

TITLE

Capture of Human Neuromesodermal and Posterior Neural Tube Axial Stem Cells

AUTHORS

Dolunay Kelle^{1#} Enes Ugur^{2,3#}, Ejona Rusha⁴, Dmitry Shaposhnikov¹, Alessandra Livigni⁵, Sandra Horschitz^{6,7}, Mahnaz Davoudi⁶, Andreas Blutke⁸, Judith Bushe⁹, Michael Sterr^{10,11}, Ksenia Arkhipova¹², Benjamin Tak¹², Ruben de Vries¹², Mazène Hochane¹², Britte Spruijt¹², Aicha Haji Ali¹, Heiko Lickert^{10,11}, Annette Feuchtinger⁹, Philipp Koch^{6,7}, Matthias Mann³, Heinrich Leonhardt², Valerie Wilson⁵, Micha Drukker^{1,4,12,13}

Co-first authorship

CORRESPONDENCE

m.drukker@lacdr.leidenuniv.nl

¹Institute of Stem Cell Research, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, 85764, Germany

²Faculty of Biology and Center for Molecular Biosystems (BioSysM), Human Biology and BioImaging, Ludwig-Maximilians-Universität München, Munich, 81377, Germany

³Department of Proteomics and Signal Transduction, Max-Planck Institute of Biochemistry, Martinsried, 82152, Germany

⁴Induced Pluripotent Stem Cell Core Facility, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, 85764, Germany

⁵Centre for Regenerative Medicine, Institute for Stem Cell Research, School of Biological Sciences, University of Edinburgh, 5 Little France Drive, Edinburgh EH16 4UU, UK

⁶Central Institute of Mental Health, University of Heidelberg/Medical Faculty Mannheim, Mannheim, 68159, Germany

⁷Hector Institute for Translational Brain Research (HITBR), Mannheim, 68159, Germany

⁸Institute for Veterinary Pathology, Ludwig-Maximilians-Universität München Veterinärstr. 13, 80539 Munich, Germany

⁹Core Facility Pathology and Tissue Analytics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, 85764, Germany

¹⁰Institute of Diabetes and Regeneration Research, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, 85764, Germany

¹¹German Center for Diabetes Research (DZD), Neuherberg, 85764, Germany

¹²Division of Drug Discovery and Safety, Leiden Academic Centre for Drug Research (LACDR), Leiden University, 2333 CC Leiden, The Netherlands

¹³Department of Cell and Chemical Biology, Leiden University Medical Center, Einthovenweg 20, 2333 ZC Leiden, The Netherlands

AUTHOR CONTRIBUTIONS

D.K., E.R., D.S., M.Dr. designed the study and M.Dr. supervised the study. D.K. established human AxSCs lines; performed screening for mouse and orangutan AxSCs culture conditions and differentiation, immunocytochemistry, RT-qPCR experiments; prepared and analyzed the proteome, chromatome, and scRNA sequencing samples and prepared the figures. E.U. prepared, measured and analyzed the proteome and chromatome samples under supervision of H.L. and M.M, prepared the related figures and programmed the web application. E.R. established additional human AxSC lines, assisted with cell line establishment and differentiation experiments, experimental design and analysis of scRNA sequencing samples. D.S. and A.H.A performed screening for human AxSCs, established the human HMGU1 AxSCs, generated the β -catenin hESC line and performed RNA sequencing. A.L. and V.W. performed in vivo injection and analysis of the mouse AxSCs and contributed to the writing of the manuscript. S.H. and M.Da. performed electrophysiology experiments under supervision of P.K. M.Dr. performed teratoma assays and A.B and J.B. performed histopathological analysis. M.S. generated libraries for scRNA sequencing under supervision of H.L. K.A. performed preprocessing of the scRNA sequencing samples and assisted with downstream analysis. B.T. generated the time-lapse videos and performed with B.S. analysis of neuromuscular junction. R.d.V. and M.H. performed Seahorse experiment. D.K. and M.Dr. wrote the manuscript.

ACKNOWLEDGEMENTS

This work was partially funded by the Lingling Wiyadharma Fonds ter bevordering van de beoefening der Natuurwetenschappen, and VolkswagenStiftung under grant A130140. We gratefully acknowledge the support of the John Templeton Foundation grant #62572. We would like to thank the German Academic Exchange Service (DAAD) for supporting DK. We acknowledge the technical support of Core Facility Genomics at Helmholtz Munich. We thank Dr. Inti Alberto De la Rosa-Velázquez for help with next generation sequencing, Dr. Peter Lichtner for help with genotyping, Dr. Thomas Walzthoeni for help with data processing, Moritz Thomas and Dr. Carsten Marr for help with data analysis. We would like to thank Dr. Thure Adler, Dr. Anna Pertek and Dr. Friederike Matheus for help with the animal experiments as well as Dr. Silvia Schirge for help with embryo dissections. We would like to acknowledge Dr. Christian Schroeter for critical reading of the manuscript.

DECLARATIONS OF INTEREST

D.K., E.R., D.S., A.H.A and M.D. are named on a patent related to the work described in this study.

ABSTRACT (150 words)

The spinal cord, nerves, and skeletal muscles arise from neuromesodermal progenitors (NMPs). We have developed a growth-factor screening strategy, utilizing ES and iPS cells, facilitating the indefinite self-renewal of two types of human axial stem cells (AxSCs), closely resembling mouse NMPs (NM-AxSCs) and posterior neural tube progenitors (N-AxSCs). Under specific regimens—Wnt/CHIR99021, FGF2, and TGF- β inhibitor SB431542 (CFS) and excluding FGF2 (CS), respectively—these AxSCs self-renew and sustain telomeres. Single cell transcriptomics and proteomics have revealed expression of posterior growth-zone and dorsoventral neural tube markers in NM-AxSCs, and correspondingly, differentiation to a wide spectrum of neural tube neurons and myocytes. N-AxSCs rapidly matured into dorsal sensory subsets and neural crest. Crucially, neither AxSC type produces teratomas, and analogous mouse NM-AxSCs integrated successfully into the neural tube and somites. Capturing of AxSCs from patient and GMP ES / iPS cells without transgenesis unveils ontogeny and promises modeling and therapy in neuropathies.

INTRODUCTION

Genes and signaling pathways that temporarily promote self-renewal of progenitor cells in human embryos can be leveraged to “capture” indefinitely proliferating stem cells with analogous differentiation capabilities both *in vivo* and *in vitro*. This concept underpins the derivation and cultivation of ground and primed state embryonic and induced pluripotent stem cells (e/iPSCs)^{1–4}, as well as formative and rosette PSCs from periimplantation embryos.^{5,6} However, ethical restrictions constrain research of human gastrulation and neural tube formation during Carnegie stages 5–9.⁷ We thus asked whether human ES and iPS cells can be utilized for the derivation of stem cells that embody gastrulation and peripheral neurogenesis, bypassing the use of embryos.

Neuromesodermal progenitors (NMPs) are multipotent cells that develop into the neural tube, trunk neural crest, spinal cord, and sensory neurons, and somitic and tailbud mesoderm.^{8–11} In mice, these cells intriguingly express stemness, neurogenesis, and mesoderm transcription factors *Sox2* and *Brachyury*.¹² Mouse NMPs reside in the caudal lateral epiblast (CLE) and the node-streak border (NSB) around E7.5 and persist in the chordo-neural hinge (CNH) until E13.5.^{10,13} While mouse NMPs manifest characteristics of stem cells, such as ability to passage through embryos¹⁴, no stable stem cell lines retaining NMP and posterior neural tube traits have been established.

Differentiation protocols that involve 2–3-day activation of Wnt/β-catenin and FGF2 treatment yield transient NMP-like progenitors from both human and mouse e/iPSCs.^{15–19} These progenitors exhibit defining markers like *SOX2*, *T* (*Brachyury*), and *CDX2*, mirroring mouse NMPs.^{10,20,21} However, human and mouse NMPs cultured *in vitro* appear to display late NMP phenotype as indicated by the presence of posterior *HOX* genes from paralogous groups 9–13.^{15,17,19} Further, late NMPs manifest markers like *Mnx1* and non-canonical Wnt/β-catenin ligands such as *Wnt5a*.^{13,21} This led us to explore strategies to harness mechanisms of temporary self-renewal of NMPs and posterior neural progenitors, aiming to derive stem cells that sustain the respective states through extended passages.

The transition between NMPs, posterior neural progenitors and their patterning along the rostrocaudal axis establishes the mammalian neural tube.^{22–24} Studies analyzing branching cell trajectories in mouse embryos demonstrated that trunk neural crest progenitors delaminate from the dorsal region of the neural tube.^{25–27} On the ventral side of the neural tube, sonic hedgehog (Shh) signaling promotes differentiation of motor neurons and ventral interneurons.²⁸ Although various differentiation protocols have been optimized to produce posterior neuronal subtypes and trunk neural crest, they lack the efficiency of the dual Smad inhibition protocol in the induction of anterior neurons and cerebral organoids.^{29,30} At present, the challenges of forced expression of neurogenic transcription factors and cellular heterogeneity obstruct the detailed study of the progenitor intermediates and lineages in the human neural tube during their differentiation from NMPs.^{29,31} As such, we aimed to capture pure NMP- and posterior neural

tube-like stem cells to dissect lineages, elucidate mechanisms underlying human neural tube development, and enhance cellular modelling capabilities.

Here, we report the establishment of AxSC lines from multiple species, with characteristics that closely mirror NMPs and neural progenitors without the need for ectopic expression of transgenes or use of human embryos. We identified two regimens: Wnt/β-catenin (CHIR99021), FGF2, and TGF-β pathway inhibitor SB431542 (abbreviated as CFS) for NM-AxSCs (neuromesodermal-like) and CS for N-AxSCs (neuronally-committed). These CFS and CS lines can be cultivated across dozens of passages, retaining telomere length. They maintain their developmental correspondence to NMPs and posterior neural tube progenitors, and capacity to differentiate to paraxial mesoderm and neural derivatives, both in vitro and when grafted into embryos. Their classification was determined using single cell transcriptomics, proteome analysis, as well as chromatin occupancy proteomics. The data can be explored using a web tool <https://axialstemcells.shinyapps.io/atlas>. Reciprocal medium exchange showed that CFS cells can convert to a CS phenotype but not vice versa.

The derivation of undifferentiated CFS and CS lines from ES and iPS cells offer insights into the particularly refractory early stages of human postimplantation development. These cells also offer exciting possibilities for GMP manufacturing and banking of a range of neural and mesodermal cell products. The differentiation protocols of CFS and CS cells are shorter and more directed than those for ES and iPS cells, enhancing the modeling of ontogeny and disease in body axis tissues. Unlike pluripotent stem cell lines, CFS and CS cells do not make teratomas, positioning them as a potentially pivotal resource for stem cell therapies targeting muscular, spinal cord, and nerve pathologies without risk of teratoma.

RESULTS

Screening human ES cell-derived long-term renewing multipotent posterior stem cells

To capture human multipotent NMP- and posterior neural tube-like stem cells, we sought to induce cell lines capable of sustaining the long-term expression of endogenous *T*(BRA) and *SOX2*, or *PAX6*, *SOX2* and *HES5* over numerous passages. These genes represent the defining characteristics of NMPs and the neuronally committed posterior progenitors that are derived from NMPs.^{12,32,33} As a first step, we aimed to identify the temporal point during human ES cell differentiation when the activation of Wnt/β-catenin promotes NMP markers without yet inducing somitogenesis and mesendoderm genes. We analyzed RNA-seq data where the Wnt/β-catenin pathway was activated via CHIR99021 treatment and by employing human ES cells harbouring stably integrated β-catenin ΔN90 variant that is responsive to doxycycline.³⁴ The time course RNA-seq revealed strong upregulation of *T* and *CDX2* at 24h, followed by more advanced developmental markers e.g. *GATA6* and *FGF17* which coincided with downregulation of *SOX2* at 48-72h (Figure S1A-C). Thus, we utilized the NMP-like state at 24h post induction with CHIR99021

as a starting point for screening the propagation of posterior stem cells including neural tube progenitors.

At the second step, we screened the formation of putative NMP-like cells by serial passaging and modulating the signalling pathways that were up-regulated during NMP induction (Figures 1A,B and S1D). We performed screening with H9 cells that were cultivated on Matrigel in 5% O₂ as hypoxia is favourable for NMPs.³⁵ The screening of Wnt/β-catenin, FGF2, TGF-β and BMP modulators resulted in the formation of cell lines in the CHIR99021/FGF2/SB-431542 (abbreviated CFS), and without FGF2 (abbreviated CS) conditions. The observed down-regulation of *OCT4* and *NANOG* in three independently derived lines indicated the non-pluripotent state of CFS and CS cell lines. The CFS cells maintained high *T* and *CDX2* expression, while CS cells exhibited high *PAX6* levels. Both types expressed *SOX2* at levels comparable to undifferentiated H9 cells. These NMP and neural characteristics persisted for a minimum of 27 passages across the three cell lines (Figure 1C). Treatment with CHIR99021 and FGF2 (abbreviated CF) or CHIR99021 alone (abbreviated CH) yielded cell lines with phenotypes that were analogous those established in CFS and CS conditions, however, they were discontinued due to colonies that were heterogeneous compared to the CFS and CS conditions (Figure S1E,F). From approximately passage 6, CFS and CS lines manifested a characteristic colony morphology, and demonstrated mutually exclusive expression of NMP and neuronal markers (Figure 1D,E). Time lapse microscopy revealed filopodia-mediated motility of both cell types, merging of cells to colonies, and pre-neural morphology, particularly of CS cells that exhibited occasional spontaneous axon growth (Video S1).

Next, we employed omics to validate the NMP and posterior neural tube markers of CFS and CS cells (Figure S2). Using chromatin aggregation capture and proteome mass spectrometry (MS) we found that the three independently derived CFS and CS lines exhibited chromatin occupancy enrichment of *T* and *CDX2* in CFS cells, *PAX6* and *HES5* in CS cells, and *SOX2* in both cell types and undifferentiated H9 ePSCs (Figure 1F). Single cell RNA sequencing (scRNA-seq) further demonstrated specific expression of *T* and *CDX2* in CFS cell lines, and mutually exclusive expression of *PAX6* and *HES5* in CS cells (Figure 1G and S1G).

A defining hallmark of PSCs is formation of teratoma-like tumors comprising multiple tissues.³⁶ To address whether CFS and CS cells underwent irreversible commitment from PSCs, we injected CFS, CS, and ePSC-H9 cells under the kidney capsule of immunodeficient mice. While teratomas were detected in all ePSC-H9 injected animals, none of the animals injected with CFS and CS cells (0/13) exhibited endoderm or teratocarcinoma tissue characteristics, rather tissues resembling cartilage and adipocytes which are derivatives of mesodermal and mesenchymal cells (Figures 1H and S1H,I).³⁷

Finally, we analysed characteristics of stemness and plasticity. We noticed chromatin enrichment and expression of essential stemness genes in CFS and CS cells including *MYCN*, *LIN28B*, *SALL4*, and *TRIM71* at levels akin to human ePSCs (*MYCN* was higher in CFS and CS

relative to ePSC H9), and depletion of pluripotency factors, namely, OCT4, NANOG, and PRDM14 (Figure 1I,J).³⁸ Since the maintenance of telomere length contributes to the extended replicative lifespan of stem cells³, we analyzed the relative telomere length of three CFS and CS cells over 20 passages. Importantly, the telomere lengths of CFS, CS and H9 ePSC lines were similarly maintained (Figure 1K). These findings suggest capture of novel posterior stem cells directly from ES cells, bypassing the need for postimplantation human embryos. Given that the CFS and CS cells consistently retained key characteristics of NMPs and posterior neural tube progenitors across numerous passages, we designated them as NMP and neuronal axial stem cells, respectively abbreviated as NM- and N-AxSCs. For the exploration of the proteome and chromatome of NM- and N-AxSC we created an online platform available at <https://axialstemcells.shinyapps.io/atlas>.

Hierarchical ordering and reproducible derivation of AxSC from iPSCs, species, and embryos

We next examined the hierarchy of the developmental potential and lineage commitment of CFS and CS cell lines. To interrogate, we performed serial passaging experiments in which the respective media of CFS and CS cells were swapped (Figure 2A). We observed that the CFS cells gave rise to cells with CS morphology following 4-6 passages, and that the markers of NMPs, *CDX2* and *T* were downregulated, and *PAX6* was upregulated. Conversion in the other direction, however, did not occur, as the three CS lines that were propagated in CFS media did not show significant upregulation of *T* and *CDX2* or stable downregulation of *PAX6*. Additionally, *SOX2* was downregulated and the cells exhibited spontaneous differentiation (Figure 2B). These results indicate that the CFS cell lines represent NM-AxSCs that are hierarchically upstream of N-AxSCs CS cell lines.

Next, we analyzed the reproducibility of deriving AxSCs. To validate the protocols across human PSCs from different (epi)genetic backgrounds, we tested them on HUES6 ES cells and HMGU1 iPS cells.^{39,40} In three rounds of derivation, we consistently observed cell lines that mirrored the morphology and expression of NMP and posterior neural markers, analogous to the H9 CFS and CS cell lines (Figures 2C and S3A). Moreover, scRNA-seq analysis revealed that the majority of the genes enriched in CFS and CS cell lines (relative to their respective undifferentiated state of H9, HUES6 and HMGU1) were consistent between cell lines.

To derive mouse AxSCs we screened Wnt/β-catenin, FGF, TGF-β, and BMP modulators with the addition of FGF8 for its role in mouse gastrulation.⁴¹ We performed protocol screening with mouse primed state EpiSCs on Matrigel or mouse embryonic fibroblast (MEF) coated plates as feeders. We utilized EpiSCs following splitting, skipping NMP induction, because a significant portion of mouse EpiSCs express *Brachyury* in the primed pluripotency state.^{42,43} The screening resulted in a successful derivation of a cell line on Matrigel using a modified CFS recipe with 3μM CHIR99021, 50ng/ml FGF2, and 10μM TGF-β inhibitor SB431542 (Figure 2D,E). Mouse CFS_1 line exhibited a strong induction of *Brachyury* and *Cdx2*, with *Sox2* levels similar to those in parental cells, while *Oct4*, *Nanog*, and *Pax6* were notably absent (Figures 2D-G and S3B). This expression

pattern persisted for a minimum of 21 passages (Figure 2G). The derivation of CFS cells with MEFs also yielded cell lines under a range of concentrations, specifically CHIR99021 (3-5 μ M), FGF2 (50-100ng/ml), with or without DMH1 (Figures 2D,E and S3C). While this confirms that NMP-like CFS cell lines can be derived from mouse PSCs, we were unable to establish neuronally-committed posterior mouse CS-like cell lines. Thus, to validate the reproducibility of neuronally committed posterior AxSCs, we applied the CS protocol to Sumatran orangutan (*Pongo abelii*) female iPSCs.⁴⁴ This yielded cell lines with both 5 μ M and 10 μ M CHIR99021 which expressed SOX2, PAX6 and HES5 while lacking NMP and pluripotency markers (Figure S3D-F).

Lastly, we investigated the potential to capture NMP-like AxSCs directly from mouse embryos. We explanted the stem zone region of 7-8 and 16-17 somite stage (SS) embryos, dissociated the tissue, and plated the cells in the mCFS condition, specifically 3 μ M CHIR99021, 50ng/ml FGF2, and 10 μ M SB431542 (Figure 2H). This yielded successful derivation of NMP-like cell lines from both stages on Matrigel and MEFs (Figure 2I,J). Intriguingly, the cell lines derived from 7-8 SS presented distinct morphological differences compared to those from the 16-17 SS mouse embryos. Specifically, the 7-8 SS bore resemblance to mouse EpiSCs (Figure 2I). The cells displayed a more epithelial-like appearance and expressed Hes5, a marker not seen in CFS cell lines derived from 16-17 somite embryos (Figure 2J). These observations suggest that the CFS condition can support stem cells of varying NMP stages. We concluded that these protocols allow capture of NM-AxSCs and N-AxSCs from a broad range of pluripotent cells and directly from embryos, and that species differences influence the conditions needed to propagate these cells.

Correspondences of CFS and CS lines with NMPs and posterior neural tube progenitors

To further investigate the developmental correspondences of CFS and CS lines we established a reference model comprising of transcription factors expressed by early and late NMPs, neural progenitors, trunk neural crest, neural tube, ventral and dorsal domains of the neural tube, as well as mesodermal genes expressed in the nascent, caudal, and paraxial mesoderm.^{16,20,21,32,45} To distinguish posterior and anterior neural tube progenitors we included central nervous system progenitor makers (Figure 3A). We used the model as a framework to analyze scRNA-seq, chromatome, and proteome of the human stem cell lines established in CFS and CS conditions.

The scRNA-seq of undifferentiated CFS cells revealed a striking similarity to mouse NMPs with a tendency to a late phenotype. This was manifested by the enrichment of MNX1 (late NMP), EVX1, RARG, MSX1, ETV5, NKX1-2, and GBX2 relative to CS lines and ES cells (Figure 3B).^{13,21} Furthermore, CFS cells exhibited a substantial enrichment of multiple HOX genes from anteriormost to A-D9-11, pointing to a resemblance to late mouse NMPs.^{15,16,21} Notably, HOX, HOX antisense, and posterior embryonic transcripts DUSP6 and WNT5 (late NMP) encompassed roughly half of the top 50 enriched genes in CFS cells (Figure S4A).^{21,46} Additionally, the somite segmentation oscillator HES7 gene^{47,48} was also expressed in CFS cells.

Conversely, undifferentiated CS cells displayed a striking phenotype of neural progenitor cells demonstrated by enrichment of PAX6, SOX1, SOX3, IRX3, IRX5, and HES5, relative to both CFS lines and ES cells (Figure 3B).^{16,21} Notably, in the top 50 genes enriched genes CS cells expressed a subset of the anterior HOX genes - specifically A2-3 and B1-4 – which were enriched in CFS cells, in addition to neurofilament genes such as TUBA1A, NES (Nestin), and VIM (Vimentin) (Figure S4A). Furthermore, transcription factors linked with neural progenitor state, including NR2F1, NR2F2, POU3F2, and POU3F3^{49,50} were uniquely enriched in CS cells, implying a role in the self-renewal of these cells. To elucidate differences from anterior neural progenitor cells (NPCs), we differentiated H9 ePSCs by the dual SMAD inhibition protocol (Figure S4B-C).⁵¹ A comparison of the scRNA-seq data between NPCs and CS cells revealed significant differences, in the expression of the ventral midbrain marker NKX6-1, and the absence of expression of HOX genes and dorsal neural tube markers, which were expressed in CS cells (Figure S4D-E).

