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,*]

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

*. 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 and integrated them with ALS genetic risk to identify biological pathways disrupted by ALS risk genes. We identified 13 gene co-expression modules, each of which represents a distinct biological process or cell type, finding enrichment of ALS genetic risk within two, SC.M4, representing genes related to RNA processing and gene regulation, and SC.M2, representing genes related to intracellular transport and autophagy. We also observe upregulation of SC.M2 in presymptomatic human motor neurons (MNs). Top hub genes of this module include ALS-implicated risk genes involved in intracellular transport and autophagy, including KPNA3, TMED2, and NCOA4, the latter of which regulates ferritin autophagy, implicating this process in ALS pathophysiology. These unbiased, genome-wide analyses confirm the importance of aberrant RNA processing, intracellular transport, and autophagy as drivers of ALS pathophysiology.

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[7,8,9,10]. 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). Furthermore, analysis of the transcriptional landscape in the spinal cord, the location of lower motor neuron degeneration in ALS, has not yet been fully explored[3].

To investigate gene expression signatures associated with ALS in the spinal cord, we generated gene co-expression networks using RNA-sequencing of 62 neurotypical human cervical spinal cord samples from the GTEx consortium[12]. We characterized these co-expression modules through gene ontology (GO) enrichment analysis, which reveals that the modules represent a diverse range of biological processes. To identify modules enriched with ALS genetic risk factors, we queried enrichment of rare, large effect risk mutations previously described in the literature, as well as common genetic variation[13,14]. The second approach is genome-wide and therefore unbiased, which focuses on the enrichment of small effect common risk variants identified from a Genome Wide Association Study (GWAS) of ALS[15]. By partitioning SNP based heritability and performing enrichment testing for genes with rare mutations, we find significant overlap of genes harboring rare and common variation in SC.M4, which represents genes involved in RNA processing and epigenetic regulation. We also find enrichment of ALS common genetic risk variants in SC.M2, which represents genes involved in intracellular transport, protein modification, and cellular catabolic processes. We extend these results by demonstrating the dysregulation of these modules in data from multiple published human ALS patient and animal model tissues[16,17,18], especially SC.M2, which changes at very early stages of disease. Taken together, our results highlight dysfunction in RNA processing, intracellular transport, and autophagy, as major processes underlying ALS pathogenesis, highlighting SC.M4 and SC.M2, the latter of which is enriched for genetic risk and dysregulated early in presymptomatic human motor neurons. This set of genes provides a rich set of targets for exploring new mechanisms for therapeutic development.

Results

Co-expression network analysis as a strategic framework for elucidating ALS susceptibility

86 We reasoned that because ALS preferentially affects spinal cord and upper motor neurons, co-
87 expression networks from spinal cords of control subjects would provide an unbiased context for
88 understanding potential convergence of ALS risk genes. We used the Genotype-Tissue
89 Expression (GTEx) Project[12] control cervical spinal cord samples to generate a co-expression
90 network as a starting point for a genome wide analysis to identify specific pathways or cell types
91 affected by ALS genetic risk. First, we identified gene co-expression modules and assigned
92 each module a biological, as well as cell-type identity[19,20]. Next, we leveraged multiple datasets
93 identifying ALS genetic risk factors[13–15,21–23] to comprehensively assess which pathways and
94 cell-types are susceptible to ALS genetic risk. Finally, we compared these predictions with
95 human ALS and animal model datasets[16–18] to develop a mechanistic understanding of ALS
96 etiology and progression (**Fig. 1**).
97
98 *Generation of human spinal cord gene co-expression network*
99 We utilized RNA-sequencing data from the GTEx consortium to generate a co-expression
100 network of human spinal cord (**Methods**). Co-expression networks are highly sensitive to
101 biological and technical confounders[20], which we controlled for by calculating the correlation of
102 biological and technical covariates with the top expression principal components (PCs)
103 (**Methods, Supplementary Fig. S1a-c**). We identified 11 technical and biological covariates:
104 seqPC1-5, RNA integrity number (RIN), Hardy's death classification scale, age, sampling
105 center, ethnicity, and ischemic time as significant drivers of expression (**Supplementary Fig.**
106 **S1a**). We regressed out the effect of these covariates using a linear model (**Methods**) and
107 verified that the regressed expression dataset was no longer correlated with technical and
108 biological covariates (**Supplementary Fig. S1b**).
109
110 We applied WGCNA to construct a co-expression network (**Fig. 2a, Supplementary Fig. S2;**
111 **Supplementary Data 1; Methods**) identifying 13 modules. To verify that the modules were not

112 driven by confounding covariates, we confirmed that the module eigengenes were not
113 correlated with the major technical and biological covariates (**Supplementary Fig. S1d**). We
114 also found that randomly sampled genes from within each module were significantly more inter-
115 correlated when compared to a background set, verifying the co-expression of genes within
116 each module (**Supplementary Fig. S1e**). To biologically annotate our co-expression modules,
117 we performed gene ontology (GO) term enrichment and identified the top hubs (**Fig. 2c-g, 3c-d,**
118 **Supplementary Fig. S3a-f**). GO analysis revealed that co-expression modules represented
119 coherent biological processes that we categorized into five distinct groupings (**Fig. 2b**),
120 encompassing morphogenesis and development, epigenetics and gene regulation, neuronal
121 signaling, immune activation, and intracellular transport, metabolism, and sensing functions.
122 Most modules were significantly enriched for markers of specific cell types, suggesting that they
123 represent gene co-expression also related to major cell classes, including SC.M1 (astrocytes),
124 SC.M2 (oligodendrocytes), SC.M3 (endothelial cells), SC.M9 and SC.M11 (neurons), SC.M10
125 (ependymal), and SC.M12 and SC.M13 (microglia) (**Supplementary Fig. S4**). These groupings,
126 both with regards to cell type and molecular pathways, represent diverse biological functions
127 that play key roles in human spinal cord biology.
128
129 *Enrichment of ALS genetic risk in intracellular transport/autophagy module SC.M2 and RNA*
130 *processing/gene regulation module SC.M4*
131 To address how these modules fit within the context of known ALS susceptibility factors, we
132 assessed whether a literature-curated set of well-known, large effect size, familial ALS risk
133 genes[13,14] were enriched within our co-expression modules (**Fig. 3a**). We found a nominally
134 significant enrichment with SC.M4 (OR = 2.52, P = 0.0124). However, the FDR adjusted p-value
135 was not significant (FDR= 0.174), likely due to the small number of familial ALS risk genes
136 identified to date. SC.M4 was highly enriched with GO terms related to RNA processing and
137 epigenetic regulation (**Fig. 3c**), consistent with the known role of several ALS risk genes in RNA

138 metabolism and function. To prioritize interacting partners to ALS risk genes found in SC.M4,
139 we leveraged a direct protein-protein interaction (PPI) network using the top 300 hub genes of
140 SC.M4, finding a significant network that includes ALS risk genes such as TARDBP and
141 EWSR1 (**Fig. 3e**, **Methods**, p = 0.001).
142
143 To improve power, we performed an unbiased genome-wide analysis, partitioning the common
144 SNP-based heritability across the co-expression modules using stratified LD score-regression[24]
145 based on summary statistics from the most recent GWAS of ALS (ALS n= 20,806; CTL n =
146 59,804)[15] (**Fig. 3b**). We found significant enrichment of ALS common risk variants in the SC.M2
147 (FDR = 0.047) and SC.M4 modules (FDR = 0.041). SC.M2 was enriched with GO terms related
148 to intracellular transport, protein modification, and cellular catabolic processes (**Fig. 3d**).
149 Additionally, SC.M2 was significantly enriched with oligodendrocyte markers (**Supplementary**
150 **Fig. S4,** Oligodendrocytes_mature: OR = 5.3, FDR < 10e-5; Oligodendrocytes_myelinating: OR
151 = 2.9, FDR < 0.05), while SC.M4 showed no specific cell type enrichment. Generation of a direct
152 protein-protein interaction (PPI) network using the top 300 hub genes of SC.M2 reveals a
153 significant network that highlights SC.M2 hub genes that interact with known ALS risk genes
154 such as TMED2 and KPNA3[22] (**Fig. 3f**, **Methods**, p = 0.018).
155
156 As a comparison, we tested for the enrichment of common risk variants from Inflammatory
157 Bowel Disorder (IBD)[25] across modules (**Fig. 3b**). We found a significant enrichment of IBD
158 common risk variants (FDR = 0.047) in module, SC.12, which is a microglial/immune module,
159 consistent with the known disease biology of IBD[26]. As a negative control, we performed
160 enrichments of common risk variants from Autism Spectrum Disorder (ASD)[27,28] and found no
161 enrichment in any module. Further, when we compared with Alzheimer's Disease (AD)[29],
162 Frontotemporal Dementia (FTD)[30], and Progressive Supranuclear Palsy (PSP)[31] (**Fig. 3b**), we
163 also found no significant enrichments for these diseases' common risk variants, demonstrating

