ALS disrupts spinal motor neuron maturation and aging pathways within gene co-expression networks

Ritchie Ho1, Samuel Sances1, Genevieve Gowing1, Mackenzie Weyandt Amoroso2, Jacqueline G O’Rourke1, Anais Sahabian1, Hynek Wichterle3,4, Robert H Baloh1,5, Dhruv Sareen1,6 & Clive N Svendsen1,6

Modeling amyotrophic lateral sclerosis (ALS) with human induced pluripotent stem cells (iPSCs) aims to reenact embryogenesis, maturation and aging of spinal motor neurons (spMNs) in vitro. As the maturity of spMNs grown in vitro compared to spMNs in vivo remains largely unaddressed, it is unclear to what extent this in vitro system captures critical aspects of spMN development and molecular signatures associated with ALS. Here, we compared transcriptomes among iPSC-derived spMNs, fetal spinal tissues and adult spinal tissues. This approach produced a maturation scale revealing that iPSC-derived spMNs were more similar to fetal spinal tissue than to adult spMNs. Additionally, we resolved gene networks and pathways associated with spMN maturation and aging. These networks enriched for pathogenic familial ALS genetic variants and were disrupted in sporadic ALS spMNs. Altogether, our findings suggest that developing strategies to further mature and age iPSC-derived spMNs will provide more effective iPSC models of ALS pathology.

ALS is a devastating neurodegenerative disorder that typically presents in adulthood, has no effective treatments, and involves paralysis and death within 3–5 years of diagnosis. While animal models of ALS have captured molecular and physiological aspects of disease onset and progression, genetic differences between humans and animals limit the interpretation and relevance of phenotypic results. A powerful in vitro system complementary to animal models of ALS is patient-derived iPSCs. iPSCs from ALS patients have the advantage of harboring the patient’s complex genetic makeup that contributes to their disease. In disease modeling with iPSCs, the goal is to recapitulate in a dish the embryonic development, maturation and aging of cell types involved with ALS pathology. spMNs are one of the primary cell types that die in ALS. While their embryonic development is understood to enable their differentiation in vitro from pluripotent cells, the fidelity of in vitro spMNs to in vivo spMNs remains unresolved. Thus, further optimizing spMN differentiation protocols from iPSCs could enhance the monitoring of disease onset and reveal molecular mechanisms to therapeutically target.

spMN progression from the embryonic to the mature and aged state, the point at which they degenerate in ALS, occurs over decades. It therefore remains unclear whether iPSC-derived motor neurons (iMNs), which are produced in vitro in several weeks, can recapitulate the decades of complex in vivo events leading to motor neuron (MN) degeneration. Notably, a systems-level comparison has not been conducted between iMNs and adult spMNs, the in vivo counterparts that iMNs attempt to model. We thus performed a direct, global transcriptional comparison of spMNs and iMNs to examine the expression kinetics of gene networks as cells progress from pluripotency to matured and aging adult states. These analyses indicated that iMNs are more similar to fetal than to adult spinal tissue and revealed a sequential activation and repression of molecular pathways in spinal motor tissue through stages of embryonic development, maturation and aging. Notably, these networks were enriched for genetic variants associated with MN disease, revealing that maturation- and age-related pathways may play roles in disease initiation. Finally, these maturation- and aging-associated gene networks were dysregulated in familial and sporadic ALS. Collectively, these findings suggest that strategies to mature and age iPSC-derived spMNs may provide more relevant iPSC models of ALS.

RESULTS
iPSC-derived MNs resemble fetal rather than adult MNs

We compared expression profiles among fibroblasts, iPSCs, fetal spinal cord, whole adult spinal cord and laser captured MNs from the spinal cords of control and ALS patients, which included those with mutations in the superoxide dismutase 1 gene, SOD1, or the charged multivesicular body protein 2B gene, CHMP2B9 (Supplementary Table 1a). Selecting profiles from the same microarray platform reduced the likelihood of confounding batch effects. An established 7-week in vitro protocol was used to generate cultures with 33–45% ChAT- and neurofilament heavy polypeptide (SMI-32)-double positive differentiated iMNs10 (Fig. 1a). We also obtained expression data for MN cultures from human embryonic stem cells containing an HB9::GFP reporter driven by the expression of a homeobox transcription factor involved with MN and pancreas development; the cells were sorted into GFP-positive and negative fractions11.
To represent immature in vivo spinal tissue during embryonic development, we generated expression data from human fetal spinal cords. Immunostaining with SMI-32 and the insulin gene enhancer transcription factor ISL1 revealed MNs in the ventral horn (Fig. 1b). After normalization of the expression for 10,605 genes in all samples (n = 43, Supplementary Table 2a), hierarchical clustering and Pearson correlation revealed that iMNs were more similar to fetal rather than adult spinal cord tissue or iPSCs from which they were differentiated (Fig. 1c). Principal component analysis (PCA) revealed that principle component (PC) 1 distinguished between pluripotent cells and adult spMNs (Supplementary Fig. 1a). PC2 distinguished non-laser-captured adult spinal tissues from all other samples, possibly as a result of strong gene expression contributed by heterogeneous adult spinal tissues that was absent in other samples. Notably, PC3 distinguished iMNs and fetal spinal cords from all other samples. When considering only PC1 and PC3 (Fig. 1d), an intuitive progression of MN development can be visualized. Regression analysis of linear models for sample traits against PC coordinates also revealed that tissue type best explained PC1 and PC3 (Supplementary Fig. 1b and Supplementary Table 1b).

Next, molecular pathways associated with PC1 and PC3 were examined. Gene set enrichment analysis (GSEA) was performed on the ranked gene loadings for PC1 and PC3 to detect gene sets that concordantly placed each sample along each component (Fig. 1e and Supplementary Table 2b). Pathways related to synaptic transmission were enriched among positive gene loadings along PC1, supporting the idea that MN maturation is characterized in part by changes in electrophysiology. Conversely, pathways related to cell cycle and DNA repair were enriched among negative gene loadings along PC1. Furthermore, pathways related to nervous system development were enriched among positive gene loadings along PC3. Genes associated with integrin binding, which regulates neuronal migration during development, were enriched among negative gene loadings along PC3. Together, these data account for differences in global expression as well as specific pathways across all samples, further supporting the idea that iMNs are molecularly restricted to a fetal-like state. By summarizing the diverse Gene Ontology (GO) terms enriched along their PC gene loadings for this expression data set and for simplification, PC1 is best described as spMN maturation, and PC3 is best described as embryonic spMN development.

Network analysis resolves development, maturation and aging
Gene co-expression network analysis is a useful method to link tightly co-expressed gene modules to phenotypic traits. Using weighted gene co-expression network analysis (WGCNA), 10,605 genes across 40 samples were hierarchically clustered based on topological overlap (Fig. 2a). This analysis identified 55 modules ranging in size from 30 to 511 genes with a median of 66 genes (Supplementary Table 2c). Altogether, 4,711 genes were assigned to modules (Fig. 2a) and 5,894 genes were not classified into any modules (Fig. 2a).

The module eigengene represented the expression of all genes classified into that module. When each eigengene expression was correlated against a sample trait using Pearson’s method (Supplementary Table 1a), only the age of the spinal tissue donor significantly correlated positively or negatively with a subset of eigengenes (Fig. 2b). In addition to sample traits, each sample analyzed by PCA was assigned a coordinate along each PC. When correlation tests were run for each eigengene expression against PC1, the component which described spMN maturation, or against PC3, the component which described embryonic spMN development, subsets of eigengenes were significantly correlated with each PC in an independent manner (Fig. 2b).

This analysis thus identified gene co-expression networks that strongly associate with age, spMN maturation and embryonic spMN development (Fig. 2c).

While neuron maturation and aging are physiological features likely involved in ALS and other late-onset neurodegenerative diseases, their distinction has been difficult to define. While regression analysis highlighted tissue type as the main covariate that correlated with PC1, the age of tissue donor slightly but significantly correlated with PC1 as well (Supplementary Fig. 1b). Notably, correlations performed on the 55 modules revealed that some modules significantly correlated with spMN maturation but not age and vice versa (Fig. 2b). Additionally, some modules correlated with both spMN maturation and age in the same direction, while other modules correlated with both in opposite directions. WGCNA was thus able to finely resolve gene co-expression networks associated with the processes of neuron maturation and aging.

