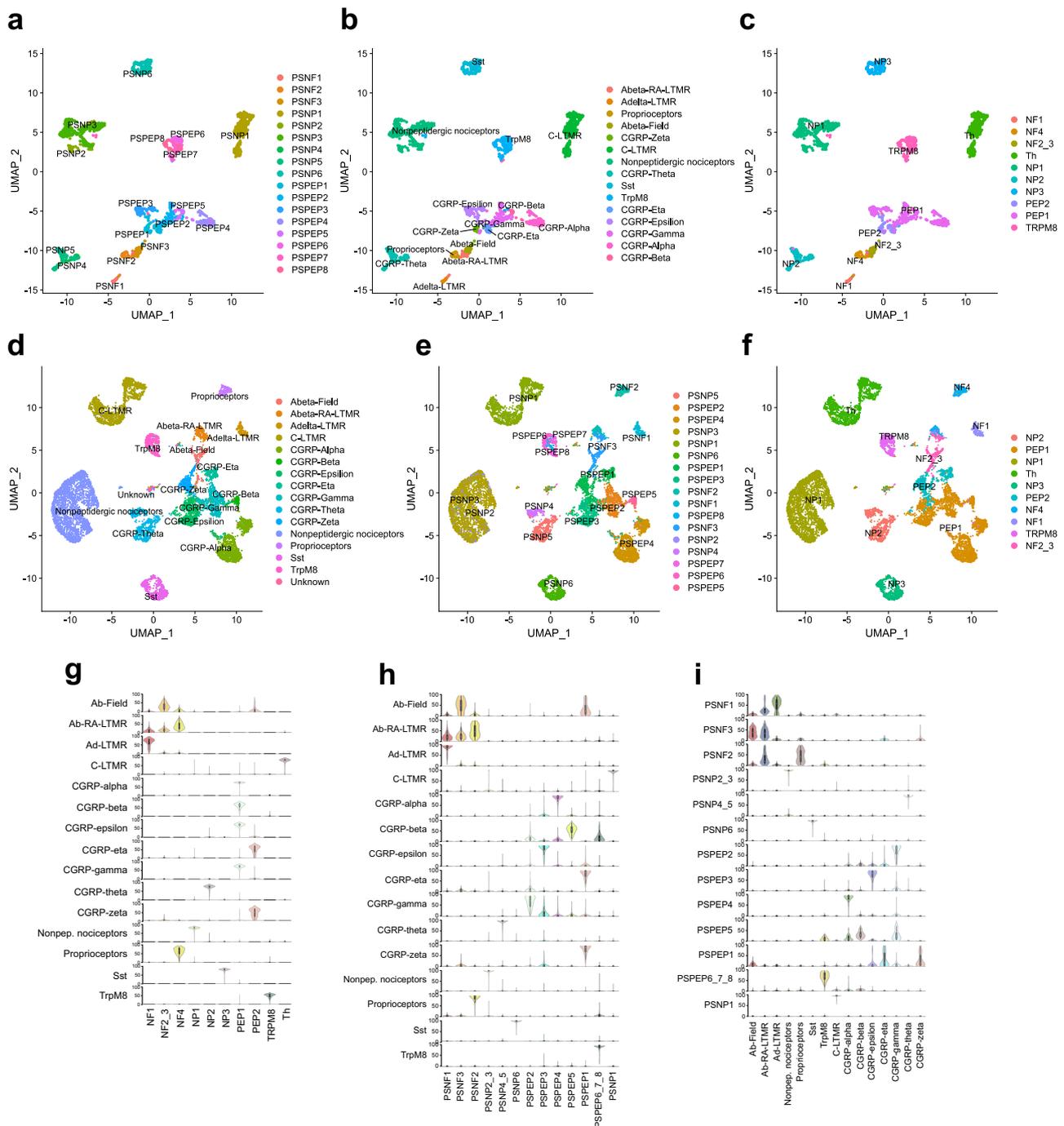
**Fig. 1** Somatosensory neuron clusters in the macaque DRG. **a** A schematic view of the workflow. **b** UMAP plots showing (left) the contribution of individual animals ( $N = 3$ ) to STRT-2i-seq clusters and (right) final cluster numbers (m = male, f = female). **c** Violin plots showing total counts of unique transcripts, detected genes, and transcript counts for neuronal and satellite-glia marker genes in the neuronal clusters. Y-axes show detected genes per cell for nFeature\_RNA; all others are raw UMI counts. Boxplot defines the median, interquartile range (IQR), and  $1.5 \times$  IQR (whiskers). **d** A hierarchically organized heatmap with the five most specific genes (by *p*-adj) for each cluster. **e** UMAPs showing mouse canonical marker gene expression in the STRT-2i-seq macaque clusters. **f** STRT-2i-seq macaque clusters named after most likely mouse counterparts. **g** Named STRT-2i-seq clusters after label transfer from the STRT-2i-seq data. Image sources: freevectors.net (human silhouette), needpix.com (microwell plates), openclipart.org (Eppendorf tube).

(Supplementary Fig. 2e, f). Tentative neuron identities for this dataset were assigned in a supervised manner using the STRT-2i-seq data as a reference (Fig. 1f, g and Supplementary Fig. 2g). Interrogation of marker gene expression used for tentative cell-type assignment showed identical patterns between the two macaque

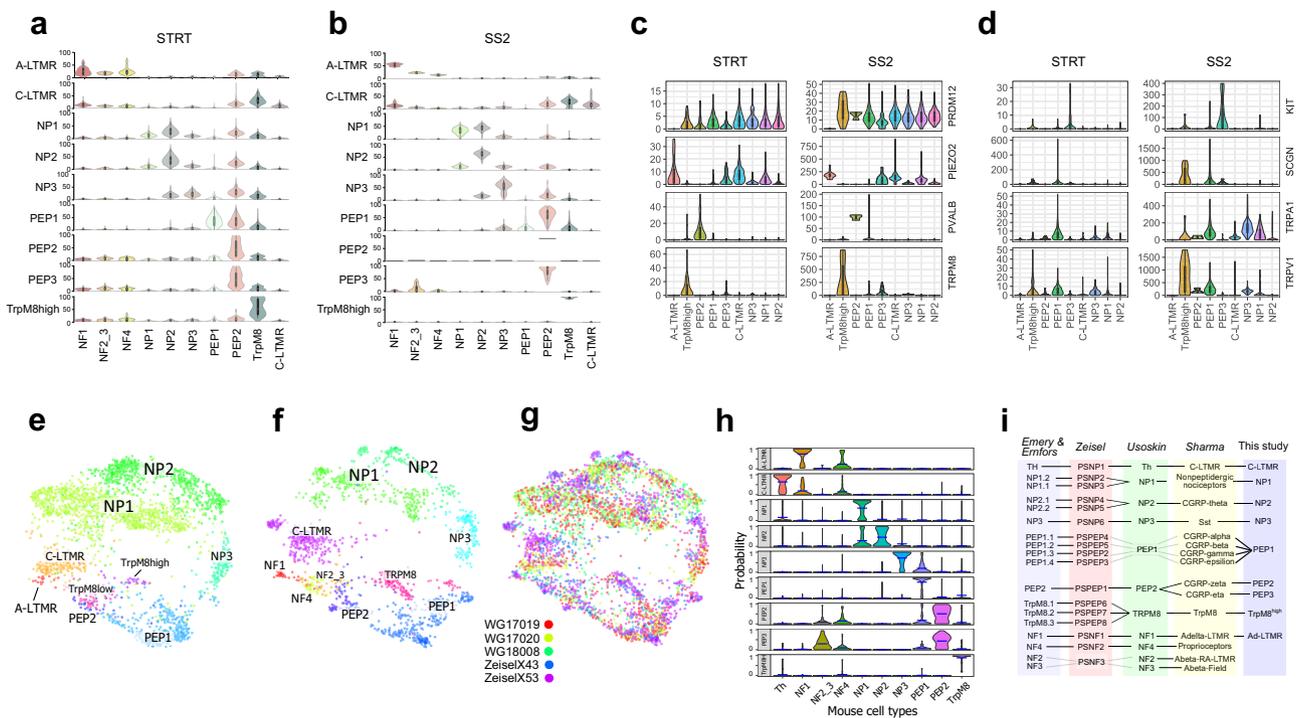
datasets (Supplementary Fig. 2h); thus, representing an independent identification of the cell types identified in the STRT-2i-seq data. For full marker lists from both scRNA-seq datasets, see Supplementary Data 1 and 2. An interactive web resource for browsing the datasets is available at <https://ernforsgroup.shinyapps.io/macaqueDRG/>.

**A consensus on mouse DRG neuron types across datasets.** We wanted to further leverage the vast knowledgebase of mouse DRG neuron types in our investigation of primate DRG neurons. For this, we used two major scRNA-seq datasets from previously published mouse studies, referred hereafter as the Zeisel and Sharma datasets<sup>7,9</sup>. These studies identify similar number of cell types, but it is not known if the same kinds of neurons were identified and furthermore, the studies use different

nomenclature. We therefore first identified the corresponding cell types between the datasets. Using the label transfer method implemented in Seurat, we transferred labels from Sharma over to the Zeisel data (Fig. 2a, b and Supplementary Fig. 3a) and then also named the Zeisel types using Usoskin nomenclature<sup>8</sup> (Fig. 2c). We then repeated the label transfer from Zeisel to Sharma data and named the Sharma types also using Usoskin nomenclature (Fig. 2d–f and Supplementary Fig. 3b). Finally, we



**Fig. 2 Consensus of mouse DRG neuron types across datasets and nomenclatures.** **a–f** UMAPs showing **a** Zeisel types with Zeisel nomenclature, **b** Zeisel data after label transfer from Sharma data, **c** Zeisel data with Usoskin nomenclature, **d** Sharma data with original nomenclature, **e** Sharma data after label transfer from Zeisel data, and **f** Sharma data with Usoskin nomenclature. **g** Probability scores of Sharma types against Zeisel trained module (performed with Usoskin nomenclature). **h** Probability scores of Sharma types against Zeisel trained module (performed with Zeisel nomenclature). **i** Probability scores of Zeisel types against Sharma trained module (performed with original nomenclature on Sharma). Boxplots in **g–i** define the median, interquartile range (IQR), and  $1.5 \times$  IQR (whiskers).



**Fig. 3 Identification of primate correlates of mouse neuron types.** **a, b** Violin plots showing prediction percent probability scores for each macaque cluster against established mouse DRG cell types trained with Usoskin nomenclature. **a** STRT-2i-seq ( $n = 3$ ), **b** Smart-seq2 ( $n = 5$ ). **c** Violin plots of markers expressed in macaque clusters. Y-axis, raw UMI counts for STRT-2i-seq (STRT) and counts for Smart-seq2 (SS2). **d** Violin plots showing genes differentially expressed between macaque PEP2 and PEP3 clusters. Boxplots in **a-d** define the median, interquartile range (IQR), and  $1.5 \times$  IQR (whiskers). **e-g** LargeViz plot showing cross-species clustering of macaque STRT-2i-seq and mouse Zeisel et al. datasets on a shared plot. **e** Macaque, **f** mouse, **g** merged plot (legend shows origin of samples, WG = macaque, Zeisel = mouse). **h** Probability plot of label propagation from mouse neuron types to macaque neuron types. Horizontal blue bars represent mean position for each distribution. **i** Synopsis of the corresponding DRG neuron types between mouse and macaque datasets and nomenclatures.

repeated the label transfer from Zeisel to Sharma both ways using Usoskin labels for the Zeisel data, and observed identical results (Supplementary Fig. 3c, d).