164 the specificity of the ALS GWAS enrichment for modules SC.M2 and SC.M4. Overall, these
165 genetic enrichment analyses implicate dysfunction in RNA processing and intracellular
166 transport, including in oligodendrocytes, as putative drivers of ALS pathophysiology in the
167 human spinal cord.
168
169 *Enrichment of ALS genetic modifiers in ribosomal-associated module SC.M6*
170 As an orthogonal approach to assess which genes and pathways are mediators of ALS
171 pathophysiology, we leveraged genetic screen datasets from three studies of modifiers of
172 C9orf72 and FUS toxicity in yeast[21–23]. We assessed our normal human spinal cord modules for
173 enrichment with modifiers of C9orf72 and FUS toxicity identified from these screens. We found
174 significant enrichment of C9orf72 and FUS toxicity suppressors in SC.M6 (**Fig. 4a**), a module
175 representing genes involved in ribosomal function (**Fig. 2e**), but without any cell type specific
176 enrichment (**Supplementary Fig. S4**). This finding is consistent with previous observations that
177 modifiers of Glycine-Arginine$_{100}$ (GR$_{100}$) DPR and FUS toxicity are enriched with gene ontology
178 terms related to ribosomal biogenesis[21,23].
179
180 We reasoned that SC.M6 could be leveraged to prioritize highly interconnected hub genes
181 within the module that may warrant additional scrutiny as points of convergence for C9orf72 and
182 FUS pathology. To address this, we first identified the interacting partners for the top 100
183 SC.M6 hub genes by generating an indirect protein-protein interaction (PPI) network (**Methods**).
184 We identified 30 modifiers of C9orf72 and FUS toxicity that were present in this indirect PPI
185 network, either as hubs or as interacting partners (**Fig. 4b**)[21–23]. These modifiers include
186 ribosomal proteins (RPL and RPS-related genes), intracellular transporters of ribosomal
187 components (TNPO1), ribosomal assembly proteins (NOB1), and RNA-binding proteins
188 (HABP4) (**Fig. 4b**), indicating specific aspects of ribosomal dysfunction shared between C9orf72
189 and FUS toxicity. Indeed, previous studies[21] have suggested that the overlap in modifiers of

190 C9orf72 and FUS toxicity could be due to the sequence similarity between the
191 arginine/glycine/glycine (RGG) domain in FUS and the GR_{100} sequence expressed by
192 C9orf72[32,33].
193
194 *Dysregulation of immune-related modules in postmortem ALS spinal cord*
195 Until this point, we had assessed co-expression modules for their relevance to ALS based on
196 enrichment of genetic risk factors. To validate if the co-expression modules identified using the
197 GTEx dataset were directly dysregulated in ALS patients, we processed bulk RNA-sequencing
198 data from a previous study of postmortem control and ALS cervical spinal cord[16]. We corrected
199 the data for potential confounders such as age, ethnicity, sex, and sequencing variability and
200 assessed module preservation (**Supplementary Fig. S5a-b**, **Methods**). We identified 10 of the
201 13 modules generated from GTEx as preserved in the postmortem control and ALS cervical
202 spinal cord ($Z_{summary} > 2$) with four (SC.M2, SC.M4, SC.M5, and SC.M8) strongly preserved
203 ($Z_{summary} > 10$). The remaining three modules (SC.M9, SC.M11, and SC.M10) were not
204 preserved ($Z_{summary} < 2$) (**Supplementary Fig. S5c**).
205
206 We adopted two approaches to assess which modules were dysregulated in ALS in the human
207 spinal cord. The first assessed enrichments of differentially expressed genes (DEGs) between
208 postmortem control and ALS cervical spinal cord samples[16] (**Methods**). We found a significant
209 enrichment of DEGs upregulated in cervical spinal cord from ALS patients in SC.M12 (FDR =
210 5.92e-39) and of downregulated DEGs in SC.M3 (FDR = 0.00152) (**Fig. 5a**). SC.12 is highly
211 enriched with microglial markers and GO terms related to immune response (**Fig. 2f,**
212 **Supplementary Fig. S4a**), whereas SC.M3 is highly enriched with endothelial markers and GO
213 terms related to inflammation and morphogenesis (**Fig. 2c, Supplementary S4a**).
214

215 For our second approach, we assessed if the spinal cord modules were themselves differentially
216 expressed between ALS and control samples by looking at the relationship between the
217 eigengenes (first principle component; **Methods**) of each module and disease status
218 (**Supplementary Fig. S5d**). We found that the SC.M1 eigengene was significantly upregulated
219 in ALS samples. SC.M1 was enriched with astrocyte markers and GO terms related to wound
220 response (**Supplementary Fig. S3a, S4a**). Overall, our analyses identify a marked upregulation
221 of neuro-immune cells in the human post-mortem spinal cord, including astrocytes and
222 microglia, similar to previous reports[16,17]. However, we note that genes differentially expressed
223 in postmortem ALS spinal cord are not enriched in causal genetic risk, suggesting that these
224 changes are not causal drivers.
225
226 *Dysregulation of SC.M2 in human motor neurons primed for degeneration*
227 A major disadvantage of utilizing postmortem human tissues is that the molecular pathology is
228 indicative of late-stage disease. In particular, immune hyperactivation and infiltration is
229 consistently seen in neurodegenerative disorders including ALS. However, it is unclear if these
230 processes are disease causing or merely a reaction to neurodegenerative cell death[34]. To
231 identify early-stage differences in human spinal cord likely to be representative of causal factors,
232 we leveraged a previously published dataset of laser capture micro-dissected motor neurons
233 deemed to be primed for degeneration[17]. In this study, the authors isolated motor neurons from
234 ALS patients' postmortem spinal cord, but in regions that did not exhibit any signs of
235 degeneration at the time of death. We processed and corrected the data for potential
236 confounders such as age, gender, postmortem interval, RNA integrity, and sequencing
237 variability[17] (**Supplementary Fig. S6a-b**). In assessing module preservation, we did not observe
238 preservation of SC.M2 and SC.M4 (Z = 1.70 & 0.30 respectively), however we observed glial
239 infiltration into the LCM sections **(Supplementary Fig. S6c, S6e**), as seen in previous findings
240 from the authors of this dataset. To mitigate the effects of glial infiltration, the authors had

241 removed the top 1000 genes along expression PC1, which reduced the number of GO terms
242 related to immune response, while still differentiating ALS from control samples. We utilized the
243 same approach and re-assessed module preservation, finding that SC.M2 was mildly preserved
244 after removal of immune-related genes (Z = 2.50) (**Supplementary Fig. S6d**). Then, we
245 assessed differential expression of the preserved modules in this dataset and observed up-
246 regulation of SC.M2 and SC.M5 in this motor neuron data set (**Fig. 5b**). This finding indicates
247 that SC.M2, a module related to intracellular transport, and enriched for ALS genetic risk
248 variants, is dysregulated in ALS-susceptible motor neurons.
249
250 *Dysregulation of immune and neuronal-related modules in SOD1-G93A (ALS) mice*
251 Finally, as an orthogonal approach to identify temporal differences in ALS, we leveraged a
252 spatial transcriptomics dataset from spinal cord of SOD1-G93A mice over four time points
253 during disease progression[18,35,36]. To assess which modules were dysregulated over the
254 progression of disease, we performed an enrichment analysis between the human spinal cord
255 modules and mouse DEGs stratified by time point (**Methods**). Consistent with our findings in
256 postmortem tissues (**Fig. 5a-b, Supplementary Fig. S6e**), immune response modules (SC.M12
257 and SC.M13) were highly enriched with up-regulated DEGs as early as postnatal day 70, which
258 is the time of clinical disease onset in this model (**Fig. 5c, 2f, Supplementary Fig. S3f**). We
259 also noted a strong enrichment of down-regulated DEGs in a neuronal signaling modules
260 (SC.M7, SC.M9, and SC.M11) (**Supplementary Fig. S3b, Fig. 2d, Supplementary Fig. S3e**)
261 by postnatal day 120 (p120 - end-stage ALS), which is consistent with motor neuron
262 degeneration. Interestingly, we found enrichment of down-regulated DEGs in SC.M2 at p100
263 (symptomatic ALS), providing evidence that SC.M2 is dysregulated at symptomatic stages of
264 ALS in this model system (**Fig. 5c, Supplementary Fig. S8**). We also noted an enrichment of
265 up-regulated DEGs in SC.M6 at p120 (end-stage ALS), suggesting dysregulation of SC.M6 in
266 the end-stages of ALS in this model system (**Fig. 5c, Supplementary Fig. S7**). Overall,

267 enrichment analyses in SOD1-G93A (ALS) mice highlight immune and neuronal-related
268 modules as key disrupted pathways in the spinal cord and implicates the dysregulation of
269 SC.M2, a module enriched with ALS genetic risk and is dysregulated in presymptomatic motor
270 neurons, and SC.M6, a module enriched with ALS genetic modifiers, in later stages of disease
271 progression (**Fig. 5d**).
272
273 **Discussion**
274 To help elucidate the biological mechanisms underlying ALS pathophysiology in an unbiased
275 genome-wide manner, we generated co-expression networks using RNA-sequencing of control
276 human cervical spinal cord samples from the GTEx Consortium. We identified 13 total co-
277 expression modules each representing distinct aspects of biology in the control spinal cord
278 including inflammation, neurotransmission, RNA processing, and cellular metabolism, as well as
279 major cell classes. We find that ALS genetic risk factors are enriched in SC.M4, a co-expression
280 module representing genes involved in RNA processing and epigenetics and SC.M2, which is
281 enriched in oligodendrocyte markers and genes involved in intracellular transport, protein
282 modification, and cellular catabolic processes. Analyses using human ALS-susceptible motor
283 neurons exhibit up-regulation of SC.M2, suggesting that the genes in the SC.M2 module are
284 directly dysregulated in early stages of disease, further highlighting their potential value as
285 targets for disease modification.
286
287 It is notable that genes within both SC.M2 and SC.M4 have previously been shown to cause
288 ALS or have been strongly linked with risk. For example, SC.M4, which is enriched in RNA
289 processing, harbors RNA binding proteins mutated in ALS including TDP-43 and FUS/TLS,
290 which are abnormally aggregated and mis-localized in ALS affected neurons, leading to the loss
291 of normal RNA binding function with incitement of neurotoxicity[37]. One of the top hubs of the
292 SC.M4 module is PTPN23, a non-receptor-type tyrosine phosphatase that is a key regulator of

