Deep learning modeling of rare noncoding genetic variants in human motor neurons defines *CCDC146* as a therapeutic target for ALS

Sai Zhang[1,2,3,12,†], Tobias Moll[4,12], Jasper Rubin-Sigler[5,12], Sharon Tu[5], Shuya Li[6], Enming Yuan[7], Menghui Liu[1], Afreen Butt[4], Calum Harvey[4], Sarah Gornall[4], Elham Alhalthli[4], Allan Shaw[4], Cleide Dos Santos Souza[4], Laura Ferraiuolo[4], Eran Hornstein[8], Tatyana Shelkovnikova[4], Charlotte H. van Dijk[9], Ilia S. Timpanaro[9], Kevin P. Kenna[9], Jianyang Zeng[6], Philip S. Tsao[10,11], Pamela J. Shaw[4], Justin K. Ichida[5], Johnathan Cooper-Knock[4,†], Michael P. Snyder[3,†]

[1]Department of Epidemiology, University of Florida, Gainesville, FL, USA
[2]J. Crayton Pruitt Family Department of Biomedical Engineering, Genetics Institute, McKnight Brain Institute, University of Florida, Gainesville, FL, USA
[3]Department of Genetics, Center for Genomics and Personalized Medicine, Stanford University School of Medicine, Stanford, CA, USA
[4]Sheffield Institute for Translational Neuroscience (SITraN), University of Sheffield, Sheffield, UK
[5]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
[6]School of Engineering, Research Center for Industries of the Future, Westlake University, Hangzhou, Zhejiang, China
[7]Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, China
[8]Department of Molecular Genetics and Department of Molecular Neuroscience, Weizmann Institute of Science, Rehovot, Israel
[9]Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands
[10]VA Palo Alto Healthcare System, Palo Alto, CA, USA
[11]Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA
[12]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 fatal and incurable neurodegenerative disease caused by the selective and progressive death of motor neurons (MNs). Understanding the genetic and molecular factors influencing ALS survival is crucial for disease management and therapeutics. In this study, we introduce a deep learning-powered genetic analysis framework to link rare noncoding genetic variants to ALS survival. Using data from human induced pluripotent stem cell (iPSC)-derived MNs, this method prioritizes functional noncoding variants using deep learning, links cis-regulatory elements (CREs) to target genes using epigenomics data, and integrates these data through gene-level burden tests to identify survival-modifying variants, CREs, and genes. We apply this approach to analyze 6,715 ALS genomes, and pinpoint four novel rare noncoding variants associated with survival, including chr7:76,009,472:C>T linked to *CCDC146*. CRISPR-Cas9 editing of this variant increases *CCDC146* expression in iPSC-derived MNs and exacerbates ALS-specific phenotypes, including TDP-43 mislocalization. Suppressing *CCDC146* with an antisense oligonucleotide (ASO), showing no toxicity, completely rescues ALS-associated survival defects in neurons derived from sporadic ALS patients and from carriers of the ALS-associated G4C2-repeat expansion within *C9ORF72.* ASO targeting of *CCDC146* may be a broadly effective therapeutic approach for ALS. Our framework provides a generic and powerful approach for studying noncoding genetics of complex human diseases.

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], however, the knowledge of genetic factors influencing ALS survival remains scarce, hindering the development and translation of therapies to prolong patient survival.

Conventional genetic analysis, including genome-wide association studies (GWAS) and rare variant association studies (RVAS), can be underpowered due to the lack of functional characterization of genetic variants[5,6]. This limitation is particularly pronounced for noncoding variants which comprise the majority of human genetic material and hold significant inherited risk for complex diseases[7]. Precisely understanding the biological function of noncoding variants requires specific cellular context because of their involvement in transcriptional regulation that defines cell identity[8]. Moreover, rare variants (referred to as variants with minor allele frequency (MAF) below 0.5% in this study) pose additional challenges in traditional studies of variant function[9,10]. Recent advances in genomic deep learning[11–16] offer an efficient instrument for querying functional effects of noncoding variants, regardless of their allele frequency, in versatile cell types and tissues.

In this study, we developed a deep learning-powered genetic analysis framework to decode the genetic basis of ALS survival. Compared to conventional methods, our approach substantially enhanced the discovery power and pinpointed four novel rare noncoding variants associated with an average reduction in ALS survival of >50%. These variants were linked to three genes via transcriptional regulation: *CCDC146*, *ZNF839*, and *NRXN1*. We provided independent evidence to support the links between expression levels of nominated genes and the severity of ALS-specific molecular or clinical phenotypes. Leveraging CRISPR-Cas9, we introduced the top variant candidate chr7:76,009,472:C>T into MNs derived from a patient carrying an ALS-associated G4C2-repeat expansion of *C9ORF72*[3], and confirmed its regulatory effect on *CCDC146*. We further demonstrated that this variant exacerbated molecular phenotypes associated with ALS, including dendritic regression and TDP-43 mislocalization. Finally, we developed an antisense oligonucleotide (ASO) for suppressing *CCDC146* in neurons, and demonstrated that this ASO rescued the ALS-associated survival deficit irrespective of genetic background or the baseline expression of *CCDC146*, and without significant toxicity. We note that CCDC146 loss-of-function (LOF) is well tolerated *in vivo*[17]. We conclude that *CCDC146* suppression could be developed as a new and broadly effective treatment for ALS. Our framework establishes a new paradigm for studying noncoding genetics for complex diseases.

Results

Overview of the framework

Focusing on human motor neurons, the primary cells affected in ALS, our deep learning-based rare variant analysis framework (**Fig. 1**) integrates MN functional genomic profiling with ALS whole-genome sequencing (WGS). In this way we have optimized both the

