Investigating the basis of lineage decisions and developmental trajectories in the dorsal spinal cord through pseudotime analyses

Sandeep Gupta[1+], Eric Heinrichs[1,2+], Bennett G. Novitch[1, 3, 4], and Samantha J. Butler[1, 3, 4*]

1. Department of Neurobiology, David Geffen School of Medicine
2. Genetics and Genomics Graduate program, University of California Los Angeles, USA
3. Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research
4. Intellectual & Developmental Disabilities Research Center

University of California, Los Angeles
Los Angeles, CA 90095

[+] Equal contributions
* Correspondence: butlersj@ucla.edu
310 206 8416
CHS South Tower 67-200J, 650 Charles E Young Drive East,
University of California, Los Angeles,
Los Angeles CA 90095

Running title: Spatial and temporal atlas of stem cell-derived spinal sensory interneurons

Key words: dorsal cord, cell fate, stem cells, sensory interneurons, single cell RNA-Seq analysis, pseudotime

Summary statement: Pseudotime analyses of embryonic stem cell-derived dorsal spinal interneurons reveals both novel regulators and lineage relationships between different interneuron populations.

Abstract

Dorsal interneurons (dIs) in the spinal cord encode the perception of touch, pain, heat, itch, and proprioception. While previous studies using genetic strategies in animal models have revealed important insights into dI development, the molecular details by which dIs arise as distinct populations of neurons remain incomplete. We have developed a resource to investigate dI fate specification, by combining a single-cell RNA-Seq atlas of mouse ESC-derived dIs with pseudotime analyses. To validate this *in silico* resource as a useful tool, we have used it to first identify novel genes that are candidates for directing the transition states that lead to distinct dI lineage trajectories, and then validated them using *in situ* hybridization analyses in the developing mouse spinal cord *in vivo*. We have also identified a novel endpoint of the dI5 lineage trajectory, and found that dIs become more transcriptionally homogenous during terminal differentiation. Together, this study introduces a valuable tool for further discovery about the timing of gene expression during dI differentiation and uses it to clarify dI lineage relationships.

Introduction

Somatosensation permits us to perceive touch, temperature, pain (nociception), and hold our bodies correctly in space (proprioception). These sensory modalities are critical for daily life, as well as emotional well-being. Sensory information is received in the periphery, and transmitted to higher-order centers in the brain, or spinal motor circuits, by sensory relay circuits in the dorsal spinal cord (Lai et al., 2016). These circuits arise from six populations of dorsal interneurons (dI1-dI6) with distinct molecular signatures, connectivity, and sensory functions (Andrews et al., 2017; Gupta and Butler, 2021). dIs emerge during embryonic development in response to multiple patterning and differentiation signals. Previous studies have shown that the bone morphogenetic protein (BMPs) and Wnt families act from the roof plate at the dorsal midline to direct the dorsal-most dI identities (dI1-dI3) (Andrews et al., 2017; Gupta et al., 2022; Hazen et al., 2012; Le Dreau et al., 2012; Lee et al., 1998; Liem et al., 1997; Liem et al., 1995; Megason and McMahon, 2002; Muroyama et al., 2002; Yamauchi et al., 2008). The fate specification process for the intermediate dI identities (dI4-dI6) is less well defined, but a recent study showed that retinoic acid (RA) is sufficient to direct these fates *in vitro* (Gupta et al., 2022), suggesting a role for RA acting from the paraxial mesoderm *in vivo*. On differentiating, dIs then migrate to the correct laminae in the adult dorsal horn to form the different sensory circuits (Koch et al., 2018).

The genetic program that directs dI differentiation remains unresolved. While a number of transcription factors have been identified that are critical for dorsal progenitor (dP) and dI identity (Lai et al., 2016), it nonetheless has remained unclear how multiple growth factors act to specify six distinct dI populations. We are assessing these mechanisms by developing embryonic stem cell (ESC) models (Andrews et al., 2017; Gupta et al., 2018; Gupta et al., 2021); most recently, we have described an improved protocol which can generate the complete complement of dIs with the correct functional and molecular signatures (Gupta et al., 2022). Stem cell models offer many advantages for mechanistic discovery including an unparalleled ability to control growth conditions and probe cellular/molecular responses in large populations of synchronously developing cells without the confounding effects of embryonic redundancy and lethality (Gaspard and Vanderhaeghen, 2010; Veenvliet et al., 2021; Zhu and Huangfu, 2013). Our studies using these models have suggested that BMPs do not establish dI fate by acting as

morphogens (Andrews et al., 2017), rather dI fates appear to be established in a series of nested choice points (Gupta et al., 2022). Spinal progenitors are initially dorsalized by RA, subdivided into multipotential dP subgroups by RA±BMP signaling, and then directed into specific dI fates by as yet unknown mechanisms.

Here, we leverage our previously acquired single-cell (sc) RNA-Seq atlas of mESC-derived dIs (Gupta et al., 2022), to develop a tool to identify novel genes that direct dI fate specification. This tool combines the scRNA-Seq atlas with pseudotime analyses to reconstruct dI-specific lineage trajectories and thereby identify the transitional states before the key choice points. *In silico*-identified candidate genes for these transitional states were then validated *in vivo* using an expression analysis in the developing mouse spinal cord. These studies have investigated the endpoint of the dI5 lineage trajectory, and the emergence of distinct dI5 subtypes. We have also observed that the different dI fates converge upon terminal differentiation, i.e., assume neuronal identities which are more transcriptionally similar than during their proceeding developmental trajectories. Taken together, this analysis provides a comprehensive understanding of dI lineage relationships and develops a resource for identifying novel developmental regulators of sensory circuit formation.

Material and Methods

Seurat data processing and integration

Cellranger output (Gupta et al., 2022) was loaded into R (4.1.3) (R Core Team, 2022) and Seurat (Hao et al., 2021) (v4.0.4 - v4.1) and separate objects were made for the RA and RA+BMP protocols with min.cells = 3 and min.features = 200. Both datasets were then filtered for quality control based on violin plots of metadata with the goal of removing outliers on both ends (RA: nFeature > 2500, nCount_RNA > 5000 and < 50000, percent.mito < 10; BMP: nFeature > 200, nCount_RNA > 5000 and < 35000, percent.mito < 7). SCTransform (V1) was run on both datasets individually. Standard dimensional reduction was followed with 40pcs, 3 UMAP dimensions, and default cluster resolution. To isolate cell types relevant to the desired differentiation and to remove unwanted byproducts and low-quality cells, the data were then subsetted to include only *Sox2*$^{+}$ or *Tubb3*$^{+}$ clusters that were also *Nanog*$^{-}$ (pluripotent stem cells) and *Sox10*$^{-}$ (neural crest cells), to remove cell types that were unrelated to the differentiation (Fig. S1A, B), and reprocessed with SCTransform pipeline. The two datasets were next integrated in Seurat (v4.2) using 3000 integration features and reciprocal principal component analysis (RPCA) integration based on PCs calculated from commonly varying genes was performed (rather than canonical correlation analysis to avoid overfitting). The combined data were dimensionally reduced and embedded into 3 UMAP dimensions using 40pcs. Clustering was performed using a resolution of 2 to obtain clusters that roughly correlated with differentiation trajectory and timepoint in differentiation. PrepSCTFindMarkers was run to correct counts from different datasets to aid in further expression analysis. All plots were generated using ggplot2 (Wickham, 2016), Plotly (Fahd Qadir, 2019; Qadir et al., 2020; Sievert, 2020), and Graphpad Prism.

Monocle pseudotemporal ordering

To find pseudotime trajectories in our combined dataset, we transferred our data to Monocle3 (Cao et al., 2019) using SeuratWrappers, and the cells were clustered using the UMAP reduction and Learn_Graph was run to ascertain the principal graph of the data. The parameters were optimized to close the overall loop of the dataset and provide sufficient branching without yielding erroneous branches (use partition = F, learn_graph_control: Euclidian distance ratio = 2,

geodesic distance ratio = 1/5, minimal branch length = 10, orthoganol project tip = F, n_center = 340, prune graph = T). Choose_cells was used to select all cells in the progenitor area up to the initial bottleneck and these were all set as the root when running order_cells. These data were then added back into Seurat as a metadata column to be accessed during further analysis.

Pseudotime and marker gene analysis

Marker gene analysis was done with Seurat v4.3. FindAllMarkers (only.pos = T) was run on the combined dataset to identify marker genes for each cluster. The Seurat object was split into 5 sub-objects, one for each of the 5 trajectories, by subsetting on clusters. To analyze expression over pseudotime, Locally Estimated Scatterplot Smoothing with a span of 0.3 was used to create a curve that fit the data and predict gene expression at every 0.1 pseudotime value. These values were then plotted in ggplot2 using geom_tile.

Analysis of the different pseudotime timepoints was achieved by splitting the data into progenitor (the cells used as the root in Monocle), dP (cells less than 7.5 Pseudotime), or dI (cells greater than or equal to 7.5 Pseudotime) groupings. The cutoff between dP and dI was picked based on when expression of most dP markers peaked, as well as where secondary splits in the trajectories occured, and UMAP expression plots. FindAllMarkers was run and any positive marker gene with an adjusted p value < 0.05 was submitted to Metascape (Zhou et al., 2019) for gene ontology (GO) and other analyses. Similarly, *Sncg* positive cells were analyzed by taking any cells in clusters 16 or 7 with a *Sncg* expression value greater than 1 (and a Pseudotime > 0.3 due to a clustering issue) and running FindMarkers against the remaining cells. Metascape-based plots were made in Prism.

In situ hybridization and immunohistochemistry

Digioxigenin (DIG)-labeled RNA probes against the 3′ untranslated regions of genes of interest were generated using the Roche RNA Labeling Kit and hybridized onto 12-14μm transverse sections of embryonic spinal cords. *In situ* hybridization signals were visualized using anti-DIG antibody conjugated with an alkaline phosphatase fragment (Roche) and nitro-blue tetrazolium and 5-bromo-4-chloro-3'-indolyphosphate substrates. Target sequences were amplified using cDNA derived from the mouse embryonic stem cell (mESC)-derived spinal cord cell types using

the primers listed in Table 1. All primers were designed with the Primer 3 program (http://primer3plus.com/) and T7 promoter sequence was added on all the reverse primers for generating antisense mRNA probes using T7 RNA polymerase (Roche). For immunohistochemistry (IHC) followed by ISH, spinal cord sections were directly treated with 1% antibody blocking solution (1% heat inactivated horse serum in 1X PBST) for 1 hour at room temperature followed by incubation with the primary antibodies overnight at 4°C. Fluorescently labelled species-specific secondary antibodies (Jackson Immunoresearch Labs) were used to detect the signal. The following primary antibodies were used: Pax2 (Rabbit, 1:500, Invitrogen, catalog number: 71-6000), Lmx1b (Guinea pig, 1:100, gift from Thomas Mueller, Dresden), Lhx2 (goat, 1:250, Santa Cruz Biotechnology, catalog number: sc-19344). Sections were then counterstained with DAPI and imaged on a Zeiss LSM800 confocal system.

