Gene co-expression network analysis in human spinal cord highlights mechanisms underlying amyotrophic lateral sclerosis susceptibility

Jerry C. Wang[1], Gokul Ramaswami[1], Daniel H. Geschwind[1,2,3,4*]

Affiliations

[1] Program in Neurogenetics, Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, United States of America

[2] Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, United States of America

[3] Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, United States of America

[4] Institute for Precision Health, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, United States of America

*. Correspondence should be addressed to D.H.G: dhg@mednet.ucla.edu

Abstract

Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disease defined by motor neuron (MN) loss. Multiple genetic risk factors have been identified, implicating RNA and protein metabolism and intracellular transport, among other biological mechanisms. To achieve a systems-level understanding of the mechanisms governing ALS pathophysiology, we built gene co-expression networks using RNA-sequencing data from control human spinal cord samples, identifying 13 gene co-expression modules, each of which represents a distinct biological process or cell type. Analysis of four RNA-seq datasets from a range of ALS disease-associated contexts reveal dysregulation in numerous modules related to ribosomal function, wound response, and leukocyte activation, implicating astrocytes, oligodendrocytes, endothelia, and microglia in ALS pathophysiology. To identify potentially causal processes, we partitioned heritability across the genome, finding that ALS common genetic risk is enriched within two specific modules, SC.M4, representing genes related to RNA processing and gene regulation, and SC.M2, representing genes related to intracellular transport and autophagy and enriched in oligodendrocyte markers. Top hub genes of this module include ALS-implicated risk genes such as KPNA3, TMED2, and NCOA4, the latter of which regulates ferritin autophagy, implicating this process in ALS pathophysiology. These unbiased, genome-wide analyses confirm the utility of a systems approach to understanding the causes and drivers of ALS.

Introduction

Amyotrophic lateral sclerosis (ALS) is an adult-onset neurodegenerative disease defined by motor neuron (MN) loss and results in paralysis and eventually death 3-5 years following initial disease presentation[1]. In 5-10% of cases, patients exhibit a Mendelian pattern of inheritance and the disease segregates within multigenerational affected families (familial ALS)[2]. However, in 90% of cases, the underlying cause of ALS is unknown and disease occurrence is classified as sporadic (sporadic ALS)[2]. Most genetic risk in ALS is polygenic, indicating a complex genetic architecture[3]. Given this complex genetic architecture with contributions of rare and common alleles, we took a systems level approach to understand the gene expression signatures associated with ALS pathophysiology.

Weighted gene co-expression network analysis (WGCNA)[4] has been a powerful method for organizing and interpreting wide-ranging transcriptional alterations in neurodegenerative diseases. In Alzheimer's disease (AD), transcriptomic and proteomic studies have revealed a pattern of dysregulation implicating immune- and microglia-specific processes[5,6]. More broadly across tauopathies including AD, frontotemporal dementia (FTD), progressive supranuclear palsy (PSP), molecular network analyses have identified gene modules and microRNAs mediating neurodegeneration, highlighting its utility in prioritizing molecular targets for therapeutic interventions[7–10]. In ALS, previous studies have studied the transcriptional landscape in sporadic and C9orf72 (familial) ALS post-mortem brain tissues, finding a pattern of dysregulation in gene expression, alternative splicing, and alternative polyadenylation that disrupted RNA processing, neuronal functioning, and cellular trafficking[11]. However, the generalizability of these findings is limited due to relatively small sample sizes (n=26). Numerous studies have previously examined the transcriptional landscape of the spinal cord, which hosts axons from the upper motor neurons (UMN) and somae of the lower motor neuron

64 (LMN) and is the location of LMN degeneration in ALS[12–17]. However, a rigorous interrogation of
65 gene co-expression networks in the spinal cord has not yet been fully explored.
66
67 To investigate gene expression signatures associated with ALS in the spinal cord, we generated
68 gene co-expression networks using RNA-sequencing of 62 neurotypical human cervical spinal
69 cord samples from the GTEx consortium[18]. We characterized these co-expression modules
70 through gene ontology (GO) enrichment analysis, which reveals that the modules represent a
71 diverse range of biological processes. To identify modules enriched with ALS genetic risk
72 factors, we queried enrichment of rare, large effect risk mutations previously described in the
73 literature, as well as common genetic variation[19,20]. The second approach is genome-wide and
74 therefore unbiased, which focuses on the enrichment of small effect common risk variants
75 identified from a Genome Wide Association Study (GWAS) of ALS[21]. By partitioning SNP based
76 heritability and performing enrichment testing for genes with rare mutations, we find significant
77 overlap of genes harboring rare and common variation in SC.M4, which represents genes
78 involved in RNA processing and epigenetic regulation. We also find enrichment of ALS common
79 genetic risk variants in SC.M2, which represents genes involved in intracellular transport,
80 protein modification, and cellular catabolic processes. We extend these results by
81 demonstrating the dysregulation of these modules in data from multiple published human ALS
82 patient and animal model tissues[12,13,22,23]. We identify modules related to ribosomal function,
83 wound response, and leukocyte activation that are dysregulated across this diverse set of ALS-
84 related datasets. The set of genes underlying these modules provides a rich set of targets for
85 exploring new mechanisms for therapeutic development.
86
87 **Results**
88 *Co-expression network analysis as a strategic framework for elucidating ALS susceptibility*
89

90 We reasoned that because ALS preferentially affects spinal cord and upper motor neurons, co-
91 expression networks from spinal cords of control subjects would provide an unbiased context for
92 understanding potential convergence of ALS risk genes. We used the Genotype-Tissue
93 Expression (GTEx) Project[18] control cervical spinal cord samples to generate a co-expression
94 network as a starting point for a genome wide analysis to identify specific pathways or cell types
95 affected by ALS genetic risk. First, we identified gene co-expression modules and assigned
96 each module a biological, as well as cell-type identity[24,25]. Next, we leveraged multiple datasets
97 identifying ALS genetic risk factors[19–21,26–28] to comprehensively assess which pathways and
98 cell-types are susceptible to ALS genetic risk. Finally, we compared these predictions with
99 human ALS and animal model datasets[12,13,22,23] to develop a mechanistic understanding of ALS
100 etiology and progression (**Fig. 1**).
101
102 *Generation of human spinal cord gene co-expression network*
103 We utilized RNA-sequencing data from the GTEx consortium to generate a co-expression
104 network of human spinal cord (**Methods**). Co-expression networks are highly sensitive to
105 biological and technical confounders[25], which we controlled for by calculating the correlation of
106 biological and technical covariates with the top expression principal components (PCs)
107 (**Methods, Supplementary Fig. S1a-c**). We identified 11 technical and biological covariates:
108 seqPC1-5, RNA integrity number (RIN), Hardy's death classification scale, age, sampling
109 center, ethnicity, and ischemic time as significant drivers of expression (**Supplementary Fig.**
110 **S1a**). We regressed out the effect of these covariates using a linear model (**Methods**) and
111 verified that the regressed expression dataset was no longer correlated with technical and
112 biological covariates (**Supplementary Fig. S1b**).
113
114 We applied WGCNA to construct a co-expression network (**Fig. 2a, Supplementary Fig. S2;**
115 **Supplementary Data 1; Methods**) identifying 13 modules. To verify that the modules were not

116 driven by confounding covariates, we confirmed that the module eigengenes were not
117 correlated with the major technical and biological covariates (**Supplementary Fig. S1d**). We
118 also found that randomly sampled genes from within each module were significantly more inter-
119 correlated when compared to a background set, verifying the co-expression of genes within
120 each module (**Supplementary Fig. S1e**). To biologically annotate our co-expression modules,
121 we performed gene ontology (GO) term enrichment and identified the top hubs (**Fig. 2c-g, 3c-d,**
122 **Supplementary Fig. S3**). GO analysis revealed that co-expression modules represented
123 coherent biological processes that we categorized into five distinct groupings (**Fig. 2b**),
124 encompassing morphogenesis and development, epigenetics and gene regulation, neuronal
125 signaling, immune activation, and intracellular transport, metabolism, and sensing functions.
126 Most modules were significantly enriched for markers of specific cell types, suggesting that they
127 represent gene co-expression within major cell classes, including SC.M1 (astrocytes), SC.M2
128 (oligodendrocytes), SC.M3 (endothelial cells), SC.M9 and SC.M11 (neurons), SC.M10
129 (ependymal), and SC.M12 and SC.M13 (microglia) (**Supplementary Fig. S4**). These groupings,
130 both with regards to cell type and molecular pathways, represent diverse biological functions
131 that play key roles in human spinal cord biology.
132
133 *Enrichment of ALS genetic risk in intracellular transport/autophagy module SC.M2 and RNA*
134 *processing/gene regulation module SC.M4*
135 To address how these modules fit within the context of known ALS susceptibility factors, we
136 assessed whether a literature-curated set of well-known, large effect size, familial ALS risk
137 genes[19,20] were enriched within our co-expression modules (**Fig. 3a**). We found a nominally
138 significant enrichment with SC.M4 (OR = 2.52, P = 0.0124). However, the FDR adjusted p-value
139 was not significant (FDR= 0.174), likely due to the small number of familial ALS risk genes
140 identified to date. SC.M4 was highly enriched with GO terms related to RNA processing and
141 epigenetic regulation (**Fig. 3c**), consistent with the known role of several ALS risk genes in RNA

142 metabolism and function. To prioritize interacting partners to ALS risk genes found in SC.M4,
143 we leveraged a direct protein-protein interaction (PPI) network using the top 300 hub genes of
144 SC.M4, finding a significant network that includes ALS risk genes such as TARDBP and
145 EWSR1 (**Fig. 3e**, **Methods**, p = 0.001).
146
147 The small number (~50) of large effect size risk genes identified to date in ALS makes the
148 detection of strong enrichments difficult. To complement the rare risk gene enrichment
149 analyses, we performed an unbiased genome-wide analysis, partitioning the common SNP-
150 based heritability across the co-expression modules using stratified LD score-regression[29]
151 based on summary statistics from the most recent GWAS of ALS (ALS n= 20,806; CTL n =
152 59,804)[21] (**Fig. 3b**). We found significant enrichment of ALS common risk variants in the SC.M2
153 (FDR = 0.047) and SC.M4 modules (FDR = 0.041). SC.M2 was enriched with GO terms related
154 to intracellular transport, protein modification, and cellular catabolic processes (**Fig. 3d**).
155 Additionally, SC.M2 was significantly enriched with oligodendrocyte markers (**Supplementary**
156 **Fig. S4,** Oligodendrocytes_mature: OR = 5.3, FDR < 10e-5; Oligodendrocytes_myelinating: OR
157 = 2.9, FDR < 0.05), while SC.M4 showed no specific cell type enrichment. Generation of a direct
158 protein-protein interaction (PPI) network using the top 300 hub genes of SC.M2 reveals a
159 significant network that highlights SC.M2 hub genes that interact with known ALS risk genes
160 such as TMED2 and KPNA3[27] (**Fig. 3f**, **Methods**, p = 0.018).
161
162 As a comparison, we tested for the enrichment of common risk variants from Inflammatory
163 Bowel Disorder (IBD)[30] (N=86,640) across modules (**Fig. 3b**). We found a significant enrichment
164 of IBD common risk variants (FDR = 0.047) in module, SC.12, which is a microglial/immune
165 module, consistent with the known disease biology of IBD[31]. As a negative control, we
166 performed enrichments of common risk variants from Autism Spectrum Disorder (ASD)[32,33]
167 (N=15,954) and found no enrichment in any module. Further, when we compared with