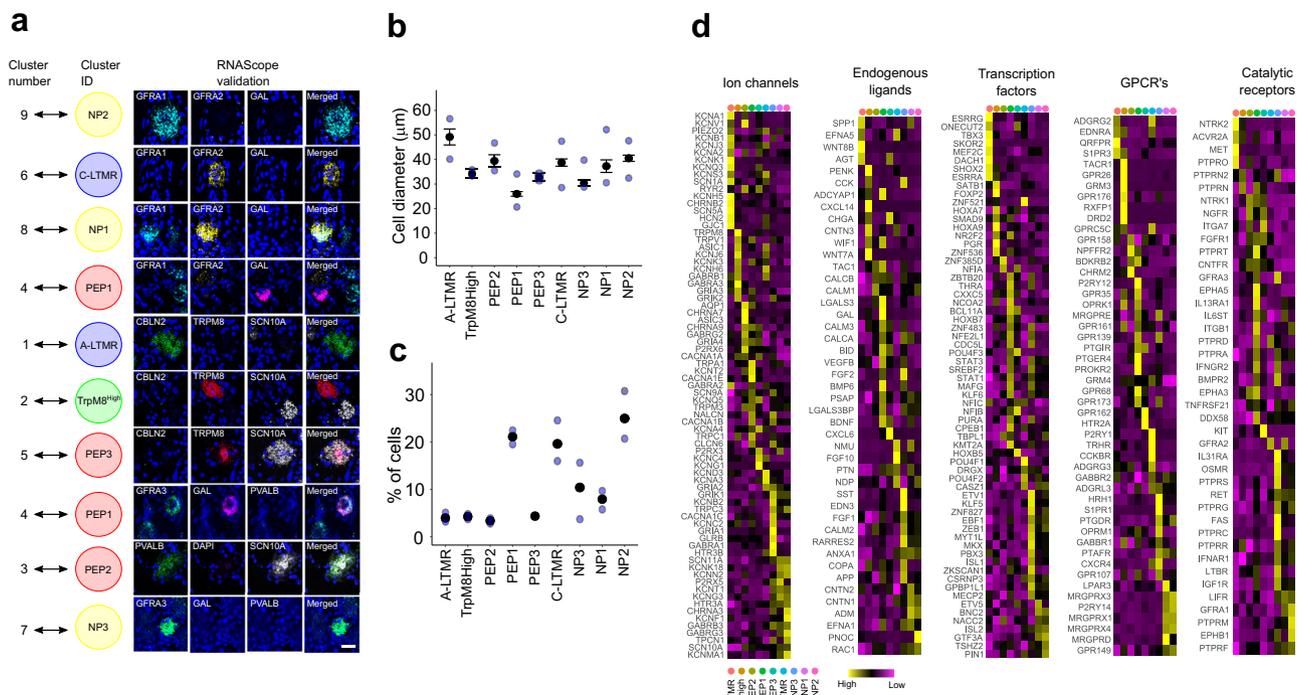
As an independent method to establish the similarity between cell types identified in the different studies, we employed neural-network-based probabilistic scoring modules for learning cell-type features between datasets. We trained the modules using both mouse datasets and using all three nomenclature versions (Zeisel and Usoskin nomenclatures for Zeisel data; Sharma nomenclature for Sharma data) (Supplementary Fig. 4a–i). We then tested the performance of these modules in finding the corresponding cell types between the datasets and found that the modules detected with high probability cell types across the different datasets and that the cell-type assignments concurred with the results obtained by the label transfer. For the rest of our analyses we used the Usoskin type nomenclature as the default naming. Taken together, the results confirmed a one-to-one relationship between nociceptor types identified in refs. 7,9 although the A $\delta$ -nociceptors of Zeisel were split into two subtypes in Sharma (Fig. 2a–i). Because Usoskin nomenclature performed with the least noise and greatest prediction scores, we used this nomenclature for the rest of our analyses.

**Overall cross-species conserved strategy for somatosensation.**

We proceeded to use the probabilistic neural-network machine-learning approach to evaluate whether the neuronal basis of somatosensation is conserved between macaque and mouse, and to validate our tentative assignment of cross-species neuron correlations. For this purpose, we generated the probability score

for each macaque cluster to the mouse neuron types using both the macaque STRT-2i-seq and Smart-seq2 datasets. Each of the macaque clusters showed similarity to the previously assigned mouse neuron types (Fig. 3a, b and Supplementary Fig. 5). These results also indicated that the macaque A-LTMR cluster consists of cells corresponding to the lightly myelinated A $\delta$ -LTMR type. Our original annotation of corresponding macaque-mouse neuron types was consistent with the expression of *PRDM12* in all nociceptors and the mechanosensory channel *PIEZO2* in A-LTMRs, C-LTMRs, and NP1 (Fig. 3c). Interestingly, the mouse proprioceptor marker *PVALB* was expressed in the macaque PEP2 neurons. PEP3, the other macaque neuron type showing similarity to mouse PEP2 differed from PEP2 neurons by expression of *TRPM8*, *PIEZO2*, *KIT*, and *SCGN*, but no or low levels of the heat-sensitive channels *TRPV1* and *TRPA1* (Fig. 3d). When compared to the mouse PEP2 subtypes in the Sharma data, macaque PEP3 showed higher probability to CGRP-eta over CGRP-zeta and PEP2 to CGRP-zeta over CGRP-eta (Supplementary Fig. 5a, b), indicating two types of A $\delta$ -nociceptors.

To gain further confidence in our cross-species analyses we performed co-integration of our macaque STRT-2i-seq data with the Zeisel mouse data using Conos. Here, the previously assigned macaque clusters showed close positioning to homologous mouse clusters on a joined cross-species clustering graph (Fig. 3e–g). This strong cross-species association was also apparent on the probability profiles after label propagation from mouse clusters to individual macaque neurons. All macaque clusters showed close to one-to-one correspondence to individual mouse clusters with PEP2 and PEP3 having the strongest association to the mouse PEP2 cluster (Fig. 3h). Combined, these results show evidence of



**Fig. 4** In vivo validation and gene family expression in the macaque DRG neuron types. **a** Naming the macaque clusters after the closest corresponding mouse cell types (left) with RNAScope validations (right). Scale bar = 20  $\mu\text{m}$ . **b** Diameter of neuronal types in lumbar DRGs, mean  $\pm$  SEM. **c** Percent of each neuronal type among neurons of all types identified in this study in lumbar DRG. In **b**, **c**, blue dots represent values for individual animals. **d** Heatmaps of gene expression profiles for selected gene families in the macaque STRT-2i-seq DRG neuron clusters. Source data for **b**, **c** are provided as a Source Data file.

a strong cross-species association of sensory neuron types, indicating that the overall cellular basis for somatosensation is conserved between mouse and macaque (Fig. 3i).

The macaque neuronal types were validated in vivo by triple in situ hybridization (Fig. 4a and Supplementary Fig. 6). *SCN10A* was used as a general marker for nociceptors together with cluster-specific markers or a combination of markers defining only one cluster. For some clusters, negative markers were used to rule out types other than the one under analysis (see “Methods”).

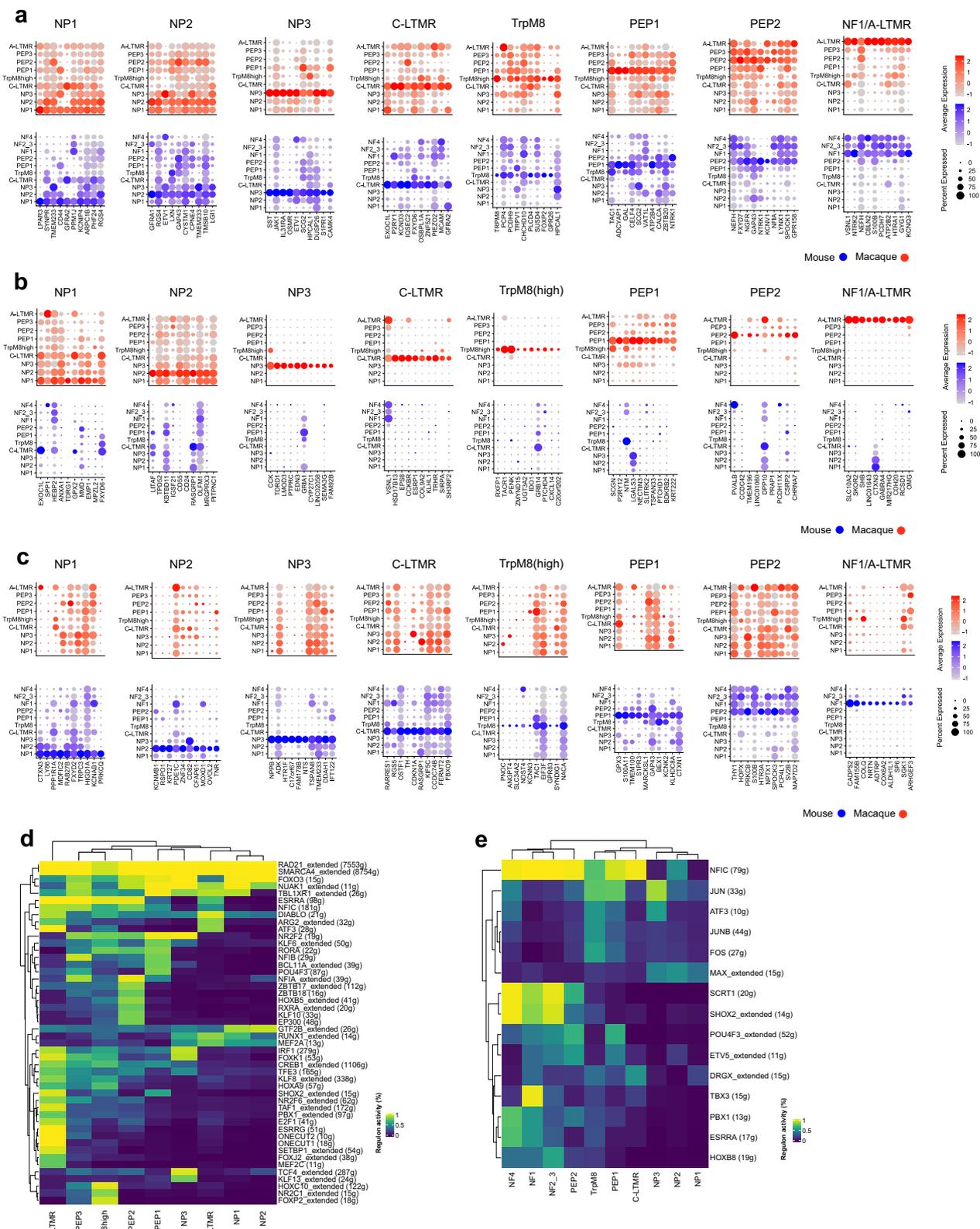
In addition, we analyzed the soma size distribution and percentage contribution of each cell type in the DRG using in situ hybridization data (Fig. 4b, c). Finally, we interrogated each of the neuron types for unique expression patterns for transcription factors, ion channels, G-protein couple receptors (GPCRs), catalytic receptors, and endogenous ligands, including neuropeptides (Fig. 4d). This revealed, for example, the expression of multiple GPCRs related to exogenous defense and cholestatic itch in NP1 and NP2 (*MRGPRX1-4*), and to histaminergic itch and inflammatory lipids in NP3 (*HRH1*, *S1PR1*) and NP1 (*LPAR3*).