Results

Creation of a single cell atlas with the full repertoire of dorsal progenitors, transition states, and dorsal interneurons

To develop a resource to identify novel genes that direct neural fate specification and differentiation in the dorsal spinal cord, we mined our previously established single cell (sc) transcriptomic dataset that represents a complete atlas of *in vitro*-derived dorsal interneurons (dI) (Gupta et al., 2022). In brief, mESCs were converted into posterior neuromesodermal progenitors (NMPs) though the addition of basic FGF and the GSK3β antagonist CHIR 99021 (CHIR) (Gouti et al., 2014). These NMPs were then differentiated into dIs through the addition of either retinoic acid (RA) or RA together with bone morphogenic protein (BMP) 4 (Fig. 1A). These two protocols respectively generate either dorsal progenitors (dP) 4 - dP6 or dP1-dP3, which then give rise to mature dI4-dI6s (RA protocol) or dI1-dI3s (RA+BMP4 protocol). These heterogenous cell populations were collected at day 9 of differentiation and processed for scRNA-Seq (Fig. 1A). Downstream analyses were performed to first compile an *in vitro*-derived single cell atlas for the dIs (Gupta et al., 2022), and then perform pseudotemporal ordering, to identify candidate genes that direct fate changes at transition points (Fig. 1B, see also methods).

Projection of both datasets using Seurat (Hao et al., 2021) into the same three-dimensional UMAP space reveals the overlap and divergence in the cell types arising from the RA (red) and RA+BMP4 (blue) protocols (Fig. 1C). The datasets generally overlap in the *Sox2*$^+$ progenitor pool, and diverge after a bottleneck point when they branch into dI-specific trajectories (Fig 1F: dI1: *Lhx9*$^+$/*Barhl2*$^+$; dI2: *Foxd3*$^+$; dI3: *Isl1*$^+$; dI4: *Pax2*$^+$; dI5: *Lmx1b*$^+$; dI6: *Dmrt3*$^+$). The dI1 and dI5 lineages immediately emerge as distinct trajectories, while dI2, dI3, and dI4 initially share a common progenitor lineage before branching (Fig. 1D, F). We did not observe a distinct trajectory for dI6, rather it arises between the endpoints of the dI2 and dI4 lineages (Fig. 1D).

Unsupervised clustering of the dataset yields 31 clusters, which further subdivide the progenitor and dI lineages (Fig. 1E, G). The progenitor domain clusters are enriched for genes regulating the cell cycle, including S-phase (*Pcna*, *Mcm4*, *Gmnn*) (Komamura-Kohno et al., 2006; Kushwaha et al., 2016; Zerjatke et al., 2017), G2/M-phase (*Cdc20*, *Aurka*) (Cazales et al., 2005; Lara-Gonzalez et al., 2019) and G1-phase (*Ccnd1*) (Wang et al., 2018) markers, as well as roof

plate markers (*Msx1*) (Liu et al., 2004) (Fig. 1G). Clusters 25 and 11 span the transition from pan-dP to dI2/dI3/dI4 identities. Cluster 25 is enriched for broadly expressed dP markers including *Neurog1/2*, *Pax3* and *Olig3*, while cluster 11 shows expression of both dP markers and dI2/dI4 markers, such as *Lhx1/5* and *Pou4f1* (previously known as *Brn3a*) (Alaynick et al., 2011). Similarly, cluster 27 represents the first transition step of dPs towards the dI5 identity, expressing both dP5 markers, such as *Ccnd1* and *Ascl1*, and post-mitotic dI5 markers, like *Tlx3* and *Lbx1*. We see little to no expression of ventral markers, such as *Pitx2*, *En1*, and *Gata2*, in this dataset, confirming the specificity of the RA±BMP4 differentiation protocol (Fig. 1G).

Pseudotime analysis identifies new transition state-specific markers for the dIs

To investigate the temporal changes in gene expression in the five differentiation trajectories (Fig. 1D), we performed pseudotemporal ordering using Monocle3 (Cao et al., 2019), with distance calculated based on the Sox2^{+} progenitor population as the starting root. The pseudotime values were then superimposed onto the UMAP atlas, to reveal the distance and trajectories over which the progenitors differentiate into post-mitotic neurons (Fig. 2A). After optimization (see methods), the monocle trajectories (Fig. 2A, Supplemental movie 1) reveal similar bifurcation points to the cluster-based trajectory assignments (Fig. 2B). First, the dI5 branch splits off from dI1-dI4 lineages, rapidly followed by the dI1 branch splitting from the dI2-dI4 lineages. The dI2/dI3/dI4 lineages continue on a common path until they bifurcate, first to yield dI4 vs. dI2/dI3, and then dI2 vs. dI3 (Fig. 2A, B). This data suggests a shared lineage relationship between the dI2/dI3 and dI4 populations. Interestingly, we also observed that the monocle and lineage trajectories converge upon terminal differentiation (Fig 1C, Fig 2A, B, Supplemental movie 1). After branching to become distinct developmental trajectories, the dI2/dI3/dI4/dI5 lineages then merge as they differentiate as if they have again become more transcriptionally similar to each other. While the dI1 trajectory remains distinct, it also curves towards the same region of statistical similarity occupied by the other terminal branches.

To follow the gene expression changes over pseudotime, we first examined the temporal distribution of known marker genes. In every case, we find that the canonical dP and dI markers have the correct lineage-specific and temporal expression patterns (Fig. 2D). Thus, dP markers start to be expressed prior to pseudotime values of 5, while the expression of dI markers tends to

peak near or after pseudotime values of 10, suggesting that pseudotime distance accurately reflects developmental time *in vivo*. We next subdivided the data into progenitor, dP, and dI populations based on their pseudotime values and trajectory divergence points (Fig. 2C), to identify genes expressed in the different differentiation states, and performed gene ontology (GO) analyses for enriched terms (Fig 2F). While the progenitors and dIs showed the predicted enrichment of cell cycle related terms, and synapse formation related terms respectively, the dPs were enriched for terms related to neural differentiation, patterning and cell fate, further supporting the conclusion that dP clusters (between pseudotime value >0-7.5) represent transitory states (Fig. 2F).

To identify candidate genes that establish these transitory states, we performed differential gene expression on all clusters. We then compared the top marker genes from the first cluster in the dI1 (clusters 15) vs. dI2/dI3/dI3 (cluster 25) vs. dI5 (cluster 27) lineages after removing common markers (Fig. 2E). We then used UMAP expression plots of each of these genes to identify genes with the most interesting expression patterns. This analysis (Figs. 2D and S2) identified both canonical transcription factors known to be important for establishing dP fates, including *Atoh1* (dP1, (Helms and Johnson, 1998)), *Neurog1* (dP2), (Gowan et al., 2001) and *Neurog2* (dP2-dP5) (Sommer et al., 1996) validating our methodology, many novel genes whose expression patterns were then assessed *in vivo*.

Validation of putative transition state markers in vivo

Nine genes showing enriched expression in the dP transition clusters were selected for further analysis. The expression of these genes was examined in transverse sections of E10.5 and E11.5 spinal cord in an *in situ* hybridization analysis (Fig. 3). Of these nine genes, one - *Gsg1l*, a known Atoh1 target gene (Lai et al., 2011), which encodes a regulatory subunit of AMPA receptor (Kamalova et al., 2021) – shows specific expression in the dI1 lineage (arrows, Fig. 3A) as predicted by the pseudotime analysis (heatmap, Fig. 3A). A further two genes - *Chrna3*, which encodes a cholinergic receptor subunit (Flora et al., 2013), and *Cbfa2t2*, a transcriptional corepressor (Tu et al., 2016) – are initially broadly expressed by differentiating neurons in the E10.5 dorsal spinal cord (Fig. 3B, C). By E11.5, the expression of both genes becomes more prominent in the dP1/dI1 lineage and the intermediate zone (IZ), the region where differentiating

dPs exit the ventricular zone (VZ) as they migrate laterally to become postmitotic neurons (Fig. 3B). Again, this distribution strikingly mirrors the predictions from the pseudotime analyses (heatmaps, Fig. 3B, C), especially for *Cbfa2t2,* which is expressed sequentially first in dP5 and then dI1, with lower expression in the rest of the dorsal IZ (arrows, Fig. 3C).

Four of the selected genes – *Fbxl7, Tfap2b, Prmt8,* and *Sstr2* – are predicted to be present in subsets of dPs based on the pseudotime analysis. Three of these genes - *Tfap2b,* an AP2 family transcription factor (Zainolabidin et al., 2017), *Prmt8,* an arginine methyltransferase (Dong et al., 2021), and *Sstr2*, somatostatin receptor 2 (Stumm et al., 2004) – show expression first in the VZ (E10.5) followed by robust increases in the IZ (E11.5) *in vivo* (arrows, Fig. 3E, F, G). The fourth gene - *Fbxl7,* part of the ubiquitin ligase complex, is expressed in the intermediate VZ, but not upregulated in the IZ (bracket, Fig. 3D) as predicted *in silico*. These genes are expressed in multiple dI lineages. For example, *Tfap2b,* is upregulated in the IZ specifically in the dI2-dI5 lineages in both the *in vitr*o-derived atlas and in the E11.5 spinal cord (Fig. 3E). The remaining two genes - *Smoc1*, a secreted calcium-binding protein (Thomas et al., 2017), and *Gsx1*, a previously idenified spinal cord transcription factor (Mizuguchi et al., 2006) - were also validated from the *in silico* data. Both genes were predicted to be expressed at the beginning of the dP5 trajectory, which was borne out in the *in viv*o analysis (arrows, Fig. 3H, I). In particular, *Smoc1* is specifically expressed in the dP5 domain in the VZ of both E10.5 and E11.5 spinal cord (arrows, Fig. 3H).

Finally, many of the nine *in silico*-identified genes - *Chrna3, Cbfa2t2, Tfap2b, Prmt8, Sstr2* - show striped expression patterns in the VZ and/or IZ, which is a hallmark of genes directing neurogenesis in a domain-restricted manner (Marklund et al., 2010; Skaggs et al., 2011). For example, both *Tfap2b* and *Prmt8* are expressed in two stripes in the VZ at E10.5 (brackets, Fig. 3E, F), while *Prtm8* and *Sstr2* are discontinuously expressed in the IZ (arrows, Fig. 3F, G). These genes are thus candidates for factors that direct multipotential progenitors into more restricted transition states before resolving into specific dI lineages.

Pseudotime analysis reveals subtype diversification of dIs.

While many clusters could be identified using canonical markers, the identity of cluster 16 could not be resolved (Fig. 1E, 4A). By eye, cluster 16 is consistent with being at the endpoint of the *Foxd3*$^+$ dI2 lineage (Fig. 1D), however, cluster 16 is derived from the RA protocol which generates mostly dI4/dI5/dI6s (Fig. 1C). Supporting this possibility, the monocle trajectories suggested that either the dI5 or dI4 lineages could contribute to cluster 16 (Fig. 4B). To resolve the identity and origin of cluster 16, marker gene analysis was used to identify *Sncg* - synuclein gamma - as the most significantly upregulated gene in cluster 16 (Fig. 4C, D). GO analysis of the *Sncg*$^+$ expressing cells vs. those exclusively in cluster 16 showed similar enrichment of terms related to axon projection and synapses, indicating that these cells may collectively participate in establishing long range connections (Fig. 4F, G).