168 Alzheimer's Disease (AD)[34] (N=455,258), Frontotemporal Dementia (FTD)[35] (N=12,928), and
169 Progressive Supranuclear Palsy (PSP)[36] (N=12,308) (**Fig. 3b**), we also found no significant
170 enrichments for these diseases' common risk variants, demonstrating the specificity of the ALS
171 GWAS enrichment for modules SC.M2 and SC.M4. Overall, these genetic enrichment analyses
172 implicate dysfunction in RNA processing and intracellular transport, including in
173 oligodendrocytes, as putative drivers of ALS pathophysiology in the human spinal cord.
174
175 *Enrichment of ALS genetic modifiers in ribosomal-associated module SC.M6*
176 As an orthogonal approach to assess which genes and pathways are mediators of ALS
177 pathophysiology, we leveraged genetic screen datasets from three studies of modifiers of
178 C9orf72 and FUS toxicity in yeast[26–28]. We assessed our human spinal cord modules for
179 enrichment with modifiers of C9orf72 and FUS toxicity identified from these screens. We found
180 significant enrichment of C9orf72 and FUS toxicity suppressors in SC.M6 (**Fig. 4a**), a module
181 representing genes involved in ribosomal function (**Fig. 2e**), but without any cell type specific
182 enrichment (**Supplementary Fig. S4**). This finding is consistent with previous observations that
183 modifiers of Glycine-Arginine$_{100}$ (GR$_{100}$) DPR and FUS toxicity are enriched with gene ontology
184 terms related to ribosomal biogenesis[26,28].
185
186 We reasoned that SC.M6 could be leveraged to prioritize highly interconnected hub genes
187 within the module that may warrant additional scrutiny as points of convergence for C9orf72 and
188 FUS pathology. To address this, we first identified the interacting partners for the top 100
189 SC.M6 hub genes by generating an indirect protein-protein interaction (PPI) network (**Methods**).
190 We identified 30 modifiers of C9orf72 and FUS toxicity that were present in this indirect PPI
191 network, either as hubs or as interacting partners (**Fig. 4b**)[26–28]. These modifiers include
192 ribosomal proteins (RPL and RPS-related genes), intracellular transporters of ribosomal
193 components (TNPO1), ribosomal assembly proteins (NOB1), and RNA-binding proteins

194 (HABP4) (**Fig. 4b**), indicating specific aspects of ribosomal dysfunction shared between C9orf72
195 and FUS toxicity. Indeed, previous studies[26] have suggested that the overlap in modifiers of
196 C9orf72 and FUS toxicity could be due to the sequence similarity between the
197 arginine/glycine/glycine (RGG) domain in FUS and the GR_{100} sequence expressed by
198 C9orf72[37,38].
199
200 *Dysregulation of immune-related modules in postmortem ALS spinal cord*
201 Until this point, we had assessed co-expression modules for their relevance to ALS based on
202 enrichment of genetic risk factors. To assess whether the co-expression modules identified
203 using the GTEx dataset were directly dysregulated in ALS patients, we processed bulk RNA-
204 sequencing data from a previous study of postmortem control and ALS cervical spinal cord[13].
205 We corrected the data for potential confounders such as age, ethnicity, sex, and sequencing
206 variability and assessed module preservation (**Supplementary Fig. S5a-b**, **Methods**). We
207 identified 10 of the 13 modules generated from GTEx as preserved in the postmortem control
208 and ALS cervical spinal cord ($Z_{summary} > 2$) with four (SC.M2, SC.M4, SC.M5, and SC.M8)
209 strongly preserved ($Z_{summary} > 10$). The remaining three modules (SC.M9, SC.M11, and SC.M10)
210 were not preserved ($Z_{summary} < 2$) (**Supplementary Fig. S5c**).
211
212 We adopted two approaches to assess which modules were dysregulated in ALS in the human
213 spinal cord. The first assessed enrichments of differentially expressed genes (DEGs) between
214 postmortem control and ALS cervical spinal cord samples[13] (**Methods**). We found a significant
215 enrichment of DEGs upregulated in cervical spinal cord from ALS patients in SC.M12 (FDR =
216 5.92e-39) and of downregulated DEGs in SC.M3 (FDR = 0.00152) (**Fig. 5a**). SC.12 is highly
217 enriched with microglial markers and GO terms related to immune response (**Fig. 2f,**
218 **Supplementary Fig. S4a**), whereas SC.M3 is highly enriched with endothelial markers and GO
219 terms related to inflammation and morphogenesis (**Fig. 2c, Supplementary S4a**).

220

221 For our second approach, we assessed if the spinal cord modules were themselves differentially
222 expressed between ALS and control samples by looking at the relationship between the
223 eigengenes (first principle component; **Methods**) of each module and disease status
224 (**Supplementary Fig. S5d**). We found that the SC.M1 eigengene was significantly upregulated
225 in ALS samples. SC.M1 was enriched with astrocyte markers and GO terms related to wound
226 response (**Supplementary Fig. S3a, S4a**). Overall, our analyses identify a marked upregulation
227 of neuro-immune cells in the human post-mortem spinal cord, including astrocytes and
228 microglia, similar to previous reports[13,22]. However, we note that genes differentially expressed
229 in postmortem ALS spinal cord are not enriched in causal genetic risk, suggesting that these
230 changes are not causal drivers.

231

232 *Dysregulation of immune and neuronal related modules in dissected human motor neurons*
233 *primed for degeneration*

234 A major disadvantage of utilizing postmortem human tissues is that the molecular pathology is
235 indicative of late-stage disease. In particular, immune hyperactivation and infiltration is
236 consistently seen in neurodegenerative disorders including ALS. However, it is unclear if these
237 processes are disease causing or merely a reaction to neurodegenerative cell death[39]. To
238 identify early-stage differences in human spinal cord potentially representative of causal factors,
239 we leveraged a previously published dataset of laser capture micro-dissected motor neurons
240 deemed to be primed for degeneration[22]. In this study, the authors isolated motor neurons from
241 ALS patients' postmortem spinal cord, but in regions that did not exhibit any signs of
242 degeneration at the time of death. We processed and corrected the data for potential
243 confounders such as age, gender, postmortem interval, RNA integrity, and sequencing
244 variability[22] (**Supplementary Fig. S6a-b**). We identified 9 of the 13 modules generated from
245 GTEx as preserved in the micro-dissected motor neuron dataset ($Z_{summary} > 2$) with two (SC.M3

246 and SC.M6) strongly preserved ($Z_{summary}$ > 10). The remaining four modules (SC.M2, SC.M4,
247 SC.M8, and SC.M10) were not preserved ($Z_{summary}$ < 2) (**Supplementary Fig. S6c**).
248
249 We assessed the enrichment of DEGs between ALS and control micro-dissected motor neurons
250 in our modules. We found enrichment of up-regulated DEGs within SC.M1, SC.M3, SC.M6, SC.
251 M12, and SC.M13, modules related to astrocyte, endothelial, and microglial functions
252 (**Supplementary Fig. S4a**). We found enrichment of down-regulated DEGs within SC.M7 and
253 SC.M9 (**Fig. 5b**), two modules related to synaptic signaling (**Fig. 2d, Supplementary Fig. S3b**),
254 possibly marking degeneration at early stages of ALS. We also assessed the relationship
255 between the module eigengene and disease status, finding significant up-regulation of SC.M1,
256 SC.M3, SC.M5, SC.M6, SC.12, and SC.M13 in the ALS samples (**Supplementary Fig. S6d**).
257 Despite performing laser-capture micro-dissection of motor neurons, our results indicate that the
258 predominant signal from this dataset highlights the infiltration of neuro-immune cells, including
259 astrocytes and microglia in the progression of ALS pathophysiology.
260
261 *Dysregulation of a ribosome-related module in iPSC-derived motor neurons from ALS8 patients*
262 To gain further insights into processes dysregulated in motor neurons and avoid the pro-
263 immune phenotypes observed in previous postmortem studies[13,22], we leveraged a RNA-seq
264 dataset generated from iPSC-derived motor neurons from ALS8 patients[23]. In this study, iPSC-
265 derived motor neurons were differentiated from five patients carrying pathogenic variants in the
266 VAPB gene and three unaffected control individuals. We processed and corrected this dataset
267 for potential confounders such as age, sex, sampled tissue, and sequencing variability
268 (**Supplementary Fig. S7a-b**). We identified 10 of the 13 modules generated from GTEx as
269 preserved in the iPSC-derived motor neuron dataset ($Z_{summary}$ > 2) with two (SC.M7 and SC.M9)
270 strongly preserved ($Z_{summary}$ > 10). The remaining three modules (SC.M5, SC.M12, and SC.M13)
271 were not preserved ($Z_{summary}$ < 2) (**Supplementary Fig. S7c**).

272

273 We assessed the enrichment of DEGs between iPSC-derived MNs from severe ALS versus
274 control individuals in our modules. We found modest enrichment of up-regulated DEGs within
275 SC.M1, SC.M7, SC.M9, and SC.M11 (OR = 1.77, 1.76, 2.09, 3.84 respectively), modules
276 related to astrocyte and neuronal functions (**Fig. 5c, Supplementary Fig. S4**). We found a
277 strong and pronounced enrichment of down-regulated DEGs within SC.M6 (OR = 27.28), a
278 ribosomal-related module (**Fig. 2e, 5c**). We also assessed the relationship of module eigengene
279 to disease status, revealing up-regulation of SC.M2 and SC.M7 and down-regulation of SC.M1
280 (**Supplementary Fig. S7d**). By analyzing gene expression in iPSC-derived motor neurons from
281 ALS patients, we identified striking downregulation of ribosomal function, represented by
282 SC.M6, as a prominent dysregulated feature specific to MNs.

283

284 *Dysregulation of immune and neuronal-related modules in SOD1-G93A (ALS) mice*
285 Finally, as an orthogonal approach to identify temporal differences in ALS, we leveraged a
286 spatial transcriptomics dataset from spinal cord of SOD1-G93A mice over four time points
287 during disease progression[12,40,41]. To assess which modules were dysregulated over the
288 progression of disease, we performed an enrichment analysis between the human spinal cord
289 modules and mouse DEGs stratified by time point (**Methods**). Consistent with our findings in
290 postmortem tissues (**Fig. 5a-b, Supplementary Fig. S6e**), immune response modules (SC.M12
291 and SC.M13) were highly enriched with up-regulated DEGs as early as postnatal day 70, which
292 is the time of clinical disease onset in this model (**Fig. 5c, 2f, Supplementary Fig. S3f**). We
293 also noted a strong enrichment of down-regulated DEGs in neuronal signaling modules (SC.M7,
294 SC.M9, and SC.M11) (**Supplementary Fig. S3b, Fig. 2d, Supplementary Fig. S3e**) by
295 postnatal day 120 (p120 - end-stage ALS), which is consistent with motor neuron degeneration.
296 Interestingly, we found enrichment of down-regulated DEGs in SC.M2 at p100 (symptomatic
297 ALS), providing evidence that SC.M2 is dysregulated at symptomatic stages of ALS in this

