The Human Motoneuron Expression Signature is Defined by ALS-Related Genes

Authors: Archana Yadav[1]†, Kaya J.E. Matson[2,3]†, Li Li[2], Isabelle Hua[2], Pallavi Gaur[1], Mor R. Alkaslasi[4,5], Saadia Hasan[6], Ahmad Galuta[7], Annemarie Dedek[7,8], Sara Ameri[7], Jessica Parnell[7,8], Mohammad M. Alshardan[7], Feras Abbas Qumqumji[7], Saud M. Alhamad[7], Alick Pingbei Wang[7], Gaetan Poulen[9], Nicolas Lonjon[9], Florence Vachiery-Lahaye[9], Mike A. Nalls[10,11,12], Yue A. Qi[13,] Michael E. Hildebrand[7,8], Pierre-Francois Mery[14], Emmanuel Bourinet[14], Luc Bauchet[9,14], Eve C. Tsai[7], Michael E. Ward[6], Claire E. Le Pichon[4], Vilas Menon[1]*, Ariel J. Levine[2]*

Affiliations:

[1] Department of Neurology, Center for Translational and Computational Neuroimmunology, Columbia University; New York, NY, USA.

[2] Spinal Circuits and Plasticity Unit, National Institute of Neurological Disorders and Stroke; Bethesda, MD, USA.

[3] Johns Hopkins University Department of Biology, Baltimore, MD, 21218, USA.

[4] Unit on the Development of Neurodegeneration, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development; Bethesda, MD, USA.

[5] Department of Neuroscience, Brown University, Providence, RI, USA.

[6] Inherited Neurodegenerative Diseases Unit, National Institute of Neurological Disorders and Stroke; Bethesda, MD, USA.

[7] Neuroscience Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada.

[8] Department of Neuroscience, Carleton University, Ottawa, ON, Canada.

[9] Department of Neurosurgery, Gui de Chauliac Hospital Hospital, and Donation and Transplantation Coordination Unit, Montpellier University Medical center, Montpellier, France.

[10] Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA.

[11] Center for Alzheimer's and Related Dementias, National Institutes of Health, Bethesda, MD, USA.

[12] Data Tecnica International LLC, Glen Echo, MD, USA.

[13] Center for Alzheimers and Related Dementias, National Institute of Aging; Bethesda, MD, USA.

[14] Institute of Functional Genomics, Montpellier University, CNRS, INSERM, Montpellier, France.

† Equal contribution.

* Co-corresponding authors. Email: vm2545@cumc.columbia.edu and ariel.levine@nih.gov

Abstract: In neurodegenerative diseases of the human spinal cord, such as amyotrophic lateral sclerosis (ALS), motoneurons are particularly vulnerable to degeneration. It is hypothesized that their large size contributes to disease susceptibility, but the link between genetic variants associated with disease and cell-type specific degeneration is not clear. We characterized human spinal cord cells using single-nucleus RNA-sequencing and protein profiling. We found that human motoneurons displayed a unique expression profile characterized by factors involved in cytoskeletal structure, cell size, and degenerative disease (including ALS-associated genes SOD1, NEFH, OPTN, TUBA4A, PRPH, and STMN2) and that protein expression of these genes correlated with larger cell size in tissue. This work suggests a motoneuron-specific signature underlies their selective vulnerability to neurodegeneration.

One-Sentence Summary: Human spinal motoneurons preferentially express ALS genes, providing an explanation for their selective vulnerability to degeneration.

Introduction:

The human spinal cord relays, processes, and transforms sensory inputs and descending cues from the brain into sensory, motor, respiratory, and autonomic outputs. These critical processes rely on a diverse array of spinal cord cell types, each with their own functions, molecular repertoires, and vulnerabilities to injury or disease. Neurodegenerative diseases that affect the spinal cord, including amyotrophic lateral sclerosis (ALS), spinal muscular atrophy (SMA), and hereditary spastic paraplegia (HSP), cause weakness, paralysis, and death. In each case, only certain spinal cell types are selectively vulnerable: in HSP, corticospinal, sensory, and spinocerebellar neurons show degeneration [1-3]; SMA is primarily a developmental disease of spinal motoneurons [4]; and in ALS, corticospinal neurons and multiple populations of ventral spinal interneurons die in addition to the signature phenotype involving degeneration of spinal motoneurons [5-12].

These vulnerable neurons often share a defining characteristic of having long axons and large cell bodies, but the molecular mechanisms that link human disease mutations to the degeneration of specific cell types are not clear. For example, ALS-related genes are thought to be widely expressed across different neuronal subtypes and non-neuronal cells [13-20]. This observation has led to a hypothesis that a cell's vulnerability or resilience in ALS is due to undefined features in its underlying transcriptomic profile [17,19,21-24]. Thus, we sought to establish the first comprehensive catalog of the adult human spinal cord at single-cell resolution, and to determine if underlying transcriptional and proteomic profiles of certain cell types could predispose them to selective vulnerability in neurodegenerative diseases. We identified 65 cell types in the human spinal cord and provide a browsable interface (https://vmenon.shinyapps.io/hsc_biorxiv/) as a resource for the field. In motoneurons, we found an enrichment of genes involved in cytoskeletal structure, cell size, and degenerative disease. This work reveals a molecular basis for the selective vulnerability of motoneurons to degeneration.

Results

Molecular atlas of human spinal cord

We created a molecular catalog of the human lumbar spinal cord to investigate how specialized transcriptional repertoires of spinal cord cell types could endow them with their particular cellular features and functions. Recent papers have used single cell sequencing to identify human spinal cord cell types from embryonic tissue and an unpublished pre-print reports a single cell sequencing characterization of adult spinal cord tissue [25-27]. However, none of these studies examined motoneurons in detail or the expression patterns of genes associated with neurodegenerative disease, leaving the question of differential vulnerability of specific cell types unaddressed.

We obtained post-mortem lumbar spinal cord tissue from seven donor transplant cases (Fig. 1a and Data File Table S1), using neuroprotective conditions, such as body chilling and perfusion with a high magnesium solution, rapid dissection, and flash freezing of tissue immediately in the operating room. Single nuclei were obtained and a dataset of 55,420 nuclei that passed quality control filtering was obtained with 2,187 genes detected per nucleus, on average. Initial clustering of all nuclei clearly distinguished the major known cell classes present in spinal cord tissue, including oligodendrocytes and their precursors and progenitors, meningeal cells, astrocytes, endothelial and pericyte cells, microglia, and neurons; the latter included glutamatergic neurons, GABAergic/glycinergic neurons, and motoneurons. Comparison of these cell classes to our prior work in the mouse spinal cord [28] revealed substantial overlap in cellular signatures as well as notable differences. For example, oligodendrocytes accounted for a larger proportion of the human dataset. Although we cannot rule out differential sensitivity during dissociation, the higher oligodendrocyte fraction is consistent with the larger ratio of white matter to gray matter area in human versus mouse spinal cords (Supplementary Fig. 1a) and could reflect the relative expansion of long axon tracts linking the brain and spinal cord in humans. Overall, these major groupings showed clear segregation of previously reported markers for these cell types, thus allowing for further investigation within each of these broad classes (Fig. 1b, Supplemental Fig. S2-6), as described below.

Fig. 1

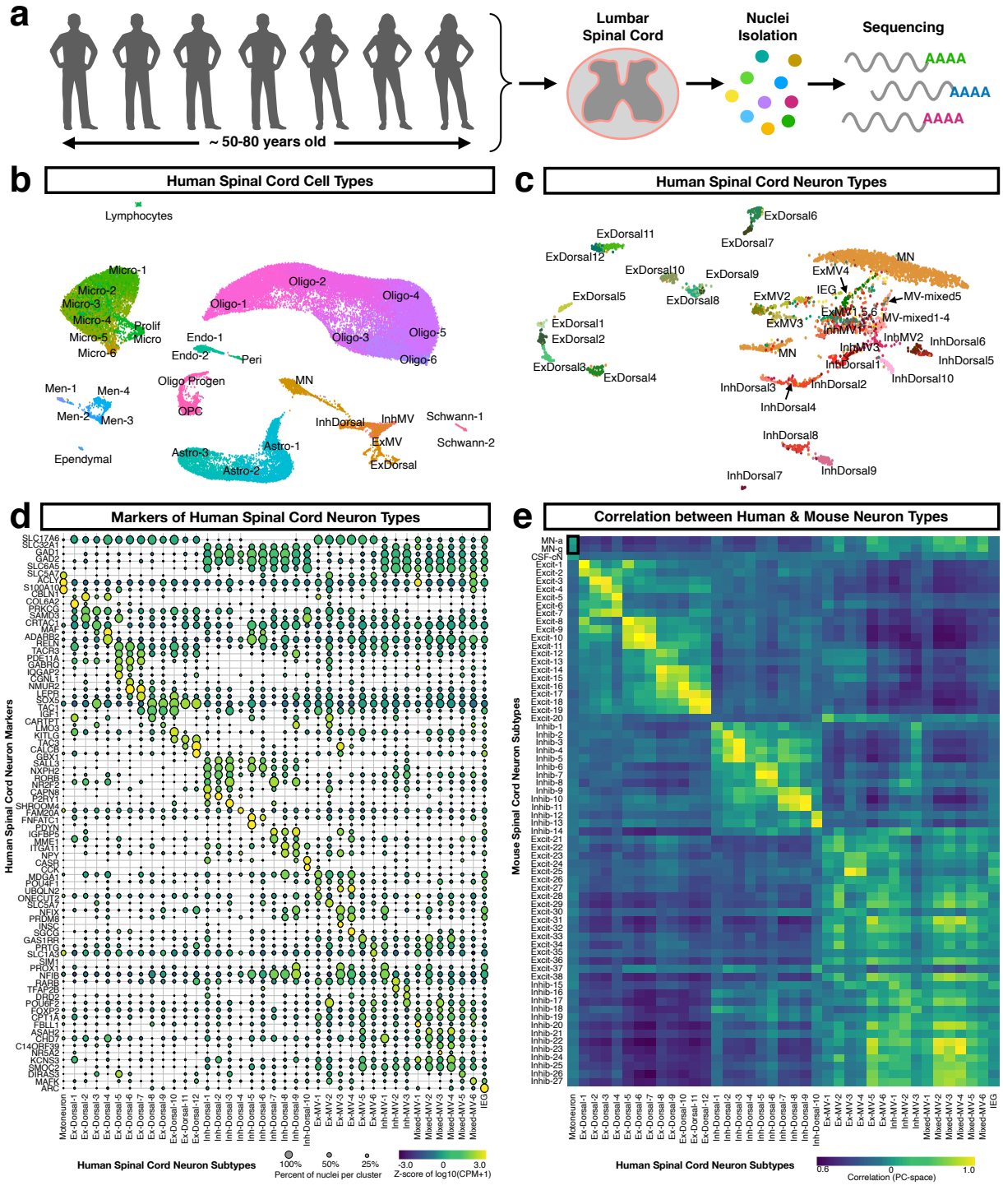


Fig. 1: A single cell catalog of the human spinal cord reveals the gene expression signature of human motoneurons. a, Lumbar spinal cord tissue was obtained from seven subjects (male and female, ~50-80 years old) and processed for single nucleus RNA sequencing. **b,** UMAP plot showing the major cell types of the human spinal cord, each in separate color. Cells of the oligodendrocyte lineage

are shown in pink/purple and include oligodendrocyte precursor cells (OPC), progenitors (Oligo Progen), six types of oligodendrocytes (Oligo-1 through Oligo-6), as well as two types of Schwann cells (Schwann-1 and –2). Microglia cells are shown in green and includes a putatively proliferating population (Prolif Micro) and six types of microglia (Micro-1 through Micro-6). Astrocytes are shown in turquoise and include three types (Astro-1 through Astro-3). Meninges are shown in blue and include four types (Men-1 through Men-4). Vascular cells are shown in teal and include two types of endothelial cells (Endo-1 and –2) and pericytes (Peri). Ependymal cells are shown in teal. Neurons are shown in orange and include five broad classes based on their neurotransmitter status and putative location: motoneurons (MN), excitatory dorsal neurons (ExDorsal), inhibitory dorsal neurons (InhDorsal), excitatory mid and ventral neurons (ExMV), and inhibitory or mixed mid and ventral neurons (InhMV). **c,** UMAP plot of human spinal neurons showing 39 refined populations. **d,** Dot plot of broad class and marker genes for the 39 neuronal populations of the adult human spinal cord. The size of each circle represents the fraction of nuclei in each cluster that express each gene and the color of each circle represents the z-score of the log normalized mean expression of counts per million of each gene in each cluster (Z-score of $\log_{10}$(CPM+1)). **e,** Heatmap correlation plot of the human spinal cord neurons compared to mouse populations from a recent harmonized analysis of multiple studies. Correlation is colored from purple to yellow and was calculated using principal components. A black box (upper left-hand corner) highlights the relatively low conservation of motoneurons between humans and mice.

Cell class-specific clustering identified subgroups within each of these broad classes. In total, we identified 2 subsets of Schwann cells, 7 subsets of oligodendrocytes (including putative progenitor cells), 4 subsets of meningeal cells, 3 subsets of astrocytes, 2 subsets of endothelia, a single cluster of pericytes, 7 subsets of microglia (including a putative proliferative type), and 39 neuronal subclusters (Fig. 1b,c and Data File Table S3, with a browsable interface: https://vmenon.shinyapps.io/hsc_biorxiv/). Comparison of human spinal cord neuronal populations to a harmonized mouse spinal cord atlas revealed strong conservation, particularly for cell types that reside in the dorsal horn of mouse [28]. This supported the prospective definition of dorsal or mid-ventral characterization for each human population, which we combined with neurotransmitter phenotype to create a nomenclature for human spinal cord neurons. Accordingly, non-motoneuron populations can be separated into glutamatergic, GABAergic/glycinergic classes, or mixed (prefixed with E/I/Mixed - excitatory/inhibitory/mixed annotation), and dorsal/mid-ventral (D/MV annotation) groups, each distinguished by a unique combination of marker genes (Fig. 1d). One neuronal cluster was defined by expression of immediate early response genes (IEG), though it is unclear whether this reflects neuronal activity/stress during the patient's life or post-mortem artifacts.

A more detailed analysis of human spinal neurons showed that the ventral clusters detected were less clearly segregated from each other as compared to dorsal clusters. These putative ventral groups had lower cluster robustness scores (Supplementary Fig. S8C-D, S9) and contained some mixed populations of VGLUT1 and VGAT-expressing neurons. In contrast, dorsal subpopulations were robustly distinguishable from each other, cleanly separated into glutamatergic or GABAergic/glycinergic sub-types, with signatures that were more highly

conserved between human and mouse (Fig. 1e). These observations likely reflect conserved spatial trends that we previously observed in the mouse spinal cord for discrete dorsal cell types versus overlapping ventral cell types [28]. Of note, human motoneurons displayed relatively low conservation with mouse motoneurons, compared with other human-mouse pairs of related cell types (Fig. 1e).

Human motoneurons are defined by genes related to cell structure, cell size, and ALS

We next focused on the gene expression profile of human motoneurons.[1] We examined the top 50 marker genes that distinguished this cluster from other human spinal neurons. Motoneuron marker genes included those involved in acetylcholine synthesis and function (SLC5A7 and ACLY), as expected, but surprisingly were dominated by three partially overlapping sets of genes: (1) those involved in cytoskeletal structure, (2) neurofilament genes related to cell size, and (3) those that are directly implicated in ALS pathogenesis (Fig. 2a).

Cytoskeletal components were the most abundant category of motoneuron marker gene and the most enriched gene ontology (GO) terms, including GO annotation clusters related to microtubules (p=0.000009) and axon structure and neurofilaments (p=0.000018) (Data File Table S4). The marker genes that were structural components of neurofilaments (NEFL, NEFM, NEFH, and PRPH) have been directly linked to cell size, axon diameter, and degeneration [29-34], providing a potential link between human motoneuron gene expression and cellular phenotype. Amongst ALS-related motoneuron marker genes, there were both cytoskeletal genes (NEFH, PRPH, TUBA4A, and STMN2), as well as genes that are not directly linked to cellular structure (SOD1, OPTN, and SPP1).

[1] Unfortunately, the human motoneuron cluster could not be divided into more refined types. This may reflect technical limits (these nuclei contained a relatively low number of genes per nucleus) or biological continua amongst motoneuron features. Co-clustering with mouse MNs from previously published datasets suggested a division into alpha/beta and gamma sub-types but these were not clearly separated by human marker genes. As a result, human motoneurons were analyzed as one group.

Fig. 2

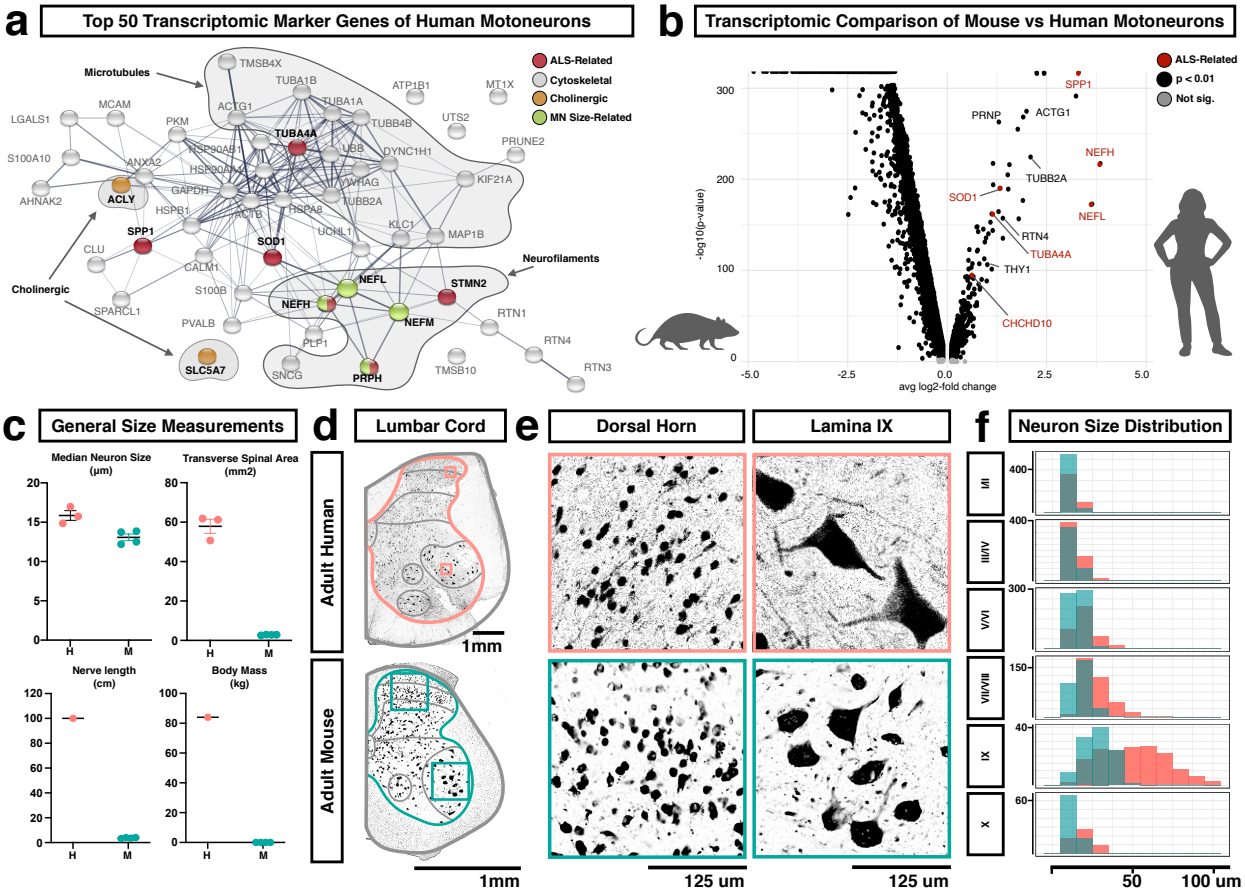


Fig. 2. Human motoneurons are characterized by genes associated with ALS, cell structure, and increased cell size. a, Association network plot constructed using the String protein database for the top 50 marker genes of human motoneurons. Genes related to cholinergic neurotransmission are shown in orange, genes related to ALS are shown in red, and genes whose over-expression in mice causes enlargement and/or degeneration of motoneurons are shown in green. Families of genes related to the microtubule or neurofilament cytoskeletal components are highlighted by gray. **b,** Volcano plot showing the distribution of genes enriched in either lumbar motoneurons from adult mice or lumbar motoneurons from adult humans, with several significant genes of interest labeled, including genes related to ALS (red). Genes are plotted by the average change in expression (avg $\log_2$-fold change) and by the statistical strength of the difference (-$\log_{10}$(p-value). Insignificant genes are shown in gray and genes that are significantly different are shown in black or red. **c,** Gross anatomical and neuronal measurements of the human (H) and mouse (M) lumbar spinal cords. Measurements include median neuron size (μm), transverse area of the spinal cord (mm^2), maximum nerve length (cm), and body mass (kg). **d,** Transverse sections of one side of the adult lumbar human (above) and mouse (below) spinal cords, with antibody labeling for NeuN. Images are representative of data from three subjects. Scale bars are 1 mm. Boxes indicate the regions shown in panel E. Gray lines indicate the laminar/regional boundaries used in panel F and were based on prior

work [35-38]. **e,** Higher magnification view of NeuN labeled spinal neurons from panel **d** in the human (above) and mouse (below). The left-side images are from the dorsal horn and the right-side images are of putative motoneurons in lamina IX. Scale bars are 125 µm. **f**, Histogram showing the count distribution of neuron feret distance (maximum caliper, similar to diameter) in human (pink) and mouse (teal) across the different lamina regions of the adult lumbar spinal cord. Measurements are given in µm and the count scale is shown at the right of each plot. Bonferroni-adjust Wilcox test p-values and Bhattacharyya Coefficients (BC) for human vs mouse distributions are as follows. I/II: p=7.5e-27, BC=0.93, III/IV: p=4.0e-12, BC=0.96, V/VI: p=3.2e-30, BC=0.89, VII/VIII: p=5.7e-49, BC=0.80, IX: p=1.6e-19, BC=0.71, X: p=9.5e-10, BC=0.92.

