Single nucleus multiomic atlas of human dorsal root ganglia reveals the contribution of non-neuronal cell types to pain

Kevin Boyer[1#], Pauline Meriau[2#], Lite Yang[2,3#], Huma Naz[1], Sarah Rosen[2], George Murray[1], Adam Dourson[3], Juliet Mwirigi[3], John Del Rosario[3], Maria Payne[3], Prashant Gupta[3], Jiwon Yi[3], Richard Slivicki[3], Zachariah Bertels[3], John Lemen[4], Yujie Chen[1], Ting Wang[1], Alexander Chamessian[3], Bryan A. Copits[3], Robert W. Gereau[2,3*], Valeria Cavalli[2*], Guoyan Zhao[1,5,6,*]

1 Department of Genetics, Washington University Pain Center, Washington University School of Medicine, St. Louis, MO, United States

2 Department of Neuroscience, Washington University Pain Center, Washington University School of Medicine, St. Louis, MO, United States

3 Department of Anesthesiology, Washington University Pain Center, Washington University School of Medicine, St. Louis, MO, United States

4 MidAmerica Transplant, St. Louis, MO, United States

5 Department of Neurology, Washington University School of Medicine, St. Louis, MO, United States

6 Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO, United States

# Equal contribution

*Corresponding author

Guoyan Zhao (gzhao@wustl.edu), Valeria Cavalli (cavalli@wustl.edu), Robert W. Gereau (gereaur@wustl.edu)

Abstract

Sensory neurons residing in dorsal root ganglia (DRG) transmit sensory information such as pain, itch, touch, pressure and bodily position to the central nervous system. The activity of sensory neurons is regulated by non-neuronal cell types in the DRG, including satellite glial cells (SGCs) and fibroblasts. Dysregulated gene expression in DRG cells contributes to sensory nervous system disorders such as chronic pain and itch. Understanding the genetic underpinnings of these conditions requires dissecting transcriptional regulation in human DRG (hDRG). In this study, we profiled transcriptomic and chromatin accessibility landscapes from postmortem hDRG samples at single-cell level. We demonstrate that sequencing depth significantly impacts downstream analysis, with deeper sequencing yielding more detected cells and features, improved data integration, refined clustering and annotation, and more accurate scientific interpretations. We identified nine major cell types, defined their molecular signatures, and mapped cis-regulatory landscapes. Integration of gene expression with chromatin accessibility enabled peak-gene association and transcriptional network analyses, revealing transcription factors, their target genes, and their regulatory elements. This approach uncovered cell types, genes, and cis-regulatory regions potentially driving pain conditions. Our unbiased genome-wide analysis confirmed known pain-related genes and highlighted novel candidates. These findings provide new insight into molecular mechanisms and candidate cell types involved in pain. Importantly, our results demonstrate that non-neuronal cell types—including endothelial cells, fibroblasts, macrophages, and SGCs—play critical roles in pain pathogenesis and should be investigated as therapeutic targets.

One Sentence Summary

snMultiomic profiling of human dorsal root ganglia revealed candidate cell types and causal variants underlying pain, underscoring the role of non-neuronal cells.

INTRODUCTION

Vertebrates sense a wide range of external and internal stimuli through functionally distinct peripheral sensory neurons whose cell bodies reside in the DRG. These neurons are tuned to respond to specific physical and chemical stimuli, enabling the perception and discrimination of various sensations, such as touch, itch, pain, temperature, and proprioception(*1*). In addition to neurons, the primary sensory system requires the integrated function of multiple cell types including SGCs, fibroblasts, macrophages, and Schwann cells. In rodent models, cellular and gene expression changes in these cells have been implicated in various painful states(*2-8*) and degenerative diseases(*9-11*). The last three decades have seen significant advances in understanding the electrochemical, cellular and molecular characteristics of neurons and non-neuronal cells in DRG of animal models(*7, 8, 12-14*). Recent advances in single-cell or single-nucleus RNA sequencing technology (sc/snRNA-seq) have greatly expanded our knowledge of DRG cell type composition and gene expression in rodent models(*15-19*). However, less is known about the cellular and molecular characteristics of human DRG (hDRG)(*20*). Recent studies have revealed important similarities and differences in molecular and cellular characteristics of DRG between animal models and human(*17, 21-26*), which have profound implications for translating data from rodent models to human pathologies, and subsequent therapeutic developments. However, how the identity of different cell types in hDRG are regulated and how they may contribute to pain condition development remains not well understood.

Chronic pain, conventionally defined as pain lasting longer than 3 months, is highly prevalent worldwide and represents a significant socioeconomic and public health burden(*27, 28*), contributing to excess mortality(*29, 30*). At present, the mechanisms underlying chronic pain in various conditions are not fully understood, and existing treatments often prove ineffective for most patients(*31-35*). The widespread use of opioids—the most commonly prescribed class of pain medication—has sparked a nationwide crisis. Therefore, a better understanding of the cellular and molecular mechanisms underlying pain conditions may pave the way for the development of novel and safer treatments for many chronic pain conditions. Twin studies have indicated that chronic pain conditions show heritability ranges from 16% to 50%(*36*). Multiple large-scale genome-wide association studies (GWAS) have revealed that chronic pain conditions are heritable, with an estimated genetic contribution account to 10.2 to 45% heritability(*27, 35*). Furthermore, shared heritability and genetic variants across multiple chronic pain conditions have been reported (*35, 37-39*), implicating a possible shared genetic basis of pain conditions. GWAS studies have identified tens to thousands of single nucleotide polymorphisms (SNPs) associated with pain conditions. However, these studies have not directly identified causal variants and/or the genes driving chronic pain conditions. Furthermore, given that over 90% of disease- and trait-associated SNPs are located in the non-coding regions of the genome(*40, 41*), the cell type and the mechanisms through which they affect gene expression in the context of pain conditions have not yet been described.

In this study, we integrated the transcriptomic and epigenetic data from single-cell multiomics analysis of human DRG with pain-associated genes and SNPs from multiple GWAS to infer susceptible cell types and causal SNPs underlying human pain disorders (**Fig 1a**). We analyzed eleven human DRG samples from seven donors (**Supplementary Table S1**), using the 10x Genomics Single Cell Multiome ATAC + Gene Expression Assay, which integrates single nucleus RNA sequencing (snRNA-seq) and single nucleus assay for transposase-accessible chromatin sequencing (snATAC-seq) data from the same cells. The joint profiling of gene expression and chromatin accessibility of the same exact cells is superior at refining cell types and revealing gene regulatory networks, providing the most direct link between cis-regulatory elements and their target genes (*42*). We identified nine major cell types, defined their molecular signatures, and unveiled transcriptional network that regulates major cell type identities. More importantly, we identified cell-type-specific pain-associated genes and cis-regulatory regions potentially driving pain conditions including known pain-related genes and potential novel candidates. Our work provides new insights into the genetics and underlying biology of chronic pain and may help to inform new treatment strategies.

RESULTS

Low abundance of neurons in the human DRG

The DRG comprises many different cell types including neurons, SGCs, myelinating and non-myelinating Schwann cells (mySC and nmSC), fibroblasts, endothelial cells, and immune cells. In fact, most cells in the DRG are non-neuronal, which presents a challenge for analyzing neurons in this tissue via single-cell modalities. Our previous work showed that although the actual percentage of neuronal cells in the mouse DRG is about 12%, the percentage of neurons detected in single cell RNA sequencing (scRNA-seq) and snRNA-seq analyses was only about 1% for mouse and 0.5% for rat samples (*17, 43*). Human DRG also contain more neurons per DRG than animals, estimated to be 70,000 neurons in young human versus 15,000 in rats(*44-46*). In comparison to rodents, human DRGs exhibit increased size and contain a greater number of cells (*20*). In a comparative immunofluorescence of human and mouse DRG sections, we confirmed these differences. DRG sections stained for the neuronal markers (NEFH for human, NeuN for mouse) and for DAPI (all nuclei) revealed that the majority of nuclei in human DRG are non-neuronal cell types (**Fig 1b**). Quantification of the number of neurons showed that neurons represent ~2% of all DRG cells (**Fig. 1c-e**), which is six times lower than the percentage observed in mice(*43*). Additionally, staining for glial cells (GFAP for human, Fabp7 for mouse) and for connective tissue (Collagen IV) demonstrated that human DRGs contain a markedly higher proportion of connective tissue compared to mouse DRGs (**Fig. 1b**), thereby highlighting key structural differences between species (*20*).

Optimization of neuronal nuclei recovery for human DRG snRNA-seq analysis

The size of neuronal soma poses unique challenges in the unbiased characterization of their molecular profiles. As the largest cells in the DRG, neurons in humans have diameters ranging from 20 to 120μm with large

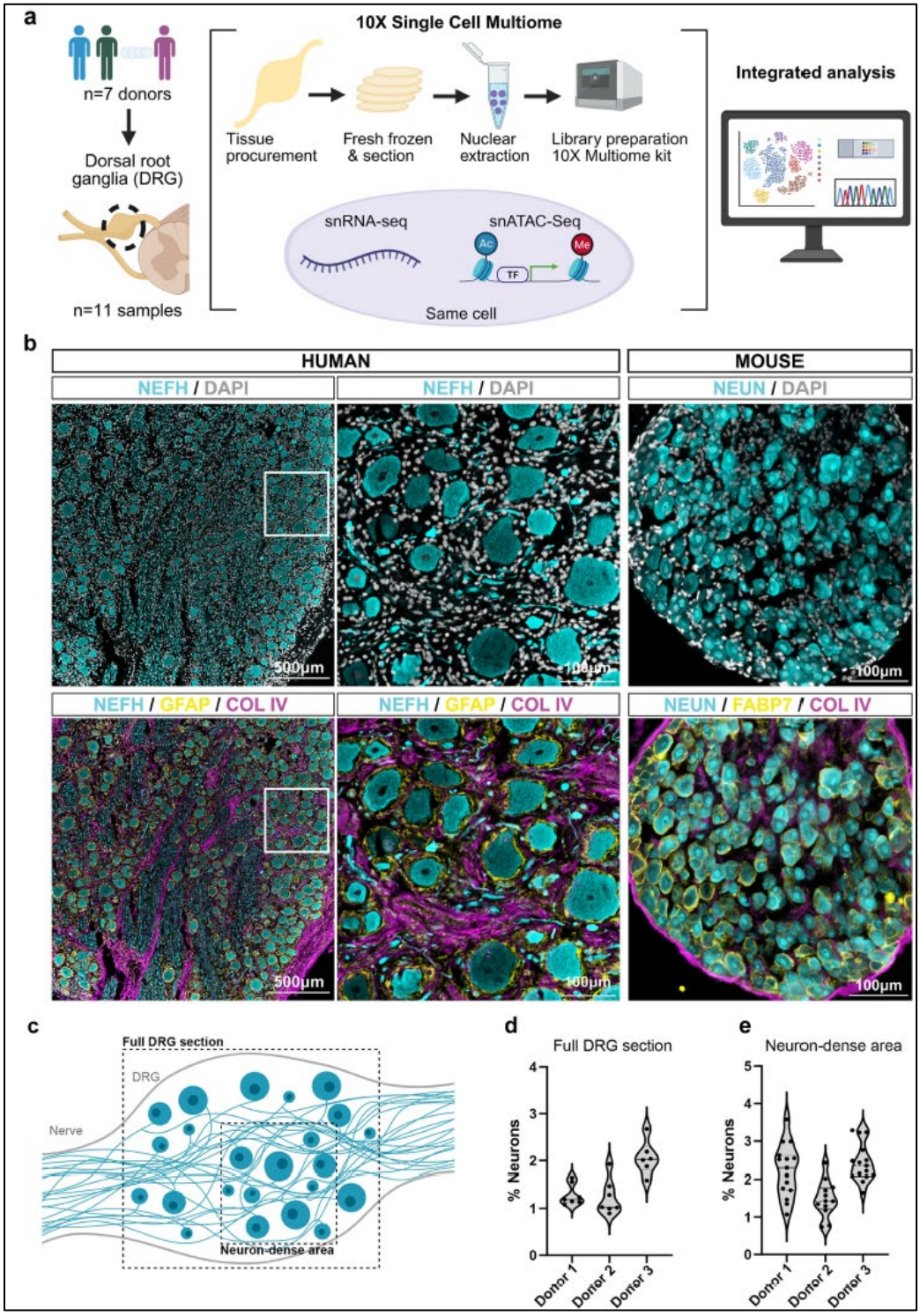


Figure 1: Study design and neuronal population quantification. (**a**) Workflow schematic representation illustrates the processing of hDRG, which enables the concurrent execution of snRNA-Seq and snATAC-Seq on a single cell, utilizing 10X Single Cell Multiome technology. (**b**) Representative images of immunofluorescence staining of human and mouse DRG sections. Neurons (NEFH for humans and NeuN for mice) are labeled in cyan and DAPI is shown in white. The glial cells (GFAP for humans and Fabp7 for mice) are labeled in yellow. The connective tissue (Collagen IV) is labeled in magenta. Scale bar: 500 mm (left panel) and 100 mm (middle and right panel). (**c**) A schematic representation of the distribution of neurons in an hDRG. (**d**) Quantification of the proportion of neurons in a full DRG section, including areas dense with neurons soma, and axon bundles. (**e**) Quantification of the proportion of neurons in neuron-soma-dense areas

nuclei(*44, 47*). Given the 10X Genomics technology, which has a maximum cell/nucleus size of 30 µm during microfluidic-based cell barcoding step, we opted for single nucleus sequencing. To determine potential limiting factors and optimize a protocol that enriches for neuronal nuclei, we first compared multiple protocols with varying extraction and treatment conditions to determine their effect on neuronal nuclei yield. We compared four nuclear dissociation methods previously described(*48*) a nuclear extraction protocol with density gradient centrifugation alone (Opti), Opti followed by FACS (Opti/FACS), a crude extraction protocol followed by FACS (Non-

