Multi-omic profiling of human and mouse dorsal root ganglia enables targeted gene delivery to nociceptors

Lily S. He[1]*, Parth Bhatia[1]*, Shamsuddin A. Bhuiyan[1]*, Evangelia Semizoglou[1], Jiaxiang Wang[1], Jia Li[1], Joohyun Nam[3], Jay X. J. Luo[1], Chayse Arnhold[1], Difei Zhu[1], Mengyi Xu[1], Dustin Griesemer[2], Hyo Jeong Yong[1], Lorna Jayne[4], Evangeline Gilmer[1], Qiyi Li[1], Katerina Pantaleo[1], Lite Yang[1], Erika K. Williams[1], Selwyn Jayakar[1], Brian J. Wainger[3], Sinisa Hrvatin[4-5], William Renthal[1]

1. Department of Neurology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
2. Department of Anesthesiology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
3. Department of Anesthesiology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA
4. Whitehead Institute for Biomedical Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA
5. Howard Hughes Medical Institute, Massachusetts Institute of Technology, 14 Cambridge, MA, USA

*equal contribution
Correspondence: wrenthal@bwh.harvard.edu

Abstract

Chronic pain conditions are often driven by hyperexcitability of nociceptors, the peripheral sensory neurons that detect noxious stimuli. Recent single-cell transcriptomic studies have begun to clarify the molecular identity of distinct peripheral sensory neuron subtypes, but tools that restrict transgene expression to nociceptors while sparing other dorsal root ganglion (DRG) subtypes remain limited. Here, we combined single-nucleus multi-omic profiling of human and mouse DRG with *in vivo* AAV enhancer screening to identify *cis*-regulatory elements that drive biased AAV expression in mouse DRG nociceptors and human iPSC-derived nociceptors. We then validated that an enhancer AAV designed to express $K_{ir}2.1$ preferentially in nociceptors reduces DRG neuronal excitability. Leveraging these multi-omic datasets, we trained a sequence-based model to decode the *cis*-regulatory logic of nociceptors, enabling both the prioritization of native candidate elements and the design of synthetic enhancers with a range of nociceptor targeting properties. Together, these cross-species multi-omic resources define conserved DRG regulatory programs and provide a viral toolkit for pain research with potential translational applications for patients with refractory pain.

Introduction

Chronic pain is often driven by maladaptive hyperexcitability of nociceptors, which are a class of peripheral sensory neurons that detect noxious stimuli. Most analgesics act on multiple cell types across central and peripheral nervous systems, which in the case of opioids, can lead to potentially dangerous side effects.[1–4] Thus, there is an urgent need for more targeted, cell-type-specific strategies to modulate nociceptor activity selectively while sparing other neuron types. Toward that goal, recent single-cell/nucleus RNA sequencing (sc/snRNA-seq) studies of DRG have revealed unique transcriptomic profiles of multiple nociceptor subtypes that are largely conserved between rodents and humans.[5,6] Transgenic mice engineered to target distinct DRG subtypes have yielded important insights into their function,[7–13] but these methods are not easily translated to other model organisms or humans. Indeed, there are currently no tools that provide subtype-selective access to nociceptors across species.

Recent advances in single-cell multi-omic technologies offer a new strategy for characterizing the gene-regulatory logic underlying DRG cell-type-specific gene expression. By jointly profiling chromatin accessibility and gene expression in the same nucleus, it is possible to identify *cis*-regulatory elements (CREs) that promote cell-type-specific gene expression within defined neuronal populations. Engineering AAVs to use cell-type-specific CREs for driving transgene expression in defined cell populations has been successfully implemented in the brain and spinal cord,[14–17] but it has been challenging to apply this approach to peripheral sensory neurons due to our limited understanding of the DRG neuronal epigenome.

Here, we constructed DRG multi-omic cell atlases in both human and mouse and utilized joint chromatin accessibility and transcriptomic data to define evolutionarily conserved *cis*-regulatory programs and CREs that are preferentially active in nociceptors. We built a suite of enhancer

AAVs and quantified their ability to drive viral expression preferentially in nociceptors *in vivo* using scRNA-seq and spatial transcriptomics. We then trained a sequence-based model to decode the *cis*-regulatory logic of nociceptors, which can be used to prioritize native CREs for enhancer AAVs and to design synthetic enhancers. Finally, we show that an enhancer AAV that preferentially expresses in mouse nociceptors relative to other DRG cell types significantly reduces DRG neuronal excitability in mice and drives strong expression in human iPSC-derived nociceptors. Together, these findings establish a cross-species DRG multi-omic resource and outline a general strategy for developing cross-species viral tools for translational pain research with potential therapeutic applications.

Results

Single-nucleus multi-omic profiling of mouse and human DRG

To characterize the *cis*-regulatory logic of mouse and human DRG cell types, we constructed multi-omic (snRNA-seq and snATAC-seq) cell atlases of DRG from 71 male and female adult mice and 66 human donors (32 male, 34 female; Table S1) (Figures 1A). This multi-omic approach allows for the simultaneous interrogation of transcriptional output and chromatin accessibility within the same nucleus, enabling the direct linkage of distal CREs to their target genes without the ambiguity of *in silico* integration.[18] As DRG neurons comprise only ~1% of total cells in the ganglion,[5,6] we optimized neuronal enrichment protocols for both mouse and human DRG prior to multi-omic profiling. In mice, we sorted genetically labeled DRG neurons (*Vglut2*-Cre+/-; *Sun1*-GFP+/-) and observed that >90% of DRG cells sequenced are neurons (Figure S1A). In humans, we sorted NeuN-stained nuclei and observed that up to 60% of DRG cells sequenced are neurons (Figure S1B). We then performed snMultiome of neuronally enriched mouse and human DRG samples, and after quality controls (see Methods, Figure S1C-D, S2A-D), we obtained 23,928 mouse DRG nuclei, including 12,272 epigenomically profiled nuclei from prior work,[19,20] and 155,008 human DRG nuclei. We next performed unsupervised clustering based on gene expression and annotated cell types using the expression of known marker genes (see Methods).[5,6] We annotated eight broad cell types in both mouse and human DRG: neurons, fibroblasts, satellite glia (SGC), myelinating Schwann Cells (mySC), non-myelinating Schwann Cells (nmSC), endothelial cells, immune cells, and pericytes (Figures 1B-C; Figure S2E-G). Gene expression and chromatin accessibility at loci of known marker genes were consistent with anchored cell type annotations (Figures 1B-C, bottom).[5,6] and largely similar between donors and sexes (Figures S3A-D). Differential gene expression between DRG cell types identified 9,856 differentially expressed genes in mice and 11,777 in humans ($\log_2$FC > 0.5, adj. p-val < 0.05, cell type compared to all others) (Table S2 and S3, see Bhuiyan et al., co-submitted for in depth gene expression analyses of human DRG cell types).

To identify regions of transposase-accessible chromatin in each DRG cell type, we performed "peak" calling (see Methods) and observed that peaks of accessible chromatin are predominantly located in intronic and non-coding regions distal to the gene body in both mouse and human (Figures S3E-F; see Methods). We then performed differential chromatin accessibility analysis for the 8 broad DRG cell types by comparing peaks of accessible

105 chromatin in one cell type to all other cell types. We observed 268,405 unique differentially
106 accessible peaks in mice and 252,806 in humans ($log_2FC > 0.5$, adj. p-val < 0.05, cell type
107 compared to all others) (Figure 1D-F; Table S4 and S5). To identify overlap in differentially
108 accessible peaks between the two species, we used the LiftOver tool to identify human regions
109 with high sequence similarity (syntenic) to the mouse differentially accessible peaks.[21,22] Among
110 these syntenic regions, we observed significant overlap (35-44%) between the differentially
111 accessible peaks in corresponding mouse and human DRG cell types relative to a null
112 distribution generated by shuffling cell type identities within the pool of top accessible regions
113 (adj. p-value $<10^{-80}$, Fisher's Exact test) (Figure 1E). This degree of overlap is consistent with
114 conservation of cell-type-specific regulatory programs between mouse and human DRG, and
115 suggests that some differentially accessible peaks may mark shared CREs across species.[23]

116 We next used SCENIC+ to characterize the gene regulatory networks of transcription factors
117 (TFs), CREs, and predicted target genes that establish cell identity in mouse and human DRG
118 cell types. SCENIC+ quantifies the activity of "eRegulons" in each cell by integrating
119 co-expression of a given transcription factor (TF) and predicted target genes with chromatin
120 accessibility and TF motif enrichment at corresponding putative CREs.[24] We observed multiple
121 TFs in which their eRegulons are preferentially active in distinct DRG cell types compared to all
122 other DRG cell types in both mouse (Figure 1G; Table S6) and human (Figure 1H; Table S7).
123 For example, *ISL1*, a transcription factor known to regulate DRG neuronal development,[25]
124 displays significantly higher eRegulon activity in DRG neurons than in non-neuronal cell types in
125 both mouse and human. This ISL1 eRegulon includes the mature neuronal marker gene, *Elavl2*,
126 and a syntenic CRE at which there is both an ISL1 consensus binding site and increased
127 chromatin accessibility in both mouse and human neurons compared to other DRG cell types
128 (Figures 1G-H). Consistent with observations in central nervous system cell types,[15,26–30] the
129 conserved, cell-type-specific regulatory activity of certain TFs between mouse and human DRG
130 cell types suggests that CREs identified in mice have the potential to recruit similar TFs and
131 confer comparable regulatory activity in humans.

132 *Cross-species cis-regulatory logic of nociceptors*

133 To identify regulatory elements that are selectively active in nociceptors across species, we next
134 subclustered 21,064 mouse and 6,893 human DRG neurons by their gene expression profiles
135 (Tables S8-9) and annotated their neuronal subtype by anchoring to previous human and mouse
136 DRG sc/snRNA-seq studies.[6,31–33] We observed the same ~18-22 DRG neuronal subtypes in
137 both mouse and human that have been reported previously (Figure 2A), which includes the
138 important C-peptidergic (C-PEP) and C-non-peptidergic (C-NP) nociceptors that are implicated
139 in chronic pain conditions.[6,31–33] After filtering the annotated mouse and human neuronal nuclei
140 for those with high quality chromatin accessibility data (Figure S1C-D; Figure S4A-B; see
141 Methods), we performed downstream epigenomic analysis on 18,238 mouse and 1,597 human
142 DRG neurons. As human DRG neurons are more challenging to obtain high quality chromatin
143 accessibility data than in mouse or even in DRG non-neuronal cells (Figure S4C-D), we
144 grouped human neuronal subtypes into 6 broader classes (A-Propr, A-LTMR, A-PEP, C-PEP,
145 C-NP, and C-Thermo) for downstream epigenomic analyses. We observed neuronal
146 subtype-specific patterns of chromatin accessibility at corresponding marker gene loci for each
147 of these neuronal classes (Figure 2B-E; Figure S4E-F). Utilizing all DRG neurons that passed

quality controls in mouse or human (Figure S4C-D), we next performed genome-wide differential chromatin accessibility analysis across distinct DRG neuronal subtypes in each species. We identified 575,462 peaks in mice and 9,370 peaks in humans that are differentially accessible ($log_2FC > 0.5$, adj. p-val < 0.05 [mouse], < 0.1 [human]) in one DRG neuronal subtype compared to the others (Figures 2C, 2E, Tables S10 and S11). The difference in numbers of differentially accessible peaks between mouse and human is largely driven by their respective neuron numbers; downsampling mouse DRG neuron numbers to match human neuron numbers yielded ~12,000 differentially accessible peaks ($log_2FC > 0.5$, adj. p-val < 0.05). Differentially accessible peaks across mouse DRG neuronal subtypes are significantly more likely to overlap with differentially accessible peaks across human DRG neuronal subtypes compared to random length-matched ENCODE CREs (Fisher's exact test, $p < 10^{-16}$) , supporting some conservation of chromatin accessibility between mouse and human DRG neurons.[34]

We next used SCENIC+ to prioritize differentially accessible peaks that may function as neuronal subtype-specific CREs in both humans and mice. In mice, SCENIC+ nominated 484 'eRegulons' and 61,044 putative CREs, and in humans, SCENIC+ nominated 53 eRegulons and 3,384 putative putative CREs (Figures 2F-G, Tables S12 and S13). Clustering DRG neurons by their eRegulon activities was sufficient to resolve known neuronal identities (Figure S4G), highlighting the potential of these inferred gene regulatory networks to identify key TFs and candidate CREs that drive neuronal subtype-specific gene expression in both mouse and human. For example, SCENIC+ analyses identified increased RUNX3 eRegulon activity in proprioceptors compared to other neuronal subtypes and a putative CRE upstream of the proprioceptor-specific gene, *Pvalb*, which is consistent with the known role of RUNX3 in proprioceptor development (Figures 2F, S4H).[35,36] In C-PEP nociceptors, SCENIC+ implicates BCL11A with increased eRegulon activity compared to other neuronal subtypes in both mouse and human and nominates a putative CRE upstream of the nociceptor-enriched gene, *Sstr2* (Figures 2F-G, S4H).

These findings suggest that our mouse and human DRG multi-omic atlases and downstream gene regulatory network analyses could help identify CREs that are capable of driving nociceptor-specific gene expression in both species, a critical requirement for developing translational viral tools with targeted activity in nociceptors. Toward that goal, we found that ~16% of all differentially accessible peaks between mouse neuronal subtypes are also linked to a SCENIC+ eRegulon and that ~18% of these have syntenic peaks in human DRG neurons (Figure 2I; Table S14). This analysis highlights a core set of CREs that are implicated in cell-type-specific gene regulation and have the potential to function similarly across species.

In vivo screening of candidate nociceptor-specific CREs

To functionally screen for CREs that drive gene expression preferentially in nociceptors *in vivo*, we constructed AAVs in which prioritized CREs are inserted upstream of a minimal promoter (mini CMV) and a reporter gene (EGFP). AAVs were then delivered to mice intracerebroventricularly (ICV), and reporter expression was quantified in each DRG neuronal subtype using fluorescent *in situ* hybridization (FISH) and/or scRNA-seq (Figure 3A). We initially selected 9 CREs (Figure 3B; see Methods) based on their differential accessibility patterns in

two important classes of nociceptors in pain transduction, C-PEP (CRE1-7) or C-NP (CRE8-9) nociceptors. We also included two controls: an AAV with only a minimal promoter driving reporter expression (No CRE) and an AAV with the ubiquitous CAG promoter driving reporter expression. The 11 AAVs (9 nociceptor CREs, 2 controls) were packaged in the AAV-PHP.S capsid[37] and injected ICV into neonatal mice to ensure broad DRG infectivity;[38–40] DRGs were collected at least 6 weeks post-injection.

We used FISH to quantify co-expression of viral eGFP with either the C-PEP nociceptor marker, *Tac1*, or the C-NP nociceptor marker, *Mrgprd*. Two of the nine tested C-PEP nociceptor CREs (CRE1, CRE6) drove reporter expression with significantly greater specificity in *Tac1*+ neurons than the CAG control (Figure 3C) (One-way ANOVA, $F(3,27) = 6.80$, $p < 0.001$; Dunnett's multiple comparisons test). We additionally confirmed that CRE1, which displayed the greatest specificity for C-PEP by FISH, drives limited expression in off-target cell types (*Mrgprd*+ C-NP nociceptors, *Hapln4*+ Aβ-LTMRs and proprioceptors, and *Fam19a4*+ C-LTMRs) (Figure S5A). We also identified one of the C-NP targeting CREs, CRE8, drives eGFP expression significantly more specifically in *Mrgprd*+ C-NP neurons than the CAG control (Figure 3C) (One-way ANOVA, $F(2,6) = 7.42$, $p = 0.024$; Dunnett's multiple comparisons test).

As neither *Tac1* nor *Mrgprd* marker genes label all C-PEP or C-NP nociceptors, respectively (Figure 3, S5A), we turned to scRNA-seq to provide an orthogonal and more quantitative assessment of AAV reporter expression for CREs that displayed promising nociceptor specificity by FISH. In particular, CRE1, CRE6, CRE8, and control (No CRE, CAG) AAVs were delivered ICV to neonates as described above, followed by whole cell dissociation of DRGs and scRNA-seq at least 6 weeks post-injection. In total, 34,899 cells were profiled by scRNA-seq across all CREs and controls, with 2-4 biological replicates per CRE. Neuronal subtype annotations were obtained by anchoring scRNA-seq data to the mouse DRG reference atlas[6] (Figure 3D) and AAV transcripts were quantified in each neuronal subtype. C-PEP target specificity ratios (AAV+ C-PEP cells / all other AAV+ cells) and C-NP target specificity ratios (AAV+ C-NP cells / all other AAV+ cells) were calculated for each CRE and compared to control AAVs. Consistent with the FISH results, we observed that CRE1 and CRE6 drive reporter expression more specifically in C-PEP nociceptors than No CRE or CAG controls (~3-5 fold, Fisher's exact test: $p < 0.001$), and CRE8 drives reporter expression more specifically in C-NP nociceptors than No CRE or CAG controls (~4-6 fold, Fisher's exact test: $p < 0.001$) (Figure 3E). The additional granularity of scRNA-seq revealed that, of C-PEP nociceptors (C-PEP.Calca+Sstr2, C-PEP.Calca+Oprk1, C-PEP.Calca+Dcn), both CRE1 and CRE6 drive expression preferentially in C-PEP.Calca+Sstr2 neurons compared to other C-PEP subtypes and that, of C-NP nociceptors (C-NP.Mrgprd, C-NP.Mrgpra3+Trpv1, C-NP.Mrgpra3+Mrgprb4), CRE8 drives reporter expression preferentially in C-NP.Mrgprd neurons compared to other C-NP subtypes (Figure S5B). These scRNA-seq results were orthogonally assessed by FISH (Figure S5C), where we observed that 56 ± 2% of CRE1 and 45 ± 13% of CRE6 driven reporter positive cells were C-PEP.Calca+Sstr2 compared to 28 ± 3% of No CRE and 36 ± 1% of CAG controls, respectively (Figure S5D) (One-way ANOVA, $F(3,8) = 2.005$, $p = 0.0404$, Dunnett's multiple comparisons test for No CRE versus CRE1. Data are presented as mean ± SEM).

Given the low success rate and high cost of screening individual CREs, we turned to a multiplexed AAV screen to functionally assess CRE activity *in vivo* more efficiently (Figure 4A).[41]

233 We selected 106 CREs predicted to drive expression in C-PEP nociceptors, C-NP nociceptors,
234 *Sst*+ pruriceptors, or *Mrgpra3*+ pruriceptors, or combinations of nociceptor classes, and
235 constructed an AAV library in which each CRE was cloned in *cis* with unique DNA barcodes
236 (see Methods). We also included a No CRE control that was uniquely barcoded.
237 Next-generation sequencing of the resulting AAV library confirmed that all CREs were
238 represented (Figure S6A). After packaging the CRE library in AAV-PHP.S, the AAV library was
239 injected ICV into three neonatal mice and scRNA-seq of their DRGs was performed >6 weeks
240 later as described above. After quality control, clustering, and cell type annotation, a total of
241 10,442 DRG cells were included for downstream analysis. Of these, 2,275 cells expressed AAV
242 transcripts covering 106 CREs (Figure 4B), of which 41 CREs were detected in at least 6 cells
243 and included in downstream analyses (Figure S6B). We observed that certain CREs drove
244 barcode expression preferentially in distinct DRG neuronal cell classes compared to the No
245 CRE control (Figure 4C), with scRNA-seq providing the resolution to identify expression within
246 specific subtypes (Figure S6C). For example, CREs accessible in peptidergic neurons (Figure
247 4C) drove expression in C-PEP (CRE70 and CRE71) or A-PEP (CRE18) nociceptors (Figure
248 S6C). Similarly, CREs accessible in non-peptidergic cells drove expression in C-NP.Mrgpra3
249 pruriceptors (CRE45 and CRE47) or C-NP.Sst pruriceptors (CRE114). This specificity extended
250 to non-nociceptive lineages, where CRE68 and CRE79 drove preferential expression in
251 A-Propr.Pvalb and C-LTMR.Th cell types, respectively.

252 We next examined the extent to which subtype-specific chromatin accessibility predicts
253 CRE-driven reporter expression across DRG neuronal subtypes. We calculated expected on/off
254 target expression ratios for each CRE based on the pattern of differential chromatin accessibility
255 at each CRE's locus (see Methods; Table S15); CRE specificity ratios ranged from 0 to >100 for
256 their respective on-target DRG neuron class. Overall, we observed a positive correlation
257 (Spearman's $r = 0.26$, $p < 0.05$) between the measured differential chromatin accessibility at
258 each CRE locus and the CRE-driven reporter expression in each DRG cell type (Figure 4D),
259 consistent with our attempt to prioritize functional CREs for *in vivo* testing. Interestingly, a subset
260 of CREs in our multiplex screen exhibited high differential accessibility in their target cell types
261 compared to their off target cell types, but their corresponding AAV barcodes were only detected
262 in off target cell types (On/off target expression ratios = 0, > 6 cells). While our goal was to
263 identify CREs that act as cell-type-specific enhancers, our findings suggest that some putative
264 CREs might instead function as cell-type-specific silencers or require additional native chromatin
265 context to enhance expression in their on-target cell types. Excluding these discordant CREs
266 significantly improved the correlation between on target accessibility and on/off target AAV
267 expression ratio from 0.26 to 0.47 ($p < 0.05$, Figure 4D).

268 Taken together, this multiplexed CRE scRNA-seq screen enabled the functional characterization
269 of dozens of CREs in parallel as well as the discovery of several new enhancer AAVs capable of
270 targeting distinct DRG neuronal subtypes.

271 *Modeling cis-regulatory logic of DRG neurons*

272 The correlation of cell-type-specific chromatin accessibility with CRE activity in DRG neurons
273 prompted us to ask if we could quantitatively model the *cis*-regulatory logic of DRG nociceptors.
274 Previous studies have leveraged machine learning approaches such as convolutional neural

275 networks (CNNs) to develop models trained on snATAC-seq data that can predict
276 cell-type-specific chromatin accessibility based on sequence alone.[42–44] We thus constructed a
277 CNN, Predicting Accessibility In Nociceptors (PAIN-net), trained on DNA sequences from our
278 mouse DRG snMultiome data, to predict both the neuronal subtype-specific chromatin
279 accessibility signal intensity and the Log_2FC of differential accessibility in each neuronal subtype
280 compared to all other subtypes (see Methods). We tested model performance on randomly
281 selected sequences comprising ~15% of the mouse DRG snMultiome data that was excluded
282 from training, and observed a median Pearson's correlation of 0.72 across cell types between
283 model-predicted and measured differential accessibility (Figure 5A; Figure S7A; Table S16).
284 This model's performance is comparable to chromatin accessibility predictive models generated
285 in other tissues.[43]

286 A key motivation for developing PAIN-net was to prioritize CREs that drive cell-type-specific AAV
287 expression in DRG. We thus asked whether PAIN-net-predicted CRE accessibility across DRG
288 cell types is associated with each CRE's *in vivo* specificity, as measured by their scRNA-seq
289 on/off target expression ratio. PAIN-net correctly predicted the nociceptor subtypes in which
290 CRE1, CRE6 and CRE8 were most active *in vivo*: C-PEP.Calca+Sstr2 (CRE1 and CRE6) and
291 C-NP.Mrgprd (CRE8) (Figure 5B). We also observed good model performance across the *in*
292 *vivo* multiplexed CRE screen; PAIN-net predictions were positively correlated with the measured
293 on/off target expression ratios measured by scRNA-seq (Spearman's $r = 0.33$; p-value <0.05;
294 Figure 5C; correlation between predicted and measured accessibility in Figure S7B). These
295 findings suggest that PAIN-net can be a valuable tool for nominating CREs likely to drive
296 neuronal subtype-specific expression in the DRG.

297 We next asked which TF motifs are enriched in the sequences PAIN-net predicts to have the
298 greatest differential accessibility in each DRG neuronal subtype compared to random
299 sequences. PAIN-net identified TFs known to contribute to sensory neuron development (e.g.
300 BCL6 in C-PEP nociceptors, RUNT family members in proprioceptors, RUNX1 in C-LTMR.Th
301 neurons, and ATF3 in Atf3 neurons),[13,45,46] as well as several new TFs that warrant further
302 investigation (Figure 5D). These results are also in line with those observed in our SCENIC+
303 analysis of chromatin accessibility data (Figures 2F-G).

304 We next employed PAIN-net to assess the independent regulatory potential of key transcription
305 factors by embedding their consensus binding motifs into random background sequences and
306 predicting the resulting chromatin accessibility across neuronal subtypes. We found that *in silico*
307 disruption of these embedded ATF3, RUNX1, and ISL1 consensus motifs significantly altered
308 predicted accessibility (FDR < 0.01), whereas mutations in the flanking random regions yielded
309 negligible effects (Figure S7C). Importantly, disruption of TF consensus motifs preferentially
310 reduced chromatin accessibility in the specific neuronal subtypes in which the corresponding TF
311 functions (e.g., ATF3 in the *Atf3* neurons) (Figure S7C). This indicates that PAIN-net has
312 successfully captured cell-type-specific *cis*-regulatory dependencies of known motifs.