We further examined the expression of a panel of ALS-related genes compiled from the literature [18,39-45] across human spinal cord cell types. In addition to the genes above, we found that CHCHD10 and KIF5A were enriched in spinal motoneurons, extending this signature profile (Supplementary Fig. S12 and S14). We also observed enriched expression of SPP1, FUS, and C9ORF72 in microglia and NEFH, SOD1, STMN2, and TUBA4A in a mid/ventral (Mixed-MV-1) population (Supplementary Fig. S13, S14). TARDBP was not detected at sufficient levels in the dataset to characterize its expression pattern.

The enriched expression of neurodegeneration-associated genes in human motoneuron transcriptomics may have been partly due to the age of the study donors. We examined expression of ALS-related genes in a dataset of human embryonic spinal cord cell types [25] and found low levels of gene expression (i.e. NEFH and TUBA4A), moderate but broad cell type expression (i.e. OPTN and PRPH), or high and ubiquitous cell type expression (i.e. SOD1 and STMN2) (Supplementary Fig. S15). Thus, the enrichment of ALS-related genes in human motoneurons was not apparent in newly formed motoneurons but emerged at some point during motoneuron maturation or aging. To test whether this expression profile reflected a non-specific enrichment of degeneration-associated genes in human motoneurons with age, we compared the expression of genes for multiple neurodegenerative diseases across human spinal cord cell types. This analysis revealed a specific association of ALS-related genes in human motoneurons.

To determine whether ALS-related genes are also enriched in motoneurons in mice, the major animal model for studying the genetic basis of neurodegenerative disease, we compared the human data to prior single nucleus sequencing data from lumbar skeletal motoneurons from adult mice [46]. We found that prominent ALS-related genes were enriched and were expressed at higher levels specifically in the human motoneurons (Fig. 2b). To determine if this enrichment is unique to motoneurons, we examined the analysis of a recent study on conservation in human brain gene expression patterns [47] and found that three genes of interest (SOD1, TUBA4A, OPTN) had a significantly higher mean human to mouse divergence score than other assayed

genes (mean score of 0.587 ± 0.19 versus 1,426 other genes with mean 0.320 ± 0.123, p=0.0002).

Why might human motoneurons be defined by genes related to cell size and structure, compared to other human neurons and mouse motoneurons? It is well established that human motoneurons are large, but to answer these relative size questions, we analyzed neuron soma size across all laminae in human and mouse lumbar spinal cord tissue. Given the obvious differences in overall body size and anatomy, we expected that most classes of human neurons would be larger than mouse neurons. Surprisingly, we found that, overall, human and mouse lumbar spinal neurons were approximately the same size, with a median Feret diameter (maximal caliper length) of 16.02 and 13.13 µm, respectively (human mean 20.3 ± 0.28 s.e.m; mouse mean 14.28 ± 0.12 s.e.m.) (Fig. 2c and Data File Table S2). Indeed, across most laminae of the spinal cord, human and mouse neurons displayed somewhat similar size distributions. By contrast, human lamina IX spinal neurons were approximately 2-fold larger than those in mouse and could be up to ~120 µm across compared to ~50 µm in mouse (Fig. 2e,f and Data File S5). These measurements are consistent with those previously reported for human and mouse spinal motoneuron soma [48-50] and the same proportion that has been observed for human and mouse motoneuron axon caliber [51-53]. Assuming that human alpha motoneurons are within the higher end of this size distribution, then they are (1) much larger than other human spinal neurons, (2) increased in scale relative to mouse motoneurons, and (3) among the among the largest vertebrate neurons, including elephant motoneurons (~85 µm) [54], human Betz corticospinal neurons (~60-100 µm) [55], subsets of human dorsal root ganglion neurons (up to 100 µm) [56] and salmon Mauthner cells (~87 µm) [57]. This notable size of human motoneurons may explain the specialized gene expression signature that we observed. Furthermore, these results raise the possibility that evolutionary pressure has driven a concerted difference related to cytoskeletal and ALS-related genes in human versus mouse motoneurons to support their increased size relative to other neurons.

Cell-targeted proteomic profiling of human motoneurons

The single nucleus sequencing data above provided a genomic scale characterization of spinal cell gene expression but could not reveal whether the unique transcriptomic profile of human motoneurons was translated into a corresponding protein network. To capture a large-scale view of the human motoneuron proteome, we used a recently developed approach, Biotinylation by Antibody Recognition, followed by mass spectroscopy (BAR-MS) [58]. This allowed us to capture proteins in close physical proximity (up to 1 micron range) to a motoneuron-enriched gene, STMN2, in comparison with proteins near a generic neuron protein, NeuN (Fig. 3a and Supplemental Figure S16). We first confirmed that STMN2 was expressed in spinal motoneurons by comparison with the cholinergic marker ChAT, and found extensive co-localization of these proteins in large cells of lamina IX. STMN2 is also present in sensory afferent fibers in the dorsal horn. Following BAR-MS, a total of 1,509 proteins were detected, of which 103 proteins were statistically significantly enriched in the STMN2 condition (paired t-test $p < 0.05$) as compared to the NeuN condition (Data File S8). This group of enriched proteins showed over-representation for the top transcriptomic motoneuron marker genes

(hypergeometric p = 0.0161). Specifically, the cholinergic marker protein ACLY, as well as ANXA2, ATP1B1, HSPB1, LGALS1, and PRPH were significantly enriched. The proteins for several other transcriptomic motoneuron markers were also higher in the STMN2 condition (Fig. 3c), although they did not reach the significance threshold (Data File S8). By contrast, among the 17 proteins enriched in the NeuN condition, none were from the list of top transcriptomic motoneuron markers. STMN2 was not detected by the BAR-MS method, likely due to lower expression levels. Thus, while proteomic profiling that was targeted to human motoneurons

detected only a fraction of the endogenous proteome, these data provided independent and complementary support for the motor neuron transcriptomic signature that we described above.

Fig. 3

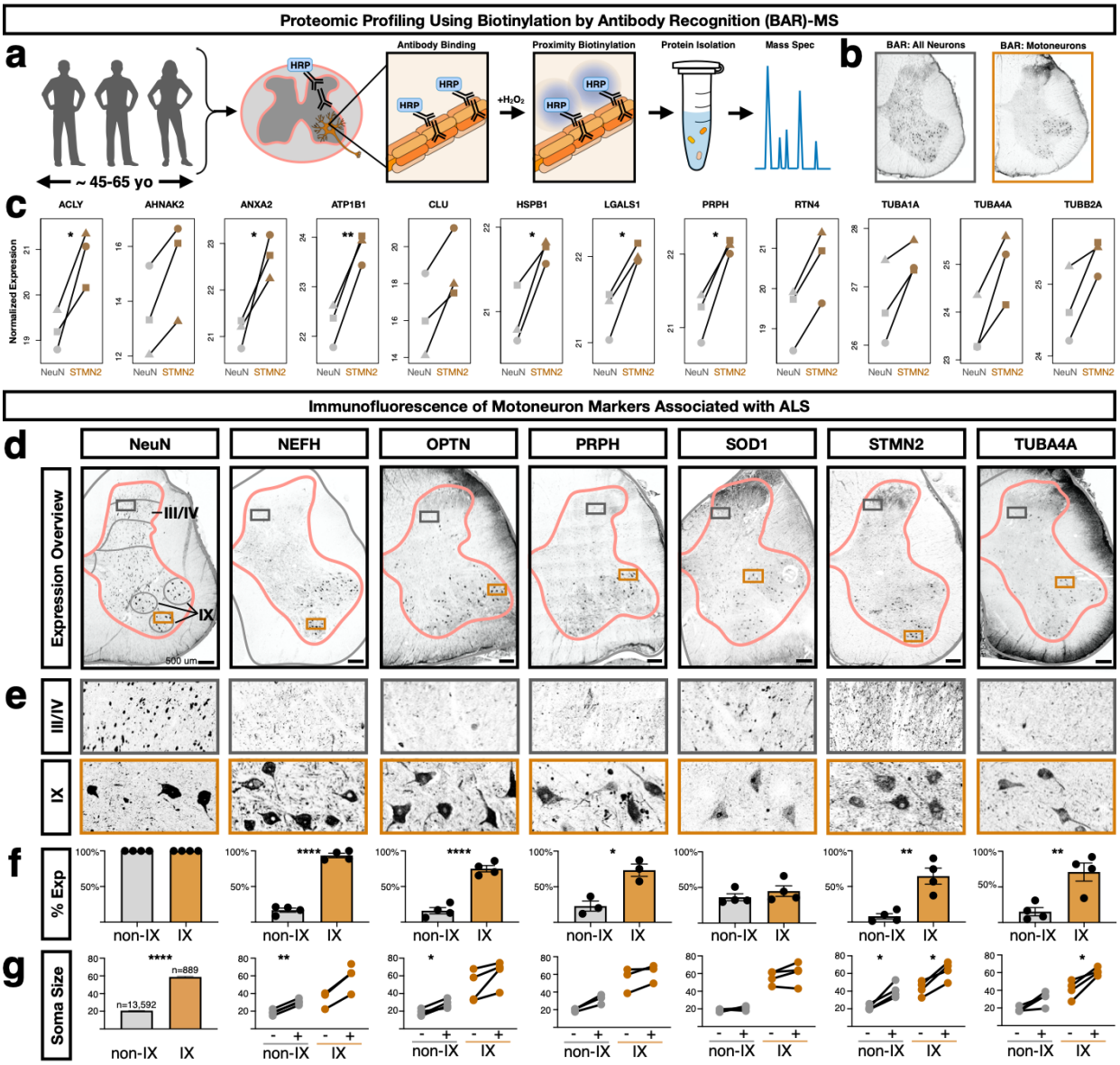


Fig. 3. ALS-related proteins are enriched in human motoneurons. a, Biotinylation by Antibody Recognition followed by mass spectroscopy (BAR-MS) was performed on lumbar spinal cord tissue from three subjects (male and female, ~45-65 years old). b, Streptavidin staining on an adult human lumbar spinal cord against NeuN (all neurons) and STMN2 (motoneurons as well as sensory afferent fibers). **c,** Protein expression for selected transcriptomic markers, as detected by BAR-MS. The plots show the Log2 fold change (Log2FC) of the mean-centered expression for NeuN (left) and STMN2 (right). Experiments from three individual subjects are shown, with lines connecting the paired

conditions within each subject. Paired t-test results are shown where * indicates $p < 0.05$ and ** indicates $p < 0.01$. **d,** Antibody staining on adult human lumbar spinal cord against NeuN (RBFOX3 gene, general neural marker) and the ALS-related genes NEFH, OPTN, PRPH, SOD1, STMN2, and TUBA4A. Gray matter outlines are shown in pink and boundaries of lamina I/II, III/IV, V/VI, VII/VIII, IX, and X are shown in gray. Boxes indicate the enlarged images in panel. **e,** Images are representative of data from three subjects (two male and one female). Scale bars are 500 µm. **e,** Inset of the images in panel **d,** from the boxed region in laminae III/IV or lamina IX. The width of the insets is 500 µm. **f,** Quantification of the percent of NeuN+ neurons that co-expressed the indicated proteins in either all neurons not in lamina IX (non-IX) or those in lamina IX. The mean ± s.e.m. are shown. The plotted values and number of cells counted in each subject and category are available in Data File Table S5). Paired t-test results are shown where * indicates $p < 0.05$, ** indicates $p < 0.005$, **** indicates $p < 0.0001$. **g,** The sizes of NeuN+ neurons are shown for each indicated protein. For NeuN, 100% of cells were positive, by definition, and the total counts and sizes (mean ± s.e.m.) are shown for neurons not in lamina IX (non-IX) or those in lamina IX. For all other indicated proteins, the feret distance sizes are shown for all neurons that did not (-) or did (+) express the indicated protein (mean feret distance in µm). Each line joins values within one subject. There is an unpaired value for NEFH because we did not detect neurons in lamina IX that did not express NEFH. The plotted values and number of cells measured in each subject and category are available in Data File Table S5. Paired two-tailed t-test p-values, after Benjamini-Hochberg FDR correction, are shown where * indicates $p < 0.05$, ** indicates $p < 0.005$. **** indicates $p < 0.0001$.

ALS-related genes are enriched in human motoneurons in situ and associated with cell size

To target key ALS-related factors in tissue and to compare protein expression and *in situ* cell size, we next analyzed the protein expression of six ALS-related genes in post-mortem lumbar spinal cord from four donors, using immunofluorescence: NEFH, OPTN, PRPH, SOD1, STMN2, and TUBA4A. We found that neurons expressing NEFH, OPTN, PRPH, STMN2, and TUBA4A proteins were all enriched within the motoneuron region (lamina IX) of the lumbar spinal cord, with limited positive cells in other regions except for scattered, large cells in lamina III/IV of the dorsal horn which may be projection neurons and smaller neurons in medial lamina VII (Fig. 3d-g and Data File Table S5). SOD1 was present in lamina IX and throughout the spinal cord in a distinct peri-nuclear distribution, in contrast to the enriched RNA expression that we detected by single nucleus RNA sequencing. To ensure the accuracy of the SOD1 expression pattern, we validated the SOD1 antibody through targeted knockdown in human iPS neurons (Supplementary Fig. S18D). Overall, these data confirm the expression of ALS-related proteins in human spinal motoneurons in tissue [14,59].

We also studied the expression of these proteins in the mouse spinal cord, using lumbar tissue from aged animals (11 months old) to approximate the advanced age of the human subjects in this study. We found that Nefh, Optn, Prph, Stmn2, and Tuba4a displayed enrichment in lamina IX, while Sod1 was expressed ubiquitously, similar to what has been previously described for Sod1 in mice (Supplementary Fig. S18 A-C) [14]. Together with the comparative transcriptomic

analysis above, this suggests that while human and mouse motoneurons are both enriched for expression of ALS-related genes, in human motoneurons the relative expression levels are higher and the enrichment of these genes as motoneuron-specific markers is greater.

Finally, we tested the relationship between expression of ALS-related genes and cell size within human spinal neurons in tissue. We measured the feret distances of human neurons expressing each ALS-related protein in comparison with non-expressing neurons. We found that neurons that expressed NEFH, OPTN, PRPH, STMN2, and TUBA4A were generally larger than non-expressing neurons, both within the motoneuron region of lamina IX and in other lamina (Fig. 3g). Within lamina IX, this likely reflects enrichment within the larger alpha motoneurons (versus gamma) and in other laminae, this may reflect expression within spinocerebellar projection neurons that degenerate in ALS [9,10] or other large cell classes. Importantly, we found that the very largest lamina IX neurons – that are known to be most susceptible to degeneration in ALS [11,51,52,60] – were the most likely to express these markers. For lamina IX neurons with a feret distance greater than 70 μm, on average 100% expressed NEFH, 81% expressed OPTN, 88% expressed PRPH, 60% expressed SOD1, 90% expressed STMN2, and 95% expressed TUBA4A (Data File Table S5). These data further link motoneuron size and vulnerability to these cytoskeletal genes that have causative roles in motoneuron size and human disease.

Discussion

Why particular subsets of neurons are selectively vulnerable in various diseases is a fundamental challenge in understanding and treating neurodegenerative and motor disorders. In the human spinal cord, a prevailing model of spinal cord motoneuron degeneration in ALS and SMA is that their large cellular size renders them uniquely susceptible to generic molecular milieu and cellular stressors. However, the underlying assumption that disease related genes are broadly expressed across vulnerable and resistant spinal cell types is based on limited evidence from human tissue. Here, to carefully characterize shared and cell type-specific signatures, we established the first comprehensive single cell atlas of the adult human spinal cord, used cell-targeted proteomic profiling, and performed in tissue marker analysis. We found that human motoneurons were selectively distinguished by factors involved in the cytoskeleton, motoneuron size, and ALS, linking gene expression, cell structure, and disease.

Although we captured all major cell types and most known subclasses of cells in this catalog, we foresee further advances as additional data sets of this type are generated. In particular, the motoneuron population itself does not clearly segregate into subgroups based on molecular profile. Although there is enrichment of known motoneuron subtype markers in subsets of these cells, their overall profiles are not distinct enough to form discrete subgroups; this may be technical, due to the overall signal-to-noise ratio of single-nucleus RNA-seq in key genes within these subgroups, or it may be that motoneuron substructure is continuous at the transcriptomic level. This also applies to the ventral neuron subgroups; here, although we identify subclusters within these neurons, they were less discrete from each other than were the dorsal neuron subgroups. As technological advances allow for higher-sensitivity transcriptomics on large numbers of ventral and motoneurons, a clearer picture of the heterogeneity within this groups may become apparent. However, this does not affect our ability to identify robust signals

distinguishing motoneurons from other classes of neurons, especially when combining single-nucleus transcriptomics with protein validation. Here, we show how these complementary methods reinforce the overall molecular features that distinguish motoneurons from other neuronal types, suggesting specific aspects underlying their selective vulnerability in ALS and other motoneuron diseases.

This work builds on recent efforts towards understanding molecular and cellular heterogeneity in the human spinal cord, particularly during development. Rayon and colleagues [25], focused on first trimester spinal cord derived from four human embryos, and identified diverse progenitor and neuronal populations, and performed a systematic comparison with the spinal cord cell types of the developing mouse spinal cord. Zhang and colleagues [26] profiled the early and mid-stages of fetal development with an important focus on glial development and cell-cell communication. For the adult human spinal cord, Zhang and colleagues performed single nucleus RNA-seq on the spinal cord from two donors and identified glial and neuronal cell types among 15,811 nuclei [27]. However, they did not characterize human motoneurons, the focus of our work here.

Our data on the transcriptomic and protein signatures of human motoneurons strengthen the case for selective vulnerability based on fundamental requirements of motoneuron cell size, axonal caliber, and microtubule stability [41,61]**.** In general, all cells require a functional cytoskeleton, raising the question of why spinal motoneurons in particular are so crucially dependent on the proper expression and function of neurofilament and microtubule networks. The neurofilament genes that were enriched in human spinal motoneurons compared with other neuronal populations – NEFL, NEFM, NEFH, and PRPH – are precisely those structural components that drive increased axon caliber and cell size [53,62-64]. Over-expression of mouse NEFL, human NEFM, human NEFH, or mouse PRPH in transgenic mice can each cause enlargement and swellings of motoneuron somas and subsequent axon degeneration [29-34], providing a direct link between human motoneuron gene expression and cellular phenotype. Relatedly, these neurofilament genes are found in other large neurons in the brain and peripheral nervous system, suggesting that they may be part of a common signature that permits increased cell size [59,65-68].

The microtubule components and regulators distinguishing human motoneurons in our study likewise suggest particular cellular requirements for stable cytoskeletal networks. Amongst numerous actin, tubulin, and kinesin components that were enriched in human motoneurons, two factors bear special mention. STMN2 is a regulator of microtubule and axonal stability that was recently found to be a rare genetic cause of ALS [42] and may be a major downstream mediator of TDP-43 function in axon degeneration [43,69,70]. TUBA4A is an unusual tubulin isoform that lacks a critical tyrosine site for regulating microtubule dynamics and is thus a constituitively stable α-tubulin variant [71] and TUBA4A variants are associated with ALS genetically [15]. A robust cytoskeleton may be important in human motoneurons to facilitate axonal transport of mitochondria, protein, and RNA in the long motoneuron axon [61] or may provide structural support to protect motoneuron axons against microtears and stretching during body movement.

