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.
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 itch1–3. 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 neurons4–9.

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 sensation1–3.

Nociception is largely conferred through unmyelinated peptidergic C-fiber neuron types and a few lightly myelinated Aδ-nociceptors, a TrpM8 expressing cluster of neurons, as well as cell types marked by expression of Mrgprd, Mrgpra3, or Sst (named small neurons in humans, but only small neurons in mouse, suggesting that heritability for musculoskeletal pain converge on two discreet DRG sensory neuron types2. 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 rodents.

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)21–25. 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 size26. Close to half of the risk of developing chronic pain are attributable to genetic factors27–29, including musculoskeletal pain conditions28. 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 neurons29. Pleiotropy of single-nucleotide polymorphisms (SNPs) among painful and non-painful conditions has also been shown30, even in human DRG31. 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 schizophrenia32,33, neuroticism34, intelligence35,36, and Alzheimer’s disease37, 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 disorders38,39 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 Seurat40. 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 Conos41, 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 gene expression; C-LTMRs (6) were assigned using Gfra2 and Znf521 (Zpf521 in the mouse); NTRK1 and GAL were used to identify Pep1 (4); Scn11a and Trpm8 suggested the identity of the Trpm8high (2) cluster (negative for Scn11a); Il31ra expression was used for naming Np3 (7); Cpe6 together with Ntrk2 was used to assign putative A-LTMRs (1), and Cpe6 together with Ntrk1 and Scn11a were used to assign the Pep2 (3) cluster. The final cluster (5) also expressed Cpe6, Ntrk1, and Scn11a, 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.
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/.

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 × 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 Smart-seq2 clusters after label transfer from the STRT-2i-seq data. Image sources: freevectors.net (human silhouette), needpix.com (microwell plates), openclipart.org (Eppendorf tube).
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. 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 (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...
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 type features 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 Aδ-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 Aδ-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 PIZEO2 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 (Supplementary Fig. 4a). This study also showed evidence of the mechanosensory channel PIZEO2 in A-LTMRs, C-LTMRs, and NP1 (Fig. 3c).

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...
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, SIPR1) 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 mouse7,9 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 SIPR1 (n = 36 genes) and for C-LTMRs P2RY1, EXOC11, KCND3, IQSEC2, OSBPL1A, and FXYD6 (n = 60 genes). The largest cell-type-specific shared gene program was found between mouse and macaque Aδ-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, HTRIF, 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 species42. 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/NFI (A6-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 SCENIC43 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 mouse9,44–46 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 project47,48, 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 pain in PEP1 neurons ($P_{FDR} = 16\%$, each), while NP2 neurons were associated most significantly with the heritability of chronic back pain and hip pain ($P_{FDR} = 5\%$ 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 ($P_{FDR} = 10\%$, 10\%, 13\%, respectively), while NP2 neurons were associated most significantly with the heritability of chronic back pain, hip pain, and knee pain ($P_{FDR} = 10\%$, 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 features13–16,49,50. 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 mouse7–9 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 sensation1. C-LTMRs involved in the affective aspect of pleasant touch5, C-cold thermoreceptors (TrpM8high), Aδ fast mechanical nociceptors involved in pinprick pain (PEP2)51–54 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\textsuperscript{55–57}. Although the neural-network prediction score was low, macaque PEP3 appears to be an $\alpha$-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 Mrgrpd receptor, but not thermal sensation. In the macaque, the
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 neurons) specifically affects histamine-dependent and histamine-independent itch, but not acute noxious heat, cold, or mechanical pain. NP2 neurons have therefore been considered to be dedicated itch neurons in rodents. 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, 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. 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. 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) and PRDM12. 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 TRP8high, myelinated nociceptors, and A-LTMRs (see https://erfnorgroup.shinyapps.io/macaquegr/ for interrogation of gene expression). NTRK1 is largely confined to macaque TRP8high, 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 Aδ-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 TRP8high 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 contribute to the heritable risk of chronic pain. Recent methods 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 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% with 7.6% for chronic back pain. 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. Furthermore, all chronic pain GWAS in the UK Biobank displayed inflated genomic control parameters (λGC ≈ 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 and furthermore, is consistent with findings that there is a shared genetic factors in conditions manifesting chronic pain except for migraine.
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 rostro-caudal axis of the mouse, including DRG7–9, jugular ganglion10, and trigeminal ganglion11, 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 rostro-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+ 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.12

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.26 In this study, we expected to capture the signal associated with the heritable risk in the different somatosensory neurons. This 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/repression 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 rational for a deeper investigation of their participation with regard to human chronic pain. Furthermore, many drugs fall into the translational chasm between mice and humans9. Our atlas of somatosensory neuron gene expression in non-human primates 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 for organ collection in an unrelated study.24 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 I.

