Antisense oligonucleotide depletion of *CCDC146* is a broad-spectrum therapeutic strategy for ALS

Sai Zhang[1,2,15,†], Tobias Moll[3,15], Jasper Rubin-Sigler[4,15], Sharon Tu[4], Shuya Li[5], Enming Yuan[6], Menghui Liu[1], Marriane King[3], Elham Alhathli[3], Connie Treanor[3], Alison Twelvetrees[3], Marcel Weinreich[3], Sam Bonsall[3], Calum Harvey[3], Sarah Gornall[3], Allan Shaw[3], Johanna Ganssauge[7], Akshay Bhinge[7], Cleide Dos Santos Souza[3], Laura Ferraiuolo[3], Eran Hornstein[8], Ophir Shalem[9], Anna Avila[3], Tatyana Shelkovnikova[3], Charlotte H. van Dijk[10], Ilia S. Timpanaro[10], Kevin P. Kenna[10], Jianyang Zeng[5], Philip van Damme[11], Philip S. Tsao[12,13], Pamela J. Shaw[3,14], Justin K. Ichida[4], Johnathan Cooper-Knock[3,14,†], Michael P. Snyder[2,†]

[1]Department of Epidemiology, University of Florida, Gainesville, FL, USA
[2]Department of Genetics, Center for Genomics and Personalized Medicine, Stanford University School of Medicine, Stanford, CA, USA
[3]Sheffield Institute for Translational Neuroscience (SITraN), University of Sheffield, Sheffield, UK
[4]Department of Stem Cell Biology and Regenerative Medicine, Eli and Edythe Broad Center for Regenerative Medicine and Stem Cell Research, University of Southern California, Los Angeles, CA, USA
[5]School of Engineering, Research Center for Industries of the Future, Westlake University, Hangzhou, Zhejiang, China
[6]Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, China
[7]Living Systems Institute, University of Exeter, Exeter, UK
[8]Departments of Molecular Genetics and Molecular Neuroscience, Weizmann Institute of Science, Rehovot, Israel
[9]Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[10]Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands
[11]Neurology Department, University of Leuven (KU Leuven), Belgium
[12]VA Palo Alto Healthcare System, Palo Alto, CA, USA
[13]Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA
[14]NIHR Sheffield Biomedical Research Centre, Sheffield, UK
[15]These authors contributed equally: Sai Zhang, Tobias Moll, and Jasper Rubin-Sigler
[†]Correspondence: sai.zhang@ufl.edu; j.cooper-knock@sheffield.ac.uk; mpsnyder@stanford.edu

Abstract

Amyotrophic lateral sclerosis (ALS) is a heritable and incurable disease defined by the degeneration of motor neurons (MNs), yet the genetics of ALS remain partially understood. Using a genomic deep learning-powered whole-genome analysis of 6,715 ALS patients, we identify four rare noncoding variants associated with patient survival, including chr7:76,009,472:C>T which is linked to a 70.6% reduction in survival. Genetic editing of this variant into iPSC-derived MNs increases *CCDC146* expression and exacerbates ALS-specific phenotypes including TDP-43 mislocalization. We reveal that CCDC146 was located within the basal body of primary cilia in human MNs, and that cilia structure and function is impaired by *CCDC146* overexpression but is restored by its depletion. Suppressing *CCDC146* using an antisense oligonucleotide (ASO) completely rescues ALS-specific survival defects in neurons derived from both sporadic and familial patients, and it extends survival and reverses TDP-43 pathology in an aggressive ALS mouse model. Taken together, CCDC146 is a new modifier of ALS survival that acts via the primary cilia of MNs. ASO targeting of *CCDC146* is a potential therapeutic approach for both sporadic and genetic forms of ALS, particularly because congenital absence of CCDC146 is well tolerated.

Introduction

Amyotrophic lateral sclerosis (ALS) is a devastating neurodegenerative disease[1] characterized by the progressive loss of motor neurons (MNs), ultimately leading to respiratory failure and death. ALS survival is markedly heterogeneous: the median survival is between 20 and 48 months, but 10-20% of ALS patients live longer than 10 years[2]. This heterogeneity persists even between patients who share the same monogenic risk factor[3]. A wide range of genetic variations and genes have been associated with ALS risk[4], but a lack of understanding of genetic factors influencing ALS progression is hindering the development of therapies to prolong patient survival.

Conventional genetic analyses, such as genome-wide association studies (GWAS) and rare variant association studies (RVAS) are likely to be underpowered because of limited functional characterization of genetic variants[5,6]. This limitation is pronounced for noncoding variants, which comprise the majority of human genetic material and complex disease heritability[7]. Understanding the biological function of noncoding variants requires specific cellular context because of their involvement in the transcriptional regulation that defines cell identity[8]. Moreover, *rare* variants pose additional challenges in traditional studies of variant function, such as the expression quantitative trait locus (eQTL) analysis[9,10], which rely on prevalent cohort-level evidence. Recent advances in sequence-to-function (seq2func) deep learning models[11–16], also known as genomic deep learning models, provide a powerful alternative method to predict at high-resolution the functional effects of noncoding variants and bypassing the requirement for individual-level data.

In this study, we developed a gDL-powered genetic analysis framework (**Fig. 1**) and performed thus-far the largest (n = 6,715) rare variant analysis for ALS survival, with a focus on the noncoding genetic variants with a functional effect in human MNs. We discovered four novel rare noncoding variants that are associated with an average reduction in ALS survival of more than 50%. We introduced the top scored variant chr7:76,009,472:C>T (where score indicates the probability of an effect on gene expression within MN) into MNs derived from an ALS patient. We confirmed an effect of chr7:76,009,472:C>T on *CCDC146* expression, and an exacerbation of phenotypes associated with ALS including neuronal survival, dendritic regression, TDP-43 localization, and disruption of primary cilia structure and function. Depletion of *CCDC146 in vitro* using an antisense oligonucleotide (ASO) rescued ALS-specific survival deficits in patient-derived neurons, irrespective of the genetic background or of baseline *CCDC146* expression, and without significant toxicity. ASO targeting of murine *Ccdc146* extended survival by over 50% in an aggressive ALS mouse model caused by TDP-43 mislocalization[17]. Indeed the ASO reversed TDP-43 mislocalization and prevented MN loss in the mouse. We localized CCDC146 within the basal body of primary cilia[18] of MNs and

87 discovered that the structure and function of primary cilia is inversely proportional to CCDC146
88 expression. Cilia dysfunction has been associated with ALS[19], but our work positions
89 *CCDC146* and MN cilia as a therapeutic target with relevance across the spectrum of ALS
90 patients.

91 **Results**

92 ***Overview of the study design***

93 We developed a functional genomics-informed genetic analysis pipeline (**Fig. 1**) to improve
94 discovery power and interpretability. Briefly, we first predicted the functional impacts of
95 noncoding variants on gene regulation in human MNs utilizing a gDL. Functional variants were
96 then mapped to their target genes based on the enhancer-promoter (E-P) interactions
97 constructed within MNs. Informed by these variant-to-gene maps, a gene burden-based
98 survival analysis was performed using 6,025 high-quality ALS genomes (out of 6,715
99 genomes), and pinpointing survival-associated rare variants and corresponding genes. Next,
100 we measured the effect of the top variant chr7:76,009,472:C>T, and changes in expression of
101 its target gene *CCDC146,* in iPSC-derived MNs to determine the underlying biology. Finally,
102 we designed ASOs targeting *CCDC146* to test whether CCDC146 depletion can rescue
103 disease-associated survival defects across different ALS genetic backgrounds, both *in vitro*
104 and i*n vivo*. This framework is generic and can be broadly applied to study rare variant genetics
105 of other complex diseases.

106 ***Constructing the regulome and predicting noncoding variant effects in human motor***
107 ***neurons***

108 To characterize the transcriptional regulation in human motor neurons, we mapped MN E-P
109 interactions by applying the activity-by-contact (ABC) model[20] (**Supplementary Fig. 1A**) to
110 the transcriptomic (RNA-seq) and epigenomic (ATAC-seq, H3K27ac ChIP-seq, and Hi-C)
111 profiling data of iPSC-derived MNs (iMN) derived from neurologically normal donors[21]
112 (**Supplementary Fig. 1B** and **1C**, **Supplementary Table 1**, **Supplementary Results**;
113 **Methods**). A large proportion of MN E-P links (89.0%) were cell-type-specific
114 (**Supplementary Fig. 1D**), and were relatively enriched with long-range interactions
115 (**Supplementary Fig. 1E**, **Supplementary Results**). As expected, transcription factors (TFs)
116 and genes involved in MN E-P links were enriched with biological pathways known to be critical
117 for MN function (**Supplementary Fig. 1F-1H** and **2A**, **Supplementary Results**).
118 Next, we sought to examine the function of noncoding variants specifically within MNs.
119 The activity of epigenomic features, such as chromatin accessibility and histone modifications,
120 varies widely across cell types and contexts, yet much of this regulation is encoded in the

121 underlying DNA sequence[11,12,14]. To learn this epigenome grammar in MNs, we developed a
122 genomic deep learning model[12,22], named NeuroNet, to predict chromatin accessibility and key
123 histone marks (H3K27ac, H3K4me1, and H3K4me3) in iMNs directly from the reference DNA
124 sequence (**Fig. 2A**; **Methods**). After training (**Methods**), NeuroNet accurately predicted these
125 epigenomic features, achieving a mean area under receiver operating characteristic curve
126 (AUROC) of 0.851 (SD = 0.10; **Fig. 2B**) and a mean area under precision-recall curve
127 (AUPRC) of 0.853 (SD = 0.063; **Fig. 2C**).

128 We then used NeuroNet to characterize the functional effects of noncoding variants in
129 MNs by querying the difference in predicted epigenomic features between reference and
130 alternative alleles (**Fig. 2A**; **Methods**). Differing from traditional approaches in modeling
131 variant effects[23], this *in-silico* strategy requires no individual-level data, which greatly increases
132 its flexibility and generalizability. NeuroNet outputs a score that defines the probability of a
133 specific variant altering a given epigenomic feature. To benchmark our variant effect
134 prediction, we compared NeuroNet predictions with those of DeepSEA[12], another deep
135 learning model which has been used to predict 2,002 epigenomic features in various tissues
136 and cell types. As expected, NeuroNet and DeepSEA predictions aligned well in neuron- and
137 brain-related tissues and cell types compared to other cell-types (H3K4me1: $P = 2.4 \times 10^{-5}$;
138 H3K4me3: $P = 4.5 \times 10^{-3}$, two-sided Fisher's exact test; **Fig. 2D** and **Supplementary Fig. 2B**).
139 The alignment was marginally significant for H3K27ac ($P = 0.059$, two-sided Fisher's exact
140 test; **Supplementary Fig. 2C**), but not significant for chromatin accessibility ($P > 0.05$, two-
141 sided Fisher's exact test; **Supplementary Fig. 2D**), probability due to the limited number ($n =$
142 3) of neurons and brains profiled by the DNase I hypersensitive sites sequencing (DNase-
143 seq). In ongoing analyses, to integrate information from all epigenomic features, we used the
144 maximum of all four variant-specific scores to derive the final NeuroNet variant score.

145 NeuroNet is designed to capture functional consequences of noncoding variants
146 including, but not limited to, TF binding disruption. To aid the biological interpretation of
147 NeuroNet predictions, we examined the overlap between NeuroNet-prioritized variants and
148 variants previously associated with changes in TF binding[24]. We then used this association to
149 determine which NeuroNet score should be used as a 'cutoff' to assign 'functional' noncoding
150 variants. Specifically, we split NeuroNet prediction scores into percentiles and examined which
151 cutoff produced the strongest enrichment with variants that disrupt TF binding[24] (**Methods**).
152 Variants with higher NeuroNet scores were more enriched with TF-binding-disrupting variants
153 (**Fig. 2E**). This demonstrates that NeuroNet has captured known TF biology, despite the fact
154 that TF motifs were *not* included in model training. The strongest enrichment with disruption
155 to known TF binding motifs was achieved at the 95% cutoff (NeuroNet score ≈ 0.06; $P = 5.5 \times$

156 10^{-4}, *t*-statistics = 3.46, two-sided *t*-test; **Fig. 2E**). Accordingly, we classified variants with a
157 NeuroNet score > 0.06 as 'functional' within MN in the remainder of our analyses.
158 To further examine the functional implication of NeuroNet prediction, we assessed the
159 distribution of prioritized variants across the genome. Notably, we found that MN cis-regularoty
160 elements (CREs), including promoters and enhancers, were enriched with variants prioritized
161 by NeuroNet (odds ratio [OR] > 1.0, two-sided Fisher's exact test; **Fig. 2F**; **Methods**). As in
162 our TF binding site analysis, this enrichment was positively correlated with the NeuroNet score
163 in all features, and peak enrichment was obtained at the 95% cutoff threshold (**Fig. 2F**).
164 Finally, we benchmarked our variant effect prediction against multiple widely-used
165 methods, including GERP++[25], GWAVA[26], CADD[27], LINSIGHT[28], and DeepSEA[12]. By
166 following the suggested cutoffs to define functional or deleterious variants for different
167 approaches, we observed that NeuroNet-defined functional variants were enriched with
168 functional or deleterious variants prioritized by other methods ($P < 0.05$, OR > 1.0, two-sided
169 Fisher's exact test; **Fig. 2G**; **Methods**). The strongest enrichment was obtained for DeepSEA,
170 probably because the model design is most similar to NeuroNet.

171 ***NeuroNet-powered rare variant analysis for ALS survival***

172 We sought to integrate NeuroNet prediction with ALS whole-genome sequencing (WGS) data
173 to identify rare noncoding variants that impact ALS risk or survival via disrupting transcriptional
174 regulation in MNs.
175 We performed variant-level and sample-level quality controls (QCs) for 9,600 whole
176 genomes generated in the Project MinE[29] (Data Freeze 2), resulting in an analysis-ready
177 cohort encompassing 6,025 ALS patients and 2,264 controls (**Methods**). Next, using
178 NeuroNet scores, we identified functional noncoding rare variants (allele frequency [AF] <
179 0.5% within gnomAD v2.1 non-Finnish European [NFE] samples, allele count [AC] ≥ 3 in the
180 analysis cohort, and NeuroNet score > 0.06) located within MN CREs. We mapped these
181 variants to their target genes based on the ABC-predicted E-P links. A gene-level rare variant
182 burden profile (i.e., the number of variants mapped to each gene) was then constructed for
183 each individual in the QC'ed cohort, ready for downstream genetic association analysis.
184 We first studied the genetics of ALS risk. We conducted a rare variant burden risk
185 analysis (ALS patients versus healthy controls) using SKAT-O[30,31] to quantify the burden of
186 NeuroNet-prioritised variants (**Methods**), with covariates including age, sex, age^2, sex x age,
187 sequencing platform, and the first 20 principal components to adjust for known confounders.

188 No gene was significantly associated with ALS risk after multiple testing correction, suggesting
189 an underpowered analysis (**Supplementary Table 2**). However, a number of genes that have
190 been linked to ALS risk achieved nominal significance, such as *KIF5A*[32,33] (P = 0.016) and
191 *NEAT1*[34] (P = 0.033).

192 Next, we performed a rare variant burden survival analysis using the Cox proportional-
193 hazards (PH) model to measure the effect of NeuroNet-prioritised variants (**Methods**). We
194 included 26 covariates: age, sex, age^2, sex x age, sequencing platform, *C9orf72* status, and
195 the first 20 principal components, to adjust for known confounders. Noncoding rare variant
196 burden was significantly associated with ALS survival for seven genes after Bonferroni multiple
197 testing correction (adjusted $P < 0.05$; **Fig. 3A** and **3B**, **Supplementary Table 3**), for which no
198 inflation was observed (λ = 0.99). The mean survival time for the carriers of these rare variants
199 ranges from 0.99 to 2.46 years, which is 2.79 to 1.32 years shorter than that of the overall
200 cohort (**Fig. 3C**).

201 Four variants were linked to three gene sets: chr7:76,009,472:C>T resides within an
202 enhancer linked to *CCDC146*, *FAM185BP*, *FGL2*, and *UPK3B* (**Fig. 3D**);
203 chr14:102,782,854:G>A and chr14:102,786,108:G>T are located within an enhancer of
204 *ZNF839* (**Fig. 3E**); chr2:51,259,489:G>C is within the promoter of *NRXN1* (**Supplementary**
205 **Fig. 3A**). As a benchmark, we performed a standard rare variant burden survival analysis
206 without variant prioritization using NeuroNet (**Methods**), but no gene passed multiple testing
207 correction (**Supplementary Fig. 3B** and **Supplementary Table 4**). These results
208 demonstrate the superiority of our deep learning-based framework in enhancing discovery
209 power and biological interpretation.

210 Lastly, we were interested in the overlap of rare variant associations between ALS
211 survival and risk. We examined this by testing the enrichment of survival-associated genes
212 within risk-associated genes under different P-value cutoffs. Notably, we observed significant
213 gene overlaps across a spectrum of P-value cutoffs ($P < 0.05$, two-sided Fisher's exact test;
214 **Supplementary Fig. 3C**), suggesting an overlapping genetic architecture underlying these
215 two ALS phenotypes within noncoding genomic regions.

216 *Reproducing chr7:76,009,472:C>T within iMNs alters CCDC146 expression and*
217 *exacerbates ALS-associated phenotypes*

218 The noncoding variant chr7:76,009,472:C>T achieved the highest NeuroNet score amongst
219 noncoding variants associated with ALS survival (**Fig. 3B**). This variant was linked to four
220 candidate target genes based on predicted E-P links (**Fig. 3D**). To experimentally verify our
221 prediction and determine the regulatory target of chr7:76,009,472:C>T, we utilized CRISPR-
222 Cas9 to reproduce this variant in iPSCs derived from an ALS patient caused by a G4C2-repeat
223 expansion within *C9orf72* (**Fig 4A**; **Methods**), where the unedited cells served as an isogenic

224 control. Of note, we observed extremely rapid disease progression in a patient who carried
225 both a *C9orf72* expansion and chr7:76,009,472:C>T in the Project MinE cohort
226 (**Supplementary Table 5**).

227 We first examined whether chr7:76,009,472:C>T altered the expression of candidate
228 target genes implicated by predicted E-P links. To achieve this, we differentiated edited iPSCs
229 carrying chr7:76,009,472:C>T, as well as isogenic control iPSCs, into mature MNs using our
230 published small molecule protocol[21] (**Supplementary Fig. 4A**; **Methods**). The ABC model
231 predicted an E-P link for four genes: *CCDC146, FAM185BP, FGL2,* and *UPK3B*. Study of
232 human tissue revealed that UPK3B is almost exclusively expressed in urothelium[35], and
233 therefore we excluded this gene from further study. We extracted RNA from mature iMNs and
234 measured the expression levels for *CCDC146, FAM185BP,* and *FGL2* (**Methods**). Expression
235 of *CCDC146* was significantly elevated in the presence of chr7:76,009,472:C>T compared to
236 isogenic control iMNs ($P = 5.6 \times 10^{-3}$, fold change [FC] = 1.35, two-sided *t*-test; **Fig. 4B**), but
237 *FAM185BP,* and *FGL2* were not differentially expressed.

238 Next, we sought to determine the effect of chr7:76,009,472:C>T on ALS-associated
239 phenotypes within iMNs. Our genetic analysis suggested that chr7:76,009,472:C>T increased
240 the rate of ALS progression, resulting in a more severe phenotype; therefore we hypothesized
241 that chr7:76,009,472:C>T would increase the severity of ALS-associated phenotypes in iMNs
242 compared to isogenic control iMNs. For an additional comparison, we also included iMNs from
243 an age- and sex-matched neurologically normal individual.

244 ALS is associated with a number of MN molecular phenotypes, including increased
245 cell death[36], reduced neurite outgrowth[37,38], and TDP-43 mislocalization[39]. We took advantage
246 of this battery of phenotypes to assess the effect of chr7:76,009,472:C>T on ALS severity *in*
247 *vitro*. iMNs were placed in culture and monitored continuously for 12 days (**Methods**). We
248 measured cell survival and neurite growth at 8 hour intervals. Notably, compared to isogenic
249 control iMNs, the presence of chr7:76,009,472:C>T was associated with a reduction in both
250 neurite length ($P < 0.05$, two-sided *t*-test; **Fig. 4C**) and the number of neurite branch points
251 (**Supplementary Fig. 4B**), and with increased neuron death (**Supplementary Fig. 4C**). iMNs
252 carrying chr7:76,009,472:C>T had a significantly higher burden of TDP-43 mislocalization at
253 the end stage ($P = 0.024$, two-sided *t*-test; **Fig. 4D** and **4E**; **Methods**), compared to isogenic
254 control iMNs.

255 Previous literature suggests that CCDC146 is located within the basal body of primary
256 cilia[18]. We confirmed that this is also the case within iMNs (**Supplementary Fig. 4D**;
257 **Methods**). Moreover, in iMNs that carried chr7:76,009,472:C>T with elevated *CCDC146*
258 expression, the length of primary cilia was reduced compared to the isogenic control iMNs (P
259 $= 0.04$, paired *t*-test; **Fig. 4F** and **4G**; **Methods**). Primary cilia function is related to CRABP1
260 expression and hence the maintenance of neuromuscular junctions (NMJs) between MNs and

261 their target muscles[40]. NMJ loss is a key phenotype of ALS[41] and reduced CRABP1 has been
262 linked to ALS[19]. Immunoblotting for CRABP1 revealed that in iMNs carrying
263 chr7:76,009,472:C>T, CRABP1 expression was reduced compared to the isogenic control
264 iMNs (P = 0.03, paired t-test; **Fig 4H** and **4I**; **Methods**). Interestingly, ALS-associated Q331K-
265 TARDBP mutations, which disrupt function of TDP-43, have been associated with large
266 downregulation of CRABP1 mRNA[42] suggesting that the link between ciliary function and
267 CRABP1 expression may be mediated via TDP-43 dysfunction.

268 We have shown that CCDC146 is located within the basal body of primary cilia and
269 that *CCDC146* levels have an effect on cilia structure and function. CCDC146 is a microtubule-
270 associated protein and cilia are formed from microtubules. Based on this, we further examined
271 microtubule dynamics. EB3 is a well-established regulator of primary cilium biogenesis,
272 facilitating the anchoring of microtubule minus-ends at the basal body and promoting vesicular
273 trafficking to the ciliary base[43,44]; EB3 depletion has been shown to reduce primary cilium
274 length[43]. Interestingly, we identified a significant reduction in EB3 velocity in
275 chr7:76,009,472:C>T-iMNs with increased *CCDC146* expression (FC = 0.8, P = 0.039,
276 unpaired t-test; **Fig. 4J** and **4K**; **Methods**), suggesting that the disrupted microtubule
277 dynamics is the link between *CCDC146* and cilia structure and function.

278 ***ASO suppression of CCDC146 corrects cilia dysfunction and rescues ALS-associated***
279 ***survival deficits in patient-derived neurons***

280 We have linked chr7:76,009,472:C>T to elevated expression of *CCDC146* and to ALS-
281 associated molecular phenotypes in iMNs, including TDP-43 mislocalization and cilia
282 dysfunction. To further test whether the expression change of *CCDC146* is the key driver of
283 this toxicity, we synthesized an antisense oligonucleotide (ASO) targeting *CCDC146* mRNA
284 for degradation.

285 We have established a survival readout for iPSC-derived neurons (iNs) whereby
286 neurons derived from patients display an increased sensitivity to glutamate-induced
287 excitotoxicity[36,45] (**Fig. 5A**; **Methods**). This reflects long standing observations regarding
288 glutamate toxicity in ALS[46–48]. We have also established this readout in a 3D spheroid format
289 (3D-iN) to enable prolonged culturing without primary glial cells[45,48] (**Fig. 5A**; **Methods**). In our
290 experiment, we treated 3D-iNs derived from multiple patients, including four healthy controls,
291 four ALS patients carrying an *C9orf72* G4C2-repeat expansion, two ALS patients carrying
292 ALS-associated mutations within TARDBP, and five sporadic ALS (sALS) patients
293 (**Supplementary Table 6**). 3D-iNs were exposed to either *CCDC146* ASO or a scrambled
294 control ASO, followed by longitudinal tracking (**Fig. 5A**; **Methods**). First, we confirmed the
295 reduction in *CCDC146* mRNA level within 3D-iNs treated with the *CCDC146* ASO compared
296 to the scrambled ASO (36.9% reduction, P = 0.031, two-sided t-test; **Fig. 5B**; **Methods**).