Our multimodal data sets support a model in which neurons with long projections such as human motoneurons and corticospinal neurons evolved a particular complement of cytoskeletal genes to sustain their extreme cellular phenotypes. Large soma size and axon caliber may be required to sustain extensive dendritic trees and axons up to a meter long, to support cell energetics, or for

firing rate and conduction parameters [72-74]. These large cells then rely critically on this protein network and are selectively vulnerable to its abnormal function. If correct, this model has direct implications for designing treatments for ALS and related neurodegenerative diseases. For example, targeting the expression of genes that are enriched in healthy motoneurons may have undesired effects on unaffected motoneurons [16].

It is important to note that large cell size in itself does not guarantee selective vulnerability; independent intrinsic factors certainly contribute to the pathogenesis and selectivity of motoneuron disease. While some types of large human neurons display similar vulnerability, others, such as peripheral proprioceptive sensory neurons, are not as susceptible to degeneration. This may reflect intrinsic cellular factors such as the ability of sensory neurons to regenerate and repair their peripheral axons, or other factors that modify vulnerability in resistant versus vulnerable cell types [19,21]. The motoneuron-specific signatures in our study, therefore, provide a clearer picture of potential interplay between multiple factors associated not just with cell size, but also cellular structure maintenance that may be dysregulated in disease.

Finally, it is critical to consider the spinal cord as a *community* of cell types that function together in normal health and disease. While our work here focused on the molecular signature of motoneurons, the single cell data set that we present provides the first comprehensive resource of the cell types of adult human spinal cord, and will have broad implications for understanding spinal cord biology. For example, in the context of ALS, non-cell autonomous influences from pre-synaptic neurons and glia play major roles in pathophysiology [17,75-78]. We observed enriched expression of several cytoskeletal ALS-related genes in a mixed ventral horn neuronal population which may correspond to the V2a and V1 populations of spinal interneurons. These cell types been shown to degenerate and withdraw their synapses from motoneurons at an early stage of ALS in mouse models and could also correlate with interneurons of the dorsomedial region of lamina VII that display particular loss and degeneration in human patients with ALS [6-8,12]. Amongst non-neuronal cells, microglia in particular contribute to ALS disease progression in a manner that depends on the genetic and disease context [79-82]. We observed enriched expression of the ALS-related genes SPP1, FUS, and C9ORF72 in microglia, suggesting specific biological processes that could be related to non-cell autonomous disease mechanisms. These preliminary findings highlight the potential of the human spinal cord atlas to parse how ubiquitous genetic alterations interact with diverse cell-type specific molecular profiles to cause and promote neurodegeneration. Ultimately, we hope that our work will serve as a broad resource and foundation for studying the wide range of cell types involved in sensory and motor function in the human spinal cord.

Materials and Methods:

Human spinal cord acquisition and preparation. Spinal cords for single nucleus RNA sequencing were obtained from brain-dead organ-donor patients (~50-80 years old, 4 men, 3 women) under the approval of the French institution for organ transplantation (Agence de la Biomédecine) or the Ottawa Health Science Network Research Ethics Board, following the template provided by the University of Ottawa and the Tri-Council Policy Statement Guidelines. Both approvals imply consent for using anonymized donor genetic information. Human lumbar spinal cords were retrieved under chilled body and neuroprotective conditions as described

previously [83-85]. The extraction procedure took 20-40 minutes and was done within three hours of cessation of circulation by aortic cross-clamp. Lumbar spinal cord tissue was flash frozen on liquid nitrogen in the operating room and stored at -80°C until nuclei isolation.

For immunohistochemistry experiments, lumbar spinal cord tissue was isolated from organ-donor patients (~55-65 years old, 3 men, 1 woman). The tissue was immediately fixed in 4% paraformaldehyde for 24-48 hours, then washed in PBS, and placed in 30% sucrose for 2-4 days at 4°C before being embedded in OCT medium for sectioning.

Mouse work and spinal cord acquisition. All procedures and experiments were approved by the Animal Care and Use Committee of NINDS (protocol #1384). Adult mice were of 50:50 mixed background from strains C57BL/6J and BALB/CJ, housed in standard conditions. For basic anatomical experiments, two male and two female mice of approximately 24 weeks old were used. For ALS marker gene expression studies, two male and one female mice of approximately 11 months old were used. To obtain spinal cord tissue, anesthetized mice were transcardially perfused with PBS followed by cold 4% paraformaldehyde (PFA). The spinal cords were harvested and post-fixed in cold 4% PFA overnight at 4°C, cryoprotected by immersion in 30% sucrose overnight at 4°C and embedded in OCT medium for sectioning.

Nuclei isolation. Nuclei were isolated from fresh frozen human spinal cords using a triton-based protocol [86]. Briefly, after removing the dura, half a segment of spinal cord was placed in a Dounce homogenizer (Kontes Dounce Tissue Grinder) containing 500 µL of lysis buffer (0.32 M sucrose, 10 mM HEPES [pH 8.0], 5 mM CaCl2, 3 mM 586 MgAc, 0.1 mM ETDA, 1 mM DTT, 0.1% Triton X-100). After douncing with 5 strokes of pestle A and 5-10 strokes of pestle B, the lysate was diluted in 3 mL of sucrose buffer (0.32 M sucrose, 10 mM 588 HEPES [pH 8.0], 5 mM CaCl2, 3 mM MgAc, 0.1 mM ETDA, 1 mM DTT) and passed over a 70 µm strainer. The filtered lysate was centrifuged at 3,200 x g for 5 min at 4°C. After centrifugation, the pellet was resuspended in 3 mL of sucrose buffer and centrifuged again at 3,200 x g for 5 min at 4°C. After centrifugation, the pellet was resuspended in 3 mL sucrose buffer and incubated for 2 min on ice. The sample was transferred to an Oak Ridge tube and homogenized for 1 min using an Ultra-Turrax Homogenizer (IKA). Then, 12.5 mL of density sucrose buffer (1 M sucrose, 10 mM HEPES [pH 8.0], 3 mM MgAc, 1 mM DTT) was layered below the sample. The tube was centrifuged at 3,200 x g for 20 min and the supernatant immediately poured off. The nuclei on the side of the tube were resuspended with 100 µL of PBS with 0.04% BSA and 0.2 U/µL RNase inhibitor. Nuclei were inspected for visual appearance and quantified with a hemocytometer before proceeding with nuclei capture and sequencing.

Single nucleus RNA sequencing. Single nucleus RNA sequencing was carried out using Single-cell gene expression 3' v3 kit on the Chromium platform (10X Genomics) according to manufacturer's instructions with one modification. Following reverse-transcription, an additional PCR cycle was added to the number of cycles for cDNA amplification to compensate for decreased cDNA abundance in nuclei compared to cells. Libraries were sequenced to a minimum depth of 20,000 reads per nucleus using an Illumina HiSeq 3000 (PE 26 – 8 – 98 bp). Raw sequencing reads were demultiplexed, aligned, and a count matrix was generated using CellRanger. For alignment, introns and exons were included in the reference genome (GRCh38).

Quality check analysis. All the 10x runs for each human sample were initially filtered with an nUMI cutoff of >1000 and then nuclei with less than 5% mitochondrial gene contamination were retained. Next, the mitochondrial genes were also removed from the matrices.

Top level UMAP and clustering. The 8 human datasets were integrated using SCTransform normalization followed by CCA based integration from Seurat 4.0 [87,88]. The integrated sets were then jointly analyzed to identify optimal Principal Component values based on ElbowPlot and PCheatmaps. PC value of 30 was used for clustering and UMAP. The clusters, obtained from resolution of 0.6, were then manually annotated based on the expression of marker genes for various cell types namely Neuron, Astrocytes, Microglia, Oligodendrocytes, OPC, Endothelial, Pericytes, Meninges. etc.

Sub clustering of major cell types. Identification of subclusters within cell types was performed for each cell type separately (Neurons, Microglia, Astrocytes) or as groups (Group 1- Oligodendrocytes, OPC, and Schwann cell; Group 2- Endothelial, pericytes, meninges and lymphocytes). In addition, the subclustering was also done in multiple rounds if any doublets or contamination of other cell types as observed (described below).

For subclustering of major cell types, the raw counts were extracted from 8 datasets, for each cell type and then re-normalized (log normalization) and scaled in order to prepare for integration. The integration of 8 datasets belonging to a particular cell type was performed based on CCA-integration workflow from Seurat 4.0. Optimal PC values were selected based on ElbowPlot and PCheatmaps for each cell type in order to be used for further sub clustering and preparation of UMAPs. Multiple resolutions were interrogated depending on cell type ranging from 0.08 to 3.

During each round, the doublets and contamination of other cell types was removed and the above steps were performed again. Doublets were identified by clusters that expressed markers for more than one cell type. All clusters were checked for doublets by their markers using wilcox and auroc tests, as well as FeatureScatter in Seurat.

Subclustering of neurons. Neurons were clustered in 2 stages, first dividing the neurons into motoneurons, ventral neurons and dorsal neurons and, second, further subclustering within motoneurons, ventral neurons and dorsal neurons.

During the first stage, the log normalization of raw counts and scaling (regressing nUMI and mitochondrial effects) of each dataset was done followed by integration based on above steps. During second stage the raw counts were again extracted from each group (MN, ventral and dorsal) and normalized using SCTransform in order to avoid dataset size related limitations and followed by integration from Seurat 4.0. In order to obtain a refined set of neuronal subpopulations, all the sub clusters were interrogated for ‘low quality’, doublets and other contamination and were subsequently removed from the analysis. All the refined clusters were then re-integrated to prepare a combined neuronal UMAP and mapped with refined subcluster annotations.

During each sub clustering, the differentially expressed genes were identified based on Wilcox Rank Sum test and ROC analysis within FindMarkers function from Seurat 4.0. The distinct subpopulations based on expression of candidate markers were manually annotated.

Cluster robustness assessment and silhouette scores. We used two approaches to assess cluster robustness: a post-hoc machine learning-based classification approach, and a silhouette score approach.

For the post-hoc machine learning approach, we built a random forest classifier for every pair of neuronal clusters, trained on 80% of the data. This classifier was then used to assign cluster membership for the remaining 20% of the cells, and the entire process repeated such that each cell in every pairwise cluster comparison was classified 100 times. A cell that was classified into its original cluster <90 times was deemed "misclassified". For every pair of clusters, we then calculated the mean percentage of cells that were misclassified among the two clusters to generate pairwise cluster robustness scores. For visualization as a constellation diagram, we only connected cluster pairs with minimum misclassification percentage >3%, representing their connections with the mean misclassification percentage.

For silhouette score evaluation, we used the 'silhouette' function from the 'cluster' library in R (https://cran.r-project.org/web/packages/cluster/index.html), where the distance matrix with 25 PCs and neuronal cell type annotations were used as input.

Cross species analysis between human spinal cord vs mouse meta-analysis datasets. Cross species comparison between human and mouse meta-analysis [28] spinal cord datasets were performed in two stages: 1. top level including all major cell types and 2. neurons only.

In both cases, the orthologous genes within mouse data matrix were converted to human homologs using biomaRt package [89] from Bioconductor and in-house scripts. The raw counts from both human and mouse datasets were then split by different samples and then re-normalized, scaled and integrated. For top-level integration, SCTransform based integration was performed whereas for neurons log normalization-based integration was performed. After that the UMAPs and correlation matrices were generated for further cross-species comparison of various cell types at top level and neuronal sub-clusters.

Cross-correlation of human and mouse cluster expression. Cross-species cluster correlation measures were calculated from PCs in the integrated space (using 20 PCs for the top level comparison of major cell classes, and 22 PCs for the neuronal subclass comparison, so as to align with the PC number used for UMAP representation). Aggregate correlation values for each pair of clusters (one mouse, one human) were calculated as the mean correlation value across all human-mouse nuclei pairs from the respective clusters.

GO analysis of human motoneuron marker genes. The top markers (based on smallest adjusted p-value) of human spinal motoneurons were determined based on the Wilcox Rank Sum test and were analyzed using DAVID 6.8 GO enrichment analysis (https://david.ncifcrf.gov/summary.jsp). The general categories of GOTERM_BP_DIRECT, GOTERM_CC_DIRECT, and GOTERM_MF_DIRECT were analyzed and functional annotation clustering was performed using default parameters including medium classification stringency.

Focused comparison of mouse and human motor neurons. Human motor neurons were compared to mouse lumbar skeletal motor neurons from a recent study [46]. Mouse MN genes were converted to human homologs using Homologene (https://CRAN.R-project.org/package=homologene). Only genes with human homologs present in both datasets were included in the analysis (13,574). Raw counts were extracted from each original dataset, normalized using SCTransform, and integrated based on integration anchors. Clustering was performed as described above (resolution = 0.4), and differentially expressed genes were identified based on Wilcox Rank Sum test and ROC analysis within FindMarkers function from Seurat 4.0.

Analysis of evolutionarily convergence/divergence scores. All available data on gene expression-based human:mouse divergence scores was downloaded from Pembroke et al [47]. Genes of interest were then extracted, yielding scores for three genes (SOD1, TUBA4A, OPTN) that overlapped with this data. We compared the mean and standard deviation of these three genes to the same metrics for the remainder of the assayed genes from the Pembroke report (N=1426 other genes) using a standard two-sided t-test.

Neurodegenerative disease gene analysis. Post-QC scRNAseq count data was extracted for seven major cell classes of interest. For each gene per cell class, mean expression was calculated across all assayed cells of that class. These means were then transformed to a Z scale to facilitate comparisons across multiple cell types. The Z scaling was carried out using the mean and standard deviations as the scaling functions as this is the common convention for this conversion. Additionally, genes that did not have any count based data available for that cell class were set to zero at Z scaling. From this large dataframe of normalized counts per cell type, candidate genes for HSP, PD and ALS were were extracted from Genomics England Expert Panel App genes audited at the "green" level of confidence [https://panelapp.genomicsengland.co.uk/]. AD genes were annotated by an expert panel and extracted [90] and the ALS list was also supplemented with genes from the literature, as described in the main text. These extracted genes were then clustermapped using the python package seaborn with Z scores greater than 7 truncated to a value of 7 for display purposes.

Proximity labeling by antibody recognition in fixed spinal cord tissue. Proximity labeling experiments were performed on free floating 30 µm cryosections of fixed human spinal cord from three adult tissue donors (2 male and 1 female). For each target primary antibody, 80 sections from each donor were used and for each negative control, 40 sections from each donor were used. Sections were incubated for 1 hour at room temperature in blocking buffer (1% IgG-free BSA, 10% normal donkey serum, 0.1% Triton-X 100 in PBS), washed three times in TBST buffer, and incubated with antibodies (NeuN 1:1000 and STMN2 1:1500) for 45 hours at 4°C. The paired negative controls were treated the same except exposure to primary antibody. The labeled slices were treated with 0.3% H_2O_2 to quench endogenous peroxidases for 30-min. After thoroughly washing off the quench, slices were incubated in horseradish peroxidase (HRP)-conjugated secondary antibody in block buffer for 1-h; followed by incubation in phenol-biotin (10mM) in 1% FBS and TBS for 1-h and treatment with 0.003% H_2O_2 in the phenol-biotin buffer for 10-min. The slices were quenched using cold TBS + 500mM sodium ascorbate + 10mM sodium azide for 5-min and were lysed in 3% SDS and 2% sodium deoxycholate in TBS by boiling at 99°C for 1-h. To homogenize the spinal cord, we sonicated the lysate for 40-s on and 40-s off for 15-min. The samples were boiled at 99°C for an additional 30-min and were spun at 18,000 g at 4°C. The supernatant was transferred into fresh clean Lo-Bind tubes (Fisher Scientific Cat #13698794).

Sample preparation for proteomics. We have previously described the enrichment of biotinylated proteins and bead titration assay [91], Briefly, total protein concentration of spinal cord lysates was measured by the DC Protein Assay (Bio-Rad Cat # 5000111). We aliquoted 1 mg proteins per replicate and incubated them with 250 µL pre-washed streptavidin magnetic beads at 4 °C for 12-h to enrich biotinylated proteins. Four sequential wash buffers were used to remove non-specific binding proteins. Wash buffer A: 2% SDS; wash buffer B: 500mM NaCl, 0.1% deoxycolic acid, 1% Triton-X, 1 mM EDTA in 50mM Tris-HCl; wash buffer C:250mM NaCl,

0.5% deoxycolic acid, 0.5% NP-40, 1 mM EDTA in 50mM Tris-HCl; wash buffer D: 2 M Urea in 50mM Tris-HCl. Magnetic beads were then resuspended in 100 µL of wash buffer D followed by protein reduction and alkylation on beads using 5 mM tris(2-chloroethyl) phosphate (1-h) and 15 mM iodoacetamide (30-min), respectively. Then, we transferred the beads to 100 µL 50 mM Tris-HCl buffer with mixed with 1 µg trypsin/Lys-C (Promega Cat# V5073) per replicate. The on-beads tryptic digestion was carried out in a ThermoMixer at 37 °C with 1200 rpm shaking for 16-h. The resulting peptides quenched with 10% trifluoroacetic acid (TFA). Waters Oasis HLB 96-well extraction plate was used for peptide desalting according to the manufacture instruction. The purified tryptic peptides were freeze dried and were reconstituted in 10 µL of 0.1% TFA in water for mass spectrometry-based proteomics analysis.

Quantitative proteomics data processing. The BAR-labeled peptides were subjected to a nano liquid chromatography (LC) for separation and were analyzed on an Orbitrap Eclipse mass spectrometer (MS) coupled with a High Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS) interface using data dependent acquisition (DDA). Briefly, the peptides were separated on a 50cm C_{18} reverse phase nano column for 120-min with linear gradient 2%-35% acetonitrile in 0.1% formic acid. To increase the proteome coverage, we applied three 3 compensation voltages (CVs), -35V, -50V, and -65V, in FAIMS-DDA analysis, and cycle time was set to 1-s per CV. The full scan used 120,000 resolution with 400-1,200 m/z ratio scan window; the fragment scan used higher-energy C-trap dissociation (HCD) in ion trap with 1.6 m/z isolation window. We set dynamic exclusion to 45-s with 10ppm mass tolerance. The acquired MS raw files were subjected to database search using proteome discoverer (v2.4.1.15) software package and Uniprot Human proteome FASTA file (UP000005060) containing 20,379 reviewed high-confident gene entries. The label free quantification was used to quantify and normalize protein abundance based on the total peptide intensity identified in the same sample. The negative controls were database searched separately to minimize the potential match-between-run interference.

Quantitative proteomics analysis. Quantified protein abundances were taken as input to inspect the number of identified proteins and missing values (dropouts) in the dataset. To impute the missing values, we implemented a standard Maximum Likelihood Estimate technique using DEP package (version 1.12.0) in RStudio (version 1.1.463). Using this set of final abundance values, we log-transformed the data to generate approximately normal distributions for protein abundances, and ran paired two-tailed t-tests (where NeuN and STMN2-enriched data for each sample was considered a pair) to assess differential abundance. To identify over-representation of differentially expressed proteins ($p<0.05$, upregulated in the STMN2-enriched samples) with our motoneuron-enriched gene list from the transcriptomic data, we used a standard hypergeometric test, using all proteomics-detected species as the background list.

SOD1 antibody validation in human iPS neurons with targeted knockdown. Previously published human inducible pluripotent stem cells (hiPSCs) were used to knock down SOD1 [92]. A SOD1 or non-targeting control sgRNA was cloned into a mU6-sgRNA EF1a-puro-T2A-2XmycNLS-BFP vector (gift from Martin Kampmann's lab; Addgene #127965). sgRNA sequences are as follows: SOD1: GAGGCACCACGACAGACCCG, non-targeting sgRNA: GAATATGTGCGTGCATGAAG. Lentivirus was produced via transduction of Lenti-X HEK 293T cells using Lipofectamine 3000 in DMEM high glucose GlutaMAX Supplement media containing 10% FBS. 24 hours post-transfection, media was replaced, including ViralBoost

Reagent (ALSTEM, #VB100). 96 hours post-transfection, media was collected and concentrated 1:10 in 1xPBS using Lenti-X concentrator (Takara Bio, #631231), aliquoted, and stored at –80°C. 100 ml of these aliquots was used to transduce 100,000 hiPSCs to generate SOD1 KD and control lines. The cells were split and replated on Matrigel (Corning Incorporated #354277) coated coverslips with the viral concentrate in E8+Y-27632 ROCK Inhibitor and allowed to incubate for 24 hours at 37°C, 5% CO_2. The media was replaced with E8 and the cells were allowed to grow for another 24 hours before fixation with 4% PFA in PBS for 10 mins at room temperature. Cells were washed with PBS 3 times and permeabilized in block (PBS + 3% donkey serum + 0.1% tritonX) for 30 mins at room temperature. Primary antibody targeting SOD1 (Sigma, HPA001401-100UL) was diluted at 1:500 in block and cells were incubated in primary overnight at 4°C on a rocker. The next day, cells were washed three times with PBST and incubated in block with secondary antibody (Jackson ImmunoResearch # 711-625-152) and Hoechst (Thermo Scientific #62249) for 1 hour at room temperature. Following 3 washes with PBST, the coverslips were mounted using ProLong Gold antifade reagent (Invitrogen #P36934). After curing, the coverslips were imaged using Nikon spinning disk confocal using laser wavelengths of 405 nm, 488 nm, 561 nm, and 640 nm at 100ms exposure and 75%, 25%, 25% and 100% power respectively. Images were edited using ImageJ.