298 model system (**Fig. 5c, Supplementary Fig. S9**). We also noted an enrichment of up-regulated
299 DEGs in SC.M6 at p120 (end-stage ALS), suggesting dysregulation of SC.M6 in the end-stages
300 of ALS in this model system (**Fig. 5c, Supplementary Fig. S8**). Overall, enrichment analyses in
301 SOD1-G93A (ALS) mice highlight immune and neuronal-related modules as key disrupted
302 pathways in the spinal cord and implicates the dysregulation of SC.M2, a module enriched with
303 ALS genetic risk, and SC.M6, a module enriched with ALS genetic modifiers, in later stages of
304 disease progression (**Fig. 5e**).
305
306 **Discussion**
307 To help elucidate the biological mechanisms underlying ALS pathophysiology in an unbiased
308 genome-wide manner, we generated co-expression networks using RNA-sequencing of control
309 human cervical spinal cord samples from the GTEx Consortium. We identified 13 total co-
310 expression modules each representing distinct aspects of biology in the control spinal cord
311 including inflammation, neurotransmission, RNA processing, and cellular metabolism, as well as
312 major cell classes. We find that ALS genetic risk factors are enriched in SC.M4, a co-expression
313 module representing genes involved in RNA processing and epigenetics and SC.M2, which is
314 enriched in oligodendrocyte markers and genes involved in intracellular transport, protein
315 modification, and cellular catabolic processes.
316
317 It is notable that genes within both SC.M2 and SC.M4 have previously been shown to cause
318 rare Mendelian forms of ALS or have been strongly linked with risk. For example, SC.M4, which
319 is enriched in RNA processing, harbors RNA binding proteins mutated in ALS including TDP-43
320 and FUS/TLS, which are abnormally aggregated and mis-localized in ALS affected neurons,
321 leading to the loss of normal RNA binding function with incitement of neurotoxicity[42]. One of the
322 top hubs of the SC.M4 module is PTPN23, a non-receptor-type tyrosine phosphatase that is a
323 key regulator of the survival motor neuron (SMN) complex, which helps assemble small

324 ribonucleoproteins particles (snRNPs) and mediates pre-mRNA processing in motor neurons[43].
325 Furthermore, two other top hubs in SC.M2, KPNA3 and TMED2, are nucleocytoplasmic
326 transporters that act as modifiers of dipeptide repeat (DPR) toxicity in cases of C9orf72
327 mutations[27]. NCOA4, another top hub in SC.M2, is a key mediator of ferritinophagy and has
328 been implicated in neurodegeneration[44,45]. Taken together, the presence of ALS-linked risk
329 genes as hubs in SC.M2 and SC.M4 further underscores the importance of aberrant RNA
330 processing[42] and intracellular transport[46] as processes likely driving ALS pathophysiology.
331
332 We also find enrichment of genetic modifiers of ALS in SC.M6, indicating the importance of
333 ribosomal proteins and elongation factors (**Fig. 4B**) in mitigating ALS pathophysiology. These
334 factors have been suggested to act by reducing translation of toxic dipeptide repeat (DPRs)
335 expressed off of the GGGGCC hexanucleotide repeat expansion found in C9orf72[26]. We found
336 SC.M6 to be up-regulated with disease status in the SOD1-G93A mouse and dissected human
337 motor neurons primed for neurodegeneration (**Fig. 5b, d**), datasets with predominant glial
338 infiltration. Interestingly, we found a striking down-regulation of SC.M6 in iPSC-derived motor
339 neurons from ALS8 patients (**Fig. 5c**), a dataset without contaminating glial signal. This
340 suggests that role of this co-expression module is complex and cell-type dependent in ALS.
341
342 Across three of four ALS tissue and model datasets analyzed, we noted the presence of a
343 strong pro-immune phenotype (**Fig. 5**). While immune dysregulation is a key molecular feature
344 of the spinal cord by mid- to late-stage ALS, it remains unclear whether this disease signature is
345 primary or secondary[47,48]. Our analyses focused on ALS genetic risk factors, which serve as
346 causal anchors to identify, in an unbiased manner, specific modules that implicate processes
347 involved at the early stages of ALS pathophysiology. Furthermore, immune dysregulation is a
348 common feature in post-mortem tissue from other neurodegenerative and neuropsychiatric
349 disorders[8,49,50]; however, the enrichment of genetic risk within upregulated immune genes is

350 complex. For neuropsychiatric disorders such as autism and schizophrenia, the strongest
351 genetic enrichments are within neuronal genes[49]. In contrast, the strongest genetic risk
352 enrichments for Alzheimer's disease are within microglia[51]. For ALS, a recent assessment of
353 genetic risk within different cell types implicates oligodendrocytes[51] which is consistent with our
354 finding of risk enrichment within the oligodendrocyte-enriched module, SC.M2 (**Fig. 3b**).
355
356 A major question still remains concerning the specificity of MN degeneration in ALS. Our results
357 support the hypothesis that ALS is a non-cell-autonomous disease[52], as evidenced by SC.M1,
358 SC.M2, SC.M3, SC.12, and SC.M13 being enriched with astrocyte, oligodendrocyte, endothelia,
359 and microglia markers (**Supplementary Fig. S4**). Oligodendrocytes have critical roles through
360 the metabolic support of neurons and neurotransmission in the central nervous system[53]. A
361 previous study has shown that oligodendrocytes harboring familial and sporadic, but not
362 C9orf72 variants of ALS, induce motor neuron death via a SOD1-dependent mechanism, but
363 can be rescued via lactate supplementation[54]. These previous observations corroborate our
364 findings that highlight the potential deleterious effect of dysregulation in the intracellular
365 transport and autophagy system on oligodendrocyte - MN dynamics in ALS. Furthermore,
366 neuro-immune responses mediated by astrocytes and microglia also play a critical role in ALS
367 pathogenesis[55,56] and disruptions of blood spinal cord barrier and damage to the endothelial
368 cells have been observed in electron micrographs from ALS[57]. Overall, our analyses support the
369 notion that ALS is multifactorial disorder, implicating a number of cell-types and mitigating
370 mechanisms, acting at different stages of disease. Invariably, understanding of cell-type specific
371 mechanisms governing MN degeneration will involve future interrogation of the transcriptome
372 and epigenome from ALS spinal cord at single-cell resolutions. Additionally, emerging RNA-
373 sequencing datasets, including from the TargetALS foundation (http://www.targetals.org),
374 contain larger sample sizes of ALS and control spinal cord that will help elucidate the cellular
375 dynamics underlying ALS pathophysiology.

376

377 **Methods**

378 *Initial processing of RNA-sequencing datasets*

379 The latest version of GTEx data (version 7) was obtained from the dbGaP accession number
380 phs000424.v7.p2 on 10/10/2017 and pre-processed as previously described[18]. Genes in the
381 GTEx dataset containing fewer than 10 samples with a TPM > 1 were excluded from
382 downstream analysis on the basis of unreliable quantification for very lowly expressed genes.
383 We removed 4 outlier samples in the GTEx dataset that had sample-sample connectivity Z-
384 scores of > 2 as previously described[58]. Our starting GTEx dataset consisted of 62 samples and
385 14577 genes.

386

387 RNA-sequencing data[13] from postmortem cervical spinal cord (Brohawn et al.) was downloaded
388 using the SRP accession number SRP064478 on 5/24/2019. RNA-sequencing data[22] from laser
389 capture microdissected (LCM) MNs primed for degeneration (Krach et al.) was downloaded
390 using the SRP accession number SRP067645 on 6/27/2019. RNA-sequencing data[23] from
391 iPSC-derived motor neurons from ALS8 patients (Oliveira et al.) was downloaded using the
392 SRP accession number SRP223674 on 10/18/2020. Reads were aligned to the reference
393 genome GRCh37 using STAR (version 2.5.2b) and quantified using RSEM (version 1.3.0) with
394 Gencode v19 annotations. Genes in the Brohawn et al., Krach et al., and Oliveira et al. datasets
395 with zero variance in expression were excluded from further analysis. All expression data was
396 quantified in TPM (transcript per million reads) and TPM values +1 were $\log_2$ transformed to
397 stabilize their variance. The starting Brohawn et al., Krach et al., and Oliveira et al. datasets
398 consisted of 33050, 37138, and 14171 genes respectively. Krach et al. metadata tables were
399 transcribed from the manuscript and missing values for PMI (postmortem interval) were imputed
400 as the mean of all available PMI values.

401

402 *Identification and regression of biological and technical confounders from gene expression*
403 *datasets*
404 Gene expression values were subjected to dimensionality reduction using principal component
405 analysis (PCA) with data centering and scaling. The top four components for the GTEx dataset,
406 which collectively captures 70.6% of the total variance, were assigned to be expression principal
407 components (PCs) 1-4. The top five components for the Brohawn et al., Krach et al., and
408 Oliveira et al. dataset, which collectively capture 55.9%, 42.3%, and 56.3% of the total variance
409 respectively, were assigned to be expression principal components (PCs) 1-5. Sequencing Q/C
410 metrics were calculated using PicardTools (version 2.5.0) and dimensionally reduced using PCA
411 with centering and scaling. Sequencing Q/C metrics were aggregated manually for Brohawn et
412 al. and Krach et al. datasets and aggregated using MultiQC[59] for the Oliveira et al. dataset
413 (**Supplementary Data 2-4**). The top five components of sequencing metrics for the GTEx,
414 Brohawn et al., Krach et al., and Oliveira et al. datasets, which collectively capture 75.1%, 91%,
415 90.4%, and 89.0% of the total variance respectively, were selected and assigned to be
416 sequencing principal components (seqPCs) 1-5.
417
418 Numerical covariates of technical or biological importance were correlated with the expression
419 PCs using Spearman's correlation. Binary covariates were numerically encoded and multi-level
420 categorical covariates of technical or biological importance were binarized using dummy
421 variables. The model fit (R^2) was used to assess correlations. We identified technical or
422 biological covariates with a meaningful influence on gene expression if they were correlated with
423 an $R > 0.3$ to any of the top expression PCs.
424
425 For the GTEx expression dataset, a linear model was fit for each gene as follows: GTEX.Expr ~
426 seqPC1 + seqPC2 + seqPC3 + seqPC4 + seqPC5 + RIN + DTHHRDY + AGE + SMCENTER +
427 TRISCHD and the effect of all covariates were regressed out. RIN indicates RNA integrity

428 number, DTHHRDY represents the Hardy Scale, SMCENTER represents the center for
429 sampling, and TRISCHD represents ischemic time. Ethnicity was not included in the linear
430 model because it was collinear with SMCENTER (**Supplementary Fig. S1c**).
431
432 For the Brohawn et al. dataset, a linear model was fit for each gene as follows: Expr ~ seqPC1
433 + seqPC2 + seqPC3 + seqPC4 +seqPC5 + Ethnicity + Sex + Age + Disease and the effect of all
434 covariates except disease were regressed out.
435
436 For the Krach et al. dataset, a linear model was fit for each gene as follows: Expr ~ Diagnosis +
437 Age + Gender + RIN + PMI + seqPC1 + seqPC2 + seqPC3 + seqPC4 + seqPC5 and the effect
438 of all covariates except diagnosis were regressed out.
439
440 For the Oliveira et al. dataset, a linear model was fit for each gene as follows: Expr ~ Disease +
441 Age + Sex + Tissue + seqPC1 + seqPC2 + seqPC3 + seqPC4 + seqPC5 and the effect of all
442 covariates except disease were regressed out.
443
444 *Weighted Gene Co-expression Network Analysis (WGCNA)*
445 We used WGCNA[4,58] to generate a gene co-expression network using the GTEx dataset. We
446 chose a soft power of 10 using the function “pickSoftThreshold” in the WGCNA R package to
447 achieve optimal scale free topology. The Topological Overlap Matrix (TOM) was generated
448 using the function “TOMsimilarity” in the WGCNA R package. A hierarchical gene clustering
449 dendrogram was constructed using the TOM-based dissimilarity and module selection was
450 tested using multiple permutations of the tree cutting parameters: minimum module size, cut
451 height, and deep split. Minimum module size (mms) was assessed at 50, 100, and 150. Module
452 merging (dcor) was assessed at 0.1, 0.2, and 0.25. Deep split (DS) was assessed at 2 and 4.
453 By visual inspection, we utilized the following parameters: minimum module size = 100, module

454 merging = 0.1, and deep split = 2 to generate the network. We assigned unique module
455 identifiers (SC.M1-13) for each module color except the grey module which represents genes
456 that were not co-expressed.
457
458 Pairwise Spearman's correlations were calculated using 1000 randomly sampled genes within
459 each module and a matched background set. If a module contained less than 1000 genes, all
460 genes in the module were sampled.
461
462 *Enrichment of gene sets in co-expression modules*
463 We analyzed a collection of gene sets generated by previous genetic studies of ALS. For large
464 effect size ALS risk genes, we obtained 50 ALS literature curated ALS risk genes from two
465 studies[19,20]. For the determination of cell-type specificity, we obtained the top 50 differentially
466 expressed genes per cell-type by fold change from single-nuclei RNA sequencing of postnatal
467 day 2 mouse spinal cord[24]. For the human ALS spinal cord differentially expressed gene (DEG)
468 enrichments, we obtained 265 DEGs (172 up- and 93 down-regulated) at FDR < 0.10 using the
469 union of DESeq2 and EdgeR analyses[13]. For the dissected motor neurons primed for
470 neurodegeneration, we obtained 1600 DEGs (1290 up- and 310 down-regulated) at FDR < 0.05
471 from the report[22]. For the iPSC-derived motor neurons from ALS patients, we chose to examine
472 the severe ALS vs. control comparison and filtered gene biotype for protein-coding genes,
473 obtaining 653 DEGs (364 up- and 289 down-regulated) at FDR < 0.01 from the report[23]. For
474 DEGs from SOD1-G93A mice, we obtained genes with a Bayes Factor > 3 (corresponding to
475 FDR = 0.1) identified through a spatial transcriptomics study[12].
476
477 Logistic regression was performed to calculate an odds ratio and p-value for enrichment of each
478 co-expression module with a test gene set. The p-values were FDR-corrected for the number of
479 modules analyzed.