**Species differences and similarities in gene expression.** Our above results confirm an overall existence of neuronal correlates between the mouse and a primate; nevertheless, important divergences could still exist when examining expression of individual genes within each of the different mouse-macaque correlates. In order to provide a more comprehensive map of the molecular conserved and divergent features of somatic sensation and pain between the mouse and the macaque, we compared gene expression patterns between the species (Fig. 5a–c and Supplementary Data 3). Because strategies to identify new molecular targets for development of analgesic drugs often are focused on genes expressed uniquely in the neuron type(s) causative of pain, we examined the presence of genes within the different neuron

types to identify conserved transcriptional programs between species as well as sets of genes that are expressed in highly species-specific manner between corresponding cell types. In such analyses, false negatives can confound the results. We therefore first examined the reliability of the individual STRT-2i-seq and Smart-seq2 datasets in side-by-side analyses of cell-type-specific expression patterns of mouse-macaque shared genes observing high reproducibility across different platforms (Supplementary Fig. 7a–c). We thereafter combined the datasets for mouse<sup>7,9</sup> and macaque (STRT-2i-seq and Smart-seq2) to obtain integrated datasets for mouse and macaque. Analysis of this dataset revealed the existence of robust conserved molecular features between mouse and non-human primate (Fig. 5a).

For example, neuronal type-specific mouse-macaque conserved features for NP3 neurons included *SST*, *JAK1*, *IL31RA*, *OSMR*, and *S1PR1* ( $n = 36$  genes) and for C-LTMRs *P2RY1*, *EXOC1L*, *KCND3*, *IQSEC2*, *OSBPL1A*, and *FXYD6* ( $n = 60$  genes). The largest cell-type-specific shared gene program was found between mouse and macaque A $\delta$ -LTMRs ( $n = 126$  genes) whereas NP1 and NP2 both had cell-type-specific conserved features of less than 30 genes ( $n = 26$  and 27, respectively). However, we also identified cell-type-specific gene expression that were species specific and these existed both in mouse and macaque (Fig. 5b, c and Supplementary Data 3). As examples, species differences included in NP3 the specific expression of *TDRD1*, *EDN3*, and *GRIA1* in macaque and *NPPB*, *HTR1F*, and *NTS* in mouse. For C-LTMRs, *TH* and *RARRES1* were specific for mouse, whereas *HSD17B13* and *CCKBR* were specific for macaque. These species-specific expression patterns need to be considered when translating results obtained in rodents to primates.

We further performed supervised computational screens to find gene families whose differential expression could reliably distinguish similar cell types within and across species<sup>42</sup>. A set of over



**Fig. 5 Comparison of gene expression profiles across species. a** Dot plots showing ten most specific shared markers of corresponding/homologous cell types between the species. **b** Dot plots showing ten most macaque specific markers for each corresponding cell-type pair. **c** Dot plots showing ten most mouse-specific markers for each corresponding cell-type pair. Dot sizes in **a–c** correspond to the percentage of cells expressing the gene in the cluster and color scale indicates log2FC. **d** Heatmap showing binary regulon activity in each of the macaque DRG cell types. **e** Heatmap showing binary regulon activity in the mouse DRG cell types. In **d, e** the number of target genes is indicated in the parenthesis.

1500 gene families from HGNC (HUGO Gene Nomenclature Committee) was used for these screens (Supplementary Data 4). Within each species the overall top performing gene families defining all sensory neurons included ion channels, G-protein coupled receptor families, cell adhesion molecules (CAMs), and others (Supplementary Fig. 8a and Supplementary Data 5 and 6). Shared mouse-macaque top performing families included voltage-gated ion channels, G-protein-coupled receptors, and neuropeptides (Supplementary Fig. 8b and Supplementary Data 7), suggesting that sensory neuron identities culminate mostly on genes of these families even across species. The highest performing gene families for correctly assigning macaque and mouse corresponding sensory neuron subtypes (Supplementary Fig. 8c and Supplementary Data 7) showed that A-LTMR/NF1 (A $\delta$ -LTMR) neurons were identified nearly perfectly by many gene families, for example, by ion channels and CAMs. For the peptidergic C-fibers (PEP1), neuropeptides/receptor ligands were the highest performing family, as expected. On the other hand, NP1 and NP2 scored poorly in comparison to all other types, suggesting that these neuron types have diverged the most in their gene expression signatures between mouse and macaque. As a final comparison between the species, we used SCENIC<sup>43</sup> to identify gene regulatory networks formed from master transcription factors and their gene targets (i.e. regulons) in all DRG cell types of both macaque and mouse (Fig. 5d, e). We found, for example, the known fate determining regulons driven by *SHOX2*, *RUNX1*, and *FOXP2* in mouse<sup>9,44–46</sup> and predicted several new species unique as well as cross-species conserved regulons determining cell-type identities (see Supplementary Data 8 and 9 for genes).

**Genetic association of neuronal types contributing to human pain states.** The identification of the cellular and molecular basis for somatosensation and pain in non-human primate enables us to determine the contribution of different primate neuron types in human chronic pain states. We therefore used human genetic data to explore how each cell type in the macaque DRG connects to painful phenotypes in humans by employing genome-wide association studies (GWAS). To do so, we used a large cohort made available by the UK Biobank project<sup>47,48</sup>, where we assessed chronic pain from self-report at eight body sites (Fig. 6a). Among the nine neuron types identified in the macaque DRG, we found that common variants associated with chronic pain sites mapped to sets of genes that were specifically expressed in two neuron types in the STRT-2i-seq dataset. We found enrichment of headaches, facial, neck and shoulder, stomach, and hip chronic pains partitioned heritability in PEP1 neurons ( $P_{\text{FDR}} = 16\%$ , each), while NP2 neurons were associated most significantly with the heritability of chronic back pain and hip pain ( $P_{\text{FDR}} = 5\%$  and  $8\%$ , respectively) (Fig. 6b and Supplementary Data 10). Thus, heritability of hip pain was significantly enriched in both PEP1 and NP2 neuron types while heritability of all other pain sites was significantly enriched in only one neuron type of the macaque DRG. Since the epidemiological prevalence of chronic pain patients to report more than one body site is high, the signal attributed to, for example, subjects with hip pain may also have back pain, so it is not clear where the association signal is deriving from. To control for the co-morbidities, we performed new full GWASes for each of the pain sites (row in Supplementary Fig. 9b, c), and then removed one by one all comorbid pain sites (column in Supplementary Fig. 9b, c) in all GWASes for PEP1 and NP2 neurons. Although some statistical power is lost in this analysis due to reduced size of chronic case participants, partitioned heritability in PEP1 was confirmed in most GWASes (facial, neck/shoulder, stomach/abdomen, and hip pain), while association with

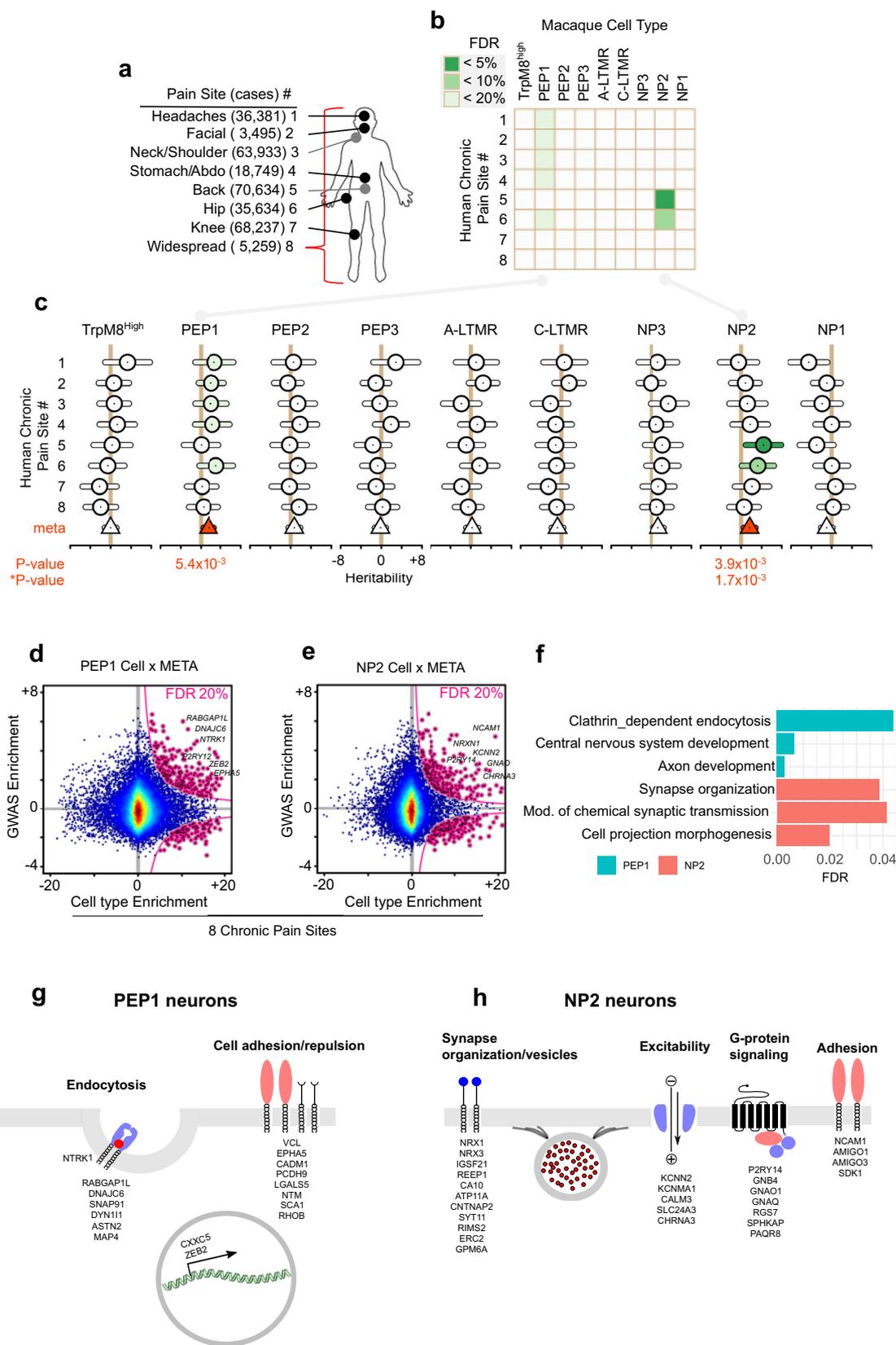
back pain remained negative (row in Supplementary Fig. 9b). However, the significance of PEP1 for headaches was lost for all GWASes. Furthermore, back pain and hip pain remained significant for NP2 neurons in all GWASes when excluding other pain sites (row in Supplementary Fig. 9c). Our results show that seven of the nine neuron types were not associated with any chronic pain sites, and hence the two neuron types, PEP1 and NP2, together represent the main enrichment of musculoskeletal pain heritability. A meta-analysis across all pain sites for each cell type (Fig. 6c) consistently found that both PEP1 ( $P = 5.4 \times 10^{-3}$ ) and NP2 ( $P = 3.9 \times 10^{-3}$ ) displayed significant enrichment across all pain sites.

