Title

The utilization of in vitro model systems to dissect the role of Smad1 and Smad5 during dorsal spinal cord development

ISBN

9798252408033

Author

Gallardo Dominguez, Salena

Publication Date

2026-06-12

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

The utilization of *in vitro* model systems to dissect the role of Smad1 and Smad5 during dorsal spinal cord development

A dissertation submitted in partial satisfaction of the requirements for the degree

Doctor of Philosophy in Molecular Biology

by

Salena Marie Gallardo Dominguez

2026

ABSTRACT OF THE DISSERTATION

The utilization of *in vitro* model systems to dissect

the role of Smad1 and Smad5 during dorsal

spinal cord development

by

Salena Marie Gallardo Dominguez

Doctor of Philosophy in Molecular Biology

University of California, Los Angeles, 2026

Professor Samantha Joanna Butler, Chair

Embryonic development relies on the coordinated action of multiple signaling pathways that guide pluripotent cells toward specific fates. Among these, Bone Morphogenetic Proteins (BMPs), are essential in the development of various tissues and organ systems. This versatility raises a fundamental question: how does a single signaling pathway generate such diverse responses? In this thesis, I investigate how downstream BMP signals are interpreted during dorsal spinal cord development, with a focus on the distinct contributions of Smad1 and Smad5. I summarize work that leverages *in vitro* differentiation systems to dissect BMP pathway mechanisms and demonstrate how these platforms can be further utilized to study dorsal interneuron (dI) specification.

First, we establish a directed differentiation protocol that allows us to generate the full complement of dorsal interneurons from mouse embryonic stem cells. This highly synchronized *in vitro* model system serves as a platform for probing BMP-dependent signaling events that are difficult to isolate during *in vivo* development of the dorsal spinal cord.

Next, I used this *in vitro* model system to investigate whether Smad1 and Smad5 transmit BMP signals in distinct ways during neural differentiation. We found that Smad5 acts

early in lineage specification, preventing bipotent neuromesodermal progenitors from adopting mesodermal fates and instead steering them toward neural derivatives. In addition, Smad5 is required for the specification of dP1–dP3 progenitors and their subsequent differentiation into dI1–dI3 interneurons. In contrast, Smad1 primarily functions to restrict the expansion of the dP1 lineage. Together, these findings demonstrate that R-Smads are not functionally interchangeable and instead contribute distinct regulatory activities during dorsal spinal cord development.

Finally, we extend these approaches to human systems by developing a neuromesodermal progenitor based differentiation protocol that produces human dIs spanning anterior–posterior identities. Our work illustrates how *in vitro* model systems can be used to interrogate developmental signaling pathways, highlights that the differential functions of R-Smads contributes to the promiscuity of the BMP signaling pathway and provides new insights into how dorsal spinal cord diversity is established.

The dissertation of Salena Gallardo Dominguez is approved.

Aparna Bhaduri

William Edward Lowry

Karen Marie Lyons

Bennett G Novitch

Samantha Joanna Butler, Committee Chair

University of California, Los Angeles

2026

Dedication

This work is dedicated to my son Augustine. May you always remain resilient in pursuit of your dreams and never impose limitations on what you believe you can achieve. Mom loves you, always.

Table of Contents

List of Figures……. viii

Acknowledgments……. x

VITA……. xii

Chapter 1- Introduction……. 1

Chapter 2 – *In vitro* atlas of dorsal spinal interneurons reveals Wnt signaling as a critical regulator of progenitor expansion……. 7

Abstract……. 7

Introduction…….. 7

Results…….. 10

Discussion……. 23

Material and Methods……. 28

Figures……... 43

Chapter 3 - Smad1 and Smad5 differentially mediate BMP signaling……. 57

Abstract……. 57

Introduction…… 57

Results…….. 60

Discussion…… 69

Material and Methods……. 73

Figures…….. 91

Chapter 4 – *In vivo* human embryonic spinal cord atlas validates stem cell–derived human dorsal interneurons and reveals ASD spinal signatures…… 108

Abstract……. 108

Introduction…….. 108

Results…… 112

Discussion…… 120

Material and Methods…….. 124

Figures…… 134

Chapter 5 – Conclusions…… 148

Appendix – Star Protocol…… 150

Bibliography…… 181

List of Figures

Figure 2-1 Distinct temporal combinations of RA and BMP4 direct different mESC identities……. 43

Figure 2-2 Single cell sequencing identifies relevant spinal function-specific modules in *in vitro*-derived-dIs…… 45

Figure 2-3 Stem cell-derived sensory interneurons resemble their endogenous counterparts……... 47

Figure 2-4 Wnt are upregulated as an immediate response to BMP4 signaling in mESC-derived NPCs……. 49

Figure 2-5 Wnt/β-catenin signaling is required for BMP4-mediated neuronal diversity…… 51

Figure 2-6 Wnt/β-catenin signaling mediates proliferation, not patterning, of mESCs derived spinal NPCs……. 53

Figure 2-7 Activating canonical Wnt signaling expands mESC-derived NPC populations... 55

Figure 3-1 Timeline of Smad signaling in the mouse dI directed differentiation protocol… 91

Figure 3-2 Generation of Smad1 and Smad5 knockout mESC lines……. 93

Figure 3-3 Timeline for bulkRNA-Seq experiment & NMP analysis…… 95

Figure 3-4 Transcriptional changes +6hr and +24hrs after Day4 BMP4 addition……. 98

Figure 3-5 The loss of Smad5 alters lineage specification during dI *in vitro* differentiation. 101

Figure 3-6 The loss of Smad1 and Smad5 differentially alters dI cell fate specification….. 103

Figure 3-7 The loss of Smad1 and Smad5 alters transcriptional dynamics……... 105

Figure 4-1 Construction of an *in vivo* human embryonic spinal cord scRNASeq reference atlas……. 133

Figure 4-2 Analysis of dI lineages within the *in vivo* human reference atlas……. 135

Figure 4-3 Specification of dI identity in the human RA±BMP4 NMP protocol…….. 137

Figure 4-4 Assessing the effect of culture duration on NMP identity along the anterior-posterior axis…….. 140

Figure 4-5 Assessing the effect of GDF11 treatment on both dorsal-ventral and anterior-posterior identity…….. 143

Figure 4-6 Comparison of *in vitro* data with the *in vivo* reference atlas reveals ASD signatures……. 145

Acknowledgments

I would like to thank Dr. Samantha Butler for her unwavering belief in my scientific capabilities. Her mentorship has been invaluable and has shaped my growth as a scientist. Most importantly, I leave her lab with my scientific curiosity not only intact but strengthened. I would also like to thank the members of my committee for their time, patience, and constructive feedback throughout my graduate studies.

I would also like extend my appreciation for the California Institute for Regenerative Medicine Bridges to Stem Cell Research program. They funded my first trainee position at the Sanford Consortium for Regenerative Medicine, affiliate of UC San Diego. This program has continued to train the next generation of stem cell biologists, some of whom I had the pleasure of training and working with during my graduate studies.

I gratefully acknowledge the many funding sources that have supported my graduate training and enabled the completion of this work. These awards have provided essential financial support, resources, and professional development opportunities that have contributed to my growth as a scientist. I sincerely thank the UCLA NIH Training Grant in Genomic Analysis and Interpretation (T32HG002536), the UCLA Whitcome Pre-Doctoral Fellowship in Molecular Biology, the Carl Storm Underrepresented Minority Fellowship, the UCLA Suzanne Eaton PhD Memorial Prize, and the UCLA Broad Center of Regenerative Medicine and Stem Cell Research Travel Award. I also acknowledge funding from the Marcus Foundation and NIH R01 NS123187 to S.J.B., which supported this research.

Chapter 2 of this dissertation is, with permission, a published manuscript:

Gupta, S., Kawaguchi, R., Heinrichs, E., **Gallardo, S**., Castellanos, S., Mandric, I., Novitch, B.G., and Butler S.J. "*In vitro* atlas of dorsal spinal interneurons reveals Wnt signaling as a critical regulator of progenitor expansion." *Cell reports* vol. 40,3 (2022): 111119. doi:10.1016/j.celrep.2022.111119

Chapter 3 of this dissertation is a manuscript that is in preparation:

Gallardo, S., et al. “Smad1 and Smad5 differentially transduce BMP signaling during *in vitro* differentiation of mouse embryonic stem cells into dorsal interneurons.

Chapter 4 of this dissertation is, with permission, a published manuscript:

Gupta, S., Heinrichs, E., Rodriguez, C., Friedman, E., **Gallardo, S**., et al. “*In vivo* human embryonic spinal cord atlas validates stem cell-derived human dorsal interneurons and reveals ASD spinal signatures” bioRxiv doi: 10.64898/2025.12.22.696129

Included in the appendix is, with permission, a published protocol:

Rodriguez, C., **Gallardo, S**., Gupta, S., & Butler, S. J. (2026). “Protocol for inducing dorsal spinal sensory interneurons from mouse embryonic stem cell-derived neuromesodermal progenitors.” *STAR Protocols*, *7*(1), 104307. doi: 10.1016/j.xpro.2025.104307

Vita

Education

B.S., Biochemistry, California State University San Marcos (CSUSM) June 2019

Research positions

Graduate Student Researcher, UCLA, Department of Neurobiology 2020- Present

Principal Investigator: Samantha Butler, PhD

Project: Investigating how bone morphogenetic proteins (BMPs) are differentially transduced by Smad1 and Smad5 during dorsal spinal cord development

Project: In vivo human embryonic spinal cord atlas validates stem cell-derived human dorsal interneurons and reveals ASD spinal signatures

Student Research Assistant, California State University San Marcos 2018-2019

Principal Investigator: Jane Kim, PhD

Project: Investigate if CRISPR-Cas9 can be used to contract/eliminate abnormally long repetitive microsatellite DNA sequences

CIRM Bridges to Stem Cell Research Internship (40hr/week), UC San Diego 2017-2018

Laboratory of Dr. Lawrence Goldstein

Project: Investigate the effect ATP binding cassette transporter subfamily A member 7 Alzheimer's Disease risk variants have on the function of human neurons, microglia, and astrocytes

Publications

Gupta, S., Heinrichs, E., Rodriguez, C., Friedman, E., **Gallardo, S**., et al. "*In vivo* human embryonic spinal cord atlas validates stem cell-derived human dorsal interneurons and reveals ASD spinal signatures" bioRxiv doi: 10.64898/2025.12.22.696129

Rodriguez, C., **Gallardo, S**., Gupta, S., Butler S.J. “Protocol for inducing dorsal spinal sensory interneurons from mouse embryonic stem cell-derived neuromesodermal progenitors” *STAR protocols.* Accepted

Gupta, S., Kawaguchi, R., Heinrichs, E., **Gallardo, S**., Castellanos, S., Mandric, I., Novitch, B.G., and Butler S.J. “*In vitro* atlas of dorsal spinal interneurons reveals Wnt signaling as a critical regulator of progenitor expansion.” *Cell reports* vol. 40,3 (2022): 111119. doi:10.1016/j.celrep.2022.111119

Select scientific presentations

Gallardo, S., Rodriguez, C., Gajardo Del Real, G., Gupta, S., Butler, S.J. Elucidating how BMP signaling is differentially transduced by Smad1 and Smad5 during dorsal spinal cord development. Gordon Research Conference – Neural Development. Luca, Italy. August 2024

Gallardo, S., Gupta, S., Butler, S.J. Investigating how BMP signaling remodels chromatin accessibility and gene expression during differentiation of mouse embryonic stem cells into spinal dorsal interneurons. National Human Genome Research Institute (NHGRI) annual trainee meeting. Salt Lake City, Utah. April 2023

Gallardo, S., Kolhnhofer, B., Goldstein LSB. Using a CRISPR-CAS9 Genetically Modified ABCA7 knockout Human IPSC Line to Model Late Onset Alzheimer’s Disease. International Society for Stem Cell Research (ISSCR) Annual Meeting. Melbourne, Australia. June 2018

Awards and Honors

Suzanne Eaton PhD Memorial Prize, 2025 (UCLA)

Carl Storm Underrepresented Minority (CSURM) Fellowship

Microscopy Image Selected- Brain Research Institute Calendar, 2023 (UCLA)

Whitcome Pre-Doctoral Fellowship in Molecular Biology, 2023 (UCLA)

Broad Center of Regenerative Medicine and Stem Cell Research Travel Award, 2023 (UCLA)

Genomic Analysis Training Program NIH-T32 Fellowship, 2022 (UCLA)

New Ventures Fellow, 2022-2024 UCLA Technology Development Group

Chapter 1 – Introduction

During early embryogenesis, signaling from distinct families of growth factors act on pluripotent cells to direct their proliferation, differentiation, and migration. The coordinated activity of these signaling pathways drives the formation of diverse tissues and organ systems. Bone Morphogenetic Proteins (BMPs), a major branch of the Transforming Growth Factor-β (TGF-β) superfamily, are integral to skeletal, nervous, and cardiovascular system development[1-3]. The reiterative use of BMP signaling reflects the pathway's capacity to mediate a wide array of context-dependent cellular behaviors. Despite this broad functional repertoire, the downstream mechanisms that generate diverse BMP-dependent responses remain incompletely defined.

The canonical BMP signaling pathway

There are over 20 known BMP ligands that bind to and activate a heteromeric complex of type I and type II serine/threonine kinase BMP receptors (BMPRs)[4]. There are three type I receptors: type 1A BMP receptor (BMPR-1A or Alk3), type 1B BMP receptor (BMPR-1B or ALK6), and type 1A activin receptor (ActR-1A or ALK2) and three type II receptors: (type 2 BMP receptor (BMPR-2), type 2 activin receptor (ActR-2A), and type 2B activin receptor (ActR-2B) that are known to interact with BMPs. Activated Bmprs phosphorylate the receptor regulated (R) Smads at a C-terminal SSXS motif, which then complex with common mediator (Co) Smads and translocate to the nucleus, whereby R-smads regulate transcriptional activity of the cell[5, 6]. The diversity of the machinery proteins involved in the BMP signaling cascade allow BMPs to act in a signal specific manner. For example, different type I Bmprs translate distinct BMP activities during development[7]. Furthermore, inhibitory (I) Smads are able to prevent the R-Smad/Co-Smad complex from translocating to the nucleus, and thus block transcriptional changes mediated by R-Smads from occurring[8].

There are several molecules that are known to modulate the BMP signaling pathway. Antagonists include Chordin, Noggin, and the DAN (differential screening-selected gene in neuroblastoma) family of proteins, attenuating the pathway by extracellularly sequestering BMP ligands, preventing them from engaging with receptors[9, 10]. Membrane level antagonists can interfere with receptor availability, activation, and receptor-Smad coupling. This includes BAMBI (bmp and activin membrane-bound inhibitor), as well as I-Smad recruitment of Smurf1/2 to BMP receptors, leading to their ubiquitination and subsequent degradation[11, 12].

Although BMP signaling consists of a diverse array of ligands and receptors, along with multiple layers of regulation, these varied inputs ultimately funnel into the activation of a limited set of R-Smads. This raises the question of how few R-Smads are capable of directing a wide range of cellular responses attributed to BMP signaling.

Receptor activated Smads (R-Smads) as mediators of BMP signaling

Upon activation and translocation to the nucleus, Smad1/5/8 can all bind to DNA, through their N-terminal MAD homology 1 (MH1) domain[13, 14]. Smad binding elements (SBEs) are sequence specific motifs that identify Smad target genes. Smad1/5 can recognize the ‘CAGAC’ SBE and has additionally shown a preference for GC-rich motifs. Smads on their own have a very low DNA binding affinity, and the interaction with other cofactors, through the conserved Mad homology 2 (MH2) domain, strengthens their interaction with DNA[15-17]. Specific combinations of Smad-cofactor complexes can simultaneously regulate a distinct group of genes, known as synexpression groups. Master transcription factors, determined by cell type or lineage specificity, can also work synergistically with activated R-Smads and open up local chromatin to make SBEs more accessible[18].

The question of how so few R-Smads are capable of directing a wide range of cellular responses is compounded by the longstanding view that these transcriptional regulators

function redundantly. Smad1, Smad5, and Smad8 each contain a proline-rich linker region connecting the MH1 and MH2 domains, but this region harbors distinct patterns of post-translational modification sites across the three R-Smads[19]. These differences suggest that Smad1/5/8 may be differentially regulated in ways that alter their stability, turnover, or nuclear shuttling. Moreover, Smad1, Smad5, and Smad8 can exhibit non-overlapping expression patterns across tissues and developmental contexts, raising the possibility that they are not fully functionally redundant in their ability to mediate BMP signaling[20, 21].

BMP signaling during dorsal spinal cord development

Dorsal spinal cord development provides an ideal window in vertebrate embryogenesis to investigate the downstream mechanisms of BMP signaling. During neurulation, the neural plate thickens and invaginates, resulting in the formation of the neural tube along the rostrocaudal axis of the embryo. The posterior region will develop into the spinal cord. Here, the roof plate serves as a major signaling center, secreting multiple BMP ligands including BMP4, BMP5, BMP6, BMP7, and GDF7 (BMP12). Contributing signals from the floor plate (Shh), and surrounding somites in the paraxial mesoderm (retinoic acid) give rise to distinct groups of neural progenitor cells (NPCs) that line the ventricular zone along the dorsal ventral axis. The multipotent NPCs then begin to divide and differentiate as they migrate laterally into the mantle layer. This process results in the segregation of distinct neuronal subtypes, patterning the different laminae of the spinal cord. The dorsal most region is comprised of six classes of dIs (dI1-dI6) which are responsible for processing and relaying somatosensory information from the periphery to the brain, as well as modulating proper motor neuron output activity[22]. The six progenitor populations, as well as the six dI subtypes that they give rise to can be identified through the combinatorial expression of transcription factors.

BMP signaling from the roof plate is essential for proper dorsal interneuron specification[23, 24]. *In vivo* chicken electroporation experiments knocking down BMP4 and BMP7 result in lowered dI1, dI3, and dI5 populations, to varying degrees[25]. We have also shown that BMPs direct dI identity in a signal-specific manner, dependent on the BMP ligand and type I BMP receptor[7]. These observations underscore the central role of BMP signaling in establishing dorsal interneuron identities, emphasizing the need to understand how this signal is interpreted at the level of R-Smads. In the developing dorsal spinal cord, the focus of our lab and of this dissertation, the relevant R-Smads are Smad1 and Smad5.

The role of Smad1 and Smad5 in dorsal progenitor and dorsal interneuron specification

Smad1 and Smad5 display distinct expression patterns during dorsal spinal cord development. Smad5 is enriched in neural progenitors within the ventricular zone, whereas Smad1 is more prominent in differentiated neurons[20]. These differences raise the possibility that Smad1 and Smad5 may perform non-redundant functions in mediating BMP signaling. However, defining their specific roles *in vivo* has proven challenging. Numerous studies in Xenopus, chicken, and mouse models have shown that depletion of Smad1/5 disrupts dorsal progenitor formation, dorsal interneuron specification, and cell-cycle dynamics, yet these experiments often yield conflicting results and do not resolve the phenotypic discrepancies observed between Smad1 and Smad5 specific loss-of-function approaches[20, 25-27]. Technical limitations—including perduring gene function, off-target RNA inhibition, and interference from background endogenous signaling—further complicate interpretation. As a result, the differential contributions of Smad1 and Smad5 to dorsal spinal cord development remain incompletely understood

Summary

BMP signaling from the roof plate is essential for proper dorsal interneuron specification and localization. The diversity of this signaling pathway, consisting of numerous BMP ligands and receptors, becomes bottlenecked at the receptor activated Smads, which have largely been attributed to act redundantly. *In vivo* studies examining Smad1/5 during spinal cord development have not clearly defined the extent to which they function uniquely or in a shared capacity. BMP signaling is a conserved regulator of tissue development across multiple species. By dissecting the independent functions of Smad1 and Smad5, you can uncover core mechanisms by which cells across diverse tissues and species interpret reiterative BMP signals.

In chapter 2 we define a directed differentiation protocol that recapitulates the *in vivo* dorsal spinal cord signaling environment, allowing us to derive bona fide dIs from mouse embryonic stem cells (mESCs). We found that mESC-derived dIs resemble endogenous dIs, with the correct molecular and functional signatures. This innovative model system allows us to investigate BMP signaling in a large population of cells behaving relatively synchronously, compared to the developing embryo.

In chapter 3 we use our *in vitro* model system to test the hypothesis that Smad1 and Smad5 differentially transduce BMP signaling during mESC to dI differentiation. We employ CRISPR-CAS9 gene editing technology to generate Smad1Δ and Smad5Δ mESC lines. Extensive bulk and scRNA seq experiments across multiple developmental timepoints within our dorsal interneuron *in vitro* differentiation protocol reveal Smad1 and Smad5 temporally drive distinct cell fate decisions.

In chapter 4 we expand our work in chapter 2 and establish a neuromesodermal progenitor-based differentiation approach that generates human dIs spanning anterior-posterior identities. We compared the transcriptomes of *in vitro* dI4/dI5 subclasses to *in vivo* dIs stemming from

human embryonic spinal cord tissue. We found not only are they similar but they harbor mechanosensory circuit signatures linked to autism spectrum disorder, implicating spinal circuits in autistic phenotypes. This extension of our work has been included to show how *in vitro* model systems can be used to further clarify how human dI diversity is established.

Chapter 2 – *In vitro* atlas of dorsal spinal interneurons reveals Wnt signaling as a critical regulator of progenitor expansion

Abstract: Restoring sensation after injury or disease requires a reproducible method for generating large quantities of bona fide somatosensory interneurons. Toward this goal, we assess the mechanisms by which dorsal spinal interneurons (dIs; dI1-dI6) can be derived from mouse embryonic stem cells (mESCs). Using two developmentally relevant growth factors, retinoic acid (RA) and bone morphogenetic protein (BMP) 4, we recapitulate the complete in vivo program of dI differentiation through a neuromesodermal intermediate. Transcriptional profiling reveals that mESC-derived dIs strikingly resemble endogenous dIs, with the correct molecular and functional signatures. We further demonstrate that RA specifies dI4-dI6 fates through a default multipotential state, while the addition of BMP4 induces dI1-dI3 fates and activates Wnt signaling to enhance progenitor proliferation. Constitutively activating Wnt signaling permits the dramatic expansion of neural progenitor cultures. These cultures retain the capacity to differentiate into diverse populations of dIs, thereby providing a method of increasing neuronal yield.

Introduction

Spinal cord injuries (SCIs) can result in the loss of both coordinated movement and somatosensation [28, 29], and the ability to perceive the environment. Without somatosensation, patients can no longer communicate through touch, sense danger through pain, or control the functions of the gut and bladder [30]. While stem cell-based cellular replacement therapies have shown promise for SCI [31, 32], the focus so far has been on the recovery of coordinated motor function [33, 34]. Less attention has been paid to regaining sensory functions, despite their critical roles ensuring physical and emotional well-being, and providing modulatory feedback to the motor system through proprioceptive pathways. Somatosensory information is received peripherally and then relayed centrally by different populations of dorsal interneurons (dIs) in the

spinal cord. Each dI population is specialized for distinct somatosensory modalities, and forms microcircuits that span different layers of the dorsal horn [35, 36]. For example, dI1s (proprioceptors) [37], dI2s and dI3s (both mechanosensors) [38, 39] and dI6 (gait) [40] are located in the deep layers of the dorsal horn, while dI4s and dI5s regulate pain, heat, and itch in the superficial layers of the dorsal horn [41]. To replace damaged spinal tissue, it will be critical to derive a complete complement of dorsal sensory interneurons that molecularly and functionally phenocopy endogenous dIs.

Towards this goal, we are working to identify the developmental mechanisms that specify sensory dorsal interneurons (dIs) in the spinal cord, and then apply these insights to develop protocols to derive dIs from stem cells [42-44]. Distinct classes of dIs arise during embryonic development, when growth factors, including retinoic acid (RA) and the bone morphogenetic proteins (BMPs), pattern six progenitor domains (dP1-dP6), which differentiate into the dI1-dI6s [36, 45]. The addition of RA and BMP4 in embryoid body (EB) protocols [46], can induce some dI fates in both mouse and human stem cell cultures [42, 44, 47]. However, it has remained unclear how the addition of RA and BMP4 specifically induces dI fates, and whether the dIs generated from these protocols mirror their endogenous counterparts. BMPs pattern multiple organ systems throughout development, including the non-neuronal cardiac mesoderm [48, 49] as well as osteogenic tissues [50]. The BMPs also have reiterative activities both patterning dPs and controlling their proliferation to ensure that the precise number of specific dIs are generated [44, 51]. While the Smad second messenger complex mediates dI patterning [52, 53] the downstream signals regulating proliferation remain unresolved.

Here, we define RA±BMP4 protocols that generate mouse embryonic stem cell (mESC)-derived dIs via a neuromesodermal progenitor (NMP) intermediate [54] that faithfully mimic the normal developmental programs of the neural tube [55-57]. This approach has successfully generated a complete atlas of *in vitro*-derived dIs. Transcriptional profiling of these dIs demonstrated that

they possess sensory function-specific gene signatures that strikingly resemble endogenous dIs, indicating that our protocols generate *bona fide* dIs. Using these protocols as an *in vitro* model for dorsal spinal cord development*,* we identified the hierarchy of fate specification decisions mediated by RA and BMP signaling. We further identify Wnt/β-catenin signaling as an immediate downstream response to BMP signaling which maintains neural progenitors in a mitotic state. Remarkably, elevating Wnt signaling, using the CHIR99021 (CHIR) agonist, can dramatically extend the proliferative capacity of dPs, while preserving their ability to differentiate into specific sensory neurons on demand.

Taken together, this study provides a mechanistic understanding of the developmental trajectories needed to generate the full complement of different dI identities. These insights were used to develop a protocol to expand the propagation of multipotential dPs, thereby paving the way for the for the production of *in vitro*-derived dIs needed for both drug screening and cellular transplantation therapies.

Results

Establishment of 2d-protocols to direct mESCs towards dorsal spinal interneuron fates.

To direct dI fates through an NMP intermediate, we adapted a mESC protocol previously used to derive ventral spinal neurons [54, 55]. In this protocol, treatment with basic fibroblast growth factor (bFGF) and the Wnt agonist CHIR (Figure 1B), results in ~ 90% of mESCs acquiring a bipotential NMP identity by day 3 in culture (Figure S1A) distinguished by the co-expression of both mesodermal (brachyury (T)) and neural progenitor cell (NPC, Sox2) markers (Figure S1A). Using the developmentally relevant growth factors (Figure 1A) previously shown to direct ESCs towards dorsal spinal fates [42, 44, 47, 58], we then treated day 3 NMPs with 100nM RA for 48 hours either alone (protocol 1, Figure 1B) or in combination with 10ng/ml BMP4 starting at either day 3 (protocol 2) or day 4 (protocol 3). Treatment with RA alone did not activate BMP signaling (distinguished by the absence of phosphorylated (p) Smad1/5/8 staining; Figure 1C) and resulted in dorsalized neural cultures, with $Pax3^+$ $Pax6^+$ $Sox2^+$ neural rosettes extending $Tuj1^+$ neurites (arrowheads, Figures 1F, 1I). When BMP4 was added in combination with RA, BMP signaling was now activated (arrowheads, Figures 1D, E). However, the timing of BMP4 addition critically affected NMP identity. When BMP4 was added concomitantly with RA at day 3, no $Pax3^+$ (Figure 1G) or $Sox2^+$ (Figure 1J) cells were observed. The aggregates exhibited Tuj1 staining, but had no apparent neuronal morphology (Figure 1J). In contrast, when RA was added at day 3, followed by BMP4 at day 4, we again observed $Pax3^+$ $Pax6^+$ $Sox2^+$ neural rosettes, which robustly extended $Tuj1^+$ neurites (arrowheads, Figures 1H, 1K).

Unbiased transcriptomic analyses identify the developmental trajectories leading to either spinal cord or cardiac fates.

To assess the transcriptional changes resulting from the timed addition of RA±BMP4, we conducted bulk RNA-Seq analyses along the timelines of the three protocols (Figure 1L). Principal component analysis (PCA) [59] revealed that PC1 and PC2 contributes 44% of the

variation in the data, and separates the least (D0, D3) from the most (D7, D9) differentiated samples (Figure S1B), while PC3 contributes 15% variance and separates the samples from different protocols (Figure S1C). A three-dimensional PCA plot combining PC1-PC3 reveals three differentiation trajectories (Figure 1M), connected by bifurcation points at day 3 and day 4. The C-branch (RA+BMP4, protocol 2) follows a distinct trajectory from the R-branch (RA, protocol 1) after bifurcating at day 3, while the B-branch (RA+BMP4, protocol 3) bifurcates from the R-branch at day 4 (Figure 1M). The R- and B- branches follow similar trajectories and terminate in adjacent PCA spaces, distinct from the C-branch, suggesting that their differentiation program is largely equivalent. Supporting this conclusion, a Pearson correlation demonstrated R- and B- branch transcriptomes were the most similar to each other (Figure 1N). We next examined whether the three protocols direct distinct differentiation outcomes, by assessing the groups of genes upregulated over time using weighted gene co-expression and network analysis (WGCNA) [60]. Distinct eigengene modules [61, 62], i.e., a collection of genes with similar expression trends, were identified for each branch (Figure S1E). Of these, the *yellow-green, dark-grey* and *green* modules show increasing expression from day 0 to 9 for protocols 1, 3, and 2 respectively (Figure 1O). A functional enrichment analysis of these modules using Enrichr [63] revealed that. both the *yellow-green* and *dark-grey* gene modules were highly enriched for nervous system-specific gene ontology (GO) terms (z-score>30) (Figure 1O), supporting the hypothesis that protocols 1 and 3 promote neural identities. In contrast, the top GO term in the *green* module is “outflow tract septum” (Figure 1O), a structure that regulate the blood flow in the heart [64]. Other cardiac-related terms, such as actomyosin, contractile actin filament, and stress fibers terms are represented in the cellular component categories (Figure S1D). We additionally observed that multiple cardiac-specific genes are upregulated by day 9 in cells derived from protocol 2 (Figure S1F). Taken together, these analyses reveal the timeline

over which RA and BMPs act sequentially to direct NMPs towards the neural lineages, rather than endocardial fates.

ScRNA-Seq of mESC-derived neurons identifies the full complement of spinal dorsal interneurons.

To assess whether protocols 1 and 3 direct distinct classes of dIs, we obtained single-cell transcriptomes from 9,704 cells using protocol 1 and 13,121 cells using protocol 3, taken at day 9 of the differentiation procedure (Figure 2A). We performed Louvain clustering followed by Uniform Manifold Approximation and Projection (UMAP) plots, to visualize groups of transcriptionally distinct cell types [65]. This pipeline identified 21 clusters of cells in both protocol 1 (Figure 2C) and 3 (Figure 2D). In both conditions, ~30% of cells identify as being stressed; they cluster separately and display <2000 RNA counts (Figures 2C, 2D, S2A, B). ~2-4% cells are pluripotent stem cells, marked by *Sox2*, *Pou5f1 (Oct4)* and *Nanog* expression, while 5-8% cells are non-neural, either of mesodermal or cardiac neural crest identity, expressing *Twist1*, *Runx2,* and *Hand2* [66-68] (Figures 2C, 2D, S2A, B). The remaining ~60% cells have a neural identity, either *Sox2*$^+$ NPCs or *Tubb3*$^+$ (class III β-tubulin, Tuj1) differentiated neurons (feature plots, Figures 2C, 2D).

To distinguish the neuronal subtypes, we subclustered the *Tubb3*$^+$ *c*ells, resulting in 14 (protocol 1) and 13 (protocol 2) transcriptionally distinct clusters (Figures 2E, 2F). We then used a well-characterized panel of transcription factors (Figure 2B) to assign dI identities to these clusters [35, 36]. This analysis revealed that protocol 1 primarily generates the dI4, dI5 and dI6s that mediate pain, itch, and heat perception, while protocol 3 most notably generates the dI1, dI2 and dI3s that regulate proprioception, gait, and mechanosensation. Specifically, in protocol 1 the majority of subclustered *Tubb3*$^+$ cells divide into two major groups of closely aligned clusters (Figures 2B, 2E, S2C): one group (clusters 7, 2, 0, and 1) expresses *Lmx1b, Prrxl1*, and *Tlx1/3* which

define dI5 identity [35]. The second group (clusters 6, 12, 13, 4, 5, 3) co-express *Pax2* and *Lhx1* which denotes dI4/dI6 identity. Of these clusters, only cluster 6 also expresses *Dmrt3*, an established marker of dI6s [69]. The neurotransmitter profile supports these fate designations: phenocopying endogenous neurons (Figure 2B) the dI4/dI6 clusters express *Gad2*, a GABA synthesizing enzyme [70], while the dI5 clusters express *SLC17a6 (vGlut2),* a glutamate transporter specific to excitatory neurons (Figure 2E) [71]. Of the remaining four clusters, cluster 8 and 9 map to dI1 and dI2 identities (Figure S2C), while cluster 10 is enriched for neural and neuroendocrine specific genes, and cluster 11 is enriched for ribosomal genes (Table S1).

In protocol 3, the subclustered *Tubb3*$^{+}$ cells divide into three groups of closely aligned clusters (Figure 2B, 2F, S2D): clusters 7, 1, 0 and 8 express *Lhx2/9* and *Barhl1/2*, denoting dI1 identity, clusters 11, 5 and 6 express the dI2 markers *Foxd3, Lhx1* and *Lhx5*, while clusters 12 and 10 express *Isl1* and *Tlx3*, indicating dI3s. As *in vivo*, all three classes of *in vitro*-derived neurons are excitatory, expressing either *Slc17a6* or *Grin2b* (Figure 2F). Of the remaining four clusters, clusters 2 and 9 express *Atoh1* and *Neurog1* respectively, which mark the dP1 and dP2 state, and suggests these cells are immature dI1 and dI2s (Figure S2D). Cluster 4 expresses dI4-specific markers (Figure S2D) and cluster 3 express multiple ribosomal and mitochondrial genes indicative of dying neurons [72] (Table S1).

In summary, protocol 1 (RA) generates ~ 33% dI4s, 40.5% dI5s, 8% dI6s, 5% dI1s, 5% dI2s, as well as 8.5% unknown neural cell types (pie chart, Figure 2E). In contrast, protocol 3 (RA+BMP4) generates ~45.5% dI1s, 29% dI2s, 8% dI3s, 8% dI4s, as well as 9.5% unknown cell types (pie chart, Figure 2F). Both immunohistochemical and qRT-PCR analyses further support these designations (Figure S2E-S2H).

GO analyses of ESC-derived dIs identifies sensory modality-specific ontology modules

We next assessed whether the functional identities of these *in vitro* derived neurons could be predicted from the transcriptomic data. *In vivo*, the dIs relay diverse somatosensory modalities either locally within the spinal cord [73, 74], or by long-range afferent connections [75]. Generally, the most dorsal populations regulate proprioception (dI1/dI2) [76, 77] and touch-induced mechanosensation (dI3s) [38] while the more intermediate classes form relay circuits for pain, itch, heat, and touch (dI4/dI5) [41, 78]. The dI6s regulate coordinated movement by forming inhibitory synapses with spinal motor neurons, while themselves receiving cholinergic and glutamatergic inputs [40].

To identify functionally relevant gene ontologies, we selected the cluster for each dI population that highly expressed the differentiation markers (>0.2 log fold change), and subjected them to the DAVID pipeline [79, 80], which identifies enriched biological processes and signaling pathways related to diseases or drugs (Figure 2G). Remarkably, each class of ESC-derived dIs displayed the relevant modality-specific signatures. For the dorsal-most neurons, we analyzed cluster 0 (dI1), cluster 5 (dI2), and cluster 10 (dI3) and specifically found enriched GO categories related to balance and walking behaviors, as well as excitatory synaptic signaling. Both dI1 and dI2 clusters additionally include proprioception-related terms (blue terms, Figure 2G, Table S2). For the more intermediate neurons, we analyzed cluster 4 (dI4), cluster 1 (dI5), and cluster 6 (dI6) The dI4/dI5 clusters include GO terms associated with pain and itch perception, specifically implicating both oxytocin, a neuropeptide that modulates pain processing [81] and β-alanine, an amino acid that induces itch [82]. As *in vivo*, the dI4 cluster contains terms related to inhibitory synaptic signaling, while the dI5 cluster is enriched for excitatory synaptic terms (red terms, Figure 2G, Table S2). In contrast, the dI6 cluster contains terms associated with coordinated movement, and synaptic signaling pathways, accurately reflecting their functional identity. Finally, we also found enriched terms in the dI clusters related to different drug addiction pathways, including amphetamine and endocannabinoids for dI1/dI2s, and amphetamine,

morphine and cocaine for dI4/dI5s (Figure 2G, S3A, Table S2). Such signatures suggest the dIs mediate the ability of psychoactive drugs to modulate pain and itch perception [83] or proprioception [84]. Other dI-subtype markers identified in this analysis include receptors, ion channels and adhesion molecules (Figure S3A, S3B; Table S4).

ESC-derived dIs are transcriptionally indistinguishable from endogenous spinal interneurons

To assess whether the ESC-derived dIs are also transcriptionally similar to their *in vivo* counterparts, we used reciprocal PCA based data integration and label transfer [85] to compare a single-cell dataset taken from embryonic stage (E) 9.5-E13.5 mouse spinal cords [57] with the *in vitro*-derived cell types in protocols 1 and 3. Projecting the combined *in vivo* and *in vitro* datasets into UMAP plots revealed a remarkable degree of overlap (Figure 3A, 3B, 3C). This overlap was only observed for the neural progenitor and neural identities; the blood, mesoderm, neural crest and skin lineages present in the *in vivo* dataset were notably absent from the *in vitro* datasets (Figure 3A). While cells from protocol 1 and 3 mapped directly on top of the spinal cord cells (Figure 3B), these three datasets were distinct from trachea, lung, and kidney cells [86] (Figure 3E).

To compare the distribution of cell types in the *in vivo* and *in vitro* datasets, we performed unsupervised clustering on an integrated dataset, yielding a total of 29 clusters (Figure 3C); 24 clusters (0-10,12-25) were common to all datasets, while 6 clusters were unique to the spinal cord dataset. A dot plot analysis of dI-specific markers within the shared 24 clusters showed striking similarities between the *in vivo* and *in vitro*-derived datasets (Figure 3D). While there are modest differences, there is overall concordance between both the number of cells expressing a given dI marker, and the strength of marker expression in the *in vivo* and *in vitro* cell clusters. This concordance continues as the neurons mature, demonstrated by the expression of neuropeptides, which modulate the intensity of sensory inputs in the adult spinal cord [35, 87]. Specific neuropeptides, including prepronociceptin (*pnoc*; dI2, dI4), neuromedin S (*nms*; dI4)

neuropeptide Y (*npy,* dI4), gastrin releasing peptide (*grp;* dI1, dI5), are enriched in the same populations of dIs in both the *in vitro* and *in vivo* datasets (Figure S3C, S3D). Thus, the ESC-derived dIs appear to proceed along the same maturation process as endogenous spinal neurons.

Finally, to assess whether the *in vitro*-derived dI populations specifically mirror their endogenous counterparts, we compared the protocol 1 vs. 3 datasets to an atlas of the neuronal classes of the spinal cord (Figure 3F), generated using the annotation provided in [57]. The *in vivo* dI1-dI3s, and dI4-dI6s, map to identical positions in protocol 3 (RA+BMP4) and protocol 1 (RA) respectively (Figure 3F). In contrast, there was no overlap between the ventral spinal neurons, such as the motor neurons (MNs), with the *in vitro* datasets (Figure 3F and data not shown). We also computationally extracted the cells expressing specific dI subtype markers from the *in vitro* and *in vivo* datasets and projected them in the same UMAP space (Figure 3G). This analysis further demonstrated that all six classes of *in vitro*-derived dIs are transcriptionally indistinguishable from their spinal cord-derived counterparts.

Taken together, these studies suggest that our *in vitro* protocols are accurately replicating the *in vivo* developmental program of the spinal cord permitting the generation of *bona fide* sensory interneurons.

RA activates a transcriptional network that specifies a dI4-dI6 default state

Using the directed differentiation protocols, we investigated the unresolved mechanisms by which RA and BMP direct dI fate specification. We first assessed the genetic basis of RA-mediated fate decisions. During the 24hr pulse of RA at day 3, *Pax3* levels gradually increase, while *T* expression concomitantly declines (Figure S5A). By day 4, ~90-95% of cells are $Pax3^{+}$ (Figure S5B), indicating NMPs have transitioned to a dorsal neural progenitor state (dP state 1) (Figure S4A). To identify the transcriptional changes occurring during this first commitment step,

we performed bulk RNA-Seq at day 3 (NMP state), day 3.25 (early transcriptional response to RA) and day 4 (dP state1) (Figure S4B).

The differential expression analysis identified 138 upregulated and 19 downregulated genes after 6hrs of RA treatment (Figure S4C). Upregulated genes include known RA-regulated factors, such as *Meis1, Meis2, Stra8*, *Cyp26a1*, and *Rarb*, confirming that cells are directly responding to RA. By 24 hours of RA exposure, 389 genes were significantly upregulated including genes present in intermediate dPs and the dI4/dI5s (Figure S4D) while 590 genes were significantly downregulated, which include mesodermal specific genes (Figure S4E-S4G). These findings support the hypothesis that RA induces NMPs towards intermediate spinal identities, while also directing them away from mesodermal fates, thereby ensuring the transition from an NMP to $Pax3^{+}$ dP state 1.

We next assessed whether the dP state1 can be defined as a function-specific interaction network, by subjecting the upregulated genes to a Metascape protein-protein interaction (PPI) analysis [88]. At 6hrs, a nascent PPI network exists, with one functional module for embryonic patterning containing the RA-target genes *Meis1* and *Meis2* (Figure S4H, inset). However, by 24hrs, the PPI network has expanded into 13 functional modules (Figures S4I, S5C). Four modules, *insulin-like signaling*, *pattern specification, sensory organ morphogenesis*, and *peptide ligand receptor,* contained genes expressed in the intermediate spinal cord (insets, Figures S4I, S5C). Notably, *Meis1* and *Meis2* interact with *Pax6, Lmx1b, Msx1* and *Msx2*, in the *pattern specification* module. Using both qRT-PCR (Figure S5D) and transcriptional profiling of untreated day 4 cells (no RA control), or after a 24-hour pulse of RA ±AGN193109, a pan-RAR inhibitor (Figure S4J), we independently validated that *Meis1* and *Meis2*, but not *Meis3*, are upregulated by RA including many patterning genes. Thus, Meis1/2 are good candidates for the key transcription factors that link RA signaling to the initiation of spinal cord patterning, i.e. dP state1.

Wnt signaling is induced as an immediate consequence of BMP4 treatment

We hypothesized that the addition of BMP4 to NPCs in dP state1, directs them to a more dorsal dP state, i.e. dP state2 (Figure S4A). To define the transcriptional changes occurring after BMP4 addition, we performed bulk RNA-Seq at day 4 (dP state1), day 4.25 (early transcriptional response to RA+BMP4) and at day 5 (dP state2) (Figure 4A). We identified 106 genes upregulated by 6hrs and 305 genes upregulated after 24hrs of BMP4 exposure (Figure S6A). 78 genes, out of 411, were common between both time points (Figures 4B, S6B). The GO analysis of these 78 common genes included the *BMP signaling pathway* (Figure 4C, Table S3), confirming that the analysis identified BMP responsive genes.

Additionally, we also found enriched GO terms related to Wnt signaling across three functional categories (Figure 4C). Both Wnt ligands, including *Wnt1, Wnt2, Wnt3, Wnt3a* and *Wnt4*, and Wnt receptors, *frizzled (Fzd) 8, Fzd3*, *Fzd10* and *Lrp6* (Figure S6C) were upregulated in response to BMP4 treatment along with Wnt signaling regulators, such as *Dkk, Lmx1a* [89] and *Bambi* [90]. (Figures 4D, 4E, S6B). Selected Wnt-signaling genes were further assessed in a qPCR analysis of day 5 NMPs to validate whether they are BMP4-responsive. We found that expression of all tested genes significantly increased after the addition of BMP4, compared to the RA alone condition (Figure 4F). The upregulation of *Wnt* expression was suppressed if Noggin, a BMP inhibitor, was concomitantly added with BMP4 on day 4 (Figure 4F).

Many of the upregulated Wnts, *Wnt1, Wnt3a,* and *Wnt4,* are expressed in the developing dorsal spinal cord *in vivo* [91, 92] and signal through the canonical Wnt signaling pathway [93], suggesting that canonical Wnt signaling mediates the ability of BMP4 to direct the dP state1 to dP state2 transition.

Inhibiting canonical Wnt signaling blocks the induction of the dorsal-most fates

To assess the role of canonical Wnt signaling in dI fate specification, we added endo-IWR1 (IWR1e), a small molecule inhibitor which blocks β-catenin function [94] to dPs from day 4-6 (Figure 5A). By day 9, the RA+BMP4 condition robustly induces dI1s-dI3s (Figure 5F-5I, 5N-5P). In contrast, the addition of IWR1e resulted in the loss of dI1s and dI2s (Figure 5J, 5K, 5N, 5O) while dI3 differentiation was less affected (Figure 5L, 5P).

BMP4 directs the dorsal-most fates (dI1-dI3), while concomitantly suppressing RA-dependent intermediate dorsal fates (dI4-dI6, Figures 5E, I, S6D-S6E). Supporting a role for the Wnts mediating this latter activity, there is no suppression of $Pax2^{+}$ dI4/dI6 and $Lmx1b^{+}$ dI5 fates when IWR1e is added to the dP cultures, along with RA+BMP4 (Figure 5M, 5Q). The addition of IWR1e did not erode $Pax3^{+}$ pan-dorsal identity or *Pax7* expression, but did suppress Olig3 (Figure S6F), which may contribute to the loss of dI1/dI2 and perdurance of dI4-dI6 fates.

Together, these results demonstrate that the BMP4-dependent dI fates require canonical Wnt/β-catenin signaling. In its absence, BMP4-mediated suppression is lost, and dPs retain their intermediate spinal cord identity.

Wnt signaling promotes BMP4-dependent fates by regulating proliferation, not patterning

Previous studies have suggested that Wnts act ether as mitogens [95, 96] or patterning factors [97, 98] in the specification of the dorsal spinal cord. We assessed these models, first by determining the effect of canonical Wnts (Wnt1, Wnt2,Wnt3a and Wnt4), a non-canonical Wnt (Wnt9b), and a small molecule Wnt agonist (CHIR99021) on dI identity in the RA±BMP4 protocols. When Wnts were added togther with RA, during the period when NPCs are competent to respond to patterning factors, they were unable to induce the dorsal-most dIs (Figure S7A). Similarly, when the Wnts were added concomitantly with BMP4, there was no synergistic effect on the specification of dorsal spinal identity (Figure 6A, 6B). Thus, Wnts are not sufficient to direct dI fate identity.

We next asked whether activating Wnt signaling affected the mitotic index of mESC-derived NPCs after the patterning period. To better visualize proliferation, mESCs were allowed to form EBs [42, 44, 47], and then treated with RA±BMP4 over the same timeline as protocol 1 and 3 (Figure 1B) to induce dorsal spinal cord patterning. EBs were then cultured with either DMSO or CHIR from day 5 to day 9 (Figure 6C), and finally pulsed with 5-Ethynyl-2′-deoxyuridine (EdU) to label S-phase cells. Strikingly, there was a ~3-fold increase in EdU^+ cells in CHIR-treated EBs (Figure 6E, 6G, 6H) compared to DMSO control (Figure 6D, 6F, 6H) in both protocols. Similarly, the CHIR-treated EBs had a ~2-4-fold increase in the number of pHistoneH3 $(pH3)^+$ cells in M-phase, compared to control (Figure 6D-6G,6I). Many of the dividing cells in the CHIR-treated cultures were $Sox2^+$ NPCs (Figure 6E, 6G), while NPCs are largely depleted by day 9 in control cultures (Figure 6D, 6F). Thus, extending the window of Wnt signaling prolong NPC proliferation without affecting their original identity, i.e., NPCs from protocol 1 expressed *Pax3* and *Pax7*, while NPCs from protocol 3 expressed *Pax3* and *Olig3* (Figure 6J-6P).

Together, these studies support the hypothesis that Wnt/β-catenin signaling functions as a mitogen for dorsal spinal NPCs, rather than specifying distinct dI identities.

Wnt/β-catenin signaling can be modulated to extend the timeline of dP patterning.

Finally, we assessed whether the mitogenic properties of Wnt signaling can be leveraged to expand the pool of mESC-derived NPCs. RA±BMP4 patterned EBs were serially passaged at 1:3 dilution every 7 days in the presence of DMSO (control) or CHIR (Figure 7A). While the number of EBs were comparable in the DMSO and CHIR conditions on the first passage, they declined sharply by the second passage in the controls (Figures 7B, 7C, 7E, 7F, 7G, 7I, S7B, S7C). By the third passage, few EBs were observed in the DMSO condition, while the number of CHIR-treated EBs had increased by ~3- to 6-fold (Figure 7E, 7I, S7B, S7C). The ability of CHIR to expand the number of EBs was observed in both RA±BMP4 conditions, but was most robust for the RA+BMP4 patterned EBs (Figure 7E, 7I). CHIR-treated EBs remained $Sox2^+$ after

multiple passages, suggesting that they maintained their NPC identity (Figure 7D, 7H, S7D, S7E).

To examine whether expanded NPCs can differentiate, we dissociated RA±BMP4 CHIR-treated EBs at passages 6 (Figure 7J) or 7 (data not shown) onto Matrigel-coated plates, and cultured them in the presence of CHIR, DMSO (i.e. CHIR withdrawal), or DAPT, a pro-differentiation agent [99] (Figure 7J). By day 10, while the CHIR-treated cultures remained $Sox2^+$ (Figure 7K, 7N, 7Q,) the DMSO-treated cultures show robust induction of post-mitotic neural markers, *neuronal nucle*i (*NeuN/Rbfox3*), and *Tubb3* (Figure 7Q) and contained many $Tuj1^+$ neurites (Figure 7L, 7O). Thus, the removal of CHIR triggers spontaneous neural differentiation. The addition of DAPT did not significantly augment *NeuN* and *Tubb3* expression compared to DMSO (Figure 7Q) although there was increased sprouting of Tuj^+ neurites (Figure 7M, 7P).

