DEVELOPMENTAL NEUROSCIENCE

Modular derivation of diverse, regionally discrete human posterior CNS neurons enables discovery of transcriptomic patterns

Nisha R. Iyer[1,2]†, Junha Shin[1]†, Stephanie Cuskey[1], Yucheng Tian[1,2], Noah R. Nicol[1,2], Tessa E. Doersch[1], Frank Seipel[1,2], Sunnie Grace McCalla[1,3], Sushmita Roy[1,3]*, Randolph S. Ashton[1,2]*

Copyright © 2022 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

Our inability to derive the neuronal diversity that comprises the posterior central nervous system (pCNS) using human pluripotent stem cells (hPSCs) poses an impediment to understanding human neurodevelopment and disease in the hindbrain and spinal cord. Here, we establish a modular, monolayer differentiation paradigm that recapitulates both rostrocaudal (R/C) and dorsoventral (D/V) patterning, enabling derivation of diverse pCNS neurons with discrete regional specificity. First, neuromesodermal progenitors (NMPs) with discrete *HOX* profiles are converted to pCNS progenitors (pCNSPs). Then, by tuning D/V signaling, pCNSPs are directed to locomotor or somatosensory neurons. Expansive single-cell RNA-sequencing (scRNA-seq) analysis coupled with a novel computational pipeline allowed us to detect hundreds of transcriptional markers within region-specific phenotypes, enabling discovery of gene expression patterns across R/C and D/V developmental axes. These findings highlight the potential of these resources to advance a mechanistic understanding of pCNS development, enhance in vitro models, and inform therapeutic strategies.

INTRODUCTION

Nervous system diversity arises in response to a complex choreography of spatiotemporally restricted cues along the elongating embryo's rostrocaudal (R/C) and dorsoventral (D/V) axes. These coordinated patterning events encode neural progenitors and postmitotic neurons with unique transcriptional signatures that define a myriad of subtypes, which in turn orchestrate the precise neural circuits that shape human behavior (*1*, *2*). While human pluripotent stem cell (hPSC)–based approaches can, in theory, provide access to all these populations, differentiation strategies have intensely focused on recapitulating spinal D/V patterning with less attention to the generation of subtypes along the R/C axis. Even so, direct differentiation protocols have been achieved for relatively few cardinal neurons, which default to hindbrain or cervical identity (*3*–*8*). Motor neuron (MN) optimization has predominated, with robust differentiation schemas allowing high yields (*3*, *5*, *8*) and some control over columnar and R/C identity (*8*, *9*), but these protocols are not designed to adapt to other phenotypes. There has been some success in recreating R/C (*10*, *11*) and D/V (*12*–*14*) signaling centers in human organoid models. However, the variability in efficiency, cell type distribution, and maturity of terminal populations, as well as the difficulty of cell recovery from organoid tissues, limits the scalability of these platforms for clinical applications.

We sought to develop a robust, modular differentiation methodology in monolayer culture to derive any posterior central nervous system (pCNS) phenotype by recapitulating the sequence of patterning events during development. Morphogenesis of the posterior neural tube, which forms the hindbrain and spinal cord (i.e., pCNS), is distinct from the anterior neural tube, which forms the brain. It begins near the primitive streak with a bipotent population of axial stem cells called neuromesodermal progenitors (NMPs) (*15*, *16*). As they proliferate, NMPs fuel R/C extension of the embryo, and their paraxial mesoderm or neuroectoderm progeny acquires a region-specific identity via combinatorial Hox gene expression (*17*). The human genome contains 39 Hox genes subdivided into 13 paralogous groups (*HOX1* to *HOX13*) arranged in four genomic clusters (*HOXA* to *HOXD*). Maintenance of NMP bipotentiality—and thus progressive, colinear Hox gene activation—is governed by the balance between Wnt/β-catenin, fibroblast growth factor (FGF), and retinoic acid (RA) signaling pathways (*15*, *17*–*21*). A shift toward RA signaling prompts an exit from the bipotent NMP state to the neural fate and terminates Hox gene progression, resulting in neuroepithelial progeny with a precisely restricted R/C position via their HOX "code" (*18*). Concurrent with folding of the neural plate to form the neural tube, D/V patterning is initiated by secretion of morphogens dorsally from the roof plate [bone morphogenetic proteins (BMPs) and Wnts] and ventrally from the floor plate [sonic hedgehog (Shh)] (*22*). These signals trigger concentration- and time-dependent expression of cross-repressive transcription factors that establish 11 discrete progenitor domains in the spinal cord, 5 ventral domains (p0 to p3 and pMN) and 6 dorsal domains (pd1 to pd6), broadly conferring locomotor (V0 to V3 and MN) and somatosensory (dI1 to dI6) phenotypes (*1*). Hindbrain patterning has been less extensively studied than the spinal cord, but analogous D/V populations are present, with five ventral domains (V0 to V2, 5HT, and MN) and eight dorsal domains (dA1 to dB4) distributed in a rhombomere-specific manner (*23*, *24*). Although primarily considered drivers of R/C patterning, *Hox* genes remain dynamic through D/V specification and become restricted to discrete dorsal or ventral domains that correlate with the formation of distinct neuronal subtypes (*2*, *25*–*28*).

[1]Wisconsin Institute for Discovery, University of Wisconsin-Madison, Madison, WI, USA. [2]Department of Biomedical Engineering, University of Wisconsin-Madison, Madison, WI, USA. [3]Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI, USA.
*Corresponding author. Email: rashton2@wisc.edu (R.S.A.); sroy@biostat.wisc.edu (S.R.)
†These authors contributed equally to this work.

Downloaded from https://www.science.org at University of Wisconsin Madison on February 18, 2025

Previously, we showed that hPSCs could be efficiently converted to NMPs with discrete HOX profiles along the R/C axis by temporal modulation of Wnt, FGF, and RA signaling (*29*). Here, we expand on that work to demonstrate an optimized transition from the NMP to the pCNS progenitor (pCNSP) state, enabling concentration- and time-dependent D/V patterning and rapid conversion to neurons with discrete regional phenotypes. We generated a single-cell RNA-sequencing (scRNA-seq) dataset comprising 59,502 cells that profile multiple points along the R/C and D/V axes, providing an expansive map of transcriptional programs that regulate neuronal specification. The novelty of our dataset also posed analytical challenges to neuronal characterization, as the reliance on known transcriptional markers determined from rodent development potentially excludes human-specific cell types. We established an unbiased cell population identification and characterization pipeline that identifies coarse-resolution primary clusters and fine-resolution subclusters corresponding to cell subtypes. Last, we developed a strategy to characterize regionally or phenotypically comparable populations by identifying genes that exhibit combinatorial patterns of expression across cell types. Our computational analyses revealed differences in marker expression between our hPSC-derived neurons and embryonic mouse and human spinal neurons, novel expression patterns in cardinal neurons corresponding to different R/C positions, and evidence that perturbations in progenitor patterning persistently alter postmitotic gene expression patterns. We anticipate that our modular differentiation paradigm and associated computational tools will be a valuable resource for biomanufacturing discrete, region-specific, pCNS populations, which will enable precise modeling of human development and disease as well as homologous cell grafts for regenerative medicine applications.

RESULTS

Smad inhibition optimizes conversion of NMPs to naive pCNSPs

We first evaluated whether applying a single ventral patterning schema to hPSC-derived NMPs from diverse R/C regions would enable consistent derivation of ventral neuronal phenotypes. Using our HOX protocol, we derived six different NMP cultures from human embryonic stem cells (hESCs) corresponding to 24 hours (H24), 48 hours (H48), 72 hours (H72), 120 hours (H120), 168 hours (H168), and 216 hours (H216) patterning periods in FGF8, CHIR, and/or GDF11 and dorsomorphin (fig. S1, A and B) (*29*). NMP cultures were exposed to RA and small-molecule Shh agonists Smoothened agonist (SAG) and purmorphamine (Pur) before the addition of DAPT (N-[N-(3, 5-difluorophenacetyl)-l-alanyl]-s-phenylglycinet-butyl ester), which induces rapid neuronal conversion to ventral neuron (vN) phenotypes. Samples were cryopreserved, thawed, and cultured overnight before immunocytochemistry and scRNA-seq analysis (fig. S1, B to D).

In agreement with our previous publication, cultures expressed increasingly caudal HOX paralogs that could be correlated to cervical (*HOX1-8*; H24-vN, H48-vN, and H72-vN), thoracic (*HOX1-9*; H120-vN), lumbar (*HOX1-11*; H168-vN), and lumbosacral (*HOX1-13*; H216-vN) spinal regions (figs. S1E and S4A) (*29*). We attributed the absence of hindbrain identities (expressing only *HOX1-4*) and similarity between H24, H48, and H72 *HOX* profiles to prolonged RA exposure during the neuronal differentiation stage, because RA alone is capable of caudalizing cells to a cervical fate (*30*). Notably, analysis at single-cell resolution revealed intrasample uniformity in *HOX* expression (fig. S4A). This illustrates our HOX protocol's ability to discretize the pCNS R/C axis, in contrast to the broad or heterogeneous *HOX* profiles observed in other direct differentiation protocols and organoid models (*6*, *10*, *13*, *31*).

Although our aim was to produce cultures with high *SNAP25*$^+$ neuronal content, cell type heterogeneity within and across samples was apparent by staining (fig. S1C) and sample (fig. S1D), cluster (fig. S1F), and gene expression (fig. S1, G and H) distributions on *t*-distributed stochastic neighbor embedding (*t*-SNE) visualizations of single-cell transcriptomic data. Neural progenitor (*SOX2*$^+$) and neuron (*SNAP25*$^+$) composition varied between 10 and 80% (fig. S1, I and J). Thus, while samples could be patterned to discrete regions on the R/C axis, direct application of ventral morphogens caused inconsistent neuronal differentiation across different NMP populations.

We hypothesized that consistent neuronal differentiation from NMPs first requires efficient induction to SOX2$^+$/PAX6$^+$ pCNSPs, akin to the formation of neural plate epithelium from tail bud progenitors during gastrulation (*15*). This process is regulated by RA and Noggin (a BMP antagonist) secreted by the somites and notochord (*18*, *32*). We derived SOX2$^+$/BRACHYURY$^+$ H120 NMPs (fig. S2A) and then exposed them to RA and/or small-molecule Smad inhibitors (SB + LDN) for up to 3 days (H120-pCNSPs, fig. S2B). Both RA and SB + LDN were required to generate SOX2$^+$/PAX6$^+$ H120-pCNSPs efficiently (fig. S2, C to Z). In the absence of one or both factors, we observed persistent PAX3$^+$ and PAX7$^+$ cells that could become mesodermal (PAX3$^+$/PAX7$^+$), myogenic (PAX3$^+$) (fig. S2, C to T), or neural crest (SOX10$^+$) progeny (fig. S2, GG to LL). Both factors were also required to prevent inadvertent dorsal (PAX6$^+$/PAX3$^+$/PAX7$^+$; AP2a$^+$) (fig. S2, O to T and GG to LL), intermediate (PAX6$^+$/PAX3$^+$) (fig. S2, O to T), or ventral (NKX6.1$^+$) (fig. S2, AA to FF) patterning. Thus, RA and SB + LDN cooperate to repress PAX3 and PAX7, which allows for the conversion of NMPs to unbiased, naive pCNSPs for subsequent D/V patterning.

Concentration-dependent differentiation of pCNSPs along D/V axis

To simplify derivation of diverse pCNSPs with precise R/C positioning, we wanted to use the same ventralizing or dorsalizing differentiation schema for all cultures. Ventral interneurons (INs) and MNs arise in response to graded Shh signaling in the developing neural tube (Fig. 1A) (*22*). Thus, we first sought to determine whether hPSC-derived pCNSPs could be efficiently patterned to ventral identities in a concentration-dependent manner. We patterned H120-pCNSPs for 4 days in either 100 nM or 1 μM RA containing SB + LDN and varying concentrations of SAG and Pur to generate ventral progenitor cultures (Fig. 1B). Sustained exposure to SB + LDN suppressed PAX3 and PAX7 expression (Fig. 1, C to F), while Shh signaling caused concentration-dependent increases in ventral progenitor markers (Fig. 1, A and G to M). Notably, reducing the concentration of RA during ventral patterning improved the potency of Shh signaling, resulting in significant increases in *NKX6.1*, *OLIG2*, and *NKX2.2* expression under optimal culture conditions (Fig. 1, K to M). Exposing ventral progenitor cultures to DAPT for 5 days induced rapid neuronal differentiation (Fig. 1B) and appropriately stratified postmitotic INs and MNs (Fig. 1, A and N to Y).

Efficient dorsal patterning of pCNS neurons in vitro has historically been difficult because of the ubiquitous roles of BMPs and Wnts elsewhere in the developing embryo. There has also been debate whether BMPs perform as morphogens (concentration dependent)

Downloaded from https://www.science.org at University of Wisconsin Madison on February 18, 2025