Further examination of ventral and dorsal neural tube markers refined our classification (Figure 3A). We observed enrichment of both ventral - with markers such as NKX6-1, SP8, and FOXB1 - and dorsal markers, including PAX3 and MSX1 in CFS cells (Figure 3B). Contrastingly, CS cells expressed only the dorsal subset of these genes with additional enrichment in other dorsal neural tube markers like PAX7, ZIC1, MSX2, and GSX1. The scRNA-seq analyses of HUES6 and HMGU1 CFS and CS cell lines demonstrated consistent phenotypes with the undifferentiated CFS and CS cells from the H9 line, with these phenotypes being maintained in early and late passage (Figure S4F). We observed higher expression levels of EN1 and SOX3 and a lower level of CDX2 in CFS HUES6 cells, potentially indicating a more advanced neural progenitor state. Taken together, these data support classifying CFS cell lines as late NMPs with characteristics of mesoderm potential, and CS cell lines as posterior neural tube progenitors with bias towards dorsal fates. This corroborates the proposed hierarchy, positioning CFS upstream of the CS cell lines.

To interrogate the stem cell states of CFS and CS cell lines, we examined their chromatomes (<https://axialstemcells.shinyapps.io/atlas>). Inspection of the arrangement of the samples in principal component analysis revealed that the global difference between CFS and CS to ES cells is greater than between CFS and CS cell states. Loading of the coefficients assigned to individual proteins, confirmed the contribution of NMP markers, dorsal markers, and multiple HOX genes to the CFS cell state, and of neural tube dorsal markers to the CS cell state (Figure 3C,D). It suggested that the regulation of stemness by SOX2 and SALL4 is common to CFS, CS and ePSCs, without the involvement of the pluripotency factors such as OCT4, NANOG, and PRDM14. Despite propagation being reliant on Wnt/β-catenin signaling in both CFS and CS cell lines, a higher enrichment was found in the association of several Wnt/β-catenin transcription factors such as LEF1, SP5, SP6, and SP8 with the chromatin of CFS cells. Conversely, transcription factors associated with retinoic acid and nuclear hormone receptors such as NR2F1, NR2F2, and CRABP1, MEIS1 and MEIS2¹⁶, were enriched in the chromatin of CS cells. Lastly, inspection of disparities between the chromatome and scRNA-seq, highlighted that PAX3 displays equal association with

chromatin in CFS and CS cells, and the neural crest marker TFAP2A is associated with chromatin of CS cells. These data aligned the NMP-like phenotype of CFS cell with the involvement of Wnt/β-catenin in NMP differentiation to paraxial mesoderm⁵², and the neural tube progenitor phenotype of CS cells with the role of retinoic acid in posterior development of the neural tube.⁵³

At the next stage, we analyzed the proteome data to gain insights about metabolic and bioenergetics regulation of CFS and CS cell lines. Unsupervised clustering revealed grouping of differentially expressed proteins in 7 clusters with clusters 1 and 2 being specific to CFS and CS cells respectively. GO terms analysis of cluster 1 supported the NMP association of CFS cells, displaying enrichment in terms related to mesoderm, nerve, axonal and glial cell development, while the neural tube progenitor association of CS cells was apparent in cluster 2 with enrichment of neuronal development terms (Figure 3E). Notably, the presence of hundreds of metabolic and mitochondrial proteins in the upregulated clusters 3 and 4 pointed to a metabolic reprogramming during the transition from pluripotency to CFS and CS cell states. Specifically, cluster 3 encompassed terms related to the TCA cycle, acetyl-CoA catabolism, and protein catabolism. Aligning with the elevated abundance of proteins implicated in oxidative phosphorylation, cluster 4 highlighted terms pertaining to mitochondrial biogenesis and mitochondrial protein translation elongation, initiation, and termination. To determine the energetics outcomes we conducted Seahorse assays, revealing that while CFS, CS, and ES cells have a similar anaerobic glycolytic capacity, CFS and CS cells produced a substantially higher total amount of ATP via oxidative phosphorylation relative to ES cells (Figure 3F). Moreover, the maximal potential for ATP production via oxidative phosphorylation of CFS and CS cell lines was found to be approximately 5- and 3-fold higher than ES cells, respectively. We observed several isoforms of PFK, which facilitate irreversible commitment of fructose-6-phosphate to glycolysis, among the most sharply downregulated proteins in CFS and CS cells (Figure S4G). This implied that to maintain a similar glycolytic capacity in CFS, CS and ES cells other enzymes must be compensating. Accordingly, candidate compensatory enzymes, FOXK1 and FOXK2, were found to be enriched in CFS and CS cells (Figure S4G). Lastly, GO terms that were associated with clusters 6 and 7 revealed the downregulation of intracellular membrane trafficking, glycoprotein biogenesis, ion transports in CFS and CS cell lines. This multifaceted regulation is likely linked to the motile behaviour of CFS and CS cells, manifested by the downregulation of tight junction proteins such as CLDN6, upregulation of N-Cadherin, and induction of EOMES, which promotes epithelial-to-mesenchymal transformation (EMT) in mesoderm progenitors (Figure S4G).⁵⁴ We concluded that the bioenergetics of human CFS and CS cells are significantly elevated compared with PSCs due to a global reprogramming of metabolic and mitochondrial proteins.

Differentiation of CFS and CS to neural tube and neuroglia

We next explored the correspondences between the NMP characteristics of CFS cells, the dorsal neural tube phenotype of CS cells, and the differentiation outcomes. We induced neuronal

differentiation by neurotrophic factors⁵⁵, conducted time-course scRNAseq, imaged cellular morphologies, and measured electrophysiological activity (Figure 4A). Cells were classified based on markers indicative of ventral and dorsal spinal cord subdomains, which included motor neurons, interneurons, trunk neural crest, preEMT cells, neuroglial cells, and sensory neurons from the dorsal root ganglion (Figure S5A).^{25,32} Time-lapse imaging revealed neuronal morphologies emerging as early as days 3-4 for the CS cell lines and by day 7 for the CFS lines (Figure 4B, Video S2). Consistently, approximately 80% of CS cells were in the G1 phase by the second day of differentiation, suggesting a transition to a postmitotic state earlier than CFS cells (Figure 4C). Moreover, the differentiation of CFS cells produced a population of CS-like cells that exhibited distribution of cell cycle phases indicating a proliferative CS population in the differentiation hierarchy (Figure 4C and Figure S5B). These findings align with the abovementioned evidence that CFS cells, as parental NMPs, possess a broader differentiation potential compared than CS cells.

We subsequently generated cell clusters in the scRNA-seq data, and analyzed the clusters according to the ventral-to-dorsal ordering of the neural tube domains and spinal cord cell types, including of the trunk neural crest (Figure 4D). In differentiated CFS cells from day 14-28 we noticed the expression of V0-V1 ventral interneuron markers, such as *EVX1* and *EVX2*, as well as *EN1*, across several clusters, but not in differentiated CS cells (Figure 4E and S5B). Additionally multiple cell clusters showed expression of the motor neuron marker *MNX1*, with some showing expression with *EN1*. This pattern suggested that the ventral differentiation of CFS lines occurs via ventral progenitor intermediate(s) which are distinct from CS cells. Assessing dorsal neural tube markers, we identified distinct clusters expressing dl2-3 and dl5-6 markers, including coexpression of *POU4F1* with *POU4F2*, *FOXD3*, or *OTP* (dl2 and dl3 respectively) on day 14 and 28 CFS cells, and expression of *LMX1B* and *DMRT3* (dl5 and dl6 respectively) in both CFS and CS cells (Figure S5C). This indicated that both CFS and CS possess the potential to differentiate into dorsal spinal cord interneurons.

Additionally, we examined markers of sensory neurons and neuroglial cells. We noticed in separate clusters coexpression of *POU4F1* with sensory neuron markers *DRGX* and *TAFA1*, or *ISL1* and *PIEZ02* (Figures 4D and S5C), thus indicating potential of CFS and CS lines to differentiate into two subtypes of sensory cells. Inspection of the coexpression of preEMT neural crest markers *BMPER* and *GDF7*, as well as the Schwann cell markers *SOX10*, *MPZ*, *S100B* and *CDH19* showed that they were detected in separate clusters of differentiating CFS cells. The cells with mesenchymal morphology that were noticed during the differentiation of CFS cells may correspond to these populations (Figures 4B,E and S5C). Additionally, we identified clusters that coexpressed the neural crest markers *ETS1* and *PRRX1* during the differentiation of both CFS and CS cells. Yet apart from EMT markers including *SNAI1*, *SNAI2*, and *TWIST1*, other well established were characterized lineage markers were absent. This suggested that these clusters might represent a transient neural crest state populated with proliferative cells, as evidenced by the

high proportion of cells in the G1/S phase (Figure 4C). We confirmed the differentiation of cells displaying motor and sensory neuron markers in additional CFS and CS cell lines using RT-qPCR and immunocytochemistry (Figure S6A,B). Taken together, these data indicate that CFS cells possess a broad differentiation potential that spans the dorsoventral axis of the spinal cord, dorsal root ganglion and neuroglia, including CS-like progenitors that are restricted to dorsal interneurons and sensory neurons.

Finally, we investigated the electrophysiological properties of CFS and CS neurons using whole-cell patch-clamp recordings. Voltage-clamp experiments at 2-3 and 7-8 weeks of differentiation revealed fast inward Na^+ currents followed by slow outward K^+ currents (Figures 4F and S6C). Further, current-clamp experiments demonstrated action potentials of consistent amplitudes when neurons were stimulated to their threshold potential (Figure 4F). Remarkably, input resistance and resting membrane potentials were similar at early and late time points, which indicate early maturation (Figure S6C). We confirmed these results by analyzing HMGU1 CFS and CS neurons which consistently generated action potentials albeit their maturation was slower (Figure S6D,E). Lastly, we noticed spontaneous excitatory postsynaptic currents in CFS neurons at the late time point, indicating network activity and formation of functional synapses (Figures 4F and S6D). In conclusion, we demonstrate that both CFS and CS neurons can achieve functional maturity in just two weeks of differentiation.

Embryonic chimerism and mesoderm potential of CFS cell lines

We first assessed the capacity for paraxial mesoderm differentiation by directing human CFS and CS cells towards skeletal myocytes and engrafting mouse CFS cells in embryos. To achieve this, we modified a 2D protocol, specifically omitting the NMP induction step (Figure 5A).⁵⁶ After a 5-week differentiation period, CFS cells demonstrated upregulation in skeletal muscle markers *MYOD* (except CFS_3), *CDH15* (M-cadherin), and *MYOG*^{57,58}, whereas the CS cells did not (Figures 5B and S7A-C). Notably, early paraxial mesoderm markers *TBX6* and *MSGN* were not expressed. Immunocytochemistry confirmed the expression of myocyte markers *MYOD*, *MYOG*, M-cadherin, and myosin heavy chain (*MYHC*), and revealed multinucleated cells in differentiated CFS cells but not in CS cells (Figures 5C and S7A-C). To evaluate muscle innervation, we cocultured C2C12 mouse myotubes with human CFS derived neurons, observing acetylcholine receptor (AChR) clusters in close proximity to myocytes, suggesting neuromuscular junction formation (Figure 5D). These findings highlight the alignment of CFS lines with NMPs, showcasing their potential for paraxial mesoderm differentiation.

Significantly, we conducted ex vivo studies in which CFS cells were functionally challenged for integration within the endogenous NMP environment and differentiation according to intrinsic cues. Mouse CFS cells that were derived from EpiSCs (mCFS_1 line) were labelled with lipophilic dyes (Dil or DiO) and then grafted into the node/streak border region of E8.5 (6-8 somite stage) mouse embryos (Figure 5E). Post injection, these embryos were cultured ex vivo

for 24-48h. At both 24h and 48h (E9.5 and E10.5), we noticed a significant contribution of graft-derived cells in the presomitic mesoderm and the neural tube (Figure 5F). This was verified by overlapping Sox2 staining with Dil (Figure 5G). To further assess the contribution of mouse CFS to somitic mesoderm structures, grafted E10.5 embryos were immunostained using a Pax3-specific antibody. Given that Pax3 is expressed both in somites and the neural tube^{32,48}, we concurrently performed immunostaining with a Sox2-specific antibody to distinguish somite integration.^{59,60} Remarkably, we observed Dil positive somites that expressed Pax3 but not Sox2 (Figure 5H). Additionally, we analyzed the sections for Meox1 expression, a known marker somite marker^{48,59}, and identified Meox1 positive cells overlapping with Dil in somites (Figure S7D). Moreover, within the tail bud region where the endogenous NMP cell population of the chordo-neural hinge (CNH) resides at E10.5, we identified intermingled cells with triple labelling of Dil, Sox2, and Brachyury, suggesting integration of grafted CFS cells in the NMP niche (Figure 5I). In all these cases, we observed similar levels of immunofluorescence in host and their equivalent graft-derived cells. Moreover, we identified in some of the embryos sections where a labelled secondary neural tube (Figure S7E), indicated that grafted cells had formed ectopic neural tube. In conclusion, our data highlight that the NMP characteristics of CFS cells functionally correspond with robust contributions to the paraxial mesoderm, neural tube, and the native niche of NMPs. Thus, CFS cell lines fulfil the classification of multipotent self-renewing *in vitro* maintained NMP-like stem cells.

DISCUSSION

We explored multipotent AxSCs that embody NMPs and posterior neural tube progenitors because knowledge about human embryogenesis around Carnegie stages 6-12, the period of gastrulation, somite, and neural tube development, is very limited. Similar to NMPs and neural tube progenitors, the pluripotent cells present in periimplantation embryos are transient. However, these cells can be converted to self-renewing ES cells by modulation of self-renewal and differentiation pathways.^{1-4,42} Our success in converting both human and animal ES and iPS cells into NMP- and neural tube-like AxSCs leverages these insights. Rather than utilizing embryos, we subjected human ES cells to a brief period of NMP induction⁶⁰, subsequently engaging in the serial passaging of cells that exhibited expression of critical NMP markers such as T, CDX2, and SOX2, but lacked expression of pluripotency factors. This likely provides a competitive advantage for the NMP- and posterior neural tube-like cells to outgrow other cell types, analogous to the growth advantage of ES cells over extraembryonic cells.⁶¹ The maintenance of telomeres is also likely crucial, suggesting a potential for unlimited expansion of undifferentiated AxSCs. Although it has not been directly proven, the derivation of CS lines (N-AxSCs) from ES and iPS cells likely occurs via one or more transient NMP states. This inference is drawn from the evidence that we were able to derive CS cell lines from CFS cells, and that CS-like cells emerged during the neural differentiation of CFS cell lines. Supporting this notion, CS cells only differentiated into the dorsal subset of neural tube cells that CFS cells are capable of differentiating into.

The discovery of N-AxSCs is paramount, as these stem cells represent posterior neural tube progenitors markedly different from those derived through the dual SMAD inhibition protocol used in differentiating anterior neurons and cerebral organoids.^{62,63} Accordingly, in terms of the molecular phenotype, we identified differences between NPCs derived through dual SMAD inhibition and AxSCs. Moreover, from a signaling perspective, the necessity for WNT/β-catenin signaling in the self-renewal of AxSCs aligns with the expression of Wnt ligands in the tailbud region.^{7,21,64} Given the high effectiveness of AxSCs for neural differentiation in terms of purity and speed, e.g. demonstrated by the emergence of electrically active cells within 14 days of differentiation, we posit that the derivation of both types of AxSCs can establish a pivotal framework for probing the mechanisms underlying both normal and aberrant human posterior development. As an initial demonstration, we have embarked on deciphering the lineages and ontogeny of the human neural tube, occurring at developmental time points where embryo analysis is severely constrained. Our current model suggests that the bifurcation of the neuroglial lineage from trunk neural crest takes place before the transition from NM-AxSCs to N-AxSCs, with the latter representing a developmental stage preceding the neural crest's delamination from the neural tube.

As N-AxSCs were derived both following NMP induction and directly from NM-AxSC lines by cultivation in CS media, we postulate that N-AxSCs act as a posterior neural tube stem cells,

hierarchically downstream to NMPs. In support of this, we detected a subset of differentiated NM-AxSCs closely resembling N-AxSCs. Furthermore, N-AxSCs, in contrast to NM-AxSCs, did not differentiate into myocytes and specifically expressed canonical neuronal progenitor markers, including anterior trunk HOX genes. This differentiation was constrained to the dorsal subsets of neurons, a pattern also observed in NM-AxSC differentiation. Such a multistep posterior hierarchy, embodied by NMP- and neural tube-like stem cell states, can offer a robust framework for exploring mechanisms underlying both normal and aberrant human posterior development.

Human NM-AxSCs express trunk HOX genes and non-canonical Wnt ligands, however akin to late mouse NMPs, they lack the posteriormost HOX genes.²¹ This raises the possibility that NM-AxSCs may emulate late NMPs. *MNX1* expression in human NM-AxSCs supports this notion, as it is specifically expressed in late NMPs.¹³ In the process of deriving mouse CFS cells from embryos, we have noted both HES5 positive and negative mouse NM-AxSCs from early and late somite stages, respectively, for the first 10 passages, suggesting that additional state(s) of AxSCs might be captured. Collectively, our results point to the potential of deriving additional AxSC states using these conditions.

While the pluripotency circuitry of OCT4 and NANOG is evidently terminated in both NM- and N-AxSCs, the stem cell characteristics manifested in these multipotent stem cells are likely connected to expression of pluripotency-and reprogramming associated factors, namely MYCN, LIN28B, and SALL4.³⁸ SALL4 may be of particular relevance here, given its crucial role in the proliferation of NMPs.⁶⁵ The exchange of chromatin factors and metabolic reprogramming may also be tied to the transition of pluripotency to AxSC states. Both NM- and N-AxSCs exhibit significantly higher TCA cycle and OXPHOS activity than ES cells. This suggests a connection between catabolism of acetyl-CoA and a decline in DNA methylation and histone acetylation, processes which are associated with exit from pluripotency.⁶⁶ Of note, the specific concentrations of CHIR99021 are crucial for induction and maintenance of AxSCs as we discovered that the derivation of human and murine AxSCs is possible within a narrow range of CHIR99021 concentration. The importance of cultivation of AxSCs under hypoxia might be connected with the enhanced stability of NMPs in vitro when maintained in hypoxic conditions.³⁵ Thereby, the capture of AxSCs lays an essential groundwork for exploring the epigenetic mechanisms and other pivotal factors that drive the differentiation of NMPs and neural tube progenitors. Such insights hold the potential to guide the direct reprogramming of NM-AxSCs from somatic cells.

Ultimately, the representation of NMP development within AxSCs could pave the way for medical breakthroughs. The identification of stem cells with a limited differentiation capacity is the cornerstone of innovative cell therapies, such as the transplantation of hematopoietic stem cells for treating blood malignancies⁶⁷, and epidermal stem cells for curing fatal skin disorders.⁶⁸ Analogously, the AxSCs that we identified, and potentially future subtypes, represent genetically unmanipulated, specific origins of the neural tube, nerves, and muscles. Thus, there is merit in exploring them in regenerative medicine, for example in spinal cord injury. Fundamentally, AxSCs

are straightforward to derive and cultivate, they are suitable for large-scale cultivation, GMP manufacturing, and cell therapy banking. Since AxSCs are obtained without transgenes, and lack teratoma formation ability, these cell lines, along with their differentiated progeny, are likely to be safer in cellular transplantation compared to cells differentiated directly from ES and iPS cells.

The developmental restriction of AxSCs also signify benefits for investigating abnormal congenital development and peripheral neurodegeneration and neuropathies. Compared to cerebral models employing for example the dual SMAD inhibition protocol⁶⁹, posterior neurulation remains under-characterized. Consequently, AxSCs may serve as pivotal tools, offering an alternative to the forced expression of transgenes or the employment of morphogen gradients^{70–72}, which pose challenges in the production of pure cellular populations. These potential applications hold particular relevance in examining mechanisms linked to spinal cord defects. In both sporadic and familial genetic instances, understanding the disrupted processes that affect the elongation of the neural tube and the formation of complex layers of interneurons, efferent motor neurons, afferent sensory neurons, and neuroglial cells is crucial. Similarly, efforts to model conditions like lower amyotrophic lateral sclerosis (ALS) and severe pain syndromes⁷³, which impair spinal cord motor and sensory neurons, respectively, are poised to benefit significantly.