We next asked whether specific dI fates are preserved in the expanded NPC pool. By day 10 (timeline in Figure 7J), passage 6/7 cultures derived from RA-patterned EBs contained many $Lmx1b^+$ dI5s, but only a few $Pax2^+$ dI4/dI6s, while cultures from the RA+BMP4-patterned EBs contained both $Isl1^+$ dI3s and $Lhx1b^+$ dI5s, but almost no $Lhx2^+$ dI1s, or $Foxd3^+$ dI2s (Figures S7I, S7J, 7R, 7T, 7U). Thus, while some patterning information is preserved over many passages, prolonged CHIR treatment erodes its fidelity, with RA+BMP4 dPs now expressing high levels of *Pax7* and low levels of *Olig3*, as if they have been returned to an intermediate spinal identity (Figure S7G, S7H). However, dorsal patterning information can be restored, by including a pulse of RA±BMP4 between day 3 and 5 during the differentiation of the expanded EBs on Matrigel (Figure 7J). The addition of RA resulted in a ~6-fold increase in the number of dI4/dI6s (Figure 7S, 7X), while the addition of RA+BMP4 induced ~6-fold more dI1s, and ~2-fold dI2s (Figure 7V, 7X), compared to control (Figure 7R, 7T, 7X). Unexpectedly, the pulse of RA+BMP4 suppressed the dI3 fate (Figure 7U, 7W, 7X). This strategy was successful at

restoring almost all dI identities in cultures taken from EBs both at the earliest passage and at the latest passage assessed (passage 7; Figure 7X).

Thus, these data show that the chemical activation of the canonical Wnt signaling can be successfully used to expand the size of mESC-derived NPC population, and thereby to derive large numbers of specific dI populations.

Discussion

Stem cell-based directed differentiation protocols result in both a source of *in vitro*-derived cell-types and a model system to dissect developmental mechanisms. We have built on previous studies that derive spinal motor neurons *in vitro* from an NMP intermediate [54, 55], to develop protocols that produce a complete *in vitro* atlas of dorsal spinal sensory interneurons. Our transcriptomic analyses suggest that these neurons are indistinguishable, both molecularly and functionally from their endogenous counterparts. We then used these protocols to investigate the mechanisms that regulate dI fate choices. These studies identified the hierarchy of dI fate specification and resolved the role of Wnt signaling as a mitogen in this process. We leveraged this mechanistic understanding to expand the numbers of dPs as a means of increasing dI yield.

Hierarchy of dI fate specification

Previous studies have shown that NMPs contribute to spinal cord and paraxial mesodermal lineages both *in vivo* [100, 101] and *in vitro* [54]. Our transcriptomic analyses revealed that the dI differentiation program conforms to the canonical Waddington model [102, 103] where NMPs are at the top of a “hill” of possible fates, and the time and sequence in which they receive RA and BMP4 signals control their differentiation into dIs vs mesodermal fates. Thus, the addition of RA at day 3 channels NMPs towards the intermediate dI fates. However, if BMP4 is added concomitantly with RA, the differentiation trajectory alters to specify endocardial lineages. BMP4 can only induce the dorsal-most dI fates after 24 hours of RA treatment, suggesting that NMPs must move through an RA-induced early competency state (dP state1), before BMP4 can relay dorsal spinal patterning information. These studies thereby inadvertently developed a directed differentiation protocol (protocol 2) that is a starting point to derive the cardiac cell types needed to regenerate the heart [104]. BMP4 may direct NMPs to form lateral plate mesoderm [105], which contributes multiple cell types to the embryonic heart. This finding also suggests that NMPs may have more pluripotency for mesodermal fates than previously realized. In this case, the use of

alternative growth factors in protocol 2, will direct NMPs towards additional mesodermal derivatives.

Our studies suggest that RA is required to establish dP state 1, before specifying the dI4-dI6 fates. The *Meis1/2* genes, which are dynamically expressed in the early spinal cord neuroepithelium [106], may be central to this competency state. *Meis1/2* are among the earliest genes to be induced by RA [107, 108], and become central to a gene interaction network that contains the genes that specify intermediate spinal identity. Thus, the upregulation of *Meis1/2* may drive the multipotential progenitor state (dP state1) that differentiates into the dI4, dI5 and dI6s. The addition of BMP4 both suppresses dI4-dI6 fates, and induces another multipotent state, dP state 2, that differentiates into dI1, dI2 and dI3. Interestingly, dP state1 may be a default state, which was first suggested by earlier studies in the intermediate spinal cord [109, 110]. When RA+BMP4 patterned progenitors are expanded, they appear to revert to the default dP state1 identity. This reversion can in turn be reversed: a pulse of RA+BMP4 remains sufficient to convert dP state1 back to dP state2 and elicit formation of the dorsal most dIs.

Extent to which in vitro-derived dIs phenocopy endogenous sensory interneurons

Our studies significantly extend previous studies seeking to generate dIs [42, 44, 47] by directly assessing the heterogeneity of the cultures, and documenting the extent to which stem cell-derived dIs resemble their endogenous counterparts. Using scRNA-Seq analyses, we identified that our protocols generate the full complement of dIs, that there is no heterogeneity beyond dorsal spinal derivatives in the *Tubb3*$^{+}$ cells, and that the transcriptomes of *in vitro*- and *in vivo*-derived dIs are indistinguishable. *In vitro*-derived dIs express the correct patterning factors and neurotransmitters, and display the relevant functional GO terms, suggesting they encode the correct sensory modalities. For example, balance-related GO terms were enriched in dI1, dI2, dI3 and dI6, which regulate proprioception and gait, while pain and itch GO terms were enriched for dI4 and dI5, which relay noxious stimuli *in vivo*. We also find GO terms related to

serotonergic and dopaminergic synapses, which may permit communication between the spinal cord and enteric nervous system. This signature is associated with the synapses formed between spinal neurons in the dorsal horn and the descending afferents from the brain stem [111] and diencephalon [112], which regulate both pain perception [113] and sensation in gut [114] and bladder [115]. These circuits are of considerable therapeutic importance, given that spinal injuries can lead to both irritable bowel syndrome [116] and loss of bladder control [115, 117]. Additionally, we identified gene signatures related to addictive psychoactive compounds like morphine, cocaine, and amphetamine, suggesting the dIs are the cellular targets for these drugs, elevating their importance for drug screening platforms.

Role and application of Wnt signaling as a mitogen

The BMPs have reiterative roles regulating dI fate specification and modulating the cell-cycle [44, 51]. Here, we demonstrate that the control of proliferation is mediated through the activation of Wnt signaling. Like the BMPs, multiple Wnt genes, especially those that activate the canonical Wnt/beta-catenin pathway, are expressed in the developing dorsal spinal cord [92]. The BMP and Wnt pathways are known to interact in other contexts, including during neural crest delamination [118] and the assignment of hematopoietic fates [119]. However, the role of the Wnts in the spinal cord has remained unresolved. Previous *in vivo* studies ascribed both dI fate specification [97, 120] and mitogenic roles [51, 96] to the dorsal Wnts. Our data unambiguously showed that the Wnts tested so far are not sufficient to pattern dI fates, rather they function as mitogens in this context, perhaps controlling the size of the dorsal-most dI populations. It remains unresolved as to why there are so many Wnts present in the spinal cord *in vivo.* One possibility is that they have signal-specific activities, like the BMPs [44], sculpting the number of cells in each dI population. This mechanistic insight enabled us to expand mESC-derived spinal cultures by adding CHIR to constitutively activate Wnt/β-catenin signaling after the patterning stages.

Remarkably this manipulation sustained dorsal spinal neural identity while enabling the dP to proliferate for more than 8 passages (the maximum number we attempted).

This protocol modification may represent a key step toward the long-sought goal of being able to supply specific neural populations in a limitless manner. Previous neurosphere studies have suggested that immortalized cells lose their neurogenic potential over time [121, 122]. We rather find, in this expansion protocol, that even the progenitors from the latest passages we tried (passage 5-7), retain the ability to differentiate as neurons. However, while neurogenesis remains stable, the patterning information does erode over time. Cell cycle duration has been shown to regulate cell fate by acting as a filter for long transcripts [123, 124]. Such mechanisms may erode cellular identity (i.e. dP state 2) when NPCs are held in an extended proliferative phase. Nonetheless, dI fates can be again restored with a 2-day pulse of RA±BMP4.

Implications of stem cell-derived sensory neurons for drug screening and regenerative therapies

Our studies have identified distinct psychoactive and analgesic drug-related signaling pathways in different dI populations. These findings raises the possibility of using stem cell-derived dIs in drug screening studies to identify additional analgesics, and to study the mechanistic action of drugs such as cocaine and amphetamine at the cellular level. Targeting a spinal cellular substrate for therapeutics, that concomitantly spare cortical activation, may result in pain relief that avoids the addictive side effects that have led to opioid abuse.

It is also now critical to assess the regenerative potential of these cells, by transplantation into an injured or diseased spinal cord. At present, it remains unresolved whether the pure neuronal populations [125], will be more functionally beneficial than replacing populations of multipotential progenitors or mixtures of neurons [126]. The identification of dI-specific cell surface markers in this study (Table S4) enables the sorting of specific neural populations, which will make it possible to distinguish between these possibilities.

In summary, these studies uncover further mechanistic insights into the process of dI differentiation *in vitro* to derive *bona fide* spinal sensory neurons in the large numbers needed for clinical and drug discovery applications. They represent a critical step towards using ESC-derived dIs for studying the effects of various addictive compounds and in *vitro* disease modeling for pain disorders.

Materials and Methods

KEY RESOURCE TABLE:

Reagents	**Source**	**Identifier/Catalogue number**
Antibodies		
Goat polyclonal anti-Lhx2	Santa Cruz	Sc-19344
Goat polyclonal anti-Isl1	R&D Systems	AF1837
Rabbit polyclonal anti-Pax2	Invitrogen	71-6000
Mouse monoclonal anti-Lhx1/5	Developmental Studies Hybridoma bank	4F2
Goat polyclonal anti-Pax3	R&D Systems	AF2457
Mouse monoclonal anti-Pax7	Developmental Studies Hybridoma bank	AB_528428
Rabbit monoclonal anti-phosphoSMAD1/5	Cell Signaling Technology	D5810
Rabbit polyclonal anti-phospho Histone H3	Cell signaling Technology	9701
Mouse monoclonal Tubulin III beta (clone-)	Covance	MMS-435P
Goat polyclonal anti-Sox2	Santa Cruz	sc-17320

Rabbit polyclonal anti-Pax6	MBL life science	PD-022
Goat polyclonal anti-T (Brachyury)	R&D Systems	AF-2085
Guinea pig polyclonal anti-FoxD3 sera	gift from Thomas Müller; [127]	
Guinea pig polyclonal anti-Lmx1b sera	gift from Thomas Müller; [127]	
Guinea pig polyclonal anti-Olig3 sera	gift from Thomas Müller; [127]	
Chemicals, reagents and peptides		
CHIR99021	Tocris Bioscience	4423
IWR1e	Millipore	681669
XAV939	Stem Cell Technologies	72674
Matrigel membrane matrix	Fisher Scientific, Corning	CB-40234
DAPT	Sigma Aldrich	565784
Human recombinant BMP4	Thermo Fisher Scientific	PHC9534
Retinoic acid	Sigma Aldrich	R2625
Human recombinant Wnt1	Thermo Fisher Scientific	PHC1804
Human recombinant Wnt2	Millipore	SRP-6560

Human recombinant Wnt3a	R&D Systems	5036-WN-010
Mouse recombinant Wnt4	R&D Systems	475WN005
Human recombinant Wnt5b	R&D Systems	7347-WN
Mouse recombinant Wnt9b	R&D Systems	3669-WN-025
Human basic FGF (bFGF)	Thermo Fisher Scientific	PHG0023
DMEM:F12	Hyclone	SH30023
Neurobasal medium	Thermo Fisher Scientific	21103049
B27 supplement	Fisher Scientific	17-504-044
N2 supplement	Thermo Fisher Scientific	17502048
Mouse LIF	Millipore	ESG1107
Trypsin Inhibitor 1x solution	Sigma-Aldrich	T7659
Paraformaldehyde 16%	Sigma-Aldrich	28906
Phosphate buffer saline (PBS) 1X	Hyclone	SH30028.02
Prolong Gold	Thermo Fisher Scientific	P36930
Critical Commercial Assays		
RNeasy plus mini kit	Qiagen	74106

Superscript IV First-Strand synthesis kit	Thermo Fisher Scientific	18091050
Dead cell removal kit	Miltenyi Biotec Inc.	130-090-101
EdU labelling assay	Invitrogen	C10338
Qia shredder	Qiagen	79654
Software		
Corel DRAW version 20	Corel DRAW	www.coreldraw.com
Zeiss Blue lite	Zeiss	https://www.zeiss.com/microscopy/us/products/microscope-software/zen-lite.html
Prism 9	Graphpad	https://www.graphpad.com/scientific-software/prism/
Image J	National Institute of Health	https://imagej.nih.gov/ij/
BioRender		https://biorender.com

R-Package V.3.2.5	The R project for Statistical computing	https://www.r-project.org/
Cell Ranger v 3.0.2	10x Genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/
Seurat v. 3.1.1	Satija Lab	https://satijalab.org/seurat/
EnrichR	Ma'ayan Lab	https://maayanlab.cloud/Enrichr/
Experimental Model: Cell lines		
MM13 mouse embryonic stem cell line	Tom Jessell Lab, Columbia University [46]	
Deposited data		
Bulk and single-cell RNA-Seq	Gene Expression Omnibus (GEO)	GSE185891
Oligonucleotides	**Mouse primers for qPCR analysis**	
	Forward primer	Reverse primer

GAPDH	GGCCTTCCGTGTTCCTAC	TGTCATCATACTTGGCAGGTT
Lhx2	cagcttgcgcaaaagacc	taaaaggttgcgcctgaact
Foxd3	ccccaacactgaccaacag	gtttgctccgccagctta
Isl1	ATCGAGTGTTTCCGCTGTGT	ATGGGTTCGGCTGCCATTT
Lhx5	CGGAGCAAAGTCTTCCACCT	GGACGACACCGAGTTGAGAC
Lbx1	GCCGAAGGCCGCGATG	TGCGTGATTTTCGCCGTTTC
Lmx1b	GCATCAAGATGGAGGAGCAC	CTCGTTGACTCGCATCAGG
Meis1-spliced variant A	ACCCTGATGGACAGCCAATG	GCATACTTTGCAGCCCTGG
Meis1-spliced variant B	CAATCGAAAGCCAGGGGGAA	CTCGGCTGTCCCATACTCAC
Meis2	GCACAGGATGCGTTTCAGTG	TGTCTAACCCATCGCCTCCA
Meis3	TGGGTTCAAAGGGAAGGAAGG	GAGGTGAAGTCCTGAAGCCC

Irx1	CGGCCGCGGTCACCT	CAGTTCATACTGA GAGCCCATCT
Irx3	ACAGGTGACAAGAAAAATATCGC	CAGCGTCCAGAT GGTTCTGT
NeuN	ATCGTAGAGGGACGGAAAATTGA	GTTCCCAGGCTT CTTATTGGTC
Tubb3	ATGGACAGTGTTCGGTCTGG	AGCACCACTCTG ACCAAAGAT
Sox2	CAAAAACCGTGATGCCGACT	CGCCCTCAGGTT TTCTCTGT
Pax3	AAACCCAAGCAGGTGACAAC	GCGTCCTTGAGC AATTTGTC
Pax7	TGGCCAAACTGCTGTTGATT	TAGGCTTGTCCC GTTTCCAC
Olig3	ACGTTAGGGCTGTTCCAAGG	GGCAAGGGGGTA GGATGAAC
Wnt1	ACTACGTTGCTACTGGCACT	GTTCACGATGCC CCACCATC
Wnt2	TGATGTAGACGCAAGGGGGT	CATGTACCACCAT GAAGAGCTGA

Wnt3a	CTACCCGATCTGGTGGTCCT	ACAGAGAATGGG CTGAGTGC
Wnt4	CAGGAAGGCCATCTTGACACAC	GTCTTTACCTCGC AGGAGCC
Wnt5b	CCGGCCACCTTTCTCTCTTA	CGGAGCTCCCAC GCTGAT
Wnt10b	ACGGTTTAAGCCCCAAGAGC	GAGTCAGTCAGC GCCTCC
Axin2	GCTGCGCTTTGATAAGGTCC	CGCGAACGGCTG CTTATTTT
Tcf7	GGACGGGAGCACACTTCG	GCCTTCAGGCCG TCCTCT

EXPERIMENTAL MODEL AND SUBJECT DETAILS

ESC maintenance and two-dimensional (2D) differentiation:

The mouse ESC line, MM13, was routinely maintained in ES cell media (DMEM+20% FBS) with 100u/ml of mouse leukocyte inducing factor (LIF). Cells were maintained on mitotically inactive irradiated mouse embryonic fibroblasts (MEFs) and were passaged at least twice before starting the differentiation. To prepare cells for the differentiation, ESC colonies were dissociated with 0.25% trypsin and plated on gelatin coated plates at 1:10 dilution to reduce MEF transfer. Cells cultured on gelatin coated plates and allowed to proliferate for 1-2 days. To initiate differentiation, cells were plated on 0.1% gelatin coated 24-well CellBIND dishes (Corning) at 20,000 cells/ml density in 0.5ml/well N2/B27 medium that contains 10ng/ml basic fibroblast

growth factor (bFGF). The N2/B27 medium contains 1:1 portions of Advanced Dulbecco's modified Eagle Media (DMEM) F12 (Hyclone) and Neurobasal media (Thermo Fisher Scientific) supplemented with 0.5x N2, 1x B27 supplement (with Vitamin A), 2mM L-glutamine, 0.1mM 2-mercaptoethanol (ME), and 1% BSA. On day 1, small colonies of cells can be observed attached to the bottom of the wells. On day 2, cells were supplemented with 10ng/ml bFGF and 5μM CHIR99021 for 24 hours to induce a neuromesodermal identity [54]. On day 3, media was changed as follows: for protocol 1, cells were exposed to 100nM RA for 48hrs. For protocol 2, 10ng/ml human recombinant BMP4 (Thermo Fisher Scientific) was added concomitantly with 100nM RA (RA+BMP4) for 48hrs. For protocol 3, cells were exposed to 100nM RA for 24hrs, followed by a pulse of RA+BMP4 for another 24 hours. To induce terminal differentiation in all three conditions, the medium containing growth factors was replaced with basic N2/B27 medium at day 5, and cell allowed to differentiate to day 9. At the end of the differentiation, the cultures were fixed for immunohistochemical analysis or lysed to obtain RNA for RNA-Seq or quantitative reverse transcriptase PCR analysis.

Bulk-RNA sequencing and data processing: Total cell lysate was obtained in buffer RLT (Qiagen) at different time points from three independent differentiations (biological replicates) conducted parallelly. RNA extraction was performed using RNeasy mini kit (Qiagen) and the quality was determined using Agilent Technologies 2100 Bioanalyzer and only samples with a RIN score >8.0 were sequenced. Stranded libraries were constructed using Universal plus mRNA sequencing kit (NuGEN) and sequenced onto 2 lanes of NovaseqS1 to generate 30 million reads/sample. Obtained reads were aligned to the latest mouse reference genome (mm10) using the STAR spliced read aligner. Pearson correlation and principal component analysis (PCA) was performed to determine the similarities among the samples and replicates. Differentially expressed genes were identified using DESeq and EdgeR package on R platform. The candidate genes were selected using cutoff false discovery rate (FDR <0.05) and log fold

change (>2). Differentially expressed gene lists were further subjected to Enrichr, Metascape [88] and DAVID functional annotation tools to identify enriched gene ontology (GO) categories.

WGCNA analysis: We used WGCNA package [60] to identify the expression dynamics of gene modules associated with differentiations for protocol 1, 2, and 3. WGCNA was performed separately on RNA-Seq samples from protocol 1 (R-branch), 2 (C-branch), and 3 (B-branch) with day 0 and day 3 as common samples. First, differentially expressed genes were extracted for each group to determine the most variable genes. The Pearson correlation matrices were calculated for all RNA pairs in each group and were transformed into adjacency matrices using the power function. A dynamic tree-cut algorithm was used to identify gene co-expression modules where modules were defines as branches cut off from the tree and labeled in unique colors. The module eigengene (ME) represents the first principal component for each module. Gene ontologies associated with the modules were determined by supplying top 100 genes to the Enrichr platform.

Protein-protein interaction (PPI) network analysis: We used the metascape algorithm (www.metascape.org) to identify PPI networks induced by 6hrs and 24hrs of RA exposure. First, significant differentially expressed (DE) genes were extracted by pair-wise comparison with the day3 dataset (day3 v/s 6hrs RA and day 3 v/s 24hr RA) using log fold change> 2 and false discovery rate (FDR)<0.01 as cut off for the upregulated genes. The entire gene list obtained from the DE analysis was then submitted to Metascape platform with Mus musculus as both input and analysis species. Both gene ontologies (GO) and PPI network were then retrieved for each gene list using the default Metascape parameters applied under the express analysis tool (minimum network size=3, maximum network size 500).

Single-cell RNA sequencing and analysis:

Preparation of single cell suspension: Day 9 cultures from both protocol1 (RA) and protocol3 (RA+BMP4) were dissociated with 0.25% cold trypsin for 5 mins. Trypsin activity was stopped

with the addition of trypsin inhibitor (Sigma-Aldrich) in 1:1 ratio. Cells were then pelleted by centrifugation at 1000RPM and resuspended in 1x PBS. The ratio of viable cells was determined using Trypan blue staining. Dead cells were removed using MACS dead cell removal kit (Miltenyi Biotec Inc.) by following the manufacturer's protocol. Eluted live cells were then suspended in 1X PBS containing 0.04% BSA solution (400µg/ml) for the library preparation.

Library preparation, and sequencing: ~10,000 live cells/conditions were used to construct single-cell specific cDNA libraries using protocol described in 10x Genomics chromium single cell 3' reagent kit (v3.1 Chemistry). Briefly, cells were partitioned into nanoliter-scale Gel-Beads-in-emulsion (GEM) using the 10x Chromium controller. Each GEM contains a unique barcode which is shared among the cDNA generated from a single cell. Cells were then lysed, and cDNA synthesis and feature barcoding were performed in the GEMs. The sequencing libraries were recovered using Magnetic separation and the quantity and quality of cDNA were assessed by Agilent 2100 expert High Sensitivity DNA Assay. cDNA samples were sequenced on 1 lane of NovaSeq 6000 S2 flow cell and reads were mapped to mouse mm10 genome using Cell Ranger v.3.0.2 to generate fasq files.

Filtering genes and cells: Seurat package v3.1.1 was used for the analysis. For each condition, genes expressing in less than 3 cells were removed and cells with > 250 features were retained. PCA was used for dimension reduction (npcs=50) with top 3000 most variable genes. Raw count data were normalized using regularized negative binomial regression with SCT transformation. Cell clustering is done using Shared Nearest Neighbor (SNN) Graph method and cluster specific markers were identified by Wilcox Rank Sum test. Each cell cluster was annotated by a combination of the following methods 1) the expression of known canonical cell-type markers 2) Gene Set Enrichment Analysis (GSEA) of cluster specific markers determined by FindMarkers function in Seurat.

Reclustering of neuronal clusters: From the entire single-cell dataset, neuronal cells were first identified by the expression of pan-neuronal markers-Tubb3 and Elavl1/2. Roughly 30% cells express these markers in both protocol 1 and protocol 3 and were thus annotated as the mature neurons. The neuronal populations were then extracted using the Seurat package in R, by clustering the cells expressing Tubb3 >2 (median normalized log expression levels) using the Shared Nearest Neighbor (SNN) Graph method. This manipulation resulted in 12 clusters in protocol 1 (RA) and 13 clusters in protocol3 (RA+BMP4). Clusters were visualized with Uniform Manifold Approximation Projection (UMAP) method and the unique cluster-specific markers were identified using FindMarker function in Seurat.

GO analysis for neuronal clusters: We used DAVID functional annotation tool [79] (https://david.ncifcrf.gov/tools.jsp) for identifying enriched biological processes and pathways associated with different neuronal clusters. For this, cluster-specific gene lists were sorted according to log fold change (FC) values and all genes with log FC = >0.2 were selected and submitted to the DAVID platform with default settings and Mus musculus as the input species. Ontological terms were then surveyed to determine if terms reflect the known functions of the dIs that the given cluster belongs to. Selected GO terms were arranged in the decreasing order of -log10 (p-value) and represented as bar plots.

Integration of single-cell datasets from different origins: The *in vivo* spinal cord single-cell dataset [57] was downloaded from https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-7320/ and loaded into R. Gene names were standardized to match our dataset using the biomaRt function [128]. This data was then loaded into Seurat (version 4.0) and processed using the SCTransform pipeline [129]. *In vivo* cell type labels were simplified and transferred to the *in vitro* datasets. The *in vivo* dataset was then integrated with the *in vitro* protocol datasets using Seurat's reciprocal PCA integration pipeline, and projected into a three-dimensional UMAP space. Tabula Muris reference datasets of kidney (2781 cells), lung (5449 cells) and trachea

(11269 cells) were downloaded from https://figshare.com/articles/dataset/Robject_files_for_tissues_processed_by_Seurat/5821263, loaded into R, and then processed through Seurat, as above. The three Tabula Muris datasets were then integrated with the *in vitro* and *in vivo* spinal cord datasets using Seurat's reciprocal PCA integration pipeline and embedded into a three-dimensional UMAP space. Code for this analysis is available upon request.

dI cell type extraction based on the marker gene expression: Cutoffs for gene expression levels when extracting cells were initially determined through the cluster-specific violin plots from the integrated in vivo and *in vitro* spinal cord datasets for the SCT assay. Cutoffs were then refined by plotting the extracted cells on a UMAP plot adjusting the values to exclude non-specific cells with background levels of the genes. The cutoffs values for each gene are as follows: dI1 = *Barhl1* > 1 (*in vivo*-417 cells, *in vitro*-1120 cells), dI2 = *Foxd3* > 1 (*in vivo*-552 cells, *in vitro*-395 cells), dI3 = *Isl1* > 1 *Tlx3* >0.5 *(in vivo*-1669 cells, *in vitro*-89 cells), dI4 = *Pax*2 > 1 and *Lhx1* > 1 (*in vivo*-1943 cells, *in vitro*-810 cells), dI5 = *Lmx1b* > 1 and *Pou4f1* > 1 (i*n vivo*-433 cells, *in vitro*- 216 cells), dI6 = *Dmrt3* > 1 (*in vivo*-73 cells, in vitro-60 cells). To extract *Tubb3*+ cells in the *in vivo* and *in vitro* combined dataset, expression by cluster was plotted, and the clusters expressing high levels of *Tubb3* were extracted as a whole. Code is available upon request.

Embryoid body (EB) culture, and expansion: To initiate EB formation, dissociated embryonic stem cells were seeded in low attachment plates (Corning) at 10^5 cells/ml density in N2/B27 medium supplemented with 10ng/ml bFGF. To induce posterior spinal cord and dorsal spinal patterning, RA and RA+BMP4 were added using the same timeline as described for the 2D differentiation, and the media was changed at every other day after day 5.

For expansion cultures, EBs were dissociated in single cell suspension at day 6 by treating with 0.25% Trypsin, followed by trituration to break up the larger cell clumps. Cells were pelleted and

passaged in N2/B27 medium containing either DMSO or 5μM CHIR at 1:3 dilution. Media was changed at every third day during the expansion culture.

Embryoid body (EB) differentiation: To induce differentiation, EBs were dissociated into smaller clumps by a 5–7-minute treatment with cold 0.25% Trypsin (Gibco), followed by trituration. Trypsin was neutralized with 1:1 addition of Trypsin inhibitor (Sigma-Aldrich) and cells were pelleted by centrifugation at 1000RPM. Cell pellets were resuspended in N2/B27 medium (without growth factors) and plated onto a Matrigel precoated plate. In our experience, the single cell suspension of 100-150 EBs provided enough cells to seed one 24-well plate. Spontaneous neural differentiation is induced when dissociated EBs are cultured in N2/B27 medium for 10 days. At the end of 10 days, multiple neural processes can be visualized by phase contrast microscopy and cultures were either fixed with 4% PFA for immunohistochemistry or lysed to extract RNA in RLT buffer (Qiagen).

Immunohistochemistry: For 2d differentiations, adherent cultures were first washed with 1xPBS and fixed with fresh cold 4% PFA for 10 minutes in the well. Following fixation, cultures were washed twice with 1xPBS to remove any remaining PFA. Cells were blocked with 1xPBS with 1% heat inactivated horse serum for 1 hour and primary antibodies are added in the blocking solution for an overnight incubation at 4C. Following washes, species appropriate secondary antibodies (Jackson Immuno Research Laboratories) were added in 1x PBT (PBS + 0.1% Triton 20) for 1hr. The cultures were then washed with 1x PBT to remove traces of secondary antibodies and counterstained with DAPI to stain nuclei. Plates were then imaged on Zeiss LSM800 inverted confocal microscope at 10x magnification.

For EBs, EBs were collected and fixed with fresh 4% PFA for 20 minutes at 4°C. EBs were then equilibrated with 30% sucrose for overnight at 4C and embedded into the OCT compound for cryo-sectioning. 12um sections were collected onto positive charged slides (VWR), incubated

with the antibody blocking solution for 1hr, and then stained as above. Sections were counterstained with DAPI, mounted in Prolong Gold and imaged using a Zeiss LSM800 inverted confocal microscope.

Quantitative (q) RT-PCR analysis: RNA was isolated with RNeasy Mini Kit (Qiagen, cat no. 74104), and converted to cDNA using Superscript IV First Strand Synthesis kit (Thermo Fisher Scientific). qRT–PCR was performed using SYBR Green PCR Master Mix (Applied Biosystems) on Roche 480 Lightcycler (Roche). The relative fold expression was calculated using 2-ΔΔCt method of comparing the expression of the target gene with that of the housekeeping gene glyceraldehyde 3-phosphate dehydrogenase (Gapdh). All experiments were repeated at least 2-3 times (technical replicates) from at least two independent differentiations (biological replicates), and data is represented as mean ± SEM. The primer sequences for all target genes are listed in the Table S5.

QUANTIFICATION AND STATISTICAL ANALYSIS

Image quantification: To count the number of nuclei, images were converted to 8bit image in ImageJ, their threshold intensity was adjusted to capture only the florescent nuclei, and the number of nuclei counted using the analyze particle tool in ImageJ. For each image, the area under the DAPI staining was determined using ImageJ and target cells were represented as % of cells in 100μm^2 DAPI$^+$ area.

Statistics: Data are represented as mean± SEM (standard error of the mean). Tests for statistical significance were performed using Prism software (version 9). Values of $p < 0.05$ were considered significant in all cases.

Figures

Figure 2-1: Distinct temporal combinations of RA and BMP4 direct different mESC identities.

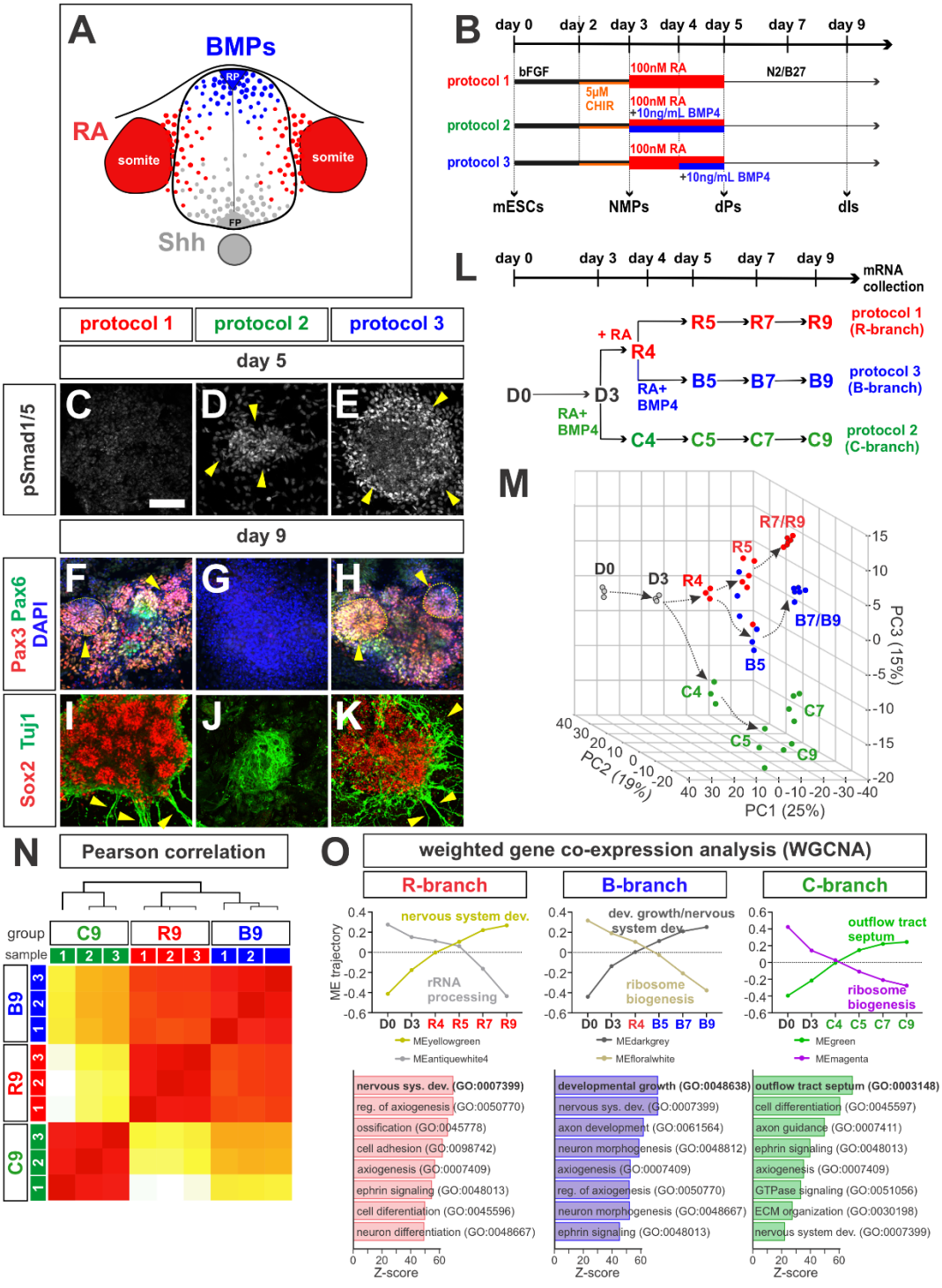

Figure 2-1: Distinct temporal combinations of RA and BMP4 direct different mESC identities.

(A) Schematic of the source of patterning signals in the neural tube.

(B) Schematic of three RA±BMP4 protocols used to direct mESCs towards dorsal spinal cord identities through an NMP intermediate..

(C-E) Protocol 2 and 3, but not in protocol 1, induce phosphorylated Smad1/5/8 (pSmad1/5/8) by day 5 showing active BMP signaling .

(F-K) Rosettes expressing Sox2 (NPCs, red, I, K), Pax3 (dPs, red, F, H), Pax6 (dPs, green, F, H), and Tuj1 (neurites, green, I, K) are observed in protocol 1 and 3, but not in protocol 2 (G, J) by day 9.

(L) Timeline for bulk-RNA seq sample acquisition (N=3).

(M) Principal component (PC) analysis identified similar trajectories for protocol 1 (red, R-branch) and 3 (blue, B-branch) which are distinct from the trajectory of protocol 2 (green, C-branch).

(N) Pearson correlation analysis showing the cells from protocol 1 (R9) and 3 (B9) are transcriptional similar, and distinct from those in protocol 2 (C9) at day 9.

(O) Weighted gene co-expression network analysis (WGCNA) identified the successive downregulation of a rRNA biogenesis module implicated in the loss of pluripotency [130, 131]. Conversely, neural modules are sequentially upregulated in the R and B branches, while a cardiac module is sequentially upregulated in the C-branch. See Figure S1E, for other modules.

Scale bar: C-K, 100μm.

Figure 2-2: Single cell sequencing identifies relevant spinal function-specific modules in *in vitro*-derived-dIs.

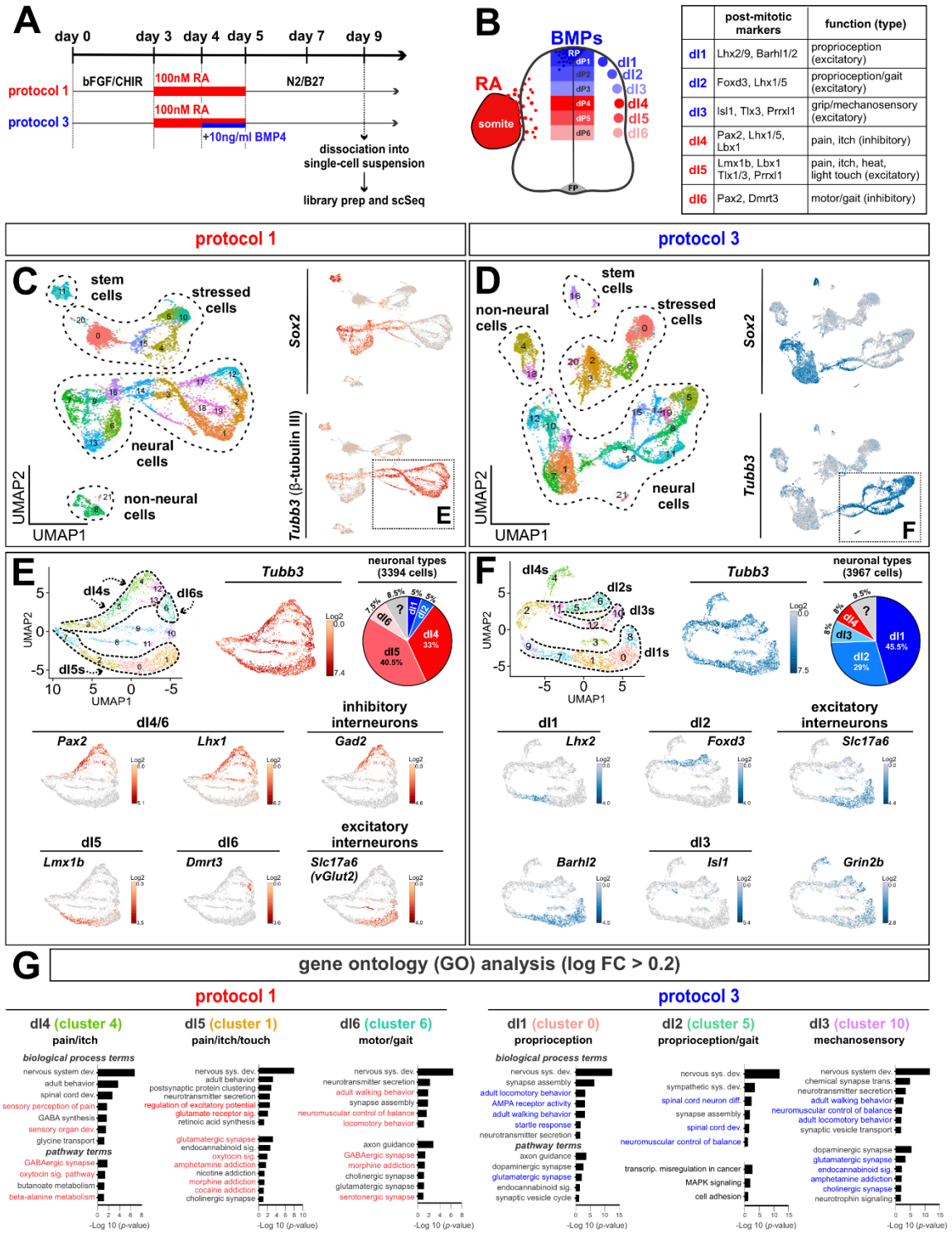

Figure 2-2: Single cell sequencing identifies relevant spinal function-specific modules in *in vitro*-derived-dIs.

(A) Timeline for single cell (sc) sequencing of protocol 1 and 3 (N=1).

(B) Schematic showing the position dPs and dIs in the spinal cord. The dI1-dI6 mediate distinct functionalities and can be distinguished by distinct transcription factors.

(C, D) UMAP projection of day 9 cells derived from RA (C) and RA+BMP4 conditions (D). Feature plots show the *Sox2*$^{+}$ NPCs and *Tubb3 (Tuj1)*$^{+}$ differentiated neurons.

(E) Protocol 1: sub-clustering of the *Tubb3*$^{+}$ neuronal clusters (C) identified 14 clusters (3394 cells). Feature plots show *Pax2*$^{+}$ *Lhx1*$^{+}$ dI4/dI6, *Lmx1b*$^{+}$ dI5 and *Dmrt3*$^{+}$ dI6 populations. The pie chart depicts the percentage of dI1-dI6s obtained, only ~8.5% of neuronal cells of unknown identity (see Table S1). Concurring with the *in vivo* functionality (B), the dI4/dI6 clusters express *Gad2* (inhibitory neurons) while the dI5 clusters express *Slc17a6* (vGlut2, excitatory neurons).

(F) Protocol 3: Sub-clustering of the *Tuj1*$^{+}$ neuronal clusters shown in D, identified 13 clusters (3967 cells). Feature plots show *Lhx2*$^{+}$ *Barhl1*$^{+}$ dI1, *Foxd3*$^{+}$ dI2 and *Isl1*$^{+}$ dI3 populations. The pie chart represent percentage of dI1-dI6s; ~9.5% cells are unknown neural identity (see Table S1, cluster 3). As in *in vivo* (B), the dI1-dI3 clusters express excitatory markers, *Slc17a2* and *Grin2b.*

(G) Gene ontology (GO) analysis of the genes upregulated (Log FC > 0.2) in the indicated clusters reveal that the relevant sensory/motor signatures are enriched in the *in vitro* derived dIs.

Figure 2-3: Stem cell-derived sensory interneurons resemble their endogenous counterparts.

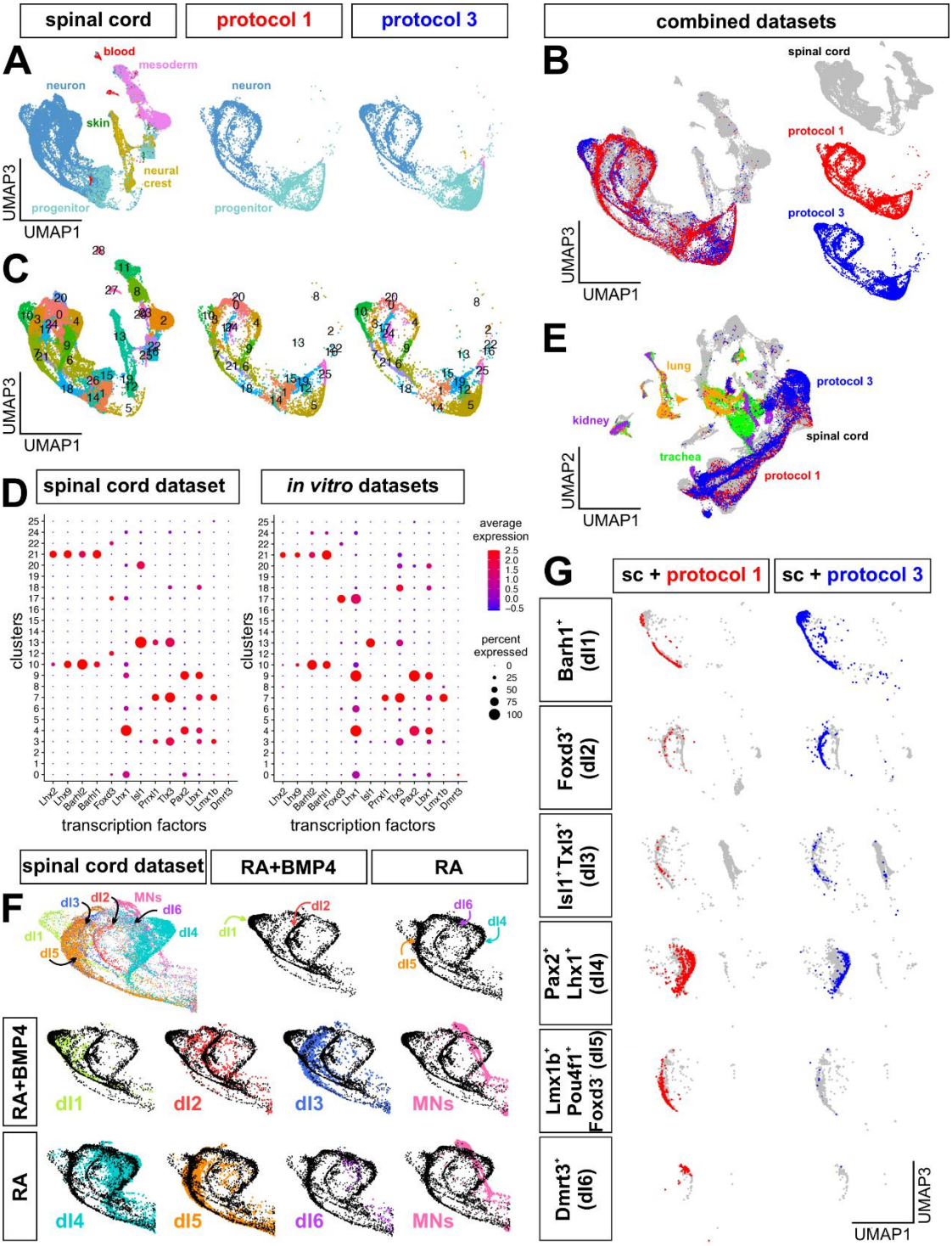

Figure 2-3: Stem cell-derived sensory interneurons resemble their endogenous counterparts.

(A) Integrated UMAP plot of the sc-Seq datasets from the *in vitro* differentiations and the *in vivo* embryonic spinal cord (E9.5-E13.5) [57]. Cell type labels from the *in vivo* dataset were projected onto the *in vitro* dataset.

(B) Integrated UMAP plot highlighting the overlap between the *in vitro* and *in vivo* cells. An atlas of these combined datasets with dI annotation can be found here (file opens after being downloaded).

(C) Clustering of the integrated dataset yields 24 shared clusters, with 6 clusters unique to the *in vivo* dataset.

(D) Dot plot of the 24 shared clusters showing the expression levels of dI marker genes.

(E) Integration of the *in vivo* spinal cord and *in vitro* datasets with lung, kidney, and trachea datasets (*Tabula Muris* Consortium) show the *in vitro* cells only overlap with spinal cord cells.

(F) UMAP plots of the neuronal portion of the integrated datasets. The *in vivo* dIs are colored using the *in vivo* dataset annotations. Each dI population was then overlaid onto the *in vitro* datasets (black) from either protocol 1 (RA) or 3 (RA+BMP4). Substantial overlap was seen for dI1-dI6, while no overlap was observed for motor neurons (pink).

(G) dI-specific cells were extracted from both the *in vitro* protocols and *in vivo* spinal cord dataset, and then plotted in the integrated UMAP space with *in vitro* cells on top of *in vivo* cells.

Figure 2-4: Wnts are upregulated as an immediate response to BMP4 signaling in mESCs-derived NPCs

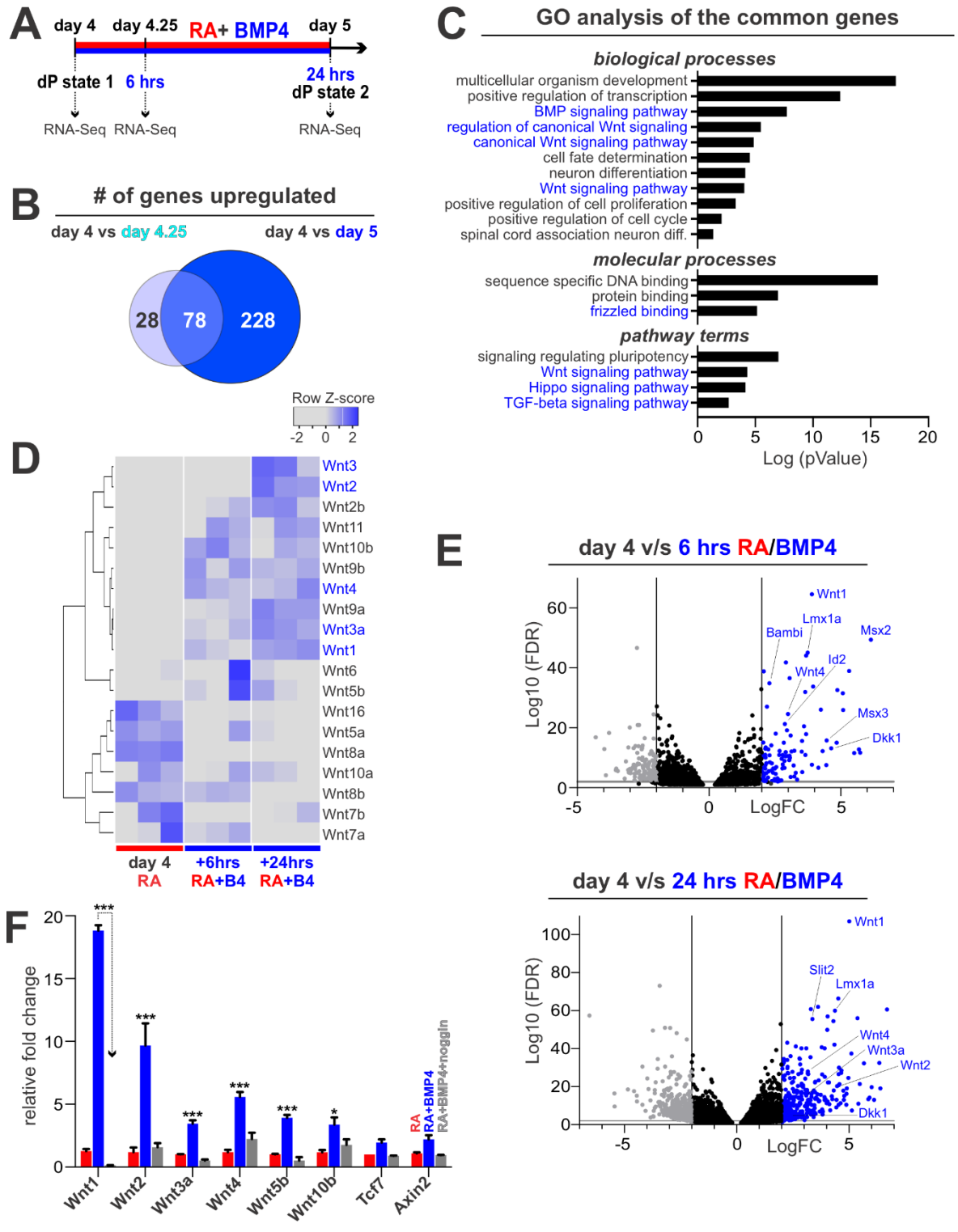

Figure 2-4: Wnts are upregulated as an immediate response to BMP4 signaling in mESCs-derived NPCs

(A) Timelines for bulk RNA-Seq analysis of protocol 3 (N=3).