297 Notably, we observed that the survival defect was completely rescued by *CCDC146*
298 suppression compared to the scrambled ASO ($P < 0.05$, two-sided *t*-test; **Fig. 5C-E** and
299 **Supplementary Fig. 5A**). The result persisted across all of the different genetic backgrounds
300 tested in our experiment (11 patient lines in total; **Fig. 5C-D** and **Supplementary Fig. 5A**).
301 Moreover, there was no evidence of toxicity mediated by the *CCDC146* ASO in control 3D-iNs
302 ($P > 0.05$, two-sided *t*-test; **Fig. 5D** and **Supplementary Fig. 5A** and **5B**). Indeed, loss-of-
303 function (LoF) of CCDC146 has been shown to be well tolerated in humans[18]. Of note, none
304 of the experimentally-tested patient-derived 3D-iN lines exhibited elevated *CCDC146*
305 expression at baseline (**Supplementary Fig. 5C**). Together, these results demonstrate that
306 *CCDC146* suppression may be a broadly applicable therapeutic strategy for different forms of
307 ALS.

308 Next, we sought to determine whether ASO suppression of *CCDC146* reverses deficits
309 in primary cilia structure and function (**Fig. 4G** and **4H**) which were associated with
310 chr7:76,009,472:C>T and increased expression of *CCDC146*. We treated iMNs derived from
311 a patient carrying a *C9orf72* G4C2-repeat expansion with *CCDC146* ASO and measured
312 primary cilia length and CRABP1 expression after 72 hours (**Methods**). Cilia length was
313 significantly increased after *CCDC146* ASO treatment compared to untreated iMNs and to
314 treatment with a scrambled ASO ($P = 0.03$, one-way ANOVA; **Fig. 5F** and **Supplementary**
315 **Fig. 5D**). Similarly, CRABP1 expression was increased after treatment with *CCDC146* ASO
316 compared to scrambled ASO ($P = 0.01$, paired *t*-test; **Fig. 5G** and **5H**).

317 We conclude that *CCDC146* suppression rescues ALS-associated survival defects *in*
318 *vitro* and reverses deficits in cilia structure and function, suggesting that ciliary function is the
319 mechanism by which *CCDC146* modifies neuronal survival.

320 ***ASO suppression of murine Ccdc146 rescues motor phenotypes and survival defects***
321 ***associated with TDP-43 mislocalization***

322 Depletion of human *CCDC146* corrected survival defects in ALS patient-derived neurons
323 irrespective of genetic background and including sporadic ALS. The majority of ALS is
324 associated with TDP-43 mislocalization[41], and therefore we hypothesised that *CCDC146*
325 depletion may be effective in an *in vivo* model of ALS caused by TDP-43 mislocalization.
326 Moreover, in iMNs the elevated *CCDC146* expression was associated with an increase in
327 TDP-43 mislocalization, suggesting that reducing *CCDC146* expression may reverse this
328 aberration. To test these hypotheses, we utilized an aggressive juvenile mouse model
329 (TAR4/4 mouse) of ALS caused by overexpression of human TDP-43 which manifests with
330 mislocalization of human and murine tdp-43[17] (**Fig. 6A**). This model has been previously used
331 to develop therapeutic strategies for ALS[49].

332 We generated an ASO against murine *Ccdc146* which successfully depleted *Ccdc146*
333 expression when injected intracerebroventricular (IVC) at D1 (P = 6.14 x 10^{-3}, two-sided *t*-test;
334 **Fig. 6B**; **Methods**). Notably, the *Ccdc146* ASO mitigated motor phenotypes compared to a
335 scrambled ASO and the difference increased over time (**Fig. 6C**); at D22 the differences were
336 significant for all measures (n > 30 in each group; gait impairment score: P = 2.7 x 10^{-4}; tremor
337 score: P = 2.2 x 10^{-4}; kyphosis score: P = 2.1 x 10^{-4}, two-sided *t*-test; **Fig. 6C**; **Methods**).
338 Moreover, survival was significantly prolonged in mice treated with the *Ccdc146* ASO (n = 55,
339 36% increase in median survival time, P = 5.5 x 10^{-6}, Cox regression; **Fig. 6D**; **Methods**),
340 compared to mice treated with the scrambled ASO. To determine whether the *Ccdc146* ASO
341 acts via reversal of MN toxicity, we performed MN counts in wild type and TAR4/4 mice treated
342 with *Ccdc146* or scrambled ASO. Treatment with *Ccdc146* ASO significantly increased the
343 average number of MNs at D22 (P = 1.6 x 10^{-4}, two-sided *t*-test; n > 10 biological replicates, n
344 = 6 fields-of-view (FOV) per mouse; **Fig. 6E** and **6F**). Indeed at D22, MN counts in *Ccdc146*
345 ASO-treated TAR4/4 mice were equivalent to wild type animals (**Fig. 6E**). Finally, we sought
346 to examine the distribution of murine TDP-43 within MNs using immunohistochemistry
347 (**Methods**). The *Ccdc146* ASO significantly increased the nuclear:cytoplasmic ratio of murine
348 TDP-43 within MNs at D22 compared to the scrambled ASO (P = 1.76 x 10^{-4}, two-sided *t*-test;
349 **Fig. 6G** and **6H**), exhibiting an almost complete rescue of TDP-43 mislocalization when
350 compared to the wild type mice (P > 0.05, two-sided *t*-test; **Fig. 6G** and **6H**).
351 We conclude that ASO suppression of *CCDC146* is an effective therapeutic strategy
352 *in vivo,* and that this activity is upstream of TDP-43 mislocalization which is the defining
353 pathology of ALS[39].

354 **Discussion**

355 We have designed ASOs targeting *CCDC146* which ameliorate ALS-associated phenotypes,
356 including survival and TDP-43 mislocalization both *in vitro* and *in vivo*. We achieved this by
357 starting with a hypothesis-free *in-silico* screen for genetic modifiers of ALS survival. Distinct
358 from previous works, our genetic screen was built upon a deep learning-based variant
359 interpretation framework that makes MN-specific functional effect predictions for an
360 exhaustive set of noncoding variants. Our model, named NeuroNet, was trained to predict
361 various MN epigenomic features (e.g., chromatin accessibility and histone modification) from
362 DNA sequence. This enables the model to make functional predictions for unseen genetic
363 variants irrespective of allele frequency. Deep learning-based seq2func modeling has been
364 widely adopted to examine the cell-type-specificity of regulatory genome function and genome
365 variation[11–16,22,50], and here we demonstrated its success in identifying potential therapeutic
366 targets for ALS. Conventional approaches for investigating variant effects, such as the
367 expression and chromatin accessibility quantitative trait loci (eQTL and caQTL) analyses,

368 require matched individual genotype and transcriptomic/epigenomic profiling data, which is
369 typically underpowered for the study of rare or ultra-rare variants.
370 Combining NeuroNet predictions of variant effects on the epigenome, together with
371 enhancer-promoter links, enabled us to build a variant-to-gene (V2G) map within MNs. The
372 result is a gene-resolved prediction of noncoding variant function. Based on this V2G, we have
373 conducted large-scale rare variant burden tests for both ALS risk and survival. Although the
374 risk analysis was underpowered due to the paucity of control samples, a NeuroNet-enabled
375 survival analysis succeeded whereas a traditional analysis was still underpowered. This
376 demonstrates the benefit of incorporating functional information into a genetic analysis.
377 We have identified four noncoding variants linked to ALS survival, and they were
378 predicted to modulate the expression of three protein-coding genes: *ZNF839*, *NRXN1*, and
379 *CCDC146*. Interestingly, although we did not identify any genome-wide significant genes
380 associated with ALS risk, we did note a significant overlap between nominally significant risk-
381 and survival-associated genes. This is consistent with known genetic variants which affect
382 both ALS risk and survival including noncoding variants associated with *UNC13A*[51] and
383 *C9orf72*[3], and provides evidence for the overlapping genetic architecture of ALS risk and
384 survival within the noncoding genome.
385 We reproduced the top hit chr7:76,009,472:C>T in iMNs carrying an ALS-associated
386 G4C2 expansion within *C9orf72*, and verfied its regulation of *CCDC146* expression. We
387 showed that chr7:76,009,472:C>T exacerbated ALS-associated phenotypes in iMNs,
388 including reduced MN survival compared to isogenic control neurons. Moreover, using an ASO
389 to reduce *CCDC146* expression completely rescued ALS-associated survival defects in iMNs,
390 irrespective of genetic background and without significant toxicity, and also irrespective of the
391 baseline *CCDC146* expression. Similarly, we observed a dramatic survival benefit in an
392 aggressive mouse model of ALS caused by mislocalization of nuclear TDP-43[17], which is the
393 key molecular pathology underlying >98% of ALS[39]. We note that a different ASO has recently
394 been approved for treatment of ALS patients carrying a pathogenic mutation within SOD1[52].
395 Unlike that treatment - Tofersen - our data suggest an ASO targeting *CCDC146* could be
396 administered to broad forms of ALS and not limited to a specific genetic subtype.
397 CCDC146 is a microtubule associated protein (MAP) and localized to sperm flagella
398 and to primary cilia in particular. Biallelic variants in *CCDC146* have been associated with
399 reduced sperm viability and non-syndromic male infertility in both mice and humans[18]. Flagella
400 are functionally mirrored by cilia and it is notable that primary cilia dyskinesia has been
401 associated with ALS[19,53,54]. Here, we have determined that CCDC146 is localized to primary
402 cilia in human MNs, in a similar manner to *C21ORF2/CFAP410*[19] which is an ALS risk gene[55]
403 and is important for cilia assembly. Our data suggest that this may be because CCDC146 is
404 a modifier of microtubule dynamics, and cilia are microtubule structures. Primary cilia occur

405 one per cell in neurons and extend into the extracellular space where they can transduce
406 intercellular signals via G-protein coupled receptors. Unlike in sperm cells, neuronal cilia are
407 non-motile structures with an important role in sensing extracellular signals and in transduction
408 of *sonic hedgehog* (shh) signalling. Elevated *CCDC146* causes shortened primary cilia which
409 would be expected to impair intraflagellar transport (IFT)[19]. IFT is the process by which
410 structural components are selected and positioned for the assembly and maintenance of cilia
411 structure; indeed IFT occurs continuously in two directions towards and away from the tip of
412 each cilium and occurs in proportion to cilium length[56].

413 We identified loss of CRABP1 in MNs overexpressing *CCDC146* and increase in
414 CRABP1 after *CCDC146* depletion by ASO. CRABP1 is an important protein for NMJ structure
415 and function within MNs. CRABP1 expression is relatively specific to the spinal cord and to
416 MN in particular[40]; indeed CRABP1 knockout mice exhibit an ALS-like phenotype, including
417 MN degeneration[40]. *CRABP1* expression has been linked to ciliary function via shh
418 signalling[57], and we also note reports in the literature of a link between TDP-43 dysfunction
419 and *CRABP1* expression[42]. Indeed, *CRABP1* is implicated in maintenance of NMJ function via
420 *CAMK2B* and *AGRN*[40], both of which are misplaced in the context of TDP-43 dysfunction[58];
421 this convergence supports the idea that this is a key pathway in driving neuronal toxicity in
422 ALS. We suggest that the positive effect of *CCDC146* on MN health is mediated via
423 modification of ciliary function leading to recovery of TDP-43 mislocalisation and elevation of
424 CRABP1 expression. TDP-43 mislocation is a key driver of >98% of ALS[39,58,59] and so our
425 proposed mechanism would explain why the *CCDC146* ASO appears to be broadly applicable
426 to patient-derived neurons. Here we have provided evidence to link *CCDC146* expression to
427 TDP-43 mislocalization both *in vitro* and *in vivo*. While further research is needed to confirm
428 and clarify the interaction between ciliary function and TDP-43 mislocalization, our work opens
429 a new avenue for understanding the role of primary cilia as a universal driver of MN death in
430 ALS.

431 It is notable that, aside from a nonsyndromic fertility defect, loss of CCDC146 is well
432 tolerated[18]; this crucial observation suggests that our ASO approach may be viable without
433 significant toxicity. Additionally, this gene is not well conserved[60] (probability of loss of function
434 intolerance (pLI) = 0). CCDC146 is expressed throughout the brain and spinal cord, and
435 particularly within neurons[61] (Human Protein Atlas). Expression of *CCDC146* has been
436 identified specifically in neurons from the human spinal cord[62]. Matching our finding of higher
437 *CCDC146* expression associated with reduced ALS survival, the expression of *CCDC146* is
438 higher in ALS patient spinal cord compared to control cervical spinal cord although, at a bulk
439 level, this difference was not statistically significant[63].

440 We investigated the potential of targeting CCDC146 through small-molecule drugs.
441 We first searched chemical and patent databases but found no active compounds targeting

442 CCDC146 and similar proteins (**Methods**). CCDC146 is also classified as a poorly explored
443 target without active small molecules according to the Pharos database[64,65]. Due to the lack
444 of experimentally solved protein structure data, we obtained the AlphaFold-predicted structure
445 of CCDC146[66] (**Supplementary Fig. 6**). The predicted structure mainly consists of coiled coils
446 linked by flexible regions, without ligand binding sites detected on its surface[67] (**Methods**),
447 suggesting that it would be challenging to design small-molecule drugs for this target and
448 underscoring the importance of alternative targeting approaches such as an ASO.

449 We also sought to validate another gene *ZNF839* in the context of disease progression.
450 In particular, we inspected gene expression in iMNs derived from an independent cohort
451 encompassing 245 ALS patients (www.answerals.org; **Methods**). We measured disease
452 progression using two metrics, including the survival time from symptom onset and the rate of
453 change in ALS Functional Rating Scale in its revised version (ALSFRS-R)[68]. ALSFRS-R is
454 widely used to measure ALS progression[69]. We found that lower expression of *ZNF839* was
455 associated with faster rate of disease progression (β = 0.85, P = 5.9 x 10^{-3}, multivariable linear
456 regression; **Supplementary Fig. 7A**; **Methods**) and shorter survival time (β = -0.79, P = 0.044,
457 Cox PH regression; **Supplementary Fig 7B**; **Methods**). The expression levels of other genes
458 we identified were not correlated with either the rate of change in the ALSFRS-R or survival,
459 probably because the AnswerALS-based analysis was underpowered.

460 Given the significant enrichment of European samples in our WGS cohort, a potential
461 limitation of our work is that our conclusions might not be generalizable to other ancestral
462 groups. However, the allele frequencies of all four survival variants are below 1% (the
463 maximum population-level allele frequency is 5.9% obtained for chr14:102,782,854:G>A)
464 across different populations according to the gnomAD whole-genome collection 2.1.1,
465 implicating the ancestral transferability of the effect of our identified variants, although further
466 confirmation is required.

467 In summary, we have developed a deep learning-powered genetic analysis framework
468 for systematic identification of cell-type-resolved genetic modifiers of ALS survival. Each of the
469 candidate genes we pinpointed represents a potential therapeutic target, but we have
470 demonstrated *in vitro* and *in vivo* that ASO knockdown of *CCDC146* is a safe, broadly
471 applicable and efficacious therapeutic strategy. NeuroNet can be applied to study the genetics
472 of other MN-specific diseases, including spinal muscular atrophy (SMA) and hereditary
473 sensory and motor neuropathy (HSMN). Our analysis framework contributes to a better
474 understanding of ALS genetic architecture, defines a novel therapeutic target for ALS, and
475 provides a generic paradigm for enhancing noncoding genetic analysis of complex human
476 diseases.

477 **Acknowledgments**

478 This work was supported by NIH (CEGS 5P50HG007735 to M.P.S., R01NS131409 and
479 R01NS097850 to J.K.I.), Wellcome Trust (216596/Z/19/Z to J.C.-K.), NIHR (NF-SI-0617-
480 10077 to P.J.S.), Motor Neurone Disease (MND) Association (894-791, 899-792 and 913-792
481 to C.H., S.G., S.Z. and J.C.-K.), John Douglas French Alzheimer's Foundation (to J.K.I.), and
482 Robert Packard Center for ALS Research (1375407 to S.Z.). We also acknowledge support
483 from the MND Association Fellowship (2323-799) and the Kingsland Fellowship (to T.M.), the
484 NIHR Sheffield Biomedical Research Centre (NIHR203321), and the NIHR Sheffield Clinical
485 Research Facility. We are very grateful to the ALS patients and control subjects who
486 generously donated biosamples. We acknowledge transcriptomic and clinical data provided
487 by the AnswerALS Consortium. Figures 1, 2a, 4A, 5A and 6A were created using
488 BioRender.com.

489 **Author contributions**

490 S.Z., J.C.-K. and M.P.S. conceived and designed the study. S.Z. and E.Y. developed
491 NeuroNet. S.Z., J.C.-K., T.M. and J. R.-S. performed statistical analyses. T.M. performed all
492 the experiments for reproducing and characterizing chr7:76,009,472:C>T in iMNs. J. R.-S.,
493 S.T. and T.M. carried out ASO-related experiments in 3D-iNs, iMNs and mice. S.Z., J.C.-K.,
494 T.M. and J. R.-S. interpreted the data with assistance from all other authors. The Project MinE
495 ALS Sequencing Consortium was involved in WGS data acquisition and analysis. S.Z., J.C.-
496 K. and M.P.S. supervised the work. S.Z., J.C.-K. and M.P.S. wrote the manuscript with
497 feedback from all other authors.

498 **Competing interests**

499 M.P.S. is a co-founder and the scientific advisory board member of Personalis, Qbio, January,
500 SensOmics, Filtricine, Akna, Protos, Mirvie, NiMo, Onza, Oralome, Marble Therapeutics and
501 Iollo. He is also on the scientific advisory board of Danaher, Genapsys, and Jupiter. J.K.I. is a
502 co-founder and a scientific advisory board member of AcuraStem, Inc. and Modulo Bio, a
503 scientific advisory board member of Synapticure and Vesalius Therapeutics. J.K.I. is also an
504 employee of BioMarin Pharmaceutical. The remaining authors declare no competing interests.
505 We have a patent application related to this work.

506 **Main figure legends**

507 **Fig. 1. Schematic of the study design.** Leveraging epigenomic and transcriptomic profiling
508 of iPSC-derived human motor neurons (iMNs), we first identified enhancer-promoter (E-P)
509 interactions using the Activity-by-Contact (ABC)[20] model (**Panel 1**, top). Next, we developed
510 a genomic deep learning model[12,14] named NeuroNet to prioritize functional noncoding

511 variants that could significantly alter MN epigenomes (**Panel 1**, bottom). Using the Project
512 MinE[29] whole genome sequencing (WGS) data (**Panel 2**), we identified risk variants and genes
513 modulating ALS survival (**Panel 3**) via the Cox proportional-hazards (PH) model, in which we
514 collapsed all functional rare noncoding variants (predicted by NeuroNet) to their target genes
515 (predicted by the ABC model). To validate our discovery, we introduced the top candidate
516 variant chr7:76,009,472:C>T into patient-derived iMNs using CRISPR-Cas9 (**Panel 4**, left),
517 and verified its regulatory effects on *CCDC146* expression and ALS-associated phenotypes
518 (**Panel 4**, right), including cell survival, neurite growth, and TDP-43 mislocalization. Finally, we
519 designed ASOs suppressing *CCDC146* (**Panel 5**, left). We demonstrated that *CCDC146*
520 ASOs rescued ALS-associated survival deficits in patient-derived neurons across different
521 genetic backgrounds, and improved motor phenotypes, survival, and MN loss in an aggressive
522 ALS mouse model (**Panel 5**, right). iPSC, induced pluripotent stem cell; CRE, cis-regulatory
523 element; TF, transcription factor; fRV, functional rare variant; ASO, antisense oligonucleotide;
524 RNase H, Ribonuclease H.

525 **Fig. 2. NeuroNet predicts the noncoding variant effects in MNs.** (**A**) Schematics of the
526 NeuroNet model. (**B-C**) Predictive performance of NeuroNet evaluated based on the receiver
527 operating characteristic (ROC) curve (**B**) and precision-recall curve (**C**). The area under curve
528 (AUC) scores were calculated for different epigenomic features and metrics. The dashed gray
529 line indicates the random prediction. (**D**) Correlation between NeuroNet and DeepSEA
530 predictions for H3K4me1. *P*-value by two-sided Fisher's exact test. Neuron- and brain-related
531 tissues and cell types are highlighted. PCC, Pearson correlation coefficient. (**E**) Enrichment of
532 TFBS-disrupting variants within NeuroNet-prioritized functional variants (various cutoffs
533 applied). Enrichment was estimated by *t*-statistics. The box plot center line, limits, and
534 whiskers represent the median, quartiles, and 1.5x interquartile range (IQR), respectively.
535 TFBS, transcription factor binding site. (**F**) Enrichment of NeuroNet-prioritized functional
536 variants (various cutoffs applied) within MN CREs. OR is annotated in the solid line and 95%
537 CI is represented in the colored area. Statistics of the 95% cutoff were highlighted. *P*-value by
538 two-sided Fisher's exact test. OR, odds ratio; CI, confidence interval. (**G**) Enrichment of
539 functional or deleterious variants predicted by traditional approaches within NeuroNet-
540 prioritized functional variants (cutoff = 0.06). *P*-value by two-sided Fisher's exact test. Dot and
541 error bar indicate OR and 95% CI, respectively. MLE, Mean -log e-value.

542 **Fig. 3. Survival analysis based on rare functional noncoding variants in motor neurons.**
543 (**A**) Quantile-quantile (QQ) plot of burden-based survival analysis. Genes sharing the same
544 set of variants were collapsed into groups. The dashed red line indicates the significance cutoff
545 based on the Bonferroni correction. (**B**) Statistics of significant tests. The dot and error bar
546 represent the Cox model coefficient and SE, respectively. SE, standard error. (**C**) Kaplan-

547 Meier (KM) estimator plots for significant tests. Carriers for *ZNF839* variants were defined as
548 individuals having at least one variant from chr14:102,782,854:G>A and
549 chr14:102,786,108:G>T. (**D-E**) Regulatory relationships between noncoding variants and
550 target genes for chr7:76,009,472:C>T (**D**) and *ZNF839* (**E**), respectively. Peak signals from
551 MNs derived from CS14-iPSCs are visualized. Significant genes after Bonferroni correction
552 (adjusted $P < 0.05$) are highlighted in red. CRE, cis-regulatory element; E-P, enhancer-
553 promoter.

554 **Fig. 4. Experimental modeling of chr7:76,009,472:C>T in *C9orf72*-ALS iPSC-derived**
555 **motor neurons.** (**A**) Schematic of CRISPR/Cas9-mediated modeling of chr7:76,009,472:C>T
556 in *C9orf72*-ALS iPSC-derived motor neurons (iMNs). (**B**) Gene expression levels measured
557 by q-RT-PCR for three genes linked to the enhancer harboring chr7:76,009,472:C>T (n = 3
558 technical replicates). The bar plot and error bar represent the mean and standard error,
559 respectively. P-value by two-sided t-test. ns, not significant; **, $P < 0.01$. (**C**) Quantification of
560 neurite length (n = 3 technical replicates) in an age- and sex-matched normal control, an
561 isogenic control carrying a G4C2-repeat expansion of *C9orf72*, and edited iMNs, across 12
562 days of tracking using the Incucyte live-cell analysis system (Sartorius). The dot and error bar
563 represent the mean and standard error, respectively. P-value by two-sided t-test. *, $P < 0.05$.
564 (**D**) Quantification of the proportion of MNs exhibiting mislocalization of nuclear TDP-43 to the
565 cytoplasm in isogenic control and edited iMNs. The bar plot and error bar represent the mean
566 and standard error, respectively. P-value by two-sided t-test. (**E**) Representative images of
567 TDP-43 staining (yellow) in edited iMNs; nucleus (DAPI, blue) and cytoplasm (MAP2, red)
568 staining are also shown. MNs with cytoplasmic mislocalization of TDP-43 are shown (top
569 panels), together with a cytoplasmic inclusion (arrow); for comparison MNs are shown where
570 TDP-43 is localized to the nucleus (bottom panels). Scale bar 50μM. (**F**) Representative
571 images of iMNs where primary cilia were demarcated with Arl13b. Scale bar: original images
572 100μm; magnified inserts 25μm. (**G**) Measurement of length of primary cilia in iMNs carrying
573 chr7:76,009,472:C>T compared to an isogenic control carrying a *C9orf72* expansion (each
574 dot is the mean length from a single field of view, n = 3 separate motor neuron differentiations).
575 P-value by two-sided paired t-test. (**H**) Measurement of CRABP1 expression in iMNs carrying
576 chr7:76,009,472:C>T compared to an isogenic control carrying a *C9orf72* expansion (n = 4
577 technical replicates). P-value by two-sided paired t-test. (**I**) Immunoblotting for CRABP1 across
578 different conditions, with TUJ1 for normalization. (**J**) Mean velocity of EB3 comets (μm/sec) in
579 iMN carrying a *C9orf72* expansion with or without chr7:76,009,472:C>T (n = 3 technical
580 replicates; ≥4 axons imaged per condition per replicate). The bar plot and error bar represent
581 the mean and standard error, respectively. P-value by two-sided t-test. (**K**) Straightened axon

582 and representative kymographs depicting distance moved (μm) by EB3 comets over time
583 (seconds) in iMN carrying a *C9orf72* expansion with or without chr7:76,009,472:C>T.