ALS genetic variants disrupt spMN maturation and aging
To test if any of the 55 modules identified by WGCNA were associated with age-related diseases, we ran enrichment analyses using the ClinVar database. Pathogenic genetic variants associated with the search terms “motor neuron” were enriched in the dark gray, purple and light green modules, and those associated with “amyotrophic lateral sclerosis” were enriched in the yellow module (Fig. 2b). Notably, these four modules were significantly correlated with spMN maturation, age or both (Fig. 2b). Variants associated with dementia were enriched in the green-yellow module, which significantly anticorrelated with age, and variants associated with autism were enriched in the dark magenta module, which significantly correlated with embryonic development. Enrichment for MN disease and ALS variants in these modules were specific, as there was no significant enrichment for genetic variants associated with other age-related disease like Alzheimer’s disease, hypertension or prostate cancer (Supplementary Fig. 2a), suggesting that the age- and spMN-maturation-correlated modules are indeed modifiers of ALS and MN disease presentation. We subsequently analyzed modules built with the expression data set that included the ALS samples to increase the sample size of spMNs, thereby increasing the robustness of the detected modules, and to provide some representation of co-expression networks as they occur in spMNs from ALS patients. We note that mutations in CHMP2B are of uncertain pathogenicity in ALS, given that there is insufficient evidence of segregation with disease in familial pedigrees and that a rare, candidate polymorphism exists in non-ALS populations.

Network analysis resolves spMN maturation and aging pathways
To gain a biological understanding of network-to-sample trait relationships, the age- or spMN-maturation-associated modules were tested for enrichment of genes belonging to GO terms. Each of the 55 modules was grouped into four classes: significant positive (PC1pos) or negative (PC1neg) correlation with spMN maturation, and positive (AGEpos) or negative (AGEneg) correlation with age (Supplementary Table 2d). Performing a four-way intersection, this revealed numbers of overlapping and distinct GO terms enriched in each class (Fig. 2d and Supplementary Fig. 2b,c).

Pathways involved in cell cycle and mitosis were exclusively enriched in PC1neg (Fig. 2d), consistent with GSEA results (Fig. 1e). Neuron projection modules positively correlated with both PC1 and age, suggesting that axonogenesis is a continual process throughout...
spMN maturation and aging. Translation and ribosomal pathways were enriched in both PC1neg and AGEneg (Fig. 2d), indicating that coregulated gene networks associated with protein production decline in cells during spMN maturation and aging. Conversely, expression of proteasome-associated genes increased with age, suggesting a role for protein degradation and catabolism in older cells. Notably, amyloid precursor processing genes were enriched in AGEneg, indicating a progressively declining ability to prevent glycoprotein aggregates in older cells. Immune response and myelin sheath genes were enriched in PC1pos and AGEneg (Fig. 2d), suggesting that these gene networks increase in expression during spMN maturation but decrease in expression during aging. Altogether, these analyses effectively

**Figure 1** iPSC-derived MNs resemble fetal rather than adult MNs. (a) Immunostaining of iMNs used in expression profiling. ChAT (red) and SMI-32 (green) positive cells indicate the presence of MNs differentiated from iPSCs. Nuclear DAPI stains are shown in blue. n = 3 independent experiments and differentiation efficiency is quantified by ChAT- and SMI-32-double positive motor neurons. Scale bar, 4.75 μm. (b) Immunofluorescence staining of a 63-day-old fetal spinal cord. Inset depicts ISL1 (red) and SMI-32 (green) positive cells in the motor horn with ventrally projecting processes, indicating the presence of MNs at this developmental stage. Nuclear DAPI stains are shown in blue. Scale bars, 200 μm. (c) Unsupervised hierarchical clustering of 10,605 mRNA transcripts in n = 43 samples. Heat map indicates Pearson’s correlation between pairwise sample comparisons, and dendrogram indicates average linkage distance between samples. ESC, embryonic stem cell. (d) Principal component analysis of 10,605 mRNA transcripts in n = 43 samples. Sample coordinates along PC1 and PC3 are shown. Total variance: PC1, 14.61%; PC3, 7.01%. For sample colors, see tissue type legend in c. Arrows depict the progression of fibroblasts (magenta) reprogrammed to iPSCs (red), the subsequent differentiation of iPSCs into either HB9::GFP-negative cells (yellow) or HB9::GFP-positive MNs (green), which project toward fetal spinal cords (cyan). Arrows also depict the progression of fetal spinal cords toward adult spinal tissues (blue and black). Black labels next to cyan data points indicate age of fetal spinal cord donors in µm. (e) Gene set enrichment analysis of ranked gene loadings from PC1 and PC3 for GO terms. Positive and Negative categories indicate GO terms for each category are listed along with family-wise error rate (FWER)-corrected P-values. Enrichment plots are shown for bolded GO terms.

For additional information, see Supplementary Figure 1 (all PCs and linear model statistics) and Supplementary Tables 1a (sample metadata), 2a (linear expression values) and 2b (full list of significantly enriched gene sets and P-values).
resolved biological processes associated with spMN maturation, aging or both. Furthermore, pathway enrichment through WGCNA recovered additional ontologies not detected using PCA with GSEA alone (Supplementary Table 2e), indicating that WGCNA is a more sensitive approach to discover pathways of interest.

Key markers assess spMN maturation and embryonic development

Having identified expression networks that significantly associated with the processes of spMN maturation and embryonic spMN development, we investigated whether the gene properties within those networks could identify key markers of both processes. A reduced list of key markers would enable a comparison among samples that were expression-profiled on different microarray platforms and provide a robust indicator of the embryonic developmental and maturation status of pluripotent cell-derived MNs. To this end, we scored each of the 10,605 genes along three properties with respect to PC1 and PC3 that reflect spMN maturation and embryonic spMN development, respectively (Fig. 1d): preferential gene loading, gene significance and intramodule gene membership. Preferential gene loading is the net contribution of each gene to one particular PC and minimal contribution to all other PCs. Gene significance is the Pearson’s correlation between the gene expression value and a sample trait, as in, in this instance, spMN maturation or embryonic development. Intramodule membership is the correlation between the gene expression value and the eigengene of the module to which it was assigned.

After generating gene scores for each of these properties, we partitioned the genes into four classes (PC1pos, PC1neg, PC3pos and PC3neg), based on the signed value of their gene loadings in PC1 or PC3, and ranked them by a score summarizing the three properties (Supplementary Table 2c-e-h). From each of these ranked lists, we selected five genes on the basis of the best gene scores as well as for diverse representation of module colors and prior knowledge of biological relevance to spMN maturation or embryonic spMN development. Together these totaled 20 genes that best predicted the sample clustering along the axes of PC1 and PC3 (Fig. 3a), and reverse transcription-qPCR validated the microarray expression patterns of these genes (Supplementary Fig. 3a and Supplementary Table 3a).

To test the robustness of this marker panel to assay spMN maturation and embryonic spMN development, expression data were downloaded from additional studies as validation data sets using a variety of microarray platforms (Supplementary Table 1a). As expected, expression values for 6,640 overlapping genes from different platforms and studies exhibited varied distribution patterns (Supplementary Fig. 3b). Quantile normalization conformed all values to a single distribution pattern (Supplementary Fig. 3b and Supplementary Table 3b).

Performing PCA on all 6,640 quantile normalized genes produced PCs that resembled the training data set (Figs. 1d and 3b and Supplementary Fig. 3c). Specifically, PC1 again best described spMN maturation, and PC7 in this instance best described embryonic spMN development. Notably, PC3 captured variations due to differences between Illumina and Affymetrix microarray platforms, despite quantile normalization minimizing these differences (Supplementary Fig. 3c,b). PCA performed using the 20 key genes reflected spMN maturation in PC1 and embryonic spMN development in PC2 (Fig. 3c and Supplementary Fig. 3d). Variations arising from different microarray platforms were not noticeably captured. A receiver-operating-characteristic analysis on the validation data set demonstrated that the 20 genes outperformed 6,640 genes with either PCA or Pearson correlation, yielding a higher area under the curve (Fig. 3d and Supplementary Fig. 3e-g). These observations support the efficacy and universal adaptability of combining PCA and WGCNA to identify key markers of spMN maturation and embryonic spMN development.

Figure 2  Network analysis resolves co-expression modules. (a) WGCNA clustered 10,605 genes across pluripotent cells, iMNs, fetal spinal tissues and adult spinal tissues based on similar network topology (n = 40 samples). Height metric on dendrogram indicates topological overlap (TO) distance between genes. A dynamic tree-cutting algorithm grouped tightly networked genes, illustrated as low-hanging branches, into 55 modules represented by arbitrary colors directly below the branches. Genes falling onto light gray are not classified into any module. (b) Upper and middle panels: heat map indicates Pearson’s correlation of module eigengenes with 5 sample traits. n = number of samples for which there is data for the indicated sample trait, and thus used in the correlation. Outlined panels indicate correlations with a Bonferroni-corrected P-value < 0.01; these modules were kept for subsequent GO analysis. Lower panel: gene variants associated with diseases in the ClinVar database were tested against each of the 55 modules for enrichment. n = number of samples for which there is data for the indicated disease, represented on the human microarray platform, and thus used in the enrichment analysis. Green heat map indicates Benjamini–Hochberg-corrected negative log10 P-value from each hypergeometric test. Corrected P-values < 0.05 are called significant and outlined in black. This data is also shown in Figure 5b. (c) For each module eigengene that significantly correlates (Positive) or anticorrelates (Negative) with age, PC1, or PC3 (Bonferroni-corrected P-value < 0.01), the module eigengene expression values are graphed in a scatter plot against either sample age or PC coordinate. Locally weighted scatterplot smoothing lines are graphed for each module. (d) Chow–Ruskey diagram illustrating the number of overlapping and distinct GO terms (Bonferroni-corrected P-value < 0.05) enriched in modules identified as significantly correlated or anticorrelated to age (AGEpos or AGEneg, respectively) or spMN maturation (PC1pos or PC1neg, respectively). Representative pathways are listed in gray boxes extending from diagram, along with the lowest Bonferroni-corrected P-values across all modules. For additional information, see Supplementary Figure 2 (additional WGCNA and Chow–Ruskey diagrams) and Supplementary Tables 2a (linear expression values), 2c (module assignments and properties), 2d,e (gene set enrichments and P-values) and 2j (module–trait correlations, P-values and ClinVar enrichment P-values).