FIGURE LEGENDS

Figure 1 | Capture and characterization of long-term renewing putative NMP and neural tube multipotent stem cell lines from human ES cells. **A**, A schematic illustrating the screening process for capturing undifferentiated posterior stem cells from human ES cells following 24h WNT/β-catenin induction (CHIR99021) and cell splitting in conjunction with cultivation with cocktails of small molecule and growth factors (as indicated) for 10 passages. Cultures were dissociated to single cells when passaged on Matrigel-coated plates in 5% CO₂ / 5% O₂, and B27 medium without vitamin A or antibiotics. **B**, The green rectangles indicate conditions that enabled successful establishment of cell lines for >10 passages, and gray rectangles represent cell lines that were discontinued after 9 passages. Crossed-out rectangles represent screened conditions that did not support the establishment of renewing cell lines. CFS and CS denote cell lines that were established in CHIR99021, FGF2, and TGF-β pathway inhibitor SB431542, or without FGF2, respectively, with passaging every 3-4 and 5-7 days respectively. The analysis of CH and CF lines is shown in Figure S1E-F. **C**, RT-qPCR characterization of H9 CFS and CS lines using pluripotency and NMP markers. n=3 independent derivations of CFS and CS lines (marked CFS/CS_1-3) induced from different passages of parental H9 ES cells at passage range 55-60 (error bars represent the SEM for three cell line per condition). **D**, Phase contrast images displaying the distinct loose colony morphologies of H9 CFS and CS lines which are markedly distinct from parental H9 ES cells (scale bars: 50 μM). **E**, Representative images of H9 CFS and CS lines (p26-28) immunostained for SOX2, T (BRA), CDX2, and PAX6 (scale bars: 40 μm). **F**, A heatmap derived from the LC-MS chromatome analysis of CFS and CS lines 1-3 (p14-15), and undifferentiated parental H9 ES cells (p61) displaying canonical chromatin bound NMP marker proteins (Figure S2). **G**, UMAP visualization of the single cell RNA-seq data of CFS_2, CS_2 (p14) cell lines, and H9 ES parental cells (10504, 12603 and 8671 single cells respectively), displaying NMP markers. **H**, Assessment of the teratoma formation potential of CFS and CS lines following injection to the kidney capsule of immunodeficient mice. Histological H&E images showing a tumor mass 4 weeks after the injection of control parental H9 ES cells (n=4), and tumor-free kidneys following H9 CFS and CS injection (n=9 and 4 respectively; scale bar: 2 mm). **I-J**, A heatmap (I) based on chromatome mass spectrometry, and a dotplot (J) based on scRNA-seq data, displaying stemness and pluripotency genes. **K**, RT-qPCR quantification of the relative telomere length in CFS_1-3, CS_1-3 lines and undifferentiated H9 ES cells over 10-20 passages.

Figure S1 | Related to Figure 1. **A-D**, Schematic representation of the bulk RNA-seq (**A**) data analysis and representative genes (**B**) following CHIR99021 treatment and doxycycline activation of β-catenin ΔN90 integrated in human H9 ES cells to pinpoint the temporal point of NMP induction. **B**, Note the upregulation of primitive streak *T (BRA)* and posterior embryo *CDX2* transcription factors at 24h, followed by the downregulation of *SOX2*. Additionally, there is a steep rise in mesoderm and endoderm-associated transcription factors and morphogens,

including *TBX6*, *FGF17*, *BMP4*, and *GATA6* between 24-48h. **C**, A heatmap displaying upregulated genes and developmental associations, i.e. ectoderm, mesoderm, neural tube and neural crest. Expression was normalized to untreated undifferentiated cells (n=2). **D**, An analysis of the over-represented signaling pathways in the differentiation time course. Activated pathways at 24h were the basis for screening pathway agonists and antagonists of Wnt/β-catenin, FGF, TGF-β, and BMP pathways. **E-F**, The characterization of putative human CH and CF cell lines, derived by 5 mM CHIR99021 (CH), or CHIR99021 and 100 ng/ml FGF2 (CF) via RT-qPCR (**E**) and immunostaining (**F**) using pluripotency and NMP markers (error bars represent SEM). **G**, UMAP visualization of the single cell RNA-seq data of CFS_2, CS_2, and H9 ES parental cell lines (Figure 1G) displaying *HES5* expression. **H-I**, Histological hematoxylin and eosin images showing representative sections of the injection sites of H9 CFS and CS cell lines (**H**), and a tumor mass displaying trilineage features of teratomas following injection of undifferentiated H9 ES cells (**I**) in the kidney capsule of immunodeficient mice (scale bars: 50 μm).

Figure S2 | Related to Figures 1 and 3. **A**, Schematic detailing the workflow of mass spectrometry proteomes of whole cells and chromatome (chromatin capture) along with the quantification strategy. Chromatome sample preparation was based on Chromatin Aggregation Capture (ChAC) protocol followed by Data-Independent MS Acquisition (DIA). **B**, The total number of identified proteins across various DIA search strategies using MS raw data analysis tools, Spectronaut 16 and DIA-NN 1.8., which facilitated matching peptide identifications between samples. The chromatomes were processed either alone or by including proteome samples, resulting in increased peptide identification through sample matching. Search strategies with an additional Data-Dependent MS Acquisition (DDA-) based peptide spectrum library and/or an additional DIA-based peptide spectrum library were used in parallel. For downstream analysis, we chose a hybrid DDA and DIA library because it yielded the largest number of identified proteins (error bars represent SEM). **C**, R2 correlations of chromatomes and proteomes across H9 ES, CFS_1-3, and CS_1-3 cell lines. **D**, A principal component analysis of the proteomes showing that component 1 and 2 together account for approximately 90% of the variance. **E**, Analysis of total number of proteins that display significant changes between the proteomes and chromatomes of the respective H9, CFS and CS cell lines (Student's t-test, p-value < 0.05, log2 fold change ≥ 1).

Figure 2: Elucidating of the irreversible commitment of CFS and CS cells, derivation from human iPS cells, and mouse CFS lines. **A**, A schematic representation of media swap experiments where CFS were transitioned to CS conditions, and reciprocally, CS were transitioned to CFS conditions. **B**, qRT-PCR analysis and cellular morphology during and following the reciprocal media swap experiments at passages 1-6 (depicted left and right). **C**, Derivation and characterization of CFS and CS cell lines from hiPSC line HMGU1.1-3 and hESC HUES6 with markers as in Figure 1C (gene expression normalized with the respective parental undifferentiated cells). **D-F** A schematic

representation of a strategy for protocol screening for the derivation of mouse posterior stem cells from EpiSCs (p38-p43, E14.1 mESC line). Cells were passaged by single cell splitting to Matrigel (route 1, **E**) or inactivated MEFs followed by transfer to Matrigel at p10 (route 2, **F**). Screening included small molecules and growth factors pertaining to the derivation of human CFS and CS derivation with the addition of mouse FGF8. Green rectangles represent successful derivations of cell lines cultured with cell splitting of >10 passages, labeled 1 through 5 (**E-F**). Phase contrast images of mouse CFS cell lines successfully derived on both Matrigel (mCFS_1) and MEFs-Matrigel (mCFS_2) are shown. **G**, RT-qPCR analysis of a representative mCFS_1 line using pluripotency, NMP, and paraxial mesoderm markers. **H**, A schematic illustration detailing the strategy for dissection of the CLE region of E8.5 mouse embryos and grouping according to somite stage 7/8 and 16/17. Cells were dissociated and plated on Matrigel in mouse CFS medium as mCFS_1 line. After p7, cells were passaged either on Matrigel or MEFs. **I**, Phase contrast images showing the distinct colony morphologies of CFS cells derived from 7/8 and 16/17 somite stage (SS) embryos (scale bars: 50 μ M). **J**, RT-qPCR analysis of mouse CFS lines derived from embryos (p11), using pluripotency, NMP, and paraxial mesoderm markers (error bars represent SEM).

Figure S3 | Related to Figure 2. **A**, Venn diagrams representing the total count of the genes expressed in >50% of the cells in human ES H9, HUES6, and iPS cell HMGU1 derived CFS (p15-18) and CS (p12-14) lines based on scRNA-seq (CFS H9: 3215, CFS HUES6: 2660, CFS HMGU1: 7059, CS H9: 6777, CS HUES6: 833, CS HMGU1: 8757 single cells). **B**, Representative immunostaining images of mCFS_1 (p21) for Sox2, Brachyury (T ortholog), and Cdx2 (scale bar: 20 μ m). **C**, RT-qPCR analysis of mouse CFS lines for pluripotency, NMP, and paraxial mesoderm markers corresponding to mCFS_2-5 lines related to Figure 2E-F (relative expression normalized to parental cells (n=4). **D**, Schematics of protocol used for the derivation of Sumatran orangutan (*Pongo abelii*) female iPS cells (Pongo) CS_1-2 cell lines. **E-F**, RT-qPCR and representative immunofluorescence images used for characterization of Pongo CS_1-2 lines (error bars represent SEM, and scale bars: 20 μ m).

Figure 3: Omics and metabolic characterization of human CFS and CS cell lines. **A**, On the left a model of an E8-E10.5 mouse embryo illustrating the approximate locations of transcription factors expressed in NMPs and axial structures, including the neural tube and somitic mesoderm along the rostrocaudal axis. On the right a model of the neural tube progenitor domains at around Carnegie stages 12-19.³² **B**, The expression of the markers categorized as per (A), is analyzed based on scRNA-seq data of CFS_2, CS_2 and ES cells H9 (10504, 12603 and 8671 single cells respectively). **C**, A principle component analysis (PCA) derived from the chromatome data of CFS_1-3, CS_1-3 and ES cells H9, highlighting component 1 and 2 which combined, account to approximately 70% of the variance. **D**, The figure presents Individual loadings of the PCA with NMP and pluripotency markers, HOX proteins, and downstream factors of the WNT and retinoic

acid pathways. **E**, A heatmap displaying seven clusters of differentially expressed proteins based on a global proteome analysis of CFS_1-3, CS_1-3 and ES cells H9. Selected subset of the GO terms corresponding to the respective clusters are showed. **F**, Figure comparing the extracellular acidification rate (ECAR) and oxygen consumption rate (OCR) to pinpoint metabolic differences and energy production pathways in CFS_2, CS_2 and H9 ES cells. ECAR and OCR rates indicate glycolytic metabolism, and mitochondrial oxidative phosphorylation metabolism together with non-mitochondrial respiration, respectively. ECAR and OCR are normalized by cell number. Error bars represent SEM; unpaired t-tests. ns, not significant; *, p<0.05; **, p<0.005; ****, p<0.0001.

Figure S4 | Related to Figure 3. **A**, The top 50 differentially expressed genes in the scRNA-seq data of CFS_2, CS_2, and ePSCs-H9, as described in Figure 3B. Markers that provide additional insights regarding CFS cell lines include MLLT3, a hematopoietic regulator and downstream effector of T(BRA)⁷⁴, SPRY4 a promoter of ESC stemness^{75,76}, and SPRY2 which is associated to with both mesodermal and neural development.⁷⁷ Markers that provide additional insights regarding CS cell lines include, PRTG, which is suggested to suppress of neural differentiation in early mouse embryos⁷⁸, ZFHX4 which is upregulated during neural commitment⁷⁹, ZFHX3 and ZIC2 both involved in the regulation of both muscle and neural development⁸⁰⁻⁸³ and CRABP1 which is a target of RA receptors. **B**, Schematics of the dual SMAD inhibition differentiation protocol of neural progenitor cells (NPCs). **C**, RT-qPCR characterization of NPCs using pluripotency and NPC markers. **D-E**, UMAP visualization (**D**) and a dotplot (**E**) of the single cell RNA-seq data of CS_2 (Figure 1G), and NPC (p3) lines displaying dorsoventral spinal cord markers and HOX genes. **F**, The expression of the markers categorized as per Figure 3A and 3B in the scRNA-seq data of human ES H9, HUES6, and iPS cell HMGU1 derived CFS and CS cell lines (Figure S3A). **G**, Protein rank based on the relative log2 fold change with respect to proteomes of CFS_1-3 versus H9 ES cells, CS_1-3 versus H9 ES cells, and CFS_1-3 versus CS_1-3. The annotation of proteins is based on CFS, CS and pluripotency markers and GO term categories in Figure 3E.

Figure 4: Single cell RNA-seq and electrophysiology of the neural differentiation of human CFS and CS cell lines. **A,B**, A schematic illustration of the neural differentiation protocol omitting NMP induction (**A**), and the corresponding phase contrast images at the respective time points (**B**) (scale bar: 50 μ m). CFS_2 and CS_1 cell lines were dissociated and subject 10x chromium scRNA-seq at the indicated time points. **C**, UMAP analysis and quantification of G1, G2M and S cell cycle phases derived from the time course scRNA-seq data. **D**, An illustration depicting marker genes of the annotated cell types along the ventral – dorsal axis of the neural tube, and arrows connecting CFS and CS cell lines with the respective progeny (preEMT NC: pre-epithelial-to-mesenchymal transition neural crest cells, NC_d: neural crest derivative). **E**, Annotations of cell clusters in CFS Day14 (upper panel), CFS Day28 (middle panel) and CS DAY14 (lower panel) within the scRNA-seq data based on markers identified in **D** and Figure 5SA. Clusters representing cells

that display similar transcriptional signatures are labelled with the same coloring scheme across samples. Expression of ventral spinal cord neuron markers (*MNX1*, *EN1*, *EVX1*, *EVX2*) in CFS datasets, and dorsal spinal cord neuron markers (*LMX1B*, *DMRT3*, *POU4F1*, *ISL1*) in CS dataset are color labelled in corresponding UMAP plots (right panel). **F**, Representative current- and voltage-clamp data illustrating passive and active membrane properties of CFS_2 and CS_2 cells, at time points 2-3 weeks and 7-8 weeks into neuronal differentiation. Displayed are voltage clamp recordings (left column), current clamp recordings of evoked action potentials (AP, middle column) and spontaneous excitatory postsynaptic currents (right column). Currents in the left column were evoked in response to increasingly depolarizing voltage (steps of 20mV; 300ms). Fast-inward Na^+ current (downwards) were followed by slow outward K^+ currents (upwards). Voltages shown in the middle column were responses to both hyperpolarizing (downward) current injection and depolarizing (upward) current injection (steps of 20 pA; 300ms), showing multiple evoked APs in response to continued stimulation. The current trace in the right column reveals spontaneous excitatory postsynaptic currents during CFS cell differentiation, indicating spontaneous neural network activity with functional synapses. **G**, Summary of peak Na^+ currents, AP peak amplitudes, and the total number of evoked APs per cell at each time point for measurements of 14-21 days CS, n=23; CFS, n=13; 49-56 days CS, n=11; CFS, n=16.

Figure S5 | Related to Figure 4. **A**, An overview of neural tube domains and their associated markers. **B**, Dotplots showing the expression of undifferentiated CS markers in clusters from CFS and CS neural differentiation (Figure 4E). In the same manner, dotplots representing ventral spinal cord neuron markers in differentiated CS cells. **C**, Dotplots showcasing the (co)expression of marker genes corresponding to the cell clusters identified in Figure 4D,E.

Figure S6 | Related to Figure 4. **A**, Replications of the neural differentiation of additional CFS and CS cell lines. RT-qPCR analysis of *ISL1*, *MNX1* and *POU4F1* at differentiation Day28 (Figure 4A) showing H9 ES cell derived CFS lines 1-3 and CS lines 1 and 2 (Figure 1C) (P: parental undifferentiated CFS and CS cells, N: neural differentiation day 28, error bars represent SEM). **B**, Confirmation of the sensory and motor neuron differentiation by immunostaining and imaging. Day28 cells were stained with BRN3A, *MNX1*, *ISL1*, and TUBB3 antibodies (scale bar: 50 μM). **C**, A summary of the peak K^+ currents, cell input resistance, and resting membrane potential at each time point for measurements of 14-21 days CS, n=23; CFS, n=13; 49-56 days CS, n=11; CFS, n=16). **D**, Representative current- and voltage-clamp data illustrating the passive and active membrane properties of HMGU1 CFS and CS cells at the indicated times of differentiation. Displayed are voltage clamp recordings (left column), current clamp recordings of evoked action potentials (AP, middle column) and spontaneous excitatory postsynaptic currents (right column). Currents shown in the left column were evoked in response to increasingly depolarizing voltage (steps of 20mV; 300ms). The fast-inward Na^+ current (downwards) was followed by slow outward K^+

currents (upwards). Voltages shown in the middle column were in response to both hyperpolarizing (downward) current injection and depolarizing (upward) current injection (steps of 20 pA; 300ms), showing multiple evoked APs in response to continued stimulation. The current trace in the right column reveals spontaneous excitatory postsynaptic currents during CFS differentiation, indicating spontaneous neural network activity with functional synapses. **E**, A summary of peak Na⁺, peak K⁺ currents, and the total number of evoked APs at each time point for measurements of HMGU1 14-21 day CS cells, n=23; CFS cells, n=6; 49-56 days CFS cells, n=28.

Figure 5: Skeletal muscle differentiation and E8.5 mouse embryo complementation assays with CFS cells. **A-C**, A schematic illustration of skeletal muscle differentiation omitting NMP induction (A) an RT-qPCR analysis (B) and immunostaining (C) of NMP, somitic mesoderm, and skeletal muscle markers at Day35-38 (P: parental undifferentiated CFS cells, SKM: skeletal muscle differentiation of CFS cells, error bars represent SEM, scale bars: 50 μm and 20 μm for high and low magnification). **D**, Immunofluorescence analysis of neuromuscular junction formation by α-bungarotoxin staining of day 21 neuronally differentiated CFS cells that were cocultured with C2C12 myoblasts for additional 12 days. **E**, An illustration depicting the microcapillary grafting of fluorescently labeled (Dil or DiO) undifferentiated mouse CFS cells (Figures 2D-E, G and S3B) into E8.5 mouse embryos in the node-streak border (NSB) region, and regions where fluorescently labeled cells were detected at the 24h or 48h follow-up time points. **F**, A table summarizing the number of embryos transplanted at E8.5, and observed Dil or DiO signal in axial structures, as follows, somites (som), neural tube (NT), presomitic mesoderm (PSM), chordo-neural hinge (CNH) and tailbud mesoderm (TBM) (Emb: embryo, Ect: ectopic, sp: somite pairs). Additionally, the notochord was labelled in 2 embryos, and labelled putative neural crest cells were observed in an additional 2 embryos. In 1 embryo a clump of labelled cells was present in the hindgut lumen. **G**, Representative images showing the transplantation site of mouse CFS cells labeled with Dil (E8.5) and incorporation into somites (som) and neural tube (NT) at the E10.5 follow-up. **H**, A transverse section of embryo 2 showing Dil labelled cells (red) in the neural tube, and Sox2 (green) immunostaining of the same section, and an overlay image. The white outline shows the position of Dil-labelled cells, noting that the Sox2 signal in the grafted cells was at a similar level to the endogenous host cells. **I**, A transverse section of embryo 4 showing Dil label in one of the somites. Maximum intensity projections of z-stack images taken from the same section with double-immunostaining of Sox2 (green) and Pax3 (magenta). **J**, A transverse section of the tail bud of embryo 2 showing Dil labelled cells in the epithelium (arrowhead) and mesenchyme (lower panel, right). Maximum intensity projections of z-stack images taken from the same section double-immunostained for Sox2 (green) and Brachyury T (red). Dil-labelled cells are outlined in white, noting that the signal of Sox2 and T were similar to neighboring endogenous unlabeled host cells.

Figure 7S | Related to Figure 5. **A**, Representative images of multinucleated cells in skeletal muscle differentiation of CFS cells at day 38 that were immunostained as in Figure 5C (scale bars: 50 μ M for phase contrast and 20 μ M for immunostaining images). **B,C**, The phase contrast images of day 35 differentiation of CS cells (scale bar: 50 μ m) (B), and RT-qPCR analysis (C) of markers as in Figure 5B at Day 35 (ND: not determined, P: parental undifferentiated CFS cells, SKM: skeletal muscle differentiation of CFS cells). **D**, A transverse section of embryo 2 (Figure 5H) showing Dil-labelled cells in the neural tube (NT) and somite (som). Immunostaining images of Sox2 (green) and Meox1 (red) are shown below. Labeled cells correctly express Sox2 and Meox1 in NT and somites respectively. **E**, Representative images showing the transplantation site of mouse CFS cells labeled with DiO at E8.5 (i) and incorporation into presomitic mesoderm (PSM), som, NT, chordo-neural hinge (CNH) and tailbud mesoderm (TBM) at E9.5 (ii). Embryos sectioned at the indicated site at E10.5 (iii), and immunostained for Sox2 (red) and Hoechst (blue). Cells arranged in auxiliary neural tube structures were derived from mouse CFS cells (green: DiO labeled) (iv-xii).