293 the survival motor neuron (SMN) complex, which helps assemble small ribonucleoproteins
294 particles (snRNPs) and mediates pre-mRNA processing in motor neurons[38]. Furthermore, two
295 other top hubs in SC.M2, KPNA3 and TMED2, are nucleocytoplasmic transporters that act as
296 modifiers of dipeptide repeat (DPR) toxicity in cases of C9orf72 mutations[22]. Taken together, the
297 presence of ALS-linked risk genes as hubs in SC.M2 and SC.M4 further underscores the
298 importance of aberrant RNA processing[37] and intracellular transport[39] as dysregulated
299 processes underlying ALS pathophysiology.
300
301 Intriguingly, PTPN23 and NCOA4, hubs in the SC.M4 and SC.M2 modules respectively, are
302 involved in the autophagy of ferritin[40,41], which may link iron metabolism and neurodegeneration
303 via these genes' involvement in iron metabolism and ferroptosis[42]. Ferritin has been found to be
304 elevated in serum and cerebrospinal fluid (CSF) from ALS patients[43–45], and recent work
305 implicates iron metabolism in muscle health, suggesting serum ferritin as a potential biomarker
306 in ALS[46]. Based on this, we hypothesize that disruptions in autophagy of ferritin are involved at
307 early stages of ALS pathophysiology. Because of recent evidence implicating markers of
308 ferroptosis in ALS progression and prognosis in patients[47], coupled with our current limited
309 understanding of ferritin autophagy in relation to ALS pathophysiology, our findings strongly
310 suggest that the roles of ferritin autophagy and ferroptosis in ALS warrant additional scrutiny.
311
312 A major question still remains concerning the specificity of MN degeneration in ALS. Our results
313 support the hypothesis that ALS is a non-cell-autonomous disease[48], as evidenced by SC.M2
314 being enriched with oligodendrocyte markers (**Supplementary Fig. S4a**). Oligodendrocytes
315 have critical roles through the metabolic support of neurons and neurotransmission in the
316 central nervous system[49]. A previous study has shown that oligodendrocytes harboring familial
317 and sporadic, but not C9orf72 variants of ALS, induce motor neuron (MN) death via a SOD1-
318 dependent mechanism, but can be rescued via lactate supplementation[50]. These previous

319 observations corroborate our findings that highlight the potential deleterious effect of
320 dysregulation in the intracellular transport and autophagy system on oligodendrocyte - MN
321 dynamics in ALS. Invariably, understanding of cell-type specific mechanisms governing MN
322 degeneration will involve future interrogation of the transcriptome and epigenome from ALS
323 spinal cord at single-cell resolutions. Additionally, emerging RNA-sequencing datasets,
324 including from the TargetALS foundation (http://www.targetals.org), contain larger sample sizes
325 of ALS and control spinal cord that will help elucidate the cellular dynamics underlying ALS
326 pathophysiology.
327
328 **Methods**
329 *Initial processing of RNA-sequencing datasets*
330 The latest version of GTEx data (version 7) was obtained from the dbGaP accession number
331 phs000424.v7.p2 on 10/10/2017 and pre-processed as previously described[12]. Genes in the
332 GTEx dataset containing fewer than 10 samples with a TPM > 1 were excluded from
333 downstream analysis on the basis of unreliable quantification for very lowly expressed genes.
334 We removed 4 outlier samples in the GTEx dataset that had sample-sample connectivity Z-
335 scores of > 2 as previously described[51]. Our starting GTEx dataset consisted of 62 samples and
336 14577 genes.
337
338 RNA-sequencing data[16] from postmortem cervical spinal cord was downloaded using the SRP
339 accession number SRP064478 on 5/24/2019. RNA-sequencing data[17] from laser capture
340 microdissected (LCM) presymptomatic MNs was downloaded using the SRP accession number
341 SRP067645 on 6/27/2019. Reads were aligned to the reference genome GRCh37 using STAR
342 (version 2.5.2b) and quantified using RSEM (version 1.3.0) with Gencode v19 annotations.
343 Genes in the Brohawn et al. and Krach et al. datasets with zero variance in expression were
344 excluded from further analysis. All expression data was quantified in TPM (transcript per million

345 reads) and TPM values +1 were log_2 transformed to stabilize their variance. The starting
346 Brohawn et al. and Krach et al. datasets consisted of 33050 genes and 37138 genes
347 respectively. Krach et al. metadata tables were transcribed from the manuscript and missing
348 values for PMI (postmortem interval) were imputed as the mean of all available PMI values.
349
350 *Identification and regression of biological and technical confounders from gene expression*
351 *datasets*
352 Gene expression values were subjected to dimensionality reduction using principal component
353 analysis (PCA) with data centering and scaling. The top four components for the GTEx dataset,
354 which collectively captures 70.6% of the total variance, were assigned to be expression principal
355 components (PCs) 1-4. The top five components for the Brohawn et al. and the Krach et al.
356 dataset, which collectively capture 55.9% and 42.3% of the total variance respectively, were
357 assigned to be expression principal components (PCs) 1-5. Sequencing Q/C metrics were
358 calculated using PicardTools (version 2.5.0) and dimensionally reduced using PCA with
359 centering and scaling. The top five components of sequencing metrics for the GTEx, Brohawn et
360 al., and Krach et al. datasets, which collectively capture 75.1%, 91%, and 90.4% of the total
361 variance respectively, were selected and assigned to be sequencing principal components
362 (seqPCs) 1-5.
363
364 Numerical covariates of technical or biological importance were correlated with the expression
365 PCs using Spearman's correlation. Binary covariates were numerically encoded and multi-level
366 categorical covariates of technical or biological importance were binarized using dummy
367 variables. The model fit (R^2) was used to assess correlations. We identified technical or
368 biological covariates with a meaningful influence on gene expression if they were correlated with
369 an $R > 0.3$ to any of the top expression PCs.
370

371 For the GTEx expression dataset, a linear model was fit for each gene as follows: GTEX.Expr ~
372 seqPC1 + seqPC2 + seqPC3 + seqPC4 + seqPC5 + RIN + DTHHRDY + AGE + SMCENTER +
373 TRISCHD and the effect of all covariates were regressed out. RIN indicates RNA integrity
374 number, DTHHRDY represents the Hardy Scale, SMCENTER represents the center for
375 sampling, and TRISCHD represents ischemic time.
376
377 For the Brohawn et al. dataset, a linear model was fit for each gene as follows: Expr ~ seqPC1
378 + seqPC2 + seqPC3 + seqPC4 +seqPC5 + Ethnicity + Sex + Age + Disease and the effect of all
379 covariates except disease were regressed out.
380
381 For the Krach et al. dataset, a linear model was fit for each gene as follows: Expr ~ Diagnosis +
382 Age + Gender + RIN + PMI + seqPC1 + seqPC2 + seqPC3 + seqPC4 + seqPC5 and the effect
383 of all covariates except diagnosis were regressed out.
384
385 *Weighted Gene Co-expression Network Analysis (WGCNA)*
386 We used WGCNA[4,51] to generate a gene co-expression network using the GTEx dataset. We
387 chose a soft power of 10 using the function “pickSoftThreshold” in the WGCNA R package to
388 achieve optimal scale free topology. The Topological Overlap Matrix (TOM) was generated
389 using the function “TOMsimilarity” in the WGCNA R package. A hierarchical gene clustering
390 dendrogram was constructed using the TOM-based dissimilarity and module selection was
391 tested using multiple permutations of the tree cutting parameters: minimum module size, cut
392 height, and deep split. Minimum module size (mms) was assessed at 50, 100, and 150. Module
393 merging (dcor) was assessed at 0.1, 0.2, and 0.25. Deep split (DS) was assessed at 2 and 4.
394 By visual inspection, we utilized the following parameters: minimum module size = 100, module
395 merging = 0.1, and deep split = 2 to generate the network. We assigned unique module

396 identifiers (SC.M1-13) for each module color except the grey module which represents genes
397 that were not co-expressed.
398
399 Pairwise Spearman's correlations were calculated using 1000 randomly sampled genes within
400 each module and a matched background set. If a module contained less than 1000 genes, all
401 genes in the module were sampled.
402
403 *Enrichment of gene sets in co-expression modules*
404 We analyzed a collection of gene sets generated by previous genetic studies of ALS. For large
405 effect size ALS risk genes, we obtained 50 ALS literature curated ALS risk genes from two
406 studies[13,14]. For the determination of cell-type specificity, we obtained the top 50 differentially
407 expressed genes per cell-type by fold change from single-nuclei RNA sequencing of postnatal
408 day 2 mouse spinal cord[19]. For the human ALS spinal cord differentially expressed gene (DEG)
409 enrichments, we obtained 265 DEGs (172 up- and 93 down-regulated) at FDR < 0.10 using the
410 union of DESeq2 and EdgeR analyses[16]. For DEGs from SOD1-G93A mice, we obtained genes
411 with a Bayes Factor > 3 (corresponding to FDR = 0.1) identified through a spatial
412 transcriptomics study[18].
413
414 Logistic regression was performed to calculate an odds ratio and p-value for enrichment of each
415 co-expression module with a test gene set. The p-values were FDR-corrected for the number of
416 modules analyzed.
417
418 *Enrichment of common ALS genetic risk variants in co-expression modules*
419 Stratified LD score regression[24] was performed to evaluate the enrichment of common risk
420 variants from GWAS studies of ALS[15], Inflammatory Bowel Disorder (IBD)[25], Autism Spectrum
421 Disorder (ASD)[27,28], Alzheimer's Disease (AD)[29], Frontotemporal Dementia (FTD)[30], and