We next tested if these results were reproducible in the independent scRNA-sequencing dataset obtained by the Smart-seq2 protocol. Consistent with the STRT-2i-seq dataset, we found enrichment of stomach, hip, and neck, and shoulder chronic pain partitioned heritability in PEP1 neurons ( $P_{\text{FDR}} = 10\%$ ,  $10\%$ ,  $13\%$ , respectively), while NP2 neurons were associated most significantly with the heritability of chronic back pain, hip pain, and knee pain ( $P_{\text{FDR}} = 10\%$ , each) (Supplementary Fig. 9d). Thus, heritability to pain was consistently assigned to the same neuronal types using the STRT-2i-seq and Smart-seq2 datasets, although significance of PEP1 to headaches and facial pain was lost in the latter.

In order to identify functional pathways and genes whereby human heritability contributes to chronic pain in the different neuronal types, genes were ranked by exclusivity of neuronal cell-type expression to establish enrichment scores. Human GWAS enrichment scores were mapped to the macaque single-cell expression enrichment to identify top genes in type-specific cells contributing to chronic pain sites in PEP1 and NP2 neurons, revealing hundreds of genes (Fig. 6d, e and Supplementary Data 10). These were thereafter used to identify cellular pathways that confer vulnerability to chronic pain (Fig. 6f and Supplementary Data 10). Combined, these results show that heritability of chronic pain at different sites is associated with each of the two major pain neuron types through different biological pathways (see Supplementary Data 10). Pathways contributing to chronic pain in PEP1 neurons included “clathrin-dependent endocytosis”, “central nervous system development”, and “axon development” with an enrichment of proteins involved in the process of endocytosis, cell adhesion as well as a few transcription factors (Fig. 6f, g). In contrast, pathways in NP2 neurons included “synapse organization”, “chemical synaptic transmission”, and “cell projection morphogenesis” and were dominated by genes associated with organization of the synaptic membrane and its vesicles, ion channels participating in excitability, G-protein signaling, and cell adhesion (Fig. 6f, h). Thus, the two neuron types contribute to chronic pain via distinct pathways.

## Discussion

The human DRG like the mouse contain neurons with different histochemical and electrophysiological features<sup>13–16,49,50</sup>. The identification of the molecular types of primate somatosensory neurons addresses the longstanding question whether cell types involved in somatosensation is conserved between rodents and primates. We conclude that the mouse<sup>7–9</sup> and the Rhesus macaque largely share molecular neuron types which using mouse genetics have been functionally identified as A-LTMRs involved in touch and proprioceptive sensation<sup>1</sup>, C-LTMRs involved in the affective aspect of pleasant touch<sup>5</sup>, C-cold thermoreceptors (TrpM8<sup>high</sup>), A $\delta$  fast mechanical nociceptors involved in pinprick pain (PEP2)<sup>51–54</sup> and mechano-heat C-nociceptors (PEP1), as well as “non-peptidergic” neuronal types (NP1, NP2, NP3) known in mouse to be involved sensing noxious



mechanical threshold and itch sensation<sup>55–57</sup>. Although the neural-network prediction score was low, macaque PEP3 appears to be an A $\delta$ -fiber fast nociceptor which correspond to mouse CGRP-eta type in Sharma nomenclature, which is a subtype of the mouse PEP2 type in the Usoskin nomenclature.

Overt differences in the overall cellular basis for nociception between mouse and macaque largely relates to NP1 and NP2 neurons. In the mouse, NP1 is involved in detecting pricking mechanical stimuli and  $\beta$ -alanine induced itch through the Mrgprd receptor, but not thermal sensation. In the macaque, the

**Fig. 6 Contribution of macaque DRG cell types to partitioned heritability of human chronic pain sites.** **a** UK Biobank chronic pain sites mapped to the human body. Number of chronic case participants shown in parentheses. **b** Heatmap of FDR-corrected  $p$  values for enrichment in partitioned heritability of each macaque DRG cell type ( $n = 3$  macaques) to each human chronic pain site. **c** Forest plot of partitioned heritability estimates for each macaque DRG cell type contributing to chronic pain. Shown are heritability coefficient estimates (circles) and their 95% confidence intervals (bars) for each pain site. Fixed-effect, standard error weighted, meta-analyzed heritability estimates (triangles) also shown, colored red when significant at Bonferroni-corrected level (corrected for nine cell types). **d, e** Top genes in type-specific cells contributing to specific chronic pain sites. Top genes highlighted in a scatter plot (pink, left). Scatter plots show human GWAS enrichment of gene ( $y$ -axis) as a function of macaque cluster enrichment ( $x$ -axis). **f** Bar plot of top pathways for PEP1 and NP2 cell types in the meta-analysis of the eight chronic pain sites. **g, h** Schematic illustrations of pathway-related genes in **g** PEP1 and **h** NP2 neuron types. Image sources: wikipedia.org (human silhouette).

expression of the heat-activated channel TRPV1 in NP1 neurons, in addition to TRPA1, suggests a broader function than in the mouse. The mouse NP2 neurons express histamine and chloroquine receptors and ablation of these neurons (Mrgpra3<sup>+</sup> neurons) specifically affects histamine-dependent and histamine-independent itch, but not acute noxious heat, cold, or mechanical pain<sup>56</sup>. NP2 neurons have therefore been considered to be dedicated itch neurons in rodents<sup>58</sup>. However, this neuronal type was recently found to code for both itch and pain, with itch behavior induced by metabotropic Gq-linked stimulation and pain behavior through fast ionotropic stimulation<sup>59</sup>, suggesting that the same neuronal populations can drive distinct sensations in a stimulus-dependent manner. In addition, based on the molecular profiles, macaque NP2 neurons appear partly different from mouse as histamine receptor HRH1 expression is low in macaque NP2 neurons. Other known functional stimuli detectors found in these cells are MRGPRX1–4, which are also expressed in NP1 neurons. MRGPRX1–4 are promiscuous low-affinity receptors involved in non-histaminergic itch, conveyed, for example, by chloroquine and pruritogenic peptides<sup>58</sup>. Primate NP1 and NP2 neurons may thus have at least partly different functions than in the mouse.

Even though the exact neuronal basis for human chronic pain is unknown, insights have been obtained through the identification of genes causing congenital insensitivity to pain<sup>22,60</sup>. While most of the genes causing painless phenotype are abundantly expressed in all DRG neuron types, some display restricted expression patterns, thus opening for linking neuronal types to phenotype. These include congenital insensitivity to pain by mutations in *SCN9A* (Nav1.7), *SCN11A* (Nav1.9), and *NTRK1* (TRKA)<sup>21–25,61</sup> and *PRDM12* (ref. 62). In contrast to mouse which display an enriched expression in nociceptors, macaque *SCN9A* is broadly expressed at similar levels in all neuronal types, while *SCN11A* expression is more similar to mouse with expression at varying levels in all unmyelinated neuronal types (C-LTMRs, PEP1, NP1–3,) with very low levels in TRPM8<sup>high</sup>, myelinated nociceptors, and A-LTMRs (see <https://ernforsgroup.shinyapps.io/macaqueDRG/> for interrogation of gene expression). *NTRK1* is largely confined to macaque TRPM8<sup>high</sup>, PEP1, PEP2, and PEP3 neuronal types, with lower levels in the other nociceptors. *PRDM12* expression is consistent with mouse, appearing in all macaque neurons except A $\delta$ -LTMRs. Thus, although it is not possible to pinpoint the exact neuronal types, it seems based on expression of these causative genes for human monogenic pain insensitivity disorders that PEP1, PEP2, PEP3, and TRPM8<sup>high</sup> represent important neuronal types for nociception. However, it cannot be excluded that neuronal types sufficient for driving chronic pain might partly involve neuron types other than those required for nociception.

Previous GWAS studies have uncovered genome-wide significant genes that contributes to the heritable risk of chronic pain. Recent methods<sup>32,33</sup> allow the integration of GWAS and scRNA-seq data to map cell types contributing to disease through testing the enrichment for cell-type-specific expression of genes

with nearby risk SNPs. Such analyses consider the heritability carried by all common SNPs, linking them to nearby genes, rather than focusing only on genome-wide significant genes. Using this methodology, we linked GWAS results of several human chronic pain sites to specific neuronal types in the primates. Significance for cell-type-specific contributions reported in previous studies<sup>32,63</sup> showed stronger association than that reported in this work (false discovery rate (FDR) in the range of 5–20%). However, the estimation for the heritability of pain ranges from 2 to 10%<sup>64–66</sup> with 7.6% for chronic back pain<sup>67</sup>. In a comparative study of heritability between different classes of diseases, it was shown that mental health disorders show higher heritability estimates than self-reported pain phenotypes<sup>66</sup>. Furthermore, all chronic pain GWAS in the UK Biobank displayed inflated genomic control parameters ( $\lambda_{GC} \gg 1$ ) while at the same time an LD regression score intercept close to 1, indicating that the inflation's origin is highly polygenic in nature. Because of these two effects combined (lower heritability and higher polygenicity), a tissue- or cell-type-specific contribution to chronic pain would be smaller in comparison with other diseases, like schizophrenia. On the contrary, it is perhaps remarkable that some specific cell types show significance, particularly since the burden also distributes to cell types other than sensory neurons. However, the strength of our analysis is that it considers the accumulated contribution of all small effect sizes in sensory neurons distributed across the human genome. Thus, we believe that the significance observed is in the range of expectation. Furthermore, we show that the association to the identified neuronal types emerge from multiple pain GWAS and in addition was reproduced in an independent dataset using a different sequencing platform.

The results show that musculoskeletal pain genetics is well represented in the DRG. Our findings reveal a connection to two of nine types of neurons to multiple chronic pain sites, namely PEP1 and NP2. However, there is an overlap of individuals being included in some of these pain groups and if a cell type is enriched in multiple pain GWAS, there is a risk of reporting the false association that is driven by phenotypic co-morbidity but not a true genetic association. Iterative removal of one pain site at a time in the GWAS and examining if the significance to the other pain site remain indicates that most of the GWAS signals observed is attributed to the annotated pain sites themselves. These results point towards a common underlying pain vulnerability regardless of the body site where pain is manifested. PEP1 for headache could not be confirmed in this analysis. Thus, assignment of contribution of PEP1 cell type to headaches might be due to co-morbidity of other pain sites with headaches. This suggest that musculoskeletal pain and headache are unique experiences caused by different neuronal mechanisms, which is in line with reports that in congenital insensitivity to pain phenotypes, there are painless cases with the only pain felt consisting in tension headaches<sup>68</sup> and furthermore, is consistent with findings that there is a shared genetic factors in conditions manifesting chronic pain except for migraine<sup>69</sup>.