480

481 *Enrichment of common ALS genetic risk variants in co-expression modules*

482 Stratified LD score regression[29] was performed to evaluate the enrichment of common risk
483 variants from GWAS studies of ALS[21] (N=80,610), Inflammatory Bowel Disorder (IBD)[30]
484 (N=86,640), Autism Spectrum Disorder (ASD)[32,33] (N=15,954), Alzheimer's Disease (AD)[34]
485 (N=455,258), Frontotemporal Dementia (FTD)[35] (N=12,928), and Progressive Supranuclear
486 Palsy (PSP)[36] (N=12,308) within each co-expression module. We used the full baseline model of
487 53 functional categories per calculation of partitioned heritability (1000 Genomes Phase 3). We
488 defined genomic regions of each co-expression module as their constituent gene bodies plus a
489 flanking region of +/- 10 KB.

490

491 *Gene Ontology Enrichments*

492 Gene Ontology enrichment analysis was performed using the gProfiler package[60] with
493 hierarchical filtering set to "moderate", domain size set to "annotated", src filter set to "GO:BP"
494 and "GO:MF", electronic annotations (IEAs) included, maximum p-value set to 0.05, correction
495 method set to "FDR", and a maximum set size of 1000. To reflect the weight of each gene within
496 a co-expression module, genes were ordered by their module membership (kME) during the
497 enrichment calculations.

498

499 *Network visualization of module hub genes*

500 For each module, an undirected and weighted network amongst the top 30 module hub genes
501 (by kME) was visualized using the igraph package in R[61].

502

503 *Protein-protein interaction (PPI) networks*

504 PPI networks were produced using the Disease Association Protein-Protein Evaluator
505 (DAPPLE)[62]. Two hundred and thirty eight of the top 300 SC.M4 hub genes, 238 of the top 300

506 SC.M2 hub genes, 84 of top 100 SC.M6 hub genes as ranked by Module Membership (kME)
507 were found in the DAPPLE database and seeded into the shown PPI network (**Fig. 3e, 3f, 4b**).
508 All networks were produced from 1,000 permutations (within-degree node-label permutation).
509 SC.M6, SC.M4, and SC.M2 contain the following network properties respectively: direct edges
510 count (p = 0.001, p =0.001, p = 0.001), seed direct degrees mean (p = 0.001, p = 0.001, p =
511 0.018), seed indirect degrees mean (p = 0.001, p = 0.006, p = 0.006), and the CI degrees mean
512 (p = 0.001, p = 0.132, p = 0.332).
513
514 *Module preservation analyses*
515 Module preservation for co-expression modules in the Brohawn et al. and Krach et al. datasets
516 were performed as previously described using the modulePreservation function in the WGCNA
517 R package[58]. We performed 200 permutation tests using a randomSeed of 1.
518
519 *Correlations of module eigengenes to biological and technical covariates*
520 To evaluate the correlations of module eigengenes to biological and technical covariates in the
521 Brohawn et al. and Krach et al. datasets, a Student p-value was calculated using the function
522 "corPvalueStudent" in the WGCNA R package. P-values were FDR corrected across the
523 modules and plotted using the function "labeledHeatmap" in the WGCNA R package. To
524 account for multiple biological replicates per patient in the Oliveira et al. dataset, a mixed effect
525 linear model was fitted for each module eigengene and each of the following covariates:
526 Disease, Age, Sex, Tissue, seqPC1, seqPC2, seqPC3, seqPC4, and seqPC5. The covariate
527 "Isolate", which uniquely identifies each individual, was used as the random effect.
528
529 **Data Availability**

Module assignments are provided as Supplementary Data 1 and the full range of values underlying the barplots and heatmaps in Figures 3a-b, 4a, 5a-c, and Supplementary Figures S4, S5d, S6d, S7d, S8, S9 are provided as a Source Data File.

Code Availability

Underlying R code to run WGCNA is available at [https://github.com/dhglab/ALS-Gene-Network-Manuscript].

537 **References**
538

539 1. Ravits, J. *et al.* Deciphering amyotrophic lateral sclerosis: What phenotype,
540 neuropathology and genetics are telling us about pathogenesis. *Amyotroph. Lateral*
541 *Scler. Front. Degener.* **14**, 5–18 (2013).
542 2. Taylor, J. P., Brown, R. H. & Cleveland, D. W. Decoding ALS: from genes to
543 mechanism. *Nature* **539**, 197–206 (2016).
544 3. Hardiman, O. *et al.* Amyotrophic lateral sclerosis. *Nat. Rev. Dis. Primer* **3**, 17071 (2017).
545 4. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network
546 analysis. *BMC Bioinformatics* **9**, 559 (2008).
547 5. Zhang, B. *et al.* Integrated Systems Approach Identifies Genetic Nodes and Networks in
548 Late-Onset Alzheimer's Disease. *Cell* **153**, 707–720 (2013).
549 6. Mostafavi, S. *et al.* A molecular network of the aging human brain provides insights into
550 the pathology and cognitive decline of Alzheimer's disease. *Nat. Neurosci.* **21**, 811–819
551 (2018).
552 7. Rexach, J., Swarup, V., Chang, T. & Geschwind, D. *Dementia risk genes engage gene*
553 *networks poised to tune the immune response towards chronic inflammatory states*.
554 http://biorxiv.org/lookup/doi/10.1101/597542 (2019) doi:10.1101/597542.
555 8. Swarup, V. *et al.* Identification of evolutionarily conserved gene networks mediating
556 neurodegenerative dementia. *Nat. Med.* **25**, 152–164 (2019).
557 9. Johnson, E. C. B. *et al.* Large-scale proteomic analysis of Alzheimer's disease brain and
558 cerebrospinal fluid reveals early changes in energy metabolism associated with
559 microglia and astrocyte activation. *Nat. Med.* **26**, 769–780 (2020).
560 10. Swarup, V. *et al.* Identification of Conserved Proteomic Networks in Neurodegenerative
561 Dementia. *Cell Rep.* **31**, 107807 (2020).
562 11. Prudencio, M. *et al.* Distinct brain transcriptome profiles in C9orf72-associated and
563 sporadic ALS. *Nat. Neurosci.* **18**, 1175–1182 (2015).
564 12. Maniatis, S. *et al.* Spatiotemporal dynamics of molecular pathology in amyotrophic
565 lateral sclerosis. *Science* **364**, 89–93 (2019).
566 13. Brohawn, D. G., O'Brien, L. C. & Bennett, J. P. RNAseq Analyses Identify Tumor
567 Necrosis Factor-Mediated Inflammation as a Major Abnormality in ALS Spinal Cord.
568 *PLOS ONE* **11**, e0160520 (2016).
569 14. D'Erchia, A. M. *et al.* Massive transcriptome sequencing of human spinal cord tissues
570 provides new insights into motor neuron degeneration in ALS. *Sci. Rep.* **7**, 10046 (2017).
571 15. Schiffer, D., Cordera, S., Cavalla, P. & Migheli, A. Reactive astrogliosis of the spinal cord
572 in amyotrophic lateral sclerosis. *J. Neurol. Sci.* **139**, 27–33 (1996).
573 16. Phatnani, H. P. *et al.* Intricate interplay between astrocytes and motor neurons in ALS.
574 *Proc. Natl. Acad. Sci.* **110**, E756–E765 (2013).
575 17. Barham, C. *et al.* RNA-Seq Analysis of Spinal Cord Tissues from hPFN1G118V
576 Transgenic Mouse Model of ALS at Pre-symptomatic and End-Stages of Disease. *Sci.*
577 *Rep.* **8**, 13737 (2018).
578 18. GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature*
579 **550**, 204–213 (2017).
580 19. Eykens, C. & Robberecht, W. The genetic basis of amyotrophic lateral sclerosis: recent
581 breakthroughs. *Advances in Genomics and Genetics* https://www.dovepress.com/the-
582 genetic-basis-of-amyotrophic-lateral-sclerosis-recent-breakthrough-peer-reviewed-
583 fulltext-article-AGG (2015) doi:10.2147/AGG.S57397.
584 20. Chia, R., Chiò, A. & Traynor, B. J. Novel genes associated with amyotrophic lateral
585 sclerosis: diagnostic and clinical implications. *Lancet Neurol.* **17**, 94–102 (2018).
586 21. Nicolas, A. *et al.* Genome-wide Analyses Identify KIF5A as a Novel ALS Gene. *Neuron*
587 **97**, 1268-1283.e6 (2018).