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 dissected out, and kept in cold NMDG-cutting solution until disassociation (NMDG-CS adopted from ref.75; concentrations in mM: 103 NMDG (10 N HCl to adjust pH to 7.4), 2.5 KCl, 1.2 NaH2PO4, 30 NaHCO3, 20 HEPES, 25 glucose, 5 sodium ascorbate, 2 thiose, 3 sodium pyruvate, 10 N-acetyl-l-cysteine, 1 mM EDTA). Disassociation procedure was modified from ref.76 and all reagents were from the same sources. All following steps were prepared in cold NMDG-CS. Connective tissue and nerve excess were removed before disassociation. 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.LS005318, Worthington) in NMDG-CS, 0.2 ml CH2O, 1 ml NMDG-CS) and run through a 10 µm pluriStrainer (cat. Mmul_10/rheMac10, ftp://hgdownload.soe.ucsc.edu/goldenPath/rheMac10/). Since macaque genes differ from those of the human orthologues or putative orthologues, transcript annotations are incomplete on the macaque genome, and many NMDG-CS. Connective tissue and nerve excess were removed before disassociation. 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.LS005318, Worthington) in NMDG-CS, 0.2 ml CH2O, 1 ml NMDG-CS) and run through a 10 µm pluriStrainer (cat. Mmul_10/rheMac10, ftp://hgdownload.soe.ucsc.edu/goldenPath/rheMac10/). Since macaque genes differ from those of the human orthologues or putative orthologues, transcript annotations are incomplete on the macaque genome, and many NMDG-CS. Connective tissue and nerve excess were removed before disassociation. 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.LS005318, Worthington) in NMDG-CS, 0.2 ml CH2O, 1 ml NMDG-CS) and run through a 10 µm pluriStrainer (cat. Mmul_10/rheMac10, ftp://hgdownload.soe.ucsc.edu/goldenPath/rheMac10/). Since macaque genes differ from those of the human orthologues or putative orthologues, transcript annotations are incomplete on the macaque genome, and many

scRNA-seq, sequencing, and sequence alignment. WaferGen chips were processed according to the STRT-2i-seq workflow described in ref.77 and FACs sorted cells were processed similarly to the Smart-Seq2 workflow described in ref.78. The resulting WaferGen (samples WG in the text) and Smart-Seq2 (samples S2 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, 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 transcriptome 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 rhesus10 were put in the final set. If either or at 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 (“MiR*”s) in hg19, where both ends were required to match for a name change to be applied. (4) In the few cases where both hg19 transcripts with the same or at end had different gene names, a rhesus10 transcript with the same end was assigned a name which is a combination (“name1/ name2/…” of the alternative names from hg19). When there was a longer or extension in the liftOver model of a human transcript compared with the over- lapping monkey transcript, the 5′ or 3′ exons of the corresponding monkey gene were adjusted accordingly. Supplementary Data 12 lists all 2445 genes for which names from Mmu10 build were replaced with hg19 gene names, as described above. The outcome of the liftOver was raw WG sample data from the following nine clusters detected per neuron—6380, total number of genes detected in at least three neurons—22160, of which 15821 are in human gene nomenclature consortium.