Apart from headache, the different pain sites suggest the involvement of one of two main neuron types PEP1 and NP2 in all cases but hip pain, which is associated with both. Thus, these results suggest that different neuron types are associated with different chronic pain sites. This opens for the question if the different types of pain could be location dependent, thus depending on segmental levels of DRGs or trigeminal ganglion neurons in the facial region rather than neuronal type dependency. However, analysis of somatosensory neurons across the rostral-caudal axis of the mouse, including DRG<sup>7–9</sup>, jugular ganglion<sup>70</sup>, and trigeminal ganglion<sup>7</sup>, reveals that the neuronal strategy of somatosensation is shared regardless of body location. Thus, similar types of somatosensory neurons exist in the DRG as in jugular and trigeminal ganglion. Because of this, we assumed that rostral-caudal differences in pain sites should not affect the identification of involved cell types. This motivated us to use macaque DRG to predict the heritable risk genes at different pain locations in the humans, including trigeminal area pain. While mouse models have not specifically addressed the contribution of NP2 neurons to chronic pain, the contribution of PEP1 neurons to pain is well established. For example, ablation of CGRP<sup>+</sup> neurons, which include the PEP1 neurons, results in the loss of noxious heat sensation as well as inflammatory and neuropathic heat hyperalgesia in the mouse<sup>71</sup>.

Largely different genes contribute to chronic pain in PEP1 and NP2 neurons. In previous human association studies for musculoskeletal chronic pain there is a marked enrichment of genes involved in neurotransmission but also in, for example, immune genes, metabolic processing, skeletal tissue differentiation, and hormone signaling pathway genes<sup>26</sup>. In this study, we expected to capture the signal associated with the heritable risk in the different somatosensory neurons. The results reveal unique vulnerability pathways at play during pain chronification in the different neuronal types, suggesting that the causal mechanisms might be different between chronic pain conditions. While there is an enrichment of risk genes in PEP1 neurons that belongs to clathrin-dependent endocytosis and axon and nervous system development, NP2 neurons display an enrichment in synaptic organization and transmission and cell projection morphogenesis. This does not inform that these pathways are the only ones present in the different sensory neurons of particular populations. Instead, it shows that expression of, for example, specific members of cell adhesion/repulsion genes carrying nearby variations with significant heritable risk to chronic pain are enriched in PEP1 neurons. Analysis of the underlying genes with enriched heritability reveal a common pattern related to a susceptibility of neuronal connectivity, although with different functional classes of genes in different neurons. Thus, we conclude that the results support the notion that the major genetic risks expressed in somatosensory neurons are carried in genes involved in structural and functional connections of the neurons, and thus impacts on neurotransmission. The most direct effects may be contributed by synaptic adhesion molecules, which are known to be involved in synaptic plasticity of sensory neurons.

We have mapped heritability to two specific types of primary sensory neurons. However, as previously mentioned, we do not explain the full heritability to musculoskeletal chronic pain in this study, since cell types other than those analyzed may also contribute to chronic pain. Furthermore, the improvement of the statistical power and functional human genomic data may add resolution. Thus, more genes could contribute to chronic pain within sensory neurons as well as in cells other than primary sensory neurons, such as, for example, immune and vascular cells for headaches. Nevertheless, the finding that two neuronal types among the variety of sensory neurons carry a significant

enrichment for the heritable risk to musculoskeletal pain indicates a contribution of these cell types to the pathophysiology of musculoskeletal chronic pain. This provides a rationale for a deeper investigation of their participation with regard to human chronic pain. Furthermore, many drugs fall into the translational chasm between mice and humans<sup>72</sup>. Our atlas of somatosensory neuron gene expression in non-human primate should be important to verify putative drug targets and can also be of help in informed strategies for development of conceptually new analgesic drugs.

## Methods

**Animals.** For WaferGen (STRT-2i-seq), DRGs from two 5-year-old female (samples WG17019 and WG17020) and one 14-year-old male (WG18008) macaques were used. Smart-Seq2 samples were prepared from five females aged 5–7 years from the same colony. All animals were healthy Indian rhesus macaques (*Macaca mulatta*) from a colony of outbred animals housed in the Astrid Fagraeus laboratory at the Karolinska Institutet. To reduce the usage of laboratory animals, all tissue samples used in this study were collected from animals euthanized for organ collection in an unrelated study<sup>73,74</sup> (ethics license N2/15) approved by the Stockholm Laboratory Animals Ethics Committee in accordance to the Swedish legislation SJVFS 2019:9 saknr L150. A full list of samples used in the study is presented in Supplementary Data 11.

**Preparation of cell suspensions.** Approximately 1 h after animal termination 6–8 lumbar DRGs (three pairs of the biggest and one pair of anterior/posterior L3–L7) were exposed, dissected out, and kept in cold NMDG-cutting solution until dissociation (NMDG-CS adopted from ref. <sup>75</sup>; concentrations in mM: 103 NMDG (10 N HCl to adjust pH to 7.4), 2.5 KCl, 1.2 NaH<sub>2</sub>PO<sub>4</sub>, 30 NaHCO<sub>3</sub>, 20 HEPES, 25 glucose, 5 sodium ascorbate, 2 thiourea, 3 sodium pyruvate, 10 *N*-acetyl-L-cysteine, 1 mM EDTA). Dissociation procedure was modified from ref. <sup>76</sup> and all reagents were from the same sources. All following steps were performed in cold NMDG-CS. Connective tissue and nerve excess were removed before dissociation. DRGs were cut once longitudinally (along the nerve) followed by chopping into ~0.5 mm slices with a tissue slicer. Samples were enzymatically treated (3.9 ml papain solution (45 units/ml, PAPL, cat.n.LS003118, Worthington) in NMDG-CS, 0.2 ml DNase I (1 mM in NMDG-CS, cat.n.LK003172, Worthington), 0.4 ml TrypLE (cat.n.12605036, Life Technologies), 0.4 ml Collagenase/Dispase (20 mg/ml, cat.n. LS004106, Worthington)) for 1 h at 37 °C with triturating every 15 min through a 1 ml tip cut to a ~5 mm opening. Then the sample was gently triturated through 3–4 cut 1 ml tips (5–2 mm opening). The resulting cell suspension was then run through 100- $\mu$ m cell strainers (pluriStrainer<sup>®</sup>, 100  $\mu$ m). The total volume of the sample (after strainer rinsing) was ~5 ml. One milliliter of 10% Optiprep (cat.n. D1556, Sigma Aldrich) in NMDG-CS was loaded under the cell suspension solution using a gel loading tip and centrifuged at 200g for 6 min with a low break. Supernatant was discarded, the cell pellet re-suspended in NMDG-CS + B27 supplement (NMDG-CS/B27), and run through a 10  $\mu$ m pluriStrainer<sup>®</sup> pre-coated with NMDG-CS/B27. The strainer was rinsed twice with 5 ml of NMDG-CS/B27 applied slowly along the wall of the strainer, turned over, and the remaining big cells were then collected by back flushing the filter twice with 5 ml of filtered NMDG-CS/B27. Five microliters of CellTracker (Green CMFDA, cat.no. C7025, 50  $\mu$ g in 10  $\mu$ l of DMSO) was added into the resulting 10 ml of cell suspension. The cells were kept on ice for 15–20 min, centrifuged for 3 min at 100g and re-suspended in 900  $\mu$ l of 0.45  $\mu$ m filtered NMDG-CS/B27/12% optiprep (optiprep was added to slow cell sinking in the solution while dispensing) and dispensed immediately into a WaferGen9600 Chip. Samples for Smart-Seq2 protocol (SS2) were dissociated in a similar way but CellTracker Orange was used instead of Green. Then samples were run through FACS (BD Influx, nozzle—200  $\mu$ m pressure—2.5 PSI), gating events for high values in both SSC and Orange fluorescent channel and sorted into 384 plates for immediate freezing for later processing in the Smart-Seq2 pipeline.

**scRNA-seq, sequencing, and sequence alignment.** WaferGen chips were processed according to the STRT-seq-2i workflow described in ref. <sup>77</sup> and FACS sorted 384-well plates were processed according to the Smart-Seq2 workflow described in ref. <sup>78</sup>. The resulting WaferGen (samples WG in the text) and Smart-Seq2 (samples SS2 in the text) were sequenced on three and one lanes, respectively, on the Illumina HiSeq 2500 platform. Reads were aligned to the *Macaca mulatta* genome build (Mmul\_10/rheMac10, <ftp://hgdownload.soe.ucsc.edu/goldenPath/rheMac10/>). Since transcript annotations are incomplete on the macaque genome, and names of macaque genes differ from these of the human orthologues or putative orthologues, we used the following procedure to extend and normalize the monkey transcript annotation set with the human hg38 annotations: (1) LiftOver (<http://hgdownload.soe.ucsc.edu/goldenPath/rheMac10/liftOver/>) was used to align the rheMac10 and hg38 genomes. (2) A transcript from hg38 that had not even the smallest overlap with any transcript in rheMac10 was added to final transcript set. (3) All transcripts

from rheMac10 were put in the final set. If either 5' or 3' end matched within a few bases to a human transcript, the gene name was changed to the human name. The exception to this was microRNAs ("MIRxxx") in hg38, where both ends were required to match for a name change to be applied. (4) In the few cases where two hg38 transcripts with the same 5' or 3' end had different gene names, a rheMac10 transcript with the same end was assigned a name which is a combination ("name1/name2/...") of the alternative names from hg38. 5) When there was a longer 5' or 3' extension in the liftOver model of a human transcript compared with the overlapping monkey transcript, the 5'/3' exons of the corresponding monkey gene were adjusted accordingly. Supplementary Data 12 lists all 2445 genes for which names from Mmul10 build were replaced with hg38 gene names, as described above. The outcome gene statistics for raw WG sample data is the following: median genes detected per neuron—6380, total number of genes detected in at least three neurons—22160, of them 15821 are in human gene nomenclature consortium.

**RNASeque, microscopy, image analysis, and quantification.** Sets of freshly dissected macaque lumbar DRG were fresh frozen in OCT over dry ice and swiftly stored at  $-80^{\circ}\text{C}$  until sectioning. Cryosections were cut at  $10\ \mu\text{m}$  thickness, the slides were dried at RT, and then stored again at  $-80^{\circ}\text{C}$  to preserve RNA integrity. RNASeque version 2.0 using the RNASeque Fluorescent Multiplex Reagent Kit (Advanced Cell Diagnostics Inc.) was used for in situ hybridization. The staining combinations listed below were used for validations. TrpM8<sup>High</sup>: TRPM8, SCN10A (negative), CBLN2 (negative); PEP3: TRPM8, SCN10A, CBLN2 (negative); A-LTMR: CBLN2, TRPM8 (negative), SCN10A (negative); NP1: GFRA1, GFRA2, GAL (negative); NP2: GFRA1, GFRA2 (negative), GAL (negative); C-LTMR: GFRA2, GFRA1 (negative), GAL (negative); PEP1: GAL, GFRA1 (negative), GFRA2 (negative); NP3: GFRA3, PVALB (negative), GAL (negative); PEP2: SCN10A, PVALB. The following probes were used to validate the clusters: Hs-SCN10A (#Cat No. 406291), Hs-TRPM8-C3 (#Cat No. 543121), Mmu-GAL-C3 (#Cat No. 830031), Hs-GFRA1-C2 (#Cat No. 435661), Hs-GFRA2 (#Cat No. 463011), Hs-CBLN2-C2 (#Cat No. 446051), Hs-GFRA3-C2 (#Cat No. 535521), and Mmu-PVALB (#Cat No. 461691). The cross-reaction of human probes was predicted computationally by Advanced Cell Diagnostics Inc. and verified in macaque tissue by the authors. A Zeiss LSM800 confocal microscope was used to capture images and ImageJ (v.1.52h) was used for image analysis. Lumbar DRG sections from two animals (females, about 5 years old) different from those used for scRNA-sequencing were processed for quantification. Quantification of the proportion each neuronal type in the DRG and measurements of cell diameter of the different neuronal types was performed by two persons (total cells analyzed = 1009 cells). Background enhancement was used to identify all neuronal profiles with a visible nucleus in each section. Signal quantification was performed automatically in imageJ on manually outlined cells. Using this data cells were classified to clusters in R using a probe signal threshold as a positive classifier. Probe signal threshold was determined for each individual probe signal using a signal distribution plot. Different animals showed similar relative diameter of each neuronal type (Pearson correlation coefficient = 0.98).