Immunohistochemistry antibodies. SOD1 (Sigma, HPA001401-100UL), OPTN (Proteintech, 10837-1-AP), NeuN (EMD Millipore, ABN90), Neurofilament H (Cell Signaling, 2836S), Chat (Millpore Sigma, AB144P), TUBA4A (Thermofiser, PA5-29546), Alexa Fluor® 647 Anti-alpha Tubulin (Abcam, ab190573), Stathmin-2/STMN2 (Novus, NBP1-49461), and Peripherin/PRPH (Millipore, AB1530).

Immunohistochemistry. Immunohistochemistry for human and mouse spinal cords were performed as previously described [93] with modifications for human spinal cords. Briefly, mouse spinal cords were cut at 50 µm and blocking buffer (1% IgG-free BSA, 10% normal donkey serum, 0.1% Triton-X 100 in PBS) for one hour, prior to incubation in blocking buffer and primary antibody for 48 hours at 4°C. Primary antibody was washed off three times in PBS before a 2-hour incubation in secondary antibody at room temperature. Secondary antibody was washed off three times in PBS before adding a coverslip.

Human spinal cords were cut at 14 µm, washed twice in TBS and placed in 0.05% sodium azide-TBS at 4°C for 3 days under a LED light to quench autofluorescence. Human spinal cords were placed in blocking buffer (1% IgG-free BSA, 10% normal donkey serum, in TBS) for one hour prior to incubation in blocking buffer and primary antibody for 48 hours at 4°C. Primary antibody was washed off three times in TBS with 0.025% triton before a 2-hour incubation in secondary antibody at room temperature. Secondary antibody was washed off three times in TBS with 0.025% triton before adding a coverslip.

Imaging. Images of immunohistochemistry samples were imaged using a Zeiss 800 LSM confocal microscope.

Image analysis and quantification. The images were overlaid in Adobe Photoshop where borders between the gray and white matter and the lamina within the gray matter were drawn. These images were then exported to ImageJ for analysis. The cells were measured manually by outlining each cell using the selection tool and adding them to groups within the ROIManager in ImageJ based on lamina. Feret diameter measurements of all the ROIs for each section were

saved in a spreadsheet. The white and gray matter of each subject were outlined in ImageJ and their areas were exported to a spreadsheet.

To identify colocalization of markers with NeuN, each neuron was first outlined with the selection tool in ImageJ and saved into different groups based on whether the cell was in lamina IX or not. Then, each cell that had co-occurrence of the markers were placed into separate groups (double positive in lamina IX and double positive outside lamina IX). Feret diameter measurements were then saved to a spreadsheet and the number of cells in each group were counted in Python.

Data and code availability: Upon publication, anonymized raw sequencing data and counts tables will be deposited in the Gene Expression Omnibus (GEO) and the raw mass spectrometry datasets will be deposited with Synapse.org. In addition, visualization of expression data at the cluster and donor level are available through a searchable web resource at https://vmenon.shinyapps.io/hsc_biorxiv/.

Supplementary Materials

Figures. S1 to S20

Acknowledgments: We gratefully acknowledge the gift of human tissue from the 10 donors included in this work and their families, whose contribution has been critical for this work. In addition, we thank Dr. Hemali Phatnani for her advice throughout this project and for comments on the manuscript and Joana Petrescu and Dr. Andy Qi for assistance with protein validation.

Funding:
This work was supported by NINDS Intramural funds through 1ZIA NS003153 (KJEM, LL, IH, AJL) and 1ZIANS003155 (SH and MEW); by NICHD Intramural funds through 1ZIAHD008966 (MRA, CELP); by the Intramural Research Program of the NIA project ZO1 AG000535 (MAN); by R01 AG06683 from NIA (AY, VM); by the Canadian Institutes of Health Research, the University of Ottawa Department of Surgery, the Ontario Neurotrauma Foundation, and the Praxis Spinal Cord Institute (AG, SA, JP, MMA, FAQ, SMA, APW, ECT, MEH); and by ANR-15-CE16-012, FRC-EET-2019 grants and CNRS/INSERM/Montpellier Hospital research supports (PFM, EB, LB).

Author contributions:
Conceptualization: V.M., A.J.L.
Methodology: A.Y., K.J.E.M., L.L., I.H., M.R.A., A.G., A.D., S.A., J.P., M.M.A., F.A.Q., S.M.A., A.P.W., G.P., N.L., F.V.L., M.A.N., S.H., M.E.H., E.B., P.F.M., L.B., E.C.T., M.E.W., C.E.L.P., V.M., A.J.L.
Investigation: A.Y., K.J.E.M., L.L., I.H., Y.A.Q., M.R.A., S.H., M.A.N., M.E.W., C.E.L.P., V.M., A.J.L.
Formal Analysis: A.Y., K.J.E.M., L.L., I.H., P.G., M.R.A., S.H., M.A.N., M.E.W., C.E.L.P., V.M., A.J.L.
Visualization: A.Y., K.J.E.M., I.H., M.R.A., V.M., A.J.L.
Writing: A.Y., K.J.E.M., M.E.W., V.M., A.J.L.
Supervision: M.E.H., E.B., P.F.M., L.B., E.C.T., M.E.W., C.E.L.P., V.M., A.J.L.

Funding acquisition: M.E.H., M.A.N., E.B., P.F.M., L.B., E.C.T., M.E.W., C.E.L.P., V.M., A.J.L.

Competing interests: M.A.N.'s participation in this project was part of a competitive contract awarded to Data Tecnica International LLC by the National Institutes of Health to support open science research, he also currently serves on the scientific advisory board for Clover Therapeutics and is an advisor to Neuron23 Inc.

References

1. Blackstone, C. Hereditary spastic paraplegia. *Handb Clin Neurol* **148,** 633–652 (2018).
2. SCHWARZ, G. A. & Liu, C. N. Hereditary (familial) spastic paraplegia; further clinical and pathologic observations. *AMA Arch Neurol Psychiatry* **75,** 144–162 (1956).
3. Bruyn, R. P., van Dijk, J. G., Scheltens, P., Boezeman, E. H. & Ongerboer de Visser, B. W. Clinically silent dysfunction of dorsal columns and dorsal spinocerebellar tracts in hereditary spastic paraparesis. *J Neurol Sci* **125,** 206–211 (1994).
4. Arnold, E. S. & Fischbeck, K. H. Spinal muscular atrophy. *Handb Clin Neurol* **148,** 591–601 (2018).
5. Ravits, J. *et al.* Deciphering amyotrophic lateral sclerosis: what phenotype, neuropathology and genetics are telling us about pathogenesis. *Amyotroph Lateral Scler Frontotemporal Degener* **14 Suppl 1,** 5–18 (2013).
6. Romer, S. H. *et al.* Accessory respiratory muscles enhance ventilation in ALS model mice and are activated by excitatory V2a neurons. *Exp. Neurol.* **287,** 192–204 (2017).
7. Allodi, I., Montañana-Rosell, R., Selvan, R., Löw, P. & Kiehn, O. Locomotor deficits in a mouse model of ALS are paralleled by loss of V1-interneuron connections onto fast motor neurons. *Nat Commun* **12,** 3251–18 (2021).
8. Stephens, B. *et al.* Widespread loss of neuronal populations in the spinal ventral horn in sporadic motor neuron disease. A morphometric study. *J Neurol Sci* **244,** 41–58 (2006).
9. Williams, C., Kozlowski, M. A., Hinton, D. R. & Miller, C. A. Degeneration of spinocerebellar neurons in amyotrophic lateral sclerosis. *Ann Neurol* **27,** 215–225 (1990).
10. Averback, P. & Crocker, P. Regular involvement of Clarke's nucleus in sporadic amyotrophic lateral sclerosis. *Arch Neurol* **39,** 155–156 (1982).
11. Kawamura, Y. *et al.* Morphometric comparison of the vulnerability of peripheral motor and sensory neurons in amyotrophic lateral sclerosis. *J Neuropathol Exp Neurol* **40,** 667–675 (1981).
12. Salamatina, A. *et al.* Differential Loss of Spinal Interneurons in a Mouse Model of ALS. *Neuroscience* **450,** 81–95 (2020).
13. Gurney, M. E. *et al.* Motor neuron degeneration in mice that express a human Cu,Zn superoxide dismutase mutation. *Science* **264,** 1772–1775 (1994).
14. Pardo, C. A. *et al.* Superoxide dismutase is an abundant component in cell bodies, dendrites, and axons of motor neurons and in a subset of other neurons. *Proceedings of the National Academy of Sciences* **92,** 954–958 (1995).
15. Smith, B. N. *et al.* Exome-wide rare variant analysis identifies TUBA4A mutations associated with familial ALS. *Neuron* **84,** 324–331 (2014).
16. Kim, G., Gautier, O., Tassoni-Tsuchida, E., Ma, X. R. & Gitler, A. D. ALS Genetics: Gains, Losses, and Implications for Future Therapies. *Neuron* **108,** 822–842 (2020).

17. Ilieva, H., Polymenidou, M. & Cleveland, D. W. Non-cell autonomous toxicity in neurodegenerative disorders: ALS and beyond. *J. Cell Biol.* **187,** 761–772 (2009).
18. Taylor, J. P., Brown, R. H. & Cleveland, D. W. Decoding ALS: from genes to mechanism. *Nature* **539,** 197–206 (2016).
19. Nijssen, J., Comley, L. H. & Hedlund, E. Motor neuron vulnerability and resistance in amyotrophic lateral sclerosis. *Acta Neuropathol* **133,** 863–885 (2017).
20. Sreedharan, J. & Brown, R. H. Amyotrophic lateral sclerosis: Problems and prospects. *Ann Neurol* **74,** 309–316 (2013).
21. Comley, L. *et al.* Motor neurons with differential vulnerability to degeneration show distinct protein signatures in health and ALS. *Neuroscience* **291,** 216–229 (2015).
22. Brockington, A. *et al.* Unravelling the enigma of selective vulnerability in neurodegeneration: motor neurons resistant to degeneration in ALS show distinct gene expression characteristics and decreased susceptibility to excitotoxicity. *Acta Neuropathol* **125,** 95–109 (2013).
23. Kaplan, A. *et al.* Neuronal matrix metalloproteinase-9 is a determinant of selective neurodegeneration. *Neuron* **81,** 333–348 (2014).
24. Allodi, I. *et al.* Modeling Motor Neuron Resilience in ALS Using Stem Cells. *Stem Cell Reports* **12,** 1329–1341 (2019).
25. Rayon, T., Maizels, R. J., Barrington, C. & Briscoe, J. Single-cell transcriptome profiling of the human developing spinal cord reveals a conserved genetic programme with human-specific features. *Development* **148,** (2021).
26. Zhang, Q. *et al.* Single-cell analysis reveals dynamic changes of neural cells in developing human spinal cord. *EMBO Rep* **22,** e52728 (2021).
27. Zhang, D. *et al.* Single-nucleus transcriptomic atlas of spinal cord neuron in human. *bioRxiv* 2021.09.28.462103 (2021). doi:10.1101/2021.09.28.462103
28. Russ, D. E. *et al.* A harmonized atlas of mouse spinal cord cell types and their spatial organization. *Nat Commun* **12,** 5722–20 (2021).
29. Xu, Z., Cork, L. C., Griffin, J. W. & Cleveland, D. W. Involvement of neurofilaments in motor neuron disease. *J Cell Sci Suppl* **17,** 101–108 (1993).
30. Xu, Z., Cork, L. C., Griffin, J. W. & Cleveland, D. W. Increased expression of neurofilament subunit NF-L produces morphological alterations that resemble the pathology of human motor neuron disease. *Cell* **73,** 23–33 (1993).
31. Gama Sosa, M. A. *et al.* Human midsized neurofilament subunit induces motor neuron disease in transgenic mice. *Exp. Neurol.* **184,** 408–419 (2003).
32. Côté, F., Collard, J. F. & Julien, J. P. Progressive neuronopathy in transgenic mice expressing the human neurofilament heavy gene: a mouse model of amyotrophic lateral sclerosis. *Cell* **73,** 35–46 (1993).
33. Marszalek, J. R. *et al.* Neurofilament subunit NF-H modulates axonal diameter by selectively slowing neurofilament transport. *J. Cell Biol.* **135,** 711–724 (1996).
34. Beaulieu, J. M., Nguyen, M. D. & Julien, J. P. Late onset of motor neurons in mice overexpressing wild-type peripherin. *J. Cell Biol.* **147,** 531–544 (1999).
35. Routal, R. V. & Pal, G. P. A study of motoneuron groups and motor columns of the human spinal cord. *J. Anat.* **195 (Pt 2),** 211–224 (1999).
36. Schoenen, J. Clinical anatomy of the spinal cord. *Neurol Clin* **9,** 503–532 (1991).
37. Watson, C., Paxinos, G. & Kayalioglu, G. *The Spinal Cord*. (Academic Press, 2009).
38. The Human Nervous System. (2004).
39. Gregory, J. M., Fagegaltier, D., Phatnani, H. & Harms, M. B. Genetics of Amyotrophic Lateral Sclerosis. *Curr Genet Med Rep* **8,** 121–131 (2020).

40. Brown, R. H. & Al-Chalabi, A. Amyotrophic Lateral Sclerosis. *N Engl J Med* **377,** 1602 (2017).
41. Castellanos-Montiel, M. J., Chaineau, M. & Durcan, T. M. The Neglected Genes of ALS: Cytoskeletal Dynamics Impact Synaptic Degeneration in ALS. *Front Cell Neurosci* **14,** 594975 (2020).
42. Theunissen, F. *et al.* Novel STMN2 Variant Linked to Amyotrophic Lateral Sclerosis Risk and Clinical Phenotype. *Front Aging Neurosci* **13,** 658226 (2021).
43. Klim, J. R. *et al.* ALS-implicated protein TDP-43 sustains levels of STMN2, a mediator of motor neuron growth and repair. *Nat Neurosci* **22,** 167–179 (2019).
44. Morisaki, Y. *et al.* Selective Expression of Osteopontin in ALS-resistant Motor Neurons is a Critical Determinant of Late Phase Neurodegeneration Mediated by Matrix Metalloproteinase-9. *Sci Rep* **6,** 27354–19 (2016).
45. Yamamoto, T., Murayama, S., Takao, M., Isa, T. & Higo, N. Expression of secreted phosphoprotein 1 (osteopontin) in human sensorimotor cortex and spinal cord: Changes in patients with amyotrophic lateral sclerosis. *Brain Res.* **1655,** 168–175 (2017).
46. Alkaslasi, M. R. *et al.* Single nucleus RNA-sequencing defines unexpected diversity of cholinergic neuron types in the adult mouse spinal cord. *Nat Commun* **12,** 2471–14 (2021).
47. Pembroke, W. G., Hartl, C. L. & Geschwind, D. H. Evolutionary conservation and divergence of the human brain transcriptome. *Genome Biol.* **22,** 52–33 (2021).
48. McHanwell, S. & Biscoe, T. J. The sizes of motoneurons supplying hindlimb muscles in the mouse. *Proc R Soc Lond B Biol Sci* **213,** 201–216 (1981).
49. Ishihara, A. *et al.* Cell body size and succinate dehydrogenase activity of spinal motoneurons innervating the soleus muscle in mice, rats, and cats. *Neurochem Res* **26,** 1301–1304 (2001).
50. Kawamura, Y. & Dyck, P. J. Lumbar motoneurons of man: III. The number and diameter distribution of large- and intermediate- diameter cytons by nuclear columns. *J Neuropathol Exp Neurol* **36,** 956–963 (1977).
51. Sobue, G. *et al.* Spinal and cranial motor nerve roots in amyotrophic lateral sclerosis and X-linked recessive bulbospinal muscular atrophy: morphometric and teased-fiber study. *Acta Neuropathol* **55,** 227–235 (1981).
52. Sobue, G., Matsuoka, Y., Mukai, E., Takayanagi, T. & Sobue, I. Pathology of myelinated fibers in cervical and lumbar ventral spinal roots in amyotrophic lateral sclerosis. *J Neurol Sci* **50,** 413–421 (1981).
53. Nguyen, M. D., Larivière, R. C. & Julien, J. P. Reduction of axonal caliber does not alleviate motor neuron disease caused by mutant superoxide dismutase 1. *Proceedings of the National Academy of Sciences* **97,** 12306–12311 (2000).
54. Hardesty, I. *Observations on the medulla spinalis of the elephant with some comparative studies of the intumescentia cervicalis and the neurones of the columna anterior*. (1902).
55. Braak, H. & Braak, E. The pyramidal cells of Betz within the cingulate and precentral gigantopyramidal field in the human brain. A Golgi and pigmentarchitectonic study. *Cell Tissue Res.* **172,** 103–119 (1976).
56. Haberberger, R. V., Barry, C., Dominguez, N. & Matusica, D. Human Dorsal Root Ganglia. *Front Cell Neurosci* **13,** 271 (2019).
57. Zottoli, S. J. Comparison of Mauthner cell size in teleosts. *The Journal of Comparative Neurology* **178,** 741–757 (1978).
58. Bar, D. Z. *et al.* Biotinylation by antibody recognition-a method for proximity labeling. *Nat Meth* **15,** 127–133 (2018).

59. Tsang, Y. M., Chiong, F., Kuznetsov, D., Kasarskis, E. & Geula, C. Motor neurons are rich in non-phosphorylated neurofilaments: cross-species comparison and alterations in ALS. *Brain Res.* **861,** 45–58 (2000).
60. McIlwain, D. L. Nuclear and cell body size in spinal motor neurons. *Advances in neurology* **56,** 67–74 (1991).
61. Clark, J. A., Yeaman, E. J., Blizzard, C. A., Chuckowree, J. A. & Dickson, T. C. A Case for Microtubule Vulnerability in Amyotrophic Lateral Sclerosis: Altered Dynamics During Disease. *Front Cell Neurosci* **10,** 204 (2016).
62. Lee, M. K. & Cleveland, D. W. Neuronal intermediate filaments. *Annu. Rev. Neurosci.* **19,** 187–217 (1996).
63. Hoffman, P. N., Griffin, J. W. & Price, D. L. Control of axonal caliber by neurofilament transport. *J. Cell Biol.* **99,** 705–714 (1984).
64. Friede, R. L. & Samorajski, T. Axon caliber related to neurofilaments and microtubules in sciatic nerve fibers of rats and mice. *Anat. Rec.* **167,** 379–387 (1970).
65. Bakken, T. E. *et al.* Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **598,** 111–119 (2021).
66. Zeisel, A. *et al.* Molecular Architecture of the Mouse Nervous System. *Cell* **174,** 999–1014.e22 (2018).
67. Nguyen, M. Q., Buchholtz, von, L. J., Reker, A. N., Ryba, N. J. & Davidson, S. Single-nucleus transcriptomic analysis of human dorsal root ganglion neurons. (2021). doi:10.7554/eLife.71752
68. Limone, F. *et al.* Single-nucleus sequencing reveals enriched expression of genetic risk factors sensitises Motor Neurons to degeneration in ALS. *bioRxiv* 2021.07.12.452054 (2021). doi:10.1101/2021.07.12.452054
69. Shin, J. E. *et al.* SCG10 is a JNK target in the axonal degeneration pathway. *Proc. Natl. Acad. Sci. U.S.A.* **109,** E3696–705 (2012).
70. Melamed, Z. *et al.* Premature polyadenylation-mediated loss of stathmin-2 is a hallmark of TDP-43-dependent neurodegeneration. *Nat Neurosci* **22,** 180–190 (2019).
71. Chen, J. *et al.* α-tubulin tail modifications regulate microtubule stability through selective effector recruitment, not changes in intrinsic polymer dynamics. *Dev Cell* **56,** 2016–2028.e4 (2021).
72. Schoenen, J. Dendritic organization of the human spinal cord: the motoneurons. *The Journal of Comparative Neurology* **211,** 226–247 (1982).
73. Manuel, M., Chardon, M., Tysseling, V. & Heckman, C. J. Scaling of Motor Output, From Mouse to Humans. *Physiology (Bethesda)* **34,** 5–13 (2019).
74. Perge, J. A., Niven, J. E., Mugnaini, E., Balasubramanian, V. & Sterling, P. Why do axons differ in caliber? *J. Neurosci.* **32,** 626–638 (2012).
75. Ragagnin, A. M. G., Shadfar, S., Vidal, M., Jamali, M. S. & Atkin, J. D. Motor Neuron Susceptibility in ALS/FTD. *Front Neurosci* **13,** 532 (2019).
76. Appel, S. H., Beers, D. R. & Zhao, W. Amyotrophic lateral sclerosis is a systemic disease: peripheral contributions to inflammation-mediated neurodegeneration. *Curr Opin Neurol* **34,** 765–772 (2021).
77. Van Harten, A. C. M., Phatnani, H. & Przedborski, S. Non-cell-autonomous pathogenic mechanisms in amyotrophic lateral sclerosis. *Trends Neurosci.* **44,** 658–668 (2021).
78. Vahsen, B. F. *et al.* Non-neuronal cells in amyotrophic lateral sclerosis - from pathogenesis to biomarkers. *Nat Rev Neurol* **17,** 333–348 (2021).
79. Boillée, S. *et al.* Onset and progression in inherited ALS determined by motor neurons and microglia. *Science* **312,** 1389–1392 (2006).