RNAseq, 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 °C until sectioning. Cryosections were cut at 10 μm thickness, the slides were dried at RT, and then stored again at −80 °C to preserve RNA integrity. RNAscope version 2.0 using the RNAscope Fluorescent Multiplex Reagent Kit (Advanced Cell Diagnostics Inc.) was used for in situ hybridization. The staining combinations listed below were used for validations. TrpR8FlH: TRPM8, SCN10A (negative), CBLN2 (negative); PEP3: TRPM8, SCN10A, CBLN2 (negative); A-LTMIR, CBLN2, TRPM8 (negative), SCN10A (negative); NP1: GFRA1, GFRA2, GAL (negative); NP2: GFRA1, GFRA2 (negative); NP3: GFRA1, GFRA2, GAL (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. 4006291), Hs-TRPM8-C3 (#Cat No. 543121), Mmu-GAL-C3 (#Cat No. 88031), Mmu-CBLN2-C2 (#Cat No. 435661), Mmu-CBLN2-C2 (#Cat No. 446051), Hs-GRF-A2 (#Cat No. 553521), and Mmu-PVALB (#Cat No. 461691). The cross-reactive hybridization probe was predicted computationally by Advanced Cell Diagnostics Inc. and verified in macaque tissue by the authors. A Zeiss ISM800 confocal microscope was used to capture images and implemented in Fiji 1.5.2b) was used for image analysis. Lumbar DRG samples from two animals (females, about 5 years old) different from those used for scRNAn sequencing were processed for quantification. Quantification of the expression of 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). Backgrond subtraction was used to identify all neuronal profiles within each visible nucleus in each section. Signal quantification was performed automatically in imageJ on manually outlines 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 dimeter 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 STR-2i-seq replicates. The data were normalized (Normalized-Data) after which 2000 most variable features were selected (FindVaria- bleFeatures). To assign match effects between replicates we used Seurat’s integrated analysis approach that transforms datasets into a shared space using a probe signal threshold as a positive classi- fication of the proportion of the alternative names from hg38. When there was a longer or extension in the liftOver model of a human transcript compared with the overlapping monkey transcript, the 5′ or 3′ exons of the corresponding monkey gene were adjusted accordingly. Supplementary Data 12 lists all 2445 genes for which names from Mmu10 build were replaced with hg19 gene names, as described above. The outcome of the liftOver was raw WG sample data from the following nine clusters detected per neuron—6380, total number of genes detected in at least three neurons—22160, of which 15821 are in human gene nomenclature consortium.

Label transfer between mouse DRG datasets. To visualize mouse DRG scRNA- seq datasets as UMAPs, the individual datasets from refs. 7,9 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, resolu- tion = 0.1) but the cell identities were set as original identities from the pub- lication (FDR). Gene enrichments in >25% of cells in the cluster with FDR < 0.05 and >0.25 average log2-fold-change were considered as marker genes for the cluster.

Conosc analysis. For Conosc41 WG macaque datasets were integrated using CCA space (#BuildGraph(k = 15, k.self = 5, space =CCA, ncomps = 30, n = 1:20, snn = 0.25, snn.F = 0.25)). Mouse scRNA-seq data were processed with the variances standardized using (Embbed- Graph(method = "UMAP", min.dist = 0.1, spread = 20, n.cores = 4, min.prob.low- er = 1e-3). For macaque (WG) and mouse (Zeisel) dataset co-integration was performed as SbuildGraph (k = 15, k.self = 5, space =CCA, ncomps = 30, n = 1:20, snn = 0.25, snn.F = 0.25). Mouse DRG data were downloaded from (http://lloon.innarslab.org/clone/Mousebrain.org/level/). L6_Peripheral_sensory_neurons.loom) and GEO (GSE139088) and the cluster annotation was modified to conform to established nomenclature8, Mouse genename were switched to corresponding human orthologs using biomaRt (v2.42).

Scoring 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 level79,80. We built a vanilla neural-net 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. The cross-type assignment was performed as described in ref. 46. 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 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 final score. Nesterov Momentum was used as a stochastic gradient descent (SGD) optimizer. To choose the appropriate 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 corre- sponding to the maximum point of learning curve reaching the accuracy plateau. 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 validate the significance of the prediction, and the p values were calculated by FDR. Data were visualized using the radial-vizual-networks model 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; species1sgenes, species2sgenes) or in both (intersect; species1genes, species2genes) were listed. A log2FC 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.
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 cluster 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. 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, with pathways sourced from Gene Ontology’s biological process database, obtained December 2019 from URL 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. The first round of standard genotyping quality control was performed by the UK Biobank consortium and is fully documented on their web portal (https://biobank.ctsu.ox.ac.uk/crystal/). For GWAS, no other cohorts with the same recruitment characteristics as the UK Biobank are available from the biobank 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.