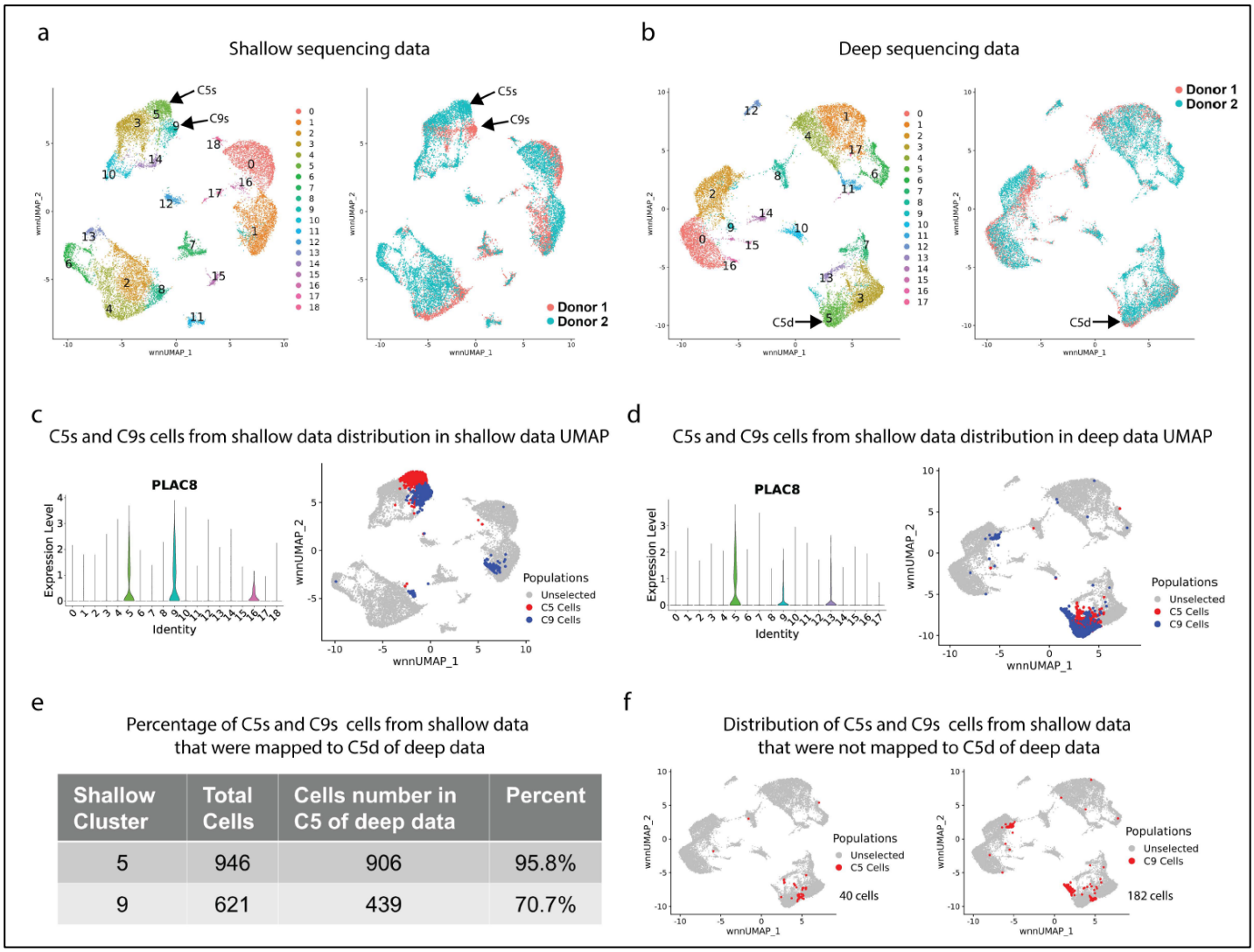
gradient/FACS), and the nuclear isolation kit developed by 10X Genomics (10X kit). We also compared two different douncing conditions and two different detergents (**Supplementary Table S2**). Because the proportion of neurons can differ significantly between donors, tissues sampled from different lumbar levels, and even different sections, we pooled multiple samples from multiple donors to use as starting material. This ensured that any observed differences in the proportion of cell populations are more likely reflecting the differences in cell isolation protocols. To assess neuronal nuclei enrichment, we performed snMultiome-seq on nuclei isolated by each protocol and performed preliminary unsupervised clustering and quality control using single cell genomic tool Seurat(*49*). We identified neuronal cell clusters based on the unique expression of neuronal marker genes *RBFOX1* and *SNAP25* and calculated the percentage of neuronal nuclei in each sample. The nuclear extraction method using gradient centrifugation (Opti-prep) yielded the most enrichment of neuronal nuclei (2.39 ± 1.03%, n=4) with stable performance across different detergents and disassociation methods, whereas non-gradient/FACS, Opti/FACS, and 10X kit protocols yielded 0%, 0.06%, and 0.56% (n=2) neuronal nuclei, respectively (**Supplementary Table S2**). Opti-prep preserved more neuronal nuclei compared to other nuclear extraction approaches and yielded a number of neurons comparable to the neuronal percentage measured in hDRG sections. Therefore, we used this approach to prepare nuclei for snMultiome-seq in this study.

Impact of sequencing depth on cell annotation and data interpretation

Next, we used the optimized Opti-prep protocol on two samples to determine the impact of sequencing depth on cell type annotation and feature detection. To maximize the differences between samples, we chose two hDRG samples (Donor 1:20220418A and Donor 2:20220805A) from different lumbar levels and from two individuals with different pain histories and distinct clinical features (**Supplementary Table S1**). We generated data with a target of 50k reads/cell in an initial shallow sequencing run (referred to as “shallow data”). We next performed additional sequencing with a target of total 150k reads/cell for snRNA-seq data and 100k reads/cell for snATAC-seq data, after combining data from both sequencing runs (referred to as “deep data”). We first compared the mapping statistics and sample quality assessment measurements generated by the 10X Cell Ranger ARC software (**Supplementary Table S3**). We also performed de novo peak calling using Model-based Analysis of ChIP-Seq (MACS2)(*50*) yielding similar number of ATAC-seq peaks. As expected, increased sequencing depth resulted in more cells and more features detected in both transcripts and ATAC peaks for both samples (**Fig. S1**). The sequencing depth had a more significant impact on the median number of features detected for the snATAC-seq than that of the snRNA-seq, doubling the number of features detected for both peak-calling methods. After removing low-quality cells, we obtained 16,888 and 18,416 nuclei for shallow and deep sequencing data respectively. Our analysis thus revealed that increased sequencing depth led to a 9% increase in the number of cells.

To further compare shallow and deep sequencing data, we analyzed these two datasets independently using the same analysis pipeline and applying identical analysis parameters. First, we performed data normalization,

batch effect removal, and integration for snRNA-seq and snATAC-seq data separately using Seurat (v5)(*51*) and Signac (v1.6.0)(*52*). Subsequently, the two single-modality datasets were projected onto a joint neighbor graph using the weighted nearest neighbor (WNN) method which enables the joint analyses of both DNA accessibility and gene expression. The shallow data resulted in 19 total clusters (**Fig. 2a**), while the deep data resulted in 18 cell clusters (**Fig. 2b**). We noticed that in shallow sequencing data, cluster 5 and 9 were largely segregated by donors with Cluster 5 (referred to as C5s) cells largely from donor 1 and C9 (referred to as C9s) cells from donor 2 (**Fig. 2a**). However, both clusters uniquely enriched for *PLAC8* expression (**Fig. 2c**), suggesting similar cell



Shallow Cluster	Total Cells	Cells number in C5 of deep data	Percent
5	946	906	95.8%
9	621	439	70.7%

Figure 2: Impact of sequencing depth on cell annotation and data interpretation. (**a**) For shallow sequencing data, cells were visualized using weighted nearest neighbor (WNN) uniform manifold approximation and projection (wnnUMAP) after quality control. (left) Cells colored by cluster identity; (right) Cells colored by donor identity. (**b**) Plots for deep sequencing data similar to (a). (**c**) Violin plot of PLAC8 expression levels (left) and the distribution of cells from C5s and C9s in shallow sequencing data wnnUMAP (right). (**d**) Violin plot of PLAC8 expression levels (left) and the distribution of cells from C5s and C9s in deep sequencing data wnnUMAP (right). (**e**) A table showing the percentage of C5s and C9s cells from shallow data that were mapped to C5d of deep sequencing data. (**f**) wnnUMAP showing the distribution of C5s (left) and C9s (right) cells from shallow data that were not mapped to C5d of deep data.

types shared by the two donors. We projected the cells from C5 and C9 (referred to as C5s+C9s cells) from the shallow data (**Fig. 2c**) to the UMAP of the deep sequencing data and found that most of the C5s+C9s cells (95.8% of C5s and 70.7% of C9s, mapped to a single cluster of C5 (referred to as C5d) in deep sequencing data (**Fig. 2d,e**), which is uniquely enriched for *PLAC8* expression. Additionally, C5s+C9s cells from the two donors were better aligned in deep sequencing (**Fig. 2b**), supporting that C5s and C9s from shallow data were the corresponding shared C5d cells in the two donors. Furthermore, we noticed that 40 (4.2%) cells from shallow C5s and 182 (~29.3%) cells from shallow C9s cluster were not mapped to the deep C5d cluster. Instead, most of these cells were mapped to the C9d cluster of the deep data, which was distributed across the UMAP space and found to be a doublet cluster (**Fig. 2f**). In summary, this analysis reveals that deep sequencing enabled the detection of more cells, more features, better data integration, better cell clustering and annotation, as well as more accurate downstream analysis and scientific interpretations. Consequently, we performed deep sequencing for the high-quality samples (S1-S7) and four additional samples to a target of 150k reads/cell to assemble a high-quality single-nucleus multiomic atlas of human DRGs (**Supplementary Table S4**).

Single-nucleus multiomic atlas of human DRGs

A total of 97,643 nuclei were obtained from 11 samples derived from 7 donors. Following the implementation of a rigorous quality control protocol, which entailed the elimination of doublets and the removal of low-quality cell clusters, a total of 66,599 high-quality transcriptome profiles were obtained. The data then underwent a series of processing steps, including data normalization, batch effect removal, integration of snRNA-seq and snATAC-seq data and projected data onto a joint neighbor graph using the WNN method as previous published(*53*) (**Fig. 3a**). We delineated 14 cell clusters with distinct gene expression profiles. By leveraging known cell type-specific marker gene expression, we identified nine major cell types including endothelial cells (*FLT1, EMCN, ANO2*)(*54*), fibroblast/mesenchymal cells (*EGFR, APOD, DCN, PDGFRA, COL1A1, FBLN1, EYA1*), macrophages (*SYK, OLR1, PALD1, PLAC8, F13A1, MRC1, LYVE1, SIGLEC1*), neurons (*SNAP25, RBFOX1, ROBO2, SYT1, THY1, TUBB3, PRPH*), myelinating Schwann cells (mySC) (*PRX, DRP2, NCMAP, MPZ*), non-myelinating Schwann cells (nmSC) (*NRXN1, SCN7A, SORCS3, SEMA3E*), satellite glia cells (SGCs) (*SHANK2, KCNJ3, CTNND2, FABP7*), pericytes/smooth muscle cells/mural cells (*PDGFRB, RGS5, MYOCD*), and T cells (*SKAP1, BCL11B, CD2, CD247*) (**Fig. 3a-b, S2a, Supplementary Table S5**)(*17, 55-58*). To assess donor- and sample-based batch effects, we examined the distribution of nuclei across cell clusters for each donor, revealing an overall even distribution across cell type clusters (**Fig. S2b**). The 11 samples were distributed uniformly across cell type clusters, exhibiting equivalent numbers of unique molecular identifiers (UMIs) and detected genes (**Fig. S2c**). SGCs were the most prevalent cell type in our dataset with 21,527 nuclei, followed by macrophages of 16,441 nuclei. In total, we obtained 1,125 neurons with an average percentage of 1.7%, consistent with our antibody staining-based quantification (**Fig 1d,e**). However, the percentage of neurons and other cell types varied significantly across samples (**Fig. S2d**), which could reflect sampling bias of neuron-sparse vs. neuron-dense areas of the DRGs or other donor variable such as age and medical history.

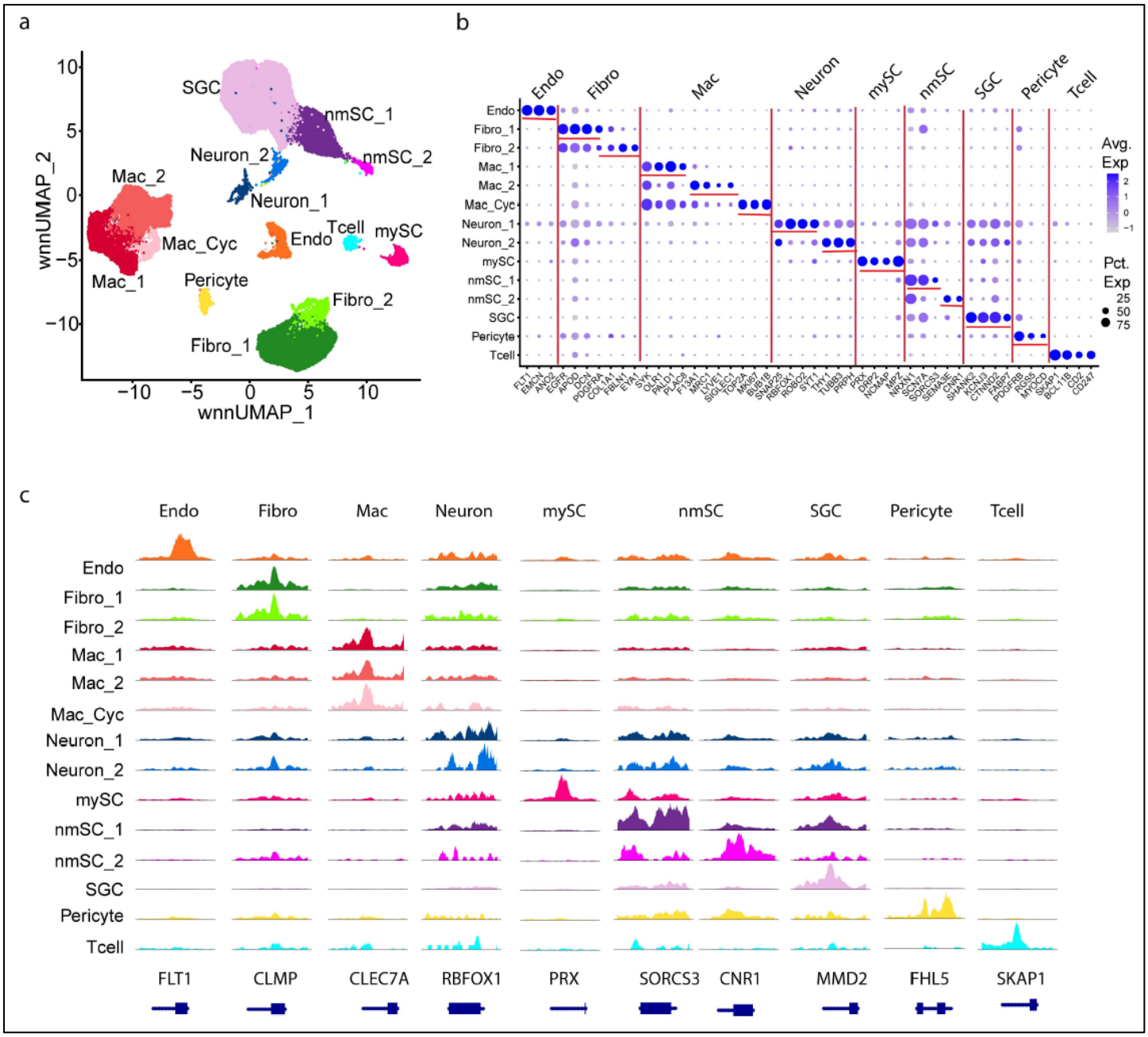


Figure 3. Cell type identification in human DRG of all the cells obtained from 11 samples after quality control. (**a**) Cells were visualized using wnnUMAP colored by cluster identity. (**b**) Dot plot of marker gene expression levels. (**c**) Genome browser views showing the aggregated and normalized snATAC-seq signals at promoter regions (±1 kb from transcription start site) of selected cell-type marker genes. Tracks are colored with the same color code as in (*a)* and (*b)*.