313 Having tested the model on known consensus sites, we next sought to map the precise
314 nucleotides driving accessibility in the nociceptor CREs we screened *in vivo*. To do this, we
315 performed *in silico* saturation mutagenesis using PAIN-net to systematically predict the impact of
316 every possible single-nucleotide substitution on chromatin accessibility across each CRE. *In*

317 *silico* mutagenesis revealed that the most impactful substitutions clustered at the center of the
318 CRE sequences (Figure 5E), coinciding with the summits of differentially accessible peaks.
319 Notably, the predicted magnitude of these substitutions was greater for a CRE's on-target cell
320 types than its off-target type (Figure 5E), suggesting that the model captures cell-type-specific
321 regulatory logic. Applying this unbiased mutagenesis to CRE1 revealed that the specific
322 nucleotides driving differential accessibility coincide with known transcription factor binding
323 motifs, including those for BCL6/11, RUNT, Homeobox, and AP-1 (Figure 5F). Collectively, these
324 findings indicate that PAIN-net captures the *cis*-regulatory logic used by DRG neuronal subtypes
325 to establish unique gene expression patterns. This framework provides a powerful tool to both
326 nominate native nociceptor CREs for testing *in vivo* and to design synthetic elements with the
327 potential to achieve greater cell-type-specificity than is observed with natural CREs.[31,35,36]

328 *Functional and translational applications of nociceptor-specific AAVs*

329 Given the long-term goal of developing cell-type-specific AAVs to inhibit nociceptor activity, we
330 next generated AAVs in which either the nociceptor-specific CRE1 or the pan-expressing CAG
331 promoter drives expression of the inward-rectifying potassium channel $K_{ir}2.1$ and mCherry
332 (Figure 6A). We also constructed a control AAV using the CAG promoter to drive expression of
333 mCherry alone. AAVs were packaged in the AAV-PHP.S capsid[37] and delivered *in vitro* to
334 dissociated DRG cells cultured for whole cell patch clamp electrophysiology. CAG-$K_{ir}2.1$- and
335 CAG-mCherry-transduced neurons were identified by bright mCherry fluorescence and
336 CAG-$K_{ir}2.1$ neurons exhibited significantly reduced excitability compared to mCherry-transduced
337 neurons as measured by an increase in the amount of injected current required to stimulate
338 action-potential firing (rheobase) (Figures 6B-C). Expression of mCherry in
339 CRE1-$K_{ir}2.1$-transduced neurons was too dim to reliably distinguish all cells from
340 autofluorescence, but we observed a population of small-to-medium diameter neurons[47,48] with
341 rheobase currents higher than any measured in mCherry control cultures ($p < 0.0001$, One-way
342 ANOVA; Figures 6B-C; Figure S8A; see Methods). Both CAG-$K_{ir}2.1$ and high rheobase
343 CRE1-$K_{ir}2.1$ neurons demonstrated hyperpolarized membrane potentials ($p < 0.0001$, One-way
344 ANOVA; Figure 6D), and larger inward-rectifying currents at hyperpolarized potentials (Figure
345 6E). The increased potassium conductance at sub-threshold voltages likely underlies this
346 observed hyperpolarization in the resting membrane potential and is consistent with the known
347 hyperpolarizing effects of $K_{ir}2.1$ overexpression.[49] Collectively, these data demonstrate that
348 CRE1-driven $K_{ir}2.1$ expression reduces DRG neuronal excitability.

349 To assess whether our nociceptor CREs are likely to drive AAV gene expression preferentially in
350 human nociceptors, we cultured human iPSCs (LiPSC-GR1.1) and differentiated them using
351 protocols that produce either nociceptor-like sensory neurons or cardiomyocytes.[50,51] We
352 transduced these cells with AAV-PHP.S vectors expressing an eGFP reporter under the control
353 of CRE1, CRE8, or the ubiquitous CAG promoter. After 3-4 weeks in culture for functional
354 maturation, we assessed reporter expression by immunofluorescence microscopy. We observed
355 robust reporter expression driven by the CAG promoter in both iPS nociceptors and
356 cardiomyocytes, but CRE1 and CRE8 only drive reporter expression in nociceptors (Figure 6F).
357 These findings support the possibility that nociceptor CREs identified and screened in mice can
358 function similarly in nociceptors across species.

Discussion

Peripheral sensory neurons comprise diverse subtypes that collectively encode touch, temperature, itch, and pain. Chronic pain is linked to nociceptor hyperexcitability, yet translating mechanistic insight into precise therapeutics has been constrained by the lack of subtype-selective access across species.[52,53] Here we built matched single nucleus multi-omic atlases of mouse and human DRG, defined conserved cell-type-specific regulatory programs, and identified CREs that drive gene expression preferentially in nociceptor subtypes *in vivo* and in human iPSC-derived nociceptor-like neurons. We also developed a sequence-based model (PAIN-net) that captures key components of nociceptor *cis*-regulatory logic, enabling both prioritization of CREs likely to function as enhancers *in vivo* and design of synthetic elements with tunable properties. Together, these resources (available at www.painseq.com) support a generalizable strategy for cell-type-specific neuromodulation in the peripheral nervous system.

Enhancer AAV tool development relies on cell-type–resolved epigenomic maps; in the CNS, such maps have enabled nomination of CREs that function *in vivo* when placed upstream of minimal promoters in AAV cassettes.[14–17] By generating matched mouse and human DRG snMultiome atlases, we extend this capability to the peripheral nervous system, where neuronal subtype diversity is well described transcriptionally but has remained difficult to access genetically without transgenic mouse lines. The overlap we observe between mouse and human DRG regulatory features provides a mechanistic rationale for using mouse *in vivo* screens as a discovery engine for elements likely to function in human nociceptors, while also establishing a reference for interpreting species-specific divergence that may influence translation.

Our multi-omic analyses extend prior transcriptomic descriptions of DRG[5,6,30–32] by defining subtype-resolved *cis*-regulatory landscapes and TF-centered regulatory networks. Gene regulatory network analyses highlight candidate regulators across sensory lineages (e.g. enrichment of RUNX-family activity in proprioceptors) and link these programs to specific putative enhancers. This framework supports CRE nomination using convergent evidence, including subtype-enriched accessibility, cross-species sequence conservation, and membership in TF-centered regulatory modules. As chromatin accessibility alone does not guarantee functional enhancer activity *in vivo,* this multi-criterion prioritization framework helps address an important bottleneck in enhancer engineering.[56,57] That said, CRE prioritization remains imperfect because selectivity and expression level can change substantially when CREs are placed in an AAV cassette context. Accordingly, functional AAV screens of prioritized CREs using quantitative, cell-type–resolved readouts remain essential.

Our functional AAV screen in mouse DRG identified a subset of nominated CREs that drive reporter expression preferentially in nociceptor classes, including enhancers biased toward peptidergic (CRE1, CRE6) and non-peptidergic populations (CRE8). These tools were successfully validated using both FISH to assay co-localization with canonical markers and scRNA-seq to quantify full subtype-resolved specificity. While three individually tested CREs showed specificity for their intended nociceptor classes, our multiplexed barcoded CRE screen enabled us to functionally rank dozens of elements within the same AAV library. This approach

identified high-performing nociceptor CREs (e.g., CRE114) as well as elements that preferentially drive expression in non-nociceptive DRG populations (e.g., thermosensory, proprioceptive, or mechanosensory lineages).

The multiplexed screen also provided mechanistic insight into how endogenous chromatin accessibility relates to cell-type expression patterns driven by an AAV. The modest but significant correlation between subtype-enriched chromatin accessibility and *in vivo* barcode expression supports accessibility as a useful guide for enhancer discovery while underscoring the contribution of additional regulatory layers. For example, some CREs showed strong on-target enrichment, whereas other elements with similar accessibility profiles produced off-target expression. We also observed a subset of highly accessible CREs that drove expression only in off-target cells, suggesting that these sequences could act as context-dependent silencers. Together, these observations motivate continued empirical screening and the development of predictive models that can both prioritize candidate CREs and guide the design of synthetic elements that improve specificity, strength, and/or cross-species portability.

To this end, we developed PAIN-net, a sequence-based model that predicts neuronal subtype–specific chromatin accessibility directly from DNA sequence and provides a quantitative framework for interpreting enhancer "grammar." PAIN-net generalizes to held-out loci and tracks measured CRE-driven expression patterns, supporting the view that a substantial component of nociceptor *cis*-regulatory logic is learnable from sequence features. Motif enrichment and *in silico* perturbation analyses further suggest that PAIN-net captures biologically meaningful dependencies on transcription factor binding sites and their arrangement. More broadly, PAIN-net brings model-guided CRE design into the peripheral sensory system and establishes a foundation for optimizing CRE specificity and strength or for combinatorial targeting strategies. In practice, such design capabilities could enable viral tools tailored to distinct experimental or therapeutic goals, including broader or more selective targeting of nociceptor subtypes, or exclusion of specific proprioceptive and mechanosensory neurons to minimize sensory side effects.

Consistent with the potential of nociceptor-biased AAVs as a future pain therapeutic, CRE1 drove sufficient expression of $K_{ir}2.1$ to reduce DRG neuron excitability *in vitro*, with electrophysiological signatures consistent with increased inward-rectifying potassium conductance and hyperpolarization. This result demonstrates that even comparatively modest enhancer-driven expression can yield meaningful functional effects when paired with a potent effector. It also highlights a design principle relevant to therapeutic translation: specificity and expression strength can often trade off. For certain effectors, sub-maximal expression may still provide robust modulation while reducing over-expression toxicity and provide a wider window to fine-tune expression.[58] The robust and specific expression of CRE1 in human iPSC-derived nociceptors over cardiomyocytes supports the translational potential that conserved *cis*-regulatory logic can be leveraged to target human nociceptors.

Several limitations define priorities for future work. First, human neuronal chromatin accessibility data were comparatively sparse, necessitating broader grouping of neuronal subtypes than in

440 mice for portions of the epigenomic analysis. Expanding high-quality human multi-omic
441 sampling across anatomical levels, ages, injury states, and diverse donors will sharpen
442 enhancer nomination and help identify regulatory programs that are human-specific or
443 state-dependent. Second, the diversity and scarcity of sensory neurons in peripheral ganglia
444 fundamentally limits CRE screening throughput. Development of nociceptor iPSC models that
445 better recapitulate *in vivo* nociceptor subtypes could enable ultra-high throughput synthetic
446 enhancer screens that are currently being performed in cell lines in other contexts.[42] Third, our
447 *in vivo* AAV screening used neonatal ICV delivery and a viral capsid optimized for broad DRG
448 transduction in mice. Translation will require validating enhancer performance across multiple
449 delivery routes, capsids, and evaluation of off-target expression across multiple tissues.

450 Despite these limitations, the cross-species DRG multi-omic atlases, predictive models, and
451 nociceptor-selective AAVs presented here provide important first steps toward a modular toolkit
452 for peripheral neuromodulation across species. This framework should extend to other
453 peripheral ganglia and autonomic circuits where genetic access remains limited.

Methods

Animals

All mouse experiments in this study were approved by the National Institutes of Health and the Harvard Medical School IACUC. Experiments followed the ethical guidelines outlined in the NIH Guide for the Care and Use of Laboratory Animals. Mice were housed under standard conditions, with no more than five animals per cage, maintained on a 12 h day/night cycle, with food and water. Both male and female mice were used for experiments, aged between 6-24 weeks, and the minimum number of animals was used for each experimental group. Animals were maintained on a C57BL/6J genetic background.

Mouse nuclear isolation

Dorsal root ganglia from all spinal levels were dissected from 50 mice (males and females) aged 6 to 24 weeks. All DRGs were directly frozen on dry ice and stored at -80C until processing for single-nucleus RNA sequencing (snRNA-seq), using a detergent-based nuclei isolation protocol optimized for neuronal tissue. All reagents were prepared on the same day and kept ice-cold unless otherwise specified. RNase-free technique was used throughout.

On the day of preparation, a swinging-bucket centrifuge was pre-cooled to 4°C. Nuclei lysis buffer (NST), resuspension buffer (ST), and 1× nuclei buffer (10x Genomics) were prepared fresh, supplemented with RNase inhibitor (RNasin Plus, Promega N2611), and maintained on ice. Dounce homogenizers and Tissue Tearor probes were thoroughly cleaned between samples using nuclease-free water and 70% ethanol to prevent cross-contamination. NST (lysis) buffer consisted of 1×stock buffer (20 mM NaCl, 20 mM Tris-HCl pH 7.5, 1 mM $CaCl_2$, 6 mM $MgCl_2$), 0.2% IGEPAL CA-630 (Sigma I8896-50ML), and 0.84% BSA in nuclease-free water. ST buffer consisted of 1×stock buffer supplemented with 0.9% BSA. All buffers were vacuum-filtered, stored at 4°C, and supplemented with RNase Inhibitor immediately prior to use.

Pooled frozen DRGs from 5 mice were processed each time. Tissue was transferred from dry ice and briefly warmed at the tube edges to facilitate homogenization. Samples were lysed in NST buffer containing IGEPAL CA-630 and BSA, supplemented with RNase inhibitor. Tissue was mechanically disrupted using a Tissue Tearor on a low-speed setting (2–3) for 5 seconds, ensuring complete coverage of the tissue. Homogenates were kept on ice immediately following disruption. The homogenate was transferred to a glass Dounce homogenizer and gently washed to recover residual tissue. Samples were further homogenized using a tight pestle with 15 slow strokes, avoiding bubble formation. The resulting homogenate was passed through a 40 µm cell strainer into a pre-chilled 15 mL conical tube, followed by an additional wash of the Dounce homogenizer with ST buffer to maximize nuclei recovery. Filtered homogenates were centrifuged at 500×g for 10 minutes at 4°C. Supernatant was carefully removed, leaving approximately 0.5 mL above the pellet. The nuclei pellet was gently resuspended in 1× nuclei buffer (10x Genomics, PN-2000207) as needed.

Nuclei suspensions were transferred for fluorescence-activated nuclei sorting (FANS) performed by the BWH Center for Cellular Profiling. Collection tubes were pre-coated with 1× nuclei buffer to reduce nuclei loss. Sorting was performed using a 70 µm nozzle. Nuclei were gated based on GFP signal (neuronal nuclei are GFP+) at forward and side scatter, exclusion of debris, and DNA staining using 7-aminoactinomycin D (7-AAD; final concentration 10 ng/µL). Negative controls were used to establish gating parameters.

Sorted nuclei were collected directly into tubes containing 100ul of nuclei buffer. To concentrate nuclei in 5ul 1× nuclei buffer for downstream 10x Genomics multiome application, concanavalin A-conjugated paramagnetic beads (CUTANA) were used as previously described.[59] For snRNA-seq experiments, nuclei were taken directly into library preparation without additional purification or centrifugation to minimize loss.

Nuclei concentration was determined using a hemocytometer following dilution with Trypan blue or DAPI staining. Counts were averaged across four quadrants and used to calculate nuclei concentration (nuclei/µL). An aliquot of nuclei was optionally inspected by fluorescence microscopy to confirm nuclear integrity and morphology.

Mouse snMultiome-seq library preparation and sequencing

Nuclei suspensions were concentrated and approximately 10,000 nuclei per library were loaded on the 10x Genomics Chromium Controller, following the manufacturer's guidelines for the Chromium Next GEM Single Cell Multiome ATAC + Gene Expression kit. Library preparation, GEM generation, and barcoding were performed according to the manufacturer's protocol.

Mouse multi-ome data processing

Read alignment and ambient RNA correction

Raw reads collected from mice DRG samples were aligned to the Mouse mm10 (GENCODE vM23/ensembl98) mouse reference genome using the Cell Ranger-ARC pipeline (v2.0.0; default settings). The reference genome was downloaded from the 10x Genomics website. To ensure high-quality data integration, RNA counts were filtered for ambient noise using CellBender[60] (default settings).

RNA modality

Seurat v5 was used to process the RNA modality by combining all preprocessed in-house and published ambient RNA corrected count matrices (18 libraries total). Nuclei with more than 1,000 detected genes and less than 25% mitochondrial counts were retained. Filtered data were normalized using log-normalization, and the top 2,000 variable genes were identified, excluding mitochondrial genes. Data were then scaled, and principal component analysis (PCA) was

537 performed using the selected variable genes. Batch correction and integration across libraries
538 were performed using the Seurat CCA method. The number of principal components (PCs) was
539 determined as the point where selected PCs cumulatively explained at least 90% of the
540 variance and additional PCs explained less than 5% of the standard deviation of highly variable
541 gene expression. We then clustered the data at a resolution of 1 and clusters were embedded in
542 UMAP space. Clusters were then annotated based on the enrichment ($Log_2FC > 0.5$, adjusted
543 p-value < 0.05 relative to all other clusters) of the following marker genes: *Snap25*, *Rbfox3*, and
544 *Pvalb,* as neurons; *Fabp7* and *Ednrb* as for SGCs; *Mpz* and *Mbp* for mySCs; *Scn7a* for nmSC;
545 *Ptprc and Ccr2*for Immune cells; *Pecam1* and *Flt1* for Endothelial cells; *Notch3* and *Kcnj8* for
546 Pericytes; *Pdgfra* and *Tbx18* for Fibroblasts. Clusters showing enrichment of both neuronal and
547 non-neuronal markers were annotated as doublets and removed.

548 Neurons were then subsetted and re-clustered. For this step, the QC threshold was increased to
549 retain nuclei with more than 1,500 detected genes and less than 15% mitochondrial gene
550 counts. We then predicted neuronal subtype identity from the mouse harmonized DRG atlas[6]
551 using Seurat's label transfer method, and excluded nuclei with a prediction score less than 0.5.

552 With the exception of three mouse libraries, all libraries had a mixture of male and female mice
553 and thus we performed an additional computational step in assigning sex labels to each
554 nucleus. Sex labels were assigned to individual nuclei based on expression of the
555 Y-chromosome gene *Uty* and the X-inactivation transcript *Xist*. Cells were classified as "Male" if
556 *Uty* expression exceeded the 75th percentile while *Xist* remained below the 75th percentile, and
557 as "Female" if *Xist* exceeded the 75th percentile while *Uty* remained below the 75th percentile;
558 cells not meeting these criteria were initially labeled as "Ambiguous". For the three libraries
559 where only one sex was sequenced, sex labels were manually assigned (mDRG_rest_day3,
560 Techameena_ctrl_F, and Techameena_ctrl_M).
561

562
563 *ATAC modality*

564 Signac (v1.16.0) was used to process the ATAC modality. Peak files (.bed) and fragment files
565 generated by Cell Ranger ARC were used to construct peak-by-cell count matrices for each of
566 the 18 libraries, which were then merged into a single object. ATAC data were subsetted to
567 retain only RNA-matched cells, and RNA-derived cell identities were transferred to the ATAC
568 modality based on shared barcodes. Nuclei with fewer than 1,000 total fragments in peaks or
569 TSS enrichment scores below 3 were excluded. Cell-type specific peaks were called using
570 MACS2 in BEDPE mode, restricted to standard chromosomes and excluding blacklist regions.
571 The resulting peak set was used for all downstream analyses. This workflow was applied both to
572 the full set of annotated cell types (neuronal and non-neuronal populations) and to a
573 neuron-focused analysis in which neuronal subtypes were considered separately.
574
575
576 *Pseudobulk-based differential accessibility of mouse DRG cell types for CRE prioritization*

To identify differentially accessible (DA) regions in mouse DRG cell types, we used a pseudobulk strategy implemented in Seurat/Signac and DESeq2. Briefly, cells were grouped by annotated cell type and randomly assigned to one of n = 10 synthetic pseudoreplicates per cell type (set.seed = 42 for reproducibility). For each cell type-pseudoreplicate combination, raw accessibility counts from the MACS2-called peak matrix were aggregated using AggregateExpression to generate a pseudobulk count matrix with one column per pseudoreplicate.

For DA testing, we performed repeated one-vs-rest comparisons. A model was fit with design ~ celltype_replicateID. Peaks with low overall signal were removed prior to modeling (retaining features with total counts ≥ 10 across all pseudobulk samples). Size factors were estimated using DESeq2's poscounts method to accommodate sparsity typical of accessibility data, and the negative binomial model was fitted using fitType = "local." Peaks were considered differentially accessible based on adjusted p-value thresholds (FDR < 0.05 for all analyses) and effect size cutoffs ($\log_2$FC > 0.5), as reported.

Gene regulatory network analysis

To infer cell type-specific gene regulatory networks for the mouse DRG, we used SCENIC+ which required two inputs: (1) the pseudobulk-based DA regions in mouse DRG cell types, (2) a raw gene expression matrix of annotated mouse nuclei and (3) the binarized topic-region associations outputted from pycisTopic (method details below).

PycisTopic[24] was used to group peaks into groups of regions with a common accessibility profile across cells (topics). Briefly, peak accessibility profiles were exported from a Signac object as a sparse peak-by-cell count matrix (MatrixMarket format) together with matching cell barcodes, peak coordinates, and per-cell metadata. These inputs were used to construct a cisTopic object, while filtering regions overlapping the ENCODE blacklist. Cell-level annotations were then added to the cisTopic object to enable downstream visualization and interpretation.

Topic modelling was performed using latent Dirichlet allocation (LDA) implemented in pycisTopic with MALLET. Multiple models were trained across a range of topic numbers (10–200 topics), using default parameters (seed=555) and multi-core parallelization (60 CPUs). Trained models were subsequently evaluated using standard topic model selection metrics (Arun_2010, Cao_Juan_2009, Mimno_2011, and model log-likelihood). Based on these diagnostics, a model with 48 topics was selected for major DRG cell classes, and for DRG neuronal subtypes 120 topics were selected. Topics were binarized to derive discrete topic-associated regulatory region sets ("candidate enhancers"), using two complementary approaches: Otsu thresholding and a rank-based strategy selecting the top 3,000 regions per topic. Binarized topic-region associations were exported as sorted BED files as the input for SCENIC+.

For the pycisTopic and SCENIC+ analysis of the mouse multi-ome data across major cell classes (neuronal and non-neuronal populations), the dataset was first randomly downsampled

618 to a maximum of 1,000 nuclei per cell type. For the pycisTopic and SCENIC+ analysis of
619 neuronal DRG subtypes, the full neuronal dataset was used.
620
621
622 **Human DRG acquisition and dissection**

623 All procedures for human tissue procurement and ethical oversight were reviewed and approved
624 by the Institutional Review Board at Brigham and Women's Hospital [2024P002201/MGB1924].
625 Protocol IRB#2017P000757 was reviewed by Mass General Brigham IRB and determined to be
626 not human research. DRGs were collected from consented donors using a rapid autopsy
627 protocol at Mass General Brigham as described in Bhuiyan et al., 2024.[6] All donors were
628 deidentified. Age and sex of donors are reported in Table S1.

629
630 **Human nuclear isolation**

631 Single nuclei of human DRGs were collected using a previously described gradient protocol at
632 Harvard Medical School.[61] Briefly, human DRGs were initially pulverized on dry ice and stored at
633 -80C until processing. Approximately pulverized tissue from 4 donors was used per experiment.
634 and ~0.5 to 1 cm^3 of powder was placed into homogenization buffer [0.25 M sucrose, 25 mM
635 KCl, 5 ml of $MgCl_2$, 20 mM tricine-potassium hydroxide (KOH; pH 7.8), 1 mM dithiothreitol
636 (DTT), actinomycin (5 µg/ml), 0.04% BSA, and ribonuclease (Rnase) inhibitor (0.1 U/µl)] for ~15
637 seconds on ice. After the brief incubation for 15 seconds on ice, samples were transferred to a
638 Dounce homogenizer for an additional 10 strokes with a tight pestle in a total volume of 5 mL of
639 homogenization buffer. IGEPAL was added to a final concentration of 0.32%, and five additional
640 strokes were performed with the tight pestle. The tissue homogenate was then passed through
641 a 40-µm filter and diluted 1:1 with a working solution. Neuronal nuclei were enriched as
642 previously described[20] and concentrated using concanavalin A-conjugated paramagnetic beads
643 (CUTANA).[59]

644
645 **Human snMultiome-seq library preparation and sequencing**

646 Nuclei were labeled with DAPI to inspect their integrity under a fluorescent microscope and
647 determine their quantity manually using a hemocytometer. Approximately, 10,000 nuclei were
648 used for each ATAC and gene expression library. Libraries were generated according to the
649 Chromium Next GEM Single-cell Multiome ATAC + Gene Expression guidelines (10x
650 Genomics). Both ATAC and RNA libraries were shipped to GENEWIZ (Azenta Life Sciences)
651 and were sequenced based on the sequencing parameters suggested by 10x Genomics on the
652 Illumina NovaSeq X Plus platform to an approximate read depth of ~25,000 reads per cell
653 (multiome) and ~20,000 reads per cell (RNA).

654
655 **Human multi-ome data processing**

656 *Read alignment and ambient RNA correction*

657 Raw reads from the human samples were aligned to the Human GRCh38 (GENCODE
658 v32/Ensembl98) using the Cell Ranger-ARC pipeline (v2.0.0; default settings). The reference

genome was downloaded from the 10x Genomics website. To ensure high-quality data integration, RNA counts were filtered for ambient noise using CellBender[60] (default settings).

Genotyping and demultiplexing

To maximize cost and efficiency, snRNA-seq was performed on batches of 1-13 donors. We used Souporcell to demultiplex sequencing reads to the level of individual donors. Briefly, Souporcell detects genetic variants within sequenced RNA, generates a cell by variant genotype matrix, and clusters cells based on their allelic assignments.[62] Souporcell version 2.5 was run using default settings. Clusters were initialized with known genotypes (imputed from DNA sequencing) provided by the -known_genotypes option. As input, we combined single cell RNA runs containing the same donors. We verified that samples contained the expected donors by cross-correlating cluster genotypes for each sample against imputed genotypes for all donors. We restricted analysis to Souporcell clusters whose genotypes were correlated with imputed genotypes with a Pearson correlation coefficient of greater than 0.7.

For genotyping, DNA purification was performed with up to 25mg of tissue using the QIAamp Fast DNA TissueKit and sent to Azenta Life Sciences for 4X WGS. Imputation was performed using GLIMPSE2 and the HGDP-1kGP reference panel. Only common variants (minor allele frequency > 5%) were used for Souporcell cluster initialization.

RNA modality

Seurat v5 was used to process the RNA modality by combining all ambient RNA corrected count matrices (33 libraries total). Nuclei with more than 500 detected genes and less than 20% mitochondrial counts were retained. Filtered data were normalized using log-normalization, and the top 2,000 variable genes were identified. Data were then scaled, and PCA was performed using the selected variable genes. Batch correction and integration across libraries were performed using the Seurat CCA method. The number of PCs was determined as the point where selected PCs cumulatively explained at least 90% of the variance and additional PCs explained less than 5% of the standard deviation of highly variable gene expression. We then clustered the data at a resolution of 0.7 and clusters were embedded in UMAP space.

Clusters were then annotated based on the enrichment ($Log_2FC > 0.5$, adjusted p-value < 0.05 relative to all other clusters) of the following marker genes: *SNAP25*, *PRPH*, and *NEFH*, as neurons; and *SPARC*, *MPZ*, and *PTPRC*, as non-neuronal populations. Clusters showing enrichment of both neuronal and non-neuronal markers were annotated as doublets and removed. Non-neuronal clusters were then further annotated using *FABP7* and *EDNRB* as markers for SGCs; *MPZ* and *MBP* for mySCs; *SCN7A* for nmSC; *PTPRC*, *CCR2* and *CD3E* for Immune cells; *PECAM1* and *FLT1* for Endothelial cells; *NOTCH3* and *KCNJ8* for Pericytes; and *PDGFRA* and *TBX18* for Fibroblasts.