Received: 24 July 2020; Accepted: 8 February 2021; Published online: 8 March 2021

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://ermforschgroup.shinyapps.io/macaqueDRG/ and Smart-seq2 dataset (dorsal root ganglia of five adult Rhesus macaques) https://ermforschgroup.shinyapps.io/macaqueSS2/. Macaca mulatta genome build (Mmml_10/rhesMac10) 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/LdPeripheral_sensory_neuronsloom. 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].
References

1. Abeira, V. E. & Ginty, D. D. The sensory neuron touch. *Neuron* 79, 618–639 (2012).
2. Dong, X. & Dong, X. Peripheral and central mechanisms of itch. *Neuron* 98, 482–494 (2018).
3. Lewis, G. R. & Moshourab, R. Mechanosensation and pain. *J. Neurobiol.* 61, 30–44 (2004).
4. Lai, Q. Population coding of somatic sensations. *Neurosci. Bull.* 28, 91–99 (2012).
5. Emery, E. C. & Ermols, 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* 16, 83–102 (2018).
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. Neurosci. 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 sensory properties of human DRG neurons. *J. Neurophysiol.* 113, 5172–5185 (2015).
15. Zhang, X., Priest, B. T., Beller, I. & Gold, M. S. Voltage-gated Na+ currents in human dorsal root ganglion neurons. *elife* 6, e23335 (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. Transcriptional 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 (2019).
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, L., Woods, W. A. & Woods, C. G. Understanding the genetic basis of congenital insensitivity to pain with anhidrosis. *Nat. Genet.* 53, 485–488 (2011).
23. Leipoldt, E. et al. A de novo gain-of-function mutation in SCN11A causes loss of pain perception. *Nat. Genet.* 45, 1399–1404 (2013).
24. Calvo, M. A. et al. The genetics of neuropathic pain from model organisms to clinical application. *Neuron* 104, 637–653 (2019).
25. 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).
26. Zorina-Lichtenwalzer, K., Meloto, C. B., Khoury, S. & Diatchenko, L. Genetic predictors of human chronic pain conditions. *Neuroscience* 338, 36–62 (2016).
27. Calvo, M. A. 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. Yeh, Y. et al. Mutations in SCN9A, encoding a sodium channel alpha subunit, in patients with primary erythralgia. *J. Med. Genet.* 41, 171–174 (2004).
30. Zorina-Lichtenwalzer, K., Meloto, C. B., Khoury, S. & Diatchenko, L. Genetic predictors of human chronic pain conditions. *Neuroscience* 338, 36–62 (2016).
31. Calvo, M. A. et al. The genetics of neuropathic pain from model organisms to clinical application. *Neuron* 104, 637–653 (2019).
32. Venehaam, 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).
33. Meloto, C. B. et al. Human pain genetics database: a resource dedicated to human pain genetics research. *Pain* 159, 749–763 (2018).
34. 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).
80. Manno, G. L. et al. Molecular diversity of midbrain development in mouse.
81. Hovestadt, V. et al. Resolving medulloblastoma cellular architecture by single-
83. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change
84. Fehrmann, R. S. N. et al. Gene expression analysis identi
85. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and ef
86. Marcandalli, J. et al. Induction of potent neutralizing antibody responses by a
designated protein nanoparticle vaccine for respiratory syncytial virus. Cell 176,
87. Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term
88. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still
89. Loh, P.-R. et al. Efficient Bayesian mixed-model analysis increases association
90. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype

does not properly render in preview.

Acknowledgements
We thank Martin Haring for advice on enzymatic treatment/dissociation strategies and
Ming-Dong Zhang for advice on dissection. We also thank Gunilla B. Karlsson Hedestam
for donating eutenized animals, Sten Linnarsson for computational support, and
Eneritz Agire 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.Y.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 R.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: 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 Wooll and the other
anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information
Request reprints from the corresponding author(s) by e-mail.

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