422 Progressive Supranuclear Palsy (PSP)[31] within each co-expression module. We used the full
423 baseline model of 53 functional categories per calculation of partitioned heritability (1000
424 Genomes Phase 3). We defined genomic regions of each co-expression module as their
425 constituent gene bodies plus a flanking region of +/- 10 KB.
426
427 *Gene Ontology Enrichments*
428 Gene Ontology enrichment analysis was performed using the gProfiler package[52] with a
429 maximum set size of 1000. To reflect the weight of each gene within a co-expression module,
430 genes were ordered by their module membership (kME) during the enrichment calculations.
431
432 *Network visualization of module hub genes*
433 For each module, an undirected and weighted network amongst the top 30 module hub genes
434 (by kME) was visualized using the igraph package in R[53].
435
436 *Protein-protein interaction (PPI) networks*
437 PPI networks were produced using the Disease Association Protein-Protein Evaluator
438 (DAPPLE)[54]. Two hundred and thirty eight of the top 300 SC.M4 hub genes, 238 of the top 300
439 SC.M2 hub genes, 84 of top 100 SC.M6 hub genes as ranked by Module Membership (kME)
440 were found in the DAPPLE database and seeded into the shown PPI network (**Fig. 3e, 3f, 4b**).
441 All networks were produced from 1,000 permutations (within-degree node-label permutation).
442 SC.M6, SC.M4, and SC.M2 contain the following network properties respectively: direct edges
443 count ($p = 0.001$, $p = 0.001$, $p = 0.001$), seed direct degrees mean ($p = 0.001$, $p = 0.001$, $p =$
444 0.018), seed indirect degrees mean ($p = 0.001$, $p = 0.006$, $p = 0.006$), and the CI degrees mean
445 ($p = 0.001$, $p = 0.132$, $p = 0.332$).
446
447 *Module preservation analyses*

448 Module preservation for co-expression modules in the Brohawn et al. and Krach et al. datasets
449 were performed as previously described using the modulePreservation function in the WGCNA
450 R package[51]. We performed 200 permutation tests using a randomSeed of 1.
451
452 *Correlations of module eigengenes to biological and technical covariates*
453 To evaluate the correlations of module eigengenes to biological and technical covariates, a
454 Student p-value was calculated using the function “corPvalueStudent” in the WGCNA R
455 package. P-values were FDR corrected across the modules and plotted using the function
456 “labeledHeatmap” in the WGCNA R package.
457
458 **Data Availability**
459 Module assignments are provided as Supplementary Data 1 and the full range of values
460 underlying the heatmaps in Figures 3a-b, 4a, 5a-c, and Supplementary Figures S4, S5d, S6e,
461 S7, S8 are provided as a Source Data File.
462

References

1. Ravits, J. *et al.* Deciphering amyotrophic lateral sclerosis: What phenotype, neuropathology and genetics are telling us about pathogenesis. *Amyotroph. Lateral Scler. Front. Degener.* **14**, 5–18 (2013).
2. Taylor, J. P., Brown, R. H. & Cleveland, D. W. Decoding ALS: from genes to mechanism. *Nature* **539**, 197–206 (2016).
3. Hardiman, O. *et al.* Amyotrophic lateral sclerosis. *Nat. Rev. Dis. Primer* **3**, 17071 (2017).
4. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
5. Zhang, B. *et al.* Integrated Systems Approach Identifies Genetic Nodes and Networks in Late-Onset Alzheimer's Disease. *Cell* **153**, 707–720 (2013).
6. Mostafavi, S. *et al.* A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. *Nat. Neurosci.* **21**, 811–819 (2018).
7. Rexach, J., Swarup, V., Chang, T. & Geschwind, D. *Dementia risk genes engage gene networks poised to tune the immune response towards chronic inflammatory states*. http://biorxiv.org/lookup/doi/10.1101/597542 (2019) doi:10.1101/597542.
8. Swarup, V. *et al.* Identification of evolutionarily conserved gene networks mediating neurodegenerative dementia. *Nat. Med.* **25**, 152–164 (2019).
9. Johnson, E. C. B. *et al.* Large-scale proteomic analysis of Alzheimer's disease brain and cerebrospinal fluid reveals early changes in energy metabolism associated with microglia and astrocyte activation. *Nat. Med.* **26**, 769–780 (2020).
10. Swarup, V. *et al.* Identification of Conserved Proteomic Networks in Neurodegenerative Dementia. *Cell Rep.* **31**, 107807 (2020).
11. Prudencio, M. *et al.* Distinct brain transcriptome profiles in C9orf72-associated and sporadic ALS. *Nat. Neurosci.* **18**, 1175–1182 (2015).
12. GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
13. Eykens, C. & Robberecht, W. The genetic basis of amyotrophic lateral sclerosis: recent breakthroughs. *Advances in Genomics and Genetics* https://www.dovepress.com/the-genetic-basis-of-amyotrophic-lateral-sclerosis-recent-breakthrough-peer-reviewed-fulltext-article-AGG (2015) doi:10.2147/AGG.S57397.
14. Chia, R., Chiò, A. & Traynor, B. J. Novel genes associated with amyotrophic lateral sclerosis: diagnostic and clinical implications. *Lancet Neurol.* **17**, 94–102 (2018).
15. Nicolas, A. *et al.* Genome-wide Analyses Identify KIF5A as a Novel ALS Gene. *Neuron* **97**, 1268-1283.e6 (2018).
16. Brohawn, D. G., O'Brien, L. C. & Bennett, J. P. RNAseq Analyses Identify Tumor Necrosis Factor-Mediated Inflammation as a Major Abnormality in ALS Spinal Cord. *PLOS ONE* **11**, e0160520 (2016).
17. Krach, F. *et al.* Transcriptome–pathology correlation identifies interplay between TDP-43 and the expression of its kinase CK1E in sporadic ALS. *Acta Neuropathol. (Berl.)* **136**, 405–423 (2018).
18. Maniatis, S. *et al.* Spatiotemporal dynamics of molecular pathology in amyotrophic lateral sclerosis. *Science* **364**, 89–93 (2019).
19. Rosenberg, A. B. *et al.* Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360**, 176–182 (2018).
20. Parikshak, N. N., Gandal, M. J. & Geschwind, D. H. Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. *Nat. Rev. Genet.* **16**, 441–458 (2015).

21. Chai, N. & Gitler, A. D. Yeast screen for modifiers of C9orf72 poly(glycine-arginine) dipeptide repeat toxicity. *FEMS Yeast Res.* **18**, (2018).
22. Jovičić, A. *et al.* Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. *Nat. Neurosci.* **18**, 1226–1229 (2015).
23. Sun, Z. *et al.* Molecular Determinants and Genetic Modifiers of Aggregation and Toxicity for the ALS Disease Protein FUS/TLS. *PLoS Biol.* **9**, e1000614 (2011).
24. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
25. Liu, J. Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979–986 (2015).
26. Xavier, R. J. & Podolsky, D. K. Unravelling the pathogenesis of inflammatory bowel disease. *Nature* **448**, 427–434 (2007).
27. Grove, J. *et al. Common risk variants identified in autism spectrum disorder*. http://biorxiv.org/lookup/doi/10.1101/224774 (2017) doi:10.1101/224774.
28. The Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol. Autism* **8**, 21 (2017).
29. Jansen, I. E. *et al.* Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* **51**, 404–413 (2019).
30. Ferrari, R. *et al.* Frontotemporal dementia and its subtypes: a genome-wide association study. *Lancet Neurol.* **13**, 686–699 (2014).
31. Chen, J. A. *et al.* Joint genome-wide association study of progressive supranuclear palsy identifies novel susceptibility loci and genetic correlation to neurodegenerative diseases. *Mol. Neurodegener.* **13**, 41 (2018).
32. Boeynaems, S. *et al.* Phase Separation of C9orf72 Dipeptide Repeats Perturbs Stress Granule Dynamics. *Mol. Cell* **65**, 1044-1055.e5 (2017).
33. Ozdilek, B. A. *et al.* Intrinsically disordered RGG/RG domains mediate degenerate specificity in RNA binding. *Nucleic Acids Res.* **45**, 7984–7996 (2017).
34. Cooper-Knock, J. *et al.* A data-driven approach links microglia to pathology and prognosis in amyotrophic lateral sclerosis. *Acta Neuropathol. Commun.* **5**, 23 (2017).
35. Gurney, M. E. *et al. Motor neuron degeneration in mice that express a human Cu/Zn superoxide dismutase mutation. Science 264: 1772–5*. (Science, 1994).
36. Ripps, M. E., Huntley, G. W., Hof, P. R., Morrison, J. H. & Gordon, J. W. Transgenic mice expressing an altered murine superoxide dismutase gene provide an animal model of amyotrophic lateral sclerosis. *Proc. Natl. Acad. Sci.* **92**, 689–693 (1995).
37. Polymenidou, M. *et al.* Misregulated RNA processing in amyotrophic lateral sclerosis. *Brain Res.* **1462**, 3–15 (2012).
38. 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).
39. Kim, H. J. & Taylor, J. P. Lost in Transportation: Nucleocytoplasmic Transport Defects in ALS and Other Neurodegenerative Diseases. *Neuron* **96**, 285–297 (2017).
40. Goodwin, J. M. *et al.* Autophagy-Independent Lysosomal Targeting Regulated by ULK1/2-FIP200 and ATG9. *Cell Rep.* **20**, 2341–2356 (2017).
41. 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).
42. Quiles del Rey, M. & Mancias, J. D. NCOA4-Mediated Ferritinophagy: A Potential Link to Neurodegeneration. *Front. Neurosci.* **13**, 238 (2019).