588 22. Krach, F. *et al.* Transcriptome–pathology correlation identifies interplay between TDP-43
589 and the expression of its kinase CK1E in sporadic ALS. *Acta Neuropathol. (Berl.)* **136**,
590 405–423 (2018).
591 23. Oliveira, D. *et al.* Different gene expression profiles in iPSC-derived motor neurons from
592 ALS8 patients with variable clinical courses suggest mitigating pathways for
593 neurodegeneration. *Hum. Mol. Genet.* **29**, 1465–1475 (2020).
594 24. Rosenberg, A. B. *et al.* Single-cell profiling of the developing mouse brain and spinal
595 cord with split-pool barcoding. *Science* **360**, 176–182 (2018).
596 25. Parikshak, N. N., Gandal, M. J. & Geschwind, D. H. Systems biology and gene networks
597 in neurodevelopmental and neurodegenerative disorders. *Nat. Rev. Genet.* **16**, 441–458
598 (2015).
599 26. Chai, N. & Gitler, A. D. Yeast screen for modifiers of C9orf72 poly(glycine-arginine)
600 dipeptide repeat toxicity. *FEMS Yeast Res.* **18**, (2018).
601 27. Jovičić, A. *et al.* Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic
602 transport defects to FTD/ALS. *Nat. Neurosci.* **18**, 1226–1229 (2015).
603 28. Sun, Z. *et al.* Molecular Determinants and Genetic Modifiers of Aggregation and Toxicity
604 for the ALS Disease Protein FUS/TLS. *PLoS Biol.* **9**, e1000614 (2011).
605 29. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide
606 association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
607 30. Liu, J. Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel
608 disease and highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979–986
609 (2015).
610 31. Xavier, R. J. & Podolsky, D. K. Unravelling the pathogenesis of inflammatory bowel
611 disease. *Nature* **448**, 427–434 (2007).
612 32. Grove, J. *et al. Common risk variants identified in autism spectrum disorder*.
613 http://biorxiv.org/lookup/doi/10.1101/224774 (2017) doi:10.1101/224774.
614 33. The Autism Spectrum Disorders Working Group of The Psychiatric Genomics
615 Consortium. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum
616 disorder highlights a novel locus at 10q24.32 and a significant overlap with
617 schizophrenia. *Mol. Autism* **8**, 21 (2017).
618 34. Jansen, I. E. *et al.* Genome-wide meta-analysis identifies new loci and functional
619 pathways influencing Alzheimer's disease risk. *Nat. Genet.* **51**, 404–413 (2019).
620 35. Ferrari, R. *et al.* Frontotemporal dementia and its subtypes: a genome-wide association
621 study. *Lancet Neurol.* **13**, 686–699 (2014).
622 36. Chen, J. A. *et al.* Joint genome-wide association study of progressive supranuclear palsy
623 identifies novel susceptibility loci and genetic correlation to neurodegenerative diseases.
624 *Mol. Neurodegener.* **13**, 41 (2018).
625 37. Boeynaems, S. *et al.* Phase Separation of C9orf72 Dipeptide Repeats Perturbs Stress
626 Granule Dynamics. *Mol. Cell* **65**, 1044-1055.e5 (2017).
627 38. Ozdilek, B. A. *et al.* Intrinsically disordered RGG/RG domains mediate degenerate
628 specificity in RNA binding. *Nucleic Acids Res.* **45**, 7984–7996 (2017).
629 39. Cooper-Knock, J. *et al.* A data-driven approach links microglia to pathology and
630 prognosis in amyotrophic lateral sclerosis. *Acta Neuropathol. Commun.* **5**, 23 (2017).
631 40. Ripps, M. E., Huntley, G. W., Hof, P. R., Morrison, J. H. & Gordon, J. W. Transgenic
632 mice expressing an altered murine superoxide dismutase gene provide an animal model
633 of amyotrophic lateral sclerosis. *Proc. Natl. Acad. Sci.* **92**, 689–693 (1995).
634 41. Gurney, M. *et al.* Motor neuron degeneration in mice that express a human Cu,Zn
635 superoxide dismutase mutation. *Science* **264**, 1772 (1994).
636 42. Polymenidou, M. *et al.* Misregulated RNA processing in amyotrophic lateral sclerosis.
637 *Brain Res.* **1462**, 3–15 (2012).

43. Husedzinovic, A. *et al.* The catalytically inactive tyrosine phosphatase HD-PTP/PTPN23 is a novel regulator of SMN complex localization. *Mol. Biol. Cell* **26**, 161–171 (2015).
44. Mancias, J. D., Wang, X., Gygi, S. P., Harper, J. W. & Kimmelman, A. C. Quantitative proteomics identifies NCOA4 as the cargo receptor mediating ferritinophagy. *Nature* **509**, 105–109 (2014).
45. Quiles del Rey, M. & Mancias, J. D. NCOA4-Mediated Ferritinophagy: A Potential Link to Neurodegeneration. *Front. Neurosci.* **13**, 238 (2019).
46. Kim, H. J. & Taylor, J. P. Lost in Transportation: Nucleocytoplasmic Transport Defects in ALS and Other Neurodegenerative Diseases. *Neuron* **96**, 285–297 (2017).
47. McCombe, P. A. & Henderson, R. D. The Role of Immune and Inflammatory Mechanisms in ALS. *Curr. Mol. Med.* **11**, 9 (2011).
48. McCombe, P. A. The Peripheral Immune System and Amyotrophic Lateral Sclerosis. *Front. Neurol.* **11**, 12 (2020).
49. Gandal, M. J. *et al.* Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science* 6 (2018).
50. Morabito, S., Miyoshi, E., Michael, N. & Swarup, V. Integrative genomics approach identifies conserved transcriptomic networks in Alzheimer's disease. *Hum. Mol. Genet.* **29**, 2899–2919 (2020).
51. Nott, A. *et al.* Brain cell type–specific enhancer–promoter interactome maps and disease risk association. *Science* **366**, 1134–1139 (2019).
52. Ilieva, H., Polymenidou, M. & Cleveland, D. W. Non–cell autonomous toxicity in neurodegenerative disorders: ALS and beyond. *J. Cell Biol.* **187**, 761–772 (2009).
53. Simons, M. & Nave, K.-A. Oligodendrocytes: Myelination and Axonal Support. *Cold Spring Harb. Perspect. Biol.* **8**, a020479 (2016).
54. Ferraiuolo, L. *et al.* Oligodendrocytes contribute to motor neuron death in ALS via SOD1-dependent mechanism. *Proc. Natl. Acad. Sci.* **113**, E6496–E6505 (2016).
55. Yamanaka, K. & Komine, O. The multi-dimensional roles of astrocytes in ALS. *Neurosci. Res.* **126**, 31–38 (2018).
56. Geloso, M. C. & D'Ambrosi, N. The Dual Role of Microglia in ALS: Mechanisms and Therapeutic Approaches. *Front. Aging Neurosci.* **9**, 10 (2017).
57. Garbuzova-Davis, S. *et al.* Impaired blood–brain/spinal cord barrier in ALS patients. *Brain Res.* **1469**, 114–128 (2012).
58. Horvath, S. *Weighted Network Analysis: Applications in Genomics and Systems Biology*. (Springer-Verlag, 2011).
59. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
60. Reimand, J. *et al.* g:Profiler—a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res.* **44**, W83–W89 (2016).
61. Csárdi, G. & Nepusz, T. The igraph software package for complex network research. in (2006).
62. Rossin, E. J. *et al.* Proteins Encoded in Genomic Regions Associated with Immune-Mediated Disease Physically Interact and Suggest Underlying Biology. *PLoS Genet.* **7**, e1001273 (2011).
63. Carbon, S. *et al.* AmiGO: online access to ontology and annotation data. *Bioinformatics* **25**, 288–289 (2009).

Acknowledgements

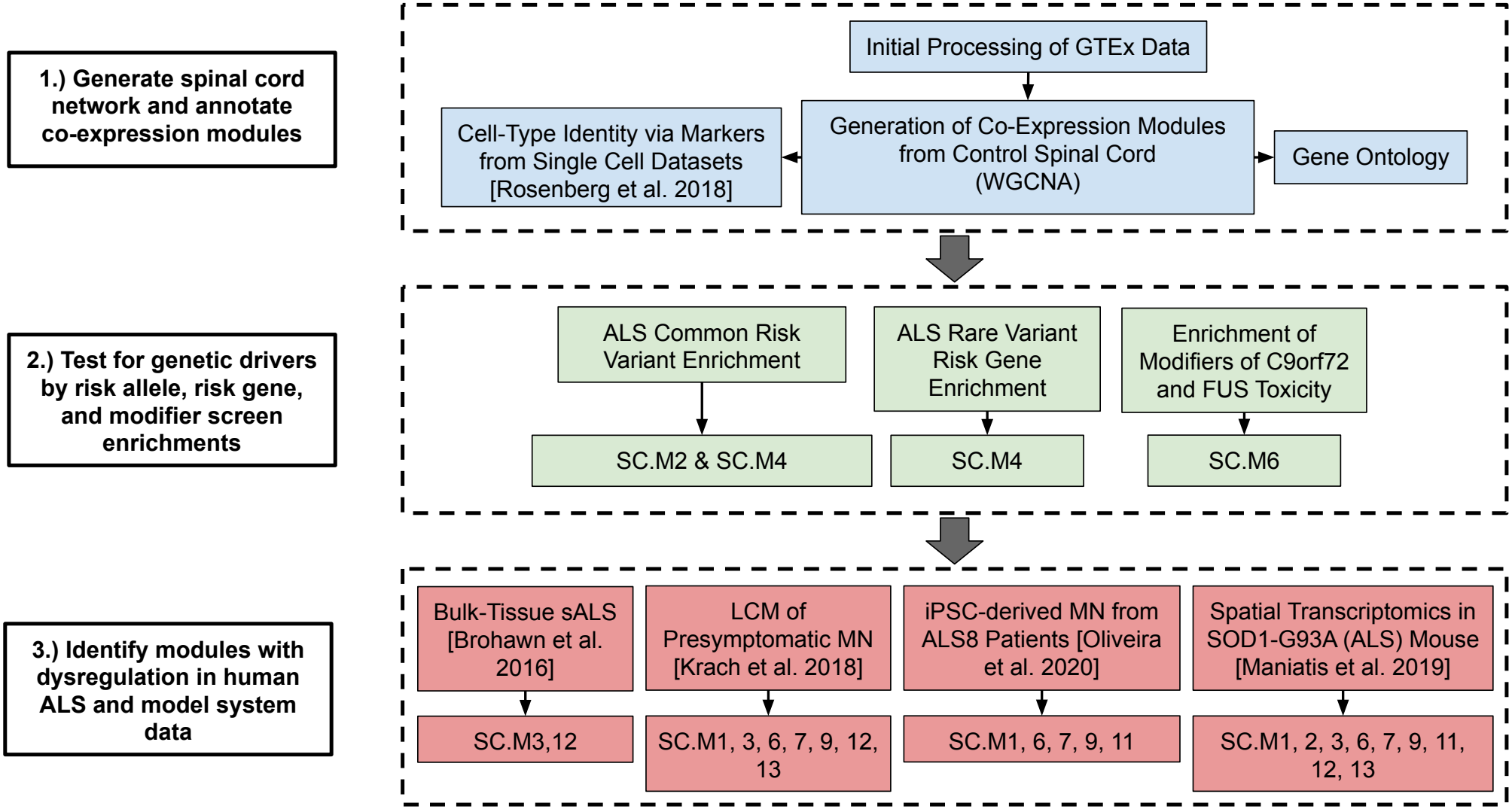
We thank Christopher Hartl, William Pembroke, Timothy S. Chang, and other members of the Geschwind Laboratory for their technical assistance and manuscript feedback. We also thank Aaron D. Gitler and members of his laboratory for their stimulating discussions and assistance with the genetic modifier screen data. Funding for this work was provided by the Boyer Endowment Award, Helga K. & Walter Oppenheimer Endowment Award, and the MacDowell Endowment Award through the UCLA Undergraduate Research Center - Sciences (to J.C.W), NIMH 1F32MH114620 (to G.R), and NIMH 5R01MH109912 (to D.H.G).

Author Contributions

J.C.W, G.R, and D.H.G planned the analyses and wrote the manuscript. Analyses were performed by J.C.W with supervision from G.R. and D.H.G.

Competing Financial Interests

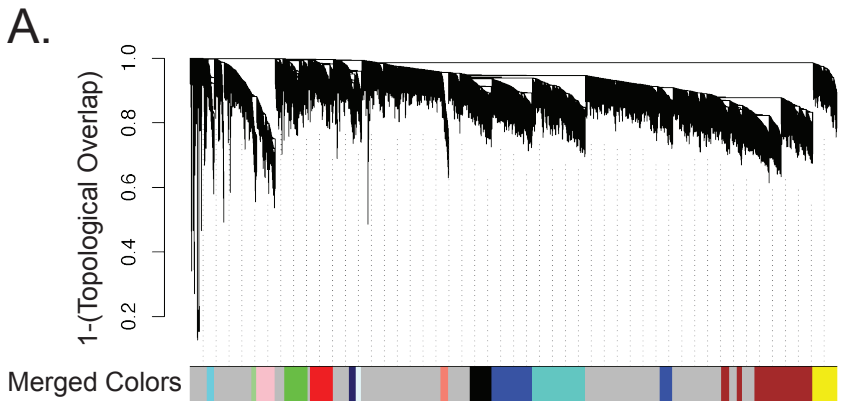
The authors declare no competing financial interests.