584 **Fig. 5. Suppression of *CCDC146* in ALS patient-derived neurons via antisense**
585 **oligonucleotide (ASO).** (**A**) Schematic of our modeling strategy which utilized patient-derived
586 GFP-labeled iPSC-derived neurons in 3D spheroids (3D-iNs). Longitudinal tracking enables
587 accurate cell counting to determine survival following a glutamate insult together with a
588 scrambled ASO or a *CCDC146*-targeting ASO treatment. (**B**) Comparison of gene expression
589 levels measured by q-RT-PCR between *CCDC146*-targeting ASO and scrambled ASO
590 treatments. The bar plot and error bar represent the mean and standard error, respectively.
591 *P*-value by two-sided *t*-test. *n* = 4 technical replicates per condition. (**C**) Kaplan-Meier (KM)
592 estimator plots for four patient cell lines. The shaded area represents 95% confidence interval
593 (CI). *P*-value by log-rank test. C9, *C9orf72*; sALS, sporadic ALS. (**D**) Comparison of hazard
594 ratio measuring the incidence of neuron loss among different conditions and treatments. The
595 bar plot and error bar indicate the mean and standard error, respectively. NC, negative control
596 (scrambled); ctrl, control; C9, *C9orf72*; sALS, sporadic ALS. *P*-value by two-sided *t*-test. ns,
597 not significant. (**E**) Representative images of 3D-iNs derived from a sporadic ALS patient
598 treated with the scrambled ASO (top panels) or the *CCDC146*-targeting ASO (bottom panels).
599 Images were taken from approximately the same location on the plat at Day 0 (left panels)
600 and Day 8 (right panels), demonstrating neuronal loss which was rescued by *CCDC146*
601 suppression. Scale bar: 200μm. (**F-G**) Measurement of length of primary cilia (**F**) and CRABP1
602 expression (**G**) in iMNs after treatment with *CCDC146* ASO compared to a scrambled ASO or
603 untreated iMNs. The bar plot and error bar represent the mean and standard error,
604 respectively. *P*-value by two-sided *t*-test. ns, not significant. (**H**) Immunoblotting for CRABP1
605 across different treatments, with TUJ1 for normalization.

606 **Fig. 6. ASO suppression of *CCDC146 in vivo.*** (**A**) Strategy for evaluating the *Ccdc146* ASO
607 in Thy1::hTDP43Tg/Tg mice[17]. Intracerebroventricular (ICV) injections of the ASO were
608 performed at postnatal day 1 (D1), with motor phenotype evaluation every two days from
609 postnatal day 14 (D14). WT, wild type; MN, motor neuron. (**B**) Mouse endogenous *Ccdc146*
610 gene expression levels measured after treatment with the *Ccdc146* ASO or a scrambled
611 control ASO at D4 (*n* = 4 technical replicates). The bar plot and error bar represent the mean
612 and standard error, respectively. *P*-value by two-sided *t*-test. (**C**) Evaluations of gait
613 impairment, tremor, and kyphosis after treatment with the *Ccdc146* ASO or a scrambled ASO.
614 The line plot and error bar indicate the mean and standard error, respectively. *P*-value by two-
615 sided *t*-test. *, <0.05, **, <0.01, ***, <0.001. (**D**) Kaplan-Meier (KM) estimator plot for mouse
616 survival after treatment with either *Ccdc146* ASO (*n* = 32) or a scrambled control ASO (*n* =
617 23). The shaded area represents 95% confidence interval (CI). *P*-value by Cox regression. (**E**)

618 Motor neuron counts in TAR4/4 and wild type mice treated with either *Ccdc146* ASO or a
619 scrambled control ASO. Mice were sacrificed at D22 for histology. The bar plot and error bar
620 represent the mean and standard error, respectively. *P*-value by two-sided *t*-test. WT, wild
621 type; ns, not significant. (**F**) Immunostaining of large TUJ1+ neurons in the ventral horn of
622 TAR4/4 mice treated with a scrambled ASO or *Ccdc146* ASO used to perform motor neuron
623 counts. (**G**) Immunostaining of total TDP-43 in large TUJ1+ neurons in the ventral horn of wild-
624 type and TAR4/4 mice treated with a scrambled ASO or *Ccdc146* ASO. Dotta lines to outline
625 the cell body/nucleus used for quantification. WT, wild type; NC, negative control (scrambled);
626 TAR4/4, Thy1::hTDP43Tg/Tg. (**H**) Nuclear:cytoplasmic TDP-43 ratio at D22 across different
627 treatments. Each data point represents the average TDP-43 nuclear:cytoplasmic ratio for one
628 mouse. The bar plot and error bar indicate the mean and standard error, respectively. *P*-value
629 by two-sided *t*-test. WT, wild type; ns, not significant.

630

631 **KEY RESOURCES TABLE**

Reagent or Resource	Source	Identifier
Antibodies		
anti-Nestin	Abcam	#ab18102
anti-Pax6	Abcam	#ab5790
anti-Beta III Tubulin	Abcam	#ab18207
anti-HB9	Abcam	#ab221884
anti-Islet 1/2	Abcam	#AB109517
anti-NeuN	Millipore	#MAB377
anti-ChAT	Millipore	#AB144P
anti-SMI-32	Biolegend	#801701
anti-phospho TDP-43 (Ser409)	Affinity Biosciences	#AF7365
anti-TDP-43	Proteintech	#12892-1-AP
anti-MAP2	Proteintech	#17490-1-AP
anti-GFP	Novus Biologicals	#AB_10128178
anti-NFH	Millipore Sigma	#AB_477272
anti-CCDC146	Sigma	#HPA020082
CoraLite® Plus 488-conjugated anti-ARL13B	Proteintech	#CL488-17711
Donkey anti-Mouse IgG Alexa Fluor 488	Thermofisher	#A-21202
Donkey anti-Rabbit IgG Alexa Fluor 568	Thermofisher	#A-10042
Donkey anti-Goat IgG Alexa Fluor 555	Thermofisher	#A-21432
Goat anti-Guinea Pig IgG Alexa Fluor 647	Thermofisher	#A-21450

Donkey anti-Mouse IgG Alexa Fluor 594	Thermofisher	#A-32744
Donkey anti-Rabbit IgG, Alexa Fluor 488	Thermofisher	#A-21206
Donkey anti-Mouse IgG Alexa Fluor 568	Thermofisher	#A-10037
Goat anti-Chicken IgY (H+L) Alexa Fluor Plus 488	Thermofisher	#A-32931
Alexa Fluor 680-AffiniPure Donkey Anti-Rabbit IgG (H+L)	Jackson ImmunoResearch	#711-625-152
Chemicals, Peptides and Recombinant Proteins		
Hoechst 33342	ThermoFisher Scientific	#62249
Triton X-100	Sigma-Aldrich	#T8787
Donkey serum	Millipore	#S30-100ML
Bovine serum albumin	Sigma	#A2153
KnockOut™ DMEM/F-12	Gibco	#12660012
Neurobasal medium	Gibco	#21103049
Penicillin-Streptomycin	Sigma	#P4333
Gibco GlutaMAX™ Supplement	ThermoFisher Scientific	#35050061
N-2 supplement	ThermoFisher Scientific	#17502048
B-27 supplement	ThermoFisher Scientific	#17504001
mTeSR-Plus Medium	StemCell Technologies	#100-0276
Accutase™ Cell Dissociation Reagent	Gibco	#11599686
ReLeSR™	StemCell Technologies	#100-0483
Matrigel®	Corning	#356230
Poly-l-ornithine	Sigma	#P4957

Purmorphamine	Tocris Bioscience	#4551
DMH-1	Tocris Bioscience	#4126
SB431542	Tocris Bioscience	#1614
CHIR99021	Tocris Bioscience	#4423
ROCK inhibitor (Y-27632 dihydrochloride)	Tocris Bioscience	#1254
Compound E	Tocris Bioscience	#6476
All-trans retinoic acid	Sigma	#R2625
Valproic acid	Sigma	#P4543
hBDNF	R&D systems	#248-BDB-010/CF
IGF-1	R&D systems	#291-G1
CNTF	R&D systems	#257-NT-010/CF
Dulbecco's Phosphate Buffered Saline	Sigma	#D8537-500ML
EDTA	Sigma	#E5134
HEPES	Sigma	#H3375
PMSF protease inhibitor	ThermoFisher Scientific	#36978
Protease inhibitor tablet	Roche	#1697498
TEMED	ThermoFisher Scientific	#17919
Ammonium Persulfate	ThermoFisher Scientific	#17874
30% Acrylamide/Bis-acrylamide solution	Sigma	#A3699-100ML
TE Buffer, RNAse-free pH 8	ThermoFisher Scientific	#AM9849
Ethidium bromide solution	Sigma	#E1510
VeriFi mix red	PCRBio	#PB10.42-01

Tri reagent	Sigma	#93289-100ML
M-MLV reverse transcriptase	ThermoFisher Scientific	#28025-013
5x First Strand buffer	ThermoFisher Scientific	#18057-018
0.1M Dithiothreitol	ThermoFisher Scientific	#707265ML
dNTP Mix	ThermoFisher Scientific	#10534823
Random hexamer primer	ThermoFisher Scientific	#SO142
SYBR Green Brilliant III master mix	Agilent	#600882
Critical Commercial Assays		
Protein Bradford Assay	Bio-Rad	#5000201
GenElute Mammalian Genomic DNA Miniprep Kit	Sigma	#G1N350
Direct-zol RNA Miniprep Kit	Zymo Research	#R2050
QIAquick PCR Purification kit	Qiagen	#28104
QIAquick Gel Extraction Kit	Qiagen	#28506
Experimental Models: Cell Lines		
GM23338 (iPSC line derived from healthy volunteer)	Coriell Institute	#CVCL_F182
KOLF2.1J (iPSC line derived from healthy volunteer)	The Jackson Laboratory	#HPSI0114i-kolf_2
C9-183 (iPSC line derived from a patient with C9orf72-ALS)	The University of Sheffield	#RRID:CVCL_UF85
C9-183_chr7:76,009,472:C>T (iPSC line derived from a patient with C9orf72-ALS with CRISPR-engineered enhancer edit [chr7:76,009,472:C>T])	The University of Sheffield	-
Software and Algorithms		

R v4.0.1	https://cran.r-project.org/mirrors.html	
PRISM 10	GraphPad	
ICE CRISPR analysis tool	https://ice.synthego.com/#/	
CFX Maestro	Bio-Rad	
Harmony Imaging Analysis Software	PerkinElmer (Revvity)	
FIJI (FIJI Is Just ImageJ)	NIH	
Incucyte® software	Sartorius	
Image Studio™ Software	LI-COR	
IGV v2.4.16	https://software.broadinstitute.org/software/igv/	
NeuroNet	https://github.com/szhang1112/neuronet	

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Michael P. Snyder (mpsnyder@stanford.edu).

Material availability

All unique/stable reagents generated in this study are available from the Lead Contact without restriction.

Data availability

The epigenome and transcriptome profiling data of iPSC-derived motor neurons (iMN) used in the ABC model and NeuroNet are available at encodeproject.org with the following link: https://www.encodeproject.org/publications/de19555b-a49f-471c-bfbc-be3b628fe9bf/. NeuroNet variant effect scores can be downloaded at https://doi.org/10.6084/m9.figshare.25444534. The Project MinE WGS data are accessible upon application approval (https://www.projectmine.com/research/data-sharing/). The

AnswerALS WGS and RNA-seq data are available upon request approval (https://dataportal.answerals.org/home).

Code availability

The NeuroNet source code and tutorial are available at https://github.com/szhang1112/neuronet.

652 **METHODS**

653 ***Prediction of enhancer-promoter interactions using the ABC model***

654 We implemented the ABC model following the guidelines provided at
655 https://github.com/broadinstitute/ABC-Enhancer-Gene-Prediction. First, we called peaks for
656 the iMN ATAC-seq profiling data[21] using MACS2, and then identified the candidate enhancer
657 elements using "makeCandidateRegions.py" with peakExtendFromSummit = 250 and
658 nStrongestPeaks = 150000. The black-listed regions generated by the ENCODE 4
659 (https://www.encodeproject.org/) were used for removing enhancers overlapping regions with
660 anomalous sequencing reads. Second, we applied "run.neighborhoods.py" to quantify the
661 enhancer activities by counting ATAC-seq and H3K27ac ChIP-seq reads in candidate
662 enhancer regions. RNA-seq profiling of iMNs was also provided to inform expressed genes.
663 Quantile normalization was applied using K562 epigenomic data as the reference. At last,
664 using "predict.py" we computed the ABC scores by combining the enhancer activities
665 (calculated by the second step) with the Hi-C profiling. Hi-C data were fit to the power-law
666 model. The suggested threshold 0.02 was used to define valid E-P links.

667 ***NeuroNet model design and training***

668 The architecture of NeuroNet follows the Beluga structure[12], a powerful convolutional neural
669 network (CNN) designed to predict 2,002 epigenomic features from the DNA sequence, except
670 the last layer of NeuroNet is tailored to predict four epigenomic features (chromatin
671 accessibility, H3K27ac, H3K4me1, and H3K4me3) in iMNs.
672 The input of NeuroNet is the one-hot encoded 2,000-base-pair (bp) DNA sequence of
673 which the centered 200 bp region overlaps peaks of at least one epigenomic feature – called
674 valid sequence in this paper. The model outputs four probabilities indicating the peak status.
675 Given the encoded input sequence X_i (i = 1, ..., N), the model f is trained to minimize the
676 following binary cross-entropy loss:

677 $$L(\Omega) = \Sigma_{ij} (y_{ij} \cdot \log \sigma (f_j (X_i)) + (1 - y_{ij}) \cdot \log (1 - \sigma (f_j (X_i)))) / 4N,$$

678 Where j is the index of four epigenomic features, σ is the sigmoid function, y_{ij} is the binary
679 label of the j-th feature of the i-th valid sequence, and Ω is the set of network parameters.
680 Genome-wide valid sequence samples were randomly split into 1270000, 4000, and
681 164080 training, validation, and testing samples, respectively. The dataset was further
682 augmented by adding reverse complement sequences. Stochastic gradient descent (SGD)
683 with learning rate = 0.1, momentum = 0.9, and weight decay = 1×10^{-6} was adopted as the

684 optimizer. Batch size was set to 256. We also applied early stopping to prevent overfitting, in
685 which the AUPRC was selected as the monitoring metric for an imbalanced dataset.
686 The network structure of NeuroNet follows the Beluga architecture which has been
687 widely used in biological sequence modeling. Following the best practice[11,12,50], we trained
688 NeuroNet based on 1,270,000 training examples and evaluated model prediction using
689 164,080 sequence-epigenome pairs held-out on chromosomes 8 and 9.

690 ***Variant effect prediction using NeuroNet***

691 After the model was trained, we utilized NeuroNet to infer the functional effects of individual
692 variants on epigenomic features. Given a genetic variant, we construct a variant-centered DNA
693 sequence based on the reference genome, along with its reverse complement sequence. The
694 model predictions for both forward and reverse complement sequences are averaged to yield
695 the variant epigenomic scores $S_{alt,j}$. Similarly, the epigenomic scores $S_{ref,j}$ are also calculated.
696 We define the variant functional scores $S_{var,j}$ per epigenomic feature as the absolute difference
697 between reference and alternative epigenomic scores, i.e., $S_{var,j} = |S_{ref,j} - S_{alt,j}|$, where j indicates
698 different epigenomic features. The final NeuroNet score S_{NN} for each variant is defined as the
699 maximum of all functional scores, i.e., $S_{NN} = \max_j S_{var,j}$. As justified in the main text, we chose
700 0.06 as the threshold to define "functional noncoding variants" in MNs, that is, a noncoding
701 variant is kept for downstream analysis if and only if its NeuroNet score satisfies $S_{NN} > 0.06$.

702 ***Transcription factor binding motif analysis***

703 The TF binding motif analysis was performed using GimmeMotifs[70]. The short-range and long-
704 range enhancers were defined as enhancers with the mean absolute E-P distance below 5%
705 and above 95% of the distribution, respectively. The differential motifs between short- and
706 long-range enhancers were identified using the "gimme maelstrom" command. This command
707 combined the results from different methods using the Stouffer's method, yielding an
708 integrated *Z*-score. A higher *Z*-score means the motif better matches the sequence of the peak
709 region in one condition than the other.

710 ***Network analysis***

711 We downloaded the human PPIs from STRING version 12.0[71], including 19,622 proteins and
712 6,857,702 protein interactions. We extracted high-confidence (combined score >700) PPIs for
713 downstream analysis, including 16,185 proteins and 236,000 protein interactions. To eliminate
714 the bias caused by hub proteins[72], we ran the random walk with restart (RWR) algorithm over
715 the PPI network, wherein the restart probability was set to 0.5, resulting in a smoothed network
716 after preserving the top 5% predicted edges (N = 6,243,766). We then applied the Louvain

717 method[73], a classic community detection algorithm searching for densely connected modules,
718 to decompose the network into different modules. After the algorithm converged, we obtained
719 1,261 modules with an average size of 13 nodes.
720 The enrichment of genes of interest within each module was tested using the
721 hypergeometric test. Modules with false discovery rate (FDR) < 0.1 after the BH procedure
722 were reported.

723 ***Gene ontology analysis***

724 The gene ontology (GO) analysis in this work was performed based on clusterProfiler[74] and
725 enrichR[75]. All enrichment analyses were focused on the biological process and molecular
726 function. The gene enrichment was evaluated based on a two-sided Fisher's exact test, and
727 only terms with adjusted $P < 0.1$ (after BH correction) were reported. To remove the GO term
728 redundancy in some cases, we applied the “simplify” function provided by clusterProfile.

729 ***Benchmarking of NeuroNet prediction against other variant scores***

730 To benchmark NeuroNet against DeepSEA, the Pearson correlation between NeuroNet
731 prediction for four MN epigenomic features and DeepSEA prediction for matched epigenomic
732 features was calculated. For each MN epigenomic feature, all correlation coefficients were first
733 sorted and evenly divided into highly- and weakly-correlated sets, and then the rank of brain-
734 and neuron-related correlations was examined using a two-sided Fisher’s exact test within
735 these two sets.
736 SNP2TFBS[24] was used to predict the variant effect on TF binding affinity. The binding
737 affinities for different TFs were averaged for each studied variant to estimate its overall effect.
738 Given a particular NeuroNet cutoff, the noncoding variants were split into two groups according
739 to their NeuroNet predictions. We compared the SNP2TFBS scores between these two groups
740 of variants and examined the difference using a two-sided *t*-test. The *t*-statistics were reported
741 to indicate the SNP2TFBS score difference between two groups. The final NeuroNet score
742 threshold was determined at the largest difference.
743 With the determined NeuroNet threshold, the correlation between NeuroNet prediction
744 and other scores was inspected by comparing scores between NeuroNet-positive and
745 NeuroNet-negative variants using a two-sided Fisher’s exact test.

746 ***Whole-genome sequencing data in Project MinE***

747 Project MinE WGS data (Data Freeze 2) consists of 6,715 ALS patients and 2,885 healthy
748 controls recruited at specialized neuromuscular centers in the UK, Belgium, Germany, Ireland,
749 Italy, Spain, Turkey, the United States and the Netherlands[76]. Patients were diagnosed with

750 possible, probable or definite ALS according to the 1994 El-Escorial criteria[77]. All controls were
751 free of neuromuscular diseases and matched for age, sex and geographical location. Sample-
752 level and variant-level quality controls (QCs) have been described previously[76]. Briefly,
753 variants with missingness > 5% were removed, as were variants out of Hardy-Weinberg
754 equilibrium (HWE) in controls ($P < 1 \times 10^{-6}$) and monomorphic variants. All samples had
755 missingness <10% across all 22 chromosomes and samples with an inbreeding coefficient >3
756 standard deviations from the mean of the distribution were excluded, as were unexpected
757 related samples. After QCs, 6,025 ALS genomes were retained for downstream analysis.

758 ***ALS survival and risk analysis based on NeuroNet-prioritized rare noncoding variants***

759 After QCs, single nucleotide variants (SNVs) meeting all the following criteria were reserved
760 for the survival analysis: (1) minor allele frequency (estimated based on non-Finnish European
761 samples in the gnomAD whole-genome collection 2.1.1) < 0.5%; (2) minor allele count ≥ 3
762 within the QCed cohort; (3) residing within MN CREs (promoters and enhancers) defined by
763 the ABC model; (4) NeuroNet score > 0.06. These criteria together defined the "functional rare
764 noncoding variants" in our study. Next, screened variants were collapsed to their target genes
765 based on the ABC E-P links and the number of total mapped variants for each gene was
766 counted (i.e., gene burden). To associate the gene burden to survival, we applied the Cox
767 proportional-hazards model implemented by the "coxph" function in the R "survival" package,
768 with 26 covariates including age, sex, age^2, sex x age, sequencing platform, *C9orf72* status
769 and the first 20 principal components. Because many genes shared the same group of
770 regulatory variants resulting in duplicate tests, we collapsed those genes into unique sets and
771 carried out the Bonferroni multiple testing correction. Tests with adjusted $P < 0.05$ were
772 reported.

773 We also performed a case-control analysis by including the control samples in Project
774 MinE. 2,264 controls were reserved for analysis after QCs. Functional rare noncoding variants
775 were identified in the same way as that in the survival analysis, and then included in the case-
776 control study. Gene burdens were calculated similarly. To associate gene burden to the
777 disease status, we applied the SKAT-O method[30] implemented in the RVTESTS[30], with 25
778 covariates including age, sex, age^2, sex x age, sequencing platform, and the first 20 principal
779 components.

780 ***Benchmarking of survival analysis***

781 The benchmarking survival analysis was conducted the same as NeuroNet-powered survival
782 analysis, except the NeuroNet score was not used for prioritizing noncoding variants.

783 ***Transcriptome analysis for the AnswerALS cohort***

784 For AnswerALS dataset, RNA-seq data of iMNs and phenotype data were obtained for 245
785 ALS patients. Gene expression was normalized for gene length and then sequencing depth to
786 produce transcripts per kilobase million (TPM) using Kallisto v0.46.0[78]. Read counts were
787 trimmed Mean of M-values (TMM) normalized[79] to account for differences in library size. For
788 each gene, based on the expression level in and outside the top 25%, patients were split into
789 two groups – high-expression and low-expression groups. We used multivariable linear
790 regression to determine the relationship between gene expression level and rate of change in
791 the ALSFRS-R, and Cox proportional-hazards regression on two patient groups to determine
792 the relationship between gene expression and survival time from symptom onset. In both
793 cases, age of disease onset, sex, site of disease onset, and *C9orf72* status were included in
794 the model as covariates. Significance testing was performed for all genes expressed in iMNs
795 as determined by >0 read counts in more than half of samples, and mean read counts across
796 all samples >25th percentile. The ALSFRS-R was scored from a maximum of 48 where a
797 reduced score indicates increasing disability[69]. We calculated the rate of change in the
798 ALSFRS-R (i.e., Δ(ALSFRS-R)) assuming a linear decline from a score of 48 at symptom
799 onset.

800 ***CRISPR-editing of survival-associated genetic variant***

801 The variant chr7:76,009,472:C>T (hg19 genome build, ENSEMBL) was introduced to
802 *C9orf72*-ALS 183 iPSCs (passage 21) (RRID:CVCL_UF85) using CRISPR/Cas9-mediated
803 homology directed repair (Synthego[80]). Donor sequence:
804 GTGCCAGTCGCAGGCTCTCCAATCAGGAGGCTGCCGTGGGGCCAGGGGTGTGGCTCC
805 GGTTCCCACGCGGGGCGAATTCCCTCTCCCGGTAGTCTCCCCAG. gRNA sequence:
806 GGGAACCGGAGCCGCGCCCC. *C9orf72*-ALS 183 iPSCs were derived from a 51 year old
807 Caucasian male with *C9orf72* G4C2 repeat expansion and a diagnosis of familial ALS at the
808 time of sampling[81,82]. Editing efficiency was confirmed via Sanger sequencing using the
809 following primers: fwd (5'-3'): CAGAAGCCCTGAGGAGTGC, rvs (5'-3'):
810 GCCTCGGAGGTTTTTCTGGA. Sequencing traces were uploaded to the Synthego ICE
811 Analysis tool (v3) (https://ice.synthego.com), and the indel efficiency within the mixed pool was
812 calculated at 91%. Editing efficiency remained stable throughout motor neuron differentiation.