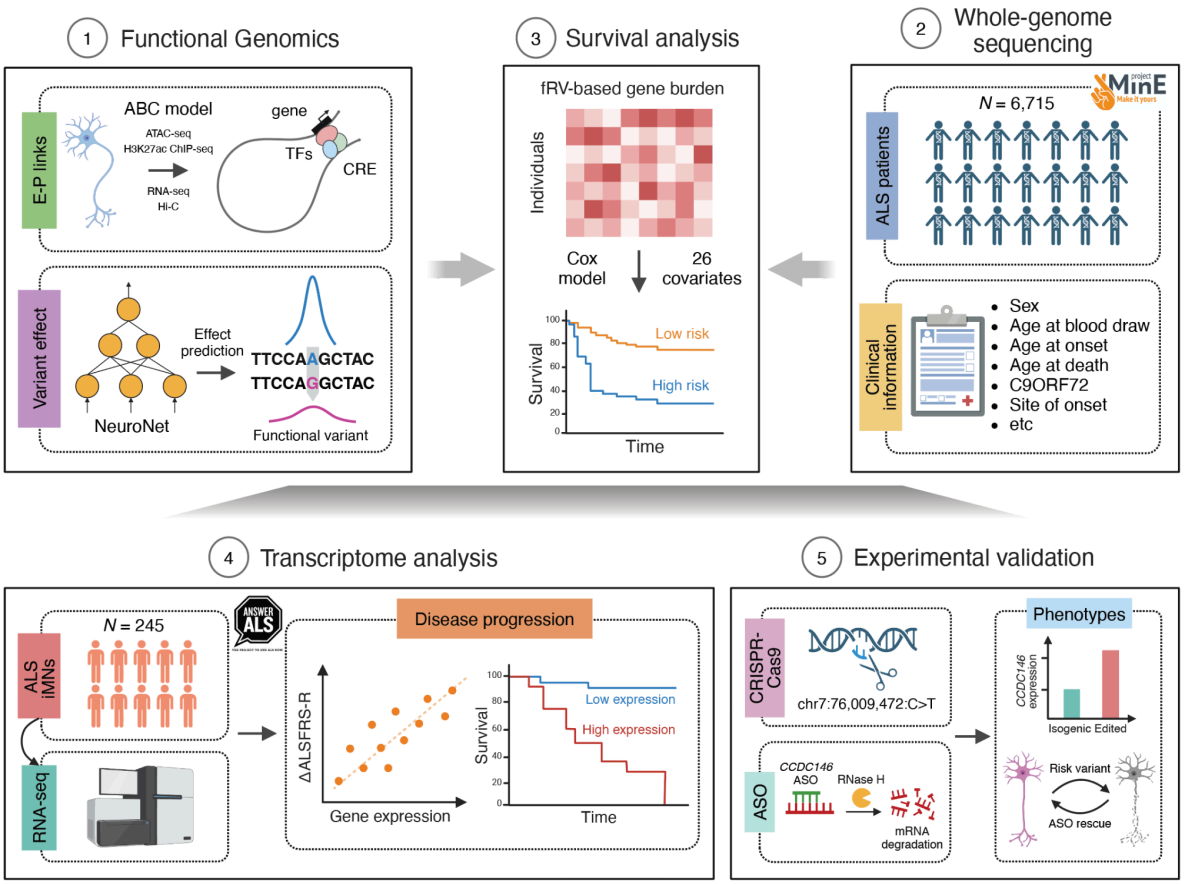


Fig. 1 | Schematic of the study design. Leveraging epigenome and transcriptome profiling in iPSC-derived motor neurons (iMNs), we first identified enhancer-promoter (E-P) interactions within MNs using the Activity-by-Contact (ABC) model (**Panel 1**, top). Next, we developed a deep learning model named NeuroNet to prioritize functional noncoding variants that could significantly alter MN epigenomes (**Panel 1**, bottom). Using the Project MinE whole genome sequencing (WGS) data (**Panel 2**), we identified risk variants and genes modulating ALS survival (**Panel 3**) utilizing the Cox proportional-hazards (PH) model, in which we collapsed all functional rare noncoding variants (predicted by NeuroNet) to their target genes (predicted by the ABC model). We validated the association between our candidate survival genes and ALS progression in an independent ALS cohort (**Panel 4**). Further, we introduced the top candidate variant chr7:76,009,472:C>T into patient-derived iMNs using CRISPR-Cas9 (**Panel 5**, top left), and verified its effects on both *CCDC146* expression and ALS-associated molecular phenotypes. Finally, utilizing an ASO to suppress *CCDC146* (**Panel 5**, bottom left), we completely rescued ALS-associated survival deficits in patient-derived neurons across different ALS genetic backgrounds, showing no significant toxicity. iPSC, induced pluripotent stem cell; CRE, cis-regulatory element; TF, transcription factor; fRV, functional rare variant; ASO, antisense oligonucleotide; RNase H, Ribonuclease H.

statistical power and biological interpretability of discovered genetic modifiers of ALS survival. Specifically, we first map the MN regulome (**Panel 1** top, **Fig. 1**), including enhancer-promoter (E-P) interactions[18], using transcriptome and epigenome (chromatin accessibility, H3K27ac histone mark, and 3D genome structure) profiling data from human induced pluripotent stem cell (iPSC)-derived MNs (iMNs)[19]. Next, we develop NeuroNet (**Panel 1** bottom, **Fig. 1**), a genomic deep learning model[12,14] to predict functional effects of noncoding variants on MN epigenomes (chromatin accessibility and histone marks). By mapping NeuroNet-prioritized functional variants to their target genes via E-P links, we then conduct a large-scale rare variant association study for ALS survival (**Panel 3**, **Fig. 1**), analyzing WGS data from 6,715 ALS patients[20] (**Panel 2**, **Fig. 1**). This integrative approach links functional noncoding variants which could alter MN epigenomes and further gene

expression, to differential ALS survival, which is verifiable using independent functional data (**Panel 4**, **Fig. 1**) and genome manipulation (**Panel 5**, **Fig. 1**).

Constructing the regulome in motor neurons

To systematically study transcriptional regulation in human motor neurons, we mapped MN enhancer-promoter (E-P) interactions by applying the activity-by-contact (ABC) model[18] (**Fig. 2a**; **Methods**) to the transcriptome (RNA-seq) and epigenome (ATAC-seq, H3K27ac ChIP-seq, and Hi-C) profiling data for iMNs derived from neurologically normal donors (N = 3)[19]. We identified 45,085 E-P interactions for 15,494 expressed genes (absolute E-P distance: mean = 188 kb, standard deviation (std) = 421 kb; **Fig. 2b** and **Supplementary Table 1**). We identified at least one non-promoter cis-regulatory element (CRE) for 85.1% of expressed genes. Of all expressed genes in iMNs, 74% were associated with more than one enhancer (mean = 2.9, std = 1.8; **Fig. 2c**). Similarly, 41% of enhancers were predicted to regulate more than one target gene (mean = 2.4, std = 3.2; **Fig. 2d**). This multi-to-multi relationship between enhancers and target genes highlights the complexity of transcriptional regulation within MNs. Importantly, 52% of E-P links skipped the closest gene (mean number of skipped genes = 2.5, std = 5.3; **Fig. 2e**), indicating the importance of long-range gene regulation within MNs.

To assess the specificity of MN E-P links, we compared E-P links between MNs and other neuronal cells based on a recent study[21]. While the majority of enhancers (63.7%) and expressed genes (65.6%) in MNs overlapped with those in other neuronal cells, a large proportion of MN E-P links (89.0%) were cell-type-specific (**Fig. 2f**). This result highlights the difference of transcriptional regulation between MNs and other neuronal cells, emphasizing the necessity for focused study of MNs. Interestingly, the distance of E-P links in MNs tended to be longer than those in other neuronal cells (median: 39 kb vs. 24 kb; $P < 2.2 \times 10^{-16}$, two-sided t-test; **Fig. 2g**).

To provide a functional characterization of gene regulation in MNs, we performed transcription factor (TF) binding motif analysis within enhancers. We compared TF binding motifs between short-range (<5% absolute E-P distance, N = 945) and long-range (>95% absolute E-P distance, N = 945) enhancers (**Fig. 2h**; **Methods**). The binding motif of SP110 was enriched in short-range enhancers; this TF has been implicated in the regulation of penetrance in an animal model of ALS[22]. TFs such as ATF1 and ATF3, whose binding motifs were enriched in long-range enhancers, are primarily involved in homeostasis in response to extracellular stimuli[23]. This is particularly important for cells such as MNs that are not only post-mitotic but also show limited functional redundancy because of their high energy demands and extreme axonal length. Similarly, the SP1 binding motif was enriched in long-range enhancers, and SP1 has been implicated in maintaining neuronal survival in the context of oxidative stress[24].

Next, we examined the biological implication of E-P links from a gene perspective. As expected, the set of genes linked to enhancers via short-range interactions (<5% absolute E-P distance, N = 775) were enriched with biological functions common to all neuronal subtypes (adjusted $P < 0.1$, Benjamini-Hochberg (BH) correction, two-sided Fisher's exact test; **Extended Data Fig. 1**; **Methods**). Genes with long-range enhancers (>95% absolute E-P distance, N = 774) were enriched with metalloendopeptidase inhibitor activity (GO:0008191; adjusted P = 0.045, BH correction, two-sided Fisher's exact test).