Human DRGs contain up to 16 neuronal subtypes (*21, 24*), and our snRNA-seq transcriptomic profile identified two neuronal populations (Neuron_1, Neuron_2), both expressing pan-neuronal markers *SNAP25*, neurofilaments (*NEFH, NEFL, NEFM*), α- and β-tubulin (*TUBA1A, TUBA1B*, *TUBB3*) (**Fig. 3b, 4a, S3a**) (*20, 24*). Both neuronal populations also expressed known peripheral sensory neuronal markers *SLC17A6* (*VGLUT2*), *SYP* (synaptophysin), and ubiquitin C-terminal hydrolase *UCHL1* (*PGP9.5*)(*24*)(**Fig. S3b**), confirming their

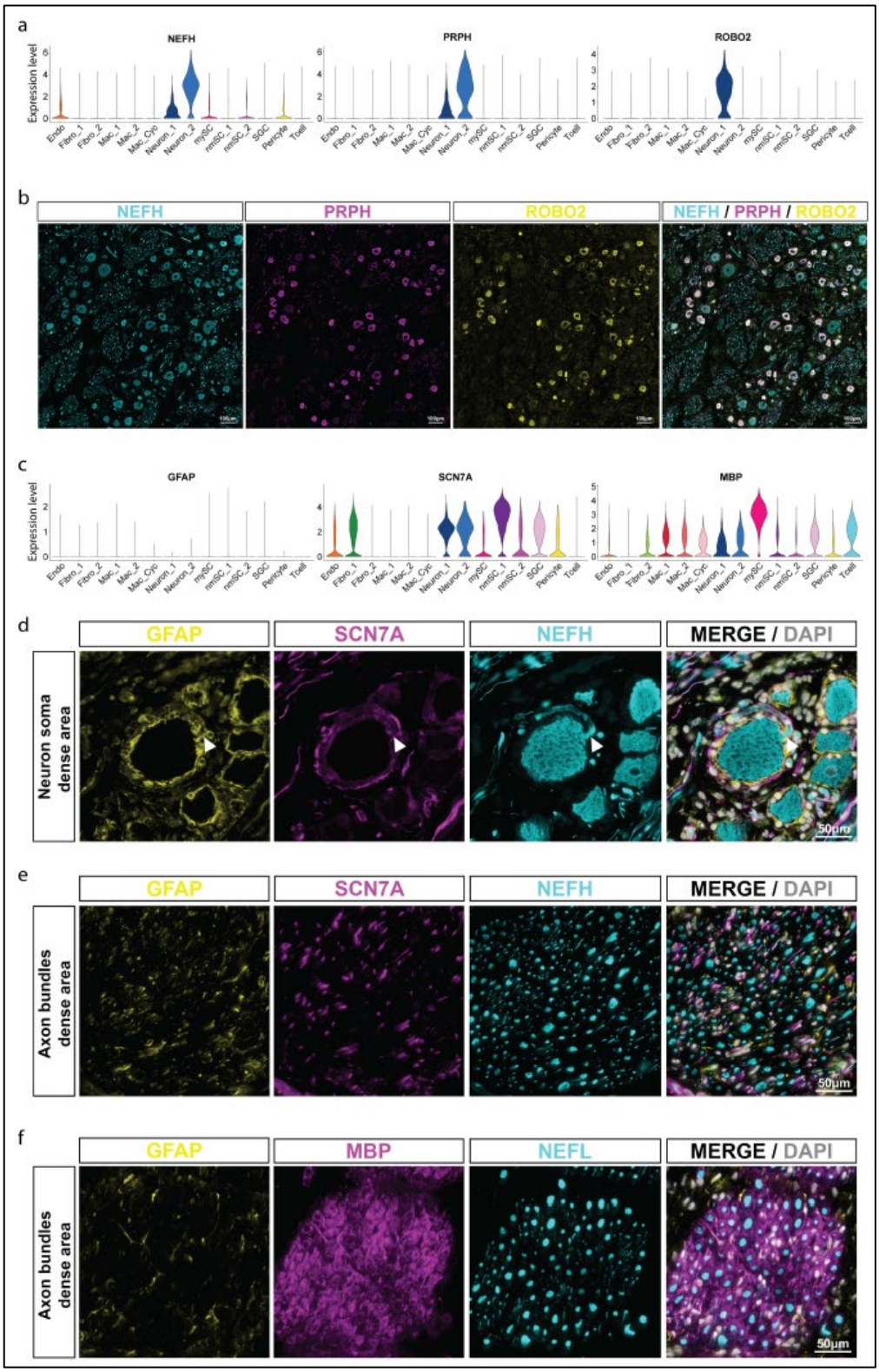


Figure 4: Examination of neuronal and glia cell marker protein expressions in human DRGs. (**a**) Violin-plots showing the expression of NEFH, ROBO2 and PRPH across the different cell populations. (**b**) Representative images of immunofluorescence staining of neurons, labeled for NEFH (cyan), PRPH (magenta) and ROBO2 (yellow). Scale bar: 100 mm. (**c**) Violin-plots showing the expression of GFAP, SCN7A and MBP across the different cell populations. Representative images of immunofluorescence staining of neuron soma dense area (**d**) or axon bundles dense area (**e**) of hDRG, labeled for GFAP (yellow), SCN7A (magenta) NEFH (cyan), and DAPI (gray). Scale bar: 50 mm. (**f**) Representative images of immunofluorescence staining of axon bundles dense area of hDRG, labeled for MBP (yellow), SCN7A (magenta) NEFH (cyan), and DAPI (gray).

sensory neuron identity. Mammalian DRG neurons are broadly classified into two major categories: myelinated large-diameter (A-fiber) and unmyelinated small-diameter (C-fiber) neurons(*59*). Neuron_1 population had higher expression of *RBFOX1, ROBO2,* and *SYT1* as well as unique expression of *CACNA1B, SRRM3, NTRK1, TMEM255A, PRDM12, TRPV1,* and *TRPA1* (**Fig. 3b, 4a, S3c**). Since *Rbfox1* and *Robo2* are enriched in small diameter neurons in rat and mouse DRGs(*60, 61*), Neuron_1 cluster likely included multiple subtypes of

unmyelinated small-diameter C-fibers neurons, since it was enriched for *PRDM12*, a transcriptional regulator critical for the development of human C-fibers and pain-sensing neurons (nociceptors)(*62*), *TRPV1* and *TRPA1,* genes known to be expressed in C-fiber nociceptors in mice required for heat and cold sensation(*1*), and *NTRK1* (TrkA), which is restricted to nociceptive DRG neurons(*63*). In contrast, Neuron_2 exhibited higher expression of neurofilament heavy chain *NEFH* and peripherin (*PRPH*) (**Fig. 4a)**. Given that NEFH, but not PRPH, is enriched in large-diameter A-fiber DRG neurons(*24*), Neuron_2 likely represents a mixture of small and large myelinated neurons. Given the limited resolution of our dataset, we were underpowered to resolve the full heterogeneity of neuronal subtypes (*21, 24*) and therefore classified them into two broad populations.

We next examined expression of markers for the two neuronal populations at the protein level, given that correlations between mRNA and protein expression are often weak or not always consistent (*64-67*). Immunofluorescence staining with antibodies against NEFH and PRPH (transcriptionally enriched in Neuron_2) and ROBO2 (transcriptionally enriched in Neuron_1) revealed striking discrepancies between protein and transcript expression. Although all neurons displayed comparable NEFH protein levels, consistent with prior report (*68*), PRPH protein distribution showed the opposite pattern from the transcriptomic data. Specifically, PRPH protein was found to be present in smaller neurons (likely C-fibers) but absent from larger neurons. Moreover, ROBO2 protein localized exclusively to PRPH-positive, smaller-diameter neurons (**Fig. 4a, b)**. This stark difference between the levels of mRNAs and proteins suggests the presence of additional post-transcriptional regulatory mechanisms at play. These findings underscore the importance of validating transcriptomic data with protein-level analysis for accurate cell population characterization.

Next, we investigated protein expression in the glial cells and their spatial relationship with sensory neurons. SGCs in humans express the markers Glial fibrillary acidic protein (GFAP) and Glutamine synthetase(*11*), although GFAP is expressed a low level in normal condition and is upregulated in SGCs under pathological conditions(*68*). We observed that GFAP protein was strongly enriched in SGCs, which formed a characteristic circular pattern around neuronal soma, even though no *GFAP* transcripts were detected (**Fig. 4c, d).** These observations are consistent with our previous studies(*17*). We noted that some of the GFAP positive SGCs also extended to the proximal axon segment, thereby covering both the soma and axon initial segment (**Fig 4d**). In close contact with the SGCs, we identified a second glial population, nmSCs characterized by the expression of SCN7A. nmSCs were positioned around the soma, forming a second larger ring, and also extended along the axon (**Fig 4d**). In the axon rich areas of the DRG, we observed both nmSCs (SCN7A positive) and mySCs, marked by MBP expression (**Fig. 4e, f)**, consistent with their role in ensheathing axons. Together, these findings demonstrate a layered organization of three distinct glial cell types around sensory neurons, with SGCs forming the innermost ring, nmSCs establishing close contact with both SGCs and axons, and mySCs localized to distal axonal compartments.

We then examined the immune cell population, which are known to contribute to pain, repair, and also monitor the vasculature in the DRG(*69*), at least in rodents. We identified three subpopulations of macrophage, all of which had highly enriched expression of Spleen Associated Tyrosine Kinase (*SYK*) (**Fig 3b**), a critical regulator of macrophage-mediated inflammatory responses(*56*). Mac_2 also had a highly unique expression of *F13A1, MRC1, LYVE1,* and *SIGLEC1*, genes characteristic of brain perivascular macrophages(*70*). The lymphatic vessel endothelial hyaluronan receptor 1 (*LYVE1*) - expressing macrophages have been shown to coat murine and human blood vessels harboring smooth muscle cells across many tissues and maintaining arterial tone through hyaluronan-mediated regulation of smooth muscle cell collagen(*71*). Cycling macrophage (Mac_Cyc) was named because of the unique expression of multiple proliferation markers such as *TOP2A* and *MKI67* and there is evidence in rodents that macrophages in DRG have the capacity to self-renew (*72*).

Mapping and characterization of human DRG open chromatin landscapes

To define the gene-regulatory programs underlying the identity and function of each DRG cell type, we identified the open chromatin regions in each of the 14 DRG cell types. Since peak calling based on the annotated cell clusters enables the identification of more cell-type-specific peaks(*53, 73*), we therefore aggregated the chromatin accessibility profiles from the nuclei comprising each cell cluster and identified the cell-type-specific open chromatin regions with MACS2 (*50*) using a previously reported approach(*74*). This resulted in a uniform peak set of 501 bp which were used for all the downstream analysis. We detected a total of 278,208 open chromatin regions (referred to as OCRs) across all 14 cell types (**Fig. S4a**). SGC had the most OCRs. We detected very few OCRs for the Neuron_2 clusters likely due to the small number of but highly heterogenous neuronal cells. Of these OCRs, 15.75% were located at annotated promoter regions, 48.75% in introns, 25.75% in the distal intergenic regions (**Fig. S4b**). The OCRs for canonical marker genes are consistent with the marker gene expression of major cell-type annotations (**Fig. 3c**). Each cell type was characterized by many specifically expressed genes and accessible OCRs (**Fig. S4c–d**).

Transcriptional regulatory network governing human DRG cell type identities

Genomic regulatory elements such as enhancers, most of which lie in non-coding regions of the genome, control gene expression in a cell-type and tissue-specific manner. To investigate transcriptional regulatory mechanisms underlying cell-type–specific gene expression in the human DRG, we applied the computational framework FigR(*75*) to link distal cis-regulatory elements with target genes and infer gene-regulatory networks (GRNs) governing cell identity and function (**Fig. 5a**). Because chromatin accessibility and gene expression were simultaneously profiled in the same cells, we performed an integrative analysis using both snATAC-seq and snRNA-seq data. FigR was first used to identify significant distal peak-to-gene expression interactions by correlating accessibility of peaks within a fixed 100 kb window around each transcription start site (TSS) with expression of that gene, followed by permutation-based testing. We identified a total of 40,607 unique chromatin accessibility peaks significantly associated with gene expression (permutation $p \leq 0.05$), spanning 12,401 genes

within 100 kb of each other. Applying the default parameters of n ≥ 7 significant peak-gene associations per gene, we identified genes with a high density of peak-gene interactions, referred to as “domains of regulatory chromatin (DORCs)”(*75*). In total, we found 1,532 DORC genes associated with 14,548 total peaks (**Fig. 5b, Supplementary Table S6**).

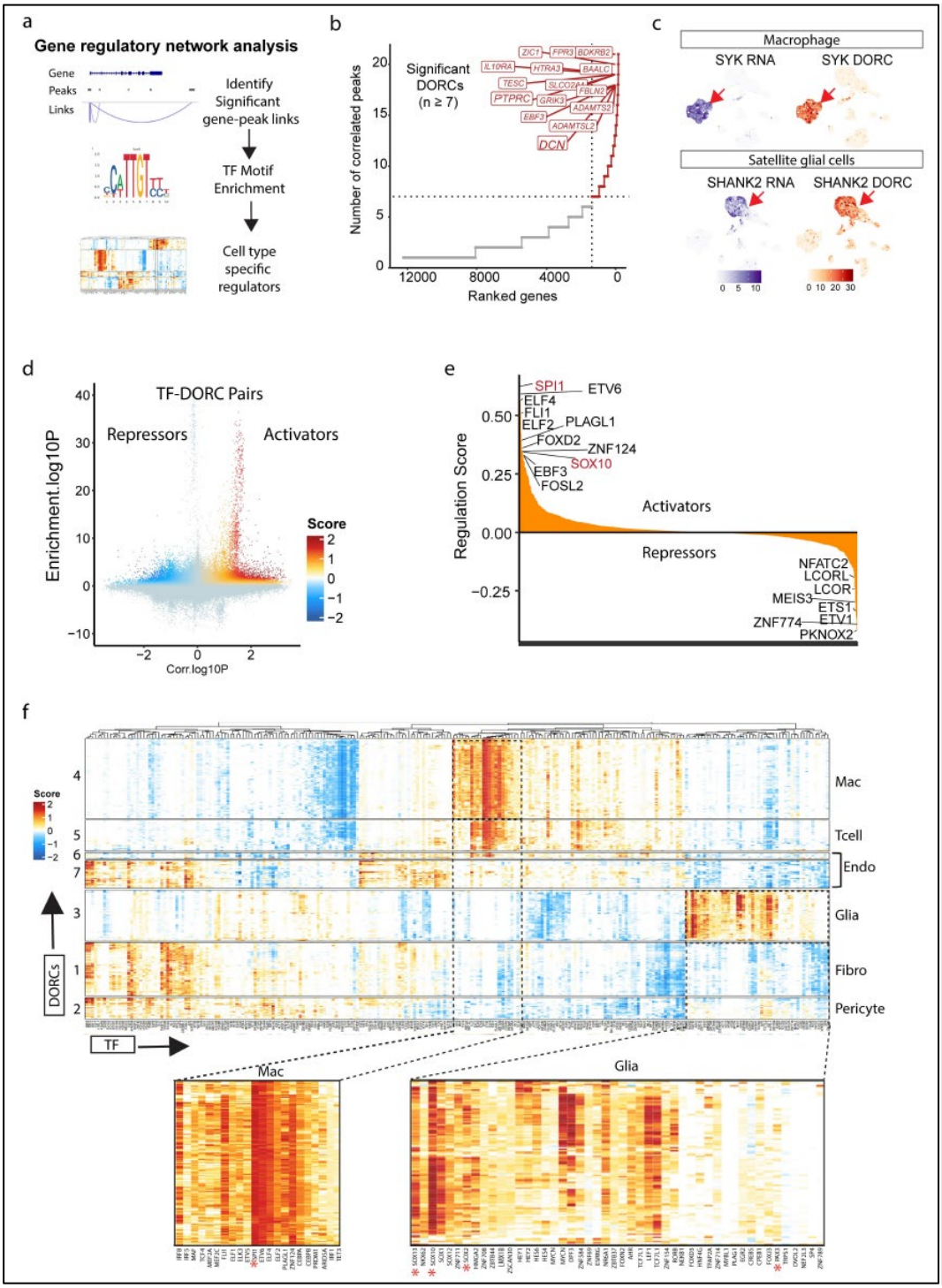


Figure 5: Integrated transcriptional regulatory network analysis identifies key regulatory modules for each cell type. (**a**) Schematic of transcriptional regulatory network analysis. (**b**) Significant DORCs identified by FigR. (**c**) UMAP of DORC accessibility scores (right) and paired RNA expression (left) for macrophage marker gene SYK and SGC marker SHANK2. (**d**) Scatterplot showing all DORC-to-TF associations, colored by the signed regulation score. (**e**) Mean regulation score (signed, −log10 scale) across all DORCs (n = 1,532) per TF, highlighting select TF activators (left skewed) versus TF repressors (right skewed). (**f**) Consensus k-mean clustering heatmap of DORC regulation scores for all significant DORC-to-TF associations. Rows indicate DORCs. Columns indicate TFs.