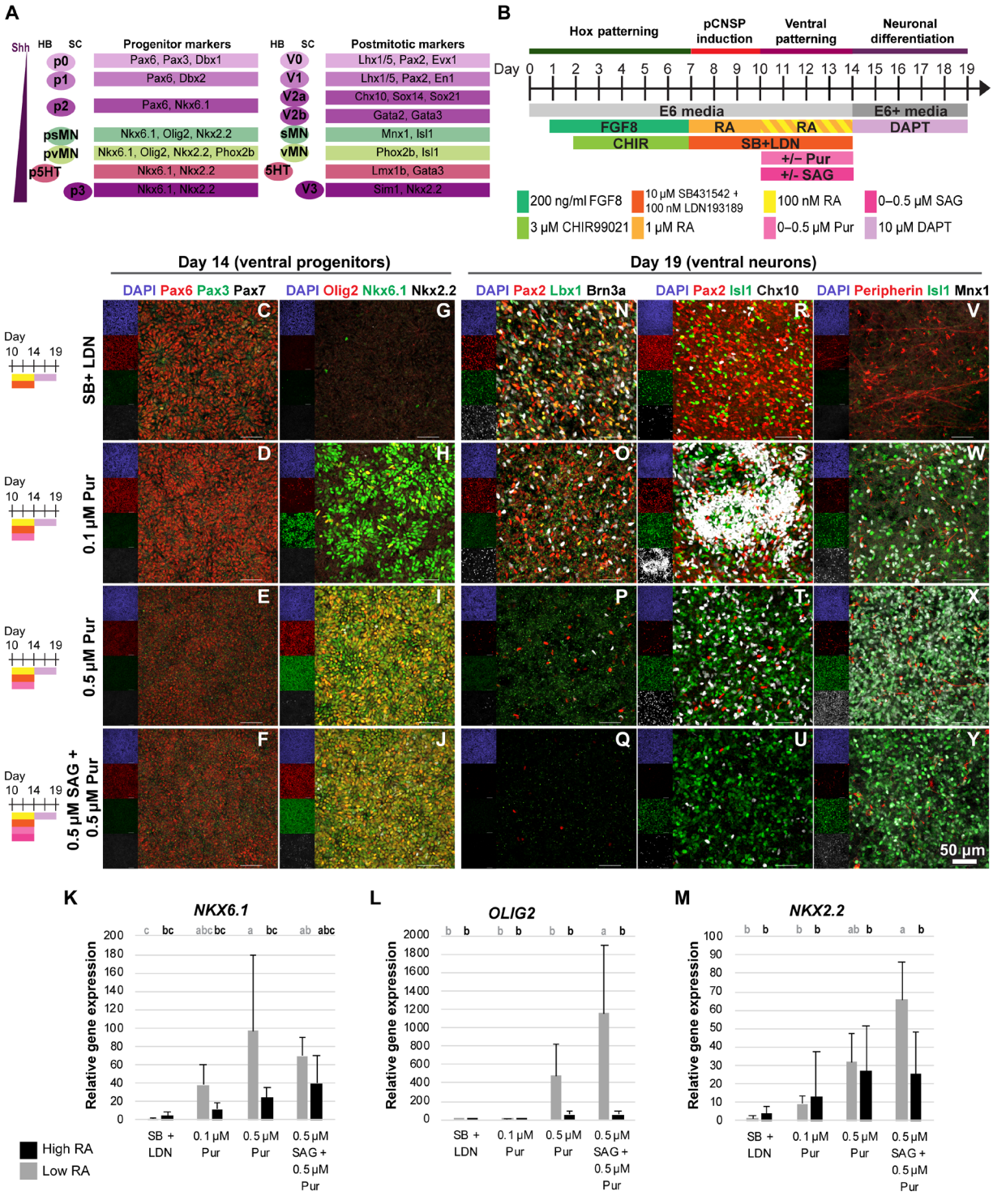

Fig. 1. Concentration-dependent Shh patterning of ventral spinal neurons. (**A**) Ventral pCNS populations, with characteristic progenitor and postmitotic transcription factor markers for the hindbrain (HB) and spinal cord (SC). (**B**) Timeline of ventral differentiation from H120-NMPs. (**C** to **F**) Immunostaining shows that cultures are uniformly PAX6$^+$/PAX3$^-$/PAX7$^-$, indicative of ventral progenitor domains p1 to p3. (**G** to **J**) As SHH agonist concentration increases, cultures shift from (G) PAX6$^+$ (p0/p1) to (H) NKX6.1$^+$ (p2) to (I) NKX6.1$^+$/OLIG2$^+$ (pMN) to (J) NKX6.1$^+$/OLIG2$^+$/NKX2.2$^+$ (pvMN/p3). (**K** to **M**) qRT-PCR in day 14 progenitor cultures. Error bars represent SD ($n = 6$ biological replicates per condition). Data shown as relative gene expression compared to 100 nM RA SB + LDN condition. Statistics were calculated by one-way analysis of variance (ANOVA) with Tukey-Kramer post hoc. Significance for the multiple pairwise comparisons is summarized through the connecting letters report, whereby samples with different letters are significantly different by at least $P < 0.05$ (*79*). (**N** to **Y**) Immunostaining in day 19 postmitotic neurons. As SHH agonist concentration increases, cultures shift from (N, R, and V) LBX1$^+$ (dI4 to dI6) and PAX2$^+$ (V0 and V1) to (O, S, and W) CHX10$^+$ (V2a) to (P, T, and X) MNX1$^+$/ISL1$^+$ (sMN) to (Q, U, and Y) MNX1$^-$/ISL1$^+$ (vMN). Scale bars, 50 μm. Subpanels separate 358 nm (blue), 555 nm (red), 488 nm (green), and 647 nm (white) fluorochrome channels.

Downloaded from https://www.science.org at University of Wisconsin Madison on February 18, 2025

or act deterministically (type of BMP defines phenotype) (*7*, *33*, *34*), complicating efforts toward a streamlined differentiation strategy. To investigate this question, we cultured H120-pCNSPs for 4 days in either 100 nM or 1 μM RA containing cyclopamine (Cyc)—an Shh antagonist—and varying concentrations and exposure durations of BMP4 to generate dorsal progenitor cultures (Fig. 2B). Termination of SB + LDN during dorsal patterning released suppression of PAX3 and PAX7 activity (Fig. 2, C to F), which were elevated in response to increased BMP signaling (Fig. 2, L and M). This occurred without significant changes in *PAX6* expression, indicating the cells' maintenance of a CNS identity (Fig. 2K). Quantitative real-time polymerase chain reaction (qRT-PCR)–assessed gene expression patterns also indicated a shift from intermediate to dorsal fates with BMP4 exposure (Fig. 2, A, N, and O). While we observed OLIG3$^+$ pd1/pd2/pd3 progenitors and the formation of some AP2α$^+$ roof plate cells under conditions with the highest BMP4 exposure, no SOX10$^+$ neural crest progeny was present (Fig. 2J). Treatment with DAPT induced rapid neuronal differentiation (Fig. 2B) and appropriately stratified postmitotic dorsal INs in direct correlation to BMP4 exposure concentration/duration (Fig. 2, A and P to AA). This indicates that BMP4 behaves as a morphogen in agreement with other recent findings (*6*). Furthermore, because BMP7 has been shown to be required for neurogenesis of dI1/dI3/dI5 INs (*34*), we wanted to determine whether adding BMP7 during the neuronal differentiation phase could push progenitors toward more dorsal postmitotic fates (Fig. 3A). Using BMP7 treatment, we observed a shift from dI4/dI5/dI6 to dI2/dI3 INs (Figs. 2A and 3, B, D, and F) and from dI2/dI3 to dI1/dI2 INs (Figs. 2A and 3, C, E, and G) in progenitors pulsed or maintained in BMP4 (20 ng/ml) over the dorsal patterning period, respectively. Collectively, the results demonstrate that our differentiation schema generates the full spectrum of D/V cell types from a single R/C position (H120), with the ability to obtain desired subtypes by optimizing morphogen exposure within discrete time frames (Fig. 3H).

Single-cell transcriptomes reveal differential population distributions after combined R/C and D/V patterning

We generated an expansive scRNA-seq dataset comprising dorsal and ventral populations differentiated from six NMP time points (H24, H48, H72, H120, H168, and H216) (Fig. 4, A and B, and figs. S3, A to C, and S4B). For dorsal differentiation, pCNSPs were exposed to 100 nM RA and Cyc and pulsed with BMP4 (20 ng/ml) during the 4-day progenitor patterning period (fig. S3, A and D). For ventral differentiation, pCNSPs received 100 nM RA, SB + LDN, and 0.5 and 0.5 μM Pur (fig. S3, B and E). In addition, in the D/V patterning stage, pCNSPs at H216 were patterned with either 1 μM (H216R) or 100 nM RA (H216) to determine whether RA further affects caudalization (fig. S3, A and B). After DAPT treatment, the resulting samples were near homogeneously neuronal (85 to 98% *SNAP25*$^+$), with trace *SOX2*$^+$ floor plate (*SHH*$^+$) and roof plate (*LMX1A*$^+$) cells and minimal expression of markers from other cell lineages, thereby demonstrating the efficiency of our modular differentiation methodology (Fig. 4C).

Analysis of *HOX* expression across all samples indicated discretization along the R/C axis such that dorsal and ventral cultures derived from the same NMPs showed globally similar *HOX* expression (Fig. 4, D and E). Compared to the previously presented scRNA-seq dataset (fig. S1), samples were rostrally shifted (fig. S4, A and B). This is likely a consequence of using SB + LDN during the pCNSP induction stage, which recapitulates the role of Noggin to abruptly terminate HOX progression in NMPs (*32*). Notably, increased RA concentration did not result in activation of more caudal *HOX* paralogs in H216R compared to H216 samples (fig. S4B), confirming that R/C patterning occurs during NMP and pCNSP differentiation and is independent of RA concentration during D/V patterning. However, increased RA under H216R conditions did cause elevated expression of *HOXB8* and *HOXA5* in dorsal and ventral samples, respectively (Fig. 4, D and E, and fig. S4B), suggesting that RA may continue to play a role in neuronal subtype specification (*2*, *25*, *27*).

We visualized simultaneous expression of all *HOX* genes by clustering (Materials and Methods; Fig. 4, F and G; and fig. S4C), which revealed inter- and intrasample *HOX* profile heterogeneity, with caudal samples unexpectedly inclusive of rostral *HOX* profile clusters (clusters 1 to 7) (Fig. 4G). These "mismatched" *HOX* profile clusters were associated with different phenotypes—including MNs and dI1, dI2, and dI3 INs (Figs. 4, F and G, and 5, A to C)—indicating neuronal subtype-specific *HOX* gene stratification in accordance with findings in vivo (*2*, *25*, *27*). Thus, while *HOX* genes can be used globally to assess a sample's R/C positional identity, nuances in *HOX* gene expression profiles of hPSC-derived neuronal subtypes also emerge with cell maturity and D/V specification. Our differentiation methodology may thus be used to explore how HOX dynamics influence pCNS neuronal specification and circuit organization (*2*).

We next assessed differentiation efficiency to various cardinal cell types. Dorsal and ventral samples showed gene expression patterns associated with appropriate transcriptional markers (Fig. 4H). The dataset was sparse in intermediate cardinal neurons corresponding to V0 (*EVX1*), V1 (*EN1*), and dI6 (*DMRT3*) INs, a consequence of using patterning conditions that yielded few DBX1$^+$ progenitors (fig. S3, D and E). When we specifically examined cardinal cell type distributions defined by nonoverlapping combinatorial transcription factor expression (table S1A), we observed increasingly dorsal or ventral character as samples were caudalized. For example, dorsal H24-dN, H48-dN, and H72-dN samples corresponding to hindbrain-rostral cervical spinal cord were ventrally shifted toward *LBX1*$^+$ dI4/dI5/dI6 INs compared to H120-dN, H168-dN, and H216-dN samples, which included primarily dI1/dI2/dI3 INs (Fig. 4, H and I). Similarly, ventral H24-vN, H48-vN, and H72-vN samples were dorsally shifted to *CHX10*$^+$ V2a and *GATA2/3*$^+$ V2b INs compared to H120-vN, H168-vN, and H216-vN samples, which had a greater proportion of MNs (Fig. 4, H and I). Given our application of a consistent D/V patterning protocol, these data suggest inherent differences in region-specific NMP differentiation potential. Moreover, although increased RA did not contribute to caudalization during progenitor patterning (Fig. 4, D and E, and fig. S4B), higher RA exposure in H216R samples caused a shift toward more intermediate cell types compared to H216 samples (Fig. 4I), reaffirming our previous observation that RA modulates morphogen potency and is involved in neuronal fate determination.

Unbiased clustering isolates cardinal cell types

Although transcription factors that define pCNS cardinal cell types are generally conserved during evolution (*24*, *35*), they could exclude potential species-specific or region-specific differences unique to our dataset. For example, if only cells expressing known cardinal markers are analyzed, 15 to 50% of cells across our samples would remain uncharacterized (Fig. 4I and table S1A). Therefore, we

Downloaded from https://www.science.org at University of Wisconsin Madison on February 18, 2025

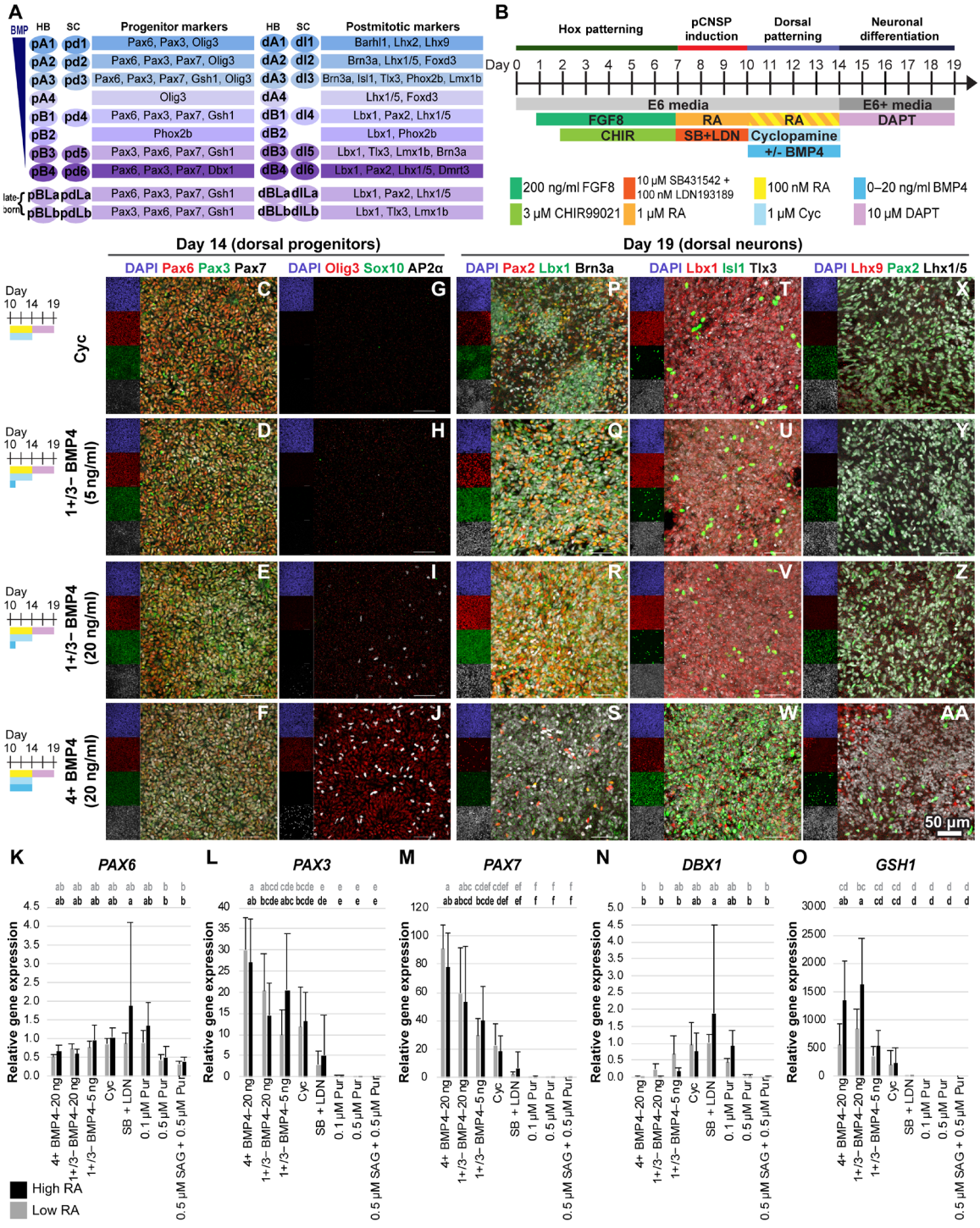

Fig. 2. Concentration- and time-dependent BMP4 patterning of dorsal spinal neurons. (**A**) Dorsal pCNS populations, with characteristic progenitor and postmitotic transcription factor markers for the hindbrain and spinal cord. (**B**) Timeline of dorsal differentiation from H120-NMPs. (**C** to **J**) Immunostaining shows that as BMP4 concentration and duration increase, cultures shift from (C and G) PAX6$^+$/PAX3$^+$ (p0) to (D, H, E, and I) PAX6$^+$/PAX3$^+$/PAX7$^+$ (pd4 to pd6) to (F and J) OLIG3$^+$ (pd1 to pd3). AP2α^+ roof plate cells, but not SOX10$^+$ neural crest progeny, were present at the highest concentrations used. (**K** to **O**) qRT-PCR in day 14 progenitors. Error bars represent SD ($n = 6$ biological replicates per condition). Data shown as relative gene expression compared to 100 nM RA SB + LDN condition. Statistics were calculated by one-way ANOVA with Tukey-Kramer post hoc. Significance for the multiple pairwise comparisons is summarized through the connecting letters report, whereby samples with different letters are significantly different by at least $P < 0.05$ (*79*). (**P** to **AA**) Immunostaining in day 19 postmitotic neurons. As BMP4 concentration and duration increase, cultures shift from (P, T, and X) LBX1$^-$/PAX2$^+$ (V0 and V1), LBX1$^+$/PAX2$^+$/LHX1$^+$ (dI6), LBX1$^+$/BRN3A$^+$/TLX3$^+$ (dI5) to (Q, U, Y, R, V, and Z) predominantly LBX1$^+$/PAX2$^+$/LHX1$^+$ (dI4) and LBX1$^+$/BRN3A$^+$/TLX3$^+$ (dI5) to (S, W, and AA) BRN3A$^+$/ISL1$^+$/TLX3$^+$ (dI3) and BRN3A$^+$/PAX2$^-$/LHX1$^+$ (dI2). Scale bars, 50 μm. Subpanels separate 358 nm (blue), 555 nm (red), 488 nm (green), and 647 nm (white) fluorochrome channels.