(B) Venn diagram showing the overlap of genes significantly upregulated after 6 and 24hrs of RA+BMP4 treatment (Log FC > 2, FDR <0.01) . 78 genes were common to both comparisons.

(C) GO analysis of the 78 common genes shows the upregulation of Wnt signaling pathway in all three GO categories: biological processes, molecular processes, and pathway categories.

(D) Heat map showing the expression (FPKM, N=3) of 19 Wnt ligands at day 4 (dP state 1), day 4.25 (immediate response to BMP4) and day 5 (dP state 2).

(E) Volcano plots showing upregulated genes at 6hrs and 24hrs of RA+BMP4 treatment. Multiple Wnt ligands and Wnt pathway genes are highly upregulated in both conditions.

(F) RT-qPCR validation of selected Wnt genes under RA, RA+BMP4 and RA+BMP4+noggin conditions at day 5. Wnt expression was significantly upregulated by BMP4, but reduced in the presence of noggin. N= 3 differentiations. The data are presented as the mean±SEM.

Significance is determined by two-way ANOVA (Tukey's multiple comparison test), *= $p < 0.05$, **$p<0$ .005, *** $p<0.0005$.

Figure 2-5: Wnt/β-catenin signaling is required for BMP4-mediated neuronal diversity

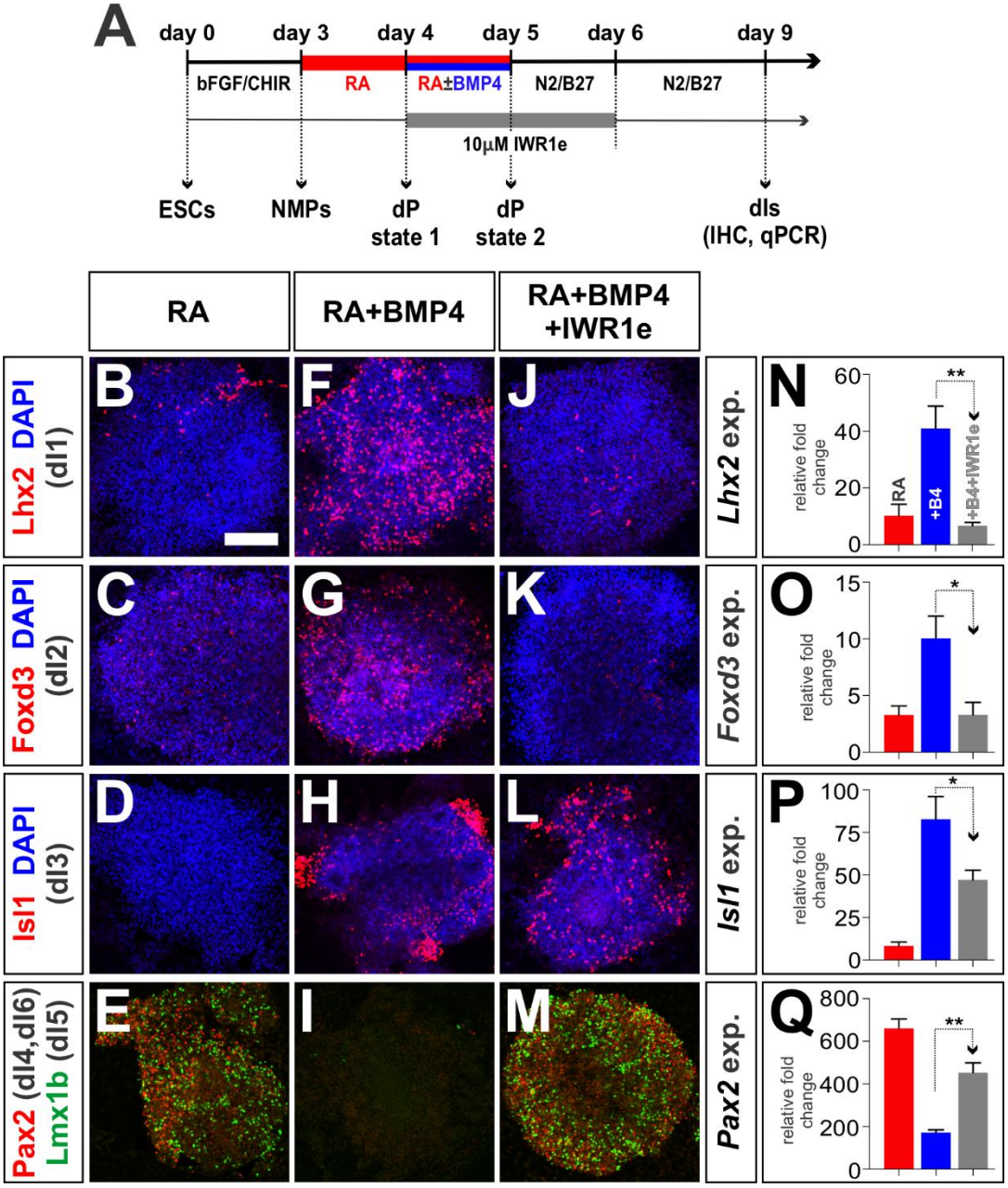

Figure 2-5: Wnt/β-catenin signaling is required for BMP4-mediated neuronal diversity

(A) Timeline to assess the requirement for Wnt signaling in the induction of dI fates .

(B-Q) Protocol 1 induces Pax2^{+} dI4/6s (E, Q) and Lmx1b^{+} dI5s (E), while protocol 3 induces the Lhx2^{+} dI1s (F, N), Foxd3^{+} dI2s (G, O) and Isl1^{+} dI3s (H, P). The addition of IWR1e to protocol 3

dramatically reduces number of dI1s and dI2s (J, K, N, O), more modestly decreases the dI3s (L, P) and concomitantly increases the dI4-dI6s (M, Q). N=3 differentiations for qRT-PCR. Significance tests: one-way ANOVA, *= $p < 0.05$, **$p<0.005$, *** $p<0.0005$. Scale bar: 100μm.

Figure 2-6: Wnt/β-catenin signaling mediates proliferation, not patterning, of mESCs derived spinal NPCs

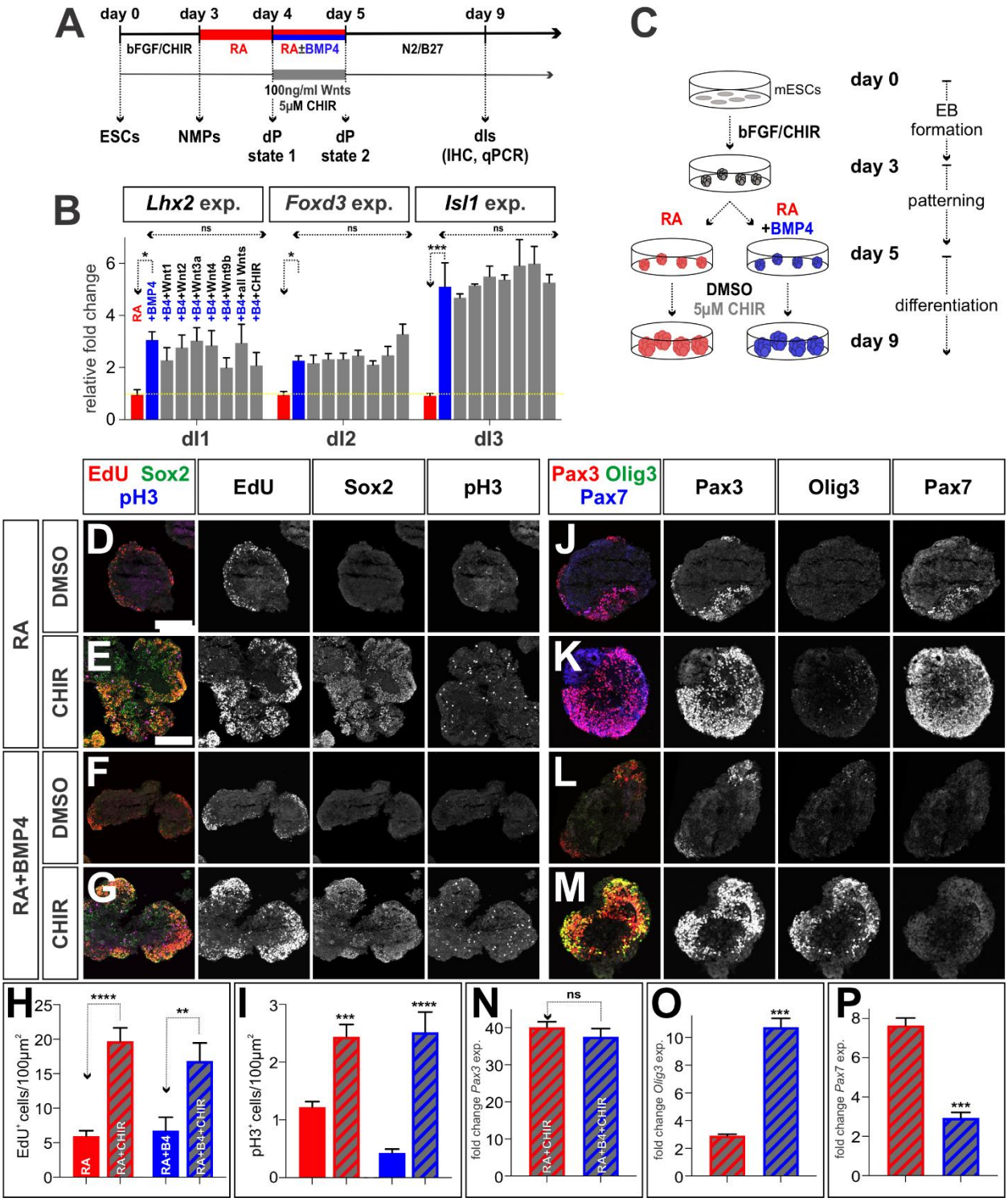

Figure 2-6: Wnt/β-catenin signaling mediates proliferation, not patterning, of mESCs derived spinal NPCs

(A) Timeline to assess whether Wnt signaling modulates the BMP-mediated dI fates though patterning activities.

(B) RT-qPCR analyses for dI1-dI3 marker genes shows that neither Wnts (100ng/ml) nor CHIR (5μM) affect (one way ANOVA) the BMP4-mediated dI fates. Expression levels were normalized to the day 0 and RA condition (N=2).

(C) Embryoid body (EB) differentiation timeline to assess the effect of activating Wnt signaling on the NPC proliferation. Samples were collected at day 9 for immunohistochemistry (IHC) and RT-qPCR.

(D-I) Both EdU incorporation (red, D-G; S-phase) and phosphorylated histone H3 staining (blue, D-G; pH3, M-phase) show that CHIR treatment increases cell proliferation by ~3 fold in RA±BMP4 patterned EBs, compared to DMSO control, resulting in increased numbers of $Sox2^{+}$ NPCs (green, D-G). Significance was determined using one-way ANOVA (Kruskal-Wallis test) (N=15-25 EBs)..

(J-P) IHC (J-M) and qRT-PCR (N-P) for Pax3 (red, J-M; all dPs), Olig3 (green, J-M; dP1-dP3) and Pax7 (blue, J; dP4-dP6) demonstrated that CHIR treatment increases the numbers of NPCs but does not significantly (unpaired t-test) alter their dorsal-ventral identities. (N=3 differentiations for qRT-PCR)

Significance values: *= $p < 0.05$, **$p<0.005$, *** $p<0.0005$. Scale bar= 200μm.

Figure 2-7: Activating canonical Wnt signaling expands mESC-derived NPC populations.

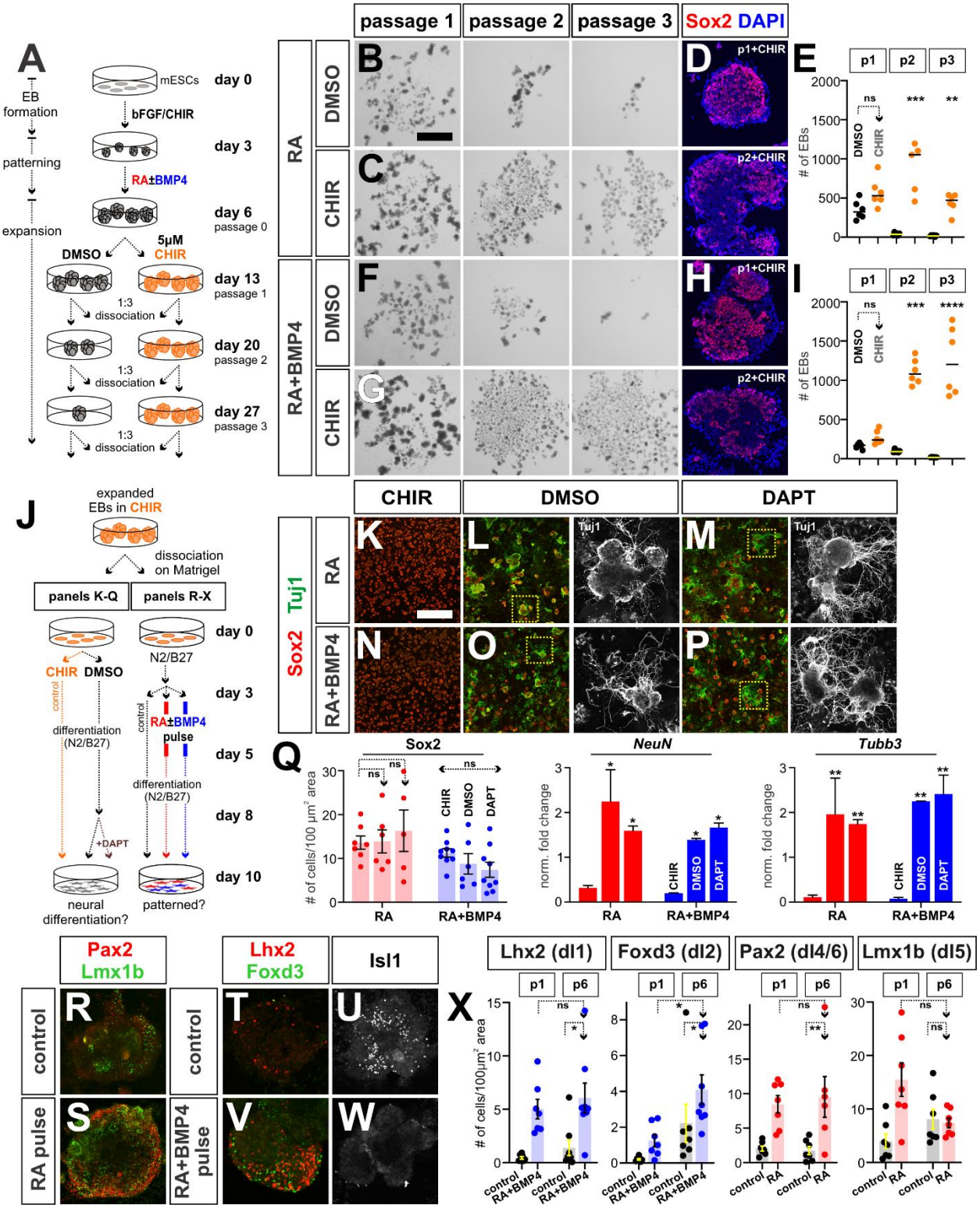

Figure 2-7: Activating canonical Wnt signaling expands mESC-derived NPC populations.

(A) Timeline for the expansion protocol for mESC-derived spinal cord progenitors.

(B-I) RA- (B-E) and. RA+BMP4-(F-I) treated EBs were passaged with DMSO (B, F) or CHIR (C, D, G, H). Bright field images (B, C, F. G) show that passaging with CHIR, but not DMSO, causes >6-fold increase in the number of EBs (E, I; one-way ANOVA Kruskal-Wallis test, N=2 independent differentiations). All CHIR-treated EBs maintain a $Sox2^+$ NPC identity after passage 1 and 2 (D, H), but the expansion of RA-patterned EBs was less robust than that of RA+BMP4-patterned EBs (E, I).

(J) Schematic of the timeline and experimental approaches to assay the differentiation potential of CHIR-treated EBs. The left timeline shows conditions to induce neural differentiation in CHIR-expanded EBs following their dissociation. The right timeline determines the extent to which dI patterning can be restored in CHIR-treated EBs, with a 2-day pulse of RA±BMP4.

(K-Q) IHC analysis for $Sox2^+$ NPCs and $Tuj1^+$ neurites on passage 6 EBs (48 days) revealed that removal of CHIR induces spontaneous neural differentiation in both RA- (L) and RA+BMP4-treated (O) EBs. The addition of DAPT had a modest effect on the number of $Tuj1^+$ neurites (M, P). RT-qPCR analyses confirm the upregulation of *Tubb3* and *NeuN* in CHIR depleted conditions (Q) However, the number of $Sox2^+$ cells remain unchanged.

(R-X) Assessing the patterning profile of expanded EBs showed they can differentiate into dI5s (R, X; RA control) or dI3s (U, RA+BMP4 control), but not dI1s, dI2s and dI4s (R, T, X). A 2-day pulse of RA±BMP4 from day 3-5 restored dI1/2 differentiation (V, X) with a ~ 5-fold increase in the number of $Pax2^+$-dI4/dI6 neurons (S, X). The differentiation potential of passage (p) 1 and p6 (X) EBs did not show significant change over time, except for dI2s, where differentiation potential modestly improves over time. (N= 6-10 EBs, Mann-Whitney test).

Significance values: *= $p < 0.05$, **$p<0.005$, *** $p<0.0005$.

Scale bar: B, C, F, G = 2mm, Q-V = 100μm; J-O = 2000 μm

Chapter 3 - Smad1 and Smad5 differentially mediate BMP signaling

Abstract: A central unresolved question in development biology is how systems of overwhelming complexity arise from relatively few families of growth factors. Compounding this issue, signaling pathways often show signal convergence, where many ligands interact with fewer receptors, which then signal through a single second messenger complex. Here we address this question in the context of bone morphogenetic protein (BMP) signaling and its role directing dorsal spinal cord development, focusing on two receptor-regulated (R) Smads, Smad1 and Smad5. Multiple models have been proposed for their mode of action from acting redundantly though combined signal strength, to having distinct activities that drive different fate outcomes. Here we address this question by generated CRISPR-edited *Smad1* and *Smad5* null mouse embryonic stem cell (ESC) lines to dissect the cell fate of activities of individual R-Smads, with a resolution not possible *in vivo.* Using a directed differentiation protocol for dorsal interneurons (dI), together with bioinformatic analyses, we have defined the roles of the R-Smads at key decision points along the dI specification timeline. Together, these findings support a model in which Smad1 and Smad5 play largely distinct roles in dorsal spinal cord development. While both R-Smads can activate canonical BMP-responsive transcriptional targets, they asymmetrically contribute to cell fate specification. Smad1 plays a restricted role, while Smad5 has a dominant role regulating dorsal progenitor transcriptional dynamics and reiteratively directing the dorsal-most dI fates.

Introduction

Bone morphogenetic proteins (BMPs) direct a remarkably wide array of cellular functions across all embryonic tissues, including proliferation, differentiation, migration, and cellular patterning [1-3]. Over 20 BMP ligands signal through ~7 type I and type II serine/threonine kinase BMP receptors, which phosphorylate, and thereby activate, three canonical second messengers, the receptor-regulated (R) Smads (Smad1/5/8) (Fig. 1B) [132]. The R-Smads complex with a co-Smad, Smad4, and translocate to the nucleus to regulate specific gene expression programs [5, 6]. The Smad

complex acts as activators or repressors depending on the cofactors present and can also reshape chromatin structure, either directly or through the engagement of histone acetyltransferase (HAT) and histone deacetylase (HDAC) complexes [133, 134].

The ability of Smads to act as transcriptional regulators is well established. However, it remains unresolved how a single second messenger complex translates a diverse ligand input through a limited complement of receptors to result in the wide range of cellular outcomes directed by the BMPs. We are assessing this question in the dorsal spinal cord, where BMPs have reiterative roles directing cell fate [22, 45] and axon guidance [135]. The dorsal spinal cord contains six classes of dorsal interneurons (dIs), dI1-dI6, which are derived from six dorsal progenitor (dP) populations, dP1-dP6. These subtypes can be distinguished by their complement of transcription factors [136]. BMP signaling is known to specify the dorsal-most spinal cord, i.e., the roof plate (RP), dI1, dI2 and dI3 [44, 137-139]. The intermediate spinal cord, i.e. dI4-dI6, is thought to be specified by retinoic acid (RA) [140]. RA is present in the surrounding paraxial mesoderm and has an earlier role directing neuralization [109, 110].

Previous studies have suggested that R-Smads function interchangeably during embryogenesis [141]. However, R-Smads are found in different compartments of the developing spinal cord: Smad5 is present in highest levels in progenitors, whereas Smad1 is enriched in post-mitotic neurons and their processes, and Smad8 is completely absent [142]. This distribution pattern is consistent with Smad1 and Smad5 having distinct, rather than redundant, roles in spinal cord development. Our loss-of-function conditional genetic approaches in mouse and overexpression studies in chicken supported this hypothesis, suggesting that Smad5 mediates the cell fate specification activity of the BMPs, while Smad1 regulates axon outgrowth [142]. However, other studies have differed in their conclusions; RNA inhibition approaches in chicken embryos suggested that Smad1 and Smad5 redundantly mediate dI fate specification [53]. This finding led to the signal strength model which proposes that the level of combined Smad1/5 activity, rather the identity of

individual Smads, determines cell fate [143]. R-Smads are phosphorylated (p) in a gradient in the spinal cord [142], consistent with the signal strength model. However, we subsequently found that different BMPs can direct similar levels of pSmad, yet result in markedly different fate outcomes [44], suggesting that R-Smad identity, rather than signal strength, is the key determinant of cell fate. However, *in vivo* approaches have inherent limitations. Acutely misexpressing genes in chicken is performed in a background of reiterative endogenous signaling, while RNA inhibition approaches can be non-specific, targeting multiple related genes. In contrast, perduring gene function in conditional knockout approaches in mice may mask phenotypes especially if protein turnover is slow. To overcome these limitations, we have developed a new mouse embryonic stem cell (ESC) model system where CRISPR/Cas9 genome editing [144] was used to generate lines lacking either Smad1 (Smad1Δ) or Smad5 (Smad5Δ). Stem cell models have been powerful systems for interrogating complex developmental patterning and transcriptional mechanisms [54, 55, 145], including ventral spinal cord development [55, 146]. They generate large populations of cells synchronously executing the same developmental decisions and permit experimental control over signaling inputs in ways that are not feasible *in vivo*.

Here we have used established directed differentiation protocols for dIs [140, 147] as a platform to dissect the individual roles of Smad1 and Smad5 in dorsal spinal cord development. These protocols generate dIs via a neuromesodermal progenitor (NMP) intermediate [54], and thereby recapitulate the endogenous program of dorsal spinal cord development [140, 148]. The RA protocol generates intermediate dIs (dI4-dI6), while the RA+BMP4 protocol directs the dorsal-most dIs (dI1-dI3) [140]. By applying bioinformatic approaches at key decision points along the dI specification timeline in the control, Smad1Δ and Smad5Δ mESC lines, we have defined the individual contributions of each R-Smad to dorsal spinal cord development. We find that while both R-Smads activate canonical BMP-responsive transcriptional targets, their contributions to cell fate decisions are asymmetric. Smad1 plays a restricted role in suppressing specific dorsal fates, while Smad5

is the dominant second messenger, directing transcriptional dynamics across the dorsal progenitor domain (dP1-dP6) and reiteratively directing the dorsal-most dI fates. Together, these findings support a model in which Smad identity, rather than signal strength alone, determines the diverse cell fate outputs of BMP signaling in the dorsal spinal cord.

Results

Generation of Smad1 and Smad5 knockout mESC lines to investigate dI specification

We have previously established two directed differentiation protocols (RA±BMP4) that derives the complete complement of dIs from mESCs (Fig.1A). These protocols generate dIs that are transcriptionally and functionally indistinguishable from their endogenous counterparts [140, 147, 149], suggesting that together they accurately recapitulate dorsal spinal cord development. Briefly, mESCs are treated with bFGF/CHIR to direct them into a biopotential neuro-mesodermal progenitor (NMP) identity by day 3. The NMPs are then neuralized with retinoic acid (RA) for two days. To generate the dorsal-most dIs, NMPs are additionally treated with BMP4 for 24 hours on day 4. By day 9, post mitotic neural markers are robustly expressed (Myt1l and Tuj1; Fig.1C, 1D, 1F, 1G), together with dP (Pax3; Fig.1C, 1F), and dI markers, which include Lhx2 (dI1), Lhx1/5 (dI2/dI4/dI6), and Lmx1b (dI5) (Fig. 1D, 1G, 1E, 1F). While both protocols generate equivalent numbers of dPs and neurons (Fig. 1I), RA+BMP4 enriches for dI1-dI3, while RA alone results in increased numbers of the intermediate spinal subtypes, dI4-dI6 (Fig. 1C -1J) [44, 140].

To begin understanding the role of second messenger signaling in dI specification, we assessed the level and activity of Smad1 and Smad5 in the directed differentiation protocols. Previous studies have shown that Smad8 is not expressed in the developing dorsal spinal cord [142]. Both Smad1 and Smad5 (Fig. 1K) are present at relatively constant levels along the RA+BMP4 timeline, suggesting BMP addition does not affect Smad protein levels. Rather, the addition of BMP4 at day 4 results in phosphorylated (p; activated) Smad1/5 within 6 hours of treatment (Fig.1K). Elevated pSmad levels persist for 24 hours through day 5. However, by day 7, pSmad

levels are comparable across the RA and RA+BMP4 protocols (Fig. 1K 1L). Together, these data define the temporal window for investigating the transcriptomic changes downstream of canonical BMP signaling that drive dI fate specification and suggest that low level BMP signaling occurs in the RA protocol.

To investigate whether Smad1 and Smad5 have distinct roles mediating BMP-dependent transcriptional programs, we generated Smad1 and Smad5 knockout (Δ) mESC lines. We employed a dual-nicking CRISPR–Cas9 strategy [150], designing paired guide RNAs to excise the start codon in exon 2 of the *Smad1* and *Smad5* loci (Fig. 2A). This approach successfully resulted excised exon 2 (Fig. 2B-2D) and thereby generated three independent protein-null mESC lines for each Smad (Fig. 2D). Loss of Smad1 or Smad5 did not alter the mESC morphology during routine passaging, or the expression of key pluripotency markers, such as Oct4, Nanog, and Sox2 (Fig. 2F).

Transcriptomic profiling of WT, Smad1Δ and Smad5Δ lines along the dI differentiation timeline

To identify the gene networks regulated by Smad1 and Smad5 during dI specification, we performed bulk RNA sequencing (Seq) in the Smad1Δ (n=3 independent lines), Smad5Δ (n=3 lines), and isogenic control (n=3 samples) mESC lines at key developmental time points (Fig. 3A) [140]. These timepoints were: [1] day 0, pluripotent mESCs; [2] day 3, NMP formation; [3] day 4, after 24 hours of RA treatment; [4] day 4.25 and [5] day 5, 6 and 24 hours respectively after BMP4 treatment, when cells enter a dorsal progenitor (dP) state; and [6] day 9, dI differentiation (n = 75 samples, after library preparation and QC, see Methods).

Principal component analysis showed that samples clustered primarily by developmental time point projecting along the expected trajectory from pluripotent to NMP to neural states (Fig. 3C). The Smad1Δ and Smad5Δ lines followed the same overall temporal progression, although their positions were consistently shifted relative to the controls (Fig. 3D), suggesting altered transcriptional states rather than a complete disruption of differentiation. Supporting this

conclusion, Pearson correlations showed that Smad1Δ, Smad5Δ, and control samples are highly correlated at day 0 (Fig. 3F), and day 3 (Fig. 3H). The samples segregate most closely according to development time point, rather than genetic background, with differences becoming most pronounced by day 9 (Fig. 3B). Volcano plots comparing day 0 mESCs from the Smad1Δ, Smad5Δ, and control lines also show relatively few differentially regulated genes (Fig. 3G).

Smad1Δ and Smad5Δ mESCs generate NMPs with altered neuromesodermal potential

We further assessed the consequences of Smad1Δ and Smad5Δ on NMP differentiation by comparing the day 3 transcriptome from each line to control day 0 mESCs. Control, Smad1Δ and Smad5Δ cells all downregulated the pluripotency genes, *Tbx3*, *Klf4*, and *Nanog*, and upregulated NMP markers, including *Nkx1.2*, *Cdx2*, *T,* and *Evx1* (Fig. 3E). Gene Ontology (GO) analysis of significantly upregulated genes at day 3 showed enrichment for terms associated with both mesodermal and neurogenic programs (Fig. 3J-3L). Together, these data indicate that the loss of Smad1 or Smad5 does not prevent the formation of NMPs.

While Smad1Δ, Smad5Δ, and control samples were highly transcriptionally correlated at day 3 (Fig. 3H), some transcriptional differences relative to control were observed (Fig. 3I). Smad1Δ cells showed increased expression of *Zic1*, *Foxg1*, and *Sox1*, genes involved in early neural development and neural progenitor maintenance. In contrast, Smad5Δ cells showed increased expression of mesoderm-associated genes, including *Myf5*, *Asb4*, and *Fabp4* (Fig. 3I). Thus, although the Smad mutants attain an NMP identity, they may be biased towards distinct lineage trajectories. Supporting this hypothesis, immunohistochemistry showed that more Smad5Δ NMPs express Cdx2 protein than the control, consistent with enhanced mesodermal priming (Fig. 3M). In contrast, Smad1Δ NMPs show decreased Cdx2 and Sox2 levels, suggesting they are primed away from a mesodermal trajectory (Fig. 3N).

Stimulating with NMPs with BMP4 during dI differentiation increases chromatin accessibility

We next sought to assess the chromatin landscape from day 4-5 to better understand the developmental events that occur after recombinant BMP4 is added to the differentiation media. Given that chromatin accessibility changes precede transcriptional outputs, this analysis also provided a framework for interpreting subsequent gene expression differences. Assay for Transposase Accessible Chromatin (ATAC)-Seq was performed at days 4, 4.25, and 5 capturing chromatin accessibility changes 6 and 24 hours after adding BMP4 (Fig. 4A). Combining peaks across all conditions and time points generated a unified ATAC-seq peak set of 141,592 peaks. ~46% of peaks mapped to intergenic regions, indicating that the largest fraction of accessible chromatin corresponds to distal regulatory elements (Fig. 4B).

To assess whether BMP4 alters chromatin accessibility, we examined the ATAC-Seq peak profile across developmental time and treatment conditions. At 6 hours, the chromatin accessibility profiles were essentially the between BMP4-treated and untreated samples, indicating minimal early remodeling. However, by day 5, the BMP4-treated samples exhibited increased accessibility relative to the RA-control (Fig. 4C). These changes were restricted to discrete regulatory elements, indicating that BMP4 induces localized rather than global chromatin remodeling (Fig. 4D).

Chromatin accessibility profiling identifies gene targets of BMP4 signaling

ATAC-seq peaks showing increased accessibility after 24 hours of BMP4 treatment compared to RA-treated controls (log2FC >0.5) were defined as BMP4-responsive regions. HOMER motif enrichment analysis identified 53 transcription factor binding motifs associated with BMP4-induced chromatin accessibility changes, 37 of which are present in >15% of target sequences, and include *Gata6*, *Gata2*, *Isl1* and *Lhx2* (Fig. 4E). *Lhx2* and *Isl1* expression is also upregulated after 6 hours of BMP4 treatment (Fig. 4F). Thus these factors, which mark dI1/dI3 identities respectively [136] are upregulated as an early response to BMP4 and may thus drive the initial regulatory programs underlying BMP-dependent chromatin remodeling and gene activation. 32

transcription factor motifs enriched in BMP4-responsive chromatin were not concurrently transcriptionally upregulated, suggesting they reflect a residual chromatin footprint of prior transcription factor activity. For example, *Cdx2* expression is highly expressed in day 3 NMPs, but downregulated by day 5 (Figs. 3M, 4F); yet its motif remains enriched 24 hours after BMP4 treatment, suggesting that BMP4 can selectively stabilized pre-existing accessible regions.

We next sought to identify genes showing concordant promoter accessibility and transcriptional upregulation after BMP4 treatment. 16 promoter regions with significantly increased accessibility in 24-hour BMP4-treated samples were identified, mapped to their closest genes, and validated against bulk RNA-Seq data from days 4 and 5 (Fig. 4G, 4H). This analysis identified *Id2*, *Msx1*, and *Wfikkn1* which all display increased chromatin accessibility proximal to the transcription start site, consistent with promoter remodeling (Fig. 4G). Notably, *Id2* and *Msx1* are established transcriptional targets of BMP signaling [151-153] supporting the model that BMP4 promotes their activation through localized, promoter-associated chromatin accessibility changes. Wfikkn1 (WAP, follistatin/kazal, immunoglobulin, kunitz, and NTR domain–containing) has not previously implicated in dorsal spinal cord development; however, it encodes a secreted multidomain protein that can bind BMP4 and antagonize GDF8 and GDF11 [154, 155]. Thus, BMP4 may dynamically induce *Wfikkn1* expression to establish a negative feedback mechanism that fine-tunes BMP ligand availability during neural differentiation.

Smad1 and Smad5 have shared and distinct roles downstream of BMP4 signaling

ATAC-Seq analysis revealed minimal chromatin remodeling at 6 hours (day 4.25) but robust, localized regulatory changes at 24 hours following BMP4 treatment. We therefore focused on this later timepoint to assess transcriptional differences in control, Smad1Δ and Smad5Δ lines in the day 5 bulk RNA-Seq data. For each genotype, BMP4 treated samples were compared to matched RA-only controls to identify BMP-dependent transcriptional responses, and the overlap

between up- or downregulated genes was determined for the Smad1Δ, Smad5Δ and control cells (Fig. 4L).

298 genes were upregulated across all three lines, accounting for 45% and 26% of the genes upregulated in Smad1Δ and Smad5Δ samples (Fig. 4L). This shared gene set included established BMP targets, such as *Id1*, *Id2*, and *Wfikkn1* (Fig. 4I, 4L), and GO analysis returned terms including “response to BMP” and “Wnt signaling pathway,” consistent with our previous findings [140] (Fig. 4M). Since either Smad1 or Smad5 are required for this response, the R-Smads appear to act redundantly to mediate these BMP-responsive transcriptional programs. In contrast, 67 and 59 genes were specifically upregulated in control/Smad1Δ cells and control/Smad5Δ cells respectively (Fig. 4L), identifying candidate targets selectively regulated by Smad1 or Smad5 (heatmaps, Fig. 4J and Fig. 4K).

We further identified 172 upregulated and 99 downregulated genes expressed in Smad1Δ and Smad5Δ, but not control, BMP4-treated cells (Fig. 4L). The presence of these genes exclusively in the Smad1Δ and Smad5Δ datasets suggests that loss of either R-Smad disrupts shared transcriptional mechanisms. A GO analysis of these genes implicated pathways involved in extracellular matrix organization, synapse organization, and neural migration (Fig. 4N). Finally, there were relatively few uniquely regulated genes in the Smad1Δ cells, which yielded minimal GO term enrichment (Fig. 4O). In contrast, there were substantially more uniquely regulated genes in Smad5Δ cells (Fig. 4L). GO analysis suggests these Smad5-dependent genes regulate non-canonical or context-dependent functions (Fig. 4P), as previously found [156]. Taken together, these findings support a model in which Smad1 and Smad5 play shared and distinct transcriptional roles mediating the transition from the NMP to dP state, with Smad5 having a broader role regulating the response to BMP4.

Smad1 and Smad5 have distinct roles in dorsal spinal lineage specification

To assess whether Smad1 and Smad5 have differential requirement for dorsal lineage specification, we generated single-cell (sc) RNA-Seq atlases for Smad1Δ, Smad5Δ, and control cells at day 9 of the RA±BMP4 differentiation protocol. To enable accurate comparison across the six datasets [RA-control (n = 13,822); RA-Smad1Δ (n = 11,157); RA-Smad5Δ (n = 10,616); RA+BMP4-control (n =8,276); RA+BMP4-Smad1Δ (n=10,875); and RA+BMP4-Smad5Δ (n=7,944)], they were also integrated into a single combined transcriptional atlas (Fig. 5A). Given the Smad-dependent alterations in NMP specification (Fig. 3), we first assessed the proportion of *Foxc1*$^{+}$, *Twist1*$^{+}$, *Meox2*$^{+}$ mesodermal and *Pax3*$^{+}$, *Myt1l*$^{+}$, *Msx1*$^{+}$ neural derivatives generated in the three lines (Fig. 5B, 5C). In the RA condition, ~90% of cells were classified as neural across genotypes (Fig. 5D). BMP4 addition resulted in ~25% of cells following a mesodermal trajectory in both the control and Smad1Δ samples (Fig. 5E). This proportion was strikingly expanded in the Smad5Δ group, which had ~40% mesodermal population (Fig. 5E). These findings support a model in which Smad5 biases NMPs towards neural fates at the expense of mesoderm induction (Fig. 5J).

We then subdivided the neural category into broad lineage divisions – cycling progenitors, roof plate, dorsal progenitors, neurons - based on the expression of canonical marker genes (Fig. 5B, 5F). In the RA+BMP4 condition, where R-Smads are activated in response to extrinsic BMP4 (Fig 1N), progenitors accumulate at the expense of neurons in both R-SmadΔ lines, compared to control (Fig. 5H). However, in the RA condition, where R-Smads are activated by intrinsic BMP signaling (Fig 1M), control and Smad5Δ lines show similar cellular proportions, while roof plate cells accumulate at the expense of neurons in the Smad1Δ line (Fig. 5G). Together, these data suggest that both R-Smads are required for progenitors to initiate the differentiation process, but then Smad1 specifically suppresses them from progressing to a roof plate fate (Fig. 5K).

Finally, we assessed the requirement for Smad1 and Smad5 in the specification of dPs and dIs. The scRNA-Seq data for dPs and neurons were subsetted and reclustered to generate an integrated neural atlas for Smad1Δ, Smad5Δ, and control cells (Fig. 6A). Annotation using established dP and dI marker genes (Fig. 6B) confirmed that the RA±BMP4 protocols direct the complete complement of dP (dP1-dP6), and dI (dI1- dI6) populations (Fig. 6C, 6F). dP phenotypes were specifically observed in the RA protocol, where the R-Smads have opposing effects. Smad5Δ results in the loss of dP1-dP3, whereas Smad1Δ increases the proportion of dP1 and dP2 (Fig. 6D, 6I), suggesting the R-Smads function antagonistically in dP specification. dI phenotypes were observed in both protocols (Fig. 6J). Most notably, the dorsal-most dIs, especially the dI1s, were markedly and consistently lost in the Smad5Δ condition relative to control, with a corresponding increase in the intermediate dI fates (Fig. 6H, 6I). dI1s were also lost in the Smad1Δ condition, but only in the RA+BMP4 protocol, perhaps reflecting a requirement for Smad1 when BMP4 signaling is amplified. The dI1 phenotype was validated by immunohistochemistry, confirming a significant reduction in $Lhx2^{+}$ dI1 cells in the Smad5Δ line (Fig. 6K, 6L). Together, these data support a model where the R-Smads play distinct roles in dI fate specification. Smad5 plays the dominant role, acting reiteratively to specify the dorsal-most dI fates (Fig. 7M).

RNA velocity analysis identifies genes governing dP state transitions

To further dissect the role of the R-Smads, we performed an RNA velocity analysis on our integrated neural atlas to infer the direction and dynamics of dI state transitions. The resulting global velocity field showed that dPs directionally flowed into their respective dI lineages in the integrated atlas, reflecting the expected developmental progression (Fig. 7A). Unexpectedly, a transcriptionally distinct progenitor cluster (Fig. 6B) was observed to give rise to dP2-dP6, suggesting this represents a pool of multipotent dorsal progenitor (mdP) cells (Fig. 7A). We next examined the velocity magnitude for the control cells (Fig. 7C). The dP populations exhibited the

highest velocity lengths, consistent with being in a transitioning state (Fig. 7C). By contrast, the dI populations have reduced transcriptional dynamics, with exception of the dI4 and early dI5 trajectories, which retain elevated velocity lengths (Fig. 7C). Thus, dI4/dI5s may remain in a dynamic state, as they activate subtype-specific transcriptional programs.

To identify genes that drive or mark these cell state transitions, we performed velocity-based gene ranking across the control clusters. This analysis revealed cluster-specific genes where their transcriptional dynamics are strongly correlated with velocity directionality. *Magi1* and *Smoc1* were identified as two of the top-ranked velocity gene. *Magi1* is broadly expressed across multiple dP clusters (Fig. 7E), and may have a general role regulating cell polarity and cell-cell adhesion [157]. Conversely, *Smoc1* expression is specific to the dP4/5 population, suggesting more specific role in this lineage (Fig. 7F). *Smoc1* has previously been identified a dP5 marker gene in an independent pseudotime analysis [149], and acts as a BMP agonist, although this activity depends on context [158, 159]. Smoc1 may thus act through BMP signaling to maintain velocity magnitude in transitioning dP5s. Together, this RNA velocity analysis captures biologically relevant transcriptional dynamics underlying dI state transitions.

Smad5 controls transcriptional dynamics across dP and dI clusters

Having established the baseline velocity dynamics within a control atlas, we investigated whether these transcriptional programs are altered in Smad1Δ or Smad5Δ cells. We quantified mean velocity length for all dorsal populations in the RA±BMP4 protocols and found that loss of the R-Smads results in reduced velocity magnitudes across the dP clusters, with the most marked decreases in Smad5Δ cells (Fig. 7G-7J). Notably, the dP4-dP6 populations, previously thought to be BMP independent, showed the most pronounced reductions in velocity length. While dI populations were not broadly affected, the dI1 trajectory was an exception, showing reduced velocity vectors in Smad1Δ and Smad5Δ cells compared to control (RA+BMP4 protocol, Fig. 7G, 7I). We sought to identify the genes driving velocity in dI1s and identified *Nav3* (neuronal navigator

3) and *Cadm2* (cell adhesion molecule 2) as the top-ranked genes with reduced velocity in both lines (Fig. 7K). Both genes have enriched expression in a subset of dI1s that are markedly lost in the Smad5Δ atlas (Fig. 7L). Nav3 [160, 161] and Cadm2 [162] regulate cytoskeletal dynamics and cellular adhesion respectively, suggesting that Smad1Δ and Smad5Δ dI1s are impaired for neurite extension, axon guidance, and synapse formation.

While both Smad1Δ and Smad5Δ trajectories exhibit reduced RNA velocity, this effect is most pronounced for the Smad5Δ line. Thus, we propose Smad5 as the principal mediator of the transcriptional programs that maintain progenitor activity and regulate their progression into differentiated dI states (Fig. 7N).

Discussion

In this study, we use an mESC to dI *in vitro* model system to investigate if Smad proteins differentially mediate downstream BMP signaling during spinal cord development. Bulk and scRNA seq datasets, derived from control, Smad1Δ, and Smad5Δ cultures across multiple developmental timepoints within our protocol, reveal Smad1 and Smad5 have distinct functions. We found the loss of either Smad1 or Smad5 does not prevent mESCs from acquiring a neuromesodermal identity, suggesting that these proteins are either dispensable or play redundant roles during the generation of bipotent NPCs (Fig.3). Beyond this stage, however, their roles diverge. Smad5 is essential for directing progenitors toward a neural lineage and preventing their differentiation into mesodermal derivatives (Fig.5). As differentiation proceeds, Smad1 restricts the expansion of dP1 progenitors, whereas Smad5 is required for the formation of dP1–dP3s and their progression into dI1–dI3 interneurons, while simultaneously limiting the expansion of more intermediate dI4–dI6 populations (Fig.6). Interestingly, Smad5 lowers velocity across all dorsal progenitor populations, suggesting a broad dampening of transcriptional programs that allow dPs to transition to dIs (Fig. 7). Finally, our experimental approach identified numerous BMP4-responsive genes and genomic regions, that may function

in other developmental contexts or tissues where BMP signaling plays a regulatory role (Fig.4). Together, this demonstrates R-Smads can function in distinct roles and offers further insight into how BMP signaling drives diverse cellular behaviors in multiple developmental systems.

Non-canonical roles of Smad5

Non-canonical functions of Smad5 have begun to emerge. Cytoplasmic, non-activated, Smad5 has been linked to modulating the bioenergetic homeostasis of cells in response to fluctuations in intracellular pH[156]. Throughout our bioinformatic datasets we consistently see that the loss of Smad5 produces a broader transcriptomic impact than loss of Smad1 (Fig. 4). Following 24 hours of BMP4 treatment, we identified 596 upregulated and 458 downregulated genes unique to the Smad5Δ atlas, underscoring the extensive disruption of gene expression when Smad5 is absent. Gene ontology analysis supported previous claims, as categories spanning 'mitochondrial ATP synthesis' and 'regulation of neuronal differentiation' (Fig. 4P) were found when querying with the downregulated genes (Fig. 4). Given that specific metabolic programs are tightly coupled to neural progenitor proliferation and differentiation[163], an altered metabolism may explain why the Smad5Δ progenitors have a decreased RNA velocity. Furthermore, our gene sets that we deem to be regulated by Smad5 are a valuable resource for those who are probing for additional non-canonical functions of R-Smads

Wfikkn1 as regulator of BMP signaling

Several well-characterized BMP antagonists bind BMP ligands and modulate signaling strength. Here, we identify *Wfikkn1* (WAP, Follistatin/Kazal, Immunoglobulin, Kunitz, and Netrin Domain–Containing 1) as a candidate BMP4-induced gene in the developing dorsal spinal cord. Wfikkn1 is a secreted protein shown to bind BMP2 and BMP4, yet it does not overtly inhibit their activity, leaving its physiological role unresolved[154, 155]. It has been suggested that Wfikkn1 may function to localize or fine-tune BMP4 signaling, acting as a context-dependent modulator rather than a

classical antagonist. This raises the intriguing idea that Wfikkn1 may participate in a negative-feedback loop, similar to how other pathways, such as FGF signaling via Sprouty proteins, induce their own attenuators to shape the spatial range and intensity of the signal[164]. This possibility is particularly compelling given that BMPs have traditionally been viewed as morphogens, yet our findings indicate that BMP4 can act in a signal-specific manner rather than through a simple concentration gradient[165]. Wfikkn1 may therefore help restrict, localize, or attenuate BMP4 activity during dorsal spinal cord development. However, additional work will be required to verify its expression and spatial localization in the developing spinal cord, before any conclusions can be drawn about its functional involvement.

Identification of driver genes governing dP state transitions

The mechanisms that pattern the dorsal spinal cord and specify distinct dorsal progenitor and interneuron populations remain incompletely defined. Through RNA velocity analysis of our control cell atlas, we identified genes that are being actively transcribed within dorsal progenitor populations. These dP-specific candidate driver genes likely play essential roles in enabling progenitors to transition into their respective dP subtypes, addressing a key gap in our understanding of dorsal spinal cord development.

This set of candidate driver genes represents a valuable resource for groups. Like our own, who are working to use directed differentiation protocols to generate defined dorsal interneuron populations for regenerative or cell-replacement applications. Current protocols often fail to produce pure or fully representative dI subtypes, with some lineages under-represented or absent. By providing insight into the transcriptional programs and developmental trajectories that guide dP progression, these driver genes may inform refinements to existing differentiation strategies.

Together, these findings bridge a key gap in our understanding of dorsal spinal cord development while offering actionable insight for improving the generation of specific dorsal interneuron populations *in vitro.*

Mechanistic implications for BMP signal transduction

Although Smad1 and Smad5 have long been considered functionally redundant R-Smads, our work demonstrates that they can operate through distinct, non-overlapping mechanisms. This challenges the prevailing assumption that BMP pathway specificity arises primarily from ligand–receptor combinations or extracellular modulators, and instead highlights differential R-Smad usage as an underappreciated source of signaling diversity. It will be important to determine whether the BMP pathway similarly relies on selective engagement of individual R-Smads in other tissues, and whether this phenomenon extends to the R-Smads that mediate TGF-β signaling, Smad2 and Smad3.

While our findings show that Smad1 and Smad5 can differentially regulate transcription, the mechanisms underlying these differences remain unclear. R-Smads have intrinsically weak DNA-binding affinity, and their genomic engagement is largely stabilized through interactions with co-factors. One possibility is that Smad1 and Smad5 differ in their co-factor recruitment profiles, resulting in altered DNA-binding affinity or site selectivity and thereby contributing to their distinct transcriptional roles. In addition, the stoichiometry of R-Smad/co-Smad4 complexes may further influence DNA-binding affinity or functional versatility in the nucleus. Such differences could explain why, 24 hours after BMP4 treatment, certain gene sets are misregulated in both Smad1Δ and SmadΔ datasets, but not in control (Fig. 4). Furthermore, the phosphorylation and ubiquitination sites that reside in the proline-rich linker region connecting the R-Smad MH1 and MH2 domains have been shown to alter R-Smad nuclear import and export[166]. These sites differ between Smad1 and Smad5, and this may be a mechanism by

which they are differentially trafficked to and from the nucleus, contributing to their distinct functional roles during BMP signal transduction.

Overall, these findings highlight Smad1 and Smad5 are not functionally interchangeable, though additional studies will be required to uncover the mechanisms that give rise to these differences.

Materials and Methods

Differentiation of mouse embryonic stem cells into dorsal interneurons

Control, Smad1Δ, and Smad5Δ mESCs were differentiated as previously described[167] (see appendix).