The list of DORC genes encompassed many known and novel cell-type specific marker genes. These included *FLT1* and *ERG* for endothelial cells; *DCN, APOD, EGFR, PDGFRA, FBLN1* for fibroblasts; *SYK, PTPRC,*

BCL11B, CD2, CSF1R, MRC1, F13A1, LYVE1, and *PLAC8* for immune cells; NCMAP for mySC, *SCN7A, SEMA3E, and SORCS3* for nmSC; *MYOCD, RGS5* for pericytes; as well as *CTTND2* and *SHANK2* for SGC (**Fig 5c, S5**). The strong correspondence between DORC accessibility scores and RNA expression levels in the same cell populations suggest that FigR successfully captured the genes and regulatory sequences defining each cell type.

Next, we applied FigR to identify putative transcription factors (TFs) regulators — both activators and repressors — of DORCs. Specifically, we examined the enrichment of different TF binding motifs within DORCs and the correlation of TF expression levels with the overall accessibility level of each DORC gene (DORC score). FigR integrates these two dimensions to compute a signed probability score called the "regulation score" (on a −log10 scale), which represents the regulatory activities of motif-enriched and RNA-correlated TFs. Positive values indicate activators and negative values indicate repressors. This analysis generated 526,084 TF-DORC associations (all tested TFs across all tested DORCs), with each point representing one TF-DORC pair (**Fig. 5d**). We identified 871 significant regulators across all cells for the given DORCs by filtering TF-DORC associations using a threshold of absolute (regulation score) ≥ 1(*76*) (**Supplementary Table S7)**. Ranking TFs by their mean regulation score across DORCs, among the top ranked activators were well-established cell identity regulators, such as *SPI1* (PU.1) for macrophage (*77*) and *SOX10* which is required for specification of all PNS glia (*78*) (**Fig. 5e**).

To uncover cooperative TFs networks, we performed consensus k-means clustering on the TF-DORC regulation scores. This analysis grouped TFs (column) that may coordinately regulate modules of DORCs (rows) (**Fig. 5f).** Notably, we identified transcription regulatory networks underlying cell-type-specific gene signature modules, including glia (module 3 :*CTNND2, SHANK2, SORCS3*), macrophage (module 4: *SYK, CLEC7A, C3, IRF8, PLAC8*), T cells (module 5: *PTPRC, CD69, STAT4*), endothelial cells (module 6 and 7: *FLT1*), and fibroblasts (module 1: *DCN, APOD, FBLN1, EGFR*), and pericytes (module 2: *EDNRA, NOTCH3*) (**Supplementary Table S8)**. The predicted TFs for each cell-type-specific gene module aligned well with established biology. For example, *SPI1* (PU.1), *JUNB, IRF8, RUNX1, MEF2A, and MEF2C* are key regulator of macrophage identity(*77*), while *SOX10, SOX2*, *SOX5, SOX6,* and *PAX3* play critical roles in glia development and specification (*78*) .

SOX10 regulatory network in the adult peripheral glial cells

Our gene regulatory network analysis identified *SOX10* as a top transcription activator regulating glial signature gene modules (**Fig. 5e, f**), with predicted binding to 103 DORCs (**Supplementary Table S9)**. These predicted targets included established glial markers and genes important for glial development and function in peripheral nervous system (PNS), such as MBP, *SORCS3, SHANK2, SOX1, SOX2, CTNND2, and PAX3*(*79-83*). To understand cell-type-specific functions of SOX10, we examined SOX10 expression and regulatory activity in individual cell types. While *SOX10* was expressed in neurons, mySC, nmSC, and SGC in the snRNA-seq data

(**Fig. 6a**), TF foot printing analysis of SOX10 binding sites (**Fig. 6b**) revealed SGC, nmSC_1, and nmSC_2 as the top three cell populations, suggesting that SOX10 exhibit strong regulatory activities in glial cells (**Fig. 6c**). To validate cell-type specificity, we performed immunofluorescence staining of SOX10 co-labeled with GFAP for SGCs, SCN7A for nmSC and MBP for mySC. SOX10 is a transcription factor and was primarily located in the nucleus as expected. SOX10-positive nuclei were distributed across the DRG, localizing both around neuronal soma and within axon-rich regions of the DRG (**Fig. 6d, e)**. Around neuronal soma, SOX10 nuclei colocalized with *GFAP* in SGCs and with SCN7A in nmSC (**Fig. 6d)**. In axon-rich areas, SOX10 colocalized with MBP in mySCs and with SCN7A in nmSC (**Fig. 6e)**. SOX10 has been shown to directly activate downstream target gene *MBP* expression(*84*) as we have predicted (**Fig. 6f**). Notably, SOX10 protein was absent from neurons, despite

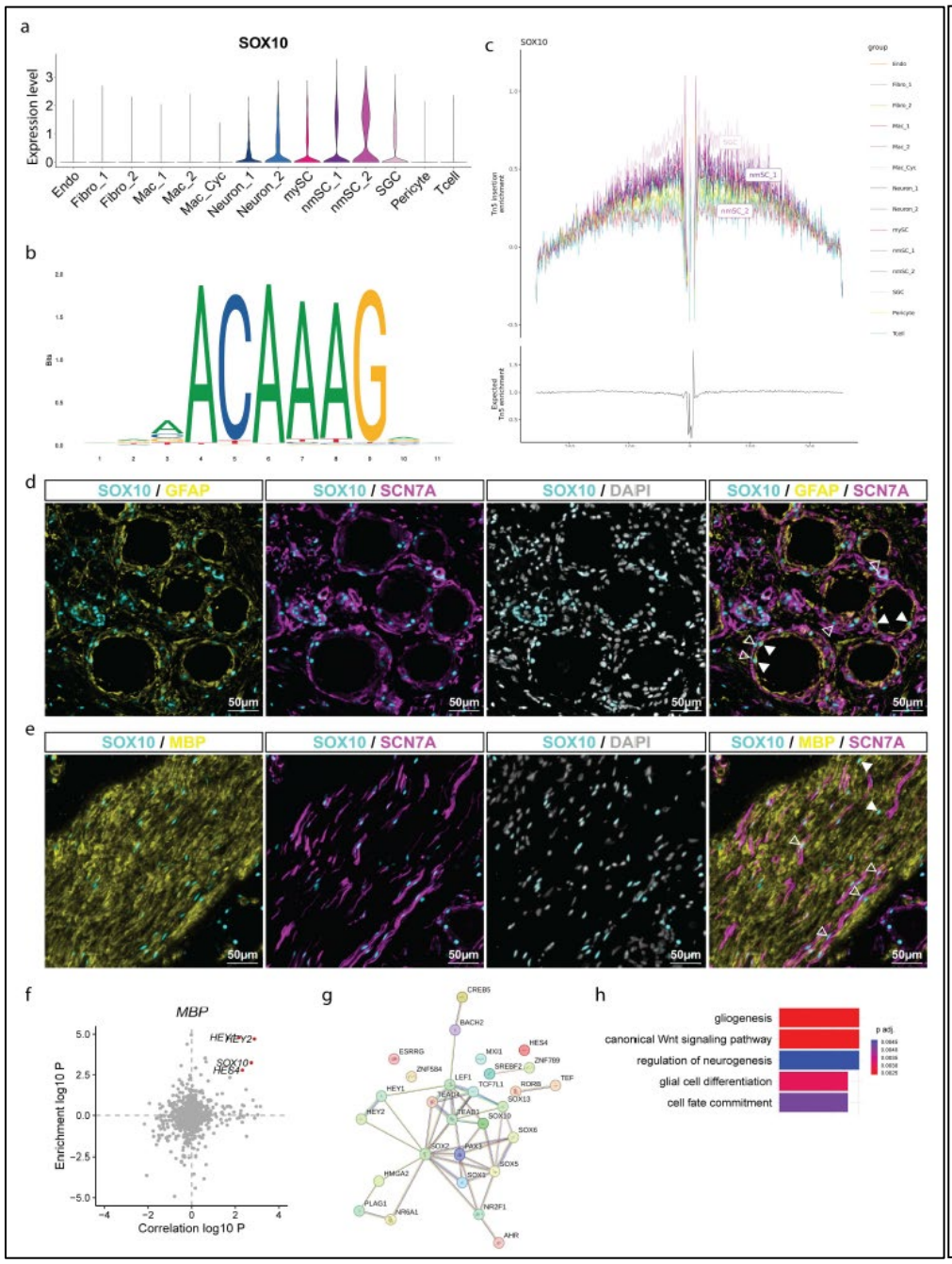


Figure 6: SOX10 regulatory network. (**a**) Violin-plots showing the expression of *SOX10* across the different cell populations. (**b**) Sequence logo for DNA sequence motif of SOX10. (**c**) TF footprinting analysis for SOX10 motifs sites. (**d, e**) Representative images of immunofluorescence staining of neuron soma dense area (d) and axon bundles dense area (e) of hDRG, labeled for SOX10 (cyan), SCN7A (magenta), DAPI (gray), and GFAP (d) or MBP (e) in yellow. (**d**) Solid arrows indicate SOX10-positive nuclei in GFAP-enriched cells. Hollow arrows indicate SOX10-positive nuclei in SCN7A-enriched cells. (**e**) Solid arrows indicate SOX10-positive nuclei in MBP-enriched cells. Hollow arrows indicate SOX10-positive nuclei in SCN7A-enriched cells. (**f**) Candidate TF regulators of MPB highlighted in red, with all other TFs shown in gray. (**g**) Protein-protein interaction network for the candidate TFs that regulate glia DORCs. (**h**) Gene ontology terms enriched in the candidate TFs that regulate glia DORCs. Scale bar: 50 μm.

detectable transcripts in the snRNA-seq data (**Fig. 6a**). This represents another example of a discrepancy between transcript and protein levels.

Our GRN analysis revealed that binding sites for many TFs were concordantly enriched in the glial DORC modules (**Fig. 5f**), suggesting cooperative regulation of glial identity and function. These included multiple high mobility group (HMG) domain transcription factors—*SOX10, SOX2*, *SOX5, SOX6,* and *PAX3*, which can interact with each other through protein-protein interaction as revealed by the String database (**Fig. 6g**). Sox10 and Pax3 physically interact with each other to activate c-RET in neural crest cells, which is critical for enteric ganglia development(*85*). Similarly, Sox10, Sox5 and Sox6 act synergistically to regulate oligodendrocyte maturation in the central nervous system (*86, 87*). Consistent with their role in glia fate specification during development in the CNS, these TFs were significantly enriched for pathways including “Cell fate commitment” and “Glial cell differentiation” (**Fig. 6h**). Our analysis suggests that these TFs may cooperate with each other and play critical roles in defining glial cell identity and function in the adult PNS.

Similarly, SPI1 was identified as the top-ranked transcription activator in macrophages (**Fig. 5e, Supplementary Table S7**). SPI1 expression was restricted to macrophage populations (**Fig. S6a**), and TF foot printing analysis identified Mac_1, Mac_2, and Mac_Cyc as the top cell populations most enriched for SPI1 binding sites (**Fig. S6b**). We identified 241 putative SPI1 target DORCs, including common macrophage marker genes such as PTPRC, *CLEC7A*, *CSF1R*, *F13A1*, *IRF8*, *PLAC8*, *SYK*, and *C3* (**Fig. S6a**, **Supplementary Table S10**), consistent with SPI1's established role in macrophage cell fate regulation (*88, 89*). IRF8 has been shown to be critical for transforming microglia (the resident macrophage of the brain and spinal cord) into a reactive phenotype, which is associated with neuropathic pain after peripheral nerve injury(*90*). Furthermore, SPI1 and IRF8 together regulate cell type-specific expression of NLRP3 inflammasome in human macrophages(*91*), consistent with our GRN analysis that binding sites for these two TFs concordantly enriched in the macrophage DORC modules (Fig. 5f).

Defining cell types and putative regulatory mechanisms critical for pain conditions

Genome-wide association studies (GWAS) have successfully identified many disease-associated loci. However, most GWAS studies are performed on bulk tissue samples —representing complex mixtures of cell types, or tissues not directly relevant to the disease of interest. On the other hand, disease-causal variants often act in specific tissues and cell types. Therefore, the causal variants, their target genes, and the relevant cell types largely remain unsovled. Single-cell multiomic datasets now offer a powerful opportunity to predict disease-relevant cell types and pinpoint causal variants by jointly examining gene expression and chromatin accessibility. To address these challenges, we first curated a comprehensive set of pain-associated SNPs and genes. In total, we collected 6,212 SNPs across seven published GWAS studies of migraine and pain conditions, along with their associated genes (**Supplementary Table 11**). The following studies were included: 1) Gormley et al.(*92*)

identified 38 migraine susceptibility loci (referred to as Migraine Gormley) through a meta-analysis of 375,000 individuals from the International Headache Genetics Consortium. 2) Johnston et al.(*27*) identified 39 SNPs associated with multisite chronic pain in the UK Biobank. 3) Hautakangas et al. (*93*) performed a GWAS of 102,084 migraine cases and identified 123 risk loci and subtype-specific risk alleles. 4) Khoury et al. (*94*) reported 983 and 43 genetic determinants of chronic single-site and multisite pain respectively, in the UK Biobank. 5) Toikumo et al. conducted a cross-ancestry meta-analysis of pain intensity in 598,339 participants in the Million Veteran Program, identifing 126 independent genetic loci, including 69 novel ones. 6) Zorina-Lichtenwalter et al. (*35*) identified 315 common genetic factors shared across 24 distinct pain conditions in over 400,000 individuals in the UK Biobank participants. 7) Wistrom et al. (*95*) compiled 242 genes linked to pain-associated behaviors through literature search spanning over three decades. 8) Lastly, we included 28 genes associated with hereditary sensory and autonomic neuropathies (HSAN) provided at the HEREDITARY SENSORY & AUTONOMIC NEUROPATHIES (HSAN) website (https://neuromuscular.wustl.edu/time/hsn.htm).

We then used snRNA-seq data to identify disease-relevant cell types in the human DRG. Using cell cluster-specific marker genes, we performed hypergeometric tests to identify cell populations that were significantly enriched for pain-associated gene sets (**Fig. 7a**). Neuron_1 marker genes were significantly enriched across nearly all pain conditions (except migraine), consistent with the finding that Neuron_1 includes nociceptor populations. Interestingly, Mac_1 marker genes were significantly enriched for all pain conditions, highlighting the critical role of macrophages in inflammation and pain processing (*6, 96*). Fibroblast, pericyte, endothelial cell and glial cell marker genes were also enriched in subsets of pain conditions.