80. Spiller, K. J. *et al.* Microglia-mediated recovery from ALS-relevant motor neuron degeneration in a mouse model of TDP-43 proteinopathy. *Nat Neurosci* **21,** 329–340 (2018).
81. Xie, M. *et al.* TREM2 interacts with TDP-43 and mediates microglial neuroprotection against TDP-43-related neurodegeneration. *Nat Neurosci* **25,** 26–38 (2022).
82. Lall, D. *et al.* C9orf72 deficiency promotes microglial-mediated synaptic loss in aging and amyloid accumulation. *Neuron* **109,** 2275–2291.e8 (2021).
83. Galuta, A. *et al.* A Guide to Extract Spinal Cord for Translational Stem Cell Biology Research: Comparative Analysis of Adult Human, Porcine, and Rodent Spinal Cord Stem Cells. *Front Neurosci* **14,** 607 (2020).
84. Bauchet, L. *et al.* Isolation and Culture of Precursor Cells from the Adult Human Spinal Cord. *Methods Mol. Biol.* **2389,** 103–110 (2022).
85. Dedek, A. *et al.* Loss of STEP61 couples disinhibition to N-methyl-d-aspartate receptor potentiation in rodent and human spinal pain processing. *Brain* **142,** 1535–1546 (2019).
86. Matson, K. J. E. *et al.* Isolation of Adult Spinal Cord Nuclei for Massively Parallel Single-nucleus RNA Sequencing. *J Vis Exp* e58413–e58413 (2018). doi:10.3791/58413
87. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20,** 296–15 (2019).
88. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* (2019). doi:10.1016/j.cell.2019.05.031
89. Durinck, S. *et al.* BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21,** 3439–3440 (2005).
90. Ramos, D. M., Skarnes, W. C., Singleton, A. B., Cookson, M. R. & Ward, M. E. Tackling neurodegenerative diseases with genomic engineering: A new stem cell initiative from the NIH. *Neuron* **109,** 1080–1083 (2021).
91. Frankenfield, A. M., Fernandopulle, M. S., Hasan, S., Ward, M. E. & Hao, L. Development and Comparative Evaluation of Endolysosomal Proximity Labeling-Based Proteomic Methods in Human iPSC-Derived Neurons. *Anal Chem* **92,** 15437–15444 (2020).
92. Tian, R. *et al.* CRISPR Interference-Based Platform for Multimodal Genetic Screens in Human iPSC-Derived Neurons. *Neuron* **104,** 239–255.e12 (2019).
93. Sathyamurthy, A. *et al.* Massively Parallel Single Nucleus Transcriptional Profiling Defines Spinal Cord Neurons and Their Activity during Behavior. *Cell Rep* **22,** 2216–2225 (2018).

Supplementary Materials for

The Human Motoneuron Expression Signature is Defined by ALS-Related Genes

Authors: Archana Yadav†, Kaya J.E. Matson†, Li Li, Isabelle Hua, Pallavi Gaur, Mor R. Alkaslasi, Saadia Hasan, Ahmad Galuta, Annemarie Dedek, Sara Ameri, Jessica Parnell, Mohammad M. Alshardan, Feras Abbas Qumqumji, Saud M. Alhamad, Alick Pingbei Wang, Gaetan Poulen, Nicolas Lonjon, Florence Vachiery-Lahaye, Mike A. Nalls, Andy Qi1, Michael E. Hildebrand, Pierre-Francois Mery, Emmanuel Bourinet, Luc Bauchet, Eve C. Tsai, Michael E. Ward, Claire E. Le Pichon, Vilas Menon*, Ariel J. Levine*

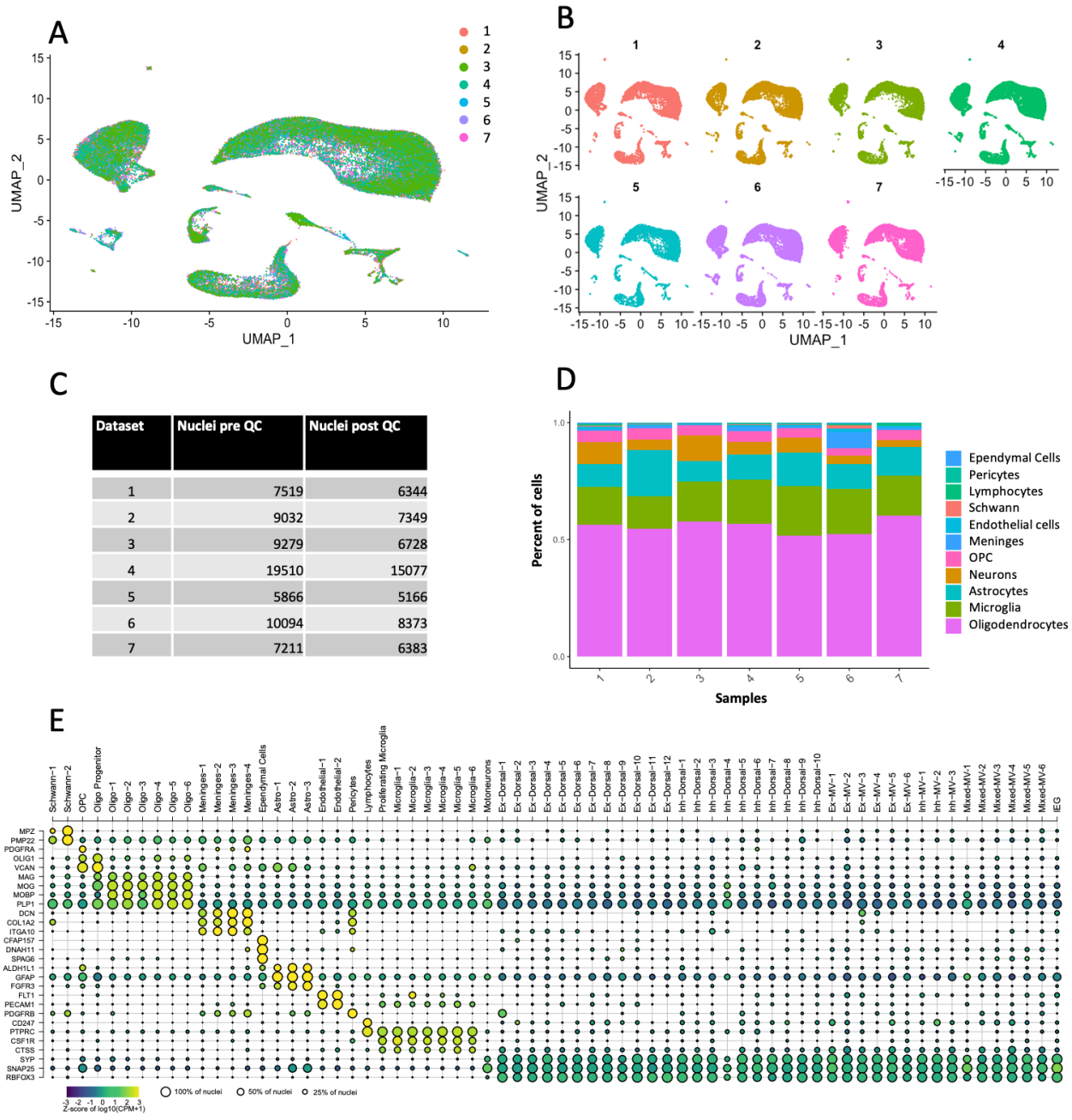
† These authors contributed equally to this work.
*Correspondence to: Email: vm2545@cumc.columbia.edu and ariel.levine@nih.gov

This PDF file includes:

Supplementary Figs. S1 to S20

Other Supplementary Materials for this manuscript include the following:

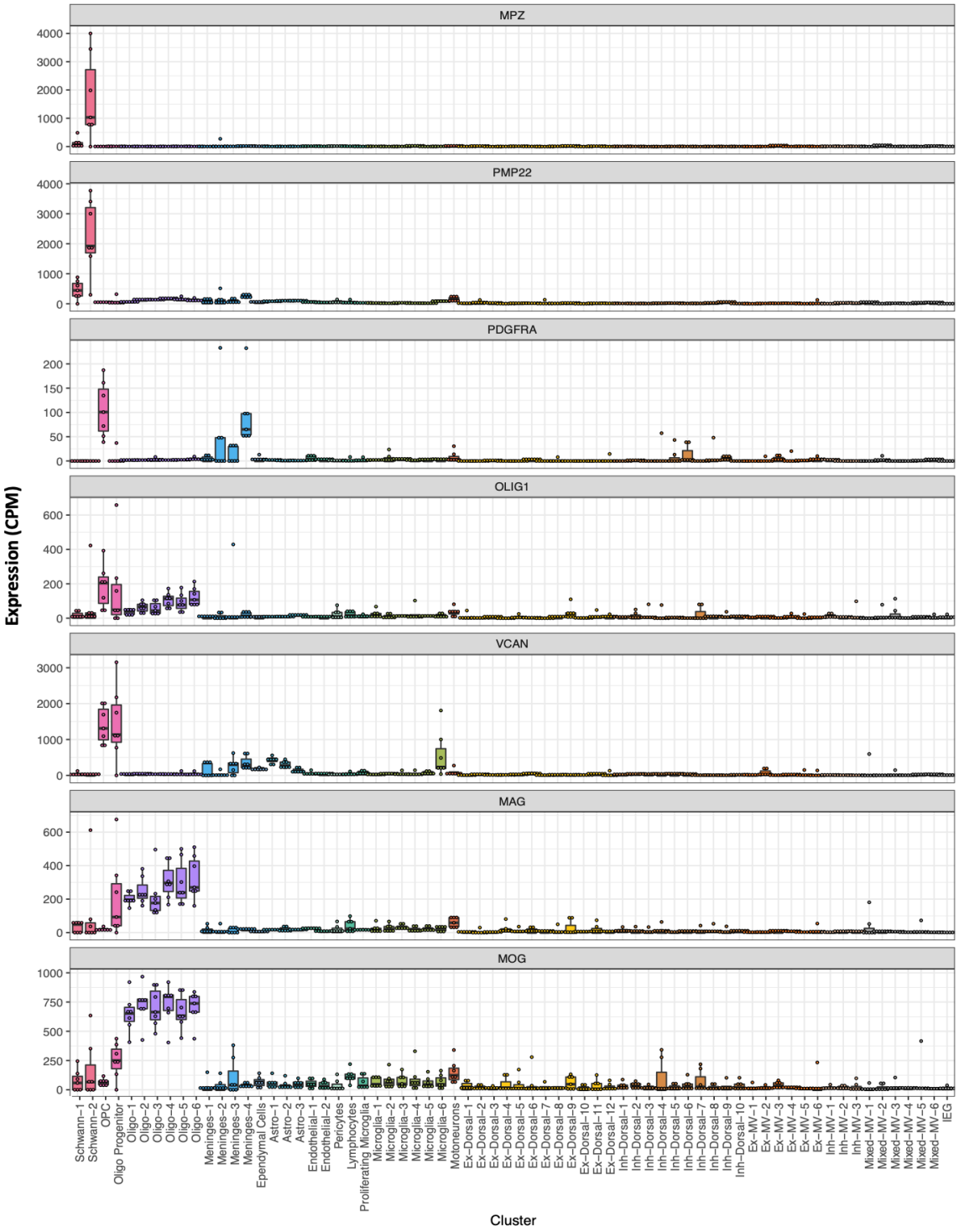
Data File Tables S1 to S8



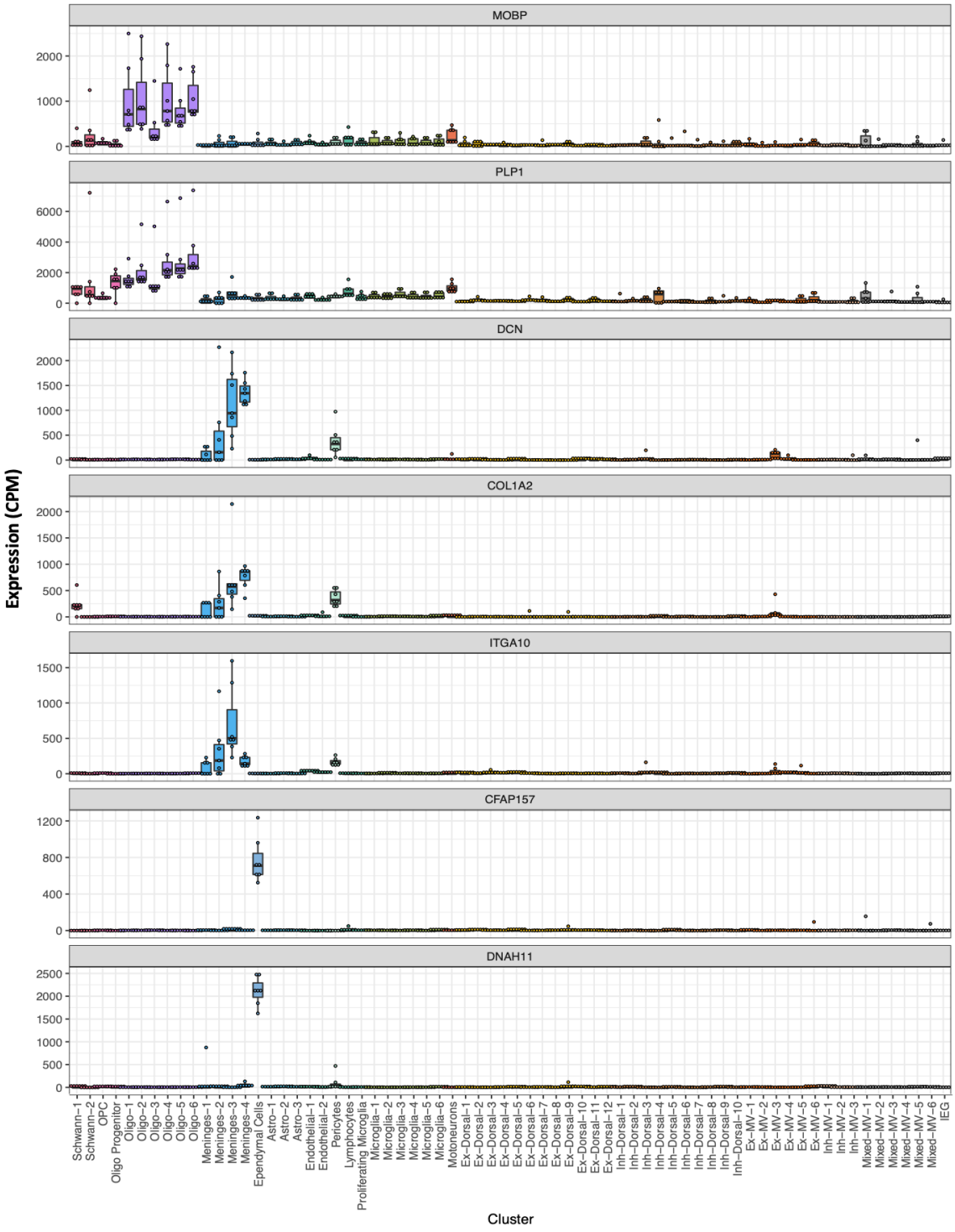
Dataset	Nuclei pre QC	Nuclei post QC
1	7519	6344
2	9032	7349
3	9279	6728
4	19510	15077
5	5866	5166
6	10094	8373
7	7211	6383

Supplementary Fig. S1. Integration at the top level and identification of major cell types. (A) UMAP representation of the 55,420 nuclei after integration of the 7 human datasets. (B) UMAP from panel A split by datasets to depict the overlap between datasets. (C) Number of nuclei before and after quality check analysis (includes removal of doublets, low quality and other cell-type based contaminations). The number of nuclei in panel A are equal to total post QC. (D) Proportion of nuclei assigned as a particular cell type per dataset, (E) Expression of 28 canonical marker genes in all the major cell types and their subclusters (also depicted as UMAP representation in figure 2B in main manuscript. Microglia-1-5, Meninges-1-4, Endothelial-1-2

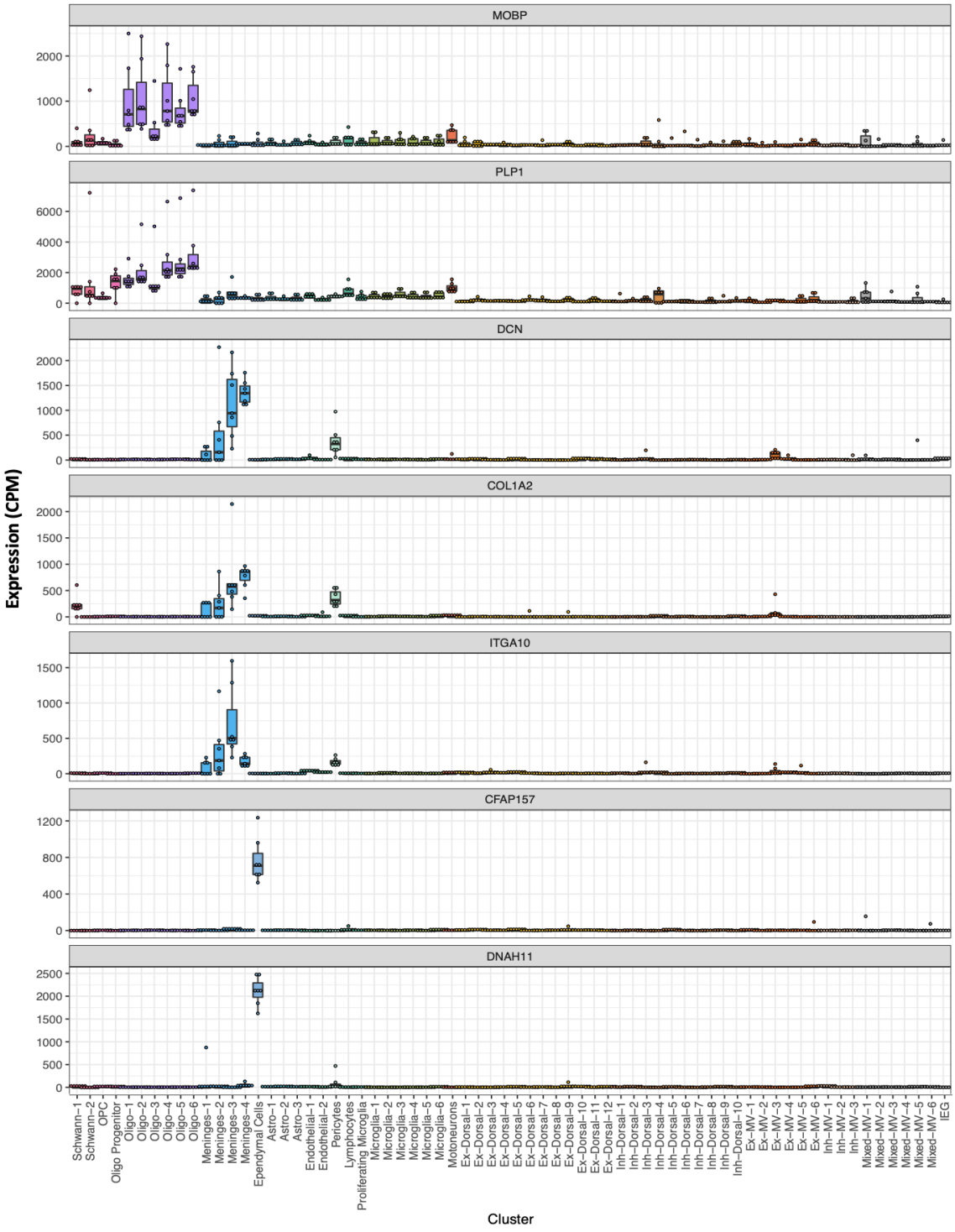
corresponds to Micro-1-5, Men-1-4, Endo1-2, respectively; in Fig1B). Ex- Excitatory, Inh-Inhibitory, MV- Mixed ventral.