813 Sequence traces are available on request. CRISPR-edited cells were engineered by Synthego
814 (EditCo Bio).

815 ***Tissue culture of iPSCs and iMNs***

816 iPSCs were cultured in mTesR Plus media (StemCell Technologies) in Matrigel-coated (8.7
817 ug/cm²) 6-well plates (Corning) and maintained at 37°C, 5% CO2. Cells were passaged when
818 ~80% confluent using ReLeSR (StemCell Technologies), according to the manufacturer's
819 instructions. iPSCs were differentiated into neural progenitor cells (NPCs) using a modified
820 version of the dual SMAD inhibition protocol[21,83]. On the day after plating (day 1), after the
821 cells had reached ~100% confluence, the cells were washed once with PBS and the medium
822 was replaced with neural induction medium (50% of KnockOut DMEM/F-12, 50% of
823 Neurobasal), 0.5x N2 supplement (ThermoFisher), 1x Gibco GlutaMAX Supplement
824 (ThermoFisher), 0.5x B-27 supplement (ThermoFisher), 50 U ml−1 penicillin and 50 mg ml−1
825 streptomycin, supplemented with SMAD inhibitors (DMH-1 2 μm; SB431542 10 μm; and
826 CHIR99021 3 μm). This medium was changed every day for 6 days, and on day 7 the medium
827 was replaced with neuralization medium supplemented with DMH-1 2 μm, SB431542 10 μm
828 and CHIR 1 μm, All-Trans Retinoic Acid 0.1 μm (RA), and Purmorphamine 0.5 μm (PMN). The
829 cells were kept in this medium until day 12 when it was possible to observe a uniform
830 neuroepithelial sheet. At this point the cells were split 1:4 with Accutase (Gibco), onto matrigel
831 substrate in the presence of 10 μm ROCK inhibitor (Y-27632 dihydrochloride, Tocris), giving
832 rise to a sheet of neural progenitor cells (NPC). After 24 hours of incubation the medium was
833 changed to neural medium supplemented with RA 0.5 μm and PMN 0.1 μm, and the medium
834 was changed every day for 6 more days. On day 19 motor neuron progenitors were split with
835 accutase into plates pre-coated with poly-l-ornithine (10 μg/ml for 1hr at RT followed by 3x
836 washes in dH_2O) followed by matrigel (8.7 ug/cm² for 1hr at RT), and the medium was replaced
837 with motor neuron differentiation medium supplemented with RA 0.5 μm, PMN 0.1 μm,
838 compound E 0.1 μm (StemCell Technologies), BDNF 10 ng/mL, CNTF 10 ng/mL and IGF-1

839 10 ng/mL (Tocris) until day 28. On day 29, the media was replaced with neuronal maturation
840 media (neurobasal media supplemented with 1% of B27, BDNF 10 ng/mL, CNTF 10 ng/mL
841 and IGF-1 10 ng/mL). The cells were then fed every 72 hrs until day 40.

842 ***Treatment of cells with antisense oligonucleotides (ASOs)***

843 iMN monocultures were treated with 5 μM of either a scrambled control
844 (GCGACTATACGCGCAATATG) or CCDC146-targeting (TGCTGACTAATTGTGGTTAC)
845 ASO (Integrated DNA Technologies) in neuronal maturation media for 72 hours prior to
846 downstream assessment of ciliary function by immunocytochemistry or immunoblotting.

847 ***Measurement of gene expression in iMNs***

848 Gene expression was determined via quantitative real-time polymerase chain reaction (qRT-
849 PCR). Primer sequences were as follows: *FGL2: Fwd: TCAAAGTGTCCCAGCCAAGAA, Rvs:*
850 *ACTCTGTAGGTCTCACTGCTTC. FAM185BP: Fwd: TCAGGGTCTGGCTGTGTAGA, Rvs:*
851 *TCCCAGACAGATCACTTTGCC. CCDC146: Fwd: ACTCCACAAAGCTCATCAGAAAG, Rvs:*
852 *TCCTGAAGTTTTGATACAATTTTGC.* RNA was extracted using a Direct-zol RNA Purification
853 Kits (Zymo) according to the manufacturer's instructions. 2μg of total RNA was converted to
854 cDNA in an initial reaction mixture containing 1μL 10mM dNTPs, 1μL 40μm random hexamer
855 primer (ThermoFisher Scientific), and DNAse/RNAse-free water to a total volume of 14μL.
856 This mixture was heated for 5 minutes at 70°C then incubated on ice for 5 minutes. 4μL of 5x
857 FS buffer, 2μL 0.1M DTT, and 1μL M-MLV reverse transcriptase (200U/μL) (ThermoFisher
858 Scientific) were then added, and cDNA conversion was performed in a PCR thermocycler
859 (37°C for 60 minutes, 85°C for 10 minutes). cDNA was amplified using RT-PCR with Brilliant
860 III SYBR Green (Agilent) as per the manufacturer's instructions. Ct analysis was performed
861 using CFX Maestro software (BioRad). β-actin was chosen as a reference gene because
862 expression is stable in iPSC-derived neurons[84]. qPCR products were sequenced to confirm
863 target specificity and sequence traces are available on request. Statistical analysis was
864 performed using GraphPad Prism version 10.0.0.

865 ***Longitudinal survival tracking of iMNs***

866 Survival of iMN was assessed from day 28-40 using an Incucyte® Live Cell Analysis System
867 (Sartorius) at 20x magnification across 3 biological repeats. Neurite length (mm/mm^2) was
868 used as a readout for cell survival. For each biological repeat, 10 wells were imaged per group
869 with 9 fields of view per well. The neurotrack module (Sartorius) was used to determine neurite
870 metrics from brightfield images. In each case statistical significance was calculated by

applying a paired *t*-test at each timepoint. Statistical analysis was performed using GraphPad Prism version 10.0.0.

Immunocytochemistry

For immunostaining, neural progenitor cells (NPC) and motor neurons (MN) were washed with phosphate-buffered saline (PBS) and fixed with 4% paraformaldehyde for 10 min at room temperature. After fixation samples were washed three times with PBS and permeabilized with 0.3% Triton X-100 diluted in PBS for 5 min. The cells were subsequently blocked in 5% Donkey serum (Millipore) for 1h at room temperature (RT). After blocking, cells were incubated with the appropriate primary antibodies: mouse anti-nestin 1:200 [Abcam #ab22035]; rabbit anti-Pax6 1:50 [Abcam #ab5790]; guinea pig anti-MAP2 1:1000 [Synaptic Systems #188004]; chicken anti-TUJ1 1:1000 [Merck Millipore #AB9354]; mouse anti-SMI32 1:1000 [Biolegend #801701]; rabbit anti-HB9 1:250 [Abcam #ab221884]; goat anti-ChAT 1:50 [Merck #AB144P]; rabbit anti-CCDC146 [Atlas Antibodies #HPA020082] diluted in PBS containing 5% of DS overnight. Cells were then washed with PBS three times. Fluorescent secondary antibodies (Alexa Fluor 488, 561, or 647, diluted 1:1000 with 5% DS) (ThermoFisher Scientific #A-21202, #A-21432, A-21450, #A-32744, #A-21206) were subsequently added to the cells and incubated for 1h at RT. The samples were washed with PBS three more times and incubated with Hoechst 33342 (Invitrogen) for nuclear staining for 5 minutes. For primary cilium staining, a CoraLite® Plus 488-conjugated ARL13B polyclonal antibody (Proteintech #CL488-17711) was applied for 1hr after all washes to limit cross-reactivity with other primary antibodies of the same species. All experiments included cultures where the primary antibodies were not added; non-specific staining was not observed in these negative controls. Images were obtained using an Opera Phenix™ High Content Screening System (Revvity) at 40x magnification to generate a high-resolution z-stack at 0.5 µm intervals throughout the entire nuclear volume of the cell. Images were analyzed using the Harmony™ Image analysis system. 488, 561, and 647 nm lasers along with the appropriate excitation and emission filters were used. These settings were kept consistent while taking images from all separate differentiations.

Measurement of cilia length

CIlia length was the longest dimension of cilia immunostaining measured in maximum projection through visualised cilia. Each motor neuron has only one cilium. A technical replicate was taken to be the mean cilium length in a single field of view from at lest three separate motor neuron differentiations per condition. A paired t-test was used to determine statistical significance and account for variability between motor neuron differentiations.

905 ***Measurement of microtubule dynamics***

906 iPS-MN were transfected with 1 µg EB3_mScarlet (Addgene #98826, kindly gifted from the
907 Dorus Gadella lab) at day 39 using Lipofectamine™ 3000 (ThermoFisher Scientific), according
908 to the manufacturer's guidelines. For live imaging by TIRF microscopy 24 hours post
909 transfection, cells were cultured in phenol red-free neurobasal media supplemented with
910 10mM Hepes, 1% B-27 supplement, and imaged using a custom-built Nikon microscope with
911 an iLas2 multi-application unit (Gataca Systems). Time-lapse imaging of EB3 comets within
912 axons was performed at 200 ms intervals for 2 minutes using a 568 nm laser. Axons were
913 straightened with the segmented line tool in FIJI, resliced, and then converted into kymographs
914 using a maximum intensity z-projection. The velocity of EB3 comets was analysed using the
915 FIJI plugin, TrackMate[85].

916 ***Western blotting***

917 iMN were lysed on ice in IP-lysis buffer (50 mM Hepes pH7.5, 150 mM NaCl, 10% glycerol,
918 0.5% Triton X-100, 1 mM EDTA, 1 mM DTT, protease inhibitor cocktail (Sigma)), passed 10x
919 through 25-guage needles and incubated on ice for 10 minutes. Samples were centrifuged at
920 17,000 xg for 5 minutes at 4°C. Protein extracts were quantified using Bradford Reagent
921 (BioRAD), resolved by SDS–PAGE (15% resolving gel; 5% stacking gel), and electroblotted
922 onto nitrocellulose membranes. Membranes were blocked in 5% BSA + TBST (20 mM Tris,
923 137 mM NaCl, 0.2% (v/v) Tween® 20, pH 7.6) for 1 hour at room temperature (RT).
924 Membranes were probed using the relevant primary antibody [anti-CRABP1 (1:1000 dilution),
925 Proteintech #12588-1-AP; anti-TUJ1 (1:1000 dilution), Abcam #ab18207] diluted in 5%
926 BSA + TBST overnight at 4°C. Membranes were washed 3x in TBST, incubated with
927 an Alexa Fluor 680-AffiniPure Donkey Anti-Rabbit IgG (H+L) secondary antibody (Jackson
928 #711-625-152) diluted in TBST for 1 hour at RT, and washed 3x in TBST. Fluorescence signal
929 intensity was analysed using the Odyssey® Fc Imaging System (LI-COR Biosciences).

930 ***Generation of iPSC-derived 3D-induced neurons***

931 To generate iPSC-derived 3D-induced neurons (3D-iNs), iPSCs were first co-transfected
932 (Lipofectamine Stem, Thermofisher) with a piggybac expression cassette containing a dox-
933 inducible h*NGN2* transgene and mApple fluorescent reporter (addgene 182308) and the
934 Super piggyBac Transposase expression vector (SBI). After a brief (24h) recovery, transfected
935 iPSCs were replated using accutase+10µm ROCK-inhibitor (Ri, Selleckchem) at sparse
936 densities where an initial selection with 20µm blasticidin (Cayman, BSD) was performed. BSD
937 concentrations were increased until a highly pure population of mApple fluorescent cells were
938 observed and subsequently used for downstream studies. To generate the 3D-iN cultures,

939 iPSCs expressing the h*NGN2* piggybac construct were dissociated with accutase and 10µm
940 ROCK-inhibitor (Selleck) and seeded into Matrigel coated 6-well plates at ~200,000 cells per
941 well in induction medium containing DMEM+Glutamax (Thermofisher), NEAA (Gibco), 1%
942 penicillin/streptomycin, N2 supplement (Gibco) , and 10ng/ml each of BDNF (R&D) and NT-3
943 (Peprotech) plus doxycycline (2µg/mL) to drive the expression of the h*NGN2* cassette. 48
944 hours after induction, induction medium was replenished and BSD (20-40µg/mL) was added
945 if non-converting, non-transgenic cells were observed. 72 (Day 3) hours after induction, BrdU
946 (Millipore Sigma) at 40µm was added to eliminate dividing cells. By Day 5 of induction, the
947 pure population of converting iNs were replated using accutase and Ri (20µM) to ultra-low
948 attachment 96-well U-bottom plates (Corning) containing neuronal maintenance medium
949 (MM): neurobasal (Thermofisher), N2 and B27 supplements (Gibco),
950 1%penicillin/streptomycin, glutamax (Thermofisher, 100x), and NEAA (Thermofisher, 100x)
951 plus BDNF and NT-3 (10ng/mL each) at densities of ~50,000-80,000 neurons per well, where
952 the cells would coalesce into 3D-iN cultures. 3D-iNs were maintained in this format until Day
953 40 for survival tracking with medium changes (MM plus BDNF and NT3) every 3 days.

954 ***3D-iN longitudinal survival tracking***

955 3D-iNs were allowed to mature until Day 30 after dox and were then transduced with a
956 *SYN1*::GFP lentivirus to enable labeling of individual neurons as previously described[45,48]. 72-
957 hours prior to the start of the survival assay (Day 37 after Dox), 3D-iNs were treated with ASO
958 gapmers designed using an in-house strategy; 10µm of either a scrambled or *CCDC146* ASO
959 were administered via gymnotic uptake. On Day 39 (24 hours prior to start of tracking), 3D-
960 iNs were embedded in 100% Matrigel (Corning) for 30 minutes to allow gelling. MM was
961 replenished after the 3D-iNs were fixed. Following a baseline image at Day 0 of the tracking
962 (Day 40 of differentiation), glutamate (L-Glutamic acid monosodium salt hydrate; 5mM
963 Milipore) was added to the medium and replenished every other day thereafter. Continuous z-
964 stacks starting from the 3D-iN surface were captured every other day (from Day 0) using a
965 Zeiss AxioZoom.v16 wide-field upright fluorescence microscope. Image processing and
966 analysis were performed using ImageJ software (NIH). To quantify neuronal death ~100
967 neurons were tracked across 4 replicates per line per condition. Hazard ratios (HR) were
968 calculated using the log-rank method; for each replicate across each condition, the HR was
969 calculated as a comparison to the aggregate survival of the scrambled ASO control spheroids.

970 ***hTDP43 mouse studies***

971 Mouse strains C57BL/6J and B6;SJL-Tg(Thy1-TARDBP)4Singh/J (stock number: 012836)
972 were purchased from the Jackson Laboratory. The murine Thy1 promoter drives expression

973 of human TDP-43 in neurons and other CNS/PNS cells. Animals were housed in a
974 temperature and humidity-controlled environment under standard conditions with access to
975 food and water ad libitum. Animals were subjected to 12-hour light/dark cycles. All animal care
976 and experimental procedures were performed in accordance with local institution guidelines
977 of the University of Southern California and approved by the Institutional Animal Care and use
978 Committee of the University of Southern California (protocol number: 11938).
979 hTDP43 transgenic mice were purchased from the Jackson Laboratory (stock number:
980 012836). hTDP43$^{Tg/Tg}$ homozygous mutant mice were generated by crossing hTDP43$^{Tg/+}$
981 heterozygous mice with hTDP43$^{Tg/+}$ heterozygous mice. Genotyping primers are as follows:
982 Common primer: (5'- TGAAATCCGGGTGGTATTGG-3'); WT primer: (5'-
983 GGTGAGTTTAACCTTCAAGGGCT-3'); Transgene primer: (5'-
984 AGCTTGCTAGCGGATCCAGAC-3'). Homozygous transgenic mice were identified by a
985 single, 500 bp band, heterozygotes with a 303 bp and 500 bp bands, and wild-type (WT) mice
986 with a single 303 bp band. Homozygous transgenic animals and their WT littermate controls
987 were used for the subsequent ASO studies.

988 ***ASO mouse treatments***

989 25µg of a negative control or Ccdc146 ASO were administered by intracerebroventricular
990 (ICV) injection at post-natal day 1 (P1). ASOs were ordered from Integrated DNA
991 Technologies and were designed with chemical modifications to improve their potency and
992 efficiency for gene knockdown. Phosphorothioate linkages were used within the sequence to
993 provide resistance to nucleases, the 2'-O-methoxy-ehtyl (MOE) modification was added to
994 enhance nuclease stability and binding affinity for the target RNA. ASOs were ordered with
995 additional HPLC and sodium/salt exchange purification steps to remove potentially toxic
996 byproducts during synthesis. Purity of the ASOs was confirmed by IDT via mass spectrometry.
997 ASO sequences used in these studies are as follows (m=2'-O-Methyl, *-phosphorothioate
998 linkage, MOE = 2'-Methoxyethyl): msNC ASO: /52MOErC/*/i2MOErC/ /i2MOErT/ /i2MOErA/
999 /i2MOErT/ A*G*G*A*C*T*A*T*C*C*/i2MOErA/ /i2MOErG/ /i2MOErG/* /i2MOErA/*/32MOErA/;
1000 msCcdc146 ASO: /52MOErC/*/i2MOErA/ /i2MOErA/ /i2MOErT/ /i2MOErC/
1001 T*G*C*T*G*C*T*C*C*A*/i2 MOErC//i2MOErT/ /i2MOErA/* /i2MOErA/*/32MOErC/

1002 ***ICV injection procedure***

1003 D1 neonatal mice were anesthetized by placing them on a paper towel lined bucket of ice to
1004 induce hypothermia. The injection site targeting the lateral ventricles was defined as 2/5 the
1005 distance from the anatomical landmark, lambda to the eye. To perform the ICV injection, a
1006 900 series Hamilton syringe (10µL) adapted with the bottom portion of Neuros model 1701

1007 RN syringe. A 33-gauge needle was inserted 2mM deep, perpendicular to the skull surface
1008 into the lateral ventricle (left). Mice received 3µL of the respective NC ASO or Ccdc146 ASO
1009 in PBS (25µg). Following the injection, mice were placed under a heat lamp in a new cage
1010 containing soiled bedding from the parental cage and were monitored until they had fully
1011 recovered. Motor phenotyping (gait, kyphosis, and tremor) was examined starting at P14 and
1012 recorded every other day until D22. Phenotype score was determined according to a rubric
1013 previously reported (PMID: 28405022). A higher score denotes a more severe phenotype
1014 (0=no phenotype, 4=most severe/end stage). The humane euthanasia endpoint was
1015 established to be when the mouse was no longer able to right itself. Hydrogel, powdered
1016 pellets, and high caloric diet gel were placed on the bottom of each cage containing the mutant
1017 mice.

1018 ***Mouse tissue collection***

1019 WT or hTDP43$^{Tg/Tg}$ were treated with a NC ASO or Ccdc146 ASO at postnatal day 1 via ICV
1020 injection. For tissue collection for, D22 mice were anesthetized and intracardiac perfusion was
1021 performed first with PBS and then with 4% PFA in PBS, the tissue was then fixed in 4% PFA
1022 for another 4 hours. Tissue was then incubated in 15% sucrose in PBS+.01% sodium azide
1023 for 24 hours and then 30% sucrose in PBS +.01% sodium azide for another 48 hours prior to
1024 embedding and immunofluorescence analysis. For RNA isolation, mice were euthanized and
1025 isolated brains were separated into two hemispheres, one hemisphere and spinal cords were
1026 lysed for RNA analysis and the other was frozen at -80°C.

1027 ***Mouse histology***

1028 Cryoprotected tissues were embedded in O.C.T (Fisher, cat: 23730571) and cryosectioned
1029 into 16µM slices using a Leica CM3050S cryostat onto Superfrost Plus Microscope slides
1030 (Fisher) and stored at -80°C until use. For spinal cord neuron counts, lumbar spinal cord
1031 sections were rehydrated in 10x PBS for >40 minutes and washed in PBS+.1% Triton X-100
1032 for 10 minutes. Slides were then washed in PBS for 2, 5-minute washes. The slides were then
1033 incubated with NeuroTrace (Fluorescent Nissl Stain, Thermo N2148) diluted 1:50 in PBS for
1034 20 minutes in the dark. After removing the stain, slides were washed for another 10 minuites
1035 in PBS+.1% Triton-X-100 for 10 minutes, 2x5 minute washes in PBS and 2 hours at room
1036 temperature/overnight at 4°C in PBS prior to mounting. To quantify the neurons in the ventral
1037 horn region of the spinal cord, 3 sections per animal were stained with NeuroTrace and image
1038 using a Zeiss LSM 800 microscope. The lateral motor column was used for analysis. The
1039 counts were quantified based on the average total neuron counts in the lumbar hemicords
1040 across all 3 sections per animal, per treatment condition. To quantify the cellular localization

1041 of TDP43/tdp43, the nuclear and cytoplasmic fluorescence intensity was measured using an
1042 antibody with reactivity for both murine and human protein (10782-2-AP, Proteintech, 1:400).
1043 The fluorescence intensity of the TDP43 staining in the cell nucleus and cytoplasm were
1044 measured using Image J and recorded as the nuclear to cytoplasmic ratio. TUJ1+ motor
1045 neurons in the ventral spinal cord were imaged using Zeiss LSM 800 with oil immersion at
1046 63x. 15–25 neurons were quantified per animal.

1047 ***Small-molecule drug search***

1048 We searched chemical activity databases ChEMBL[86] and PubChem[87] for CCDC146 and its
1049 similar proteins, but did not obtain any active compounds for these proteins. Here, similar
1050 proteins were defined as those with at least 50% sequence identities to the canonical
1051 sequence of human CCDC146 protein, provided by the UniProt database[88]. We also searched
1052 patent databases including WIPO (https://patentscope.wipo.int/), EPO
1053 (https://www.epo.org/en/searching-for-patents) and USPTO
1054 (https://www.uspto.gov/patents/search), using search terms "CCDC146" and "coiled-coil
1055 domain-containing protein 146", to confirm that there is currently no available drugs for this
1056 target.

1057 ***Protein structure and binding site prediction***

1058 The predicted protein structure of CCDC146 protein was obtained from the AlphaFold Protein
1059 Structure Database[89] (https://alphafold.ebi.ac.uk/). To evaluate the druggability of CCDC146,
1060 we predicted potential ligand binding sites on its surface and obtained negative prediction
1061 results using P2Rank[67], which is an accurate ligand binding site prediction tool.

1062 **Supplementary Figures**

1063 **Supplementary Fig. 1. Enhancer-promoter (E-P) interactions in human motor neurons.**
1064 (**A**) Graphical illustration of the ABC model. TSS, transcription start site; E, enhancer; G, gene.
1065 (**B**) Distribution of distance from enhancer to TSS. (**C**) Statistics of E-P links, including number
1066 of enhancers per gene, number of genes per enhancer, and number of genes skipped by
1067 enhancer. enh, enhancer. (**D**) Overlap between E-P links in MNs and those in neuronal cells
1068 ($NeuN^+$). (**E**) Distribution comparison of E-P links between MNs and neuronal cells. Density
1069 curves were added. Dashed vertical lines represent distribution medians. *P*-value by two-side
1070 *t*-test. (**F**) Differential motif enrichment between long-range and short-range E-P links. LR,
1071 long-range; SR, short-range. (**G**) Network module enriched with genes targeted by short-range
1072 E-P links. *P*-value by hypergeometric test. (**H**) Gene ontology (GO) enrichment for M56 genes.

1073 Redundant GO terms were removed using the “simplify” function provided by “clusterProfiler”.
1074 GO terms with adjusted $P < 0.1$ were visualized. *P*-value by two-sided Fisher’s exact test.

1075 **Supplementary Fig. 2. Additional results for E-P links and NeuroNet prediction.** (**A**) Gene
1076 ontology (GO) enrichment for genes with short-range enhancers in iPSNs. GO terms with
1077 adjusted $P < 0.1$ (BH correction) were shown. *P*-value by two-sided Fisher’s exact test. (**B-D**)
1078 Pearson correlations between NeuroNet and DeepSEA predictions for H3K4me3 (**B**),
1079 H3K27ac (**C**), and ATAC (**D**). *P*-value by two-sided Fisher’s exact test. Neuron- and brain-
1080 related tissues and cell types were highlighted. PCC, Pearson correlation coefficient.

1081 **Supplementary Fig. 3. Additional results for rare variant burden analyses.** (**A**) Regulatory
1082 relationships between chr2:51,259,489:G>C and target genes. Peak signals from MNs derived
1083 from CS14-iPSCs were visualized. Significant genes in NeuroNet-informed survival analysis
1084 were highlighted in red. CRE, cis-regulatory element. (**B**) Quantile-quantile (QQ) plot of a
1085 standard rare variant burden-based survival analysis without NeuroNet-based variant filtering.
1086 Genes sharing the same set of variants were collapsed into groups. (**C**) Overlap of genes
1087 between survival and risk analysis with different *P*-value cutoffs. Overlap was examined by
1088 two-sided Fisher’s exact test. The dot and error bar indicate the odds ratio and 95% CI,
1089 respectively. The dashed gray line represents odds ratio = 1. CI, confidence interval.