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit anti-Sox2	Cell Signaling Technology	Cat#: 2738
Rabbit anti-Brachyury	Abcam	Cat#: ab209665; RRID:AB_2750925
Goat anti-Brachyury	R&D Systems	Cat#: AF2085; RRID:AB_2200235
Rabbit anti-CDX2	Cell Signaling Technology	Cat#: 12306; RRID:AB_2797879
Rabbit anti-Pax6	Biolegend	Cat#: 901301; RRID:AB_2565003
Rabbit anti-Islet 1	Abcam	Cat#: ab20670; RRID:AB_881306
Mouse anti-Brn3a	Santa cruz	Cat#: sc-8429; RRID:AB_626765
Mouse anti-Mnx1	Santa cruz	Cat#: sc-515769
Chicken anti-B3 tubulin	Novus	Cat#: NB100-1612; RRID:AB_10000548
Mouse anti-Myod	Life Technologies	Cat#: MA1-41017; RRID:AB_2282434
Rabbit anti-M-Cadherin	Cell Signaling Technology	Cat#: 40491S; RRID:AB_2799178
Mouse anti-Myogenin	Life Technologies	Cat#: 14-5643-82; RRID:AB_1907431
Mouse anti-Sox2	R&D Systems	Cat#: MAB2018; RRID:AB_358009
Mouse anti-Pax3	DSHB	RRID:AB_528426
Goat anti-Meox1	Santa cruz	Cat#: sc-10185; RRID:AB_2143425
Mouse anti-neurofilament	DSHB	RRID: AB_531793
Rabbit IgG isotype control	GeneTex	Cat#: GTX35035; RRID:AB_10623175
Mouse IgG isotype control	GeneTex	Cat#: GTX35009; RRID:AB_10618187
Goat anti-Chicken IgY (H L) Secondary Antibody, Alexa Fluor 594	Life Technologies	Cat#: A11042; RRID:AB_2534099
Goat anti-rabbit IgG (H+L), Alexa Fluor 488, highly cross-absorbed	Life Technologies	Cat#: A11034; RRID:AB_2576217
Goat anti-mouse IgG (H+L) Alexa Fluor Plus 647, highly cross-absorbed	Life Technologies	Cat#: A32728; RRID:AB_2633277
Donkey anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 488	Life Technologies	Cat#: A21202; RRID:AB_141607

Donkey anti-Goat IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 594	Life Technologies	Cat#: A11058; RRID:AB_142540
F(ab')2-Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 647	Life Technologies	Cat#: A21246; RRID:AB_1500778
Alexa Fluor® 647 AffiniPure Goat Anti-Mouse IgG (H+L)	Jackson	Cat#: 115-605-146; RRID:AB_2338912
Goat anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 488	Life Technologies	Cat#: A11001; RRID:AB_2534069
Cy™3 AffiniPure Goat Anti-Mouse IgG, F(ab') ₂ fragment specific	Jackson	Cat#: 115-165-006; RRID:AB_2338681
Chemicals and recombinant proteins		
DMEM/F12	Life Technologies	11320074
DMEM, high glucose, pyruvate	Life Technologies	41966029
StemMACS iPS Brew XF	Miltenyi Biotech	130-104-368
KnockOut DMEM	Life Technologies	10829018
Neurobasal medium	Life Technologies	21103049
Neurobasal-A medium	Life Technologies	10888022
RPMI-1640 with L-Glutamine	Life Technologies	11875093
StemFit Basic02	Nippon Genetics	Basic02
B27 supplement	Life Technologies	5001207
B27 supplement without insulin	Life Technologies	A1895601
B27 supplement without vitamin A	Life Technologies	12587001
Insulin-Transferrin-Selenium (ITS -G)	Life Technologies	41400045
Insulin-Transferrin-Selenium-Ethanolamine (ITS-X)	Life Technologies	51500056
N2 supplement	Life Technologies	17502048
Recombinant Human/Murine/Rat Activin A	Peprotech	120-14E
L-Ascorbic acid 2-phosphate sesquimagnesium salt hydrate	Sigma-Aldrich	A8960
Recombinant Human/Murine/Rat BDNF	Peprotech	450-02
Recombinant Human BMP-4	R&D Systems	314-BP
Adenosine 3',5'-cyclic monophosphate sodium salt monohydrate	Sigma-Aldrich	A6885
CHIR 99021 trihydrochloride	Biotechne	4953
Compound E	Biotechne	6476
DMH-1	Biotechne	4126
Purmorphamine	R&D Systems	4551
Recombinant Human FGF-basic	Peprotech	100-18B
Recombinant Human FGF-8a	Peprotech	100-25A
Recombinant Human GDNF	Peprotech	450-10
Recombinant Murine HGF	Peprotech	315-23
Recombinant Murine IGF-I	Peprotech	250-19
LDN-193189 (hydrochloride)	Biomol	Cay19396
Recombinant mouse LIF protein	Millipore	ESG1106
Retinoic acid	Sigma-Aldrich	R2625
SB431542	R&D Systems	1614
Recombinant Human Sonic Hedgehog	Peprotech	100-45
Recombinant Human TGF-β1	Peprotech	100-21

Y-27632 dihydrochloride	R&D Systems	1254
Collagen I	BD Corning	354236
Collagen IV	Santa Cruz	sc-29010
Fibronectin human	BD Corning	FALC356008
Matrigel	BD Corning	11543550
Poly-D-lysine	Sigma-Aldrich	A-003-E
Poly-DL-ornithine hydrobromide	Sigma-Aldrich	P8638
0.05% Trypsin	Life Technologies	25300054
Accutase	Sigma-Aldrich	A6964
Passaging solution XF	Miltenyi Biotech	130-104-688
TrypLE Express	Life Technologies	12605010
Versene	Life Technologies	15040-066
Collagenase IV	Life Technologies	17104019
Papain	Sigma-Aldrich	P3125
Bovine Albumin Fraction V	Life Technologies	15260037
Fetal Bovine Serum	HyClone	SH30071.03
Fetal Bovine Serum	PAN-Biotech	P30-3301
Horse Serum	Sigma	H1138
Goat Serum	Cell Signaling Technologies	5425S
GlutaMAX	Life Technologies	35050038
Knockout Serum Replacement	Life Technologies	10828028
Non-Essential Amino Acid	Life Technologies	11140050
Penicillin/Streptomycin	Life Technologies	15070063
2-Mercaptoethanol	Life Technologies	31350-010
16% Formaldehyde (w/v), Methanol-free	Thermo Fisher	28908
Paraformaldehyde solution 4% in PBS	Santa cruz	sc-281692
Ethanol	Carl Roth	9065.2
Hydrochloric acid 32%	Merck Millipore	100319
Isopropanol	Carl Roth	6752.2
Methanol	Serva Electrophoresis	45631.02
NP-40	Sigma-Aldrich	NP40S
cOmplete EDTA-free Protease Inhibitor	Roche	11873580001
Deoxyribonuclease II from bovine spleen	Sigma-Aldrich	D8764-150KU
Trypan Blue Stain (0.4%)	Life Technologies	T10282
Sodium Chloride Solution	Sigma-Aldrich	59222C
Magnesium Chloride Solution	Sigma-Aldrich	M1028
Trizma Hydrochloride Solution pH 7.4	Sigma-Aldrich	T2194-100ML
DAPI	Sigma-Aldrich	10236276001
Hoechst 33342	Thermo Fisher	H3570
α -bungarotoxin	Invitrogen	B13422
ActinRed 555 ReadyProbes reagent	Invitrogen	R37112
Triton X-100	Sigma-Aldrich	X100
Tween-20	Sigma-Aldrich	P9416
Glycine	Sigma-Aldrich	23391.02
HEPES	Thermo Fisher	15630056
Urea	Sigma-Aldrich	U1250

EDTA	Carl Roth	8043.2
UltraPure SDS Solution	Invitrogen	15553027
Sera-Mag beads	Sigma-Aldrich	GE45152105050250
Acetonitrile	Merck Millipore	1000301000
Trifluoroacetic acid	Thermo Fisher	85183
Tris(2-carboxyethyl)phosphine	Thermo Fisher	PG82080
Chloroacetamide	Sigma-Aldrich	C0267
Guanidinium chloride	Thermo Fisher	24110
Ammonium hydroxide	Supelco	1054280250
Formic acid	Sigma-Aldrich	94318
LysC Protease	Thermo Fisher	90051
Trypsin Protease	Thermo Fisher	90058
SYBR Green Mastermix	Applied Biosystems	4309155
Taqman Gene Expression Mastermix	Applied Biosystems	4369514
ProLong Glass Antifade Mountant with DAPI	Life Technologies	P36935
TURBO Dnase	Life Technologies	
Commercial assays		
RNeasy Mini Kit	Qiagen	74106
SuperScript III First-Strand Synthesis System	Life Technologies	18080051
Wizard SB Genomic DNA Purification System	Promega	A2361
Pierce™ BCA Protein Assay Kit	Thermo Fisher	23225
Relative Human Telomere Length Quantification qPCR Assay Kit	ScienCell	8908-SC
Chromium Next GEM Single Cell 3' GEM, Library & Gel Bead Kit v3.1	10X Genomics	1000121
Chromium Next GEM Single Cell 3' GEM, Single Index Kit T Set A	10X Genomics	1000213
Chromium Next GEM Single Cell 3' GEM, Dual Index Kit TT Set A, 96 rxns	10X Genomics	1000215
HumanCytoSNP-12 v2.1 BeadChip Kit	Illumina	WG-320-2101
Agilent High Sensitivity kit	Agilent	5067-4626
SP Reagent Kit v1.5/ 100 cycles	Illumina	20028401
S4 Reagent Kit v1.5 /200 cycles	Illumina	20028313
RNeasy Minelute RNA cleanup kit	Qiagen	74204
Agilent 2100 Bioanalyzer with RNA Pico 6000 kit	Agilent	5067-1513
RiboZero Gold (Human/Mouse/Rat) kit	Illumina	RS-122
TruSeq Stranded total RNA LT kit	Illumina	20020596
Agilent 2100 Bioanalyzer DNA 1000 kit	Agilent	5067-1504
Qubit dsDNA HS Assay Kit	Life Technologies	Q32854
Deposited data		
Human developing spinal cord single cell RNA sequencing	Rayon et al. ³²	GEO: GSE171892
Human reference genome GRCh38	Genome Reference Consortium	https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001405.26/

Experimental models: Cell lines		
H9	WiCELL Research Institute	https://www.wicell.org
HUES6	Rao et al. ⁴⁰	https://pubmed.ncbi.nlm.nih.gov/26748419/
HMGU1	Kunze et al. ³⁹	https://pubmed.ncbi.nlm.nih.gov/29119608/
EpiSCs	Kurek et al. ⁴³	https://pubmed.ncbi.nlm.nih.gov/25544567/
Orangutan iPSC	Geuder et al. ⁴⁴	https://pubmed.ncbi.nlm.nih.gov/33568724/
H9-ΔN90-β-catenin	This paper	
C2C12	ATCC	https://www.atcc.org/products/crl-1772
Oligonucleotides		
Primers	Sigma-Aldrich	Table 1
Probes	Thermo Fisher	Table 1
Software and algorithms		
Quantstudio	Thermo Fisher	https://www.thermofisher.com/de/en/home/technical-resources/software-downloads/quantstudio-12k-flex-real-time-pcr-system.html
ZEN Microscopy Software	Zeiss	https://www.zeiss.com/microscopy/en/products/software/z_eiss-zen-lite.html
Fiji	National Institutes of Health	https://imagej.net/software/fiji/downloads
Python	Python Software Foundation	https://www.python.org/
GraphPad Prism	GraphPad Software	https://www.graphpad.com/features
Adobe Illustrator	Adobe	http://www.adobe.com
Demuxlet	Kang et al. ⁸⁴	https://github.com/statgen/demuxlet

CellRanger v6.1.2	10X Genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/installation
Scanpy v1.6.1	Wolf et al. ⁸⁵	https://github.com/scverse/scanpy
Solo	Bernstein et al. ⁸⁶	https://github.com/calico/solo
CarDEC	Lakkis et al. ⁸⁷	https://github.com/jlakkis/CarDEC
pandas v1.1.3		https://github.com/pandas-dev/pandas
numpy v1.19.5		https://numpy.org/
louvain 0.7.0		https://github.com/traft/Python-louvain
R	The R Foundation	https://www.r-project.org/
shiny v1.7.1		https://github.com/rstudio/shiny
shinydashboard v0.7.2		https://github.com/rstudio/shinydashboard
shinyHeatmaply v0.2.0		https://github.com/cran/shinyHeatmaply
plotly v4.10.0		https://github.com/plotly/plotly.R
heatmaply v1.3.0		https://github.com/talgalili/heatmaply
tidyverse v1.3.1		https://github.com/tidyverse/tidyverse
tidyverse v1.2.0		https://github.com/tidyverse/tidyverse
dplyr v1.0.9		https://github.com/tidyverse/dplyr
ggplot2 v3.3.6		https://github.com/tidyverse/ggplot2
Kallisto v0.43.0_3	Bray et al. ⁸⁸	http://pachterlab.github.io/kallisto/
Tximport v1.10.1	Soneson et al. ⁸⁹	http://bioconductor.org/packages/release/bioc/html/tximport.html

DESeq2 v1.22.2	Love et al. ⁹⁰	http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html
Genomatix software suite	Precigen Bioinformatics	http://mygga.genomatix.de
Enrichr	Kuleshov et al. ⁹¹	https://maayanlab.cloud/Enrichr/
JENSEN tissue expression database	Santos et al. ⁹²	https://tissues.jensenlab.org

DATA AVAILABILITY

Proteome and chromatome data have been deposited in ProteomeXchange with identifier PXD046087, can also be explored at the following integrative website <https://axialstemcells.shinyapps.io/atlas>. Single cell RNA sequencing data is available on NCBI Gene Expression Omnibus under accession number GSE247215. The publicly available bulk RNA sequencing dataset on NCBI Gene Expression Omnibus with accession number GSE130381 was used in this study.

METHOD DETAILS

Maintenance of pluripotent stem cells

All cell culture of all species was maintained at 37°C with 5% CO₂ and 5% O₂ on Matrigel (1:100, BD Corning) coated plates unless otherwise specified. After passaging cells, 10 µM Y-27632 (R&D Systems) was always added to the medium for the first 24 hours. Medium was refreshed daily, unless otherwise specified. All cell culture components were purchased from Life Technologies unless specified otherwise.

Human pluripotent stem cells (hPSCs), including H9 (WiCell Research Institute), HUES6⁴⁰, and HMGU#1³⁹, were cultured using StemMACS iPS-Brew XF medium (Miltenyi Biotec). Cells were split using StemMACS Passaging Solution XF (Miltenyi Biotec) at a ratio of 1:10.

E14 mouse embryonic stem cells (mESCs)-based epiblast stem cells (EpiSCs)⁴³ were cultured in 1:1 DMEM/F12 and Neurobasal medium, supplemented with 0.5X N2, 0.5X B27, 0.033% BSA, 50 µM β-mercaptoethanol, 1X GlutaMAX, 1X penicillin/streptomycin, 20 ng/ml Activin A (Peprotech), 12 ng/ml FGF2 (Peprotech), and 2 µM IWP2 (Santa Cruz). The cells were split using 1X TrypLE Express at a ratio of 1:20-1:40.

Pongo abelii Sumatran orangutan induced pluripotent stem cells (iPSCs)⁴⁴ were cultured in StemFit Basic02 medium (Nippon Genetics) supplemented with 1X penicillin/streptomycin and 100 ng/ml FGF2. The cells were regularly split using Versene at a ratio of 1:10.

Maintenance of C2C12 myoblast cells

C2C12 myoblast cells (ATCC) were cultured in DMEM with 4.5 g/L D-glucose, 4 mM L-glutamine and 1mM pyruvate (Thermo Fisher) containing 10% fetal bovine serum (FBS) and 0.5% penicillin/streptomycin. Cells were cultured 37°C with 5% CO₂. Medium was replenished every 2 days. C2C12 cells were passaged to uncoated culture dishes when cells reach 70-80% confluence using 1X TrypLE Express. Myoblasts differentiation was induced at 90% confluence by changing the medium to DMEM supplemented with 2% horse serum (HS) and 0.5% penicillin/streptomycin. Differentiation medium was refreshed every other day.

Generation ΔN90 β-catenin hESC line

To generate tetracycline-inducible overexpression of constitutively active β-catenin (with a deletion of the first 90 amino acids from the N-terminus), H9 cells were transfected with the PB-GFP-P2A-ΔNβCAT plasmid and a Piggybac transposase-coding plasmid using the P3 primary cell 4D nucleofector kit (Lonza). After 48 hours, the cells were transferred to a 10-cm dish and subjected to selection with 50 µg/ml Hygromycin B for a duration of 2 weeks. Subsequently, the polyclonal stable line was maintained following the same protocol as the parental H9 line, but with the addition of 25 µg/ml Hygromycin B.

Induction of WNT signaling and time course analysis

Cells were split using Accutase (Sigma) at a density of 2.5x10⁵ cells per well in mTesR1 medium. After 24 hours, the medium was replaced with differentiation medium composed of RPMI-1640 with L-Glutamine and 1X B27 without insulin, supplemented with 10 µM CHIR99021 (Tocris). For beta-catenin overexpression the CHIR99021 was substituted with 1 µg/ml doxycycline (Clontech).

Derivation of human axial stem cells

H9 and HUES6 PSCs were adapted to single cell splitting through passaging with TrypLE Express at 37°C for 6 minutes, for a minimum of 3 passages. Subsequently, PSCs (between passage 55-60) were plated at a density of 71.5x10³ cells/cm² in StemMACS iPS-Brew XF. After 24 hours the cells were washed with PBS and incubated for 24 hours in RPMI-1640 supplemented with 1X B27 without insulin and 10 µM CHIR99021. Next, the cells were dissociated using TrypLE Express for 5 minutes at 37°C and centrifuged at 300 g for 3 minutes. The pellet was resuspended either in human CFS medium, (RPMI-1640, 1X NEAA, 1X B27 without vitamin A, 100 ng/ml FGF2, 5 µM CHIR99021, and 10 µM SB43152), or in human CS medium (RPMI-1640, 1X NEAA, 1X B27 without vitamin A, 5 µM CHIR99021, and 10 µM SB43152). For the first 7 passages, the cells were split twice a week at a ratio between 1:5-1:20 using TrypLE Express for 4-5 minutes for CFS and 4 minutes for CS at 37°C. After passage 7, both cell lines were split at a ratio between 1:10-1:20, with CS cells being split once a week and CFS cells being split twice a week, on average.

HMGU1 AxSCs were derived using a similar method with minor differences. The cells were split at a ratio of 1:20 - 1:30 in either human CFS or CS maintenance medium. Routine splitting was performed at a 1:10-1:20 dilution using either StemMACS Passaging Solution XF or Versene.

Derivation of mouse axial stem cells

EpiSCs were split using TrypLE Express and plated at a dilution of 1:20-1:40 directly in mouse CFS medium (RPMI-1640, 1X NEAA, 1X B27 without vitamin A, 50 ng/ml FGF2, 3 μ M CHIR99021, and 10 μ M SB43152). The cells were continuously maintained on either a Matrigel-coated plate (1:100) or on mitotically inactivated fibroblast feeders seeded at a density of 1×10^5 cells/cm² until passage 10. After that, they were maintained on Matrigel-coated plates (1:100). The cells were split two or three times per week by incubating them with TrypLE Express at 37°C for 4 minutes and then seeded at a dilution of 1:20-1:40.

Stem zone regions of Foxa2VenusFusion x mTmG litters were dissected at E8.5. Mice were kept and experiments performed at the central facilities at Helmholtz Munich German Research Center of Environmental Health in accordance with the German animal welfare legislation and acknowledged guidelines of the Society of Laboratory Animals (GV-SOLAS) and of the Federation of Laboratory Animal Science Associations (FELASA). Mice are kept under SPF conditions in animal rooms with light cycle of 12/12 hours, temperature of 20-24°C and humidity of 45-65%. Mice receive sterile filtered water and standard diet for rodents ad libitum. For embryo generation mice at the age of ≥ 8 weeks were used. Embryonic tissues were centrifuged at 400 g for 4 minutes. The supernatant was discarded, and the pellet was dissolved in 0.25% Trypsin, then incubated at 37°C for 15 minutes. After incubation, FBS was added, followed by centrifugation at 400 g for 4 minutes. The pellet was then resuspended in mouse CFS medium and seeded on Matrigel-coated plates. For the first three passages 1X penicillin/streptomycin was added to the medium. The cells were split twice a week by incubating them with TrypLE Express at 37°C for 4 minutes, and then seeded at a dilution of 1:20-1:40. After passage 7, the cells were plated on feeder cells or Matrigel-coated plates.

Derivation of orangutan axial stem cells

Sumatran orangutan iPSCs were treated for 24 hours with an induction medium comprising RPMI-1640, 1X B27 without insulin, and either 5 or 10 μ M CHIR99021. Next, the cells were dissociated using StemMACS Passaging Solution XF and resuspended in human CS medium as described earlier. The cells were split at a dilution of 1:10-1:20 once or twice a week, using StemMACS Passaging Solution XF.

Derivation of human neural progenitor cells

H9 cells were rinsed with PBS twice and incubated with Collagenase IV (2 mg/ml) for 30 minutes at 37°C followed by plating 1:1 onto low-attachment 6-well plate in a medium consisting of

DMEM/F12, 20% KSR, 1% GlutaMAX, 1% NEAA, 1 μ M DMH-1, 10 μ M SB431542, 3 μ M CHIR99021, 0.5 μ M Purmorphamine (PMA), 10 μ M Y-27632. The next 2 days, fresh medium was applied without Y-27632. On Day3 and Day4, the medium was replaced with 1:1 DMEM/F12 and Neurobasal-A medium, 1:100 B27 without vitamin A, 1:200 N2, 1% GlutaMAX, 1 μ M DMH-1, 10 μ M SB431542, 3 μ M CHIR99021, 0.5 μ M PMA. On Day5, the medium was changed to 1:1 DMEM/F12 and Neurobasal-A medium, 1:100 B27 without vitamin A, 1:200 N2, 1% GlutaMAX, 1 μ M DMH-1, 50 μ g/ml Ascorbic acid (AA). On Day6, the Day5 medium with 5 ng/ml FGF2 was added. On Day7, cells were plated on a Matrigel-coated (1:100) 6-well plate at 1:1 ratio in a medium consisting of 1:1 DMEM/F12 and Neurobasal-A medium, 1:100 B27 without vitamin A, 1:200 N2, 1% GlutaMAX, 1 μ M DMH-1, 50 μ g/ml AA, 20 ng/ml FGF2 (NPC medium). NPC medium was changed daily until Day14 and then every other day. On Day14 and Day21, cells were washed with PBS and treated with Collagenase IV (2mg/ml) at 37°C for 15 minutes and TrypLE Express in 37°C for 4 minutes respectively. Cells were plated on Matrigel-coated (1:100) plates at a ratio between 1:4-1:6. After Day21, NPCs were passaged using TrypLE Express when the confluence reached >70%. Cells were maintained at 37°C, 5% CO₂ and 5% O₂.

Neural differentiation from human axial stem cells

Plates were coated with 20 μ g/ml Poly-D-lysine (Sigma) and 20 μ g/ml Poly-ornithine (Sigma) in PBS and incubated overnight at 37°C. The following day, the coating solution was replaced with 20 μ g/ml Collagen I (BD Corning) and 20 μ g/ml Fibronectin (BD Corning) in PBS and incubated overnight at 37°C. The next day, the coating solution was substituted with 10 μ g/ml Collagen IV (Santa cruz) in 0.05M HCl (Merck) and incubated for 2-4 hours at 37°C, followed by 2 washes with PBS. For immunostaining experiments, cells were directly differentiated on either coverslips or chamber slides coated as described above.

Human CFS and CS lines were dissociated using TrypLE Express and plated on the pre-coated plates in the respective medium, at a density of 28.5x10³ cells/cm² for CFS and 71.5x10³ cells/cm² for CS. The next day, the medium was replaced with differentiation medium consisting of a 1:1 mixture of DMEM/F12 and Neurobasal-A medium, 1X B27, 1X N2, 0.1 μ M Retinoic acid (RA) (Sigma), 100 ng/ml SHH (Peprotech), 100 μ M cAMP (Sigma), 10 ng/ml GDNF (Peprotech), 10 ng/ml BDNF (Peprotech), 10 ng/ml IGF (Peprotech), and 0.1 μ M Compound E (Biotechne). The medium was changed every other day.