Western Blots across differentiation timeline

Cells were washed with 1x PBS and cold Pierce® RIPA buffer (Thermo Scientific) supplemented with a 1x protease inhibitor cocktail (Roche) and 1x phosphatase inhibitor cocktail (Roche) was directly added to multiple wells of a 24 well differentiation plate. The lysed cell suspensions were pooled, held on ice for 15 minutes, then centrifuged while at 4°C for 15 mins. at 14,000 x g. A BCA assay was performed on the supernatant to determine protein concentrations. 40ug of protein was used as input for SDS-PAGE, using a gradient 4-12% Bis-Tris SDS gel (GenScript) followed by transfer onto PVDF membranes (Millipore Sigma). The membranes were blocked using 5% non-fat dry milk for 1 hour (Bio-Rad Laboratories). The blocked membranes were then incubated with the primary antibodies (key resources table) at 4°C overnight. Thereafter, the membranes were washed three times with TBST (20 mM Tris, 150 mM NaCl, 0.1% Tween 20), incubated with species-specific horseradish peroxidase-conjugated secondary antibodies (Jackson ImmunoResearch Laboratories) for an hour at room temperature and then washed twice with TBST and once with TBS. The chemiluminescent bands were analyzed using the Pierce Femto Chemiluminescence kit on the ChemiDoc™ imaging system (Bio-Rad).

Immunohistochemistry

Cell cultures were washed with 1x PBS and fixed with fresh cold 4% PFA for 8 min in the well. Following fixation, cultures were washed twice with 1xPBS to remove any remaining PFA. Cells were blocked with 1xPBS with 1% heat inactivated horse serum for 1 h and primary antibodies are added in the blocking solution for an overnight incubation at 4°C. Following washes, species appropriate secondary antibodies (Jackson Immuno Research Laboratories) were added in 1x PBST (PBS +0.1% Triton 20) for 1hr. The cultures were then washed with 1x PBST to remove traces of secondary antibodies and counterstained with DAPI to stain nuclei. Plates were then imaged on Zeiss LSM800 inverted confocal microscope at 20× magnification.

Generating Smad1 and Smad5 deficient mESC lines

A CRISPR-mediated genomic deletion strategy was utilized as described[150], to generate Smad1 and Smad5 deficient mESC lines. Briefly, guide rnas (gRNAs) targeting upstream and downstream of exon2 in either the *Smad1* or *Smad5* gene locus were cloned into the pSpCas9(BB)-2A-Puro V2.0 plasmid. Various combinations of gRNA constructs were then delivered to the cell via the Lonza 4D-Nucleofector® X Unit. Nucleofected cells were plated at a single cell density onto tissue culture plates containing mESC media supplemented with puromycin. Colonies that grew in the presence of puromycin were isolated, expanded, and further characterized to confirm the correct genomic deletion.

Characterization of Smad1 and Smad5 exon2 genomic excision

We confirmed successful excision of the exon 2 locus by PCR, Sanger sequencing, and western blotting. Candidate Smad1Δ and Smad5Δ mESC lines were grown on 0.1% gelatin-coated plates and passaged at 75% confluency. After two passages, cells were washed with 1x PBS and incubated with 0.25% trypsin at 37°C for 3 min. Trypsin dissociation activity was neutralized with mESC media in a 1:1 ratio and cells were collected and centrifuged at 1,000 RPM for 5 min. Cell pellets were resuspended in 1x PBS and evenly partitioned for downstream analysis.

PCR and Sanger sequencing: Genomic DNA was isolated from the resuspended cell pellet using the QIAamp® DNA Mini kit and following the manufacturer's protocol. PCR amplification across the CRISPR-targeted exon 2 genomic region was performed using primers designed to anneal outside the deleted interval. PCR conditions were: 95 °C for 2 min; 30 cycles of 95 °C for 15 s, 55.7 °C for 30 s, and 72 °C for 3 min; followed by 72 °C for 5 min and held at 4 °C. Amplicons were resolved on either a 1.25% (Smad1Δ) or 2% (Smad5Δ) agarose gel. After visually confirming the presence of a large genomic deletion, the remaining PCR products were cleaned up using the MinElute™ PCR purification Kit (Qiagen) and submitted for Sanger sequencing. PCR amplification of genomic DNA from the SF20 mESC line resulted in two different sized PCR products; each band was gel extracted using the Monarch® DNA Gel Extraction Kit (New England Biolabs) and submitted separately for Sanger sequencing. Sequence traces were aligned to the GRCm38 *Mus musculus* genome using the ApE (A Plasmid Editor) program to determine the size and boundaries of the genomic deletion.

Western Blot: Cell suspension was centrifuged for 5 min. at 1000 RPM and resuspended in 1x Pierce® RIPA buffer (Thermo Scientific) supplemented with a 1x protease inhibitor cocktail (Roche). The lysed protein suspension was held on ice for 15 mins. then centrifuged while at 4°C for 15 mins. at 14,000 x g. The supernatant was subjected to SDS-PAGE and the remaining steps were carried out as previously described in the above methods section.

Bulk-RNA sequencing and data processing

Sample collection, library preparation, and sequencing: Total cell lysate was obtained in buffer RLT (Qiagen) at different time points from three independent differentiations (biological replicates) conducted parallelly, for each genotype: control (mm13), Smad1Δ (C2, C34, C36) and Smad5Δ (SF20, SF27, SF43). RNA extraction was performed using RNeasy mini kit (Qiagen) and the quality was determined using Agilent Technologies 2100 Bioanalyzer. Stranded libraries were constructed using Universal plus mRNA sequencing kit (NuGEN) and

sequenced onto 1.5 lanes of NovaseqS4 to generate a minimum of 40 million reads/sample (with the exception of D4-C34 at 24M reads). Notably, 4 samples failed library prep (D0-SF20, D0-SF27, D3-SF43, D9-B4-ControlA) and 3 samples did not generate an adequate library for sequencing (D3-SF27, D5-RA-SF27, D9-RA-SF20). A fourth differentiation was performed to replace the D0 through D5 samples, resulting in a total of 79 samples that were sequenced.

Data processing: Raw sequencing reads were aligned to the *Mus musculus* GRCm38 reference genome using the STAR aligner and gene-level counts were generated with featureCounts (Subread package). Genes with fewer than 10 total reads across all samples were removed prior to normalization. Quality control was performed using Pearson correlation and principal component analysis (PCA) to assess replicate similarity and genotype clustering. Differential expression analysis was conducted using the DESeq2 package in R. Differentially expressed genes were extracted using DESeq2 contrast statements corresponding to: (1) time effects within WT, Smad1Δ, and Smad5Δ, and (2) genotype differences (Smad1Δ vs WT and Smad5Δ vs WT) at D0, D3, and D4. For the later differentiation stages, samples collected at D4, D4.25, D5, and D9 under RA or RA+BMP4 media conditions were analyzed using a grouped DESeq2 design. Differential expression analyses were performed to compare RA vs RA + BMP4 at matched timepoints (D4.25 and D5) within each genotype. Significant genes were defined as those with an adjusted p-value (padj) < 0.05 and absolute log2FoldChange >1. Venn-like overlap analyses were performed to identify genes uniquely or commonly regulated across WT, Smad1Δ, and Smad5Δ backgrounds at D4.25 and D5. Notably, 3 samples were excluded due to a file level read error encountered during FASTQ data retrieval (D0-ControlB, D4.25-B4-ControlC, D9-RA-C34).

Gene Ontology (GO) enrichment analysis: Gene Ontology enrichment analysis was performed using the clusterProfiler package. Biological process enrichment was tested using the

Benjamini–Hochberg method for multiple-testing correction, with significance thresholds of p-adjusted < 0.05 and q-value < 0.2.

Generation of heatmaps: Variance-stabilized expression values were averaged within each timepoint–genotype–media group, row-centered by Z-scaling, and visualized as heatmaps using the pheatmap package in R.

ATAC-sequencing

We performed parallel differentiations of three control (biological replicates) mESC lines and collected samples for ATAC sequencing at D0, D3, D4, D4.25 (6 hours ± the addition of BMP4) and D5 (24 hours ± the addition of BMP4).

Sample preparation and sequencing: Samples were prepared as previously described[168]. Cultures were dissociated with 0.25% cold trypsin for 5 min. Trypsin activity was stopped with the addition of trypsin inhibitor (Sigma-Aldrich) in 1:1 ratio. Cells were then pelleted by centrifugation at 1000RPM and resuspended in 1x PBS. Cells were counted and 75,000 cells were collected and centrifuged at 500 x g for 5 mins at 4°C. The cell pellet was resuspended in 50 uL of cold lysis buffer (10mM Tris-Cl pH7.4, 10mM NaCl, 3mM $MgCl_2$, 0.1% IGEPAL CA-630) and immediately centrifuged at 500 x g for 10 min at 4°C. Nuclei were resuspended in transposition reaction mix containing TDE1 Tn5 Transposase and 1x TD reaction buffer (Illumina) and incubated at 37°C for 30 min. Transposed DNA was purified using the MinElute™ PCR Purification Kit (Qiagen) per the manufacturer's instructions. Libraries were sequenced on a NovaSeq S4 (XP workflow, paired-end 2×100 bp), yielding an average of ~120 million reads per sample.

Preprocessing: Nextera adapter sequences were trimmed and low-quality bases removed using Trimmomatic. Trimmed reads were aligned to the GRCm39 mouse reference genome using Bowtie2. Resulting SAM files were converted to BAM format, coordinate-sorted, and indexed

using SAMtools. PCR duplicates were identified and removed using Picard MarkDuplicates. Reads mapping to ENCODE blacklisted regions were excluded using BEDtools intersect with the mm39 blacklist (v2). Filtered BAM files were merged across biological replicates per condition, re-sorted, and re-indexed using SAMtools. Open chromatin regions were identified by peak calling with MACS3. Normalized coverage tracks were generated using deeptools bamCoverage with counts per million (CPM) normalization and a bin size of 25 bp.

Unified Peak Set and ATAC Signal Quantification: A unified set of regulatory regions was generated by centering fixed-width windows on peak summits identified across all conditions, resulting in a unified peak set of 141,592 peaks. Genomic annotation of the unified peak set was performed using annotatePeaks.pl from the HOMER suite against the mm39 reference genome. ATAC signal over the unified peaks was quantified using computeMatrix (deepTools) in reference-point mode, anchored at the center of each peak with 500 bp flanking regions on each side. Chromatin accessibility was visualized using plotHeatmap and plotProfile from deepTools, displaying mean CPM signal centered on peak summits across samples.

Motif enrichment analysis: Treatment-responsive regulatory elements at 24 hours were defined as unified ATAC-seq peak regions exhibiting a $\log_2$ fold-change > 0.5 in normalized accessibility signal between treated (+BMP4) and untreated (RA) cells. Transcription factor motif enrichment analysis was performed on these peaks using HOMER (findMotifsGenome.pl) against the mm39 reference genome.

Promoter-Centric Analysis: Promoter regions were defined as ±2 kb windows centered on the transcription start site (TSS) of each annotated gene. BMP4 responsive ATAC peaks at 24 hours were intersected with promoter regions using BEDtools intersect, identifying 16 promoter regions with increased chromatin accessibility. These regions were mapped to their nearest gene and intersected with bulk RNA-seq data from the same timepoint comparison (D5 ±

BMP4). Genome browser tracks displaying CPM-normalized ATAC signal at the promoters of candidate genes (*Id2*, *Msx1*, and *Wfikkn1*) were visualized over a ~±2 kb window centered on the annotated TSS of each gene.

Single-cell RNA sequencing

We performed parallel differentiations of control (mm13), Smad1Δ (C36) and Smad5Δ (SF43) mESC lines and collected samples for single-cell sequencing at the end of our mESC to dI differentiation protocol.

Preparation of single cell suspension: Control, Smad1Δ, and Smad5Δ day 9 cultures, from both RA and RA + BMP4 protocols, were dissociated with 0.25% cold trypsin for 5 min. Trypsin activity was stopped with the addition of trypsin inhibitor (Sigma-Aldrich) in 1:1 ratio. Cells were then pelleted by centrifugation at 1000RPM and resuspended in 1x PBS. Dead cells were removed using MACS dead cell removal kit (Miltenyi Biotec Inc.) by following the manufacturer's protocol. Eluted live cells were then suspended in 1X PBS containing 0.04% BSA solution (400µg/ml) for the library preparation.

Library preparation and sequencing: ~10,000 live cells/conditions were used to construct single-cell specific cDNA libraries using protocol described in 10x Genomics chromium single cell 3′ reagent kit (v3.1 Chemistry). Briefly, cells were partitioned into nanoliter-scale Gel-Beads-in-emulsion (GEM) using the 10x Chromium controller. Each GEM contains a unique barcode which is shared among the cDNA generated from a single cell. Cells were then lysed, and cDNA synthesis and feature barcoding were performed in the GEMs. The sequencing libraries were recovered using Magnetic separation and the quantity and quality of cDNA were assessed by Agilent 2100 expert High Sensitivity DNA Assay. Libraries were sequenced across two lanes of an Illumina NovaSeq X Plus using a 10B flow cell.

Preprocessing and filtering: Raw 10x Genomics count matrices were imported into Seurat. For each sample, cells were filtered based on thresholds determined from QC plots. Specifically, cells were retained if they met sample-specific criteria for gene complexity (nFeature_RNA > 250, with upper limits of 7,300–8,000 depending on sample), total UMI counts (800 < nCount_RNA < 38,000–50,000), and mitochondrial content (percent.mt < 10%). These thresholds were applied independently to each genotype–condition dataset (WT, Smad1Δ, and Smad5Δ under RA or BMP4 conditions) to remove low-quality cells, empty droplets, and potential doublets. Cells failing to meet any of these criteria were excluded prior to library-size normalization, dimensionality reduction, and subsequent downstream analysis.

Data integration: Following per-sample quality control, the six filtered Seurat objects (WT, Smad1Δ, and Smad5Δ under BMP4 and RA conditions) were merged into a single Seurat object and split by genotype–treatment group using SplitObject(). Each sub-object was independently normalized with SCTransform(), and the top 3,000 variable features were selected across all groups using SelectIntegrationFeatures() and prepared for integration with PrepSCTIntegration(). Cross-dataset integration anchors were identified with FindIntegrationAnchors() and used to generate a single batch-corrected object via IntegrateData(), both using SCT normalization.

Dimensionality reduction, clustering and Cell Type annotation: Following integration, principal component analysis (PCA) was performed on the corrected integrated assay using RunPCA(). Uniform Manifold Approximation and Projection (UMAP) was computed using the top 30 principal components as input via RunUMAP(). Unsupervised clustering was performed using a shared nearest-neighbor graph constructed with FindNeighbors() across the same 30 PCs, followed by Louvain community detection at a resolution of 0.8 using FindClusters().

Cell type identities were assigned to each Seurat cluster using the ScType automated annotation framework[169]. ScType was run on the SCT-normalized, scaled assay with a custom major lineage marker gene file. Cluster-level annotations generated by ScType were used to label cells in a final UMAP visualization, in which each cell was colored by its ScType-assigned major lineage identity. To visualize and quantify the proportion of NMP derivatives, the progenitor, roof plate, dorsal progenitor, and neuron clusters were grouped into a single "neural" category, while mesoderm and vascular clusters were grouped into a single "mesoderm" category.

To enable fair visual comparison of cell type composition across samples, cells were down sampled prior to plotting. The minimum cell count across all genotype–treatment groups was determined, and each group was randomly subsampled to this number using a fixed random seed (seed = 123) to ensure reproducibility. A down sampled Seurat object was generated from the retained cells and used exclusively for visualization (Fig.5 D,E,G,H).

Cell type proportion analysis: Cell type proportions were quantified per genotype–treatment group by tabulating the number of cells assigned to each cluster within each sample. Proportions were expressed as a percentage of total cells per group and visualized as pie charts for each of the six genotype–treatment combinations. Category labels were displayed within each slice only when the proportion exceeded 5% of total cells.

Marker gene dot plot: To validate ScType cluster annotations, a dot plot was generated using a curated panel of known marker genes across all annotated cell type clusters. Expression was visualized from the SCT-normalized assay using Seurat's DotPlot() function. Dot size reflects the percentage of cells within each cluster expressing a given gene, and dot color reflects the scaled mean expression level. Genes absent from the dataset were excluded prior to plotting.

dP and dI subset, reclustering, and annotation: Cells annotated as dorsal progenitor or neuron in the full integrated object were extracted as a subset. PCA, UMAP, and unsupervised clustering were performed as previously described. Cell type annotation of the filtered neural subset was performed using ScType with a curated dP and dI marker panel. Upon inspection of dorsal progenitor marker gene expression, the dP3 population was not captured as a discrete cluster. To recover this population, cells initially annotated as dP2 that expressed *Ascl1* above a normalized expression threshold of 0.25 were manually relabeled as dP3. This revised annotation was stored as a separate metadata field and used for all downstream visualization and analyses of the dP/dI subset. Noteably, prior to clustering, an anomalous neural cluster was identified upon inspection and its relative contribution to each genotype–treatment group was quantified. As this cluster represented a minor and inconsistent population, it was excluded from all subsequent analyses. Down sampled UMAPS were generated as previously described (Fig. 6 D,E,G,H). Cell type proportion analysis was performed as previously described. Marker gene dot plot was generated as previously described.

RNA velocity analysis

Data preparation and preprocessing: Spliced and unspliced RNA count matrices were generated for each sample using velocyto and loaded into Python using the anndata library. Each sample was then filtered to retain only cells present in the complete neural atlas (Fig. 6A). The six filtered samples were concatenated into a single AnnData object. To maintain consistency with the Seurat analysis, UMAP coordinates and cluster annotations were transferred directly from the neural atlas Seurat object by loading the exported embeddings and cluster label metadata and mapping them to the AnnData object.

Preprocessing for velocity estimation was performed using scVelo. Genes were filtered and normalized using a minimum shared spliced/unspliced count threshold of 20 and a minimum total count of 20. A nearest-neighbor graph was constructed using 30 neighbors, and first- and

second-order moments of the spliced and unspliced count distributions were computed across the neighborhood graph in preparation for velocity estimation.

Analysis of integrated neural atlas: RNA velocity was estimated using the stochastic model implemented in scVelo. The spliced/unspliced read proportions were inspected per cluster to confirm adequate unspliced RNA recovery across cell types. Velocity vectors were projected onto the UMAP embedding as a streamline plot, with cells colored by cluster identity, to visualize the directionality and continuity of transcriptional trajectories across the dP/dI populations.

Identification of driver genes in control cells: To characterize baseline transcriptional dynamics, velocity length, confidence, and confidence transition were assessed in control cells stemming from both RA and RA+BMP4 protocols. Velocity genes were ranked across clusters in WT cells using scv.tl.rank_velocity_genes(), with cells grouped by cluster identity and a minimum correlation threshold of 0.3. Phase portraits were generated for selected top velocity genes including *Magi1* and *Smoc1*.

Velocity analysis in Smad1Δ and Smad5Δ data sets: Velocity streamline plots were generated for each of the six groups (Control, Smad1Δ, and Smad5Δ under RA and RA + BMP4 protocols). Velocity magnitude was quantified across groups and clusters by computing the mean velocity length, velocity confidence, and velocity confidence transition per group–cluster combination.

dI1 cells from RA+BMP4 treated samples were isolated and the mean absolute velocity magnitude was computed per gene for each group. Genes with negligible velocity signal across all groups were excluded. For each remaining gene, pairwise comparisons between control and each mutant were performed using the Wilcoxon rank-sum test, and p-values were corrected for multiple testing using the Benjamini–Hochberg false discovery rate method. Candidate genes with significantly reduced velocity in mutant relative to control cells (FDR < 0.05, delta > 0) were identified separately for Smad1Δ and Smad5Δ comparisons and ranked by effect size. *Nav3*

and *Cadm2* were identified as genes with significantly reduced mean velocity magnitude in Smad1Δ and Smad5Δ cells relative to control. The normalized expression of *Nav3* and *Cadm2* was visualized on the dI1 UMAP for each RA+BMP4 treated group.

Quantification and statistical analysis

Image quantification: To count the number of nuclei, images were converted to 8bit image in ImageJ, their threshold intensity was adjusted to capture only the florescent nuclei, and the number of nuclei counted using the analyze particle tool in ImageJ.

Statistics: Data are represented as mean ± SEM (standard error of the mean). Tests for statistical significance were performed using Prism software. Values of $p < 0.05$ were considered significant in all cases.

Key Resource Table:

Reagents	**Source**	**Identifier/Catalogue number**
Antibodies		
Goat polyclonal anti-Lhx2	Santa Cruz	Sc-19344
Mouse monoclonal anti-Oct-4	Cell Signaling Technology	75463
Rabbit monoclonal anti-Nanog	Cell Signaling Technology	4903S
Goat polyclonal anti-Sox2	R&D systems	AF2018
Rabbit polyclonal anti-Myt1l	Proteintech	25234-1-AP

Mouse monoclonal anti-Pax3	R&D systems	MAB2457
Rabbit polyclonal anti-Tubulin β-3	BioLegend	802001
Guinea pig polyclonal anti- Lmx1b sera	gift from Thomas Müller	N/A
Mouse monoclonal anti-Lhx1/5	Developmental Studies Hybridoma bank	4F2
Rabbit monoclonal anti-PSmad1/5/9	Cell Signaling Technology	11971
Rabbit polyclonal anti-Smad5	Proteintech	12167-1-AP
Goat polyclonal anti-Smad1	R&D Systems	AF2039
Goat polyclonal anti-Brachyury	R&D Systems	AF2085
Rabbit monoclonal anti-Cdx2	Cell Signaling Technology	D11D10
Commercial Assays		
QIAamp® DNA Mini kit	Qiagen	51304
MinElute™ PCR purification Kit	Qiagen	28004
Monarch® DNA Gel Extraction Kit	New England Biolabs	T1020

SuperSignal™ West Femto Maximum Sensitivity Substrate	Thermo Scientific	34096
Illumina Tagment DNA Enzyme and Buffer Small Kit	Illumina	20034197
RNeasy plus mini kit	Qiagen	74106
Dead cell removal kit	Miltenyi Biotec Inc.	130-090-101
Chemicals, reagents and peptides		
RIPA Lysis and Extraction Buffer	Thermo Scientific	PI89900
Precast polyacrylamide gels	GenScript	M00653
PhosSTOP™	Roche	4906845001
cOmplete™ Protease Inhibitor Cocktail	Roche	11697498001
CHIR99021	Tocris Bioscience	4423
Human recombinant BMP4	Thermo Fisher Scientific	PHC9534
Retinoic acid	Sigma Aldrich	R2625
Human basic FGF (bFGF)	Thermo Fisher Scientific	PHG0023
DMEM:F12	Cytiva	SH30023.01
Neurobasal medium	Thermo Fisher Scientific	21103049

B27 supplement	Fisher Scientific	17-504-044
N2 supplement	Thermo Fisher Scientific	17502048
Mouse LIF	Millipore	ESG1107
Trypsin Inhibitor 1x solution	Sigma-Aldrich	T7659
0.25% Trypsin-EDTA	Thermo Fisher Scientific	25200–056
BD Pharmigen DAPI solution	BD Biosciences	564907
DMEM high glucose	Cytiva	SH30022.02
EmbryoMax nucleosides (100x)	Sigma-Aldrich	ES-008-D
0.1% Gelatin in water	STEMCELL Technologies	07903
2-Mercaptoethanol	Thermo Fisher Scientific	21985023
Embryonic stem cell-qualified fetal bovine serum (FBS)	Thermo Fisher Scientific	16141079
GlutaMAX supplement (100x)	Thermo Fisher Scientific	35-050-061
Penicillin-streptomycin 100x (10,000 U/mL)	Thermo Fisher Scientific	15140–122
Paraformaldehyde 16%	Sigma-Aldrich	28906
DPBS, no calcium, no magnesium (1x)	Thermo Fisher Scientific	14190–250

Blotting Grade Blocker Non Fat Dry Milk	Bio Rad	1706404XTU
Immobilon® -P PVDF Membrane	Millipore	IPVH00010
pSpCas9(BB)-2A-Puro (PX459) V2.0	Addgene	62988
Software and algorithms		
DESeq2 v1.38.3	Love Lab	https://github.com/thelovelab
clusterProfiler v4.6.2	Guangchuang Yu	https://yulab-smu.top/biomedical-knowledge-mining-book/
STAR v2.7.10b	Alex Dobin	https://github.com/alexdobin/STAR
R v4.2.1	The R project for Statistical computing	https://www.r-project.org/
Cell Ranger v1.1.0	10x Genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/
Seurat v5.0.3	Satija Lab	https://satijalab.org/seurat/
Subread		https://subread.sourceforge.net/
Velocyto v0.17.17		https://velocyto.org/
Scvelo v0.3.4	Theis Lab	https://scvelo.readthedocs.io/en/stable/index.html#
Jypyter Lab v4.4.6	Project Jupyter	https://jupyter.org/

Pheatmap v1.0.12	Raivo Kolde	https://cran.rstudio.com/web/packages/pheatmap/index.html
Corel DRAW version 20	Corel DRAW	www.coreldraw.com
Zeiss Blue lite	Zeiss	https://www.zeiss.com/microscopy/us/products/microscope-software/zen-lite.html
Prism 9	Graphpad	https://www.graphpad.com/scientific-software/prism/
Image J	National Institute of Health	https://imagej.nih.gov/ij/
ApE v.3.1.3	M. Wayne Davis	https://jorgensen.biology.utah.edu/wayned/ape/
Experimental model: Cell lines		
MM13 mouse embryonic stem cell line	Tom Jessell Lab, Columbia University	N/A
Mouse Embryonic Fibroblasts (MEFs)	Thermo Fisher Scientific	A34181
Oligonucleotides – mouse PCR primers	**Forward**	**Reverse**
Smad1	AGGTTTCCCCCAATCTGTTC	TGAGGACTGTCAGCACCAAG
Smad5	GACCAGGTATTTCGCAGCCT	GGCCTCCCTCTCAACTCCTA
Oligonucleotides – gRNAs	**Forward**	**Reverse Complement**
Smad5-sgRNA-F1	CACCgaactatttggttccgaaga	AAACtcttcggaaccaaatagttc
Smad5-sgRNA-B2	CACCgaactaggcttattagcgaaa	AAACtttcgctaataagcctagttc

Smad1-sgRNA-F1	CACCgCCACTCTGCTCGACT GTATT	AAACaatacagtcgagcagagtggc
Smad1-sgRNA-B1	CACCgTACAGGTGCTTGTTA CAGCG	AAACCGCTGTAACAAGCACCT GTAc

Figures

Figure 3-1. Timeline of Smad1/5 signaling in the mouse dI directed differentiation protocol

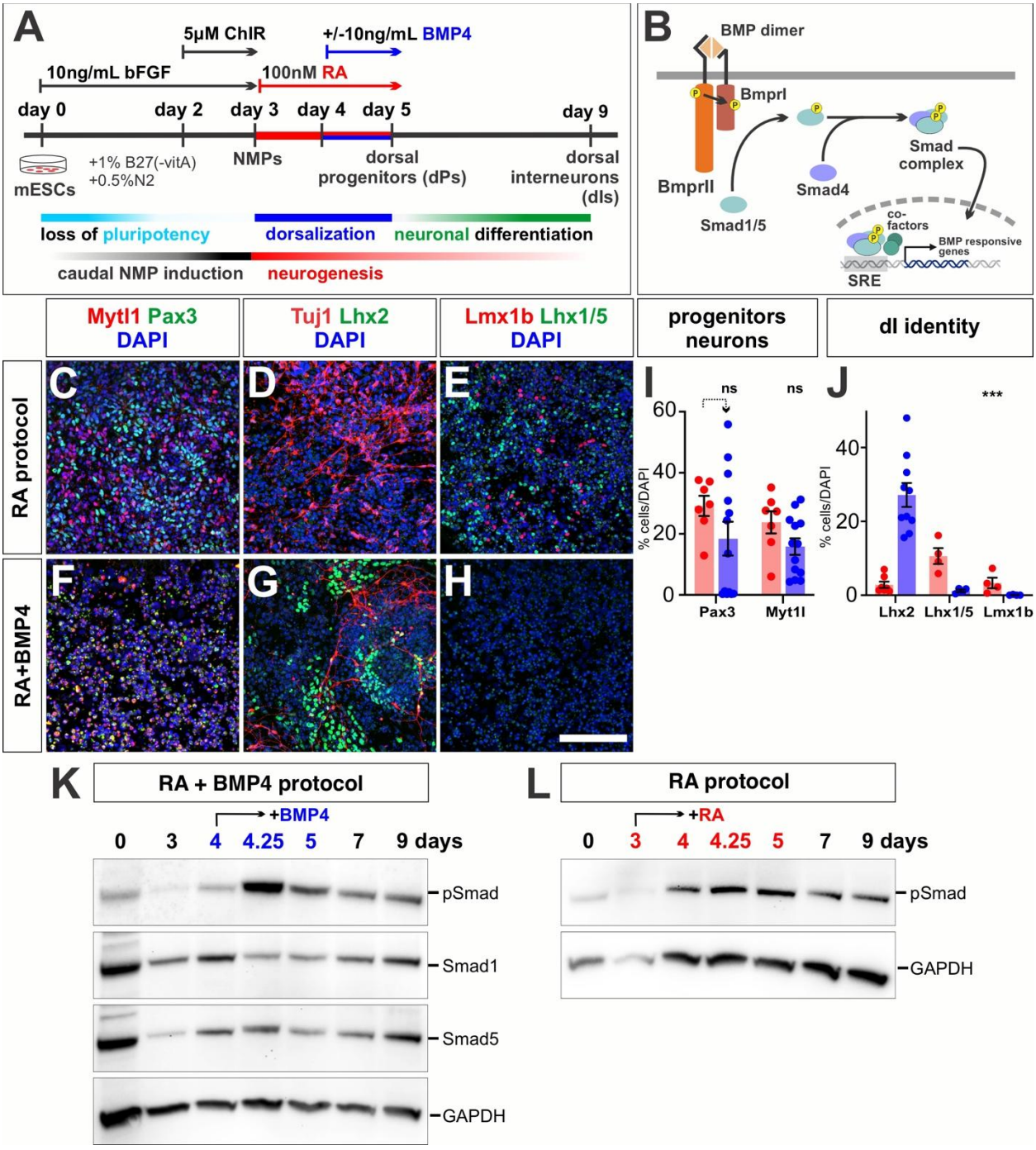

Figure 3-1. Timeline of Smad1/5 signaling in the mouse dI directed differentiation protocol

(A) Schematic timeline of the 9-day RA±BMP4 *in vitro* directed differentiation protocols to derive dIs from mESCs. mESCs are first induced to lose pluripotency and enter the neuromesodermal progenitor (NMP) state. The addition of RA from day 3-5 directs NMPs towards a dorsal progenitor (dP) identity, which resolves into dI4s, dI5s and dI6s. Adding BMP4 concomitantly with RA from day 4-5, directs dPs towards dI1s, dI2s and dI3s.

(B) Schematic of the canonical BMP signaling pathway. Upon BMP receptor binding, Smad1/5 are phosphorylated (p) and activated to serve as the key downstream mediators of BMP signaling in dorsal spinal cord development.

(C-J) Day 9 *in vitro* cell cultures are effectively neutralized, containing both $Pax3^{+}$ dPs and $Myt1l^{+}$ post mitotic neurons (C, F, and I; RA protocol n=7, RA + BMP4 protocol n=13 images acquired across two technical replicates, 1 differentiation). As previously demonstrated [140, 147] The RA protocol induces the expression of $Lhx1/5^{+}$ dI4s and $Lmx1b^{+}$ dI5s (E, H and J; RA protocol n=7 (Lhx2) and n=4 (Lhx1/5, Lmx1b), RA + BMP4 protocol n=10 (Lhx2) and n=4 (Lhx1/5, Lmx1b) images acquired across two technical replicates, 1 differentiation), while the RA+BMP4 protocol induces the expression of $Lhx2^{+}$ dI1s (D, G and J).

(K,L) Western analyses assessed both Smad1 and Smad5 levels, and BMP signaling activity (p-Smad) across the timeline of the RA±BMP4 protocols. Smad1 and Smad5 are ubiquitously expressed throughout the RA + BMP4 protocol (K). BMP signaling, as measured by p-Smad immunoreactivity, is consistently activated at low levels from day 4 onward in the RA protocol (L), but is markedly upregulated within 6 hours of recombinant BMP4 treatment in the RA+BMP4 protocol (K).

Probability of similarity between control and experimental groups: *** $p<0.0005$; two-way ANOVA

Scale bar: 100μm

Figure 3-2. Generation of Smad1 and Smad5 knockout mESC lines

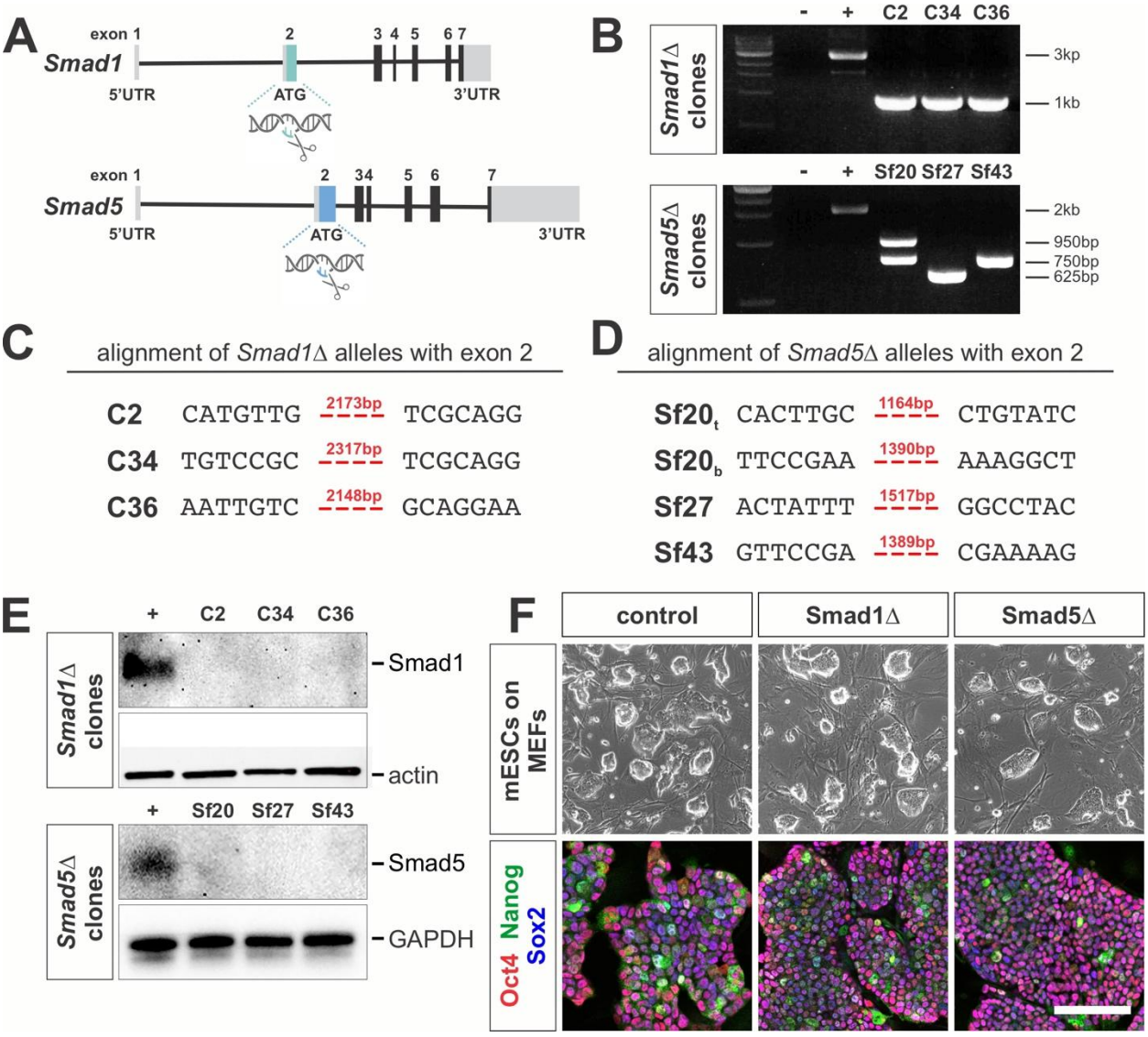

Figure 3-2. Generation of Smad1 and Smad5 knockout mESC lines

(A) CRISPR-Cas9 double nicking gene editing strategy. Guide RNAs flanking exon 2 were used to delete (Δ) the start codon in either *Smad1* or *Smad5* transcript, to thereby disrupt protein translation.

(B) Agarose gel electrophoresis of PCR amplicons spanning the targeted region in three independent Smad1Δ lines (C2, C34, C36) and three independent Smad5Δ lines (Sf20, Sf27, Sf43), demonstrated genomic deletions at the exon 2 locus.

(C) Exon 2 deletion was confirmed by sequencing and aligning the PCR amplicons to the GRCm38 *Mus musculus* genome, showing that the deletion breakpoints correspond precisely to the gRNA-targeted cut sites.

(E) Western analyses validated that all Smad1Δ or Smad5Δ lines contain null alleles.

(F) Phase contrast imaging shows that Smad1Δ and Smad5Δ mESCs are morphologically comparable to the isogenic control line (top, F) and continue to express pluripotency markers Oct4, Nanog, and Sox2 (bottom, F).

Scale bar: 100μm

Figure 3-3. Smad1/5 play opposing roles in NMP fate commitment toward neural and mesodermal lineages at day 3

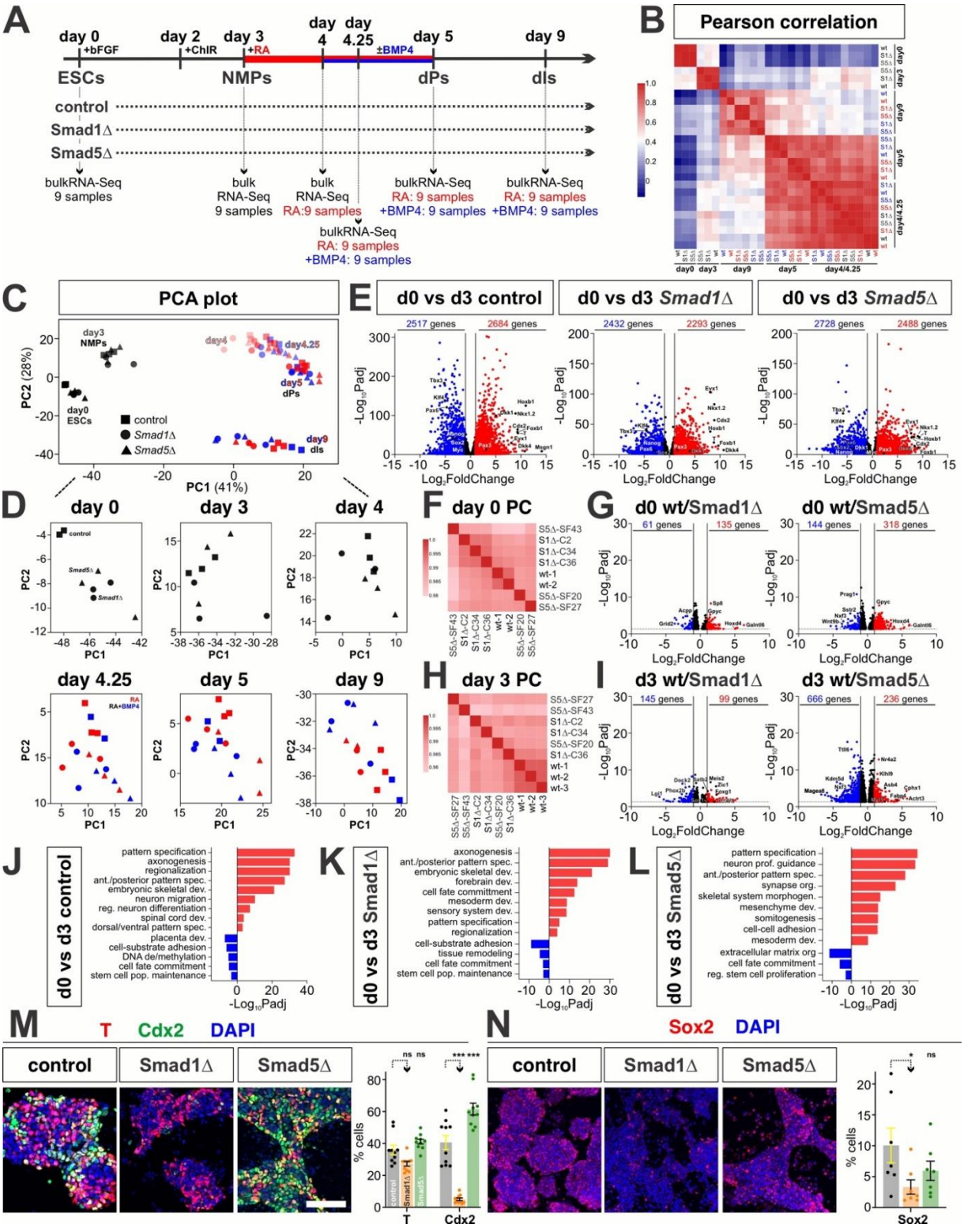

Figure 3-3. Smad1/5 play opposing roles in NMP fate commitment toward neural and mesodermal lineages at day 3

(A) Timeline for bulk RNA-sequencing experiment (81 samples total). RNA was collected at the indicated developmental time points along the protocol. Three biological replicates were used for each genotype, i.e., from the three independently generated *Smad1*Δ and *Smad5*Δ mESC lines.

(B) Pearson correlation of averaged biological replicates at each time point and condition. Samples clustered primarily by developmental stage, with a block of high correlation from 6–24 hours after BMP4 addition, indicating highly similar transcriptional states during this time window.

(C, D) Principal component analysis (PCA, C) revealed a shared trajectory as mESCs (day 0) transition towards dorsal interneurons (dIs, day 9). The factor loadings (D) show that Smad1Δ and Smad5Δ samples progressively deviate from the control lines over time, consistent with transcriptional disruptions accumulating as differentiation proceeds.

(E) Volcano plots comparing day 0 to day 3 samples show consistent upregulation of NMP-associated genes (*T*, *Cdx2*, *Nkx1.2*) and downregulation of pluripotency genes (*Tbx3*, *Klf4*, *Nanog*) across control, Smad1Δ, and Smad5Δ lines, suggesting all lines successfully acquire NMP fate.

(F-I) Pearson correlations (F, G) and volcano plots (H, I) comparing Smad1Δ and Smad5Δ lines to the control at days 0 (F, G) and 3 (H, I). All lines remain highly transcriptionally correlated at both time points; however, differential expression analysis revealed that Smad1Δ cells show increased expression of neural-associated genes (*Zic1*, *Sox1*, *Foxg1*), whereas Smad5Δ cells upregulate mesoderm-associated markers (*Myf5*, *Asb4*, *Fabp4*) relative to controls.

(J–L) Gene Ontology (GO) analyses of the significantly up- and down-regulated genes at day 3 (relative to day 0) revealed enrichment for neural and mesodermal biological processes.

(M–N) As validation, cultures were immunostained for Brachyury (T, red), Cdx2 (green), Sox2 (red) and DAPI (blue). Decreased numbers of Smad1Δ cells express Cdx2 (M, n=10 images acquired from two technical replicates, 1 differentiation) and Sox2 (N, n=7 images acquired from two technical replicates, 1 differentiation). In contrast, increased number of Smad5Δ cells express T, suggesting a mesodermal identity.

Scale bar: 100μm

Significance threshold: $\log_2$ fold-change $-1 >$ or $1 <$ and adjusted p-value < 0.05. Probability of similarity between control and experimental groups: * $p<0.05$, *** $p<0.0005$; two-way ANOVA

Figure 3-4. ATAC-Seq analyses from days 4- 5 reveal divergent epigenomic and transcriptional roles for Smad1 and Smad5 downstream of BMP4 treatment

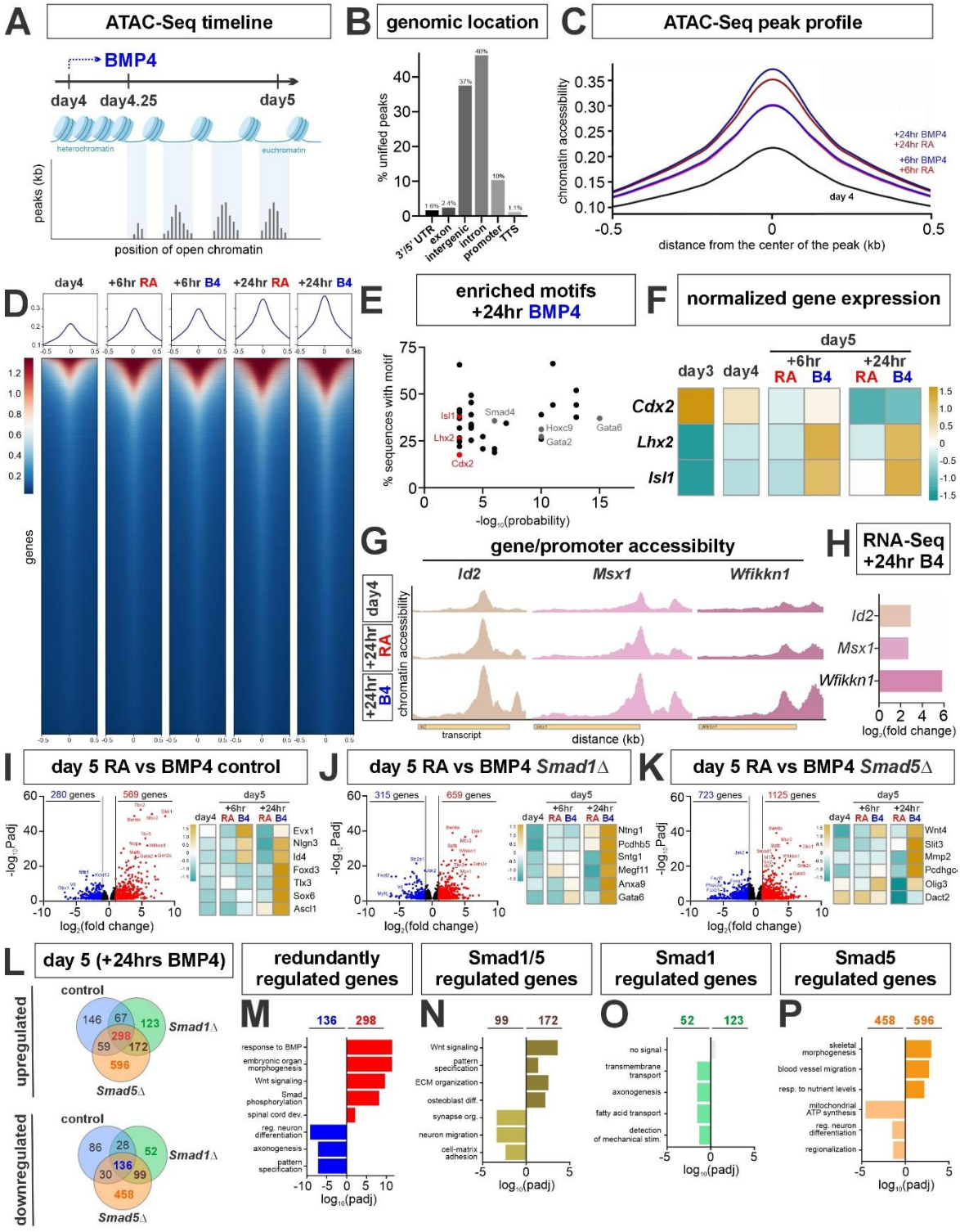

Figure 3-4. ATAC-Seq analyses from days 4- 5 reveal divergent epigenomic and transcriptional roles for Smad1 and Smad5 downstream of BMP4 treatment

(A) Schematic timeline of ATAC-seq studies to assess the chromatin accessibility changes before (day 4) and after (day 4.25 and 5) BMP4 treatment. Three biological replicates were used for each time point (control line only; n=9 samples).

(B) ATAC-seq peaks across samples were merged to form a unified peak set (n = 141,592). The majority of peaks fall in intronic (46%) and intergenic (37%) regions and together account for ~83% of all peaks.

(C, D) Accessibility profiles of differentially accessible regions showed progressive increases in chromatin accessibility from day 4 to day 5. While there were no significant differences between BMP4-treated and control samples at 6 hours, clear accessibility changes emerge by 24 hours.

(E) Motif enrichment analysis identified 37 transcription factor motifs significantly enriched in regions with increased chromatin accessibility following 24 hours of BMP4 treatment.

(F) Heatmap examining the expression of key transcription factors identified in (E), shows BMP4 treatment rapidly upregulates *Lhx2* (dI1), and *Isl1* (dI3) expression. However, *Cdx2* (mesoderm) is progressively downregulated, even though Cdx2 motifs remain accessible 24 hours after BMP4 treatment. Thus, BMP4 treatment may selectively stabilize pre-existing accessible regions.

(G, H) Similarly, by 24 hours after BMP4 addition, chromatin accessibility is increased at the promotors of established BMP-regulated genes, such as *Id2* and *Msx1*, with concomitant transcriptional upregulation. This analysis also identified novel regulators of dorsal fate specification, such as *Wfikkn1,* a secreted BMP antagonist whose upregulation may represent a negative feedback mechanism to dampen BMP signaling.

(I–K) Building on the localized chromatin remodeling observed 24 hours after BMP4 treatment, differential gene expression was examined in the control, Smad1Δ, and Smad5Δ bulk RNA-Seq datasets at day 5. Volcano plots show that the most upregulated genes across the lines are

shared canonical BMP target genes. Heatmaps further identify genes uniquely regulated in each line, revealing both Smad-specific and shared transcriptional responses to BMP4.

(L) Venn diagram showing the overlap of genes up or downregulated after 24 hours of BMP4 treatment between Smad1Δ, Smad5Δ and control cells. Loss of Smad5 consistently affects a larger subset of genes.

(M-P) GO analyses of subsets of genes taken from the Venn diagram (L). The 298 genes redundantly upregulated by Smad1 and Smad5 show enrichment for BMP signaling pathway terms (M). Shared transcriptional changes in Smad1Δ and Smad5Δ cells (172 upregulated, 99 downregulated genes) are associated with increased extracellular matrix organization and decreased neuron migration and cell adhesion (N). Strikingly, there are relatively few transcriptional changes unique to the loss of Smad1 (O), whereas Smad5Δ leads to more extensive transcriptional remodeling associated with non-canonical BMP functions, such as altered regulation of bioenergetics (P).