Next, we used snATAC-seq data to identify putative causal variants through a disease SNP prioritization framework (Mangetti-Goncalves et. al. unpublished) This framework integrates cis-regulatory modules (CRMs) predicted by the Vertebrate Regulatory MOdule Detector (**VRMOD**) with ten orthogonal human datasets of experimentally defined cis-regulatory modules as well as genomic and epigenomic datasets (details in Method). Gene transcription is regulated by TFs binding to cognate transcription factor binding sites (**TFBSs**), which often cluster into **CRMs**, to ensure precise spatiotemporal gene expression(*97, 98*). Importantly, GWAS-reported SNPs are usually tag-SNPs—representative variants that capture the signal of nearby correlated SNPs through linkage disequilibrium. Because tag-SNPs may not be causal, it is essential to prioritize functional non-coding variants located within regulatory modules. VRMOD accurately predicts functional CRMs directly from genomic sequences alone. CRMs supported by both VRMOD predictions and orthogonal experimental datasets represent the strongest candidates for harboring transcriptional regulatory activities. For integration, we included ten human experimental datasets including experimentally defined regulatory modules as well as genomic and epigenomic datasets representing putative regulatory sequences from six sources: 1) 60 experimentally defined human CRMs curated from literature; 2) 998 enhancers from VISTA Enhancer Browser defined by transgenic

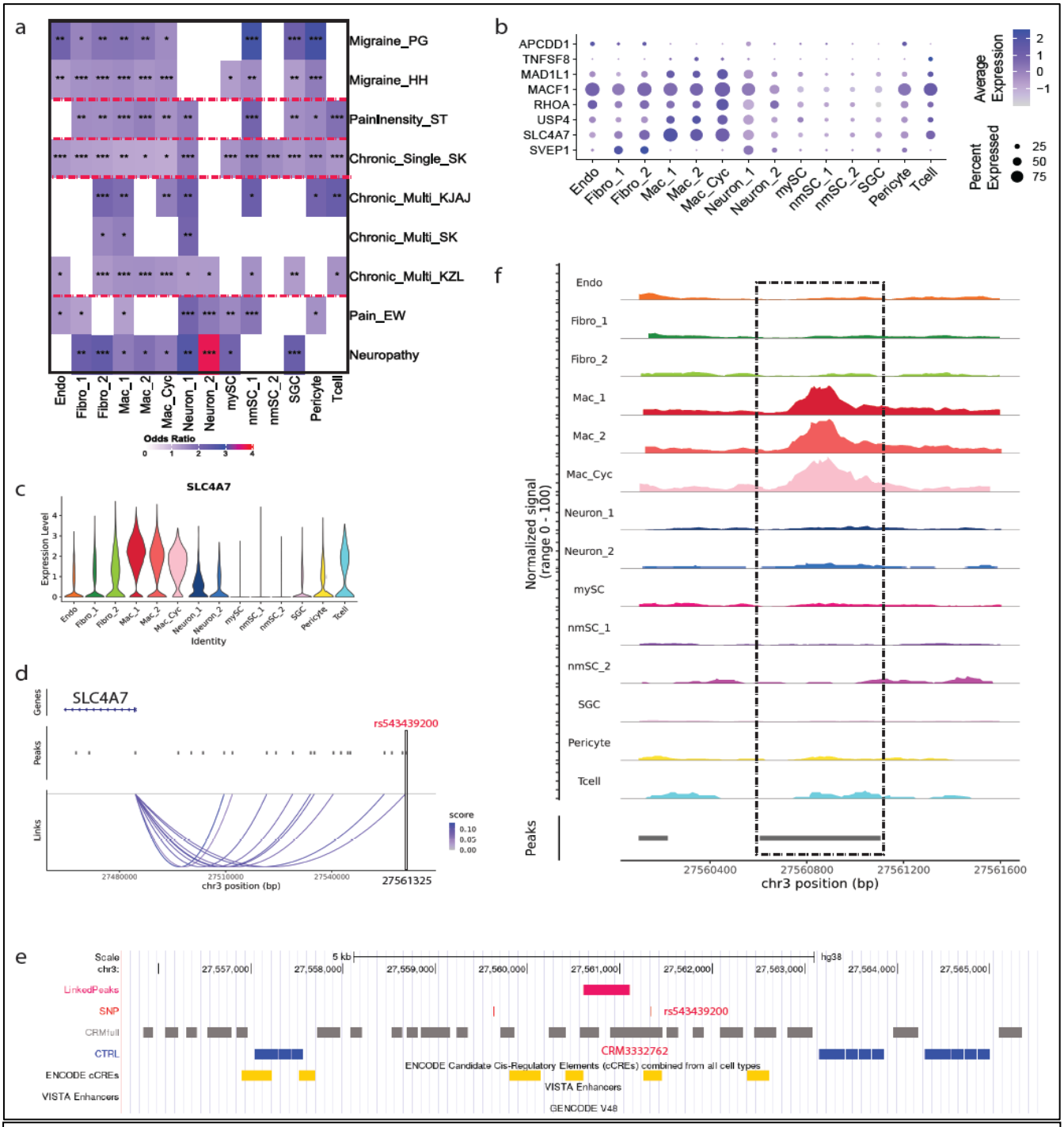


Figure 7: Defining cell types and putative regulatory mechanisms critical for pain conditions. (**a**) Heatmap showing cell types whose marker genes significantly enriched for pain-associated genes. (**b**) Dot plot showing relative expression of all pain-associated genes significantly linked to snATAC-seq peaks. (**c**) Violin-plot showing the expression levels of SLC4A7 across all cell populations. (**d**) peak–gene links for SLC4A7 and the location of pain-associated SNP rs543439200. (**e**) Genome view of the snATAC-seq peak significantly associated with SLC4A7, the location of pain-associated SNP rs543439200, VRMOD predicted CRM3332762, and ENCODE cCREs. (**f**) snATAC-seq peak signals.

mouse experiments(*99*). 3) Predicted enhancers from the Enhancer Atlas 2.0.; 4) Chromatin accessibility (ATAC-

seq and DNAse-seq) peaks and histone modification ChIP-seq peaks (H3K27ac, H3K4me1, and H3K4me3) from the Cistrome database; 5) ENCODE candidate cis-regulatory elements (cCREs)(*100*); and 6) scATAC-seq peaks from 45 adult and fetal human tissues(*101*). By integrating GWAS pain-associated SNPs with our single cell ATAC-seq data through this framework, we prioritized genomic regions most likely to harbor true functional non-coding variants that are physically linked to GWAS tag-SNPs within the same CRM.

First, we identified VRMOD-predicted CRMs that contained at least one pain-associated SNP (hereafter referred to as **pain-associated CRMs**). We then linked these CRMs to potential target genes by overlapping them with peaks significantly associated with genes located within 100kb of a transcription start site (TSS), based on the genome wide peak-gene linkage analysis described above. In total, we found eleven peaks that overlapped (≥ 1bp) with ten pain-associated CRMs. These peaks were linked to eight unique genes: *SVEP1, SLC4A7, USP4, RHOA, MACF1, MAD1L1, TNFSF8, APCDD1* (**Fig. 7b, Supplementary Table 12**). Although these genes were expressed across all cell types, they were generally de-enriched in glial cells, with some showing cell-type-specific enrichment. Notably, *SLC4A7* was strongly enriched in macrophages (**Fig. 7b, c**). *SLC4A7* encodes the electroneutral sodium/HCO3 cotransporter NBCn1, which plays important physiological and pathophysiological roles in diverse cell types. *SLC4A7* was identified as a gene associated with a single-site chronic pain SNP reported in Khoury et. al(*94*). Located on human chromosome 3, *SLC4A7* had 12 peaks significantly associated with it within 100 kb of its TSS (**Fig. 7d**), suggesting complex transcriptional regulation. Among these, we identified a VRMOD predicted module, CRM3332762, that encompassed the pain-associated SNP rs543539200 and overlapped with both an ATAC-seq peak (chr3:27560605-27561105) and an ENCODE cCRE (**Fig. 7e**). These lines of evidence support its role as a regulatory element modulating *SLC4A7* expression and potentially contributing to pain conditions. Interestingly, the ATAC-seq peak chr3:27560605-27561105 overlapping CRM3332762 was unique to macrophages, further suggesting that SNP rs543539200 may regulate *SLC4A7* expression in a macrophage-specific manner.

Both Ras homolog gene family member A (*RHOA*) and Ubiquitin-specific protease 4 (*USP4*) were associated with pain-linked SNP rs4955440, a multi-site chronic pain SNP (*94*) located between the two genes. Both genes were broadly expressed across nearly all cell types, with varying expression levels, except non-myelinating Schwann cells (**Fig. 8a,b**). *RHOA* encodes a small GTPase of the Ras homolog (Rho)-family that regulates cytoskeleton dynamics and cell adhesion. *USP4*, a member of the largest subfamily of cysteine protease deubiquitinating enzymes (DUBs), removes mono- and polyubiquitinated chains from target proteins. We identified two significant peaks associated with *RHOA* and three with *USP4* within 100kb of their respective TSS (**Fig. 8c**). The pain-associated SNP rs4955440 was located within VRMOD-predicted CRM3377162, which overlapped with an snATAC-seq peak (chr3:49350847-49351347), and an ENCODE cCRE (**Fig. 8d**), supporting its putative regulatory function. Rho-associated coiled-coil-containing protein kinase (ROCK) is the primary downstream effector of RhoA(*102*). Dysregulated activation of the RhoA/ROCK signaling pathway has been

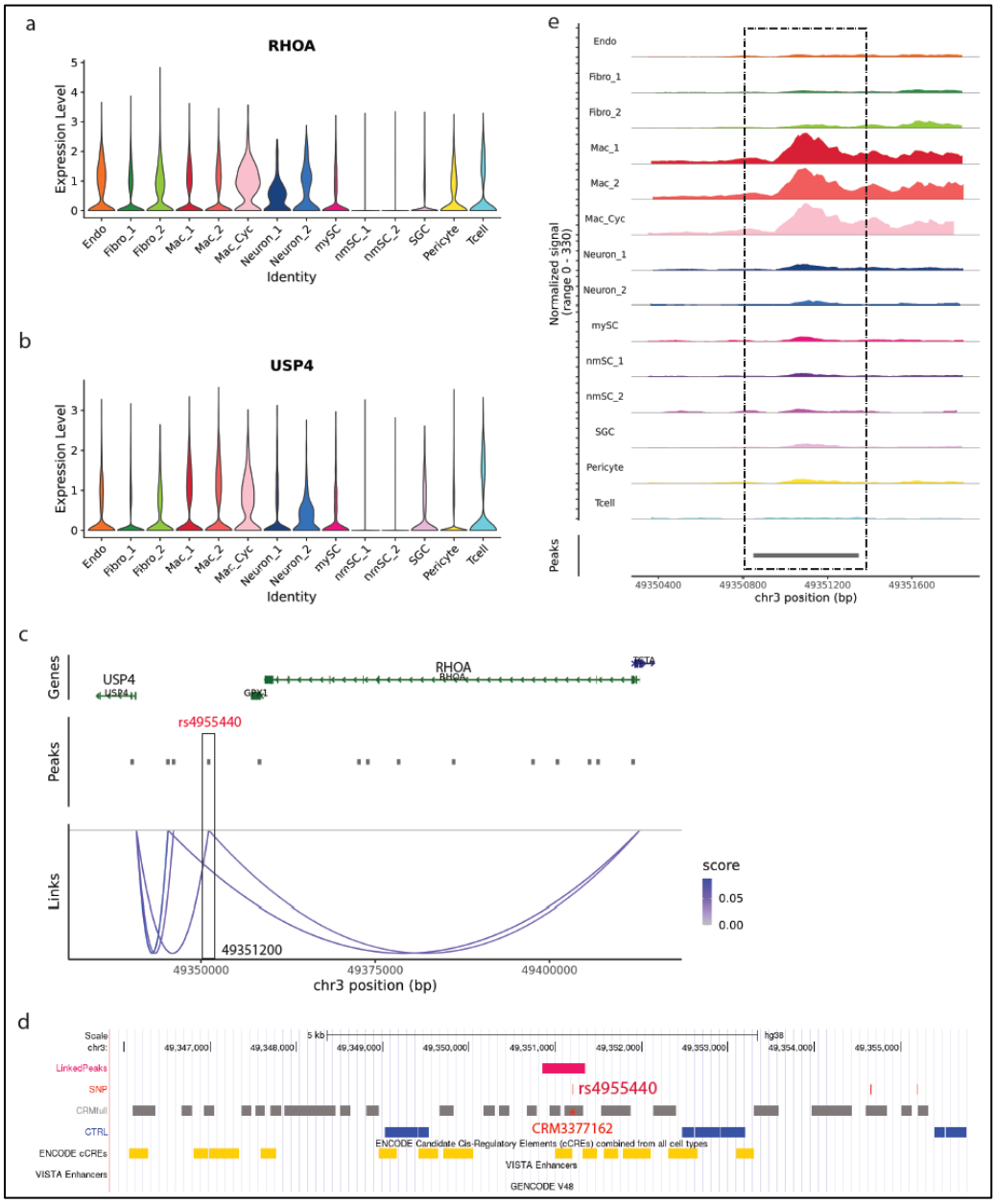


Figure 8: Macrophage-specific association of pain-associated SNP rs4955440 linked to RHOA and USP4. (**a, b**) Violin-plots showing the expression of RHOA (a) and USP4 (b) across different cell populations. (**c**) Peak–gene links for RHOA and USP4 and the location of pain-associated SNP rs4955440. (d) Genome view of the snATAC-seq peak significantly associated with RHOA and USP4, the location of pain-associated SNP rs4955440, VRMOD predicted CRM3377162, and ENCODE cCREs. (**f**) snATAC-seq peak signals.

reported in the DRG and spinal cord in rodent pain models, and pharmacological inhibition of this pathway alleviates hyperalgesia and allodynia(*103*). *USP4,* in contrast, exhibits context-dependent roles in injury and inflammation. Following spinal cord injury, *USP4* expression is downregulated, promoting microglial activation and subsequent neuronal inflammation(*104*). Conversely, in a liver injury model, *USP4* levels were elevated, where it functions to reduce hepatocyte apoptosis and liver inflammation(*105*). These findings underscore the cell-type- and context-specific regulation of USP4. Interestingly, the snATAC-seq peak chr3:49350847-49351347, overlapping CRM3377162, was unique to macrophages (**Fig. 8e**), suggesting a potential macrophage-specific regulatory role of rs4955440 in pain susceptibility.

In addition to the identification of genes whose dysregulation is known to be associated with pain, we also identified new putative pain-associated genes. For example, Sushi, von Willebrand factor type A, EGF and

pentraxin domain containing 1 (*SVEP1*) is an extracellular matrix protein that causally promotes vascular disease(*106*) and associates with multiple other diseases including hypertension, type 2 diabetes, and glaucoma(*107, 108*). *SVEP1* is uniquely expressed in fibroblasts and Neuron_1, the small diameter sensory neuron population (**Fig. S7a).** Interestingly, many pain-associated SNPs were identified within the first intron of *SVEP1* (**Fig. S7c**). However, only two SNPs -- rs77748855 and rs80253943(*94*)., were located within VRMOD-predicted CRMs (CRM5626436) that also overlapped with snATAC-seq peak was (chr9:110575075-110575575), suggesting their regulatory function in DRG cell types. The snATAC-seq peak had the strongest signal in fibroblast with lower activities in endothelial cells, neurons, and nmSCcells (**Fig. S7d**). Dysregulation of extracellular matrix genes has been associated with neuropathic pain(*109*). Both SNPs were located within snATAC-seq peak chr9:110575075-110575575 (**Fig. S7c**), which had stronger activity in fibroblasts and neurons (**Fig. S7d**). MAD1L1 encodes the spindle assembly checkpoint (SAC) protein MAD1, which is an essential component of the mitotic checkpoint. MAD1L1 mutations cause a new variant of mosaic variegated aneuploidy with systemic inflammation and unprecedented tumor susceptibility(*110*). Whether *SVEP1* or *MAD1L1* play any role in pain sensation remains to be tested.