1.) Generate spinal cord network and annotate co-expression modules
Initial Processing of GTEx Data
Cell-Type Identity via Markers from Single Cell Datasets [Rosenberg et al. 2018]
Generation of Co-Expression Modules from Control Spinal Cord (WGCNA)
Gene Ontology
2.) Test for genetic drivers by risk allele, risk gene, and modifier screen enrichments
ALS Common Risk Variant Enrichment
ALS Rare Variant Risk Gene Enrichment
Enrichment of Modifiers of C9orf72 and FUS Toxicity
SC.M2 & SC.M4
SC.M4
SC.M6
3.) Identify modules with dysregulation in human ALS and model system data
Bulk-Tissue sALS [Brohawn et al. 2016]
LCM of Presymptomatic MN [Krach et al. 2018]
iPSC-derived MN from ALS8 Patients [Oliveira et al. 2020]
Spatial Transcriptomics in SOD1-G93A (ALS) Mouse [Maniatis et al. 2019]
SC.M3,12
SC.M1, 3, 6, 7, 9, 12, 13
SC.M1, 6, 7, 9, 11
SC.M1, 2, 3, 6, 7, 9, 11, 12, 13

701 **Figure 1**. Overview of spinal cord co-expression network analysis.

702 Flow diagram of analyses in this study. A co-expression network was generated using the GTEx
703 cervical spinal cord expression data and Weighted Gene Co-Expression Network Analysis
704 (WGCNA). Modules were annotated by their associated biological processes, hub genes, and
705 cell-type identities. Modules were then assessed for their enrichment of ALS genetic risk factors
706 and analyzed for differential expression in RNA-seq data from human ALS tissues and model
707 system studies[12,13,22,23].

708



B.

Molecular Themes for Co-expression Modules

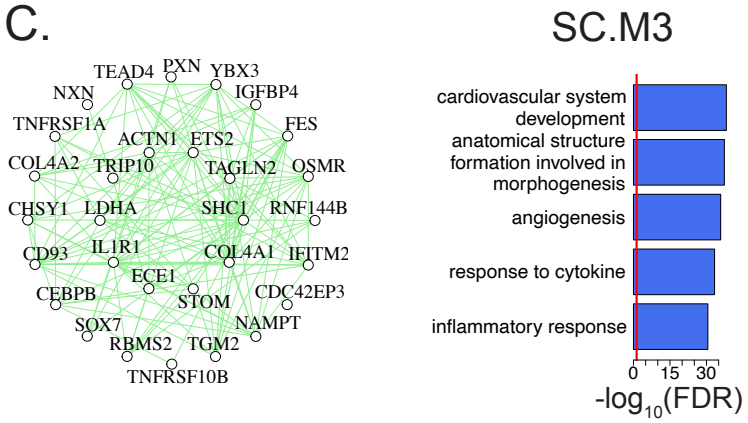
Morphogenesis & Development
Red (SC.M1)
Green (SC.M3)

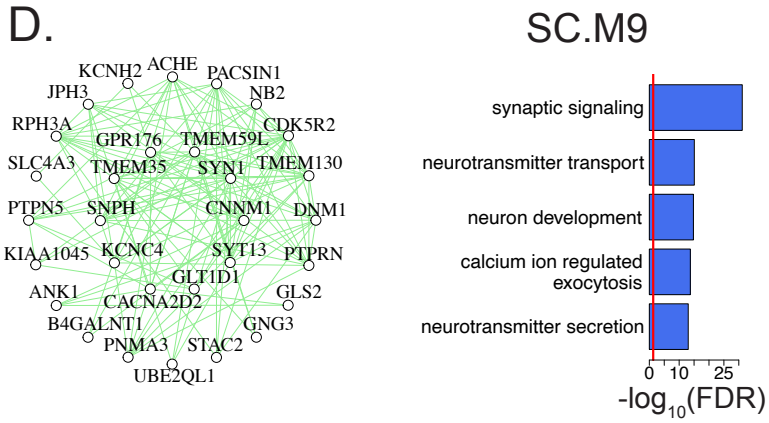
Neuronal Signaling
Yellow (SC.M7)
Pink (SC.M9)
Light Green (SC.M11)

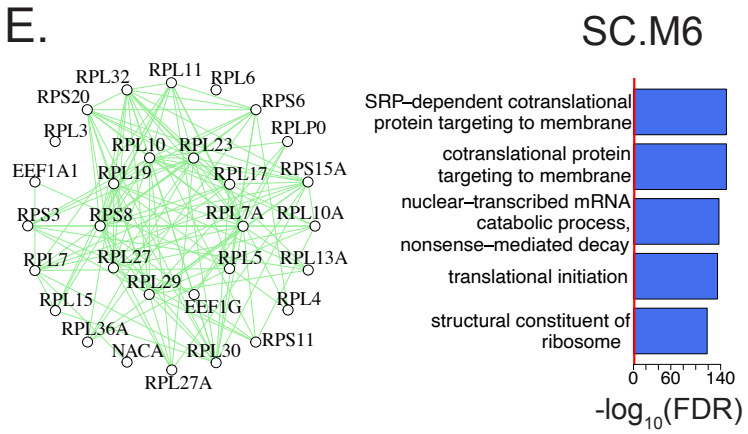
Epigenetics & Gene Regulation
Blue (SC.M4)
Salmon (SC.M6)
Black (SC.M8)

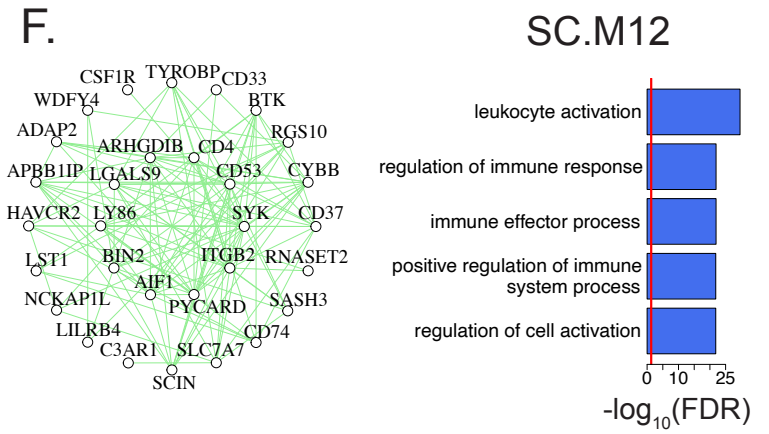
Immune Activation
Midnight Blue (SC.M12)
Light Cyan (SC.M13)

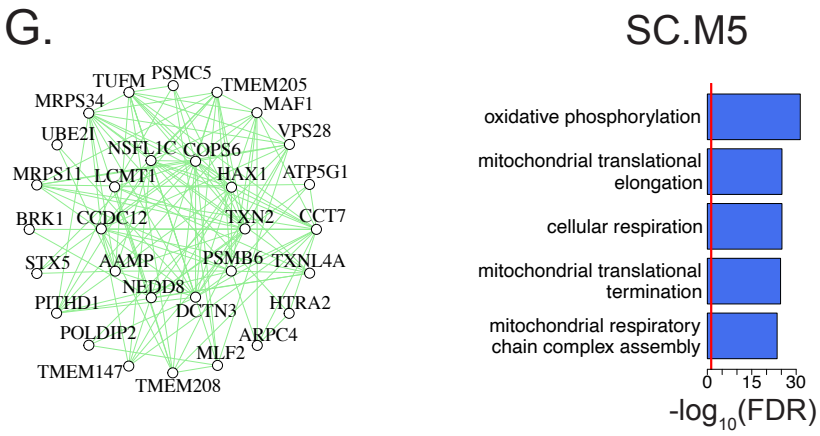
Intracellular Transport, Metabolism, & Sensing
Brown (SC.M2)
Turquoise (SC.M5)
Cyan (SC.M10)











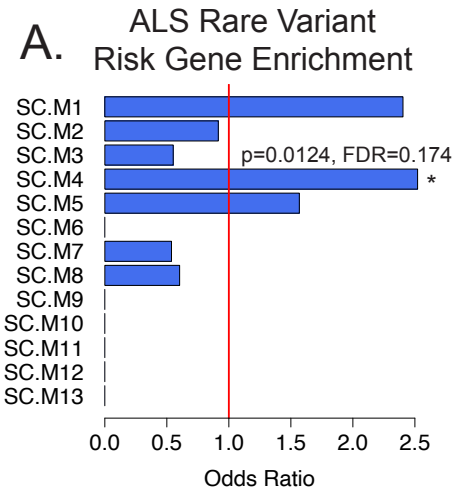
709 **Figure 2**. Construction of co-expression network in control cervical spinal cord.

710 (A) Dendrogram of the topological overlap matrix for gene co-expression dissimilarity. Co-
711 expression module assignments are shown in the “Merged Colors” track (See Supplementary
712 Fig. S2 for optimization of tree-cutting parameters). Genes that failed to be clustered into a co-
713 expression module were grouped within the grey module.

714 (B) Thirteen co-expression modules were categorized into five broad molecular themes based
715 on Gene Ontology (GO) term enrichment and assigned a unique module identifier (SC.M1-13).

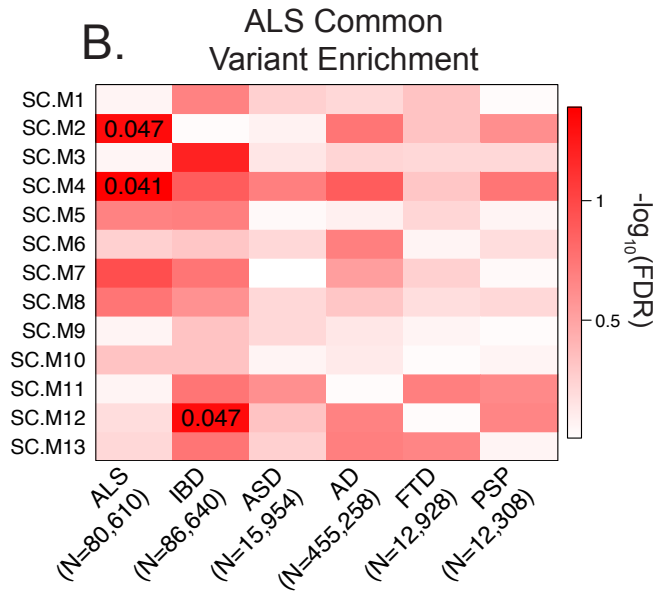
716 (C, D, E, F, G) Top 30 hub genes and 300 connections for SC.M3, SC.M9, SC.M6, SC.M12,
717 and SC.M5 respectively (left panels). Top 5 enriched GO terms for each of the respective
718 modules (right panels). Red lines mark an FDR corrected p-value threshold of 0.05.