Neurons were then subsetted for focused subtype analysis. We then predicted neuronal subtype identity using Seurat's label transfer method by anchoring the neuronal subset to a human DRG atlas.[5] For this step, anchor features were defined as genes shared between the dataset and the atlas, after removing non-coding RNAs and unknown genes. Nuclei with prediction scores greater than 0.5 were retained for downstream analysis.

To assign nuclei to respective human donors, Souporcell[62] was used in genotype aware mode for all samples, except H32_NeuN for which genotype sequencing data was not available.

ATAC modality

For quality control, data were filtered using ArchR[63] (v1.0.3) (number of fragments per cell > 1000, TSS Enrichment > 3). Signac (v1.16.0) was used to process the ATAC modality. Peak files (.bed) and fragment files generated by Cell Ranger ARC were used to construct peak-by-cell count matrices for each of the 18 libraries, which were then merged into a single object. ATAC data were subsetted to retain only RNA-matched cells, and RNA-derived cell identities were transferred to the ATAC modality based on shared barcodes. Cell-type-specific peaks were called using MACS2 in BEDPE mode, restricted to standard chromosomes and excluding blacklist regions. For neurons, the MACS2 peak calling was done with the "broad" parameter set to TRUE due to higher sparsity of the data. The resulting peak set was used for all downstream analyses. This workflow was applied both to the full set of annotated cell types (neuronal and non-neuronal populations) and to a neuron-focused analysis in which neuronal subtypes were considered separately.

Pseudobulk-based differential accessibility of human DRG cell types for CRE prioritization

Differential chromatin accessibility was assessed on a per–cell type basis using a pseudobulk strategy implemented in Seurat/Signac and DESeq2. Briefly, cells were grouped by annotated cell type and randomly assigned to one of n = 10 synthetic pseudoreplicates per cell type (set.seed = 42 for reproducibility). For each cell type–pseudoreplicate combination, raw accessibility counts from the MACS2-called peak matrix were aggregated using AggregateExpression to generate a pseudobulk count matrix with one column per pseudoreplicate.

For DA testing, we performed repeated one-vs-rest comparisons. A DESeq2 model was fit with design ~ celltype_replicateID. Peaks with low overall signal were removed prior to modeling (retaining features with total counts ≥ 10 across all pseudobulk samples). Size factors were estimated using DESeq2's poscounts method to accommodate sparsity typical of accessibility data, and the negative binomial model was fitted using fitType = "local". Peaks were considered differentially accessible based on adjusted p-value thresholds (FDR < 0.05 for analyses for DRG cell types; FDR < 0.1 for DRG neuronal subtypes) and effect size cutoffs (log2FC > 0.5), as reported.

Gene regulatory network analysis

To infer cell type-specific gene regulatory network for the human DRG, we used SCENIC+, which required three inputs: (1) the pseudobulk-based DA regions in human DRG cell types, (2) the raw gene expression matrix of annotated human nuclei, and (3) the binarized topic-region associations outputted from pycisTopic. For pycisTopic, we used the same protocol as we did with mice, with the exception of using the ENCODE black list for humans and no downsampling was performed for any analyses.

Differential gene expression across sexes

To identify sex-specific transcriptomic differences, we performed pseudobulk differential expression analysis using the Seurat and DESeq2 packages in R. We aggregated single-cell RNA counts by sample, sex, and cell type using AggregateExpression to create pseudobulk profiles, subsequently extracting metadata to reconstruct the experimental design. For each cell type with sufficient replication (minimum 2 replicates per sex), we performed differential expression analysis between male and female samples using FindMarkers with the DESeq2 test (test.use = "DESeq2"). To visualize the results, we generated volcano plots using ggplot2 and ggrepel, identifying significant genes (adjusted p-value < 0.05 and $|\log_2 FC| > 0.5$).

CRE prioritization

To prioritize candidate CREs for functional screening, we first identified genomic regions exhibiting significant differential chromatin accessibility in target neuronal subtypes (e.g., peptidergic or non-peptidergic nociceptors) compared to all other DRG cell types ($\text{Log}_2FC > 0.5$, adj. p-value < 0.05). These candidate peaks were further filtered for evolutionary conservation by mapping mouse sequences to the human genome using the liftOver tool, ensuring the selection of elements with conserved regulatory potential. We additionally prioritized CREs predicted to function as active enhancers by SCENIC+, specifically those linked to cell-type-specific eRegulons or containing binding motifs for key lineage-defining transcription factors (e.g., RUNX1, BCL11A). Final candidate enhancer sequences were confirmed using custom tracks visualized via the UCSC Genome Browser. Genomic coordinates for peaks of interest (e.g., mm10) were located, and DNA sequences were extracted directly from the browser. For optimal cloning efficiency, enhancers were selected based on a size constraint of less than 1 kb in length.

Construction of enhancer-driven AAV expression vectors

Enhancer sequences (see Table S17) were identified and analyzed for unique restriction sites to facilitate directional cloning. Enhancer templates were generated via genomic DNA extraction (QIAGEN 69504) or as synthetic gBlocks (IDT). Templates were amplified using Q5 High-Fidelity DNA Polymerase (NEB M0492S) under standard cycling conditions, verified by agarose gel electrophoresis, and purified (QIAquick, QIAGEN 28104).

783 Both the AAV backbone and enhancer inserts were digested with appropriate restriction
784 enzymes (e.g., *MluI*, *SalI-HF*, or *XhoI*). To minimize self-ligation, vectors were dephosphorylated
785 with Antarctic Phosphatase (NEB M0289S). Digested products were gel-purified and ligated
786 using T4 DNA Ligase (NEB M0202S) at a 1:5 vector-to-insert molar ratio. Following
787 transformation into *E. coli* (DH5α) (NEB C2987H), successful clones were screened via colony
788 PCR or restriction analysis and validated by Sanger or whole-plasmid sequencing (Azenta Life
789 Sciences). Validated constructs were produced using endotoxin-free Maxiprep (Thermo Fisher
790 A31231) and submitted for AAV packaging.
791
792
793
794 **Multiplexed CRE library design**
795
796 A multiplexed library was constructed containing 106 prioritized CREs, a mini CMV enhancer
797 control, and a No CRE control (see Table S17). For each element, a 275bp sequence flanking
798 the CRE summit was synthesized as an oPool (IDT) with the following architecture:

799 5' InFusion overhang - CRE - restriction sites (SpeI, FseI) - barcode - 3' InFusion overhang

800 Unique barcodes were assigned using the R package DNAbarcodes, following the logic
801 described in Hrvatin et al. 2019.[41] The oPool was PCR-amplified and cloned into the AAV
802 backbone as previously described.[64] A functional reporter cassette (containing a minimal CMV
803 promoter, mGreenLantern-KASH, and WPRE) was subsequently inserted into the library
804 restriction sites (*SpeI/FseI*) as described[41] with the following architecture:

805 SpeI–minimal CMV promoter--mGreenLantern-KASH--WPRE–FseI

806 The structure of the final AAV library is:

807 AAV ITR-CRE–minimal promoter–mGreenLantern-KASH–WPRE-barcode-polyA-AAV ITR
808
809
810 **Amplification of CRE library and barcoding**
811
812 The CRE library was generated with an approach similar to that described in Hrvatin et al.
813 2019.[41] In brief, the CRE library was generated by PCR amplification of a synthetic oPool (IDT)
814 using Q5 High-Fidelity Master Mix (NEB). Amplicons (approx. 350 bp) were purified and verified
815 via agarose gel electrophoresis. The AAV backbone was linearized through double restriction
816 digestion and gel-purified to remove template DNA.

817 The purified oPool inserts were cloned into the linearized vector using In-Fusion Snap Assembly
818 (Takara Bio 638947). To ensure high transformation efficiency, the assembly products were
819 concentrated and desalted prior to electroporation into MegaX DH10B T1R cells (Invitrogen
820 C6400-03). To maintain library representation, we ensured >100X coverage by plating the
821 library on 245-mm BioAssay dishes (Corning 431111). Transformation efficiency was verified via
822 serial dilution plates, and individual clones were validated through Sanger or whole-plasmid
823 sequencing.

824 *Functional cassette insertion and validation*
825
826 A secondary cloning step was performed to insert the reporter cassette (minimal
827 promoter-Cre-fluorophore-WPRE) into the barcoded library backbone. The cassette was
828 PCR-amplified and inserted via restriction/ligation. To prevent backbone self-ligation, the library
829 was treated with Shrimp Alkaline Phosphatase (rSAP) (NEB M0371S).

830 The final library was expanded via electroporation as described above. Library diversity and
831 barcode distribution were confirmed by Next-Generation Sequencing (Illumina MiSeq, 2x250bp)
832 of a 150–300 bp amplicon covering the barcode region (see Figure S6A) with an approach
833 previously described.[64] Validated pooled libraries were then processed for AAV packaging.
834
835
836 **AAV preparation**

837 The pooled CRE library or individual AAV constructs were packaged into AAV-PHP.S at the
838 Boston Children's Hospital Viral Core or with PackGene Biotech, Inc. The titers (2–1000 × 10^{12}
839 genome copies/mL) were determined by qPCR and adjusted to order requirements. The purity
840 of the viruses was assessed via SDS-PAGE and the viruses were validated to be endotoxin-free
841 via the Limulus Amebocyte Lysate (LAL) test.
842
843
844 **I.C.V. injections**

845 For the intracerebroventricular (ICV) injection procedures in mice, AAV viruses at a dose of 1-4
846 × $10^{10\text{-}11}$ gc/ml along with Trypan Blue at a total injection volume of about 3ul were prepared.
847 Neonatal pups at postnatal day 0-4 (P0-4) were cryo-anesthetized for 1-2 minutes and injected
848 with the viruses to cerebral lateral ventricles using a Hamilton Gastight syringe (Hamilton
849 7653-01) with 32G needle. At least 4 weeks after AAV administration, DRG tissues were
850 collected.
851
852 **RNAscope *in situ* hybridization**

853 The mouse DRG dissections were performed as previously reported[65,66] with minor
854 modifications. In brief, mice were anesthetized with isoflurane for about 5 minutes and
855 decapitated. The skin and muscles were removed from the dorsal aspect of the mice to expose
856 the spinal cord (SC), then the vertebral columns of the SC were exposed and placed on a tray
857 of ice. DRG were exposed after ½ vertebral columns were removed and the surrounding tissues
858 were cut away. Finally, the collected DRGs were transferred to 4% PFA for a 24h incubation at
859 4°C, followed by a 24h incubation in 30% sucrose dissolved in PBS at 4°C. For sample
860 preparation, DRG were stored at -80°C in Tissue-Tek O.C.T. compound (Electron Microscopy
861 Sciences 62550-01) until they were cryosectioned. DRG were sectioned at 14 µm thickness, and
862 mRNAs were detected by RNAscope (Advanced Cell Diagnostics, RNAscope® Fluorescent
863 Multiplex Reagent Kit v1) using the manufacturer's protocol. The following probes were used:
864 *eGFP* (Cat# 400281), *Tac1* (Cat# 410351-C2), *Mrgprd* (417921-C2), *Rbfox3* (Cat# 313311-C3).

865 *EGFP*, *Sstr2, Calca, Tubb3 Images*

866 The mouse DRG dissections were performed as previously reported[67] with minor modifications.
867 In brief, mice were anesthetized with isoflurane for about 5min before cervical dislocation and
868 decapitation. Fur was doused with 70% EtOH to restrict contamination. Pelt, skin, muscles, and
869 viscera surrounding the spinal column were removed, and the spinal column was isolated,
870 bisected, and placed in a dish with HBSS (Gibco 14175-095) on a tray of ice. Lumbar DRG
871 were dissected and placed into cryomolds that were subsequently filled with Tissue-Tek O.C.T.
872 compound (Electron Microscopy Sciences 62550-01) and allowed to solidify on dry ice for 5
873 minutes. DRG were stored at -80°C prior to cryosectioning. Lumbar DRGs were cryosectioned
874 at 16 µm thickness and adhered to SuperFrost slides (Fisherbrand 12-550-15) and frozen at
875 -80°C for storage or immediately processed. mRNAs were detected by RNAscope (Bio-Techne
876 Advanced Cell Diagnostics, RNAscope® Fluorescent Multiplex Reagent Kit v2) using the
877 manufacturer's protocol. The following probes were used: *eGFP* (Cat# 400281-C1), *Sstr2* (Cat#
878 437681-C2), *Calca* (Cat# 557031-C3), and *Tubb3* (Cat# 423391-C4).
879
880
881 **Imaging and Quantification**

882 Following FISH, sections were imaged with a Zeiss LSM 710 confocal microscope using the 20x
883 objective for cell quantification and the 40x oil objective for showing cell markers at higher
884 resolution. Cell quantifications were performed manually by ImageJ (Fiji, 2.3.0) using regions of
885 interest (ROIs) to define the quantified area. Cell counts were confirmed by labeling with nuclei
886 marker DAPI and pan-neuronal marker *Rbfox3*. AAV+ neurons were counted as cells with more
887 than five *GFP*+ punctate signals per ROI. *Tac1*+ or *Mrgprd*+ cells were counted as cells with
888 more than five *Tac1*+ or *Mrgprd*+ punctate signals, respectively, per ROI. The number of
889 neurons that were either *Tac1*+*GFP*+ or *Mrgprd*+*GFP*+ were summed and then divided by the
890 total number of *GFP*+ cells to obtain either %*Tac1*+*GFP*+ / *GFP*+ cells or %*Mrgprd*+*GFP*+ /
891 *GFP*+ cells for each image per replicate. The values for either %*Tac1*+*GFP*+ / *GFP*+ cells or
892 %*Mrgprd*+*GFP*+ / *GFP*+ cells were averaged across 2-5 images from a given animal and
893 reported as a replicate. Replicates were obtained from 3-6 different individual mice.

894 *EGFP*, *Sstr2*, *Calca*, *Tubb3* Images

895 Following FISH, sections were imaged with Andor DragonFly 600 Spinning Disk confocal
896 microscope using the water immersion 25x objective. Cell quantifications were performed
897 manually by ImageJ (Fiji, 2.17.0) using regions of interest (ROIs) to define the quantified area.
898 Cell counts were confirmed by labeling with nuclei marker DAPI and pan-neuronal marker
899 *Tubb3*. AAV+ neurons were counted as cells with more than five *EGFP*+ punctate signals per
900 ROI. *Calca*+ or *Sstr2*+ cells were counted as cells with more than five *Calca*+ or *Sstr2*+ punctate
901 signals, respectively, per ROI. The number of neurons that were either *EGFP*+*Calca*+*Sstr2*+ or
902 *Calca*+*Sstr2*+ were summed and then divided by the total number of *EGFP*+ or *Calca*+ cells
903 respectively, to obtain %*EGFP*+*Calca*+*Sstr2*+ / *EGFP*+ cells or %*Calca*+*Sstr2*+ / *Calca*+ cells
904 for each image per replicate. The values for either %*EGFP*+*Calca*+*Sstr2*+ / *EGFP*+ cells or
905 %*Calca*+*Sstr2*+ / *Calca*+ cells were averaged across 3-6 images from a given animal and
906 reported as a replicate. Replicates were obtained from 3 different individual mice.

907 **DRG cell dissociation from enhancer AAV injected animals**

908 Adult C57BL/6 mice were euthanized according to institutional guidelines. Spinal columns were
909 dissected, bisected, and DRGs were carefully extracted from all axial levels in ice-cold Hank's
910 Balanced Salt Solution (HBSS; ThermoFisher Scientific, 14175103). Peripheral and central
911 nerve roots were removed to isolate sensory ganglia, which were collected in Dulbecco's
912 Modified Eagle Medium (DMEM; ATCC, 30-2002) supplemented with 1% penicillin-streptomycin
913 (P/S; Thermo Fisher, 15140122) on ice.

914
915 Ganglia were enzymatically dissociated using a Collagenase/Type I Trypsin/DNase I mixture
916 (Sigma C9891, T8003; ThermoFisher Scientific 18047019) in DMEM at 37°C. Incubation times
917 were adjusted based on age: 30 min for 4–6-week-old animals and 45 min for animals ≥8 weeks.
918 Digestion was quenched with Neurobasal (NB) medium supplemented with 5% heat-inactivated
919 fetal bovine serum (HI-FBS; Gibco, 16140063), 1% P/S, and 1X B27 supplement (Gibco,
920 17504044). Ganglia were triturated sequentially with fire-polished glass pipettes of decreasing
921 bore size until a single-cell suspension was achieved, followed by filtration through a 70 µm
922 nylon mesh (Pluristrainer, 43-10070-60). Cells were layered onto a 15% bovine serum albumin
923 (BSA; Sigma, A9205) gradient and centrifuged at 130 × g for 10 min to remove debris. The
924 pelleted cells were gently resuspended in NB medium, ensuring a final BSA concentration ≤2%
925 to maintain compatibility with downstream single-cell applications (10x Genomics). Cell viability
926 was assessed by trypan blue exclusion and hemocytometer counting. Preparations with >80%
927 viable cells and minimal debris were used for downstream applications, including single-cell
928 RNA sequencing.
929
930
931 **scRNA-seq of enhancer AAV libraries**

932 Cell suspensions were sequenced using 10x Genomics assays, resuspended, and loaded into
933 the 10x Chromium device for snRNA-seq (10x Genomics v3.1).
934
935 *Targeted cDNA enrichment*
936 Custom enrichment of targeted transcripts was performed using cDNA generated from the 10x
937 Genomics Chromium Next GEM Single Cell 3' RNA v3.1 workflow. For each enrichment
938 reaction, 2 ng of cDNA template (obtained post-cDNA amplification) was amplified using Q5
939 High-Fidelity 2X Master Mix (New England Biolabs M0492S). The reaction was performed in a
940 25µL volume with a final concentration of 500 nM for each of the following primers:

941 - Forward (Partial Read 1): 5'-CTACACGACGCTCTTCCGATCT-3'
942 - Reverse: 5'-TGTGCACTGTGTTTGCTGACGC-3'
943
944 The PCR amplification was conducted with an initial denaturation at 98°C for 30 seconds,
945 followed by 15 cycles of denaturation at 98°C for 10 seconds, annealing at 69°C for 20 seconds,
946 and extension at 72°C for 30 seconds. A final extension was performed at 72°C for 2 minutes.

Post-amplification, the enriched product was purified using the QIAquick PCR Purification Kit (Qiagen) and eluted in a volume of 22µL EB. Product integrity and size distribution were assessed via Agilent TapeStation. The purified, enriched product was integrated back into the standard 10x Genomics workflow starting at Step 3.1 (Gene Expression Library Construction). To account for the specific nature of the enriched fragments, the enzymatic fragmentation time was reduced from 5 minutes to 2 minutes. All subsequent steps, including adapter ligation and index PCR, were completed according to the Chromium Next GEM Single Cell 3' Reagent Kits v3.1 (Dual Index) User Guide.

All other libraries were prepared for scRNA-seq according to the manufacturer's protocol. Libraries were sequenced on an Illumina NextSeq 2000 or shipped to GENEWIZ (Azenta Life Sciences) for sequencing on the Illumina NovaSeq X Plus platform to an approximate read depth of ~20,000 reads per cell. Our scRNA-seq analysis pipeline followed a protocol similar to our previous studies;[6,61] the key steps for the single virus and AAV CRE libraries are detailed below.

Single-cell RNA-seq processing of CRE libraries

All single-cell libraries (single-virus CREs and AAV CRE libraries) were aligned to the mm10 genome (modified to include the viral vector backbone sequence) using 10x Genomics CellRanger v9. Ambient RNA was removed using CellBender v0.3.0 (single-virus libraries: v0.3.0), using default parameters for both CellRanger and CellBender. For the AAV CRE library, viral barcode detection was performed from CellRanger BAM outputs. Viral barcoded reads were extracted using custom Python scripts, and annotated by matching read sequences to the known CRE barcode list (allowing a single base-pair mismatch).

For both the single-virus and AAV CRE libraries, ambient RNA-corrected filtered gene count matrices were processed in Seurat v4. Low quality cells were excluded (detected genes > 500; mitochondrial reads < 20%). Data were then split by sequencing library and integrated across runs using Seurat's standard workflow as described in Bhuiyan et al., 2024. Clustering was performed at resolution = 1.0, and clusters were annotated using established marker genes for neuronal (*Snap25*, *Rbfox3*) and non-neuronal populations (e.g., *Sparc*, *Ptprc*, *Ptgfra*).

For single-virus libraries, a second round of integration was performed. The neuronal population was subsetted and reintegrated using the same pipeline with a stricter QC threshold (detected genes > 1,000). Final neuronal identities were assigned by anchoring to the Bhuiyan et al., 2024[6] mouse DRG reference atlas using Seurat v4 label transfer; predicted labels were used as final annotations.

For the AAV CRE library, transcriptomic processing, filtering, integration, clustering, and marker-based annotation were performed as above, followed by Seurat v4 label transfer to the Bhuiyan et al., 2024[6] DRG mouse neuron-only reference atlas. Predicted labels were used as final annotations, except for clusters enriched for *Sparc* and *Mpz*, which were classified as non-neuronal.

To quantify CRE activity across functional neuronal classes, a cell class specificity ratio was computed for each CRE as: number of CRE barcode–positive cells in a given DRG cell class divided by number of barcode–positive cells across all other cell classes. DRG cell classes were defined by aggregating transcriptomic clusters into: C-PEP (Calca+Sstr2, Calca+Dcn, Calca+Oprk1, Calca+Adra2a), C-NP.Mrgpra3 (Mprgra3+Trpv1, Mrgpra3+Mrgprb4), C-NP.Mrgprd, C-NP.SST, A-LTMR (Ntrk3^{high}+Ntrk2, Ntrk3^{high}+S100a16, Ntrk3^{low}+Ntrk2), A-PEP (Calca+Smr2, Calca+Bmpr1b), A-Propr, C-LTMR.Th, C-Thermo (Trpm8 and Rxfp1), and non-neurons.

To define the “on-target” cell type for each CRE, candidate on-target sets (e.g., C-NP.Sst, C-Thermo.Trpm8) were assigned based on accessible peak evidence from UCSC Genome Browser tracks (downsampled to 270 cells per cell type). A DA score was then calculated by comparing chromatin accessibility in the proposed on-target versus off-target cell types. DA was computed using a logistic regression test (test.use = "LR") with peak counts as a latent variable (latent.vars = "nCount_MACS"). On-target assignments were designated “highly likely” (Table S16) when the DA value (log2FC on-target vs off-target) was positive.

PAIN-net development

To decode the regulatory logic determining cell-type specificity, we developed PAIN-net, a deep convolutional neural network (CNN) framework that predicts chromatin accessibility and differential expression directly from DNA sequence. We chose this architecture because CNNs can learn non-linear relationships, both between input sequences and between input sequences and the predicted variable(s), and CNNs do not require explicit featurization. Consequently, previous studies have shown that CNNs sufficiently infer CRE function from sequences.[43,69]

Model architecture

PAIN-net utilizes a 1D Residual Network (ResNet) architecture to process 800-bp input sequences. The ResNet architecture was chosen for its ability to learn complex, high-order motif combinations while mitigating gradient degradation in deep models.[70,71] The network is composed of an initial convolutional layer followed by five stacked residual blocks. To integrate sequence information across broader genomic contexts, we employed dilated convolutions within these blocks with exponentially increasing dilation rates. This design expands the receptive field to capture long-range interactions while preserving single-nucleotide resolution, as demonstrated in previous genomic models like Basenji.[71] The feature maps are aggregated via global average pooling and fed into a shared dense layer. This layer bifurcates into two task-specific heads: one predicting differential expression (log2-fold change) and the other predicting chromatin accessibility (ATAC-seq signal from bigwig files) across cell types.

Data preprocessing and encoding

To identify subtype-specific regulatory elements, we performed differential accessibility analysis on single-cell ATAC-seq data using the FindAllMarkers function in Signac.[72] We compared each neuronal subtype against all other subtypes to identify differentially accessible sequences. The

analysis was conducted using default parameters with specific adjustments to account for technical covariates and capture a broad range of fold-changes: the logistic regression test (test.use = "LR") was employed with peak counts as a latent variable (latent.vars = "nCount_MACS") and the log2-fold change threshold set to zero (log2fc.threshold = 0). The resulting log2-fold change (log_2FC) values served as the primary regression targets for the differential expression head of the model. Additionally, to capture absolute chromatin accessibility levels, each sequence was annotated with raw fragment number per base.

Genomic sequences were centered on the identified peaks and fixed to a length of 800 base pairs (bp). DNA sequences were one-hot encoded into an 800 x 4 binary matrix. To improve model invariance to strand orientation and increase the effective training set size, data augmentation was performed by including the reverse complement of all training sequences. Target labels (log_2FC and ATAC signal) were standardized to zero mean and unit variance per cell type to ensure gradient stability during training.

Model training and optimization

The model was trained using a multi-objective loss function. We minimized a hybrid loss combining Mean Squared Error (MSE) and Pearson correlation (r) for the expression head (L = MSE - r). This ensured that the model prioritized both the magnitude of fold-changes and the relative expression ranking across cell types. A standard MSE loss was applied to the chromatin accessibility head as an auxiliary task to regularize the shared representation.

To mitigate the sparsity of rare neuronal subtypes, we implemented a dynamic weighting strategy. Loss contributions were balanced by the inverse frequency of available data per cell type. Furthermore, we applied targeted boosting weights (2x) to biologically critical but underrepresented clusters (i.e peptidergic subtypes) while dampening dominant signals to prevent overfitting to majority classes (0.5x; i.e Atf3). Training was performed using the Adam optimizer with a learning rate of 10^{-3}. We utilized early stopping based on the mean Pearson correlation of the validation set to prevent overfitting.

In silico mutagenesis

We performed *in silico* saturation mutagenesis (ISM) on CRE1-115 to quantify position- and nucleotide-specific contributions to predicted enhancer activity across all modeled cell types. For each CRE, we extracted an 800-bp sequence window centered on the peak summit by taking 400 bp of flanking sequence upstream and 400 bp downstream of the peak center, and one-hot encoded each sequence into an 800×4 matrix (A/C/G/T). The PAIN-net model was used to generate predictions, and target-specific output scalers (loaded from a pickled dictionary) were applied to inverse-transform model outputs back to the original scale (reported as log_2FC). For each sequence, a baseline prediction was computed for the wild-type input, followed by generation of all single-nucleotide variants at every position (3 substitutions per valid base; 800 positions; 18 cell types). For each cell type, an ISM matrix of shape 4×800 was constructed where each entry represents the change in predicted activity (Δlog2FC) for substituting a given base at a given position relative to the wild-type prediction. Per-CRE outputs were saved into a hierarchical directory structure (CRE → target), including (i) fixed-scale heatmaps of the ISM

matrix (−1 to 1), (ii) per-mutation scatterplots of Δlog2FC versus position colored by mutant base (−1 to 1), and (iii) attribution-style sequence logos derived by assigning the wild-type base at each position the negative mean effect of its alternative substitutions (−1 to 1). In addition, for each target we computed an "importance profile" defined as the maximum absolute Δlog2FC across the four bases at each position, and aggregated these profiles across all CREs to produce global summary heatmaps (0 to 1) showing position-wise importance per sequence for each target.