To further investigate these populations, we performed a marker gene analysis on *Sncg* positive cells in the terminal region (clusters 16 and 7). This analysis identified *synaptotagmin* (*Syt*) *4* and *Syt13* (Fig. S3A) which are expressed in the Phox2a$^+$ dI5 subtype that relays pain and itch to the thalamus (Roome et al., 2020), suggesting that cluster 16 is at the endpoint of the dI5 lineage (Fig. 4E). We further examined this hypothesis by analyzing *Sncg* expression in the E11.5, E12.5 and E13.5 mouse spinal cord *in vivo*, in combination with immunohistochemistry for Pax2, which labels dI4 and dI6 (Gross et al., 2002), and Lmx1b, which decorates dI5s (Ding et al., 2004) (Fig. 5). At both stage E11.5 and E12.5, *Sncg* expression colocalizes with Lmx1b$^+$, but not Pax2$^+$, cells (inset Fig. 5A, B), supporting the hypothesis that cluster 16 represents a dI5 subtype (Fig. 4E). The anatomical position of this *Sncg*$^+$ Lmx1b$^+$ cluster is also consistent with the Phox2a$^+$ dI5 population identified by lineage tracing (Roome et al., 2020). However, Phox2a is not robustly expressed in cluster 16 (Fig. S3B), and the co-localization of *Sncg* transcripts within Lmxb1$^+$ cells does not persist into stage E13.5 (Fig 5C). Previous studies have suggested that Phox2a$^+$ dI5s migrate tangentially to the upper laminae of the dorsal horn starting at E13.5 (Roome et al., 2020). While there are a few dispersed *Sncg*-expressing cells amidst the Lmx1b$^+$ neurons at E13.5, the majority of the *Sncg*$^+$ cells remain in a deep layer (Fig. 5C, inset), suggesting that *Sncg* is either transiently expressed in Phox2a$^+$ dI5s or represents a distinct subset of neurons.

Online tool to investigate gene expression dynamics during dI differentiation in vitro.

We have created an online data tool available at https://samjbutler.shinyapps.io/Data_Viewer/ to permit the straightforward visualization of any gene present in the UMAP reduction, together with the pseudotime trajectory heatmaps (Supplemental movie 2). A variety of parameters can be adjusted using the menus including the span size for fitting a smooth curve, i.e., loess regression, and size of the plots. This tool provides a starting point for researchers to discover new dI-associated genes in our dataset, by visualizing their expression patterns and temporal changes, and thereby inferring function. This resource, and its associated data sets, may further serve as a roadmap for the derivation of dIs from both mouse and human ESCs and iPSCs and enable functional comparisons across species.

Discussion

We have established a tool to track dI fate specification both spatially and temporally *in silico* by adding pseudotemporal ordering to our scRNA-Seq atlas of the developing dorsal spinal cord. We have used this tool to identify the lineage relationships between different trajectories and predict the identity of novel regulators of dI fates. Our *in vivo* data validate this approach and further illustrate how RA±BMP4 directed differentiation protocols can be used to investigate dI cell fate decisions. We anticipate that this tool will serve as a valuable resource ro explore the genes and molecular pathways that contribute to the development of the somatosensory system in the spinal cord.

A spatial and temporal map of dI differentiation

Our analyses (Fig. 4E) distinguish five distinct trajectories corresponding to the dI1-dI5 populations. In contrast, dI6s emerge at the end of the dI4 trajectory (Fig. 1D). This observation may stem from the inefficient generation of dI6s *in vitro*, resulting in too few cells to delineate the dI6 lineage trajectory. Alternatively, it may rather accurately reflect endogenous development in the spinal cord, i.e., that dI6s arise from a common progenitor pool with dI4s and become specialized post-mitotically. Supporting the latter possibility, dI6s and dI4s express many of the same transcription factors, including *Pax2*, *Lbx1* and *Lhx1/5* (Lai et al., 2016). An additional possibility is that the dI6 population arises from both dorsal and ventral progenitors. These studies have identified that a subset of dI6s shares some characteristics with dI4. Previous studies have suggested that dI6s *in vivo* arise in part from more ventrally located p0 progenitors (Griener et al., 2017). It is notable that the relative numbers of dIs generated *in vitro* is similar to that *in vivo*. For example, the dI6 and dI3 populations are smaller than the other classes of dIs *in vivo*, similarly ~8% of the cells in RA±BMP4 protocols were dI3/dI6s, compared to 30-45% dI1/dI2/dI4/dI5s (Gupta et al., 2022). Thus, the mechanism for determining how many dIs are produced, may be an intrinsic feature of dP populations.

These studies also suggest a lineage relationship between the dI2 and dI4 populations, which have some common elements to their transcriptional code, i.e., *Neurog2*$^{+}$ dPs which resolve into *Lhx1/5*$^{+}$ dIs (Lai et al., 2016). *In vitro*, both dI2s and dI4s arise from the same dP lineage, which branches when dPs express either *Foxd3* (dI2) or *Pax2* (dI4) (Fig. 1D, Fig. 2A-C). This shared

relationship may explain why the RA protocol makes small numbers of dI2s, while the RA+BMP4 protocol has a small dI4 population (Gupta et al., 2022). The discontinuous lineage relationship between dIs has been observed in *in vivo*: BMP7 depletion may reduce the number of dI1, dI3 and dI5s in mouse (Le Dreau et al., 2012) while BMP4 overexpression in chicken dramatically increases the number of dI2s, potentially at the expense of dI4/dI6s (Andrews et al., 2017).

Our analysis of undifferentiated progenitors also reveals novel genes that may regulate dP fate specification. For example, *insulin growth factor binding protein (Igfbp) 5,* but not *Igfbp3*, is expressed in cycling spinal progenitors (Fig. 1G), suggesting they respond selectively to insulin signaling. Insulin signaling, acting though *Igfbp5,* regulates proliferation in other progenitor populations, such the olfactory bulb subventricular zone (Vicario-Abejón et al., 2003) and smooth muscle progenitors (Ahmad et al., 2020; Ren et al., 2008). Thus, modulation of insulin signaling could expand specific dP populations.

dI transcriptional identities converge during terminal differentiation

Both the UMAP and pseudotime analyses suggest that dI lineage trajectories are at their most divergent during the period when dIs are assuming their unique fates. However, at more mature pseudotime values, the dI2-dI6s converge back together into a series of clusters (clusters 6, 7, 8, 13 and 16, Fig. 4A) with similar transcriptomes (Fig. 4B, E). While there are some genetic differences between these clusters, their close proximity in UMAP space suggests that these dIs have become more transcriptionally homogenous. One possibility to explain this observation is that significant differences in gene expression are required to direct dI cell fate specification, migration, and axon guidance. However, once dIs have completed the differentiation process, then they require lower transcriptional heterogeneity because they function similarly to each other. This phenomenon has also been observed during the development of dorsal root ganglia (Faure et al., 2020), oligodendrocytes (Marques et al., 2018), *Drosophila* brain (Michki et al., 2021) and the *C. elegans* nervous system (Hobert and Kratsios, 2019). Thus, the loss of transcriptional heterogeneity upon maturation may be a common theme for neural differentiation and may be used as a mechanism to induce diverse subtypes with shared functional identities (Osseward et al., 2021).

Identification of novel regulators of dI differentiation

Our *in silico* analysis of *in vitro*-derived dI trajectories has identified novel putative regulatory genes including ion channels (*Gsg1l, Cacna2d1*), receptors (*Chrna3, Sstr2, EphA5*), actin binding proteins (*Svil, Mtcl1, Tagln3*), enzymes (*Prmt8, Fbxl7*) and transcriptional factors (*Tfap2b, Zeb1, Ted1*). Like previously identified genes that direct transitory states in dPs, i.e., *Neurog1, Neurog2* and *Ascl1,* several of the *in silico* identified genes such as *Chrna3, Prmt8, Cbfa2t2,* and *Tfap2b,* are expressed in stripes of progenitors during E10.5 spinal cord patterning (Fig. 3B. C, E, F). By E11.5, these genes, together with *Sstr2* (Fig. 3G) shift their expression to the IZ, i.e., in dPs that are exiting the cell cycle and differentiating into post-mitotic neurons. Taken together, these analyses suggest these factors regulate dI patterning and neurogenesis. Previous studies support this hypothesis. Both *Chrna3* and *Sstr2* are part of a *NeuroD1*-interacting network, during *NeuroD1* mediated reprograming of astrocytes to neurons (Ma et al., 2022). *Tfap2b* expression is regulated by *Ptf1a* during cerebellar and retinal neurogenesis (Jin et al., 2015; Zainolabidin et al., 2017). In the spinal cord, *Ptf1a* is expressed specifically in the dP4 domain, suggesting it may activate *Tfap2b* to regulate dI4/dI6 differentiation. *Tfap2b* may more generally regulate the differentiation of GABAergic neurons, since it is expressed in the inhibitory neurons in the E18.5 spinal cord (Wildner et al., 2013) and cerebellum (Jin et al., 2015).

Subtype diversification of dIs in vitro

Our analysis of the terminally differentiated clusters has identified *Sncg*, as a new marker for the dI5 subtype. Previously, *Sncg* expression was reported to be present in spinal motor neurons and dorsal root ganglia (Ninkina et al., 2003). The lateral dI5 cluster of *Sncg*$^+$ cells in the developing spinal cord initially coincides with the Phox2a$^+$ dI5 subtype, which contributes to the anterolateral system (ALS) relaying pain and itch to the thalamus (Price and Dubner, 1977; Roome et al., 2020). By E13.5, lineage tracing has shown that Phox2a$^+$ dI5s migrate away from this lateral position along a tangential route, and ultimately populate the surface of the dorsal horn (Roome et al., 2020). However, it is not clear whether *Sncg*$^+$ cells follow this route. There are very few *Sncg*$^+$ cells migrating dorsally at E13.5; rather, the cluster of *Sncg*$^+$ cells persists in its lateral position (inset, Fig. 5C). Thus, the ALS dI5s appear to express both *Phox2a* and *Sncg,*

and then downregulate *Sncg* as they migrate dorsally. The identity of the persisting *Sncg*$^+$ cluster is unresolved. It may represent a novel dI5 subtype which migrates dorsally at a later time point, since *Sncg* is ultimately expressed in some cells in the dorsal-most laminae in the adult spinal cord (arrows Fig. S3C). Interestingly, *Phox2a* was not upregulated in our dataset (Fig. S3B). This discrepancy may reflect the transient expression of *Phox2a* in dI5s *in vivo* (Roome et al., 2020)(Fig. S3B), which could not be captured *in vitro*.