Figure 3-5. The loss of Smad5 alters lineage specification during dI *in vitro* differentiation

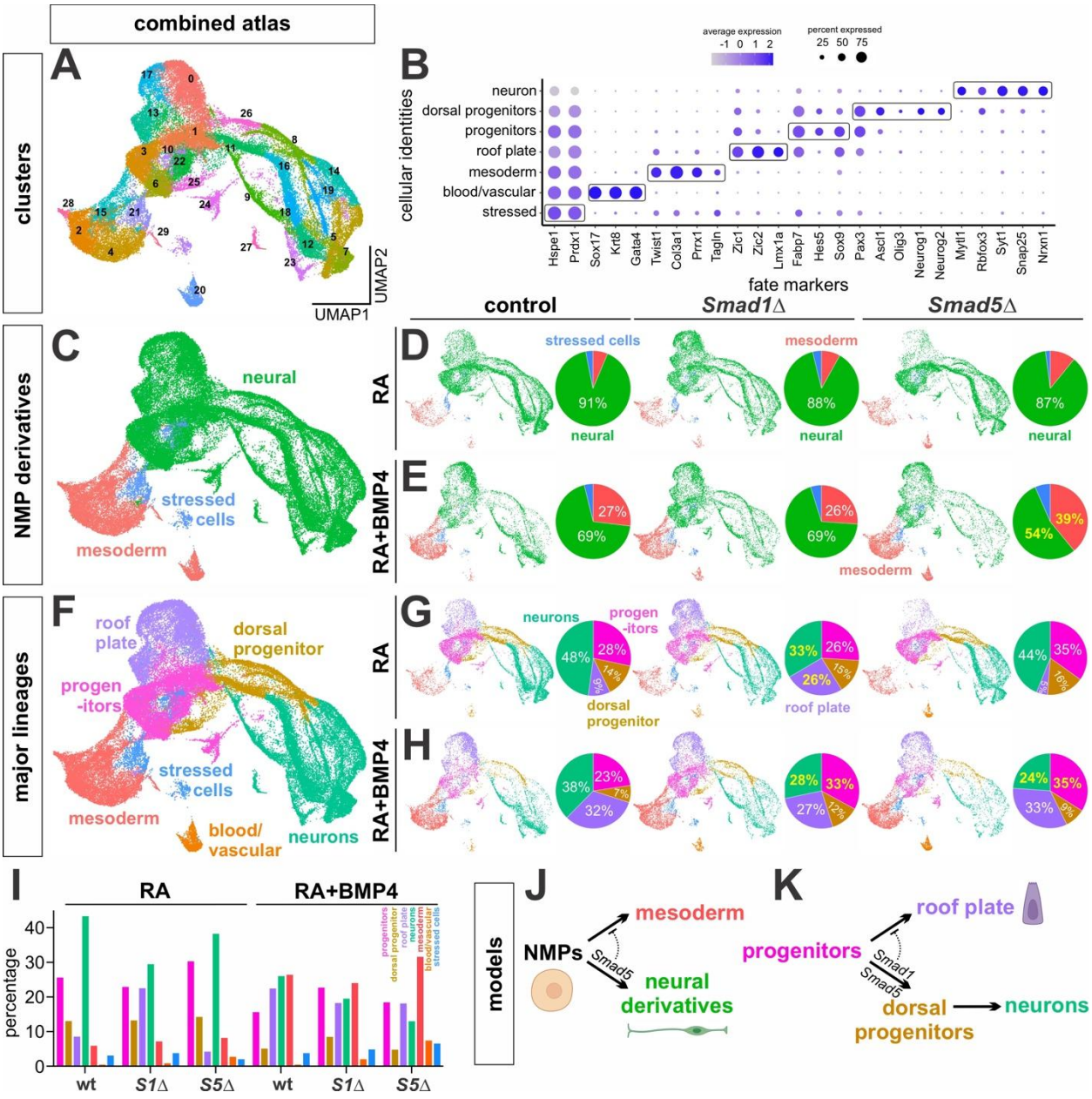

Figure 3-5. The loss of Smad5 alters lineage specification during dI *in vitro* differentiation

(A) Integrated UMAP atlas of scRNA-Seq datasets from RA±BMP4 directed differentiations of the Smad1 (C36 line), Smad5 (SF43 line) and isogenic control lines at day 9 (n=6 datasets). Clustering of the data yields 30 transcriptionally distinct cell clusters.

(B) Dot plot showing the expression levels and identity of marker genes used to define the different lineage identities derived from NMPs.

(C) Integrated UMAP atlas depicting the mesodermal (coral) and neural (green) lineage trajectories. Stressed cells are labeled in blue.

(D, E) Pie charts and accompanying sample-specific UMAP atlases quantifying the proportion of neural and mesodermal cell derivatives across sample groups. There are no major cell fate specification differences in the RA protocol, however in the RA+BMP4 protocol, the mesodermal population is expanded specifically in the Smad5Δ line.

(F) Integrated UMAP atlas identifying the six major lineage groups, and the stressed cell population.

(G, H) Pie charts and accompanying sample-specific UMAP atlases quantifying the proportions of progenitor, roof plate, dorsal progenitor, and neural cell populations. In the RA protocol, the roof plate population is expanded at the expense of the neural population in the Smad1Δ line. In the RA+BMP4 protocol, both the Smad1Δ and Smad5Δ lines have the same phenotype: the progenitor population is increased, and the number of neurons concomitantly decreased.

(I) Bar chart summarizing the percentage of the major cell lineages differentiated from control, Smad1Δ and Smad5Δ line using the RA ± BMP4 protocols.

(J, K) Models summarizing the roles of Smad1 and Smad5 directing the specification of NMP derivatives. Smad5 plays the earliest role in lineage specification, preventing NMPs from entering a mesodermal trajectory (J). As differentiation proceeds, both Smad1 and 5 are required for progenitors to progress towards a neural fate (K); Smad1 additionally prevents progenitor cells from being directed towards a roof plate fate.

Figure 3-6. Smad1 and Smad5 are differentially required to specify dI cell fates

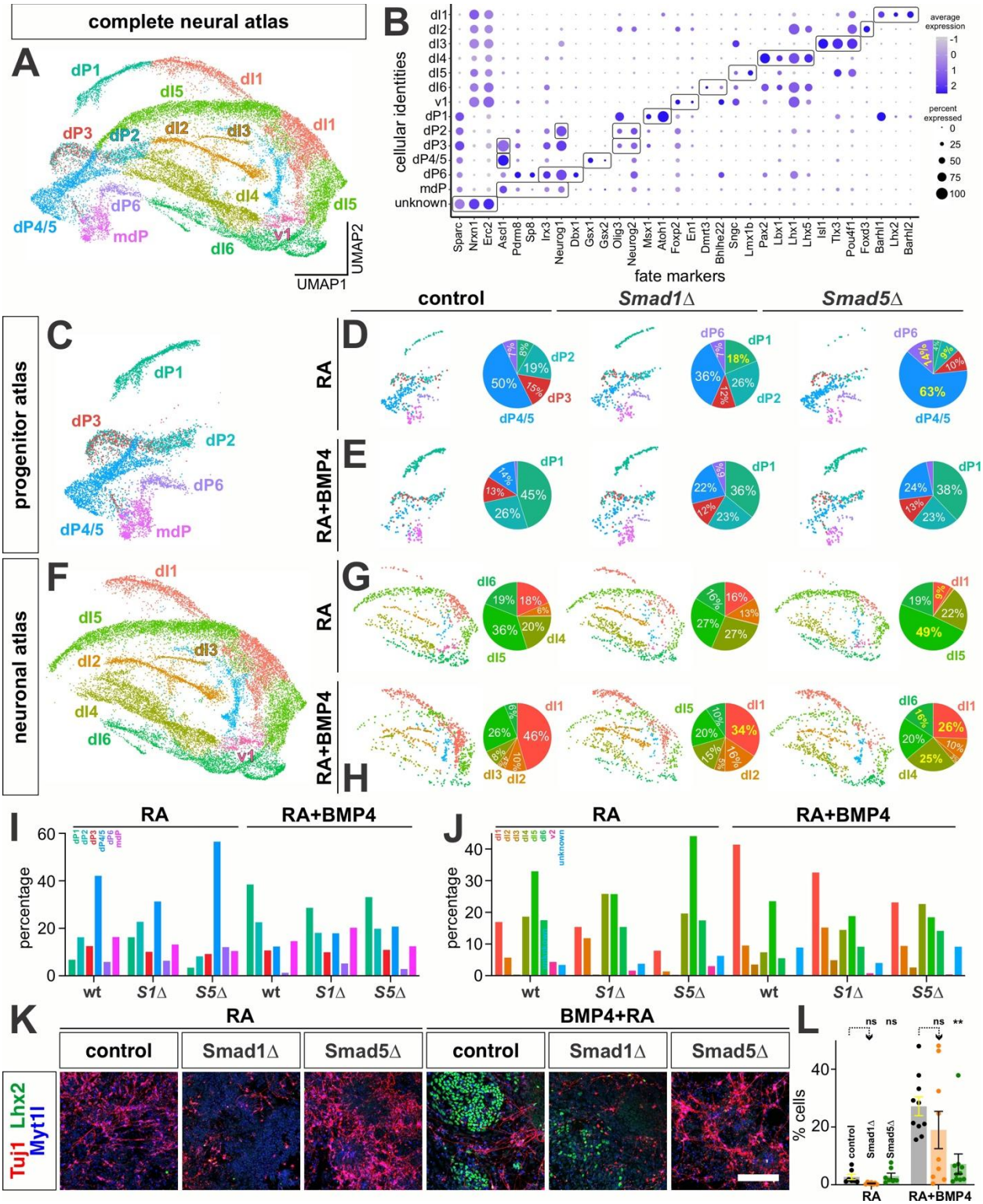

Figure 3-6. Smad1 and Smad5 are differentially required to specify dI cell fates

(A) Cells annotated as dorsal progenitors and neurons were re-clustered to generate an integrated neural UMAP atlas, capturing the dP1–dP6 and dI1–dI6 trajectories. Based on RNA velocity analyses (Figure 7), a population located at the periphery of the atlas, that appears to give rise to dP2–dP6, was annotated as "mdP" (multipotent dorsal progenitor).

(B) Dot plots showing the expression levels and identities of marker genes used to assign neural identities

(C-H) UMAPs of the integrated dP (C) and dI (F) atlases, and sample-specific UMAP atlases for the control, Smad1Δ and Smad5Δ lines in the RA ± BMP4 protocols (D, E, G, H). The RA protocol robustly generates dP/dIs from the intermediate spinal cord (dI4-dI6), with more modest numbers of dorsal-most dP/dIs, i.e. dI1-dI3. Under these conditions, the Smad5Δ line shows impaired formation of dP1s and dP2s (D), with correspondingly reduced numbers of dI1s and dI2s (G). In contrast, the Smad1Δ line shows increased numbers of dP1/dP2s (D), but no corresponding change in the number of dI1/dI2s (G). In the RA+BMP4 protocol, both the Smad1Δ and Smad5Δ lines show decreased numbers of dP1/dP2s (E) and dI1/dI2s (H).

(I, J) Bar charts summarizing the percentage of dP (I) and dI (J) found in control, Smad1Δ and Smad5Δ cell populations using the RA ± BMP4 protocols.

(K, L) As validation, cultures were immunostained for Tuj1 (axons, red), Lhx2 (dI1, green), and Myt1l (neurons, blue). As previously observed [44, 140], significant numbers of dI1s are only generated in the RA+BMP4 protocol (L). Supporting the transcriptional analysis, $Lhx2^{+}$ dI1s are most reduced in the Smad5Δ differentiation (n=10 images acquired from 2 technical replicates, 1 differentiation), compared to either the Smad1Δ (n=9 images acquired from 2 technical replicates, 1 differentiation) or control (n=10 images acquired from 2 technical replicates, 1 differentiation) differentiation.

Scale bar: 100μm

Probability of similarity between control and experimental groups: ** $p<0.005$; two-way ANOVA

Figure 3-7 Smad5 is required for normal transcriptional dynamics across all dl lineages

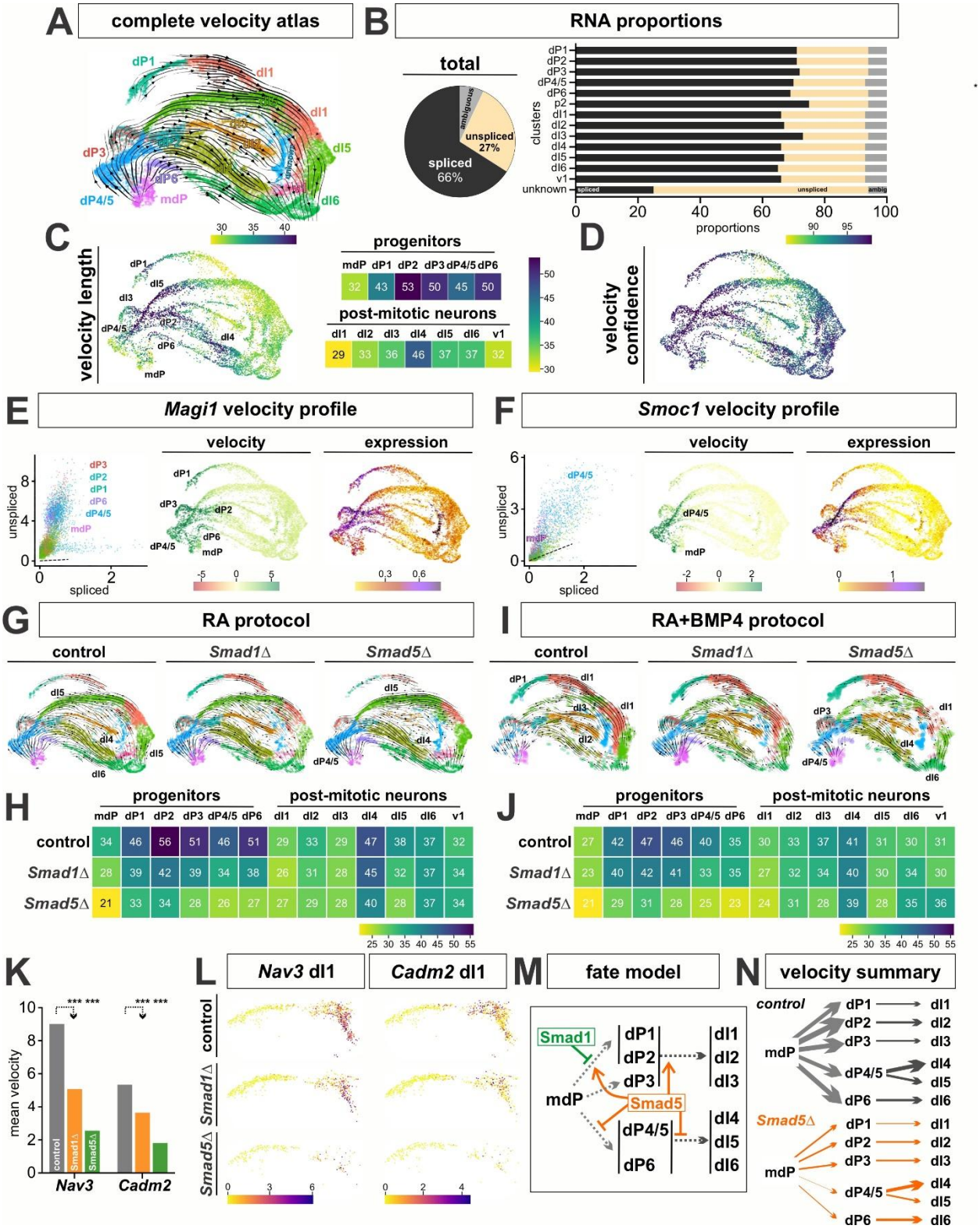

Figure 3-7 Smad5 is required for normal transcriptional dynamics across all dI lineages

(A) RNA velocity analysis of the integrated neural atlas was performed to infer the direction and dynamics of dP cell state transitions. The streamlined velocity vector field shows that each dP/dI subtype follows a distinct developmental trajectory. The dI2-dI6 subtypes stem from a common mdP population, which bifurcates into specific dP subtypes that resolve into their respective dI populations.

(B) RNA velocity metrics showing the proportion of spliced vs unspliced transcripts in the complete atlas as well as the individual clusters.

(C, D) Velocity magnitude (C) and velocity confidence (D) values from the control RA±BMP4 dataset projected onto the UMAP embedding show generally increased RNA velocity in the dP subtypes relative to the dI populations. A notable exception is the dI4 subtype, which maintains high RNA velocity along its entire trajectory.

(E-F) Phase portraits of two genes exhibiting differential RNA velocity in the control RA±BMP4 dataset. *Magi1* expression shows positive velocity across all dP populations (E), while *Smoc1* is specifically positive in the dP4/dP5 population.

(G-J) Streamlined representation of the velocity vector fields in control, Smad1Δ and Smad5Δ UMAP atlases in RA±BMP4 protocols (G, I). Heatmaps of the average velocity-length values demonstrates that Smad5Δ results in dramatically reduced velocity in all dP subtypes, in both RA±BMP4 protocols (H, J).

(K) Velocity magnitude per gene was quantified for the dI1 population across the control, Smad1Δ and Smad5Δ lines. *Nav3* and *Cadm2* exhibited the highest reductions in velocity, which was most pronounced for the Smad5Δ line.

(L) *Nav3* and *Cadm2* expression is progressively diminished in the Smad1Δ and Smad5Δ dI1s, suggesting the reduced velocity magnitude reflects impaired progression of $Nav3^{+}$/$Cadm2^{+}$ dI1s.

(M) Smad5 is required reiteratively across the specification of dI trajectories. It is required for the specification of the dorsal-most progenitor (dP1/dP2) and neuronal (dI1-dI3) populations, at the expense of the intermediate dI populations (dI4–dI6). Smad1 acts more narrowly, limiting the pool of dP1s.

(N) RNA velocity analysis shows that dP clusters have broadly increased velocity compared to dIs, with the exception of dI4 lineage. We propose Smad5 is the principal mediator of the transcriptional programs that maintain progenitor activity and regulate their progression into differentiated dI states.

Chapter 4 – *In vivo* human embryonic spinal cord atlas validates stem cell–derived human dorsal interneurons and reveals ASD spinal signatures

Abstract: Spinal cord injuries (SCI) result in the loss of motor and sensory function. We are working towards restoring sensation by developing directed differentiation protocols to generate dorsal spinal interneurons (dIs; dI1-dI6) from human embryonic stem cells (hESCs). Here, we present an improved method that produces human dIs via a neuromesodermal progenitor state, the physiological intermediate for spinal cord development. We show that retinoic acid (RA), bone morphogenetic protein 4 (BMP4) and growth differentiation factor (GDF) 11 direct dI identity, while GDF11 and extended time in culture promotes posterior spinal identities. Together, these protocols generate the full complement of dorsal subtypes along the entire anterior-posterior axis of the spinal cord. To benchmark *in vitro*-derived dIs, we constructed a single-cell RNA-Seq atlas of the human embryonic spinal cord and used it to show that hESC-derived dIs closely match their endogenous counterparts. The atlas also reveals that the dI4/dI5 populations dramatically expand in comparison with the other spinal lineages. Moreover, they have mechanosensory circuit signatures linked to autism spectrum disorder, implicating spinal circuits in autistic phenotypes.

Introduction: Spinal cord injury (SCI) is a debilitating condition lacking curative therapies[170]. Although loss of motor function is the most widely recognized consequence of SCI, disruption of somatosensory perception - including pain, touch, and proprioception - is equally detrimental for patients' quality of life[28, 171]. Compared to the recovery of motor function, recovery of sensory function is variable, and many individuals with severe injuries experience persistent deficits in sensory perception[172]. Thus, the restoration of somatosensory circuitry represents a critical but understudied challenge in regenerative neuroscience.

Cell replacement strategies offer a promising approach to repair damaged spinal circuits[173, 174]. Advances in stem cell biology now permit the generation of neuronal subtypes capable of integrating into host tissue[175-177]. However, successful circuit reconstruction in the spinal cord is

likely to require precise matching of neuronal identity to anatomical position along the anterior–posterior axis. As the spinal cord is organized into segmental domains that connect distinct peripheral organs and tissues[35], restoring function requires generating neuronal subtypes with matching regional and sensory-modality specific identities. For example, restoration of bladder and bowel function depends on reconstruction of lumbosacral circuits that regulate autonomic control[178]. Achieving such specificity requires first defining the differentiation strategies that reproduce developmental programs governing spinal cord regionalization.

The signals that determine regional identity have been well characterized in the developing spinal cord[179, 180]. Anterior/posterior identity is established by the expression of *HOX* transcription factors[180-182]. *HOX* genes are organized into four chromosomal clusters (*HOXA*, *HOXB*, *HOXC,* and *HOXD)* and activated in a temporally progressive manner[183-185]: anterior identities are specified by 3' HOX paralogs while more posterior identities emerge through later expression of the 5' HOX genes[185]. In parallel, signaling centers within the neural tube establish dorsal-ventral patterning by secreting growth factors[45]. In the dorsal spinal cord, six dorsal progenitor (dP) populations generate six classes of dorsal interneurons (dI1-dI6) that process somatosensory inputs including touch, pain, temperature, and proprioception[22]. Ventral progenitor domains produce motor neurons (MNs) and ventral interneurons that coordinate locomotor output[186, 187]. While these patterning mechanisms broadly govern sensorimotor circuit assembly, it remains unresolved how temporal and spatial programs interact to generate the cellular diversity of the dIs.

Directed differentiation protocols have successfully generated spinal MNs from pluripotent stem cells that achieved functional integration in experimental SCI models[34, 188, 189]. However, the restoration of somatosensory circuitry will also require the robust generation of dIs spanning multiple axial levels. Early differentiation protocols recapitulated aspects of spinal patterning by inducing anterior neuroepithelium followed by posteriorization toward spinal identities[42, 43, 190, 191].

While these approaches generate MNs and a subset of dI populations with brachial-thoracic identities, they do not efficiently produce neurons corresponding to the most posterior spinal regions[42, 190]. *In vivo*, much of the posterior spinal cord arises from bipotential neuromesodermal progenitors (NMPs)[192], which give rise to the lumbar-sacral circuits essential for bladder and bowel control[192, 193]. Importantly, NMP-based differentiation strategies show improved outcomes; a recent mouse embryonic stem cell (ESC) protocol that derived dIs through an NMP intermediate produced the full complement of dIs that transcriptionally and functionally resemble their endogenous counterparts[140, 194]. However, comparable approaches for generating human dIs along the full anterior-posterior axis have yet to be developed.

A major barrier to evaluating stem cell-derived spinal neurons is the lack of a comprehensive human developmental reference atlas. Directed differentiation protocols typically produce heterogenous neuronal populations, and the validation of cellular identity requires comparison to endogenous developmental trajectories[42]. Advances in single-cell transcriptomics and sophisticated computational methods have enabled the integration of large-scale datasets to reconstruct developmental lineages across tissues[195, 196]. While such approaches have been applied broadly in other organ systems, the application to human embryonic spinal cord development has been more limited. Thus, the existing datasets for the human embryonic spinal cord provide an incomplete view of neuronal diversity and lineage relationships, limiting the ability to benchmark *in vitro*-derived cell types or identify previously unrecognized neuronal subclasses. A comprehensive reference atlas of the developing human spinal cord would therefore provide an important tool for defining neuronal identity, uncovering human-specific mechanisms of neural diversity, and guiding the rational design of differentiation strategies for regenerative applications. Here, we integrated six publicly available single-cell datasets[185, 197-201] to generate a transcriptomic reference atlas of the developing human spinal cord spanning gestational weeks 4-25. The atlas reveals extensive diversification of dI populations, including a marked expansion in the dI4 and

dI5 subclasses associated with somatosensory processing of pain, touch, and itch[35]. Using this resource as a benchmark, we developed a human NMP-based differentiation protocol designed to generate dIs across the anterior-posterior axis. By modulating NMP culture duration and exposure to the posteriorizing factor GDF11[202-204], we produced interneuron populations that align with endogenous developmental trajectories and exhibit molecular signatures consistent with somatosensory circuit specialization. Comparative analysis of *in vivo* and *in vitro* dI4/dI5 subclasses reveal conserved sensory-associated gene networks, including the early emergence of gene networks associated with autism spectrum disorder (ASD) within the mechanosensory lineages. Taken together, these findings establish a developmental framework that links temporal and axial patterning to the organization of human somatosensory circuitry and establish approaches for generating regionally specified spinal interneurons for disease modeling and regenerative strategies.

Results

A comprehensive fetal human spinal cord atlas resolves the developmental trajectories underlying neuronal diversification.

To establish a reference dataset for human spinal cord development, we integrated transcriptomic datasets spanning fetal development to generate a harmonized single-cell atlas of the human spinal cord (Figure 1B). We compiled 192 publicly available single cell (sc) -and single-nucleus (sn) RNA-Seq samples from six studies[185, 197-201], yielding an integrated dataset of approximately 1.7 million cells spanning gestational weeks 4-25 (Figure S1A-S1C). Datasets were preprocessed using Seurat[205] and integrated using the deep generative modeling program SCVI[206] to enable robust cross-study alignment while preserving biological structure (Figure 1A) (see Materials and Methods for a complete description of the computational pipeline). The atlas can be viewed online here: https://cells.ucsc.edu/?ds=spinal-cord-asd. (Figure S1B).

Clustering analysis using the sctype package[207] identified 71 transcriptionally distinct populations that were assigned to 18 major spinal cord cell types using curated marker gene sets and combinatorial classification approaches (Figure 1C, 1D). These cell classes included neural progenitors, neurons, glia, and mesoderm-derived populations. Canonical lineage markers segregated expected developmental populations with minimal overlap between unrelated cell types (Figure 1E) and the relative proportion of major cell classes across developmental time recapitulated the expected ontogenetic progression[208, 209], including early enrichment of progenitor populations followed by sequential emergence of neuronal, astroglial, and oligodendroglial lineages (Figure 1F). The resulting atlas provides a reference framework for examining lineage diversification and regional specification within the developing human spinal cord.

The neuronal atlas reveals the temporal expansion and diversification of dI4 and dI5 sensory interneuron populations

To investigate how neural diversity emerges, we isolated spinal neuronal lineages and reconstructed lineage relationships across developmental time. Integration and reclustering of the neuronal cells generated a high-resolution neuronal atlas containing 308,947 cells, including 56,674 progenitors (Figure 2A). Using AUCell-based module scoring[210] and curated gene sets (Figure 2B, Tables S1-S2), we assigned all cells dorsal or ventral neuronal identities corresponding to the canonical spinal neuron classes: dorsal interneurons (dI1-dI6), ventral interneurons (v0-v3), and MNs (Figure 2A).

Trajectory reconstruction identified twelve major spinal neuronal lineages, each characterized by distinct transcriptional programs and lineage relationships (Figure 2A-2C; Figure S3A). The dI4 and dI5 populations emerged as the predominant neuronal classes with the dorsal spinal cord. Quantitative analysis revealed that dI4 neurons comprised approximately 40% of neurons (~50% excluding progenitors), while dI5 neurons accounted for approximately 22% (~30% excluding progenitors) of the neuronal compartment (Figure 2A). These populations arise alongside other dorsal lineages at approximately gestational week 5 but undergo temporally-restricted expansion between weeks 7-11 (Figure S2), suggesting preferential expansion of these interneuron populations during human somatosensory circuit development.

Subclustering analysis revealed substantial heterogeneity within the dI4/dI5 population, identifying 33 transcriptionally distinct clusters distributed along continuous developmental trajectories (Figure 2C). Comparative analysis with a recently described mouse embryonic spinal cord atlas[211] revealed alignment between human dI4/dI5 populations and mouse late-born dILA/dILB populations[212], supporting the conservation of developmental logic across species (Figure 2D). Of the 33 human dI4/dI5 clusters, approximately 24 corresponded to mouse counterparts, while nine clusters lacked clear mouse counterparts, suggesting the existence of human-specific subtypes. These findings indicate that the developing human spinal cord exhibits both conserved and species-specific diversification of dorsal sensory interneurons.

dI4 and dI5 subclasses are defined by modality-specific transcriptional signatures

We next examined whether the transcriptional heterogeneity within dI4/dI5 populations corresponded to predicted functional specialization. Differential gene expression analysis across dI4/dI5 subclusters followed by gene ontology (GO) enrichment revealed molecular programs associated with multiple sensory modalities, including nociception, mechanosensation, temperature perception, proprioception, and neuromodulatory responses (Figure 2E, 2F).

Gene ontology analysis further revealed both shared and subtype-specific sensory programs within dI4 and dI5 populations. Shared functions included pain perception, detection of mechanical stimulus (i.e. touch), and regulation of movement, consistent with established roles of dorsal interneurons in somatosensory processing. Several dI4 subclasses exhibited transcriptional signatures associated with temperature sensation (cluster 5), startle response and sensory perception of sound (cluster 4 and 26), suggesting diversification of modality-specific sensory interneuron programs during human spinal cord development. Notably, two dI4 subtypes (clusters 11 and 36) exhibited enrichment in gene programs associated with response to cocaine, supporting our previous observation that embryonic dI4/dI5s may be responsive to psychoactive compounds[140]. These findings suggest that molecular programs associated with neuromodulatory responsiveness are established early during spinal cord development and may represent intrinsic properties of emerging sensory-processing interneuron circuits.

To further resolve functional specialization among dI4/dI5 subclasses, we curated marker gene sets associated with neurotransmission and sensory signaling (Figure 2I). Several clusters expressed genes with established roles in sensory circuit function. For example, subsets of dI4 populations expressed *SSTR2* (cluster 5, overlapping with mouse dILA5), which modulates neurotransmission[213] and itch signaling[214], while *NPY*-expressing populations (cluster 12, overlapping with mouse dILA1/2), encode a peptide neurotransmitter known to inhibit pain

perception[215]. Additional subclasses expressed the mechanosensitive ion channel PIEZO2 (cluster 15, overlapping with mouse dILB2) which mediates light touch, proprioception, and mechanical pain[216]. Taken together, these transcriptional signatures indicate progressive specialization of dI4/dI5 interneuron subclasses, including populations associated with neuromodulatory and pharmacologically responsive signaling pathways.

Human neuromesodermal progenitors (NMPs) exhibit temporally regulated acquisition of posterior axial identity

Given our previous studies with mouse ESCs[140], we next investigated how differences in developmental timing influences the types of dIs generated from human pluripotent stem cells could, and whether the strategy generates dIs spanning major axial levels of the spinal cord. Thus, we adapted our NMP-based protocol for human ESCs to examine the effect of extended culture duration and signaling factors such as GDF11 [202, 203, 217] on dI patterning along the anterior-posterior and dorsal-ventral axes of the spinal cord.

Human NMPs were generated by exposing the H9 ESC line to bFGF and the Wnt agonist CHIR[54, 140] (Figure 3A). To examine whether temporal progression influences axial identity *in vitro*, hESCs were exposed to bFGF/CHIR for 2, 4, or 10 days to generate NMP populations with increasingly extended developmental competence (Figure 3A, 3B). Immunocytochemical analysis demonstrated expression of canonical NMP markers SOX2, T, and CDX2 across all timepoints (Figure 3B, 3C), consistent with bipotential neuromesodermal progenitor identity.

Single-cell and bulk RNA sequencing revealed greater transcriptional divergence from pluripotent states as NMP culture duration increased (Figure 3D-3F, 4E-4O). GO analyses indicated gradual enrichment of mesoderm-associated transcriptional programs and reduction of anterior neural signatures with extended culture duration (Figure 4G-4N). Consistent with progressive posteriorization over time, HOX gene expression increased with longer NMP culture duration, with 2-day NMPs expressing anterior HOX genes such as HOXC5, while 4- and 10-day NMPs

expressed progressively more posterior HOX paralogs, including HOXA10, HOXD11, and HOXD13 (Figure 4O). These findings indicate that human NMPs can acquire posterior competence through temporal progression in culture.

Temporally-patterned human NMPs generate the full complement of dI identities across the anterior-posterior axis.

We next examined whether temporally patterned NMPs could generate the full range of dI subtypes. NMPs were aggregated into embryoid bodies (EBs) and neuralized with retinoic acid (RA) to promote spinal neural identity (Figure 3A). qPCR and IHC analyses demonstrated robust generation of intermediate dI populations (dI4-dI6) across all NMP timepoints (Figure 3G-3K, S 4E). Addition of BMP4 during neuralization promoted specification of dorsal-most interneuron populations (dI1-dI3), consistent with the known developmental signaling logic[140]. Optimization of BMP4 timing defined day 2–14 as an effective window for dorsal specification (Figures 3A, S4A-4D). Combined RA+BMP4 treatment generated dI1 and dI3 populations across NMP conditions (Figures 3G-3K, S4E), while scRNA-seq analyses further identified dI2s not readily detected using qPCR or IHC (Figures 5G, 6A). Collectively, these results demonstrate that temporally patterned human NMP cultures generate the full complement of dI classes spanning multiple axial identities and reproduce key features of *in vivo* HOX-dependent spinal cord patterning.

GDF11 reinforces posterior identity and selectively influences dI specification

We next investigated whether GDF11, a regulator of posterior HOX gene expression in the neural tube[204], contributes to axial patterning or dI specification in human NMP-derived cultures[202, 203, 217]. To identify whether GDF11 acts in a stage-specific manner, GDF11 was applied either during the NMP stage, during dorsal progenitor (dP) differentiation, or at both developmental stages (Figure 5A, B). Bulk RNA-Seq and qPCR analysis revealed that GDF11 applied during the NMP stage had minimal impact on global transcriptional identity (Figure 4E, 4K), although subtle

differences in *HOX* gene expression profiles were observed (Figure 4O). In contrast, GDF11 applied during dorsal progenitor differentiation increased the expression of posterior HOX genes including *HOXA9*, *HOXC10* and *HOXA13*, consistent with reinforcing lumbosacral identity (Figure 5C).

GDF11 treatment also modulates dI specification, in a temporally restricted manner. While GDF11 did not significantly alter generation of dorsal-most interneurons, dI1–dI3, it selectively increased expression of markers associated with the dI4 and dI5 populations, compared to the RA condition (Figure 5D-5F, S7K). GDF11 treatment at the dorsal progenitor, but not NMP, stage increased the number of LMX1B$^+$ dI5s in both IHC and scRNA-Seq analyses (Figure 5E-5G). GDF11-dP can also drive the formation of more PAX2$^+$ dI4s (Figure S7K), although this result is inconsistent across replicates (Figure 5F). GDF11-dP suppresses *DMRT3* expression, the canonical marker for dI6s, suggesting reduced numbers of dI6s population (Figure 5D, 5G). For all conditions, there is no synergistic effect of adding GDF11 at both the NMP and dP stages, and GDF11 generally acts to suppress the effect of BMP4 addition (Figure S5). Taken together, these findings suggest that GDF11 functions as a stage-dependent modulatory signal that reinforces posterior identity and promotes the selective specification of dI subtypes.

Stem cell-derived dIs recapitulate endogenous developmental trajectories

To determine whether stem cell-derived dIs recapitulate endogenous developmental programs, neurons generated under RA, RA+BMP4 and RA+GDF11 conditions were analyzed by scRNA-Seq at day 43 (Figure 6A-6C). All differentiation conditions generated substantial populations of dI4 (50%) and dI5s (~10%), consistent with the lineage distribution observed in the *in vivo* atlas. RA+ BMP4 treatment increased representation of dorsal-most interneurons, including dI2 populations, while small populations of ventral interneurons and MNs were also detected, suggesting emergence of intrinsic patterning centers within the EBs.

Using scVI and scArches[218], we projected the *in vitro* single-cell datasets onto the full *in vivo*

reference atlas (Figure 1B) and the *in vivo* neuronal-specific atlas (Figures 2A, 6D, 6E). This approach revealed strong alignment of stem cell-derived neurons with endogenous spinal neuron trajectories (Figure 6D, 6E). Label transfer analysis demonstrated the preferential alignment of *in vitro*-derived cells with neuronal populations rather than non-neuronal lineages, supporting faithful recapitulation of spinal neuronal identities. Specifically, the day 27 and day 43 *in vitro*-derived dIs map along the corresponding *in vivo* lineage trajectories (Figure 2A, 6E). Consistent with day 43 dataset being slightly more mature, the number of dI4/dI5s are increased compared to the day 27 dataset and they extend further along their trajectories (Figure 6E). Taken together, this data suggests that the stem-cell derived dIs are transcriptionally indistinguishable from their endogenous counterparts.

While the *in vitro* dataset overlaps primarily with regions present in the earliest time points in the atlas (i.e., weeks 4-7), it is underrepresented in the dI4/dI5s specified from week 8 and onward (Figure 6F, S2). To more formally assess the maturation state of the *in vitro* neurons, we trained a classifier on the week metadata in the scVI space of the neural atlas and transferred these labels to our *in vitro* dataset. Comparison of developmental stage signatures indicated that day 43 *in vitro* neurons most closely resembled gestational week 6 populations (Figure 6G-6I), consistent with expected maturation timelines. These findings indicate that human stem cell-derived dIs align with endogenous lineage trajectories and developmental stage**.**

Mechanosensory interneuron networks enriched for ASD-associated genes emerge during early spinal cord development

We next investigated whether modality-specific transcriptional programs associated with dI4/5 subclasses provide insight into the functional organization of the developing sensory circuits. Projection of datasets from *in vitro*-derived neurons onto the neural atlas enabled assignment of cells to defined dI4/dI5 clusters, facilitating comparison of predicted functional programs across populations (Figure 1C, 6F).

GO analyses identified conserved molecular programs associated with pain perception, mechanosensation, neurotransmitter regulation, and balance across *in vitro* and *in vivo* datasets (Figure 6J, 6K). Protein interaction network analysis using STRING[219] identified conserved molecular modules that encode pain perception (dI4s, Figure S8A, S8D, S8E, S8H, S8I; dI5s, Figure S10A, S10B, S10C, S10F), the response to cocaine (dI4s, Figure S8B, S8C, S8F), perception of mechanical stimulus (dI4s, Figures S8E, S9A, S9B, S9D, S9E; dI5s, Figures S10E) mechanosensory behavior (dI4s, Figure 7L, 7N) and balance (dI4s, Figure 7O, Figure S9C; dI5s, Figure 7M). While day 43 *in vitro* neurons are relatively immature (Figure 6H, 6I), several networks share molecular signatures with *in vivo* networks encoding the same function. For example, networks encoding pain perception in cluster 20 share a protein loop centered on NPY1R, thought to mediate neuropathic pain[220] (Figure S8D, S8E), while a dI4 touch circuit in cluster 4 (Figure S9B) resembles an *in vivo* dI4 touch circuit within the corresponding cluster (Figure S9A).

Several clusters associated with mechanosensory behaviors (Figure 6N) and balance (Figure 6O) exhibited enrichment of genes with high SFARI scores[182], including NGLN2[221, 222], NRXN1/2[221, 223], SHANK1/3[224], DLG3[225], CNTNAP2[226]. Identical transcriptional signatures were observed in both *in vivo* and *in vitro* dI4/dI5 populations and emerged at very early developmental stages, i.e., gestational week 6.

Given the frequent occurrence of altered sensory processing and balance phenotypes in individuals with ASD, these findings suggest that molecular programs associated with neurodevelopmental disorders are present within developing spinal cord sensory circuits. These results raise the possibility that sensory phenotypes associated with neurodevelopmental disorders may arise from distributed circuit-level developmental programs spanning both spinal and supraspinal regions.

Discussion

In this study, we define a developmental framework underlying human dI diversification by integrating high-resolution single cell atlas and stem cell-modeling. We establish multiple human NMP–based directed differentiation strategies that generate dorsal spinal interneurons across axial levels and benchmark their identities against a temporally resolved single-cell atlas of the developing human spinal cord. Integration of *in vitro* and *in vivo* datasets demonstrates that NMP–derived dIs recapitulate endogenous transcriptional identities and developmental trajectories, permitting subtype-level comparison between stem cell-derived and fetal spinal neurons. The atlas further reveals extensive diversification of dI4 and dI5 populations, identifies subtype-specific molecular programs associated with sensory-processing functions, and uncovers transcriptional networks enriched for high-confidence ASD risk genes. Together, these findings provide a framework for understanding how human dorsal spinal interneurons acquire positional and functional identity as they assemble into somatosensory circuits.

Temporal competence and signaling context jointly shape posterior identity

Our results support a model in which posterior spinal cord identity emerges through the interaction between developmental time and signaling environment rather than via a single instructive cue. Extending NMP culture duration promotes the sequential acquisition of posterior competence, consistent with previous work showing that temporal progression contributes to HOX gene activation[202]. However, we also find that extended time in culture alters NMP identity in more complex ways, such as inducing the progressive enrichment of mesodermal gene signatures (Figure 4). Thus, while posterior competence increases with time, prolonged NMP maintenance may simultaneously bias lineages, underscoring a tradeoff between axial maturation and progenitor plasticity.

We observe broad and overlapping *HOX* gene expression profiles, consistent with previous work[202, 227], suggesting that posterior identity in human neural progenitors reflect combinatorial

HOX activity rather than sequential co-linear activation of posterior *HOX* paralogs (Figure 4). Similar deviations from classical HOX co-linearity have been observed in analyses of human fetal spinal cord development[185], supporting the idea that human axial patterning involves more flexible regulatory logic than inferred from model organisms.

Our analysis also reveals stage-dependent effects of GDF11. While GDF11 applied during NMP induction has modest effects on transcription, GDF11 applied during dorsal progenitor differentiation reinforces posterior HOX expression and selectively influences dI subtype specification. These findings suggest that GDF11 acts as a modulatory signal that stabilizes lineage decisions once progenitors have acquired sufficient temporal competence. Such temporally gated responsiveness may reflect prolonged developmental plasticity characteristic of human neural progenitors[228] and highlights how human spinal cord patterning may differ from signaling hierarchies derived from rapidly developing model organisms with shorter gestation periods.

Expansion of dI4 and dI5 populations reveals diversification of human somatosensory interneurons

The integrated human spinal cord atlas provides a temporally resolved view of interneuron diversification and reveals a striking expansion of dI4 and dI5 populations at later gestational stages (Figure 2A, S2). Comparative analysis with mouse embryonic spinal cord atlases[211] indicates substantial overlap between human dI4/dI5 clusters and late-born mouse dIL_A/dIL_B populations, suggesting conserved developmental origins.

dI4s and dI5s have transcriptional signatures that are associated with molecular pathways implicated in mechanosensation, nociception, and neuromodulatory signaling. For example, subsets of dI5 populations express PIEZO2[68], a mechanosensitive ion channel required for touch and proprioception, while subsets of dI4 neurons expresses NPY, a neuropeptide associated with inhibitory modulation of pain pathways. We also identify transcriptionally dI4/dI5 subclusters with

human specific sensory functions, consistent with observed species-specific differences in somatosensory modalities[229]. Moreover, the mechanistic basis for individual somatosensory differences remains an intriguing and understudied area. Subtle biases within dI4/dI5 populations could potentially explain wide-ranging differences between individuals, such as pain sensitivity or proprioceptive acuity. Such diversification underscores the importance of human developmental atlases for identifying cell types and regulatory programs not completely captured in model systems.

Developmental origins of sensory phenotypes associated with neurodevelopmental disorders

An unexpected finding of the atlas is the enrichment of ASD-associated gene networks within the dI4/dI5 interneuron populations (Figures 6L-6O, S8-S10). Both *in vitro* and *in vivo* dI4/dI5 clusters exhibit gene networks associated with mechanosensation, balance, and neuromodulatory signaling that include genes strongly linked to ASD risk, including NRXN1[230], SHANK1/3[231], DLG4[232], CNTNAP2[233], and NLGN2[234]. Although ASD is primarily studied in the context of cortical development, sensory processing differences are a prominent feature of the disorder.

The presence of ASD-associated molecular signatures in developing spinal interneurons raises the possibility that sensory phenotypes associated with ASD (such as differences in touch and balance processing) may arise from developmental programs spanning both spinal and supraspinal circuits. Notably, these transcriptional signatures are detectable in stem cell-derived interneurons at early developmental stages, consistent with the idea that ASD-associated molecular differences emerge during embryonic circuit formation[235]. While further functional studies are required to establish causal relationships, these findings highlight the potential contribution of spinal sensory circuitry to neurodevelopmental phenotypes conventionally attributed primarily to brain-specific mechanisms.

Implications for regenerative strategies targeting sensory circuits

The ability to generate dI subclasses matched for axial identity and developmental stage has

implications for regenerative medicine and SCI research. Sensory dysfunction following SCI reflects both neuronal loss and disruption of dI circuits that integrate mechanosensory information. The reference atlas indicates that dI4 and dI5 populations represent major components of human dorsal sensory circuitry and our differentiation protocols recapitulate this lineage distribution.

The close transcriptional correspondence between stem cell-derived and endogenous dIs suggests that NMP-based differentiation strategies can produce cell populations that approximate physiologically relevant sensory interneuron identities. Such lineage-matched populations provide a rational framework for reconstructing dorsal spinal circuits damaged by injury. More broadly, the ability to generate defined interneuron subclasses provides a foundation for mechanistic studies of human sensory circuit assembly and for testing cell replacement strategies designed to restore sensory function.

Limitations and future directions

Several questions remain regarding the specification and maturation of human dIs. Some populations, including dI2s, remain difficult to detect using conventional marker-based approaches and may represent transient developmental states or require additional patterning cues. In addition, the *HOX* gene expression patterns *in vitro* do not follow a simple anterior–posterior hierarchy, suggesting that axial identity in human neural progenitors may be governed by combinatorial or context-dependent regulatory mechanisms. Further studies will be required to determine how maturation state influences the functional integration of stem cell-derived interneurons and whether these cells can restore circuit function in models of spinal cord injury. Together, our findings highlight principles governing specification of human dorsal spinal interneurons and establish principles for engineering sensory circuit cell types with defined positional and functional identities.

Materials and Methods

Construction of the human atlas

Integration of single-cell datasets from different origins

Data from six publications[185, 197-201] were loaded into Seurat v5[205] in R. For each dataset, all samples were combined into one object, and each sample was split into a separate layer. Quality control was performed automatically by removing cells for each sample that fell outside of three median absolute deviations (both sides for nCount_RNA and nFeature_RNA, and higher for percent.mt). The Anderson dataset did not include mitochondrial genes and thus was only filtered on the two other metrics. After QC, all datasets were combined into one large Seurat object and the data was normalized and variable features were calculated. The object was then converted to anndata and integration was performed using SCVI[206] on the counts layer using only the variable features (n_layers = as.integer(2), n_latent = as.integer(30), gene_likelihood = "nb"). The latent representation was then brought back into the full Seurat object and clusters and UMAP were calculated. Cluster identities were assigned using sctype[207] and a custom list of marker genes collated from various publications.

Reclustering of neuronal clusters

To analyze the neuronal population, clusters identified as neurons, peripheral neurons, sensory neuron progenitors, and progenitors-neurons were isolated from the atlas and reintegrated using newly calculated variable features and new scVI parameters as recommended by scArches[218] (use_layer_norm = "both", use_batch_norm = "none", encode_covariates = T, n_layers = as.integer(2), dropout_rate = as.numeric(0.2)). After dimensional reduction, a few undesired populations were still left, so clusters that were identified either as peripheral populations, or unidentified neuronal populations (mostly identified by a lack of distinct marker genes) were removed and the dataset was reintegrated.

dI cell type extraction based on the marker gene expression

To identify the various populations of cells within the neuronal atlas, AUCell[210] was used with a custom list of marker genes to create a module score for each cell, for each possible cell type. Cutoffs were manually assigned for scores for each cell type, and every cell was assigned 0, 1, or multiple identities. Cells with multiple identities were then rectified using a kNN classifier (k=10) to change their identity to match those of surrounding cells. Then the same was done with cells assigned no identity to extrapolate to nearby populations not expressing high levels of any module.

dI4/dI5 subcluster analysis

Differential gene expression analysis across clusters was performed using the FindAllMarkers function from the Seurat R toolkit. The results were filtered, to only include clusters identified as containing large portions of either dI4 (2, 4, 5, 7, 9, 11, 12, 14, 18, 19, 20, 23, 26, 27, 29, 36, 37, 43, 44, 48, 52, 65) or dI5 (0, 10, 13, 15, 16, 17, 21, 30, 49, 54, 61). Of note is that due to the AUCell classification process, clusters are not made up exclusively of one cell type, but each cluster identified here has a large portion of cells identified as either dI4/5. Gene Ontology (GO) enrichment analysis was performed on significantly upregulated genes in each cluster (avg_log2FC > 0.25) using the enrichGO function from the clusterProfiler R package. The ontology category was restricted to Biological Process (BP), multiple testing correction was applied using the Benjamini-Hochberg method, and terms were considered significant with an adjusted p-value of 0.05 and q-value of 0.2. GO categories for dot plot visualization were selected based on their association with various sensory modalities and overlap, or lack thereof, amongst clusters. The STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) resource was utilized to explore known and predicted protein-protein interaction networks from genes that were making up GO categories of interest.

Single-cell RNA sequencing and analysis of in vitro cultures

Preparation of single cell suspension

Single cell suspension from EBs is generated via papain dissociation method using the Worthington papain dissociation kit (refer to manual). Briefly, EBs are first collected in 15mL centrifuge tubes and washed twice with DPBS. The papain solution is prepared by mixing papain with DNAse (supplied in the kit), and the solution is allowed to activate for 10 minutes at 37ºC before adding to the EBs followed by triturating 10 times with a p1000 pipette. EBs are then preferably placed on an orbital shaker at 37ºC and 5% CO_2 for 1 hour and 15 minutes and triturated 10 times every 5 minutes or until no visible EB clumps remain. Single-cell suspension of EBs are then centrifuged at 300 x g for 7 minutes. The cell pellet is resuspended in 2mLs of Ovomucoid protease inhibitor solution to stop the papain digestion and passed through a 40µm cell strainer to remove undigested organoid debris. Single cells are then centrifuged once more at 300 x g for 7 minutes and the cell pellet is then resuspended in 1mL of 0.04% BSA in DPBS for GEM preparation using 10X genomics platform (the concentration of BSA can be increased to reduce cell stickiness/chance of doublets).