Electrophysiological data acquisition and analysis

Primary sensory neuron culture

35-mm culture dishes were sequentially coated with Poly-D-Lysine (1–2 h, 37°C) and laminin (2.5 µg/mL), then air-dried prior to plating. Dorsal root ganglia (DRGs; ~35 per mouse) were harvested from wild-type mice (≥postnatal day 10) into ice-cold DMEM/F12. Tissue was enzymatically digested in HBSS containing Collagenase A (2 mg/mL) and Dispase (2 mg/mL) for 60 min at 37°C, followed by mechanical trituration to a single-cell suspension. To eliminate myelin and debris, the suspension was centrifuged (200 rcf, 12 min) through a 30% BSA density gradient. The resulting neuronal pellet was resuspended in complete Neurobasal medium (B27, Glutamax, Pen/Strep) and seeded at ~100,000 cells/dish. For experiments requiring $K_{ir}2.1$ expression, neurons were transduced *in vitro* immediately following dissociation. Control neurons were not transduced. AAV vectors (CAG-$K_{ir}2.1$-mCherry, or CRE1-$K_{ir}2.1$-mCherry) were applied at a titer of 1 x 10^{13} vg/mL. Specifically, 1–2 microliters of viral stock (1–2 x 10^{10} vg) was added to 5 x 10^5 cells in a 200 µL volume. Cultures were incubated overnight, and electrophysiological recordings were performed 24–48 hours post-dissociation to ensure robust ion channel expression.

Recording configurations and quality control

Whole-cell patch-clamp recordings were performed 24–48 h post-dissociation using an EPC 10 amplifier (HEKA) at room temperature. Small-diameter nociceptive neurons (<25µm, capacitance<15pF) were targeted. Recording electrodes (5–6 MΩ) were filled with a K-gluconate-based internal solution (see below). Following giga-ohm seal formation, series resistance (10–15 MΩ) was compensated by ≥80%. Data were sampled at 20 kHz and low-pass filtered at 5 kHz.

Electrophysiological solutions

The standard extracellular bath solution contained (in mM): 155 NaCl, 3.5 KCl, 1.5 $CaCl_2$, 1 $MgCl_2$, 10 HEPES, and 10 glucose (pH 7.4 adjusted with NaOH). The intracellular pipette solution consisted of (in mM): 140 K-gluconate, 13.5 NaCl, 1.8 $MgCl_2$, 0.09 EGTA, 9 HEPES, 14 creatine phosphate (Tris salt), 4 MgATP, and 0.3 Tris-GTP (pH 7.2 adjusted with KOH).[73]

Stimulation protocols and data analysis

The resting membrane potential (RMP) was measured in current-clamp mode immediately after achieving the whole-cell configuration without current injection. Cell capacitance was

determined by the amplifier's automated compensation circuit. In voltage-clamp mode, total ionic currents were elicited from a holding potential of −60 mV using 500-ms voltage steps ranging from −120 to -40 mV in 10 mV increments, with a 1s interpulse interval. Neuronal excitability was assessed in current-clamp mode using 2-s depolarizing steps (10–50 pA increments). Rheobase was defined as the minimum current required to elicit an action potential.

While CAG-Kir2.1-mCherry neurons were identified by robust fluorescence, CRE1-driven expression was often indistinguishable from autofluorescence. To objectively identify successfully transduced cells in the CRE1 group, we utilized a functional inclusion criterion based on neuronal excitability. Based on an initial characterization of control neurons, which exhibited a consistent rheobase (Mean: 55.5 pA), we established a threshold for "High Rheobase" cells. Cells in the experimental groups (CRE1-K_{ir}2.1-mCherry) were categorized as CRE1-K_{ir}2.1-mCherry cells in Figure 6 if they exhibited a rheobase >200 pA (Figure S8A). This stringent threshold (representing a >3.5-fold increase over the control mean) served as a physiological proxy for successful K_{ir}2.1 expression, as the increased membrane leak conductance provided by K_{ir}2.1 is known to shunt depolarizing currents and increase the threshold for action potential firing.[74]

Current densities (pA/pF) were calculated by normalizing peak or steady-state currents to cell capacitance. Data were analyzed using custom MATLAB scripts and OriginPro, with statistical comparisons (unpaired t-tests) performed in GraphPad Prism 9. Results are presented as mean ± SEM.

Human iPS-derived nociceptors and cardiomyocytes cell culture assay

The human iPS cell-derived nociceptors and cardiomyocytes cell culture were performed as described here: iPS cell-derived nociceptors were plated at 5,000/well density and co-cultured with 2,000 glial cells in 384-well plate, iPS cell-derived cardiomyocytes were plated at 10,000/well without co-culture. Human iPS cells were infected by directly adding concentrated AAV viral particles with multiple titers: $1x10^9$, $1x10^{10}$, $1x10^{11}$, $1x10^{12}$ to a well plate, two well plate replicates per group. Cells were re-fed every two days with maintenance medium (Nociceptors: DMEM/F-12 Glutamax (ThermoFisher Scientific 10565018), N-2 Supplement (ThermoFisher Scientific A1370701), B-27 Supplement (ThermoFisher Scientific A3353501), BDNF (Bio-Techne R&D Systems 248-BDB-050/CF), GDNF (Bio-Techne R&D Systems 212-GD-050/CF), ß-NGF (Bio-Techne R&D Systems 256-GF-100/CF), NT-3 (Bio-Techne R&D Systems 267-N3-025/CF); Cardiomyocytes: RPMI 1640 medium (ThermoFisher Scientific 11835055), 1x B27 supplement with insulin (ThermoFisher Scientific 17504-044)). After culturing for two weeks, cells were fixed with 4% PFA for 0.5h and rinsed with PBS three times. Following that, the images were acquired with a Zeiss LSM 710 confocal microscope using the 10x objective and 40x objective for zoom-in images.

DATA AND CODE AVAILABILITY

All custom code used to integrate data is available on Git (https://github.com/Renthal-Lab/painseq-multiome). Raw sequencing data and final objects will be deposited into GEO.

Figure Legends

Figure 1: snMultiome of mouse and human DRG

A: Workflow displays schematic of experimental design.

B: (*Top, Left*) UMAP projections displaying 23,928 mouse DRG nuclei colored by cell type or (*Top, Right*) log-normalized *Snap25* expression. (*Bottom, Left*) Dot plot displaying log-normalized gene expression (red) or log-normalized ATAC gene activity scores (blue) for DRG cell type marker genes. ATAC gene activity scores are defined as the sum of ATAC fragments falling within the gene body and the 2 kb region upstream of the transcription start site (TSS). Dot size represents the percentage of nuclei expressing the marker gene (RNA) or the percentage of nuclei with accessible chromatin at the marker gene locus (ATAC). (*Bottom, Right*) Genomic tracks displaying normalized chromatin accessibility at the *Snap25* locus for each mouse DRG cell type.

C: (*Top, Left*) UMAP projections displaying 155,008 human nuclei colored by log-normalized *SNAP25* expression or (*Top, Right*) cell type. (*Bottom, Left)* Genomic tracks displaying normalized chromatin accessibility at the *SNAP25* locus for each human DRG cell type. (*Bottom, Right*) Dot plot displaying log-normalized gene expression (red) or log-normalized ATAC gene activity scores (blue) for canonical marker genes. ATAC gene activity scores are defined as the sum of ATAC fragments falling within the gene body and the 2 kb region upstream of the TSS. Dot size represents the percentage of nuclei expressing the marker gene (RNA) or the percentage of nuclei with accessible chromatin at the marker gene locus (ATAC).

D: Heatmap displays the average (Z-scaled Transverse Frequency-Inverse Document Frequency [TF-IDF] normalized mean) number of fragments for all nuclei across the top 2,000 differentially accessible peaks (columns) for each mouse DRG cell type ($\log_2$FC > 0.5, adj. p-val < 0.05 relative to all other cell types). Rows represent DRG cell types.

E: Bar chart displays the mean proportion of mouse differentially accessible (DA) peaks ($\log_2$FC > 0.5, adj. p-val < 0.05 relative to all other mouse cell types) that are syntenic with human DRG differentially accessible peaks ($\log_2$FC > 0.5, adj. p-val < 0.05, relative to all other human cell types) in the corresponding DRG cell type. Values are estimated by bootstrapping (100 iterations), sampling 50 peaks per iteration from the top 500 mouse DA peaks for each cell type. Error bars indicate standard error across iterations. Gray bars indicate control test done using differentially accessible peaks with randomly shuffled cell type labels (*** = adj.p-value < 0.0001, Z-test comparing distribution of overlap of differentially accessible peaks with peaks randomly shuffled across cell types).

F: Heatmap displays the average (Z-scaled TF-IDF-normalized mean) number of fragments for all nuclei across the top 2,000 differentially accessible peaks (columns) for each human DRG cell type ($\log_2$FC > 0.5, adj. p-val < 0.05 relative to all other cell types). Rows represent DRG cell types.

G: (*Top, Left*) Dotplot displays the eRegulon activity scores (dot color) of transcription factors in mouse DRG cell types. Dot size represents the percentage of nuclei with non-zero eRegulon

scores. (*Top, Right*) Diagram displays schematic of what is factored in the calculation of an eRegulon activity score - transcription factor (TF) expression, peak region accessibility and target gene expression. (*Bottom, Left*) Genomic tracks display normalized chromatin accessibility at the *Elavl2* locus for each mouse DRG cell type. Highlighted region indicates a candidate *cis*-regulatory element (CRE) with a predicted ISL1 binding motif. Loops show Pearson correlation (p-val. < 0.05) of CRE accessibility and RNA expression of *Elavl2.* (*Bottom, Right*) UMAP projection displays ISL1 eRegulon activity scores (color scale) in mouse DRG cell types (downsampled to 2,414 nuclei for visualization purposes).

H: (*Top*) Dotplot displays the eRegulon activity scores (dot color) of transcription factors expressed in human DRG cell types. Dot size represents the percentage of nuclei with non-zero eRegulon scores. (*Bottom, Left*) UMAP projection displays ISL1 eRegulon activity scores (color scale) across 95,515 human nuclei and includes all cell types found in Fig. 1C. (*Bottom, Right*) Genomic tracks display normalized chromatin accessibility at the *ELAVL2* locus for each human DRG cell type. The highlighted region indicates a candidate CRE with a predicted ISL1 binding motif. Loops show Pearson correlation (p-val. < 0.05) of CRE accessibility and RNA expression of *ELAVL2.*

Figure 2: snMultiome of mouse and human DRG neuronal subtypes

A: UMAP projection displays snMultiome data from 21,064 mouse (*Left*) and 6,893 human (*Right*) DRG neurons. Nuclei are projected into the mouse[6] or human[5] reference atlas space. (*Middle*) Nuclei are colored by neuronal subtype, according to the key between the UMAPs. Transcriptomically corresponding neuronal subtypes between mouse and human are connected by lines.

B: Dot plot displaying log-normalized gene expression (red) or log-normalized ATAC gene activity scores (blue) for marker genes of mouse DRG neuronal subtypes. ATAC gene activity scores are defined as the sum of ATAC fragments falling within the gene body and the 2 kb region upstream of the TSS. Dot size represents the percentage of nuclei expressing the marker gene (RNA) or the percentage of nuclei with accessible chromatin at the marker gene locus (ATAC).

C: Heatmap displays the average (Z-scaled TF-IDF-normalized mean) number of fragments for all nuclei across the top 2,000 differentially accessible peaks (rows) for each mouse neuronal subtype ($\log_2$FC > 0.5, adj. p-val < 0.05 relative to all other neuronal subtypes). Columns represent DRG neuronal subtypes.

D: Dot plot displaying log-normalized gene expression (red) or log-normalized ATAC gene activity scores (blue) for canonical marker genes for human neuronal subtypes. ATAC gene activity scores are defined as the sum of ATAC fragments falling within the gene body and the 2 kb region upstream of the TSS. Dot size represents the percentage of nuclei expressing the marker gene (RNA) or the percentage of nuclei with accessible chromatin at the marker gene locus (ATAC).

1243 **E:** Heatmap displays the average (Z-scaled TF-IDF-normalized mean) number of fragments for
1244 all nuclei across the top 150 differentially accessible peaks (rows) for each human neuronal
1245 subtype ($log_2FC > 0.5$, adj. p-val < 0.1 relative to all other neuronal subtypes). Columns
1246 represent DRG neuronal subtypes.

1247 **F-G:** Heatmap displays SCENIC+ eRegulon activity scores (z-scaled) of TFs (columns) in
1248 mouse (F) or human (G), averaged across DRG neuron subtypes (rows).

1249 **H:** UMAP projections display log-normalized *Bcl11a* gene expression (first UMAP from the left)
1250 and BCL11A eRegulon activity scores (second UMAP) in mouse DRG neuronal nuclei. UMAP
1251 projections displaying log-normalized *BCL11A* expression (third UMAP) and BCL11A SCENIC+
1252 eRegulon activity scores (fourth UMAP) for human DRG neuronal nuclei are also shown.

1253 **I:** Sankey plot displays differentially accessible peaks in mouse for each neuronal subtype
1254 ($log_2FC > 0.5$, adj. p-value < 0.05 relative to all other neuronal subtypes) that are also predicted
1255 to comprise SCENIC+ eRegulons or which have syntenic peaks in the human DRG neurons.
1256 Bars and ribbons are colored by mouse neuronal subtypes. Green vertical bars indicate overlap,
1257 red vertical bars indicate no overlap.
1258
1259
1260 **Figure 3: Screening candidate CREs for nociceptor specificity**

1261 **A:** Schematic of the experimental strategy used to screen AAV expression patterns in DRG
1262 driven by candidate CREs.

1263 **B:** Genomic tracks display chromatin accessibility at the loci of candidate CREs in distinct
1264 mouse DRG neuronal subtypes.

1265 **C:** (*Top, Left*) Viral cargo schematic (mP: minimal promoter, teal box: reporter construct, blue
1266 box: P2A) and representative 20X images of mouse DRG labeled by RNAscope *in situ*
1267 hybridization for *Tac1* (magenta) and *EGFP* (green) mRNA (first 4 rows), or for *Mrgprd* (blue)
1268 and *EGFP* (green) mRNA (bottom row). White arrowheads indicate either *Tac1* and *EGFP* or
1269 *Mrgprd* and *EGFP* signal colocalization. (*Bottom, Left*) Quantification of *Tac1* and *EGFP* mRNA
1270 colocalization for each candidate CRE. (One-way ANOVA, $F(3,27) = 6.80$, $p < 0.001$; Dunnett's
1271 multiple comparisons test; **p = 0.001, CRE1 vs. CAG, * p < 0.05, CRE2 vs. CAG). (*Bottom,*
1272 *Right*) Quantification of *Mrgprd* and *EGFP* mRNA colocalization for CRE8 (One-way ANOVA,
1273 $F(2,6) = 7.42$, $p = 0.024$; Dunnett's multiple comparisons test; *p < 0.05 CRE8 vs. CAG). Data
1274 are presented as mean ± SEM. Each point represents a biological replicate. n = 2-5 animals per
1275 CRE). Scale bar, 25 µm.

1276 **D:** UMAP projection displays 34,899 mouse DRG neurons, 1,032 of which expressed AAV
1277 transcripts (>1 AAV transcript/cell; colored by CREs). Cells without detected AAV transcripts are
1278 gray. Projections and cell type annotations use the mouse reference UMAP space.[6]

1279 **E:** Quantification of C-PEP (*Left*) or C-NP (*Right*) target ratios (AAV transcripts in C-PEP or
1280 C-NP subtype / AAV transcripts in all other subtypes) for candidate CRE1, CRE6, and CRE8 as
1281 measured by scRNA-seq. Cell class specificity target ratios are normalized to the CAG control
1282 ratio. Error bars indicate the standard deviation of randomly sampled cells in each group (n = 5

iterations). Significance was determined by Fisher's exact test (ns: not significant, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$). C-PEP subtypes: C-PEP.Calca+Sstr2, C-PEP.Calca+Dcn, C-PEP.Calca+Oprk1; C-NP subtypes: C-NP.Mrgprd, C-NP.Mrgpra3+Trpv1, C-NP.Mrgpra3+Mrgprb4.

Figure 4: Multiplexed *in vivo* AAV screen to functionally assess CRE activity in DRG

A: Schematic of the experimental strategy used for multiplexed CRE screen. mGL: mGreenLantern; bc: Barcode.

B: UMAP projection displays scRNA-seq of 10,442 DRG cells from mice infected with a multiplexed AAV CRE library. Cells are colored if AAV barcodes were detected (>1 barcode/cell; 2,275 cells). Projections are in the mouse reference UMAP space.[6]

C: (*Bottom*) Genomic tracks display chromatin accessibility profiles in mouse DRG at loci for candidate CREs included in specificity analysis across 5 DRG cell classes. (*Middle*) Dotplot displays cell class specificity ratios across 5 DRG cell classes. The cell class specificity ratio (y-axis) is defined as the number of CRE barcode-positive cells in a specific cell class divided by the number of barcode-positive cells in all other cells. Dot color represents cell class specificity ratio (Z-scaled), and dot size corresponds to the number of a given CRE's barcodes that were detected per cell type. Ratios are normalized to the negative control (NC: No CRE). (*Top*) The bar plot displays the on/off target expression ratio for each CRE, where on-target cell types are defined by the chromatin accessibility at each CRE locus. The on/off target expression ratio is calculated as the number of CRE barcodes detected in on-target cell types cell types divided by the number of CRE barcodes cells in off-target (inaccessible) cell types, normalized to the negative control (No CRE). Only CREs with viral barcodes detected in > 6 cells were analyzed.

D: Scatter plot displays $\log_2$FC differential accessibility of on-target cell types compared to off target cells (x-axis) versus the on/off target expression ratio for CRE barcode expression (y-axis). On-target cell type definition and ratio are the same as C. Dashed line indicates the correlation between the on/ff target expression ratio and on-target accessibility of all 41 CREs (Spearman's $r = 0.26$); solid line indicates the same correlation after removing CREs with viral barcodes detected only in off-target cell types (Spearman's $r = 0.47$).

Figure 5: Machine learning prioritization of candidate CREs

A: (*Left*) Workflow schematic displays PAIN-net model training and validation. PAIN-net was trained on sequences of differentially accessible peaks (in each neuronal subtype compared to all others) to predict cell type-specific accessibility ($\log_2$FC in each neuronal subtype compared to all others) and peak signal. Training was performed on 85% of the data, while testing was performed on the held-out 15%. (*Right*) Heatmap displays the Pearson correlation between predicted and measured differential accessibility for each neuronal subtype ($\log_2$FC compared to all other neuronal subtypes). Results are shown for the held-out 15% data (median r = 0.72) and shuffled sequences (median r = -0.07). Color bars indicate broader neuronal classes and the

order of subtypes within those classes is from the top to bottom: Red: Atf3, Blue: C-PEP (C-PEP.Calca.Adra2a, C-PEP.Calca.Dcn, C-PEP.Calca.Oprk1, C-PEP.Calca.Sstr2), Green: A-PEP (A-PEP.Calca.Smr2, A-PEP.Calca.Bmpr1b), Purple: C-NP (C-NP.Mrgpra3.Mrgprb4, C-NP.Mrgpra3.Trpv1, C-NP.Mrgprd, C-NP.Sst), Orange: A-LTMR (A-LTMR.Ntrk3^{high}.Ntrk2, A-LTMR.Ntrk3^{high}.S100a16, A-LTMR.Ntrk3^{low}.Ntrk2), Yellow: A-Propr.Pvalb, Brown: C-LTMR.Th, Pink: C-Thermo (C-Thermo.Rxfp1, C-Thermo.Trpm8).

B: PAIN-net recapitulates cell type specificity of CRE1, 6, and 8. Predicted log_2FC was calculated for CRE1, CRE6, CRE8 and CAG control using PAIN-net for C-PEP.Calca+Sstr2 (*Left*) and C-NP.Mrgprd (*Right*) populations. All scores are normalized to the predicted log2FC of CAG. Predicted log2FC is defined as the PAIN-net predicted accessibility for a given neuronal subtype compared to all other neuronal subtypes. Orange bar (on-target): Predicted log2FC of C-PEP.Calca+Sstr2 (*left;* target subtype for CRE1 and CRE6) or C-NP.Mrgprd (*right;* target subtype for CRE8). Blue bar (off-target): Sum of predicted log2FC for all cell types aside from C-PEP.Calca+Sstr2 (*left*) or C-NP.Mrgprd (*right*).

C: PAIN-net predictions correlate with CRE driven AAV barcode expression. Scatter plot shows PAIN-net predicted on-target differential accessibility versus the on/off target expression ratio of CRE barcode expression driven in the multiplexed AAV scRNA-seq screen (Spearman's $r = 0.33$, $p < 0.05$). On-target cell types are defined as having accessible chromatin at the CRE locus (Table S15; see Methods). On-target differential accessibility is defined as the accessibility of on-target cell types compared to off target cell types.

D: Sequence logos display motifs significantly enriched (q-value < 0.05) in the top PAIN-net predicted sequences for each neuronal subtype, relative to GC-matched backgrounds. Top sequences are defined as the 1,000 regions with the highest predicted differential accessibility (Log_2FC) compared to all other cell types. Each motif is labeled with the best-matching transcription factor predicted to bind. Sequence logo size is based on nucleotide frequency.

E: Plot displays the cumulative impact of *in silico* mutations across an 800bp sequence window on PAIN-net predicted accessibility, averaged over 43 CREs. Predicted ΔLog_2FC (y-axis) is calculated as the sum of the absolute Δlog_2FC values across on target cell types (red line) or off-target cell types (blue line). The gray line serves as a negative control derived from random DNA sequences.

F: *In silico* saturation mutagenesis of CRE1 identifies key bases driving differential accessibility in C-PEP.Calca+Sstr2. Vertical grey bands mark putative transcription factor binding sites that coincide with high-impact regions. (*Top*) Attribution sequence logos illustrate the contribution of each nucleotide to C-PEP.*Calca+Sstr2* differential accessibility. Nucleotide height represents the predicted negative average impact of mutation on accessibility. (*Bottom*) Saturation mutagenesis scatter plot depicting the predicted impact (Δlog_2FC compared to original sequence) of every possible single-nucleotide substitution. Points are colored by the substituted base.

Figure 6: Functional and translational applications of nociceptor-specific AAVs

1365 **A:** (*Top*) Workflow for DRG dissociation, viral transduction, and whole-cell patch-clamp
1366 recording. (*Top, Middle*) Representative image of CAG infected cultured DRG neurons (red).
1367 (*Bottom*) Construct maps for the experimental vectors expressing K_{ir}2.1 under the ubiquitous
1368 CAG promoter or CRE1 and mP (minimal promoter).

1369 **B:** Representative voltage responses to current injection steps (step protocol shown above)
1370 from Control (*Left*), CAG-K_{ir}2.1-mCherry (*Middle*), and CRE1-K_{ir}2.1-mCherry (*Right*) cells.
1371 Control neurons exhibit repetitive firing. In contrast, CAG and CRE1 neurons exhibit a failure to
1372 fire at low current steps, with red traces indicating the increased rheobase.

1373 **C:** Summary data showing the minimum current required to elicit an action potential (rheobase).
1374 Neurons expressing K_{ir}2.1 via CAG and CRE1 showed significantly higher rheobase compared
1375 to controls (Control, n = 20; CAG, n = 9; CRE1, n = 17). Data are presented as mean ± SEM.
1376 One-way ANOVA with multiple comparisons (* $p < 0.05$, ** $p < 0.01$, **** $p < 0.0001$) ($F(2, 43) =$
1377 76.72, $p < 0.0001$). Post hoc analysis using Dunnett's multiple comparisons test revealed that
1378 rheobase was significantly increased in both the CAG group ($p < 0.0001$) and the CRE1 group
1379 ($p < 0.0001$) compared to control. Sample sizes (n) represent individual neurons recorded
1380 across at least 2 independent preparations.

1381 **D:** Quantification of resting membrane potential (RMP) showing a significant hyperpolarization
1382 of neurons expressing K_{ir}2.1 both via CAG and CRE1 compared to control cells (Control, n = 20;
1383 CAG, n = 8; CRE1, n = 17). Data are presented as mean ± SEM. One-way ANOVA with multiple
1384 comparisons (* $p < 0.05$, ** $p < 0.01$, **** $p < 0.0001$) ($F(2, 42) = 40.08$, $p < 0.0001$). Post hoc
1385 analysis using Dunnett's multiple comparisons test showed that RMP was significantly
1386 hyperpolarized in the CAG group compared to control ($p < 0.0001$), as well as in the CRE1
1387 group ($p = 0.0014$). Sample sizes (n) represent individual neurons recorded across at least 2
1388 independent preparations.

1389 **E:** Current-voltage (I-V) relationships for cells in Control, CAG-K_{ir}2.1-mCherry, and
1390 CRE1-K_{ir}2.1-mCherry groups. Steady-state current density (pA/pF) plotted against membrane
1391 potential (mV). Neurons expressing K_{ir}2.1 via CAG and CRE1 exhibit characteristic
1392 inward-rectifying currents at hyperpolarized potentials compared to Control cells, consistent with
1393 expression of K_{ir}2.1 (Control, n = 6; CAG, n = 6; CRE1, n = 11).

1394 **F:** Representative immunofluorescence images of iPSC-derived cardiomyocytes and
1395 nociceptors. Cells were transduced with AAV vectors expressing EGFP driven by the CAG
1396 promoter, CRE1, or CRE8. Images are representative of n = 3-5 independent experiments. Blue
1397 = DAPI; Green = EGFP. Scale bars = 100 µm.
1398

1399
1400 **Figure S1: Analysis pipeline and RNA-level quality control metrics for mouse and human**
1401 **DRG cell types**

1402 **A:** Barplot displays the fraction of neurons and non-neurons obtained from the multi-omic
1403 sequencing of *Sun1-GFP+/-;vGlut2-Cre+/-* sorted mouse DRG nuclei. Error bars represent
1404 standard error across libraries.