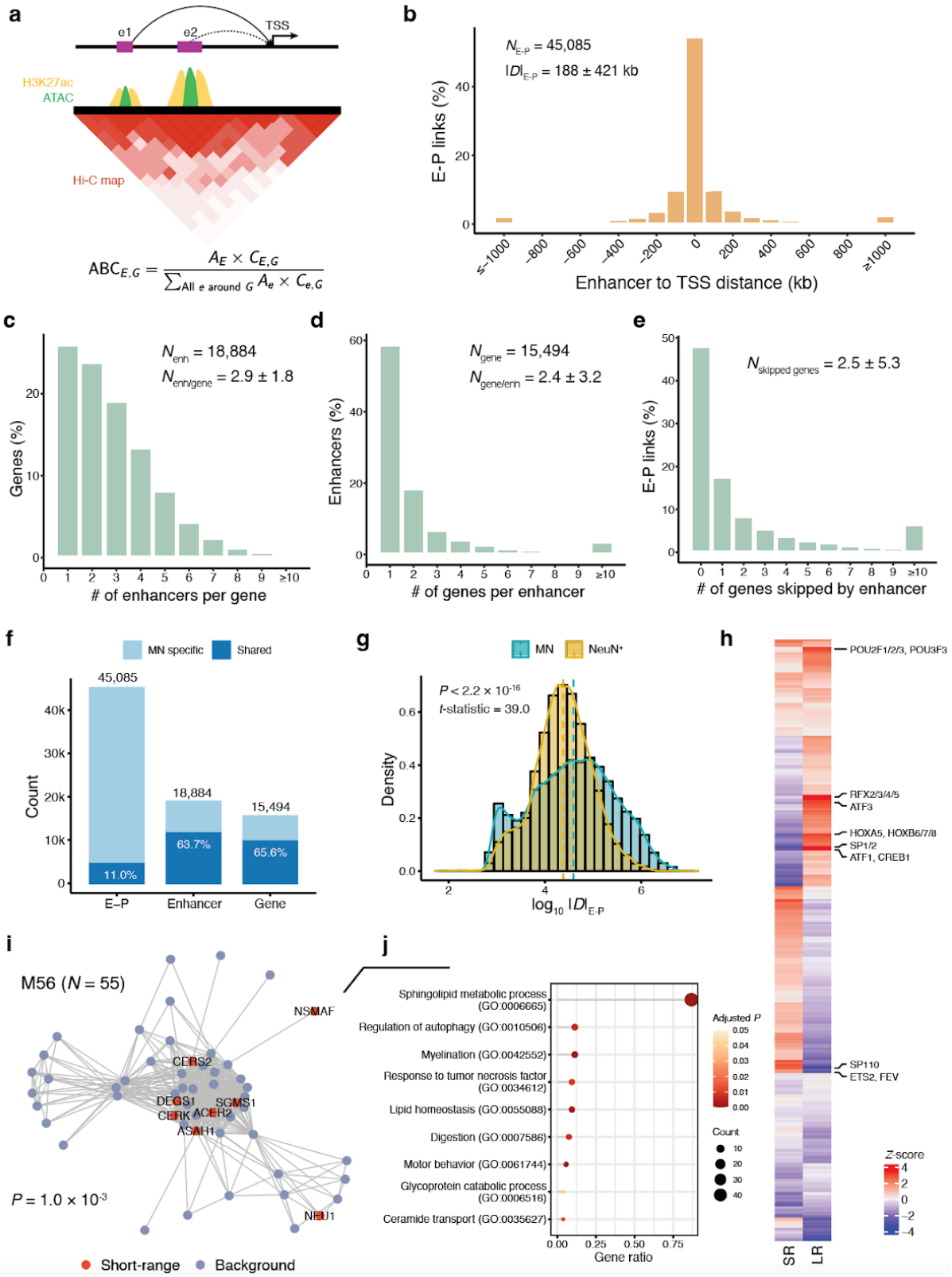


Fig. 2 | Enhancer-promoter (E-P) interactions in human motor neurons. a, Graphical illustration of the Activity-by-Contact (ABC) model, including factors used to derive E-P links. TSS, transcription start site; E, enhancer; G, gene. **b**, Distribution of distance from enhancer to TSS. D, distance in base pair (bp). **c-e**, Statistics of E-P links, including number of enhancers per gene (**c**), number of genes per enhancer (**d**), and number of genes skipped by an E-P link (**e**). enh, enhancer; #, number. **f**, Overlap between E-P links in MNs and those in neuronal cells (NeuN$^+$). **g**, Comparison of E-P links between MNs and neuronal cells. Density curves were added. Dashed vertical lines represent distribution medians. *P*-value by two-sided *t*-test. **h**, Differential motif

enrichment between long-range and short-range E-P links. LR, long-range; SR, short-range. **i**, Network module (M56) enriched with genes (highlighted in red) targeted by short-range enhancers. *P*-value by hypergeometric test. **j**, Gene ontology (GO) enrichment for M56 genes. Redundant GO terms were removed using the "simplify" function provided by "clusterProfiler"[25]. GO terms with adjusted $P < 0.1$ were visualized. *P*-value by two-sided Fisher's exact test.

Metalloendopeptidase acts extracellularly and it is notable that reduction in extracellular matrix (ECM) proteins exacerbates ALS neurodegeneration[26]. Moreover, the loss of metalloendopeptidase in neurons has been linked to impaired motor activities and cognitive deficits by altering axonal maturation and myelination in the central nervous system (CNS)[27].

To provide a systems view, we further performed a network analysis (**Methods**) based on human protein-protein interactions[28] (PPIs). We discovered one module (M56, $N = 55$; **Fig. 2i**) which was significantly enriched with genes targeted by enhancers in short range ($N = 8$; $P = 1.0 \times 10^{-3}$, hypergeometric test). Interestingly, genes within this module were enriched with biological pathways that are fundamental to both MN function and ALS pathogenesis (adjusted $P < 0.1$, BH correction, two-sided Fisher's exact test; **Fig. 2j**), including neuromuscular function (GO0061744), autophagy[29] (GO:0010506), and sphingolipid (GO:0006665) and ceramide transport (GO:0035627) which are important for organization of MN function and signaling[30]. No module was found to be enriched with long-range targeted genes. Altogether, our results reveal the complexity and specificity of gene regulation in MNs, highlighting the unique regulatory patterns for CREs and genes of functional significance in MNs and ALS.

Predicting functional effects of noncoding variants in motor neurons

The function and regulation of epigenomes (e.g., chromatin accessibility and histone marks), which vary widely across different cell types and contexts, are largely encoded in the DNA sequence[11,12,14]. To learn this epigenome grammar in MNs, we developed a convolutional neural network (CNN)-based deep learning model – NeuroNet – to predict chromatin accessibility (measured by ATAC-seq) and histone marks (H3K27ac, H3K4me1, and H3K4me3, measured by ChIP-seq) within iMNs from DNA sequence (**Fig. 3a**; **Methods**). The network structure of NeuroNet follows the Beluga architecture which has been widely used in biological sequence modeling[12,31]. Following the best practice[11,12,32], we trained NeuroNet based on 1,270,000 training examples and evaluated model prediction using 164,080 sequence-epigenome pairs held-out on chromosomes 8 and 9. NeuroNet achieved excellent prediction performance for all four epigenomic features (area under receiver operating characteristic curve (AUROC): mean = 0.851, standard deviation (SD) = 0.10; area under precision-recall curve (AUPRC): mean = 0.853, SD = 0.063; **Fig. 3b,c**; **Methods**).

Next, we used the trained NeuroNet model to characterize functional effects of noncoding genetic variants by querying the difference in predicted epigenomic features between reference and alternative alleles (**Fig. 3a**; **Methods**). Differing from traditional approaches in inferring variant effects, such as the quantitative trait loci (QTL) analysis[9,33,34], this *in-silico* strategy requires no individual-level data, which greatly increases its flexibility and generalizability. To validate our variant effect prediction, we compared NeuroNet predictions with those of DeepSEA[12], another deep learning model which has been used to predict 2,002 epigenomic features in various tissue and cell types. We estimated the Pearson correlation between NeuroNet and DeepSEA predictions for each cell type and epigenomic feature and, as expected, strong correlations were enriched with neuron- and brain-related