Downloaded from https://www.science.org at University of Wisconsin Madison on February 18, 2025

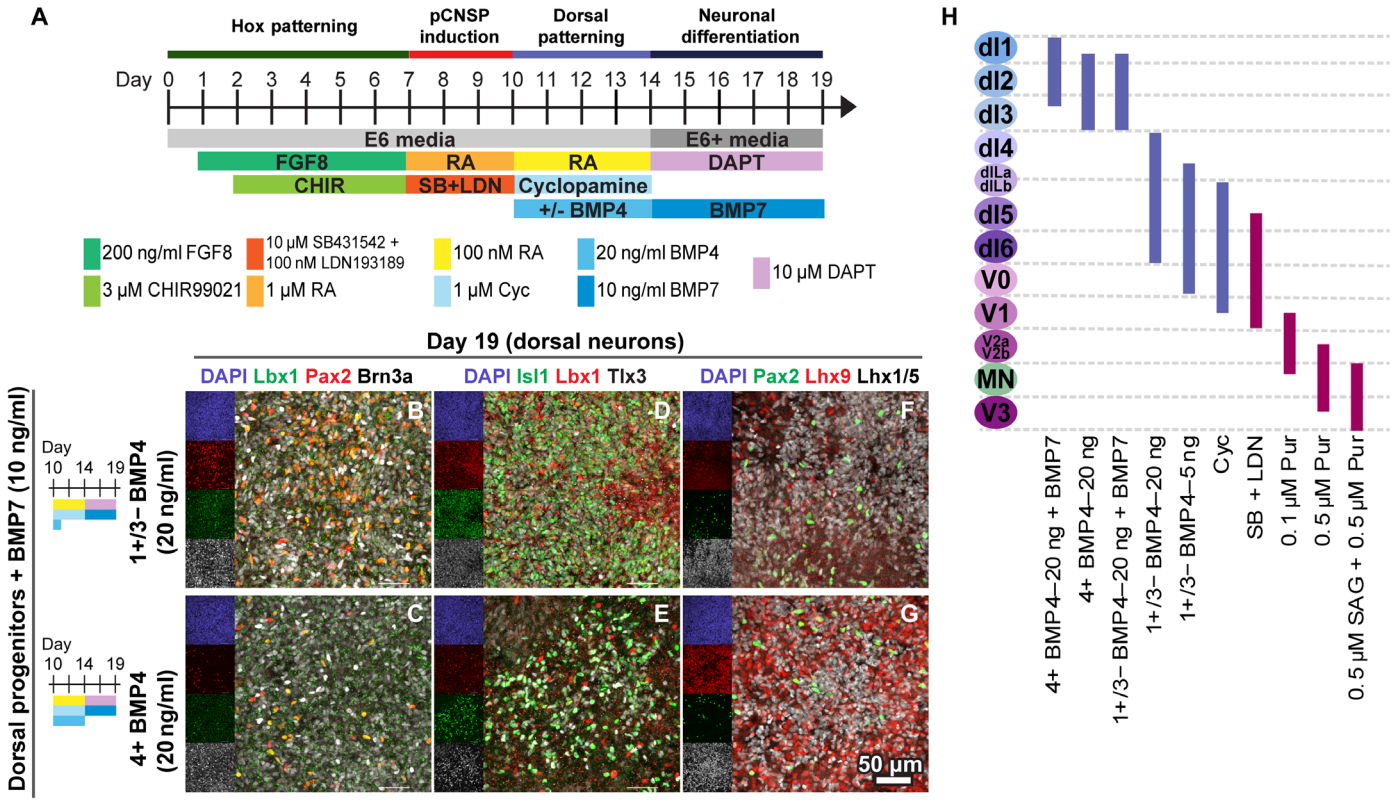

Fig. 3. Addition of BMP7 during neuronal differentiation further dorsalizes postmitotic population. (**A**) Timeline of dorsal differentiation from H120-NMPs, with BMP7 added during neuronal differentiation phase from days 14 to 19. (**B** to **G**) Immunostaining in day 19 postmitotic cultures shows that DAPT treatment rapidly converts progenitors to dorsally shifted postmitotic phenotypes compared to Fig. 2. With BMP7, cultures shift from (B, D, and F) BRN3A$^+$/ISL1$^+$/TLX3$^+$ (dI3) and BRN3A$^+$/PAX2$^-$/LHX1$^+$ (dI2) to (C, E, and G) LHX9$^+$ (dI1). (**H**) Schematic of differentiation conditions corresponding to postmitotic cardinal cell types. Subpanels separate 358 nm (blue), 555 nm (red), 488 nm (green), and 647 nm (white) fluorochrome channels.

applied a clustering method based on sparse nonnegative matrix factorization (NMF) (*36*) to define 25 "primary clusters" (Materials and Methods and Fig. 5A). We assigned clusters to hindbrain or spinal cord based on sample identities' global *HOX* expression (Fig. 4, D and E) and assessed the composition of cardinal neurons in each cluster based on expression of known markers (Fig. 5, A to C). To determine how our hPSC-derived clusters compared to in vivo neuronal populations, we performed a correlation analysis against recently available embryonic human (*35*) and mouse (*37*) neural tube scRNA-seq datasets across multiple gestational time points (Fig. 5D, fig. S5, table S2, and Materials and Methods). Both of these datasets relied on strict transcriptional definitions to define their cardinal neurons, a consequence of the sparsity of cells available for adequate clustering. Despite disparate approaches to cell type identification, we observed good similarity (Pearson correlation coefficient > 0.5) between our clusters and the neuronal populations defined by Rayon *et al.* (*35*) or Delile *et al.* (*37*) using either known markers or their sets of annotated transcription factors (Fig. 5D, top, and fig. S5). A direct comparison between cell types defined by known markers found similar concordance between in vitro and in vivo cell types as between the in vivo mouse and in vivo human cell types (table S2 and fig. S5C) (*35*, *37*). Thus, we validated that our hPSC-derived populations were comparable to human and mouse neurons in vivo.

We then determined at what stage of in vivo development our hPSC-derived populations might belong by comparing our clusters to the Carnegie stage–matched (CS12 to CS19) or mouse embryonic day–matched (E9.5 to E13.5) samples from these studies (Fig. 5E, top). Dorsal clusters (C1 to C10 and C25) showed higher correlation to samples from CS17 and CS19 (gestational days 42 to 51), compared to ventral clusters (C11 to C18 and C20 to C23), which showed similarity to samples from CS12 and CS14 (gestational days 26 to 35). A comparable trend was observed in the mouse data. These correlation patterns were in accordance with the sequential emergence of ventral and dorsal neurons in vivo, wherein ventral populations are patterned earlier in development than dorsal populations (*24*, *35*). It is notable that our hPSC-derived neurons were derived in parallel and in fewer than 38 days, which suggests an accelerated differentiation of cells in vitro compared to endogenous populations. Together, these findings validate our data-driven approach to characterizing pCNS neuron diversity and present an opportunity to detect novel neuronal markers otherwise obscured by a priori transcriptional definitions or limited by the availability of embryonic tissue.

Differential gene expression in subpopulation analysis of primary clusters

While many primary clusters comprised a single cardinal population, others were made up of closely related cell types. For example, C9 and C25 included multiple inhibitory and excitatory LBX1$^+$ populations (dI4/dI5/dI6), and C14 contained both CHX10$^+$ V2a INs and SIM1$^+$ V3 INs, which are both glutamatergic ventral INs

Downloaded from https://www.science.org at University of Wisconsin Madison on February 18, 2025

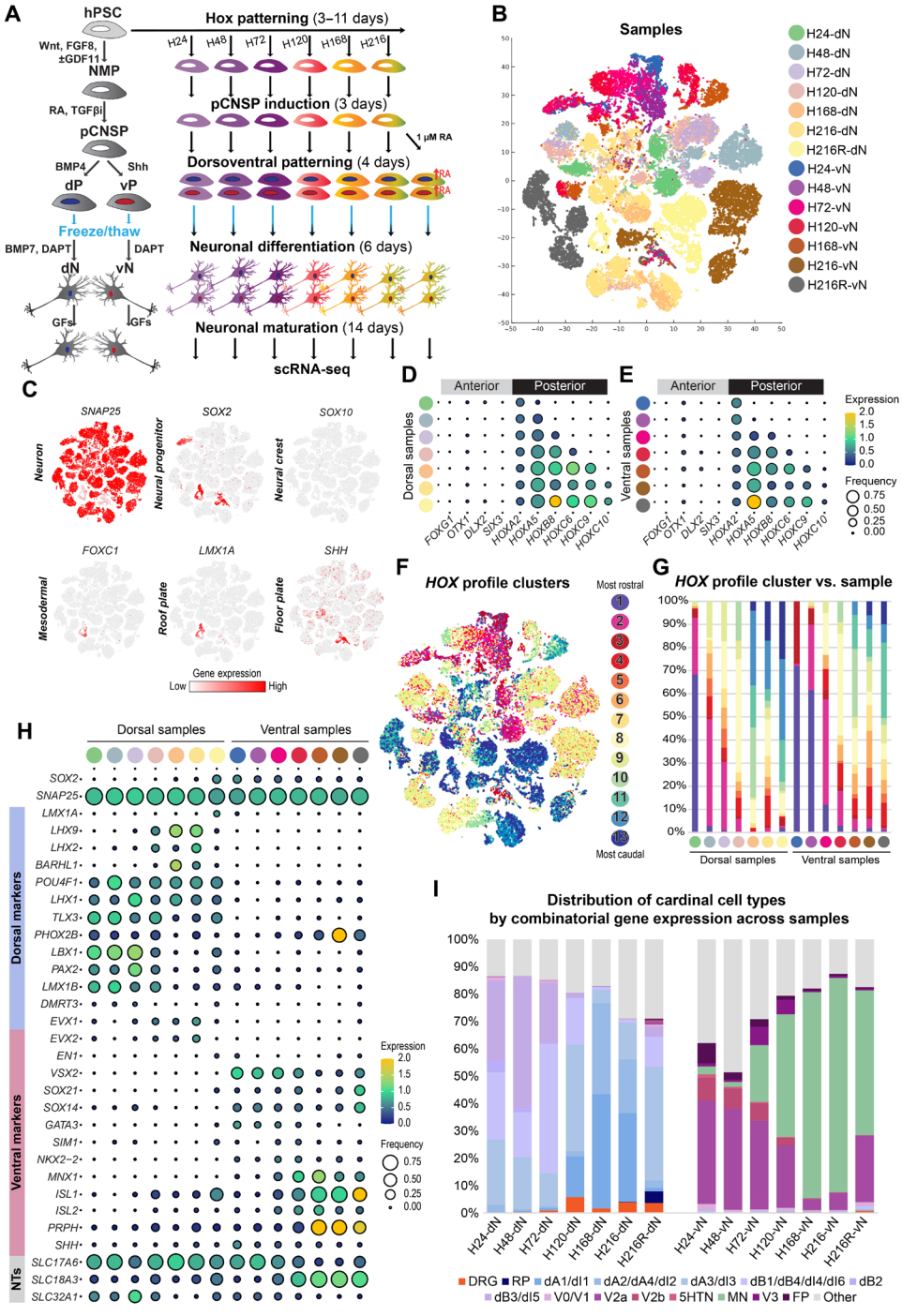

Fig. 4. scRNA-seq characterization of dorsal and ventral samples differentiated from discrete regions along R/C axis. (**A**) Timeline of differentiation from region-specific NMPs, to discrete pCNSPs, dorsal (dP) and ventral (vP) progenitors, and postmitotic neurons. (**B**) *t*-SNE plot with seven dorsal samples and seven ventral samples ($n = 46{,}959$ cells). (**C**) *t*-SNE heatmaps showing highly neuronal (*SNAP25*$^+$) cells, with few neural progenitor (*SOX2*$^+$/*SNAP25*$^-$), mesoderm (*FOXC1*$^+$), or neural crest (*SOX10*$^+$) cells. *SOX2*$^+$ progenitors are primarily floor plate (*SHH*$^+$) and roof plate (*LMX1A*$^+$). (**D** and **E**) Dot plot displaying genes associated with anterior or pCNS identity across dorsal (D) and ventral (E) samples. The size of each circle reflects the fraction of cells where the gene is detected, and the color reflects the average expression level within each cluster (blue, low expression; yellow, high expression). (**F**) *t*-SNE plot showing *HOX* profile clusters. (**G**) Distribution of *HOX* profile clusters across samples. (**H**) Dot plot displaying genes associated with dorsal or ventral neuronal phenotypes. (**I**) Distribution of cardinal pCNS neurons, peripheral dorsal root ganglion (DRG) neurons, floor plate (FP), and roof plate (RP) cells as defined by nonoverlapping combinatorial transcription factor expression across dorsal and ventral samples. "Other" includes cells that were not classified by the knowledge matrix in table S1.

Downloaded from https://www.science.org at University of Wisconsin Madison on February 18, 2025

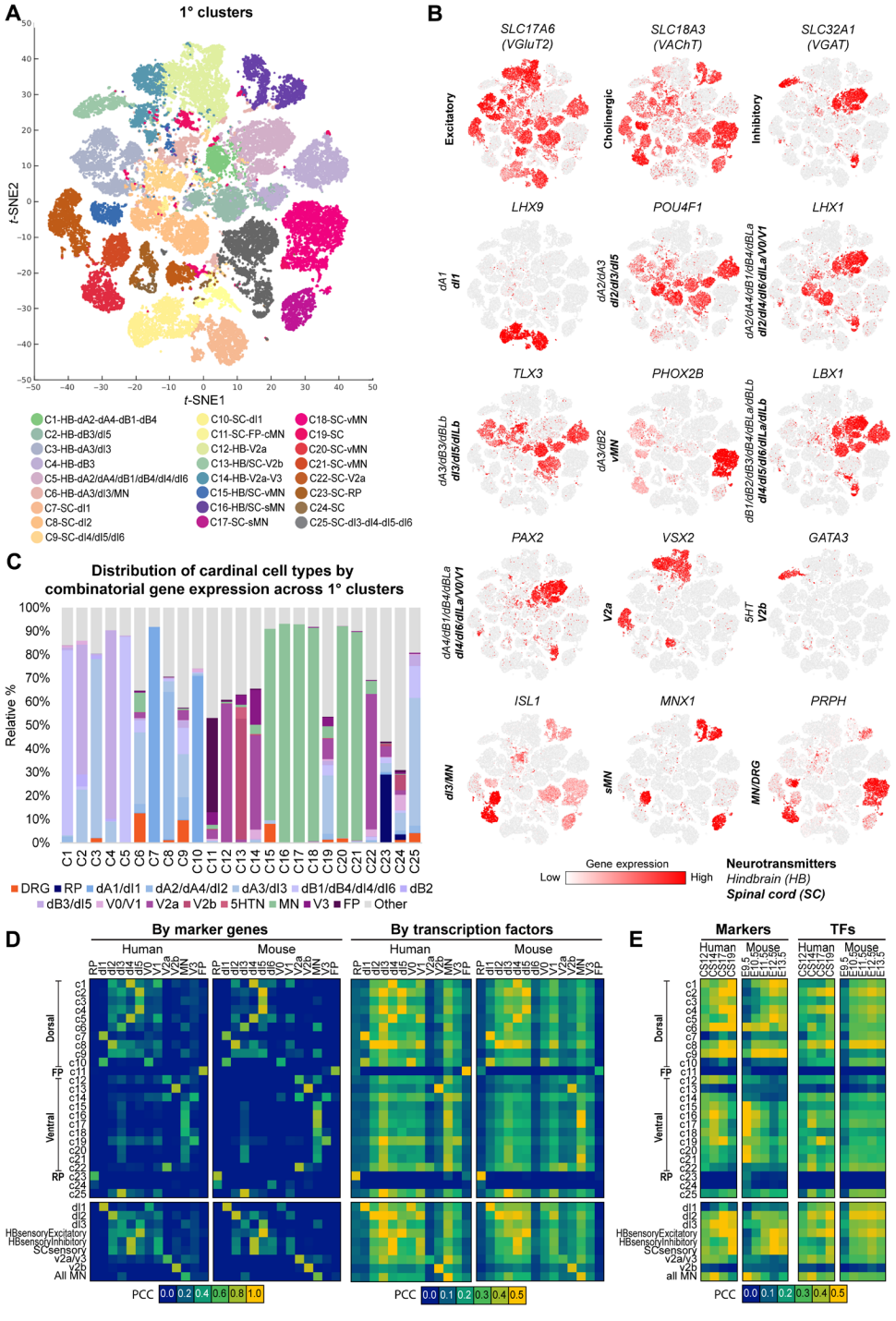

Fig. 5. Unbiased NMF-based clustering validated by global similarities between hPSC-derived and in vivo scRNA-seq datasets. (**A**) *t*-SNE plot with 25 primary clusters broadly divides our dataset by pCNS cardinal neuron identity. Legend labels indicate whether the cluster was presumed to be from the hindbrain (HB) or spinal cord (SC). (**B**) *t*-SNE heatmaps showing that characteristic cardinal transcription factors (hindbrain or spinal cord) and neurotransmitters (indicated by text legend) map to primary clusters. (**C**) Distribution of cardinal pCNS neurons, peripheral dorsal root ganglion (DRG) neurons, floor plate (FP), and roof plate (RP) cells as defined by nonoverlapping combinatorial transcription factor expression across primary clusters. "Other" includes cells that were not classified by the knowledge matrix in table S1. (**D** and **E**) Heatmap of Pearson's correlation coefficient (PCC) values matrix comparing primary clusters (top) or 9 of the 17 subpopulation groupings (bottom) against in vivo human (*36*) and mouse (*37*) embryonic (D) cardinal cell types and (E) developmental stages. Marker genes ($n = 77$ for human; $n = 55$ for mouse) were defined by the knowledge matrix provided in (*35*, *37*), and transcription factors ($n = 1463$ for human; $n = 1775$ for mouse) were defined by annotations from PANTHER and GO.