Our molecular analysis of the *Sncg*$^+$ population in our scRNA-Seq atlas also suggests a mechanism for dI diversification. Among the most highly expressed genes in the *Sncg*$^+$ cluster are *D930028M14Rik, Onecut1, Onecut2, and Onecut3*. *Onecut* genes have been shown to represses the expression of *Pou2f2*, and thereby regulate the distribution of the dI2-dI6 populations (Masgutova et al., 2019). *D930028M14Rik* is also antisense to *Pou2f2* (O'Leary et al., 2016), suggesting it could contribute to *Pou2f2* repression. The downregulation of *Pou2f2* via *Onecut1/2* and *D930028M14Rik*, could represent a genetic switch allowing the *Sncg* population to diverge from the Phox2a$^+$ dI5 population. Taken together, this analysis reveals that the *in vitro*-derived dIs diversify into subtypes also found *in vivo*, underscoring how closely the RA±BMP4 directed differentiation protocols recapitulate endogenous programs of dI cell fate specification.

Acknowledgements

We would like to thank Riki Kawaguchi and members of the Butler and Novitch labs for discussions.

Funding statement

This work was supported by the UCLA Broad Stem Cell Research Center (BSCRC) postdoctoral training grant (to S.G) and grants from the National Institutes of Health (NIH) (R01 NS123187 and R01 NS085097 to S.J.B.; R21 NS115012 and R01 NS085227 to B.G.N.), and awards from the BSCRC (to B.G.N. and S.J.B.).

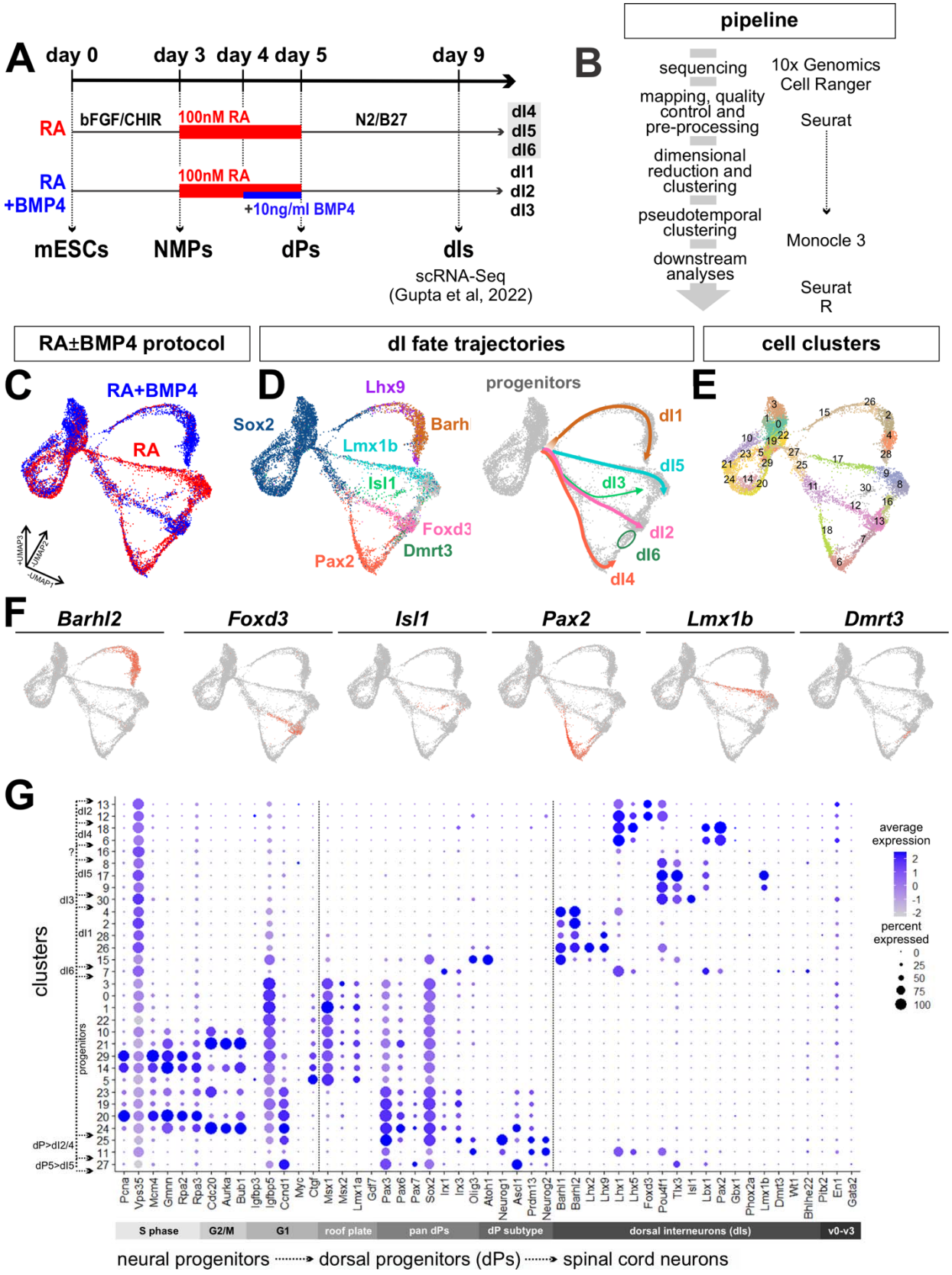
Data Accessibility

Sequencing data is available at Gene Expression Omnibus accession number GSE185891 (Samples GSM5625332 and GSM5625333). Processed data and code available on request.

Statement on competing interests

There were no competing interests.

Figure 1: Single-cell analysis pipeline to identify dorsal interneuron lineages.



(A) Schematic timeline for the derivation of dorsal interneurons (dIs) from mouse embryonic stem cells (mESCs). On day 9 of the differentiation, cells were dissociated and subjected to single-cell RNA sequencing (Gupta et al., 2022).

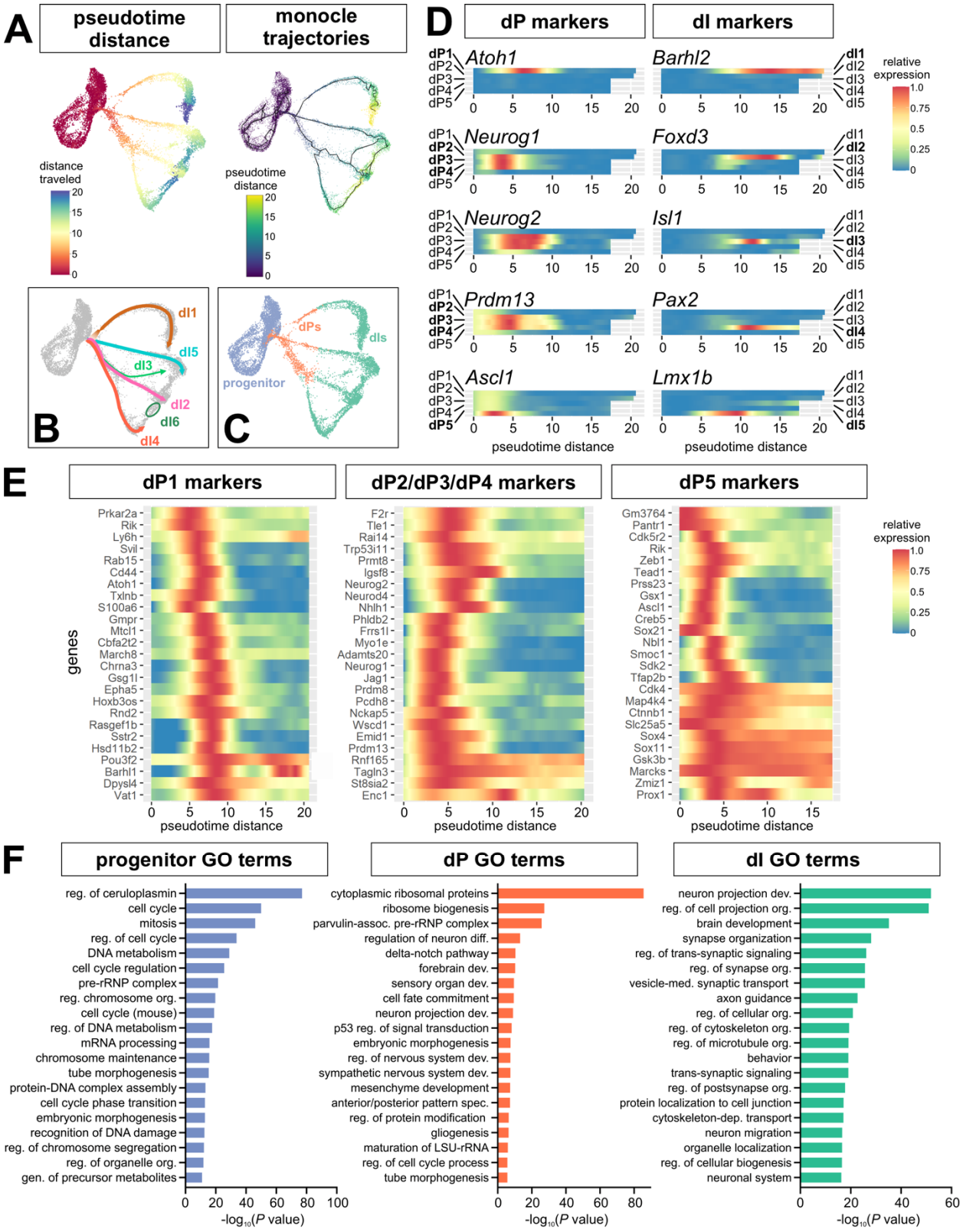
(B) Overview of the pipeline for analysis of the single-cell transcriptomic data.

(C-E) UMAP plots depict the combined cell types derived through the RA±BMP4 protocols (C), and the distinct dI lineages, as designated by marker analysis (D). Unsupervised clustering results in 31 distinct transcriptional clusters (E).

(F) UMAP feature plots showing the expression of cardinal markers for all six classes of dIs.

(G) Dot plot analysis showing the expression of various genes, which groups clusters in a continuous stream of progenitors to dorsal sensory interneurons (dIs).

Figure 2: Using pseudotime to identify dI-specific trajectories and markers *in vitro*.



(A) Pseudotime analysis identifies both the distance and trajectories over which progenitors transition to dPs and then differentiated dIs.

(B) Putative dI trajectories based on the marker analysis in Fig. 1D.