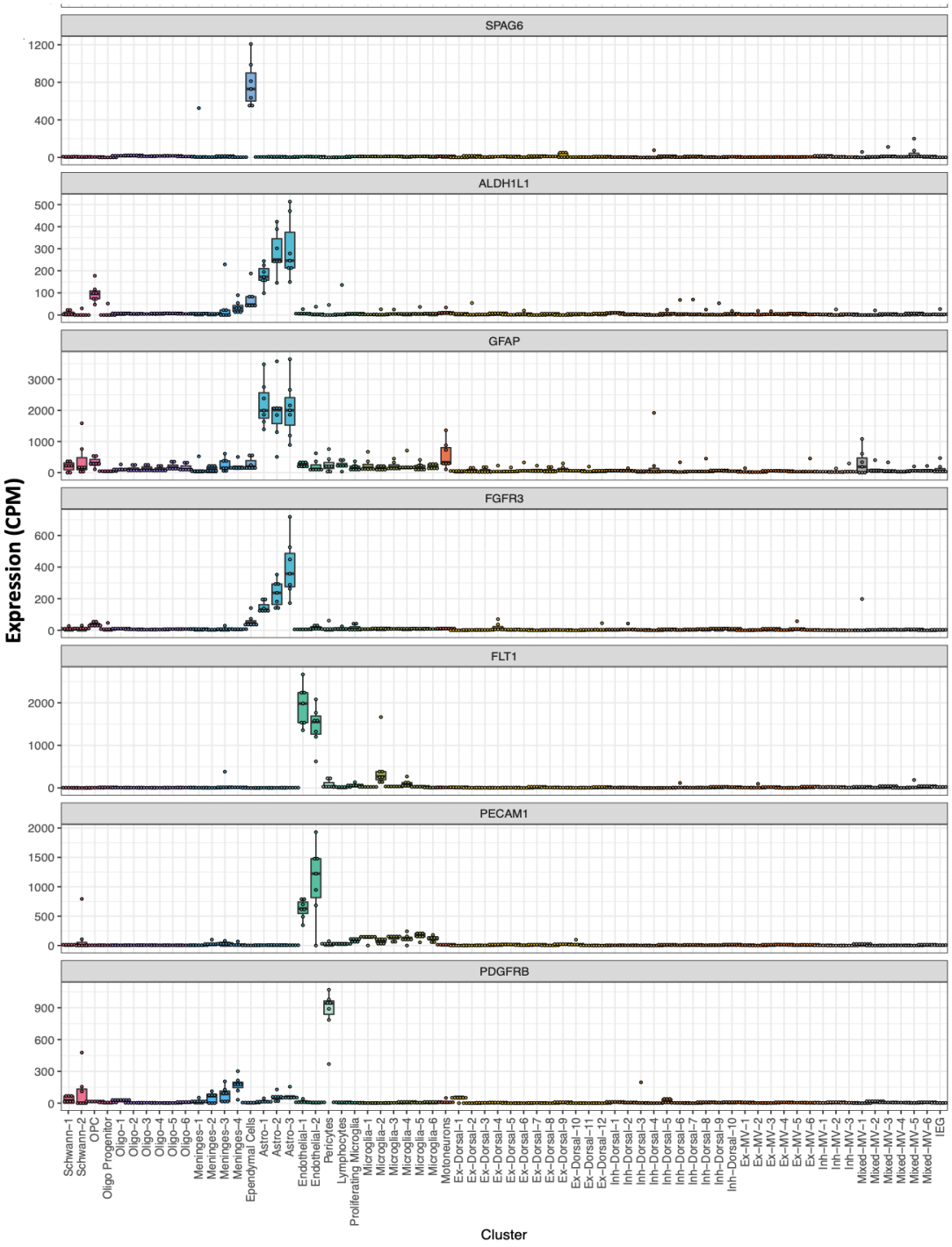
Supplementary Fig. S2. Expression of marker genes in all human spinal cord cell. Box plot representation of per-cluster and per-sample expression (Counts per Million) of top-level marker genes in all cell types. Ex- Excitatory, Inh- Inhibitory, MV- Mixed ventral.



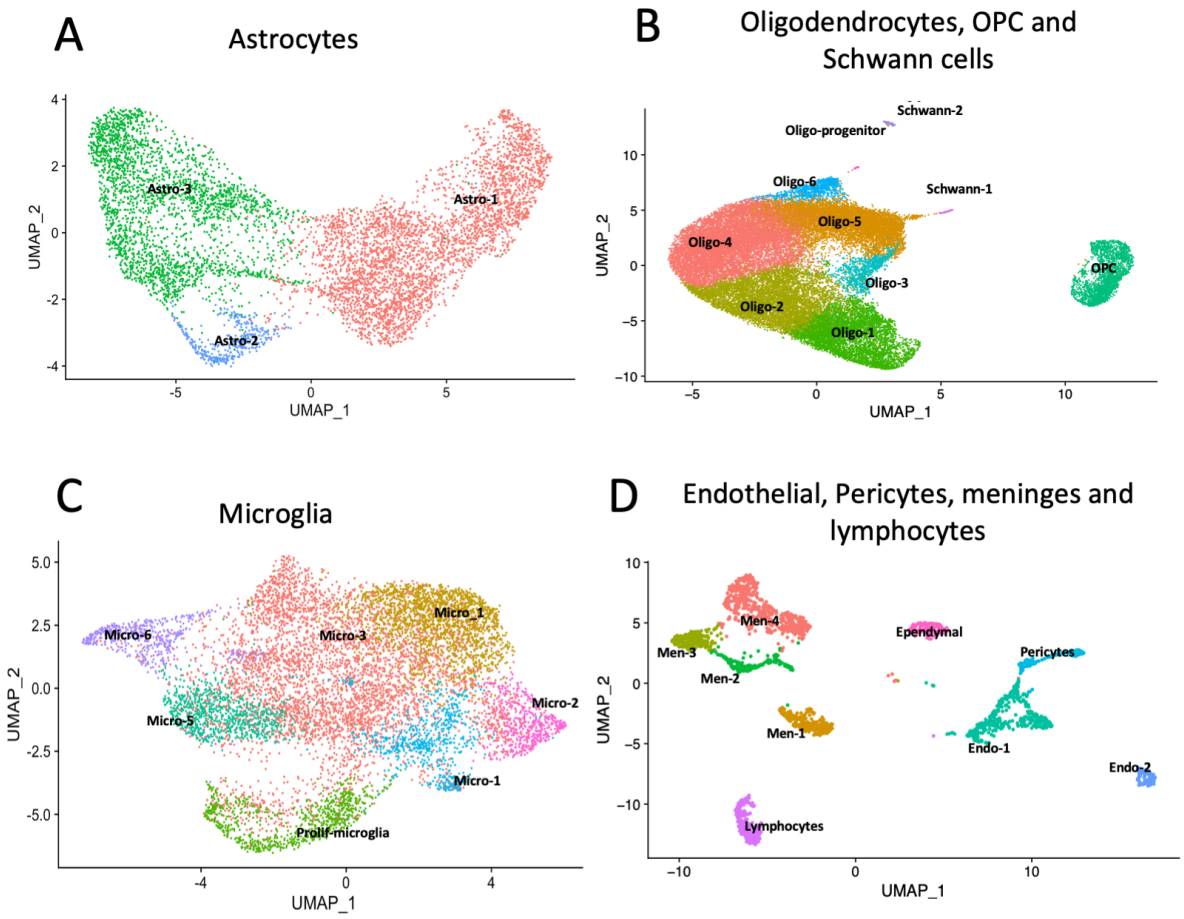
Supplementary Fig. S3. Expression of marker genes in all human spinal cord cell. Box plot representation of per-cluster and per-sample expression (Counts per Million) of top-level marker genes in all cell types.



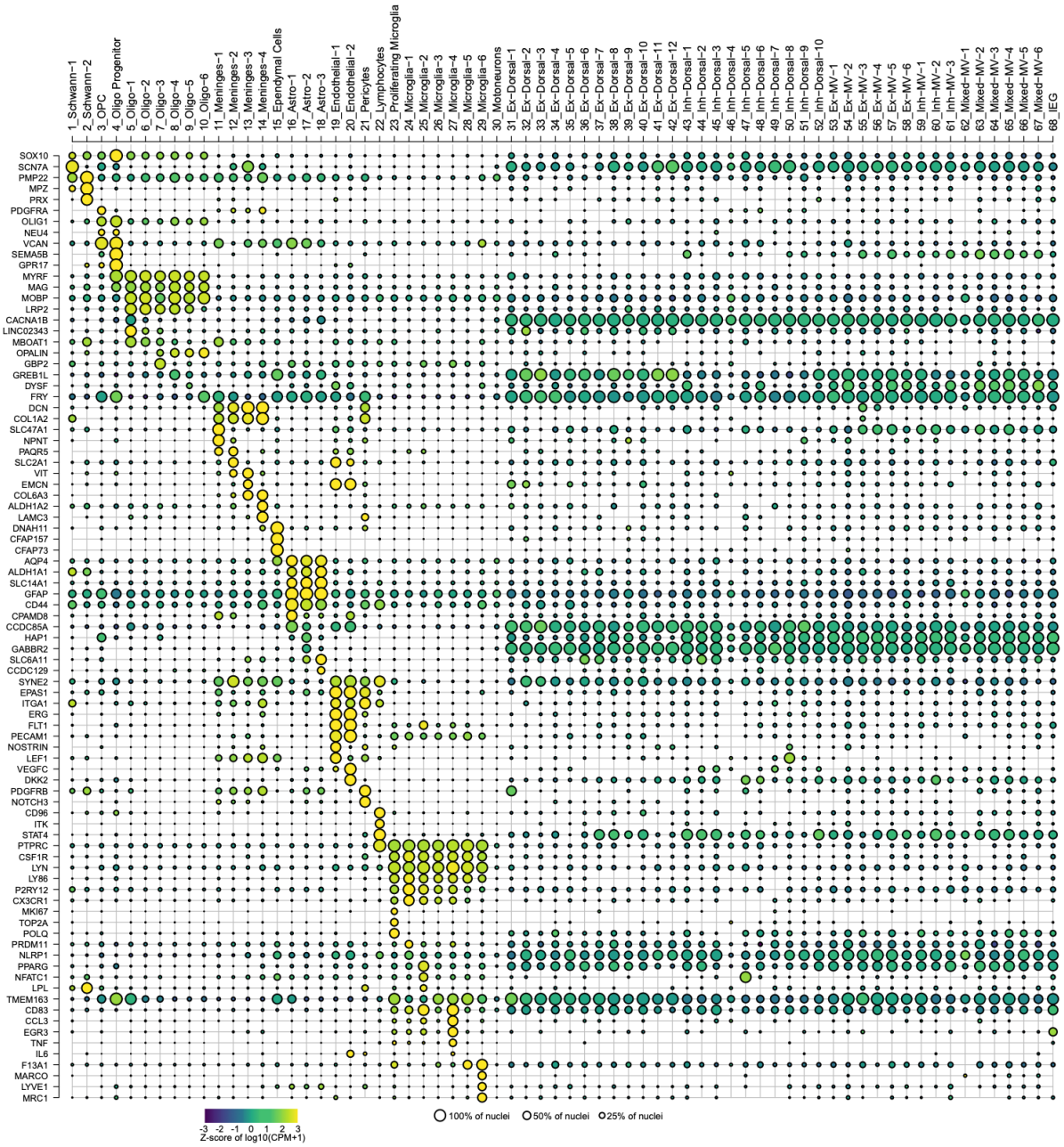
Supplementary Fig. S4. Expression of marker genes in all human spinal cord cell. Box plot representation of per-cluster and per-sample expression (Counts per Million) of top-level marker genes in all cell types.



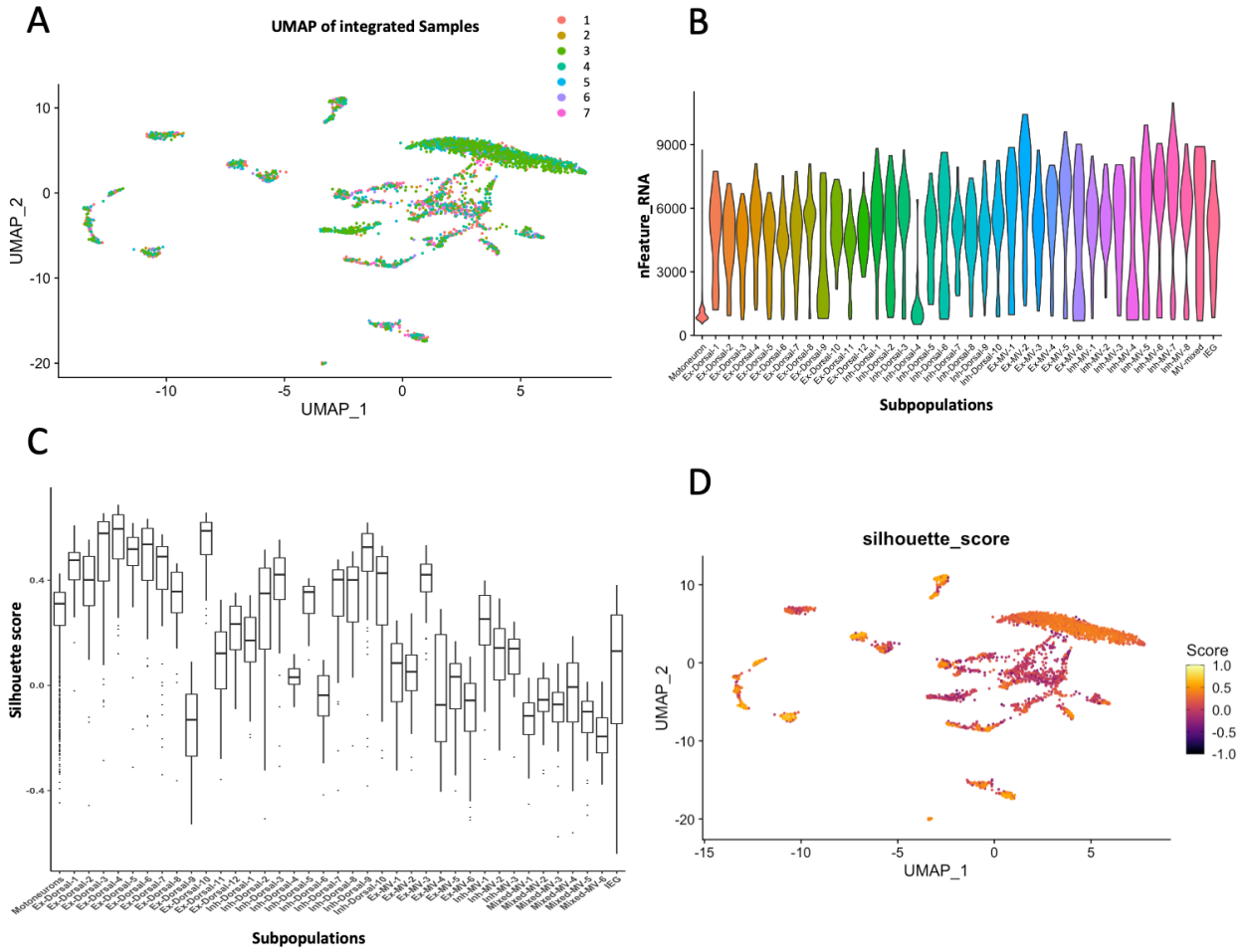
Supplementary Fig. S5. Expression of marker genes in all human spinal cord cell. Box plot representation of per-cluster and per-sample expression (Counts per Million) of top-level marker genes in all cell types.



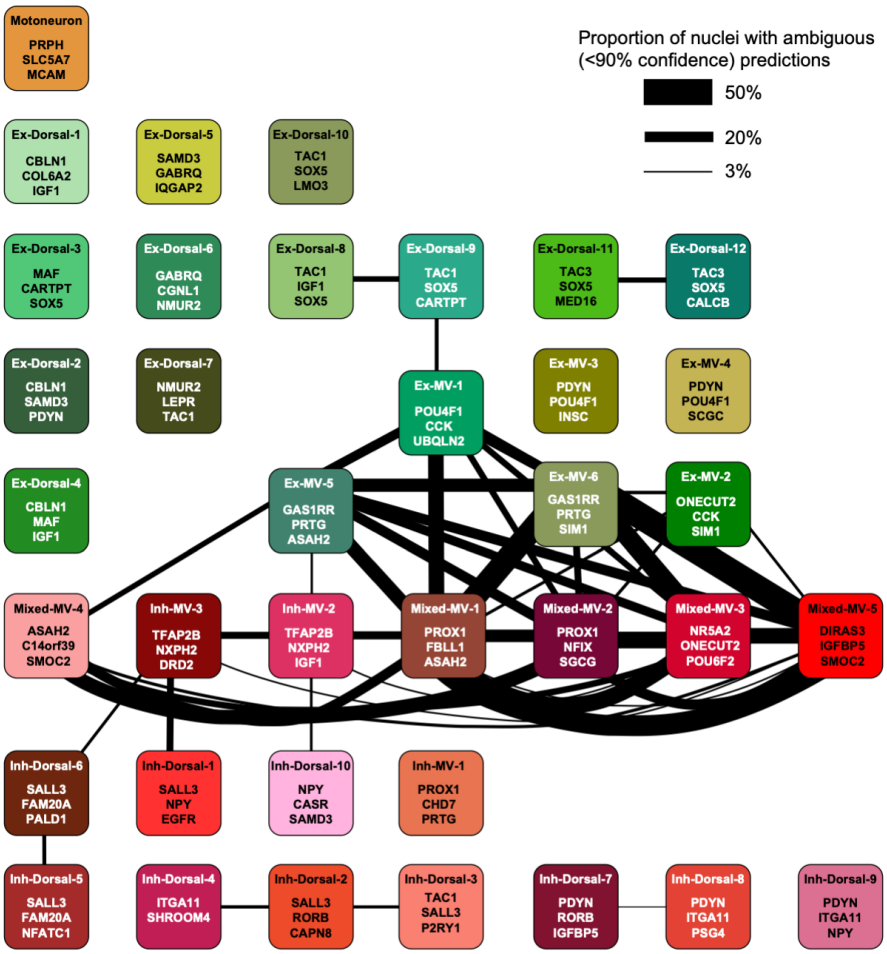
Supplementary Fig. S6. UMAP representation of various subclusters observed in Glial cell types. (A) Astrocytes (B) Oligodendrocytes, OPC and Schwann cells (C) Microglia (D) Endothelial, Pericytes, meninges and lymphocytes. The annotations match the ones used in Figure 1B.



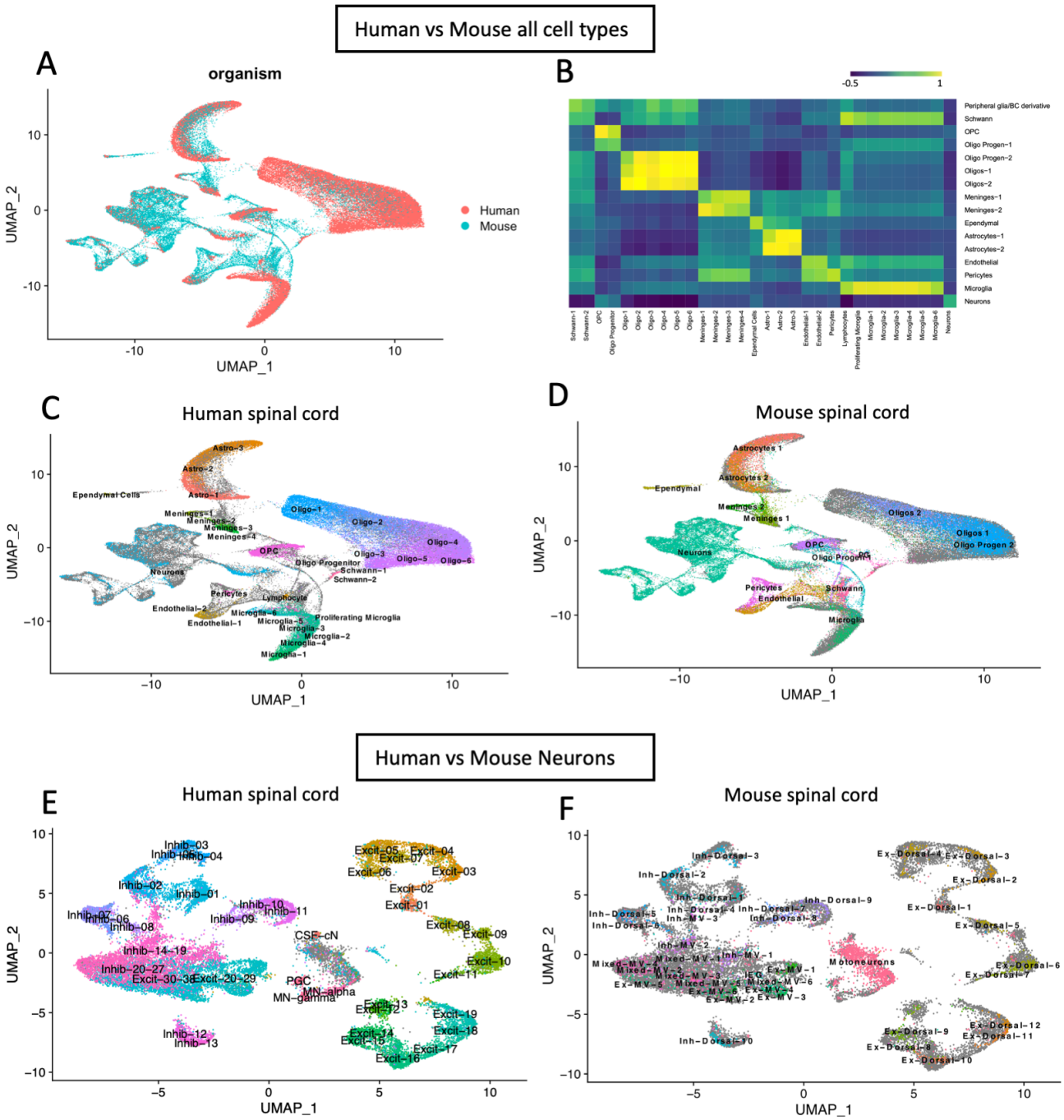
Supplementary Fig. S7. Dot plot depicting expression of various glial marker genes in all spinal cord cell types (total 68). The size of each circle represents the fraction of nuclei in each cluster that express each gene and the color of each circle represents the z-score of the log normalized mean expression of counts per million of each gene in each cluster (Z-score of log10(CPM+1)). Microglia-1-5, Meninges-1-4 and Endothelial-1-2 corresponds to Micro-1-5, Men-1-4, Endo-1-2, respectively; in Fig1B. Ex- Excitatory, Inh- Inhibitory, MV- Mixed ventral.