43. Goodall, E. F., Haque, M. S. & Morrison, K. E. Increased serum ferritin levels in amyotrophic lateral sclerosis (ALS) patients. *J. Neurol.* **255**, 1652–1656 (2008).
44. Zheng, Y., Gao, L., Wang, D. & Zang, D. Elevated levels of ferritin in the cerebrospinal fluid of amyotrophic lateral sclerosis patients. *Acta Neurol. Scand.* **136**, 145–150 (2017).
45. Ward, R. J., Zucca, F. A., Duyn, J. H., Crichton, R. R. & Zecca, L. The role of iron in brain ageing and neurodegenerative disorders. *Lancet Neurol.* **13**, 1045–1060 (2014).
46. Halon-Golabek, M., Borkowska, A., Herman-Antosiewicz, A. & Antosiewicz, J. Iron Metabolism of the Skeletal Muscle and Neurodegeneration. *Front. Neurosci.* **13**, 165 (2019).
47. Devos, D. *et al.* A ferroptosis–based panel of prognostic biomarkers for Amyotrophic Lateral Sclerosis. *Sci. Rep.* **9**, 2918 (2019).
48. Ilieva, H., Polymenidou, M. & Cleveland, D. W. Non–cell autonomous toxicity in neurodegenerative disorders: ALS and beyond. *J. Cell Biol.* **187**, 761–772 (2009).
49. Simons, M. & Nave, K.-A. Oligodendrocytes: Myelination and Axonal Support. *Cold Spring Harb. Perspect. Biol.* **8**, a020479 (2016).
50. Ferraiuolo, L. *et al.* Oligodendrocytes contribute to motor neuron death in ALS via SOD1-dependent mechanism. *Proc. Natl. Acad. Sci.* **113**, E6496–E6505 (2016).
51. Horvath, S. *Weighted Network Analysis: Applications in Genomics and Systems Biology*. (Springer-Verlag, 2011).
52. Reimand, J. *et al.* g:Profiler—a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res.* **44**, W83–W89 (2016).
53. Csárdi, G. & Nepusz, T. The igraph software package for complex network research. in (2006).
54. 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).
55. 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

603 J.C.W, G.R, and D.H.G planned the analyses and wrote the manuscript. Analyses were

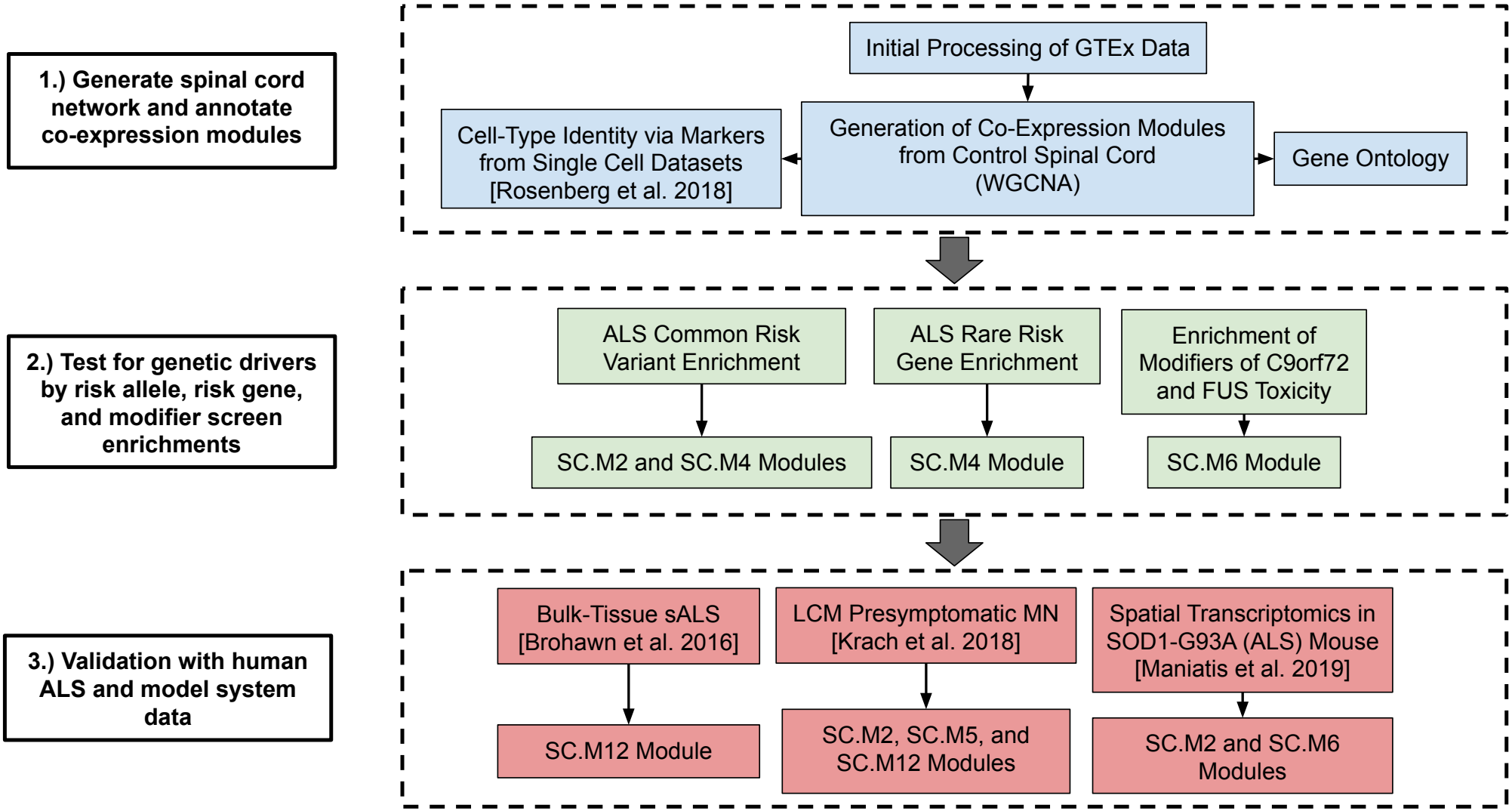
604 performed by J.C.W with supervision from G.R. and D.H.G.

605

606 **Competing Financial Interests**

607 The authors declare no competing financial interests.

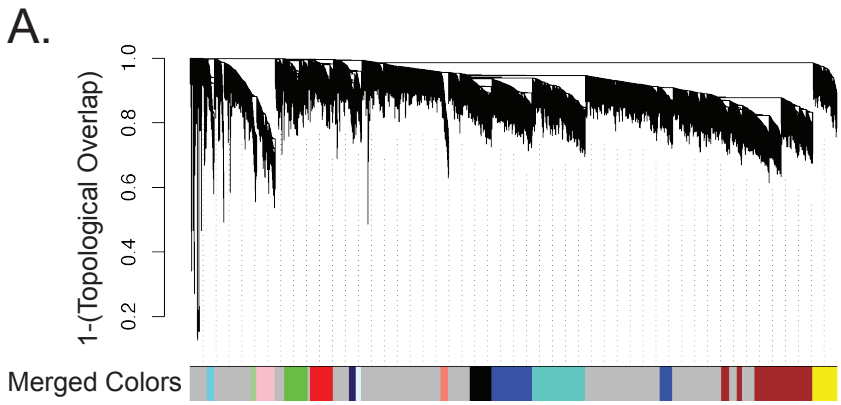
608


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 Risk Gene Enrichment
Enrichment of Modifiers of C9orf72 and FUS Toxicity
SC.M2 and SC.M4 Modules
SC.M4 Module
SC.M6 Module
3.) Validation with human ALS and model system data
Bulk-Tissue sALS [Brohawn et al. 2016]
LCM Presymptomatic MN [Krach et al. 2018]
Spatial Transcriptomics in SOD1-G93A (ALS) Mouse [Maniatis et al. 2019]
SC.M12 Module
SC.M2, SC.M5, and SC.M12 Modules
SC.M2 and SC.M6 Modules

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

610 Flow diagram of analyses in this study. A co-expression network was generated using the GTEx
611 cervical spinal cord expression data and Weighted Gene Co-Expression Network Analysis
612 (WGCNA). Modules were annotated by their associated biological processes, hub genes, and
613 cell-type identities. Modules were then assessed for their enrichment of ALS genetic risk factors
614 and analyzed for differential expression in RNA-seq data from human ALS tissues and model
615 system studies[16–18].

616



B.