1090 **Supplementary Fig. 4. Additional results for CRISPR/Cas9-mediated modeling of**
1091 **chr7:76,009,472:C>T in iPSC-derived *C9orf72*-ALS MNs.** (**A**) Representative images of
1092 iPSC-derived MNs (iMN). Top panels: isogenic control; bottom panels: edited. iMNs were
1093 studied at day 40 post-iPSC-stage when they expressed a battery of neuronal markers,
1094 including SMI-32, MAP2, and TUJ1. Scale bar: 50μm. (**B**) Quantification of the number of
1095 neurite branch-points in an age- and sex-matched normal control, an isogenic control carrying
1096 a G4C2-repeat expansion of *C9orf72*, and chr7:76,009,472:C>T-edited iMNs across 12 days
1097 of tracking using an Incucyte live-cell analysis system (Sartorius). The dot and error bar
1098 represent the mean and standard error, respectively. (**C**) Representative images from day 1
1099 and day 11 for isogenic control neurons carrying a G4C2-repeat expansion of *C9orf72* (top
1100 panels) and chr7:76,009,472:C>T-edited neurons (bottom panels). Cell soma are coloured
1101 yellow and neurites are coloured purple. Scale bar 1.6mm. (**D**) Immunocytochemistry for
1102 CCDC146 and primary cilia stained with Arl13b in iMNs derived from a neurologically healthy
1103 control. Scale bar: 50μm.

1104 **Supplementary Fig. 5. Additional results for ASO suppression of *CCDC146* in ALS**
1105 **patient-derived neurons.** (**A**) Kaplan-Meier (KM) estimator plots of 3D-iN survival starting
1106 (day 0) from day 40 after initiation of iPSC differentiation. The shaded area represents 95%

1107 confidence interval (CI). *P*-value by log-rank test. C9, *C9orf72*; sALS, sporadic ALS; ctrl,
1108 control. (**B**) Representative images of control iNs treated with the scrambled ASO (top panels)
1109 or the *CCDC146* ASO (bottom panels). Images were taken from approximately the same
1110 location on the plat at Day 0 (left panels) and Day 8 (right panels) showing no obvious toxicity.
1111 Scale bar: 200μm. Ctrl, control. (**C**) Quantification of *CCDC146* mRNA expression via q-RT-
1112 PCR in n = 1 control line and n = 8 ALS patient-derived iN lines. n = 3 technical replicates per
1113 cell line. The central bar and error bar represent the mean and standard error, respectively.
1114 The red dashed line indicates control expression level. *P*-value by the Dunnett's test. **,
1115 adjusted $P < 0.01$; ****, adjusted $P < 0.0001$; ns, not significant; #, patient cell lines used in
1116 ASO experiments. (**D**) Immunocytochemistry for primary cilia stained with Arl13b across
1117 different treatments. Scale bar: 50μm.

1118 **Supplementary Fig. 6. The AlphaFold-predicted three-dimensional structure of**
1119 **CCDC146 protein.** (**A**) The carton illustration of the predicted protein structure of CCDC146,
1120 obtained from https://alphafold.ebi.ac.uk/entry/Q8IYE0. pLDDT, predicted local distance
1121 difference test. (**B**) Protein surface of the predicted structure, illustrated using PyMol
1122 (https://pymol.org/).

1123 **Supplementary Fig. 7. Relationship between *ZNF839* expression and clinical markers**
1124 **of disease progression.** (**A**) Correlation between the expression of *ZNF839* and rate of
1125 disease progression as measured by clinical symptoms. The ALSFRS-R is scored from a
1126 maximum of 48 with a reduction in the score signifying increased disability. The red line
1127 indicates the regression line; the gray area represents the standard error. *P*-value by
1128 multivariable linear regression with sex, age of onset, site of onset, and *C9orf72* status as
1129 covariates. (**B**) Kaplan-Meier (KM) estimator plots of patients with high and low *ZNF839*
1130 expression. Gene expression level in the top 25% was defined as "high expression" (n = 59),
1131 otherwise as "low expression" (n = 176). The difference in mean survival between two groups
1132 was 0.75 years. *P*-value by Cox proportional-hazards (PH) regression.

1133 **Supplementary Tables**

1134 **Supplementary Table 1.** Enhancer-promoter links in iPSC-derived motor neurons

1135 **Supplementary Table 2.** Summary statistics of SKAT-O test for ALS risk

1136 **Supplementary Table 3.** Summary statistics of NeuroNet-powered rare variant burden test
1137 for ALS survival

1138 **Supplementary Table 4.** Summary statistics of rare variant burden test for ALS survival
1139 without variant filtering

1140 **Supplementary Table 5.** Clinical phenotypes of sequenced ALS patients carrying survival-
1141 associated rare noncoding variants

1142 **Supplementary Table 6.** Clinical phenotypes of patient iPSC lines

1143

1144

1145 **References**

1146 1. Feldman, E.L., Goutman, S.A., Petri, S., Mazzini, L., Savelieff, M.G., Shaw, P.J., and
1147 Sobue, G. (2022). Amyotrophic lateral sclerosis. Lancet *400*, 1363–1380.
1148 2. Chiò, A., Logroscino, G., Hardiman, O., Swingler, R., Mitchell, D., Beghi, E., Traynor,
1149 B.G., and Eurals Consortium (2009). Prognostic factors in ALS: A critical review.
1150 Amyotroph. Lateral Scler. *10*, 310–323.
1151 3. Cooper-Knock, J., Hewitt, C., Highley, J.R., Brockington, A., Milano, A., Man, S.,
1152 Martindale, J., Hartley, J., Walsh, T., Gelsthorpe, C., et al. (2012). Clinico-pathological
1153 features in amyotrophic lateral sclerosis with expansions in C9ORF72. Brain *135*, 751–
1154 764.
1155 4. Al-Chalabi, A., van den Berg, L.H., and Veldink, J. (2017). Gene discovery in
1156 amyotrophic lateral sclerosis: implications for clinical management. Nat. Rev. Neurol.
1157 *13*, 96–104.
1158 5. Cooper-Knock, J., Harvey, C., Zhang, S., Moll, T., Timpanaro, I.S., Kenna, K.P.,
1159 Iacoangeli, A., and Veldink, J.H. (2021). Advances in the genetic classification of
1160 amyotrophic lateral sclerosis. Curr. Opin. Neurol. *34*, 756–764.
1161 6. Li, X., Li, Z., Zhou, H., Gaynor, S.M., Liu, Y., Chen, H., Sun, R., Dey, R., Arnett, D.K.,
1162 Aslibekyan, S., et al. (2020). Dynamic incorporation of multiple in silico functional
1163 annotations empowers rare variant association analysis of large whole-genome
1164 sequencing studies at scale. Nat. Genet. *52*, 969–983.
1165 7. Zhang, F., and Lupski, J.R. (2015). Non-coding genetic variants in human disease.
1166 Hum. Mol. Genet. *24*, R102–R110.
1167 8. Almeida, N., Chung, M.W.H., Drudi, E.M., Engquist, E.N., Hamrud, E., Isaacson, A.,
1168 Tsang, V.S.K., Watt, F.M., and Spagnoli, F.M. (2021). Employing core regulatory circuits
1169 to define cell identity. EMBO J. *40*, e106785.
1170 9. GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis
1171 Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx
1172 (eGTEx) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA,
1173 Biospecimen Collection Source Site—NDRI, et al. (2017). Genetic effects on gene
1174 expression across human tissues. Nature *550*, 204–213.
1175 10. Li, X., Kim, Y., Tsang, E.K., Davis, J.R., Damani, F.N., Chiang, C., Hess, G.T., Zappala,
1176 Z., Strober, B.J., Scott, A.J., et al. (2017). The impact of rare variation on gene
1177 expression across tissues. Nature *550*, 239–243.
1178 11. Zhou, J., and Troyanskaya, O.G. (2015). Predicting effects of noncoding variants with
1179 deep learning–based sequence model. Nat. Methods *12*, 931–934.
1180 12. Zhou, J., Theesfeld, C.L., Yao, K., Chen, K.M., Wong, A.K., and Troyanskaya, O.G.

1181 (2018). Deep learning sequence-based ab initio prediction of variant effects on
1182 expression and disease risk. Nat. Genet. *50*, 1171–1179.
1183 13. Kelley, D.R., Reshef, Y.A., Bileschi, M., Belanger, D., McLean, C.Y., and Snoek, J.
1184 (2018). Sequential regulatory activity prediction across chromosomes with convolutional
1185 neural networks. Genome Res. *28*, 739–750.
1186 14. Avsec, Ž., Agarwal, V., Visentin, D., Ledsam, J.R., Grabska-Barwinska, A., Taylor, K.R.,
1187 Assael, Y., Jumper, J., Kohli, P., and Kelley, D.R. (2021). Effective gene expression
1188 prediction from sequence by integrating long-range interactions. Nat. Methods *18*,
1189 1196–1203.
1190 15. Yuan, H., and Kelley, D.R. (2022). scBasset: sequence-based modeling of single-cell
1191 ATAC-seq using convolutional neural networks. Nat. Methods *19*, 1088–1096.
1192 16. Agarwal, V., and Shendure, J. (2020). Predicting mRNA Abundance Directly from
1193 Genomic Sequence Using Deep Convolutional Neural Networks. Cell Rep. *31*, 107663.
1194 17. Wils, H., Kleinberger, G., Janssens, J., Pereson, S., Joris, G., Cuijt, I., Smits, V.,
1195 Ceuterick-de Groote, C., Van Broeckhoven, C., and Kumar-Singh, S. (2010). TDP-43
1196 transgenic mice develop spastic paralysis and neuronal inclusions characteristic of ALS
1197 and frontotemporal lobar degeneration. Proc. Natl. Acad. Sci. U. S. A. *107*, 3858–3863.
1198 18. Muroňová, J., Kherraf, Z.-E., Giordani, E., Eckert, S., Cazin, C., Amiri-Yekta, A.,
1199 Lambert, E., Chevalier, G., Martinez, G., Neirijnck, Y., et al. (2023). Lack of CCDC146,
1200 a ubiquitous centriole and microtubule-associated protein, leads to non-syndromic male
1201 infertility in human and mouse. bioRxiv, 2023.02.27.530236.
1202 https://doi.org/10.1101/2023.02.27.530236.
1203 19. De Decker, M., Zelina, P., Moens, T.G., Beckers, J., Contardo, M., Dittlau, K.S., Van
1204 Schoor, E., Ronisz, A., Eggermont, K., Moisse, M., et al. (2024). *C21ORF2* mutations
1205 point towards primary cilia dysfunction in amyotrophic lateral sclerosis. Brain.
1206 https://doi.org/10.1093/brain/awae331.
1207 20. Fulco, C.P., Nasser, J., Jones, T.R., Munson, G., Bergman, D.T., Subramanian, V.,
1208 Grossman, S.R., Anyoha, R., Doughty, B.R., Patwardhan, T.A., et al. (2019). Activity-by-
1209 contact model of enhancer-promoter regulation from thousands of CRISPR
1210 perturbations. Nat. Genet. *51*, 1664–1669.
1211 21. Zhang, S., Cooper-Knock, J., Weimer, A.K., Shi, M., Moll, T., Marshall, J.N.G., Harvey,
1212 C., Nezhad, H.G., Franklin, J., Souza, C.D.S., et al. (2022). Genome-wide identification
1213 of the genetic basis of amyotrophic lateral sclerosis. Neuron *110*, 992–1008.e11.
1214 22. Richter, F., Morton, S.U., Kim, S.W., Kitaygorodsky, A., Wasson, L.K., Chen, K.M.,
1215 Zhou, J., Qi, H., Patel, N., DePalma, S.R., et al. (2020). Genomic analyses implicate
1216 noncoding de novo variants in congenital heart disease. Nat. Genet. *52*, 769–777.
1217 23. GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects

1218 across human tissues. Science *369*, 1318–1330.
1219 24. Kumar, S., Ambrosini, G., and Bucher, P. (2017). SNP2TFBS - a database of regulatory
1220 SNPs affecting predicted transcription factor binding site affinity. Nucleic Acids Res. *45*,
1221 D139–D144.
1222 25. Davydov, E.V., Goode, D.L., Sirota, M., Cooper, G.M., Sidow, A., and Batzoglou, S.
1223 (2010). Identifying a High Fraction of the Human Genome to be under Selective
1224 Constraint Using GERP++. PLoS Comput. Biol. *6*, e1001025.
1225 26. Ritchie, G.R.S., Dunham, I., Zeggini, E., and Flicek, P. (2014). Functional annotation of
1226 noncoding sequence variants. Nat. Methods *11*, 294–296.
1227 27. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J.
1228 (2014). A general framework for estimating the relative pathogenicity of human genetic
1229 variants. Nat. Genet. *46*, 310–315.
1230 28. Huang, Y.-F., Gulko, B., and Siepel, A. (2017). Fast, scalable prediction of deleterious
1231 noncoding variants from functional and population genomic data. Nat. Genet. *49*, 618–
1232 624.
1233 29. Project MinE: study design and pilot analyses of a large-scale whole-genome
1234 sequencing study in amyotrophic lateral sclerosis (2018). Eur. J. Hum. Genet. *26*, 1537–
1235 1546.
1236 30. Lee, S., Emond, M.J., Bamshad, M.J., Barnes, K.C., Rieder, M.J., Nickerson, D.A.,
1237 NHLBI GO Exome Sequencing Project—ESP Lung Project Team, Christiani, D.C.,
1238 Wurfel, M.M., and Lin, X. (2012). Optimal unified approach for rare-variant association
1239 testing with application to small-sample case-control whole-exome sequencing studies.
1240 Am. J. Hum. Genet. *91*, 224–237.
1241 31. Zhan, X., Hu, Y., Li, B., Abecasis, G.R., and Liu, D.J. (2016). RVTESTS: an efficient and
1242 comprehensive tool for rare variant association analysis using sequence data.
1243 Bioinformatics *32*, 1423–1426.
1244 32. Nicolas, A., Kenna, K.P., Renton, A.E., Ticozzi, N., Faghri, F., Chia, R., Dominov, J.A.,
1245 Kenna, B.J., Nalls, M.A., Keagle, P., et al. (2018). Genome-wide Analyses Identify
1246 KIF5A as a Novel ALS Gene. Neuron *97*, 1268–1283.e6.
1247 33. Baron, D.M., Fenton, A.R., Saez-Atienzar, S., Giampetruzzi, A., Sreeram, A.,
1248 Shankaracharya, Keagle, P.J., Doocy, V.R., Smith, N.J., Danielson, E.W., et al. (2022).
1249 ALS-associated KIF5A mutations abolish autoinhibition resulting in a toxic gain of
1250 function. Cell Rep. *39*, 110598.
1251 34. Li, C., Wei, Q., Hou, Y., Lin, J., Ou, R., Zhang, L., Jiang, Q., Xiao, Y., Liu, K., Chen, X.,
1252 et al. (2023). Genome-wide analyses identify NEAT1 as genetic modifier of age at onset
1253 of amyotrophic lateral sclerosis. Mol. Neurodegener. *18*, 77.
1254 35. Kuriyama, S., Tamiya, Y., and Tanaka, M. (2017). Spatiotemporal expression of UPK3B

1255 and its promoter activity during embryogenesis and spermatogenesis. Histochem. Cell
1256 Biol. *147*, 17–26.
1257 36. Shi, Y., Lin, S., Staats, K.A., Li, Y., Chang, W.-H., Hung, S.-T., Hendricks, E., Linares,
1258 G.R., Wang, Y., Son, E.Y., et al. (2018). Haploinsufficiency leads to neurodegeneration
1259 in C9ORF72 ALS/FTD human induced motor neurons. Nat. Med. *24*, 313–325.
1260 37. Fiesel, F.C., Schurr, C., Weber, S.S., and Kahle, P.J. (2011). TDP-43 knockdown
1261 impairs neurite outgrowth dependent on its target histone deacetylase 6. Mol.
1262 Neurodegener. *6*, 64.
1263 38. Klim, J.R., Williams, L.A., Limone, F., Juan, I.G.S., Davis-Dusenbery, B.N., Mordes,
1264 D.A., Burberry, A., Steinbaugh, M.J., Gamage, K.K., Kirchner, R., et al. (2019). ALS-
1265 implicated protein TDP-43 sustains levels of STMN2, a mediator of motor neuron growth
1266 and repair. Preprint, https://doi.org/10.1038/s41593-018-0300-4
1267 https://doi.org/10.1038/s41593-018-0300-4.
1268 39. Neumann, M., Sampathu, D.M., Kwong, L.K., Truax, A.C., Micsenyi, M.C., Chou, T.T.,
1269 Bruce, J., Schuck, T., Grossman, M., Clark, C.M., et al. (2006). Ubiquitinated TDP-43 in
1270 frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science *314*, 130–
1271 133.
1272 40. Lin, Y.-L., Nhieu, J., Liu, P.-Y., Le, G., Lee, D.J., Wei, C.-W., Lin, Y.-W., Oh, S.-H.,
1273 Lowe, D., and Wei, L.-N. (2022). CRABP1-CaMKII-Agrn regulates the maintenance of
1274 neuromuscular junction in spinal motor neuron. Cell Death Differ. *29*, 1744–1756.
1275 41. Cooper-Knock, J., Jenkins, T., and Shaw, P.J. (2013). Clinical and Molecular Aspects of
1276 Motor Neuron Disease. Preprint,
1277 https://doi.org/10.4199/c00093ed1v01y201309gmm004
1278 https://doi.org/10.4199/c00093ed1v01y201309gmm004.
1279 42. Lin, Y.-H., Dodd, J.E., Cutillo, L., Castelli, L.M., Mihaylov, S.R., Norris, K., Higginbottom,
1280 A., Walsh, M.J., Cooper-Knock, J., Robin Highley, J., et al. (2024). GRASPS: a simple-
1281 to-operate translatome technology reveals omics-hidden disease-associated pathways
1282 in TDP-43-related amyotrophic lateral sclerosis. bioRxiv, 2024.03.04.583294.
1283 https://doi.org/10.1101/2024.03.04.583294.
1284 43. Schrøder, J.M., Larsen, J., Komarova, Y., Akhmanova, A., Thorsteinsson, R.I.,
1285 Grigoriev, I., Manguso, R., Christensen, S.T., Pedersen, S.F., Geimer, S., et al. (2011).
1286 EB1 and EB3 promote cilia biogenesis by several centrosome-related mechanisms. J
1287 Cell Sci *124*, 2539–2551.
1288 44. Zhang, Y., He, M., and Pan, J. (2025). Axonemal microtubule dynamics in the assembly
1289 and disassembly of cilia. Biochem Soc Trans *53*. https://doi.org/10.1042/BST20240688.
1290 45. Linares, G.R., Li, Y., Chang, W.H., Rubin-Sigler, J., Mendonca, S., Hong, S., Eoh, Y.,
1291 Guo, W., Huang, Y.H., Chang, J., et al. (2023). SYF2 suppression mitigates

1292 neurodegeneration in models of diverse forms of ALS. Cell Stem Cell *30*.
1293 https://doi.org/10.1016/j.stem.2023.01.005.
1294 46. Shaw, P.J., and Ince, P.G. (1997). Glutamate, excitotoxicity and amyotrophic lateral
1295 sclerosis. J. Neurol. *244 Suppl 2*, S3–S14.
1296 47. Van Den Bosch, L., Van Damme, P., Bogaert, E., and Robberecht, W. (2006). The role
1297 of excitotoxicity in the pathogenesis of amyotrophic lateral sclerosis. Biochim. Biophys.
1298 Acta *1762*, 1068–1082.
1299 48. Bowles, K.R., Silva, M.C., Whitney, K., Bertucci, T., Berlind, J.E., Lai, J.D., Garza, J.C.,
1300 Boles, N.C., Mahali, S., Strang, K.H., et al. (2021). ELAVL4, splicing, and glutamatergic
1301 dysfunction precede neuron loss in MAPT mutation cerebral organoids. Cell *184*.
1302 https://doi.org/10.1016/j.cell.2021.07.003.
1303 49. Becker, L.A., Huang, B., Bieri, G., Ma, R., Knowles, D.A., Jafar-Nejad, P., Messing, J.,
1304 Kim, H.J., Soriano, A., Auburger, G., et al. (2017). Therapeutic reduction of ataxin-2
1305 extends lifespan and reduces pathology in TDP-43 mice. Nature *544*, 367–371.
1306 50. Chen, K.M., Cofer, E.M., Zhou, J., and Troyanskaya, O.G. (2019). Selene: a PyTorch-
1307 based deep learning library for sequence data. Nat. Methods *16*, 315–318.
1308 51. Diekstra, F.P., van Vught, P.W.J., van Rheenen, W., Koppers, M., Pasterkamp, R.J.,
1309 van Es, M.A., Schelhaas, H.J., de Visser, M., Robberecht, W., Van Damme, P., et al.
1310 (2012). UNC13A is a modifier of survival in amyotrophic lateral sclerosis. Neurobiol.
1311 Aging *33*, 630.e3–e8.
1312 52. Miller, T.M., Cudkowicz, M.E., Genge, A., Shaw, P.J., Sobue, G., Bucelli, R.C., Chiò, A.,
1313 Van Damme, P., Ludolph, A.C., Glass, J.D., et al. (2022). Trial of Antisense
1314 Oligonucleotide Tofersen for SOD1 ALS. N. Engl. J. Med. *387*, 1099–1110.
1315 53. Yamashita, S., Migita, A., Hayashi, K., Hirahara, T., Kimura, E., Maeda, Y., Hirano, T.,
1316 and Uchino, M. (2010). Amyotrophic lateral sclerosis in a patient with Kartagener
1317 syndrome. Amyotroph. Lateral Scler. *11*, 402–404.
1318 54. De Decker, M., Zelina, P., Moens, T.G., Eggermont, K., Moisse, M., Veldink, J.H., Van
1319 Den Bosch, L., Jeroen Pasterkamp, R., and Van Damme, P. (2022). C21orf2 mutations
1320 found in ALS disrupt primary cilia function. bioRxiv, 2022.02.28.482239.
1321 https://doi.org/10.1101/2022.02.28.482239.
1322 55. van Rheenen, W., Shatunov, A., Dekker, A.M., McLaughlin, R.L., Diekstra, F.P., Pulit,
1323 S.L., van der Spek, R.A.A., Võsa, U., de Jong, S., Robinson, M.R., et al. (2016).
1324 Genome-wide association analyses identify new risk variants and the genetic
1325 architecture of amyotrophic lateral sclerosis. Nat. Genet. *48*, 1043–1048.
1326 56. Ishikawa, H., and Marshall, W.F. (2017). Intraflagellar transport and ciliary dynamics.
1327 Cold Spring Harb. Perspect. Biol. *9*, a021998.
1328 57. Lin, Y.-L., Lin, Y.-W., Nhieu, J., Zhang, X., and Wei, L.-N. (2020). Sonic hedgehog-Gli1

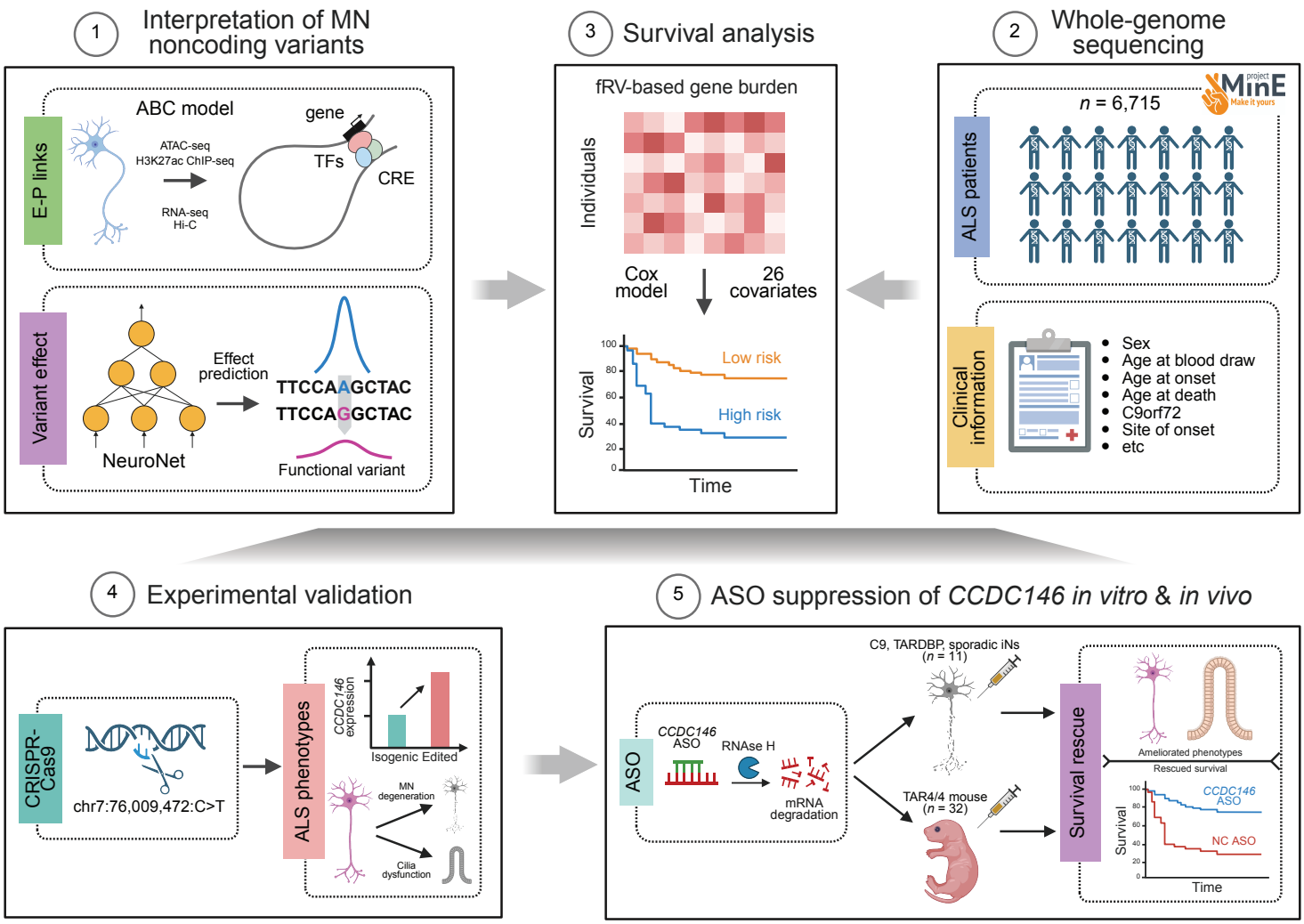
signaling and cellular retinoic acid binding protein 1 gene regulation in motor neuron differentiation and diseases. Int. J. Mol. Sci. *21*, 4125.