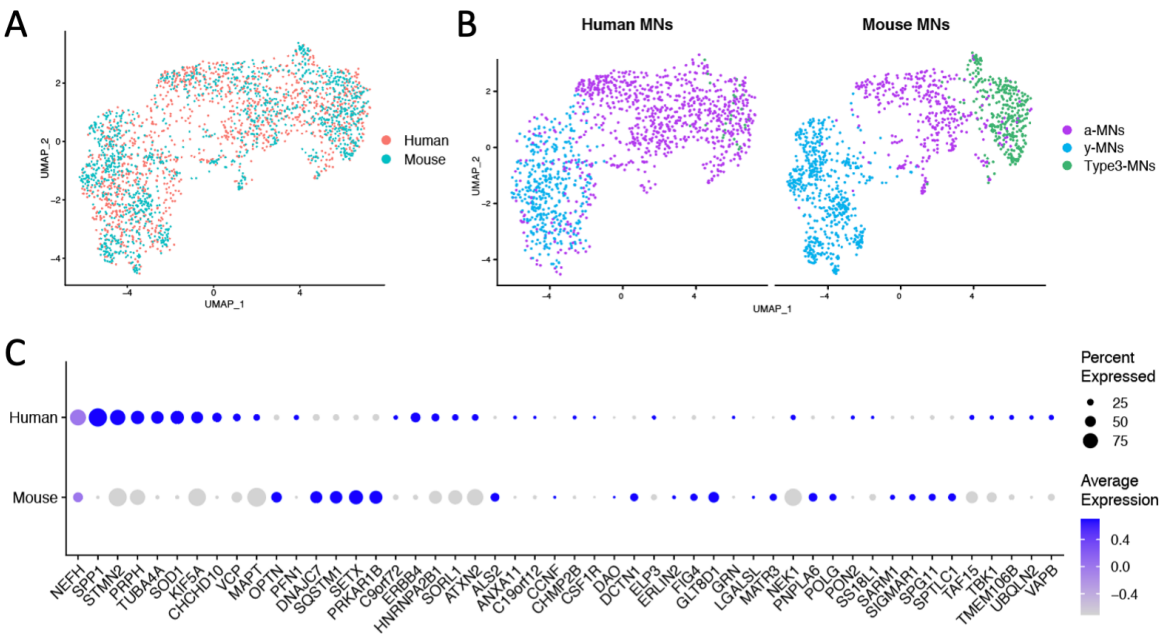
Supplementary Fig. S8. Integration and sub clustering of Neuronal sub-types in human spinal cord. (A) UMAP representation of the integration of 7 human spinal cord datasets that were identified as neurons in top level clustering (B) Violin plot depicting number of unique genes detected per nucleus per neuronal subtype. 'nFeature_RNA' is the number of genes detected in each nucleus. (C) Box plot showing distribution of silhouette scores per nucleus per neuron population in order to assess cluster robustness. The dots represent outliers. (D) UMAP plot of all neuronal nuclei colored based on their silhouette scores.



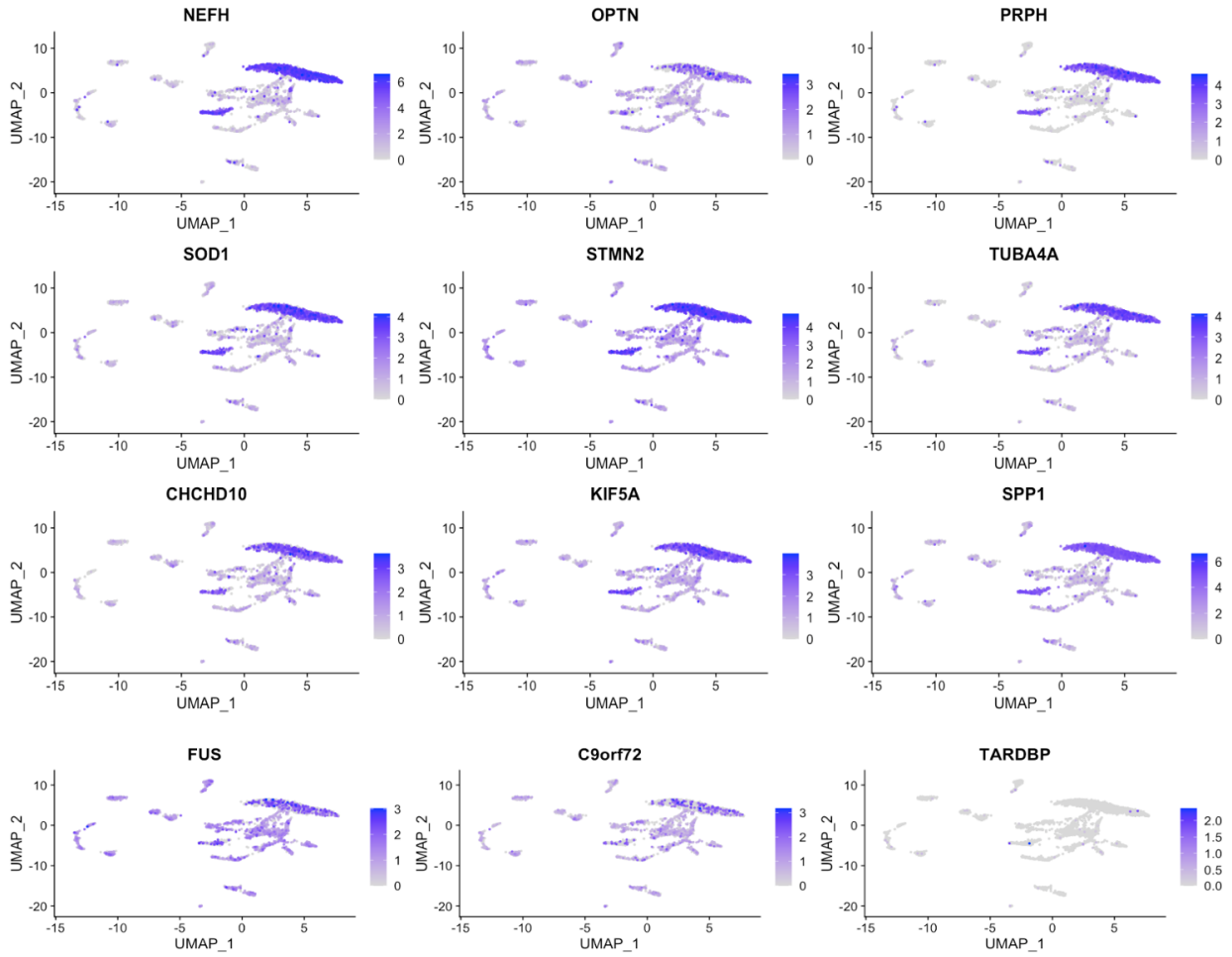
Supplementary Fig. S9. Constellation plot depicting the connections between different neuronal clusters, based on 100 iterations of post-hoc classification. The nodes are different clusters and the edges correspond to proportion of nuclei that were ambiguously predicted and are shared between the clusters during multiple iterations. The genes in each box represent a unique combination of markers to identify each cluster**.** Ex- Excitatory, Inh- Inhibitory, MV- Mixed ventral.



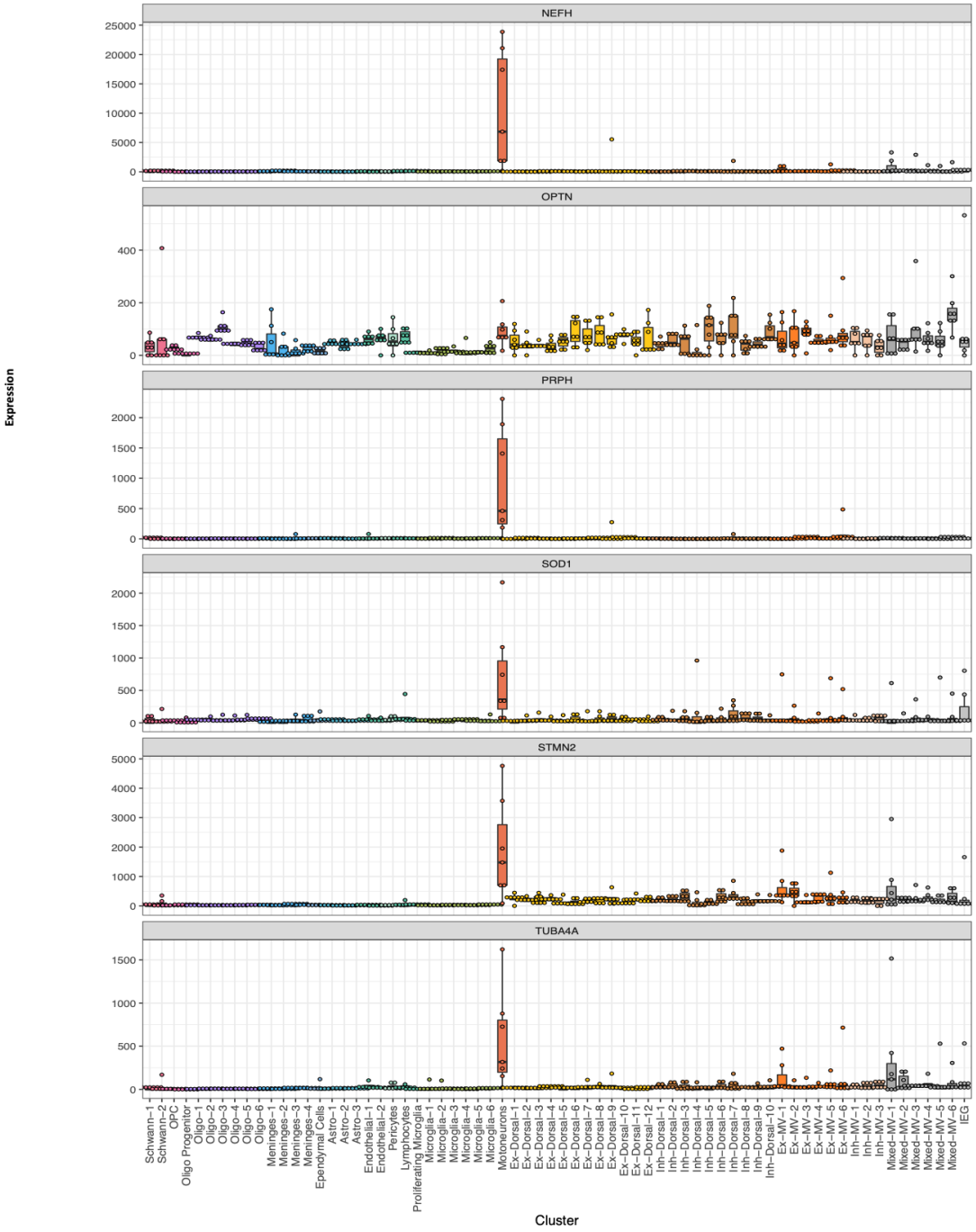
Supplementary Fig. S10. Cross species analysis between Human spinal cord and harmonized mouse spinal cord atlas (*28*). (A) Integration of the Human and mouse spinal cord datasets at the top level (includes all cell types) (B) Plot showing correlation between mouse and human cell types. Correlation is colored from purple to yellow and was calculated using principal components. (C) All the human spinal cord cell types overlayed on integrated cross-species UMAP (D) All the mouse spinal cord cell types overlayed on integrated cross-species UMAP. (E) and (F) Integration of only Neurons and overlaying of (E) human and (F) mouse annotations.



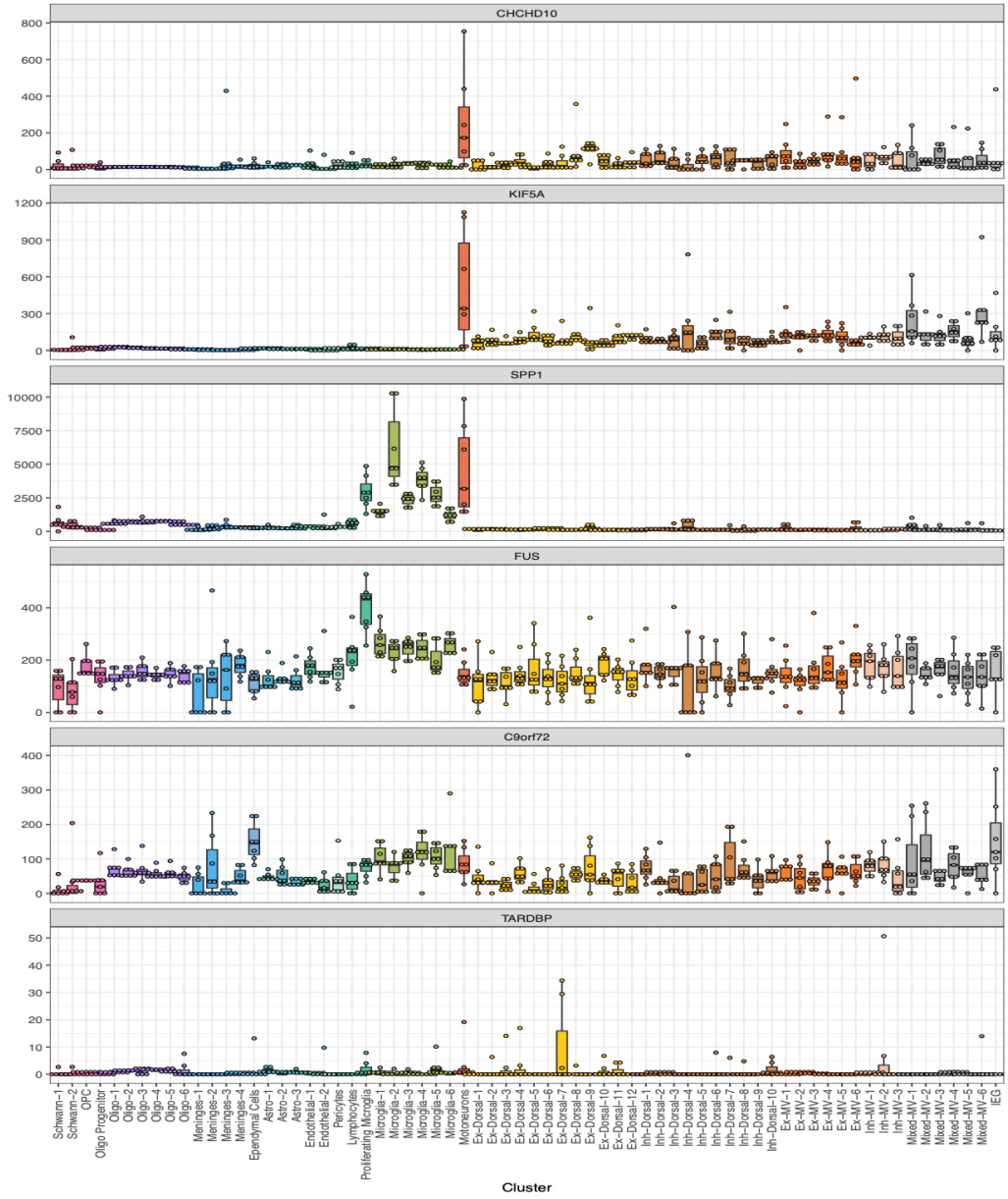
Supplementary Fig. S11. **Human and mouse motor neurons differentially express ALS risk genes**. (A) UMAP representation of integrated human and mouse MN data (*46*) by dataset. (B) UMAP representations of co-clustering of human and mouse MN data revealing potential alpha and gamma MN subtypes in human. (C) Dot plot showing expression of known ALS risk genes across human and mouse MNs. The size of the dot corresponds to the percentage of cells that belong to particular category. The color corresponds to Average expression across all cells for a particular class.



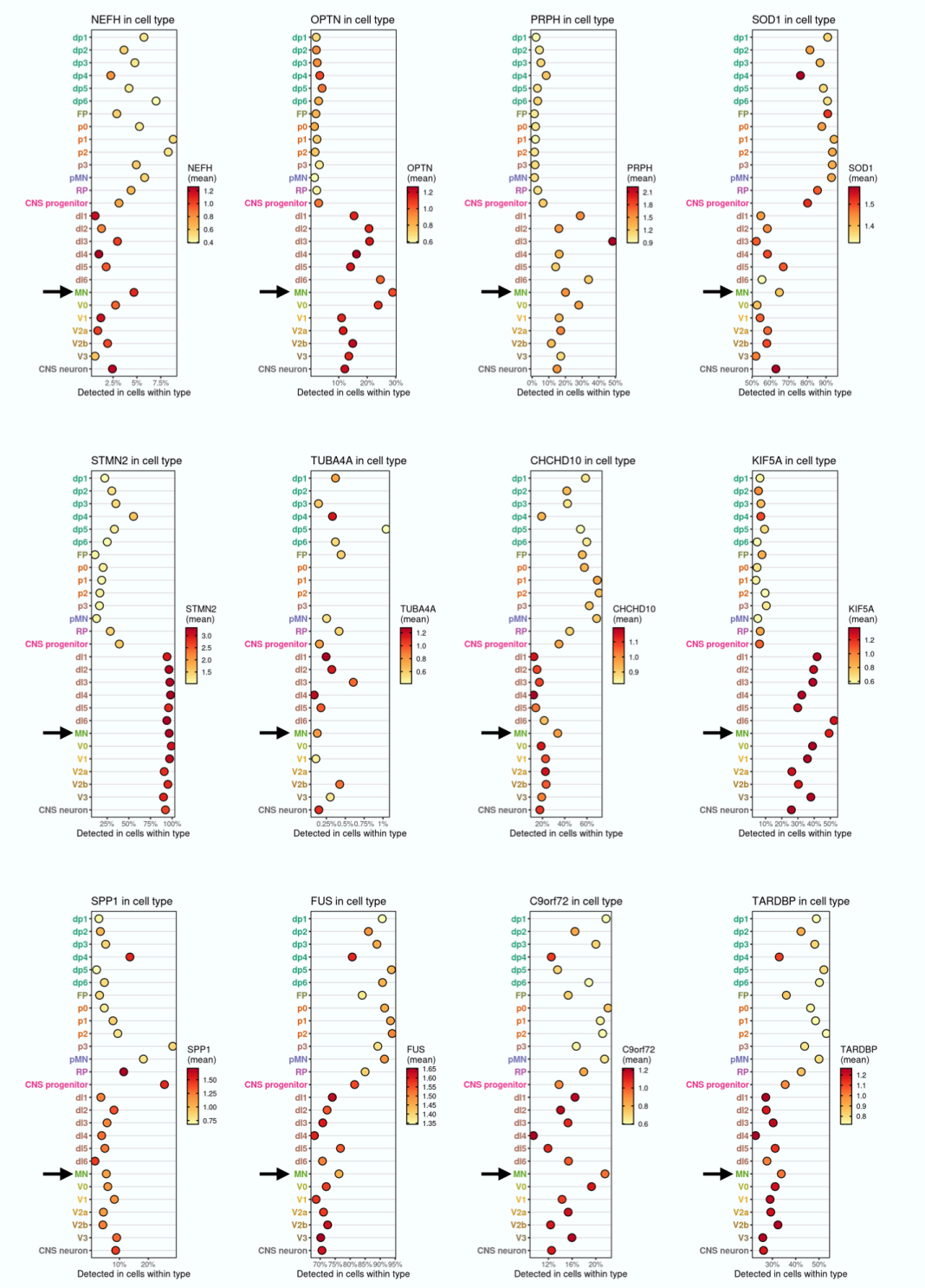
Supplementary Fig. S12. Expression of ALS-related genes in human spinal cord neurons. Log-normalized expression of NEFH, OPTN, PRPH, SOD1, STMN2, TUBA4A, CHCHD10, KIF5A, SPP1, FUS, C9orf72, TARDBP in the neurons represented in a UMAP plot. Color intensity from grey to dark blue corresponds to the amount of log normalized expression with dark blue being highest and grey being the lowest expression.