For myotube-neuronal co-culture, 1.25x10⁴ cells/cm² of CFS cells were plated on Matrigel coated dishes in CFS medium. After 24h, medium was changed to neuronal differentiation medium. Differentiated neuron clusters were harvested after 18 days of differentiation by 10 min incubation with TrypLE Express. Cells were replated on top of three-day old myotubes, which were differentiated from myoblasts separately. Myoblasts were cultured in DMEM with 10% FBS. Cells were grown until 80-90% confluence after which medium was changed to DMEM with 2%

HS. After replating the neural cells on top of the myotubes medium was changed to co-culture medium consisting of DMEM with 2% HS, 10 ng/ml BDNF, 10 ng/ml GDNF.

Skeletal muscle differentiation from human axial stem cells

Human CFS and CS lines were dissociated using TrypLE Express and seeded on Matrigel (incubated for >4 hours). The cells were seeded in their respective medium at a density between 1.5-2.5x10⁴ cells/cm². The following day, the medium was changed to DMEM/F12 supplemented with 1X NEAA, 1X ITS-G, 3 µM CHIR99021, 0.5 µM LDN193189 (Biomol), 5 ng/ml FGF2, and 10 nM RA. On Day 2, FGF2 and RA were replaced with 34 ng/ml SHH. On Day 6, the medium was changed to DMEM/F12 with 1X NEAA, 1X ITS-G, 10 ng/ml FGF2, and 10 ng/ml HGF. On Day 10, the cells were split by washing with PBS and incubating with TrypLE Express for 6-7 minutes at 37°C. The cells were then harvested, centrifuged at 400 g for 4 minutes, and the supernatant was removed. The pellet was resuspended in DMEM/F12 with 1X NEAA, 1X ITS-G, 10 ng/ml FGF2, and 10 ng/ml HGF (Peprotech) and plated at a density of 2-2.8x10⁴ cells/cm². From Day 12 onwards, the medium was changed to DMEM/F12 with 1X NEAA, 1X ITS-X, and 10 ng/ml HGF. Fresh medium was applied every other day during the differentiation process.

Quantitative RT-PCR

RNA extraction was conducted using the RNeasy Mini Kit from Qiagen following the manufacturer's instructions. The extracted RNA samples were then reverse-transcribed using the SuperScript III First-Strand Synthesis System (Life Technologies), according to the manufacturer's instructions. RT-qPCR was performed in 384-well plates using either the Power SYBR Green PCR Mastermix or the Taqman Gene Expression Assay Mastermix (Applied Biosystems), with a total volume of 10 µl. The reaction was carried out under the following conditions: for Power SYBR Green Mastermix, 2 minutes at 50°C, 10 minutes at 95°C, 40 cycles of 15 seconds at 95°C and 1 minute at 60°C; and for Taqman Gene Expression Assay Mastermix, 2 minutes at 50°C, 10 minutes at 95°C, 40 cycles of 15 seconds at 95°C and 1 minute at 60°C. The primers and probes used (Thermo) are listed in Table 1. For the analysis of orangutan AxSCs, human primers were used, except for PAX6. The Delta-Delta Ct method was employed to calculate the relative expression levels, which were normalized to the GAPDH housekeeping gene. The data were visualized using GraphPad Prism 8.0.

Immunocytochemistry

For the goat anti-Brachyury antibody (R&D Systems, Table 2), the following protocol was employed. Cells cultured on either coverslips or chamber slides were washed with 0.1% BSA in PBS and fixed with 4% paraformaldehyde (Thermo) for 20 minutes at room temperature (RT). After two washes with 0.1% BSA in PBS, permeabilization was carried out using 0.2% Triton X-100 (Sigma) in PBS for 10 minutes. Samples were blocked with 5% BSA in PBS for 1 hour at RT and

incubated overnight at 4°C with primary antibodies diluted in 0.5% BSA/1% Tween-20 (Sigma) in PBS. Following two washes with 1% Tween-20 in PBS for 5 minutes each, the cells were treated with secondary antibodies diluted 1:750 in 0.5% BSA/1% Tween-20 in PBS for 3 hours at RT. After two additional washes with 1% Tween-20 in PBS for 5 minutes each and one wash for 10 minutes, coverslips or chamber slides were mounted using Prolong Gold Antifade and incubated overnight at RT. For other antibodies, cells cultured on either coverslips or chamber slides were washed twice with PBS, followed by fixation with 4% formaldehyde in PBS for 15 minutes at RT. After washing with PBS, cells were permeabilized with 0.2% Triton X-100 in PBS for 5 minutes at RT. Subsequently, the cells were blocked with 3% BSA for 30 minutes at RT and incubated overnight at 4°C with primary antibodies diluted in 0.1% Triton X-100/3% BSA in PBS. The next day, cells were washed three times with PBS for 5 minutes each and incubated with secondary antibodies diluted 1:1000 in 0.1% Triton X/3% BSA in PBS for 1 hour at RT. After washing with PBS three times for 10 minutes each, coverslips or chamber slides were mounted using Prolong Gold Antifade and incubated overnight at RT. The primary and secondary antibodies used are listed in Table 2. Imaging was performed using an Axio Observer Z1 microscope.

Immunocytochemistry for neuromuscular junction analysis

Culture medium was removed, and cells were washed once with 1X PBS. Cells were fixed with 2% paraformaldehyde (Santa cruz) in PBS for 20 min on ice. Cells were washed once with PBS and subsequently permeabilized and blocked with 0.1% Triton X-100 (Sigma) and 1% goat serum (Cell Signaling Technologies) in PBS for 15 min on RT. ActinRed 555 ReadyProbes reagent (Invitrogen) was used for fluorescent staining of actin. Primary (mouse anti-neurofilament (DSHB)) and secondary antibodies (Alexa Fluor® 647 AffiniPure Goat Anti-Mouse IgG (Jackson), Goat anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibody (Life Technologies), Alexa Fluor™ 488, Cy™3 AffiniPure Goat Anti-Mouse IgG, F(ab')₂ fragment specific (Jackson)) were diluted in 0.1% Triton X-100 and 1% goat serum in PBS (Table 2). Primary antibodies were incubated overnight at 4°C, and were washed off after washing three times with PBS (3x10min). Cells were incubated with the secondary antibodies and 2 µg/ml α-bungarotoxin (Invitrogen) for 2 hours at RT. Secondary antibodies were washed off by 3x10 min washing with PBS. Nuclei were stained with Hoechst (1:10000) (Invitrogen) for 10 min in PBS. Cells were imaged using a Nikon Eclipse Ti confocal microscope.

Genomic DNA extraction and telomere length assay

DNA extraction was carried out using the Wizard SB Genomic DNA Purification System kit with a minor modification. Prior to elution, to the 250 µl nuclease-free water, 2 µl of RNase A solution was added, followed by a 10-minute incubation at RT. For the telomere length assay, a total reaction volume of 20 µl was prepared consisting of either 2 µl of telomere primers or 2 µl of SCR primers from the Relative Human Telomere Length Quantification qPCR kit (ScienCell), 1 µl of

genomic DNA template (5 ng/μl), 10 μl of Power SYBR Green PCR Mastermix, and 7 μl of nuclease-free water. The reaction was run under the following cycling conditions: 2 minutes at 50°C, 10 minutes at 95°C, followed by 40 cycles of 15 seconds at 95°C and 1 minute at 60°C. The relative expression levels of telomeres, normalized to the SCR, were calculated using the Delta-Delta Ct method.

RNA sequencing

1. *Sample preparation and sequencing*

3 μg of total RNA were treated with TURBO DNase (Life Technologies) and purified using RNeasy Minelute RNA cleanup kit (Qiagen). RNA quality was assessed using microcapillary electrophoresis on Agilent 2100 Bioanalyzer with RNA Pico 6000 kit (Agilent) and only RNA with RIN values >8 was processed further. Per RNA-seq library, 1 μg of DNase-treated RNA was treated with RiboZero Gold (Human/Mouse/Rat) kit (Illumina) to remove rRNAs, followed by RNA cleanup using the RNeasy Minelute RNA cleanup kit. Sequencing libraries were prepared from equal quantities of rRNA-depleted RNA using TruSeq Stranded total RNA LT kit (Illumina) according to manufacturer's instructions using 11 cycles of enrichment PCR. The quality of the libraries was assessed using Agilent 2100 Bioanalyzer with the DNA 1000 kit (Agilent). Library concentration was measured using Qubit dsDNA HS Assay Kit (Life Technologies). Multiplexing of libraries was performed according to manufacturer's instructions. Multiplexed libraries were sequenced using a NextSeq 500 (Illumina) to generate 75-nt single-end reads. Sequencing depth was 20–40 Mio reads per library.

2. *Processing and downstream analysis*

The reads were pseudo-aligned to human transcriptome (Ensembl version GRCh38.86) using *kallisto* (version 0.43.0_3).⁸⁸ Resulting estimated transcript-level abundances were aggregated into counts per gene and exported using Tximport (version 1.10.1) pipeline.⁸⁹ Differential gene expression analysis was performed using DESeq2 package (version 1.22.2)⁹⁰ in R by comparing gene expression at each stimulation time point to undifferentiated parental cell line. Normalized counts were extracted from DESeq2 object. PCA plot was constructed from rlog-transformed raw counts.

Analysis of the over-represented pathways was performed using Genomatix Pathway System (GePS) tool from the Genomatix software suite (<http://www.genomatix.de>) by providing significantly up-regulated genes from each time point and stimulation as input. Tissue and cell type associations were analysed using Enrichr tool⁹¹ and JENSEN tissue expression database (<https://tissues.jensenlab.org>). Gene lists associated with selected cell types were extracted and their normalized expression values (from DESeq2 analysis) were used to construct heatmap of the Figure S1C.

Single cell RNA sequencing

1. Preparation of single cell suspensions

All parental lines, including H9 and AxSCs derived from H9, HMGU1, and HUES6, as well as Day2 neural differentiation samples, were dissociated into single cells using TrypLE Express. The harvested cells were then centrifuged at 300 g for 3 minutes, and the resulting pellet was resuspended in a solution of 0.04% BSA in PBS. The cell density was adjusted to 1000 cells/µL. For Day 14 and Day 28 CFS-derived cells, as well as Day 28 CS-derived cells, dissociation was carried out using 0.05% Trypsin. After centrifugation at 300 g for 3 minutes, the pellet was resuspended in a solution of 0.04% BSA in PBS. The cells were then passed through 40 µm strainers, and the cell density was adjusted to 1000 cells/µL. For Day 14 CS-derived cells, we adapted a protocol that uses Accutase and Papain as dissociation agents.⁹³ The cells were first rinsed with PBS and then incubated at 37°C for 25 minutes in a dissociation buffer composed of 50% Accutase (Sigma) and 10% Papain (Sigma) in PBS. Without removing the dissociation buffer, a wash buffer containing DMEM/F12, 33.3 µg/ml Dnase II (Sigma), and 10 µM Y-27632 was added. The resulting cell suspension was filtered through a 70 µm strainer and centrifuged at 300 g for 3 minutes. The pellet was then resuspended in neural differentiation medium and filtered once again using 40 µm strainers.

2. Library preparation and sequencing

Single cell sequencing of parental human AxSC lines:

Live cells were counted using trypan blue stain (Life Technologies) with the Countess 3 automated cell counter (Thermo Fisher Scientific). Samples with viability above 85% were used to prepare cell suspensions containing 1000 cells/µL, which were subsequently used for the single cell library preparation. The leftover cells were pelleted via centrifugation and collected for DNA isolation for genotyping purposes with HumanCytoSNP-12 v2.1 BeadChip Kit (Illumina). The single cell library preparation was done using Chromium Controller instrument, Chromium Next GEM Single Cell 3' GEM, Library & Gel Bead Kit v3.1 (10X Genomics) and the Single Index Kit T Set A (10X Genomics). The final sample loaded for the 10X library prep contained a mixture of 7000 cells/cell line for a total of 21000 cells. Quality control for the library was performed using Agilent High Sensitivity kit (Agilent) following manufacturer instructions. The final pools were sequenced on NovaSeq 6000 platform using SP Reagent Kit v1.5/ 100 cycles (Illumina) with 28x8x91 configuration run leading to 40000 reads/cell.

Single cell sequencing of neural differentiation time course from human AxSC lines:

Counting was performed with Trypan blue for live – dead cell discrimination using Countess automated counter. Samples with viability above 81% were used to prepare cell suspensions containing 500-1000 cells/µL. The single cell library preparation was performed using Chromium Controller instrument, Chromium Next GEM Single Cell 3' GEM, Library & Gel Bead Kit v3.1 (10X

Genomics) and Dual Index Kit TT Set A, 96 rxns (10X Genomics). Quality control for the library was performed using Agilent High Sensitivity kit (Agilent) following manufacturer instructions. The final pool contained 16 samples each with 10000 cells per time point and they were sequenced on NovaSeq 6000 platform using NovaSeq 6000 S4 Reagent Kit v1.5 /200 cycles (Illumina) with 28x10x10x90 configuration run leading to 70000 reads/cell.

3. Processing of the datasets

For human parental human AxSC lines, SNP sequencing was performed using Illumina GSA v.3 for cells of each cell line. Next, demuxlet⁸⁴ was used to confidently assign each barcode to a specific cell line. Briefly, demuxlet harnesses genetic variation to identify the genetic origin of individual cells in a mixed sample. Ambiguous cells that could not be confidently matched to one of the three cell lines were removed.

4. Downstream analysis

Downstream analysis was performed using the Scanpy (version 1.6.1), Pandas (version 1.1.3), Numpy (version 1.19.5), and Louvain (version 0.7.0) packages. The datasets were visualized using UMAP embedding with the `pl.umap()` function. Differential expression analysis for H9, CFS, and CS datasets was conducted using the `tl.rank_genes_groups()` function with default parameters. The list of differentially expressed genes (DEGs) was extracted using the `get.rank_genes_groups()` function, setting 'reference' as "rest" and a log2 fold change threshold of 2. The top 50 DEG genes were visualized using the `pl.dotplot()` function. Genes associated with axial development, neural/mesodermal lineages (Fig. 3a), were summarized based on published scRNA sequencing studies conducted on in vivo cells.^{16,20,21,94} To assess the expression of dorsoventral spinal cord progenitor genes, the study by Rayon et al.³² was used as a benchmark.

For the time course neural differentiation experiment, each dataset was analyzed individually. Cell cycle phases were identified using the `tl.score_genes_cell_cycle()` function with default parameters, and the percentage of cells per phase was calculated using the `value_counts()` function. Clustering for each dataset was performed using the `tl.louvain()` function with a 'resolution' set to 2. Marker genes that map to each domain of spinal cord neurons identified by Rayon et al.³², as well as the accompanying database (<https://shiny.crick.ac.uk/scviewer/neuraltube/>), and marker genes for neural crest development identified by Soldatov et al.²⁵ were utilized for cluster annotation and summarized in Figure 5A. To evaluate specific cell types, single genes known to be specific to only one domain in the database mentioned above were used for analysis. For cell populations that could not be identified by a single gene, the coexpression of several genes was assessed. Clusters were manually merged and annotated based on the expression of marker genes in the datasets. Clusters representing the same cellular population across datasets were shown in the same color.

Chromatome and proteome analysis

1. Nuclei isolation and sample preparation for chromatome analysis

AxSCs and H9 cells were dissociated into single cells using TrypLE Express. The cells were incubated for 4-6 minutes at 37°C. The resulting cell suspensions were centrifuged at 300 g for 3 minutes, and the cell number was counted. For each derivation of CS and H9, 5x10⁶ cells were used, while for each derivation of CFS cells, 3x10⁶ cells were used. The cells were washed twice with PBS by centrifugation at 2300 g for 5 minutes at RT. The supernatant was discarded, and the cell pellet was resuspended in 1 ml of ice-cold lysis buffer (3 mM MgCl₂, 10 mM NaCl, 10 mM Tris pH 7.4, 1% NP-40, and freshly added 1X cComplete protease inhibitor (Sigma)). To ensure thorough dissolution of NP-40, the lysis buffer was pre-incubated on an orbital shaker at mild agitation at 4°C before being used for resuspending the cell pellet. The pellet was homogenized by pipetting up and down and then incubated on ice for 20 minutes. The suspension was centrifuged at 2300 g for 5 minutes at 4°C, and the supernatant was discarded. The pellet was resuspended and incubated in 3.3 ml of PBS with 1% methanol-free formaldehyde for 10 minutes on a rotating wheel at mild agitation at RT. The reaction was quenched by incubating the suspension with 125 mM Glycine for an additional 5 minutes on a rotating wheel. The nuclei suspension was then centrifuged at 2300 g for 5 minutes at 4°C and washed twice with ice-cold PBS. After another centrifugation step at 2300 g for 5 minutes at 4°C, the supernatant was discarded, and the pellets were snap-frozen in liquid nitrogen for 15 seconds and then stored at -80°C. Chromatin was released from crosslinked nuclei by dissolving 300 µl of SDS buffer (50 mM HEPES pH 7.4, 10 mM EDTA pH 8.0, 4% UltraPure™ SDS Solution from Invitrogen, along with a newly added 1X cComplete™ EDTA-free Protease Inhibitor Cocktail) using gentle pipetting. This mixture was left to incubate at RT for 10 minutes before adding 900 µl of freshly prepared Urea buffer (10 mM HEPES pH 7.4, 1 mM EDTA pH 8.0, 8 M urea from Sigma). The solution was then carefully inverted seven times before being centrifuged at RT for 30 min at 20000 g. The supernatant was removed, taking care not to disturb the pellet. Two additional wash steps were performed (one wash with SDS and Urea, and one wash with only SDS). The final pellet was then dissolved in 300 µl of Sonication buffer (10 mM HEPES pH 7.4, 2 mM MgCl₂, with a freshly added 1× cComplete™ EDTA-free Protease Inhibitor Cocktail). The chromatin specimens were sonicated using a Bioruptor® Plus at 4°C for 15 cycles (30 s on, 30 s off). The protein concentration was determined using the Pierce™ BCA Protein Assay Kit.

Subsequently, Protein Aggregation Capture (PAC) was performed. In this step, Sera-Mag™ beads (1 mg from Sigma) were washed three times with 70% acetonitrile for every 100 µg of chromatin solution. After the final wash, 300 µl of the chromatin solution corresponding to 100 µg was added to the beads, followed by 700 µl of 100% acetonitrile. The chromatome-bead mixtures were then vortexed and left to rest on a bench for 10 minutes. The samples were vortexed again and placed into a magnetic rack. The samples were then washed with 700 µl of 100% acetonitrile, followed by 1 ml of 95% acetonitrile, and finally with 1 ml of 70% ethanol. The

remaining ethanol was allowed to evaporate, and the beads were resuspended in 400 μ l of 50 mM HEPES pH 8.5, supplemented with freshly prepared 5 mM TCEP and 5.5 mM CAA. The samples were then left to incubate at RT for half an hour. Protease digestion was initiated by adding LysC (protease to protein ratio of 1:200) and Trypsin (1:100) and allowing the mixture to incubate overnight at 37°C under constant agitation at 1100 rpm. From this point forward, the samples were handled in the same way as the total proteome samples.

2. Sample preparation for total proteome analysis

All proteomic experiments were conducted in triplicate. AxSCs and H9 cells were dissociated into single cells using TrypLE Express. The cells were incubated for 4-6 minutes at 37°C. After incubation, the cell suspensions were centrifuged at 300 g for 3 minutes, and the cell number was determined. For each derivation of CS and H9, 2×10^6 cells were used, while for each derivation of CFS cells, 1.5×10^6 cells were used. These cells were then centrifuged at 300 g for 3 minutes. The supernatant was removed, and the cell pellets were snap-frozen in liquid nitrogen for 15 seconds, followed by storage at -80°C. Frozen cells were dissolved in 200 μ l of lysis buffer (containing 6 M guanidinium chloride, 100 mM Tris-HCl with a pH of 8.5, and 2 mM DTT) and subjected to heating for 10 minutes at 99°C with a constant agitation rate of 1400 rpm. The sonication of samples was then performed at 4°C using a Bioruptor® Plus sonication device (Diagenode) at high-intensity settings, with 30 seconds on/off intervals for 15 rounds. If the sample viscosity was adequately reduced, protein concentrations were determined; if not, sonication was repeated. Protein concentrations were assessed using the Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific) in accordance with the manufacturer's instructions. After incubating for at least 20 minutes with 40 mM chloroacetamide, 30 μ g of each proteome sample was diluted in a 30 μ l lysis buffer supplemented with chloroacetamide and DTT. These samples were further diluted in 270 μ l of digestion buffer (containing 10% acetonitrile, 25 mM Tris-HCl at pH 8.5, 0.6 μ g Trypsin/sample (Pierce™ Trypsin Protease, Thermo Fisher Scientific), and 0.6 μ g/sample LysC (Pierce™ LysC Protease, Thermo Fisher Scientific). Proteins were then digested for 16 hours at 37°C with constant shaking at 1100 rpm.

To halt protease activity, 1% (v/v) trifluoroacetic acid (TFA) was added the following day and samples were loaded onto homemade StageTips composed of three layers of SDB-RPS matrix (Empore)⁹⁵, previously equilibrated with 0.1% (v/v) TFA. After loading, two washes with 0.1% (v/v) TFA were performed, and peptides were eluted with 80% acetonitrile and 2% ammonium hydroxide. After the eluates were evaporated in a SpeedVac centrifuge, the samples were resuspended in 20 μ l 0.1% TFA and 2% acetonitrile. The peptides were completely solubilized by constant shaking for 10 minutes at 2000 rpm, and peptide concentrations were determined on a Nanodrop™ 2000 spectrophotometer (Thermo Fisher Scientific) at 280 nm.

3. Nanoflow LC-MS/MS analysis for proteomes and chromatomes

Peptide separation before MS was accomplished using liquid chromatography on an Easy-nLC 1200 (Thermo Fisher Scientific) with in-house packed 50 cm columns of ReproSilPur C18-AQ 1.9- μ m resin. A binary buffer system was used (buffer A: 0.1% formic acid and buffer B: 0.1% formic acid in 80% acetonitrile), with a gradual increase in buffer B concentration (from 5% initially to 95% at the end) to elute the peptides over a 120-minute period at a steady flow rate of 300 nl/min. The peptides were then introduced into an Orbitrap ExplorisTM 480 mass spectrometer (Thermo Fisher Scientific) via a nanoelectrospray source. Each set of triplicates were followed by a washing step while the column temperature was constantly at 55°C.