mtSOD1 distinctively affects spMN and iMN networks

The expression of modules associated with age, spMN maturation and embryonic spMN development were next investigated in spMN and iMN samples comparing a familial form of ALS caused by mutant SOD1 (mtSOD1) to controls (Fig. 4, Supplementary Tables 1a and 4a,b). mtSOD1 significantly affected several modules in both classes. Notably, modules correlated to age were downregulated by mtSOD1 in spMNs (Fig. 4a). These age-associated modules in mtSOD1 spMNs were dysregulated in the opposite manner in which they are endogenously expressed during aging. This suggests that the aging process enacts homeostatic gene expression programs that may protect spMNs, which otherwise die due to misexpression of these transcriptional programs in the mtSOD1 condition. Notably, all of the embryonic spMN development-associated modules were significantly upregulated in mtSOD1 spMNs (Fig. 4c), indicating that some embryonic pathways are also affected in mtSOD1-induced ALS.

Mapping the expression of these trait-associated modules in an independent data set for mtSOD1 and control iMNs (Supplementary Tables 1a and 4a,b) revealed patterns that were similar to spMNs in some aspects but distinct in others. We compared module expression levels between mtSOD1 and control iMNs and found that the modules correlated with age, although they were not as highly expressed...
in iMNs as in spMNs, were significantly downregulated (Fig. 4a,d). These observations suggest that age-associated gene expression networks affected by mtSOD1 ALS in spMNs may be recapitulated to some extent in iMNs. Yet, the expression of modules correlated to spMN maturation was not significantly different in iMNs (Fig. 4b,e). These observations suggest that iMNs do not faithfully recapitulate...
Figure 3 Principal component and network analyses reveal key spMN maturation and embryonic development markers. (a) Gene loadings for 20 genes selected to best represent spMN maturation and embryonic spMN development. Colored arrows and gene labels depict gene loadings and module assignments for the 20 genes along PC1 and PC3. Open gray circles represent gene loadings for 10,585 genes not selected for the panel. PCA was performed on n = 43 samples depicted in Figure 1d. (b) PCA performed on 6,640 quantile normalized gene expression values across 120 samples. Samples are plotted by their coordinates along PC1 and PC7. Colors of data points indicate similar sample types, and shapes of data points indicate the tissue than adult spMNs. (c) Gene label as described in Supplementary Figure 3 (validation data) and Supplementary Tables 1a (sample metadata), 2c (module assignments and properties), 2f-i (gene scoring properties), 3a (qPCR primer sequences) and 3b (normalized linear expression values used in validation analyses).

The expression profile of mature spMNs affected by mtSOD1. Lastly, modules that were correlated or anticorrelated to embryonic spMN development were respectively highly or weakly expressed overall in iMNs (Fig. 1f). Thus, the expression pattern of these embryonic spMN development-associated modules in an independent expression data set supports the idea that iMNs are more similar to fetal spinal tissue than adult spMNs.

Sporadic ALS dysregulates spMN maturation and age modules
WGCNA with the composite data set analyzed thus far (hereinafter referred to as the ‘iMN expression data set’; Supplementary Table 2a) did not identify modules significantly correlated to ALS conditions (data not shown). This was likely due to an insufficiently large number of samples associated with ALS in the data set, perhaps also compounded by distinct expression changes induced by the two genotypes of ALS represented, mtSOD1 and mtCHMP2B. Thus, a separate analysis was performed to identify gene networks that significantly associated with the ALS condition in spMNs. We analyzed an independent transcriptomic data set of 15,614 genes that specifically focused on comparing expression between sporadic ALS (sALS) and control spMNs
(n = 22; age range 47–81 years, median 73 years; Supplementary Tables 1a and 5a). This data set, also used in the gene reduction validation analysis (Fig. 3b), similarly compared the transcriptional profiles of laser-captured spMNs between 12 sALS patients and 10 control subjects. However, this data set was profiled on a different microarray platform than was used for the composite iMN expression data set. Therefore, PCA and WGCNA were separately performed on this expression data set (hereinafter referred to as the ‘sALS expression data set’) to independently validate co-expression networks and observe whether their expression differed between sALS and control conditions. PCA and linear regression analysis revealed that PCA was best distinguished by sALS and control spMNs (Supplementary Fig. 4a,b and Supplementary Table 1c), and therefore, this PC was best described as the sALS component.

GSEA was performed on the ranked gene loadings for this component to explore pathways and gene ontologies enriched among genes that were dysregulated in sALS spMNs (Fig. 5a and Supplementary Table 5b). Notably, pathways related to extracellular matrix were enriched among genes that were upregulated in sALS, consistent with previously described gene expression changes21. Conversely, pathways related to mitochondrial ion and electron transport were enriched among genes that were downregulated in sALS. These observations were not previously described for this data set21, highlighting the sensitivity of GSEA using PCA gene loadings to detect enriched pathways without thresholding on an arbitrarily assigned fold change cut-off.

Next, gene co-expression networks were built using only expression data from the sALS expression data set. WGCNA on this data set identified 52 modules ranging in size from 31 to 401 genes with a median size of 51 genes (Supplementary Table 5c). Altogether,
Figure 4 Gene expression networks are distinctively affected by familial ALS in spMNs and iMNs. (a-f) Scatter and density plots depicting expression levels and mtSOD1-induced fold changes of age-, spMN maturation-, or embryonic spMN development-associated module genes as defined by WGCNA performed on the expression data set lacking ALS spMNs (Supplementary Fig. 2a). This was done to prevent modules from being defined by mtSOD1-induced gene expression changes in spMNs. Grey data points, no membership in module categories; red data points, membership in correlated modules; blue data points, membership in anticorrelated modules. Grey data points are plotted behind red and blue data points in all panels. Red data points are plotted in front of blue data points in a, b; blue data points are plotted in front of red data points in c-f. 8,830 overlapping genes represented in both spMN and iMN data sets are shown in each plot. x-axis, log2 average RNA expression values among all mtSOD1 and control (CTRL) samples. y-axis, log_{10} fold change in average RNA expression when comparing mtSOD1 to control samples; some outlier data points are cut off to visualize distribution shifts. The density plots account for all data points in the expression set, and straight lines mark the median value of each distribution. Asterisks indicate Kolmogorov-Smirnov two-sided P-values for like-colored categories tested against the gray distribution, and their locations indicate the direction of the shifted distributions. a–c, mtSOD1 (n = 3) versus control (n = 3) spMNs; d–f, mtSOD1 (n = 2) versus control (n = 3) iMNs.

4,444 genes were assigned to modules (Supplementary Fig. 4c), and 11,170 genes were not classified into any modules (Supplementary Fig. 4c). Each eigengene was then correlated using Pearson’s method against sex, sALS disease status, site of ALS onset, age of tissue donor, disease course and postmortem interval (Supplementary Table 1a). Additionally, each eigengene was correlated against the iMN component. Among these sample traits, only sALS disease status and the iMN component were significantly correlated or anticorrelated with a subset of modules (Supplementary Fig. 4d).

Given this observation, we hypothesized that these gene modules were also involved with spMN maturation and age. Therefore, the extents of overlap were examined between modules defined in the iMN and sALS expression data sets, hereinafter referred to as ‘iMN modules’ and ‘sALS modules,’ respectively. To this end, each sALS module was systematically tested for enrichment of each iMN module (Fig. 5b and Supplementary Table 5g). This analysis demonstrated that some iMN modules significantly overlapped with one or more sALS modules, and vice versa.

In addition to crosstabulating module assignments between the two data sets, the stability of networks defined in each data set was examined using module preservation z-statistics. These measures indicated the likelihood that the network structures of iMN modules also occurred in the sALS data set by random chance. Consistent with results of the overlap analysis, the most significantly overlapping modules were also those whose network structures were most significantly preserved across data sets (Fig. 5b, Supplementary Fig. 5, and Supplementary Table 5d).