**scRNA-seq analysis of Macaque data.** R (version 4.0.2) and Seurat (version 3.2.2) were used for the scRNA-seq analysis. Three objects were created from the individual biological STRT-2i-seq replicates. The data were normalized (*NormalizeData*) after which 2000 most variable features were selected (*FindVariableFeatures*). To mitigate batch effects between replicates we used Seurat's integrated analysis approach that transforms datasets into a shared space using pairwise correspondences (or "anchors")<sup>40</sup>. Anchors were first defined using *FindIntegrationAnchors* (dims = 1:40) and the data were then integrated (*IntegrateData*) and scaled (*ScaleData*), followed by principal component analysis (PCA) (*RunPCA*, npcs = 100). For clustering, the following parameters were used: *RunUMAP*, reduction = pca, dims = 1:20; *FindNeighbors*, reduction = pca, dims = 1:20; *FindClusters*, resolution = 1. The clusters showing minimal expression of a neuronal markers (*SNAP25*, *RBFOX3*) and/or low number of detected genes were removed. The following round of clustering (resolution = 3) produced a total 29 clusters, of which three expressing markers of injured cells (e.g. *ATF3*) satellite glia (e.g. *APOE*) or other non-neuronal cells (e.g. *EMCN* [endothelial cells]) were removed. The cells were then reclustered (resolution = 2) and remaining cells with ambiguous identities were removed. After this, a residual signal coming from satellite-glia marker genes was scored for each cluster (*AddModuleScore*; features = "SPARC", "PMP2", "APOE", "SPARCL1", "PLP1", "ABCA8", "SI00A1", "LPL", "IFITM4P", name = "sgc\_score"). The cells were then clustered for the final time (dims = 1:30, resolution = 2) and the satellite-glia signal was regressed out (*ScaleData*; vars.to.regress = "sgc\_score1"). After this, highly similar clusters without clearly distinguishable markers were merged to produce the final nine clusters. A dendrogram was then built (*BuildClusterTree*) using dimensions 1:30. For Smart-seq2, the individual datasets for each animal were initially integrated and clustered similarly to the STRT-2i-seq data. After the first round, clusters showing a non-neuronal identity and/or low levels of detected genes were removed. After this, the cells were cluster a second time (dims = 1:20, resolution = 1.5). To assign identities to the clusters we used label transfer between the STRT-2i-seq and Smart-seq2 datasets (*FindTransferAnchors*; reference = STRT-2i, query = Smart-seq2, dims = 1:30; *TransferData*; anchorset = anchors, refdata =

usoskin\_id, dims = 1:30). Wilcoxon rank sum test was used to identify cluster-enriched genes, reporting two-sided adjusted *P* values based on Bonferroni's correction (FDR). Genes detected in >25% of cells in the cluster with FDR < 0.05 and >0.25 average log<sub>2</sub>-fold-change were considered as marker genes for the cluster.

**Label transfer between mouse DRG datasets.** To visualize mouse DRG scRNA-seq datasets as UMAPs, the individual datasets from refs.<sup>7,9</sup> were clustered with Seurat (*ScaleData*, vars.to.regress=c("orig.ident", "nCount\_RNA"); *RunPCA*, npcs = 100; *RunUMAP*, dims = 1:20; *FindNeighbors*, dims = 1:20; *FindClusters*, resolution = 0.1) but the cell identities were set as original identities from the publications. Label transfers between the datasets were then performed using the different combinations of reference and query datasets and nomenclatures (*FindTransferAnchors*; reference = reference\_dataset, query = query\_dataset, dims = 1:30; *TransferData*; anchorset = anchors, refdata = used\_nomenclature, dims = 1:30).

**Conos analysis.** For Conos<sup>41</sup> WG macaque datasets were integrated using CCA space (*\$buildGraph*(k = 15, k.self=5, space = "CCA", ncomps=30, n.odgenes=2000, snn=F, snn.k = 50), followed by UMAP embedding (*\$embedGraph*(method = "UMAP", min.dist=0.1, spread=20, n.cores=4, min.prob.lower=1e-3). For macaque (WG) and mouse (Zeisel) dataset co-integration was performed as *\$buildGraph*(k = 15, k.self=5, space = "CCA", ncomps=30, n.odgenes=2000, snn=T, snn.k = 15), followed by largeVis embedding: *\$embedGraph*(alpha = 0.1, sgd\_batched=1e6, seed = 1). Label propagation (*\$propagateLabels*) was run using method "diffusion".

**Mouse DRG scRNA-seq data.** Mouse DRG data were downloaded from ([http://loom.linnarssonlab.org/clone/Mousebrain.org.level6/L6\\_Peripheral\\_sensory\\_neurons.loom](http://loom.linnarssonlab.org/clone/Mousebrain.org.level6/L6_Peripheral_sensory_neurons.loom)) and GEO (GSE139088) and the cluster annotation was modified to conform to established nomenclature<sup>8</sup>. Mouse gene names were switched to corresponding human orthologs using biomaRt (v2.42).

**Scoring analysis of cell-type identity.** For this analysis, our goal was to score the probabilistic cell identity of each cell relative to the defined cell types at the transcriptional level<sup>79,80</sup>. We built a vanilla neural-network model for classification tasks in PyTorch framework with CUDA support for GPU computation, and trained the model to learn the general prototypes of defined cell types. To train the model, we obtained the over-dispersed genes by estimating the mean and coefficients of variation. The over-dispersed genes were further ranked by two heuristics for cell-type specificity of both fold-change and enrichment score-change<sup>9</sup>. The cross-species alignment was performed as described in ref.<sup>80</sup>. Subsequently, the ranked marker genes of defined cell types were log-transformed and scaled by Minmax normalization, and then used for the neural-network model. The neural-network model contains an input layer with the number of neuron nodes similar to the number of marker genes, a hidden layer with the number of neuron nodes similar to 20% of marker gene numbers, and an output layer with the number of neuron nodes similar to the number of defined cell types. Linear regression was performed between each layer, and the 30% threshold for dropouts was set to reduce the overfitting. Rectified linear units were used as the activation function of the hidden layer, and Softmax was used for the output layer to evaluate the probabilities. Nesterov Momentum was used as a stochastic gradient descent (SGD) optimizer. To choose the adequate regularization strength, the classifier accuracy and the loss value were inspected against epoch numbers. The classifier accuracy was estimated by a *k*-fold cross-validation, of which the dataset was randomly split (*k* = 3). The learning rate, epoch number, and momentum were chosen corresponding to the maximum point of learning curve reaching the accuracy plateaus. The ready model was used to predict the probabilities of each cell belonging to each trained reference cell types. The permutation test of dataset was applied to quantify the significance of the prediction, and the *p* values were calculated by FDR. Data were visualized using the radar plot<sup>81</sup> consisting of a sequence of equiangular polygon spokes with the distal vertex representing each trained reference cell type. The distance between the polygon center and each vertex of the polygon represents the relative probabilities of each trained reference cell assigning to the defined reference cell types and the position of each predicting cell was calculated as a linear combination of the probabilities against all reference cell types, and then visualized as the relative position to all vertices of the polygon.

**Comparison of transcriptional signatures between species.** Individual Seurat objects were formed for each species from the two mouse datasets and from the two macaque datasets (Seurat function *merge*). Within each object, the genes expressed in each cluster were first filtered (*FindMarkers*; only.pos=TRUE, logfc.threshold=0). Then, genes found specifically in one species (*setdiff*; species1\$genes, species2\$genes) or in both (*intersect*; species1\$genes, species2\$genes) were listed. A log<sub>2</sub>FC threshold of 0.25 was used for finding genes that were expressed above baseline for a corresponding cluster in both species. A similar threshold was used when defining species unique cluster markers.

**Scenic.** SCENIC<sup>43</sup> was used to infer active transcription factors and their target genes. SCENIC was performed following the authors' vignette: [https://github.com/aertslab/SCENIC/blob/master/vignettes/SCENIC\\_Running.Rmd](https://github.com/aertslab/SCENIC/blob/master/vignettes/SCENIC_Running.Rmd). First, gene sets co-expressed with transcription factors were identified and modules kept if the transcription factor motif was enriched among its targets. Then, target genes without direct motif support were removed. Finally, regulon activity was scored and binarized to determine whether the genes in each regulon are enriched in each cell.

**Computational genomics screen for gene families.** For the supervised computational screen using Metaneighbor<sup>82</sup>, a set of 1500+ gene families was downloaded from the HGNC website (<https://www.genenames.org/download/custom/>). The gene families were filtered only to include families with three or more genes expressed in the analyzed data. The analysis was run following the vignette provided by the authors (<https://github.com/maggiacrow/MetaNeighbor>). In brief, this is a machine-learning-based analysis scores the performance to correctly connect cells of known corresponding identity based on the similarity of their transcriptional profiles across a given set of genes<sup>42</sup>. For intra-species analyses, comparisons were run between 1 vs 2 individual macaques (for STRT-2i-seq data) and between two biological replicates for the mouse data. For cross-species analyses, comparisons were run with "species" as the identifier. To improve readability, the lists of highest scoring families were curated for redundancy for the figures (Supplementary Fig. 8), showing only the highest scoring family from a group of highly similar families. The full tables can be found in Supplementary Data 5, 6 and 7.

**Cell-type-specific partitioned heritability.** We followed the protocol established by Finucane et al.<sup>32</sup>. Three steps of analysis were required to address our hypothesis. The first step consisted in identifying the top 10% genes most associated with a macaque DRG cluster cell type. For this, we extracted from the expression matrix the raw counts of genes in each single-cell sample. Sample size factors were estimated using DESeq2's estimateSizeFactors<sup>83</sup>. Then, for each gene, we regressed gene expression of all samples to cluster membership (yes = 1, no = 0), using size factors and animal's sex as co-variables. We then sorted the genes by decreasing regression's test statistics and retained the top 10% ( $n = 1322$ ) most associated genes to a cluster. The second step consisted in creating cluster-specific annotations following instructions given at URL <https://github.com/bulik/ldsc/wiki/LD-Score-Estimation-Tutorial>, under the "Partitioned LD Scores" section. The control track was made up of the same genes used in Franke's control dataset<sup>84</sup> by Finucane et al.<sup>32</sup>. The third and final step consisted of screening various pain-related summary GWAS from the UK Biobank against these annotations to uncover cluster-specific contributions to partitioned heritability. Statistical significance was established at up to the 20% FDR level on a per-cluster basis, i.e., the genetic contribution for each pain site was tested in each cluster. We chose FDR instead of Bonferroni as epidemiological data establish co-morbidity of pain sites report. Meta-analyses of partitioned heritability coefficients were performed using the inverse standard error approach suggested in METAL<sup>85</sup>.