B: (*Left*) Barplot displays fraction of neurons and non-neurons obtained from multi-omic sequencing of human DRG nuclei. Error bars represent the standard error across samples. (*Right*) Barplot displays the fraction of neuronal enrichment in each sample that was sorted either for a high 7AAD signal or both high 7AAD and NeuN signals. Lines connect measurements from the same human DRG samples (dots) using each sorting strategy.

C: Flowchart displays the bioinformatic pipeline detailing how snMultiome data was processed for mouse DRG. See Methods for details.

Abbreviations: ATAC (assay for transposase-accessible chromatin); QC (quality control); nFeature (number of detected genes per cell); percent.mt (percentage of mitochondrial reads); FrIP (fraction of reads in peaks); nFrags (number of ATAC fragments per cell); TSS (transcription start site); UMAP (uniform manifold approximation and projection); eRegulon (enhancer-driven regulatory module)

D: Flowchart displays the bioinformatic pipeline detailing how snMultiome data was processed for human DRG. See Methods for details.

Figure S2: snMultiome level (RNA+ATAC) quality control metrics for mouse and human DRG cell types

A: Plots display RNA-level snMultiome metrics for mouse DRG nuclei that passed quality controls and were included in downstream analyses. (*Top*) Stacked bar plots display the number of neuronal (blue) and non-neuronal (gray) nuclei that passed quality controls by sample. (*Middle*) Box plots display the number of unique genes detected per nucleus in each sample. (*Bottom*) Box plot displays the percentage of mitochondrial reads detected per nucleus in each sample. The median value is indicated by the horizontal black line inside the box. Boundaries of the boxes mark quartiles, and whiskers represent 1.5 times the interquartile range (Q1–Q3). Krauter and Techameena refer to previous publications, Krauter et al., 2025[20] and Techameena et al., 2024,[19] respectively.

B: Plots display RNA-level snMultiome metrics for human DRG nuclei that passed quality controls and were included in downstream analyses. (*Top*) Stacked bar plots display the number of neuronal (blue) and non-neuronal (gray) nuclei that passed quality controls by sample. (*Middle*) Box plots display the number of unique genes detected per nucleus in each sample. (*Bottom*) Box plot displays the percentage of mitochondrial reads detected per nucleus in each sample. The median value is indicated by a black line inside the box. Boundaries of the boxes mark quartiles, and whiskers represent 1.5 times the interquartile range (Q1–Q3).

C: Plots display ATAC-level snMultiome metrics for mouse DRG nuclei that passed quality controls and were included in downstream analyses. (*Top*) Bar plot displays the number of nuclei that passed quality controls, colored by sample. (*Middle*) Box plots display TSS enrichment scores per nuclei for each sample. (*Bottom*) Box plots display unique fragments detected per nuclei in each sample. The median value is indicated by a black line inside the box. Boundaries of the boxes mark quartiles, and whiskers represent 1.5 times the interquartile range (Q1–Q3). (*Bottom, Left*) Histogram displays the fragment size distribution of ATAC-seq

fragments for each sample, with the y-axis representing the percentage of fragments per sample. Lines are colored by sample. (*Bottom, Right*) Tn5 insertion profile displays normalized insertion signal around transcription start sites for each sample (x-axis shows distance from TSS center in bp; y-axis shows normalized insertion profile).

D: Plots display ATAC-level snMultiome metrics for human DRG nuclei that passed quality controls and were included in downstream analyses. (*Top*) Bar plot displays the number of nuclei that passed quality controls, colored by sample. (*Middle*) Box plots display the distribution of TSS enrichment scores per cell for each sample. (*Bottom*) Box plots display the distribution of unique fragments detected per cell for each sample. The median value is indicated by a black line inside the box. Boundaries of the boxes mark quartiles, and whiskers represent 1.5 times the interquartile range (Q1–Q3). (*Bottom, Left*) Histogram displays the fragment size distribution of ATAC-seq fragments for each sample, with the y-axis representing the percentage of fragments per sample. Lines are colored by sample. (*Bottom, Right*) Tn5 insertion profile displays the normalized insertion signal around transcription start sites for each sample (x-axis shows distance from TSS center in bp; y-axis shows normalized insertion profile).

E: Heatmap displays z-scaled, log-normalized expression of marker genes (columns) for each cell type (rows) in mouse DRG. Marker genes are defined as differentially expressed in each cell type compared to all other ($Log_2FC > 0.5$, adj. p-val < 0.05).

F: Heatmap displays z-scaled, log-normalized expression of marker genes (columns) for each cell type (rows) in human DRG. Marker genes are defined as differentially expressed in each cell type compared to all others ($Log_2FC > 0.5$, adj. p-val < 0.05).

G: Stacked barplot displays the distribution of DRG cell types annotated in each donor.

Figure S3: Donor-level integration and sex-specific differences in human DRG cell types

A: (*Left*) UMAP projection clustered by gene expression (same as in Figure 1C) displays 145,562 human DRG nuclei that passed RNA quality controls. Nuclei are colored by donor ID. DRG cell types are outlined (dashed outlines). (*Right*) UMAP projection clustered by chromatin accessibility displays 67,025 human DRG nuclei that passed ATAC quality controls. Nuclei are colored by donor ID. DRG cell types are outlined (dashed outlines) and annotated.

B: (*Left*) UMAP projection clustered by gene expression (same as in Figure 1C) displays 145,562 human DRG nuclei that passed RNA quality controls. Nuclei are colored by sex. DRG cell types are as in A. (*Right*) UMAP projection clustered by chromatin accessibility displays 67,025 human DRG nuclei that passed ATAC quality controls. Nuclei are colored by sex. DRG cell types are as in A.

C: Volcano plots display the log_2FC (x-axis) and adj. p-values (y-axis) of genes differentially expressed in mouse males compared to females for each DRG cell type ($log_2FC > 0.5$, adj. p-value < 0.05; threshold indicated by dotted black lines).

D: Volcano plots display the $\log_2$FC (x-axis) and adj. p-values (y-axis) of genes differentially expressed in human males compared to females for each DRG cell type ($\log_2$FC > 0.5, adj. p-value < 0.05; threshold indicated by dotted black lines).

E: Pie chart displays the fraction of MACS2-called peaks in mice DRG that are located in distinct genomic regions, as annotated by ChIPSeeker.[75] Peaks within the promoter window were defined as those falling −1 kb to +100 bp relative to the annotated TSS. Peaks were annotated as Distal if they lay within ±200 kb of a TSS and are not promoters, or as Intergenic if they were ≥200 kb from the nearest TSS.

F: Pie chart displays the fraction of MACS2-called peaks that are located in distinct genomic regions, as annotated by ChIPSeeker. Genomic regions defined as in E.

Figure S4: QC plots for snMultiome analysis of mouse and human DRG neuronal subtypes

A: Pie chart displays the fraction of MACS2-called peaks from mouse DRG neurons located in distinct genomic regions, as annotated by ChIPseeker. Promoter peaks are defined as those falling within −1 kb to +100 bp of the annotated TSS. Distal peaks are defined as regions within ±200 kb of a TSS but are not promoters, while Intergenic peaks are those ≥200 kb from the nearest TSS.

B: Pie chart displays the fraction of MACS2-called peaks from human DRG neurons located in distinct genomic regions, as annotated by ChIPseeker. Genomic regions are defined as in A.

C: Boxplot displays TSS enrichment scores per nucleus in mouse DRG neurons and non-neurons that passed all QC filters. TSS enrichment scores are defined as the fold change in read density at transcription start sites relative to the average background signal in the 100 bp distal flanking regions.

D: Boxplot displays TSS enrichment scores per nucleus in human DRG neurons and non-neurons that passed all QC filters. TSS enrichment scores are defined as the fold change in read density at transcription start sites relative to the average background signal in the 100 bp distal flanking regions.

E: Genomic tracks display chromatin accessibility at loci of canonical marker genes for each mouse DRG neuronal subtype. Accessibility at each locus (y-axis) is scaled to the max value across all cell types (rows). Gene bodies for each marker gene are displayed along the x-axis.

F: Genomic tracks of chromatin accessibility at loci of marker genes for each human DRG neuronal subtype. Accessibility at each locus (y-axis) is scaled to the max value across all cell types (rows). Gene bodies for each marker gene are displayed along the x-axis.

G: UMAP projection displays 18,238 mouse (*Top*) and 1,597 human (*Bottom*) nuclei clustered by their eRegulon scores (see Methods). Nuclei are colored by cell type annotations (as in Figure 2).

1525 **H:** Genomic tracks display chromatin accessibility of a candidate SCENIC+ CRE upstream of
1526 *Sstr2* in mice (left) and *PVALB* in humans (right). Boxed region indicates that the CREs are
1527 predicted binding sites for *BCL6* (mice) and *RUNX3* (humans). Loops show Pearson correlation
1528 (p-val. < 0.05) of enhancer accessibility and RNA expression of *BCL6* or *RUNX3.*

1529 **Figure S5: Screening candidate CREs for C-PEP.Calca+Sstr2 specificity**

1530 **A:** Representative images of mouse DRG neurons labeled by RNAscope *in situ* hybridization
1531 respectively for *Tac1* (magenta), *Mrgprd* (blue), *Hapln4* (white), *Fam19a4* (red), and viral *EGFP*
1532 (green) mRNA for CRE1, with insets illustrating colocalization of *EGFP* and *Tac1* for CRE1.
1533 Scale bar, 25 µm.

1534 **B:** Stacked bar plots of the fraction of each DRG neuronal subtype that express the respective
1535 CRE barcodes (>1 AAV transcript/cell). Colors correspond to C-NP or C-PEP subtypes; all other
1536 cell types are gray.

1537 **C:** Representative images of mouse DRG neurons infected with AAVs using NoCRE, CAG,
1538 CRE1, or CRE6 to drive *EGFP* expression. DRG neurons are labeled by RNAscope *in situ*
1539 hybridization for *Calca* (blue), *Sstr2* (magenta) and *EGFP* (green) mRNA. Scale bar, 25 µm.
1540 Quantification in D.

1541 **D:** Quantification of *Calca, Sstr2,* and *EGFP* mRNA colocalization per *EGFP*+ cell for No CRE,
1542 CAG, CRE1, and CRE6 (n = 3-4 lumbar DRGs per mouse from 3 animals). One-way ANOVA,
1543 $F(3,8) = 2.005$, $p = 0.0404$, Dunnett's multiple comparisons test for No CRE versus CRE1. Data
1544 are presented as mean ± SEM.

1545 **E:** Quantification of C-PEP.Calca+Sstr2 cells as a percentage of *Calca*+ cells that are also
1546 *Sstr2*+. Data are presented as mean ± SEM.
1547

1548
1549 **Figure S6: Characterization of CRE Library Complexity, Barcode Recovery, and**
1550 **Chromatin Accessibility**

1551 **A:** Distribution of CRE barcodes detected in the library AAV multiplexed CRE library as
1552 measured by next-generation DNA sequencing. Reads were binned by their associated CRE
1553 barcode (left y-axis) and their frequency distribution is plotted (dark gray). Cumulative
1554 distribution of detected CRE barcodes indicates coverage of all 106 prioritized CREs, the mini
1555 CMV enhancer, and no CRE negative controls (light gray, right y-axis) in the library.

1556 **B:** Histogram of number of DRG cells expressing viral barcodes from the multiplexed CRE
1557 library as measured by scRNA-seq (106 CREs + 2 controls). Only CREs detected in at least 6
1558 cells were used (red line). No CRE control was over-represented in the library to ensure robust
1559 comparison for each CRE.

1560 **C:** Dotplot displays cell class specificity ratios across 10 distinct DRG cell classes. The cell class
1561 specificity ratio (x-axis) is defined as the number of CRE barcode-positive cells in a specific cell
1562 class divided by the number of barcode-positive cells in all other cells. Dot color represents cell

class specificity ratio (Z-scaled), and dot size corresponds to the number of viral barcodes detected per cell type. Ratios are normalized to the negative control (NC: No CRE).

Figure S7: Constructing PAIN-net to aid in the prioritization of DRG CREs

A: Scatter plots compare experimentally measured (x-axis) versus predicted (y-axis) $\log_2$FC of each neuronal subtype compared to all others. Data points are all held-out validation sequences (15%). Each subplot reports the Pearson correlation (*r*) and sequence count (*n*). The dashed diagonal line (y=x) indicates perfect prediction accuracy.

B: PAIN-net predicted on-target accessibility of candidate CRE sequences correlates with their measured on-target accessibility. On-target cell types are defined by the accessibility at each CRE locus in the multiplexed CRE library. CREs were designed to target multiple neuronal subtypes, so predicted and measured on-target accessibilities were summed across their on-target subtypes (Table S16).

C: PAIN-net predicts bases that drive cell type-specific differential accessibility in motifs of cell-type-specific transcription factors. Split violin plots display the distributions of predicted mutation impacts (delta $\log_2$FC in each cell type compared to all other cell types) for the core consensus motif (purple) and the 10 bp flanking sequence (gray), aggregated across 1,000 randomized genomic backgrounds. Plots represent the aggregate results for *Atf3, Runx1, Jun,* and *Isl1* across DRG cell types (sorted by effect size). The X-axis displays the mean change in accessibility across the three possible mutations.

Figure S8: Functional applications of nociceptor-specific AAVs

A: Summary data showing the minimum current required to elicit an action potential (rheobase). Data is the same as in Figure 6C but also includes data from all CRE1-K_{ir}2.1 neurons without functional stratification (>200 pA rheobase, red dashed line) to confirm K_{ir}2.1 expression. Statistical significance was determined using a one-way ANOVA with multiple comparisons (* $p < 0.05$, ** $p < 0.01$, **** $p < 0.0001$). Sample sizes (Control, n = 20; CAG, n = 9; CRE1 total, n = 45) represent individual neurons recorded across at least 2 independent preparations.

Supplemental Tables

1. Donor information - SuppTable1_Sample_Donor_Sex.csv
2. Mouse DRG DE List - SuppTable2_DE_Mm_DRG.csv
3. Human DRG DE List - SuppTable3_DE_Hg_DRG.csv
4. Mouse DRG DA List - SuppTable4_DA_Mm_DRG.csv
5. Human DRG DA List - SuppTable5_DA_Hg_DRG.csv
6. Mouse DRG SCENIC+ Table - SuppTable6_Scenicplus_Mm_DRG.csv
7. Human DRG SCENIC+ Table - SuppTable7_Scenicplus_Hg_DRG.csv
8. Mouse DRG Neurons DE List - SuppTable8_DE_Mm_DRG_Neurons.csv
9. Human DRG Neurons DE List - SuppTable9_DE_Hg_DRG_Neurons.csv
10. Mouse DRG Neurons DA List - SuppTable10_DA_Mm_DRG_Neurons.csv
11. Human DRG Neurons DA List - SuppTable11_DA_Hg_DRG_Neurons.csv
12. Mouse DRG Neurons SCENIC+ Table - SuppTable12_SCENIC_eregulons_Mouse_DRG_Neurons.csv
13. Human DRG Neurons SCENIC+ Table - SuppTable13_SCENIC_eregulons_Human_DRG_Neurons.csv
14. Enhancer prioritization Table - SuppTable14_Mouse_DA_Peaks_Conservation.csv
15. Opool on-target definitions - SuppTable15_oPool_ontargetness.csv
16. PAIN-net prediction values - SuppTable16_validation_predictions.csv
17. List of all CREs and details - SuppTable17_CRE_key_mm10_sequences.csv

Acknowledgements

We would like to thank donors and donors' families, without whom much of this study would not be possible. We also thank the NIH HEAL PRECISION Human Pain Network for its support and for fostering a highly collaborative research environment.

We thank Junning Case from the BWH Center for Cellular Profiling for assistance in experimental design and protocol optimization. We would also like to thank Rajesh Krishnan from The ARCND Flow Cytometry Core Facility for assistance in protocol optimization. We also thank the BWH Neurotechnology Studio. The authors thank Jesse Gillis from the University of Toronto for his advice on how to build PAIN-net. Schematics in Figures 1, 3, 4, and 6 were adapted using icons provided by BioRender.com. Finally, the authors would like to thank members of the Renthal Lab for their advice throughout the process and Hannah MacMillan for assistance with figure diagrams.

Funding: This work was supported by the Burroughs Wellcome Fund (W.R.), Rita Allen Foundation (W.R.), the Edwards PhD Studentship in Pain Research (M.X.), and the National Institute of Neurological Disorders and Stroke [U19NS130617 (W.R.), R01NS119476 (W.R.)]. W.R. also receives support from the National Institute of Drug Abuse (DP1DA054343), National Eye Institute (U01EY034709), and MGB Gene and Cell Therapy Institute. D.S.G was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number T32GM007592.

Competing interests: W.R. has received research funding from Pfizer and Eli Lilly. The authors declare that they have no other competing interests.

Data and Code Availability: All custom code used to integrate data is available on Git (https://github.com/Renthal-Lab/painseq-multiome). Raw sequencing data and final objects will be deposited into GEO after manuscript acceptance.

Author contributions: L.S.H., P.B., S.A.B., and W.R. designed, conceptualized, wrote, and reviewed the study. **J.W., L.J., M.X., L.Y., and L.S.H.** constructed AAVs. **P.B., S.A.B., and D.Z.** performed bioinformatics analyses. **E.S.** performed mouse and human DRG multi-omic profiling. **J.L., M.X., L.Y., E.G., C.A., L.S.H., K.P., E.K.W., and J.W.** performed and analyzed *in situ* hybridization. **L.S.H. and J.W.** performed single-cell sequencing. **J.L. and S.J.** performed iPSC experiments. **L.S.H., J.L., J.W., H.J.Y., and Q.L.** performed *in vivo* injections and procedures. **J.N.** and **B.W.** performed electrophysiology and associated analysis. **P.B., J.W., J.X.J.L., S.H. and W.R.** performed enhancer prioritization and library design. **D.G.** performed donor DNA sequencing and demultiplexing. **L.S.H., P.B., S.A.B., and W.R.** wrote the manuscript with input from all authors. **W.R.** supervised the research. All authors reviewed and approved the manuscript.

References

1. The Lancet Regional Health – Americas. Opioid crisis: addiction, overprescription, and insufficient primary prevention. *Lancet Reg. Health - Am.* **23**, 100557 (2023).
2. Benyamin, R. Opioid Complications and Side Effects. *Pain Physician* **2s;11**, S105–S120 (2008).
3. Calvo, M. *et al.* The Genetics of Neuropathic Pain from Model Organisms to Clinical Application. *Neuron* **104**, 637–653 (2019).
4. Renthal, W. Seeking Balance Between Pain Relief and Safety: CDC Issues New Opioid-Prescribing Guidelines. *JAMA Neurol.* **73**, 513 (2016).
5. Bhuiyan, S. A. *et al.* A Reference Atlas of the Human Dorsal Root Ganglion. *bioRxiv* 2025.11.05.686654 (2025) doi:10.1101/2025.11.05.686654.
6. Bhuiyan, S. A. *et al.* Harmonized cross-species cell atlases of trigeminal and dorsal root ganglia. *Sci. Adv.* **10**, (2024).
7. Li, L. *et al.* The Functional Organization of Cutaneous Low-Threshold Mechanosensory Neurons. *Cell* **147**, 1615–1627 (2011).
8. Qi, L. *et al.* A mouse DRG genetic toolkit reveals morphological and physiological diversity of somatosensory neuron subtypes. *Cell* **187**, 1508-1526.e16 (2024).
9. Bai, L. *et al.* Genetic Identification of an Expansive Mechanoreceptor Sensitive to Skin Stroking. *Cell* **163**, 1783–1795 (2015).
10. Rutlin, M. *et al.* The Cellular and Molecular Basis of Direction Selectivity of Aδ-LTMRs. *Cell* **159**, 1640–1651 (2014).
11. Elias, L. J. *et al.* Touch neurons underlying dopaminergic pleasurable touch and sexual receptivity. *Cell* **186**, 577-590.e16 (2023).
12. Voisin, T. *et al.* The $CysLT_2$ R receptor mediates leukotriene C_4 -driven acute and chronic itch. *Proc. Natl. Acad. Sci.* **118**, e2022087118 (2021).
13. Renthal, W. *et al.* Transcriptional Reprogramming of Distinct Peripheral Sensory Neuron Subtypes after Axonal Injury. *Neuron* **108**, 128-144.e9 (2020).
14. Mich, J. K. *et al.* Functional enhancer elements drive subclass-selective expression from mouse to primate neocortex. *Cell Rep.* **34**, 108754 (2021).
15. Ben-Simon, Y. *et al.* A suite of enhancer AAVs and transgenic mouse lines for genetic access to cortical cell types. Preprint at https://doi.org/10.1101/2024.06.10.597244 (2024).
16. Graybuck, L. T. *et al.* Enhancer viruses for combinatorial cell-subclass-specific labeling. *Neuron* **109**, 1449-1464.e13 (2021).
17. Dimidschstein, J. *et al.* A viral strategy for targeting and manipulating interneurons across vertebrate species. *Nat. Neurosci.* **19**, 1743–1749 (2016).
18. Lee, M. Y. Y., Kaestner, K. H. & Li, M. Benchmarking algorithms for joint integration of unpaired and paired single-cell RNA-seq and ATAC-seq data. *Genome Biol.* **24**, 244 (2023).
19. Techameena, P., Feng, X., Zhang, K. & Hadjab, S. The single-cell transcriptomic atlas iPain identifies senescence of nociceptors as a therapeutical target for chronic pain treatment. *Nat. Commun.* **15**, 8585 (2024).
20. Krauter, D. *et al.* Spatial organization, chromatin accessibility and gene-regulatory programs defining mouse sensory neurons. *Commun. Biol.* **8**, 908 (2025).

21. Kuhn, R. M., Haussler, D. & Kent, W. J. The UCSC genome browser and associated tools. *Brief. Bioinform.* **14**, 144–161 (2013).
22. Hinrichs, A. S. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* **34**, D590–D598 (2006).
23. Andrews, G. *et al.* Mammalian evolution of human cis-regulatory elements and transcription factor binding sites. *Science* **380**, eabn7930 (2023).
24. Bravo González-Blas, C. *et al.* SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nat. Methods* **20**, 1355–1367 (2023).
25. Sun, Y. *et al.* A central role for Islet1 in sensory neuron development linking sensory and spinal gene regulatory programs. *Nat. Neurosci.* **11**, 1283–1293 (2008).
26. Nagy, M. A. *et al.* Cis-regulatory elements driving motor neuron-selective viral payload expression within the mammalian spinal cord. *Proc. Natl. Acad. Sci.* **121**, e2418024121 (2024).
27. Lawler, A. J. *et al.* Machine learning sequence prioritization for cell type-specific enhancer design. *eLife* **11**, e69571 (2022).
28. Hunker, A. C. *et al.* Enhancer AAV toolbox for accessing and perturbing striatal cell types and circuits. Preprint at https://doi.org/10.1101/2024.09.27.615553 (2024).
29. Zemke, N. R. *et al.* Conserved and divergent gene regulatory programs of the mammalian neocortex. *Nature* **624**, 390–402 (2023).
30. Furlanis, E. *et al.* An enhancer-AAV toolbox to target and manipulate distinct interneuron subtypes. Preprint at https://doi.org/10.1101/2024.07.17.603924 (2024).
31. Usoskin, D. *et al.* Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. *Nat. Neurosci.* **18**, 145–153 (2015).
32. Sharma, N. *et al.* The emergence of transcriptional identity in somatosensory neurons. *Nature* **577**, 392–398 (2020).
33. Tavares-Ferreira, D. *et al.* Spatial transcriptomics of dorsal root ganglia identifies molecular signatures of human nociceptors. *Sci. Transl. Med.* **14**, eabj8186 (2022).
34. Moore, J. E. *et al.* An expanded registry of candidate cis-regulatory elements. *Nature* https://doi.org/10.1038/s41586-025-09909-9 (2026) doi:10.1038/s41586-025-09909-9.
35. Levanon, D. The Runx3 transcription factor regulates development and survival of TrkC dorsal root ganglia neurons. *EMBO J.* **21**, 3454–3463 (2002).
36. Orlovsky, K. *et al.* Runx3, Brn3a and Isl1 interplay orchestrates the transcriptional program in the early stages of proprioceptive neuron development. *PLOS Genet.* **20**, e1011401 (2024).
37. Chan, K. Y. *et al.* Engineered AAVs for efficient noninvasive gene delivery to the central and peripheral nervous systems. *Nat. Neurosci.* **20**, 1172–1179 (2017).
38. Kim, J.-Y., Grunke, S. D., Levites, Y., Golde, T. E. & Jankowsky, J. L. Intracerebroventricular Viral Injection of the Neonatal Mouse Brain for Persistent and Widespread Neuronal Transduction. *J. Vis. Exp.* 51863 (2014) doi:10.3791/51863.
39. Kawasaki, H. *et al.* Intracerebroventricular and Intravascular Injection of Viral Particles and Fluorescent Microbeads into the Neonatal Brain. *J. Vis. Exp.* 54164 (2016) doi:10.3791/54164.
40. Glascock, J. J. *et al.* Delivery of Therapeutic Agents Through Intracerebroventricular (ICV) and Intravenous (IV) Injection in Mice. *J. Vis. Exp.* 2968 (2011) doi:10.3791/2968.