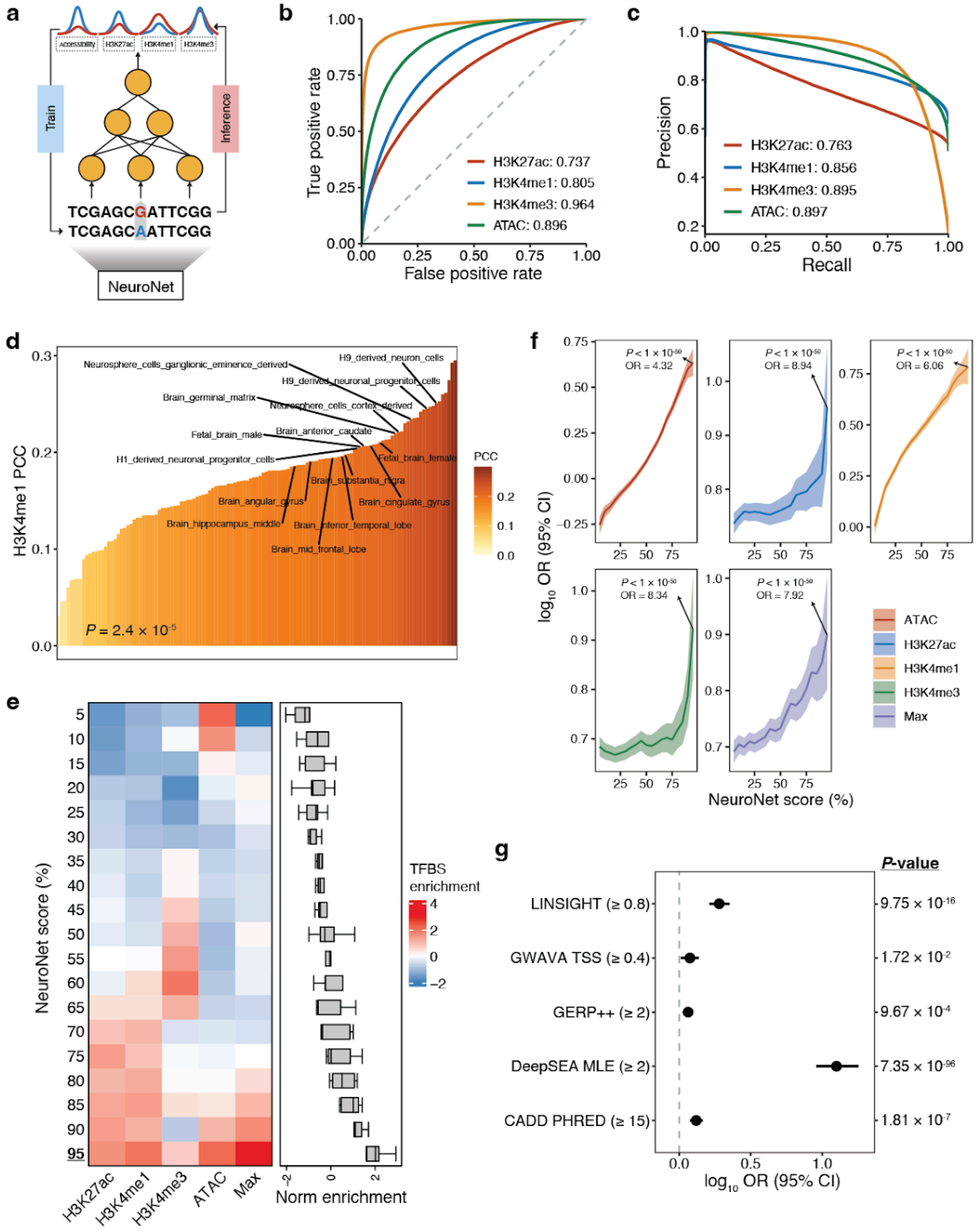


Fig. 3 | NeuroNet predicts functional effects of noncoding variants in MNs. a, Schematic of *in-silico* modeling of noncoding variant effects based on NeuroNet. **b-c**, Prediction performance of NeuroNet evaluated based on the receiver operating characteristic (ROC) curve (**b**) and the precision-recall curve (**c**). The area under curve (AUC) scores were calculated for different epigenomic features and metrics. The dashed gray line indicates random prediction. **d**, Pearson correlations between NeuroNet and DeepSEA predictions for H3K4me1. *P*-value by two-sided Fisher's exact test. Neuron- and brain-related tissues and cell types are highlighted. PCC, Pearson correlation coefficient. **e**, Enrichment of TFBS-disrupting variants within NeuroNet-prioritized variants (various cutoffs applied). Enrichment was quantified by *t*-statistics. The box plot center line, limits, and whiskers represent the median, quartiles, and 1.5x interquartile range (IQR), respectively. TFBS, transcription factor binding site; norm, normalized. **f**, Enrichment of NeuroNet-prioritized variants (various cutoffs applied) within MN CREs.

P-value by two-sided Fisher's exact test. OR is annotated with the solid line and 95% CI is represented by the colored area. Statistics of the 95% cutoff are highlighted. OR, odds ratio; CI, confidence interval. **g**, Enrichment of functional or deleterious variants predicted by other approaches within NeuroNet-prioritized functional variants (cutoff = 0.06). *P*-value by two-sided Fisher's exact test. Dot and error bar indicate OR and 95% CI, respectively. TSS, transcription start site; MLE, mean -log e-value.

cell types and tissues compared to other cell types (H3K4me1: $P = 2.4 \times 10^{-5}$; H3K4me3: $P = 4.5 \times 10^{-3}$, two-sided Fisher's exact test; **Fig. 3d** and **Extended Data Fig. 1b**). The enrichment was marginally significant for H3K27ac ($P = 0.059$, two-sided Fisher's exact test; **Extended Data Fig. 1c**), but not significant for chromatin accessibility ($P > 0.05$, two-sided Fisher's exact test; **Extended Data Fig. 1d**), probability due to the limited number ($N = 3$) of neuron and brain samples profiled by the DNase I hypersensitive sites sequencing (DNase-seq). These observations demonstrate the cell-type-specificity of variant effects captured by our NeuroNet modeling.

NeuroNet was designed to capture functional consequences of variants including, but not limited to, TF binding disruption. To aid the biological interpretation of NeuroNet predictions, we examined the overlap between NeuroNet-prioritized variants and variants associated with changes in TF binding. We then used this association to determine the NeuroNet score threshold for defining functional noncoding variants. Specifically, we split NeuroNet prediction scores into percentiles and examined which cutoff produced the strongest enrichment with variants that disrupt TF binding[35] (**Methods**). Interestingly, we observed that variants with higher NeuroNet scores were more enriched with variants that disrupt TF binding (**Fig. 3e**). This demonstrates that NeuroNet captured known TF biology despite the fact that TF motifs were not included in model training. To integrate the information from all epigenomic features, we took the maximum of four effect scores to derive the final NeuroNet variant effect score. The strongest enrichment of TF binding disruption was achieved at the 95% cutoff (NeuroNet score ≈ 0.06; $P = 5.5 \times 10^{-4}$, *t*-statistics = 3.46, two-sided *t*-test; **Fig. 3e**). Consequently, we adopted 0.06 as the threshold to define the functional variants (NeuroNet score > 0.06) in MNs.

As a further validation of the functional implication associated with NeuroNet prediction, we assessed the distribution of NeuroNet-prioritized variants across the genome. Specifically, we examined whether the MN CREs (i.e., promoters and enhancers) were enriched with NeuroNet-prioritized variants compared to other genomic regions. Notably, we found that MN CREs were enriched with functional variants predicted by NeuroNet across a wide range of cutoffs and for all epigenomic features (odds ratio (OR) > 1.0, two-sided Fisher's exact test; **Fig. 3f**; **Methods**). As for our TF binding site analysis, the enrichment was positively correlated with the NeuroNet scores in all features, and the peak enrichment was obtained at the 95% cutoff (**Fig. 3f**). This validates the biological relevance of our NeuroNet prediction and reinforces our threshold choice.

Finally, we benchmarked our prediction of functional noncoding variants against other widely-used methods, including GERP++[36], GWAVA[37], CADD[38], LINSIGHT[39], and DeepSEA[12]. By following the suggested thresholds to define functional or deleterious variants for different approaches, we observed that NeuroNet-defined functional variants were enriched with variants identified as functional or deleterious by all other methods ($P < 0.05$, OR > 1.0, two-sided Fisher's exact test; **Fig. 3g**; **Methods**). Interestingly, the strongest enrichment was obtained for DeepSEA, probably because it shares the largest similarity with NeuroNet in model design. Taken together, our analyses exhibit the cell-type-specificity and

biological relevance of our NeuroNet prediction on variant effects, and demonstrate its consistency with existing methods.

NeuroNet-powered rare variant analysis for ALS survival

We designed NeuroNet for inferring functional effects of noncoding variants in MNs, and applied the ABC model to map MN CREs to their target genes. Next, we sought to integrate these results with ALS WGS data to identify rare noncoding variants that modify ALS risk or survival via disrupting transcriptional regulation in MNs.

We first performed variant-level and sample-level quality controls (QCs) for 9,600 WGS samples generated in the Project MinE[20] (Data Freeze 2), resulting in an analysis-ready cohort encompassing 6,025 ALS patients and 2,264 controls (**Methods**). Next, we identified functional noncoding rare variants (allele frequency (AF) < 0.5% in gnomAD v2.1 non-Finnish European (NFE) samples, NeuroNet score > 0.06) located within MN CREs (promoters and enhancers). We mapped these variants onto their target genes based on the ABC-predicted E-P links. A gene-level rare variant burden profile was constructed for each sample based on the total number of mapped variants for individual genes.

To study the rare variant genetics of ALS risk, we first conducted a rare variant burden *risk* analysis (ALS patients vs. healthy controls) using SKAT-O[40,41] (**Methods**), with multiple covariates including age, sex, age^2, sex x age, sequencing platform, and the first 20 principal components to adjust for known confounders. No gene was significantly associated with ALS risk after Bonferroni multiple testing correction, suggesting that this analysis was underpowered (**Supplementary Table 2**). Interestingly, a number of genes whose protein function has been quantitatively linked to ALS risk achieved nominal significance, such as *KIF5A*[42,43] (P = 0.016) and *NEAT1*[44] (P = 0.033).

Next, we performed a rare variant burden *survival* analysis using the Cox proportional-hazards (PH) model (**Methods**), with 26 covariates including age, sex, age^2, sex x age, sequencing platform, *C9ORF72* status, and the first 20 principal components, to adjust for known confounders. Noncoding rare variant burden was significantly associated with ALS survival for seven genes after Bonferroni multiple testing correction (adjusted P < 0.05; **Fig. 4a,b** and **Supplementary Table 3**), for which no inflation was observed (λ = 0.99; **Fig. 4a**). Four rare noncoding variants were linked to three gene sets, and two variants (chr14:102,782,854:G>A and chr14:102,786,108:G>T) were linked to a single gene *ZNF839* (**Fig. 4b**). The mean survival time for the carriers of these rare variants ranges from 0.99 to 2.46 years, 2.79 to 1.32 years shorter than that of the overall cohort (**Fig. 4c**).