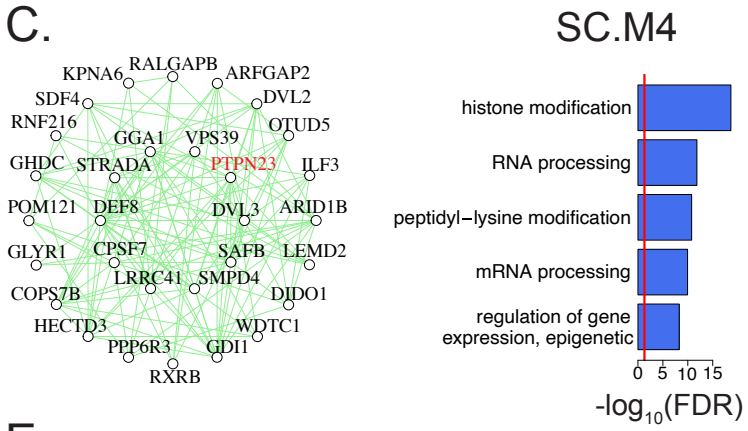
719

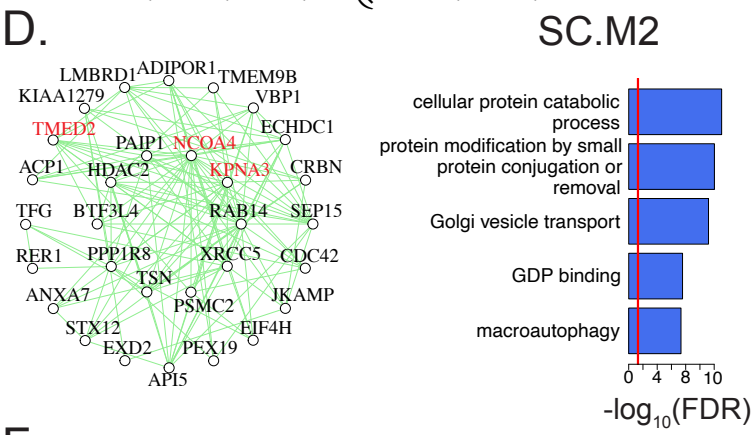


Overlapping ALS Risk Genes in SC.M4

FUS	C21orf2
SPG11	EWSR1
SETX	SS18L1
TARDBP	ZNF512B
ATXN2	



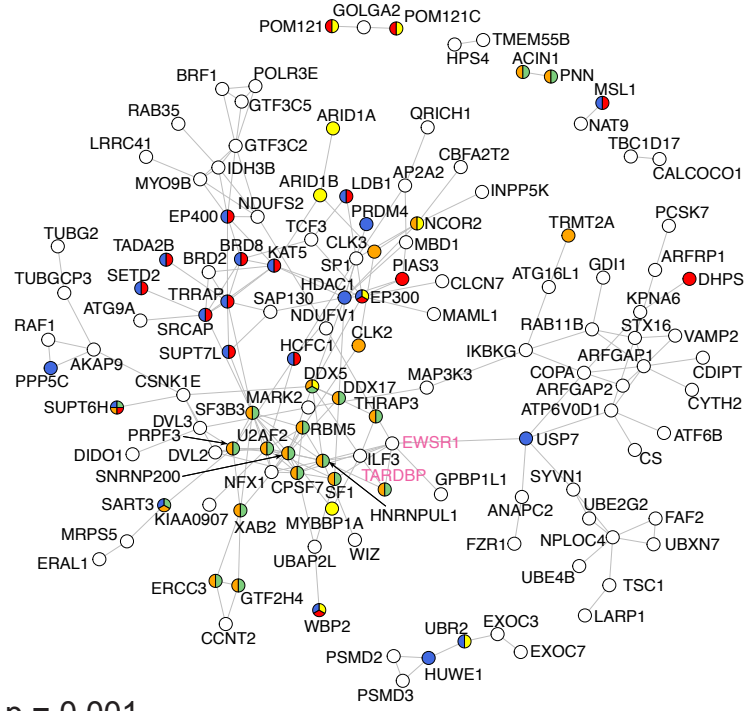




E.

Key

Blue	histone modification
Orange	RNA processing
Red	peptidyl-lysine modification
Green	mRNA processing
Yellow	regulation of gene expression, epigenetic

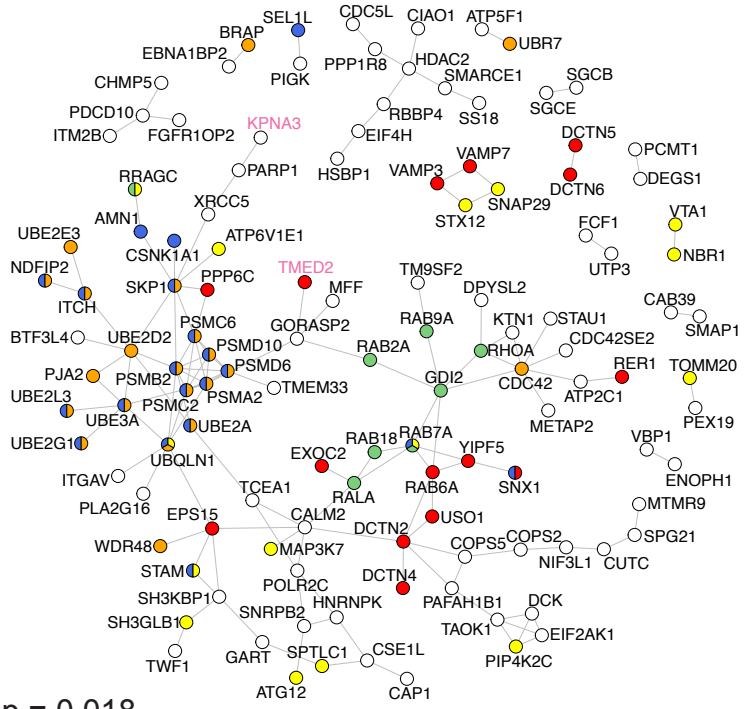


p = 0.001

F.

Key

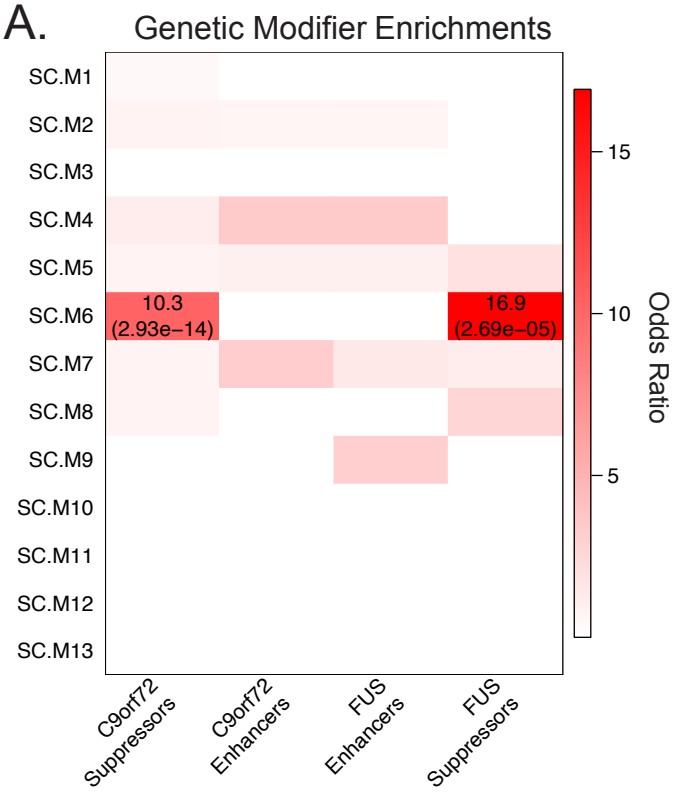
Blue	cellular protein catabolic process
Orange	protein modification by small protein conjugation or removal
Red	golgi vesicle transport
Green	GDP binding
Yellow	macroautophagy



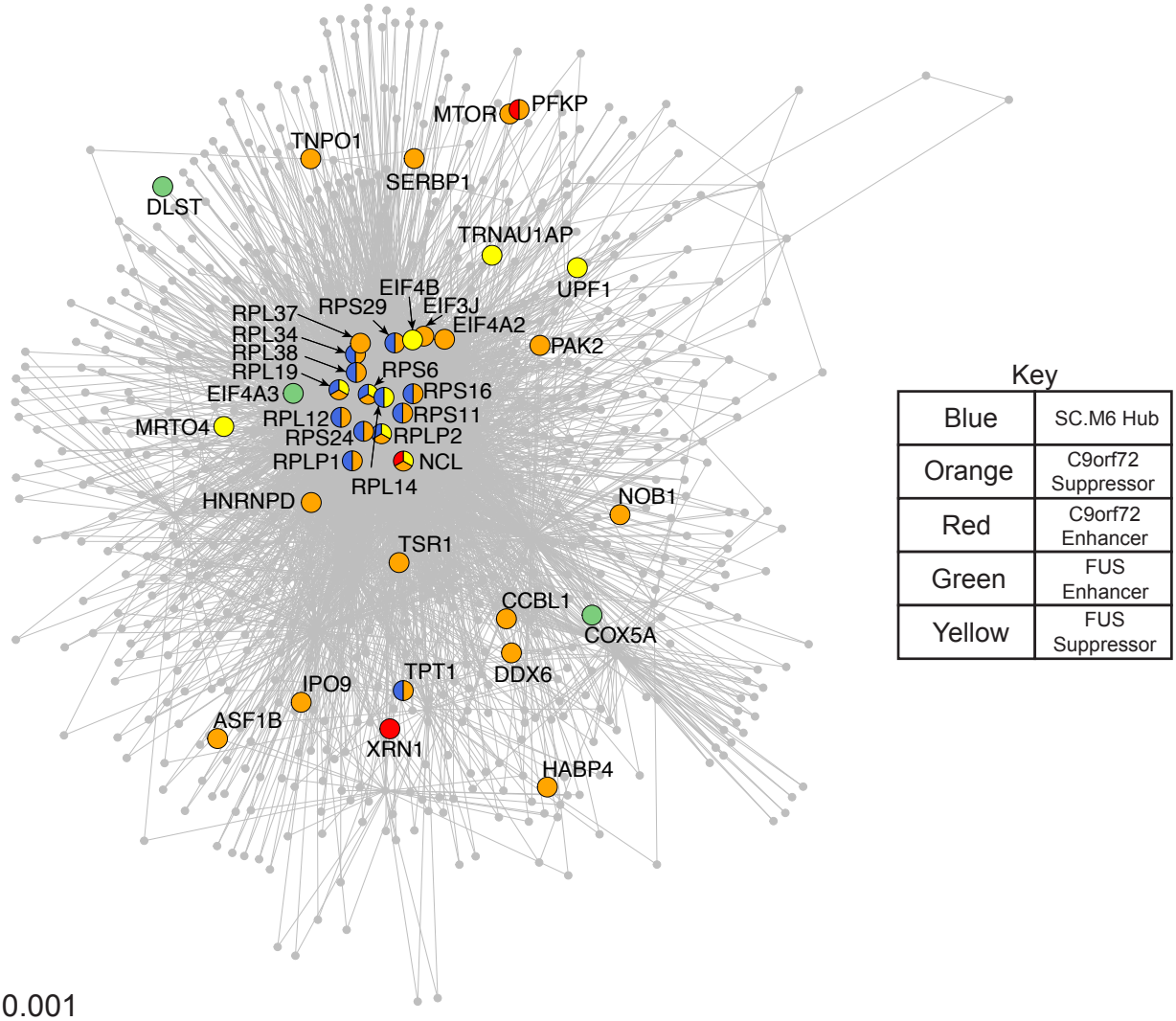
p = 0.018

720 **Figure 3.** Enrichment of ALS genetic risk factors in intracellular transport/autophagy module
721 SC.M2 and RNA processing/gene regulation module SC.M4.
722 (A) Enrichment of literature curated large-effect ALS risk genes[19,20] within co-expression
723 modules (left panel). Red line marks a baseline odds ratio of 1. SC.M4 is marked with an
724 asterisk as having a nominal enrichment p-value < 0.05. Literature-curated large-effect ALS risk
725 genes within the SC.M4 module (right panel).
726 (B) Partitioned heritability enrichments of common risk variants for ALS[21], IBD[30], ASD[32,33], AD[34],
727 FTD[35], and PSP[36] within co-expression modules. Text values of enrichments with FDR
728 corrected p-value < 0.05 are shown.
729 (C, D) Top 30 hub genes and 300 connections for SC.M4 and SC.M2 respectively (left panels).
730 Highlighted in red are genes that have been previously implicated in ALS pathophysiology[27,43–
731 45]. Top 5 enriched GO terms for each of the respective modules (right panels). Red lines mark
732 an FDR corrected p-value threshold of 0.05.
733 (E, F) Direct protein-protein interaction (PPI) network of the top 300 hub genes from modules
734 SC.M4 and SC.M2 respectively. Gene vertices are color coded according to their Gene
735 Ontology terms as shown[63]. P-value of the PPI network is derived from 1,000 permutations.
736 Highlighted in pink are genes that have been previously implicated in ALS pathophysiology[27,43–
737 45].
738

A.