41. Hrvatin, S. *et al.* A scalable platform for the development of cell-type-specific viral drivers. *eLife* **8**, e48089 (2019).
42. Gosai, S. J. *et al.* Machine-guided design of cell-type-targeting cis-regulatory elements. *Nature* **634**, 1211–1220 (2024).
43. Kempynck, N. *et al.* CREsted: modeling genomic and synthetic cell type-specific enhancers across tissues and species.
44. Minnoye, L. *et al.* Cross-species analysis of enhancer logic using deep learning. *Genome Res.* **30**, 1815–1834 (2020).
45. Chen, A. I., De Nooij, J. C. & Jessell, T. M. Graded Activity of Transcription Factor Runx3 Specifies the Laminar Termination Pattern of Sensory Axons in the Developing Spinal Cord. *Neuron* **49**, 395–408 (2006).
46. Lou, S., Duan, B., Vong, L., Lowell, B. B. & Ma, Q. Runx1 Controls Terminal Morphology and Mechanosensitivity of VGLUT3-expressing C-Mechanoreceptors. *J. Neurosci.* **33**, 870–882 (2013).
47. Petruska, J. C., Napaporn, J., Johnson, R. D., Gu, J. G. & Cooper, B. Y. Subclassified Acutely Dissociated Cells of Rat DRG: Histochemistry and Patterns of Capsaicin-, Proton-, and ATP-Activated Currents. *J. Neurophysiol.* **84**, 2365–2379 (2000).
48. Zheng, Y. *et al.* Deep Sequencing of Somatosensory Neurons Reveals Molecular Determinants of Intrinsic Physiological Properties. *Neuron* **103**, 598-616.e7 (2019).
49. Ma, C., Rosenzweig, J., Zhang, P., Johns, D. C. & LaMotte, R. H. Expression of Inwardly Rectifying Potassium Channels by an Inducible Adenoviral Vector Reduced the Neuronal Hyperexcitability and Hyperalgesia Produced by Chronic Compression of the Spinal Ganglion. *Mol. Pain* **6**, 1744-8069-6–65 (2010).
50. Sharma, A., Toepfer, C. N., Schmid, M., Garfinkel, A. C. & Seidman, C. E. Differentiation and Contractile Analysis of GFP-Sarcomere Reporter hiPSC-Cardiomyocytes. *Curr. Protoc. Hum. Genet.* **96**, (2018).
51. Deng, T. *et al.* Scalable generation of sensory neurons from human pluripotent stem cells. *Stem Cell Rep.* **18**, 1030–1047 (2023).
52. Walters, E. T., Crook, R. J., Neely, G. G., Price, T. J. & Smith, E. S. J. Persistent nociceptor hyperactivity as a painful evolutionary adaptation. *Trends Neurosci.* **46**, 211–227 (2023).
53. Reichling, D. B. & Levine, J. D. Critical role of nociceptor plasticity in chronic pain. *Trends Neurosci.* **32**, 611–618 (2009).
54. Kupari, J. *et al.* Single cell transcriptomics of primate sensory neurons identifies cell types associated with chronic pain. *Nat. Commun.* **12**, 1510 (2021).
55. Yu, H. *et al.* Leveraging deep single-soma RNA sequencing to explore the neural basis of human somatosensation. *Nat. Neurosci.* **27**, 2326–2340 (2024).
56. Johansen, N. J. *et al.* Evaluating Methods for the Prediction of Cell Type-Specific Enhancers in the Mammalian Cortex. *bioRxiv* 2024.08.21.609075 (2025) doi:10.1101/2024.08.21.609075.
57. Leone, M. J. *et al.* Combining Machine Learning and Multiplexed, In Situ Profiling to Engineer Cell Type and Behavioral Specificity. *bioRxiv* 2025.06.20.660790 (2025) doi:10.1101/2025.06.20.660790.

1778 58. McDiarmid, T. A. *et al.* Large-scale discovery of neural enhancers for cis-regulation
1779 therapies. *bioRxiv* 2025.11.04.686611 (2025) doi:10.1101/2025.11.04.686611.
1780 59. Li, L., Xu, B. & Liu, C. Sample enrichment for single-nucleus sequencing using
1781 concanavalin A-conjugated magnetic beads. *STAR Protoc.* **4**, 102595 (2023).
1782 60. Fleming, S. J. *et al.* Unsupervised removal of systematic background noise from
1783 droplet-based single-cell experiments using CellBender. *Nat. Methods* **20**, 1323–1335 (2023).
1784 61. Yang, L. *et al.* Human and mouse trigeminal ganglia cell atlas implicates multiple cell
1785 types in migraine. *Neuron* **110**, 1806-1821.e8 (2022).
1786 62. Heaton, H. *et al.* Souporcell: robust clustering of single-cell RNA-seq data by genotype
1787 without reference genotypes. *Nat. Methods* **17**, 615–620 (2020).
1788 63. Granja, J. M. *et al.* ArchR is a scalable software package for integrative single-cell
1789 chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021).
1790 64. Joung, J. *et al.* Genome-scale CRISPR-Cas9 knockout and transcriptional activation
1791 screening. *Nat. Protoc.* **12**, 828–863 (2017).
1792 65. DuBreuil, D. M. *et al.* A high-content platform for physiological profiling and unbiased
1793 classification of individual neurons. *Cell Rep. Methods* **1**, 100004 (2021).
1794 66. Perner, C. & Sokol, C. L. Protocol for dissection and culture of murine dorsal root ganglia
1795 neurons to study neuropeptide release. *STAR Protoc.* **2**, 100333 (2021).
1796 67. Sleigh, J. N., Weir, G. A. & Schiavo, G. A simple, step-by-step dissection protocol for the
1797 rapid isolation of mouse dorsal root ganglia. *BMC Res. Notes* **9**, 82 (2016).
1798 68. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21
1799 (2019).
1800 69. Kaplow, I. M. *et al.* Relating enhancer genetic variation across mammals to complex
1801 phenotypes using machine learning. *Science* **380**, eabm7993 (2023).
1802 70. Avsec, Ž. *et al.* Base-resolution models of transcription-factor binding reveal soft motif
1803 syntax. *Nat. Genet.* **53**, 354–366 (2021).
1804 71. Kelley, D. R. *et al.* Sequential regulatory activity prediction across chromosomes with
1805 convolutional neural networks. *Genome Res.* **28**, 739–750 (2018).
1806 72. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin
1807 state analysis with Signac. *Nat. Methods* **18**, 1333–1341 (2021).
1808 73. Zhang, H.-X. B. & Bean, B. P. Cannabidiol Inhibition of Murine Primary Nociceptors:
1809 Tight Binding to Slow Inactivated States of Na_v 1.8 Channels. *J. Neurosci.* **41**, 6371–6387
1810 (2021).
1811 74. Johns, D. C., Marx, R., Mains, R. E., O'Rourke, B. & Marbán, E. Inducible Genetic
1812 Suppression of Neuronal Excitability. *J. Neurosci.* **19**, 1691–1697 (1999).
1813 75. Yu, G., Wang, L.-G. & He, Q.-Y. ChIPseeker: an R/Bioconductor package for ChIP peak
1814 annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).

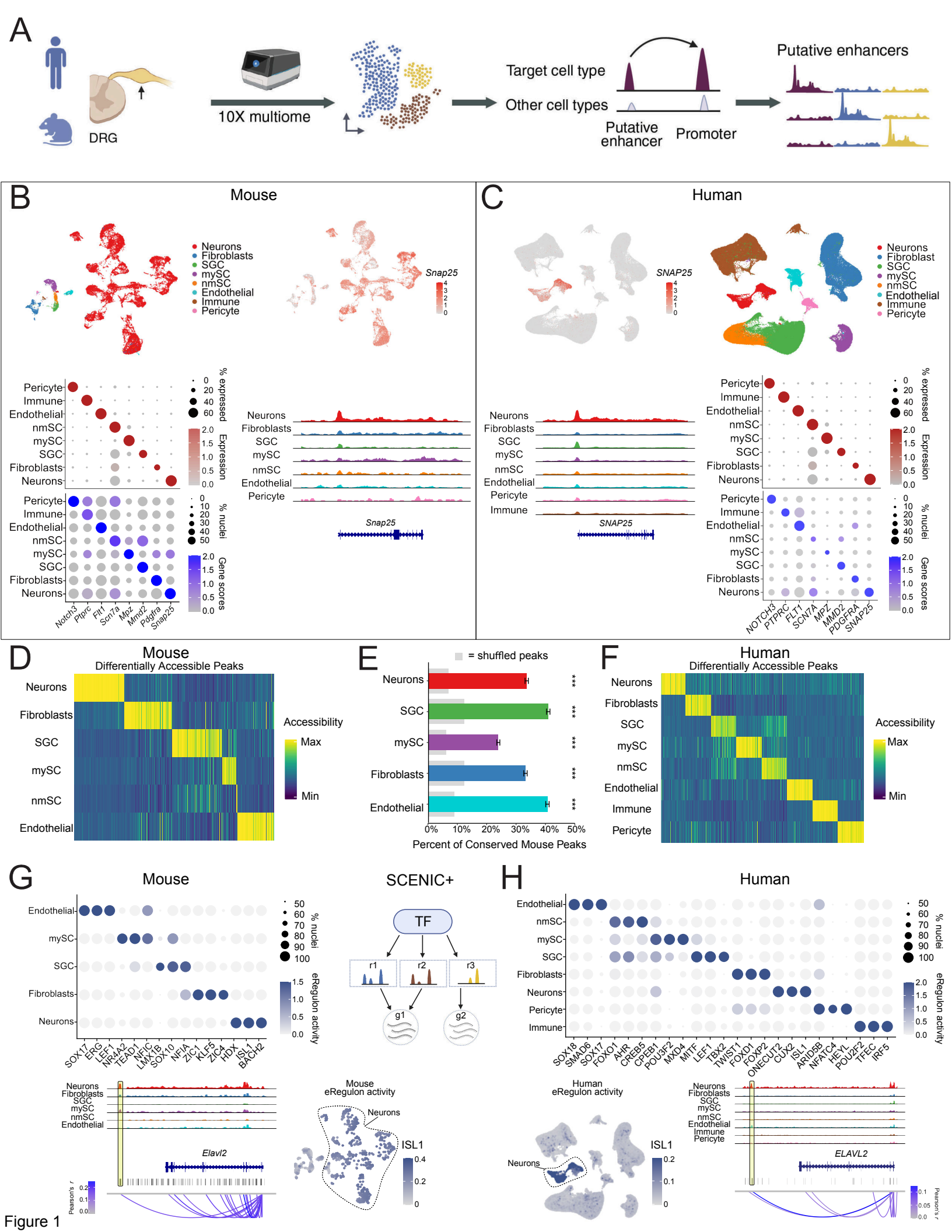
A
DRG
10X multiome
Target cell type
Other cell types
Putative enhancer
Promoter
Putative enhancers
B
Mouse
Neurons
Fibroblasts
SGC
mySC
nmSC
Endothelial
Immune
Pericyte
Snap25
% expressed
Expression
% nuclei
Gene scores
Notch3
Ptprc
Flt1
Scn7a
Mpz
Mmd2
Pdgfra
Snap25
C
Human
SNAP25
NOTCH3
PTPRC
FLT1
SCN7A
MPZ
MMD2
PDGFRA
SNAP25
D
Mouse
Differentially Accessible Peaks
Accessibility
Max
Min
E
= shuffled peaks
Percent of Conserved Mouse Peaks
F
Human
Differentially Accessible Peaks
G
Mouse
SCENIC+
TF
r1
r2
r3
g1
g2
eRegulon activity
SOX17
ERG
LEF1
NR4A2
TEAD1
NFIC
LMX1B
SOX10
NFIA
ZIC1
KLF5
ZIC4
HDX
ISL1
BACH2
Elavl2
Mouse
eRegulon activity
ISL1
H
Human
SOX18
SMAD6
SOX17
FOXO1
AHR
CREB5
CPEB1
POU3F2
MXD4
MITF
LEF1
TBX2
TWIST1
FOXD1
FOXP2
ONECUT2
CUX2
ISL1
ARID5B
NFATC4
HEYL
POU2F2
TFEC
IRF5
ELAVL2
Human
eRegulon activity
Pearson's r

Figure 1

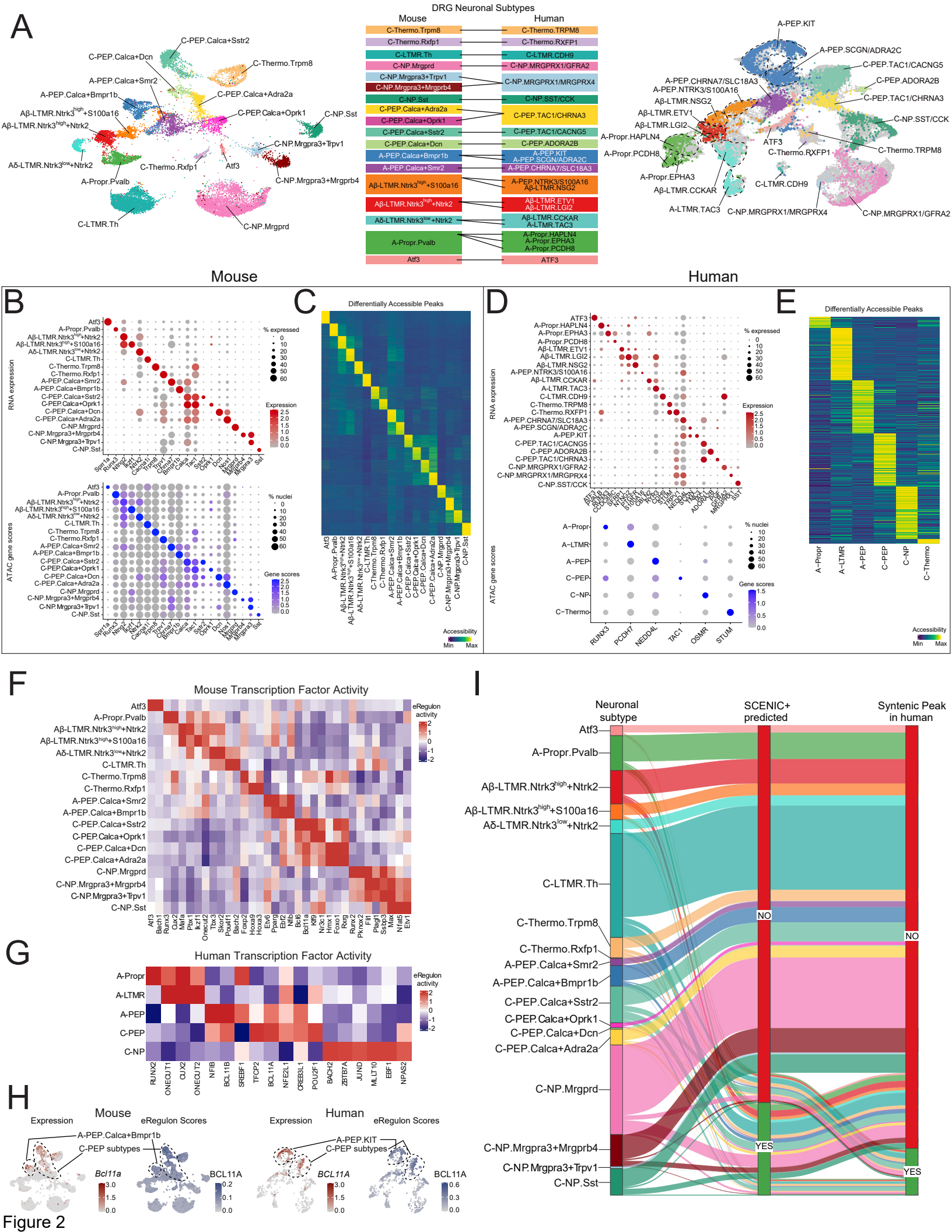
A
DRG Neuronal Subtypes
Mouse
Human
C-PEP.Calca+Sstr2
C-PEP.Calca+Dcn
C-Thermo.Trpm8
A-PEP.Calca+Smr2
A-PEP.Calca+Bmpr1b
C-PEP.Calca+Adra2a
Aβ-LTMR.Ntrk3high+S100a16
C-PEP.Calca+Oprk1
C-NP.Sst
Aβ-LTMR.Ntrk3high+Ntrk2
C-NP.Mrgpra3+Trpv1
Aδ-LTMR.Ntrk3low+Ntrk2
C-Thermo.Rxfp1
Atf3
C-NP.Mrgpra3+Mrgprb4
A-Propr.Pvalb
C-LTMR.Th
C-NP.Mrgprd
C-Thermo.Trpm8
C-Thermo.Rxfp1
C-LTMR.Th
C-NP.Mrgprd
C-NP.Mrgpra3+Trpv1
C-NP.Mrgpra3+Mrgprb4
A-PEP.Calca+Smr2
A-PEP.Calca+Bmpr1b
C-PEP.Calca+Sstr2
C-PEP.Calca+Oprk1
C-PEP.Calca+Sstr2
C-PEP.Calca+Dcn
A-PEP.Calca+Bmpr1b
A-PEP.Calca+Smr2
Aβ-LTMR.Ntrk3high+S100a16
Aβ-LTMR.Ntrk3high+Ntrk2
Aδ-LTMR.Ntrk3low+Ntrk2
A-Propr.Pvalb
Atf3
C-Thermo.TRPM8
C-Thermo.RXFP1
C-LTMR.CDH9
C-NP.MRGPRX1/GFRA2
C-NP.MRGPRX1/MRGPRX4
C-NP.SST/CCK
C-PEP.TAC1/CHRNA3
C-PEP.TAC1/CACNG5
C-PEP.ADORA2B
A-PEP.KIT
A-PEP.SCGN/ADRA2C
A-PEP.CHRNA7/SLC18A3
A-PEP.NTRK3/S100A16
Aβ-LTMR.NSG2
Aβ-LTMR.ETV1
Aβ-LTMR.LGI2
Aβ-LTMR.CCKAR
A-LTMR.TAC3
A-Propr.HAPLN4
A-Propr.EPHA3
A-Propr.PCDH8
ATF3
Mouse
Human
B
RNA expression
% expressed
Expression
ATAC gene scores
% nuclei
Gene scores
C
Differentially Accessible Peaks
Accessibility
Min
Max
D
RNA expression
% expressed
Expression
ATAC gene scores
% nuclei
Gene scores
A−Propr
A−LTMR
A−PEP
C−PEP
C−NP
C−Thermo
RUNX3
PCDH7
NEDD4L
TAC1
OSMR
STUM
E
Differentially Accessible Peaks
Accessibility
Min
Max
F
Mouse Transcription Factor Activity
eRegulon activity
G
Human Transcription Factor Activity
eRegulon activity
A-Propr
A-LTMR
A-PEP
C-PEP
C-NP
RUNX2
ONECUT1
CUX2
ONECUT2
NFIB
BCL11B
SREBF1
TFCP2
BCL11A
NFE2L1
CREB3L1
POU2F1
BACH2
ZBTB7A
JUND
MLLT10
EBF1
NPAS2
I
Neuronal subtype
SCENIC+ predicted
Syntenic Peak in human
Atf3
A-Propr.Pvalb
Aβ-LTMR.Ntrk3high+Ntrk2
Aβ-LTMR.Ntrk3high+S100a16
Aδ-LTMR.Ntrk3low+Ntrk2
C-LTMR.Th
C-Thermo.Trpm8
C-Thermo.Rxfp1
A-PEP.Calca+Smr2
A-PEP.Calca+Bmpr1b
C-PEP.Calca+Sstr2
C-PEP.Calca+Oprk1
C-PEP.Calca+Dcn
C-PEP.Calca+Adra2a
C-NP.Mrgprd
C-NP.Mrgpra3+Mrgprb4
C-NP.Mrgpra3+Trpv1
C-NP.Sst
NO
YES
NO
YES
H
Mouse
Human
Expression
eRegulon Scores
A-PEP.Calca+Bmpr1b
C-PEP subtypes
Bcl11a
BCL11A
A-PEP.KIT
C-PEP subtypes
BCL11A
BCL11A

Figure 2

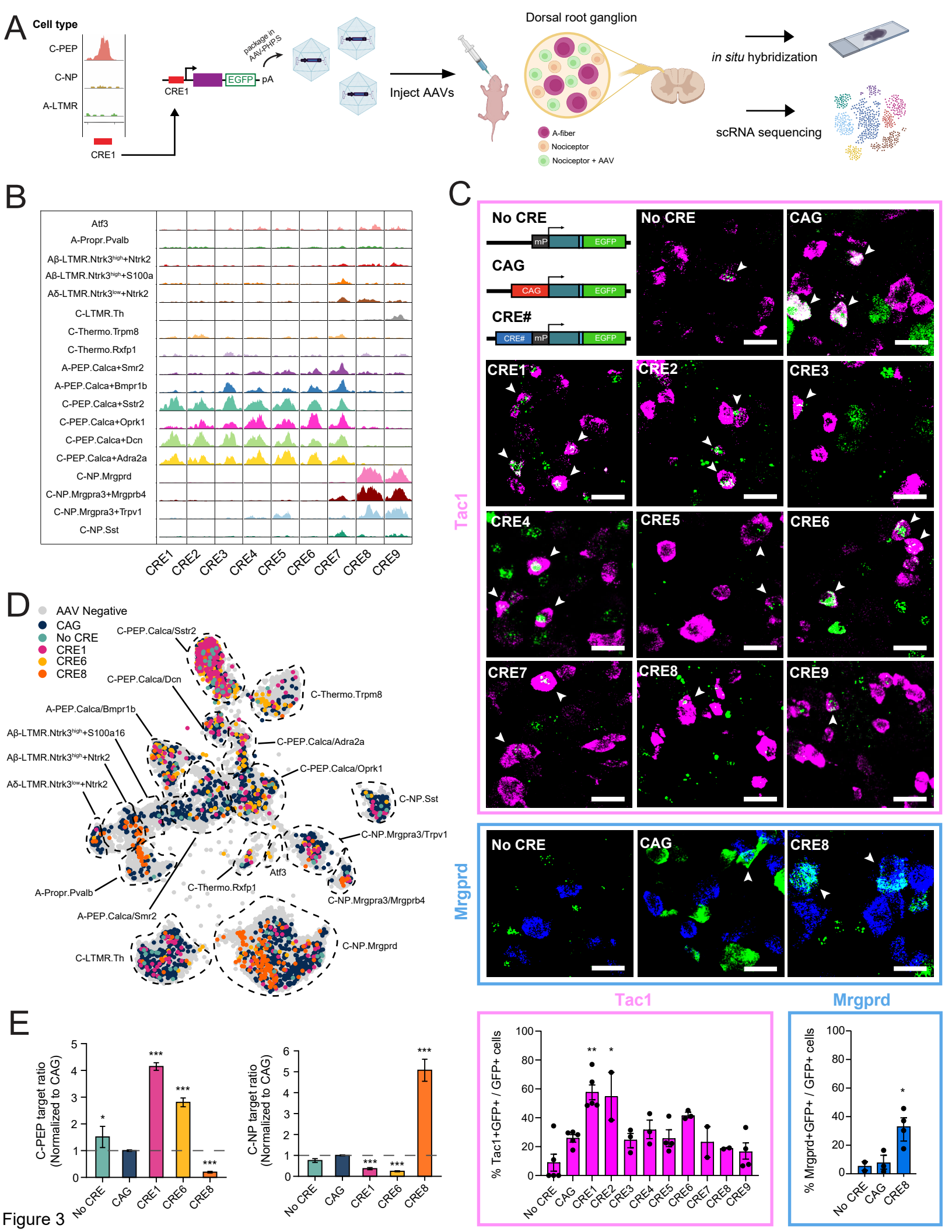
A
Cell type
C-PEP
C-NP
A-LTMR
CRE1
CRE1
EGFP
pA
package in AAV-PHP.S
Inject AAVs
Dorsal root ganglion
A-fiber
Nociceptor
Nociceptor + AAV
in situ hybridization
scRNA sequencing
B
Atf3
A-Propr.Pvalb
Aβ-LTMR.Ntrk3high+Ntrk2
Aβ-LTMR.Ntrk3high+S100a
Aδ-LTMR.Ntrk3low+Ntrk2
C-LTMR.Th
C-Thermo.Trpm8
C-Thermo.Rxfp1
A-PEP.Calca+Smr2
A-PEP.Calca+Bmpr1b
C-PEP.Calca+Sstr2
C-PEP.Calca+Oprk1
C-PEP.Calca+Dcn
C-PEP.Calca+Adra2a
C-NP.Mrgprd
C-NP.Mrgpra3+Mrgprb4
C-NP.Mrgpra3+Trpv1
C-NP.Sst
CRE1 CRE2 CRE3 CRE4 CRE5 CRE6 CRE7 CRE8 CRE9
C
No CRE
mP
EGFP
CAG
CRE#
Tac1
Mrgprd
CRE1 CRE2 CRE3 CRE4 CRE5 CRE6 CRE7 CRE8 CRE9
D
AAV Negative
CAG
No CRE
CRE1
CRE6
CRE8
C-PEP.Calca/Sstr2
C-PEP.Calca/Dcn
C-Thermo.Trpm8
A-PEP.Calca/Bmpr1b
Aβ-LTMR.Ntrk3high+S100a16
Aβ-LTMR.Ntrk3high+Ntrk2
Aδ-LTMR.Ntrk3low+Ntrk2
C-PEP.Calca/Adra2a
C-PEP.Calca/Oprk1
C-NP.Sst
C-NP.Mrgpra3/Trpv1
Atf3
C-Thermo.Rxfp1
C-NP.Mrgpra3/Mrgprb4
A-Propr.Pvalb
A-PEP.Calca/Smr2
C-NP.Mrgprd
C-LTMR.Th
E
C-PEP target ratio (Normalized to CAG)
C-NP target ratio (Normalized to CAG)
No CRE CAG CRE1 CRE6 CRE8
Tac1
% Tac1+GFP+ / GFP+ cells
No CRE CAG CRE1 CRE2 CRE3 CRE4 CRE5 CRE6 CRE7 CRE8 CRE9
Mrgprd
% Mrgprd+GFP+ / GFP+ cells
No CRE CAG CRE8