Downloaded from https://www.science.org at University of Wisconsin Madison on February 18, 2025

(Fig. 5, A to C). We organized related primary clusters into 17 different groups (Fig. 6A and table S3A), which also exhibited good concordance with previous datasets (Fig. 5D, bottom, and fig. S5). We then developed and applied a consensus clustering-based approach with the goal of defining robust subclusters representing subtypes of known cardinal populations (Materials and Methods and fig. S7, A to C). Consensus clustering was shown to improve cluster stability without sacrificing cluster quality (fig. S7D and Materials and Methods). Each subpopulation was divided into four to nine "subclusters," to which we assigned an R/C positional identity based on sample identities' global *HOX* expression (Figs. 4, D and E, and 6B). We examined the relatedness of these subclusters using hierarchical clustering and found that subclusters were generally organized by region (Fig. 6B). Analysis of differentially expressed genes (DEGs) across subclusters uncovered hundreds of genes up-regulated in region-specific cardinal subtypes (table S4 and Materials and Methods). Here, we focus on the MN, dI1, and V2a/V3 groups, highlighting findings that emphasize how unbiased clustering enables the discovery of novel region-specific markers different from or difficult to detect in transcriptionally defined populations (fig. S8A) and in available in vivo datasets (fig. S8, B to D). Similar analyses for other groups are available [table S4 and online resource (see "Data and materials availability" in the Acknowledgements)].

MNs constitute the most widely studied neurons in the spinal cord, with significant evidence of Hox-dependent specification along the R/C axis and in the development of precise motor pools (*2*, *20*, *38*–*42*). MNs in our dataset clustered into *MNX1*$^+$/*ISL2*$^+$ somatic MNs (sMNs; MN-*c1*, MN-*c3*, MN-*c7*, and MN-*c8*), which innervate skeletal muscle, and preganglionic *PHOX2B*$^+$ visceral MNs (vMNs; MN-*c2*, MN-*c5*, MN-*c6*, and MN-*c9*), which are responsible for autonomic function (Fig. 6C) (*42*–*44*). The latter population is a particularly rich target for novel findings. For example, Rayon *et al.* (*35*) report scarce expression of *TBX20* in the human spinal cord as a notable difference between human and mouse vMNs, but hPSC-derived *PHOX2B*$^+$ vMN subclusters in our dataset clearly express *TBX20* (Fig. 6C and fig. S8B). This suggests that *TBX20* is conserved between species and that our in vitro hPSC-derived dataset can help validate or invalidate conclusions based on sparser in vivo datasets. *LINC00682* also emerged as a characteristic marker of vMNs in our dataset but was up-regulated in only p3 progenitors using the classification scheme developed by Rayon *et al.* (*35*). Reassessment of the published dataset using our transcriptional definitions (table S1A and fig. S8A) revealed that *LINC00682* is also abundant in human vMNs in vivo (fig. S8B). Although poorly understood, lincRNAs are abundant in the CNS and play multiple roles in development, neural plasticity, neurodegeneration, and sex-specific disease phenotypes (*45*–*47*). Given the origin of vMNs on the pMN/p3 border (*24*), we hypothesized that *LINC00682* could be an important regulator for vMN specification. Knockdown of *LINC00682* during differentiation repressed *PHOX2B*, but not *MNX1* or *ISL1* expression, suggesting vMN-specific fate regulation and affirming the value of these data for future gene regulatory network analysis (fig. S8, E to G). In addition, although sMNs and vMNs were proportionally divided within samples (Figs. 4H and 6C), *HOX* profiles appeared hindbrain-like in vMNs but sMNs maintained spinal *HOX* profiles correlative to their sample identity. *Phox2B* is known to be a direct target of several *Hox* genes (*2*, *48*) and may contribute to this differential expression. Given that the persistence of Hox activity is thought to coincide with the development of downstream synaptic targets (*2*, *38*), it is also possible that the proximity of ganglionic targets, both spatially and developmentally, causes early down-regulation of *Hox* gene expression in vMNs compared to sMNs, a subject for future investigation.

The dI1 IN population is derived from the dorsal-most progenitor domain of the spinal cord (Fig. 2A) and migrates to the deep dorsal horn, where they have roles in proprioception (*49*, *50*). DI1 subclusters were divided into ipsilateral-projecting dI1i (*BARHL1/2*$^+$; dI1-*c2* and dI1-*c4*) (*51*, *52*) and contralateral-projecting dI1c (*LHX2*$^+$; dI1-*c1*, dI1-*c3*, and dI1-*c5*) (*53*) subtypes (Fig. 6D). The *LHX2*$^+$ population also strongly expressed *EVX1/2*, which classically identify V0 INs (*24*, *54*, *55*). In contrast to our dataset, *EVX1/2* are not expressed in dI1 INs in mouse or human scRNA-seq data (fig. S8C). Furthermore, while hPSC-derived dI1 cells uniformly expressed *LHX9* (*50*, *56*), they seldom coexpress *POU4F1* (Figs. 3, F and G, and 5B), which is characteristic in mouse (*1*, *37*) but may not be a consistent feature of the human population (*24*, *35*). *HOX* profiles appeared hindbrain-like in the *LHX2*$^+$/*EVX1*$^+$ population, but not the *BARHL1/2*$^+$ population, which showed persistent caudal *HOX* profiles, despite comparable sample compositions (Fig. 6D). In particular, *Evx1* is regulated by *Hox2* paralogs (*2*, *57*), so its expression may suggest a potential role for *Hox* genes in gene regulatory pathways responsible for ipsilateral/contralateral projection patterns in dI1 neurons.

CHX10$^+$ (*VSX2*$^+$) V2a INs have multiple roles in locomotor coordination and breathing (*58*–*60*) and are one of the few spinal interneuron populations to have been characterized by spinal segment (*26*). Of the cardinal populations in our dataset, the V2a INs show the best continuous representation throughout all ventral samples (Fig. 4I). They also express *HOX* profiles commensurate with their sample identities (Fig. 6E). Region-specific markers associated with hindbrain (V2a/3-*c2*, V2a/3-*c5*, V2a/3-*c6*, and V2a/3-*c7*) or spinal V2a INs (V2a/3-*c1* and V2a/3-*c3*) were apparent in subcluster DEGs, but these differed from markers identified with scRNA-seq by Hayashi *et al.* This is likely because that dataset comprised fluorescence-activated cell sorting (FACS)–sorted Chx10:tdTomato$^+$ cells in p0 mouse cervical and lumbar tissues, which are developmentally advanced compared to our hPSC-derived cells (*26*). Spinal V2a INs in our dataset also expressed *SLC18A3* (*VACHT*) and *LINC02303*, which were not observed in comparable mouse or human scRNA-seq data (fig. S8D), although cholinergic character in V2a INs has previously been observed in zebrafish (*61*). Last, hindbrain V2a INs atypically expressed *NK1R* (*TACR1*) (Fig. 6E), which is normally expressed in pre-Botzinger complex (pre-BotC) respiratory neurons (*62*) and dorsal horn neurons (*63*, *64*). V2a INs in the rodent hindbrain are adjacent to the pre-BotC but do not express NK1R (*60*), indicating a potential species-specific difference in rhythmic breathing organization.

Together, differences emergent in our hPSC-derived scRNA-seq dataset reveal the power of this differentiation platform to generate novel, region-specific spinal subpopulations detectable by standard DEG analysis. Whether novel markers are bona fide, evidence of accelerated maturation of cells in vitro compared to in vivo (*65*) or artifacts of in vitro differentiation is subject to future investigation.

Arboretum analysis reveals complex gene expression patterns across subclusters

While standard DEG analysis identifies strong differences between subclusters, it is restricted to pairwise comparisons. Combinatorial

Downloaded from https://www.science.org at University of Wisconsin Madison on February 18, 2025

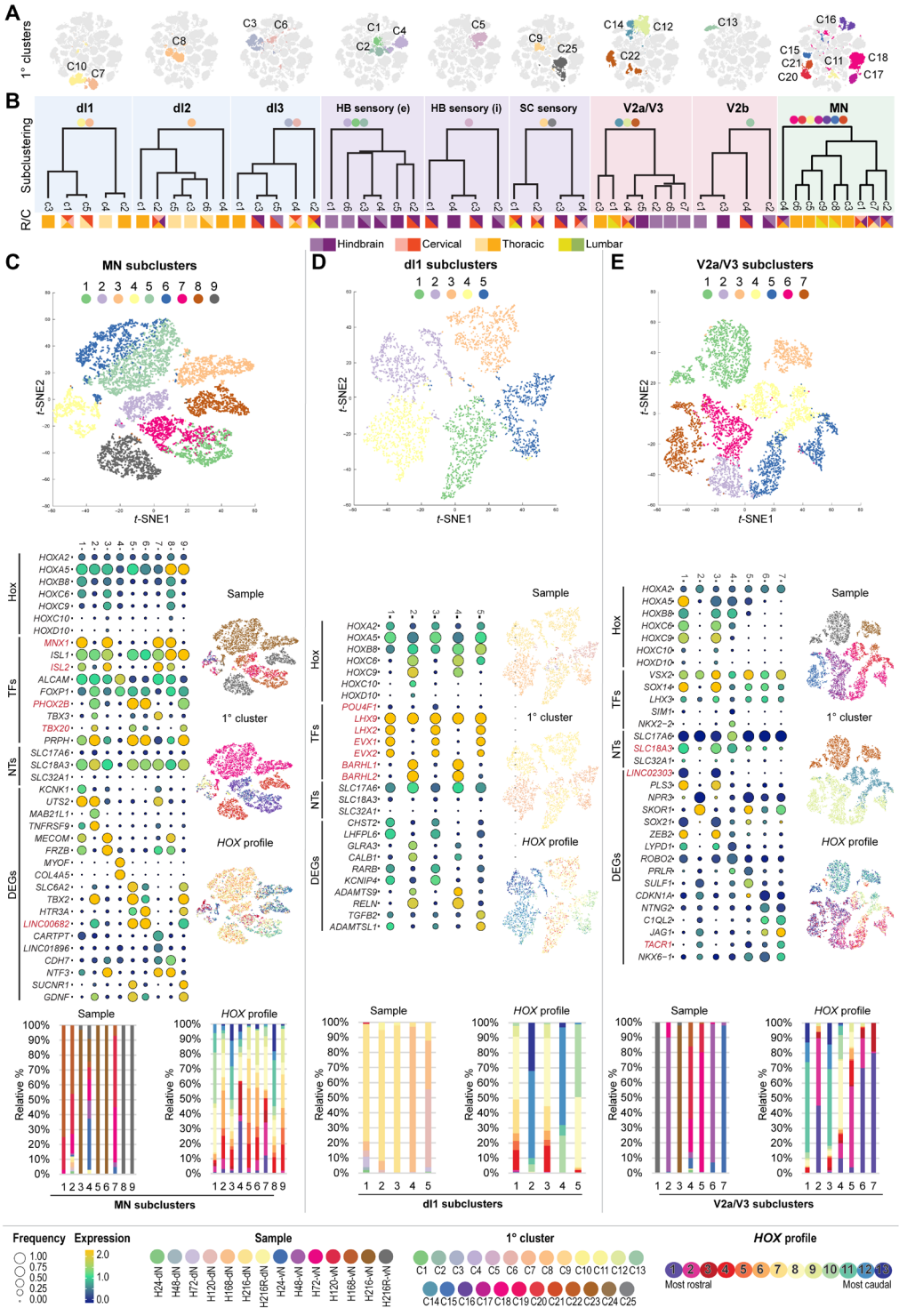

Fig. 6. Subcluster analysis reveals subpopulations with neuronal phenotype and regional specificity. (**A**) *t*-SNE plots showing primary cluster compositions in 9 of the 17 subgroups. (**B**) Hierarchal organization of subclusters with pictorial representations of estimated R/C location. Dual colors in key refer to rostral (pale) or caudal (dark) segments of hindbrain or spinal cord regions. (**C** to **E**) Subcluster analyses for (C) all MN populations, including sMN (C16, C17, and C21), vMN (C18, C20, and C21), and cranial MN (cMN; C11) clusters, (D) dI1 clusters (C7 and C10), and (E) V2a/V3 mixed clusters (C12, C14, and C22). *t*-SNEs show subclusters ($n = 5$ to 9) defined by consensus clustering and distributions of the samples, primary clusters, and Hox profiles. Dot plots display *HOX* gene expression, appropriate transcription factors (TFs) associated with the subpopulation grouping, neurotransmitters (NTs), and a selection of markers from the top 10 DEGs for each subcluster. The size of each circle reflects the fraction of cells where the gene is detected, and the color reflects the average expression level within each cluster (blue, low expression; yellow, high expression). Sample and *HOX* profile cluster distributions across subclusters are also visualized in stacked histograms.

Downloaded from https://www.science.org at University of Wisconsin Madison on February 18, 2025

patterns of gene expression spanning more than two subclusters could capture nuanced expression differences and thus more comprehensively characterize novel cell types. To this end, we developed a computational pipeline that first applied Arboretum (*66*), a multitask clustering algorithm, to assign genes into expression states based on their pseudo-bulk mean expression. Then, we identified "transitioning" gene sets, which exhibit coordinated changes in expression states across subclusters (Materials and Methods, Fig. 7A, and online resource). We interpreted these gene sets based on subclusters' regional and phenotypic identities. Here, we focus on the V2a/V3 and the high/low RA ventral groupings to demonstrate how this analysis can be used to detect patterns of interest in subpopulations along the R/C axis and to identify gene modules representing combinatorial expression changes across multiple cardinal populations in response to differentiation perturbations.