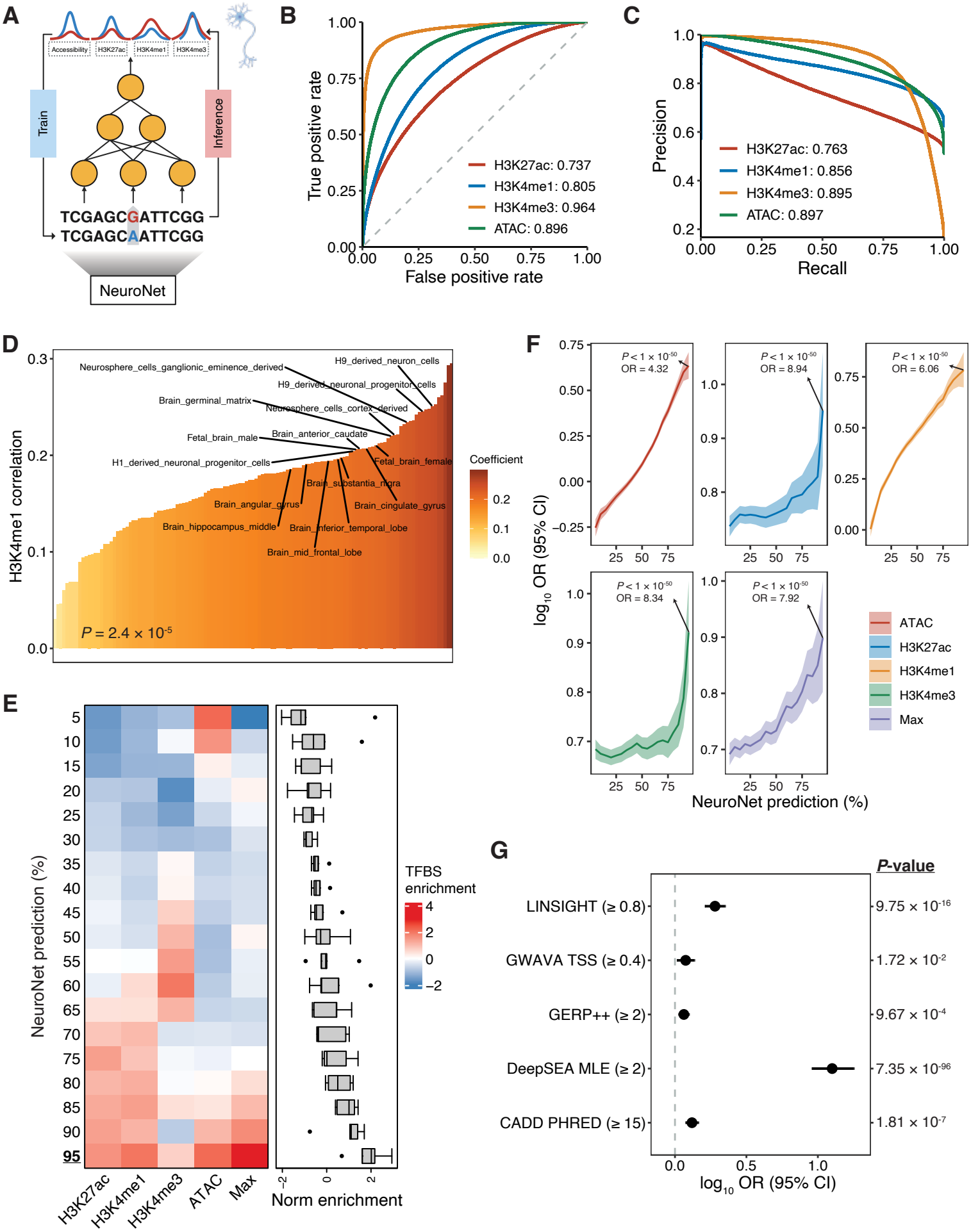
58. Brown, A.-L., Wilkins, O.G., Keuss, M.J., Hill, S.E., Zanovello, M., Lee, W.C., Bampton, A., Lee, F.C.Y., Masino, L., Qi, Y.A., et al. (2022). TDP-43 loss and ALS-risk SNPs drive mis-splicing and depletion of UNC13A. Nature *603*, 131–137.
59. Ma, X.R., Prudencio, M., Koike, Y., Vatsavayai, S.C., Kim, G., Harbinski, F., Briner, A., Rodriguez, C.M., Guo, C., Akiyama, T., et al. (2022). TDP-43 represses cryptic exon inclusion in the FTD–ALS gene UNC13A. Nature *603*, 124–130.
60. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443.
61. Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. Science *347*, 1260419.
62. Yadav, A., Matson, K.J.E., Li, L., Hua, I., Petrescu, J., Kang, K., Alkaslasi, M.R., Lee, D.I., Hasan, S., Galuta, A., et al. (2023). A cellular taxonomy of the adult human spinal cord. Neuron *111*, 328–344.e7.
63. Humphrey, J., Venkatesh, S., Hasan, R., Herb, J.T., de Paiva Lopes, K., Küçükali, F., Byrska-Bishop, M., Evani, U.S., Narzisi, G., Fagegaltier, D., et al. (2023). Integrative transcriptomic analysis of the amyotrophic lateral sclerosis spinal cord implicates glial activation and suggests new risk genes. Nat. Neurosci. *26*, 150–162.
64. Nguyen, D.-T., Mathias, S., Bologa, C., Brunak, S., Fernandez, N., Gaulton, A., Hersey, A., Holmes, J., Jensen, L.J., Karlsson, A., et al. (2017). Pharos: Collating protein information to shed light on the druggable genome. Nucleic Acids Res. *45*, D995–D1002.
65. Kelleher, K.J., Sheils, T.K., Mathias, S.L., Yang, J.J., Metzger, V.T., Siramshetty, V.B., Nguyen, D.-T., Jensen, L.J., Vidović, D., Schürer, S.C., et al. (2023). Pharos 2023: an integrated resource for the understudied human proteome. Nucleic Acids Res. *51*, D1405–D1416.
66. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589.
67. Krivák, R., and Hoksza, D. (2018). P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. J. Cheminform. *10*, 39.
68. Cedarbaum, J.M., Stambler, N., Malta, E., Fuller, C., Hilt, D., Thurmond, B., and Nakanishi, A. (1999). The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. J. Neurol. Sci. *169*, 13–21.

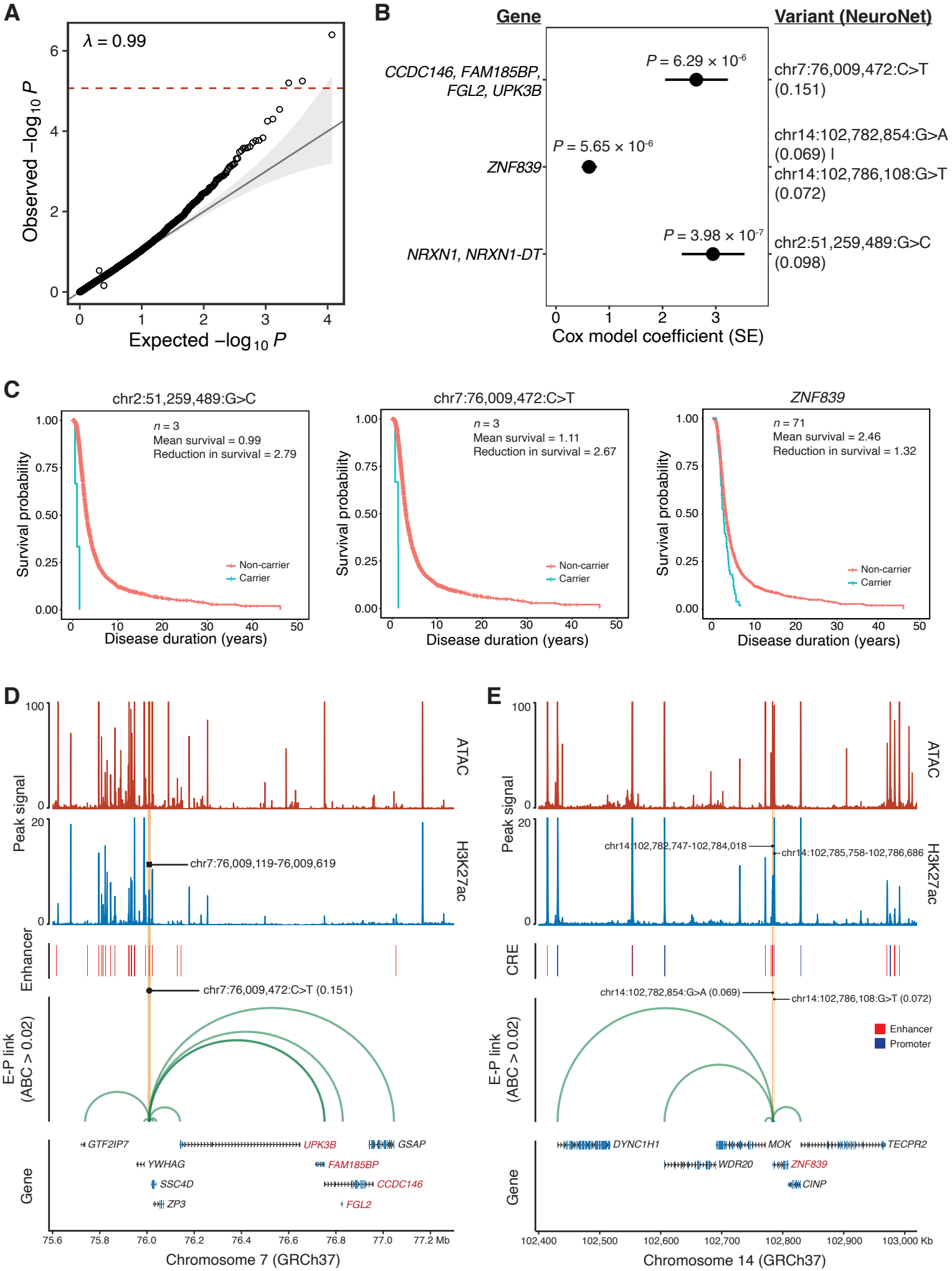
1366 69. Proudfoot, M., Jones, A., Talbot, K., Al-Chalabi, A., and Turner, M.R. (2016). The
1367 ALSFRS as an outcome measure in therapeutic trials and its relationship to symptom
1368 onset. Amyotroph. Lateral Scler. Frontotemporal Degener. *17*, 414–425.
1369 70. Bruse, N., and van Heeringen, S.J. (2018). GimmeMotifs: an analysis framework for
1370 transcription factor motif analysis. bioRxiv, 474403. https://doi.org/10.1101/474403.
1371 71. Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable,
1372 A.L., Fang, T., Doncheva, N.T., Pyysalo, S., et al. (2023). The STRING database in
1373 2023: protein-protein association networks and functional enrichment analyses for any
1374 sequenced genome of interest. Nucleic Acids Res. *51*, D638–D646.
1375 72. Krishnan, A., Zhang, R., Yao, V., Theesfeld, C.L., Wong, A.K., Tadych, A., Volfovsky,
1376 N., Packer, A., Lash, A., and Troyanskaya, O.G. (2016). Genome-wide prediction and
1377 functional characterization of the genetic basis of autism spectrum disorder. Nat.
1378 Neurosci. *19*, 1454–1462.
1379 73. Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of
1380 communities in large networks. J. Stat. Mech. *2008*, P10008.
1381 74. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan,
1382 L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics
1383 data. Innovations *2*. https://doi.org/10.1016/j.xinn.2021.100141.
1384 75. Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and
1385 Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment
1386 analysis tool. BMC Bioinformatics *14*, 1–14.
1387 76. Project MinE ALS Sequencing Consortium (2018). Project MinE: study design and pilot
1388 analyses of a large-scale whole-genome sequencing study in amyotrophic lateral
1389 sclerosis. Eur. J. Hum. Genet. *26*, 1537–1546.
1390 77. Brooks, B.R. (1994). El Escorial World Federation of Neurology criteria for the diagnosis
1391 of amyotrophic lateral sclerosis. Subcommittee on Motor Neuron Diseases/Amyotrophic
1392 Lateral Sclerosis of the World Federation of Neurology Research Group on
1393 Neuromuscular Diseases and the El Escorial "Clinical limits of amyotrophic lateral
1394 sclerosis" workshop contributors. J. Neurol. Sci. *124 Suppl*, 96–107.
1395 78. Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic
1396 RNA-seq quantification. Nat. Biotechnol. *34*, 525–527.
1397 79. Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential
1398 expression analysis of RNA-seq data. Genome Biol. *11*, R25.
1399 80. Ludwik, K.A., Telugu, N., Schommer, S., Stachelscheid, H., and Diecke, S. (2023).
1400 ASSURED-optimized CRISPR protocol for knockout/SNP knockin in hiPSCs. STAR
1401 Protoc *4*, 102406.
1402 81. Varcianna, A., Myszczynska, M.A., Castelli, L.M., O'Neill, B., Kim, Y., Talbot, J., Nyberg,

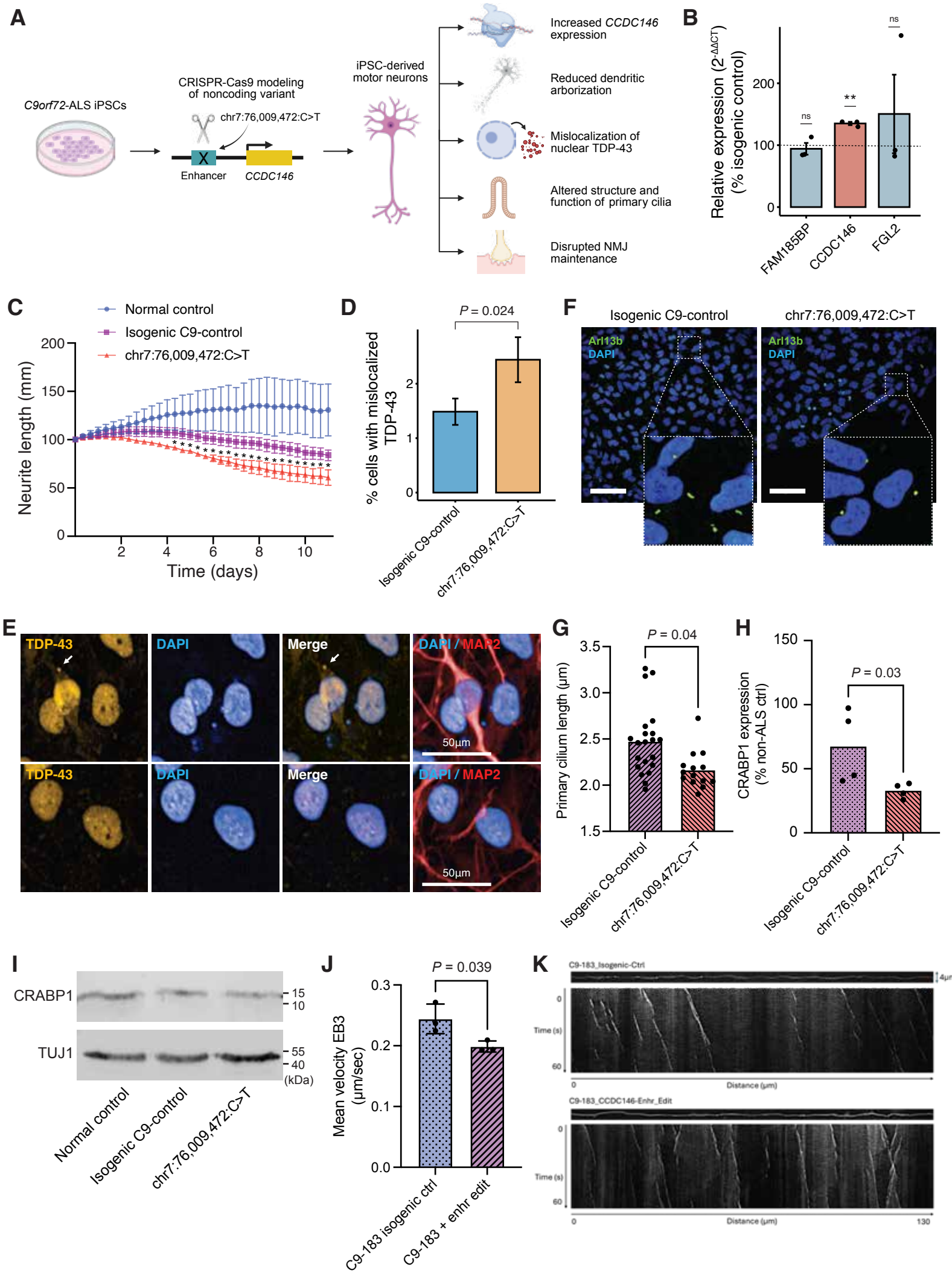
S., Nyamali, I., Heath, P.R., Stopford, M.J., et al. (2019). Micro-RNAs secreted through astrocyte-derived extracellular vesicles cause neuronal network degeneration in C9orf72 ALS. Preprint, https://doi.org/10.1016/j.ebiom.2018.11.067 https://doi.org/10.1016/j.ebiom.2018.11.067.

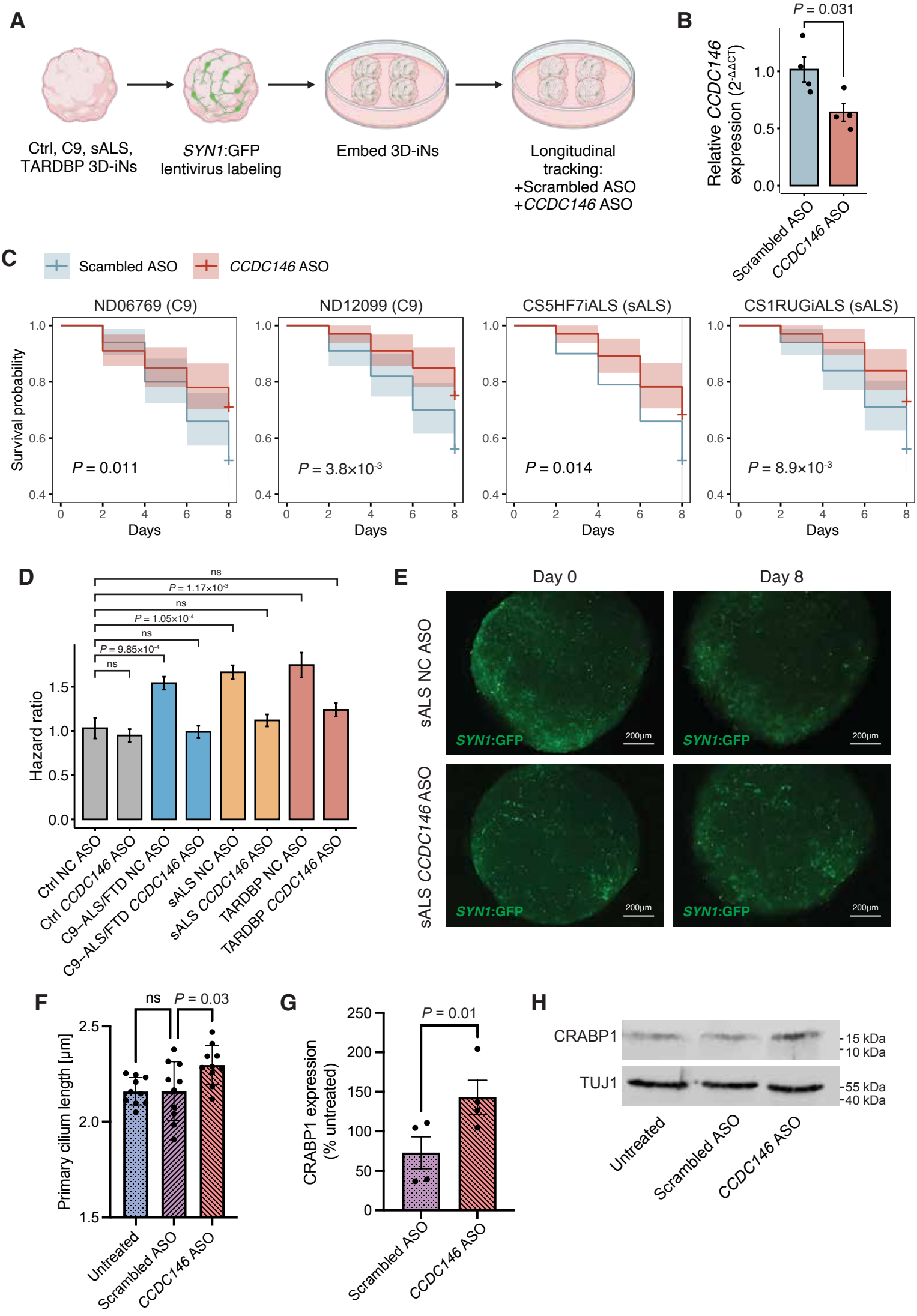
82. Castelli, L.M., Cutillo, L., Souza, C.D.S., Sanchez-Martinez, A., Granata, I., Lin, Y.-H., Myszczynska, M.A., Heath, P.R., Livesey, M.R., Ning, K., et al. (2021). SRSF1-dependent inhibition of C9ORF72-repeat RNA nuclear export: genome-wide mechanisms for neuroprotection in amyotrophic lateral sclerosis. Mol. Neurodegener. *16*, 53.
83. Du, Z.-W., Chen, H., Liu, H., Lu, J., Qian, K., Huang, C.-L., Zhong, X., Fan, F., and Zhang, S.-C. (2015). Generation and expansion of highly pure motor neuron progenitors from human pluripotent stem cells. Nat. Commun. *6*, 6626.
84. Artyukhov, A.S., Dashinimaev, E.B., Tsvetkov, V.O., Bolshakov, A.P., Konovalova, E.V., Kolbaev, S.N., Vorotelyak, E.A., and Vasiliev, A.V. (2017). New genes for accurate normalization of qRT-PCR results in study of iPS and iPS-derived cells. Gene *626*, 234–240.
85. Tinevez, J.-Y., Perry, N., Schindelin, J., Hoopes, G.M., Reynolds, G.D., Laplantine, E., Bednarek, S.Y., Shorte, S.L., and Eliceiri, K.W. (2017). TrackMate: An open and extensible platform for single-particle tracking. Methods *115*, 80–90.
86. Zdrazil, B., Felix, E., Hunter, F., Manners, E.J., Blackshaw, J., Corbett, S., de Veij, M., Ioannidis, H., Lopez, D.M., Mosquera, J.F., et al. (2024). The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. Nucleic Acids Res. *52*, D1180–D1192.
87. Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B.A., Thiessen, P.A., Yu, B., et al. (2023). PubChem 2023 update. Nucleic Acids Res. *51*, D1373–D1380.
88. UniProt Consortium (2023). UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. *51*, D523–D531.
89. Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., et al. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. *50*, D439–D444.