Figure 3

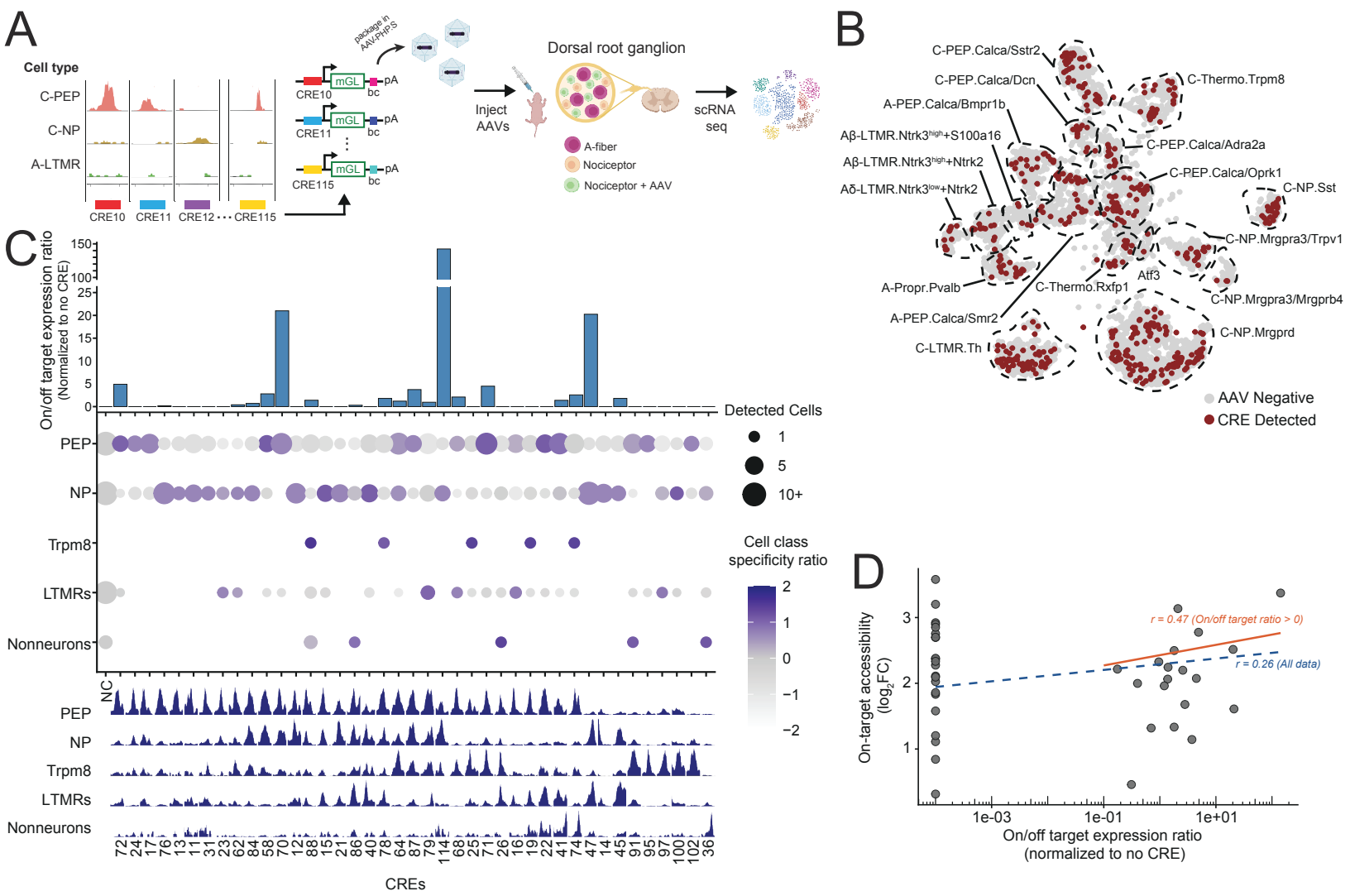
A
Cell type
C-PEP
C-NP
A-LTMR
CRE10
CRE11
CRE12
CRE115
package in AAV-PHP.S
mGL
pA
bc
Inject AAVs
Dorsal root ganglion
A-fiber
Nociceptor
Nociceptor + AAV
scRNA seq
B
C-PEP.Calca/Sstr2
C-PEP.Calca/Dcn
C-Thermo.Trpm8
A-PEP.Calca/Bmpr1b
Aβ-LTMR.Ntrk3high+S100a16
C-PEP.Calca/Adra2a
Aβ-LTMR.Ntrk3high+Ntrk2
C-PEP.Calca/Oprk1
Aδ-LTMR.Ntrk3low+Ntrk2
C-NP.Sst
C-NP.Mrgpra3/Trpv1
Atf3
A-Propr.Pvalb
C-Thermo.Rxfp1
C-NP.Mrgpra3/Mrgprb4
A-PEP.Calca/Smr2
C-NP.Mrgprd
C-LTMR.Th
AAV Negative
CRE Detected
C
On/off target expression ratio (Normalized to no CRE)
Detected Cells
1
5
10+
PEP
NP
Trpm8
LTMRs
Nonneurons
Cell class specificity ratio
NC
CREs
D
On-target accessibility (log2FC)
r = 0.47 (On/off target ratio > 0)
r = 0.26 (All data)
On/off target expression ratio (normalized to no CRE)

Figure 4

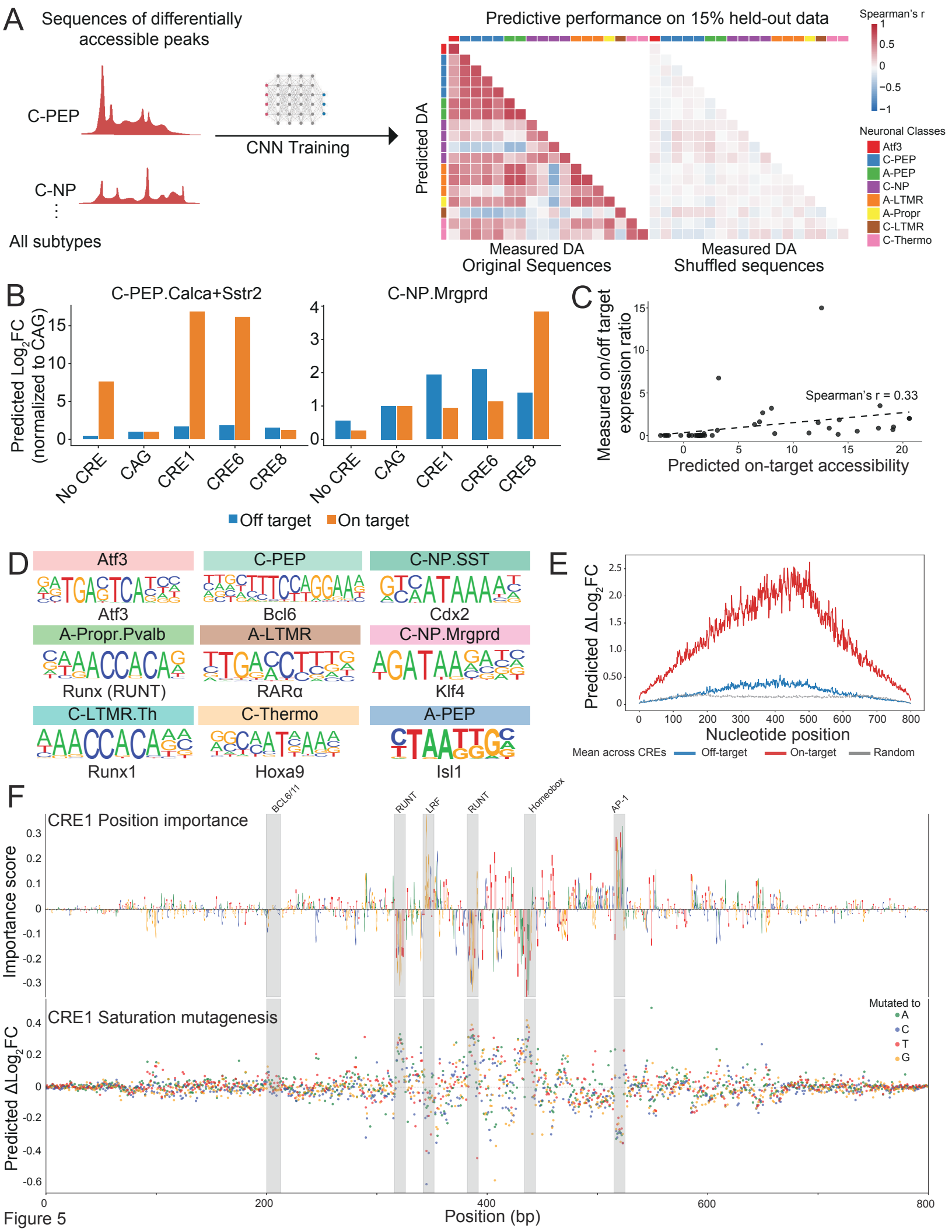
A
Sequences of differentially accessible peaks
C-PEP
C-NP
All subtypes
CNN Training
Predictive performance on 15% held-out data
Predicted DA
Measured DA
Original Sequences
Measured DA
Shuffled sequences
Spearman's r
Neuronal Classes
Atf3
C-PEP
A-PEP
C-NP
A-LTMR
A-Propr
C-LTMR
C-Thermo
B
C-PEP.Calca+Sstr2
C-NP.Mrgprd
Predicted Log2FC (normalized to CAG)
No CRE
CAG
CRE1
CRE6
CRE8
Off target
On target
C
Measured on/off target expression ratio
Spearman's r = 0.33
Predicted on-target accessibility
D
Atf3
C-PEP
C-NP.SST
A-Propr.Pvalb
A-LTMR
C-NP.Mrgprd
C-LTMR.Th
C-Thermo
A-PEP
Atf3
Bcl6
Cdx2
Runx (RUNT)
RARα
Klf4
Runx1
Hoxa9
Isl1
E
Predicted ΔLog2FC
Nucleotide position
Mean across CREs
Off-target
On-target
Random
F
BCL6/11
RUNT
LRF
RUNT
Homeobox
AP-1
CRE1 Position importance
Importance score
CRE1 Saturation mutagenesis
Predicted ΔLog2FC
Mutated to
A
C
T
G
Position (bp)

Figure 5

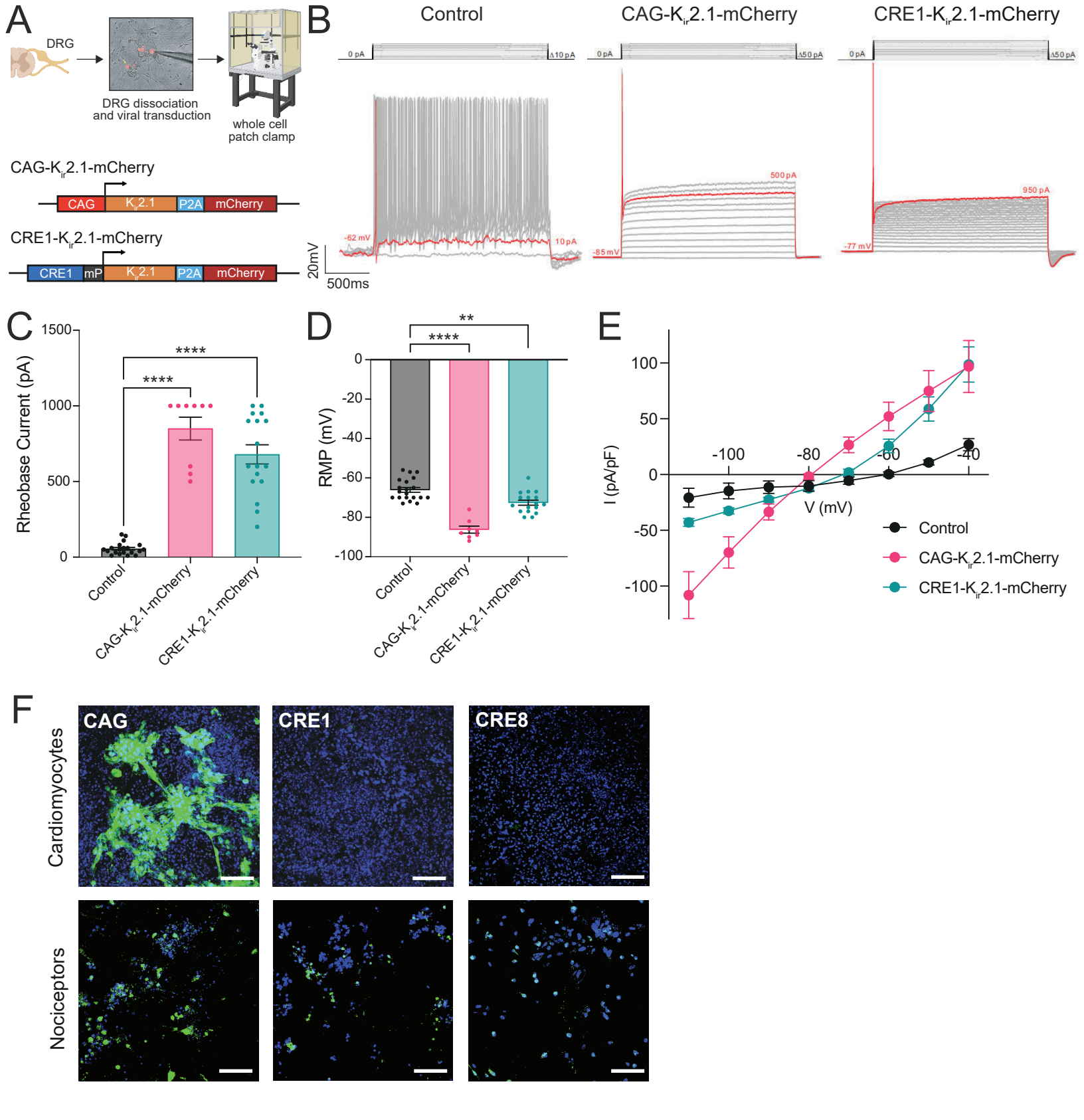
A
DRG
DRG dissociation and viral transduction
whole cell patch clamp
CAG-Kir2.1-mCherry
CAG
Kir2.1
P2A
mCherry
CRE1-Kir2.1-mCherry
CRE1
mP
B
Control
CAG-Kir2.1-mCherry
CRE1-Kir2.1-mCherry
0 pA
Δ10 pA
Δ50 pA
-62 mV
10 pA
500 pA
-85 mV
950 pA
-77 mV
20mV
500ms
C
Rheobase Current (pA)
1500
1000
500
0
****
****
Control
CAG-Kir2.1-mCherry
CRE1-Kir2.1-mCherry
D
RMP (mV)
0
-20
-40
-60
-80
-100
**
****
E
I (pA/pF)
100
50
0
-50
-100
-100
-80
-60
-40
V (mV)
Control
CAG-Kir2.1-mCherry
CRE1-Kir2.1-mCherry
F
CAG
CRE1
CRE8
Cardiomyocytes
Nociceptors

Figure 6

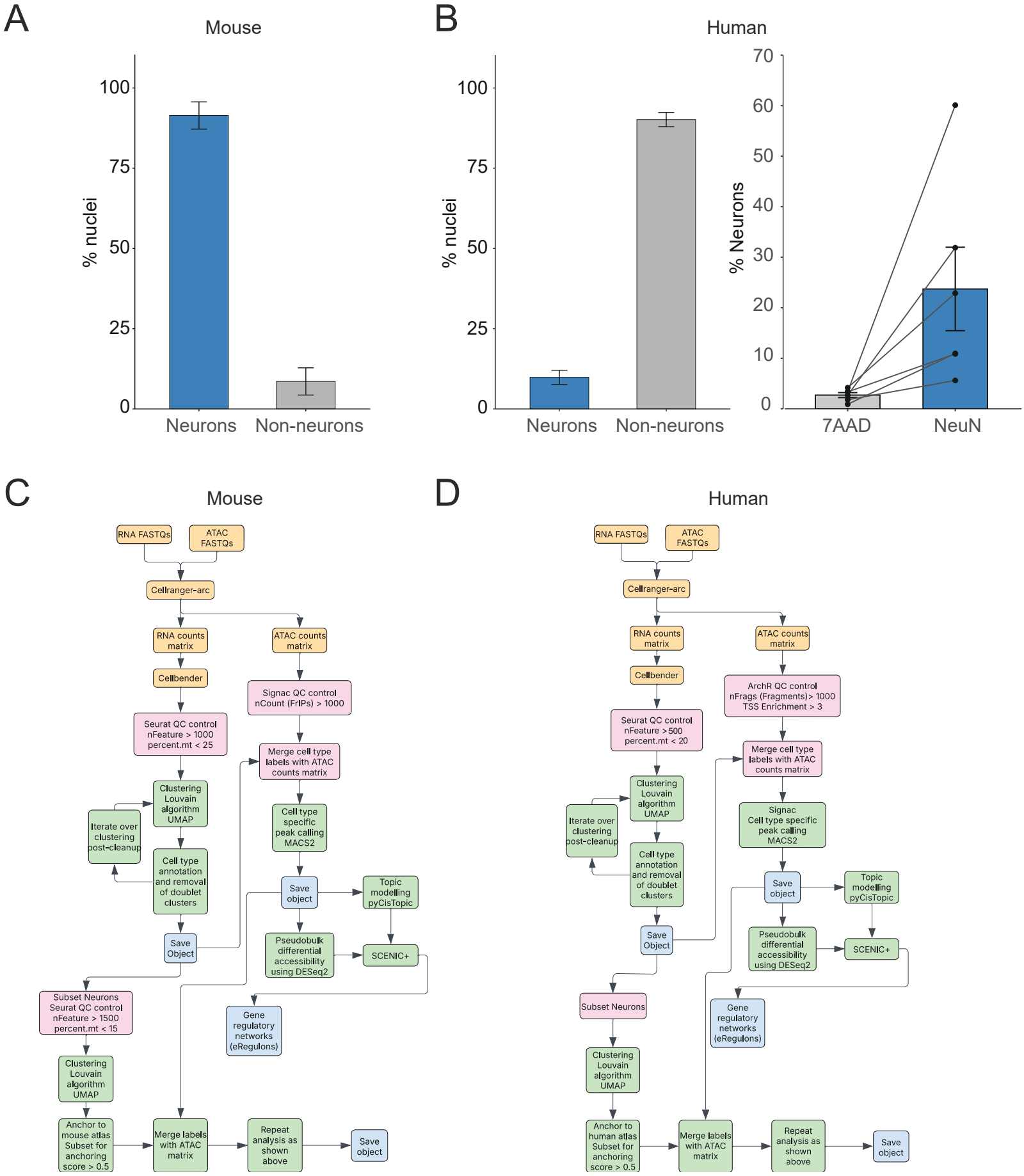
A
Mouse
% nuclei
100
75
50
25
0
Neurons
Non-neurons
B
Human
% nuclei
100
75
50
25
0
Neurons
Non-neurons
% Neurons
70
60
50
40
30
20
10
0
7AAD
NeuN
C
Mouse
RNA FASTQs
ATAC FASTQs
Cellranger-arc
RNA counts matrix
ATAC counts matrix
Cellbender
Signac QC control nCount (FrIPs) > 1000
Seurat QC control nFeature > 1000 percent.mt < 25
Merge cell type labels with ATAC counts matrix
Clustering Louvain algorithm UMAP
Iterate over clustering post-cleanup
Cell type specific peak calling MACS2
Cell type annotation and removal of doublet clusters
Save object
Topic modelling pyCisTopic
Save Object
Pseudobulk differential accessibility using DESeq2
SCENIC+
Subset Neurons Seurat QC control nFeature > 1500 percent.mt < 15
Gene regulatory networks (eRegulons)
Clustering Louvain algorithm UMAP
Anchor to mouse atlas Subset for anchoring score > 0.5
Merge labels with ATAC matrix
Repeat analysis as shown above
Save object
D
Human
RNA FASTQs
ATAC FASTQs
Cellranger-arc
RNA counts matrix
ATAC counts matrix
Cellbender
ArchR QC control nFrags (Fragments)> 1000 TSS Enrichment > 3
Seurat QC control nFeature >500 percent.mt < 20
Merge cell type labels with ATAC counts matrix
Clustering Louvain algorithm UMAP
Iterate over clustering post-cleanup
Signac Cell type specific peak calling MACS2
Cell type annotation and removal of doublet clusters
Save object
Topic modelling pyCisTopic
Save Object
Pseudobulk differential accessibility using DESeq2
SCENIC+
Subset Neurons
Gene regulatory networks (eRegulons)
Clustering Louvain algorithm UMAP
Anchor to human atlas Subset for anchoring score > 0.5
Merge labels with ATAC matrix
Repeat analysis as shown above
Save object

Figure S1

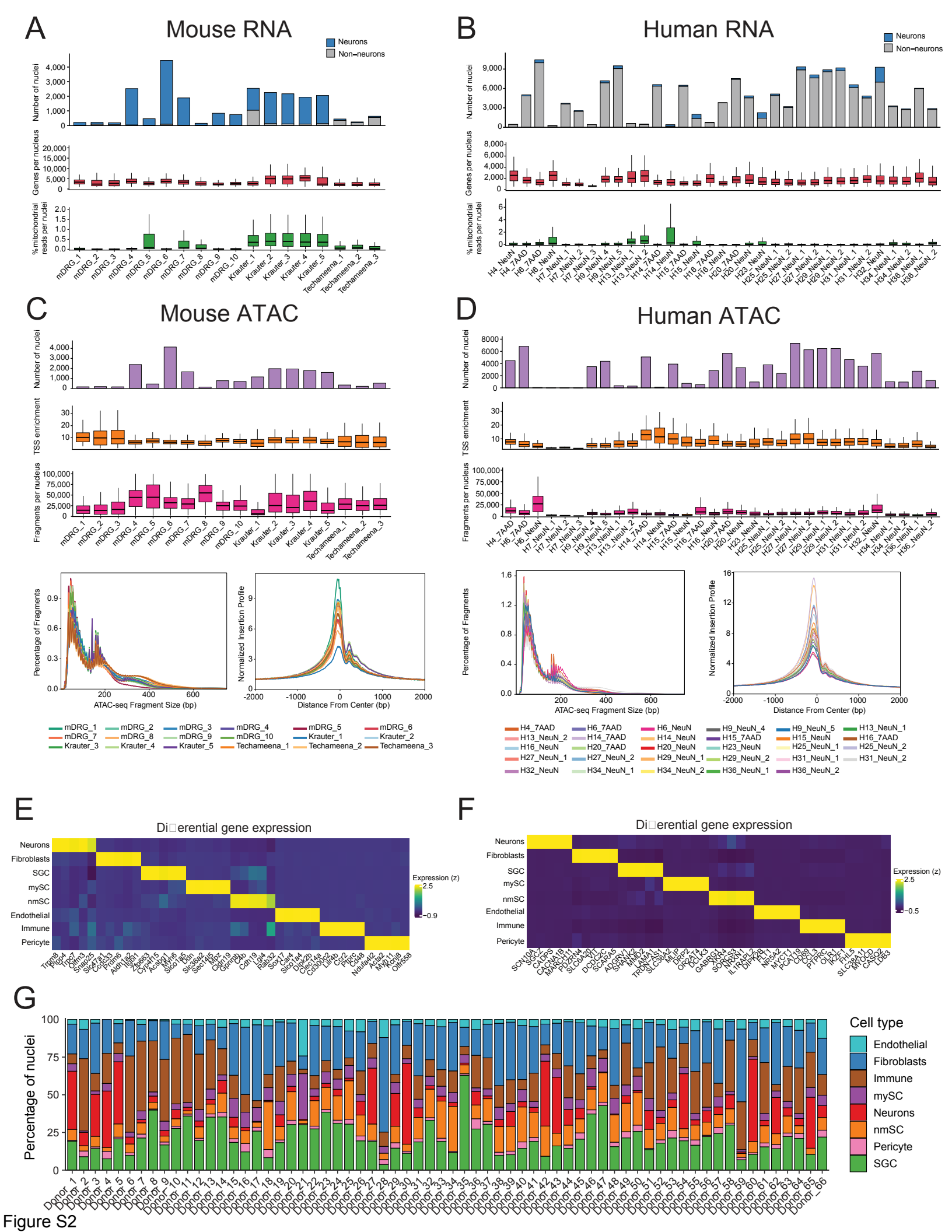
A
Mouse RNA
Neurons
Non−neurons
Number of nuclei
Genes per nucleus
% mitochondrial reads per nuclei
B
Human RNA
C
Mouse ATAC
Number of nuclei
TSS enrichment
Fragments per nucleus
D
Human ATAC
Percentage of Fragments
ATAC-seq Fragment Size (bp)
Normalized Insertion Profile
Distance From Center (bp)
E
Differential gene expression
Expression (z)
F
G
Percentage of nuclei
Cell type
Endothelial
Fibroblasts
Immune
mySC
Neurons
nmSC
Pericyte
SGC

Figure S2

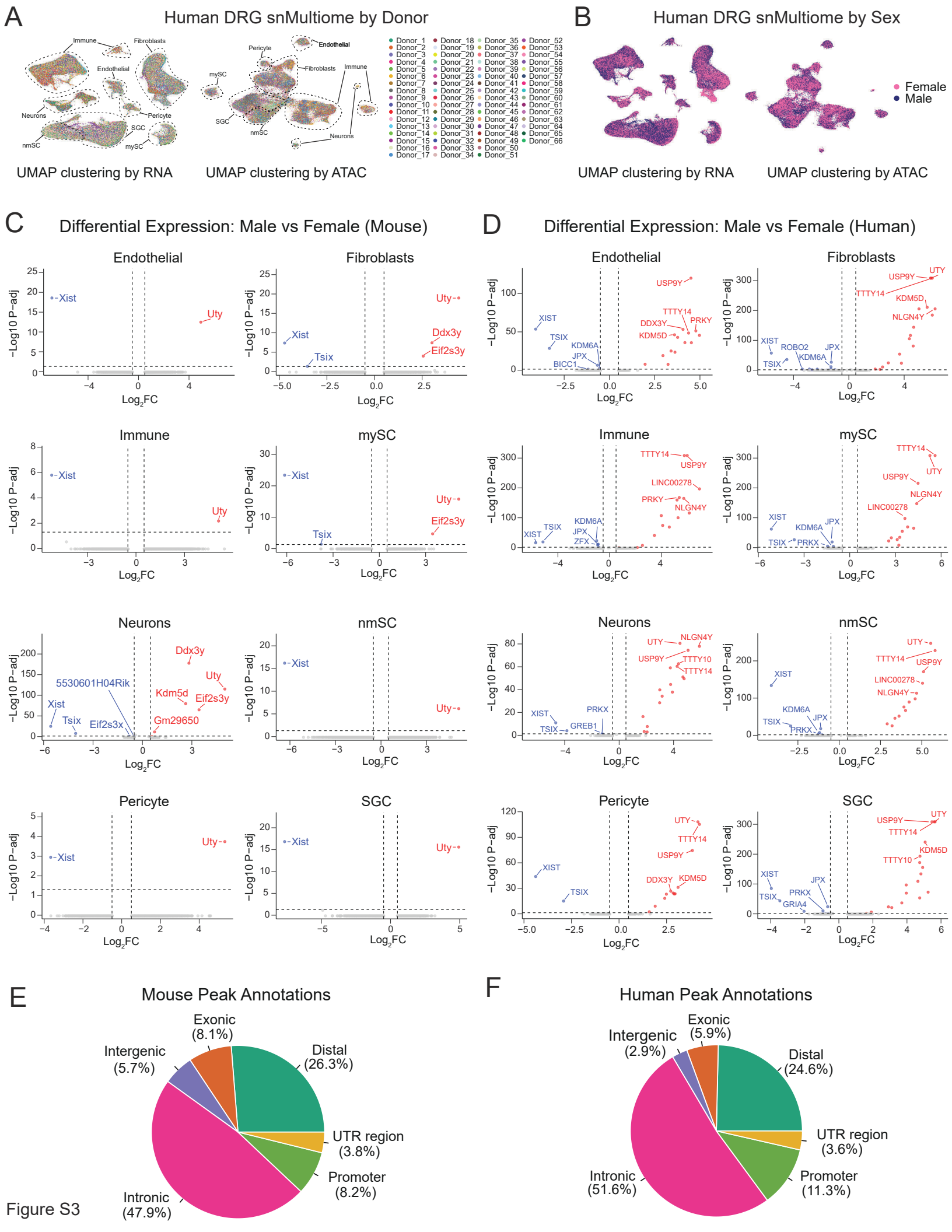
A
Human DRG snMultiome by Donor
Immune
Fibroblasts
Endothelial
Neurons
Pericyte
SGC
nmSC
mySC
UMAP clustering by RNA
UMAP clustering by ATAC
B
Human DRG snMultiome by Sex
Female
Male
UMAP clustering by RNA
UMAP clustering by ATAC
C
Differential Expression: Male vs Female (Mouse)
Endothelial
Fibroblasts
Immune
mySC
Neurons
nmSC
Pericyte
SGC
−Log10 P−adj
Log2FC
Xist
Uty
Tsix
Ddx3y
Eif2s3y
5530601H04Rik
Kdm5d
Gm29650
D
Differential Expression: Male vs Female (Human)
Endothelial
Fibroblasts
Immune
mySC
Neurons
nmSC
Pericyte
SGC
−Log10 P−adj
Log2FC
XIST
TSIX
KDM6A
JPX
BICC1
USP9Y
TTTY14
DDX3Y
KDM5D
PRKY
ROBO2
UTY
NLGN4Y
LINC00278
ZFX
PRKX
GREB1
TTTY10
GRIA4
E
Mouse Peak Annotations
Exonic (8.1%)
Distal (26.3%)
Intergenic (5.7%)
UTR region (3.8%)
Promoter (8.2%)
Intronic (47.9%)
F
Human Peak Annotations
Exonic (5.9%)
Intergenic (2.9%)
Distal (24.6%)
UTR region (3.6%)
Promoter (11.3%)
Intronic (51.6%)