Molecular Themes for Co-expression Modules

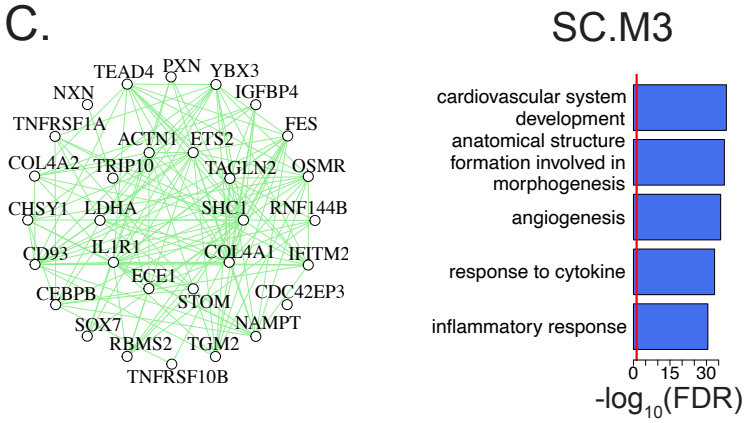
Morphogenesis & Development
Red (SC.M1)
Green (SC.M3)

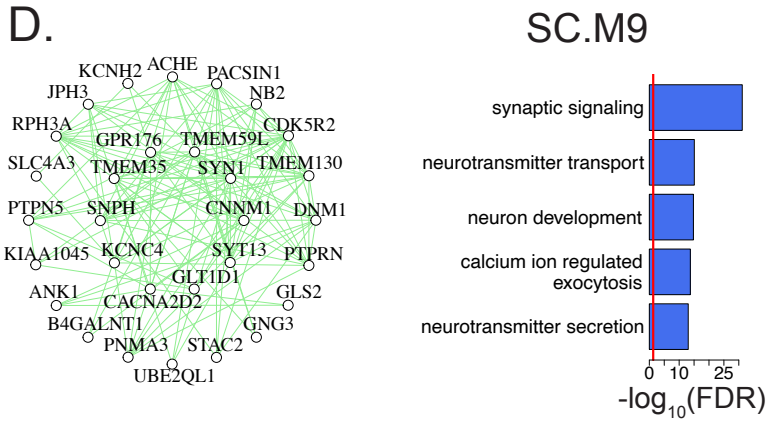
Neuronal Signaling
Yellow (SC.M7)
Pink (SC.M9)
Light Green (SC.M11)

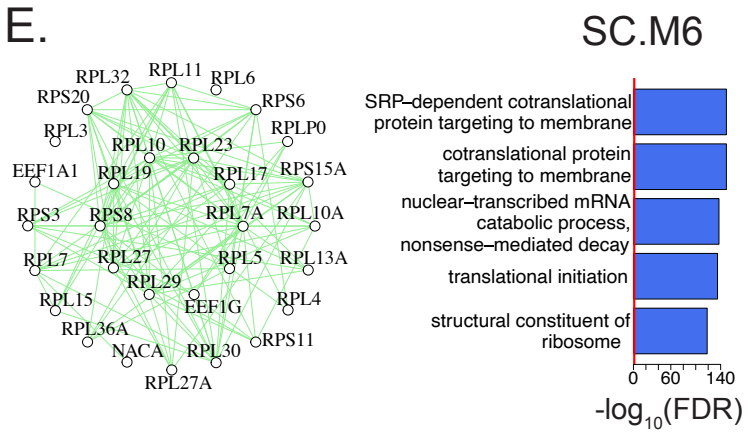
Epigenetics & Gene Regulation
Blue (SC.M4)
Salmon (SC.M6)
Black (SC.M8)

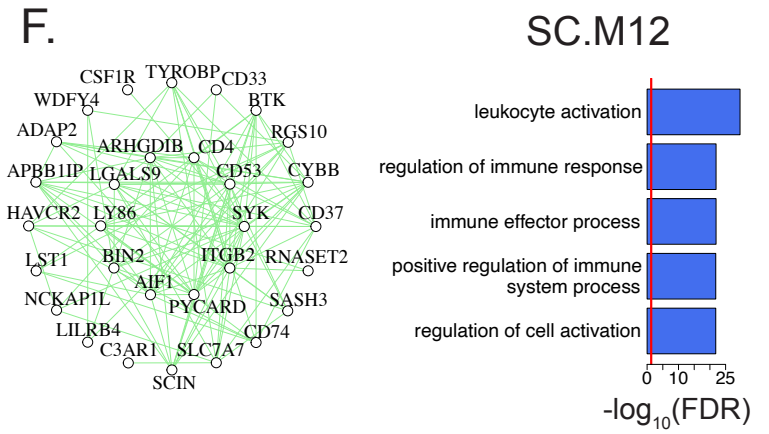
Immune Activation
Midnight Blue (SC.M12)
Light Cyan (SC.M13)

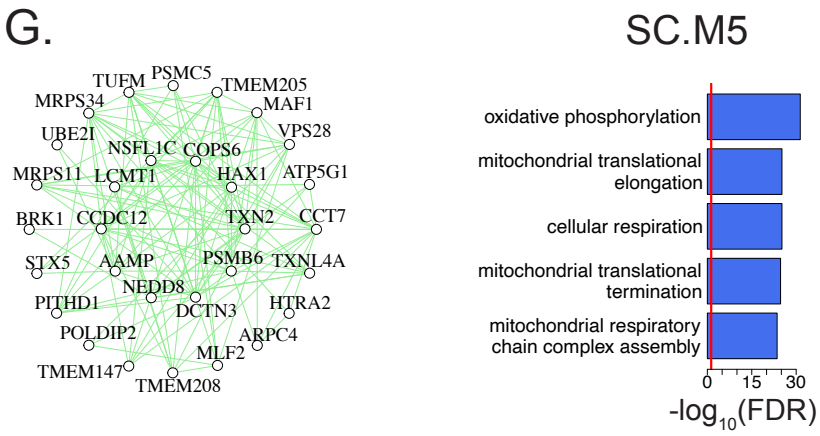
Intracellular Transport, Metabolism, & Sensing
Brown (SC.M2)
Turquoise (SC.M5)
Cyan (SC.M10)











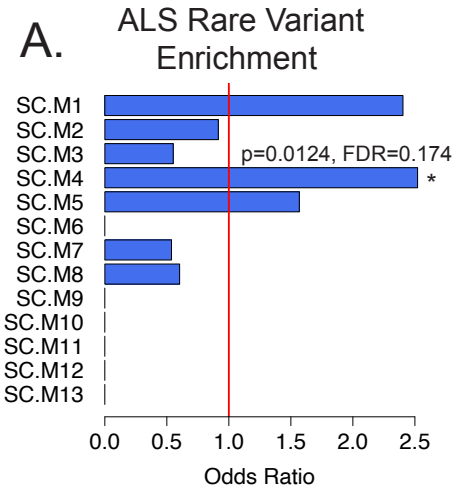
617 **Figure 2**. Construction of co-expression network in control cervical spinal cord.

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

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

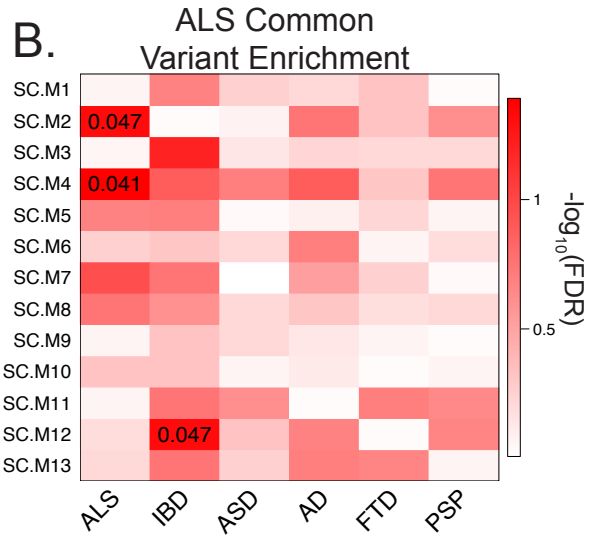
624 (C, D, E, F, G) Top 30 hub genes and 300 connections for SC.M3, SC.M9, SC.M6, SC.M12,
625 and SC.M5 respectively (left panels). Top 5 enriched GO terms for each of the respective
626 modules (right panels). Red lines mark an FDR corrected p-value threshold of 0.05.