Because the V2a/V3 grouping was divided into subclusters corresponding to R/C regions from the hindbrain through thoracolumbar spinal cord (Fig. 6E), the transitioning genes from Arboretum indicated region-specific expression patterns (Fig. 7B). Akin to DEG analysis, we found patterns of expression specific to a single subcluster, like V2a/3-*#184*, which shows elevated expression in V2a INs (c7) of the rostral hindbrain. We also found shared patterns of expression between multiple subclusters, like V2a/3-*#173* or V2a/3-*#202*, which show genes up-regulated in spinal or hindbrain subtypes, respectively. These gene sets included factors of potential interest involved in binding HOX proteins (*MEIS2*), mitochondrial activity (*RPS4X*, *NDUFA8*, and *MDH2*), cell adhesion (*PCDH9*), surface biomarkers (*CD24*), neurite outgrowth (*NRN1*), and neuron-specific alternative splicing (*NOVA1*). We also found patterns representing gradual changes in gene expression that may identify region-specific changes that emerge as a gradient along the R/C axis or between subclusters, as has previously been observed for V2a INs (*26*). For example, V2a/V3-*#116* and V2a/V3-*#197* show gradual decrease in gene expression from hindbrain to spinal V2a INs, while V2a/V3-*#164* shows gradual increase in expression from hindbrain to spinal V2a INs. Last, we identified numerous gene sets that correspond to nuanced gene expression patterns, like V2a/V3-*#204*, which exhibit high levels of gene expression in the spinal cord and rostral hindbrain, but not the caudal hindbrain. The Arboretum pipeline and associated analysis is thus a valuable resource for curation of novel gene expression patterns that can be examined with targeted in vivo studies, compared to standard DEG analysis, which fails to contextualize or detect nuanced gene expression differences between subclusters.

Next, we used Arboretum to determine whether changing RA during D/V patterning had an impact on terminal gene expression. We observed that changing RA concentration during dorsal differentiation significantly changed the distribution of postmitotic cardinal populations (Fig. 4I), but while ventral populations were slightly shifted, both H216-vN and H216R-vN samples contained V2a INs, sMNs, and vMNs (Fig. 7C). This allowed for direct comparison between cardinal populations. Arboretum identified gene sets comprising commonly up-regulated (H216/R-vN-*#54*) or down-regulated (H216/R-vN-*#51*) genes in response to the increase in RA concentration (Fig. 7D). Gene set #54 includes *HOXA5*, which validates the role of RA in activation of specific *Hox* genes and mimics occurrences in vivo. Constitutive activation of RA signaling during development was found to disrupt digit-innervating MN development (*41*), and precise retinoid levels are required for digit and tendon development (*67*). Notably, the annotation of genes in H216/R-vN-*#51* (online resource) indicated that ventral patterning with 1 μM RA compared to 100 nM RA persistently suppressed signaling pathways involved in mitochondrial electron transport, mitochondrial respiration, and oxidation-reduction in postmitotic neurons matured 20 days beyond progenitor patterning. This finding could have significant implications for in vitro modeling of neurological disorders associated with mitochondrial pathologies and cell survival after transplantation. Moreover, it highlights the need for more thorough characterization of differentiation protocols used for prospective cell therapies, which may optimize for a particular cell phenotype without considering how subtle changes in morphogen concentrations can affect long-term transplant efficacy.

DISCUSSION

By implementing a modular differentiation paradigm that explicitly decouples R/C from D/V patterning, we demonstrate the ability to direct hPSCs to any neuronal phenotype in the pCNS. We show that all D/V phenotypes—particularly dorsal INs—can be effectively generated under monolayer culture conditions. This is in contrast to the previously requisite organoid and spheroid cultures, which exhibit batch-to-batch variability and may rely on the formation of signaling centers for D/V patterning (*6*, *7*, *12*–*14*). Moreover, our patterning schema enables deeper investigation of the role of *Hox* genes, retinoids, and other signaling molecules in the development of anatomically and therapeutically relevant cell types.

Our scRNA-seq data also highlight the power of hPSCs in providing broad access to embryonic pCNS tissues. While scRNA-seq datasets from primary embryonic rodent (*37*, *68*) and human spinal cords (*35*) are invaluable resources, they have limitations. Because of the physical difficulties associated with early embryonic tissue acquisition and dissection, these datasets fail to discretize neuronal phenotypes across different pCNS R/C regions. They also sparsely sample individual cell subtypes, a consequence of poor neuronal yield and subtype rarity. By comparison, our modular protocol enables unlimited sampling of any phenotype from any differentiation time point across any R/C region. As a result, our scRNA-seq dataset spans multiple discrete regions from the hindbrain through the thoracolumbar spinal cord, improving the ability to detect nuanced transcriptional programs that potentially regulate lineage specification.

The multiregional nature of our dataset posed challenges to systematically define cell clusters. While known cell markers are used commonly to define single-cell populations (*35*, *37*), a large proportion of cells remain unlabeled. In contrast, a clustering-based approach offers a more comprehensive strategy but remains challenging especially when there are a large number of unknown populations (*69*). Our two-step approach based on sparse NMF and consensus clustering allowed a biologically meaningful grouping of cells that recapitulated known as well as novel cell types. Furthermore, our Arboretum-based approach allowed us to uncover previously unknown patterns of expression that can inform functional experiments for in-depth characterization of these cell populations. We anticipate that these platforms will enable rigorous interrogation of gene regulatory pathways responsible for neuron diversification and synaptic targeting in the pCNS. Future studies encompassing other pCNS populations including those from patient induced pluripotent stem cells will enable investigation of spatiotemporal gene expression

Downloaded from https://www.science.org at University of Wisconsin Madison on February 18, 2025

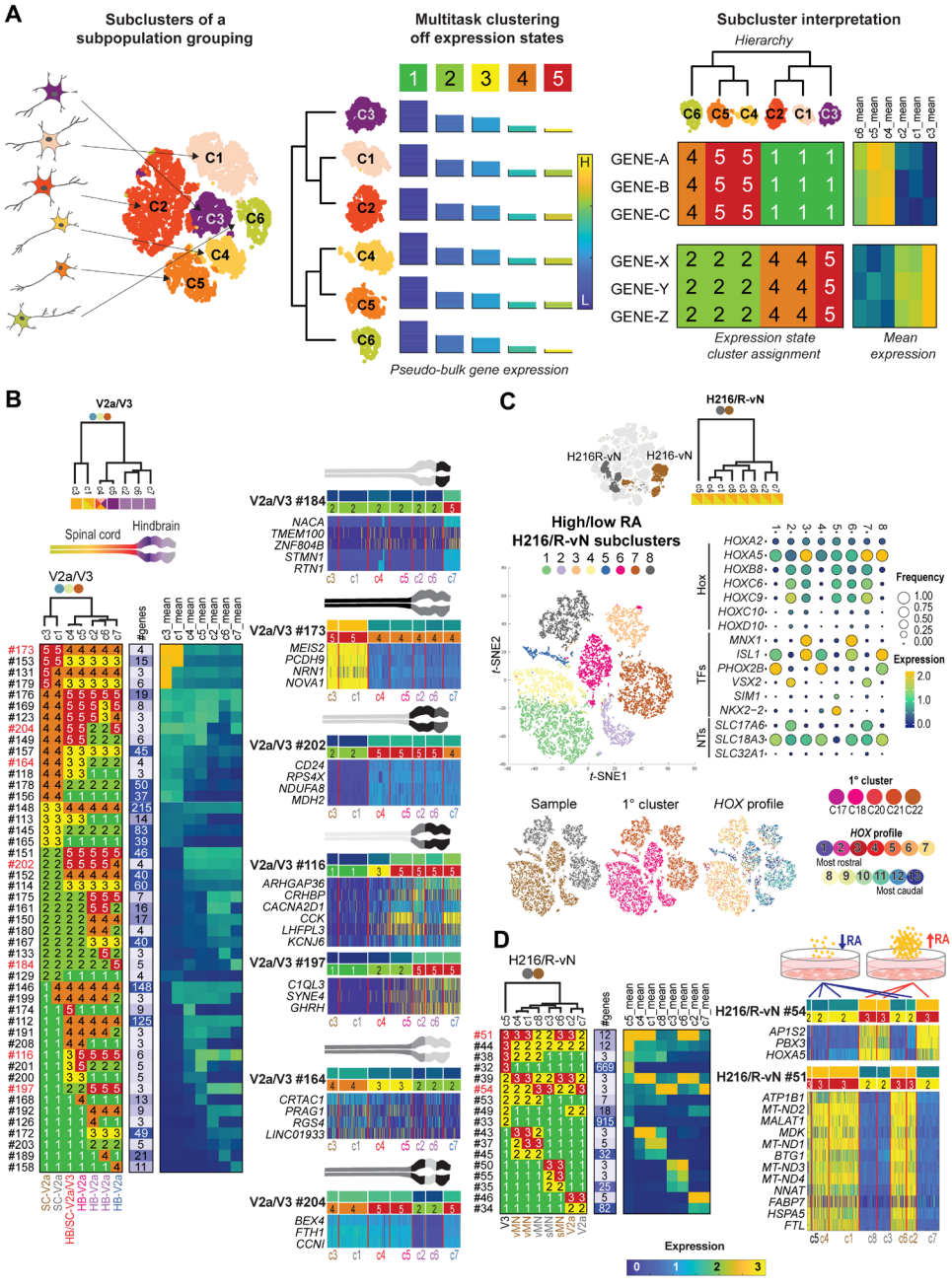

Fig. 7. Multitask clustering enables discovery of novel and nuanced gene expression patterns across subclusters. (**A**) Schematic procedure of Arboretum-based identification of combinatorial expression patterns. Arboretum, a multitask clustering method, clusters genes using their pseudo-bulk expression in each cell subcluster with consideration of relationship structure between subclusters. Each gene cluster is associated with an expression state. Subsequent interpretation of genes is made by grouping genes into sets that change their expression state across cell subclusters. (**B**) Expression state assignment patterns and mean expression levels for gene sets identified for V2a/V3 subgroup. Subclusters colored by predominant primary cluster identity (top); R/C positioning [spinal cord (SC) or hindbrain (HB)] and cardinal cell type colored by sample identity (bottom). A selection of gene sets highlights potential region-specific patterns of gene expression across the pCNS, represented in schematic images (right). (**C**) Subpopulation analyses for grouping of ventral samples exposed to high (H216R-vN) or low (H216-vN) RA. *t*-SNE plots show subclusters ($n = 8$) defined by consensus clustering and distributions of the samples (primary cluster population and high/low RA subpopulation) and *HOX* profiles. Hierarchical organization of subpopulation subclusters with pictorial representation of comparable thoracolumbar identity. (**D**) Expression state assignment patterns and mean expression levels for gene sets identified for high/low ventral subpopulation grouping. Subclusters (top) and cardinal cell type composition (bottom) colored by sample identity. A selection of gene sets highlights shared genes across multiple cardinal populations up-regulated in response to high or low RA (right).

Downloaded from https://www.science.org at University of Wisconsin Madison on February 18, 2025

dynamics during development and disease. These include diseases with pathologies exhibiting bulbar versus spinal onset and spinal cord injury, where the site and magnitude of trauma is patient specific. Improved understanding of region-specific pCNS circuitry and neurodegenerative susceptibility will inform pharmacological, cell transplantation, and gene therapy strategies, advancing the field toward personalized medicine.

MATERIALS AND METHODS

Stem cell maintenance

Experiments were conducted using the HUES3::Hb9-GFP line (Di Giorgio *et al.*) (Harvard Stem Cell Institute) or H9 (WA09, WiCell) hESC lines under xeno-free, feeder-free conditions. hESCs were maintained at 37°C in 5% CO_2 in Essential 8 (E8) medium on Matrigel (WiCell)–coated six-well tissue culture–treated plates and were passaged when 70 to 80% confluent. Briefly, cells were washed once with phosphate-buffered saline (PBS) (Invitrogen) and then incubated at 37°C in Versene (Invitrogen) for 6 min. Versene was aspirated, and the cells were gently dissociated from the well with fresh E8 and replated at a 1:12 seeding ratio. The medium was replenished daily (Lippmann *et al.*).

NMP differentiation

To initiate NMP differentiation, hESCs were washed once with PBS, incubated at 37°C in Accutase (Invitrogen) for 5 min, singularized by gentle trituration, and quenched with one volume of E8 medium. Following centrifugation for 5 min at 300*g*, hESCs were replated onto 35-mm Matrigel-coated plates at a density of 1.5×10^5 cells/cm^2 in E8 medium with 10 μM ROCK inhibitor (Y27632, Tocris). The medium was replaced with E6 medium (*70*) on the following day (day 0) and then changed to E6 supplemented with FGF8b (200 ng/ml; PeproTech) 24 hours later (day 1). On day 2, Hox propagation was initiated by activation of Wnt signaling using NMP medium consisting of E6 medium supplemented with FGF8b (200 ng/ml) and 3 μM CHIR99021 (Tocris). This constitutes the "Hox time point" of 0 hours. At various time points, NMPs were collected or differentiated to pCNSPs for scRNA-seq experiments and at 120 hours for D/V optimization experiments. For NMPs collected within 24 hours, the NMP medium was applied directly. Else, cells were subcultured at a 2:3 ratio. Briefly, cells were washed once with PBS, incubated in Accutase for 1.5 to 2 min, and removed from the surface by gentle pipetting. After centrifugation, cells were gently resuspended in NMP medium containing 10 μM Y27632 and seeded on 35-mm Matrigel-coated plates. The NMP medium was replenished on day 4. For NMPs collected between H72 and H96, the NMP medium was changed directly on day 5; else, cells were subcultured again at a 2:3 ratio. The medium was replenished daily on days 7 to 10 with the NMP medium containing GDF11 (30 ng/ml; PeproTech) and 1 μM dorsomorphin (Tocris) to stimulate caudal NMP development, with subculture on day 9 at a 1:1 ratio.

pCNSP differentiation

To initiate pCNSP differentiation, H9-derived NMPs were cultured for 1 day in pCNSP medium, consisting of E6 medium supplemented with 1 μM RA (Sigma-Aldrich), 10 μM SB-431542 (Abcam), and 100 nM LDN-193189 (Stemgent). Cells were singularized and replated at 5×10^5 cells/cm^2 in pCNSP medium containing 10 μM Y27632 for an additional 2 days. The medium was replenished daily.

D/V differentiation

pCNSPs were exposed to morphogens for 4 days to initiate D/V patterning. Dorsal progenitors were cultured in E6 medium containing 100 μM RA, 1 μM Cyc, and BMP4 (PeproTech) at different concentrations and durations. Ventral progenitors were cultured in E6 medium containing 100 nM RA, 10 μM SB-431542, 100 nM LDN-193189, Pur (Tocris), and SAGs (Calbiochem) at different concentrations. "High" RA conditions were cultured in 1 μM RA instead of 100 nM RA. Progenitors underwent neuronal differentiation for immunocytochemistry and qPCR studies by switching to maturation medium for 5 to 7 days. Maturation medium consisted of E6 containing 1× N2 supplement (Thermo Fisher Scientific), 50× B27 supplement (Thermo Fisher Scientific), 1 μM adenosine 3′,5′-monophosphate (cAMP) (Sigma-Aldrich), glial cell line–derived neurotrophic factor (GDNF) (10 ng/ml), brain-derived neurotrophic factor (BDNF) (10 ng/ml), NT-3 (10 ng/ml; PeproTech), and 10 μM DAPT (Tocris). As appropriate, BMP7 (10 ng/ml; PeproTech) was added to the maturation medium for additional dorsalization. The medium was replenished daily.