Genetic Modifier Enrichments
SC.M1
SC.M2
SC.M3
SC.M4
SC.M5
SC.M6
SC.M7
SC.M8
SC.M9
SC.M10
SC.M11
SC.M12
SC.M13
10.3
(2.93e−14)
16.9
(2.69e−05)
C9orf72 Suppressors
C9orf72 Enhancers
FUS Enhancers
FUS Suppressors
15
10
5
Odds Ratio

B.


MTOR
PFKP
TNPO1
SERBP1
DLST
TRNAU1AP
UPF1
EIF4B
RPS29
EIF3J
EIF4A2
RPL37
RPL34
RPL38
RPL19
PAK2
RPS6
EIF4A3
RPS16
RPS11
MRTO4
RPL12
RPS24
RPLP2
RPLP1
NCL
RPL14
HNRNPD
NOB1
TSR1
CCBL1
COX5A
TPT1
DDX6
IPO9
ASF1B
XRN1
HABP4
Key
Blue
SC.M6 Hub
Orange
C9orf72 Suppressor
Red
C9orf72 Enhancer
Green
FUS Enhancer
Yellow
FUS Suppressor

p = 0.001

739 **Figure 4**. Modifiers of C9orf72 and FUS toxicity converge on top hubs of ribosomal-associated

740 module SC.M6.

741 (A) Enrichment of genetic modifiers of C9orf72 and FUS toxicity[26–28] within co-expression

742 modules. Odds ratios of enrichments with FDR corrected p-value < 0.05 are shown. FDR

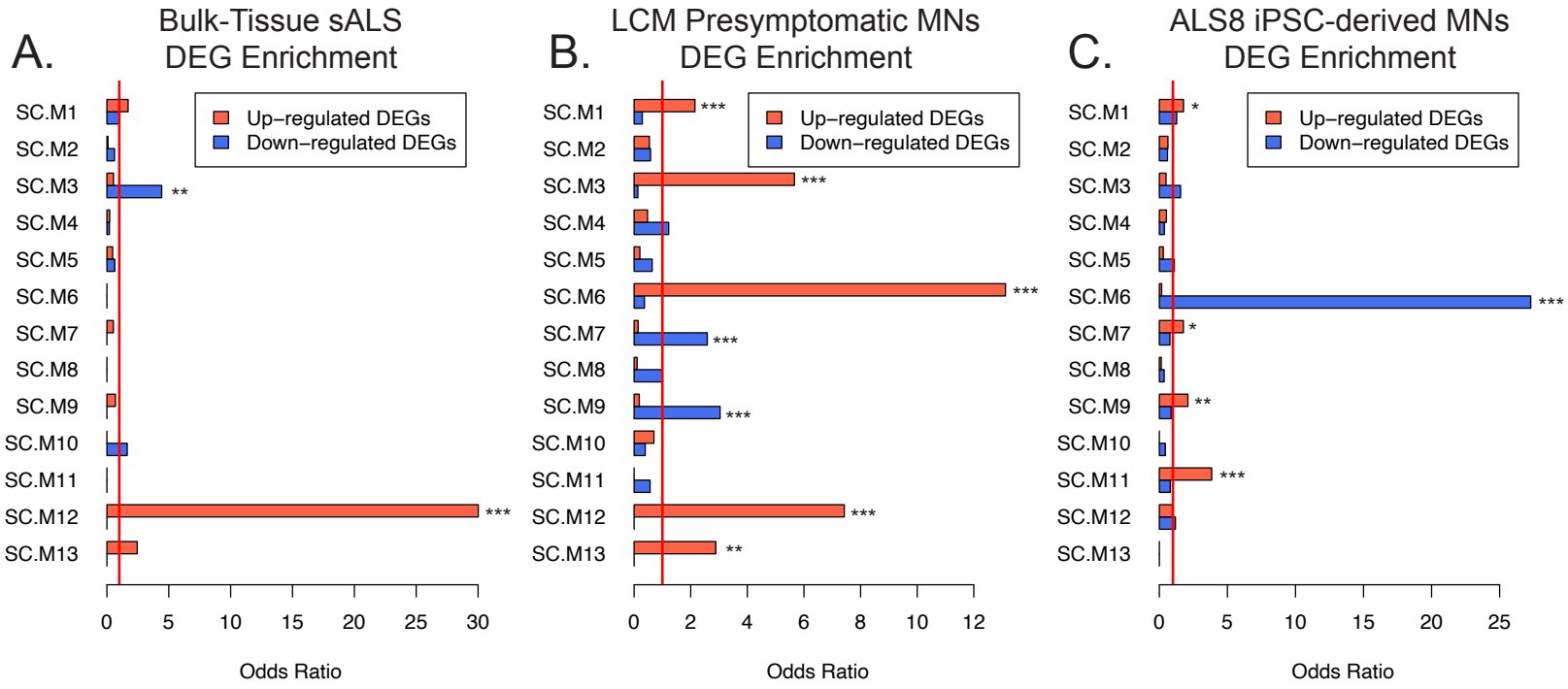
743 corrected p-values themselves are shown within parentheses.

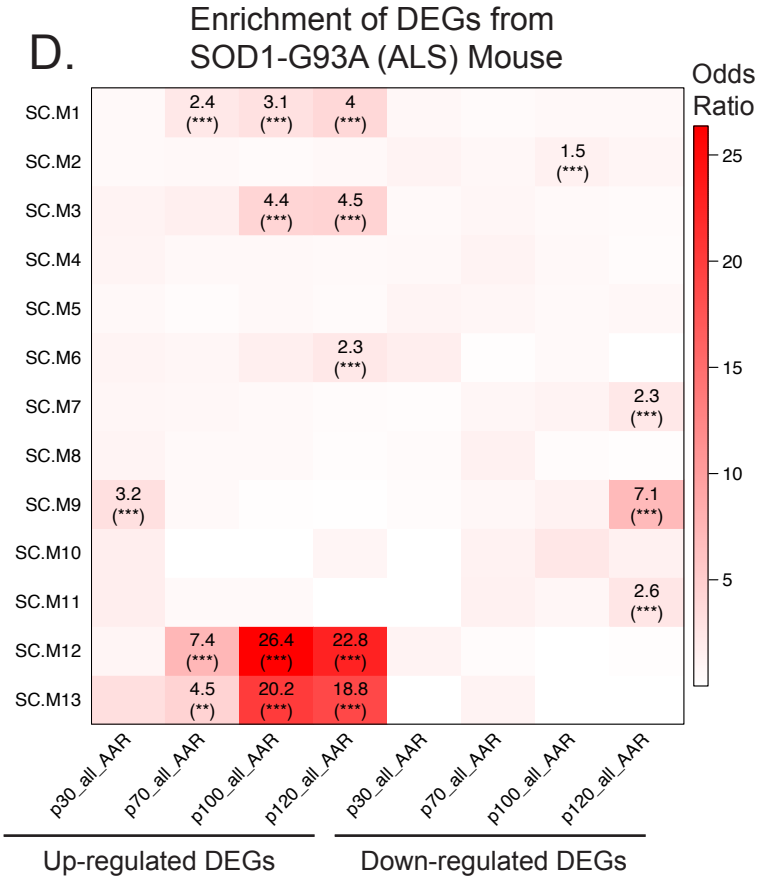
744 (B) Direct protein-protein interaction (PPI) network of the top 100 hub genes from module

745 SC.M6. Gene vertices are color coded as shown. P-value of the PPI network is derived from

746 1,000 permutations.

747





E.

Module	Number of Genes	Functional Annotations	Cell Type Enrichments	Genetic Enrichments	Tissue Dysregulation	Direction of Change	Stage of Change
SC.M1	522	Response to wounding, tube development, morphogenesis	Astrocyte	None	Laser microdissected presymptomatic human ALS MNs	Up-regulated	Presymptomatic
					iPSC-derived motor neurons from ALS8 patients	Up-regulated	N/A
					SOD1-G93A (ALS) mouse	Up-regulated	Onset - end stage
SC.M2	1610	Autophagy, intracellular transport, oligodendrocyte identity	Oligodendrocyte	Common ALS risk variants	SOD1-G93A (ALS) mouse	Down-regulated	Symptomatic
SC.M3	530	Cardiovascular development, morphogensis, angiogenesis	Endothelial	None	Bulk Tissue sALS	Down-regulated	End stage
					Laser microdissected presymptomatic human ALS MNs	Up-regulated	Presymptomatic
					SOD1-G93A (ALS) mouse	Up-regulated	Symptomatic - end stage
SC.M4	1199	RNA processing, gene regulation	None	Rare and common ALS risk variants	None	None	None
SC.M6	168	Ribosomal-related	None	Genetic modifiers of C9orf72 and FUS cytotoxicity	Laser microdissected presymptomatic human ALS MNs	Up-regulated	Presymptomatic
					iPSC-derived motor neurons from ALS8 patients	Down-regulated	N/A
					SOD1-G93A (ALS) mouse	Up-regulated	End stage
SC.M12	157	Leukocyte activation, immune response	Microglia	None	Bulk Tissue sALS	Up-regulated	End stage
					Laser microdissected presymptomatic human ALS MNs	Up-regulated	Presymptomatic
					SOD1-G93A (ALS) mouse	Up-regulated	Onset - end stage

748 **Figure 5.** Validation of co-expression module dynamics in human ALS and model system
749 datasets.
750 (A) Enrichments of 112 up-regulated and 70 down-regulated DEGs from bulk-tissue RNA-
751 sequencing of postmortem ALS cervical spinal cord[13] within co-expression modules. Red line
752 marks a baseline odds ratio of 1. * (FDR corrected p-value < 0.05). ** (FDR corrected p-value <
753 0.01). *** (FDR corrected p-value < 0.001).
754 (B) Same as Fig. 5a except for using 1290 up-regulated and 310 down-regulated DEGs from
755 RNA-sequencing of laser capture microdissections of motor neurons primed for degeneration [22].
756 (C) Same as Fig. 5a except for using 364 up-regulated and 289 down-regulated DEGs from
757 RNA-sequencing of iPSC-derived motor neurons from ALS8 patients[23].
758 (D) Enrichment of DEGs within each disease stage of the SOD1-G93A mouse[12] within co-
759 expression modules. p30, p70, p100, and p120 represent postnatal age in days at time of
760 spatial transcriptomic sequencing. Odds ratios of enrichments with FDR corrected p-value <
761 0.05 are shown. * (FDR corrected p-value < 0.05). ** (FDR corrected p-value < 0.01). *** (FDR
762 corrected p-value < 0.001).
763 (E) Summary of the key findings from this study regarding modules relevant to ALS
764 pathophysiology.