Despite an imperfect one-to-one module correspondence, overlapping iMN and sALS modules tended to significantly correlate or anticorrelate to sample traits in a consistent way. For instance, iMN modules that significantly correlated with age and spMN maturation tended to have strong overlap with sALS modules that significantly correlate or anticorrelate with age and spMN maturation.
Figure 5 spMN maturation and age modules are dysregulated in sporadic ALS. (a) Gene set enrichment analysis of 15,614 ranked gene loadings from PC1 for pathways and GO terms. Positive and Negative categories indicate gene sets enriched among genes whose loadings contribute most to the respective positive or negative direction of the sALS component (PC1 in a PCA performed on \( n = 22 \) samples). Enriched gene sets for each category are listed along with FWER-corrected \( P \)-values. Enrichment plots are shown for bold gene sets. (b) For each of the 52 sALS modules, a hypergeometric test was performed to detect enrichment for genes from each of the 55 iMN modules. Upper panel: iMN modules and the sample traits with which they are significantly correlated, as in Figure 2b. Enrichment for ClinVar pathogenic variants in motor neuron disease or ALS is also shown. The \( z \)-summary value for each iMN module measures the extent of module preservation in the sALS data set. For the likelihood of module preservation: \( z \)-summary > 10, indicates strong evidence; 10 > \( z \)-summary > 2, moderate to weak evidence; 2 > \( z \)-summary, no evidence. Bar graphs above indicate the number of genes assigned to each iMN module that were also represented by probe sets on the Affymetrix Human Exon 1.0 ST Array. Left panel: sALS modules and the sample traits with which they are significantly correlated or anticorrelated, as in Supplementary Figure 4d. The \( z \)-summary value for each sALS module measures the extent of module preservation in the iMN data set. Bar graphs, left, indicate the number of genes assigned to each sALS module that were also represented by probe sets on the Affymetrix GeneChip Human Genome U133 Plus 2.0 Array. A matrix of \( P \)-values from hypergeometric tests performed for each iMN and sALS module overlap were corrected by the Benjamini–Hochberg method, and subsequent \( P \)-values < 0.05 are marked as black squares in the matrix diagram. For additional information, see Supplementary Figures 4 (all PCs and linear model statistics, WGCNA), 5 (module preservation statistics and intramodule membership of genetic variants) and 6 (Chow-Ruskey diagrams) and Supplementary Tables 2c (iMN module assignments and properties), 5a (linear expression values of sALS data set), 5c (sALS module assignments and properties), 5d (module preservation statistics), 5b,e,f (full lists of significantly enriched gene sets and \( P \)-values) and 5g (hypergeometric test \( P \)-values).

Anticorrelated with the sALS component. Conversely, iMN modules that significantly anticorrelated with age and spMN maturation tended to have strong overlap with sALS modules that significantly correlated with the sALS component. Additionally, three of the four modules enriched for genetic variants associated with MN disease or ALS significantly overlapped with at least one sALS module that in turn significantly associated with the sALS component or sALS disease status. These observations thus support the idea that networks involved in spMN maturation and age are also affected in sALS. To gain a biological understanding of these module-to-trait relationships, enrichment analysis was performed on each significantly associated module for GO terms. Modules were then classified into two groups: those that significantly correlated with the sALS component positively (sALSpos) or negatively (sALSneg) (Supplementary Table 5c). Like the comparisons in the iMN data set, pathway enrichment using WGCNA in the sALS data set recovered additional ontologies not detected using PCA with GSEA alone (Supplementary Table 5f). We conducted four-way interaction analyses with either AGEpogs and AGEneg, PC1pos and PC1neg, or PC3pos and PC3neg (Fig. 2d, Supplementary Fig. 2b,c); these analyses revealed numbers of overlapping and distinct GO terms enriched in each of the different groups (Supplementary Fig. 6). Pathways such as RNA processing and translational elongation, which were enriched within AGEneg and PC1neg, were also enriched within sALSpos (Supplementary Fig. 6a,b). Notably, the immune response pathway, which was enriched within PC1pos and AGEneg, was also enriched in sALSpos. Conversely, pathways such as neuron differentiation and cell projection, which were enriched within AGEpogs and PC1pos, were also enriched within sALSneg. To a lesser extent, some transcriptional pathways, which were enriched within PC3pos, were also enriched within sALSpos (Supplementary Fig. 6c).

spMN maturation, aging and sALS converge on hub genes

An alternative approach was taken to explore the network properties of genes affected by maturation, aging and ALS. Four module classes were defined: modules that (i) increased or (ii) decreased expression with aging and maturation (AGEpogs and PC1pogs or AGEneg and PC1neg, respectively) and modules that (iii) increased or (iv) decreased with sALS (sALSpos or sALSneg, respectively). Consistent with overlap comparisons performed on a module-by-module basis (Fig. 5b), this comparison of combined modules demonstrates that age and spMN maturation modules had extensive overlap with sALS modules in opposite directions (Fig. 5a). Having identified which genes within the module classes are commonly affected (hereinafter referred to as ‘overlaps’), the gene expression networks in both the iMN and sALS expression data sets were observed for any properties that distinguished overlaps from genes that were uniquely found in either of the module classes (hereinafter referred to as ‘nonoverlaps’).

When comparing the network property of gene significance, overlaps had more biologically significant roles than nonoverlaps. This was particularly the case for genes classified as AGEpogs, PC1pogs, and sALSneg (Fig. 6b and Supplementary Table 6) but not the case for genes classified as AGEneg, PC1neg, and sALSpos (Fig. 6c and Supplementary Table 6). Additionally, overlaps tended to have higher intramodule membership within AGEpogs and PC1pos modules as well as within sALSneg modules (Fig. 6d and Supplementary Table 6). Furthermore, pathogenic genetic variants associated with ALS tended to have higher intramodule membership than genes not affected by ALS variants (Supplementary Fig. 5e). Together, these observations further underscore the vulnerability of cellular systems to disruptions targeting network hub genes in both the familial and sporadic ALS conditions. Lastly, intermodule membership for each gene measures its expression correlation to the eigengenes of other modules that significantly correlate or anticorrelate with the same sample trait. This metric can be used to identify hub genes that connect multiple modules that act on distinct but coordinated pathways. This comparison showed that overlaps tended to have higher intermodule membership within AGEpogs and PC1pos modules as well as within sALSneg modules (Fig. 6e and Supplementary Table 6), suggesting that they are functionally conserved as intermodule hub genes.

These gene network properties can be simultaneously visualized as network maps (Fig. 6f–i). Within AGEpogs and PC1pos modules, a network map filtered for the strongest 1% of gene-to-gene connections comprised multiple interconnected modules (Fig. 6f). This network contained far more overlaps than nonoverlaps, and overlaps tended to have greater gene significance with respect to age and also tended to have greater intramodule membership. Nonoverlaps demonstrated a lower degree of intramodule membership compared to overlaps, as they occupied more peripheral positions within clusters (Fig. 6f). The same trends were observed in sALSneg modules (Fig. 6g), AGEneg and PC1neg modules (Fig. 6h) and sALSpos modules (Fig. 6i). Altogether, these data highlight the importance of hub genes within networks associated with MN maturation and aging, as evidenced by the preservation of their network structures within the strongest connections detected in the sALS data set. Furthermore, the high gene significance measures of these hub genes suggest they are the most responsive genes to perturbations caused by sALS.
DISCUSSION

Many iPSC differentiated tissues represent an immature or fetal stage of development. This is presumably due to resetting of the epigenetic state to that of an embryo during iPSC production, which can occur even with samples over 90 years old. Our study demonstrates that, based on their transcriptomes, iPSC-derived MNs are more similar to their fetal in vivo counterparts than to adult MNs.

Through an unbiased, genome-wide approach, we identified network hub genes described as markers or functional drivers of neuronal development or maturation. For example, DCX, NEFH and SNAP25