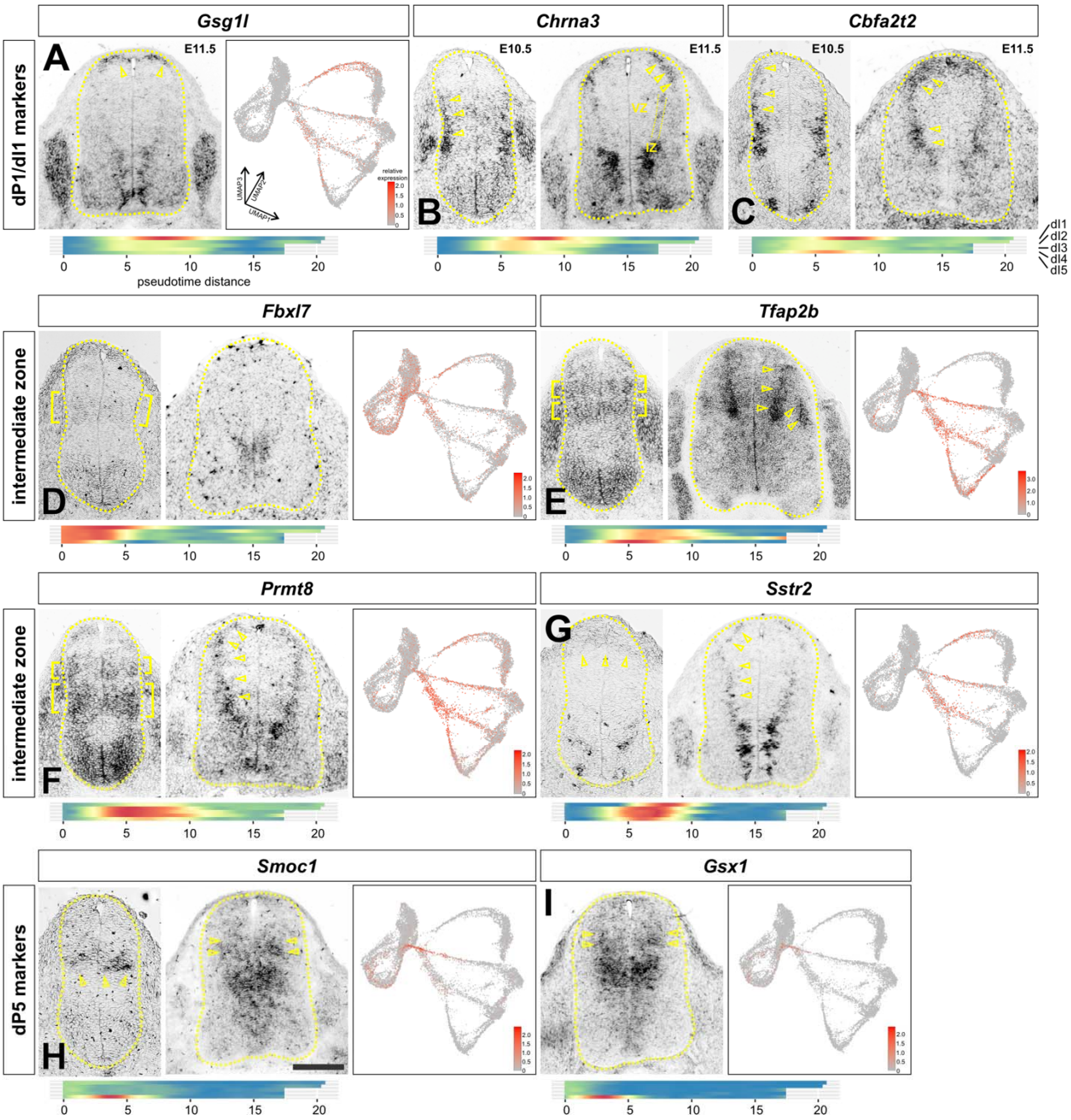
(C) Based on gene expression, we classified progenitors to be the root cells used in Monocle3, dPs as being from >0-7.5 pseudotime distance, and differentiated dIs from ≥7.5 – 20 pseudotime distance.

(D) Heatmaps drawn from the dI1-dI5 pseudotime trajectories show the correct temporal distribution of cardinal dP and dI markers.

(E) Heatmaps showing the expression of marker genes expressed in the three major dP clusters, i.e., dP1, dP2/dP3/dP4 and dP5, many of which are novel.

(F) Gene ontology (GO) analysis shows enriched biological processes in the clusters assigned to the progenitor, dP, and differentiated dI identities, as shown in panel C.

Figure 3: *In vivo* validation of dI lineage markers identified *in vitro*



Gene expression was assessed in E10.5 and E11.5 lumbar and thoracic spinal cord sections by *in situ* hybridization and compared to the predicted distribution and timing of expression from the UMAP reduction and pseudotime ordering (heatmap).

(A) *Gsg1l* is specifically expressed in dP1s, both *in vivo* and *in vitro*.

(B, C) *Chrna3* and *Cbfa2t2* are expressed in newly differentiating dIs, i.e., in the intermediate

zone (IZ, dotted lines), with highest levels in the dI1s. The heatmap shows similar enriched expression in the transitory region of the dI trajectories.

(D) *Fbxl7* expressed at low levels in intermediate dPs in the ventricular zone (VZ) at E10.5. Expression then diminishes by E11.5, as predicted *in vitro*.

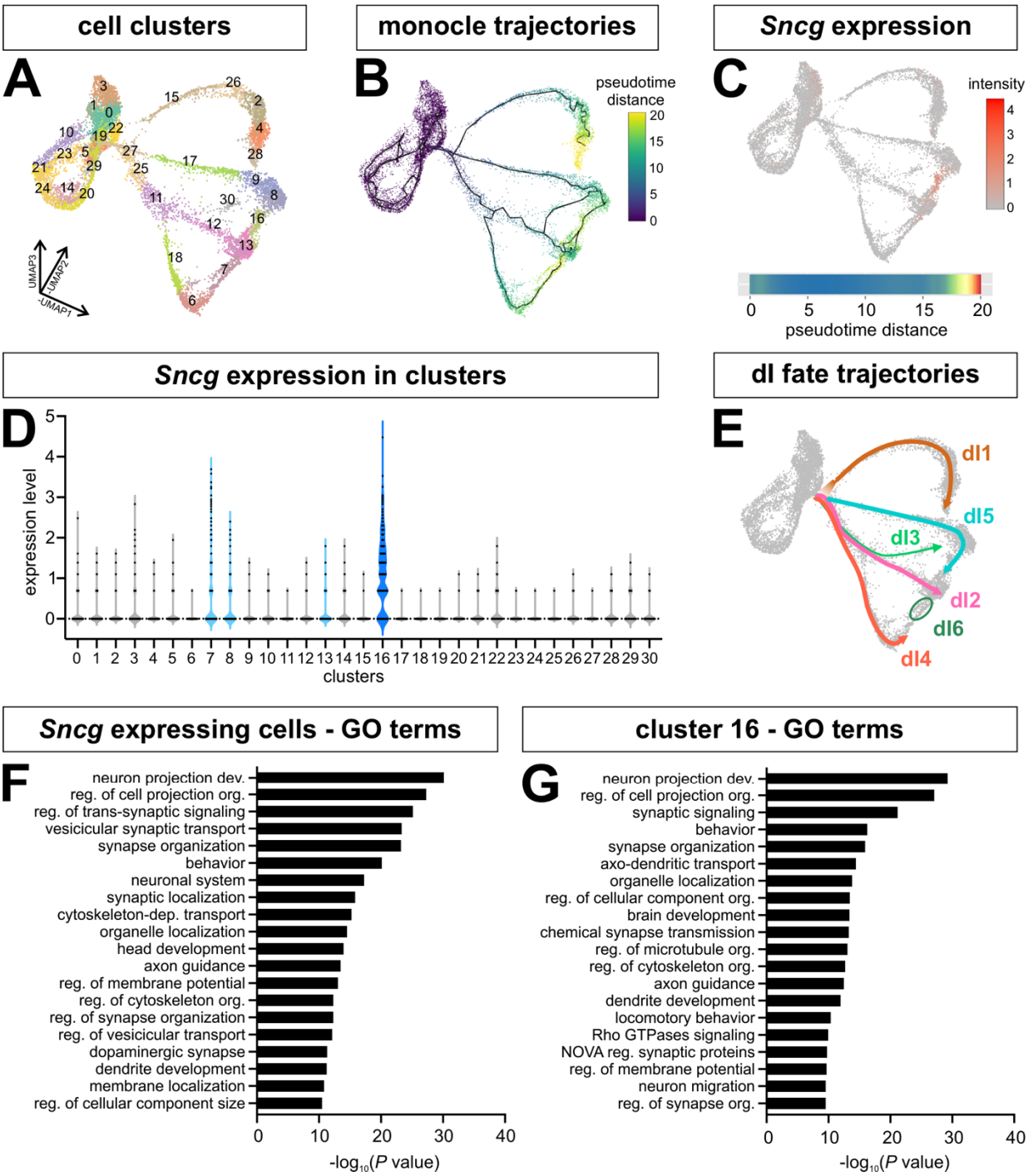
(E, F) *Tfap2b* and *Prmt8* are expressed in stripes in dPs at E10.5. Expression resolves to the newly differentiating dIs in the IZ at E11.5, as predicted by the heatmaps.

(G) *Sstr2* is expressed in dorsal dPs at E10.5 and differentiating dIs in the IZ at E11.5, corresponding to the transitory clusters in the heatmap.

(H-I) *Smoc1* and *Gsx1* are expressed broadly in dPs, with highest expression in the dP5 domain.

Scale bar: E10.5 images: 50µm; E11.5 images 100 µm

Figure 4: Characterization of Synuclein-γ (Sncg)$^{+}$ cluster 16.



(A) Unsupervised clustering resolves the dI trajectories into 31 clusters.

(B) The Monocle3 derived pseudotime trajectories suggest that dI5, dI4 and dI2 could converge on cluster 16.