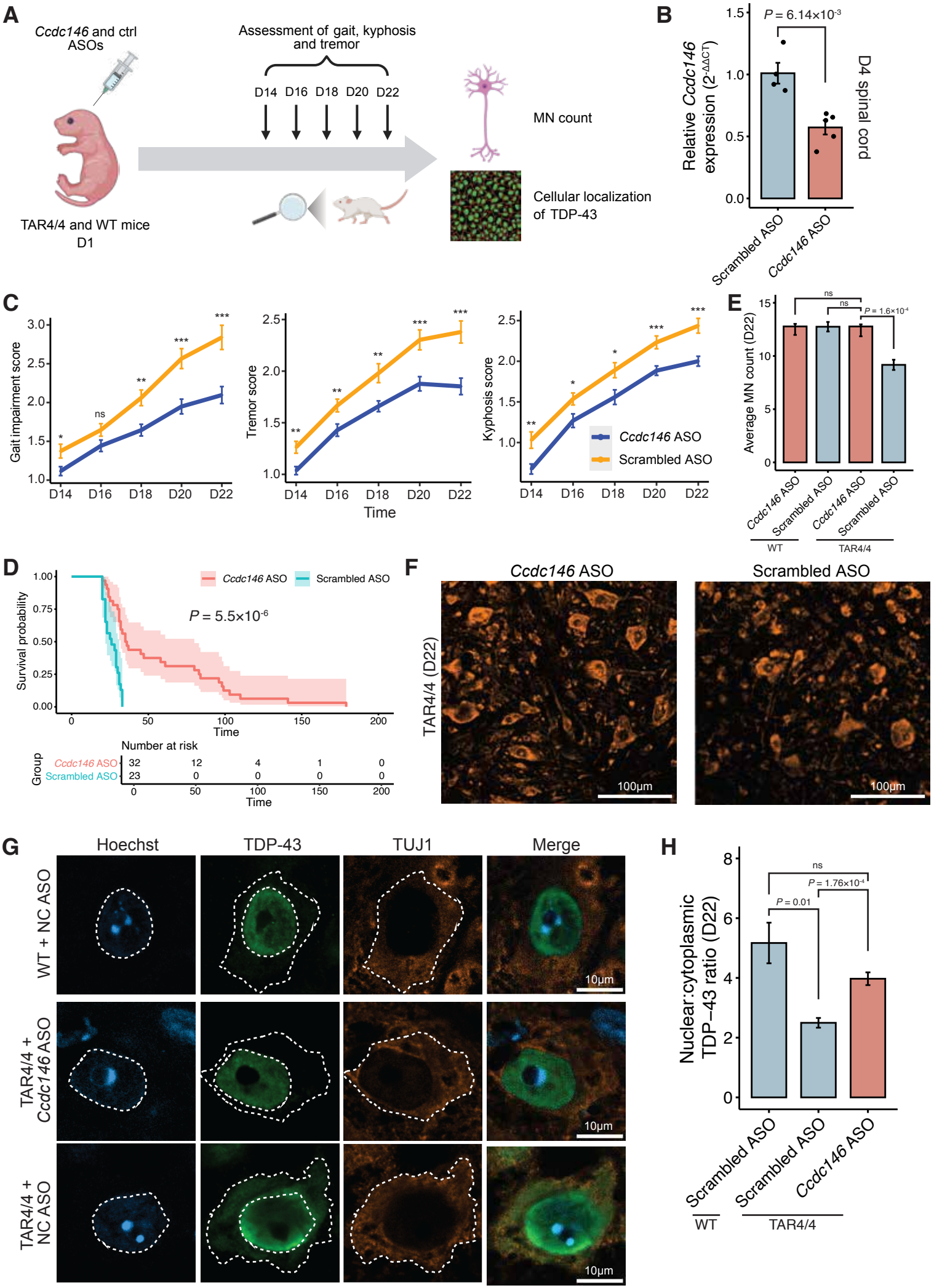

1 Interpretation of MN noncoding variants
E-P links
ABC model
ATAC-seq
H3K27ac ChIP-seq
RNA-seq
Hi-C
gene
TFs
CRE
Variant effect
NeuroNet
Effect prediction
TTCCAAGCTAC
TTCCAGGCTAC
Functional variant
3 Survival analysis
fRV-based gene burden
Individuals
Cox model
26 covariates
Survival
Low risk
High risk
Time
2 Whole-genome sequencing
n = 6,715
ALS patients
Clinical information
Sex
Age at blood draw
Age at onset
Age at death
C9orf72
Site of onset
etc
4 Experimental validation
CRISPR-Cas9
chr7:76,009,472:C>T
ALS phenotypes
CCDC146 expression
Isogenic Edited
MN degeneration
Cilia dysfunction
5 ASO suppression of CCDC146 in vitro & in vivo
ASO
CCDC146 ASO
RNAse H
mRNA degradation
C9, TARDBP, sporadic iNs (n = 11)
TAR4/4 mouse (n = 32)
Survival rescue
Ameliorated phenotypes
Rescued survival
CCDC146 ASO
NC ASO
Survival
Time

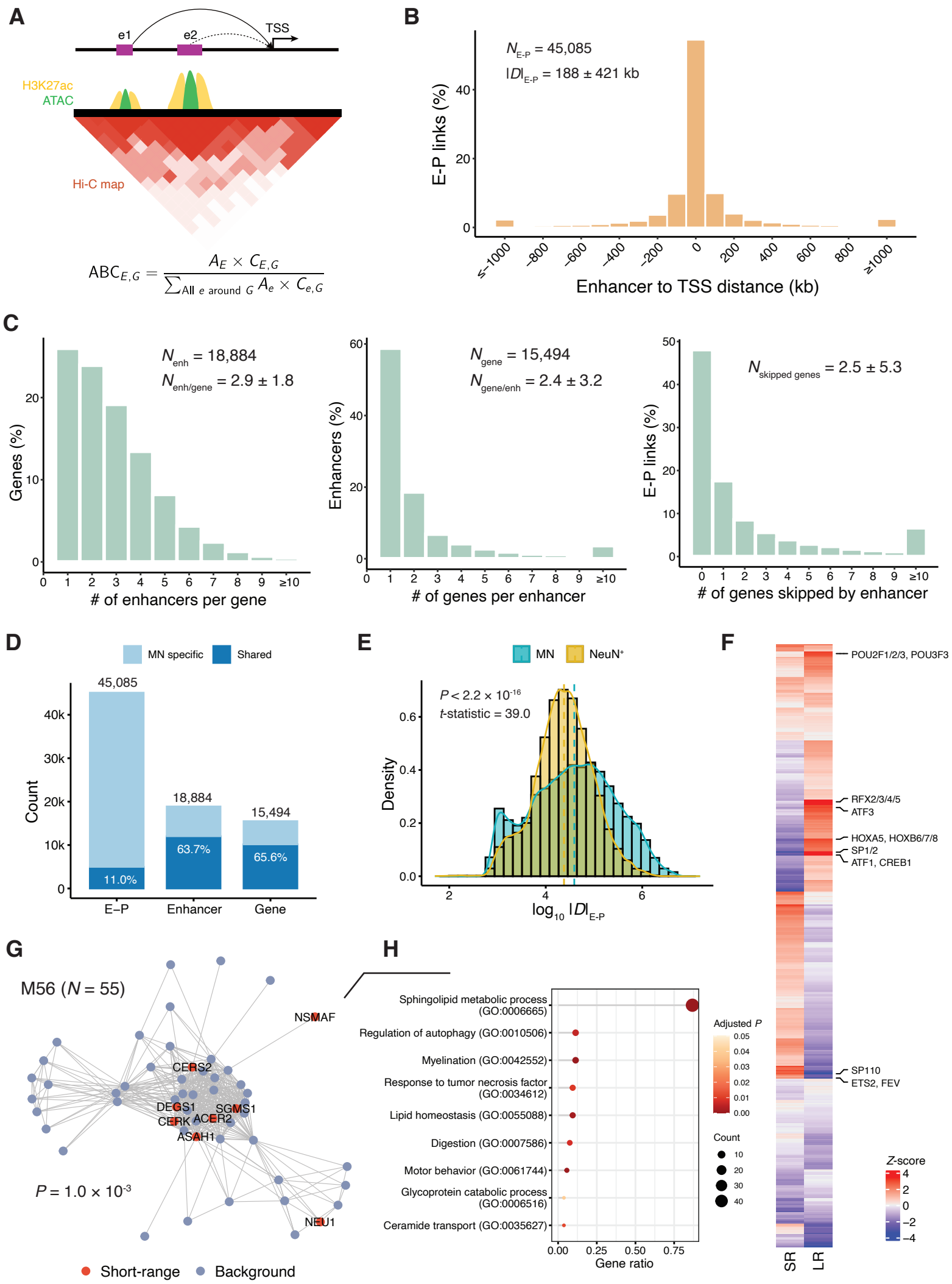

A
Accessibility
H3K27ac
H3K4me1
H3K4me3
Train
Inference
TCGAGCGATTCGG
TCGAGCAATTCGG
NeuroNet
B
True positive rate
False positive rate
H3K27ac: 0.737
H3K4me1: 0.805
H3K4me3: 0.964
ATAC: 0.896
C
Precision
Recall
H3K27ac: 0.763
H3K4me1: 0.856
H3K4me3: 0.895
ATAC: 0.897
D
H3K4me1 correlation
Neurosphere_cells_ganglionic_eminence_derived
H9_derived_neuron_cells
H9_derived_neuronal_progenitor_cells
Brain_germinal_matrix
Neurosphere_cells_cortex_derived
Fetal_brain_male
Brain_anterior_caudate
H1_derived_neuronal_progenitor_cells
Fetal_brain_female
Brain_substantia_nigra
Brain_angular_gyrus
Brain_cingulate_gyrus
Brain_hippocampus_middle
Brain_inferior_temporal_lobe
Brain_mid_frontal_lobe
Coefficient
P = 2.4 × 10-5
E
NeuroNet prediction (%)
H3K27ac
H3K4me1
H3K4me3
ATAC
Max
Norm enrichment
TFBS enrichment
F
log10 OR (95% CI)
P < 1 × 10-50
OR = 4.32
OR = 8.94
OR = 6.06
OR = 8.34
OR = 7.92
ATAC
H3K27ac
H3K4me1
H3K4me3
Max
NeuroNet prediction (%)
G
P-value
LINSIGHT (≥ 0.8)
GWAVA TSS (≥ 0.4)
GERP++ (≥ 2)
DeepSEA MLE (≥ 2)
CADD PHRED (≥ 15)
9.75 × 10-16
1.72 × 10-2
9.67 × 10-4
7.35 × 10-96
1.81 × 10-7
log10 OR (95% CI)

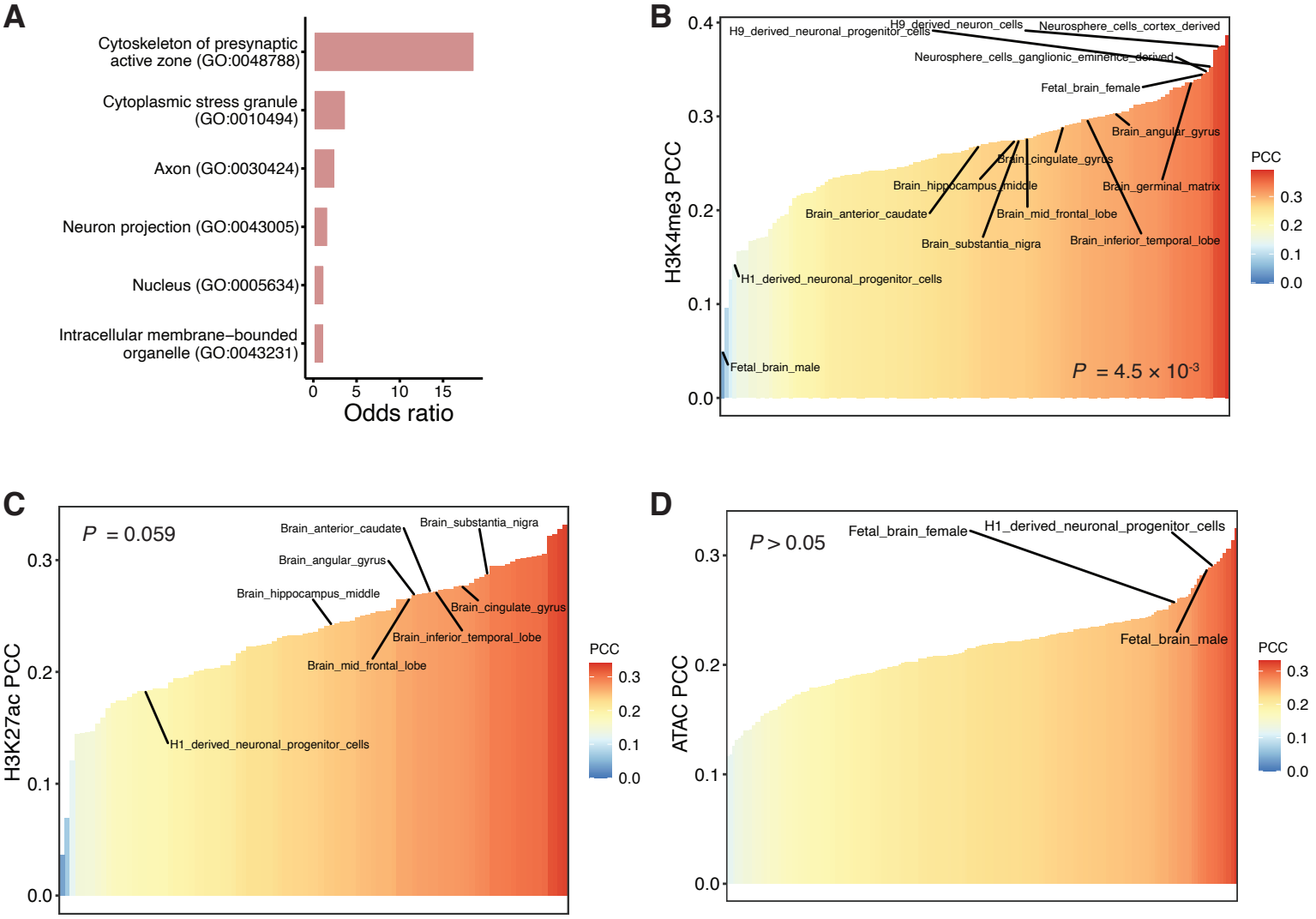

A
λ = 0.99
Observed −log10 P
Expected −log10 P
B
Gene
Variant (NeuroNet)
CCDC146, FAM185BP, FGL2, UPK3B
P = 6.29 × 10-6
chr7:76,009,472:C>T (0.151)
ZNF839
P = 5.65 × 10-6
chr14:102,782,854:G>A (0.069) |
chr14:102,786,108:G>T (0.072)
NRXN1, NRXN1-DT
P = 3.98 × 10-7
chr2:51,259,489:G>C (0.098)
Cox model coefficient (SE)
C
chr2:51,259,489:G>C
n = 3
Mean survival = 0.99
Reduction in survival = 2.79
chr7:76,009,472:C>T
n = 3
Mean survival = 1.11
Reduction in survival = 2.67
ZNF839
n = 71
Mean survival = 2.46
Reduction in survival = 1.32
Survival probability
Non-carrier
Carrier
Disease duration (years)
D
Peak signal
ATAC
H3K27ac
chr7:76,009,119-76,009,619
Enhancer
chr7:76,009,472:C>T (0.151)
E-P link (ABC > 0.02)
Gene
GTF2IP7
UPK3B
GSAP
YWHAG
FAM185BP
SSC4D
CCDC146
ZP3
FGL2
Chromosome 7 (GRCh37)
E
chr14:102,782,747-102,784,018
chr14:102,785,758-102,786,686
CRE
chr14:102,782,854:G>A (0.069)
chr14:102,786,108:G>T (0.072)
Enhancer
Promoter
DYNC1H1
MOK
TECPR2
WDR20
ZNF839
CINP
Chromosome 14 (GRCh37)

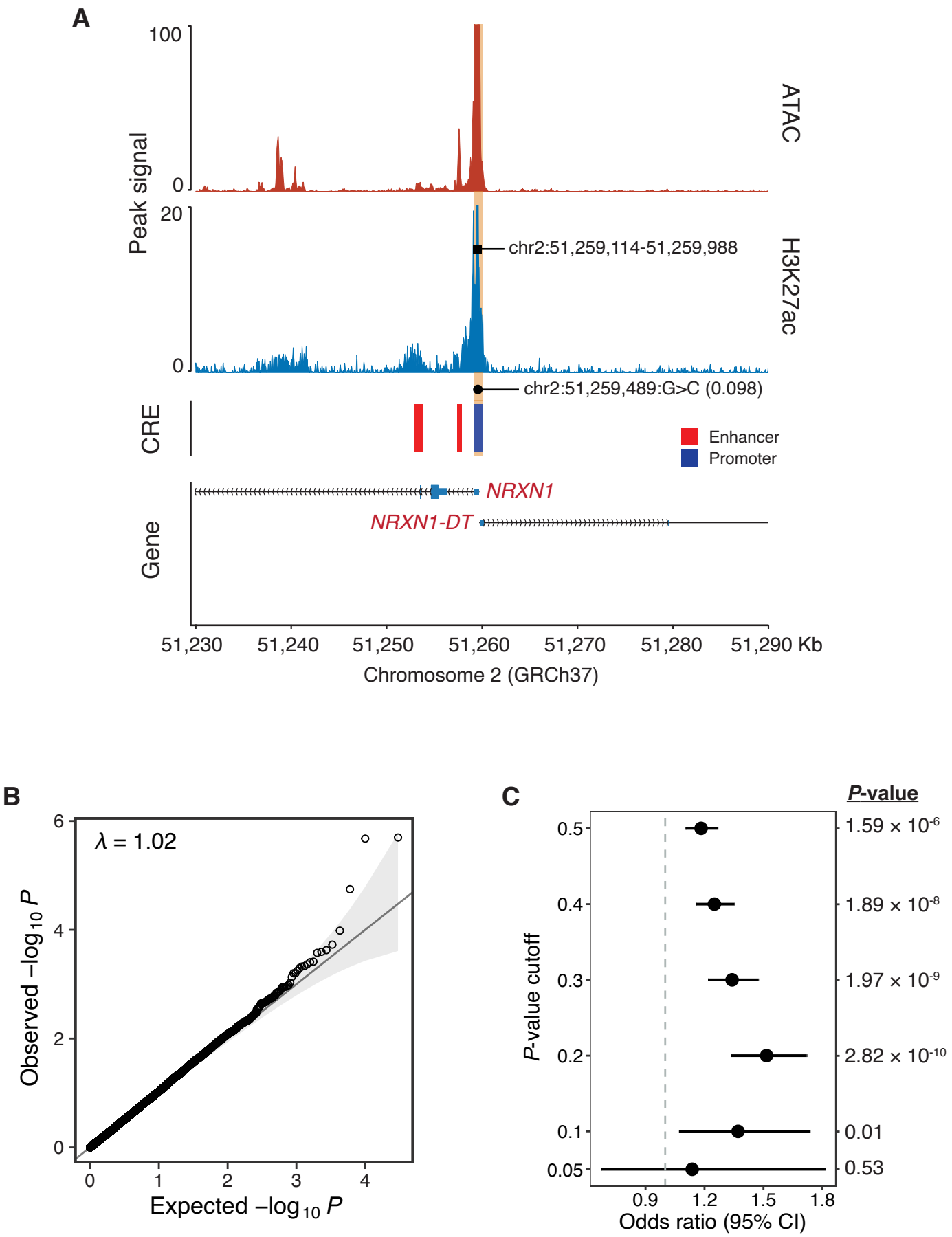

A
C9orf72-ALS iPSCs
CRISPR-Cas9 modeling of noncoding variant
chr7:76,009,472:C>T
Enhancer
CCDC146
iPSC-derived motor neurons
Increased CCDC146 expression
Reduced dendritic arborization
Mislocalization of nuclear TDP-43
Altered structure and function of primary cilia
Disrupted NMJ maintenance
B
Relative expression (2-ΔΔCT) (% isogenic control)
ns
**
ns
FAM185BP
CCDC146
FGL2
C
Normal control
Isogenic C9-control
chr7:76,009,472:C>T
Neurite length (mm)
Time (days)
D
P = 0.024
% cells with mislocalized TDP-43
Isogenic C9-control
chr7:76,009,472:C>T
F
Isogenic C9-control
chr7:76,009,472:C>T
Arl13b
DAPI
E
TDP-43
DAPI
Merge
DAPI / MAP2
50µm
G
P = 0.04
Primary cilium length (µm)
Isogenic C9-control
chr7:76,009,472:C>T
H
P = 0.03
CRABP1 expression (% non-ALS ctrl)
Isogenic C9-control
chr7:76,009,472:C>T
I
CRABP1
TUJ1
15
10
55
40
(kDa)
Normal control
Isogenic C9-control
chr7:76,009,472:C>T
J
P = 0.039
Mean velocity EB3 (µm/sec)
C9-183 isogenic ctrl
C9-183 + enhr edit
K
C9-183_Isogenic-Ctrl
Time (s)
Distance (µm)
C9-183_CCDC146-Enhr_Edit
Time (s)
Distance (µm)

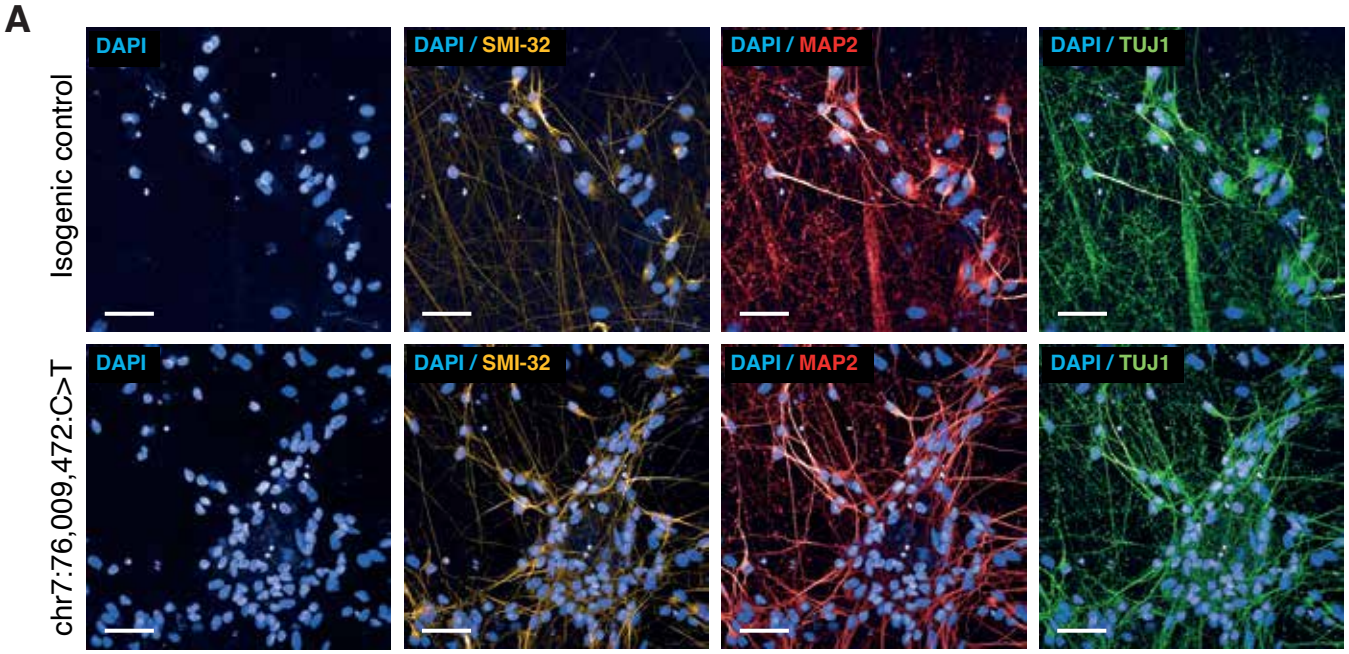

A
Ctrl, C9, sALS, TARDBP 3D-iNs
SYN1:GFP lentivirus labeling
Embed 3D-iNs
Longitudinal tracking: +Scrambled ASO +CCDC146 ASO
B
P = 0.031
Relative CCDC146 expression (2-ΔΔCT)
Scrambled ASO
CCDC146 ASO
C
Scambled ASO
CCDC146 ASO
ND06769 (C9)
ND12099 (C9)
CS5HF7iALS (sALS)
CS1RUGiALS (sALS)
Survival probability
Days
P = 0.011
P = 3.8×10-3
P = 0.014
P = 8.9×10-3
D
Hazard ratio
ns
P = 1.17×10-3
P = 1.05×10-4
P = 9.85×10-4
Ctrl NC ASO
Ctrl CCDC146 ASO
C9–ALS/FTD NC ASO
C9–ALS/FTD CCDC146 ASO
sALS NC ASO
sALS CCDC146 ASO
TARDBP NC ASO
TARDBP CCDC146 ASO
E
Day 0
Day 8
sALS NC ASO
sALS CCDC146 ASO
SYN1:GFP
200µm
F
Primary cilium length [µm]
ns
P = 0.03
Untreated
Scrambled ASO
CCDC146 ASO
G
CRABP1 expression (% untreated)
P = 0.01
Scrambled ASO
CCDC146 ASO
H
CRABP1
TUJ1
15 kDa
10 kDa
55 kDa
40 kDa
Untreated
Scrambled ASO
CCDC146 ASO