Differentiation for preoptimized scRNA-seq

Using the HUES3-Hb9-GFP hESC line, which fluorescently reports Hb9 (MNX1)$^+$ MNs (*71*), we differentiated NMPs from six time points corresponding to H24, H48, H72, H120, H168, or H216 of Hox patterning. Cultures were then switched to E6 medium containing 1 μM RA, 2 μM SAG, and 2 μM Pur for 3 days. Progenitors were subcultured at a 1:3 ratio, gently resuspended in E6 medium supplemented with 1 μM RA, 100 nM SAG, 100 nM Pur, and 10 μM Y27632, and plated on 35-mm Matrigel-coated well plates for an additional 3 days. Then, the medium was switched to E6 medium supplemented with 1 μM RA, 100 nM SAG, 100 nM Pur, and 5 μM DAPT for an additional 5 days and then cryopreserved. The medium was replenished daily.

Differentiation for optimized scRNA-seq

NMPs from six time points corresponding to H24, H48, H72, H120, H168, or H216 of Hox patterning were differentiated to pCNSPs. Dorsal progenitors were generated by culturing in E6 medium containing 100 nM RA, 1 μM Cyc, and BMP4 (20 ng/ml) for 1 day and then E6 medium containing 100 μM RA and 1 μM Cyc for three additional days. Ventral progenitors were generated by culturing in E6 medium containing 100 nM RA, 10 μM SB-431542, 100 nM LDN-193189, 0.5 μM Pur, and 0.5 μM SAG for 4 days. For H216R conditions, the RA concentration was increased to 1 μM for both dorsal and ventral differentiations. Progenitors were cryopreserved to synchronize cultures. For neuronal differentiation and maturation before sequencing, cells were thawed in maturation medium containing 10 μM Y27632 at 5×10^5 cells/cm^2 overnight, with daily medium changes for 6 days. The medium was switched to Neurobasal medium (Gibco) containing 1× N2 supplement, 50× B27 supplement, 1× GlutaMAX (Thermo Fisher Scientific), 1× penicillin-streptomycin (Invitrogen), laminin (1 ng/ml; Thermo Fisher Scientific), 1 μM cAMP, GDNF (10 ng/ml), BDNF (10 ng/ml), and NT-3 (10 ng/ml) for 14 days. Two days before sequencing, the medium was supplemented with 10 μM AraC (Sigma-Aldrich) to eliminate proliferating cells.

Cryopreservation

To create cryopreserved cell banks for further differentiation or scRNA-seq analysis, cells were dissociated in Accutase at 37°C for

Downloaded from https://www.science.org at University of Wisconsin Madison on February 18, 2025

30 min on an orbital shaker, quenched with one volume of E6 medium, centrifuged, and gently resuspended in 10% dimethyl sulfoxide in E6 medium with 10 μM Y27632. The cells were aliquoted at 1 ml per cryovial and cryopreserved with a CryoMed controlled rate freezer (Thermo Fisher Scientific) using a stepwise cooling program: rapid cooling from room temperature to 4°C, 1°C/min until reaching −60°C, and 10°C/min until reaching −100°C. Cryovials were transferred to a liquid nitrogen dewer for long-term storage.

siRNA knockdown validations

NMPs corresponding to H120 of Hox patterning were thawed from cryopreserved stocks and seeded at 5×10^5 cells/cm^2 in pCNSP medium containing 10 μM Y27632 for 2 days. Ventral progenitors were generated by culturing in E6 medium containing 100 nM RA, 10 μM SB-431542, 100 nM LDN-193189, 0.5 μM Pur, and 0.5 μM SAG for 4 days. Cultures were differentiated to neurons in maturation medium for an additional 5 days. Knockdown was performed using a Lipofectamine RNAiMAX Transfection Reagent (Invitrogen) according to the manufacturer's protocols using 10 nM small interfering RNA (siRNA) assays (table S4).

Quantitative real-time polymerase chain reaction

Total RNA was isolated using a TRIzol reagent (Invitrogen), and complementary DNA (cDNA) was synthesized using the SuperScript IV First-Strand Synthesis System (Invitrogen) according to the manufacturer's instructions. TaqMan Gene Expression Assays (table S5) and TaqMan Gene Expression Master Mix (Applied Biosystems) were used on a Bio-Rad CFX96 thermocycler with the following protocol: 50°C for 2 min; 95°C for 10 min; 40 cycles of 95°C for 15 s and 60°C for 1 min. Target genes were normalized to RPS18 expression, and relative gene expression was calculated using the comparative ΔC_t method. Fold differences in relative mRNA expression levels of target genes are reported for each gene with SDs (n = 6 biological replicates for each group). Statistical analysis was conducted using JMP-Pro 13 software. Significance was determined using a one-way analysis of variance (ANOVA) with Tukey-Kramer honestly significant difference post hoc for multiple comparisons with a 95% confidence threshold.

Immunocytochemistry

Cells were fixed in 4% paraformaldehyde for 10 min, washed thrice in PBS, and blocked in tris-buffered saline (TBS) containing 0.3% Triton X-100 and 5% normal donkey serum (TBSDT) for at least 1 hour. The cells were incubated in primary antibodies (table S5) diluted in TBSDT overnight at 4°C. After three 15-min washes in TBS containing 0.3% Triton X-100, the cells were incubated with Alexa Fluor secondary antibodies (Invitrogen) at a 1:500 dilution in TBSDT for 1 hour at room temperature. Cells were washed twice in TBS for 15 min each, counterstained with 300 nM 4′,6-diamidino-2-phenylindole (DAPI) for 10 min, and washed once more in TBS before mounting with Prolong Gold Antifade Reagent (Life Technologies) as necessary. Images were acquired using a Nikon A1R confocal microscope with Nikon NIS-Elements software and analyzed with NIS-Elements and ImageJ.

Single-cell dissociation of neurons

Cells were singularized for scRNA-seq by dissociation with papain (Worthington). Briefly, cells were washed once with PBS, incubated in papain at 37°C for 1 hour on an orbital shaker, and then triturated vigorously with a wide-bore pipette. The cell suspension was centrifuged at 300*g* for 5 min and then quenched with ovomucoid solution for 10 min at room temperature. Quenched cells were centrifuged, gently resuspended in PBS containing 0.2% bovine serum albumin and 10 μM Y27632, and then passed through a 40-mm cell strainer (Mitenyi Biotec) to remove debris. Cells were quantified and diluted to 700 cells/ml for sequencing.

Single-cell RNA sequencing

Directly after thaw or singularization, ~3000 to 5000 cells were targeted for capture from each sample. Transcriptomic profiling was performed using the Chromium Single Cell Gene Expression system (10X Genomics), according to the manufacturer's recommendations using the Single Cell 30 Reagent v2/v3 kits (10X Genomics). Post–GEM-RT (gel bead-in emulsion–reverse transcription) and post-cDNA amplification cleanup were performed using Dynabeads MyOne silane beads (Thermo Fisher Scientific) and SPRIselect (Beckman Coulter) kits, respectively. Successful library preparation was confirmed using Agilent Bioanalyzer (High Sensitivity DNA kit) and Qubit Fluorometer (High Sensitivity dsDNA kit). Experimental data were demultiplexed using the Cell Ranger Single Cell Software Suite, mkfastq command wrapped around Illumina's bcl2fastq. The MiSeq balancing run was quality-controlled using calculations based on UMI (unique molecular identifier)–tools (*72*). Sample libraries were balanced for the number of estimated reads per cell and run on an Illumina HiSeq 2500 or NovaSeq 6000 system. Cell Ranger software was then used to perform demultiplexing, alignment, filtering, barcode counting, UMI counting, and gene expression estimation for each sample according to the 10X Genomics documentation (https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger). The gene expression estimates from each sample were then aggregated using Cell Ranger (cellranger aggr) and processed through our data preprocessing pipeline to obtain filtered and normalized expression data.

Data preprocessing

For each of the 6 samples from direct differentiation (GSE186696) and 14 samples from modular differentiation (GSE186697) from our R/C time-series experiment, we filtered out genes that were expressed in fewer than five cells and cells with fewer than 5000 UMIs from the dataset. For all analysis in this study, we set our threshold for expression as 0, i.e., a gene needs to have a count >0 to be called as expressed in a cell. Each cell's expression value was depth-normalized to a depth of 5000, followed by variance stabilizing normalization as implemented in the pagoda2 package (*73*). We merged the gene expression matrices from each sample into a single matrix while taking the union of the genes from each matrix. The combined matrix is [12,543 cells × 20,598 genes] for the direct differentiation dataset and [49,959 cells × 23,941 genes] for the multiple generation dataset. We transformed the values of these matrices by taking their square root and standardizing each cell's expression profile by dividing by the mean expression of a gene in each cell for the subsequent clustering analysis.

Clustering of single cells by *HOX* gene profile

We obtained the expression values of 33 *HOX* gene paralogs from our normalized matrix to define the *HOX* profile of each cell (fig. S4, D to G, and table S6). We clustered the cells based on their *HOX* profile using two different clustering algorithms. The first approach binarized the *HOX* profiles based on nonzero expression of a *HOX* gene

Downloaded from https://www.science.org at University of Wisconsin Madison on February 18, 2025

in a cell and applied *k*-means clustering with *k* in {5,6,7,8,9,10,11,12, 13,14,15,16,17,18,19,20} on these binary profiles. The second approach applied Louvain clustering on knn graphs of cells with edges weighted by the Euclidean distances between the binary *HOX* expression profiles of each cell. We used four different values for the number of neighbors (*n*) in the knn graph, *n* in {20, 30, 40, 50}, and searched the resolution parameter of Louvain clustering (which controls the number of clusters) to identify 5 to 20 clusters. A resolution of 1 was applied initially for each desired *k*. Let k' be the number of clusters obtained at a resolution of r'. If $k' > k$, we decreased the resolution by $(0.5)^i$; else, we increased the resolution by $(0.5)^i$, where *i* is the search iteration. This process was repeated until the desired *k* was reached. There were four different clusterings (for the four values of *n*) for each *k* (the number of clusters), and the optimal *n* was selected on the basis of the lowest Pearson's correlation between the cluster means (fig. S4D). Our rationale was that the Pearson's correlation would be lowest for the most distinct clusters. We computed silhouette index (SI) for each Louvain clustering and compared this with the *k*-means clusters. Louvain clusters were used for the following analysis finally because they had better SI evaluation scores (fig. S4E). To determine *k*, we examined the patterns of *HOX* genes in each cluster in addition to the SI. On the basis of SI measures, a *k* value of 13 or 14 was optimal. We next annotated each cluster based on the pattern of expression of *HOX* genes, e.g., a cluster was annotated with a pattern "*HOXA4*$^+$/*HOXA5*-" if *HOXA4* was expressed, while *HOXA5* was not. We finally determined the number of clusters to be 13, as it had among the highest SI and the most distinct annotation patterns capturing most of the known *HOX* colinear expression patterns (fig. S4, F and G). Cluster IDs are rearranged manually based on the observed composition of nine *HOXA* genes within each cluster such that lower cluster IDs were more rostral (e.g., higher expression of *HOXA1* and *HOXA 2*), while higher cluster IDs were more caudal (e.g., higher expression of *HOXA13*).

Identification of primary clusters

We applied NMF implemented with alternating nonnegative constrained least squares (NMF-ANLS) and the active set method (*37*) with sparsity on the gene space for the identification of primary clusters. NMF decomposes an input matrix $X \in R^{m \times n}$ into two lower dimensional factors, *U* and *V* as $\|X - UV\|_F^2$, where $U \in R^{m \times k}$ and $V \in R^{k \times n}$. Here, *m* is the number of cells and *n* is the number of genes. In NMF-ANLS, the objective is defined as $\|X - UV\|_F^2 + \alpha\|U\|_F^2 + \beta(\sum_{j=1}^{n} \|V(:, j)\|_1^2)$, where the regularization parameter α controls the magnitude of *U* and β is used to tune the extent of sparsity (*36*). We used the mean value of all values of the input matrix as α and β. This implementation of NMF was shown to have a faster convergence and be more computationally efficient compared to the multiplicative update algorithm originally developed for NMF (*74*). Furthermore, on the basis of our comparisons of this algorithm to the ordinary least squares implementation in MATLAB (NMF-OLS), this produced more stable solutions (fig. S6, A and B). We performed NMF on our merged normalized [cells × genes] matrix, with the number of factors/lower dimensions, *k*, to be 5, 10, 15, 20, 25, 30, and 35. *K*-means clustering was performed on the *U* matrix, and *k* was designated to be the same as the number of the factors for improved clusters. As NMF results in different solutions depending on the starting seed, we applied NMF with 20 different random initializations and assessed the stability based on Jaccard index (JI) of the cluster assignments. Overall, the NMF factorizations were stable (JI ranging from 0.64 to 0.99). For each *k*, we took the most stable initialization based on the maximum average JI between each initialization and the remaining ones (fig. S6B). We used two metrics to determine the number of clusters. First, we extracted 23 well-known neural marker genes (table S1C) from our data matrix and calculated the SI of each clustering solution (fig. S6C). Second, we tested the significance of the difference in expression profiles for each pair of clusters. Briefly, for each *k*, we first obtained the pseudo-bulk expression of each cluster by taking the mean expression value of a gene across all cells in a cluster. Next, we investigated whether the expression vector of one cluster was significantly different from another cluster (two-sided paired *t* test, $P < 0.05$) and counted the proportion of pairs that were significant (fig. S6D). We determined $k = 25$ to be optimal based on SI and the proportion of pairs that were significantly different. To assign cell type identities, we used sample composition of each cluster (fig. S6E), and the relationship between these clusters and the clusters was defined using the HOX expression (fig. S6F) to help determine the hindbrain/spinal cord identity. Cluster IDs were rearranged manually to preserve the temporal order of the samples and similarity of sample composition of a cluster. Our code for the NMF analysis is available at https://doi.org/10.5281/zenodo.6505441.

Subpopulation clustering analysis

We regrouped our 25 primary cell clusters into 17 subgroups based on similarity of the cell types assigned to each cluster: all MNs, somatic MN, vMN, floor plate–cranial MN, hindbrain–spinal cord (HB-SC) sensory excitatory, HB-SC sensory inhibitory, HB sensory excitatory, HB sensory inhibitory, SC sensory, HB-SC proprioceptive, dl1, dl2, dlA3/dl3, V2a/V3, V2b, RA–dorsal neurons, and RA-vNs. Each subpopulation had between 1084 and 11,965 cells (table S3). For each group of clusters, we aimed to identify robust, high-confidence, and fine-grained cell subpopulations indicative of a specific cell type. We then developed a novel clustering pipeline consisting of three steps: (i) ensemble of clusterings, (ii) consensus graph generation, and (iii) consensus clustering (fig. S7A). We used the V2a/V3 and V2b groups to optimize our pipeline and applied the steps to the remaining 15 groups. Our code for the consensus clustering pipeline is available at https://doi.org/10.5281/zenodo.6505441.

For the first step, we generated a number of clustering solutions to be used for consensus clustering. We compared two different types of clustering approaches for generating the ensemble of clusterings. First, we applied NMF (with the ANLS algorithm), followed by *k*-means clustering on *U* factor matrix. Second, we applied Louvain clustering with the knn graph estimated using two approaches: (i) a knn graph from pairwise Euclidean distance estimated from an NMF-reduced space of 50 dimensions, and (ii) a knn graph estimated using fuzzy simplicial set, used in UMAP (uniform manifold approximation and projection) and scanpy (*75*). NMF was applied with the number of factors, *k*, to be 3 to 10, each with 10 different random initializations, resulting in 80 different clustering solutions. For both Louvain clustering approaches, we obtained knn graphs with $k = \{10, 20, 30, 40, 50\}$, each with eight different resolutions, {0.01, 0.05, 0.1, 0.3, 0.5, 0.7, 1.0, 1.5}, which in total resulted in 80 different clusterings.

In the second step, we created a consensus graph of cell co-clustering relationship. For every pair of cells, we counted the proportion of times the two cells were in the same cluster (across any of our three clustering approaches, *k*, and resolution) and generated a

Downloaded from https://www.science.org at University of Wisconsin Madison on February 18, 2025

weighed graph of cells with weights corresponding to this proportion. We generated three types of consensus graphs: one based on NMF only clusterings, one based on Louvain only clusterings, and one combining both NMF and Louvain.