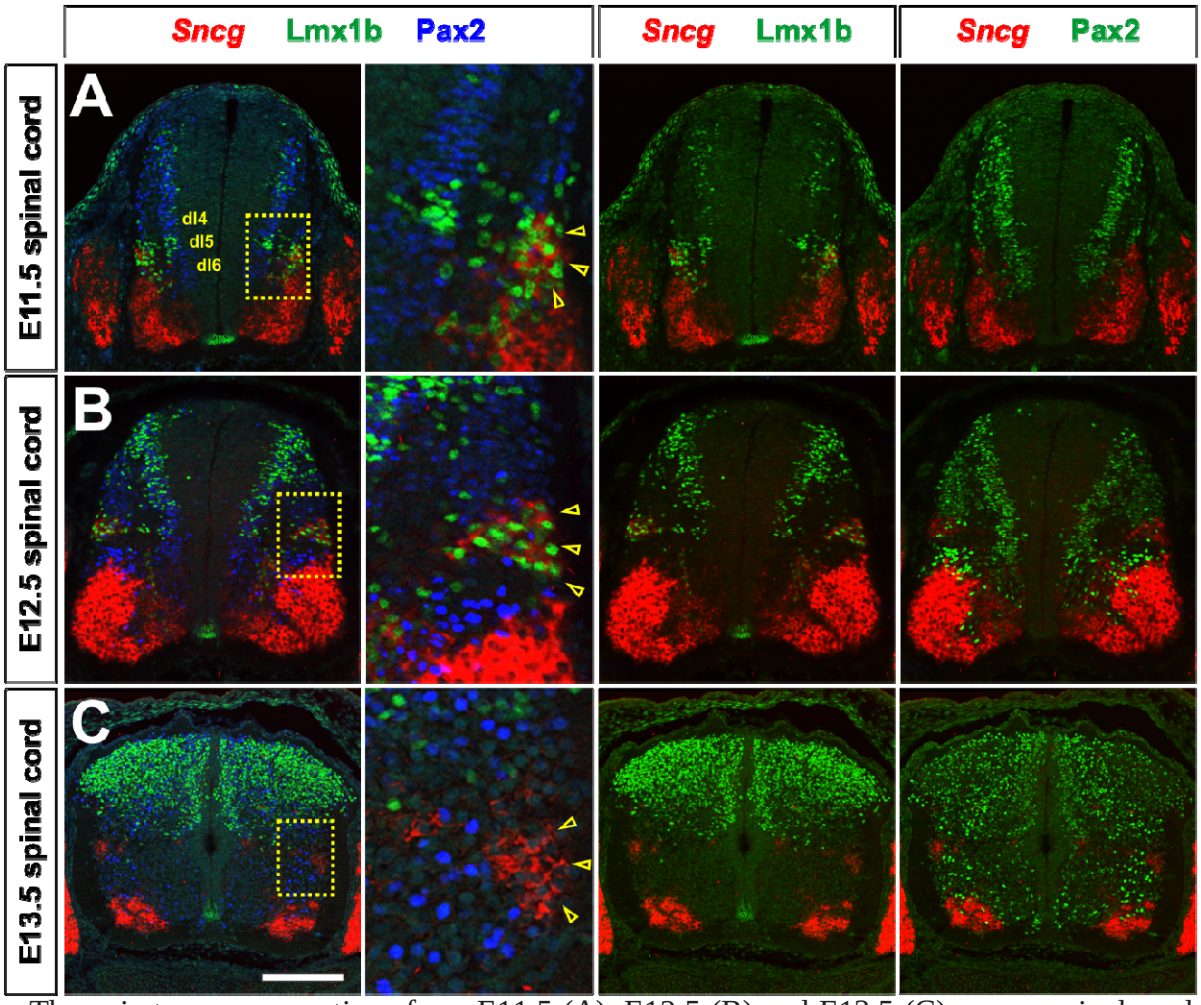
(C-D) Synuclein-γ (Sncg) is most robustly expressed in cluster 16, with some expression in

cluster 7, an adjacent cluster. Expression in the other two adjacent clusters, clusters 8 and 13, is not above background.

(E) The dI fate trajectories were redrawn from Fig.1D, 2B to depict cluster 16 as the end point of the dI5 lineage.

(F, G) Similar GO terms, related to axon guidance and synapse formation, were enriched in the Sncg expressing cells (F) and cluster 16 (G).

Figure 5: Identification of *Sncg*-expressing cells as a dI5 subtype.



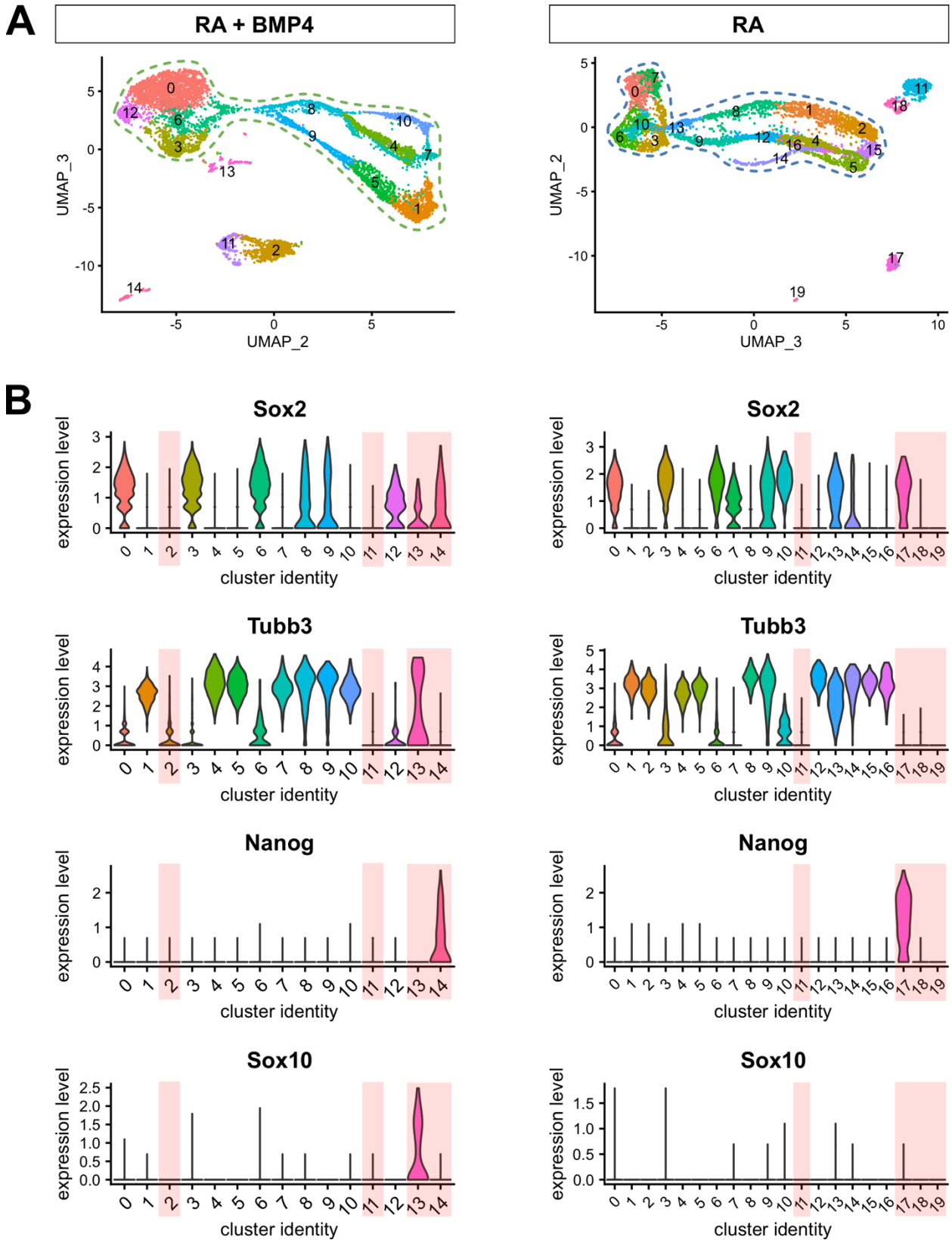
Thoracic transverse sections from E11.5 (A), E12.5 (B) and E13.5 (C) mouse spinal cords were subjected to both *in situ* hybridization to detect *Sncg (*red) and immunohistochemistry to detect Pax2 (dI4 blue, green) and Lmx1b (dI5, green).

(A, B) At both E11.5 and E12.5, *Sncg* expression is observed in a cluster of Lmx1b^{+} cells (inset), and not in Pax2^{+} neurons, in the dorsal spinal cord. The lateral position of *Sncg*-expressing cells is coincident with the Phox2a^{+} dI5 subtype (Roome et al., 2020).

(C) By E13.5, this lateral population of cells continues to express *Sncg*, but not Lmx1b, or Pax2 (inset).

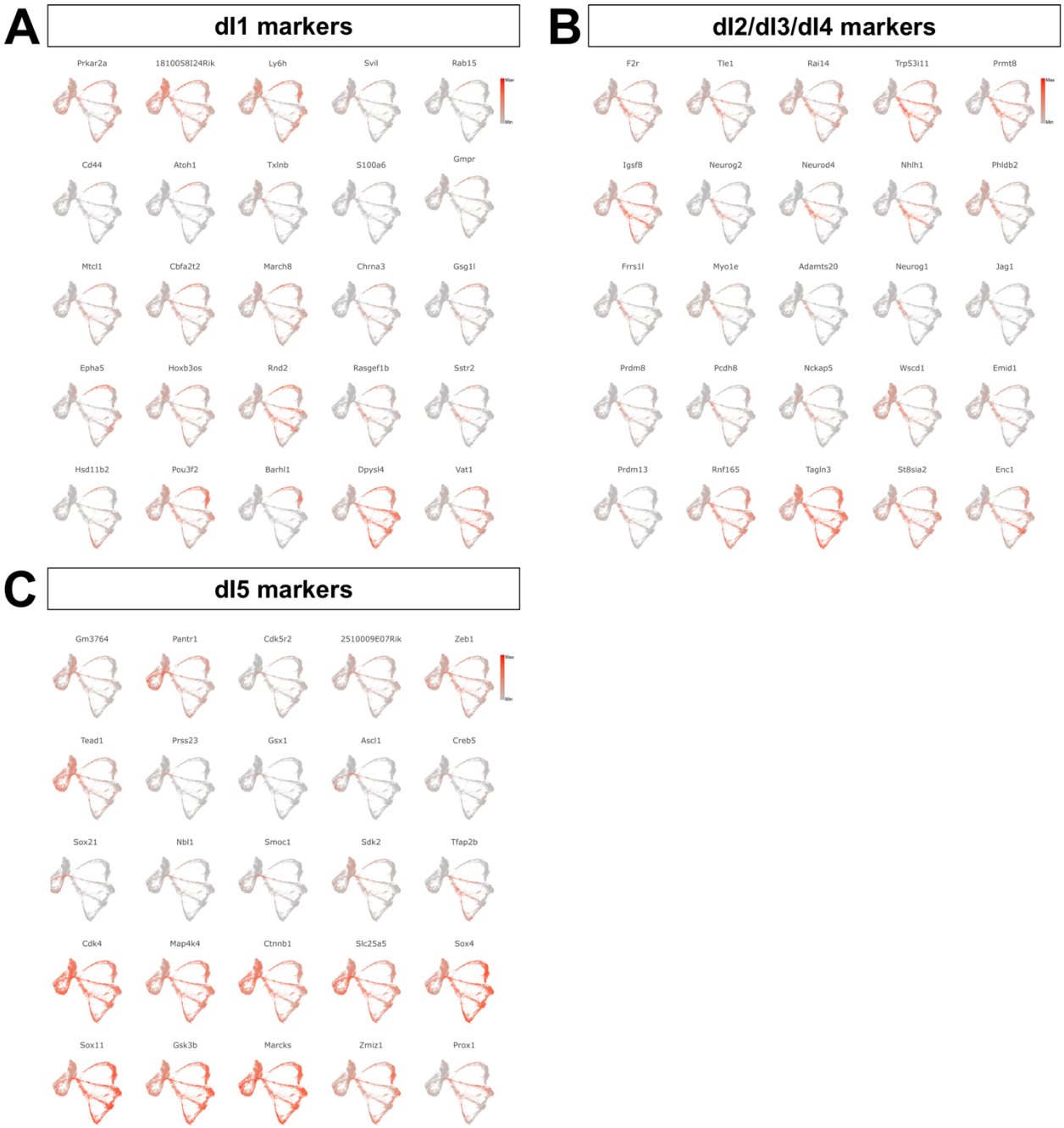
Scale bar: (A, B) 150µm, (C) 200µm

Figure S1: Inclusion and exclusion criteria for cells included in the trajectory analyses.