The four rare variants identified by our analysis contained both promoter and enhancer variants. In particular, chr2:51,259,489:G>C is located within the promoter (chr2:51,259,114-51,259,988) of *NRXN1* and the related gene *NRXN1-DT* (**Extended Data Fig. 2a**). The variant chr7:76,009,472:C>T resides within an enhancer (chr7:76,009,119-76,009,619) which is shared across several genes, including four significant genes in our survival analysis: *CCDC146*, *FAM185BP*, *FGL2*, and *UPK3B* (**Fig. 4d**). The two variants, chr14:102,782,854:G>A and chr14:102,786,108:G>T, are located within an enhancer (chr14:102,782,747-102,784,018) and the promoter (chr14:102,785,758-102,786,686) of *ZNF839* (**Fig. 4e**). This multi-to-multi relationship between noncoding variants and their regulatory targets further highlights the complexity of

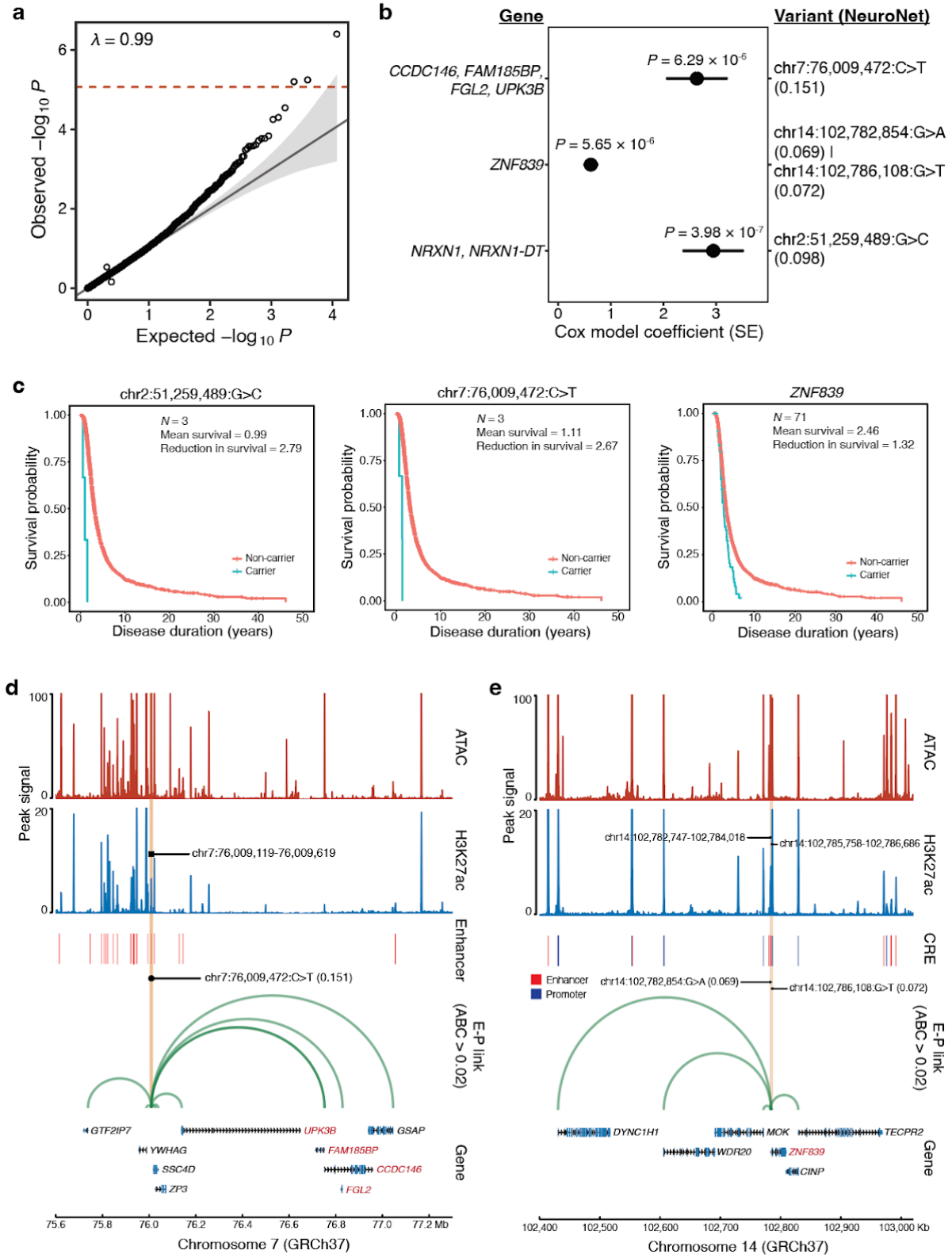


Fig. 4 | Survival analysis based on NeuroNet-prioritized rare noncoding variants. a, Quantile-quantile (Q-Q) plot of burden-based survival analysis. Genes sharing the same set of variant(s) were collapsed into a single test. The dashed red line indicates the significance cutoff based on the Bonferroni correction. The diagonal line represents the expected -$\log_{10}$(*P*-value) under the null hypothesis. The gray shaded areas represent the 95% confidence intervals under the null hypothesis. **b**, Statistics of significant tests after Bonferroni multiple testing correction. SE, standard error. **c**, Kaplan-Meier (KM) estimator plots for significant tests. Carriers for *ZNF839* variants were defined as individuals having at least one variant from chr14:102,782,854:G>A and

chr14:102,786,108:G>T. **d-e**, Regulatory relationships between variants and target genes for chr7:76,009,472:C>T (**d**) and *ZNF839* variants (**e**), respectively. Peak signals from iMNs derived from the donor CS14 were visualized. Genes from significant tests are highlighted in red.

transcriptional regulation in MNs. As a benchmark, we performed a standard rare variant burden survival analysis without prioritizing noncoding variants using NeuroNet; in this analysis no gene passed multiple testing correction (**Extended Data Fig. 2b** and **Supplementary Table 4**; **Methods**). These results demonstrate the superiority of our deep learning-based framework in enhancing discovery power and biological interpretation.

Lastly, we examined the overlap of rare variant associations between ALS survival and risk. We analyzed this by testing the enrichment of survival-associated genes within risk-associated genes under different *P*-value cutoffs. Interestingly, we observed significant gene overlaps across a spectrum of *P*-value cutoffs between survival and risk analyses ($P < 0.05$ for all cutoffs except 0.05, two-sided Fisher's exact test; **Extended Data Fig. 2c**), suggesting an overlapping genetic architecture underlying these two ALS phenotypes, especially within noncoding genomic regions.

Expression of ALS survival genes are correlated with disease progression in patient motor neurons

We next sought to validate the functional role of our candidate genes in the context of disease progression. In particular, we inspected gene expression in iMNs derived from an independent cohort[45] encompassing 245 ALS patients (**Methods**). We measured disease progression using two metrics, including the survival time from symptom onset and the rate of change in ALS Functional Rating Scale in its revised version[46] (ALSFRS-R). ALSFRS-R is widely used to measure ALS progression[47]. Of the seven candidate genes, the lower expression of *ZNF839* was associated with faster rate of disease progression ($\beta = 0.85$, $P = 5.9 \times 10^{-3}$, multivariable linear regression; **Fig. 5a**; **Methods**) and shorter survival time ($\beta = -0.79$, $P = 0.044$, Cox PH regression; **Fig 5b**; **Methods**). For other genes whose expression was not correlated with either the rate of change in ALSFRS-R or survival, we speculated whether our analysis was underpowered and therefore we sought more direct means of studying the function of our candidate variants and genes via experimental manipulation within patient-derived iMNs.

Reproducing NeuroNet-prioritized variant within iMNs alters target gene expression and exacerbates molecular phenotypes

The noncoding genetic variant chr7:76,009,472:C>T achieved the highest NeuroNet score among all four variants associated with ALS survival (**Fig. 4b**). Moreover, this variant was linked to four candidate target genes based on predicted E-P links (**Fig. 4d**). To experimentally verify our predictions and determine the regulatory target of chr7:76,009,472:C>T, we utilized CRISPR-Cas9 to reproduce this variant in iPSCs derived from a patient who suffered ALS caused by a G4C2-repeat expansion within *C9ORF72* (**Fig 5c**; **Methods**); the unedited cells served as an isogenic control. Of note, we observed extremely rapid disease progression in a patient who carried both a *C9ORF72* expansion and chr7:76,009,472:C>T in the Project MinE cohort (**Supplementary Table 5**).