DISCUSSION

The mechanisms underlying chronic pain in various conditions are not fully understood, and existing treatments often prove ineffective for most patients(*31-35*). There is thus an urgent need for advancements in our understanding of the mechanisms regulating pain transmission, which could generate new pivotal insights for the treatment of chronic pain. DRGs harbor the soma of sensory neurons, which integrate signals from the body's internal and external environment, thus representing the first step in pain signal transduction. Our study provides a deeper understanding of the transcriptional and regulatory landscapes within all cell types in hDRG, emphasizing the critical roles played by non-neuronal cell types, such as endothelial cells, fibroblasts, macrophages, and glial cells in the pathogenesis of pain. Furthermore, our study underscores the importance of sequencing depth in enhancing data resolution and accuracy, which in turn facilitates refined cell-type clustering and more precise scientific interpretations. Our study also highlights the need for cell-type validation at the protein level, because of the weak correlation between gene and protein expression.

Due to considerable differences between animal models and humans, there is an urgent need to elucidate the molecular profile of human sensory neurons and non-neuronal cells to develop meaningful strategies to treat pain. With the ability to acquire and study hDRG through organ procurement organizations, the last decade has seen significant advances in characterizing the cellular and molecular features of cells in human DRG (*17, 21, 22, 24, 26, 109, 111-116*). Most efforts have focused on transcriptional profiling to identify neuron and non-neuronal cell types, determine interspecies similarities and differences, and probe mechanisms underlying acute and chronic pain(*17, 21, 22, 111, 112*). Here, we extend beyond transcriptomics to examine the chromatin landscape within the same cells. By integrating the transcriptomic and epigenetic data from single-cell multiomics

analysis of human DRG with pain-associated genes and SNPs identified in multiple GWAS, we infer susceptible cell types and candidate causal variants underlying human pain disorders.

The large size of human neurons and their low abundance within DRG tissue represent a major obstacle to defining their molecular profiles. Our results further emphasize that even with single-nuclei sequencing, which circumvents challenges posed by the large neuronal soma in microfluidic-based single-cell RNA-seq technologies, neuronal recovery was limited, yielding only two broad populations. These findings highlight the need for profiling substantially larger cohorts of hDRG to capture sufficient neuronal diversity and enable accurate molecular characterization. Nonetheless, by leveraging a curated comprehensive set of pain-associated SNPs and genes, we found that cell types contributing to Neuron_1, but not Neuron_2, express marker genes significantly enriched across nearly all pain conditions (with the exception of migraine, as expected). This enrichment is consistent with the notion that Neuron_1 includes nociceptor populations.

Multiple non-neuronal cell types have emerged to contribute to pain development in animal models(*2-8*), underscoring the importance of characterizing the molecular profiles of non-neuronal cells in human DRG. Supporting this, our analysis of pain-associated SNPs derived from GWAS studies of various pain conditions, revealed enrichment not only for neuronal marker genes but also for marker genes of all three glial types in subsets of pain conditions. Our transcriptional profiling of glial cell types identified the three main glial types expected in DRG, namely SGCs, mSC and nmSC. Our gene regulatory network analysis identified *SOX10* as a transcription activator regulating glial signature gene modules, which was expressed in all three glial cell types. Interestingly, nmSCs labeled for *SCN7A*, occupied two distinct spatial domains within the hDRG. *SCN7A* positive nmSCs were found around the axon as expected but also, surprisingly, around SGCs. Whether the two nmSC populations we identified by unbiased clustering correspond to these two spatially distinct nmSC subtypes remains to be determined.

Our analysis of SNPs within cis-regulatory modules provided further support for non-neuronal cells in pain, especially immune cells and fibroblasts. While the role of immune cells in pain has been well established in rodent models(*117*), our study highlight a potential role for macrophages in human pain conditions. Notable, *SLC4A7*—enriched in macrophages with additional expression in fibroblasts—was identified as a gene associated with a single-site chronic pain SNP (*94*). *SLC4A7* encodes the electroneutral sodium/HCO3 cotransporter NBCn1, which plays important physiological and pathophysiological roles in diverse cell types. in In animal models, multiple lines of evidence suggest that SLC4A7 is linked to sensory dysfunctions. In mice, genetic disruption of *Slc4a7/*NBCn1 leads to pronounced loss of hair cells and degeneration of sensory receptors in the eye and inner ear causing blindness and deafness(*118, 119*). In macrophages, SLC4A7 expression increases during differentiation from monocytes and is important for proper phagosome function in

macrophages(*120*). Together, these findings suggest that regulation of SLC4A7 expression may be critical for sensory function and pain-associated symptoms.

Fibroblasts in the DRG remain poorly characterized; however, emerging evidence suggests that extracellular matrix (ECM) produced in part by fibroblasts contribute to neuronal signaling in both mice and humans, and that dysregulation of ECM genes is associated with neuropathic pain(*109, 121*). Our analysis uncovered multiple pain-associated SNPs within the first intron of SVEP1, an ECM protein previously implicated in vascular and chronic diseases such as hypertension, type 2 diabetes, and glaucoma. SVEP1 expression was detected in fibroblasts and the Neuron_1 population, further highlighting the role of ECM in pain processing. Additionally, our analysis also points to MAD1L1, a gene involved in the spindle assembly checkpoint, where mutations lead to mosaic variegated aneuploidy featuring systemic inflammation and heightened tumor susceptibility. Whether SVEP1 and MAD1L1 directly contribute to pain signaling in the DRG remains to be determined. Together, these findings underscore the potential importance of these newly identified genes in pain pathways and advocate for further investigation into their roles, which may offer new therapeutic targets for pain management.

In summary, this work provides new insight into the transcriptional regulation of divers human DRG cell types and demonstrates that the non-neuronal microenvironment in the DRG as a critical component in pain pathogenesis. Our findings highlight candidate cell types and putative molecular mechanisms underlying pain sensation, offering a foundation for future studies aimed at therapeutic development.

MATERIALS AND METHODS

Human tissue

Human DRG were obtained from organ donors with full legal consent for use of tissue in research and in compliance with procedures approved by Mid-America Transplant. The Human Research Protection Office at Washington University in St. Louis provided an institutional review board waiver.

Tissue preparation and immunohistochemistry

For immunohistochemistry studies, after isolation of mouse DRG, the tissue was postfixed using 4% paraformaldehyde for 24 hours at 4 degrees. Tissue was then washed in PBS and cryoprotected using 30% sucrose solution at 4℃ overnight. Next, the human and mouse tissues were embedded in optimal cutting temperature, frozen, and mounted for cryosectioning. Human and mouse frozen sections were cut at 7 µm and 10 µm respectively mounted on Superfrost glass slides and stored at - 80°C. The day of the experiment, slides equilibrate at -20 degrees for 1h, then a 4degrees for 10 minutes. Human sections were fixed with 4% paraformaldehyde for 30 minutes at 4 degrees. Sections were washed 3 times for 15 min each in 0.02 M PBS at room temperature. Block slide with 3% BSA, 0.3% Triton-X in DPBS for 45 min at room temperature. Sections

were washed 3 times for 15 min each in 0.02 M PBS at room temperature. Human and mouse sections were incubated overnight in 0.02 M PBS containing 0.3% Triton X100, 0.5% BSA and primary antibodies at room temperature in a humid chamber. The following day, sections were washed 3 times for 15 min each in 0.02 M PBS and incubated for 2 h at room temperature with secondary antibodies, diluted in 0.02 M PBS containing 0.3% Triton X100 and 0.5% BSA. Then, sections were washed 3 times for 15 min in 0.02 M PBS and mounted using Vectashield medium containing DAPI (VectorLabs, H-2000-10). Images were acquired at 310 or 320 using a Nikon TE2000E inverted microscope. Stained sections with only secondary antibody were used as controls. Antibodies were as follows: NEFH antibody (BioTechne, NB300-217), NEFL (CST, 10768), GFAP (Addgene, 1934450), PRPH (Abcam, ab269861), ROBO2 (Biotechne, AF3147), COL IV (Abcam, ab6586), FABP7 (Biotechne, AF3166), NeuN (SYSY,266 011), SCN7A (Biotechne, NBP1-87075), MBP (Millipore Sigma, AB9348), SOX10 (BioTechne, AF2864).

Neuron proportion quantification

To determine the percentage of neurons in adult human DRGs, we quantified the proportion of neurons in hDRG sections from three donors using pan-neuronal maker TUJ1 to label neurons, GFAP to label glial cells, and DAPI to label all nuclei**.** Due to the uneven distribution of neurons in the DRG, neuron counting was performed at two distinct magnifications. Low magnification (4X) was utilized to quantify the number of neurons in six to seven complete DRG sections per donor. Higher magnification (20X) facilitated the quantification of neurons in 12–15 neuron-dense areas per donor.

scMultiomic protocol comparison

We compared four nuclear dissociation methods previously described: a nuclear extraction protocol with density gradient centrifugation alone (Opti), Opti followed by FACS (Opti/FACS), a crude extraction protocol followed by FACS (Non-gradient/FACS), and the nuclear isolation kit developed by 10X Genomics (10X kit). Opti or Opti/FACS were preformed according to the protocol described previously with minor modifications (Cite https://bio-protocol.org/en/bpdetail?id=4102&type=0&searchid=BP1707277373849445&sort=1). Human DRG sections were minced by scissors in homogenization buffer (0.25 M sucrose, 25 mM KCl, 5 mM $MgCl_2$, 10 mM Tris-HCl, pH 8.0, 5 ug/mL actinomycin, 1% BSA, and 0.08 U/ul RNase inhibitor, 0.01% NP40) on ice. Samples were transferred to a dounce homogenizer for 15 strokes with the loose pestle in a total volume of 1 mL, followed by five additional 15 strokes with the tight pestle. The tissue homogenate was then passed through a 50 um filter and diluted 1:1 with working solution (50% iodixanol, 25 mM KCl, 5 mM $MgCl_2$, and 10 mM Tris-HCl, pH 8.0). Nuclei were layered onto an iodixanol gradient after homogenization and ultracentrifuged as described previously. After ultracentrifugation, nuclei were collected between the 30 and 40% iodixanol layers and diluted with resuspension buffer (1xPBS with 1% BSA, and 0.08 U/ul RNase inhibitor). Nuclei were centrifuged at 500 g for 10 min at 4°C and resuspended in resuspension buffer. For Opti, nuclei were processed and loaded into microfluid chip according to the manual of 10X Genomics Multiome assay. For Opti/FACS, nuclei were stained

with 7-AAD and 7-AAD+ events were sorted using 100 um nozzle on a BD FACSARIA II into a 1.5 mL microcentrifuge tube containing 100 ul of resuspension buffer. FACS-sorted nuclei were processed and loaded into microfluid chip according to the manual of 10X Genomics Multiome assay. For Non-gradient/FACS, human DRG sections were minced and homogenized as described in Opti. The tissue homogenate was then passed through a 50 um filter, and centrifuged at 500 g for 10 min at 4°C. Pellet was resuspended in resuspension buffer with 7-AAD. 7-AAD+ events were sorted using 100 um nozzle on a BD FACSARIA II into a 1.5 mL microcentrifuge tube containing 100 ul of resuspension buffer. FACS-sorted nuclei were processed and loaded into microfluid chip according to the manual of 10X Genomics Multiome assay. For 10X kit, nuclei were extracted following the manual of 10X Genomics Nuclear Isolation Kit, and processed and loaded into microfluid chip according to the manual of 10X Genomics Multiome assay.

To assess the enrichment of neuronal nuclei, we performed snMultiome-seq on nuclei isolated by each protocol and sequenced the libraries with target sequencing depth at 50k reads per cell for the snRNA-seq library. We then performed clustering based on the snRNA-seq data using R package Seurat V4.3.0 and identified marker genes for each cluster. A cluster is identified as neuronal if nuclei in that cluster showed increased level of *RBFOX1* (avg_log2FC > 2, FDR < 0.01) compared to all other clusters. We then calculated the percentage of neuronal nuclei for each sequencing library. Opti yielded the most neuronal nuclei (2.39 ± 1.03%, n=4), while 10X, Opti/FACS, and non-gradient/FACS yielded 0.54 ± 0.07% (n=2), 0.06% and 0% neuronal nuclei, respectively. Opti-prep preserves neuronal nuclei, whereas neuronal nuclei are lost with the other methods.

Opti method was used for all additional samples. All samples were sequenced deeply targeting 150,00 reads per cell for snRNA-seq and 100,000 reads per cell for snATAC-seq.

scRNA-seq data analysis

snRNA-seq data were analyzed using Seurat (v5.0.1). We assessed QC on a per sample basis before integration, uniformly removing all cells with nCount < 1000 and nFeature < 500, as well as cells with >10% of genes mapped to mitochondria. An additional cutoff for nCount was used to remove outlier cells which was determined by individual sample quality. We performed normalization using SCTransform on all samples, and feature selection and integration was performed per Seurat tutorial using default parameters. Dimensional reduction and cell clustering were performed using high dimensionality and resolution to identify low quality cells and doublets. After repeated rounds of doublet/low quality cell removal and re-integration, a final object was generated for downstream analysis.

snATAC-seq data analysis

snATAC-seq data were analyzed using Signac (v1.6.0). We calculated or obtained the following quality metrics from CellRanger ARC outputs: nucleosome signals, Transcriptional start site (TSS) enrichment score, total

number of fragments in peaks, fraction of fragments in peaks, and ratio reads in genomic blacklist regions. We removed cells with < 500 nFeatures and < 1000 nCounts, TSS enrichment score < 2, nucleosome signal > 2.
To generate uniform peak sets we first performed genome-wide chromatin accessibility peaks calling for each cell cluster using MACS v2 (MACS2), with the following flags explicitly set: --nomodel --call-summits --SPMR -q 0.01; generating a list of peak summit calls per cell clusters. Uniform peak sets were generated using MACS2 with arguments --nomodel --call-summits --SPMR -q 0.01. Peak summits were then extended by 250 bases on either end to generate 501 bp peaks. The peaks for all cell clusters were then pooled and ranked by the FDR corrected p-values of the summit (from MACS2). Overlapping peaks were iteratively removed keeping the peak with the lowest p-value until no overlapping peaks remained. Single cell counts for reads in peaks were generated by Signac as the total number of unique fragment cut sites overlapping a given peak window tallied for each unique cell barcode detected in the data, producing a matrix of single cell chromatin accessibility counts in peaks (rows) by cells (columns). This uniform peak set was then added to the final object and used for all downstream analyses. We performed Normalization, Feature selection, linear dimensional reduction, cell clustering following the tutorial of the Signac package. We create gene activity matrix, integrated with snRNA-seq data, and detected differentially accessible peaks between cell types as described in the Signac tutorial with default parameters.