Data-Dependent Acquisition (DDA) runs used a top12 shotgun proteomics method within a range of 300–1650 m/z, a default charge state of 2, and a maximum injection time of 25 ms. Full scan resolution was set at 60000 and the normalized AGC target at 300%. For MS2 scans, the orbitrap resolution was set at 15000 and the normalized AGC target at 100%, with a maximum injection time of 28 ms.

Data-Independent Acquisition (DIA) runs used an orbitrap resolution of 120000 for full scans in a scan range of 350–1400 m/z, with a maximum injection time of 45 ms. For MS2 acquisitions, the mass range was set to 361–1033 with isolation windows of 22.4 m/z. A default window overlap of 1 m/z was used. The orbitrap resolution for MS2 scans was set at 30000, the normalized AGC target at 1000%, and the maximum injection time at 54 ms.

5. *MS data quantification*

Raw MS data acquired in DIA mode was analyzed using DIA-NN version 1.8.1.⁹⁶ A hybrid approach utilizing a dedicated DDA and DIA library was employed; the DDA library was generated using SpectroMineTM (Pulsar) and the DIA library using DIA-NN. Cross-run normalization was conducted in an RT-dependent manner. Missed cleavages were set to 1, N-terminal methionine excision was activated, and cysteine carbamidomethylation was set as a fixed modification. Proteins were grouped using the additional command ‘–relaxed-prot-inf’. Match-between runs was enabled, and the precursor FDR was set to 1%.

5. *Statistical analyses*

Raw data outputs were analyzed downstream with Perseus (version 1.6.0.9).⁹⁷ CVs were calculated by filtering out proteins or precursors with fewer than 2 out of 3 valid values. Downstream analyses were conducted after imputation of missing values, which was done based on a Gaussian distribution with respect to the standard deviations of measured values (width of 0.2 and a downshift of 1.8 standard deviations). GO enrichment analyses of differentially enriched proteins were performed against the background of total identified proteins using a Benjamini-Hochberg FDR-corrected Fisher's Exact test of 0.05. These analyses were carried out individually for each cluster. Student's t-tests were calculated with a permutation-based FDR of 0.05 and an s0 value of 1, unless stated otherwise. For the multiple sample test based on an

ANOVA the FDR was set to 0.05 and the s0 value to 2. ANOVA tests of normalized chromatomes were conducted likewise after calculating pairwise differences of ChAC-DIA and total proteome values.

6. Web application development

Significant changes in chromatome, proteome and relative chromatin binding during axial stem cell differentiation were displayed in an interactive heatmap as mean row differences of log2 intensities.

The web application was programmed using R Shiny with the following libraries besides base R packages for data processing and visualization: shiny (1.7.1), shinydashboard (0.7.2), shinyHeatmaply (0.2.0), plotly (4.10.0), heatmaply (1.3.0) and png (0.1–7). From the tidyverse (1.3.1) family we further utilized tidyr (1.2.0), dplyr (1.0.9), and ggplot2 (3.3.6).

Electrophysiological analysis

1. Patch-clamp recordings

The whole-cell patch-clamp technique was performed to record from cells either derived from hiPSCs (CS and CFS) or hESCs (labeled 04.06) recorded at an early (day 14–21) and late time point (day 49–57) after start of differentiation. During recordings at RT, neurons grown on glass coverslip were continuously perfused by artificial cerebrospinal fluid (ACSF) saturated with 95% O₂ / 5% CO₂ at a flow rate of 2 ml/min (Ismatec IPC). ACSF contained 125 mM NaCl, 25 mM NaHCO₃, 2.5 mM KCl, 1.25 mM NaH₂PO₄, 10 mM D-glucose, 2 mM CaCl₂, and 1 mM MgCl₂ (pH, 7.3; osmolality, 310 mOsm/kg).

Whole-cell voltage-clamp recordings were performed with a patch-clamp amplifier (EPC 9, HEKA Elektronik). Individual cells were visualized using infrared differential interference contrast video microscopy on a Zeiss Axioskop microscope and were identified based on a spherical and bright cell body bearing 2 or more neurites. Borosilicate glass electrodes with filaments (1B150F-3, 0.84 x 1.5 x 76mm, Science Products) were pulled with a horizontal pipette puller (Model P-97 Flaming Brown Micropipette Puller, Sutter Instrument). Pipettes had resistances between 4–6 MΩ when filled with solution containing (in mM): 115 mM K-gluconate, 20 mM KCl, 10 mM Phosphocreatine-Na, 4 mM Mg-ATP, 0.3 mM GTP, 0.2 mM EGTA, and 10 mM HEPES (pH, 7.3; 290 mOsm).

In voltage-clamp, input resistance was determined at -70 mV with voltage steps of -3 mV before inward and outward currents were evoked with increasing depolarizing steps (20 mV; 300 ms). In current-clamp, the resting membrane potential (RMP) was determined and APs were evoked with increasing current steps (20 pA; 300 ms). The amplitude of the first evoked AP was analyzed. Spontaneous excitatory postsynaptic currents (sEPSCs) were recorded at -70 mV.

2. Data analysis

Offline analysis was done with FitMaster (HEKA Elektronik). Spontaneous synaptic activity was analyzed using Igor Pro (WaveMetrics, Lake Oswego, OR) and MiniAnalysis (Synaptosoft). GraphPad Prism (version 8.2.1) was used to plot the data as mean +/- SEM and to test for statistical differences among means of groups via one-way analysis of variance (ANOVA).

Teratoma analysis

1. *Injection*

Animal experiments were performed following the experimental guidelines within the animal protocol approved by the government of Upper Bavaria with nr. ROB-55.2-2532.Vet_02-14-41. In brief immunodeficient SCID mice were injected under the kidney capsule with 1×10^6 cells resuspended in 50% Matrigel. The mice were kept under observation for the duration of 6 weeks.

2. *Histopathology*

The left and right kidney were harvested and immediately fixed with 4 % (w/v) neutrally buffered formalin for a maximum of 48 hours. Kidneys were cut in a standardized manner resulting in 4 cross-sections and subsequently routinely embedded in paraffin. Sections of 3 μ m were stained with hematoxylin and eosin (H&E) and scanned with an AxioScan.Z1 digital slide scanner (Zeiss) equipped with a 20x and a 40x magnification objective. Images were obtained using the commercially available software netScope Viewer Pro (Net-Base Software GmbH).

Energy metabolism analysis

The cells (H9, CS and CFS) were dissociated using TrypLE select and seeded in cell culture microplate (Agilent Seahorse XF96) at a density of 50000 cells/well in presence of their full respective media, supplemented by 10 μ M of Y-27632. The next day the cells (90% confluence) were incubated with RPMI medium (Sigma) supplemented with L-Glutamine (2 mM) with D-(+)-Glucose (10 mM, for the measurement of Oxidative phosphorylation, or without Glucose (for glycolysis assay) and transferred to an incubator (37 °C) without CO₂ for one hour before the start of the measurement. For Mito Stress assay, the following treatments were used, respectively: Oligomycin at 1 μ M (Cayman Chemical) which inhibits the ATP synthase, Carbonyl cyanide 4-(trifluoromethoxy) phenylhydrazone (FCCP, 1 μ M, Sigma), an uncoupling agent that disrupts mitochondrial ATP synthesis by interfering with proton gradient across the mitochondrial inner membrane⁹⁸ and a combination of rotenone (Sigma) and antimycin A (Sigma) inhibiting complex I and complex III of the electron transport chain, respectively, each at a concentration of 1 μ M. For glycolysis stress test, the cells Glucose (10 mM) was first injected followed by Oligomycin (1 μ M) then 2-Deoxy-D-glucose at 10 mM (Sigma). At least 30 wells were assayed for each cell type. The measurements values were then normalized to the total protein content in each well measured by the SRB assay. The data were then visualized using Wave software and analyzed using GraphPad Prism 8.0.

Embryo grafting and assessment of contribution

1. *Mouse strains, staging and husbandry*

Wild-type C57BL/6J mice were maintained on a 12 h light/12 h dark cycle. For timed matings, noon on the day of finding a vaginal plug was designated as E0.5. Staging of early mouse embryos was carried out according to Downs and Davies.⁹⁹ All animal experiments were performed under the UK Home Office project license PEEC9E359, approved by the Animal Welfare and Ethical Review Panel of the University of Edinburgh and within the conditions of the Animals (Scientific Procedures) Act 1986.

2. *Embryo grafting*

Approximately 70% confluent mouse CFS cells were grafted.^{15,100} The cells were topically labelled with Dil or DiO in a 100 microlitre droplet of M2 medium (Sigma) by expelling approximately 1 microlitre of labelling solution (Thermo Fisher V22886 and C7001) to the drop. After 30 sec-1 minute labelling, cells were washed 2-4x in 100 microlitre droplets, special care was taken for Dil to ensure all crystals were washed off. For embryos older than 6 somites, where the hindgut covers the primitive streak, the transfer pipette was introduced to the hindgut lumen by expelling a small volume of medium for better visualisation and to avoid damage to the hindgut. Cells were grafted just dorsal to the hindgut end, which approximates to the node-streak border or anterior primitive streak of these embryos.

3. *Immunohistochemistry*

Embryo culture, cryosectioning and immunofluorescence were performed as described previously.¹⁰¹ To prevent loss of lipophilic dyes, images were acquired before immunostaining counterstaining nuclei with Hoechst 33342 (1:2000, Thermo Fisher). Pre- and post-immunostaining images of each cryosection were registered using the Hoechst as a reference channel. Imaging was performed using a Nikon Eclipse TiE widefield Inverted microscope and Plan Apo 20x/0.8 lens. Images were processed using Photoshop (Adobe) and ImageJ software (NIH).

Tables

Table 1: List of probes and primers

Probes	Cat no
CDH15	Hs00979297_m1
CDX2	Hs01078080_m1

GAPDH	Hs02758991_g1		
ISL1	Hs00158126_m1		
MNX1	Hs00907365_m1		
MSGN	Hs03405514_s1		
MYOD	Hs00159528_m1		
MYOG	Hs01072232_m1		
NANOG	Hs02387400_g1		
PAX6	Hs01088114_m1		
POU4F1	Hs00366711_m1		
POU5F1 (OCT4)	Hs00999632_g1		
SOX2	Hs01053049_s1		
TBX6	Hs00365539_m1		
TBXT	Hs00610080_m1		
Primers	Sequence	Primers	Sequence
Human ASCL1	F: TTCACCAACTGGTTCTGAG R: TAAAGATGCAGGTTGTGCG	Mouse Gapdh	F: CATCACTGCCACCCAGAAGACTG R: ATGCCAGTGAGCTTCCGTTAG
Human CDX2	F: TCGGCAGCCAAGTGAA R: GATGGTGATGTAGCGACTGTAG	Mouse Hes5	F: AAACACAGCAAAGCCTTCG R: CAGGGTCAGGAACTGTACC
Human DCX	F: GCCAGGGAGAACAAAGGACTTT R: CACCCCACTGCGGATGA	Mouse Mnx1	F: CAGCACCTTCAACTGGACCAG R: TTCGGCACTCCCCAAGAGGTT
Human GAPDH	F: CGCTCTCTGCTCCTCCTGTT R: CCATGGTGTCTGAGCGATGT	Mouse Nanog	F: GAAATCCCTCCCTGCCAT R: CAGGCATTGATGAGGCGTTC
Human HES5	F: TCAGCTACCTGAAGCACAG R: AGTAGCCTCGCTGTAGTC	Mouse Nkx1-2	F: CTCAGCACGGAACTCCAAA R: AGCCCACCAACAACACTGCAT

Human NANOG	F: CCTTCCTCCATGGATCTGCTT R: CTTGACCGGGACCTTGTCTTC	Mouse Oct4	F: GGCTTCAGACTTCGCCTC R: AGCTTAGCCAGGTTGAGGAT
Human POU5F1	F: CAATTTGCCAAGCTCCTGAAG R: AAAGCGGCAGATGGTCGTT	Mouse Otx2	F: CTGGGCTAACATTCCAGTT R: GTCCATTCAAGGTTGCTGGT
Human SOX1	F: GAGAACCCAAAGATGCACAA R: CCTCGGACATGACCTTCCA	Mouse Pax6	F: CTGAGGAACCAGAGAACAGG R: CATGGAACCTGATGTGAAGGAGG
Human SOX2	F: CCTCCGGGACATGATCAGCATGT R: GCAGTGTGCCGTTAACGGCGTG	Mouse Sox1	F: CGGATCTCTGGTCAAGTCGG R: GGGACCTCGGTACAAAGTCG
Human TBX6	F: CGTGTGAAGAGGAAACTGCG R: GACTACACTCACCTCCGCTC	Mouse Sox2	F: CAAAAACCGTGATGCCGACT R: CGCCCTCAGGTTTCTCTGT
Human TBX7	F: ATGAGCCTCGAACATCCACATAG R: CTGTGATCTCCTCGTTCTGATAA	Mouse Tbx6	F: TCAGCCTGAGCTTGGAGAACCA R: CCAGTGACTGATACTCGGCAAG
Mouse Brachyury	F: GAAGGGAGACCCCACCGAA R: TTACCTTCAGCACCGGGAAC	Orangutan PAX6	F: CAGATCTGCCACTTCCCTG R: TCACTCCGCTGTGACTGTT
Mouse Cdx2	F: CTGGACAAGGACGTGAGCAT R: ACTCGGAGGACTGACAAAG		

Table 2: List of antibodies

Primary antibodies	Cat no	Manufacturer	Dilution
Rabbit anti-Sox2	2738	Cell Signaling Technology	1:400
Rabbit anti-Sox2	ab92494	Abcam	1:200
Mouse anti-Brachyury	14-9770-82	Thermo	1:100
Goat anti-Brachyury	AF2085	R&D Systems	1:100
Rabbit anti-CDX2	12306	Cell Signaling Technology	1:200
Rabbit anti-Pax6	901301	Biolegend	1:100
Rabbit anti-Zic2	ab150404	Abcam	1:100
Rabbit anti-Hes5	NBP2-56999	Novus	1:100

Primary antibodies	Cat no	Manufacturer	Dilution
Rabbit anti-Islet 1	ab20670	Abcam	1:200
Mouse anti-Brn3a	sc-8429	Santa cruz	1:100
Mouse anti-Mnx1	sc-515769	Santa cruz	1:100
Chicken anti-B3 tubulin	NB100-1612	Novus	1:100
Mouse anti-Myod	MA1-41017	Life Technologies	1:100
Rabbit anti-M-Cadherin	40491S	Cell Signaling Technology	1:100
Mouse anti-Myogenin	14-5643-82	Life Technologies	1:100
Mouse anti-Sox2	MAB2018	R&D Systems	1:100
Mouse anti-Pax3	AB_528426	DSHB	1:20
Goat anti-Meox1	sc-10185	Santa cruz	1:50
Mouse anti-neurofilament	AB_531793	DSHB	2 µg/ml
Rabbit IgG isotype control	GTX35035	GeneTex	1 µg/ml
Mouse IgG isotype control	GTX35009	GeneTex	1 µg/ml
Secondary antibodies	Cat no	Manufacturer	Dilution
Goat anti-Chicken IgY (H L) Secondary Antibody, Alexa Fluor 594	A11042	Life Technologies	1:1000
Goat anti-rabbit IgG (H+L), Alexa Fluor 488, highly cross-absorbed	A11034	Life Technologies	1:1000
Goat anti-mouse IgG (H+L) Alexa Fluor Plus 647, highly cross-absorbed	A32728	Life Technologies	1:1000
Donkey anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 488	A21202	Life Technologies	1:1000

Donkey anti-Goat IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 594	A11058	Life Technologies	1:1000
F(ab')2-Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 647	A21246	Life Technologies	1:1000
Alexa Fluor® 647 AffiniPure Goat Anti-Mouse IgG (H+L)	115-605-146	Jackson	1:500
Goat anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 488	A11001	Life Technologies	1:500
Cy™3 AffiniPure Goat Anti-Mouse IgG, F(ab') ₂ fragment specific	115-165-006	Jackson	1:500

REFERENCES

1. Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A* **78**, 7634–7638. [10.1073/pnas.78.12.7634](https://doi.org/10.1073/pnas.78.12.7634).
2. Takahashi, K., and Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126**, 663–676. [10.1016/j.cell.2006.07.024](https://doi.org/10.1016/j.cell.2006.07.024).
3. Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell* **131**, 861–872. [10.1016/j.cell.2007.11.019](https://doi.org/10.1016/j.cell.2007.11.019).
4. Thomson, J.A. (1998). Embryonic stem cell lines derived from human blastocysts. *Science* (1979) **282**, 1145–1147. [10.1126/science.282.5391.1145](https://doi.org/10.1126/science.282.5391.1145).
5. Neagu, A., van Genderen, E., Escudero, I., Verwegen, L., Kurek, D., Lehmann, J., Stel, J., Dirks, R.A.M., van Mierlo, G., Maas, A., et al. (2020). In vitro capture and characterization of embryonic rosette-stage pluripotency between naive and primed states. *Nat Cell Biol* **22**, 534–545. [10.1038/s41556-020-0508-x](https://doi.org/10.1038/s41556-020-0508-x).
6. Kinoshita, M., Barber, M., Mansfield, W., Cui, Y., Spindlow, D., Stirparo, G.G., Dietmann, S., Nichols, J., and Smith, A. (2021). Erratum: Capture of Mouse and Human Stem Cells with Features of Formative Pluripotency (Cell Stem Cell (2021) 28(3) (453–471.e8), (S1934590920305439), (10.1016/j.stem.2020.11.005)). *Cell Stem Cell* **28**, 2180. [10.1016/j.stem.2021.11.002](https://doi.org/10.1016/j.stem.2021.11.002).
7. Wymeersch, F.J., Wilson, V., and Tsakiridis, A. (2021). Understanding axial progenitor biology in vivo and in vitro. *Development (Cambridge)* **148**, dev180612. [10.1242/dev.180612](https://doi.org/10.1242/dev.180612).
8. Cambray, N., and Wilson, V. (2007). Two distinct sources for a population of maturing axial progenitors. *Development* **134**, 2829–2840. [10.1242/dev.02877](https://doi.org/10.1242/dev.02877).
9. Garriock, R.J., Chalamalasetty, R.B., Kennedy, M.W., Canizales, L.C., Lewandoski, M., and Yamaguchi, T.P. (2015). Lineage tracing of neuromesodermal progenitors reveals novel wnt-dependent roles in trunk progenitor cell maintenance and differentiation. *Development (Cambridge)* **142**, 1628–1638. [10.1242/dev.111922](https://doi.org/10.1242/dev.111922).
10. Wymeersch, F.J., Huang, Y., Blin, G., Cambray, N., Wilkie, R., Wong, F.C.K., and Wilson, V. (2016). Position-dependent plasticity of distinct progenitor types in the primitive streak. *eLife* **5**, e10042. [10.7554/eLife.10042](https://doi.org/10.7554/eLife.10042).
11. Shaker, M.R., Lee, J.H., Kim, K.H., Ban, S., Kim, V.J., Kim, J.Y., Lee, J.Y., and Sun, W. (2021). Spatiotemporal contribution of neuromesodermal progenitor-derived neural cells in the elongation of developing mouse spinal cord. *Life Sci* **282**, 119393. [10.1016/j.lfs.2021.119393](https://doi.org/10.1016/j.lfs.2021.119393).
12. Henrique, D., Abranches, E., Verrier, L., and Storey, K.G. (2015). Neuromesodermal progenitors and the making of the spinal cord. *Development (Cambridge)* **142**, 2864–2875. [10.1242/dev.119768](https://doi.org/10.1242/dev.119768).
13. Wymeersch, F.J., Skylaki, S., Huang, Y., Watson, J.A., Economou, C., Marek-Johnston, C., Tomlinson, S.R., and Wilson, V. (2019). Transcriptionally dynamic progenitor populations organised around a stable niche drive axial patterning. *Development (Cambridge)* **146**, dev168161. [10.1242/dev.168161](https://doi.org/10.1242/dev.168161).

14. Cambray, N., and Wilson, V. (2002). Axial progenitors with extensive potency are localised to the mouse chordoneural hinge. *Development* *129*, 4855–4866. 10.1242/dev.129.20.4855.
15. Gouti, M., Tsakiridis, A., Wymeersch, F.J., Huang, Y., Kleinjung, J., Wilson, V., and Briscoe, J. (2014). In vitro generation of neuromesodermal progenitors reveals distinct roles for wnt signalling in the specification of spinal cord and paraxial mesoderm identity. *PLoS Biol* *12*, e1001937. 10.1371/journal.pbio.1001937.
16. Gouti, M., Delile, J., Stamatakis, D., Wymeersch, F.J., Huang, Y., Kleinjung, J., Wilson, V., and Briscoe, J. (2017). A Gene Regulatory Network Balances Neural and Mesoderm Specification during Vertebrate Trunk Development. *Dev Cell* *41*, 243–261. 10.1016/j.devcel.2017.04.002.
17. Frith, T.J.R., Granata, I., Wind, M., Stout, E., Thompson, O., Neumann, K., Stavish, D., Heath, P.R., Ortmann, D., Hackland, J.O.S., et al. (2018). Human axial progenitors generate trunk neural crest cells in vitro. *eLife* *7*, e35786. 10.7554/eLife.35786.
18. Edri, S., Hayward, P., Baillie-Johnson, P., Steventon, B.J., and Arias, A.M. (2019). An epiblast stem cell-derived multipotent progenitor population for axial extension. *Development (Cambridge)* *146*, dev168187. 10.1242/dev.168187.
19. Wind, M., Gogolou, A., Manipur, I., Granata, I., Butler, L., Andrews, P.W., Barbaric, I., Ning, K., Guerracino, M.R., Placzek, M., et al. (2021). Defining the signalling determinants of a posterior ventral spinal cord identity in human neuromesodermal progenitor derivatives. *Development (Cambridge)* *148*, dev194415. 10.1242/dev.194415.
20. Pijuan-Sala, B., Griffiths, J.A., Guibentif, C., Hiscock, T.W., Jawaid, W., Calero-Nieto, F.J., Mulas, C., Ibarra-Soria, X., Tyser, R.C.V., Ho, D.L.L., et al. (2019). A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* *566*, 490–495. 10.1038/s41586-019-0933-9.
21. Guillot, C., Djeffal, Y., Michaut, A., Rabe, B., and Pourquié, O. (2021). Dynamics of primitive streak regression controls the fate of neuromesodermal progenitors in the chicken embryo. *eLife* *10*, e64819. 10.7554/eLife.64819.
22. Koch, F., Scholze, M., Wittler, L., Schifferl, D., Sudheer, S., Grote, P., Timmermann, B., Macura, K., and Herrmann, B.G. (2017). Antagonistic Activities of Sox2 and Brachyury Control the Fate Choice of Neuro-Mesodermal Progenitors. *Dev Cell* *42*, e64819. 10.1016/j.devcel.2017.07.021.
23. Tzouanacou, E., Wegener, A., Wymeersch, F.J., Wilson, V., and Nicolas, J.F. (2009). Redefining the Progression of Lineage Segregations during Mammalian Embryogenesis by Clonal Analysis. *Dev Cell* *17*, 365–376. 10.1016/j.devcel.2009.08.002.
24. Gribaudo, S., Robert, R., van Sambeek, B., Mirdass, C., Lyubimova, A., Bouhali, K., Ferent, J., Morin, X., van Oudenaarden, A., and Nedelec, S. (2023). Self-organizing models of human trunk organogenesis recapitulate spinal cord and spine co-morphogenesis. *Nat Biotechnol*. 10.1038/s41587-023-01956-9.
25. Soldatov, R., Kaucka, M., Kastriti, M.E., Petersen, J., Chontorotzea, T., Englmaier, L., Akkuratova, N., Yang, Y., Häring, M., Dyachuk, V., et al. (2019). Spatiotemporal structure of cell fate decisions in murine neural crest. *Science (1979)* *364*, eaas9536. 10.1126/science.aas9536.