Library preparation, and sequencing

~ 10,000 live cells/condition were used to generate single-cell cDNA libraries using protocol described by the 10X Genomics 3'GEX library preparation kit. Cells were first partitioned into GEMs (gel-beads-in-emulsion) containing using the 10x Genomics Chromium controller. Further, 3' mRNA-seq gene libraries were prepared using the Chromium Single Cell 3′ Library & Gel Bead Kit v2 (10x Genomics), according to the manufacturer's instructions. Library quality control was assessed using Agilent and Qubit assays. If quality control is within acceptable conditions (e.g., library concentration of 20-40ng/ul), libraries were sequenced on the Illumina NovaSeq X Plus 10B. For 3'GEX libraries targeting 10,000 cells 1,800 million reads/sample were obtained. The GEO accession numbers for all the datasets can be found in Table S3.

Integration of single-cell in vitro datasets

Single-cell libraries were processed with the 10x Genomics Cell Ranger Count pipeline (v9.0.1) and loaded into Seurat v5 in R. Within the Seurat object, quality control was performed for each sample by removing cells that fell outside of three median absolute deviations for total UMI counts (nCount_RNA) and number of detected genes (nFeature_RNA), and above three median absolute deviations for percent mitochondrial reads (percent.mt). After QC, data were normalized and variable features calculated.

For day 43 samples, separate Seurat objects were created for the RA/BMP series without GDF11 and the RA ± GDF11 experimental series. Following QC, data were normalized and variable features calculated independently before the two objects were merged into a single Seurat object for subsequent integration.

For nonlinear batch integration, the Seurat object was subset to variable features and converted to an AnnData object. Integration was performed with SCVI, using a model trained on raw counts with default settings (batch_key = "sample"), to obtain a low-dimensional latent embedding. The latent representation was imported back into Seurat as a dimensional reduction, and was used to compute nearest-neighbor graphs, clusters, and UMAP embeddings. To provide an independent annotation of broad neuronal versus non-neuronal populations in the in vitro cultures, we applied sctype with a custom list of marker genes collated from various publications. For downstream atlas-based analyses, the dataset was restricted to cells classified by sctype as neuron, peripheral neuron, sensory neuron progenitor, or progenitor–neuron. Populations with overwhelmingly stressed marker genes or a lack of distinct marker genes were additionally removed. Examples of scRNA-Seq datasets before the neural lineage restriction are included in Figure S7A-S7J.

Projection of in vitro cells onto the human atlas

Projection of the in vitro cells onto the human spinal cord atlas and its neuronal subset was performed using pretrained SCVI reference models built on the whole in vivo atlas and the

neuronal atlas respectively. The in vitro Seurat object was subset so that its gene set matched the genes present in the reference AnnData object. After conversion of the object to AnnData, reference mapping followed the scArches workflow (SCVI.prepare_query_anndata followed by SCVI.load_query_data), and the query model was fine-tuned for 200 epochs to align the in vitro dataset to the reference latent space. The resulting query-aligned latent representation was imported into Seurat and projected into the appropriate reference UMAP space using the saved UMAP models stored in the atlas objects.

To transfer annotations from the *in vivo* neuronal atlas to the in vitro cells, we trained supervised machine-learning models using atlas SCVI embeddings as predictors. K-nearest neighbors (k = 10) was used for dI cell-type classification (based on AUCell-derived identities; see dI cell type extraction based on the marker gene expression) and prediction of developmental week, while random forest (ranger) was used for cluster identity transfer. Cluster-specific markers for cells assigned to in vivo dI4 or dI5 enriched clusters were identified using FindAllMarkers. These gene sets were used to assess if in vitro cells recapitulate *in vivo* dI4/dI5 subtype transcriptional programs (see dI4/dI5 subcluster analysis).

Bulk-RNA sequencing and data processing

RNA preparation

NMP cultures subjected to bulk-RNA sequencing were first washed with 1X DPBS to remove the media. Cells were then lysed by directly adding 600µL of RLT buffer in the culture wells, and the lysate is then collected and flash frozen with liquid nitrogen to be stored at -80ºC. The total RNA was then purified using Qiagen's RNeasy kit and eluted in 50µL of RNAse-free water. Using a microvolume spectrophotometer, RNA concentrations were determined and diluted to 1µg/µL. RNA samples were further subjected to quality control assessment using Agilent Technologies 2100 Bioanalyzer and only samples with a RIN score >8.0 were used for library construction using Universal plus mRNA sequencing kit (NuGEN). Libraries were then sequenced on an Illumina

Novaseq X Plus to generate ~30 million reads per sample. The GEO accession numbers for all the datasets can be found in Table S3.

Processing and differential expression analysis

Reads were obtained as raw FASTQ files, assessed for quality with FASTQC, and aligned to the human reference genome (GRCh38, GENCODE v29 annotation) using STAR. Gene-level counts were obtained with featureCounts (exon-level, paired-end read counting). Differential gene expression was tested using DESeq2. For each reference level of interest (hESC, D2NMP, D4NMP, D10NMP), a re-leveled model was built and processed separately using the design formula ~ condition, to obtain pairwise contrasts across all conditions. Global transcriptional relationships among samples were assessed using VST-normalized counts from the hESC-referenced model for PCA analysis, sample-to-sample Euclidean distance calculations, and heatmaps.

Shrinkage of $\log_2$ fold-changes for volcano plots and Gene Ontology (GO) enrichment analyses was performed on contrasts using apeglm. For each contrast, significantly upregulated and downregulated gene sets (padj < 0.05, $|\log_2$ fold change| > 1) were analyzed separately with clusterProfiler (enrichGO) using the background of all detected genes in the experiment. The ontology category was restricted to Biological Process (BP), p-values were adjusted using Benjamini–Hochberg, and terms with adjusted $p < 0.05$ ($q < 0.2$) were considered significant.

Cell culture

ESC maintenance and culture

H9 ESCs were maintained in mTeSR plus media and passaged at least twice post-thaw onto Matrigel coated plates before NMP induction. ESCs were passaged when they reached 70% confluency using ReLeSR and were plated at a 1:4 dilution ratio. The newly passaged hESC cultures reached 70% confluency within 4 days.

NMP induction

To induce 2-, 4-, and 10-day NMPs, hESCs were cultured in a N2/B27-based SAND media without vitamin A (IMDM, N-2 1%, B-27 without vitamin A 2%, GlutaMAX 1x, NEAA 1x) under adherent culture conditions. Briefly, NMP induction was initiated by first pulsing the hESCs with 10ng/mL of bFGF for one day followed by the combination of bFGF + Wnt agonist, CHIR 99021. For example, to derive 2-, 4-, or 10-day NMPs, the differentiating hESCs received one day of 10ng/mL bFGF, followed by bFGF + 3μM CHIR 99021 for one day (2-day NMPs), three days (4-day NMPs), or nine days (10-day NMPs) respectively.

Embryoid body (EB) differentiation

Upon the completion of NMP induction in adherent cultures, cell sheets containing NMPs were cut with an EZ passage tool and transferred to the ultra-low attachment plates which allowed NMP clumps to self-aggregate into EBs. EBs were cultured with SAND media containing vitamin A (SAND + B27 with vitA) and further supplemented with 1μM RA for two days to induce neural patterning. From the 2nd day until the 14th day, EBs were treated with various patterning factors to induce relevant cell types. For example, to induce dI4-6 populations, EBs were treated with 1μM RA ± GDF11 up to 14 days in suspension cultures. For inducing dI1-3 cell types, EBs were bathed in SAND media supplemented with 1μM RA + 10ng/mL BMP4, and ± GDF11 (50/100ng/mL). On the 14th day EBs were switched into maturation media until day 20, day 43, or day 130 and sampled for quality control checkpoints during maturation. Media changes were performed by collecting EBs in 15mL conical tubes and allowing them to sink to the bottom of the tube in a pellet. Exhausted media was aspirated and replaced with fresh media corresponding to condition. EBs were then transferred back to the plates using glass borosilicate pipettes which prevented EB loss due to sticking to the walls of the plastic pipettes.

Immunocytochemistry

NMPs:Immunocytochemistry (IHC) was performed on NMPs to determine the expression of NMP markers such as SOX2 and Brachyury (T). Briefly, at the end of NMP culture durations, the sheet of NMP cells was cut into pieces by EZ Passage tool and transferred onto Matrigel coated 8-well Ibidi slide chambers. In Ibidi slides, NMPs were allowed to grow for another day in their respective condition media and were processed for IHC the following day. NMPs were washed once with 1X DPBS and then fixed on ice with 4% PFA for 8–10 minutes. PFA solution was removed by three 5-minute washes with 1X DPBS, followed by incubation in the antibody blocking solution (1X DPBS, 1% HIHS, 0.1% Triton X-100, and 0.05% sodium azide) for 1 hour at 2–8ºC. Primary antibodies were diluted in antibody blocking solution and added to the NMPs for the overnight incubation at 2–8ºC. The following dilution was used: SOX2 (mouse; Santa Cruz, sc-365823; 1:500), BRACHYURY (goat; R&D systems, AF2085; 1:500), and CDX2 (rabbit, Cell Signaling Technology, D11D10; 1:500). NMPs were washed twice with 1X DPBS with 0.1% Triton X-100 (PBT) for 5 minutes before secondary antibody incubation. Corresponding species-specific secondary antibodies (Jackson ImmunoResearch Laboratories) were diluted 1:400 in 1X PBT and allowed to incubate in the dark at room temperature for 1 hour followed by counterstaining with nuclei stain DAPI (1:500 dilution in 1X PBT) for 15 minutes and replaced with 1X DPBS for imaging.

Embryoid bodies (EBs): EBs were first collected by the sinking method, and the media was removed by two 5-minute washes with 1X DPBS, followed by fixing in 4% PFA on ice for 10 minutes on rocker. EBs were then washed three times for 5 minutes before allowing them to sink in 30% sucrose in 1X PB solution for at least 1 hour or until all the EBs had sunk to the bottom of the tube, indicating adequate equilibrium with sucrose had been reached. EBs were then embedded in OCT on dry ice blocks and then sectioned at 14µm thick sections using Leica Cryostat onto positively charged microscope slides. Slides with EB sections were then washed with 1X DPBS, blocked with the antibody blocking solution for 1 hour at room temperature before

receiving primary antibodies. The following antibodies were used: LHX2 (mouse; DSHB, PCRP-LHX2-1C11; 1:50), FOXD3 (guinea pig, a gift from Thomas Mueller, Germany, 1:10,000), ISL1 (goat; R&D systems, AF1837; 1:500), PAX2 (rabbit; Invitrogen, 71-6000; 1:500), LMX1B (guinea pig; Thomas Muller Lab), LHX1/5 (mouse; DSHB, 4F2; 1:50) for 12–14 hours (overnight) at 2–8ºC in a humidified chamber. Slides were washed twice with PBT before receiving species-specific secondary antibodies (Jackson ImmunoResearch) 1:400 in the dark for 1 hour at 20ºC. Slides were then washed once with 1X PBT for 5 minutes and incubated with DAPI for 15 minutes to counterstain the nuclei. Slides were then mounted with Prolong Diamond antifade medium and imaged using a Zeiss LSM800 inverted confocal microscope.

Quantitative (q) RT-PCR analysis

Samples of interest were lysed with ~300–500µL RLT buffer and total RNA was purified using Qiagen RNeasy Kit. Total RNA was quantified on a microvolume spectrophotometer, and 500ng/µL RNA was used to generate cDNAs using Superscript IV First Strand Synthesis kit. The gene-specific primers (forward and reverse) were used at 0.5µM final concentration in a 10µL reaction with SYBR Green and qRT-PCR was performed on Roche 480 Light-cycler (Roche). The relative fold expression was determined using the ΔΔCT method to compare the expression of the target gene against the expression of a housekeeping gene glyceraldehyde 3-phosphate dehydrogenase (Gapdh).

Figures

Figure 4-1: Construction of an *in vivo* human embryonic spinal cord scRNASeq reference atlas

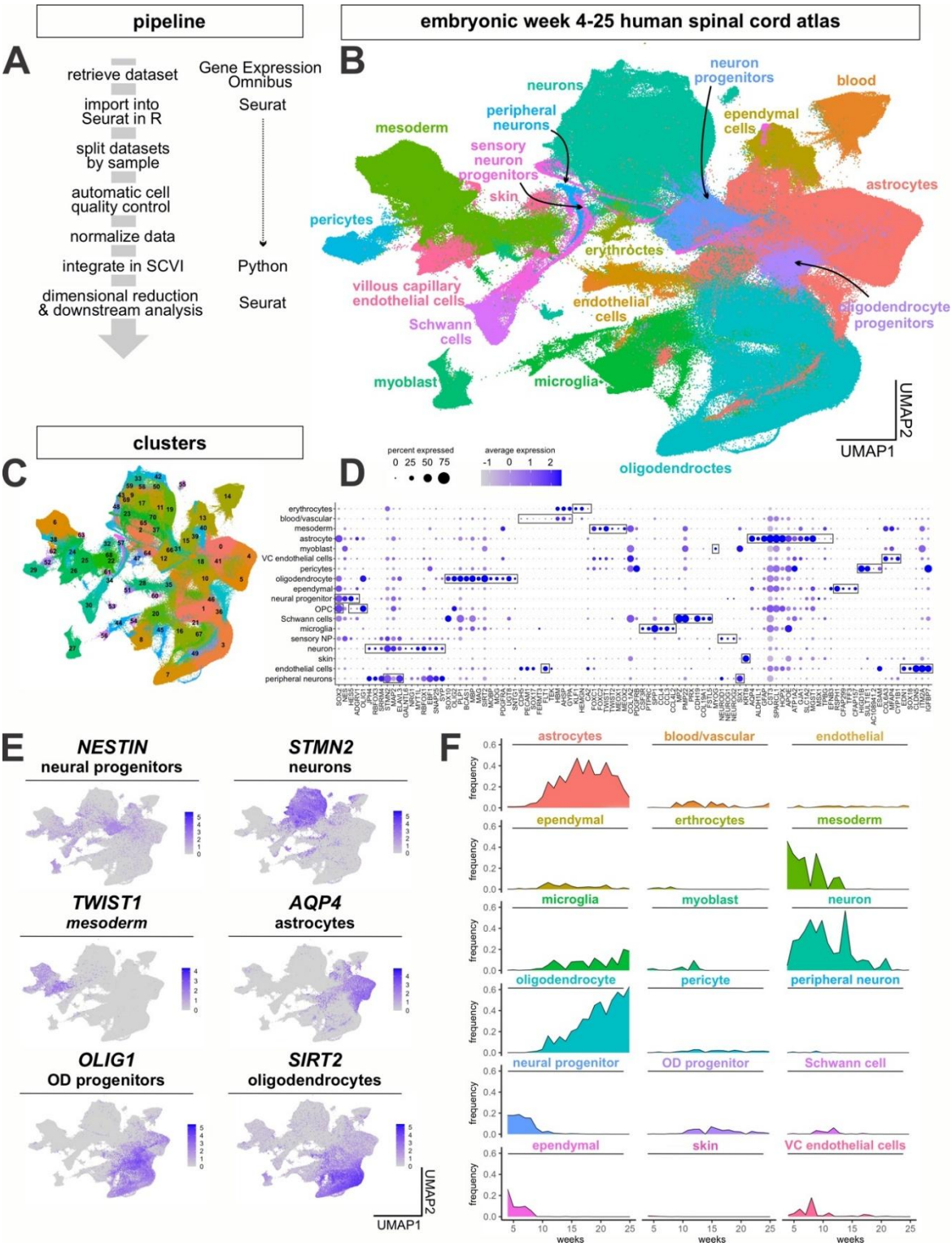

Figure 4-1: Construction of an *in vivo* human embryonic spinal cord scRNASeq reference atlas

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

(B) The *in vivo* human embryonic spinal cord single cell atlas. Integration of 1.7 million cells from 192 samples spanning gestational weeks 4 to 25 reveals 18 broad categories of cell types present in the spinal cord and surrounding tissues.

(C, D) Unsupervised clustering yielded 71 distinct transcriptional clusters (C) that could be grouped into 18 cell types (B), based on their expression of canonical marker genes (boxed genes, D).

(E) UMAPs depicting the expression of representative marker genes in the *in vivo* human embryonic spinal cord single-cell atlas.

(F) Comparison of the frequency by which the 18 cell types emerge over the timeline of the atlas demonstrates the early proliferation of mesodermal and neuronal cell types, with the later proliferation of astrocytes and oligodendrocytes. The apparent increase in neurons at week 14 likely reflects that all W14 datasets consist of single nuclei samples, whereas datasets from other time points are predominantly single-cell samples.

Figure 4-2: Analysis of dI lineages within the *in vivo* human reference atlas

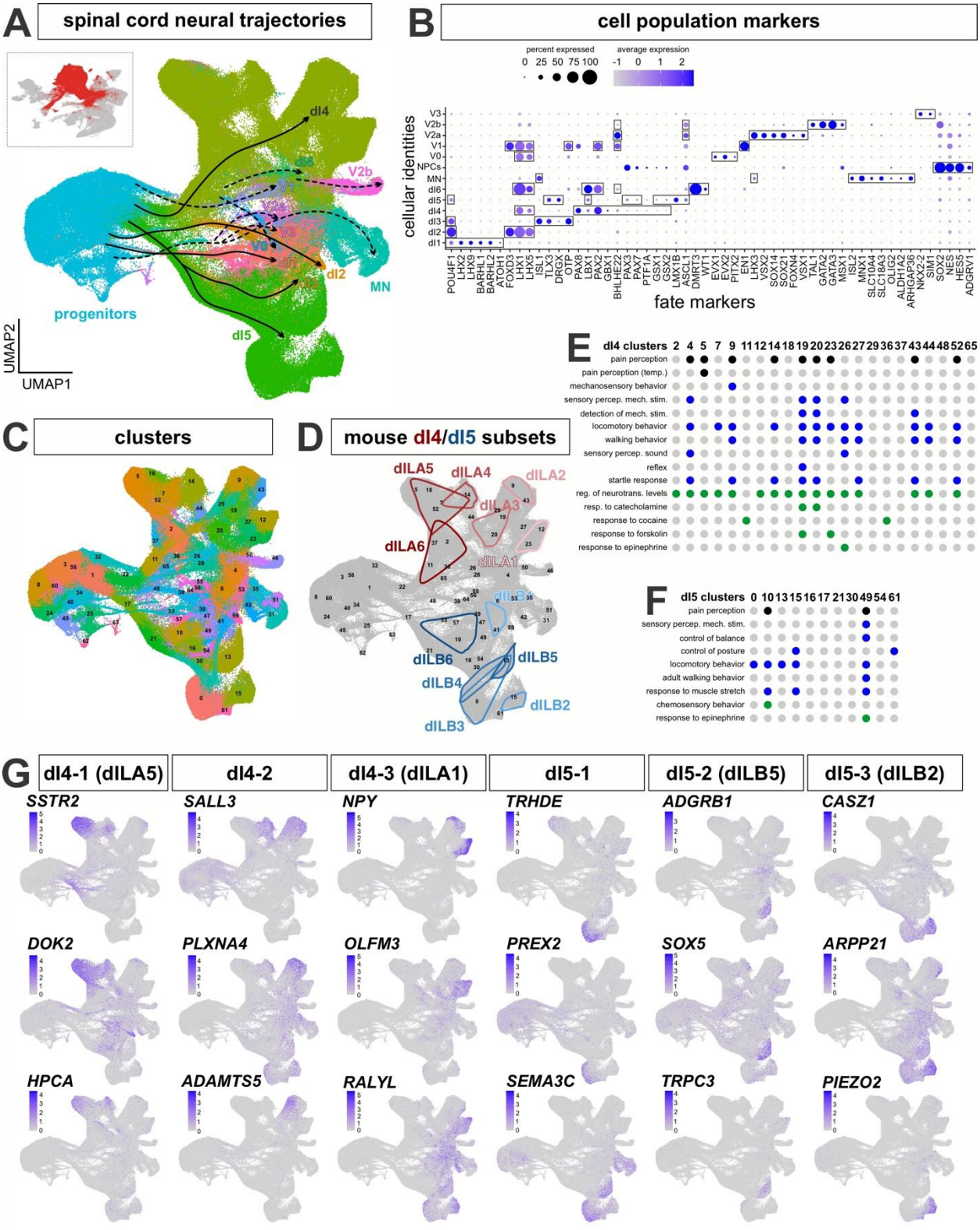

Figure 4-2: Analysis of dI lineages within the *in vivo* human reference atlas

(A) ~300,000 neuronal lineage cells were extracted from the overall atlas (inset) and then plotted as 2d UMAP reduction. Labeling cells taken from across the complete timeline with the AUCell classification system identified all the dorsal (dI1-dI6) and ventral (v0-v3, MNs) neural identities in the spinal cord and revealed a vast expansion of the dI4 and dI5 lineages. The differentiation trajectories were hand annotated using gene expression patterns.

(B) Dotplot of genes used for classification of each cell type shows expression restricted to expected cell types.

(C) Clustering of the neuronal atlas yielded 66 clusters.

(D) The expansion of human dI4/dI5 in comparison was assessed in comparison to an embryonic mouse spinal cord single cell atlas[211]. In the mouse study, dI4/dI5 subtypes were designed as late born dILs, dILA1-6 (Figure S4A) and dILB1-6 (Figure S4B).

(E, F) Gene Ontology (GO) enrichment analysis was performed on significantly upregulated genes (avg_log2FC > 0.25) for all clusters identified as either dI4 or dI5 subtypes. Dot plots of GO Biological Process terms found across dI4 (E) and dI5 (F) clusters include various behavioral functions, largely aligning with the canonical roles dI4 and dI5s play in processing sensory information. Black dots identify GO terms related to the sensory perception of pain and temperature; blue dots signify GO terms from different sensory modalities including touch, sound, movement and balance; green dots represent GO terms associated with a response to chemical stimulants.

(G) Unique gene markers were identified late-stage subsets of dI4 (dI4-1, dI4-2 and dI4-3) and dI5 (dI5-1, dI5-2 and dI5-3) to further demarcate additional distinct cellular subtypes within the dI4 and dI5 lineages. These genes were selected for their specificity to the cluster, with low to no expression in other clusters.

Figure 4-3: Specification of dI identity in the human RA±BMP4 NMP protocol

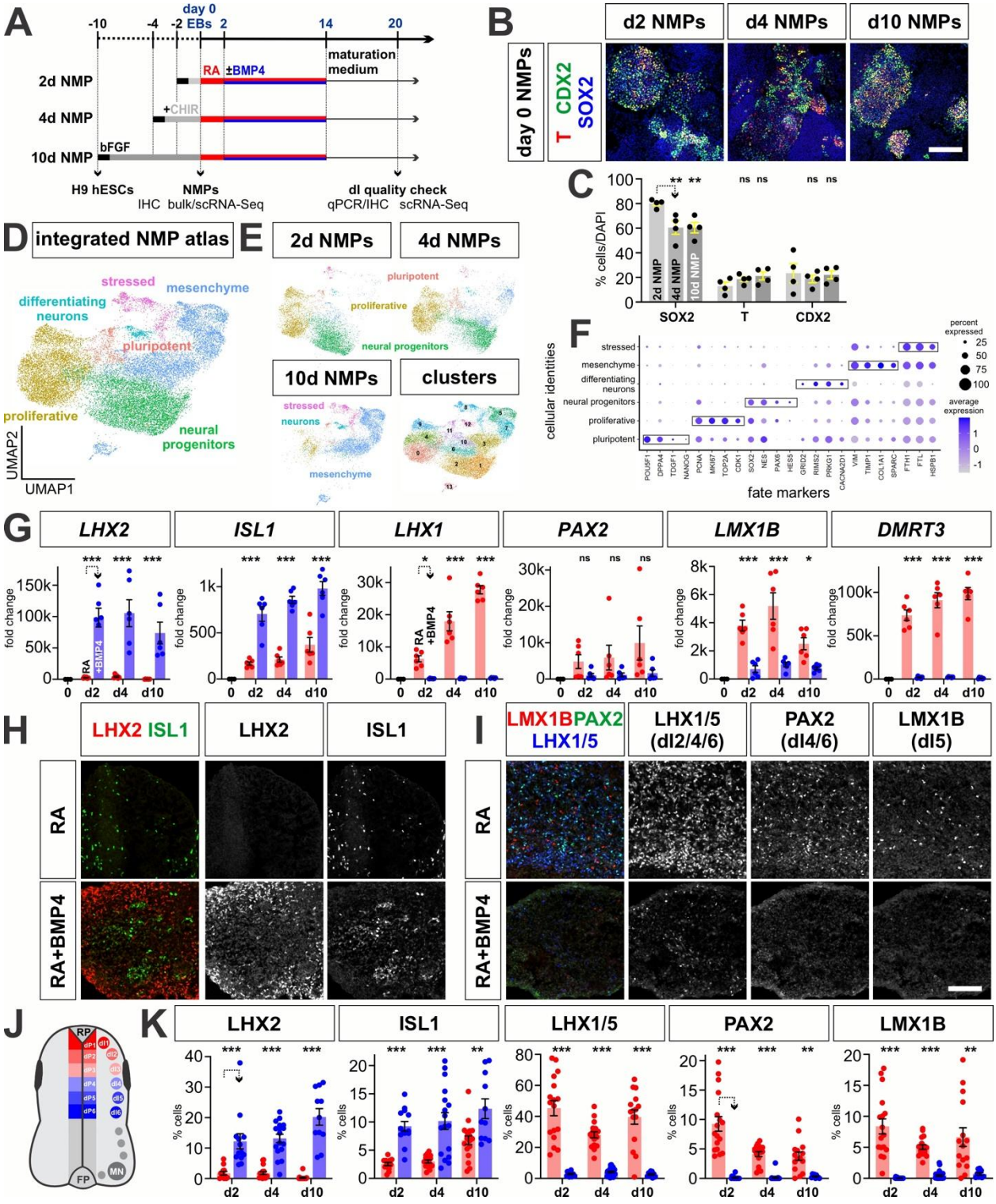

Figure 4-3: Specification of dI identity in the human RA±BMP4 NMP protocol

(A) Overview of the experimental timeline/workflow for the RA±BMP4 NMP protocols. H9 hESCs were cultured with bFGF/CHIR to induce day (d) 2, 4 and 10 NMPs under 2-dimensional (2D) culture conditions. At day 0, NMPs were sent for both bulk and single cell (sc) RNA sequencing (RNA-Seq) for transcriptional analyses. The NMPs were then converted into 3-dimensional (3D) embryoid bodies (day 0 EBs) and then treated with a 2-day pulse of RA, followed by RA+BMP4 for 12 days to induce dorsal patterning. qPCR and immunohistochemical (IHC) quality checks were performed on day 20.

(B, C) IHC analysis of d2 d4 and d10 day 0 NMPs with antibodies directed against T (BRACHYURY, red), CDX2 (green) and SOX2 (blue). Quantification suggests that the number of NMPs is equivalent across the three conditions.

(D) UMAP projection of the complete complement of day 0 NMPs, i.e. 2d, 4d and 10d, demonstrates that the cells segregate into neural progenitor and mesenchymal progenitor populations.

(E) UMAPs split by culture duration prior to NMP induction depict a progressive shift from neural progenitor enrichment in 2d and 4d NMPs towards mesenchymal identity in 10d NMPs.

(F) Dot plot of different NMP fates showing the expression levels of marker genes.

(G-I, K) Both qPCR (G) and IHC (H, I, K) analyses of canonical dI markers in day 20 EBs, demonstrate that the RA and RA+BMP4 protocols, using d2, d4 and d10 NMPs, generate the predicted complement of dIs, i.e., RA→dI4, dI5, dI6 and RA+BMP4 → dI1 and dI3.

(J) Schematic representation of the developing spinal cord *in vivo*, depicting the location of the six populations of dorsal progenitors (dP) and dorsal interneurons (dIs).

Scale bar: 100μm (B); 75μm (H, I)

Probability of similarity between control and experimental groups: *= $p < 0.05$, **$p<0.005$ *** $p<0.0005$; two-way ANOVA.

Figure 4-4: Assessing the effect of culture duration on NMP identity along the anterior-posterior axis

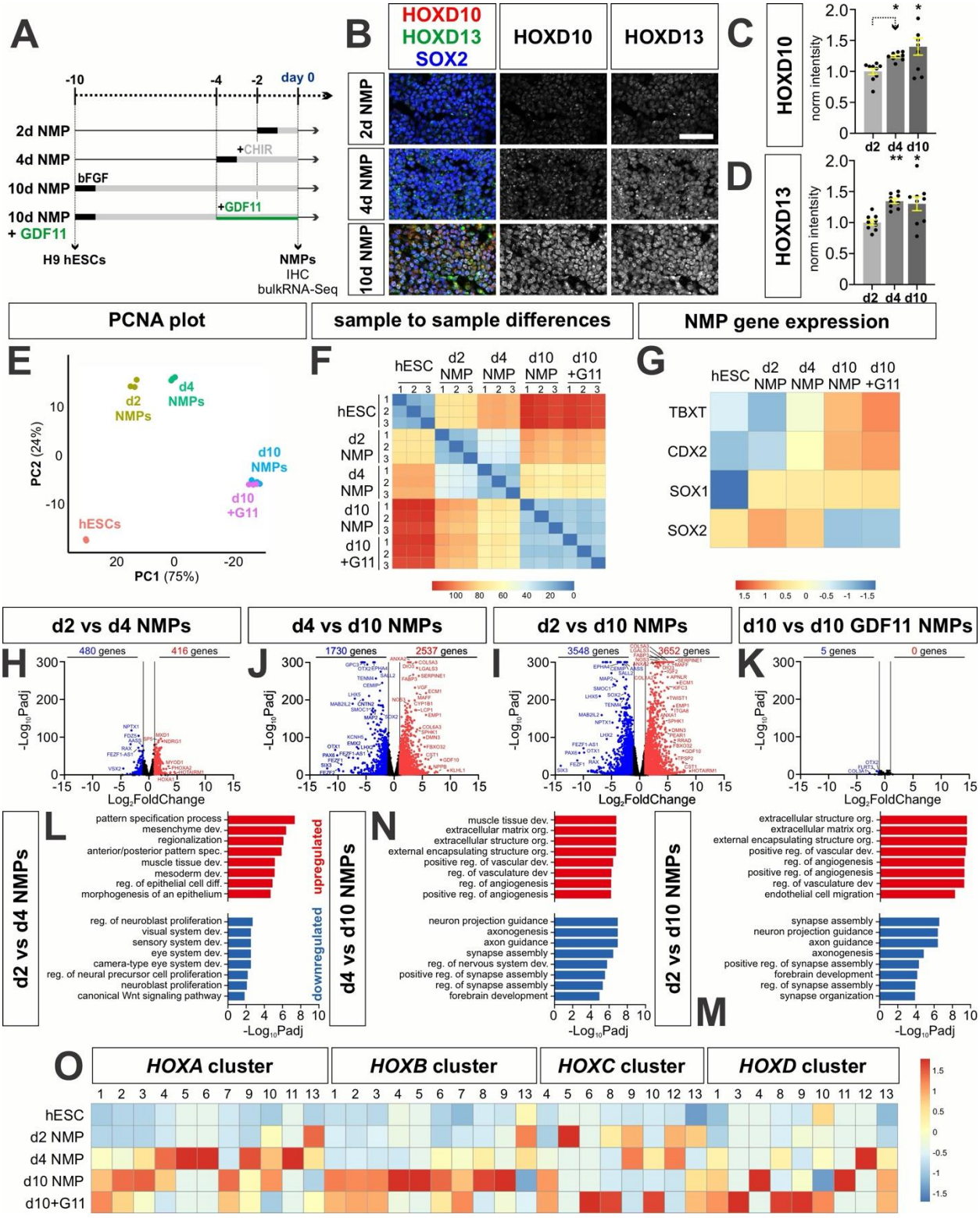

Figure 4-4: Assessing the effect of culture duration on NMP identity along the anterior-posterior axis

(A) Overview of the experimental timeline/workflow for the RA±BMP4 NMP protocols.

(B -D) IHC analyses of NMPs immediately before EB formation also suggests that prolonging the culture time progressively posteriorizes NMPs. Two posterior HOX genes - HOXD10, and HOXD13 - have significantly higher expression in d10 NMPs, compared to d2 NMPs. SOX2 levels are ~constant across the d2, d4 and d10 NMP conditions, indicating that the NMP identity is maintained.

(E) Principal-component analysis (PCA) of bulk RNA-seq expression profiles from hESCs, d2, d4, d10, and d10+GDF11 NMPs shows that samples segregate along PC1 (75%) according to the duration of NMP induction, suggesting culture time is the primary driver of transcriptional variation.

(F) Heatmap of pairwise Euclidean distances between samples. hESCs are most dissimilar from d10 and d10 + GDF11 NMPs, while d10 and d10+GDF11 samples are highly transcriptionally similar.

(H-K) Volcano plots comparing the differentially expressed genes (DEGs) in the d2, d4 and d10 NMP cultures ($|\log_2$ fold change$| \geq 1$, Benjamini-Hochberg-adjusted $p < 0.05$). The magnitude and number of transcriptional differences increase with greater separation in induction time, with the largest number of DEGs observed between d2 and d10 NMPs, and minimal differences between d10 and d10+GDF11 NMPs, indicating largely shared transcriptional states.

(L-M) Gene Ontology (GO) enrichment analysis of significantly upregulated and downregulated genes in contrasts. Genes upregulated in d4 NMPs are enriched for patterning and morphogenetic processes, while d10 NMPs show strong enrichment for extracellular matrix organization and vasculature-associated pathways. Downregulated gene sets suggest that forebrain-associated transcriptional programs are progressive reduced with increased culture

time.

(O) Heatmap of average *HOX* gene expression across conditions, scaled per gene. Robust *HOX* expression emerges in d4 NMPs, predominantly in the HOXA cluster. d10 NMPs show strong *HOXB* and *HOXD* gene expression, while the addition of GDF11 increases expression in the *HOXC* and *HOXD* clusters, suggesting that prolonged culture duration and GDF11 exposure play a role modulating axial identity through differential activation of *HOX* clusters.

Scale bar: 20μm

Probability of similarity: *= $p < 0.05$, **$p<0.005$; two-way ANOVA.

Figure 4-5: Assessing the effect of GDF11 treatment on both dorsal-ventral and anterior-posterior identity

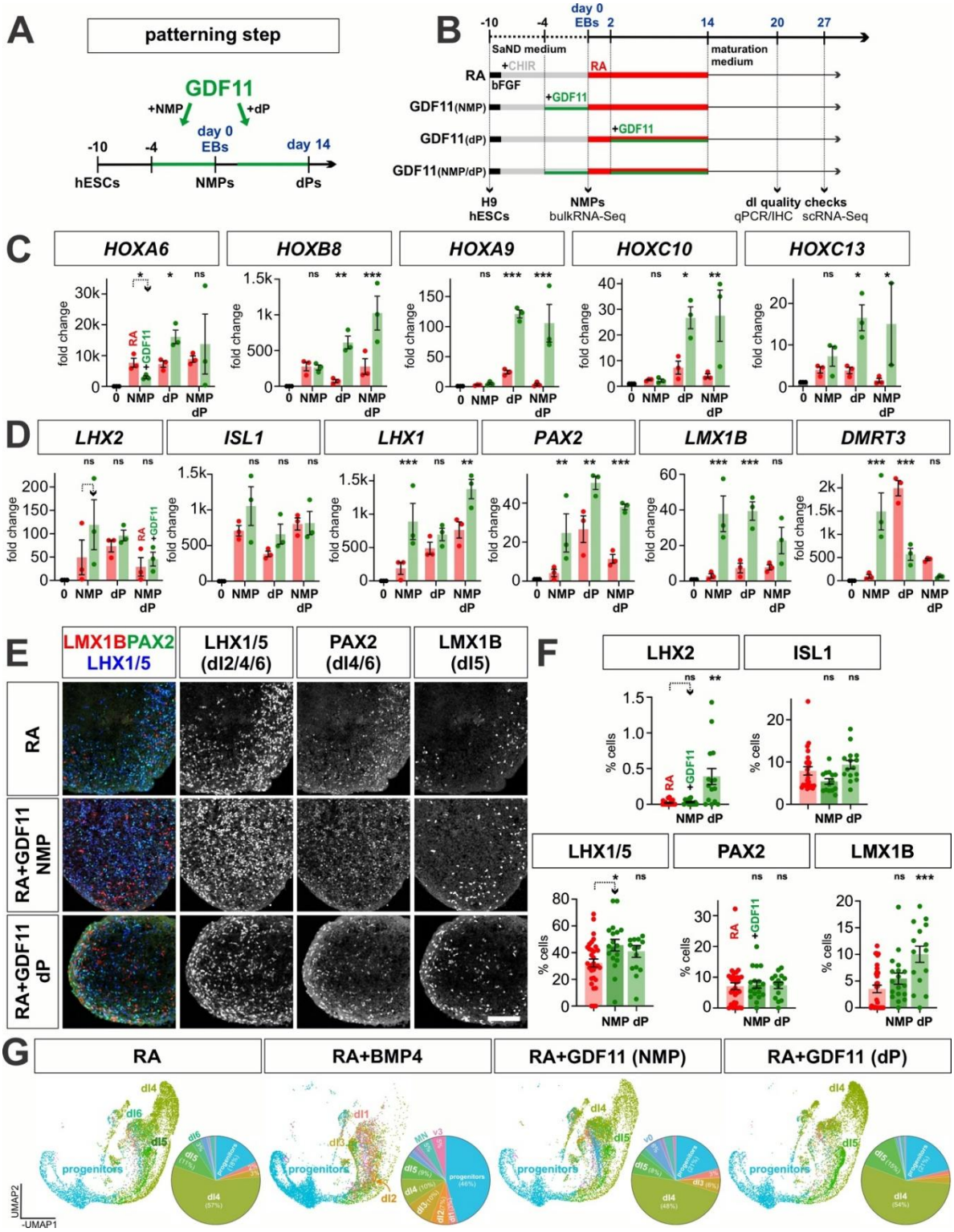

Figure 4-5: Assessing the effect of GDF11 treatment on both dorsal-ventral and anterior-posterior identity

(A, B) Overview of the experimental timeline/workflow for the RA±GDF11 NMP protocols. The effect of GDF11 treatment was assessed on NMP patterning, dP patterning and both NMP and dP patterning (A).

(C) qPCR analyses of a range of *HOX* gene expression suggests that GDF11 addition at the dP stage is the most effective at driving the expression of the posterior most HOX genes (*HOXA9*, *HOXC10* and *HOXC13)*.

(D-F) qPCR analyses of day 20 EBs (D), suggest that the addition of GDF11 at either the NMP or dP stage significantly increases expression of *PAX2* and *LMX1B*, the canonical genes that mark the dI4/dI5 identities. However, the IHC analysis suggest the effect of GDF7 on protein levels is more subtle, with the most significant effect observed when GDF11 added at the dP stage on dI5 identities.

(G) UMAPs of day 27 *in vitr*o-derived cells generated using the RA, RA+BMP4, and RA+GDF11 (NMP and dP) protocols, using neural spinal cord cell-type identity transferred from the *in vivo* neuronal atlas (Figure 2a). The pie charts summarize the proportions of neuronal cells generated by each protocol. The RA and RA+GDF11 conditions predominantly yield dI4-dI6 identities, whereas the addition of BMP4 shifts patterning towards the dorsalmost dI1-dI3 identities. Adding GDF11 during NMP induction has minimal effects on patterning compared to RA alone, while GDF11 treatment during dP formation increases in number of dI5s, apparently at the expense of the dI6 population.

Scale bar: 75μm

Probability of similarity between control and experimental groups: *= $p < 0.05$, **$p<0.005$ *** $p<0.0005$; two-way ANOVA

Figure 4-6: Comparison of *in vitro* data with the *in vivo* reference atlas reveals ASD signatures

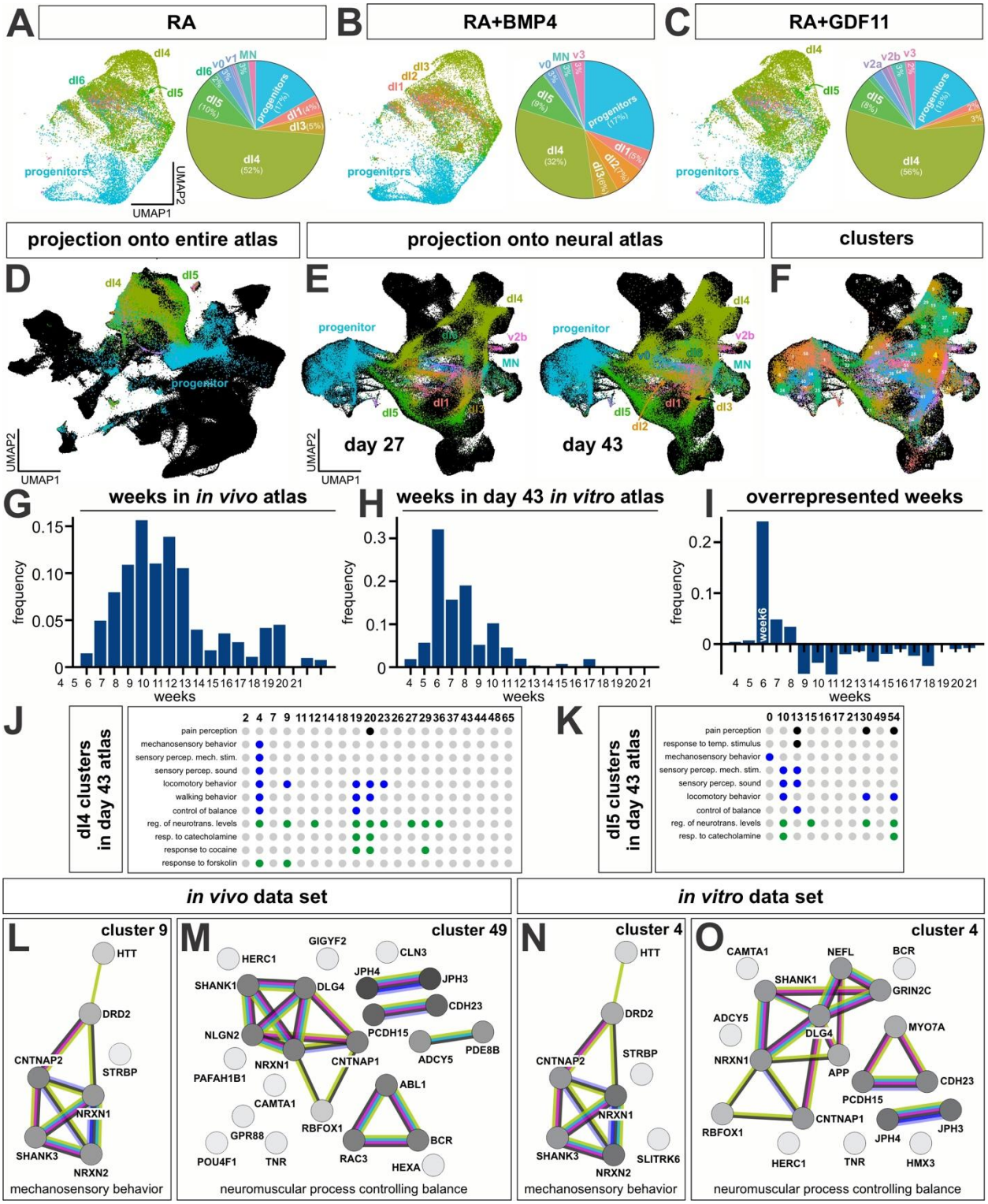

Figure 4-6: Comparison of *in vitro* data with the *in vivo* reference atlas reveals ASD signatures

A-C) UMAPs of day 43 *in vitro*-derived cells generated using the RA, RA+BMP4, and RA+GDF11 (NMP) protocols, using neural spinal cord cell-type identity transferred from the *in vivo* neuronal atlas (Figure 2a). The pie charts summarize the proportions of neuronal cells generated by each protocol. The RA condition predominantly yields dI4-dI6 identities, while the RA+BMP4 protocol shifts patterning toward dI1-dI3 populations. Addition of GDF11 prior to NMP induction does not alter resultant cell-type composition compared to RA alone.

(D) Projection of day 43 *in vitro* cells onto the complete *in vivo* spinal cord reference (shown as a black silhouette). I*n vitro* neural progenitors and neurons map directly onto their *in vivo* counterparts, selectively recapitulating endogenous spinal neural transcriptional identities.

(E, F) Projection of day 27 and 43 *in vitro*-derived cells onto the neuronal subset of the *in vivo* spinal cord atlas. The *in vitro* dIs map along the correct lineage trajectories (E) suggesting they are transcriptionally indistinguishable from endogenous neurons. The *in vitro* dIs are colored by transferred neuronal cell-type identity (E) or atlas cluster identity (F).

(G-I) The day 43 *in vitro* datasets (G) align most closely with week 6 samples in the *in vivo* reference atlas (H). Using the relative frequency of the developmental weeks in the *in vivo* atlas as a reference, we assessed for overrepresentation and found that week 6 profiles were enriched in the *in vitro* dataset compared with the *in vivo* atlas (I).

(J, K) Gene Ontology (GO) enrichment analysis was performed on significantly upregulated genes (avg_log2FC > 0.25) for any stem cell derived clusters assigned to a dI4 or dI5 atlas cluster, as performed previously for the *in vivo* dI4/dI5 clusters. Dot plots of enriched GO Biological Process terms in the dI4 (J) and dI5 (K) clusters include behavioral categories that

overlap with those observed in the in vivo datasets (Figure 2e, f). Black dots are related to the sensory perception of pain and temperature, blue dots correspond to sensory modalities including touch, sound, movement and balance, while green dots represent a response to chemical stimulants.

(L-O) To assess the extent to which *in vitro* and *in vivo* dIs may be functionally similar, the genes making up each GO term were subjected to a STRING analysis, thereby identifying the protein interaction networks that encode specific sensory behavioral modalities. These networks include a mechanosensory circuit derived from *in vivo* cluster 9 (L), as well as a balance regulating circuit derived from *in vivo* cluster 49 (M), include proteins with variants that was strongly associated with autism spectrum disorder: NLGN2, NRXN1, SHANK1/3, DLG4 and CNTNAP2. These circuits were recapitulated in the *in vitro* cluster 4 mechanosensory behavior protein interaction network (N) and the *in vitro* cluster 4 balance regulating protein interaction network (O).

For STRING networks (L-O), nodes represent proteins, edges represent protein-protein associations. Edge color indicates known interactions; experimentally determined (purple), from curated databases (teal), predicted interactions; gene neighborhood (green), gene fusions (red), gene co-occurrence (blue) and others; textmining (yellow), co-expression (black), protein homology (periwinkle)

Chapter 5 – Conclusions

By leveraging the resolution afforded by *in vitro* differentiation systems, we have defined the distinct contributions of two downstream BMP effectors, Smad1 and Smad5, at key decision points along the dorsal interneuron specification timeline. These findings carry implications that extend well beyond the developing spinal cord and advance our understanding of how a single signaling pathway is capable of directing diverse cellular outcomes during development.

Our chapter 3 findings add a new layer of complexity to an ongoing debate in the field regarding the respective roles of Smad1 and Smad5 during dorsal spinal cord development. We observe that Smad5 exerts a dominant and reiterative influence over dorsal progenitor identity, governing dP1–dP3 specification, and dI1–dI3 differentiation, while Smad1 plays a more circumscribed role, primarily restraining dP1 expansion. These are not interchangeable activities and cannot easily be reconciled with a simple redundancy model.

A separate and compelling model, proposed by Le Dréau and colleagues examining the developing chicken spinal cord, posits that it is the combined strength of SMAD1/5 activity, rather than the identity of the individual R-Smad, that dictates the mode of neural stem cell division[27]. While our findings are not necessarily incompatible with the concept of signal strength as a regulatory variable, they raise the possibility that what has been interpreted as a quantitative signal-strength effect may in fact reflect the qualitatively distinct contributions of Smad1 and Smad5 acting on different cellular processes at different developmental stages.

Interestingly, in both mouse and chicken embryos Smad1 and Smad5 have differential expression patterns[20, 25].This spatial and temporal segregation of expression suggests that the two R-Smads are poised to act in fundamentally different cellular contexts, consistent with the distinct phenotypes we observe upon their individual loss. Together, the divergent expression profiles and the non-overlapping functional requirements for Smad1 and Smad5 point toward a

model in which BMP signaling diversity is achieved, at least in part, through the deployment of distinct R-Smad effectors rather than through uniform activation of a shared second messenger complex.

Furthermore, BMP-Smad signaling is increasingly recognized as a driver of pathological processes in multiple disease contexts[236-238]. For this reason, therapeutic strategies targeting Smad1 and Smad5 have emerged as a means to counteract disease-associated BMP-Smad dysregulation[236, 239]. This therapeutic avenue is compounded by the question of functional redundancy among R-Smads. If Smad1 and Smad5 are assumed to be interchangeable, as their shared biochemical activities and overlapping target genes might suggest, then a drug designed to inhibit one may inadvertently spare or compensate through the other, with unanticipated consequences. Conversely, if the two R-Smads perform distinct functions, as our work demonstrates in the context of neural differentiation, then targeting one may selectively disrupt biological processes that the other cannot rescue. Therefore, continuing to define the extent to which Smad1 and Smad5 act redundantly versus distinctly is a prerequisite for developing safe and effective BMP pathway therapeutics.

Taken together, the body of work presented in this dissertation demonstrates that *in vitro* differentiation platforms are powerful tools for dissecting downstream signaling mechanisms that are refractory to *in vivo* interrogation. The mESC and hESC-based differentiation systems we have developed and validated here not only provide a reproducible and scalable source of dorsal spinal interneurons with translational potential for sensory circuit repair, but also serve as a generalizable experimental framework for probing how developmental signals are decoded. As BMP pathway components continue to emerge as drivers of human disease and candidates for pharmacological intervention, the imperative to understand how individual downstream effectors contribute to pathway output, and where their functions diverge, will only grow.

Appendix:

Star Protocols: Protocol for inducing dorsal spinal sensory interneurons from mouse embryonic stem cell-derived neuromesodermal progenitors.

Summary

Here, we present a protocol comprising two differentiation techniques to derive dorsal sensory spinal interneurons (dIs) from mouse embryonic stem cells (mESCs) through a neuromesodermal progenitor (NMP) precursor. First, we describe the steps for using retinoic acid (RA) to induce dI4, dI5 and dI6s which mediate pain, itch, touch and sensorimotor integration. Second, we detail the procedures for generating dI1, dI2 and dI3s, which mediate proprioception and mechanosensation, by using RA together with bone morphogenetic protein (BMP) 4.