| Gene set                                      | Number of genes | Normalized enrichment score | FWER P-value |
|-----------------------------------------------|-----------------|------------------------------|--------------|
| PID integrin1 pathway                         | 60              | 2.45                         | 0.00         |
| PID AP1 pathway                               | 69              | 2.32                         | 0.00         |
| KEGG complement and coagulation cascades      | 65              | 2.30                         | 0.00         |
| Proteinaceous extracellular matrix            | 86              | 2.27                         | 0.00         |
| Extracellular matrix                          | 88              | 2.25                         | 0.00         |
| REACTOME ion channel transport                | 48              | -2.39                        | 0.00         |
| REACTOME ligand gated ion channel transport   | 19              | -2.37                        | 0.00         |
| Mitochondrial membrane part                   | 48              | -2.34                        | 1.00 × 10⁻³  |
| Mitochondrial inner membrane                  | 60              | -2.27                        | 1.00 × 10⁻³  |
| REACTOME respiratory electron transport       | 57              | -2.22                        | 2.00 × 10⁻³  |
Figure 6  Genes associated with spMN maturation, aging, and sALS tend to be hub genes. (a) Overlap of module genes in the iMN or sALS expression data sets. To be included into the AGEpos and PC1pos group, iMN module genes must be classified as AGEpos, PC1pos, or both. Additionally, their modules must have significant overlap with at least one sALS module that anticorrelates with the sALS component (Fig. 5b). These criteria are likewise for the AGEneg and PC1neg group overlapping with sALSpos modules. Modules built from the sALS data set are similarly classified. Venn diagrams indicate the number of genes assigned to each class. (b) Boxplots of gene significance for overlapping and nonoverlapping genes from the comparison depicted in a, left. Asterisks indicate P-values from the Wilcoxon rank–sum test. The y-axis indicates gene significance for module genes against the respective sample traits in their respective expression data sets. Center line, median; lower and upper limits of box, respectively, first and third quartile; lower and upper whiskers, respectively, minimum or maximum values of the distribution up to 1.5 times the interquartile range; open circles, outliers. (c) As in b, except for genes in a, right. (d,e) Boxplots of intramodule (d) and intermodule (e) membership for genes in a. (f) Cytoscape network maps for AGEpos and PC1pos iMN module genes. The top 1% of gene-to-gene connections (627 nodes and 3,197 edges) is plotted using the prefuse force layout method. Node color, module assignment; circular node, overlap; triangular node, nonoverlap. Relative node size reflects gene significance towards age in the iMN expression data set. (g) As in f, for sALSneg module genes (323 nodes and 2,886 edges). Larger nodes in this instance have gene significance values closer to −1 and are therefore more anticorrelated to the sALS component. (h) As in f, for AGEneg and PC1neg iMN module genes (669 nodes and 6,175 edges). Larger nodes have gene significance values closer to −1 and are therefore more anticorrelated to age. For better visibility, the three distinct, contiguous network clusters were rotated and repositioned relative to each other. Therefore, node-to-node spatial relationships within, but not across, contiguous network clusters are accurate. Edges intersecting with dashed lines were shortened and repositioned to scale the diagram. No other modifications were made. (i) As in g, for sALSpos module genes (312 nodes and 878 edges). Larger nodes have greater gene significance toward the sALS component. For additional information, see Supplementary Tables 2c (iMN module assignments and properties), 2d (full lists of significantly enriched GO terms and P-values), 2e (comparison of gene set enrichments), 5c (sALS module assignments and properties), 5e (full lists of significantly enriched GO terms and P-values), 5f (comparison of gene set enrichments) and 6 (Wilcoxon rank–sum test statistics).

have been shown as histological markers of early, intermediate and late neuronal maturation, respectively26. Additionally, loss-of-function mutations in SCN1A are implicated in a failure of interneurons to develop mature action potentials in Dravet syndrome27, and ASCL1 is the key pioneering transcription factor driving direct fibroblast conversion to neurons28, which subsequently demonstrate slower maturation kinetics29. Using only 20 markers, we reduced the number of genes tested and thereby circumvented microarray-platform-specific biases that confound accurate comparison of maturation states across samples. This technique will be extremely helpful in developing methods to efficiently mature spMNs in culture, and can be applied to other iPSC-derived cell types such as upper MN models of ALS30, given the transcriptomic data available for in vivo brain tissues at different developmental stages31,32.
Using MNs to model neurological disease may require their aging in the dish. Attempts to simulate aspects of aging in iPSC-derived neurons include prolonging time in culture or introducing agents of cellular stress by chemical or genetic means33-37. Notably, these strategies can concomitantly bring about age-related molecular features along with phenotypes associated with neurodegeneration. However, our data indicate that molecular processes of neuronal aging and neuronal maturation are distinct, substantiating a perspective held by others38. Therefore, it is unclear whether these techniques accelerate iPSC progression from the embryonic to a mature neuronal state, or simply induce isolated aging pathways in immature cells. Since cellular age is collectively a multifaceted, syndromic condition7, it is most effectively assayed in a systems-wide manner. Thus, the global gene expression and network analysis we have employed is well suited to assay cellular maturity and age and can provide a comprehensive method to assess iPSC-derived tissue models under stress-induced conditions. One caveat is that the adult spinal cord samples used in our composite data set are younger than the adult laser-captured MNs. We therefore recognize that some genes in the age-associated modules may actually be more associated with different sample types. Because a large number of in vivo spinal cords profiled on the same microarray platform along with detailed clinical data were difficult to obtain, we combined all available expression profiles in order to achieve more robust statistical power. Therefore, our data set provides a resource to generate hypotheses; however, the burden will be on follow-up investigations to eliminate such confounding factors.

Studies using direct neuronal reprogramming from aged adult fibroblasts have shown that the mature epigenetic state may be maintained39. While MNs have been directly converted from embryonic and fetal fibroblasts40,41, it would be interesting to see if similar techniques performed on aged fibroblasts maintain a more mature state through the transition process. Nevertheless, we provide data and methods that future studies can apply in addressing these questions.

There are increasing genetic variants reported to be associated with several human diseases in the ClinVar database17. Notably, some modules correlated with maturation or age (or both) are enriched for variants pathogenically linked to ALS or MN disease. This lends strong evidence for maturation- and age-associated networks and pathways acting as causative effectors of late onset diseases. Future characterization of novel genetic variants classified as risk factors may explain how they act collectively to modify the penetrance or onset time of disease in individual patients.42

The mitochondrial free radical theory of aging posits that as cells age, their electron transport chain stability declines and accumulating reactive oxygen species oxidatively damage nuclear and mitochondrial DNA43. Our analysis revealed that respiratory chain components decreased expression during spMN maturation, consistent with observations in ALS patients’ blood44. Additionally, DNA repair pathways decreased expression as spMNs mature, suggesting a reduced ability to mitigate oxidative DNA damage. Our analysis also revealed that sALS further downregulated mitochondrial respiratory chain genes. Thus, the combination of spMN maturation and sALS can exacerbate a condition in which spMNs are vulnerable to oxidative damage. Notably, these mitochondrial components were reported to be downregulated in mitSOD1 iMNs, indicating that even in a fetal context, some key pathways can already be affected.

We have highlighted a dynamic expression pattern of gene networks associated with antigen presentation and immune response. Whereas these processes increased expression as spMNs matured, they decreased expression as spMNs aged. This model is consistent with the role of microglia in pruning synaptic connections during neuronal maturation and their hyperactivity in late-onset diseases45. The immune response and complement activation pathways were also upregulated in sALS, similar to what has been observed in spinal cords of C9orf72-deficient mouse models of ALS46. In summary, sALS antagonizes the endogenous expression pattern of these immune activation pathways and that of protein translation and degradation. These observations therefore support the idea that expression kinetics for these pathways serve a homeostatic, protective role in aging spMNs, which can be derailed by ALS, resulting in neurodegeneration.

Our analysis also provided another dimension to the role of genes within maturation and aging expression networks. The scale-free architectures of natural systems are robustly tolerant of perturbations to the majority of its nodes, but at the cost of being vulnerable to disruptions targeting hub nodes47,48. Our observation that ALS preferentially disrupts hub genes within maturation and aging expression networks underscores its devastation to critical cellular systems. Understanding which central genes are the most vulnerable to ALS will guide effective therapies aimed at rescuing the function of these hub genes.

The stem cell modeling and ALS communities can apply the data presented here toward their own research in several ways. This resource includes a variety of transcriptomic data tables composed from and normalized across multiple studies; an atlas of gene networks that can be examined for common or distinct roles in maturation, aging or ALS in other tissues; enriched gene sets and pathways among those gene networks; and a panel of marker genes that can accurately indicate spMN maturity either by reverse transcription-qPCR or microarrays. Notably, this study describes an analytical framework for cross-study transcriptomic comparisons, and this methodology can be broadly applied to other models of disease.

Altogether, our findings support a strong interaction between gene networks and pathways associated with spMN maturation and aging and those affected in ALS. This suggests that reenacting the endogenous spMN maturation and aging pathways in iMNs can sensitize them to ALS-induced dysfunction. Nevertheless, it is possible that achieving a mature and aged state in iMNs is superfluous to effective ALS modeling. Disease-relevant phenotypes have been gleaned from immature iMNs, including RNA foci, protein inclusion bodies, altered electrophysiology, nuclear pore deficits and cell death.10,13,20,34,49,50 While some of these phenotypes have spurred attempts at therapeutic strategies in ALS patients, it remains to be seen whether they are indeed relevant to disease etiology in adults. For instance, if ALS patients treated with retigabine (an antiepileptic drug that reduces hyperexcitability in mtSOD1 iMNs) demonstrate a positive response to the treatment, this would validate the efficacy of current iMN models in predicting events in adult spMNs. Otherwise, maturation and aging pathways should be considered as modifiers of disease presentation, and our present findings lay the groundwork for future efforts to achieve a higher fidelity model of the molecular and pathological events of ALS as they occur in vivo.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

**Accession codes.** GEO: microarray data, GSE75701.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.
ACKNOWLEDGMENTS
The authors gratefully acknowledge B. Shelley, L. Garcia and L. Ornelas for assistance with experiments and reagent organization; B. Berman and D. Rushton for statistical and programming advice; and B. Shelley, L. Garcia and L. Ornelas for critical reading and comments on the manuscript. This work was supported by the following grants: NIH/NINDS (U54NS091046-01) (C.N.S.) and the ALS Society (R.H. and C.N.S.). Project ALS supported work done by H.W. and M.W.A.