In the final step, our goal was to estimate robust cell clusters by clustering the consensus graph. We considered two clustering approaches: one based on NMF and another based on Louvain clustering. For NMF, we considered the number factors in the range of 3 to 15 and defined cell clusters l based on the factor with the largest value for the cell, i.e., l = arg max (U_{j1}, U_{j2}, ⋯, U_{jl}), where $U \in R^{mXl}$,$1 \leq j \leq m$. We repeated this procedure 10 times and picked the initialization with the highest Jaccard coefficient with the other clustering solutions. For Louvain clustering, we extracted the knn graph from the full weighted consensus graph matrix with {10, 20, 30, 40, 50} nearest neighbors and applied clustering at five different resolutions as {0.1, 0.3, 0.5, 0.7, 1.0}. We used a metric delta consensus count (DCC) for measuring the quality of the clusters on the graph. DCC is defined as DCC $= \frac{1}{l}\sum_{i=1}^{l} \mathrm{In}_i - \mathrm{Out}_i$ for l clusters, where In_i is the average of edge weights within a cluster i and Out_i is the average of edge weights between nodes of cluster i to nodes not in cluster i. On the basis of DCC values, NMF was optimal across all three steps, and we applied the same procedure to all other subpopulations (fig. S7B).

Having determined NMF to be the optimal algorithm for our consensus clustering approach, we generated NMF-based consensus clusters for each subpopulation. For all but MNs, we considered k for NMF in step 1, to range from 3 to 10, with 10 different random initializations, resulting in 80 different clusterings. For the MNs, we used a higher range of k (3 to 30) because MNs are known to be more complex than others, resulting in 280 clustering solutions. After the consensus graph was generated, a second round of optimization was performed to select the optimal k for the clustering. We used a combination of quantitative and qualitative methods. For the quantitative methodology, we used the summation of three different evaluation metrics, SI, DCC, and stability score (average JI for each pair of clusterings). The top three to five best results were short-listed and subsequently examined using our qualitative method (table S7A). Here, we manually inspected the block-diagonalness of the clustered consensus graph matrix to avoid over- or underclustering (fig. S7A, iii). On the basis of this procedure, the 17 groups were subdivided into 4 to 9 fine cell subclusters (table S3). The main paper presents the results of nine of these groups, and the remaining are available in our online resource (https://doi.org/10.5281/zenodo.6506221). The regional specificities of subclusters were addressed by the observations of sample compositions and *HOX* cluster compositions (fig. S7C).

Comparative analysis against previous human and mouse in vivo studies

We compared our scRNA-seq dataset to two previous in vivo studies that used human [Rayon *et al.* (*35*)] and mouse [Delile *et al.* (*37*)] cells. The raw data from the human and mouse single-cell expression studies were downloaded from the Gene Expression Omnibus (GSE171890) and ArrayExpress (E-MTAB-7320), respectively. Each dataset was preprocessed using the same procedure described above and finally merged into a single matrix, resulting in 23,179 genes by 47,089 cells for the human dataset and 17,335 genes by 27,725 cells for the mouse dataset. Comparative analysis was restricted to only *SNAP25*$^{+}$ neuronal cells in all datasets, which resulted in 6026 cells in the mouse dataset, 8050 cells in the human dataset, and 44,487 cells in our dataset. For each dataset, the cells were first grouped into cell types based on the expression of marker genes in the Rayon *et al.* (*35*) knowledge matrix. We compared our 25 primary clusters (Fig. 5D), 17 subgroups (fig. S5A), and 11 cell types [fig. S5B; defined using the knowledge matrix from Rayon *et al.* (*35*)] to cell types defined in the mouse [Delile *et al.* (*37*)] and human [Rayon *et al.* (*35*)] datasets. For all comparisons, we used 77 marker genes in the knowledge matrix provided by Rayon *et al.* (*35*) [using 55 mouse orthologs for Delile *et al.* (*37*) obtained from Mouse Genome Informatics (*76*)] and transcription factors (1463 genes for human-human and 1775 genes for human-mouse comparisons) defined by PANTHER and Gene Ontology (GO). For each type of cell grouping (NMF, subgroup, or cell types), we obtained a pseudo-bulk expression profile of all marker genes using the mean expression across cells within a group. The similarity between any pair of cell groupings was estimated using the Pearson's correlation of each group's pseudo-bulk profiles (see table S2 for correlation values and corresponding P values). A pair of cell groupings was considered matched if there was a high correlation between each row group to one or a few column groups. The best concordance was obtained using the cell type definition of cell groups.

Identification of DEGs

For the robustness of the DEGs, we used the intersection of three statistical tests. We first defined DEGs per subcluster of each subgroup as the genes that are expressed in >50% of the cells in a subcluster, while the ratio $= \frac{E(x \mid \mathrm{In}_i)}{E(x \mid \mathrm{Out}_i)}$ is more than 1.25 in tandem, where $E(x \mid \mathrm{In}_i)$ is defined as the number of cells expressing gene x in cluster i and $E(x \mid \mathrm{Out}_i)$ is the number of cells expressed in cells not in cluster i. Then, we computed the statistical significance of overlap between all cells expressing the gene x and all cells in the subcluster based on the hypergeometric test. We additionally tested the Welch's t test and Mann-Whitney rank test (Wilcoxon rank sum test) to assess for the differential expression of genes in a subcluster compared to the complementary part of the cluster in the subpopulation. Function "de.test.t_test" and "de.test.rank_test" of the Python package "diffxpy" of the scanpy suite (*75*) were used in this calculation, respectively. Last, we kept only the DEGs that were significant in all three tests ($P < 0.05$) to create a stringent set of DEGs (table S4).

Arboretum-based identification of subcluster-specific genes

We adapted a previously developed multitask clustering framework Arboretum (*66*) to find gene modules with similar expression patterns across subclusters of any of the 17 subgroups. Arboretum is used to jointly cluster multiple hierarchically related gene expression datasets such that cluster assignments for more similar datasets are more similar. Such relationships could be obtained from phylogenies or other hierarchical clustering. The Arboretum framework is based on a generative probabilistic process and has two components: The emission model generates the observed expression measurements at the tips of the tree and is formulated as a mixture of k Gaussians, where k is the number of clusters, and the clusters are related via transition probabilities that model the probabilistic propagation of module assignments from the root of the tree to the tips. In our application of Arboretum to scRNA-seq datasets, we first generated pseudo-bulk profiles for each cell subcluster, used

Downloaded from https://www.science.org at University of Wisconsin Madison on February 18, 2025

these to define hierarchies (described next), and finally applied Arboretum to these data with varying values of k. The component of the Gaussian mixture corresponds to an expression state. Although the original application of Arboretum models multidimensional expression matrices, we used one-dimensional pseudo-bulk representation of each cell subcluster for computational efficiency.

To obtain the relationship structure of the cell subclusters, we performed hierarchical clustering based on pairwise distances between pseudo-bulk vectors. For each of the 17 subgroups, we considered unweighted average distance (UPGMA) with different distance metrics including Euclidean distance, Pearson's correlation, and cosine distance. We picked the best structure based on the cophenetic correlation coefficient. Different groups were best described by trees from different distance functions (table S7B).

We tested k to be {3,4,5} in the Arboretum clustering of each group as the numbers of gene expression states. The best k was determined using the optimal value across three metrics: penalized log-likelihood scores, Bayesian information criterion (BIC) penalized score, and Akaike information criterion (AIC) penalized score (table S7C). After clustering, each gene is assigned a cluster assignment in each cell subcluster, which is represented by a vector of discretized expression values across subclusters.

To identify gene sets with combinatorial patterns of expression across the subclusters, we obtained genes that change their cluster assignment across subclusters and applied our previously developed tool adapted to single-cell datasets, scFindTransitioning (https://doi.org/10.5281/zenodo.6506151), which is based on hierarchical clustering of gene cluster assignment profiles. The scFindTransitoning tool takes a parameter for determining the cluster height to cut the dendrogram. For this analysis, we used the height of 0.05, which was selected on the basis of our previous experience with this tool on other datasets. We interpreted these gene sets based on their expression trends as well as known annotations from PANTHER (*77*) and GO databases (*78*). The transitioning gene sets and associated enrichment analysis results are available at https://doi.org/10.5281/zenodo.6506221.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.abn7430

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. D. C. Lu, T. Niu, W. A. Alaynick, Molecular and cellular development of spinal cord locomotor circuitry. *Front. Mol. Neurosci.* **8**, 25 (2015).
2. P. Philippidou, J. S. Dasen, Hox genes: Choreographers in neural development, architects of circuit organization. *Neuron* **80**, 12–34 (2013).
3. M. W. Amoroso, G. F. Croft, D. J. Williams, S. O'Keeffe, M. A. Carrasco, A. R. Davis, L. Roybon, D. H. Oakley, T. Maniatis, C. E. Henderson, H. Wichterle, Accelerated high-yield generation of limb-innervating motor neurons from human stem cells. *J. Neurosci.* **33**, 574–586 (2013).
4. J. C. Butts, N. Iyer, N. White, R. Thompson, S. Sakiyama-Elbert, T. C. McDevitt, V2a interneuron differentiation from mouse and human pluripotent stem cells. *Nat. Protoc.* **14**, 3033–3058 (2019).
5. Z.-W. Du, H. Chen, H. Liu, J. Lu, K. Qian, C.-L. Huang, X. Zhong, F. Fan, S.-C. Zhang, Generation and expansion of highly pure motor neuron progenitors from human pluripotent stem cells. *Nat. Commun.* **6**, 6626 (2015).
6. N. Duval, C. Vaslin, T. C. Barata, Y. Frarma, V. Contremoulins, X. Baudin, S. Nedelec, V. C. Ribes, BMP4 patterns Smad activity and generates stereotyped cell fate organization in spinal organoids. *Development* **146**, dev175430 (2019).
7. S. Gupta, D. Sivalingam, S. Hain, C. Makkar, E. Sosa, A. Clark, S. J. Butler, Deriving dorsal spinal sensory interneurons from human pluripotent stem cells. *Stem Cell Rep.* **10**, 390–405 (2018).
8. Y. Maury, J. Come, R. A. Piskorowski, N. Salah-Mohellibi, V. Chevaleyre, M. Peschanski, C. Martinat, S. Nedelec, Combinatorial analysis of developmental cues efficiently converts human pluripotent stem cells into multiple neuronal subtypes. *Nat. Biotechnol.* **33**, 89–96 (2015).
9. V. Mouilleau, C. Vaslin, R. Robert, S. Gribaudo, N. Nicolas, M. Jarrige, A. Terray, L. Lesueur, M. W. Mathis, G. Croft, M. Daynac, V. Rouiller-Fabre, H. Wichterle, V. Ribes, C. Martinat, S. Nedelec, Dynamic extrinsic pacing of the HOX clock in human axial progenitors controls motor neuron subtype specification. *Development* **148**, dev194514 (2021).
10. A. R. G. Libby, D. A. Joy, N. H. Elder, E. A. Bulger, M. Z. Krakora, E. A. Gaylord, F. Mendoza-Camacho, J. C. Butts, T. C. McDevitt, Axial elongation of caudalized human organoids mimics aspects of neural tube development. *Development* **148**, dev198275 (2021).
11. N. Moris, K. Anlas, S. C. van den Brink, A. Alemany, J. Schröder, S. Ghimire, T. Balayo, A. van Oudenaarden, A. Martinez Arias, An in vitro model of early anteroposterior organization during human development. *Nature* **582**, 410–415 (2020).
12. J. Andersen, O. Revah, Y. Miura, N. Thom, N. D. Amin, K. W. Kelley, M. Singh, X. Chen, M. V. Thete, E. M. Walczak, H. Vogel, H. C. Fan, S. P. Paşca, Generation of functional human 3D cortico-motor assembloids. *Cell* **183**, 1913–1929.e26 (2020).
13. T. Ogura, H. Sakaguchi, S. Miyamoto, J. Takahashi, Three-dimensional induction of dorsal, intermediate and ventral spinal cord tissues from human pluripotent stem cells. *Development* **145**, dev162214 (2018).
14. Y. Zheng, X. Xue, A. Resto-Irizarry, Z. Li, Y. Shao, G. Zhao, J. Fu, Dorsal-ventral patterned neural cyst from human pluripotent stem cells in a neurogenic niche. *Sci. Adv.* **5**, eaax5933 (2019).
15. D. Henrique, E. Abranches, L. Verrier, K. G. Storey, Neuromesodermal progenitors and the making of the spinal cord. *Development* **142**, 2864–2875 (2015).
16. F. J. Wymeersch, Y. Huang, G. Blin, N. Cambray, R. Wilkie, F. C. Wong, V. Wilson, Position-dependent plasticity of distinct progenitor types in the primitive streak. *eLife* **5**, e10042 (2016).
17. J. Deschamps, D. Duboule, Embryonic timing, axial stem cells, chromatin dynamics, and the Hox clock. *Genes Dev.* **31**, 1406–1416 (2017).
18. R. Diez del Corral, I. Olivera-Martinez, A. Goriely, E. Gale, M. Maden, K. Storey, Opposing FGF and retinoid pathways control ventral neural pattern, neuronal differentiation, and segmentation during body axis extension. *Neuron* **40**, 65–79 (2003).
19. J. P. Liu, The function of growth/differentiation factor 11 (Gdf11) in rostrocaudal patterning of the developing spinal cord. *Development* **133**, 2865–2874 (2006).
20. J. P. Liu, E. Laufer, T. M. Jessell, Assigning the positional identity of spinal motor neurons: Rostrocaudal patterning of Hox-c expression by FGFs, Gdf11, and retinoids. *Neuron* **32**, 997–1012 (2001).
21. R. Neijts, S. Amin, C. van Rooijen, S. Tan, M. P. Creyghton, W. de Laat, J. Deschamps, Polarized regulatory landscape and Wnt responsiveness underlie Hox activation in embryos. *Genes Dev.* **30**, 1937–1942 (2016).
22. L. Wilson, M. Maden, The mechanisms of dorsoventral patterning in the vertebrate neural tube. *Dev. Biol.* **282**, 1–13 (2005).
23. L. R. Hernandez-Miranda, T. Muller, C. Birchmeier, The dorsal spinal cord and hindbrain: From developmental mechanisms to functional circuits. *Dev. Biol.* **432**, 34–42 (2016).
24. U. Marklund, Z. Alekseenko, E. Andersson, S. Falci, M. Westgren, T. Perlmann, A. Graham, E. Sundström, J. Ericson, Detailed expression analysis of regulatory genes in the early developing human neural tube. *Stem Cells Dev.* **23**, 5–15 (2014).
25. J. S. Dasen, T. M. Jessell, Chapter Six Hox networks and the origins of motor neuron diversity. *Curr. Top. Dev. Biol.* **88**, 169–200 (2009).
26. M. Hayashi, C. A. Hinckley, S. P. Driscoll, N. J. Moore, A. J. Levine, K. L. Hilde, K. Sharma, S. L. Pfaff, Graded arrays of spinal and supraspinal V2a interneuron subtypes underlie forelimb and hindlimb motor control. *Neuron* **97**, 869–884.e5 (2018).
27. C. Nolte, R. Krumlauf, *Expression of Hox Genes in the Nervous System of Vertebrates* (Landes Bioscience, 2013); www.ncbi.nlm.nih.gov/books/NBK6519/.
28. L. B. Sweeney, J. B. Bikoff, M. I. Gabitto, S. Brenner-Morton, M. Baek, J. H. Yang, E. G. Tabak, J. S. Dasen, C. R. Kintner, T. M. Jessell, Origin and segmental diversity of spinal inhibitory interneurons. *Neuron* **97**, 341–355.e3 (2018).
29. E. S. Lippmann, C. E. Williams, D. A. Ruhl, M. C. Estevez-Silva, E. R. Chapman, J. J. Coon, R. S. Ashton, Deterministic HOX patterning in human pluripotent stem cell–derived neuroectoderm. *Stem Cell Rep.* **4**, 632–644 (2015).
30. E. O. Mazzoni, S. Mahony, M. Peljto, T. Patel, S. R. Thornton, S. McCuine, C. Reeder, L. A. Boyer, R. A. Young, D. K. Gifford, H. Wichterle, Saltatory remodeling of Hox chromatin in response to rostrocaudal patterning signals. *Nat. Neurosci.* **16**, 1191–1198 (2013).
31. M. Gouti, A. Tsakiridis, F. J. Wymeersch, Y. Huang, J. Kleinjung, V. Wilson, J. Briscoe, In vitro generation of neuromesodermal progenitors reveals distinct roles for Wnt signalling in the specification of spinal cord and paraxial mesoderm identity. *PLOS Biol.* **12**, e1001937 (2014).
32. J. A. McMahon, S. Takada, L. B. Zimmerman, C. M. Fan, R. M. Harland, A. P. McMahon, Noggin-mediated antagonism of BMP signaling is required for growth and patterning of the neural tube and somite. *Genes Dev.* **12**, 1438–1452 (1998).