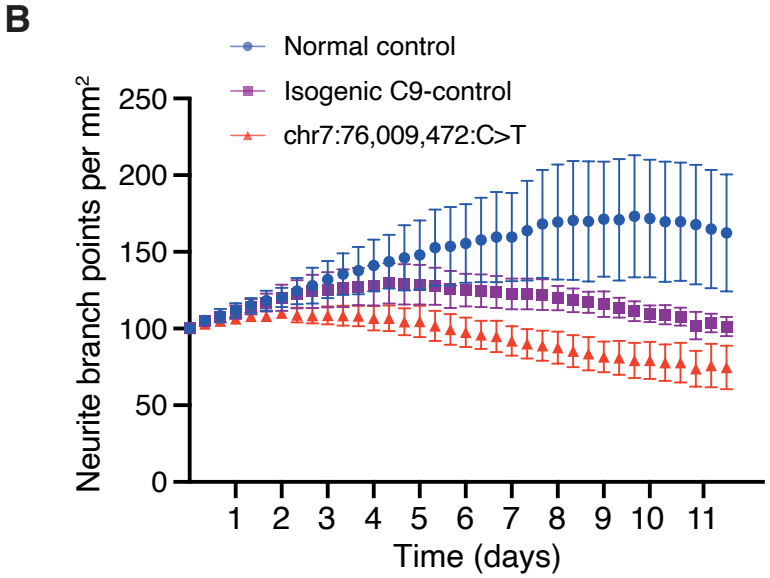

A
Ccdc146 and ctrl ASOs
TAR4/4 and WT mice
D1
Assessment of gait, kyphosis and tremor
D14 D16 D18 D20 D22
MN count
Cellular localization of TDP-43
B
P = 6.14×10-3
Relative Ccdc146 expression (2-ΔΔCT)
D4 spinal cord
Scrambled ASO
Ccdc146 ASO
C
Gait impairment score
Tremor score
Kyphosis score
Time
Ccdc146 ASO
Scrambled ASO
E
Average MN count (D22)
ns
P = 1.6×10-4
WT
TAR4/4
D
Survival probability
P = 5.5×10-6
Number at risk
Group
F
Ccdc146 ASO
Scrambled ASO
TAR4/4 (D22)
100μm
G
Hoechst
TDP-43
TUJ1
Merge
WT + NC ASO
TAR4/4 + Ccdc146 ASO
TAR4/4 + NC ASO
10μm
H
Nuclear:cytoplasmic TDP-43 ratio (D22)
P = 0.01
P = 1.76×10-4

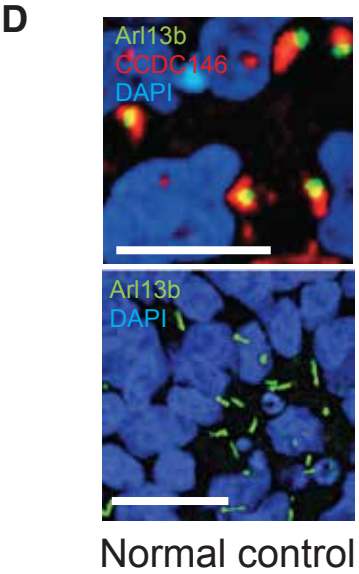

A
e1
e2
TSS
H3K27ac
ATAC
Hi-C map
$ABC_{E,G} = \frac{A_E \times C_{E,G}}{\sum_{\text{All } e \text{ around } G} A_e \times C_{e,G}}$
B
$N_{E\text{-}P}$ = 45,085
$|D|_{E\text{-}P}$ = 188 ± 421 kb
E-P links (%)
Enhancer to TSS distance (kb)
C
N_{enh} = 18,884
$N_{enh/gene}$ = 2.9 ± 1.8
Genes (%)
# of enhancers per gene
N_{gene} = 15,494
$N_{gene/enh}$ = 2.4 ± 3.2
Enhancers (%)
# of genes per enhancer
$N_{skipped\ genes}$ = 2.5 ± 5.3
E-P links (%)
# of genes skipped by enhancer
D
MN specific
Shared
45,085
18,884
15,494
11.0%
63.7%
65.6%
Count
E–P
Enhancer
Gene
E
MN
NeuN+
$P < 2.2 \times 10^{-16}$
t-statistic = 39.0
Density
$\log_{10} |D|_{E\text{-}P}$
F
POU2F1/2/3, POU3F3
RFX2/3/4/5
ATF3
HOXA5, HOXB6/7/8
SP1/2
ATF1, CREB1
SP110
ETS2, FEV
Z-score
SR
LR
G
M56 (N = 55)
NSMAF
CERS2
DEGS1
SGMS1
CERK
ACER2
ASAH1
NEU1
$P = 1.0 \times 10^{-3}$
Short-range
Background
H
Sphingolipid metabolic process (GO:0006665)
Regulation of autophagy (GO:0010506)
Myelination (GO:0042552)
Response to tumor necrosis factor (GO:0034612)
Lipid homeostasis (GO:0055088)
Digestion (GO:0007586)
Motor behavior (GO:0061744)
Glycoprotein catabolic process (GO:0006516)
Ceramide transport (GO:0035627)
Gene ratio
Adjusted P
Count

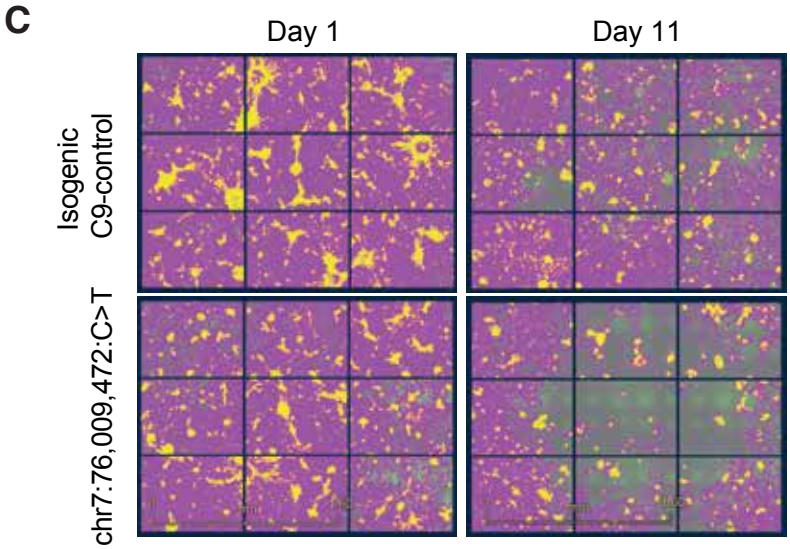

A
Cytoskeleton of presynaptic active zone (GO:0048788)
Cytoplasmic stress granule (GO:0010494)
Axon (GO:0030424)
Neuron projection (GO:0043005)
Nucleus (GO:0005634)
Intracellular membrane−bounded organelle (GO:0043231)
0
5
10
15
Odds ratio
B
H3K4me3 PCC
0.4
0.3
0.2
0.1
0.0
H9_derived_neuronal_progenitor_cells
H9_derived_neuron_cells
Neurosphere_cells_cortex_derived
Neurosphere_cells_ganglionic_eminence_derived
Fetal_brain_female
Brain_angular_gyrus
Brain_cingulate_gyrus
Brain_hippocampus_middle
Brain_germinal_matrix
Brain_anterior_caudate
Brain_mid_frontal_lobe
Brain_substantia_nigra
Brain_inferior_temporal_lobe
H1_derived_neuronal_progenitor_cells
Fetal_brain_male
P = 4.5 × 10-3
PCC
0.3
0.2
0.1
0.0
C
P = 0.059
H3K27ac PCC
0.3
0.2
0.1
0.0
Brain_anterior_caudate
Brain_substantia_nigra
Brain_angular_gyrus
Brain_hippocampus_middle
Brain_cingulate_gyrus
Brain_inferior_temporal_lobe
Brain_mid_frontal_lobe
H1_derived_neuronal_progenitor_cells
PCC
0.3
0.2
0.1
0.0
D
P > 0.05
ATAC PCC
0.3
0.2
0.1
0.0
Fetal_brain_female
H1_derived_neuronal_progenitor_cells
Fetal_brain_male
PCC
0.3
0.2
0.1
0.0

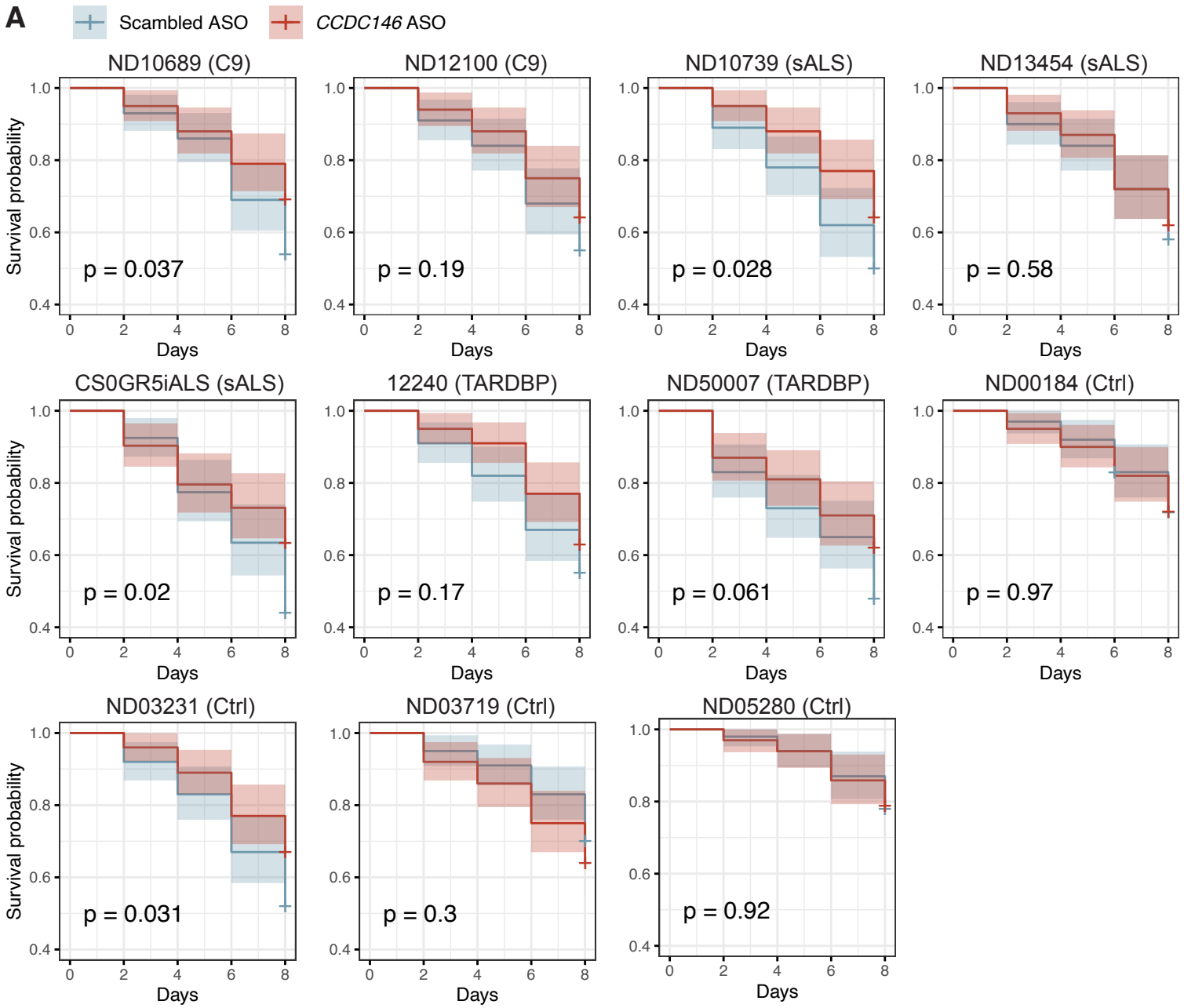

A
100
0
Peak signal
ATAC
20
0
chr2:51,259,114-51,259,988
H3K27ac
chr2:51,259,489:G>C (0.098)
CRE
Enhancer
Promoter
NRXN1
NRXN1-DT
Gene
51,230
51,240
51,250
51,260
51,270
51,280
51,290 Kb
Chromosome 2 (GRCh37)
B
λ = 1.02
Observed −log10 P
Expected −log10 P
C
P-value
P-value cutoff
0.5
0.4
0.3
0.2
0.1
0.05
1.59 × 10-6
1.89 × 10-8
1.97 × 10-9
2.82 × 10-10
0.01
0.53
0.9
1.2
1.5
1.8
Odds ratio (95% CI)

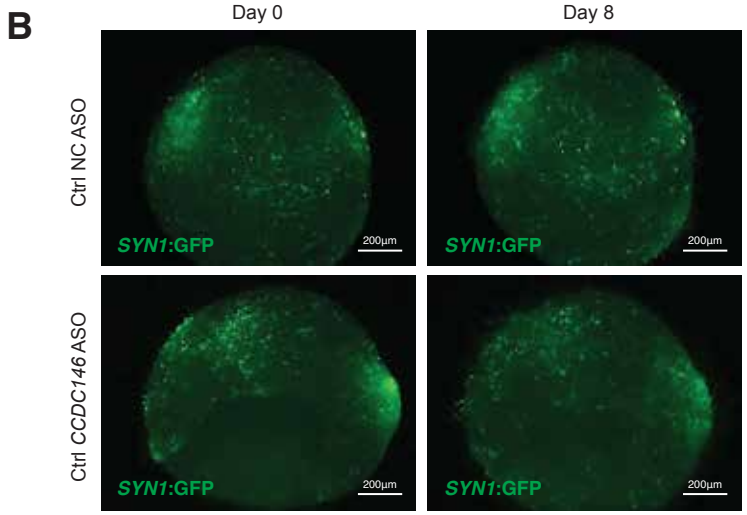

A
Isogenic control
chr7:76,009,472:C>T
DAPI
DAPI / SMI-32
DAPI / MAP2
DAPI / TUJ1

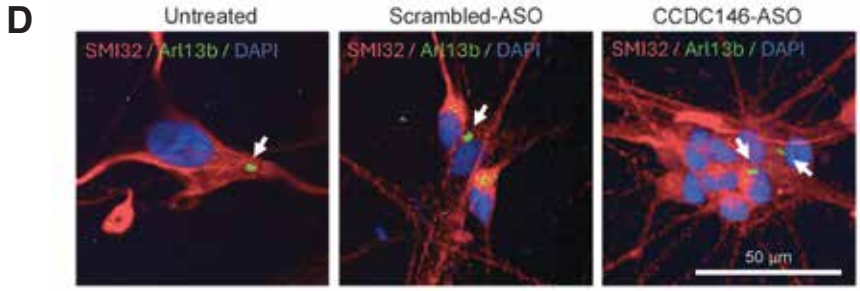

B
Normal control
Isogenic C9-control
chr7:76,009,472:C>T
Neurite branch points per mm²
250
200
150
100
50
0
1 2 3 4 5 6 7 8 9 10 11
Time (days)

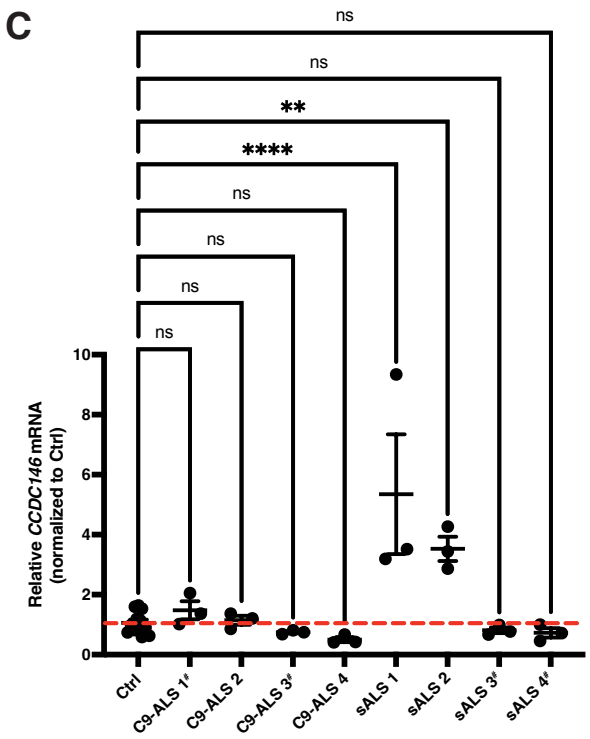

D
Arl13b
CCDC146
DAPI
Arl13b
DAPI
Normal control

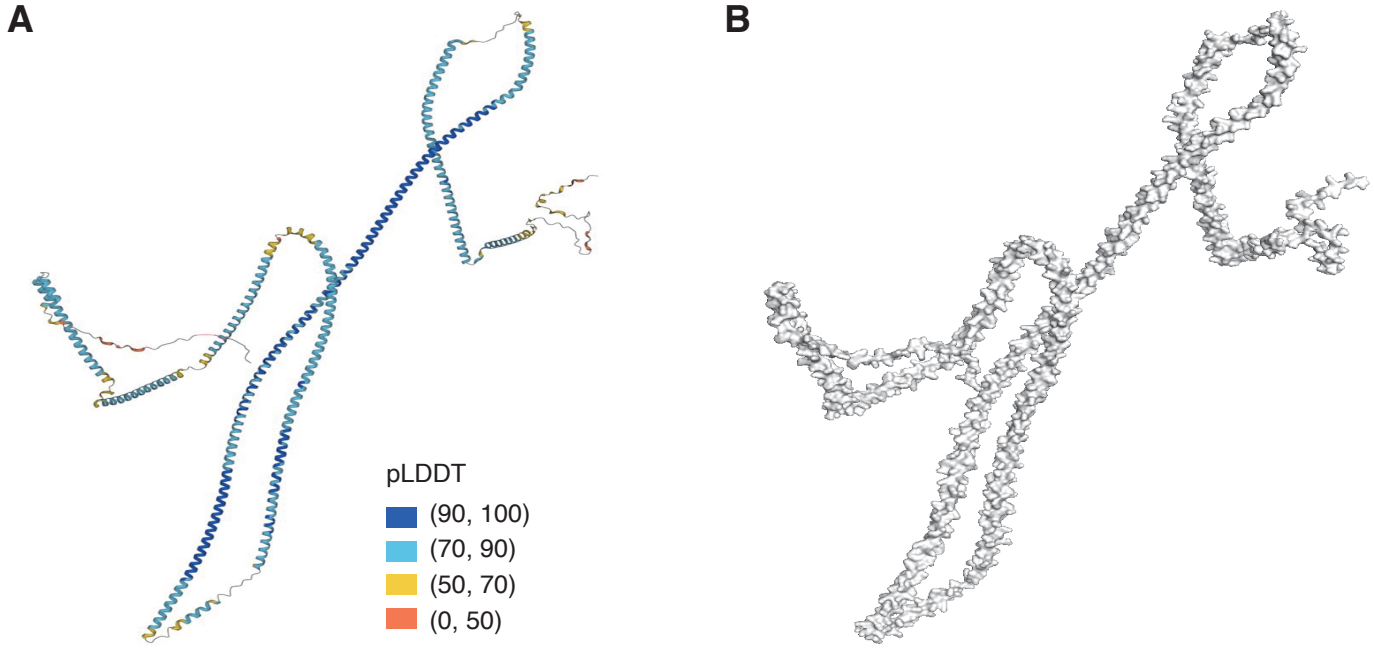

C
Day 1
Day 11
Isogenic C9-control
chr7:76,009,472:C>T

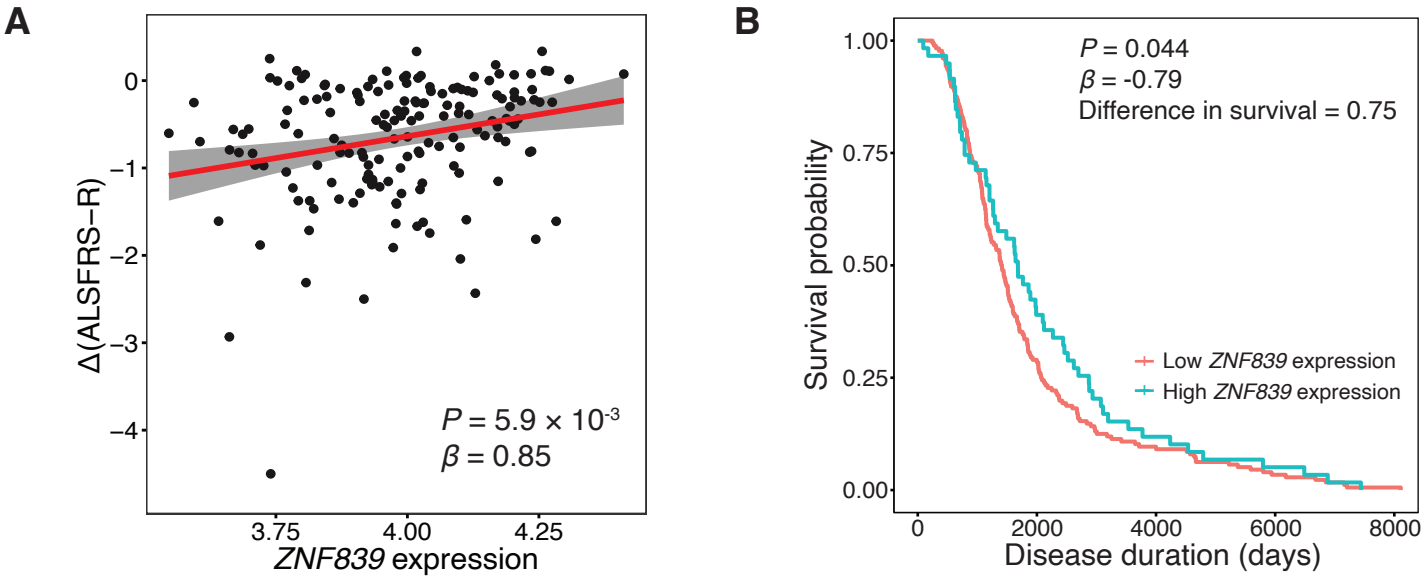

A
Scambled ASO
CCDC146 ASO
ND10689 (C9)
p = 0.037
ND12100 (C9)
p = 0.19
ND10739 (sALS)
p = 0.028
ND13454 (sALS)
p = 0.58
CS0GR5iALS (sALS)
p = 0.02
12240 (TARDBP)
p = 0.17
ND50007 (TARDBP)
p = 0.061
ND00184 (Ctrl)
p = 0.97
ND03231 (Ctrl)
p = 0.031
ND03719 (Ctrl)
p = 0.3
ND05280 (Ctrl)
p = 0.92
Survival probability
Days


B
Day 0
Day 8
Ctrl NC ASO
Ctrl CCDC146 ASO
SYN1:GFP
200µm


D
Untreated
Scrambled-ASO
CCDC146-ASO
SMI32 / Arl13b / DAPI
50 µm


C
ns
**
****
Relative CCDC146 mRNA (normalized to Ctrl)
Ctrl
C9-ALS 1*
C9-ALS 2
C9-ALS 3*
C9-ALS 4
sALS 1
sALS 2
sALS 3*
sALS 4*


A
B
pLDDT
(90, 100)
(70, 90)
(50, 70)
(0, 50)


A
0
−1
−2
−3
−4
Δ(ALSFRS−R)
P = 5.9 × 10-3
β = 0.85
3.75
4.00
4.25
ZNF839 expression
B
1.00
0.75
0.50
0.25
0.00
Survival probability
P = 0.044
β = -0.79
Difference in survival = 0.75
Low ZNF839 expression
High ZNF839 expression
0
2000
4000
6000
8000
Disease duration (days)