AUTHOR CONTRIBUTIONS
Conceptualization: R.H. and C.N.S.; methodology: R.H.; software: R.H.; formal analysis: R.H.; investigation: R.H., S.S., G.G., M.W.A., J.G.O'R., A.S., D.S. and C.N.S.; resources: H.W., R.H.B., D.S. and C.N.S.; data curation: R.H.; writing—original draft: R.H. and C.N.S.; review and editing: R.H., S.S., M.W.A. and C.N.S.; visualization: R.H.; supervision: C.N.S.; funding acquisition: C.N.S.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Hardiman, O., van den Berg, L.H. & Kiernan, M.C. Clinical diagnosis and management of amyotrophic lateral sclerosis. Nat. Rev. Neurol. 7, 639–649 (2011).
2. Dimos, J.T. et al. Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. Science 321, 1218–1221 (2008).
3. Sternerkötter, J.L., Reinhardt, P. & Schöler, H.R. Investigating human disease using stem cell models. Nat. Rev. Genet. 15, 625–639 (2014).
4. Jessell, T.M. Neuronal specification in the spinal cord: inductive signals and transcriptional codes. Nat. Rev. Genet. 1, 20–29 (2000).
5. Wichterle, H., Lieberam, I., Porter, J.A. & Jessell, T.M. Directed differentiation of embryonic stem cells into motor neurons. Cell 110, 385–397 (2002).
6. Sances, S. et al. Modeling ALS with motor neurons derived from induced pluripotent stem cells. Nat. Neurosci. 16, 542–553 (2013).
7. Arbab, M., Baars, S. & Geisjen, N. Modeling motor neuron disease: the matter of time. Trends Neurosci. 37, 642–652 (2014).
8. Cox, L.E. et al. Mutations in CHMP2B in lower motor neuron predominant amyotrophic lateral sclerosis (ALS). PLoS One 5, e9872 (2010).
9. Kirby, J. et al. Phosphatase and tensin homologue/protein kinase B pathway linked to motor neuron survival in human superoxide dismutase 1-related amyotrophic lateral sclerosis. Brain 134, 506–517 (2011).
10. Sareen, D. et al. Targeting RNA foci in iPS-derived motor neurons from ALS patients with a C9ORF72 repeat expansion. Sci. Transl. Med. 5, 208ra149 (2013).
11. Amoroso, M.W. et al. Accelerated high-yield generation of limb-innervating motor neurons from human stem cells. J. Neurosci. 33, 574–586 (2013).
12. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA 102, 15545–15550 (2005).
13. Devlin, A.C. et al. Human iPSC-derived motorneurons harbouring TARDBP or C9ORF72 ALS mutations are dysfunctional despite maintaining viability. Nat. Commun. 6, 5999 (2015).
14. Wojcik-Stanaszek, L., Gregor, A. & Zalewska, T. Regulation of neurogenesis by extracellular matrix and integrins. Acta Neurobiol. Exp. (Warsz) 71, 103–112 (2011).
15. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559 (2008).
16. Das, M.M. & Svendsen, C.N. Astrocytes show reduced support of motor neurons with aging that is accelerated in a rodent model of ALS. Neurobiol. Aging 36, 1130–1139 (2015).
17. Landrum, M.J. et al. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 42, D980–D989 (2014).
18. Ince, P.G. et al. Molecular pathology and genetic advances in amyotrophic lateral sclerosis: an emerging molecular pathway and the significance of glial pathology. Acta Neuropathol. 122, 657–671 (2011).
19. Schynck, J.C., Talbot, K. & Traynor, B.J. Genetics of sporadic amyotrophic lateral sclerosis. Hum. Mol. Genet. 16 Spec No. 2: R233–R242 (2007).
20. Kiskinis, E. et al. Pathways disrupted in human ALS motor neurons identified through genetic correction of mutant SOD1. Cell Stem Cell 14, 781–795 (2014).
21. Rabin, S.J. et al. Sporadic ALS has compartment-specific aberrant exon splicing and altered cell-matrix adhesion biology. Hum. Mol. Genet. 19, 313–328 (2010).
22. Langfelder, P., Luo, R., Oldham, M.C. & Horvath, S. Is my network module preserved and reproducible? PLoS Comput. Biol. 7, e1001057 (2011).
23. Abdelalim, E.M. & Emara, M.M. Advances and challenges in the differentiation of pluripotent stem cells into pancreatic β cells. World J. Stem Cells 7, 174–181 (2015).
24. Batalov, I. & Feinberg, A.W. Differentiation of cardiomyocytes from human pluripotent stem cells using monolayer culture. Biomark. Insights 10 (suppl. 1), 71–76 (2015).
25. Lapasset, L. et al. Rejuvenating senescent and centenarian human cells by reprogramming through the pluripotent state. Genes Dev. 25, 2248–2253 (2011).
26. Samat, H.B. Clinical neuropathology practice guide 5-2013: markers of neuronal maturation. Clin. Neuropathol. 32, 340–369 (2013).
27. Bender, A.C., Morse, R.P., Scott, R.C., Holmes, G.L. & Lenck-Santini, P.P. SCN1A mutations in Dravet syndrome: impact of interneuron dysfunction on neural networks and cognitive outcome. Epilepsy Behav. 23, 177–186 (2012).
28. Shaw, C. et al. Hierarchical reprogramming systems for direct reprogramming of fibroblasts to neurons. Cell 155, 621–635 (2013).
29. Chanda, S. et al. Generation of induced neuronal cells by the single reprogramming factor ASCL1. Stem Cell Rep. 3, 282–296 (2014).
30. Shi, Y., Kirwan, P. & Livesey, F.J. Directed differentiation of human pluripotent stem cells to cerebral cortex neurons and neural networks. Nat. Protoc. 7, 1836–1846 (2012).
31. Hawrylycz, M. et al. Canonical genetic signatures of the adult human brain. Nat. Neurosci. 18, 1832–1844 (2015).
32. Shi, Y. et al. A human stem cell model of early Alzheimer’s disease pathology in Down syndrome. Sci. Transl. Med. 4, 124ra29 (2012).
33. Studer, L., Vera, E. & Cornachio, D. Programming and reprogramming cellular age in the era of induced pluripotency. Cell Stem Cell 16, 591–600 (2015).
34. Bertens, J. et al. Direct reprogramming of human neurons retain aging-associated transcriptomic signatures and reveal age-related nucleocytoplasmic dysfunctions. Cell Stem Cell 17, 705–718 (2015).
35. Liu, M.L. et al. Small molecules enable neurogrenin 2 to efficiently convert human fibroblasts into cholinergic neurons. Nat. Commun. 4, 2183 (2013).
36. Son, E.Y. et al. Conversion of mouse and human fibroblasts into functional spinal motor neurons. Cell Stem Cell 9, 205–218 (2011).
37. Cady, J. et al. Amyotrophic lateral sclerosis onset is influenced by the burden of rare variants in known amyotrophic lateral sclerosis genes. Ann. Neurol. 77, 100–113 (2015).
38. Kennedy, S.R., Loeb, L.A. & Herr, A.J. Somatic mutations in aging, cancer and neurodegeneration. Mech. Aging Dev. 133, 118–126 (2012).
39. Din, J. et al. Specific electron transport chain abnormalities in amyotrophic lateral sclerosis. J. Neurool. 256, 774–789 (2002).
40. Chung, W.S., Welsh, C.A., Barres, B.A. & Stevens, B. Do glia drive synaptic and cognitive impairment in disease? Nat. Neurosci. 18, 1539–1545 (2015).
41. O’Rourke, J.G. et al. C9orf72 is required for proper macrophage and microglial function in mice. Science 351, 1324–1329 (2016).
42. Albert, R., Jeong, H. & Barabasi, A.L. Error and attack tolerance of complex networks. Nature 406, 378–382 (2000).
43. Langfelder, P., Mischel, P.S. & Horvath, S. When is hub gene selection better than standard meta-analysis? PLoS One 8, e61505 (2013).
44. Wainger, B.J. et al. Intrinsic membrane hyperexcitability of amyotrophic lateral sclerosis patient-derived motor neurons. Cell Rep. 7, 1–11 (2014).
45. Zhang, K. et al. The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. Nature 525, 56–61 (2015).
ONLINE METHODS

Tissue culture and processing. iMNs differentiation was performed as previously described. Briefly, mTeSR1 medium was removed from confluent iPSC cultures and replaced with Iscove’s modified Dulbecco’s medium supplemented with 2% B27–vitamin A and 1% N2 (neural induction medium) for six days. The cells were then Accutase-treated to single cell suspension, and centrifuged in 384-well PCR plates in the presence of Matrigel and Neuralbasal induction medium, which was further supplemented with 0.1 μM all-trans retinoic acid. After suspension culture for 9 d, 1 μM purmorphamine was added to the medium and aggregates were cultured for another 8 d. Thereafter, dissociated aggregates were plated onto polyornithine/laminin-coated coverslips and cultured in Dulbecco’s modified Eagle’s medium (DMEM)/F12 supplemented with 2% B27, 0.1 μM all-trans retinoic acid, 1 μM purmorphamine, 1 μM dibutyryl cyclic adenosine monophosphate, 200 ng/ml ascorbic acid, 10 ng/ml brain-derived neurotrophic factor, and 10 ng/ml glial cell line-derived neurotrophic factor for 7 weeks. HB9::GFP positive MNs and HB9::GFP negative cell samples were processed directly from their published study. Specifically, the samples analyzed in this study were differentiated using 200 ng/ml SHH as a centralizing factor.