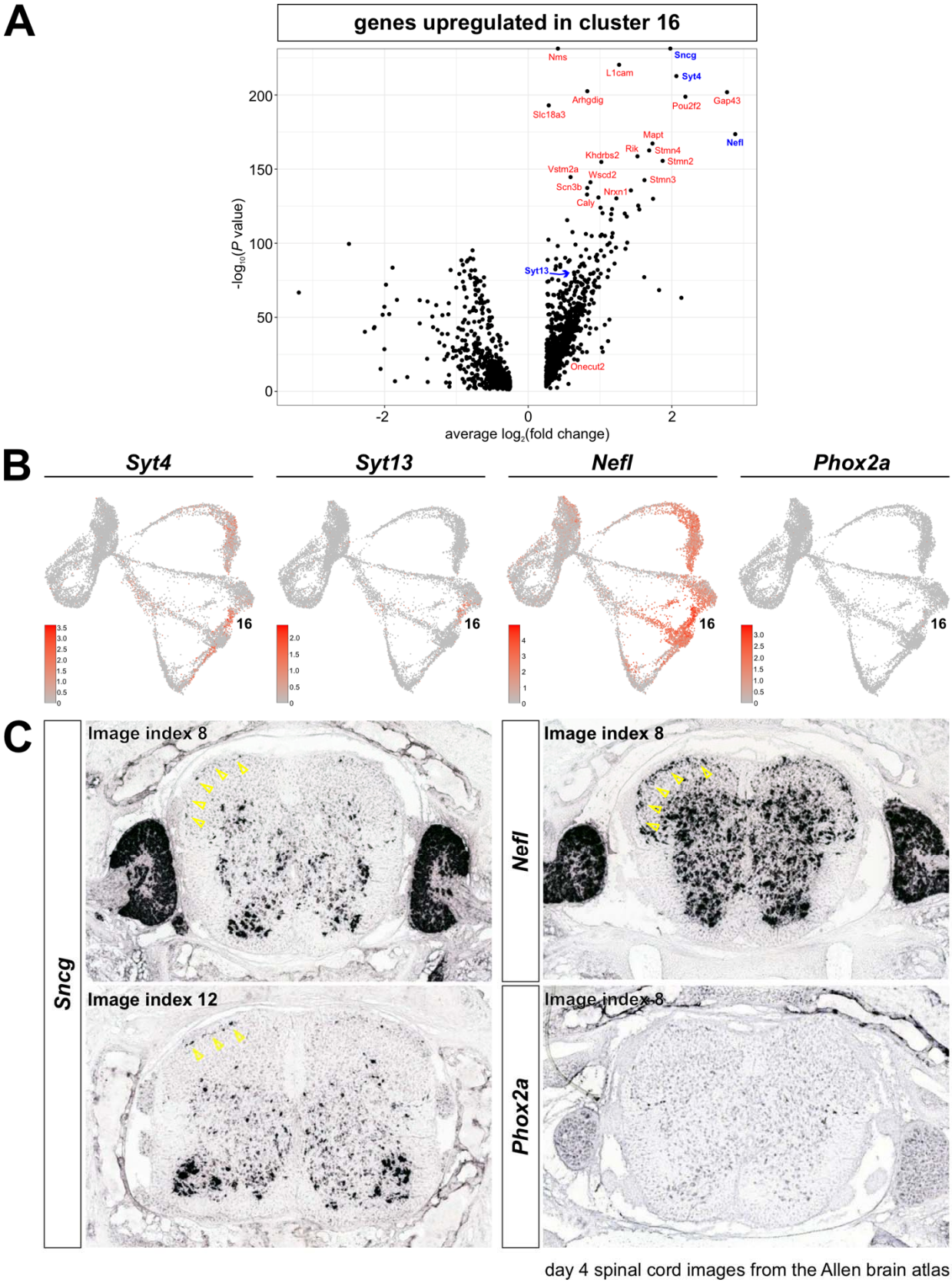
(A) UMAPs displaying the major cell clusters present in cultures resulting from the RA+BMP4 and RA-only protocols. Dashed lines illustrate cells that displayed neural progenitor or neuronal characteristics and were thereby included in the subsequent developmental trajectory analyses. (B) Violin plots of gene expression in each cluster. Clusters (pink shading) that showed either: a) no *Sox2* or *Tubb3* expression or b) expression of *Nanog*, a pluripotent stem cell marker, or *Sox10*, a neural crest progenitor marker, were excluded from the datasets.

Figure S2: UMAPs of novel genes expressed in the three major dP clusters: dP1, dP2/dP3/dP4 and dP5.



(A-C) UMAPs displaying the relative gene expression (minimal to maximal) of different genes that were enriched during the differentiation of dPs into dIs, as outlined in Fig. 2E.

Figure S3: Assessing the identity of cluster 16



(A) Volcano plot showing the genes upregulated in cluster 16. These genes include *Syt4* and *Syt13* which have previously been identified as upregulated in the dI5 subtype (Roome et al., 2020)

(B) UMAP projections of genes associated with the thalamus innervating (ALS) dI5 subtype. These genes are also upregulated in *Sncg*$^{+}$ cluster 16, with the exception of *Phox2a*.

(C) *In situ* hybridization images of postnatal day 4 spinal cord, taken from the Allen brain atlas. Two cluster 16 markers - *Sncg* and *Nefl* – are present in cells on the surface of the dorsal horn (arrows). The location of these cells is consistent with that of the ALS dI5s. Note that *Phox2a* is only transiently present in ALS dI5 subtypes during development, and is thus not expressed in postnatal spinal cord.

Supplemental Movie 1

Supplementary movie showing Monocle3 trajectories. At later pseudotime values, the trajectories merge back together, with endpoints in clusters 16 and 7, which both express high levels of *Sncg*.

Supplemental Movie 2

Supplementary movie illustrating how the interactive tool can be used to visualize clusters, pseudotime distance, and expression patterns in the dataset. Multiple genes can be followed at once, to compare the relative intensity of gene expression.

References

Ahmad, S. S., Ahmad, K., Lee, E. J., Lee, Y. H. and Choi, I. (2020). Implications of Insulin-Like Growth Factor-1 in Skeletal Muscle and Various Diseases. *Cells* **9**.

Alaynick, W. A., Jessell, T. M. and Pfaff, S. L. (2011). SnapShot: spinal cord development. *Cell* **146**, 178-178 e171.

Andrews, M. G., Del Castillo, L. M., Ochoa-Bolton, E., Yamauchi, K., Smogorzewski, J. and Butler, S. J. (2017). BMPs direct sensory interneuron identity in the developing spinal cord using signal-specific not morphogenic activities. *eLife* **6**.

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D. M., Hill, A. J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F. J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496-502.

Cazales, M., Schmitt, E., Montembault, E., Dozier, C., Prigent, C. and Ducommun, B. (2005). CDC25B phosphorylation by Aurora-A occurs at the G2/M transition and is inhibited by DNA damage. *Cell Cycle* **4**, 1233-1238.

Ding, Y. Q., Yin, J., Kania, A., Zhao, Z. Q., Johnson, R. L. and Chen, Z. F. (2004). Lmx1b controls the differentiation and migration of the superficial dorsal horn neurons of the spinal cord. *Development* **131**, 3693-3703.

Dong, R., Li, X. and Lai, K. O. (2021). Activity and Function of the PRMT8 Protein Arginine Methyltransferase in Neurons. *Life (Basel)* **11**.

Fahd Qadir, S. S., & Juan Domínguez-Bendala (2019). 3D Plotting of scRNAseq data using Seurat objects. Zenodo.

Faure, L., Wang, Y., Kastriti, M. E., Fontanet, P., Cheung, K. K. Y., Petitpre, C., Wu, H., Sun, L. L., Runge, K., Croci, L., et al. (2020). Single cell RNA sequencing identifies early diversity of sensory neurons forming via bi-potential intermediates. *Nature communications* **11**, 4175.

Flora, A. V., Zambrano, C. A., Gallego, X., Miyamoto, J. H., Johnson, K. A., Cowan, K. A., Stitzel, J. A. and Ehringer, M. A. (2013). Functional characterization of SNPs in CHRNA3/B4 intergenic region associated with drug behaviors. *Brain Res* **1529**, 1-15.

Gaspard, N. and Vanderhaeghen, P. (2010). Mechanisms of neural specification from embryonic stem cells. *Curr Opin Neurobiol* **20**, 37-43.

Gouti, M., Tsakiridis, A., Wymeersch, F. J., Huang, Y., Kleinjung, J., Wilson, V. and Briscoe, J. (2014). In vitro generation of neuromesodermal progenitors reveals distinct roles for wnt signalling in the specification of spinal cord and paraxial mesoderm identity. *PLoS biology* **12**, e1001937.

Gowan, K., Helms, A. W., Hunsaker, T. L., Collisson, T., Ebert, P. J., Odom, R. and Johnson, J. E. (2001). Crossinhibitory activities of Ngn1 and Math1 allow specification of distinct dorsal interneurons. *Neuron* **31**, 219-232.

Griener, A., Zhang, W., Kao, H., Haque, F. and Gosgnach, S. (2017). Anatomical and electrophysiological characterization of a population of dI6 interneurons in the neonatal mouse spinal cord. *Neuroscience* **362**, 47-59.

Gross, M. K., Dottori, M. and Goulding, M. (2002). Lbx1 specifies somatosensory association interneurons in the dorsal spinal cord. *Neuron* **34**, 535-549.

Gupta, S. and Butler, S. J. (2021). Getting in touch with your senses: Mechanisms specifying sensory interneurons in the dorsal spinal cord. *WIREs Mech Dis* **13**, e1520.

Gupta, S., Kawaguchi, R., Heinrichs, E., Gallardo, S., Castellanos, S., Mandric, I., Novitch, B. G. and Butler, S. J. (2022). In vitro atlas of dorsal spinal interneurons reveals Wnt signaling as a critical regulator of progenitor expansion. *Cell reports* **40**, 111119.

Gupta, S., Sivalingam, D., Hain, S., Makkar, C., Sosa, E., Clark, A. and Butler, S. J. (2018). Deriving Dorsal Spinal Sensory Interneurons from Human Pluripotent Stem Cells. *Stem cell reports* **10**, 390-405.

Gupta, S., Yamauchi, K., Novitch, B. G. and Butler, S. J. (2021). Derivation of dorsal spinal sensory interneurons from human pluripotent stem cells. *STAR Protoc* **2**, 100319.

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e3529.

Hazen, V. M., Andrews, M. A., Umans, L., Crenshaw, E. B., 3rd, Zwijsen, A. and Butler, S. J. (2012). BMP receptor-activated Smads direct diverse functions during the development of the dorsal spinal cord. *Dev Biol* **367**, 216-227.

Helms, A. W. and Johnson, J. E. (1998). Progenitors of dorsal commissural interneurons are defined by MATH1 expression. *Development* **125**, 919-928.