For complete details on the use and execution of this protocol, please refer to Gupta et al (2022).

Background information: For further scientific rationale about the development of these protocols please refer to Gouti et al. 2014[54], Andrews et al., 2017[44], Gupta et al., 2018[43], Gupta et al. 2021[22] and Gupta et al., 2022[140]. Briefly, these studies describe the first generation of directed differentiation protocols to derive mouse[44] and human[22, 43] dorsal interneurons (dIs) via neural ectodermal stem cell derivatives, the development of a neuromesodermal (NMP) intermediate stem cell protocol for the ventral spinal cord[54], and then the second generation of protocols to derive mouse dIs[140] via NMP stem cell derivatives.

Innovation: This improved protocol describes a more robust and biologically faithful method for generating *bona fide* dIs from mouse embryonic stem cells (mESCs) via an NMP intermediate state. NMPs are bipotent progenitor cells that endogenously generate the spinal cord during embryogenesis[192]. Thus, directing stem cells to differentiate along an NMP to dI trajectory recapitulates the developmental pathway that generates the spinal cord *in vivo*. It differs from our

previous approaches[22, 43, 44], which induced dIs from anterior neuroectoderm which was subsequently posteriorized and dorsalized. This first generation of protocols resulted in dIs with predominantly upper cervical/brachial identities and limited subtype diversity. In contrast, the current protocols first posteriorize stem cells to induce NMPs, and then neuralize/dorsalize them to generate dIs. This differentiation strategy generates all six classes of dIs with the correct transcription factor codes and sensory-specific functional modules. These cells are also transcriptionally indistinguishable from their *in vivo* embryonic counterparts[140]. A further innovation is the use of a two-dimensional (2d) differentiation platform rather than 3d embryoid body systems. The 2d format ensures uniform patterning, facilitates straightforward immunostaining, and enables scalable, and the reproducible assessment of dI identities. The homogeneity also makes the system suitable for high-throughput drug screening and mechanistic studies of spinal cord biology. Overall, this protocol offers a rapid, reproducible, and developmentally faithful strategy for generating authentic spinal sensory dIs. Importantly, the same principles can be applied to human pluripotent stem cells, offering a translational foundation for regenerative therapies focused on sensory regeneration.

Timing: [2 days – 2 weeks]

1. Maintain pluripotent mESC cell line on mouse embryonic fibroblast (MEF)-supported conditions for at least two passages post-thaw before beginning differentiation.
2. Passage mESCs onto a 0.1% gelatin coated plates 1-2 days before planned start date.
3. Obtain reagents listed in the key resource table.
4. Make stock solutions at the correct working concentrations.
5. Media should be prepared on demand, sterile filtered, and stored at 2-8ºC for up to two weeks.

CRITICAL: All procedures are performed in a BSL-2 certified class II type A2 biosafety cabinets. Cultures are grown and maintained at 37°C, in incubators supplemented with 5% CO_2.

Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Goat polyclonal IgG anti-Brachyury (1:500)	R&D systems	AF2085 RRID:AB_2200235
Rabbit monoclonal IgG anti-Cdx2 (1:500)	Cell Signaling Technology	D11D10
Goat polyclonal IgG anti-Isl1 (1:500)	R&D systems	AF1837 RRID:AB_2126324
Mouse monoclonal $mIgG_{2a}$ anti-Lhx2 (1:50)	DSHB	PCRP-LHX2-1C11
Rabbit polyclonal IgG anti-Pax2 (1:500)	Invitrogen	71-6000 RRID:AB_2533990
Mouse monoclonal IgG_{2a} anti-Pax3 (1:500)	R&D Systems	MAB2457 RRID:AB_2159398
Rabbit polyclonal IgG anti-Pax6 (1:500)	Proteintech	12323-1-AP
Goat polyclonal IgG anti-Sox1 (1:500)	Santa Cruz	sc-17318 RRID:AB_2195365
Mouse monoclonal IgG anti-Sox2 (1:500)	Santa Cruz	sc-365823 RRID:AB_10842165
Rabbit polyclonal IgG anti-Tubulin β-3 (1:1000)	Biolegend	802001
Oligonucleotides		
Gapdh F-GGCCTTCCGTGTTCCTAC R-TGTCATCATACTTGGCAGGTT	Gupta et al.[140]	n/a
Lhx2 F-CAGCTTGCGCAAAAGACC R-TAAAAGGTTGCGCCTGAACT	Gupta et al.[140]	n/a

Foxd3 F-CCCCAACACTGACCAACAG R-GTTTGCTCCGCCAGCTTA	Gupta et al.[140]	n/a
Isl1 F-ATCGAGTGTTTCCGCTGTGT R-ATGGGTTCGGCTGCCATTT	Gupta et al.[140]	n/a
Lhx5 F-CGGAGCAAAGTCTTCCACCT R-GGACGACACCGAGTTGAGAC	Gupta et al.[140]	n/a
Lbx1 F-GCCGAAGGCCGCGATG R-TGCGTGATTTTCGCCGTTTC	Gupta et al.[140]	n/a
Lmx1b F-GCATCAAGATGGAGGAGCAC R-CTCGTTGACTCGCATCAGG	Gupta et al.[140]	n/a
NeuN F-ATCGTAGAGGGACGGAAAATTGA R-GTTCCCAGGCTTCTTATTGGTC	Gupta et al.[140]	n/a
Tubb3 F-ATGGACAGTGTTCGGTCTGG R-AGCACCACTCTGACCAAAGAT	Gupta et al.[140]	n/a
Sox2 F-CAAAAACCGTGATGCCGACT R-CGCCCTCAGGTTTTCTCTGT	Gupta et al.[140]	n/a
Pax3 F-AAACCCAAGCAGGTGACAAC R-GCGTCCTTGAGCAATTTGTC	Gupta et al.[140]	n/a
Pax7 F-TGGCCAAACTGCTGTTGATT R-TAGGCTTGTCCCGTTTCCAC	Gupta et al.[140]	n/a
Olig3 F-ACGTTAGGGCTGTTCCAAGG R-GGCAAGGGGGTAGGATGAAC	Gupta et al.[140]	n/a
Chemicals, peptides, and recombinant proteins		
0.1% Gelatin in water	STEMCELL Technologies	07903

0.25% Trypsin-EDTA	Thermo Fisher Scientific	25200-056
2-Mercaptoethanol	Thermo Fisher Scientific	21985023
B-27 Supplement (50x)	Thermo Fisher Scientific	17504044
BD Pharmigen DAPI solution	BD Biosciences	564907
Bovine serum albumin (BSA)	Cell Signaling Technology	9998S
CHIR 99021	Tocris	4423
Dimethyl sulfoxide (DMSO)	MilliporeSigma	D2438
Distilled water	Thermo Fisher Scientific	10977023
DMEM high glucose	Cytiva	SH30022.02
DMEM-F12 1:1	Cytiva	SH30023.01
DPBS, no calcium, no magnesium (1x)	Thermo Fisher Scientific	14190-250
EmbryoMax nucleosides (100x)	Sigma-Aldrich	ES-008-D
Embryonic stem cell-qualified fetal bovine serum (FBS)	Thermo Fisher Scientific	16141079
Ethyl alcohol, pure	MilliporeSigma	E7023
GlutaMAX supplement (100x)	Thermo Fisher Scientific	35-050-061

Heat inactivated horse serum (HIHS)	Thermo Fisher Scientific	26050070
Human basic FGF recombinant protein (bFGF)	Thermo Fisher Scientific	PHG0023
Hydrochloric acid	Thermo Fisher Scientific	A144-212
Human BMP4 recombinant protein	Thermo Fisher Scientific	PHC9534
Matrigel Matrix	Corning	354277
Mouse LIF (mLIF)	MilliporeSigma	ESG1107
N-2 Supplement (100x)	Thermo Fisher Scientific	17502-048
Neurobasal medium	Thermo Fisher Scientific	21103-049
Paraformaldehyde (PFA) 16%	Thermo Fisher Scientific	NC1537886
Penicillin-streptomycin 100x (10,000 U/mL)	Thermo Fisher Scientific	15140-122
Retinoic acid (all trans)	Sigma-Aldrich	R2625
Sodium azide	Sigma-Aldrich	71289
Triton X-100	Thermo Fisher Scientific	BP151
Trypan blue solution 0.4%	Thermo Fisher Scientific	15250061
Trypsin inhibitor (1x)	Sigma Aldrich	T7659

Experimental models: cell lines		
MM13 mouse embryonic stem cell line	Wichterle et al, 2002[46]	-
Mouse Embryonic Fibroblasts (MEFs)	Thermo Fisher Scientific	A34181
Commercial assays		
RNeasy Mini Kit	Qiagen	74106
Superscript IV First-Strand Synthesis System	Thermo Fisher Scientific	18091050
Roche Diagnostics LIGHTCYCLER 480 SYBR Green	Thermo Fisher Scientific	50-720-3233
Software and algorithms		
ImageJ	NIH	https://imagej.net/ij/
Other		
24-well CellBIND flat bottom plates	Corning	3337
Adhesive PCR plate seals	Thermo Fisher Scientific	AB0558
Axygen LightCycler 480 PCR plates	Thermo Fisher Scientific	14-223-390
Centrifuge Tubes with CentriStar Cap (15mL conical)	Corning	05-538-59B
Cell scraper	Thermo Fisher Scientific	07000553

Cryogenic tubes (cryovials)	Thermo Fisher Scientific	12-565-171N
LightCycler 480 Instrument II	Roche	05015243001
LSM800 inverted confocal microscope	Zeiss	-
Mastercycler nexus GX2	Eppendorf	6336000015
Microcentrifuge 1.5mL tubes	Thermo Fisher Scientific	05-408-129
Nanodrop One Spectrophotometer	Thermo Fisher Scientific	13-400-525
Stericup Quick Release-GP Sterile Vacuum Filtration	MilliporeSigma	S2GPU05RE
T25 flask	Corning	10-126-28
Parafilm	Thermo Fisher Scientific	13-374

Materials and equipment setup

Growth factor/cytokine stock solutions

Retinoic acid (RA) 100mM stock solution

Reagent	**Final concentration**	**Amount**
Retinoic acid	100mM	50mg
DMSO	n/a	1.66mL
Total	n/a	1.66mL

Critical: Retinoic acid (RA) is a light sensitive compound. Work under low-light conditions when working with RA. To make a less concentrated solution, dilute RA stock in 100% ethanol. The

100 mM stock can be stored at −80°C for up to 2 years. The diluted stock can be stored at 4°C in the dark and must be used within a week.

Critical: Retinoic acid (RA) is a potent teratogen. Dispose concentrated stock solutions responsibly according to the manufacturer's recommendations.

Bone morphogenetic protein 4 (BMP4) stock solution (10μg/mL)

Reagent	**Final concentration**	**Amount**
BMP4	10μg/mL	10μg
4mM HCl with 0.1% BSA in DPBS	n/a	1 mL
Total	**n/a**	**1mL**

Critical: Make 25 μL aliquots and avoid multiple freeze-thaw cycle. Aliquots can be stored at −20°C for up to 1 year.

CHIR 99021 stock solution (10mM)

Reagent	**Final concentration**	**Amount**
CHIR 99021	10mM	10mg
DMSO	n/a	2.15 mL
Total	**n/a**	**2.15mL**

Critical: Solid can be stored at -20ºC for 6 months. The 10mM stock solution can be stored at -20ºC for up to one month.

Basic fibroblast growth factor (bFGF) stock solution (100μg/mL)

Reagent	**Final concentration**	**Amount**
bFGF	100μg/mL	1mg
Distilled water	n/a	2mL

0.1% BSA in DPBS	n/a	8mL
Total	n/a	**10mL**

Critical: bFGF is unstable at 37ºC. Once reconstituted with carrier protein, such as 0.1% BSA, aliquots can be stored at -20ºC for up to one year. Avoid excessive freeze-thaw cycles.

Note: Initially reconstitute bFGF in distilled water then dilute further with 0.1% BSA in DPBS.

Medium recipes

MEF medium

Reagent	**Final concentration**	**Amount**
DMEM High Glucose	n/a	220mL
FBS	10% (v/v)	25mL
GlutaMAX supplement (100x)	1x	2.5mL
Penicillin-streptomycin (100x)	1x	2.5mL
Total	n/a	**250mL**

Critical: Store medium at 2-8ºC for up to two weeks.

mESC medium

Reagent	**Final concentration**	**Amount**
DMEM High Glucose	n/a	204.5mL
FBS	15% (v/v)	37.5mL
GlutaMAX supplement (100x)	1x	2.5mL

Penicillin-streptomycin (100x)	1x	2.5mL
Nucleosides (100x)	1x	2.5mL
2-Mercaptoethanol (55mM stock)	0.1mM	0.455mL
Total	n/a	**250mL**

Critical: Store medium at 2-8ºC for up to two weeks. Warm 100x nucleosides before use.

Critical: mESC medium is specific for mouse stem cells and must be supplemented with mLIF (10,000x stock solution), immediately before use.

N2/B27 medium

Reagent	**Final concentration**	**Amount**
DMEM F12 1:1	n/a	240mL
Neurobasal medium	n/a	240mL
N-2 supplement (100x)	0.5x	2.5mL
B-27 (50x)	0.5x	5mL
GlutaMAX (100x)	1x	5mL
Penicillin-streptomycin (100x)	1x	5mL
20% BSA	0.08% (w/v)	2mL
2-Mercaptoethanol (55mM)	0.1mM	0.91mL
Total	n/a	**500mL**

Note: 20% BSA can be made by weighing out 5 grams of BSA and mixing with DMEM F12 1:1 to reach a final volume of 25mL. Vortex or mix until fully dissolved. Store 2-8ºC for up to two weeks.

Critical: Store N2/B27 medium at 2-8ºC for up to two weeks. Be sure to use B-27 supplement that contains vitamin A.

Antibody blocking solution

Reagent	**Final concentration**	**Amount**
DPBS	n/a	489.5mL
HIHS	1% (v/v)	5mL
Triton X-100	0.1% (v/v)	0.5mL
5% Sodium azide	0.05% (v/v)	5mL
Total	n/a	**500mL**

Critical: Store antibody blocking solution at 2-8ºC for up to one month.

Note: A stock solution of 5% (w/v) sodium azide is initially prepared in distilled water. Sodium azide is toxic and should be handled with care in a fume hood using proper PPE. Dispose of sodium azide waste accordingly**.**

STEP-BY-STEP METHOD DETAILS

Culturing MEF feeder-supported mESCs prior to differentiation

Timing: [14 days]

This step first establishes a monolayer of inactivated (irradiated) mouse embryonic fibroblasts (MEFs), as a feeder layer that supports the healthy culture of mESCs prior to differentiation. mESCs should be passaged at least twice onto feeder monolayers when a confluency of about 70% is reached, prior to beginning the differentiation.

Note: The thawing of cells stored in liquid nitrogen should be performed both carefully, i.e. wearing insulated gloves to retrieve the cells given that liquid nitrogen rapidly burns on contact and swiftly for improved cell viability. The cells should be thawed as quickly as possible, i.e., <2 minutes with the entire process from step 2 to starting step 4 taking no more than 10 minutes.

Note: MEFs should be thawed onto 25cm^2 T25 flasks 1 day prior to mESC thawing/passaging. MEFs are naturally adherent cell types, thus no coating with any matrix (such as gelatin) is required for MEF plating.

1) Prepare 9mL of warm MEF medium in a 15mL centrifuge tube prior to thawing MEFs.
2) Thaw one 1mL vial of gamma irradiated or mitomycin C inactivated MEFs (containing 1x10^6 cells/vial) by warming in a 37ºC water bath until there is a diminishing, but still present, pellet of frozen cells.
    a) Use 1mL of the warm MEF medium aliquoted in step 1 to fully thaw the remainder of the frozen cell pellet.

Note: One vial of MEFs is sufficient for two T25 flasks.

3) Transfer MEFs from the cryovial to the 15mL tube in step 1, which will now contain 10mL medium.

Note: To reduce cryo-preservation damage and improve cell survival, thawed cells should be diluted 10-fold in warm MEF medium.

4) Centrifuge for 5 minutes at 200 x g to pellet cells.
5) Aspirate supernatant to remove freezing medium that contains DMSO.
6) Gently resuspend cell pellet into 10mL warm MEF medium.

Note: Seeding density of MEFs should be determined by user. Here, 5x10^5 cells are used to seed one T25 flask**.**

7) Dispense 5mL of cell suspension into each T25 flask. This step prepares two T25 flasks for the first plating of mESCs and subsequent passage of mESCs.
8) Place at 37ºC in an incubator supplemented with 5% CO_2.
    a) Move flasks in a star pattern to evenly distribute MEFs.

Next day

9) On the next day, aliquot 16mL of mESC medium into a 50mL conical tube and supplement with mLIF (10,000x stock), diluted to 1x, i.e., 5µL of mLIF is required for 50ml of mESC media.

Critical: Add mLIF to the mESC medium immediately before use, to prevent the mLIF from degrading over time. mLIF in media is stable in incubator conditions for up to one week. Any medium supplemented with mLIF must be stored at 2-8°C after use.

Note: One vial of mESCs containing $1x10^6$ cells can be used to seed one T25 flask with a pre-prepared MEF feeder monolayer.

10) Follow steps 2-5 as described above, using mESC medium to thaw mESCs.
11) Resuspend mESC cell pellet in 5mL of mESC medium.
12) Aspirate the medium from one T25 flask containing MEFs and replace with the 5mL of the mESC cell resuspension.
13) Feed mESCs daily with mESC medium containing mLIF until they reach 70% confluency (about 3-5 days).

Note: At this point, passage the cells onto the second MEF containing T25 flask prepared in step 7. Alternatively, prepare T25 flasks with a MEF feeder monolayer the day before planning to split.

Note: Avoid colony overcrowding, which can cause the loss of pluripotency and increase the chance of spontaneous differentiation.

14) Passage mESCs1:4 onto monolayers of MEFs using trypsin-EDTA 0.25% as follows:

a) Rinse mESC-containing flask to be passaged with 5mL of DPBS.

b) Aspirate DPBS and add 3mL of 0.25% trypsin-EDTA.

c) Place in a 37ºC incubator for 2-3 minutes or until cells begin to lift off.

Critical: Incubation time in 0.25% trypsin-EDTA should be kept to a minimum and no longer than 5 minutes as this can result in reduced cell viability.

d) Triturate trypsin-cell mixture to generate a single cell suspension.

e) Inactivate the trypsin reaction with 5mL of mESC medium (contains 15% FBS) or alternatively with 1:1 volume of trypsin-EDTA inhibitor.

f) Collect the cell resuspension in a 15mL tube and centrifuge 200 x g for 5 minutes.

g) Aspirate supernatant and resuspend the cell pellet in 1mL of mESC medium.

h) Replace the MEF medium in the pre-prepared MEF T25 flask with 5mL mESC medium.

i) Seed the T25 flask with 250µL of the mESC cell resuspension for a 1:4 split.

j) Place flask into 37ºC incubator and replace medium daily

Note: Healthy mESC colonies have round, smooth and well-defined borders and are highly nucleated (Figure 2A). Spontaneous differentiation is rare and can be distinguished by eye by individual cells budding off from the ESC colonies; these cells have flatter morphology and separate themselves from the colony[240].

15) When the mESCs reach 70% confluency after the second passage, they are ready to differentiate. Passage mESCs onto gelatinized T25 flask at a ratio of 1:2 and allow to proliferate for 1-2 days.

Note: This step prepares mESCs for the differentiation procedure and reduces the transfer of MEFs by plating onto a gelatin substrate.

a) Coat a T25 flask with 3mL of 0.1% gelatin in water and leave in biosafety cabinet (20-25ºC) for ≥20 minutes.
b) Proceed with passaging protocol (see step 14 a-g).
c) Before seeding the flask, aspirate the gelatin solution and let dry for 1-2 minutes.
d) Dispense 5mL of mESC medium into the gelatin coated flask.
e) Seed with 500µL of the mESC cell resuspension for a 1:2 split.
f) Place into a 37ºC incubator supplemented with 5% CO_2 and assess confluency 1-2 days later.

Note: Alternatively, two rounds of MEF panning can be performed[241], which is a differential adhesion protocol used to separate mESCs from feeder MEFs.

Two-dimensional differentiation procedure to generate NMP-derived dIs

Timing: [9 days]

This procedure directs the differentiation of mESCs first into an NMP state using bFGF/CHIR (day 3, Figure 3A). The NMPs are then directed towards different populations of dIs (day 9, Figure 3A) using RA to induce primarily dI4-dI6 and RA+BMP4 to induce dI1-dI3. Directing mESCs first towards NMPs, rather than a neuro-ectodermal intermediate as in previous directed differentiation protocols for dIs[42-44], better mimics the endogenous developmental trajectory of the spinal cord.

Day 0

16) Coat one (or the number required) 24-well CELLBIND plate with 0.5mL of 0.1% gelatin or Matrigel/well.

Critical: Always keep Matrigel at 4ºC during the procedure. Anything that will contact Matrigel must be pre-chilled to prevent rapid polymerization/gelling. Once surface(s) have been coated with Matrigel leave in incubator at least 1 hour and use the same day.

Note: The dilution factor for Matrigel matrix is lot-specific and needs to be determined by user. It will be provided on the lot-specific certificate of analysis document provided by the manufacturer. In this series of experiments, Matrigel (Lot 17124003) was used at a ratio of 1:89.3 and diluted with N2/B27 medium. These guidelines provide more information on Matrigel handling.

Note: This protocol uses 24-well plates, but other sizes of plates can also be used if the seeding density is then adjusted appropriately.

Note: The gelatin coating works best within 16 hours of coating, while Matrigel coated plates can be stored by sealing the plate with parafilm and keeping it at 4ºC for 1-2 weeks.

Note: One T25 flask of ~70% confluency mESCs cultured on 0.1% gelatin yields ≥4x10^6 cells.

17) Warm N2/B27 medium and add bFGF to a final concentration of 10ng/mL.

Note: The bFGF stock is 100µg/mL and should be added to the media at a 1:10,000 dilution to reach a concentration of 10ng/mL.

Critical: Avoid prolonged warming (>30 minutes) of this medium before use, given that bFGF is unstable at 37ºC.

18) Perform the trypsin dissociation described in step 14a-f.

19) Aspirate supernatant and resuspend cell pellet in 1mL of warm N2/B27 medium containing bFGF.

20) Proceed with cell counting, with the goal of seeding each well with 10,000 cells.

a) To count cells using a hemocytometer, serially dilute the sample to 1:100.

i) Aliquot 90µL of N2/B27 medium into two 1.5mL tubes.

ii) Perform a 1:10 dilution by adding 10µL from the mESC resuspension to the first tube and mixing by pipetting.

iii) Take 10µL from the 1:10 dilution, add to the second 1.5mL tube and mix by pipetting for a 1:100 dilution.

Note: Use this second solution for the cell count.

Critical: Cells will settle at the bottom of the tube if left undisturbed, so be sure to mix gently before proceeding with any pipetting step. Performing a serial dilution increases the accuracy of cell counting.

Note: Cell viability can also be performed during cell counting using Trypan blue. In this case, first combine 45µL of the N2/B27 medium with 45µL of Trypan blue, to give the 90µL medium to start the serial dilution. Trypan blue stains dead cells blue. Cellular viability should be ≥95%.

b) Use the following formula for cell counting in a standard hemocytometer (0.1mm depth or 10^{-4}mL):

$$\text{cell concentration } \left(\frac{\text{cells}}{\text{mL}}\right) = \frac{\text{average cell count of the corner squares x dilution factor (100)}}{0.0001\text{mL}}$$

c) Use the following formula to calculate the amount of the 1mL cell resuspension to add into a given volume of media (here, 13mL is chosen to account for any pipetting errors) to reach the desired cell concentration (here, 20,000 cells/mL):

$$\text{volume (mL)} = \frac{\text{desired media volume (13mL) x desired cell concentration } \left(\frac{20{,}000 \text{ cells}}{\text{mL}}\right)}{\text{cell concentration } \left(\frac{4\text{x}10^6 \text{cells}}{\text{mL}}\right)}$$

Note: The four corners of a standard hemocytometer are comprised of a 4x4 grid of squares. For each corner, count all the cells in the entire grid (n=4), and then average them. Further explanation of this formula, and a useful calculator can be found here.

21) Use the information from the cell count to dilute the 1mL mESC suspension with N2/B27 medium containing bFGF to a concentration of 20,000 cells/mL.

Note: This working solution permits the distribution of 10,000 cells/well in a volume of 0.5mL.

22) Aspirate gelatin (or Matrigel) solution and replace with 0.5mL of the mESC suspension.

Note: To ensure uniform seeding across wells make sure to mix gently, but well before each pull.

Critical: Maintain a 24-hour interval between feedings to ensure the correct timing of exposure to growth factor/small molecules.

Day 1 (Figure 1A)

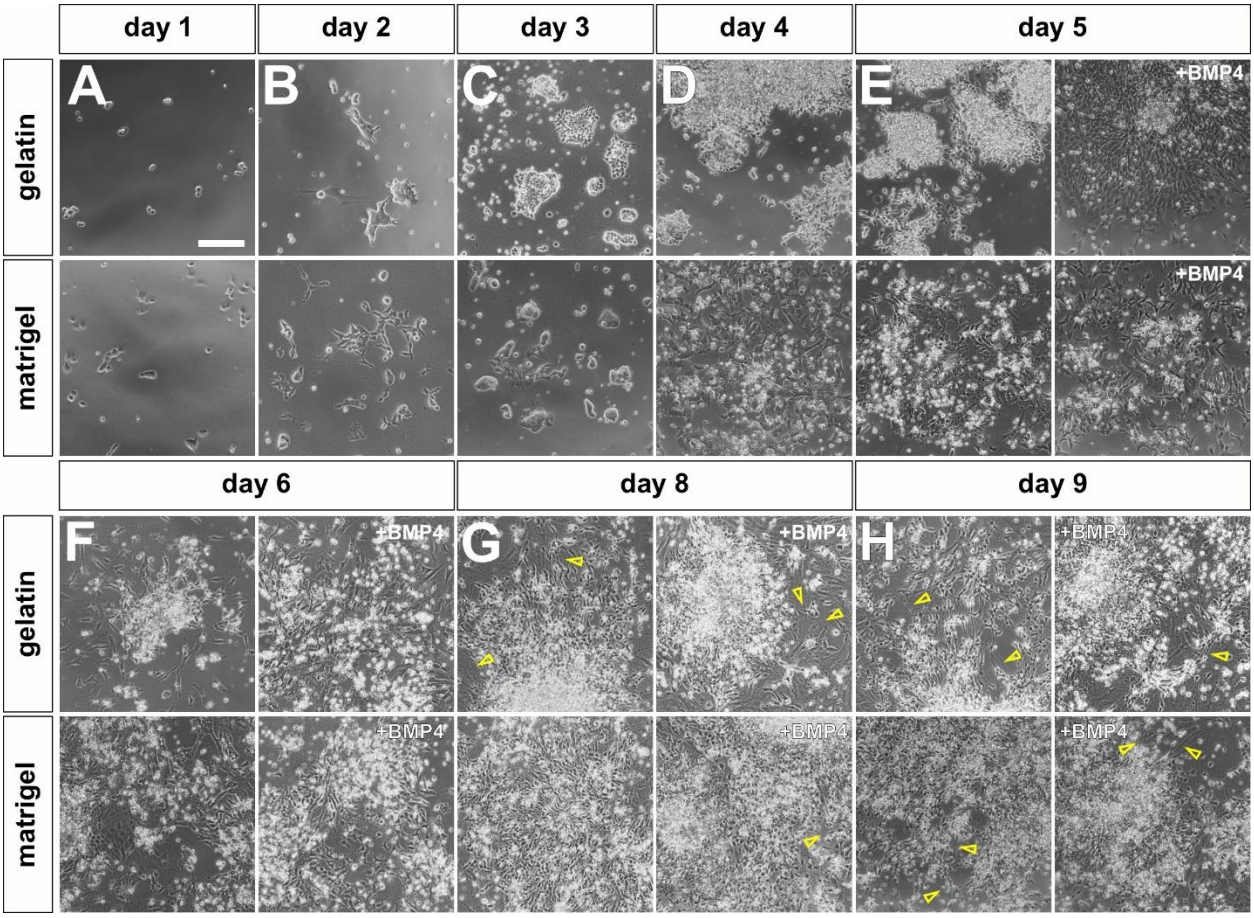

Figure 1 Images of cell morphologies from along the differentiation timeline

(A–H) Phase contrast images of cells grown on either 0.1% gelatin, or Matrigel at day 1 (A), day 2 (B), day 3 (C), day 4 (D), day 5 (E), day 6 (F), day 8 (G) and day 9 (H). Note that mESCs initially form small clones on the gelatin at the beginning of the differentiation which rapidly expand when treated with FGF/CHIR (A–C). However, as differentiation progresses (day 4/5 onwards), the cells change shape becoming flatter, elongated and pyramidal in shape, and begin extending neuronal processes starting from day 7 in both the RA and RA+BMP4 conditions (arrowheads, G, H). The abundant small, bright round cells are either cells that have detached from the gelatin substrate, or floating dead cells and can be removed from the culture when passaging. Culturing the cells either on gelatin vs Matrigel, or in RA vs RA+BMP4 does not markedly change cellular morphology. Scale bar: 100μm.

23) Observe cells. No medium change is required.

Day 2 (Figure 1B)

24) Replace medium with N2/B27 supplemented with 10ng/mL of bFGF and 5μM of CHIR 99021.

Note: Addition of the Wnt agonist CHIR is necessary for cells to be directed towards the NMP fate (Figure 2).

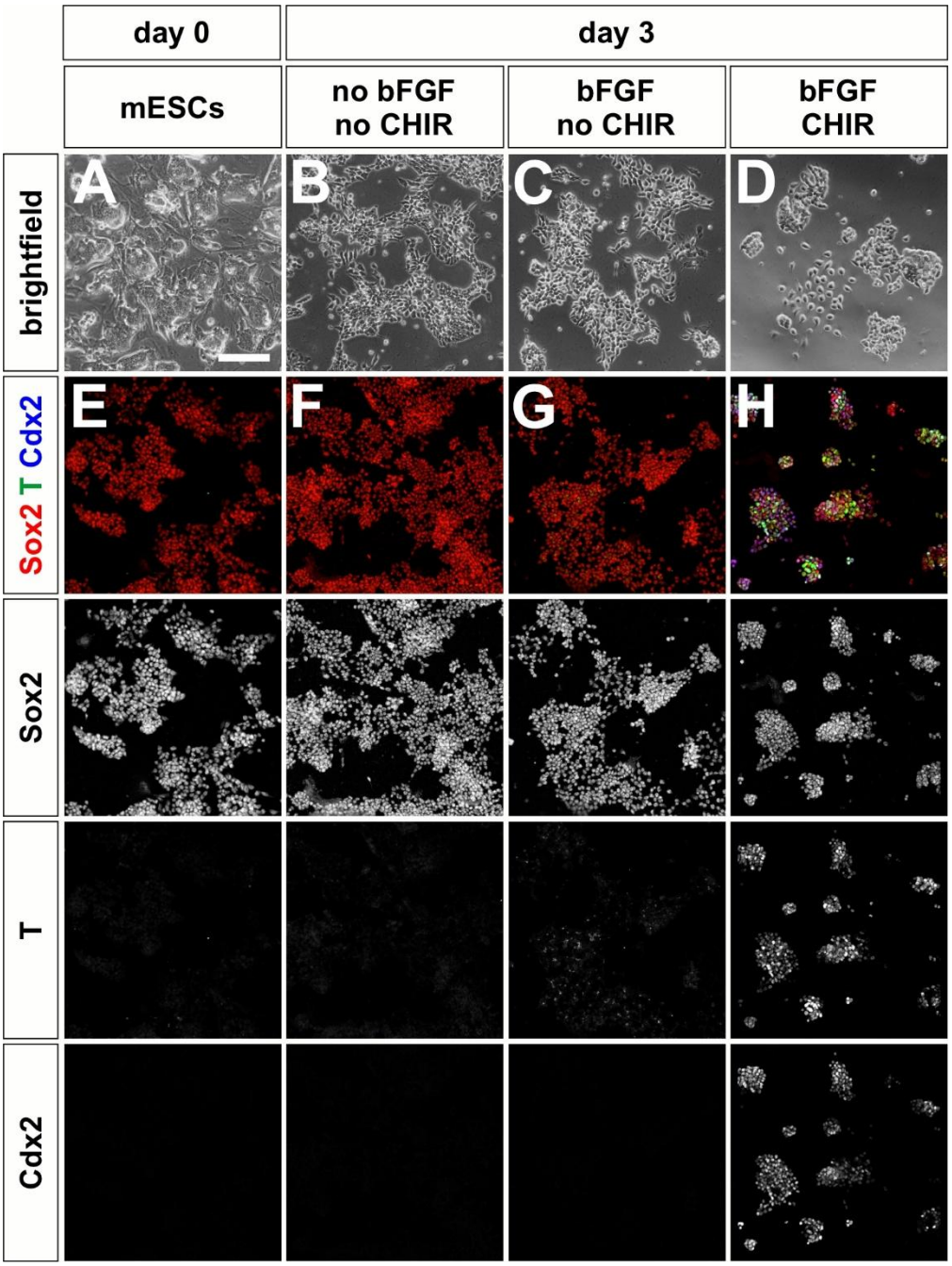

Figure 2 NMP formation requires CHIR addition

(A–D) Healthy mESC cultures have colonies with smooth edges. In contrast, cells cultured in N2/B27 medium with or without FGF/CHIR appear as nucleated, flat cell clones (B and C). NMPs are recognized by presence of small clones containing nucleated cells (D).

(E–H) Day 3 mESC cultures lacking CHIR addition only express Sox2, similar to day 0 mESCs, demonstrating they have not progressed to the NMP state. In contrast, day 3 mESC cultures that have been treated with both bFGF and CHIR express both neural and mesodermal (Brachyury (T)/Cdx2) markers. Scale bar: 50μm.

Day 3 (Figure 1C)

25) Feed cells with N2/B27 medium supplemented with 100nM RA.
    a) To reach a final concentration of 100nM RA, prepare a working stock of 1mM RA by diluting 100mM RA 100-fold in pure ethanol.
    b) 1mM RA stock is used at a final concentration of 100nM in the media (1:10,000 from 1mM stock).

Note: RA is light sensitive. Turn off the light in the biosafety cabinet for this diluting step.

Day 4 (Figure 1D)

26) To induce the dI1, dI2 and dI3 fates, feed the cells with N2/B27 medium supplemented 100nM RA + 10ng/mL BMP4. To induce the dI4-dI6 fates, continue feeding with N2/B27 medium supplemented with 100nM RA.

Note: The BMP4 stock is 10μg/mL and so should be added at 1:1,000 dilution in the media to reach a concentration of 10ng/mL.

Day 5-8 (Figure 1E-1G)

27) Feed both conditions (RA or RA + BMP4) with N2/B27 (no growth factors/small molecules added) medium every other day or daily if the media appears exhausted, i.e., yellow in color.

Note: DPBS rinses can be performed in between feedings to remove excessive cellular waste.

Day 9 (Figure 1H)

28) On the 9th day of the differentiation, cells can either be fixed directly in the well for immunocytochemistry (step 28) or processed for RNA analysis (step 29).

Quality control strategies

Two approaches can be used to confirm the gene expression profile of differentiating cultures. First, the different populations of dIs can be distinguished by immunohistochemistry (IHC) using antibodies directed against the canonical transcription factors specific to dI subtypes[136]. Second, relative gene expression can be assessed using quantitative (q) RT-PCR analysis using dI marker-specific primers (see key resources table).

29) Immunocytochemical analyses of protein localization

Note: The cultures should be physically manipulated with care, e.g., gentle pipetting along the wall of the well, to not disrupt the neuronal processes.

Critical: The reagents used to fix the cells should be at the same temperature to reduce any large fluctuations in temperatures that can damage neuronal processes.

Note: Steps a-c below should be performed on ice (4ºC) or on temperature-controlled surface maintained at 2-8ºC.

a) Rinse the well(s) to be stained with DPBS.
b) Replace the DPBS with freshly made cold 4% PFA (16% PFA stock diluted in DPBS to 4%), and fix for 8-10 minutes.

Note: The PFA fixation time can be adjusted if high non-specific background is observed in the IHC.

c) Perform three, 5-minute washes with DPBS.

Note: Proceed with antibody staining or store plates at 2-8ºC in DPBS, sealed with Parafilm to prevent the wells from drying out. Plates should be processed for staining within one week otherwise the cells will begin detaching from the plate.

d) Replace final DPBS wash with antibody blocking solution for 1 hour at 2-8ºC.
e) Add species appropriate primary antibodies diluted in antibody blocking solution to the well(s) and store at 2-8ºC for 12-14 hours (overnight).
f) Wash two times with 0.1% Triton X-100 in DPBS (PBT) for 5 minutes.
g) Add corresponding species appropriate secondary antibodies in antibody blocking solution to well (s) and incubate for 1 hour in the dark at 20ºC (room temperature).

Note: Plates can be kept in the dark, by wrapping them in aluminum foil or storing them in a closed drawer.

h) Wash once with PBT for 5 minutes.
i) Wash again for 15 minutes, adding DAPI (1:500) to the PBT solution.

j) Replace with DPBS and proceed with imaging.

k) If needed, store plates after imaging at 2-8ºC.

30) qPCR analyses of RNA expression, using Qiagen RNeasy mini kit

Note: The RNeasy handbook/protocol for purifying total RNA from lysates can be found here. Follow the instructions for “Purification of Total RNA from Animal Cells”, which is on pages 15-20 in the 2020 handbook.

a) Rinse wells to be collected for RNA analysis with DPBS.

b) Collect and lyse cell samples using 350µL of the RLT buffer provided in the Qiagen RNeasy kit. A cell scraper or pipette tip can be used to detach adherent cultures.

c) Proceed with total RNA purification or flash freeze and store at -80ºC.

Note: For each biological replicate, i.e., directed differentiation procedure, collect 3 technical replicates from 3 separate wells of each given condition.

d) Following the RNeasy kit instructions, elute RNA in 30µL of water once.

Note: Samples can be flash frozen in liquid nitrogen and stored at -80ºC. If the user does not have access to liquid nitrogen, place the samples at -80ºC as quickly as possible upon completion of the purification step.

Note: If desired, the purified total RNA can be sent for bulk RNA-Sequencing.

Critical: All samples should be assessed using a microvolume spectrophotometer, to permit cDNA to be made from the same concentration (ng/µL) of template RNA. This step reduces variability in threshold cycle (Ct) values especially when comparing to a housekeeping gene using the ΔΔCT method.

e) Following the RNeasy kit instructions, proceed with cDNA synthesis.

f) Generate cDNA with the SuperScript IV first strand synthesis kit using this protocol (Refer to pages 3-4).

Note: cDNA synthesis is made using Oligo (dT) primers. The optional step of removing RNA is also performed.

g) Perform a 10µL qPCR reaction per well in a 384-well plate, using 7.5µL qPCR master mix (Table 1), and 2.5µL cDNA, diluted 20-fold.

Table 1. qPCR setup and cycling conditions

qPCR reaction master mix

Reagent	**Amount for 1 reaction/well**
Distilled water	1.5µL
5µM oligo mix for gene of interest (forward and reverse primers)	1.0µL
SYBR green	5.0µL
Total	7.5µL

qPCR cycling conditions

Steps	**Temperature**	**Time**	**Cycles**
Initial denaturation	92 °C	2 min	1
Denaturation	92 °C	15 sec	40 cycles
Annealing	60 °C	1 min	
Extension	72 °C	1 min	
Hold	4 °C	Indefinitely	

Note: To prepare newly synthesized oligonucleotides, resuspend them in water to a final concentration of 100μM. Prepare a combined forward and reverse primer mix by diluting each primer 1:20, to yield a 5μM working stock. From this working stock, use 1.0μL per qPCR reaction resulting in 0.5μM final concentration per primer.

h) Analyze qPCR results using the ΔΔCT method[242].

Expected outcomes

After 3 days in culture with bFGF/CHIR, mESCs start to express bipotential NMP markers (Sox2, Brachyury (T), Cdx2) (Figure 2H). CHIR is required for this step; in the absence of CHIR, there is no initiation of mesodermal marker (T, Cdx2) expression (Figure 2F, 2G).

The addition of RA at day 3 both directs NMP towards both neural and dorsal spinal fates. By day 7, many cells express Pax3 and Pax6 (Figure 3C, 3D), which connate a general dorsal progenitor (dP) identity. The cells have also started to terminally differentiate as $Sox1^+$ neurons, which extend $Tuj1^+$ neuronal processes (Figure 3E-3H). Neurite extension starts in cultures as early as day 7 and becomes progressively more robust by day 8 (arrowheads, Figures 1G,3F) and day 9 (arrowheads, Figures 1G,3F).

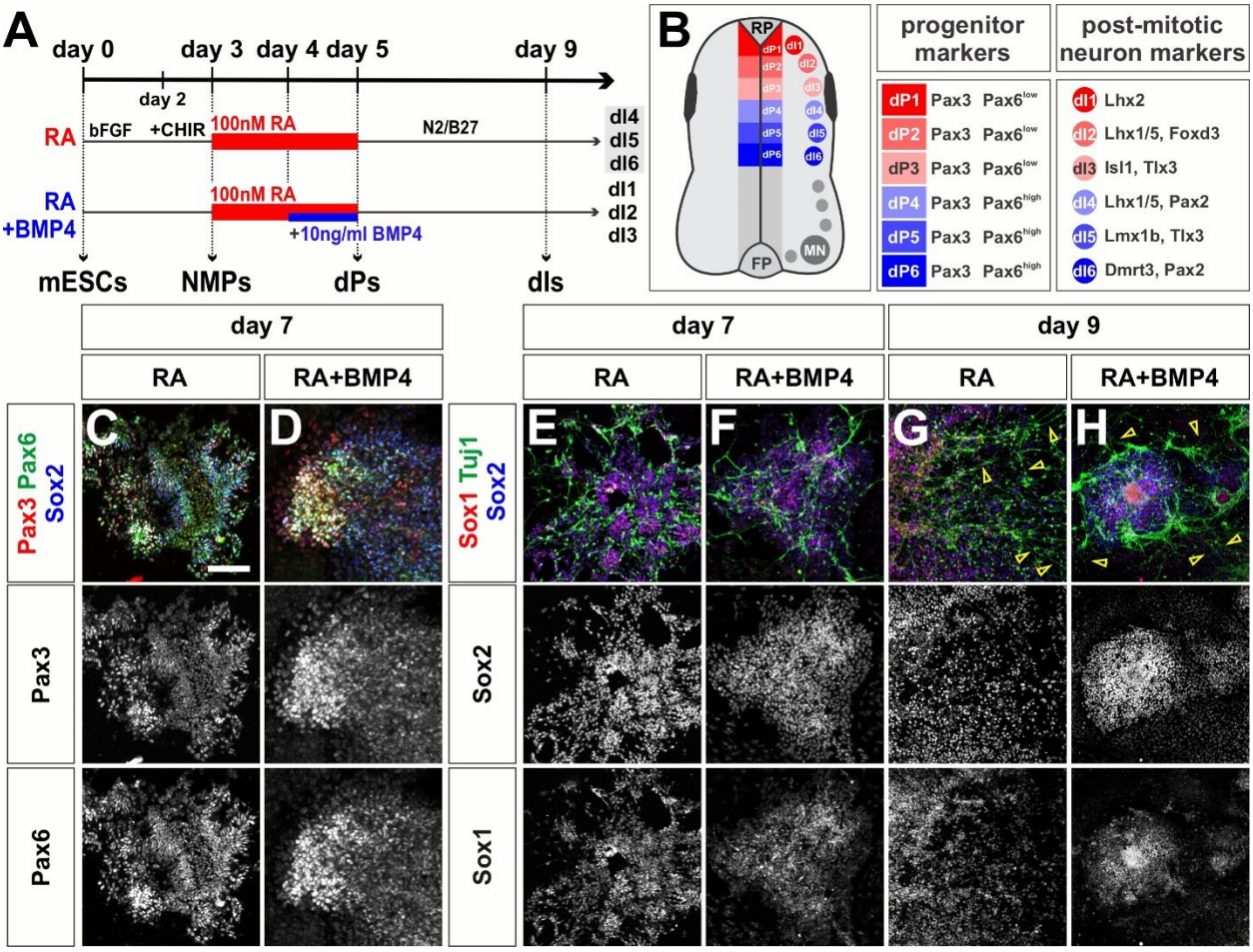

Figure 3 Immunostaining to detect dorsal spinal progenitors and neurons

(A) Schematic timeline of the RA±BMP4 mESC protocols.

(B) Schematic depicting the location of the dorsal progenitor (dP) domains and the dIs, and some canonical nuclear markers. For a complete atlas of markers, see.[11,13]

(C and D) By day 7, both RA±BMP4 protocols are making Pax3^{+} Pax6^{+} dorsal spinal tissue.

(E–H) By day 7, the RA±BMP4 protocols are both making Sox1^{+} neurons, that are extending Tuj1^{+} neural process. The Tuj1^{+} process become more extensive by day 9 (arrowheads, G, H).

Scale bar: 50μm (C and D); 100μm (E, F, G).

By day 9, cultures treated with RA will be enriched for intermediate dorsal identities, i.e., Pax2^{+} dI4, Lmx1b^{+} dI5 and Dmrt3^{+} Pax2^{+} dI6s (Figure 3B, Figure 4). Our single cell (sc) RNA-Seq

analyses have suggested that the neural populations are up to 30-40% dI4 and dI5 and ~7% dI6s[140, 194]. In contrast, cultures treated with RA+BMP4 will be enriched for the dorsal-most dorsal identities, i.e., Lhx2^{+} dI1, Foxd3^{+} dI2 and Isl1^{+} dI2s (Figure 3B, Figure 4). scRNA-Seq analyses suggest conversion rates of 45% dI1, 30% dI2 and ~8% dI3s[140, 194]. Note that dI2 and dI6s are challenging to detect by immunocytochemistry. Complete lists of canonical and novel markers for spinal cord analyses can be found in Gupta et al., 2024[194] and Alaynick et al., 2011[136].

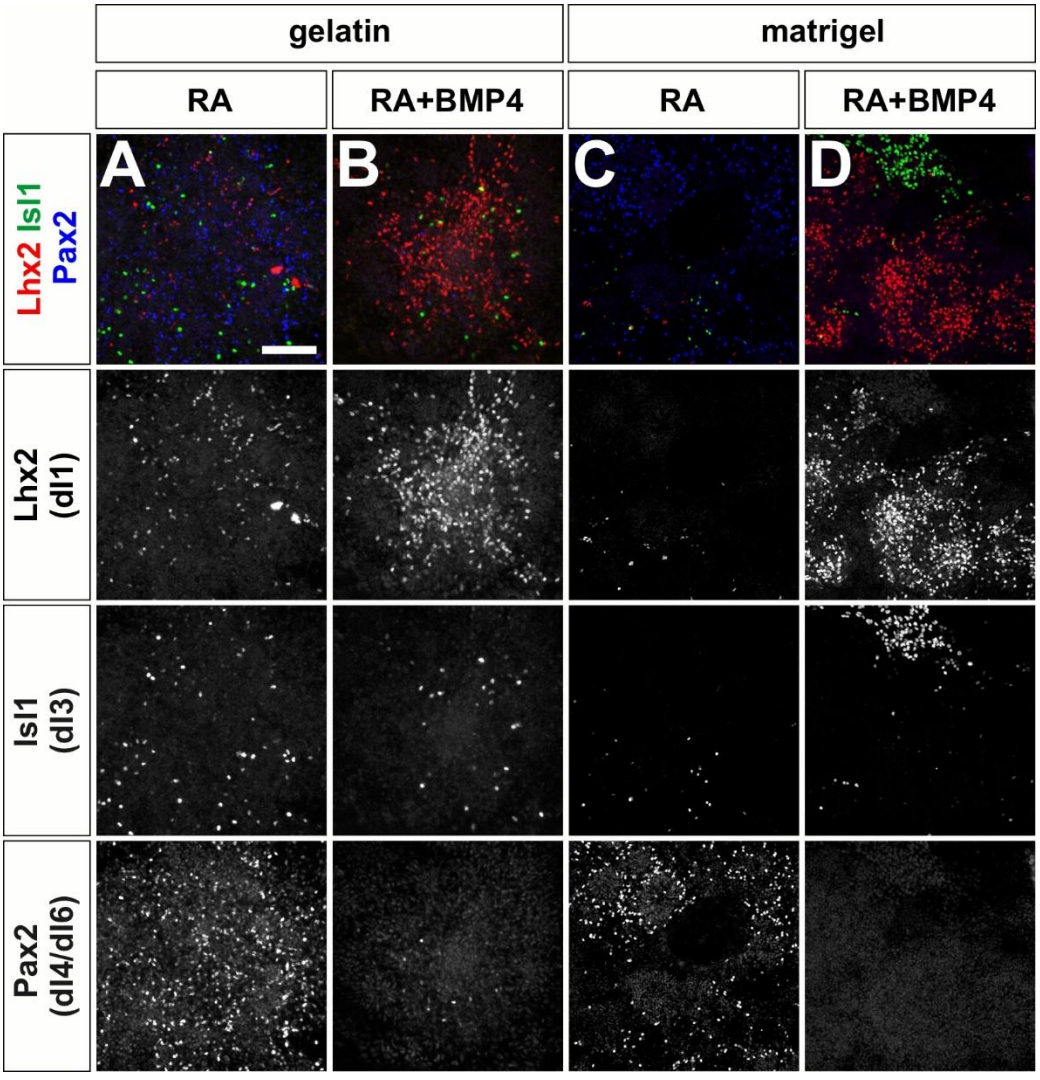

Figure 4 Immunostaining to detect different dI populations

(A and C) By day 9, the cultures derived from the RA protocol will contain a mixture of Pax2^{+} Lhx1/5^{+} dI4/dI6s and Lmx1b^{+} dI5s (not shown).

(B and C) In contrast, the cells derived from the RA+BMP4 contain a mixture of Lhx2^{+} dI1s and Isl1^{+} dI3s (B and D). dI2s are present in these cultures but are most effectively detected by scRNA-Seq.[1] Scale bar: 50μm.