Multimodal Clustering

We performed a joint neighborhood analysis integrating the snRNA-seq data and snATAC-seq data using the weighted nearest neighbor (WNN) method in Seurat v5. We performed clustering using principal component analysis (PCA) on dimensions 1:30 for the RNA and dimensions 2:30 on the LSI components for the ATAC, followed by the functions “RunUMAP” and “FindClusters” with a resolution of 0.2. Cell types were then annotated using conserved and unique markers, as well as known markers with evidence in literature.

Differentially Expressed Genes (DEGs) and Differentially Accessible Regions (DARs)

DEGs and DARs were calculated using the Seurat v5 “FindMarkers” function to find transcripts present in at least 20% of cells. An absolute log fold change ≥ 0.1 and Bonferonni adjusted p-values *P*.adj < 0.05 were used to determine statistical significant genes and peaks.

Computing peak-to-gene links

To calculate linkage scores between RNA transcripts and ATAC peaks, the “LinkPeaks” function from the Signac package was used. Peaks within 100kb of the TSS for all genes were evaluated and scored based on correlation between gene expression and chromatin accessibility. Significant peak-to-gene links were determined using default parameters for correlation coefficient (> 0.05) and p-value (< 0.05), as well as only considering links where the gene and peak are present in at least 10 cells.

FigR gene regulatory network analysis

Peak-gene *cis*-regulatory correlation analysis were performed to determine high density domains of regulatory chromatin (DORCs) using FigR. The default 100 kb window around the TSS was used.

TF motif accessibility scoring

Single cell accessibility scores for TF motifs were computed using chromVAR[49] as previously described[12,13]. For TF motif accessibility scores, the peak by TF motif overlap annotation matrix was generated using a list of human TF motif PWMs from the chromVARmotifs package in R (https://github.com/GreenleafLab/chromVARmotifs), and used along with the scATAC-seq reads in peaks matrix to generate accessibility Z-scores across all scATAC-seq cells passing filter.

Pain-associated cell types and putative GWAS causal variant identification

We integrated GWAS pain-associated SNPs together with our single cell ATAC-seq data through the VRMOD SNP and CRM prioritization framework. First, curated lists of SNPs and genes from GWAS studies for different pain conditions and diseases were collected from publications (**Supplemental Table 11**). For SNPs where associated genes were not reported by the publication, we used ChipSeeker to find the genes most closely associated with the reported SNPs. In total, we curated seven lists of SNPs from six publications with a total of 6,212 pain-associated SNPs.

Vertebrate Regulatory MOdule Detector (VRMOD) (Mangetti-Goncalves et. al, unpublished) relies solely on genomic sequences of a given organism for predicting cis-regulatory modules (CRMs), without using any empirical data or phylogenetic conservation information at any stage of CRM identification. As a result, any available empirical dataset can serve as orthogonal evidence, providing independent validation of predicted CRMs. For each CRM, we provide independent lines of evidence supporting its functionality through the integration of VRMOD prediction and ten genomic and epigenomic datasets representing putative regulatory sequences. The ten datasets are described below.

1) A total of 60 literature-curated human ExpCRMs driving gene expression in a diverse type of tissues, cell types, and developmental stages including neurons, glia and immune cells.
2) A total of 998 *in vivo* validated functional enhancer elements for human from the VISTA Enhancer Browser
3) A total of 201,032 putative human enhancers from Enhancer Atlas 2.0, annotated in 534 different tissues/cell types(*122*).
4) DHS peaks: DHS regions contain a variety of regulatory elements including enhancers, silencers, insulators, and locus control regions(*100*). Human DNase-seq data (2,401,388 peaks) were downloaded from the Cistrome database.

5) Bulk ATAC-seq peaks: The Cistrome database contained 1,217,002 human ATAC-seq peaks from 1,059 samples and 25 different tissues and cell types.
6) ChIP-seq data of the histone marks H3K4me1 and H3K27ac represent putative active enhancers. Human H3K4me1 (1,565,782) and H3K27ac (1,697,483) ChIP-seq data were downloaded from the Cistrome database.
7) ChIP-seq data of the H3K4me3, which represents putative promoters, were downloaded from the Cistrome database with 2,081,430 peaks for human.
8) scATAC-seq detects cell-type-specific open chromatin regions potentially representing cell-type-specific regulatory elements functioning in hundreds of cell types from different tissues and developmental stages. Human scATAC-seq data provided ~1.2 million elements from 45 adult/fetal tissues and 222 cell types(*101*).
9) ENCODE cCREs are candidate cis-Regulatory Elements (cCREs) obtained through the integration of ENCODE epigenomics data. Currently there are two different versions of ENCODE cCREs. Our analysis was conducted before the release of ENCODE cCREs V3 (labeled as ENCODE cCREs herein). We used the ENCODE V2 version (referred to as ENCODE_V2 herein) to perform the analysis which included 1,063,878 cCREs from 1,518 tissues/cell types for the human genome(*100*).

To identify putative pain-causal SNPs, we converted pain-associated SNPs into bed format and identified VRMOD predicted CRMs that contain at least one pain-associated SNPs (referred to as **pain-associated CRMs**). Next, we identified genes potentially regulated by **pain-associated CRMs** using peaks that were significantly linked to genes within 100kb of the genes transcription start site (TSS) determined via genome wide peak-gene link analysis described above. Because these SNPs were located within cis-regulatory sequences with empirical data supporting the regulatory function they are likely SNPs that will affect target gene expression, therefore potentially causal to pain condition. We found ten unique peaks which overlapped with at least 1bp with ten pain-associated CRMs. Eight unique genes were significantly linked with these ten peaks.

List of Supplementary Materials

Fig S1 to S7 for multiple supplementary figures
Tables S1 to S12 for multiple supplementary tables

References

1. A. E. Dubin, A. Patapoutian, Nociceptors: the sensors of the pain pathway. *J Clin Invest* **120**, 3760-3772 (2010).
2. M. Hanani, Role of satellite glial cells in gastrointestinal pain. *Front Cell Neurosci* **9**, 412 (2015).
3. L. Jasmin, J. P. Vit, A. Bhargava, P. T. Ohara, Can satellite glial cells be therapeutic targets for pain control? *Neuron Glia Biol* **6**, 63-71 (2010).
4. R. A. Warwick, M. Hanani, The contribution of satellite glial cells to chemotherapy-induced neuropathic pain. *Eur J Pain* **17**, 571-580 (2013).
5. A. J. Shepherd *et al.*, Angiotensin II Triggers Peripheral Macrophage-to-Sensory Neuron Redox Crosstalk to Elicit Pain. *J Neurosci* **38**, 7032-7057 (2018).
6. X. Yu *et al.*, Dorsal root ganglion macrophages contribute to both the initiation and persistence of neuropathic pain. *Nat Commun* **11**, 264 (2020).
7. P. Meriau, R. Kuruvilla, V. Cavalli, Satellite glial cells: Shaping peripheral input into the brain-body axis? *Neuron*, (2025).
8. M. Hanani, D. C. Spray, Emerging importance of satellite glia in nervous system function and dysfunction. *Nat Rev Neurosci* **21**, 485-498 (2020).
9. D. Simon *et al.*, Friedreich ataxia mouse models with progressive cerebellar and sensory ataxia reveal autophagic neurodegeneration in dorsal root ganglia. *J Neurosci* **24**, 1987-1995 (2004).
10. Y. Shan *et al.*, Frataxin deficiency leads to defects in expression of antioxidants and Nrf2 expression in dorsal root ganglia of the Friedreich's ataxia YG8R mouse model. *Antioxid Redox Signal* **19**, 1481-1493 (2013).
11. A. H. Koeppen, R. L. Ramirez, A. B. Becker, J. E. Mazurkiewicz, Dorsal root ganglia in Friedreich ataxia: satellite cell proliferation and inflammation. *Acta Neuropathol Commun* **4**, 46 (2016).
12. R. V. Haberberger, J. Kuramatilake, C. M. Barry, D. Matusica, Ultrastructure of dorsal root ganglia. *Cell Tissue Res* **393**, 17-36 (2023).
13. K. Jang, S. M. Garraway, A review of dorsal root ganglia and primary sensory neuron plasticity mediating inflammatory and chronic neuropathic pain. *Neurobiol Pain* **15**, 100151 (2024).
14. J. X. Yang *et al.*, Potential Neuroimmune Interaction in Chronic Pain: A Review on Immune Cells in Peripheral and Central Sensitization. *Front Pain Res (Lausanne)* **3**, 946846 (2022).
15. Q. Chen *et al.*, Unveiling adcyap1 as a protective factor linking pain and nerve regeneration through single-cell RNA sequencing of rat dorsal root ganglion neurons. *BMC Biol* **21**, 235 (2023).
16. R. Tonello *et al.*, Single-cell analysis of dorsal root ganglia reveals metalloproteinase signaling in satellite glial cells and pain. *Brain Behav Immun* **113**, 401-414 (2023).
17. O. Avraham *et al.*, Profiling the molecular signature of satellite glial cells at the single cell level reveals high similarities between rodents and humans. *Pain* **163**, 2348-2364 (2022).

18. A. A. Mapps *et al.*, Diversity of satellite glia in sympathetic and sensory ganglia. *Cell Rep* **38**, 110328 (2022).
19. O. Avraham *et al.*, Satellite glial cells promote regenerative growth in sensory neurons. *Nat Commun* **11**, 4891 (2020).
20. R. V. Haberberger, C. Barry, N. Dominguez, D. Matusica, Human Dorsal Root Ganglia. *Front Cell Neurosci* **13**, 271 (2019).
21. S. A. Bhuiyan *et al.*, Harmonized cross-species cell atlases of trigeminal and dorsal root ganglia. *Sci Adv* **10**, eadj9173 (2024).
22. D. Tavares-Ferreira *et al.*, Spatial transcriptomics of dorsal root ganglia identifies molecular signatures of human nociceptors. *Sci Transl Med* **14**, eabj8186 (2022).
23. M. Jung *et al.*, Cross-species transcriptomic atlas of dorsal root ganglia reveals species-specific programs for sensory function. *Nat Commun* **14**, 366 (2023).
24. H. Yu *et al.*, Leveraging deep single-soma RNA sequencing to explore the neural basis of human somatosensation. *Nat Neurosci* **27**, 2326-2340 (2024).
25. D. Usoskin *et al.*, Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. *Nat Neurosci* **18**, 145-153 (2015).
26. M. Q. Nguyen, L. J. von Buchholtz, A. N. Reker, N. J. Ryba, S. Davidson, Single-nucleus transcriptomic analysis of human dorsal root ganglion neurons. *Elife* **10**, (2021).
27. K. J. A. Johnston *et al.*, Genome-wide association study of multisite chronic pain in UK Biobank. *PLoS Genet* **15**, e1008164 (2019).
28. H. G. Stunnenberg, C. International Human Epigenome, M. Hirst, The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. *Cell* **167**, 1897 (2016).
29. C. Global Burden of Disease Study, Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **386**, 743-800 (2015).
30. L. J. Hocking *et al.*, Heritability of chronic pain in 2195 extended families. *Eur J Pain* **16**, 1053-1063 (2012).
31. N. B. Finnerup *et al.*, Pharmacotherapy for neuropathic pain in adults: a systematic review and meta-analysis. *Lancet Neurol* **14**, 162-173 (2015).
32. L. E. Chaparro, P. J. Wiffen, R. A. Moore, I. Gilron, Combination pharmacotherapy for the treatment of neuropathic pain in adults. *Cochrane Database Syst Rev* **2012**, CD008943 (2012).
33. M. Balanaser *et al.*, Combination pharmacotherapy for the treatment of neuropathic pain in adults: systematic review and meta-analysis. *Pain* **164**, 230-251 (2023).
34. W. H. van der Gaag, P. D. Roelofs, W. T. Enthoven, M. W. van Tulder, B. W. Koes, Non-steroidal anti-inflammatory drugs for acute low back pain. *Cochrane Database Syst Rev* **4**, CD013581 (2020).

35. K. Zorina-Lichtenwalter *et al.*, Genetic risk shared across 24 chronic pain conditions: identification and characterization with genomic structural equation modeling. *Pain* **164**, 2239-2252 (2023).
36. C. S. Nielsen, G. P. Knudsen, O. A. Steingrimsdottir, Twin studies of pain. *Clin Genet* **82**, 331-340 (2012).
37. J. Vehof, H. M. S. Zavos, G. Lachance, C. J. Hammond, F. M. K. Williams, Shared genetic factors underlie chronic pain syndromes. *Pain* **155**, 1562-1568 (2014).
38. M. Gasperi *et al.*, Chronic prostatitis and comorbid non-urological overlapping pain conditions: A co-twin control study. *J Psychosom Res* **102**, 29-33 (2017).
39. C. B. Meloto *et al.*, Human pain genetics database: a resource dedicated to human pain genetics research. *Pain* **159**, 749-763 (2018).
40. F. Zhang, J. R. Lupski, Non-coding genetic variants in human disease. *Hum Mol Genet* **24**, R102-110 (2015).
41. M. T. Maurano *et al.*, Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190-1195 (2012).
42. S. Ma *et al.*, Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. *Cell* **183**, 1103-1116 e1120 (2020).
43. O. Avraham *et al.*, Profiling sensory neuron microenvironment after peripheral and central axon injury reveals key pathways for neural repair. *Elife* **10**, (2021).
44. K. Nagashima, K. Oota, A histopathological study of the human spinal ganglia. 1. Normal variations in aging. *Acta Pathol Jpn* **24**, 333-344 (1974).
45. T. Tandrup, Unbiased estimates of number and size of rat dorsal root ganglion cells in studies of structure and cell survival. *J Neurocytol* **33**, 173-192 (2004).
46. C. A. West, A. McKay Hart, G. Terenghi, M. Wiberg, Sensory neurons of the human brachial plexus: a quantitative study employing optical fractionation and in vivo volumetric magnetic resonance imaging. *Neurosurgery* **70**, 1183-1194; discussion 1194 (2012).
47. J. Pearson, B. A. Pytel, N. Grover-Johnson, F. Axelrod, J. Dancis, Quantitative studies of dorsal root ganglia and neuropathologic observations on spinal cords in familial dysautonomia. *J Neurol Sci* **35**, 77-92 (1978).
48. L. Yang *et al.*, Human and mouse trigeminal ganglia cell atlas implicates multiple cell types in migraine. *Neuron* **110**, 1806-1821 e1808 (2022).
49. T. Stuart, R. Satija, Integrative single-cell analysis. *Nat Rev Genet* **20**, 257-272 (2019).
50. Y. Zhang *et al.*, Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
51. Y. Hao *et al.*, Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol* **42**, 293-304 (2024).
52. T. Stuart, A. Srivastava, S. Madad, C. A. Lareau, R. Satija, Single-cell chromatin state analysis with Signac. *Nat Methods* **18**, 1333-1341 (2021).