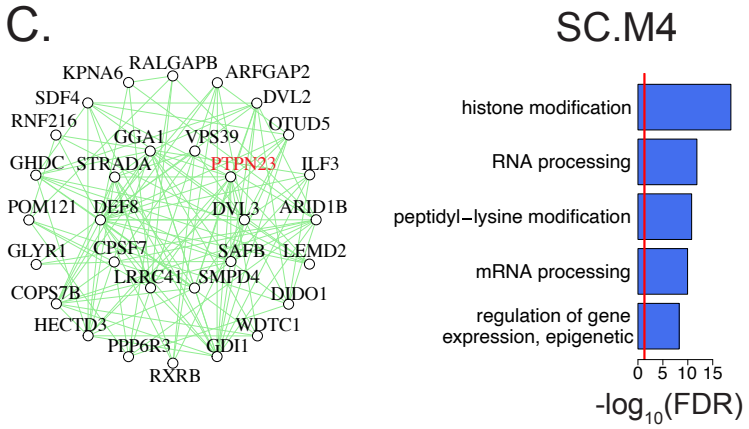
627

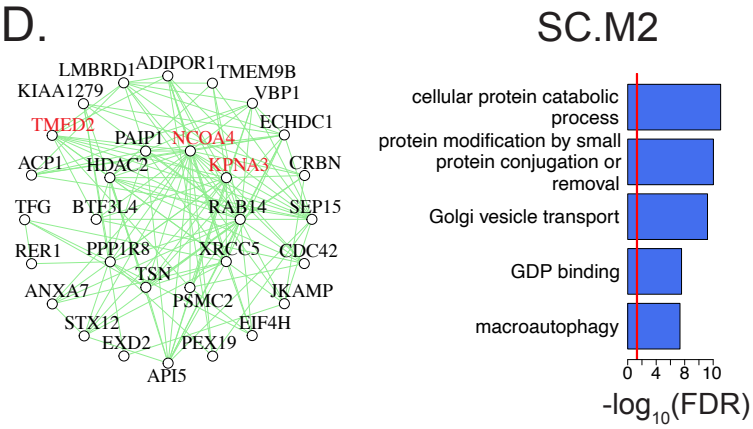


Overlapping ALS Risk Genes in SC.M4

FUS	C21orf2
SPG11	EWSR1
SETX	SS18L1
TARDBP	ZNF512B
ATXN2	



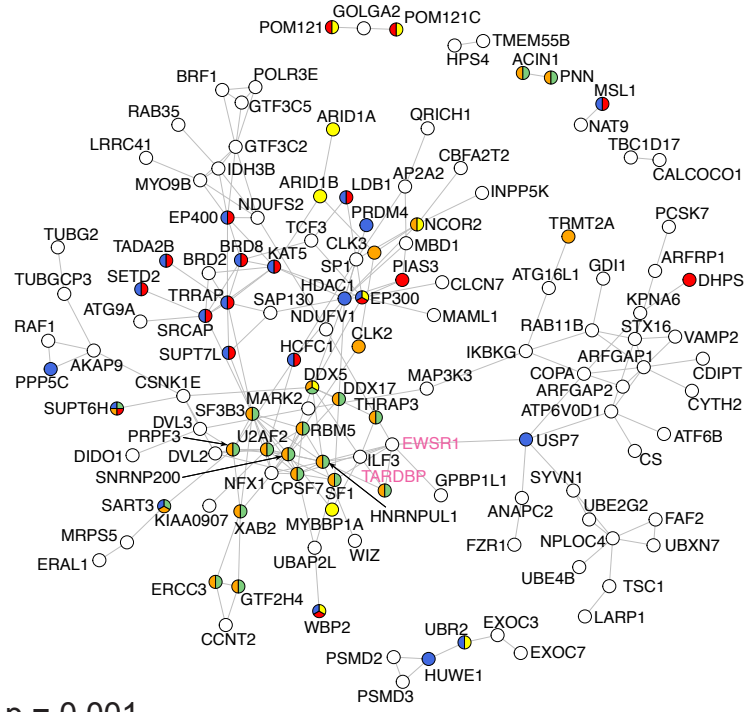




E.

Key

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

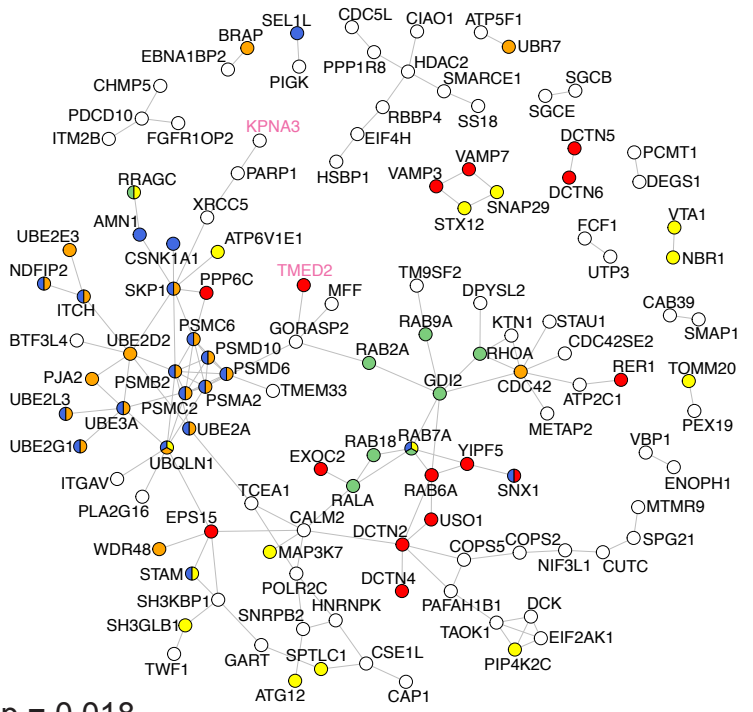


p = 0.001

F.

Key

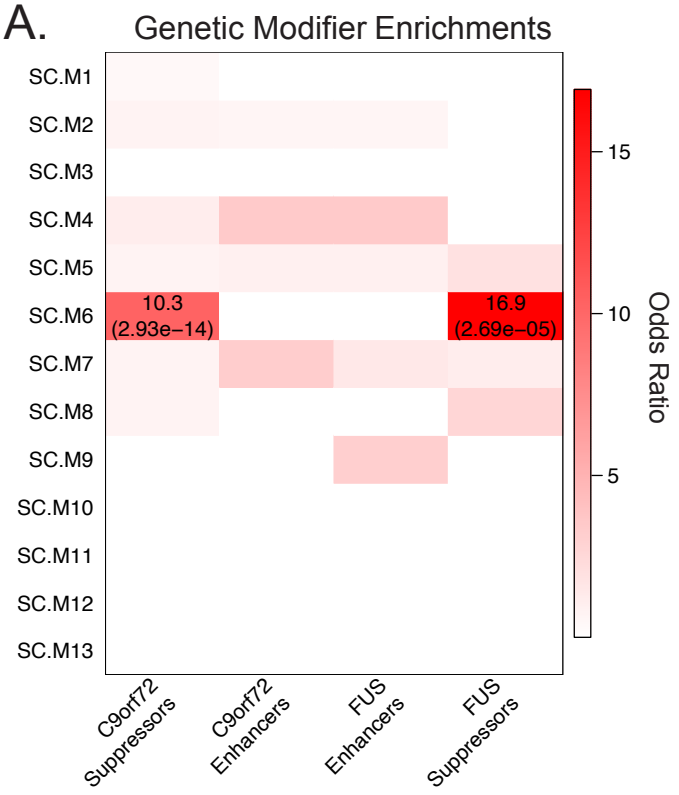
Blue	cellular protein catabolic process
Orange	protein modification by small protein conjugation or removal
Red	golgi vesicle transport
Green	GDP binding
Yellow	macroautophagy