26. Serbedzija, G.N., Bronner-Fraser, M., and Fraser, S.E. (1994). Developmental potential of trunk neural crest cells in the mouse. *Development* **120**, 1709–1718. 10.1242/dev.120.7.1709.
27. Park, J.S., Hsiung, H.A., Khven, I., Manno, G. La, and Lutolf, M.P. (2022). Self-organizing in vitro mouse neural tube organoids mimic embryonic development. *Development (Cambridge)* **149**, dev201052. 10.1242/dev.201052.
28. Ericson, J., Rashbass, P., Schedl, A., Brenner-Morton, S., Kawakami, A., Van Heyningen, V., Jessell, T.M., and Briscoe, J. (1997). Pax6 controls progenitor cell identity and neuronal fate in response to graded Shh signaling. *Cell* **90**, 169–180. 10.1016/S0092-8674(00)80323-2.
29. Limone, F., Guerra San Juan, I., Mitchell, J.M., Smith, J.L.M., Raghunathan, K., Meyer, D., Ghosh, S.D., Couto, A., Klim, J.R., Joseph, B.J., et al. (2023). Efficient generation of lower induced motor neurons by coupling Ngn2 expression with developmental cues. *Cell Rep* **42**, 111896. 10.1016/j.celrep.2022.111896.
30. Lee, J.H., Shin, H., Shaker, M.R., Kim, H.J., Park, S.H., Kim, J.H., Lee, N., Kang, M., Cho, S., Kwak, T.H., et al. (2022). Production of human spinal-cord organoids recapitulating neural-tube morphogenesis. *Nat Biomed Eng* **6**, 435–448. 10.1038/s41551-022-00868-4.
31. Iyer, N.R., Shin, J., Cuskey, S., Tian, Y., Nicol, N.R., Doersch, T.E., Seipel, F., McCalla, S.G., Roy, S., and Ashton, R.S. (2022). Modular derivation of diverse, regionally discrete human posterior CNS neurons enables discovery of transcriptomic patterns. *Sci Adv* **8**, eabn7430. 10.1126/sciadv.abn7430.
32. Rayon, T., Maizels, R.J., Barrington, C., and Briscoe, J. (2021). Single-cell transcriptome profiling of the human developing spinal cord reveals a conserved genetic programme with human-specific features. *Development (Cambridge)* **148**, dev199711. 10.1242/DEV.199711.
33. Anand, G.M., Megale, H.C., Murphy, S.H., Weis, T., Lin, Z., He, Y., Wang, X., Liu, J., and Ramanathan, S. (2023). Controlling organoid symmetry breaking uncovers an excitable system underlying human axial elongation. *Cell* **186**, 497–512. 10.1016/j.cell.2022.12.043.
34. Liu, Z., Guo, J., Wang, Y., Weng, Z., Huang, B., Yu, M.K., Zhang, X., Yuan, P., Zhao, H., Chan, W.Y., et al. (2017). CFTR-β-catenin interaction regulates mouse embryonic stem cell differentiation and embryonic development. *Cell Death Differ* **24**, 98–110. 10.1038/cdd.2016.118.
35. López-Anguita, N., Gassaloglu, S.I., Stötzel, M., Bolondi, A., Conkar, D., Typou, M., Buschow, R., Veenvliet, J. V, and Bulut-Karslioglu, A. (2022). Hypoxia induces an early primitive streak signature, enhancing spontaneous elongation and lineage representation in gastruloids. *Development (Cambridge)* **149**, dev200679. 10.1242/dev.200679.
36. Tang, C., Lee, A.S., Volkmer, J.P., Sahoo, D., Nag, D., Mosley, A.R., Inlay, M.A., Ardehali, R., Chavez, S.L., Pera, R.R., et al. (2011). An antibody against SSEA-5 glycan on human pluripotent stem cells enables removal of teratoma-forming cells. *Nat Biotechnol* **29**, 829–834. 10.1038/nbt.1947.
37. Isern, J., García-García, A., Martín, A.M., Arranz, L., Martín-Pérez, D., Torroja, C., Sánchez-Cabo, F., and Méndez-Ferrer, S. (2014). The neural crest is a source of mesenchymal stem

cells with specialized hematopoietic stem cell niche function. *Elife* 3, e03696. 10.7554/eLife.03696.

38. Yilmaz, A., and Benvenisty, N. (2019). Defining Human Pluripotency. *Cell Stem Cell* 25, 9–22. 10.1016/j.stem.2019.06.010.

39. Kunze, C., Börner, K., Kienle, E., Orschmann, T., Rusha, E., Schneider, M., Radivojkov-Blagojevic, M., Drukker, M., Desbordes, S., Grimm, D., et al. (2018). Synthetic AAV/CRISPR vectors for blocking HIV-1 expression in persistently infected astrocytes. *Glia* 66, 413–427. 10.1002/glia.23254.

40. Rao, J., Pfeiffer, M.J., Frank, S., Adachi, K., Piccini, I., Quaranta, R., Araúzo-Bravo, M., Schwarz, J., Schade, D., Leidel, S., et al. (2016). Erratum: Stepwise Clearance of Repressive Roadblocks Drives Cardiac Induction in Human ESCs (Cell Stem Cell (2016) 18 (341-353)). *Cell Stem Cell* 18, 554–556. 10.1016/j.stem.2016.03.008.

41. Tyser, R.C. V., Mammadov, E., Nakanoh, S., Vallier, L., Scialdone, A., and Srinivas, S. (2021). Single-cell transcriptomic characterization of a gastrulating human embryo. *Nature* 600, 285–289. 10.1038/s41586-021-04158-y.

42. Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., and McKay, R.D.G. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* 448, 196–199. 10.1038/nature05972.

43. Kurek, D., Neagu, A., Tastemel, M., Tüysüz, N., Lehmann, J., Werken, H.J.G. Van De, Philipsen, S., Linden, R. Van Der, Maas, A., Ijcken, W.F.J. Van, et al. (2015). Endogenous WNT signals mediate BMP-induced and spontaneous differentiation of epiblast stem cells and human embryonic stem cells. *Stem Cell Reports* 4, 114–128. 10.1016/j.stemcr.2014.11.007.

44. Geuder, J., Wange, L.E., Janjic, A., Radmer, J., Janssen, P., Bagnoli, J.W., Müller, S., Kaul, A., Ohnuki, M., and Enard, W. (2021). A non-invasive method to generate induced pluripotent stem cells from primate urine. *Sci Rep* 11, 3516. 10.1038/s41598-021-82883-0.

45. Amadei, G., Handford, C.E., Qiu, C., Jonghe, J. De, Greenfeld, H., Tran, M., Martin, B.K., Chen, D.Y., Aguilera-Castrejon, A., Hanna, J.H., et al. (2022). Embryo model completes gastrulation to neurulation and organogenesis. *Nature* 610, 143–153. 10.1038/s41586-022-05246-3.

46. Aulehla, A., and Pourquié, O. (2010). Signaling gradients during paraxial mesoderm development. *Cold Spring Harb Perspect Biol* 2, a000869. 10.1101/cshperspect.a000869.

47. Diaz-Cuadros, M., Wagner, D.E., Budjan, C., Hubaud, A., Tarazona, O.A., Donelly, S., Michaut, A., Tanoury, Z. Al, Yoshioka-Kobayashi, K., Niino, Y., et al. (2020). In vitro characterization of the human segmentation clock. *Nature* 580, 113–118. 10.1038/s41586-019-1885-9.

48. Yamanaka, Y., Hamidi, S., Yoshioka-Kobayashi, K., Munira, S., Sunadome, K., Zhang, Y., Kurokawa, Y., Ericsson, R., Mieda, A., Thompson, J.L., et al. (2023). Reconstituting human somitogenesis in vitro. *Nature* 614, 509–520. 10.1038/s41586-022-05649-2.

49. Rhodes, C.T., Thompson, J.J., Mitra, A., Asokumar, D., Lee, D.R., Lee, D.J., Zhang, Y., Jason, E., Dale, R.K., Rocha, P.P., et al. (2022). An epigenome atlas of neural progenitors within the embryonic mouse forebrain. *Nat Commun* 13, 4196. 10.1038/s41467-022-31793-4.

50. Lodato, M.A., Ng, C.W., Wamstad, J.A., Cheng, A.W., Thai, K.K., Fraenkel, E., Jaenisch, R., and Boyer, L.A. (2013). SOX2 Co-Occupies Distal Enhancer Elements with Distinct POU Factors in ESCs and NPCs to Specify Cell State. *PLoS Genet* **9**, e1003288. 10.1371/journal.pgen.1003288.
51. Reinhardt, P., Glatza, M., Hemmer, K., Tsytsyura, Y., Thiel, C.S., Höing, S., Moritz, S., Parga, J.A., Wagner, L., Bruder, J.M., et al. (2013). Derivation and Expansion Using Only Small Molecules of Human Neural Progenitors for Neurodegenerative Disease Modeling. *PLoS One* **8**, e59252. 10.1371/journal.pone.0059252.
52. Kennedy, M.W., Chalamalasetty, R.B., Thomas, S., Garriock, R.J., Jailwala, P., Yamaguchi, T.P., and Nusse, R. (2016). Sp5 and Sp8 recruit β -catenin and Tcf1-Lef1 to select enhancers to activate Wnt target gene transcription. *Proc Natl Acad Sci U S A* **113**, 3545–3550. 10.1073/pnas.1519994113.
53. Patel, N.S., Rhinn, M., Semprich, C.I., Halley, P.A., Dollé, P., Bickmore, W.A., and Storey, K.G. (2013). FGF Signalling Regulates Chromatin Organisation during Neural Differentiation via Mechanisms that Can Be Uncoupled from Transcription. *PLoS Genet* **9**, e1003614. 10.1371/journal.pgen.1003614.
54. Arnold, S.J., Hofmann, U.K., Bikoff, E.K., and Robertson, E.J. (2008). Pivotal roles for eomesodermin during axis formation, epithelium-to-mesenchyme transition and endoderm specification in the mouse. *Development* **135**, 501–511. 10.1242/dev.014357.
55. Qu, Q., Li, D., Louis, K.R., Li, X., Yang, H., Sun, Q., Crandall, S.R., Tsang, S., Zhou, J., Cox, C.L., et al. (2014). High-efficiency motor neuron differentiation from human pluripotent stem cells and the function of Islet-1. *Nat Commun* **5**, 3449. 10.1038/ncomms4449.
56. Mavrommatis, L., Jeong, H.-W., Kindler, U., Gomez-Giro, G., Kienitz, M.-C., Stehling, M., Psathaki, O.E., Zeuschner, D., Bixel, M.G., Han, D., et al. (2023). Human skeletal muscle organoids model fetal myogenesis and sustain uncommitted PAX7 myogenic progenitors. *RP87081*. 10.7554/elife.87081.1.
57. Chal, J., Al Tanoury, Z., Hestin, M., Gobert, B., Aivio, S., Hick, A., Cherrier, T., Nesmith, A.P., Parker, K.K., and Pourquié, O. (2016). Generation of human muscle fibers and satellite-like cells from human pluripotent stem cells in vitro. *Nat Protoc* **11**, 1833–1850. 10.1038/nprot.2016.110.
58. McGrew, M.J., Sherman, A., Lillico, S.G., Ellard, F.M., Radcliffe, P.A., Gilhooley, H.J., Mitrophanous, K.A., Cambray, N., Wilson, V., and Sang, H. (2008). Localised axial progenitor cell populations in the avian tail bud are not committed to a posterior Hox identity. *Development* **135**, 2289–2299. 10.1242/dev.022020.
59. Guibentif, C., Griffiths, J.A., Imaz-Rosshandler, I., Ghazanfar, S., Nichols, J., Wilson, V., Göttgens, B., and Marioni, J.C. (2021). Diverse Routes toward Early Somites in the Mouse Embryo. *Dev Cell* **56**, 141–153. 10.1016/j.devcel.2020.11.013.
60. Sambasivan, R., and Steventon, B. (2021). Neuromesodermal Progenitors: A Basis for Robust Axial Patterning in Development and Evolution. *Front Cell Dev Biol* **8**, 607516. 10.3389/fcell.2020.607516.
61. Ohinata, Y., and Tsukiyama, T. (2014). Establishment of trophoblast stem cells under defined culture conditions in mice. *PLoS One* **9**, e107308. 10.1371/journal.pone.0107308.

62. Galiakberova, A.A., and Dashinimaev, E.B. (2020). Neural Stem Cells and Methods for Their Generation From Induced Pluripotent Stem Cells in vitro. *Front Cell Dev Biol* **8**, 815. 10.3389/fcell.2020.00815.
63. Broccoli, V., Giannelli, S.G., and Mazzara, P.G. (2014). Modeling physiological and pathological human neurogenesis in the dish. *Front Neurosci* **8**, 183. 10.3389/fnins.2014.00183.
64. Turner, D.A., Hayward, P.C., Baillie-Johnson, P., Rué, P., Broome, R., Faunes, F., and Arias, A.M. (2014). Wnt/β-catenin and FGF signalling direct the specification and maintenance of a neuromesodermal axial progenitor in ensembles of mouse embryonic stem cells. *Development (Cambridge)* **141**, 4243–4253. 10.1242/dev.112979.
65. Tahara, N., Kawakami, H., Chen, K.Q., Anderson, A., Peterson, M.Y., Gong, W., Shah, P., Hayashi, S., Nishinakamura, R., Nakagawa, Y., et al. (2019). Sall4 regulates neuromesodermal progenitors and their descendants during body elongation in mouse embryos. *Development (Cambridge)* **146**, dev177659. 10.1242/DEV.177659.
66. Moussaieff, A., Rouleau, M., Kitsberg, D., Cohen, M., Levy, G., Barasch, D., Nemirovski, A., Shen-Orr, S., Laevsky, I., Amit, M., et al. (2015). Glycolysis-mediated changes in acetyl-CoA and histone acetylation control the early differentiation of embryonic stem cells. *Cell Metab* **21**, 392–402. 10.1016/j.cmet.2015.02.002.
67. Morgan, R.A., Gray, D., Lomova, A., and Kohn, D.B. (2017). Hematopoietic Stem Cell Gene Therapy: Progress and Lessons Learned. *Cell Stem Cell* **21**, 574–590. 10.1016/j.stem.2017.10.010.
68. De Rosa, L., Carulli, S., Cocchiarella, F., Quaglino, D., Enzo, E., Franchini, E., Giannetti, A., De Santis, G., Recchia, A., Pellegrini, G., et al. (2014). Long-term stability and safety of transgenic cultured epidermal stem cells in gene therapy of junctional epidermolysis bullosa. *Stem Cell Reports* **2**, 1–8. 10.1016/j.stemcr.2013.11.001.
69. Chavali, V.R.M., Haider, N., Rathi, S., Vrathasha, V., Alapati, T., He, J., Gill, K., Nikonov, R., Duong, T.T., McDougald, D.S., et al. (2020). Dual SMAD inhibition and Wnt inhibition enable efficient and reproducible differentiations of induced pluripotent stem cells into retinal ganglion cells. *Sci Rep* **10**, 11828. 10.1038/s41598-020-68811-8.
70. Vierbuchen, T., Ostermeier, A., Pang, Z.P., Kokubu, Y., Südhof, T.C., and Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**, 1035–1041. 10.1038/nature08797.
71. Bar-Nur, O., Gerli, M.F.M., Di Stefano, B., Almada, A.E., Galvin, A., Coffey, A., Huebner, A.J., Feige, P., Verheul, C., Cheung, P., et al. (2018). Direct Reprogramming of Mouse Fibroblasts into Functional Skeletal Muscle Progenitors. *Stem Cell Reports* **10**, 1505–1521. 10.1016/j.stemcr.2018.04.009.
72. Huang, P., Zhang, L., Gao, Y., He, Z., Yao, D., Wu, Z., Cen, J., Chen, X., Liu, C., Hu, Y., et al. (2014). Direct reprogramming of human fibroblasts to functional and expandable hepatocytes. *Cell Stem Cell* **14**, 370–384. 10.1016/j.stem.2014.01.003.
73. Wainger, B.J., Buttermore, E.D., Oliveira, J.T., Mellin, C., Lee, S., Saber, W.A., Wang, A.J., Ichida, J.K., Chiu, I.M., Barrett, L., et al. (2015). Modeling pain in vitro using nociceptor neurons reprogrammed from fibroblasts. *Nat Neurosci* **18**, 17–24. 10.1038/nn.3886.

74. Lolas, M., Valenzuela, P.D.T., Tjian, R., and Liu, Z. (2014). Charting Brachyury-mediated developmental pathways during early mouse embryogenesis. *Proc Natl Acad Sci U S A* **111**, 4478–4483. 10.1073/pnas.1402612111.
75. Felfly, H., and Klein, O.D. (2013). Sprouty genes regulate proliferation and survival of human embryonic stem cells. *Sci Rep* **3**, 2277. 10.1038/srep02277.
76. Lee, J.-Y., Park, S., Kim, K.-S., Ko, J.-J., Lee, S., Kim, K.P., and Park, K.-S. (2016). Novel Function of Sprouty4 as a Regulator of Stemness and Differentiation of Embryonic Stem Cells. *Dev Reprod* **20**, 171–177. 10.12717/dr.2016.20.2.171.
77. Chambers, D., and Mason, I. (2000). Expression of sprouty2 during early development of the chick embryo is coincident with known sites of FGF signalling. *Mech Dev* **91**, 361–364. 10.1016/S0925-4773(99)00288-9.
78. Wong, Y.H., Lu, A.C., Wang, Y.C., Cheng, H.C., Chang, C., Chen, P.H., Yu, J.Y., and Fann, M.J. (2010). Protophenin defines a transition stage during embryonic neurogenesis and prevents precocious neuronal differentiation. *Journal of Neuroscience* **30**, 4428–4439. 10.1523/JNEUROSCI.0473-10.2010.
79. Hemmi, K., Ma, D., Miura, Y., Kawaguchi, M., Sasahara, M., Hashimoto-Tamaoki, T., Tamaoki, T., Sakata, N., and Tsuchiya, K. (2006). A homeodomain-zinc finger protein, ZFHX4, is expressed in neuronal differentiation manner and suppressed in muscle differentiation manner. *Biol Pharm Bull* **29**, 1830–1835. 10.1248/bpb.29.1830.
80. Berry, F.B., Miura, Y., Mihara, K., Kaspar, P., Sakata, N., Hashimoto-Tamaoki, T., and Tamaoki, T. (2001). Positive and Negative Regulation of Myogenic Differentiation of C2C12 Cells by Isoforms of the Multiple Homeodomain Zinc Finger Transcription Factor ATBF1. *Journal of Biological Chemistry* **276**, 25057–25065. 10.1074/jbc.M010378200.
81. Jung, C.G., Kim, H.J., Kawaguchi, M., Khanna, K.K., Hida, H., Asai, K., Nishino, H., and Miura, Y. (2005). Homeotic factor ATBF1 induces the cell cycle arrest associated with neuronal differentiation. *Development* **132**, 5137–5145. 10.1242/dev.02098.
82. Russ, D.E., Cross, R.B.P., Li, L., Koch, S.C., Matson, K.J.E., Yadav, A., Alkasla, M.R., Lee, D.I., Le Pichon, C.E., Menon, V., et al. (2021). A harmonized atlas of mouse spinal cord cell types and their spatial organization. *Nat Commun* **12**, 5722. 10.1038/s41467-021-25125-1.
83. Inoue, T., Ota, M., Mikoshiba, K., and Aruga, J. (2007). Zic2 and Zic3 synergistically control neurulation and segmentation of paraxial mesoderm in mouse embryo. *Dev Biol* **306**, 669–684. 10.1016/j.ydbio.2007.04.003.
84. Kang, H.M., Subramaniam, M., Targ, S., Nguyen, M., Maliskova, L., McCarthy, E., Wan, E., Wong, S., Byrnes, L., Lanata, C.M., et al. (2018). Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat Biotechnol* **36**, 89–94. 10.1038/nbt.4042.
85. Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol* **19**, 15. 10.1186/s13059-017-1382-0.
86. Bernstein, N.J., Fong, N.L., Lam, I., Roy, M.A., Hendrickson, D.G., and Kelley, D.R. (2020). Solo: Doublet Identification in Single-Cell RNA-Seq via Semi-Supervised Deep Learning. *Cell Syst* **11**, 95–101. 10.1016/j.cels.2020.05.010.
87. Lakkis, J., Wang, D., Zhang, Y., Hu, G., Wang, K., Pan, H., Ungar, L., Reilly, M.P., Li, X., and Li, M. (2021). A joint deep learning model enables simultaneous batch effect correction,

denoising, and clustering in single-cell transcriptomics. *Genome Res* 31, 1753–1766. 10.1101/gr.271874.120.

88. Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* 34, 525–527. 10.1038/nbt.3519.

89. Soneson, C., Love, M.I., and Robinson, M.D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* 4, 1521. 10.12688/f1000research.7563.1.

90. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550. 10.1186/s13059-014-0550-8.

91. Kuleshov, M. V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* 44, W90–W97. 10.1093/nar/gkw377.

92. Santos, A., Tsafou, K., Stolte, C., Pletscher-Frankild, S., O'Donoghue, S.I., and Jensen, L.J. (2015). Comprehensive comparison of large-scale tissue expression datasets. *PeerJ* 2015, e1054. 10.7717/peerj.1054.

93. Jerber, J., Haldane, J., Steer, J., Pearce, D., and Patel, M. (2020). Dissociation of neuronal culture to single cells for scRNA-seq (10x Genomics). *protocols.io*.

94. Qiu, C., Cao, J., Martin, B.K., Li, T., Welsh, I.C., Srivatsan, S., Huang, X., Calderon, D., Noble, W.S., Disteche, C.M., et al. (2022). Systematic reconstruction of cellular trajectories across mouse embryogenesis. *Nat Genet* 54, 328–341. 10.1038/s41588-022-01018-x.

95. Rappaport, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* 2, 1896–1906. 10.1038/nprot.2007.261.

96. Demichev, V., Messner, C.B., Vernardis, S.I., Lilley, K.S., and Ralser, M. (2020). DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat Methods* 17, 41–44. 10.1038/s41592-019-0638-x.

97. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* 13, 731–740. 10.1038/nmeth.3901.

98. Benz, R., and McLaughlin, S. (1983). The molecular mechanism of action of the proton ionophore FCCP (carbonylcyanide p-trifluoromethoxyphenylhydrazone). *Biophys J* 41, 381–398. 10.1016/S0006-3495(83)84449-X.

99. Downs, K.M., and Davies, T. (1993). Staging of gastrulating mouse embryos by morphological landmarks in the dissecting microscope. *Development* 118, 1255–1266. 10.1242/dev.118.4.1255.

100. Huang, Y., Wilkie, R., and Wilson, V. (2015). Methods for precisely localized transfer of cells or DNA into early postimplantation mouse embryos. *Journal of Visualized Experiments* 2015, e53295. 10.3791/53295.

101. Huang, Y., Osorno, R., Tsakiridis, A., and Wilson, V. (2012). In Vivo Differentiation Potential of Epiblast Stem Cells Revealed by Chimeric Embryo Formation. *Cell Rep* 2, 1571–1578. 10.1016/j.celrep.2012.10.022.

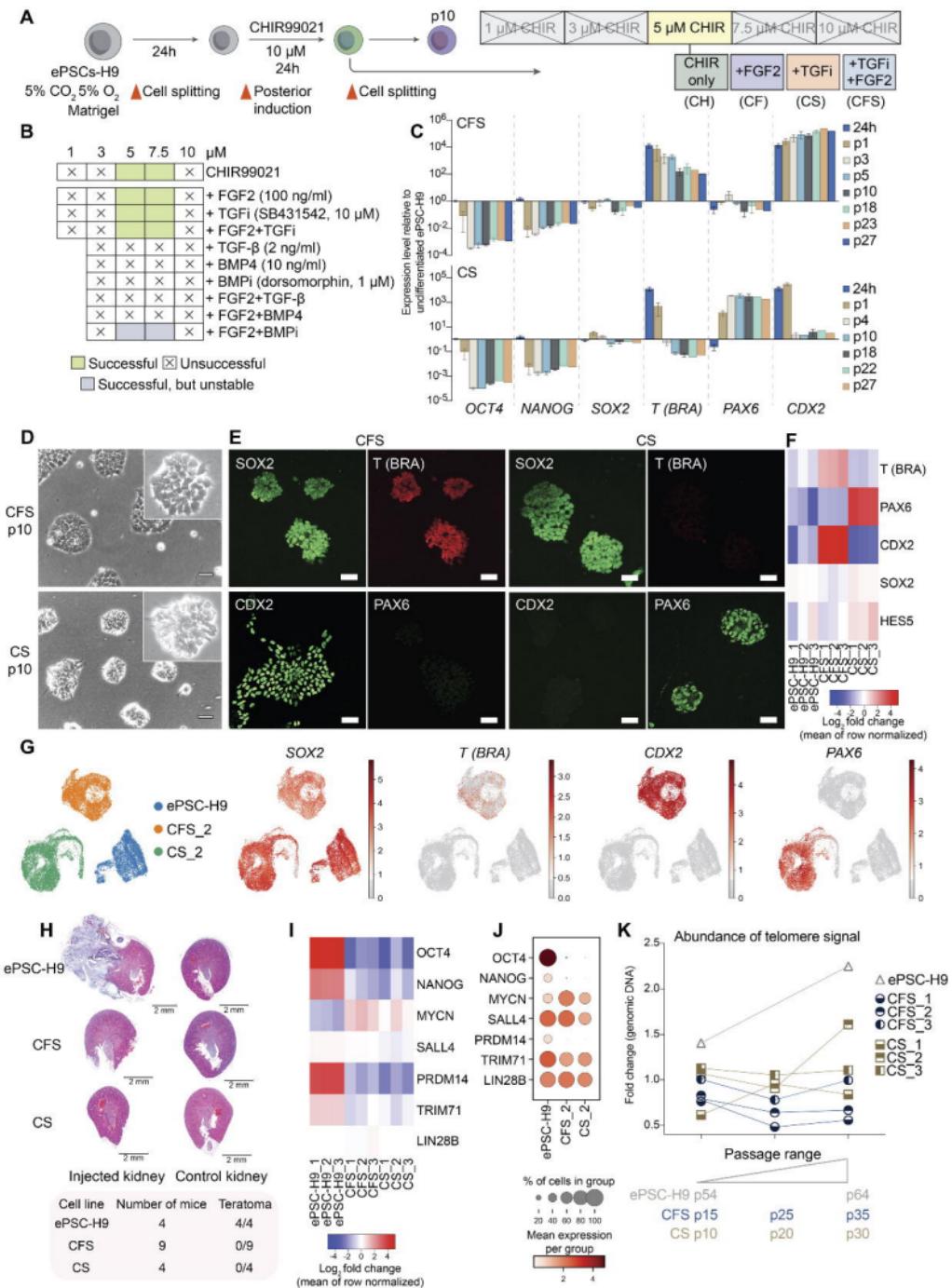


Figure 1

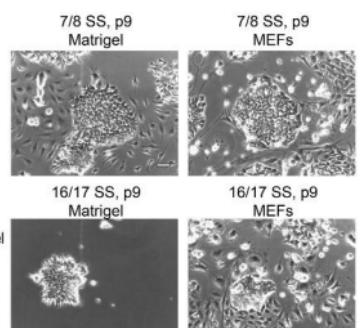
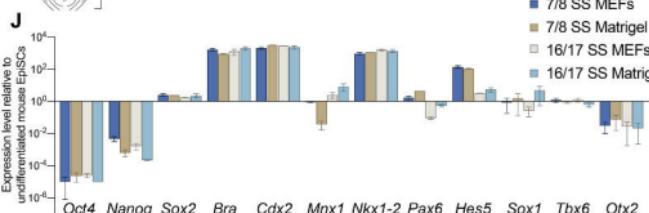
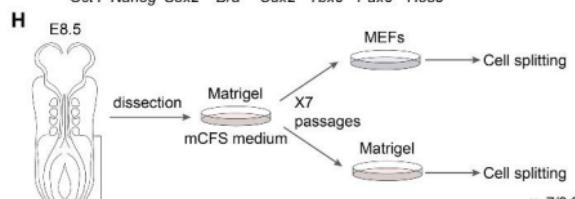
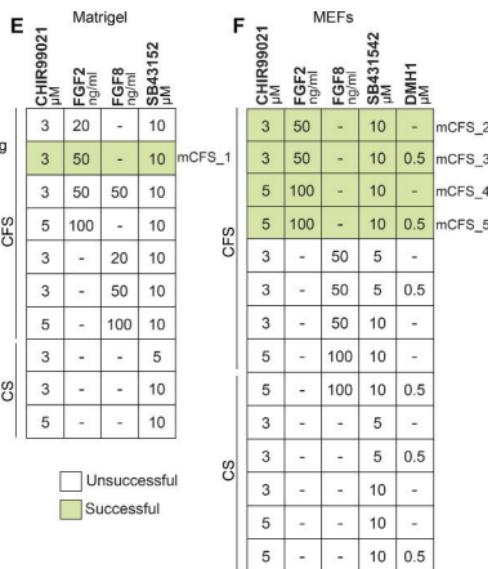
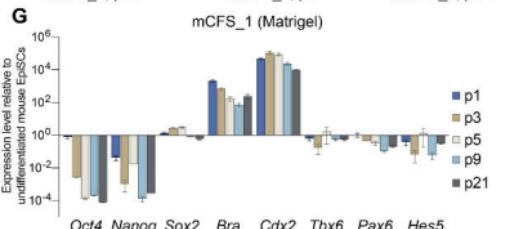
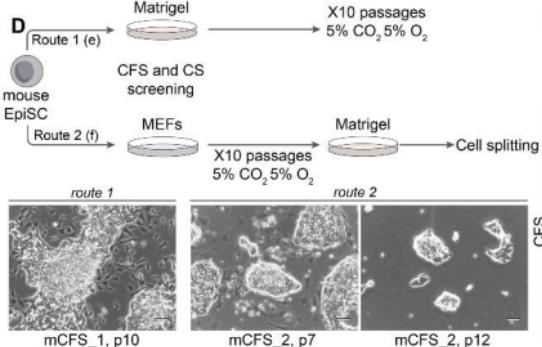
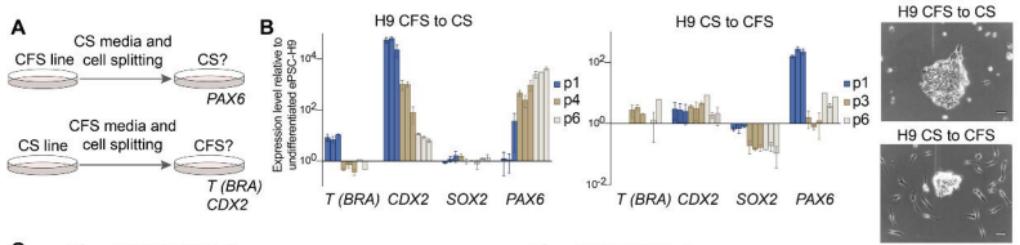


Figure 2

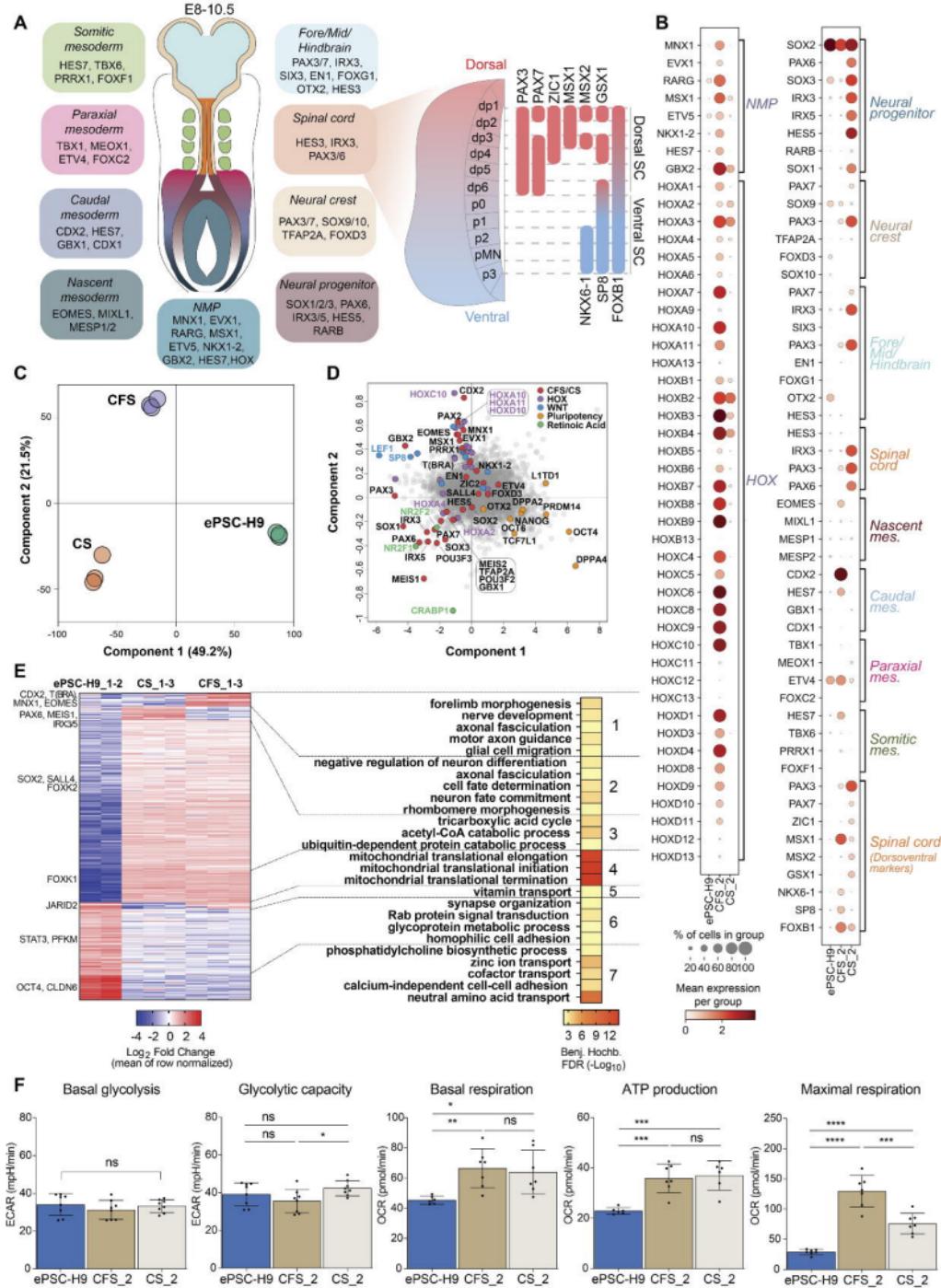


Figure 3

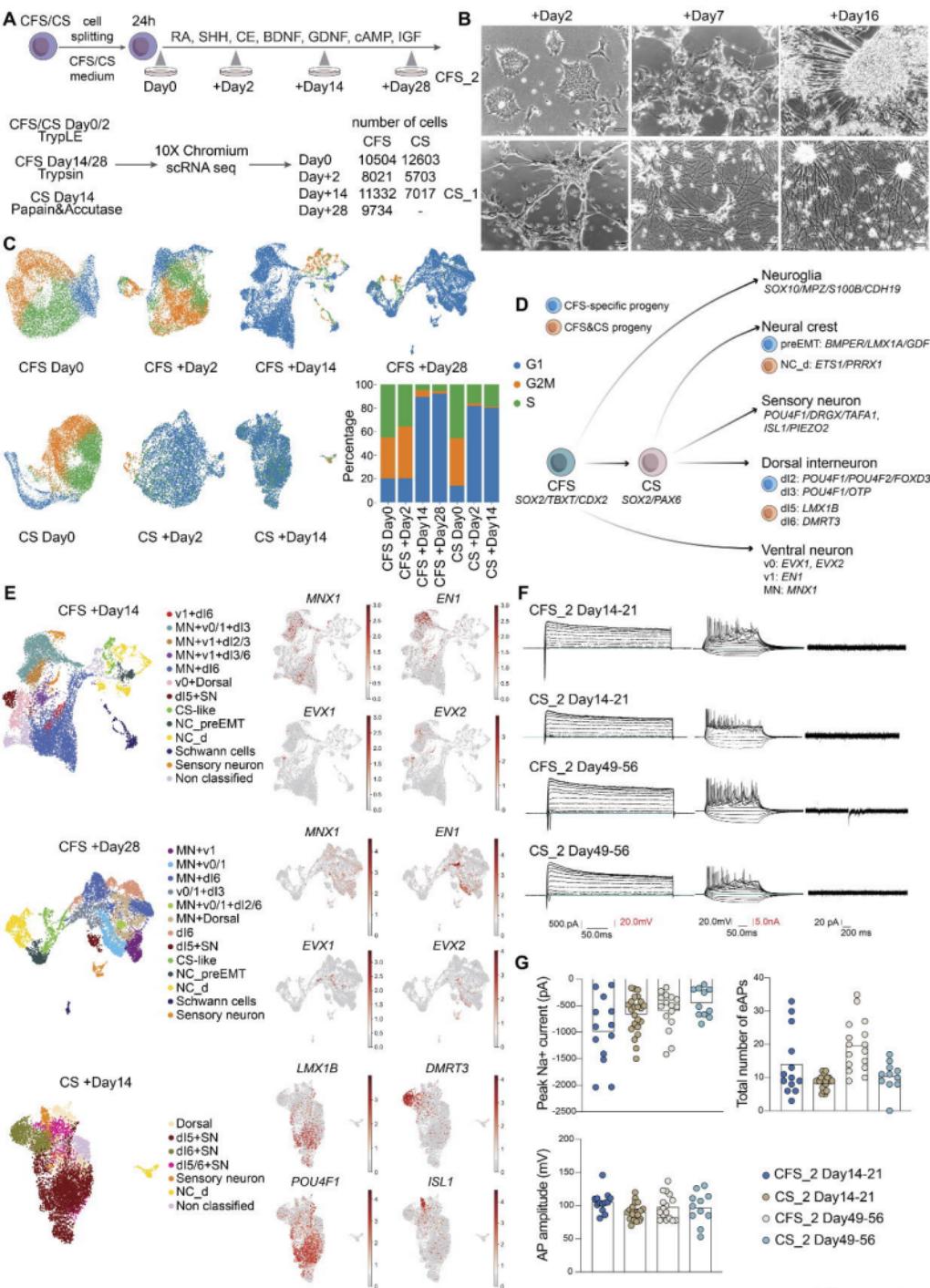


Figure 4

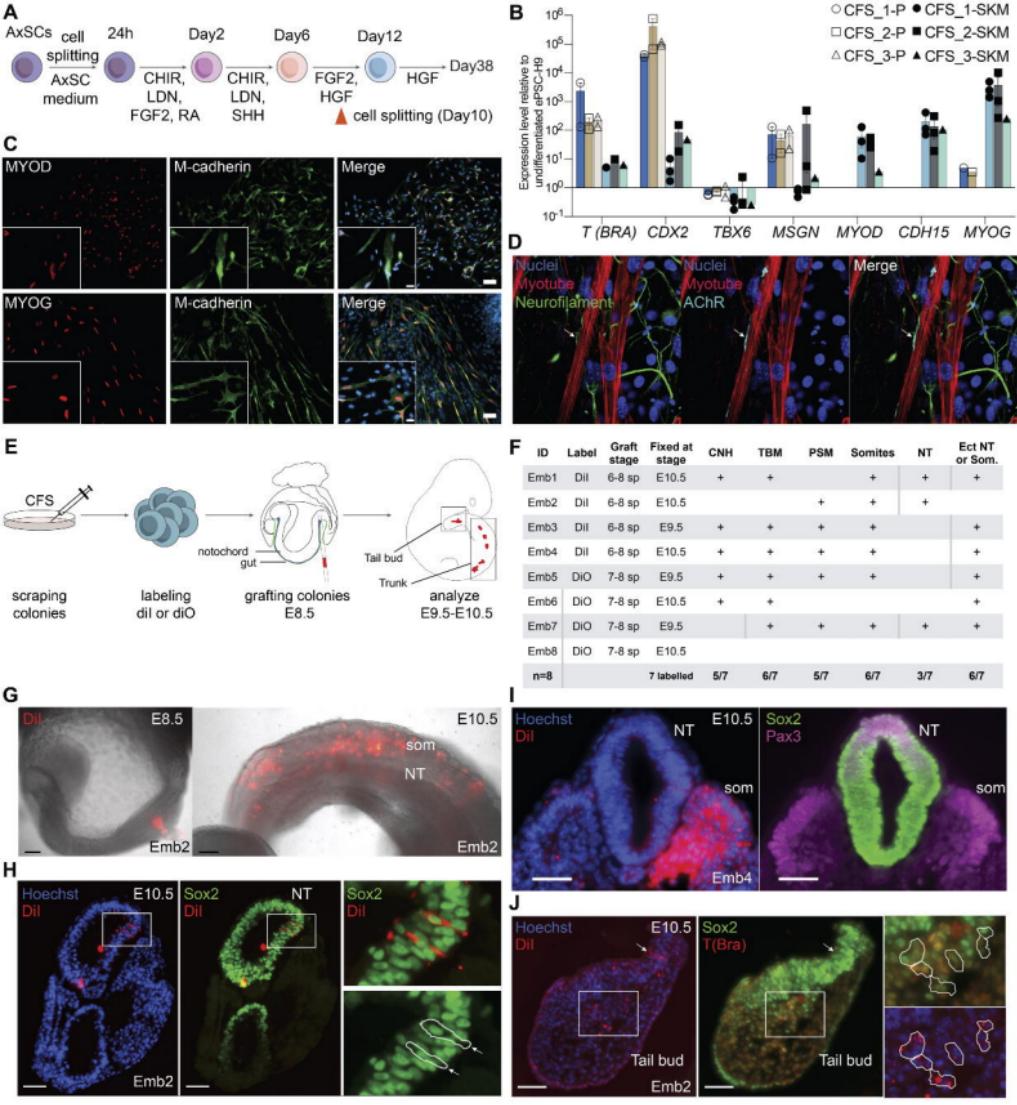


Figure 5