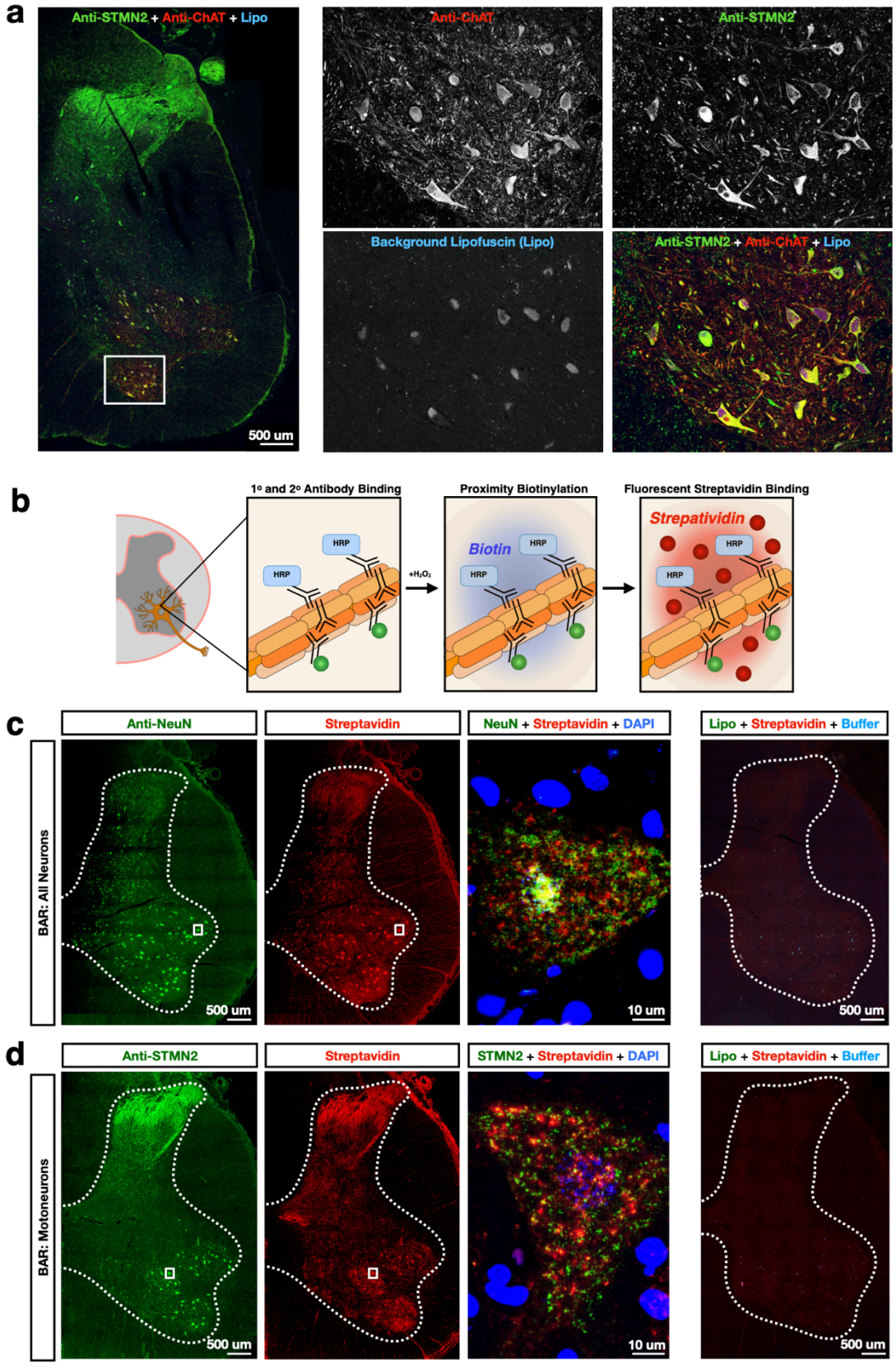
Supplementary Fig. S13. Expression of ALS-related genes in all human spinal cord cell types. Box plot shows per-cluster and per-sample expression (Counts per Million) of ALS-related genes namely, NEFH, OPTN, PRPH, SOD1, STMN2, TUBA4A, CHCHD10, KIF5A, SPP1, FUS, C9orf72, TARDBP, in order to examine consistency/variability across subjects.



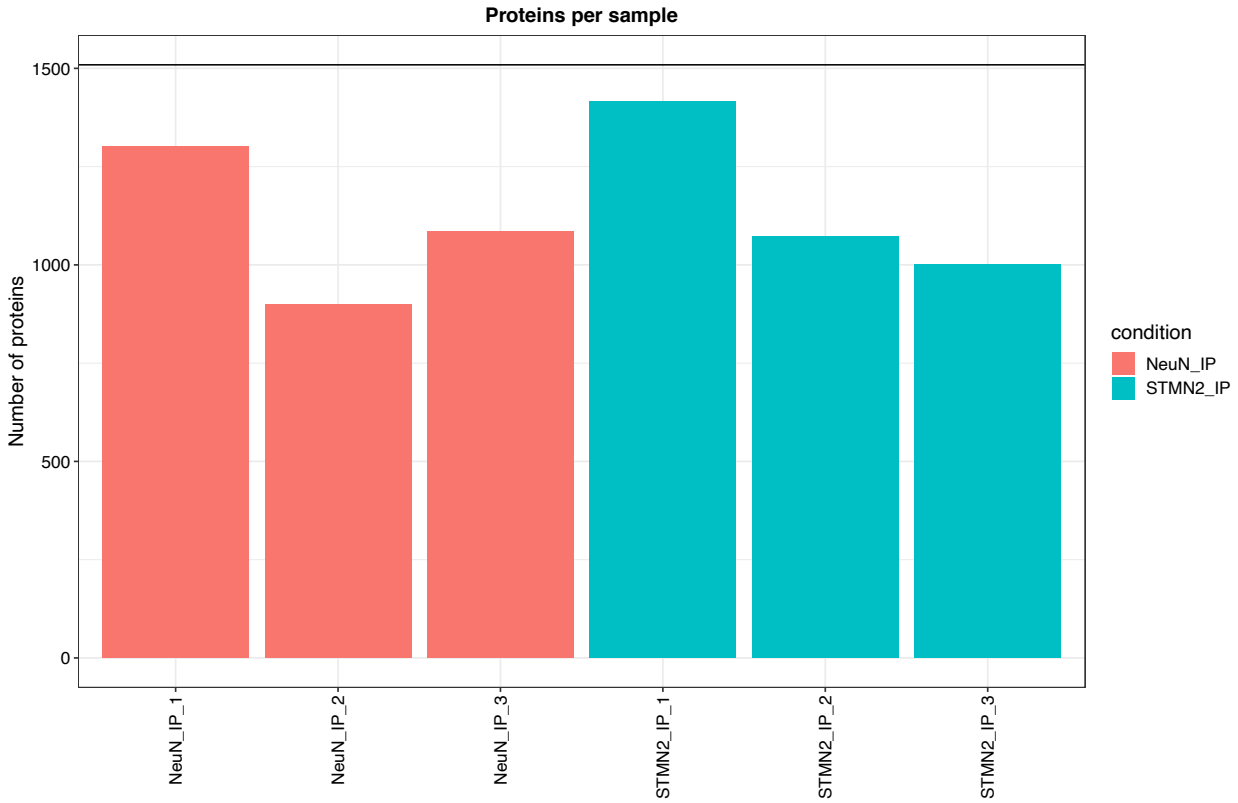
Supplementary Fig. S14. Expression of ALS-related genes in all human spinal cord cell types. Box plot shows per-cluster and per-sample expression (Counts per Million) of ALS-related genes namely, NEFH, OPTN, PRPH, SOD1, STMN2, TUBA4A, CHCHD10, KIF5A, SPP1, FUS, C9orf72, TARDBP, in order to examine consistency/variability across subjects.



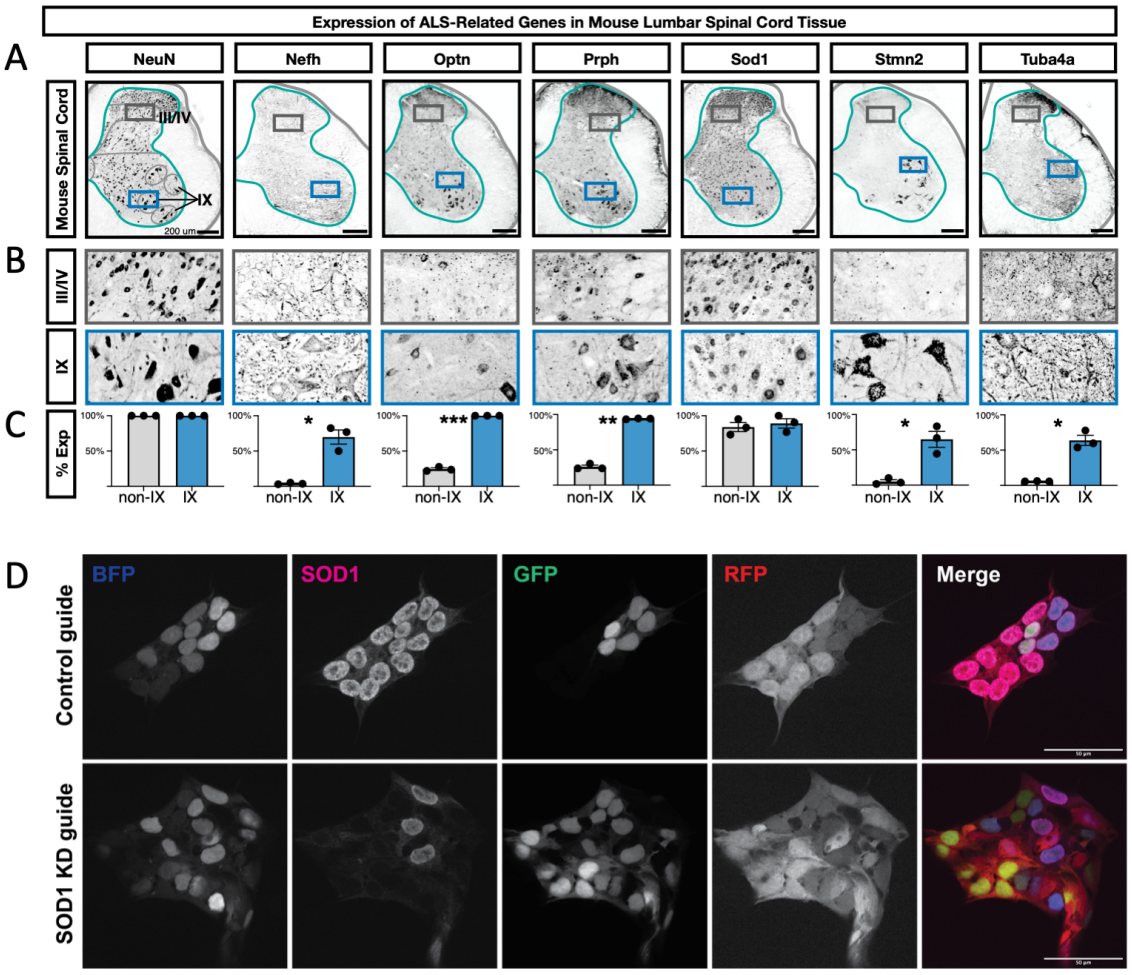
Supplementary Fig. S15. Expression of ALS-related genes in embryonic human spinal cord. Plot showing the level (color) and percent expression (location on x-axis) of 12 selected ALS-related genes in human embryonic progenitor and post-mitotic cell-types, based on https://shiny.crick.ac.uk/scviewer/neuraltube/ from (*25*). Motoneurons (MN) are indicated by a black arrow.



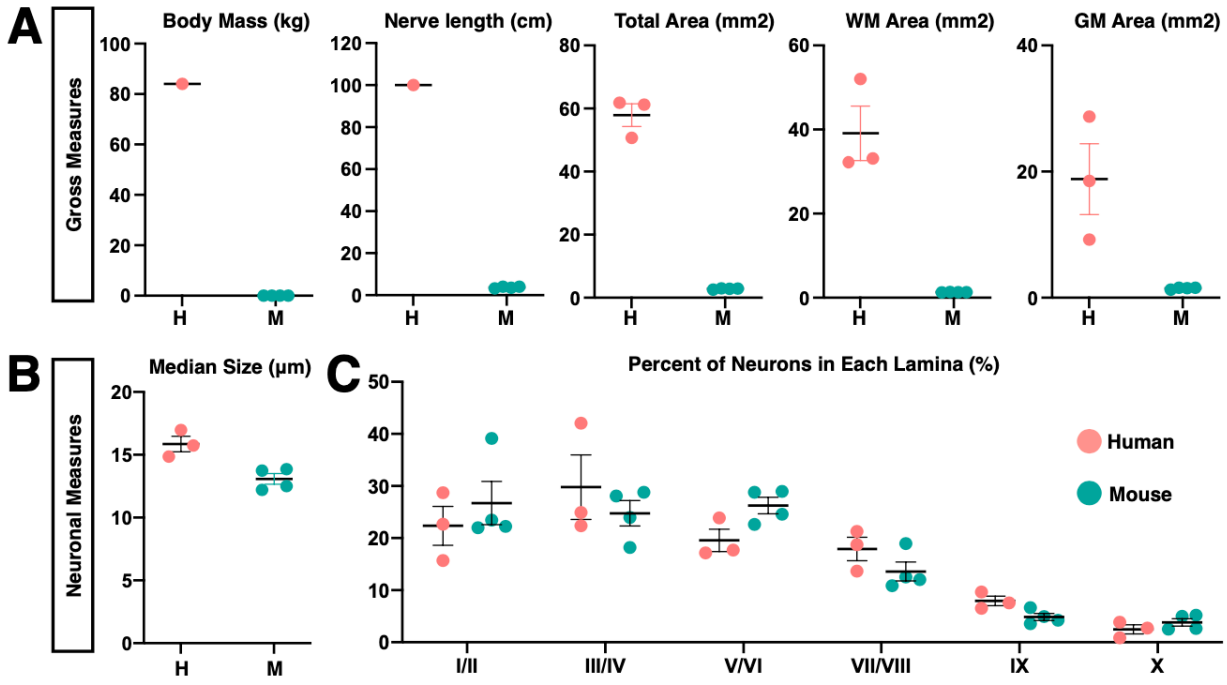
Supplementary Fig. S16. (**A**) Immunofluoresence validation of the expression of STMN2 in spinal motoneurons in a half-section of human lumbar spinal cord. Antibodies to STMN2 (Anti-STMN2, green) co-localized with antibodies to the cholinergic marker ChAT (Anti-ChAT, red) in large ventral cells that are putative motoneurons. Both proteins are detected above the background lipofuscin (Lipo, blue) signal. The scale bar on the half-section on the left is 500 um and higher magnification views of the boxed region are shown on the right. **(B)** Summary of the biotinylation by antibody recognition (BAR) experimental flow for labeling controls. Human spinal cord sections are incubated with primary antibodies to specific protein targets and then with secondary antibodies coupled to either a fluorescent label (green circles) or to horseradish peroxidase (HRP). The fluorescent label is only used for control purposes. After application of hydrogen peroxide, the HRP enzyme creates a biotin cloud in the proximate vicinity of the antibodies. The subsequent application of fluorescently labeled streptavidin (red circles) reveals the distribution of labeled protein in the spinal cord tissue. **(C)** Half-sections of adult human spinal cord sections with BAR controls for NeuN primary antibody which labels all neurons, showing the secondary fluorescent antibody (green), the fluorescent streptavidin (red), and a high magnification view of the merge with DAPI (blue). A control section with no primary antibody but only secondary HRP antibody and fluorescent streptavidin shows background lipofuscin (Lipo). Scale bars are 500 mm in the full section and 10 mm in the high magnification view. **(D)** Half-sections of adult human spinal cord sections with BAR controls for STMN2 primary antibody which labels spinal motoneurons (as well as sensory fibers in the dorsal horn), showing the secondary fluorescent antibody (green), the fluorescent streptavidin (red), and a high magnification view of the merge with DAPI (blue). A control section with no primary antibody but only secondary HRP antibody and fluorescent streptavidin shows background lipofuscin (Lipo). Scale bars are 500 um in the full section and 10 um in the high magnification view.



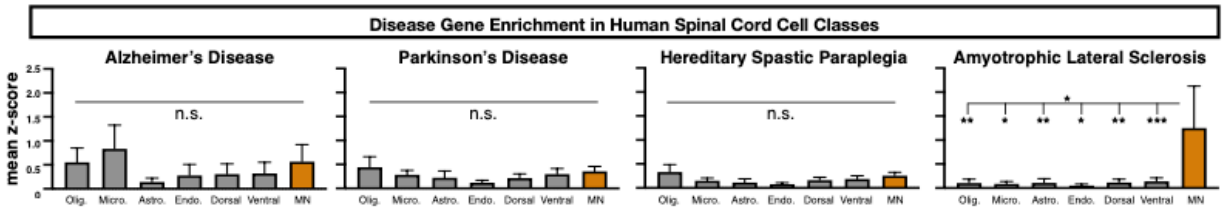
Supplementary Fig. S17. Barplot showing the number of identified proteins per sample obtained from BAR-MS proteomics pipeline. The line at 1509 depicts the total proteins identified.



Supplementary Fig. S18. Expression of ALS-related genes in mouse lumbar spinal cord tissue. (A) Antibody staining on lumbar spinal cord from aged mice (11 months old) for the orthologous proteins to those shown in Fig3C in main manuscript**.** Gray matter outlines are shown in teal and boundaries of lamina I/II, III/IV, V/VI, VII/VIII, IX, and X are shown in gray. The boxes indicate the enlarged images in panel B. (B) Inset of the images in panel A from the boxed region in laminae III/IV or lamina IX. Scale bars are 200 µm and the width of the enlarged images is 200 µm. (C) Quantification of the percent of NeuN+ neurons that co-expressed the indicated proteins in either all neurons not in lamina IX (non-IX) or those in lamina IX. The mean +/-s.e.m.are shown. The plotted values and number of cells counted in each subject and category are available in Supplementary Table 5. Paired t-test results are shown where * indicates $p < 0.05$, ** indicates $p < 0.005$. (D) Representative images of human inducible pluripotent stem cells (hiPSCs) with SOD1 knockdown and control guides 2 days after knockdown. Cells have nuclear-localizing GFP (green) and cytosolic RFP (red). BFP (blue) signifies guide uptake. Cells were stained for SOD1 (magenta). Scales bar are 50 mm.



Supplementary Fig. S19. Gross anatomical and neuronal measurements of the human and mouse lumbar spinal cords. (A) Measures of body mass, nerve length, total area, white matter (wm) area and grey matter (gm) area in the human (pink)and mouse (teal) lumbar spinal cord. Sources for human body mass (https://www.cdc.gov/nchs/fastats/body-measurements.htm) and for human nerve length (***56***). (B) Median size of human and mouse neurons (μm). (C) Percent of lumbar spinal cord neurons that reside in a given Rexed lamina.



Supplementary Fig. S20. Comparison of the z-scores (mean ± s.e.m.) for genes associated with the degenerative diseases Alzheimer's disease, Parkinson's disease, HSP, and ALS in seven different broad classes of cells: oligodendrocytes (Olig.), microglia (Micro.), astrocytes (Astro.), endothelial cells (Endo.), dorsal horn neurons (Dorsal), ventral horn neurons (Ventral), and motoneurons (MN, orange). Gene lists for each disease are available in Data File Table S6. One-way non-parametric Friedman's test was used to determine whether any cell types varied within each panel of disease genes and subsequently, non-parametric Wilcoxon tests were used to test each pair of cell types for significant differences. Friedman's test $p = 0.0278$. * indicates $p < 0.05$, ** indicates $p < 0.005$.