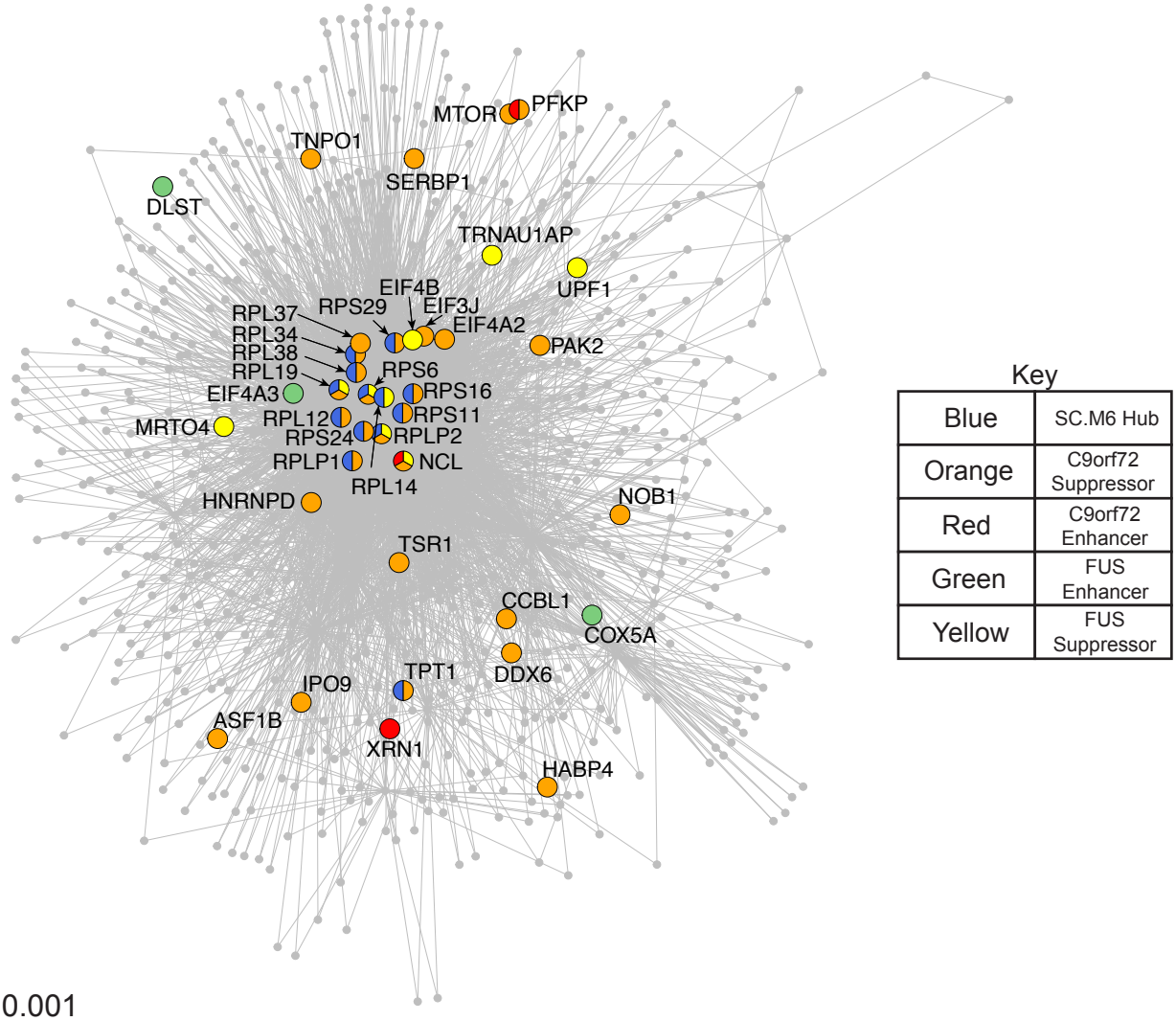
p = 0.018

628 **Figure 3.** Enrichment of ALS genetic risk factors in intracellular transport/autophagy module
629 SC.M2 and RNA processing/gene regulation module SC.M4.
630 (A) Enrichment of literature curated large-effect ALS risk genes[13,14] within co-expression
631 modules (left panel). Red line marks a baseline odds ratio of 1. SC.M4 is marked with an
632 asterisk as having a nominal enrichment p-value < 0.05. Literature-curated large-effect ALS risk
633 genes within the SC.M4 module (right panel).
634 (B) Partitioned heritability enrichments of common risk variants for ALS[15], IBD[25], ASD[27,28], AD[29],
635 FTD[30], and PSP[31] within co-expression modules. Text values of enrichments with FDR
636 corrected p-value < 0.05 are shown.
637 (C, D) Top 30 hub genes and 300 connections for SC.M4 and SC.M2 respectively (left panels).
638 Highlighted in red are genes relevant to ALS pathophysiology. Top 5 enriched GO terms for
639 each of the respective modules (right panels). Red lines mark an FDR corrected p-value
640 threshold of 0.05.
641 (E, F) Direct protein-protein interaction (PPI) network of the top 300 hub genes from modules
642 SC.M4 and SC.M2 respectively. Gene vertices are color coded according to their Gene
643 Ontology terms as shown[55]. P-value of the PPI network is derived from 1,000 permutations.
644 Highlighted in pink are genes relevant to ALS pathophysiology.
645

A.

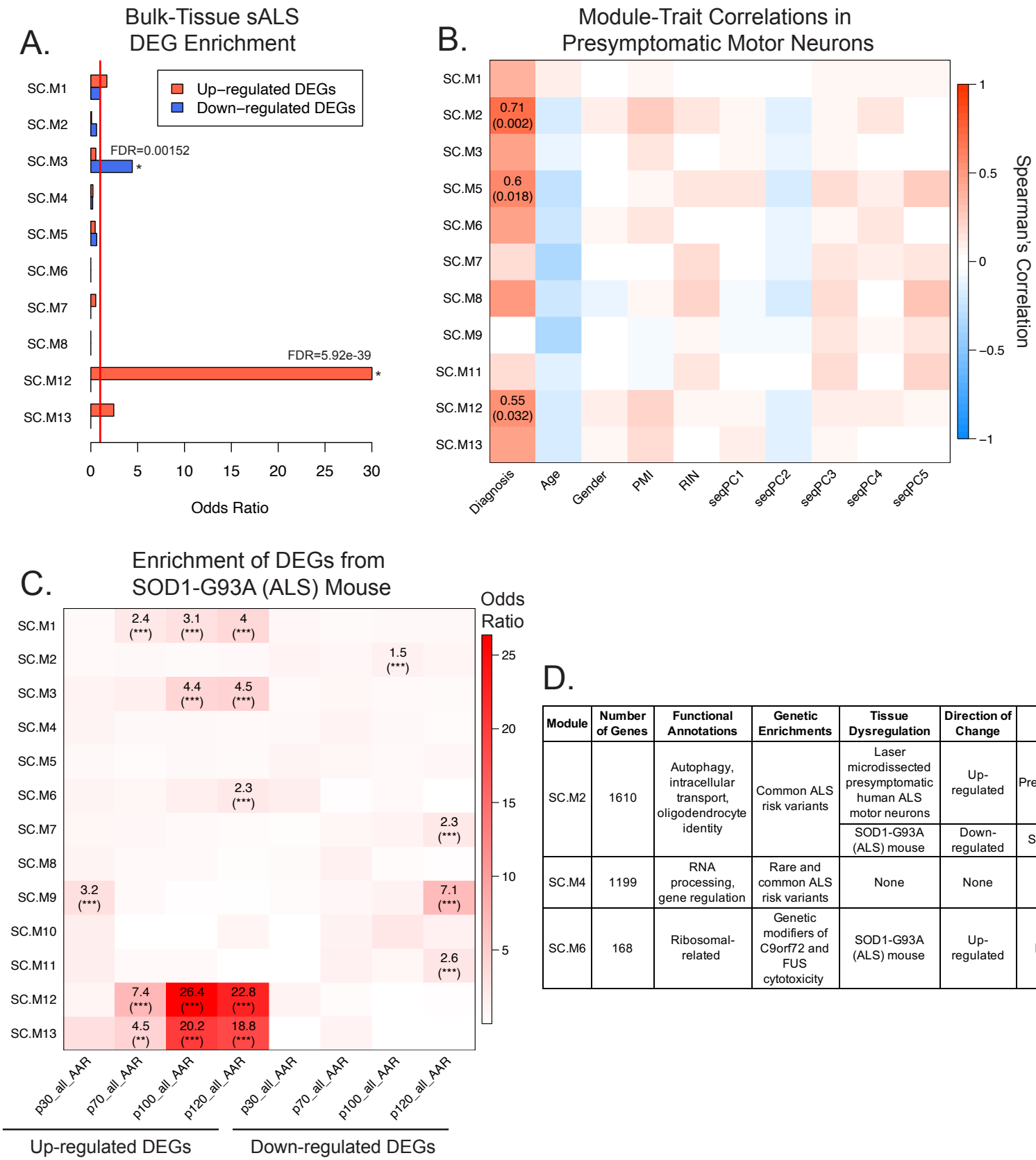

Genetic Modifier Enrichments
SC.M1
SC.M2
SC.M3
SC.M4
SC.M5
SC.M6
SC.M7
SC.M8
SC.M9
SC.M10
SC.M11
SC.M12
SC.M13
10.3
(2.93e−14)
16.9
(2.69e−05)
C9orf72 Suppressors
C9orf72 Enhancers
FUS Enhancers
FUS Suppressors
15
10
5
Odds Ratio

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

646 **Figure 4**. Modifiers of C9orf72 and FUS toxicity converge on top hubs of ribosomal-associated
647 module SC.M6.
648 (A) Enrichment of genetic modifiers of C9orf72 and FUS toxicity[21–23] within co-expression
649 modules. Odds ratios of enrichments with FDR corrected p-value < 0.05 are shown. FDR
650 corrected p-values themselves are shown within parentheses.
651 (B) Direct protein-protein interaction (PPI) network of the top 100 hub genes from module
652 SC.M6. Gene vertices are color coded as shown. P-value of the PPI network is derived from
653 1,000 permutations.
654



D.

Module	Number of Genes	Functional Annotations	Genetic Enrichments	Tissue Dysregulation	Direction of Change	Stage of Change
SC.M2	1610	Autophagy, intracellular transport, oligodendrocyte identity	Common ALS risk variants	Laser microdissected presymptomatic human ALS motor neurons	Up-regulated	Presymptomatic
				SOD1-G93A (ALS) mouse	Down-regulated	Symptomatic
SC.M4	1199	RNA processing, gene regulation	Rare and common ALS risk variants	None	None	None
SC.M6	168	Ribosomal-related	Genetic modifiers of C9orf72 and FUS cytotoxicity	SOD1-G93A (ALS) mouse	Up-regulated	End-stage

655 **Figure 5.** Validation of co-expression module dynamics in human ALS and model system
656 datasets.
657 (A) Enrichments of 112 up-regulated and 70 down-regulated DEGs from bulk-tissue RNA-
658 sequencing of postmortem ALS cervical spinal cord[16] within co-expression modules. Red line
659 marks a baseline odds ratio of 1. Asterisks mark enrichments with an FDR corrected p-value <
660 0.05.
661 (B) Spearman's correlations of module eigengenes with traits in RNA-sequencing of
662 presymptomatic motor neurons from laser capture microdissection (LCM) of postmortem ALS
663 tissues[17]. Significance was assessed using a Student's t-test. FDR corrected enrichment p-
664 values < 0.05 are shown within parentheses.
665 (C) Enrichment of DEGs within each disease stage of the SOD1-G93A mouse[18] within co-
666 expression modules. p30, p70, p100, and p120 represent postnatal age in days at time of
667 spatial transcriptomic sequencing. Odds ratios of enrichments with FDR corrected p-value <
668 0.05 are shown. * (FDR corrected p-value < 0.05). ** (FDR corrected p-value < 0.01). *** (FDR
669 corrected p-value < 0.001).
670 (D) Summary of the key findings from this study regarding modules relevant to ALS
671 pathophysiology: SC.M2, SC.M4, and SC.M6.