**Sample size.** For GWAS in the UK Biobank, the sample size was determined by the number of available participants once genotyping quality controls have been performed followed by our own study exclusion criteria (described below). GWAS genotyping quality controls have been performed by the UK Biobank themselves, and are described in this document: [https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/genotyping\\_qc.pdf](https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/genotyping_qc.pdf).

**Data exclusions.** GWAS participants were excluded based on: (1) poor genotyping quality, (2) voluntary retraction from the UK Biobank study, (3) of ancestries other than Caucasian (UK Biobank field 22006), (4) participants who answered "prefer not to answer" for the touchscreen question "In the last month have you experienced any of the following that interfered with your usual activities?" (UK Biobank field 6159).

**Replication.** For GWAS, no other cohorts with the same recruitment characteristics (cross-sectional) and matching pain questionnaires (self-declared at eight body sites, pain experienced for more than 3 months) were available for replication. The UK Biobank is one of the largest cohorts available for genetics studies. The UK Biobank cohort was used as a discovery cohort to test the hypothesis of involvement of genetic loci in genes expressed in specific cell types in primate dorsal root ganglia in pain phenotypes.

**Randomization.** No randomization of subjects in the UK Biobank for GWAS purposes was needed.

**Blinding.** The UK Biobank study was cross-sectional in nature, and so is agnostic to self-declared pain status in participants. We did not do any recruitment as part of the UK Biobank project. Study protocols were approved by UK Biobank's own "Ethics Advisory Committee". We analyze the data granted by UK biobank application number 20802.

**Cluster and GWAS enrichment.** To uncover genes simultaneously strongly associated with cluster specificity and chronic pain phenotypes, we estimated the area spanned by a given gene using a scatter plot in which both cell type and GWAS enrichments were tracked. Gene-level test statistics were obtained from summary chronic pain GWAS using MAGMA<sup>86</sup>. We retained genes for which cell-type specificity was positive, indicating an increased gene expression in the cell type of interest, with enrichment increasing with cell-type specificity.  $P$  value for a gene's area was estimated from a fit of the density of genes to a Bessel function (appropriate to model a product of two normal distributions), and then integrating from minus infinity to minus the absolute value of the area of the gene. The Bessel estimator was such that integrating over from minus infinity to plus infinity yielded a value of 1 (the estimator was normalized). All genes up to FDR 20% were retained for pathway analyses. Pathway analyses were conducted using the hypergeometrical test for overrepresentation<sup>87</sup>, with pathways sourced from Gene Ontology's biological processes<sup>88</sup>, obtained December 2019 from URL [http://download.baderlab.org/EM\\_Genesets/](http://download.baderlab.org/EM_Genesets/). We tested pathways that featured a number of genes between 10 and 1000.

**UK Biobank.** The UK Biobank is a large genetic study comprising half a million participants aged between 40 and 69 years old<sup>47,48</sup>. The first round of standard genotyping quality control was performed by the UK Biobank consortium and is fully documented on their web portal (<http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=155580>). Samples were discarded based on failed genotyping QC (heterozygosity rate, genotyping rate, etc.), genetic vs declared sex mismatch, voluntary retraction, and of non-"white British" ancestry. Genome-wide association study was performed using the full release 500K cohort version. BOLT-LMM version 2.3 performed association tests, with consideration of kinship among participants<sup>89</sup>. Age, age squared, sex, dummy-coded recruitment sites, genotyping platform, and the top 40 principal components were used as co-variables. The phenotype was defined as chronic pain at a body site (e.g. back). Cases were defined as those who answered "yes" at the touchscreen question "Have you had back pains for more than 3 months?" and that had answered "yes" at the question "In the last month have you experienced any of the following that interfered with your usual activities?" (field 6159). Control individuals were those who answered "None of the above" at field 6159 ( $n = 163,825$ ). A total of eight pain sites were available: headaches (UK Biobank field 3799; 36,381 cases), facial (field 4067; 3495 cases), neck and shoulder (field 3404; 63,933 cases), stomach and abdominal (field 3741; 18,749 cases), back (field 3571; 70,634 cases), hip (field 3414; 35,634 cases), knee (field 3773; 68,237 cases), and widespread (field 2956; 5259 cases). To control for overlap of chronic pain sites in individuals, we also performed additional GWASes for pain sites in which cases who reported a particular additional pain site we removed, one site at a time, from the analysis. A total of  $8 \times 8 = 56$  such GWASes were performed. Post-GWAS SNP quality control included: (1) be part of the Haplotype Reference Consortium<sup>90</sup>, (2) minor allele frequency >0.1%, (3) Hardy-Weinberg  $P$  value > $10^{-12}$ , (4) INFO score >0.9. In all, this yielded a total of about 8.238 million SNPs per GWAS. The minimum effective count of minor alleles was  $0.1\% \times 0.9 \times 163,835 \times 2 \sim 295$ . This work was done under the UK biobank application number 20802.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The raw and processed datasets for the scRNA-seq of macaque DRG neurons reported in this study (STRT-2i-seq and Smart-seq2) have been deposited in the Gene Expression Omnibus (GEO) under the accession GSE165569. Processed data for browsing gene expression in the different macaque neuron types can be accessed at the following links: STRT-2i-seq dataset (lumbar dorsal root ganglia of three adult Rhesus macaques) <https://ernforsgroup.shinyapps.io/macaqueDRG/> and Smart-seq2 dataset (dorsal root ganglia of five adult Rhesus macaques) <https://ernforsgroup.shinyapps.io/macaqueSS2/>. Macaca mulatta genome build (Mmul\_10/rheMac10) is available at: <ftp://hgdownload.soe.ucsc.edu/goldenPath/rheMac10/>. Mouse DRG data (Zeisel) is available at [http://loom.linnarssonlab.org/clone/Mousebrain.org.level6/L6\\_Peripheral\\_sensory\\_neurons.loom](http://loom.linnarssonlab.org/clone/Mousebrain.org.level6/L6_Peripheral_sensory_neurons.loom). Mouse DRG data (Sharma) is available at <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE139088>. We analyzed the data granted by UK biobank application number 20802, as the data from the UK Biobank is available from the biobank as part of individual agreements. Source data are provided with this paper.

## Code availability

Any custom code and data are available from the authors upon request. All analyses are based on previously published code and software (see Reporting summary).

Received: 24 July 2020; Accepted: 8 February 2021;