Figure S3

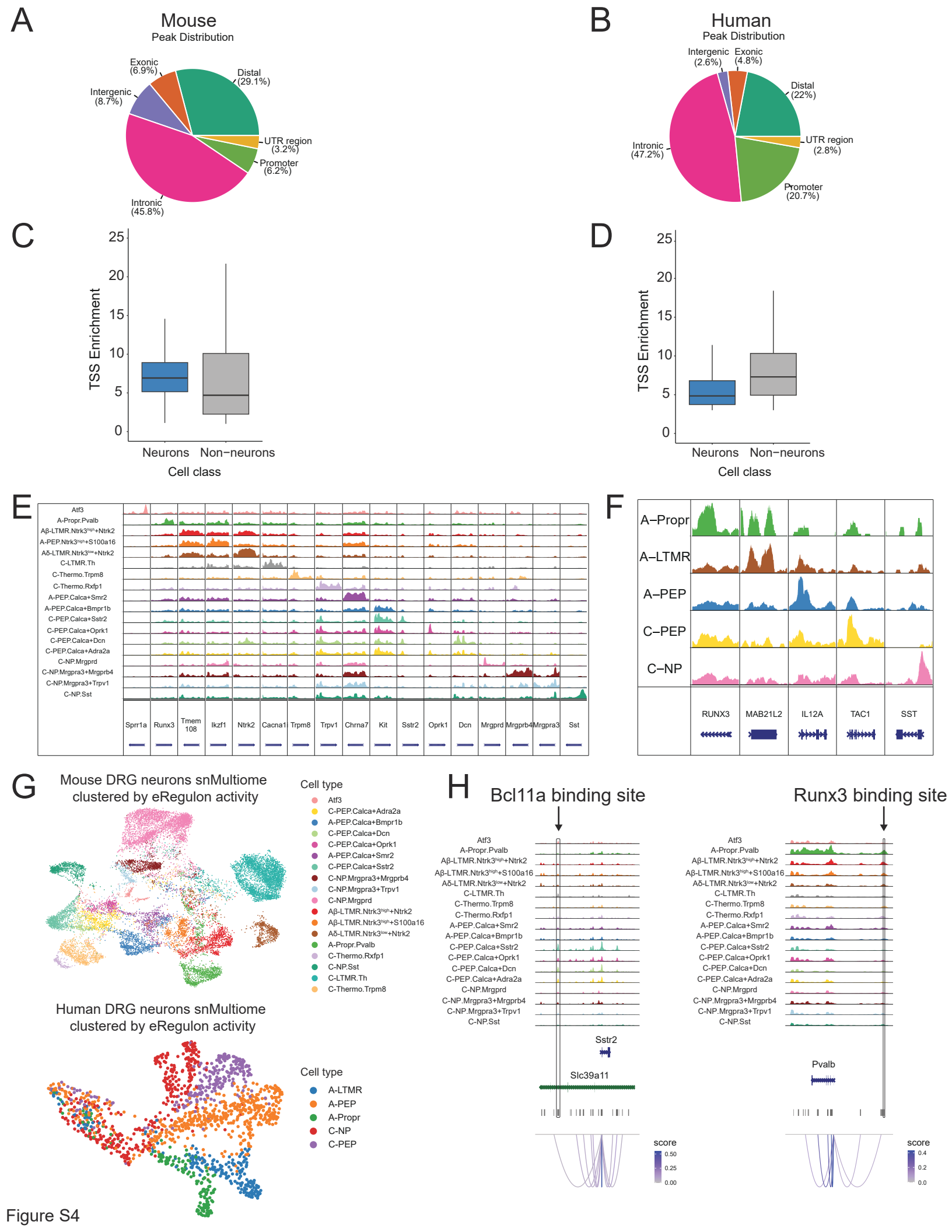
A
Mouse
Peak Distribution
Exonic (6.9%)
Distal (29.1%)
Intergenic (8.7%)
UTR region (3.2%)
Promoter (6.2%)
Intronic (45.8%)
B
Human
Peak Distribution
Intergenic (2.6%)
Exonic (4.8%)
Distal (22%)
Intronic (47.2%)
UTR region (2.8%)
Promoter (20.7%)
C
TSS Enrichment
Neurons
Non−neurons
Cell class
D
TSS Enrichment
Neurons
Non−neurons
Cell class
E
Atf3
A-Propr.Pvalb
Aβ-LTMR.Ntrk3high+Ntrk2
A-PEP.Ntrk3high+S100a16
Aδ-LTMR.Ntrk3low+Ntrk2
C-LTMR.Th
C-Thermo.Trpm8
C-Thermo.Rxfp1
A-PEP.Calca+Smr2
A-PEP.Calca+Bmpr1b
C-PEP.Calca+Sstr2
C-PEP.Calca+Oprk1
C-PEP.Calca+Dcn
C-PEP.Calca+Adra2a
C-NP.Mrgprd
C-NP.Mrgpra3+Mrgprb4
C-NP.Mrgpra3+Trpv1
C-NP.Sst
Sprr1a
Runx3
Tmem 108
Ikzf1
Ntrk2
Cacna1
Trpm8
Trpv1
Chrna7
Kit
Sstr2
Oprk1
Dcn
Mrgprd
Mrgprb4
Mrgpra3
Sst
F
A−Propr
A−LTMR
A−PEP
C−PEP
C−NP
RUNX3
MAB21L2
IL12A
TAC1
SST
G
Mouse DRG neurons snMultiome clustered by eRegulon activity
Cell type
Atf3
C-PEP.Calca+Adra2a
A-PEP.Calca+Bmpr1b
C-PEP.Calca+Dcn
C-PEP.Calca+Oprk1
A-PEP.Calca+Smr2
C-PEP.Calca+Sstr2
C-NP.Mrgpra3+Mrgprb4
C-NP.Mrgpra3+Trpv1
C-NP.Mrgprd
Aβ-LTMR.Ntrk3high+Ntrk2
Aβ-LTMR.Ntrk3high+S100a16
Aδ-LTMR.Ntrk3low+Ntrk2
A-Propr.Pvalb
C-Thermo.Rxfp1
C-NP.Sst
C-LTMR.Th
C-Thermo.Trpm8
Human DRG neurons snMultiome clustered by eRegulon activity
Cell type
A-LTMR
A-PEP
A-Propr
C-NP
C-PEP
H
Bcl11a binding site
Runx3 binding site
Sstr2
Slc39a11
Pvalb
score
0.50
0.25
0.00
score
0.4
0.2
0.0

Figure S4

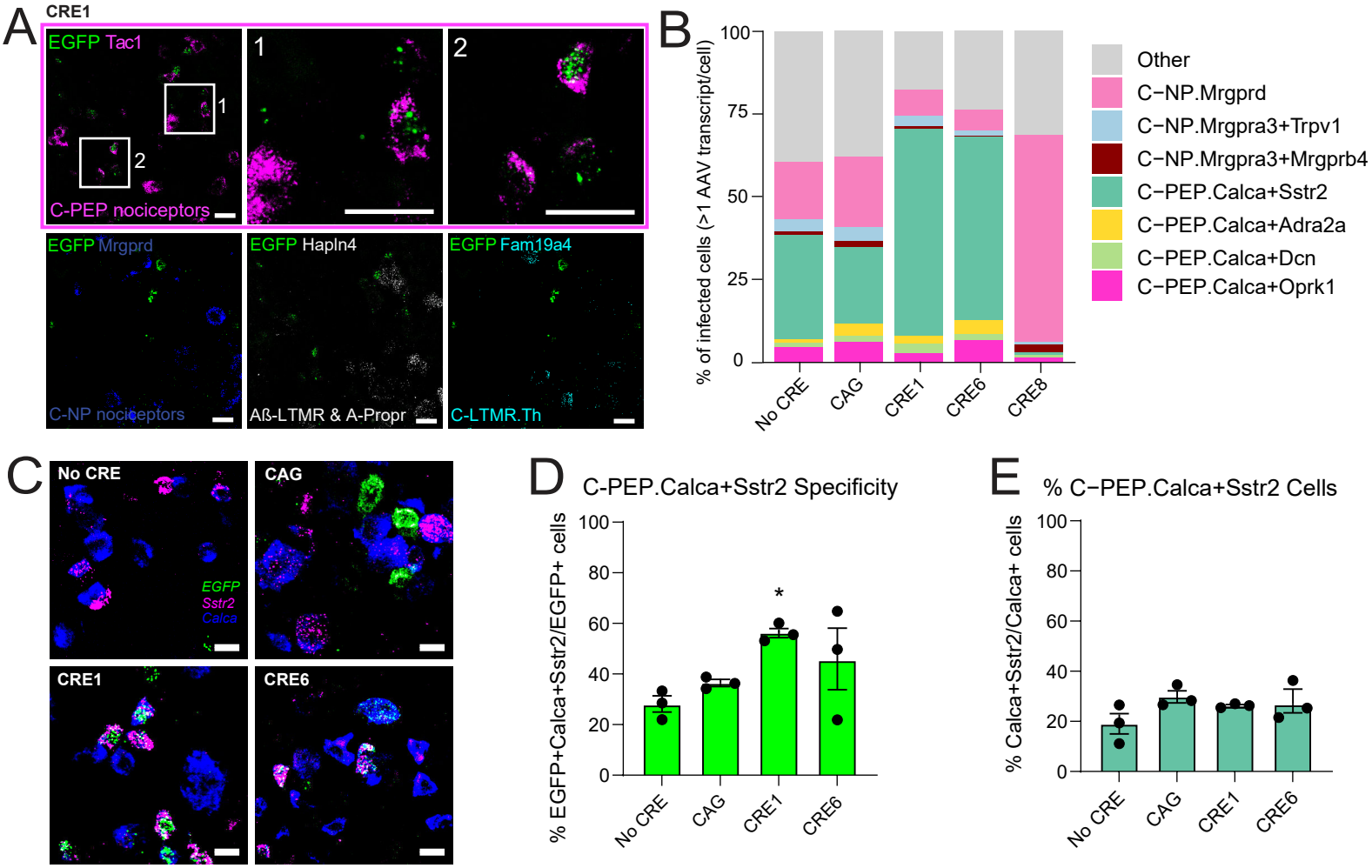
A
CRE1
EGFP Tac1
1
2
C-PEP nociceptors
EGFP Mrgprd
C-NP nociceptors
EGFP Hapln4
Aß-LTMR & A-Propr
EGFP Fam19a4
C-LTMR.Th
B
% of infected cells (>1 AAV transcript/cell)
100
75
50
25
0
No CRE
CAG
CRE1
CRE6
CRE8
Other
C−NP.Mrgprd
C−NP.Mrgpra3+Trpv1
C−NP.Mrgpra3+Mrgprb4
C−PEP.Calca+Sstr2
C−PEP.Calca+Adra2a
C−PEP.Calca+Dcn
C−PEP.Calca+Oprk1
C
No CRE
EGFP
Sstr2
Calca
CAG
CRE1
CRE6
D
C-PEP.Calca+Sstr2 Specificity
% EGFP+Calca+Sstr2/EGFP+ cells
100
80
60
40
20
0
*
No CRE
CAG
CRE1
CRE6
E
% C−PEP.Calca+Sstr2 Cells
% Calca+Sstr2/Calca+ cells
100
80
60
40
20
0
No CRE
CAG
CRE1
CRE6

Figure S5

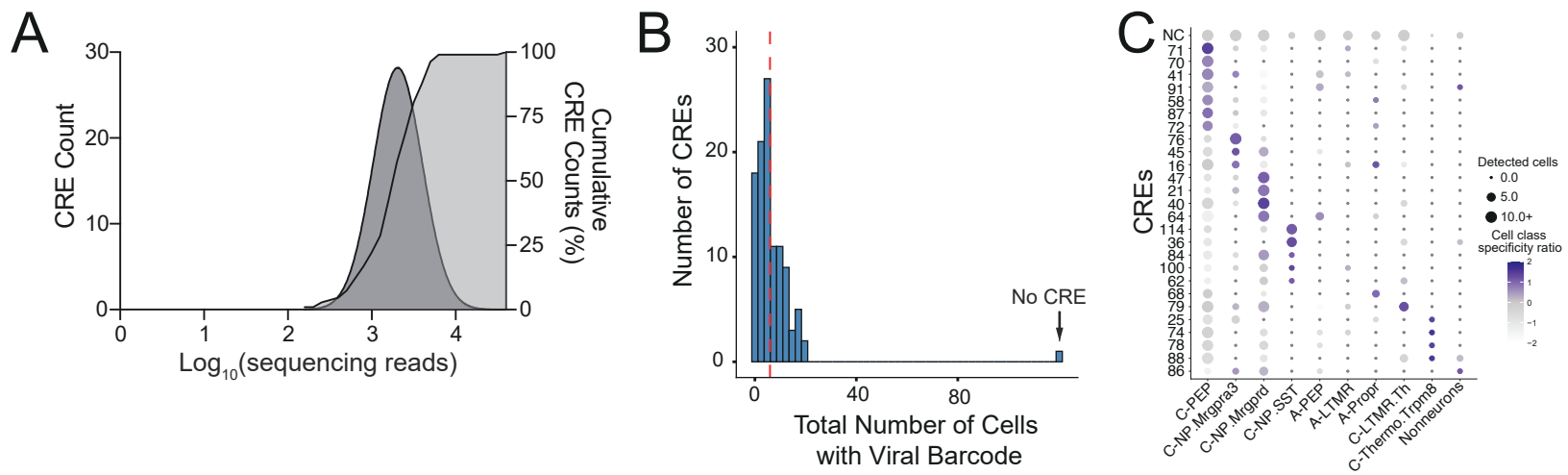
A
CRE Count
Cumulative CRE Counts (%)
Log10(sequencing reads)
B
Number of CREs
No CRE
Total Number of Cells with Viral Barcode
C
CREs
NC
C-PEP
C-NP Mrgpra3
C-NP Mrgprd
C-NP SST
A-PEP
A-LTMR
A-Propr
C-LTMR.Th
C-Thermo.Trpm8
Nonneurons
Detected cells
0.0
5.0
10.0+
Cell class specificity ratio

Figure S6

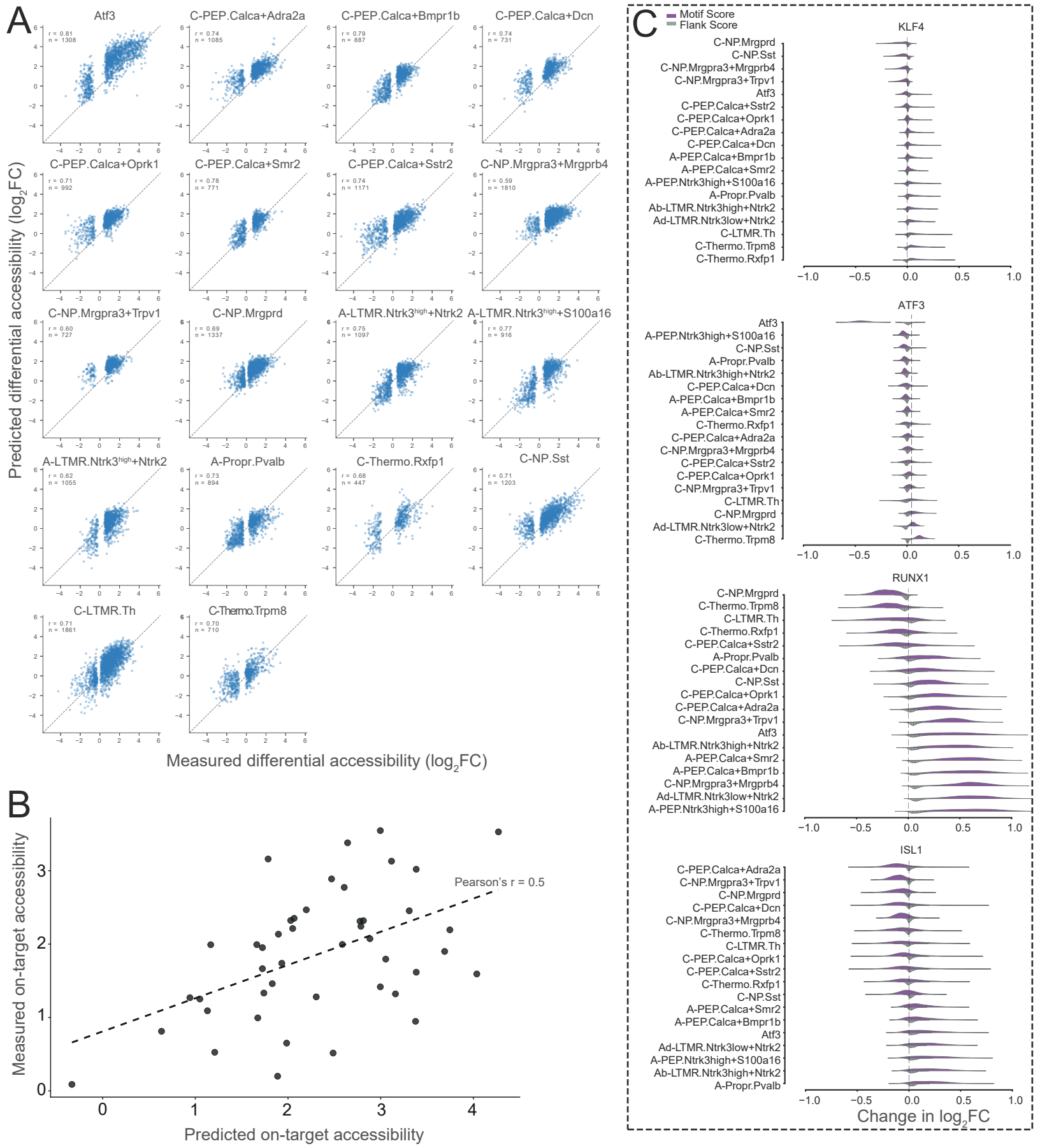
A
Atf3
r = 0.81
n = 1308
C-PEP.Calca+Adra2a
r = 0.74
n = 1085
C-PEP.Calca+Bmpr1b
r = 0.79
n = 887
C-PEP.Calca+Dcn
r = 0.74
n = 731
C-PEP.Calca+Oprk1
r = 0.71
n = 992
C-PEP.Calca+Smr2
r = 0.78
n = 771
C-PEP.Calca+Sstr2
r = 0.74
n = 1171
C-NP.Mrgpra3+Mrgprb4
r = 0.59
n = 1810
C-NP.Mrgpra3+Trpv1
r = 0.60
n = 727
C-NP.Mrgprd
r = 0.69
n = 1337
A-LTMR.Ntrk3high+Ntrk2
r = 0.75
n = 1097
A-LTMR.Ntrk3high+S100a16
r = 0.77
n = 916
A-LTMR.Ntrk3high+Ntrk2
r = 0.62
n = 1055
A-Propr.Pvalb
r = 0.73
n = 894
C-Thermo.Rxfp1
r = 0.68
n = 447
C-NP.Sst
r = 0.71
n = 1203
C-LTMR.Th
r = 0.71
n = 1861
C-Thermo.Trpm8
r = 0.70
n = 710
Predicted differential accessibility (log2FC)
Measured differential accessibility (log2FC)
B
Pearson's r = 0.5
Measured on-target accessibility
Predicted on-target accessibility
C
Motif Score
Flank Score
KLF4
C-NP.Mrgprd
C-NP.Sst
C-NP.Mrgpra3+Mrgprb4
C-NP.Mrgpra3+Trpv1
Atf3
C-PEP.Calca+Sstr2
C-PEP.Calca+Oprk1
C-PEP.Calca+Adra2a
C-PEP.Calca+Dcn
A-PEP.Calca+Bmpr1b
A-PEP.Calca+Smr2
A-PEP.Ntrk3high+S100a16
A-Propr.Pvalb
Ab-LTMR.Ntrk3high+Ntrk2
Ad-LTMR.Ntrk3low+Ntrk2
C-LTMR.Th
C-Thermo.Trpm8
C-Thermo.Rxfp1
ATF3
Atf3
A-PEP.Ntrk3high+S100a16
C-NP.Sst
A-Propr.Pvalb
Ab-LTMR.Ntrk3high+Ntrk2
C-PEP.Calca+Dcn
A-PEP.Calca+Bmpr1b
A-PEP.Calca+Smr2
C-Thermo.Rxfp1
C-PEP.Calca+Adra2a
C-NP.Mrgpra3+Mrgprb4
C-PEP.Calca+Sstr2
C-PEP.Calca+Oprk1
C-NP.Mrgpra3+Trpv1
C-LTMR.Th
C-NP.Mrgprd
Ad-LTMR.Ntrk3low+Ntrk2
C-Thermo.Trpm8
RUNX1
C-NP.Mrgprd
C-Thermo.Trpm8
C-LTMR.Th
C-Thermo.Rxfp1
C-PEP.Calca+Sstr2
A-Propr.Pvalb
C-PEP.Calca+Dcn
C-NP.Sst
C-PEP.Calca+Oprk1
C-PEP.Calca+Adra2a
C-NP.Mrgpra3+Trpv1
Atf3
Ab-LTMR.Ntrk3high+Ntrk2
A-PEP.Calca+Smr2
A-PEP.Calca+Bmpr1b
C-NP.Mrgpra3+Mrgprb4
Ad-LTMR.Ntrk3low+Ntrk2
A-PEP.Ntrk3high+S100a16
ISL1
C-PEP.Calca+Adra2a
C-NP.Mrgpra3+Trpv1
C-NP.Mrgprd
C-PEP.Calca+Dcn
C-NP.Mrgpra3+Mrgprb4
C-Thermo.Trpm8
C-LTMR.Th
C-PEP.Calca+Oprk1
C-PEP.Calca+Sstr2
C-Thermo.Rxfp1
C-NP.Sst
A-PEP.Calca+Smr2
A-PEP.Calca+Bmpr1b
Atf3
Ad-LTMR.Ntrk3low+Ntrk2
A-PEP.Ntrk3high+S100a16
Ab-LTMR.Ntrk3high+Ntrk2
A-Propr.Pvalb
Change in log2FC

Figure S7

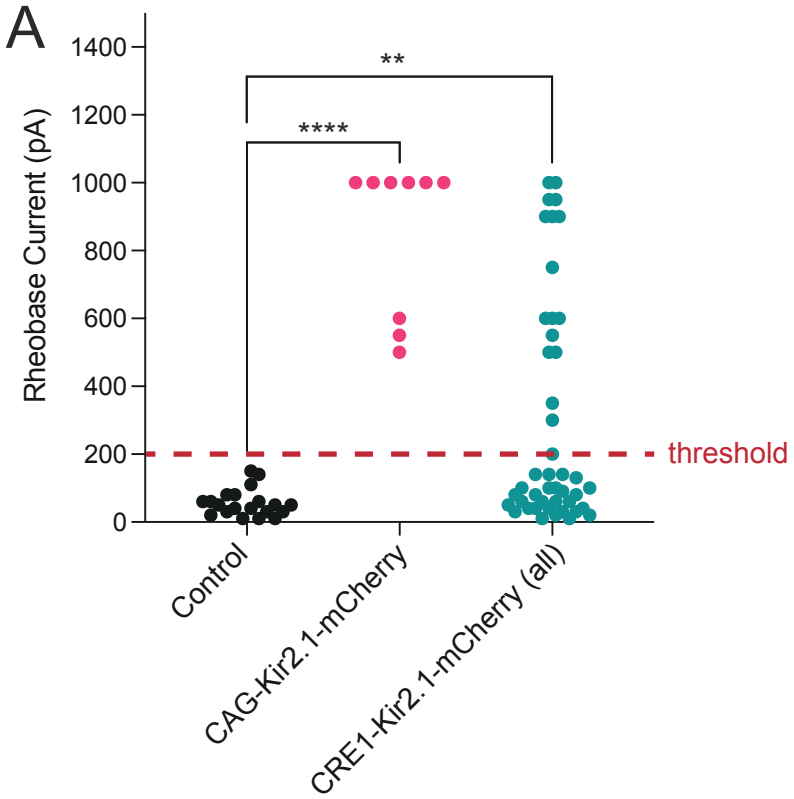
A
Rheobase Current (pA)
1400
1200
1000
800
600
400
200
0
**
****
threshold
Control
CAG-Kir2.1-mCherry
CRE1-Kir2.1-mCherry (all)

Figure S8