Quantification and statistical analysis

The conversion efficiency of NMP-derived dIs can be assessed using IHC, by comparing the number of positive cells for a canonical nuclear marker of dI identity[136] (Figure 4) to number of the DAPI^{+} cells, i.e. total number of cells, and expressed as a percentage. The quantification of each nuclear stain can be automated in ImageJ using thresholding.

Bibliography

1. Kobayashi T, Lyons KM, McMahon AP, Kronenberg HM. BMP signaling stimulates cellular differentiation at multiple steps during cartilage development. Proceedings of the National Academy of Sciences. 2005;102(50):18023–7.

2. Hemmati-Brivanlou A, Thomsen GH. Ventral mesodermal patterning in Xenopus embryos: Expression patterns and activities of BMP-2 and BMP-4. Developmental Genetics. 1995;17(1):78–89.

3. Hogan BL. Bone morphogenetic proteins: multifunctional regulators of vertebrate development. Genes Dev. 1996;10(13):1580–94.

4. Riaz Z, Hussain M, Parveen S, Sultana M, Saeed S, Ishaque U, et al. In Silico Analysis: Genome-Wide Identification, Characterization and Evolutionary Adaptations of Bone Morphogenetic Protein (BMP) Gene Family in Homo sapiens. Molecular Biotechnology. 2024;66(11):3336–56.

5. Wrana JL, Attisano L, Wieser R, Ventura F, Massagué J. Mechanism of activation of the TGF-beta receptor. Nature. 1994;370(6488):341–7.

6. Zhang Y, Feng X, We R, Derynck R. Receptor-associated Mad homologues synergize as effectors of the TGF-beta response. Nature. 1996;383(6596):168–72.

7. Yamauchi K, Varadarajan SG, Li JE, Butler SJ. Type Ib BMP receptors mediate the rate of commissural axon extension through inhibition of cofilin activity. Development (Cambridge, England). 2013;140(2):333–42.

8. Lebrun JJ, Takabe K, Chen Y, Vale W. Roles of pathway-specific and inhibitory Smads in activin receptor signaling. Mol Endocrinol. 1999;13(1):15–23.

9. Stanley E, Biben C, Kotecha S, Fabri L, Tajbakhsh S, Wang CC, et al. DAN is a secreted glycoprotein related to Xenopus cerberus. Mechanisms of Development. 1998;77(2):173–84.

10. Larraín J, Bachiller D, Lu B, Agius E, Piccolo S, De Robertis EM. BMP-binding modules in chordin: a model for signalling regulation in the extracellular space. Development. 2000;127(4):821–30.

11. Sekiya T, Adachi S, Kohu K, Yamada T, Higuchi O, Furukawa Y, et al. Identification of BMP and activin membrane-bound inhibitor (BAMBI), an inhibitor of transforming growth factor-beta signaling, as a target of the beta-catenin pathway in colorectal tumor cells. J Biol Chem. 2004;279(8):6840–6.

12. Murakami G, Watabe T, Takaoka K, Miyazono K, Imamura T. Cooperative inhibition of bone morphogenetic protein signaling by Smurf1 and inhibitory Smads. Mol Biol Cell. 2003;14(7):2809–17.

13. Shi Y, Wang YF, Jayaraman L, Yang H, Massagué J, Pavletich NP. Crystal structure of a Smad MH1 domain bound to DNA: Insights on DNA binding in TGF-β signaling. Cell. 1998;94(5):585–94.

14. Weiss A, Charbonnier E, Ellertsdóttir E, Tsirigos A, Wolf C, Schuh R, et al. A conserved activation element in BMP signaling during Drosophila development. Nat Struct Mol Biol. 2010;17(1):69–76.

15. Budi EH, Duan D, Derynck R. Transforming Growth Factor-β Receptors and Smads: Regulatory Complexity and Functional Versatility. Trends in Cell Biology. 2017;27(9):658–72.

16. Yoon SJ, Wills AE, Chuong E, Gupta R, Baker JC. HEB and E2A function as SMAD/FOXH1 cofactors. Genes and Development. 2011;25(15):1654–61.

17. Koinuma D, Tsutsumi S, Kamimura N, Imamura T, Aburatani H, Miyazono K. Promoter-wide analysis of Smad4 binding sites in human epithelial cells. Cancer Science. 2009;100(11):2133–42.

18. Trompouki E, Bowman TV, Lawton LN, Fan ZP, Wu DC, Dibiase A, et al. Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. Cell. 2011;147(3):577–89.

19. Attisano L, Lee-Hoeflich ST. The Smads. Genome Biol. 2001;2(8):Reviews3010.

20. Hazen VM, Andrews MG, Umans L, Crenshaw EB, Zwijsen A, Butler SJ. BMP receptor-activated Smads confer diverse functions during the development of the dorsal spinal cord. Developmental Biology. 2012;367(2):216–27.

21. Islam SS, Mokhtari RB, Kumar S, Maalouf J, Arab S, Yeger H, et al. Spatio-Temporal Distribution of Smads and Role of Smads/TGF-β/BMP-4 in the Regulation of Mouse Bladder Organogenesis. PLOS ONE. 2013;8(4):e61340.

22. Gupta S, Butler SJ. Getting in touch with your senses: Mechanisms specifying sensory interneurons in the dorsal spinal cord. WIREs Mech Dis. 2021;13(5):e1520.

23. Lee KJ, Dietrich P, Jessell TM. Genetic ablation reveals that the roof plate is essential for dorsal interneuron specification. Nature. 2000;403(6771):734–40.

24. Lee KJ, Mendelsohn M, Jessell TM. Neuronal pattering by BMPs: a requirement for GDF7 in the generation of discrete class of commisural interneurons in the mouse spinal cord. Genes and Development. 1998;12:3394–407.

25. le Dréau G, Garcia-Campmany L, Angeles Rabadán M, Ferronha T, Tozer S, Briscoe J, et al. Canonical BMP7 activity is required for the generation of discrete neuronal populations in the dorsal spinal cord. Development. 2012;139(2):259–68.

26. Wei CY, Wang HP, Zhu ZY, Sun YH. Transcriptional factors Smad1 and Smad9 act redundantly to mediate zebrafish ventral specification downstream of Smad5. Journal of Biological Chemistry. 2014;289(10):6604–18.

27. Le Dréau G, Saade M, Gutiérrez-Vallejo I, Martí E. The strength of SMAD1/5 activity determines the mode of stem cell division in the developing spinal cord. Journal of Cell Biology. 2014;204(4):591–605.

28. Ionta S, Villiger M, Jutzeler CR, Freund P, Curt A, Gassert R. Spinal cord injury affects the interplay between visual and sensorimotor representations of the body. Sci Rep. 2016;6:20144.

29. Lenggenhager B, Pazzaglia M, Scivoletto G, Molinari M, Aglioti SM. The sense of the body in individuals with spinal cord injury. PLoS One. 2012;7(11):e50757.

30. Furusawa K, Tokuhiro A, Sugiyama H, Ikeda A, Tajima F, Genda E, et al. Incidence of symptomatic autonomic dysreflexia varies according to the bowel and bladder management techniques in patients with spinal cord injury. Spinal Cord. 2011;49(1):49–54.

31. Antonic A, Sena ES, Lees JS, Wills TE, Skeers P, Batchelor PE, et al. Stem cell transplantation in traumatic spinal cord injury: a systematic review and meta-analysis of animal studies. PLoS Biol. 2013;11(12):e1001738.

32. Kong D, Feng B, Amponsah AE, He J, Guo R, Liu B, et al. hiPSC-derived NSCs effectively promote the functional recovery of acute spinal cord injury in mice. Stem Cell Res Ther. 2021;12(1):172.

33. Faravelli I, Bucchia M, Rinchetti P, Nizzardo M, Simone C, Frattini E, et al. Motor neuron derivation from human embryonic and induced pluripotent stem cells: experimental approaches and clinical perspectives. Stem Cell Res Ther. 2014;5(4):87.

34. Trawczynski M, Liu G, David BT, Fessler RG. Restoring Motor Neurons in Spinal Cord Injury With Induced Pluripotent Stem Cells. Front Cell Neurosci. 2019;13:369.

35. Lai HC, Seal RP, Johnson JE. Making sense out of spinal cord somatosensory development. Development. 2016;143(19):3434–48.

36. Gupta S, Butler SJ. Getting in touch with your senses: Mechanisms specifying sensory interneurons in the dorsal spinal cord. WIREs Mech Dis. 2021;13(5):e1520.

37. Lai HC, Klisch TJ, Roberts R, Zoghbi HY, Johnson JE. In vivo neuronal subtype-specific targets of Atoh1 (Math1) in dorsal spinal cord. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2011;31(30):10859–71.

38. Bui TV, Akay T, Loubani O, Hnasko TS, Jessell TM, Brownstone RM. Circuits for grasping: spinal dI3 interneurons mediate cutaneous control of motor behavior. Neuron. 2013;78(1):191–204.

39. Haimson B, Hadas Y, Kania A, Daley M, Cinnamon Y, Lev-Tov A, et al. A subpopulation of spinocerebellar tract neurons regulates the stability of bipedal stepping. bioRxiv. 2020:2020.01.07.898072.

40. Perry S, Larhammar M, Vieillard J, Nagaraja C, Hilscher MM, Tafreshiha A, et al. Characterization of Dmrt3-Derived Neurons Suggest a Role within Locomotor Circuits. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2019;39(10):1771–82.

41. Koch SC, Acton D, Goulding M. Spinal Circuits for Touch, Pain, and Itch. Annu Rev Physiol. 2018;80:189–217.

42. Gupta S, Sivalingam D, Hain S, Makkar C, Sosa E, Clark A, et al. Deriving Dorsal Spinal Sensory Interneurons from Human Pluripotent Stem Cells. Stem cell reports. 2018;10(2):390–405.

43. Gupta S, Yamauchi K, Novitch BG, Butler SJ. Derivation of dorsal spinal sensory interneurons from human pluripotent stem cells. STAR Protoc. 2021;2(1):100319.

44. Andrews MG, Del Castillo LM, Ochoa-Bolton E, Yamauchi K, Smogorzewski J, Butler SJ. BMPs direct sensory interneuron identity in the developing spinal cord using signal-specific not morphogenic activities. Elife. 2017;6.

45. Andrews MG, Kong J, Novitch BG, Butler SJ. New perspectives on the mechanisms establishing the dorsal-ventral axis of the spinal cord. Current topics in developmental biology. 2019;132:417–50.

46. Wichterle H, Lieberam I, Porter JA, Jessell TM. Directed differentiation of embryonic stem cells into motor neurons. Cell. 2002;110(3):385–97.

47. Duval N, Vaslin C, Barata TC, Frarma Y, Contremoulins V, Baudin X, et al. BMP4 patterns Smad activity and generates stereotyped cell fate organization in spinal organoids. Development. 2019;146(14).

48. Ladd AN, Yatskievych TA, Antin PB. Regulation of avian cardiac myogenesis by activin/TGFbeta and bone morphogenetic proteins. Dev Biol. 1998;204(2):407–19.

49. Kattman SJ, Witty AD, Gagliardi M, Dubois NC, Niapour M, Hotta A, et al. Stage-specific optimization of activin/nodal and BMP signaling promotes cardiac differentiation of mouse and human pluripotent stem cell lines. Cell Stem Cell. 2011;8(2):228–40.

50. Kawaguchi J, Mee PJ, Smith AG. Osteogenic and chondrogenic differentiation of embryonic stem cells in response to specific growth factors. Bone. 2005;36(5):758–69.

51. Ille F, Atanasoski S, Falk S, Ittner LM, Märki D, Büchmann-Møller S, et al. Wnt/BMP signal integration regulates the balance between proliferation and differentiation of neuroepithelial cells in the dorsal spinal cord. Dev Biol. 2007;304(1):394–408.

52. Hazen VM, Andrews MG, Umans L, Crenshaw EB, 3rd, Zwijsen A, Butler SJ. BMP receptor-activated Smads confer diverse functions during the development of the dorsal spinal cord. Dev Biol. 2012;367(2):216–27.

53. Le Dreau G, Garcia-Campmany L, Rabadan MA, Ferronha T, Tozer S, Briscoe J, et al. Canonical BMP7 activity is required for the generation of discrete neuronal populations in the dorsal spinal cord. Development. 2012;139(2):259–68.

54. Gouti M, Tsakiridis A, Wymeersch FJ, Huang Y, Kleinjung J, Wilson V, et al. In vitro generation of neuromesodermal progenitors reveals distinct roles for wnt signalling in the specification of spinal cord and paraxial mesoderm identity. PLoS Biol. 2014;12(8):e1001937.

55. Sagner A, Gaber ZB, Delile J, Kong JH, Rousso DL, Pearson CA, et al. Olig2 and Hes regulatory dynamics during motor neuron differentiation revealed by single cell transcriptomics. PLoS Biol. 2018;16(2):e2003127.

56. Rodrigo Albors A, Halley PA, Storey KG. Lineage tracing of axial progenitors using Nkx1-2CreER(T2) mice defines their trunk and tail contributions. Development (Cambridge, England). 2018;145(19).

57. Delile J, Rayon T, Melchionda M, Edwards A, Briscoe J, Sagner A. Single cell transcriptomics reveals spatial and temporal dynamics of gene expression in the developing mouse spinal cord. Development. 2019;146(12).

58. Wilson L, Gale E, Chambers D, Maden M. Retinoic acid and the control of dorsoventral patterning in the avian spinal cord. Dev Biol. 2004;269(2):433–46.

59. Son K, Yu S, Shin W, Han K, Kang K. A Simple Guideline to Assess the Characteristics of RNA-Seq Data. Biomed Res Int. 2018;2018:2906292.

60. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.

61. Farhadian M, Rafat SA, Panahi B, Mayack C. Weighted gene co-expression network analysis identifies modules and functionally enriched pathways in the lactation process. Sci Rep. 2021;11(1):2367.

62. Panahi B, Farhadian M, Hejazi MA. Systems biology approach identifies functional modules and regulatory hubs related to secondary metabolites accumulation after transition from autotrophic to heterotrophic growth condition in microalgae. PLoS One. 2020;15(2):e0225677.

63. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013;14:128.

64. Webb S, Qayyum SR, Anderson RH, Lamers WH, Richardson MK. Septation and separation within the outflow tract of the developing heart. J Anat. 2003;202(4):327–42.

65. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol. 2018.

66. Vincentz JW, Firulli BA, Lin A, Spicer DB, Howard MJ, Firulli AB. Twist1 controls a cell-specification switch governing cell fate decisions within the cardiac neural crest. PLoS Genet. 2013;9(3):e1003405.

67. Soldatov R, Kaucka M, Kastriti ME, Petersen J, Chontorotzea T, Englmaier L, et al. Spatiotemporal structure of cell fate decisions in murine neural crest. Science. 2019;364(6444).

68. Han X, Feng J, Guo T, Loh YE, Yuan Y, Ho TV, et al. Runx2-Twist1 interaction coordinates cranial neural crest guidance of soft palate myogenesis. Elife. 2021;10.

69. Andersson LS, Larhammar M, Memic F, Wootz H, Schwochow D, Rubin CJ, et al. Mutations in DMRT3 affect locomotion in horses and spinal circuit function in mice. Nature. 2012;488(7413):642–6.

70. Pillai A, Mansouri A, Behringer R, Westphal H, Goulding M. Lhx1 and Lhx5 maintain the inhibitory-neurotransmitter status of interneurons in the dorsal spinal cord. Development. 2007;134(2):357–66.

71. Cheng L, Arata A, Mizuguchi R, Qian Y, Karunaratne A, Gray PA, et al. Tlx3 and Tlx1 are post-mitotic selector genes determining glutamatergic over GABAergic cell fates. Nat Neurosci. 2004;7(5):510–7.

72. Ilicic T, Kim JK, Kolodziejczyk AA, Bagger FO, McCarthy DJ, Marioni JC, et al. Classification of low quality cells from single-cell RNA-seq data. Genome Biol. 2016;17:29.

73. Fernandes EC, Luz LL, Mytakhir O, Lukoyanov NV, Szucs P, Safronov BV. Diverse firing properties and Aβ-, Aδ-, and C-afferent inputs of small local circuit neurons in spinal lamina I. Pain. 2016;157(2):475–87.

74. Todd AJ. Identifying functional populations among the interneurons in laminae I-III of the spinal dorsal horn. Mol Pain. 2017;13:1744806917693003.

75. Wercberger R, Basbaum AI. Spinal cord projection neurons: a superficial, and also deep, analysis. Curr Opin Physiol. 2019;11:109–15.

76. Yuengert R, Hori K, Kibodeaux EE, McClellan JX, Morales JE, Huang TP, et al. Origin of a Non-Clarke's Column Division of the Dorsal Spinocerebellar Tract and the Role of Caudal Proprioceptive Neurons in Motor Function. Cell Rep. 2015;13(6):1258–71.

77. Sakai N, Insolera R, Sillitoe RV, Shi SH, Kaprielian Z. Axon sorting within the spinal cord marginal zone via Robo-mediated inhibition of N-cadherin controls spinocerebellar tract formation. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2012;32(44):15377–87.

78. Todd AJ. Neuronal circuitry for pain processing in the dorsal horn. Nat Rev Neurosci. 2010;11(12):823–36.

79. Huang dW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44–57.

80. Huang dW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009;37(1):1–13.

81. Boll S, Almeida de Minas AC, Raftogianni A, Herpertz SC, Grinevich V. Oxytocin and Pain Perception: From Animal Models to Human Research. Neuroscience. 2018;387:149–61.

82. Liu Q, Sikand P, Ma C, Tang Z, Han L, Li Z, et al. Mechanisms of itch evoked by β-alanine. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2012;32(42):14532–7.

83. Lipman ZM, Yosipovitch G. Substance use disorders and chronic itch. J Am Acad Dermatol. 2021;84(1):148–55.

84. Downey LA, Tysse B, Ford TC, Samuels AC, Wilson RP, Parrott AC. Psychomotor Tremor and Proprioceptive Control Problems in Current and Former Stimulant Drug Users: An Accelerometer Study of Heavy Users of Amphetamine, MDMA, and Other Recreational Stimulants. J Clin Pharmacol. 2017;57(10):1330–7.

85. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184(13):3573–87.e29.

86. Consortium TM, coordination O, coordination L, processing Oca, sequencing Lpa, analysis Cd, et al. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature. 2018;562(7727):367–72.

87. Russ DE, Cross RBP, Li L, Koch SC, Matson KJE, Yadav A, et al. A harmonized atlas of mouse spinal cord cell types and their spatial organization. Nat Commun. 2021;12(1):5722.

88. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun. 2019;10(1):1523.

89. Hoekstra EJ, von Oerthel L, van der Heide LP, Kouwenhoven WM, Veenvliet JV, Wever I, et al. Lmx1a encodes a rostral set of mesodiencephalic dopaminergic neurons marked by the Wnt/B-catenin signaling activator R-spondin 2. PLoS One. 2013;8(9):e74049.

90. Zhao HJ, Chang HM, Klausen C, Zhu H, Li Y, Leung PCK. Bone morphogenetic protein 2 induces the activation of WNT/β-catenin signaling and human trophoblast invasion through up-regulating BAMBI. Cell Signal. 2020;67:109489.

91. Daneman R, Agalliu D, Zhou L, Kuhnert F, Kuo CJ, Barres BA. Wnt/beta-catenin signaling is required for CNS, but not non-CNS, angiogenesis. Proc Natl Acad Sci U S A. 2009;106(2):641–6.

92. Agalliu D, Takada S, Agalliu I, McMahon AP, Jessell TM. Motor neurons with axial muscle projections specified by Wnt4/5 signaling. Neuron. 2009;61(5):708–20.

93. MacDonald BT, Tamai K, He X. Wnt/beta-catenin signaling: components, mechanisms, and diseases. Dev Cell. 2009;17(1):9–26.

94. Chen B, Dodge ME, Tang W, Lu J, Ma Z, Fan CW, et al. Small molecule-mediated disruption of Wnt-dependent signaling in tissue regeneration and cancer. Nat Chem Biol. 2009;5(2):100–7.

95. Dickinson ME, Krumlauf R, McMahon AP. Evidence for a mitogenic effect of Wnt-1 in the developing mammalian central nervous system. Development. 1994;120(6):1453–71.

96. Megason SG, McMahon AP. A mitogen gradient of dorsal midline Wnts organizes growth in the CNS. Development. 2002;129(9):2087–98.

97. Muroyama Y, Fujihara M, Ikeya M, Kondoh H, Takada S. Wnt signaling plays an essential role in neuronal specification of the dorsal spinal cord. Genes Dev. 2002;16(5):548–53.

98. Alvarez-Medina R, Cayuso J, Okubo T, Takada S, Marti E. Wnt canonical pathway restricts graded Shh/Gli patterning activity through the regulation of Gli3 expression. Development. 2008;135(2):237–47.

99. Crawford TQ, Roelink H. The notch response inhibitor DAPT enhances neuronal differentiation in embryonic stem cell-derived embryoid bodies independently of sonic hedgehog signaling. Dev Dyn. 2007;236(3):886–92.

100. Attardi A, Fulton T, Florescu M, Shah G, Muresan L, Lenz MO, et al. Neuromesodermal progenitors are a conserved source of spinal cord with divergent growth dynamics. Development. 2018;145(21).

101. Tzouanacou E, Wegener A, Wymeersch FJ, Wilson V, Nicolas JF. Redefining the progression of lineage segregations during mammalian embryogenesis by clonal analysis. Dev Cell. 2009;17(3):365–76.

102. Rajagopal J, Stanger BZ. Plasticity in the Adult: How Should the Waddington Diagram Be Applied to Regenerating Tissues? Dev Cell. 2016;36(2):133–7.

103. Waddington CH. The strategy of the genes. George Allen & Unwin Ltd.1957.

104. Duelen R, Sampaolesi M. Stem Cell Technology in Cardiac Regeneration: A Pluripotent Stem Cell Promise. EBioMedicine. 2017;16:30–40.

105. Row RH, Pegg A, Kinney BA, Farr GH, Maves L, Lowell S, et al. BMP and FGF signaling interact to pattern mesoderm by controlling basic helix-loop-helix transcription factor activity. Elife. 2018;7.

106. Sánchez-Guardado L, Irimia M, Sánchez-Arrones L, Burguera D, Rodríguez-Gallardo L, Garcia-Fernández J, et al. Distinct and redundant expression and transcriptional diversity of MEIS gene paralogs during chicken development. Dev Dyn. 2011;240(6):1475–92.

107. Gouti M, Delile J, Stamataki D, Wymeersch FJ, Huang Y, Kleinjung J, et al. A Gene Regulatory Network Balances Neural and Mesoderm Specification during Vertebrate Trunk Development. Dev Cell. 2017;41(3):243–61.e7.

108. Oulad-Abdelghani M, Chazaud C, Bouillet P, Sapin V, Chambon P, Dollé P. Meis2, a novel mouse Pbx-related homeobox gene induced by retinoic acid during differentiation of P19 embryonal carcinoma cells. Dev Dyn. 1997;210(2):173–83.

109. Novitch BG, Wichterle H, Jessell TM, Sockanathan S. A requirement for retinoic acid-mediated transcriptional activation in ventral neural patterning and motor neuron specification. Neuron. 2003;40(1):81–95.

110. Diez del Corral R, Olivera-Martinez I, Goriely A, Gale E, Maden M, Storey K. Opposing FGF and retinoid pathways control ventral neural pattern, neuronal differentiation, and segmentation during body axis extension. Neuron. 2003;40(1):65–79.

111. Schwaller F, Kanellopoulos AH, Fitzgerald M. The developmental emergence of differential brainstem serotonergic control of the sensory spinal cord. Sci Rep. 2017;7(1):2215.

112. Skagerberg G, Lindvall O. Organization of diencephalic dopamine neurones projecting to the spinal cord in the rat. Brain Res. 1985;342(2):340–51.

113. Gautier A, El Ouaraki H, Bazin N, Salam S, Vodjdani G, Bourgoin S, et al. Lentiviral vector-driven inhibition of 5-HT synthesis in B3 bulbo-spinal serotonergic projections - Consequences on nociception, inflammatory and neuropathic pain in rats. Exp Neurol. 2017;288:11–24.

114. Travagli RA, Hermann GE, Browning KN, Rogers RC. Brainstem circuits regulating gastric function. Annu Rev Physiol. 2006;68:279–305.

115. Hou S, DeFinis JH, Daugherty SL, Tang C, Weinberger J, de Groat WC. Deciphering Spinal Endogenous Dopaminergic Mechanisms That Modulate Micturition Reflexes in Rats with Spinal Cord Injury. eNeuro. 2021;8(4).

116. Holmes GM, Blanke EN. Gastrointestinal dysfunction after spinal cord injury. Exp Neurol. 2019;320:113009.

117. Taweel WA, Seyam R. Neurogenic bladder in spinal cord injury patients. Res Rep Urol. 2015;7:85–99.

118. Burstyn-Cohen T, Stanleigh J, Sela-Donenfeld D, Kalcheim C. Canonical Wnt activity regulates trunk neural crest delamination linking BMP/noggin signaling with G1/S transition. Development. 2004;131(21):5327–39.

119. Lengerke C, Schmitt S, Bowman TV, Jang IH, Maouche-Chretien L, McKinney-Freeman S, et al. BMP and Wnt specify hematopoietic fate by activation of the Cdx-Hox pathway. Cell Stem Cell. 2008;2(1):72–82.

120. Zechner D, Muller T, Wende H, Walther I, Taketo MM, Crenshaw EB, 3rd, et al. Bmp and Wnt/beta-catenin signals control expression of the transcription factor Olig3 and the specification of spinal cord neurons. Dev Biol. 2007;303(1):181–90.

121. Shen Q, Wang Y, Dimos JT, Fasano CA, Phoenix TN, Lemischka IR, et al. The timing of cortical neurogenesis is encoded within lineages of individual progenitor cells. Nat Neurosci. 2006;9(6):743–51.

122. Olivos-Cisneros L, Ramírez-Santos J, Gutiérrez-Ospina G. Proliferation rate and differentiation potential are independent during the transition from neurogenesis to gliogenesis in the mouse embryonic spinal cord. IBRO Neurosci Rep. 2021;10:75–82.

123. Abou Chakra M, Isserlin R, Tran TN, Bader GD. Control of tissue development and cell diversity by cell cycle-dependent transcriptional filtering. Elife. 2021;10.

124. Singh AM, Chappell J, Trost R, Lin L, Wang T, Tang J, et al. Cell-cycle control of developmentally regulated transcription factors accounts for heterogeneity in human pluripotent cells. Stem cell reports. 2013;1(6):532–44.

125. Fortin JM, Azari H, Zheng T, Darioosh RP, Schmoll ME, Vedam-Mai V, et al. Transplantation of Defined Populations of Differentiated Human Neural Stem Cell Progeny. Sci Rep. 2016;6:23579.

126. Kawai M, Imaizumi K, Ishikawa M, Shibata S, Shinozaki M, Shibata T, et al. Long-term selective stimulation of transplanted neural stem/progenitor cells for spinal cord injury improves locomotor function. Cell Rep. 2021;37(8):110019.

127. Muller T, Anlag K, Wildner H, Britsch S, Treier M, Birchmeier C. The bHLH factor Olig3 coordinates the specification of dorsal neurons in the spinal cord. Genes Dev. 2005;19(6):733–43.

128. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc. 2009;4(8):1184–91.

129. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. 2019;20(1):296.

130. Woolnough JL, Atwood BL, Liu Z, Zhao R, Giles KE. The Regulation of rRNA Gene Transcription during Directed Differentiation of Human Embryonic Stem Cells. PLoS One. 2016;11(6):e0157276.

131. Watanabe-Susaki K, Takada H, Enomoto K, Miwata K, Ishimine H, Intoh A, et al. Biosynthesis of ribosomal RNA in nucleoli regulates pluripotency and differentiation ability of pluripotent stem cells. Stem Cells. 2014;32(12):3099–111.

132. Wang RN, Green J, Wang Z, Deng Y, Qiao M, Peabody M, et al. Bone Morphogenetic Protein (BMP) signaling in development and human diseases. Genes Dis. 2014;1(1):87–105.

133. Akizu N, Estarás C, Guerrero L, Martí E, Martínez-Balbás MA. H3K27me3 regulates BMP activity in developing spinal cord. Development. 2010;137(17):2915–25.

134. Kim SW, Yoon SJ, Chuong E, Oyolu C, Wills AE, Gupta R, et al. Chromatin and transcriptional signatures for Nodal signaling during endoderm formation in hESCs. Developmental Biology. 2011;357(2):492–504.

135. Augsburger A, Schuchardt A, Hoskins S, Dodd J, Butler S. BMPs as mediators of roof plate repulsion of commissural neurons. Neuron. 1999;24(1):127–41.

136. Alaynick WA, Jessell TM, Pfaff SL. SnapShot: spinal cord development. Cell. 2011;146(1):178– e1.

137. Lee KJ, Dietrich P, Jessell TM. Genetic ablation reveals that the roof plate is essential for dorsal interneuron specification. Nature. 2000;403(6771):734–40.

138. Lee KJ, Mendelsohn M, Jessell TM. 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. 1998;12(21):3394–407.

139. Wine-Lee L, Ahn KJ, Richardson RD, Mishina Y, Lyons KM, Crenshaw EB, 3rd. Signaling through BMP type 1 receptors is required for development of interneuron cell types in the dorsal spinal cord. Development. 2004;131(21):5393–403.

140. Gupta S, Kawaguchi R, Heinrichs E, Gallardo S, Castellanos S, Mandric I, et al. In vitro atlas of dorsal spinal interneurons reveals Wnt signaling as a critical regulator of progenitor expansion. Cell reports. 2022;40(3):111119.

141. Arnold SJ, Maretto S, Islam A, Bikoff EK, Robertson EJ. Dose-dependent Smad1, Smad5 and Smad8 signaling in the early mouse embryo. Dev Biol. 2006;296(1):104–18.

142. Hazen VM, Andrews MA, Umans L, Crenshaw EB, 3rd, Zwijsen A, Butler SJ. BMP receptor-activated Smads direct diverse functions during the development of the dorsal spinal cord. Dev Biol. 2012;367(2):216–27.

143. Le Dreau G, Saade M, Gutierrez-Vallejo I, Marti E. The strength of SMAD1/5 activity determines the mode of stem cell division in the developing spinal cord. J Cell Biol. 2014;204(4):591–605.

144. Jiang F, Doudna JA. CRISPR-Cas9 Structures and Mechanisms. Annual review of biophysics. 2017;46:505–29.

145. Panchision DM, Pickel JM, Studer L, Lee SH, Turner PA, Hazel TG, et al. Sequential actions of BMP receptors control neural precursor cell production and fate. Genes Dev. 2001;15(16):2094–110.

146. Mazzoni EO, Mahony S, Peljto M, Patel T, Thornton SR, McCuine S, et al. Saltatory remodeling of Hox chromatin in response to rostrocaudal patterning signals. Nat Neurosci. 2013;16(9):1191–8.

147. Rodriguez C, Gallardo S, Gupta S, Butler SJ. Protocol for inducing dorsal spinal sensory interneurons from mouse embryonic stem cell-derived neuromesodermal progenitors. STAR Protoc. 2026;7(1):104307.

148. Gupta S, Heinrichs E, Rodriguez C, Friedman E, Gallardo S, Dermirjie T, et al. In vivo human embryonic spinal cord atlas validates stem cell-derived human dorsal interneurons and reveals ASD spinal signatures. bioRxiv. 2025.

149. Gupta S, Heinrichs E, Novitch BG, Butler SJ. Investigating the basis of lineage decisions and developmental trajectories in the dorsal spinal cord through pseudotime analyses. Development. 2024;151(10):dev202209.

150. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nature Protocols. 2013;8(11):2281–308.

151. Hollnagel A, Oehlmann V, Heymer J, Ruther U, Nordheim A. Id genes are direct targets of bone morphogenetic protein induction in embryonic stem cells. J Biol Chem. 1999;274(28):19838–45.

152. Le Dreau G, Escalona R, Fueyo R, Herrera A, Martinez JD, Usieto S, et al. E proteins sharpen neurogenesis by modulating proneural bHLH transcription factors' activity in an E-box-dependent manner. Elife. 2018;7.

153. Liem KF, Jr., Tremml G, Roelink H, Jessell TM. Dorsal differentiation of neural plate cells induced by BMP-mediated signals from epidermal ectoderm. Cell. 1995;82(6):969–79.

154. Kondas K, Szlama G, Trexler M, Patthy L. Both WFIKKN1 and WFIKKN2 have high affinity for growth and differentiation factors 8 and 11. J Biol Chem. 2008;283(35):23677–84.

155. Szlama G, Kondas K, Trexler M, Patthy L. WFIKKN1 and WFIKKN2 bind growth factors TGFbeta1, BMP2 and BMP4 but do not inhibit their signalling activity. FEBS J. 2010;277(24):5040–50.

156. Orlowski J. SMAD5 signaling: more than meets the nuclei. Cell Res. 2017;27(9):1075–6.

157. Stetak A, Hajnal A. The C. elegans MAGI-1 protein is a novel component of cell junctions that is required for junctional compartmentalization. Dev Biol. 2011;350(1):24–31.

158. Thomas JT, Eric Dollins D, Andrykovich KR, Chu T, Stultz BG, Hursh DA, et al. SMOC can act as both an antagonist and an expander of BMP signaling. eLife. 2017;6.

159. DeGroot MS, Williams B, Chang TY, Maas Gamboa ML, Larus IM, Hong G, et al. SMOC-1 interacts with both BMP and glypican to regulate BMP signaling in C. elegans. PLoS biology. 2023;21(8):e3002272.

160. Martínez-López MJ, Alcántara S, Mascaró C, Pérez-Brangulí F, Ruiz-Lozano P, Maes T, et al. Mouse neuron navigator 1, a novel microtubule-associated protein involved in neuronal migration. Mol Cell Neurosci. 2005;28(4):599–612.

161. van Haren J, Draegestein K, Keijzer N, Abrahams JP, Grosveld F, Peeters PJ, et al. Mammalian Navigators are microtubule plus-end tracking proteins that can reorganize the cytoskeleton to induce neurite-like extensions. Cell Motility. 2009;66(10):824–38.

162. Frei JA, Andermatt I, Gesemann M, Stoeckli ET. The SynCAM synaptic cell adhesion molecules are involved in sensory axon pathfinding by regulating axon-axon contacts. J Cell Sci. 2014;127(Pt 24):5288–302.

163. Kim D-Y, Rhee I, Paik J. Metabolic circuits in neural stem cells. Cellular and Molecular Life Sciences. 2014;71(21):4221–41.

164. Minowada G, Jarvis LA, Chi CL, Neubüser A, Sun X, Hacohen N, et al. Vertebrate Sprouty genes are induced by FGF signaling and can cause chondrodysplasia when overexpressed. Development. 1999;126(20):4465–75.

165. Andrews MG, del Castillo LM, Ochoa-Bolton E, Yamauchi K, Smogorzewski J, Butler SJ. BMPs direct sensory interneuron identity in the developing spinal cord using signal-specific not morphogenic activities. eLife. 2017;6:1–24.

166. Itoh S, Itoh F, Goumans M-J, ten Dijke P. Signaling of transforming growth factor-β family members through Smad proteins. European Journal of Biochemistry. 2000;267(24):6954–67.

167. Rodriguez C, Gallardo S, Gupta S, Butler SJ. Protocol for inducing dorsal spinal sensory interneurons from mouse embryonic stem cell-derived neuromesodermal progenitors. STAR Protocols. 2026;7(1):104307.

168. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol. 2015;109:21.9.1–.9.9.

169. Ianevski A, Giri AK, Aittokallio T. Fully-automated and ultra-fast cell-type identification using specific marker combinations from single-cell transcriptomic data. Nature Communications. 2022;13(1):1246.

170. Mensah EO, Chalif JI, Johnston BR, Chalif E, Parker T, Izzy S, et al. Traumatic spinal cord injury: a review of the current state of art and future directions - what do we know and where are we going? N Am Spine Soc J. 2025;22:100601.

171. Bursell JP, Little JW, Stiens SA. Electrodiagnosis in spinal cord injured persons with new weakness or sensory loss: central and peripheral etiologies. Arch Phys Med Rehabil. 1999;80(8):904–9.

172. Kirshblum S, Snider B, Eren F, Guest J. Characterizing Natural Recovery after Traumatic Spinal Cord Injury. J Neurotrauma. 2021;38(9):1267–84.

173. Sugai K, Nakamura M, Okano H, Nagoshi N. Stem cell therapies for spinal cord injury in humans: A review of recent clinical research. Brain Spine. 2025;5:104207.

174. Bydon M, Qu W, Moinuddin FM, Hunt CL, Garlanger KL, Reeves RK, et al. Intrathecal delivery of adipose-derived mesenchymal stem cells in traumatic spinal cord injury: Phase I trial. Nat Commun. 2024;15(1):2201.

175. Pașca SP. Constructing human neural circuits in living systems by transplantation. Cell. 2024;187(1):8–13.

176. Revah O, Gore F, Kelley KW, Andersen J, Sakai N, Chen X, et al. Maturation and circuit integration of transplanted human cortical organoids. Nature. 2022;610(7931):319–26.

177. Piao J, Zabierowski S, Dubose BN, Hill EJ, Navare M, Claros N, et al. Preclinical Efficacy and Safety of a Human Embryonic Stem Cell-Derived Midbrain Dopamine Progenitor Product, MSK-DA01. Cell Stem Cell. 2021;28(2):217–29.e7.

178. Marino RJ, Barros T, Biering-Sorensen F, Burns SP, Donovan WH, Graves DE, et al. International standards for neurological classification of spinal cord injury. J Spinal Cord Med. 2003;26 Suppl 1:S50–6.

179. Butler SJ, Bronner ME. From classical to current: analyzing peripheral nervous system and spinal cord lineage and fate. Dev Biol. 2015;398(2):135–46.

180. Sagner A, Briscoe J. Establishing neuronal diversity in the spinal cord: a time and a place. Development. 2019;146(22).

181. Di Bonito M, Glover JC, Studer M. Hox genes and region-specific sensorimotor circuit formation in the hindbrain and spinal cord. Dev Dyn. 2013;242(12):1348–68.

182. Banerjee-Basu S, Packer A. SFARI Gene: an evolving database for the autism research community. Dis Model Mech. 2010;3(3-4):133–5.

183. Miller A, Dasen JS. Establishing and maintaining Hox profiles during spinal cord development. Semin Cell Dev Biol. 2024;152-153:44–57.

184. Gouveia A, Marcelino HM, Gonçalves L, Palmeirim I, Andrade RP. Patterning in time and space: HoxB cluster gene expression in the developing chick embryo. Cell Cycle. 2015;14(1):135–45.

185. Lawrence JEG, Roberts K, Tuck E, Li T, Mamanova L, Balogh P, et al. HOX gene expression in the developing human spine. Nature communications. 2024;15(1):10023.

186. Pituello F. Neuronal specification: generating diversity in the spinal cord. Curr Biol. 1997;7(11):R701–4.

187. Price SR, Briscoe J. The generation and diversification of spinal motor neurons: signals and responses. Mech Dev. 2004;121(9):1103–15.

188. Lee H, Lee HY, Lee BE, Gerovska D, Park SY, Zaehres H, et al. Sequentially induced motor neurons from human fibroblasts facilitate locomotor recovery in a rodent spinal cord injury model. Elife. 2020;9.

189. Son D, Zheng J, Kim IY, Kang PJ, Park K, Priscilla L, et al. Human induced neural stem cells support functional recovery in spinal cord injury models. Exp Mol Med. 2023;55(6):1182–92.

190. Amoroso MW, Croft GF, Williams DJ, O'Keeffe S, Carrasco MA, Davis AR, et al. Accelerated high-yield generation of limb-innervating motor neurons from human stem cells. J Neurosci. 2013;33(2):574–86.

191. Briggs JA, Li VC, Lee S, Woolf CJ, Klein A, Kirschner MW. Mouse embryonic stem cells can differentiate via multiple paths to the same state. Elife. 2017;6.

192. Henrique D, Abranches E, Verrier L, Storey KG. Neuromesodermal progenitors and the making of the spinal cord. Development. 2015;142(17):2864–75.

193. Jin H, Liu Z, Mou J, Tang M, Huang X, Liu K, et al. Dual genetic tracing demonstrates the heterogeneous differentiation and function of neuromesodermal progenitors in vivo. Proc Natl Acad Sci U S A. 2025;122(14):e2402305122.

194. Gupta S, Heinrichs E, Novitch BG, Butler SJ. Investigating the basis of lineage decisions and developmental trajectories in the dorsal spinal cord through pseudotime analyses. Development. 2024;151(10).

195. Mead BE, Ordovas-Montanes J, Braun AP, Levy LE, Bhargava P, Szucs MJ, et al. Harnessing single-cell genomics to improve the physiological fidelity of organoid-derived cell types. BMC Biol. 2018;16(1):62.

196. Allison T, Langerman J, Sabri S, Otero-Garcia M, Lund A, Huang J, et al. Defining the nature of human pluripotent stem cell-derived interneurons via single-cell analysis. Stem Cell Reports. 2021;16(10):2548–64.

197. Rayon T, Maizels RJ, Barrington C, Briscoe J. Single-cell transcriptome profiling of the human developing spinal cord reveals a conserved genetic programme with human-specific features. Development. 2021;148(15).

198. Li X, Andrusivova Z, Czarnewski P, Langseth CM, Andersson A, Liu Y, et al. Profiling spatiotemporal gene expression of the developing human spinal cord and implications for ependymoma origin. Nat Neurosci. 2023;26(5):891–901.

199. Zhang Q, Wu X, Fan Y, Jiang P, Zhao Y, Yang Y, et al. Single-cell analysis reveals dynamic changes of neural cells in developing human spinal cord. EMBO Rep. 2021;22(11):e52728.

200. Andersen J, Thom N, Shadrach JL, Chen X, Onesto MM, Amin ND, et al. Single-cell transcriptomic landscape of the developing human spinal cord. Nat Neurosci. 2023;26(5):902–14.

201. Shi Y, Huang L, Dong H, Yang M, Ding W, Zhou X, et al. Decoding the spatiotemporal regulation of transcription factors during human spinal cord development. Cell research. 2024;34(3):193–213.

202. Lippmann ES, Williams CE, Ruhl DA, Estevez-Silva MC, Chapman ER, Coon JJ, et al. Deterministic HOX patterning in human pluripotent stem cell-derived neuroectoderm. Stem cell reports. 2015;4(4):632–44.

203. Cooper F, Gentsch GE, Mitter R, Bouissou C, Healy LE, Rodriguez AH, et al. Rostrocaudal patterning and neural crest differentiation of human pre-neural spinal cord progenitors in vitro. Stem cell reports. 2022;17(4):894–910.

204. Liu JP, Laufer E, Jessell TM. Assigning the positional identity of spinal motor neurons: rostrocaudal patterning of Hox-c expression by FGFs, Gdf11, and retinoids. Neuron. 2001;32(6):997–1012.

205. Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat Biotechnol. 2024;42(2):293–304.

206. Gayoso A, Lopez R, Xing G, Boyeau P, Valiollah Pour Amiri V, Hong J, et al. A Python library for probabilistic analysis of single-cell omics data. Nat Biotechnol. 2022;40(2):163–6.

207. Ianevski A, Giri AK, Aittokallio T. Fully-automated and ultra-fast cell-type identification using specific marker combinations from single-cell transcriptomic data. Nature communications. 2022;13(1):1246.

208. Rabadan MA, Cayuso J, Le Dreau G, Cruz C, Barzi M, Pons S, et al. Jagged2 controls the generation of motor neuron and oligodendrocyte progenitors in the ventral spinal cord. Cell Death Differ. 2012;19(2):209–19.

209. Tien AC, Tsai HH, Molofsky AV, McMahon M, Foo LC, Kaul A, et al. Regulated temporal-spatial astrocyte precursor cell proliferation involves BRAF signalling in mammalian spinal cord. Development. 2012;139(14):2477–87.

210. Aibar S, Gonzalez-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, et al. SCENIC: single-cell regulatory network inference and clustering. Nat Methods. 2017;14(11):1083–6.

211. Roome RB, Yadav A, Flores L, Puarr A, Nardini D, Richardson A, et al. Ontogeny of the spinal cord dorsal horn. bioRxiv. 2025.

212. Wildner H, Muller T, Cho SH, Brohl D, Cepko CL, Guillemot F, et al. dILA neurons in the dorsal spinal cord are the product of terminal and non-terminal asymmetric progenitor cell divisions, and require Mash1 for their development. Development. 2006;133(11):2105–13.

213. Dal Monte M, Petrucci C, Cozzi A, Allen JP, Bagnoli P. Somatostatin inhibits potassium-evoked glutamate release by activation of the sst(2) somatostatin receptor in the mouse retina. Naunyn-Schmiedeberg's archives of pharmacology. 2003;367(2):188–92.

214. Huang J, Polgar E, Solinski HJ, Mishra SK, Tseng PY, Iwagaki N, et al. Circuit dissection of the role of somatostatin in itch and pain. Nat Neurosci. 2018;21(5):707–16.

215. Diaz-delCastillo M, Woldbye DPD, Heegaard AM. Neuropeptide Y and its Involvement in Chronic Pain. Neuroscience. 2018;387:162–9.

216. Coste B, Mathur J, Schmidt M, Earley TJ, Ranade S, Petrus MJ, et al. Piezo1 and Piezo2 are essential components of distinct mechanically activated cation channels. Science. 2010;330(6000):55–60.

217. Shi Y, Liu JP. Gdf11 facilitates temporal progression of neurogenesis in the developing spinal cord. J Neurosci. 2011;31(3):883–93.

218. Lotfollahi M, Naghipourfar M, Luecken MD, Khajavi M, Buttner M, Wagenstetter M, et al. Mapping single-cell data to reference atlases by transfer learning. Nat Biotechnol. 2022;40(1):121–30.

219. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic acids research. 2021;49(D1):D605–D12.

220. Nelson TS, Sinha GP, Santos DFS, Jukkola P, Prasoon P, Winter MK, et al. Spinal neuropeptide Y Y1 receptor-expressing neurons are a pharmacotherapeutic target for the alleviation of neuropathic pain. Proc Natl Acad Sci U S A. 2022;119(46):e2204515119.

221. Sudhof TC. Neuroligins and neurexins link synaptic function to cognitive disease. Nature. 2008;455(7215):903–11.

222. Singh SK, Eroglu C. Neuroligins provide molecular links between syndromic and nonsyndromic autism. Science signaling. 2013;6(283):re4.

223. Tromp A, Mowry B, Giacomotto J. Neurexins in autism and schizophrenia-a review of patient mutations, mouse models and potential future directions. Mol Psychiatry. 2021;26(3):747–60.

224. Monteiro P, Feng G. SHANK proteins: roles at the synapse and in autism spectrum disorder. Nature reviews. 2017;18(3):147–57.

225. Kantojarvi K, Kotala I, Rehnstrom K, Ylisaukko-Oja T, Vanhala R, von Wendt TN, et al. Fine mapping of Xq11.1-q21.33 and mutation screening of RPS6KA6, ZNF711, ACSL4, DLG3, and IL1RAPL2 for autism spectrum disorders (ASD). Autism Res. 2011;4(3):228–33.

226. Penagarikano O, Abrahams BS, Herman EI, Winden KD, Gdalyahu A, Dong H, et al. Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. Cell. 2011;147(1):235–46.

227. Iyer NR, Shin J, Cuskey S, Tian Y, Nicol NR, Doersch TE, et al. Modular derivation of diverse, regionally discrete human posterior CNS neurons enables discovery of transcriptomic patterns. Sci Adv. 2022;8(39):eabn7430.

228. Dady A, Davidson L, Halley PA, Storey KG. Human spinal cord in vitro differentiation pace is initially maintained in heterologous embryonic environments. Elife. 2022;11.

229. Leitch DB, Catania KC. Structure, innervation and response properties of integumentary sensory organs in crocodilians. J Exp Biol. 2012;215(Pt 23):4217–30.

230. Cooper JN, Mittal J, Sangadi A, Klassen DL, King AM, Zalta M, et al. Landscape of. J Clin Med. 2024;13(7).

231. Jiang YH, Ehlers MD. Modeling autism by SHANK gene mutations in mice. Neuron. 2013;78(1):8–27.

232. Tokunaga S, Shimomura H, Taniguchi N, Yanagi K, Kaname T, Okamoto N, et al. A novel DLG4 variant causes DLG4-related synaptopathy with intellectual regression. Hum Genome Var. 2024;11(1):1.

233. Jang WE, Park JH, Park G, Bang G, Na CH, Kim JY, et al. Cntnap2-dependent molecular networks in autism spectrum disorder revealed through an integrative multi-omics analysis. Mol Psychiatry. 2023;28(2):810–21.

234. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature. 2014;515(7526):209–15.

235. Courchesne E, Gazestani VH, Lewis NE. Prenatal Origins of ASD: The When, What, and How of ASD Development. Trends Neurosci. 2020;43(5):326–42.

236. Andre P, Joshi SR, Briscoe SD, Alexander MJ, Li G, Kumar R. Therapeutic Approaches for Treating Pulmonary Arterial Hypertension by Correcting Imbalanced TGF-β Superfamily Signaling. Front Med (Lausanne). 2021;8:814222.

237. Zhou J, Lee PL, Tsai CS, Lee CI, Yang TL, Chuang HS, et al. Force-specific activation of Smad1/5 regulates vascular endothelial cell cycle progression in response to disturbed flow. Proc Natl Acad Sci U S A. 2012;109(20):7770–5.

238. Lin Z, Zhuang J, He L, Zhu S, Kong W, Lu W, et al. Exploring Smad5: a review to pave the way for a deeper understanding of the pathobiology of common respiratory diseases. Mol Med. 2024;30(1):225.

239. Tavakol S. The Twofold Role of Osteogenic Small Molecules in Parkinson's Disease Therapeutics: Crosstalk of Osteogenesis and Neurogenesis. Biomed Res Int. 2022;2022:3813541.

240. Bentz K, Molcanyi M, Hess S, Schneider A, Hescheler J, Neugebauer E, et al. Neural differentiation of embryonic stem cells is induced by signalling from non-neural niche cells. Cell Physiol Biochem. 2006;18(4-5):275–86.

241. Bahrami SB, Veiseh M, Boudreau NJ. Isolation and expansion of endothelial progenitor cells derived from mouse embryonic stem cells. Methods Mol Biol. 2012;916:81–96.

242. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001;25(4):402–8.