We first tested whether chr7:76,009,472:C>T altered the expression of candidate genes as implicated by the ABC model and NeuroNet. To achieve this, we differentiated edited iPSCs

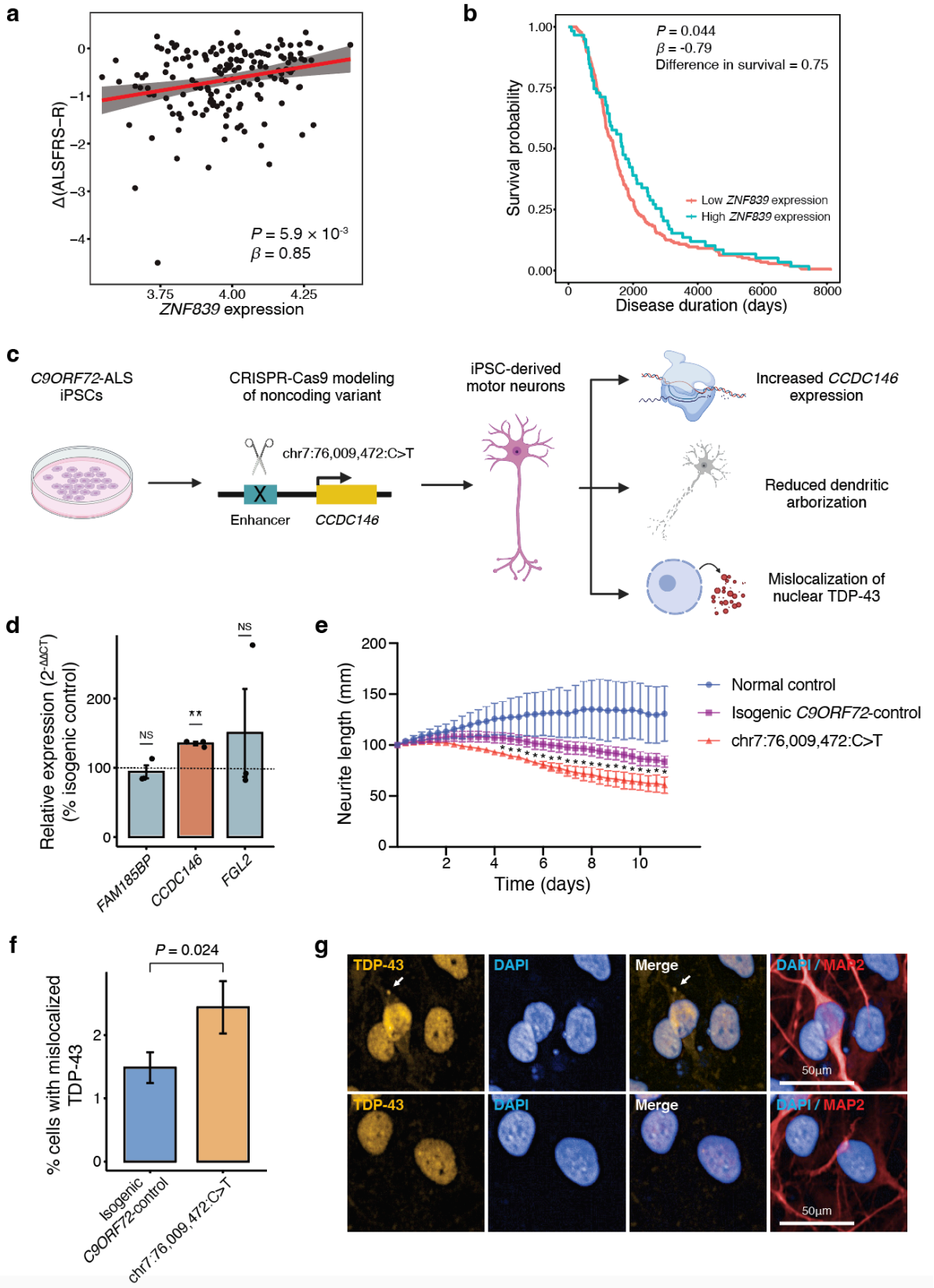


Fig. 5 | ALS survival-associated noncoding variants and target genes are correlated with clinical and molecular markers of disease progression. a, Correlation between the expression of *ZNF839* and the rate of disease progression as measured by clinical symptoms. The ALSFRS-R was scored from a maximum of 48 with a reduction in the score signifying increased disability. Red line indicates the regression line; gray areas represent the standard error. *P*-value by multivariable linear regression with sex, age of onset, site of onset, and

C9ORF72 status as covariates. **b**, Kaplan-Meier (KM) estimator plots of patients with high and low *ZNF839* expression. Gene expression level in the top 25% was defined as "high expression" (N = 59), otherwise as "low expression" (N = 176). P-value by Cox proportional-hazards (PH) regression. **c**, Schematic of CRISPR/Cas9-mediated modeling of chr7:76,009,472:C>T in *C9ORF72*-ALS iPSC-derived motor neurons (iMNs). **d**, Gene expression levels measured by q-RT-PCR for three genes linked to the enhancer harboring chr7:76,009,472:C>T. N = 3 technical replicates. P-value by two-sided t-test. NS, not significant; **, $P < 0.01$. **e**, Quantification of neurite length in an age- and sex-matched control, an isogenic control carrying a *C9ORF72* variant, and edited iMNs across 12 days of tracking using the Incucyte live-cell analysis system (Sartorius). N = 3 technical replicates. P-value by two-sided t-test. Error bars represent the standard error. *, $P < 0.05$. **f**, Quantification of the proportion of iMNs exhibiting mislocalization of nuclear TDP-43 to the cytoplasm in isogenic control and edited iMNs. Analysis was performed blinded to editing status. P-value by two-sided t-test. **g**, Representative images of TDP-43 staining (yellow) in edited iMNs. Nuclear (DAPI, blue) and cytoplasmic (MAP2, red) staining is also shown. iMNs with cytoplasmic mislocalization of TDP-43 are shown (top panels), together with a cytoplasmic inclusion (arrows). For comparison, iMNs are shown where TDP-43 was localized to the nucleus (bottom panels). Scale bar: 50μm.

carrying chr7:76,009,472:C>T, as well as isogenic control iPSCs, into mature MNs using our published small molecule protocol[19] (**Methods**). iMNs were characterized at day 40 after initiation of differentiation when we confirmed the expression of neuronal markers (**Extended Data Fig. 3a**). The ABC model predicted an E-P link for four genes: *CCDC146, FAM185BP, FGL2,* and *UPK3B*. Study of human tissue revealed that UPK3B is almost exclusively expressed in urothelium[48] and therefore we excluded this gene from further study. We extracted RNA from mature MNs and measured the expression levels for *CCDC146, FAM185BP,* and *FGL2* (**Methods**). Interestingly, we discovered that the expression level of *CCDC146* was significantly elevated in the presence of chr7:76,009,472:C>T compared to isogenic control iMNs ($P = 5.6 \times 10^{-3}$, fold change (FC) = 1.35, two-sided t-test; **Fig. 5d**).

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

ALS is associated with a number of MN molecular phenotypes, including increased cell death[49], reduced neurite outgrowth[50,51], and most importantly, TDP-43 mislocalization[52]. We took advantage of this battery of phenotypes to assess the effect of chr7:76,009,472:C>T on ALS severity *in vitro*. iMNs were placed in culture and monitored continuously for 12 days (**Methods**). We measured cell survival and neurite growth at 8 hour intervals. Notably, compared to isogenic control iMNs, the presence of chr7:76,009,472:C>T was associated with a reduction in both neurite length ($P < 0.05$, two-sided t-test; **Fig. 5e**) and the number of neurite branch points (**Extended data Fig. 3b**), along with increased neuron death (**Extended Data Fig. 3c**). Importantly, iMNs carrying chr7:76,009,472:C>T had a significantly higher burden of TDP-43 mislocalization at the end stage ($P = 0.024$, two-sided t-test; **Fig. 5f,g**; **Methods**), compared to isogenic control iMNs. Taken together, we provide multiple layers of evidence for a deleterious effect of chr7:76,009,472:C>T on MN health including ALS-specific phenotypes, agreeing well with our predictions.

Suppressing CCDC146 with an antisense oligonucleotide rescues ALS-associated survival deficit in patient-derived neurons