53. Y. Chen *et al.*, Common and specific gene regulatory programs in zebrafish caudal fin regeneration at single-cell resolution. *Genome Res* **35**, 202-218 (2025).
54. A. Bryant *et al.*, Endothelial Cells Are Heterogeneous in Different Brain Regions and Are Dramatically Altered in Alzheimer's Disease. *J Neurosci* **43**, 4541-4557 (2023).
55. J. Xu *et al.*, Human striatal glia differentially contribute to AD- and PD-specific neurodegeneration. *Nat Aging* **3**, 346-365 (2023).
56. Y. S. Yi *et al.*, Functional roles of Syk in macrophage-mediated inflammatory responses. *Mediators Inflamm* **2014**, 270302 (2014).
57. S. B. Siems *et al.*, Proteome profile of peripheral myelin in healthy mice and in a neuropathy model. *Elife* **9**, (2020).
58. T. F. Barrett *et al.*, Single-cell multi-omic analysis of the vestibular schwannoma ecosystem uncovers a nerve injury-like state. *Nat Commun* **15**, 478 (2024).
59. V. E. Abraira, D. D. Ginty, The sensory neurons of touch. *Neuron* **79**, 618-639 (2013).
60. X. N. Yi *et al.*, Dynamic changes in Robo2 and Slit1 expression in adult rat dorsal root ganglion and sciatic nerve after peripheral and central axonal injury. *Neurosci Res* **56**, 314-321 (2006).
61. L. He *et al.*, Rbfox1 regulates alternative splicing of Nrcam in primary sensory neurons to mediate peripheral nerve injury-induced neuropathic pain. *Neurotherapeutics* **21**, e00309 (2024).
62. Y. C. Chen *et al.*, Transcriptional regulator PRDM12 is essential for human pain perception. *Nat Genet* **47**, 803-808 (2015).
63. X. Fang *et al.*, trkA is expressed in nociceptive neurons and influences electrophysiological properties via Nav1.8 expression in rapidly conducting nociceptors. *J Neurosci* **25**, 4868-4878 (2005).
64. A. M. Barry *et al.*, Multi-omic integration with human dorsal root ganglia proteomics highlights TNFalpha signalling as a relevant sexually dimorphic pathway. *Pain*, (2025).
65. T. Maier, M. Guell, L. Serrano, Correlation of mRNA and protein in complex biological samples. *FEBS Lett* **583**, 3966-3973 (2009).
66. R. de Sousa Abreu, L. O. Penalva, E. M. Marcotte, C. Vogel, Global signatures of protein and mRNA expression levels. *Mol Biosyst* **5**, 1512-1526 (2009).
67. A. L. M. Caldwell *et al.*, Aberrant astrocyte protein secretion contributes to altered neuronal development in multiple models of neurodevelopmental disorders. *Nature neuroscience* **25**, 1163-1178 (2022).
68. C. Rostock, K. Schrenk-Siemens, J. Pohle, J. Siemens, Human vs. Mouse Nociceptors - Similarities and Differences. *Neuroscience* **387**, 13-27 (2018).
69. H. Lund *et al.*, CD163+ macrophages monitor enhanced permeability at the blood-dorsal root ganglion barrier. *J Exp Med* **221**, (2024).
70. W. Wen, J. Cheng, Y. Tang, Brain perivascular macrophages: current understanding and future prospects. *Brain* **147**, 39-55 (2024).

71. H. Y. Lim *et al.*, Hyaluronan Receptor LYVE-1-Expressing Macrophages Maintain Arterial Tone through Hyaluronan-Mediated Regulation of Smooth Muscle Cell Collagen. *Immunity* **49**, 326-341 e327 (2018).
72. R. Feng, V. Muraleedharan Saraswathy, M. H. Mokalled, V. Cavalli, Self-renewing macrophages in dorsal root ganglia contribute to promote nerve regeneration. *Proc Natl Acad Sci U S A* **120**, e2215906120 (2023).
73. B. Ober-Reynolds *et al.*, Integrated single-cell chromatin and transcriptomic analyses of human scalp identify gene-regulatory programs and critical cell types for hair and skin diseases. *Nat Genet* **55**, 1288-1300 (2023).
74. Y. E. Li *et al.*, A comparative atlas of single-cell chromatin accessibility in the human brain. *Science* **382**, eadf7044 (2023).
75. V. K. Kartha *et al.*, Functional inference of gene regulation using single-cell multi-omics. *Cell Genom* **2**, (2022).
76. R. T. Mayer, A. C. Bridges, J. Cocke, R. Meola, J. K. Olson, Fluorescent insect growth regulators (FIGR's): a new tool for insect physiologists. *J Insect Physiol* **22**, 515-520 (1976).
77. S. Ghisletti *et al.*, Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. *Immunity* **32**, 317-328 (2010).
78. M. Wegner, C. C. Stolt, From stem cells to neurons and glia: a Soxist's view of neural development. *Trends Neurosci* **28**, 583-588 (2005).
79. M. Wegner, C. C. Stolt, From stem cells to neurons and glia: a Soxist's view of neural development. *Trends in neurosciences* **28**, 583-588 (2005).
80. O. Avraham *et al.*, Profiling the molecular signature of satellite glial cells at the single cell level reveals high similarities between rodents and humans. *Pain* **163**, 2348-2364 (2022).
81. M. Delfino-Machin *et al.*, Sox10 contributes to the balance of fate choice in dorsal root ganglion progenitors. *PLoS One* **12**, e0172947 (2017).
82. K. Kuhlbrodt, B. Herbarth, E. Sock, I. Hermans-Borgmeyer, M. Wegner, Sox10, a novel transcriptional modulator in glial cells. *Journal of Neuroscience* **18**, 237-250 (1998).
83. H. F. Bahmad *et al.*, Clinical significance of SOX10 expression in human pathology. *Current Issues in Molecular Biology* **45**, 10131-10158 (2023).
84. Q. Wei, W. K. Miskimins, R. Miskimins, Sox10 acts as a tissue-specific transcription factor enhancing activation of the myelin basic protein gene promoter by p27Kip1 and Sp1. *J Neurosci Res* **78**, 796-802 (2004).
85. D. Lang, J. A. Epstein, Sox10 and Pax3 physically interact to mediate activation of a conserved c-RET enhancer. *Hum Mol Genet* **12**, 937-945 (2003).
86. M. Finzsch, C. C. Stolt, P. Lommes, M. Wegner, Sox9 and Sox10 influence survival and migration of oligodendrocyte precursors in the spinal cord by regulating PDGF receptor alpha expression. *Development* **135**, 637-646 (2008).

87. T. Baroti *et al.*, Transcription factors Sox5 and Sox6 exert direct and indirect influences on oligodendroglial migration in spinal cord and forebrain. *Glia* **64**, 122-138 (2016).
88. G. Jego *et al.*, Dual regulation of SPI1/PU. 1 transcription factor by heat shock factor 1 (HSF1) during macrophage differentiation of monocytes. *Leukemia* **28**, 1676-1686 (2014).
89. A. Zakrzewska *et al.*, Macrophage-specific gene functions in Spi1-directed innate immunity. *Blood, The Journal of the American Society of Hematology* **116**, e1-e11 (2010).
90. T. Masuda *et al.*, IRF8 is a critical transcription factor for transforming microglia into a reactive phenotype. *Cell Rep* **1**, 334-340 (2012).
91. T. Yashiro *et al.*, PU.1 and IRF8 Modulate Activation of NLRP3 Inflammasome via Regulating Its Expression in Human Macrophages. *Front Immunol* **12**, 649572 (2021).
92. P. Gormley *et al.*, Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. *Nat Genet* **48**, 856-866 (2016).
93. H. Hautakangas *et al.*, Genome-wide analysis of 102,084 migraine cases identifies 123 risk loci and subtype-specific risk alleles. *Nat Genet* **54**, 152-160 (2022).
94. S. Khoury *et al.*, Genome-wide analysis identifies impaired axonogenesis in chronic overlapping pain conditions. *Brain* **145**, 1111-1123 (2022).
95. E. Wistrom, R. Chase, P. R. Smith, Z. T. Campbell, A compendium of validated pain genes. *WIREs Mech Dis* **14**, e1570 (2022).
96. R. Domoto, F. Sekiguchi, M. Tsubota, A. Kawabata, Macrophage as a Peripheral Pain Regulator. *Cells* **10**, (2021).
97. M. S. Halfon, Studying Transcriptional Enhancers: The Founder Fallacy, Validation Creep, and Other Biases. *Trends Genet* **35**, 93-103 (2019).
98. S. Preissl, K. J. Gaulton, B. Ren, Characterizing cis-regulatory elements using single-cell epigenomics. *Nat Rev Genet* **24**, 21-43 (2023).
99. A. Visel, S. Minovitsky, I. Dubchak, L. A. Pennacchio, VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Res* **35**, D88-92 (2007).
100. E. P. Consortium *et al.*, Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699-710 (2020).
101. K. Zhang *et al.*, A single-cell atlas of chromatin accessibility in the human genome. *Cell* **184**, 5985-6001 e5919 (2021).
102. T. Kalpachidou, L. Spiecker, M. Kress, S. Quarta, Rho GTPases in the Physiology and Pathophysiology of Peripheral Sensory Neurons. *Cells* **8**, (2019).
103. N. Chen *et al.*, Exploring the Role of RhoA/ROCK Signaling in Pain: A Narrative Review. *Aging Dis*, (2025).
104. X. Jiang *et al.*, Downregulation of USP4 Promotes Activation of Microglia and Subsequent Neuronal Inflammation in Rat Spinal Cord After Injury. *Neurochem Res* **42**, 3245-3253 (2017).

105. J. Zhou *et al.*, USP4 deficiency exacerbates hepatic ischaemia/reperfusion injury via TAK1 signalling. *Clin Sci (Lond)* **133**, 335-349 (2019).
106. I. H. Jung *et al.*, SVEP1 is a human coronary artery disease locus that promotes atherosclerosis. *Sci Transl Med* **13**, (2021).
107. Coding Variation in ANGPTL4, LPL, and SVEP1 and the Risk of Coronary Disease. *N Engl J Med* **374**, 1898 (2016).
108. P. Gharahkhani *et al.*, Genome-wide meta-analysis identifies 127 open-angle glaucoma loci with consistent effect across ancestries. *Nat Commun* **12**, 1258 (2021).
109. R. Vroman *et al.*, Analysis of matrisome expression patterns in murine and human dorsal root ganglia. *Front Mol Neurosci* **16**, 1232447 (2023).
110. C. Villarroya-Beltri *et al.*, Biallelic germline mutations in MAD1L1 induce a syndrome of aneuploidy with high tumor susceptibility. *Sci Adv* **8**, eabq5914 (2022).
111. S. Davidson *et al.*, Human sensory neurons: Membrane properties and sensitization by inflammatory mediators. *Pain* **155**, 1861-1870 (2014).
112. M. V. Valtcheva *et al.*, Surgical extraction of human dorsal root ganglia from organ donors and preparation of primary sensory neuron cultures. *Nature protocols* **11**, 1877-1888 (2016).
113. C. Rostock, K. Schrenk-Siemens, J. Pohle, J. Siemens, Human vs. Mouse Nociceptors - Similarities and Differences. *Neuroscience*, (2017).
114. A. Serrano *et al.*, Differential expression and pharmacology of native P2X receptors in rat and primate sensory neurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **32**, 11890-11896 (2012).
115. R. Y. North *et al.*, Electrophysiological and transcriptomic correlates of neuropathic pain in human dorsal root ganglion neurons. *Brain : a journal of neurology* **142**, 1215-1226 (2019).
116. S. Shiers *et al.*, ACE2 and SCARF expression in human dorsal root ganglion nociceptors: implications for SARS-CoV-2 virus neurological effects. *Pain* **161**, 2494-2501 (2020).
117. A. Kavelaars, C. J. Heijnen, Immune regulation of pain: Friend and foe. *Sci Transl Med* **13**, eabj7152 (2021).
118. D. Bok *et al.*, Blindness and auditory impairment caused by loss of the sodium bicarbonate cotransporter NBC3. *Nat Genet* **34**, 313-319 (2003).
119. I. A. Lopez *et al.*, Time course of auditory impairment in mice lacking the electroneutral sodium bicarbonate cotransporter NBC3 (slc4a7). *Brain Res Dev Brain Res* **160**, 63-77 (2005).
120. V. Sedlyarov *et al.*, The Bicarbonate Transporter SLC4A7 Plays a Key Role in Macrophage Phagosome Acidification. *Cell Host Microbe* **23**, 766-774 e765 (2018).
121. S. Zhang *et al.*, Fibroblastic SMOC2 Suppresses Mechanical Nociception by Inhibiting Coupled Activation of Primary Sensory Neurons. *J Neurosci* **42**, 4069-4086 (2022).

122. T. Gao, J. Qian, EnhancerAtlas 2.0: an updated resource with enhancer annotation in 586 tissue/cell types across nine species. *Nucleic Acids Res* **48**, D58-D64 (2020).

ACKNOWLEDGEMENTS

We thank all participants and their families for their commitment and dedication to advancing research of diagnosis and treatment for chronic pain, and the staff at Mid-America Transplant. We thank the Genome Technology Access Center at the McDonnell Genome Institute at Washington University School of Medicine for help with genomic analysis. The Center is partially supported by NCI Cancer Center Support Grant #P30 CA91842 to the Siteman Cancer Center from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research.

Funding

National Institutes of Health grant 1U19NS130607 (RWG, VC, GZ). This publication is solely the responsibility of the authors and does not necessarily represent the official view of the National Institute of Health. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Competing Interests Statement

The authors have no conflicts of interest or financial ties to disclose.

Figure legends

Supplemental Figure 1. Sequencing statistics of shallow sequencing and deep sequencing data from Donor 1 and Donar 2.

Supplemental Figure 2. Marker gene expression and cohort information for the 11 hDRG samples. (a) Violin plot of marker gene expression levels. (b) Cells were visualized using wnnUMAP split by donors. (c) Cells were visualized using wnnUMAP split by samples (top) and violin plot of the distribution of total number of features and UMIs for each sample. (d) Stacked bar plot of proportion of cells in each sample.

Supplemental Figure 3. (a) Violin plot of pan-neuronal marker gene expression levels. (b) Violin plot of sensory neuron marker gene expression levels. (c) Violin plot of Neuron_1 marker gene expression levels.

Supplemental Figure 4. Statistics of open chromatin regions and top marker genes and peaks. (a) Bar plot of open chromatin regions (OCRs) identified in each cell type. (b) Distribution of OCRs in relation to genomic features. (c) Heatmap of top 20 marker genes for each cell type. (d) Heatmap of top 20 OCR markers for each cell type.

Supplemental Figure 5. UMAP of DORC accessibility scores (right) and paired RNA expression (left) for marker genes of each cell type.

Supplemental Figure 6. Transcriptional regulatory network analysis for macrophages. (**a**) Violin-plots showing the expression of macrophage-associated DORC genes. (**b**) TF footprinting analysis for SPI1 motifs sites.

Supplemental Figure 7. Cell-type specific SVEP1-associated risk of pain. (a) Violin-plot showing the expression levels of SVEP1 across all cell populations. (b) peak–gene links for SVEP1 and the location of pain-associated SNPs. (c) Genome view of the snATAC-seq peak significantly associated with SVEP1, the location of pain-associated SNPs, VRMOD predicted CRM5626436, and ENCODE cCREs. (d) snATAC-seq peak signal.

Supplemental Table 1. Donor metadata information.

Supplemental Table 2. Preliminary quantification of percentage of neuronal cells in human DRGs.

Supplemental Table 3. 10x cellranger output statistics of shallow and deep sequencing data.

Supplemental Table 4. 10X sequencing output statistics of the 11 hDRG samples.

Supplemental Table 5. Top cluster marker genes.

Supplemental Table 6. DORCs identified by FigR analysis.

Supplemental Table 7. Transcription factors predicted by FigR.

Supplemental Table 8. Gene list for each regulatory module identified by consensus k-means clustering analysis.

Supplemental Table 9. Target genes of SOX10 predicted by FigR.

Supplemental Table 10. Target genes of SPI1 predicted by FigR**.**

Supplemental Table 11. Curated pain-associated SNPs and genes.

Supplemental Table 12. Putative pain-causal SNPs and linked genes.