Fetal tissue was obtained from the Birth Defects Research Laboratory at the University of Washington under their approved IRB, consent and privacy guidelines. All protocols were performed in accordance with the Institutional Review Board’s guidelines at the Cedars-Sinai Medical Center under the auspice IRB-SCRO Protocol Pro00021505. Upon receipt, tissue samples were renamed D25, D32, D53 and D97 to reflect their estimated gestational stage. Samples arrived as fully or partially intact spinal columns, and spinal columns were opened before shipment. Vertebrae were removed and were partitioned into cervical, thoracic and lumbar sections. Since only spinal columns were received, the exact anatomical reference for each somite could not be accurately determined, therefore the labeling of cervical, thoracic and lumbar sections were estimated.

RNA extraction and processing. Total RNA was isolated from all frozen spinal sections using the RNeasy kit (Qiagen) with on-column DNase digestion. For each fetal spinal cord, equal amounts of total RNA from each section were pooled for expression profiling. RNA from iMNs and from ESC-derived MNs was obtained directly from samples previously reported. Prior to expression profiling, all RNA samples were run through RNeasy kit columns with on-column DNase digestion to produce similarly sized products as the fetal spinal cord samples. Reverse transcription-qPCR was performed using the Real-Time PCR System (Promega A5300), SYBR Select Master Mix (Life Technologies 4472908) and the Bio-Rad CFX384 Real-Time System. Primer sequences are listed in Supplementary Table 3a. Triplicate Ct values for each gene primer were normalized against averaged triplicate GAPDH Ct values and referenced against one transcript of that gene set at 40 Ct. RNA expression profiling was performed on the Affymetrix GeneChip Human Genome U133 Plus 2.0 Arrays at the UCLA microarray core facility. A list of all expression data sets used (either generated or downloaded) in this study is given in Supplementary Table 1a.

Immunofluorescence. For D63, spinal cord was isolated as before and fixed in 4% paraformaldehyde for 48 h. Tissue was placed in 30% sucrose for an additional 24 h, and 25 μm sections were taken using a cryostat (Leica) at ~20 °C and directly mounted on glass slides (Fisher Scientific). Lumbar tissue sections were blocked in PBS containing 5% normal donkey serum (Sigma, D9663) and 0.25% Triton-X for 1.5 h, then transferred to primary antibody solution containing mouse anti-SM1-32 (Covance, SM1-32P-100, 1:1,000) and goat anti-Isl1-1 (R&D, AF1387, 1:500) and incubated overnight at 4 °C. Samples were then incubated for 1 h at room temperature in donkey anti-mouse Alexa Fluor 488 and donkey anti-goat 594 secondary antibodies (Life Technologies, A21202 and A21289, 1:1,000 each). Samples were mounted in Fluormount-G (SouthernBiotech, 1000-01) and acquired at 10× using automated stitching on a Leica DM 6000 microscope. iMNs were fixed in 4% paraformaldehyde, rinsed with PBS, and blocked in 5% donkey serum and 0.2% Triton-X. Samples were incubated overnight at 4 °C in primary antibody solution containing goat anti-ChAT (Millipore, AB141P, 1:200) and mouse anti-SM-32 (Covance, SM1-32P-100, 1:1,000), rinsed in PBS, and incubated for 1 h at room temperature in donkey anti-mouse Alexa Fluor 488 and donkey anti-goat 594 secondary antibodies (Life Technologies, A21202 and A21289, 1:1,000 each). The immunostaining shown in Figure 1a was observed in INM lines 001, 141 and 831, which were derived from three individual patients. The immunostaining shown in Figure 1b was observed in at least 20 sections of the D52, D53, D63 and D97 fetal spinal cords.

Expression data preprocessing. Previously published mRNA microarray expression data from human fibroblasts (n = 2), embryonic stem cells (ESCs; n = 2) and iPSCs (n = 3) were chosen to represent cell types relevant to human somatic cell reprogramming. To represent mature in vivo whole spinal cord, we obtained previously published mRNA microarray expression data from adult spinal cords (n = 8; age range 23 – 53 years, median 36.5 years) and for mature in vivo spMNs, data from laser capture microdissected spMNs from familial ALS and control patients. We also included expression data for spMN and oculomotor neurons from non-ALS patients. These 17 in vivo laser-captured MN specimens came from individuals with an age range from 40 to 80 years (median age, 63 years).

All CEL files considered for use in this study were submitted to http://ArrayAnalysis.org to inspect RNA and microarray hybridization quality, and samples that failed to meet the recommended standards were removed from further analysis. Affymetrix GeneChip Human Genome U133 Plus 2.0 Array CEL files downloaded from GEO as well as produced in this study were normalized together with robust multichip analysis (RMA) using the affy package in Bioconductor R. The GEO accession number for the microarray data produced in this study is GSE75701. Noninformative probe sets were then filtered out using the pvac package in R (ref. 55). This is an unbiased, intuitive method of filtering that is more objective than setting an arbitrary probe set fluorescence intensity cut-off, which penalizes weakly expressed transcripts that are nonetheless important. A probe set can be considered noninformative due to being weakly expressed. In addition, a highly expressed probe set can also be considered noninformative if the probes constituting the probe set are highly discordant. As all probes in a probe set are designed to target the same transcript or a transcript cluster, these probes should largely perform concordantly when gene expression is measured. Probe set discordance may arise due to the post-mortem collection of RNA from human tissue, where heat or exonuclease activity can degrade RNA at either the 5’ or 3’ ends of the transcript. Inclusion of these probes would adversely affect the interpretation of PCA, thereby justifying a method to remove these probe sets. pvac is the proportion of variation accounted for by the first principal component. It measures consistency among probes belonging to the same probe set. The higher the pvac score, the more concordance is observed among probes within a probe set. The pvac package in R generates an empirical distribution of pvac scores for all probe sets in an Affymetrix expression set. It also generates an empirical distribution of pvac scores for all probe sets called absent by the MA55 command on the affy package. These probe sets are postulated to have low concordance among their probes, and therefore their distribution of pvac scores serves as the threshold for filtering discordant and informative probe sets. The application of pvac filtering on our data set serves two purposes. It can remove probe sets that are highly expressed but discordant, ruling out inaccurately quantified transcripts that have possibly been degraded in post-mortem samples. It can also remove weakly expressed probe sets that have a degree of discordance comparable to those probe sets deemed absent by the MA55 algorithm. Conversely, weakly expressed probe sets can also be retained if their probes have a sufficiently high pvac score. These probe sets may have otherwise been discarded if an arbitrary intensity threshold was applied to the expression set.

Filtered probe sets were then annotated with their HGNC symbols using the Affymetrix annotation file for the GeneChip Human Genome U133 Plus 2.0 Array Release 35, summarized at the gene level by taking the probe set with the maximum expression value to represent the maximal transcriptional activity associated with that gene, and the resulting 10,605 gene expression values were quantile normalized on the linear scale using the normalize.quantiles function in the preprocessCore package. The processed Affymetrix Human Genome U133 Plus 2.0 Array expression values for 10,605 genes are listed in Supplementary Table 2a. For RNA-seq data from mtSOD1 and control iMNs, normalized counts for each sample as they were provided through GEO accession series GSE54409 were compiled into an expression table for 14,422 Ensembl identifiers, re-annotated with HGNC symbols based on gene symbols selected through Ensembl BioMart, and summarized at the gene level by taking the transcripts with the maximum expression value to represent the maximal transcriptional activity associated with each gene. This list of genes was intersected with the list...
of 10,605 genes used for the mtSOD1 and control spMN samples to produce the expression matrix for 8,830 genes listed in Supplimentary Table 4.

For the sALS data set, CEL files downloaded from GEO accession series GSE18920 and normalized with RMA using the oligo package in Bioconductor R with the pd.hex1.0.st.v2 array library. Expression values for transcript clusters were re-annotated to Affymetrix Human Genome U133 Plus 2.0 Array probe sets and summarized at the gene level by taking the probe set with the maximum expression value to represent the maximal transcriptional activity associated with that gene, and the resulting gene expression values were quantile normalized on the linear scale using the preprocessCore package. This produced the expression table for 15,614 genes listed in Supplementary Table 5a. Power analysis was not calculated before determining a justified sample size. Composite data sets were created to compare as many fetal and adult spinal tissues as possible, limited by the available data generated on a common microarray platform to minimize batch effects. For network modeling using WGCNA, data sets with similar ratios of sample types, such as pluripotent, fetal and adult cells or control to sporadic ALS tissue, were composed to provide equal representation of the relevant cellular states. There was no randomization of the data assigned to classes, and no blinding when performing data collection and analysis of the experiments. For analysis comparing mtSOD1 to control spMNs in Figure 4a – c, of the 5 control spMNs were selected for analysis based on the highest Pearson correlation among all 5 samples to match the sample size of mtSOD1 spMNs (n = 3).