Hobert, O. and Kratsios, P. (2019). Neuronal identity control by terminal selectors in worms, flies, and chordates. *Curr Opin Neurobiol* **56**, 97-105.

Jin, K., Jiang, H., Xiao, D., Zou, M., Zhu, J. and Xiang, M. (2015). Tfap2a and 2b act downstream of Ptf1a to promote amacrine cell differentiation during retinogenesis. *Mol Brain* **8**, 28.

Kamalova, A., Futai, K., Delpire, E. and Nakagawa, T. (2021). AMPA receptor auxiliary subunit GSG1L suppresses short-term facilitation in corticothalamic synapses and determines seizure susceptibility. *Cell Rep* **34**, 108732.

Koch, S. C., Acton, D. and Goulding, M. (2018). Spinal Circuits for Touch, Pain, and Itch. *Annu Rev Physiol* **80**, 189-217.

Komamura-Kohno, Y., Karasawa-Shimizu, K., Saitoh, T., Sato, M., Hanaoka, F., Tanaka, S. and Ishimi, Y. (2006). Site-specific phosphorylation of MCM4 during the cell cycle in mammalian cells. *FEBS J* **273**, 1224-1239.

Kushwaha, P. P., Rapalli, K. C. and Kumar, S. (2016). Geminin a multi task protein involved in cancer pathophysiology and developmental process: A review. *Biochimie* **131**, 115-127.

Lai, H. C., Klisch, T. J., Roberts, R., Zoghbi, H. Y. and Johnson, J. E. (2011). In vivo neuronal subtype-specific targets of Atoh1 (Math1) in dorsal spinal cord. *J Neurosci* **31**, 10859-10871.

Lai, H. C., Seal, R. P. and Johnson, J. E. (2016). Making sense out of spinal cord somatosensory development. *Development* **143**, 3434-3448.

Lara-Gonzalez, P., Moyle, M. W., Budrewicz, J., Mendoza-Lopez, J., Oegema, K. and Desai, A. (2019). The G2-to-M Transition Is Ensured by a Dual Mechanism that Protects Cyclin B from Degradation by Cdc20-Activated APC/C. *Dev Cell* **51**, 313-325.e310.

Le Dreau, G., Garcia-Campmany, L., Rabadan, M. A., Ferronha, T., Tozer, S., Briscoe, J. and Marti, E. (2012). Canonical BMP7 activity is required for the generation of discrete neuronal populations in the dorsal spinal cord. *Development* **139**, 259-268.

Lee, K. J., Mendelsohn, M. and Jessell, T. M. (1998). Neuronal patterning by BMPs: a requirement for GDF7 in the generation of a discrete class of commissural interneurons in the mouse spinal cord. *Genes Dev* **12**, 3394-3407.

Liem, K. F., Jr., Tremml, G. and Jessell, T. M. (1997). A role for the roof plate and its resident TGFbeta-related proteins in neuronal patterning in the dorsal spinal cord. *Cell* **91**, 127-138.

Liem, K. F., Jr., Tremml, G., Roelink, H. and Jessell, T. M. (1995). Dorsal differentiation of neural plate cells induced by BMP-mediated signals from epidermal ectoderm. *Cell* **82**, 969-979.

Liu, Y., Helms, A. W. and Johnson, J. E. (2004). Distinct activities of Msx1 and Msx3 in dorsal neural tube development. *Development* **131**, 1017-1028.

Ma, N. X., Puls, B. and Chen, G. (2022). Transcriptomic analyses of NeuroD1-mediated astrocyte-to-neuron conversion. *Developmental neurobiology* **82**, 375-391.

Marklund, U., Hansson, E. M., Sundstrom, E., de Angelis, M. H., Przemeck, G. K., Lendahl, U., Muhr, J. and Ericson, J. (2010). Domain-specific control of neurogenesis achieved through patterned regulation of Notch ligand expression. *Development* **137**, 437-445.

Marques, S., van Bruggen, D., Vanichkina, D. P., Floriddia, E. M., Munguba, H., Varemo, L., Giacomello, S., Falcao, A. M., Meijer, M., Bjorklund, A. K., et al. (2018). Transcriptional Convergence of Oligodendrocyte Lineage Progenitors during Development. *Developmental cell* **46**, 504-517 e507.

Masgutova, G., Harris, A., Jacob, B., Corcoran, L. M. and Clotman, F. (2019). Pou2f2 Regulates the Distribution of Dorsal Interneurons in the Mouse Developing Spinal Cord. *Front Mol Neurosci* **12**, 263.

Megason, S. G. and McMahon, A. P. (2002). A mitogen gradient of dorsal midline Wnts organizes growth in the CNS. *Development* **129**, 2087-2098.

Michki, N. S., Li, Y., Sanjasaz, K., Zhao, Y., Shen, F. Y., Walker, L. A., Cao, W., Lee, C. Y. and Cai, D. (2021). The molecular landscape of neural differentiation in the developing Drosophila brain revealed by targeted scRNA-seq and multi-informatic analysis. *Cell reports* **35**, 109039.

Mizuguchi, R., Kriks, S., Cordes, R., Gossler, A., Ma, Q. and Goulding, M. (2006). Ascl1 and Gsh1/2 control inhibitory and excitatory cell fate in spinal sensory interneurons. *Nat Neurosci* **9**, 770-778.

Muroyama, Y., Fujihara, M., Ikeya, M., Kondoh, H. and Takada, S. (2002). Wnt signaling plays an essential role in neuronal specification of the dorsal spinal cord. *Genes Dev* **16**, 548-553.

Ninkina, N., Papachroni, K., Robertson, D. C., Schmidt, O., Delaney, L., O'Neill, F., Court, F., Rosenthal, A., Fleetwood-Walker, S. M., Davies, A. M., et al. (2003). Neurons expressing the highest levels of gamma-synuclein are unaffected by targeted inactivation of the gene. *Mol Cell Biol* **23**, 8233-8245.

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* **44**, D733-745.

Osseward, P. J., 2nd, Amin, N. D., Moore, J. D., Temple, B. A., Barriga, B. K., Bachmann, L. C., Beltran, F., Jr., Gullo, M., Clark, R. C., Driscoll, S. P., et al. (2021). Conserved genetic signatures parcellate cardinal spinal neuron classes into local and projection subsets. *Science* **372**, 385-393.

Price, D. D. and Dubner, R. (1977). Neurons that subserve the sensory-discriminative aspects of pain. *Pain* **3**, 307-338.

Qadir, M. M. F., Alvarez-Cubela, S., Klein, D., van Dijk, J., Muniz-Anquela, R., Moreno-Hernandez, Y. B., Lanzoni, G., Sadiq, S., Navarro-Rubio, B., Garcia, M. T., et al. (2020). Single-cell resolution analysis of the human pancreatic ductal progenitor cell niche. *Proc Natl Acad Sci U S A* **117**, 10876-10887.

R Core Team (2022). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Ren, H., Yin, P. and Duan, C. (2008). IGFBP-5 regulates muscle cell differentiation by binding to IGF-II and switching on the IGF-II auto-regulation loop. *J Cell Biol* **182**, 979-991.

Roome, R. B., Bourojeni, F. B., Mona, B., Rastegar-Pouyani, S., Blain, R., Dumouchel, A., Salesse, C., Thompson, W. S., Brookbank, M., Gitton, Y., et al. (2020). Phox2a Defines a Developmental Origin of the Anterolateral System in Mice and Humans. *Cell Rep* **33**, 108425.

Sievert, C. (2020). *Interactive Web-Based Data Visualization with R, plotly, and shiny*: Chapman and Hall/CRC.

Skaggs, K., Martin, D. M. and Novitch, B. G. (2011). Regulation of spinal interneuron development by the Olig-related protein Bhlhb5 and Notch signaling. *Development* **138**, 3199-3211.

Sommer, L., Ma, Q. and Anderson, D. J. (1996). neurogenins, a novel family of atonal-related bHLH transcription factors, are putative mammalian neuronal determination genes that reveal progenitor cell heterogeneity in the developing CNS and PNS. *Mol Cell Neurosci* **8**, 221-241.

Stumm, R. K., Zhou, C., Schulz, S., Endres, M., Kronenberg, G., Allen, J. P., Tulipano, G. and Höllt, V. (2004). Somatostatin receptor 2 is activated in cortical neurons and contributes to neurodegeneration after focal ischemia. *J Neurosci* **24**, 11404-11415.

Thomas, J. T., Eric Dollins, D., Andrykovich, K. R., Chu, T., Stultz, B. G., Hursh, D. A. and Moos, M. (2017). SMOC can act as both an antagonist and an expander of BMP signaling. *Elife* **6**.

Tu, S., Narendra, V., Yamaji, M., Vidal, S. E., Rojas, L. A., Wang, X., Kim, S. Y., Garcia, B. A., Tuschl, T., Stadtfeld, M., et al. (2016). Co-repressor CBFA2T2 regulates pluripotency and germline development. *Nature* **534**, 387-390.

Veenvliet, J. V., Lenne, P. F., Turner, D. A., Nachman, I. and Trivedi, V. (2021). Sculpting with stem cells: how models of embryo development take shape. *Development* **148**.

Vicario-Abejón, C., Yusta-Boyo, M. J., Fernández-Moreno, C. and de Pablo, F. (2003). Locally born olfactory bulb stem cells proliferate in response to insulin-related factors and require endogenous insulin-like growth factor-I for differentiation into neurons and glia. *J Neurosci* **23**, 895-906.

Wang, Q., He, G., Hou, M., Chen, L., Chen, S., Xu, A. and Fu, Y. (2018). Cell Cycle Regulation by Alternative Polyadenylation of CCND1. *Sci Rep* **8**, 6824.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*: Springer-Verlag New York.

Wildner, H., Das Gupta, R., Brohl, D., Heppenstall, P. A., Zeilhofer, H. U. and Birchmeier, C. (2013). Genome-wide expression analysis of Ptf1a- and Ascl1-deficient mice reveals new markers for distinct dorsal horn interneuron populations contributing to nociceptive reflex plasticity. *J Neurosci* **33**, 7299-7307.

Yamauchi, K., Phan, K. D. and Butler, S. J. (2008). BMP type I receptor complexes have distinct activities mediating cell fate and axon guidance decisions. *Development* **135**, 1119-1128.

Zainolabidin, N., Kamath, S. P., Thanawalla, A. R. and Chen, A. I. (2017). Distinct Activities of Tfap2A and Tfap2B in the Specification of GABAergic Interneurons in the Developing Cerebellum. *Front Mol Neurosci* **10**, 281.

Zerjatke, T., Gak, I. A., Kirova, D., Fuhrmann, M., Daniel, K., Gonciarz, M., Müller, D., Glauche, I. and Mansfeld, J. (2017). Quantitative Cell Cycle Analysis Based on an Endogenous All-in-One Reporter for Cell Tracking and Classification. *Cell Rep* **19**, 1953-1966.

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., Benner, C. and Chanda, S. K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **10**, 1523.

Zhu, Z. and Huangfu, D. (2013). Human pluripotent stem cells: an emerging model in developmental biology. *Development* **140**, 705-717.