Published online: 08 March 2021

## References

1. Abraira, V. E. & Ginty, D. D. The sensory neuron touch. *Neuron* **79**, 618–639 (2013).
2. Dong, X. & Dong, X. Peripheral and central mechanisms of itch. *Neuron* **98**, 482–494 (2018).
3. Lewin, G. R. & Moshourab, R. Mechanosensation and pain. *J. Neurobiol.* **61**, 30–44 (2004).
4. Ma, Q. Population coding of somatic sensations. *Neurosci. Bull.* **28**, 91–99 (2012).
5. Emery, E. C. & Ernfors, P. In *The Oxford Handbook of the Neurobiology of Pain* (ed. Wood, J. N.) (Oxford University Press, 2018).
6. Li, C.-L. et al. Somatosensory neuron types identified by high-coverage single-cell RNA-sequencing and functional heterogeneity. *Cell Res.* **26**, 83–102 (2016).
7. Sharma, N. et al. The emergence of transcriptional identity in somatosensory neurons. *Nature* **577**, 392–398 (2020).
8. Usoskin, D. et al. Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. *Nat. Neurosci.* **18**, 145–153 (2015).
9. Zeisel, A. et al. Molecular architecture of the mouse nervous system. *Cell* **174**, 999–1014.e22 (2018).
10. Holford, L. C., Case, P., Lawson, S. N. & Substance, P. neurofilament, peripherin and SSEA4 immunocytochemistry of human dorsal root ganglion neurons obtained from post-mortem tissue: a quantitative morphometric analysis. *J. Neurocytol.* **23**, 577–589 (1994).
11. Josephson, A. et al. GDNF and NGF family members and receptors in human fetal and adult spinal cord and dorsal root ganglia. *J. Comp. Neurol.* **440**, 204–217 (2001).
12. Pearson, J., Pytel, B. A., Grover-Johnson, N., Axelrod, F. & Dancis, J. Quantitative studies of dorsal root ganglia and neuropathologic observations on spinal cords in familial dysautonomia. *J. Neurol. Sci.* **35**, 77–92 (1978).
13. Davidson, S. et al. Human sensory neurons: membrane properties and sensitization by inflammatory mediators. *Pain* **155**, 1861–1870 (2014).
14. Han, C. et al. Human Nav1.8: enhanced persistent and ramp currents contribute to distinct firing properties of human DRG neurons. *J. Neurophysiol.* **113**, 3172–3185 (2015).
15. Zhang, X., Priest, B. T., Belfer, I. & Gold, M. S. Voltage-gated Na<sup>+</sup> currents in human dorsal root ganglion neurons. *eLife* **6**, e23235 (2017).
16. North, R. Y. et al. Electrophysiological and transcriptomic correlates of neuropathic pain in human dorsal root ganglion neurons. *Brain* **142**, 1215–1226 (2019).
17. Flegel, C. et al. RNA-Seq analysis of human trigeminal and dorsal root ganglia with a focus on chemoreceptors. *PLoS ONE* **10**, e0128951 (2015).
18. Sapio, M. R., Goswami, S. C., Gross, J. R., Mannes, A. J. & Iadarola, M. J. Transcriptomic analyses of genes and tissues in inherited sensory neuropathies. *Exp. Neurol.* **283**, 375–395 (2016).
19. Ray, P. et al. Comparative transcriptome profiling of the human and mouse dorsal root ganglia: an RNA-seq-based resource for pain and sensory neuroscience research. *Pain* **159**, 1325–1345 (2018).
20. Haberberger, R. V., Barry, C., Dominguez, N. & Matusica, D. Human dorsal root ganglia. *Front. Cell. Neurosci.* **13**, 271 (2019).
21. Cox, J. J. et al. An SCN9A channelopathy causes congenital inability to experience pain. *Nature* **444**, 894–898 (2006).
22. Drissi, I., Woods, W. A. & Woods, C. G. Understanding the genetic basis of congenital insensitivity to pain. *Br. Med. Bull.* **133**, 65–78 (2020).
23. Indo, Y. et al. Mutations in the TRKA/NGF receptor gene in patients with congenital insensitivity to pain with anhidrosis. *Nat. Genet.* **13**, 485–488 (1996).
24. Leipold, E. et al. A de novo gain-of-function mutation in SCN11A causes loss of pain perception. *Nat. Genet.* **45**, 1399–1404 (2013).
25. Yang, Y. et al. Mutations in SCN9A, encoding a sodium channel alpha subunit, in patients with primary erythralgia. *J. Med. Genet.* **41**, 171–174 (2004).
26. Zorina-Lichtenwalter, K., Meloto, C. B., Khoury, S. & Diatchenko, L. Genetic predictors of human chronic pain conditions. *Neuroscience* **338**, 36–62 (2016).
27. Calvo, M. et al. The genetics of neuropathic pain from model organisms to clinical application. *Neuron* **104**, 637–653 (2019).
28. Diatchenko, L., Fillingim, R. B., Smith, S. B. & Maixner, W. The phenotypic and genetic signatures of common musculoskeletal pain conditions. *Nat. Rev. Rheumatol.* **9**, 340–350 (2013).
29. Veluchamy, A., Hébert, H. L., Meng, W., Palmer, C. N. A. & Smith, B. H. Systematic review and meta-analysis of genetic risk factors for neuropathic pain. *Pain* **159**, 825–848 (2018).
30. Meloto, C. B. et al. Human pain genetics database: a resource dedicated to human pain genetics research. *Pain* **159**, 749–763 (2018).
31. Parisien, M. et al. Effect of human genetic variability on gene expression in dorsal root ganglia and association with pain phenotypes. *Cell Rep.* **19**, 1940–1952 (2017).
32. Finucane, H. K. et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
33. Skene, N. G. et al. Genetic identification of brain cell types underlying schizophrenia. *Nat. Genet.* **50**, 825–833 (2018).
34. Nagel, M. et al. Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nat. Genet.* **50**, 920–927 (2018).
35. Coleman, J. R. I. et al. Biological annotation of genetic loci associated with intelligence in a meta-analysis of 87,740 individuals. *Mol. Psychiatry* **24**, 182–197 (2019).
36. Savage, J. E. et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat. Genet.* **50**, 912–919 (2018).
37. Jansen, I. E. et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer’s disease risk. *Nat. Genet.* **51**, 404–413 (2019).
38. Mogil, J. S. Animal models of pain: progress and challenges. *Nat. Rev. Neurosci.* **10**, 283–294 (2009).
39. Woolf, C. J. Overcoming obstacles to developing new analgesics. *Nat. Med.* **16**, 1241–1247 (2010).
40. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
41. Barkas, N. et al. Joint analysis of heterogeneous single-cell RNA-seq dataset collections. *Nat. Methods* **16**, 695–698 (2019).
42. Paul, A. et al. Transcriptional architecture of synaptic communication delineates GABAergic neuron identity. *Cell* **171**, 522–539.e20 (2017).
43. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
44. Scott, A. et al. Transcription factor short stature homeobox 2 is required for proper development of tropomyosin-related kinase B-expressing mechanosensory neurons. *J. Neurosci.* **31**, 6741–6749 (2011).
45. 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).
46. Zhong, J., Pevny, L. & Snider, W. D. ‘Run’ing towards sensory differentiation. *Neuron* **49**, 325–327 (2006).
47. Sudlow, C. et al. UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
48. Allen, N. E., Sudlow, C., Peakman, T. & Collins, R. Biobank, on behalf of U. UK Biobank data: come and get it. *Sci. Transl. Med.* **6**, 224ed4–224ed4 (2014).
49. Baumann, T. K., Burchiel, K. J., Ingram, S. L. & Martenson, M. E. Responses of adult human dorsal root ganglion neurons in culture to capsaicin and low pH. *Pain* **65**, 31–38 (1996).
50. Li, Y. et al. The cancer chemotherapeutic paclitaxel increases human and rodent sensory neuron responses to TRPV1 by activation of TLR4. *J. Neurosci.* **35**, 13487–13500 (2015).
51. Arcout, A. et al. Touch receptor-derived sensory information alleviates acute pain signaling and fine-tunes nociceptive reflex coordination. *Neuron* **93**, 179–193 (2017).
52. Knowlton, W. M. et al. A sensory-labeled line for cold: TRPM8-expressing sensory neurons define the cellular basis for cold, cold pain, and cooling-mediated analgesia. *J. Neurosci.* **33**, 2837–2848 (2013).
53. Kobayashi, K. et al. Distinct expression of TRPM8, TRPA1, and TRPV1 mRNAs in rat primary afferent neurons with aδ/c-fibers and colocalization with trk receptors. *J. Comp. Neurol.* **493**, 596–606 (2005).
54. Pogorzala, L. A., Mishra, S. K. & Hoon, M. A. The cellular code for mammalian thermosensation. *J. Neurosci.* **33**, 5533–5541 (2013).
55. Cavanaugh, D. J. et al. Distinct subsets of unmyelinated primary sensory fibers mediate behavioral responses to noxious thermal and mechanical stimuli. *Proc. Natl Acad. Sci. USA* **106**, 9075–9080 (2009).
56. Han, L. et al. A subpopulation of nociceptors specifically linked to itch. *Nat. Neurosci.* **16**, 174–182 (2013).
57. Liu, Q. et al. Mechanisms of itch evoked by β-alanine. *J. Neurosci.* **32**, 14532–14537 (2012).
58. Meixiong, J., Dong, X. & Mas-Related, G. Protein-coupled receptors and the biology of itch sensation. *Annu. Rev. Genet.* **51**, 103–121 (2017).
59. Sharif, B., Ase, A. R., Ribeiro-da-Silva, A. & Séguéla, P. Differential coding of itch and pain by a subpopulation of primary afferent. *Neurons Neuron* **106**, 940–951.e4 (2020).
60. Schwartzlow, C. & Kazamel, M. Hereditary sensory and autonomic neuropathies: adding more to the classification. *Curr. Neurol. Neurosci. Rep.* **19**, 52 (2019).
61. Woods, C. G., Babiker, M. O. E., Horrocks, I., Tolmie, J. & Kurth, I. The phenotype of congenital insensitivity to pain due to the Nav1.9 variant p.L811P. *Eur. J. Hum. Genet.* **23**, 561–563 (2015).

62. Chen, Y.-C. et al. Transcriptional regulator PRDM12 is essential for human pain perception. *Nat. Genet.* **47**, 803–808 (2015).
63. Tansey, K. E. & Hill, M. J. Enrichment of schizophrenia heritability in both neuronal and glia cell regulatory elements. *Transl. Psychiatry* **8**, 7 (2018).
64. Johnston, K. J. A. et al. Genome-wide association study of multisite chronic pain in UK Biobank. *PLoS Genet.* **15**, e1008164 (2019).
65. Smith, S. B. et al. Genome-wide association reveals contribution of MRAS to painful temporomandibular disorder in males. *Pain* **160**, 579–591 (2019).
66. Ge, T., Chen, C.-Y., Neale, B. M., Sabuncu, M. R. & Smoller, J. W. Phenome-wide heritability analysis of the UK Biobank. *PLoS Genet.* **13**, e1006711 (2017).
67. Suri, P. et al. Genome-wide meta-analysis of 158,000 individuals of European ancestry identifies three loci associated with chronic back pain. *PLoS Genet.* **14**, e1007601 (2018).
68. Danziger, N. & Willer, J.-C. Tension-type headache as the unique pain experience of a patient with congenital insensitivity to pain. *Pain* **117**, 478–483 (2005).
69. Vehof, J., Zavos, H. M. S., Lachance, G., Hammond, C. J. & Williams, F. M. K. Shared genetic factors underlie chronic pain syndromes. *Pain* **155**, 1562–1568 (2014).
70. Kupari, J., Häring, M., Agirre, E., Castelo-Branco, G. & Ernfors, P. An atlas of vagal sensory neurons and their molecular specialization. *Cell Rep.* **27**, 2508–2523.e4 (2019).
71. McCoy, E. S. et al. Peptidergic CGRP $\alpha$  primary sensory neurons encode heat and itch and tonically suppress sensitivity to cold. *Neuron* **78**, 138–151 (2013).
72. Borsook, D., Hargreaves, R., Bountra, C. & Porreca, F. Lost but making progress—where will new analgesic drugs come from? *Sci. Transl. Med.* **6**, 249sr3 (2014).
73. Marcandalli, J. et al. Induction of potent neutralizing antibody responses by a designed protein nanoparticle vaccine for respiratory syncytial virus. *Cell* **176**, 1420–1431.e17 (2019).
74. Thompson, E. A. et al. TLR-adjuvanted nanoparticle vaccines differentially influence the quality and longevity of responses to malaria antigen Pfs25. *JCI Insight* **3**, e120692 (2018).
75. Ting, J. T., Daigle, T. L., Chen, Q. & Feng, G. In *Patch-Clamp Methods and Protocols* (eds. Martina, M. & Taverna, S.) 221–242 (Springer, 2014).
76. Häring, M. et al. Neuronal atlas of the dorsal horn defines its architecture and links sensory input to transcriptional cell types. *Nat. Neurosci.* **21**, 869–880 (2018).
77. Hochgerner, H. et al. STRT-seq-2i: dual-index 5' single cell and nucleus RNA-seq on an addressable microwell array. *Sci. Rep.* **7**, 16327 (2017).
78. Picelli, S. et al. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–1098 (2013).
79. Malta, T. M. et al. Machine learning identifies stemness features associated with oncogenic dedifferentiation. *Cell* **173**, 338–354.e15 (2018).
80. Manno, G. L. et al. Molecular diversity of midbrain development in mouse, human, and stem cells. *Cell* **167**, 566–580.e19 (2016).
81. Hovestadt, V. et al. Resolving medulloblastoma cellular architecture by single-cell genomics. *Nature* **572**, 74–79 (2019).
82. Crow, M., Paul, A., Ballouz, S., Huang, Z. J. & Gillis, J. Characterizing the replicability of cell types defined by single cell RNA-sequencing data using MetaNeighbor. *Nat. Commun.* **9**, 1–12 (2018).
83. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
84. Fehrmann, R. S. N. et al. Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat. Genet.* **47**, 115–125 (2015).
85. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
86. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
87. Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257–258 (2007).
88. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330–D338 (2019).
89. Loh, P.-R. et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).
90. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).

## Acknowledgements

We thank Martin Häring for advice on enzymatic treatment/dissociation strategies and Ming-Dong Zhang for advice on dissection. We also thank Gunilla B. Karlsson Hedestam for donating euthenized animals, Sten Linnarsson for computational support, and Eneritz Agirre for help with the Shiny-app and running SCENIC. We acknowledge the Eukaryotic Single Cell Genomics (ESCG) Facility and the Mass Cytometry National Facility at the Science for Life Laboratory, Sweden, for the scRNA-seq and for FACS, respectively. This work was supported by the Swedish Medical Research Council, Knut and Alice Wallenbergs Foundation (Wallenberg Scholar and Wallenberg project grant), SFO grant (StratNeuro), Wellcome Trust (Pain Consortium), European Research Council advanced grant (PainCells 740491), and Karolinska Institutet (to P.E.), EMBO fellowship to M.F., and by the Canadian Excellence Research Chairs (CERC) Program (to L.D.).

## Author contributions

Design of experiments: J.K., D.U., and P.E.; computational analysis: J.K., D.U., P.V.K., and Y.H.; LiftOver, data quality assessment, and gene annotation: N.B., P.K., and P.L.; design of RNAScope experiments: J.K., M.F., and D.U.; RNAScope experiments: D.L. and D.U.; single-cell suspensions: D.U.; provision of primate tissues: K.L.; dissections: D.U., M.S., and B.E.; human genetic studies: M.P. performed bioinformatics analyses and S.K. and L.D. interpreted the bioinformatics results. Writing—review & editing: J.K. and P.E., with input from all authors; and supervision and funding: J.K. and P.E.

## Funding

Open access funding provided by Karolinska Institute.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-21725-z>.

**Correspondence** and requests for materials should be addressed to L.D. or P.E.

**Peer review information** *Nature Communications* thanks Clifford Woolf and the other anonymous reviewer(s) for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021