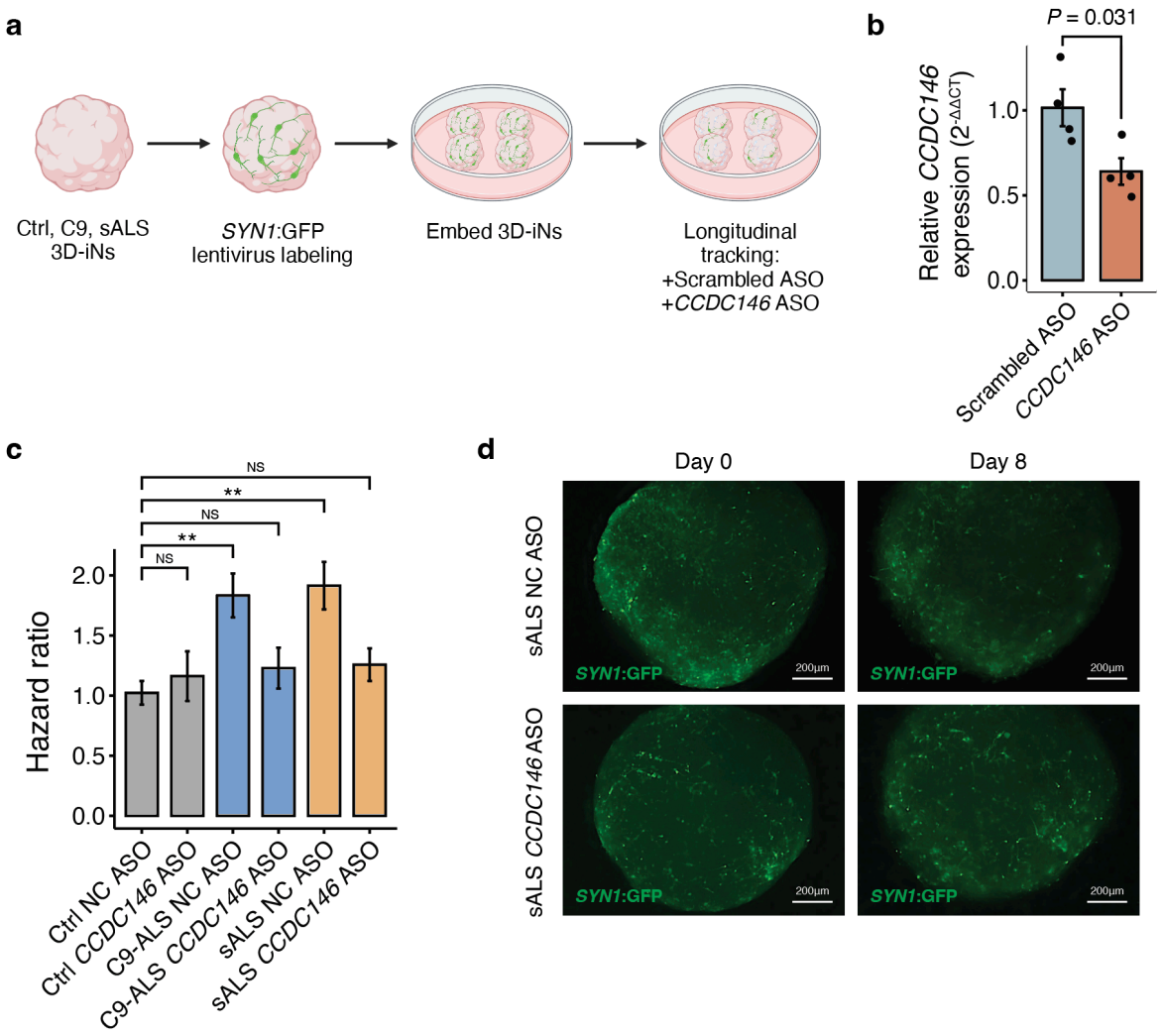


Fig. 6 | Suppression of *CCDC146* in ALS patient-derived neurons via antisense oligonucleotide (ASO) rescues survival deficits without significant toxicity. a, Schematic of our modeling strategy which utilizes patient-derived GFP-labeled iPSC-derived neurons in 3D spheroids (3D-iNs). Longitudinal tracking enables accurate cell counting to determine survival following a glutamate insult together with a scrambled ASO or a *CCDC146*-targeting ASO treatment. Ctrl, control; C9, C9ORF72; sALS, sporadic ALS. **b**, Comparison of gene expression levels measured by q-RT-PCR between *CCDC146*-targeting ASO and scrambled ASO treatments. *P*-value by two-sided *t*-test. $N = 4$ technical replicates per condition. **c**, Comparison of hazard ratio measuring the incidence of neuron loss among different conditions and treatments. $N = 2$ donors; $N = 200$ neurons, $N = 4$ spheroids per condition, per donor. *P*-value by Dunnett's test to correct for multiple testing and incorporate the control condition common to all tests. **, adjusted $P < 0.01$; NS, not significant; Ctrl, control; NC, negative control. **d**, Representative images of 3D-iNs derived from a sporadic ALS patient treated with the scrambled ASO (top panels) or the *CCDC146*-targeting ASO (bottom panels). Images were taken from approximately the same location on the plat at Day 0 (left panels) and Day 8 (right panels), demonstrating neuronal loss which was rescued by *CCDC146* suppression. Scale bar: 200µm.

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

We have established a survival readout for iPSC-derived neurons (iNs) whereby neurons derived from patients display an increased sensitivity to glutamate-induced excitotoxicity[49,53]

(**Fig. 6a**; **Methods**). This reflects long standing observations regarding glutamate toxicity in ALS[54–56]. We have also established this readout in a 3D spheroid format (3D-iN) to enable prolonged culturing without primary glial cells[53,56] (**Fig. 6a**; **Methods**); this is important for a direct assessment of interventions targeted at neuronal health. In our experiment, we treated 3D-iNs derived from multiple patients, including a healthy control, two ALS patients carrying an *C9ORF72* G4C2-repeat expansion, and two sporadic ALS (sALS) patients, using our designed *CCDC146* ASO and a scrambled control ASO, followed by longitudinal tracking (**Fig. 6a**; **Methods**). First, we confirmed the reduction in *CCDC146* mRNA level within 3D-iNs treated by the *CCDC146* ASO compared to the scrambled ASO (36.9% reduction, $P = 0.031$, two-sided *t*-test; **Fig. 6b, Methods**). Notably, we observed that the survival defect was completely rescued by *CCDC146* suppression compared to the scrambled ASO (Adjusted $P < 0.01$, Dunnett's test; **Fig. 6c,d** and **Extended Data Fig. 4a-d**). The result persisted across different genetic backgrounds including *C9ORF72* G4C2-repeat expansion and sALS (**Fig. 6c** and **Extended Data Fig. 4a-d**). Moreover, there was no evidence of toxicity mediated by the *CCDC146* ASO in control 3D-iNs (Adjusted $P > 0.05$, Dunnett's test; **Fig. 6c**, **Extended Data Fig. 4e,f**). Indeed, LOF of CCDC146 has been shown to be well tolerated in humans[17]. Of note, none of the experimentally-tested patient-derived 3D-iN lines exhibited elevated *CCDC146* expression at baseline (**Extended Data Fig. 4g**). These results demonstrate that *CCDC146* suppression may be a broadly applicable therapeutic strategy for various forms of ALS.

Discussion

Single-cell technologies have been widely applied nowadays to chart cellular diversity and molecular heterogeneity across different contexts and conditions[57]. However, it remains technically challenging to enrich human motor neurons via single-cell profiling due to their sparsity in related tissues[58,59]. This makes stem cell differentiation as the standard method for exploring motor neuron biology and related conditions thus far[19,60].

Comprehensive functional annotation of genetic variation within the noncoding genome is an unmet challenge, largely because of the cell-type-specificity of regulatory genome function. A growing body of literature is approaching this problem via training cell-type- and tissue-resolved deep learning models to predict epigenomic and transcriptomic features from DNA sequence[11–16,31,32]. This body of work gains advantages over conventional approaches in mapping variant genome function, such as analysis of expression and chromatin accessibility quantitative trait loci (eQTL and caQTL) which requires access to matched individual genotype and transcriptomic/epigenomic data and lacks power to study rare or ultra-rare variants. Here, we have applied a customized deep learning model to predict epigenome effects of noncoding variants in a specific cell type, the motor neuron, which is the hallmark cell type for ALS pathogenesis. This strategy enables us to precisely characterize the functional effects of any noncoding variants of interest, irrespective of allele frequency. NeuroNet scores are correlated with multiple existing variant scoring methods built upon evolutionary or multi-tissue/cell-type information, but many of the discoveries we reported here would only have been possible with NeuroNet.

The complexity of gene regulation challenges the delineation of enhancer-promoter interactions in specific cellular context. The ABC model[18] has become the preferred method to infer E-P links, and we have applied this method to our iMN profiling data. Our statistical analyses uncovered various unique patterns and functions associated with the MN

regulome, highlighting the need for cell-type-specific study of variant genome function. Long-range E-P interactions are particularly specific to MNs in our analysis and are associated with expression of modifiers of metalloendopeptidase, in line with previous studies linking ECM structure[26], and metalloendopeptidase[27] in particular, to neuronal function and neurodegeneration.

Combining NeuroNet predictions of noncoding variant effects on the MN epigenome, together with enhancer-promoter links, enabled us to build a variant-to-gene (V2G) map within human MNs. Based on this V2G, we have performed large-scale rare variant burden tests for both ALS risk and survival. Although the risk analysis was underpowered probably due to the paucity of control samples, we boosted the statistical power of survival analysis using the NeuroNet-aided V2G strategy, demonstrating the benefit of incorporating functional information into our genetic analysis.

We have identified four noncoding rare variants linked to ALS survival, and they were predicted to modulate the expression of three protein-coding genes: *ZNF839*, *NRXN1*, and *CCDC146*. Interestingly, although we did not identify any genome-wide significant genes associated with ALS risk, we did notice a significant overlap between nominally significant risk- and survival-associated genes. This is consistent with known genetic variants that affect both ALS risk and survival, including noncoding variants associated with *UNC13A*[61] and *C9ORF72*[3], and provides evidence for the genetic architecture overlap between ALS risk and survival within the noncoding genome.

The expression of *ZNF839* was predicted to be regulated by two different enhancer variants that are associated with shorter ALS survival. Lower *ZNF839* expression within patient-derived iMNs was correlated with reduced survival and faster clinical progression. *ZNF839* is poorly characterized in the literature, but has been associated with ALS risk via GWAS[62]. The expression of *NRXN1* was predicted to be modulated by an enhancer variant that is associated with shorter survival. *NRXN1* has been linked to ALS in the literature[63]. In particular, *NRXN1* expression is reduced by TDP-43 dysfunction, and structural variations of *NRXN1* are linked to ALS risk and severity (data not shown).