Downloaded from https://www.science.org at University of Wisconsin Madison on February 18, 2025

33. M. G. Andrews, L. M. Del Castillo, E. Ochoa-Bolton, K. Yamauchi, J. Smogorzewski, S. J. Butler, BMPs direct sensory interneuron identity in the developing spinal cord using signal-specific not morphogenic activities. *eLife* **6**, e30647 (2017).
34. G. Le Dreau, L. Garcia-Campmany, M. A. Rabadan, T. Ferronha, S. Tozer, J. Briscoe, E. Marti, Canonical BMP7 activity is required for the generation of discrete neuronal populations in the dorsal spinal cord. *Development* **139**, 259–268 (2012).
35. T. Rayon, R. J. Maizels, C. Barrington, J. Briscoe, Single-cell transcriptome profiling of the human developing spinal cord reveals a conserved genetic programme with human-specific features. *Development* **148**, dev199711 (2021).
36. H. Kim, H. Park, Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method. *SIAM J. Matrix Anal. Appl.* **30**, 713–730 (2008).
37. J. Delile, T. Rayon, M. Melchionda, A. Edwards, J. Briscoe, A. Sagner, Single cell transcriptomics reveals spatial and temporal dynamics of gene expression in the developing mouse spinal cord. *Development* **146**, dev173807 (2019).
38. C. Catela, M. M. Shin, D. H. Lee, J. P. Liu, J. S. Dasen, Hox proteins coordinate motor neuron differentiation and connectivity programs through *Ret/Gfrα* genes. *Cell Rep.* **14**, 1901–1915 (2016).
39. J. S. Dasen, J. P. Liu, T. M. Jessell, Motor neuron columnar fate imposed by sequential phases of Hox-c activity. *Nature* **425**, 926–933 (2003).
40. J. S. Dasen, A. De Camilli, B. Wang, P. W. Tucker, T. M. Jessell, Hox repertoires for motor neuron diversity and connectivity gated by a single accessory factor, FoxP1. *Cell* **134**, 304–316 (2008).
41. A. I. Mendelsohn, J. S. Dasen, T. M. Jessell, Divergent hox coding and evasion of retinoid signaling specifies motor neurons innervating digit muscles. *Neuron* **93**, 792–805.e4 (2017).
42. N. Stifani, Motor neurons and the generation of spinal motor neuron diversity. *Front. Cell. Neurosci.* **8**, 293 (2014).
43. S. Guthrie, Patterning and axon guidance of cranial motor neurons. *Nat. Rev. Neurosci.* **8**, 859–871 (2007).
44. M.-C. Tiveron, M.-R. Hirsch, J.-F. Brunet, The expression pattern of the transcription factor Phox2 delineates synaptic pathways of the autonomic nervous system. *J. Neurosci.* **16**, 7649–7660 (1996).
45. O. Issler, Y. Y. van der Zee, A. Ramakrishnan, J. Wang, C. Tan, Y.-H. E. Loh, I. Purushothaman, D. M. Walker, Z. S. Lorsch, P. J. Hamilton, C. J. Peña, E. Flaherty, B. J. Hartley, A. Torres-Berrío, E. M. Parise, H. Kronman, J. E. Duffy, M. S. Estill, E. S. Calipari, B. Labonté, R. L. Neve, C. A. Tamminga, K. J. Brennand, Y. Dong, L. Shen, E. J. Nestler, Sex-specific role for the long non-coding RNA LINC00473 in depression. *Neuron* **106**, 912–926.e5 (2020).
46. Z. Quan, D. Zheng, H. Qing, Regulatory roles of long non-coding RNAs in the central nervous system and associated neurodegenerative diseases. *Front. Cell. Neurosci.* **11**, 175 (2017).
47. A. Wang, J. Wang, Y. Liu, Y. Zhou, Mechanisms of long non-coding RNAs in the assembly and plasticity of neural circuitry. *Front. Neural. Circuits* **11**, 76 (2017).
48. O. A. Samad, M. J. Geisen, G. Caronia, I. Varlet, V. Zappavigna, J. Ericson, C. Goridis, F. M. Rijli, Integration of anteroposterior and dorsoventral regulation of Phox2b transcription in cranial motoneuron progenitors by homeodomain proteins. *Development* **131**, 4071–4083 (2004).
49. N. A. Bermingham, B. A. Hassan, V. Y. Wang, M. Fernandez, S. Banfi, H. J. Bellen, B. Fritzsch, H. Y. Zoghbi, Proprioceptor pathway development is dependent on MATH1. *Neuron* **30**, 411–422 (2001).
50. A. W. Helms, J. E. Johnson, Progenitors of dorsal commissural interneurons are defined by MATH1 expression. *Development* **125**, 919–928 (1998).
51. A. Bulfone, E. Menguzzato, V. Broccoli, A. Marchitiello, C. Gattuso, M. Mariani, G. G. Consalez, S. Martinez, A. Ballabio, S. Banfi, *Barhl1*, a gene belonging to a new subfamily of mammalian homeobox genes, is expressed in migrating neurons of the CNS. *Hum. Mol. Genet.* **9**, 1443–1452 (2000).
52. Q. Ding, P. S. Joshi, Z. Xie, M. Xiang, L. Gan, BARHL2 transcription factor regulates the ipsilateral/contralateral subtype divergence in postmitotic dI1 neurons of the developing spinal cord. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 1566–1571 (2012).
53. S. I. Wilson, B. Shafer, K. J. Lee, J. Dodd, A molecular program for contralateral trajectory: Rig-1 control by LIM homeodomain transcription factors. *Neuron* **59**, 413–424 (2008).
54. L. Moran-Rivard, T. Kagawa, H. Saueressig, M. K. Gross, J. Burrill, M. Goulding, Evx1 is a postmitotic determinant of v0 interneuron identity in the spinal cord. *Neuron* **29**, 385–399 (2001).
55. A. Pierani, L. Moran-Rivard, M. J. Sunshine, D. R. Littman, M. Goulding, T. M. Jessell, Control of interneuron fate in the developing spinal cord by the progenitor homeodomain protein Dbx1. *Neuron* **29**, 367–384 (2001).
56. K. J. Lee, M. Mendelsohn, T. M. Jessell, 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 (1998).
57. M. Davenne, M. K. Maconochie, R. Neun, A. Pattyn, P. Chambon, R. Krumlauf, F. M. Rijli, Hoxa2 and Hoxb2 control dorsoventral patterns of neuronal development in the rostral hindbrain. *Neuron* **22**, 677–691 (1999).
58. E. Azim, J. Jiang, B. Alstermark, T. M. Jessell, Skilled reaching relies on a V2a propriospinal internal copy circuit. *Nature* **508**, 357–363 (2014).
59. S. A. Crone, K. A. Quinlan, L. Zagoraiou, S. Droho, C. E. Restrepo, L. Lundfald, T. Endo, J. Setlak, T. M. Jessell, O. Kiehn, K. Sharma, Genetic ablation of V2a ipsilateral interneurons disrupts left-right locomotor coordination in mammalian spinal cord. *Neuron* **60**, 70–83 (2008).
60. S. A. Crone, J.-C. Viemari, S. Droho, A. Mrejeru, J.-M. Ramirez, K. Sharma, Irregular breathing in mice following genetic ablation of V2a neurons. *J. Neurosci.* **32**, 7895–7906 (2012).
61. A. Pedroni, K. Ampatzis, Large-scale analysis of the diversity and complexity of the adult spinal cord neurotransmitter typology. *iScience* **19**, 1189–1201 (2019).
62. P. A. Gray, W. A. Janczewski, N. Mellen, D. R. McCrimmon, J. L. Feldman, Normal breathing requires preBötzinger complex neurokinin-1 receptor-expressing neurons. *Nat. Neurosci.* **4**, 927–930 (2001).
63. T. D. Sheahan, C. A. Warwick, L. G. Fanien, S. E. Ross, The neurokinin-1 receptor is expressed with gastrin-releasing peptide receptor in spinal interneurons and modulates itch. *J. Neurosci.* **40**, 8816–8830 (2020).
64. A. J. Todd, Anatomy of primary afferents and projection neurones in the rat spinal dorsal horn with particular emphasis on substance P and the neurokinin 1 receptor. *Exp. Physiol.* **87**, 245–249 (2002).
65. A. Dady, L. Davidson, P. A. Halley, K. G. Storey, Human spinal cord differentiation proceeds rapidly in vitro and only initially maintains differentiation pace in a heterologous environment. bioRxiv 2021.02.06.429972 [Preprint]. 9 February 2021. https://doi.org/10.1101/2021.02.06.429972.
66. S. Roy, I. Wapinski, J. Pfiffner, C. French, A. Socha, J. Konieczka, N. Habib, M. Kellis, D. Thompson, A. Regev, Arboretum: Reconstruction and analysis of the evolutionary history of condition-specific transcriptional modules. *Genome Res.* **23**, 1039–1050 (2013).
67. M. Rodriguez-Guzman, J. A. Montero, E. Santesteban, Y. Gañan, D. Macias, J. M. Hurle, Tendon-muscle crosstalk controls muscle bellies morphogenesis, which is mediated by cell death and retinoic acid signaling. *Dev. Biol.* **302**, 267–280 (2007).
68. D. E. Russ, R. B. P. Cross, L. Li, S. C. Koch, K. J. E. Matson, A. Yadav, M. R. Alkaslasi, D. I. Lee, C. E. Le Pichon, V. Menon, A. J. Levine, A harmonized atlas of mouse spinal cord cell types and their spatial organization. *Nat. Commun.* **12**, 5722 (2021).
69. V. Y. Kiselev, T. S. Andrews, M. Hemberg, Challenges in unsupervised clustering of single-cell RNA-seq data. *Nat. Rev. Genet.* **20**, 273–282 (2019).
70. E. S. Lippmann, M. C. Estevez-Silva, R. S. Ashton, Defined human pluripotent stem cell culture enables highly efficient neuroepithelium derivation without small molecule inhibitors. *Stem Cells* **32**, 1032–1042 (2014).
71. F. P. Di Giorgio, G. L. Boulting, S. Bobrowicz, K. C. Eggan, Human embryonic stem cell-derived motor neurons are sensitive to the toxic effect of glial cells carrying an ALS-causing mutation. *Cell Stem Cell* **3**, 637–648 (2008).
72. T. Smith, A. Heger, I. Sudbery, UMI-tools: Modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.* **27**, 491–499 (2017).
73. V. P. Barkas, N. P. Kharchenko, E. Biederstedt, pagoda2: Single cell analysis and differential expression. R package version 1.0.5 (2021).
74. D. D. Lee, H. S. Seung, Learning the parts of objects by non-negative matrix factorization. *Nature* **401**, 788–791 (1999).
75. F. A. Wolf, P. Angerer, F. J. Theis, SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
76. C. J. Bult, J. A. Blake, C. L. Smith, J. A. Kadin, J. E. Richardson; Mouse Genome Database Group, Mouse Genome Database (MGD) 2019. *Nucleic Acids Res.* **47**, D801–D806 (2019).
77. P. D. Thomas, M. J. Campbell, A. Kejariwal, H. Mi, B. Karlak, R. Daverman, K. Diemer, A. Muruganujan, A. Narechania, PANTHER: A library of protein families and subfamilies indexed by function. *Genome Res.* **13**, 2129–2141 (2003).
78. M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
79. H.-P. Piepho, An algorithm for a letter-based representation of all-pairwise comparisons. *J. Comput. Graph. Stat.* **13**, 456–466 (2004).

Acknowledgments: We thank the University of Wisconsin-Madison Biotechnology Center Gene Expression Center and DNA Sequencing Facility for providing single-nuclei library preparation and next-generation sequencing services. scRNA-seq data were analyzed by the UW Bioinformatics Resource Center. We also thank C. Birchmeier and T. Müller for the gifts of Lbx1, Tlx3, and Lmx1b antibodies. We thank N. Fedorchak for technical and editing assistance. **Funding:** This work was supported by U.S. Environmental Protection Agency grant 83573701 (to R.S.A. and S.R.), NIH NIGMS grant R01 GM117339 (to S.R.), NIH NCATS grant UG3 TR003150 (to R.S.A.), NSF CAREER Award #1651645 (to R.S.A.), University of Wisconsin-Madison startup funds (to R.S.A. and S.R.), Innovation in Regulatory Science Award from BWF (to R.S.A.), UW

Downloaded from https://www.science.org at University of Wisconsin Madison on February 18, 2025

Data Science Foundation grant (to S.R.), UW Wisconsin Stem Cell and Regenerative Medicine Center postdoctoral fellowship (to N.R.I.), NIH NINDS postdoctoral fellowship F32 NS106740 (to N.R.I.), and James McDonell Foundation grant 3194-133-349500-4-AAB5159 (to J.S.).

Author contributions: Conceptualization: N.R.I., R.S.A., and S.R. Methodology: N.R.I., J.S., R.S.A., and S.R. Software: J.S. and S.R. Formal analysis: N.R.I. and J.S. Investigation: N.R.I., J.S., S.C., Y.T., T.E.D., N.R.N., F.S., and S.G.M. Resources: N.R.I., R.S.A., and S.R. Data curation, N.R.I., J.S., S.R., and R.S.A. Writing—original draft: N.R.I., J.S., R.S.A., and S.R. Writing—review and editing: N.R.I., J.S., R.S.A., and S.R. Supervision: R.S.A. and S.R. Project administration: R.S.A. and S.R.

Competing interests: R.S.A. is a co-founder of Neurosetta LLC, which uses the Hox protocol implemented in NMP derivation and caudalization. R.S.A. is also an inventor on U.S. patent application no. 14/496796, which claims discovery of the Hox protocol for NMP derivation and caudalization. All other authors declare that they have no other competing interests.

Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. scRNA-seq data are available through the GEO repository (GSE186698). Codes used to analyze the data are available at the following links: NMF-ANLS, Louvain clustering, the consensus clustering pipeline (https://doi.org/10.5281/zenodo.6505441), Arboretum (https://doi.org/10.5281/zenodo.6506161), and scFindTransitioning (https://doi.org/10.5281/zenodo.6506151). Files associated with the clustering analysis (including HOX profiles, primary clusters, and 17 subclusters) are available at the Mendeley Data repository (http://dx.doi.org/10.17632/fdnrn8br54.1) and can also be browsed at https://roy-lab.github.io/subcluster_analysis/ (https://doi.org/10.5281/zenodo.6506221).

Submitted 16 December 2021
Accepted 16 August 2022
Published 30 September 2022
10.1126/sciadv.abn7430

Downloaded from https://www.science.org at University of Wisconsin Madison on February 18, 2025