Gene expression analysis. Pearson correlations, statistical tests and multiple testing corrections were performed in R. Unsupervised hierarchical clustering was performed in Cluster 3.0 and heat maps were visualized using Java Treeview. PCA was performed in Cluster 3.0 or R. The signed values for principal component coordinates of samples and gene loadings were reversed as necessary to maintain consistency across analyses. PCA plots as well as scatter, density and box plots were visualized in R with the basic R plotting tools. To generate univariate linear models, the lm function was used in R. Principal component coordinates or log-transformed principal component coordinates were used as the dependent variable in order to satisfy the assumption of normally distributed standardized residuals. Sample traits were used as independent variables. P-values determined from the F-statistic were Bonferroni-corrected based on the number of principal components tested against the number of sample traits. Venn and Chow-Ruskey diagrams were generated using the Vennerable package in R. GSEA was performed on preranked lists generated using PCA gene loadings with 1,000 permutations of the gene set to generate a null distribution. GO enrichment was performed using DAVID, and genes represented on the Affymetrix GeneChip Human Genome U133 Plus 2.0 Array (iMN) and Human Exon 1.0 ST Array (sALS) were called significant and kept for GO term enrichment analysis through DAVID, as shown in Supplementary Tables 2d (iMN) and 5e (sALS). Nonlinear associations between module eigengenes and sample traits or principal component coordinates and sample traits were tested using Spearman’s method for age and postmortem interval; the Kruskal–Wallis test for the categorical variables tissue type, sex, study, disease status, site of ALS onset and sample processing; and Cox regression for ALS disease duration. With one exception, these tests (subjected to multiple testing corrections) revealed no significant associations that were not also shown to be significant using linear models. The exception was that, by Spearman’s but not Pearson’s correlation, the paleturquoise module antecorrelates with age. This observation did not impact the main interpretations, and therefore we based our discussions on results from the linear correlations.

We note that the gene sets and pathways enriched by using principal component gene loadings and GSEA serve as confirmatory indicators of sensible pathways describing the differences between samples, while the pathways enriched by using DAVID on modules serve as more discovery and hypothesis generators. WGCNA is more sensitive than PCA, because WGCNA accounts for subtractor- but consistent gene-to-gene co-expression relationships that need not have larger gene loadings per principal component to be detected. These larger gene loadings are rather rewarded in GSEA at the cost of failing to detect significantly correlated gene networks with lower gene loadings in embryonic spMN development or maturation that nonetheless may also be physiologically important.

To identify the 20 key marker genes, fibroblasts were removed from the iMN data set and PCA was applied to the resulting ‘training data set’ of 10,605 genes in 41 samples to produce 41 principal components. We considered three properties of each gene to select for the most effective markers of spMN maturation and embryonic development: gene significance, intramodule membership and preferential gene loading. As an example, HOXB8 has a gene significance value of 0.78 along embryonic spMN development (PC3). However, the gene with the highest gene significance along embryonic spMN development is CXXC4 (0.92). Among all the genes classified as PC3pos, CXXC4 has the highest gene significance score (0.92/0.92 = 1.00), and HOXB8 has a gene significance score of 0.78/0.92 = 0.85. A gene with high intramodule membership best represents the expression of all genes within the module to which it is assigned. HOXB8 belongs to the paleturquoise module and has an intramodule membership of 0.94. However, CAMK2N1, which belongs to the darkolivegreen module, has an intramodule membership of 0.97, which is the highest intramodule membership value among all the genes classified into the category PC3pos. Therefore, HOXB8 has an intramodule membership score of 0.94/0.97 = 0.97. Among all 10,605 genes in the iMN data set, HOXB8 has the highest gene loading along PC3 (0.06). However, HOXB8 may also have high gene loadings in the other 40 principal components. Therefore, to quantify to what extent HOXB8 preferentially contributes to PC3, the average gene loading for HOXB8 in the other 40 principal components is subtracted from its gene loading in PC3. Before performing this operation, the gene loadings of HOXB8 in all principal components were squared to convert them into positive values. These values needed to be scaled by the percentage contribution of each principal component to the total variance. Dividing each eigenvalue by the sum of all 41 eigenvalues and multiplying by 100 produced the percentage contribution of each eigenvalue to the total variance, and the squared gene loading for each component was multiplied by this percentage contribution. Because principal components 35, 36, etc. have lower eigenvalues,
a large gene loading for HOXB8 in these components are less detrimental to the preferential gene loading score. As observed in Supplementary Table 2i, HOXB8 has the highest preferential PC3pos gene loading score, because it has the highest gene loading in PC3 and relatively low gene loadings in all other principal components. Normalizing the scores by the maximum value in each category for each of the gene properties evenly weights their contribution in determining the best marker genes for each of the four categories PC1pos, PC1neg, PC3pos and PC3neg. The total gene scores were summed from all three properties and ranked from largest to smallest. These values are listed in Supplementary Table 2f–i.

The ROCR package was used to calculate and plot the Receiver Operator Characteristics and Area Under the Curve analyses. Predictions were based on correct classification of pluripotent stem cells, fetal-like cells and adult spinal cord cells. Tests were performed using either (i) the Pearson correlation between the expression values in the test data set and the median expression values among those cell types in the training data set or (ii) the negative threshold below PC1 coordinate for pluripotent stem cells, positive threshold above PC1 coordinate for adult spinal cord cells, and positive threshold above PC7 coordinate (6,640 genes) or PC2 coordinate (20 genes) for fetal-like cells. PC coordinates were generated by performing PCA with 6,640 or 20 genes on all samples from both the training and test data sets, but the prediction table used to generate the ROC curve and AUC values are without the training data set.

Hypergeometric tests for module enrichment were performed using custom scripts applying the dhyper function in R, and the resulting hypergeometric P-values were corrected with the Benjamini-Hochberg method. For ClinVar enrichment in the iMN modules, 22,608 genes represented on the Affymetrix GeneChip Human Genome U133 Plus 2.0 Array were used as the background number of genes. For iMN module enrichment in the sALS modules, 21,279 genes represented on both the Affymetrix GeneChip Human Genome U133 Plus 2.0 and Human Exon 1.0 ST Arrays were used as the background number of genes.

Module preservation was performed using the WGCNA package in R (ref. 22). 8,715 overlapping genes were represented in both the iMN and sALS expression data sets and were thus used in the preservation analysis. The expression values for these 8,715 genes were loaded as two data tables into one multiExpr object: one data table of expression values from the iMN data set, and the other listing the module color assignment for the 8,715 genes when WGCNA was performed using 10,605 genes in the iMN data set, and the other listing the module color assignment for the 8,715 genes when WGCNA was performed using 15,614 genes in the sALS data set. The command modulePreservation was called on the multiExpr and multiColor objects, first by using the iMN data set as the reference and the sALS data set as the test, and then vice versa. 200 permutations were used to randomly assign genes to module colors in the test data. This generates a distribution of preservation statistics under the null hypothesis that no modules are preserved across the reference and test data sets. z-statistics are defined for preservation statistics such as module density and module connectivity. The z-summary summarizes all individual z-statistics. Since z-summary is heavily influenced by module size, the median rank measure for each module is more appropriate when comparing the relative preservation among differently sized modules.

Cytoscape 3.0 was used to visualize network topology for each of the four classes of human networks depicted in Figure 6f–i. For each class, the 99th percentile of edge weights was filtered, followed by filtering for source nodes with at least 30 edges with target nodes. Edge and node attributes as well as graphical layout constraints are described in the figure legends.

Data availability. The data that support the findings of this study are available in the Supplementary Figures, Supplementary Tables and Supplementary Methods Checklist. The programming scripts written for the analysis performed in this study are compiled into a text document and are available at https://github.com/ritchieho/NN-RS551188_R_scripts.git.

A Supplementary Methods Checklist is available.

51. Chinn, M.H. et al. Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. Cell Stem Cell 5, 111–123 (2009).
52. Maherali, N. et al. A high-efficiency system for the generation and study of human induced pluripotent stem cells. Cell Stem Cell 3, 340–345 (2008).
53. Roth, R.B. et al. Gene expression analyses reveal molecular relationships among 20 regions of the human CNS. Neurogenetics 7, 67–80 (2006).
54. 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).
55. Lu, J., Kems, R.T., Peddada, S.D. & Bushel, P.R. Principal component analysis-based filtering improves detection for Affymetrix gene expression arrays. Nucleic Acids Res. 39, e86 (2011).
56. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. Stat. Appl. Genet. Mol. Biol. 4, e17 (2005).