The variant chr7:76,009,472:C>T was linked to four genes in MNs via the ABC model: *CCDC146*, *UPK3B*, *FAM185BP*, and *FGL2*. To resolve this uncertainty, we reproduced chr7:76,009,472:C>T in iMNs carrying an ALS-associated G4C2 expansion within *C9ORF72*. Interestingly, only *CCDC146* expression was altered by chr7:76,009,472:C>T. In support of our analysis, we observed that chr7:76,009,472:C>T exacerbated ALS-associated phenotypes, including TDP-43 mislocalization, and reduced MN survival, compared to isogenic control neurons. Moreover, reduction of *CCDC146* expression targeted by an ASO completely rescued ALS-associated survival defects in iMNs, irrespective of genetic background and without significant toxicity, and also irrespective of the baseline *CCDC146* expression. We note that a different ASO has recently been approved for treatment of ALS patients carrying a pathogenic mutation within SOD1[64]. Unlike that treatment - Tofersen - our data suggest that a *CCDC146*-targeted ASO could be administered to ALS patients of diverse forms, providing compelling evidence for the ASO we have developed as a potential new therapeutic for ALS.

CCDC146 is a microtubule associated protein (MAP) and localized to sperm flagella. Biallelic variants in *CCDC146* have been associated with reduced sperm viability and male infertility

530 in both mice and humans[17]. Dysfunction of MAP is a recognized upstream driver of ALS
531 pathogenesis[65]. Flagella are functionally mirrored by cilia and it is notable that primary cilia
532 dyskinesia has been associated with ALS[66,67]. We note that frank primary ciliary dyskinesia
533 has not been observed in CCDC146 knockout mice[17], but we speculate a more subtle
534 phenotype may be present and should be the subject of further research. The variants we
535 have identified are all rare in the population; we have previously speculated that variants that
536 modify ALS pathogenesis might be rare, despite the fact that ALS onset is typically later than
537 reproductive age, because they have a pleiotropic effect on gamete function[5].

538 It is notable that, aside from the fertility defect, loss of CCDC146 is well tolerated, suggesting
539 our ASO approach may be viable without significant toxicity. Indeed, loss-of-function (LOF)
540 variants within CCDC146 cause male infertility but are otherwise well tolerated in both mice
541 and humans[17]. Additionally, this gene is not well conserved[68] (probability of being
542 loss-of-function intolerant (pLI) = 0). CCDC146 is expressed throughout the brain and spinal
543 cord, and particularly within neurons[69] (Human Protein Atlas). Expression of *CCDC146* has
544 been identified specifically in neurons from the human (as opposed to mouse) spinal cord[58].
545 Matching our finding of higher *CCDC146* expression associated with reduced ALS survival,
546 the expression of *CCDC146* is higher in ALS patient spinal cord compared to control cervical
547 spinal cord although, at a bulk level, this difference was not statistically significant[70]. Future
548 work will focus on the mechanism of CCDC146 neurotoxicity in ALS and in particular, how
549 elevated *CCDC146* expression produces the exacerbation in TDP-43 mislocalization as we
550 observed in our data.

551 We investigated the potential of targeting CCDC146 through small-molecule drugs. We first
552 searched chemical and patent databases but found no active compounds targeting
553 CCDC146 and its similar proteins (**Methods**). CCDC146 is also classified as a poorly
554 explored target without active small molecules according to the Pharos database[71,72]. Due to
555 the lack of experimentally solved protein structure data, we obtained the AlphaFold-predicted
556 structure of CCDC146[73] (**Extended Data Fig. 5**). The predicted structure mainly consists of
557 coiled coils linked by flexible regions, without ligand binding sites detected on its surface[74]
558 (**Methods**), suggesting that it would be challenging to design small-molecule drugs for this
559 target. As we already discussed, in the cellular environment the CCDC146 protein is likely to
560 be located in centriole and microtubule[17], which may further limit its accessibility for ligand
561 binding. Taken together, CCDC146 could be difficult-to-drug via traditional small-molecule
562 drug discovery strategies, underscoring the importance of alternative targeting approaches
563 such as ASO.

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

571 In summary, we have developed a deep learning-powered genetic analysis framework for
572 identifying cell-type-resolved genetic modifiers of ALS survival. Each of the candidate genes
573 we pinpointed represents a potential therapeutic target. NeuroNet can be applied to study
574 the genetics of other MN-specific diseases, including spinal muscular atrophy (SMA) and

hereditary sensory and motor neuropathy (HSMN). Our analysis framework provides a generic paradigm for boosting noncoding genetic analysis of complex human diseases.

Data availability

The epigenome and transcriptome profiling data of iPSC-derived motor neurons 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

We used the ABC model (https://github.com/broadinstitute/ABC-Enhancer-Gene-Prediction/tree/master) to construct the E-P links in iMNs. Statistical analyses were conducted using corresponding R and Python packages. The NeuroNet software and tutorial can be found at https://github.com/szhang1112/neuronet.

Acknowledgments

This work was supported by NIH (CEGS 5P50HG007735 to M.P.S., R01NS131409 and R01NS097850 to J.K.I.), Wellcome Trust (216596/Z/19/Z to J.C.-K.), NIHR (NF-SI-0617-10077 to P.J.S.), Motor Neurone Disease (MND) Association (899-792 and 913-792 to C.H., S.G. and J.C.-K.), and John Douglas French Alzheimer's Foundation (to J.K.I.). We also acknowledge support from the MND Association Fellowship (2323-799) and the Kingsland Fellowship (to T.M.), the NIHR Sheffield Biomedical Research Centre for Translational Neuroscience (IS-BRC-1215-20017), and the NIHR Sheffield Clinical Research Facility. We are very grateful to the ALS patients and control subjects who generously donated biosamples. We acknowledge transcriptomic data provided by the AnswerALS Consortium. Figures 1, 3a, 5c, and 6a were created using BioRender.com.

Author contributions

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

Competing interests

M.P.S. is a co-founder and the scientific advisory board member of Personalis, Qbio, January, SensOmics, Filtricine, Akna, Protos, Mirvie, NiMo, Onza, Oralome, Marble Therapeutics and Iollo. He is also on the scientific advisory board of Danaher, Genapsys, and Jupiter. J.K.I. is a co-founder and a scientific advisory board member of AcuraStem, Inc.

and Modulo Bio, a scientific advisory board member of Synapticure and Vesalius Therapeutics. J.K.I. is also an employee of BioMarin Pharmaceutical. The remaining authors declare no competing interests.

Methods

Prediction of enhancer-promoter interactions using the ABC model

We implemented the ABC model following the guidelines provided at https://github.com/broadinstitute/ABC-Enhancer-Gene-Prediction. First, we called peaks for the iMN ATAC-seq profiling data[19] using MACS2, and then identified the candidate enhancer elements using "makeCandidateRegions.py" with peakExtendFromSummit = 250 and nStrongestPeaks = 150000. The black-listed regions generated by the ENCODE 4 (https://www.encodeproject.org/) were used for removing enhancers overlapping regions with anomalous sequencing reads. Second, we applied "run.neighborhoods.py" to quantify the enhancer activities by counting ATAC-seq and H3K27ac ChIP-seq reads in candidate enhancer regions. RNA-seq profiling of iMNs was also provided to inform expressed genes. Quantile normalization was applied using K562 epigenomic data as the reference. At last, using "predict.py" we computed the ABC scores by combining the enhancer activities (calculated by the second step) with the Hi-C profiling. Hi-C data were fit to the power-law model. The suggested threshold 0.02 was used to define valid E-P links.

NeuroNet model design and training

The architecture of NeuroNet follows the Beluga structure[12], a powerful convolutional neural network (CNN) designed to predict 2,002 epigenomic features from the DNA sequence, except the last layer of NeuroNet is tailored to predict four epigenomic features (chromatin accessibility, H3K27ac, H3K4me1, and H3K4me3) in iPSC-derived MNs.

The input of NeuroNet is the one-hot encoded 2,000-base-pair (bp) DNA sequence of which the centered 200 bp region overlaps peaks of at least one epigenomic feature – called valid sequence in this paper. The model outputs four probabilities indicating the peak status. Given the encoded input sequence X_i (i = 1, ..., N), the model f is trained to minimize the following binary cross-entropy loss:

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

Where j is the index of four epigenomic features, σ is the sigmoid function, y_{ij} is the binary label of the j-th feature of the i-th valid sequence, and Ω is the set of network parameters.

Genome-wide valid sequence samples were randomly split into 1270000, 4000, and 164080 training, validation, and testing samples, respectively. The dataset was further augmented by adding reverse complement sequences. Stochastic gradient descent (SGD) with learning rate = 0.1, momentum = 0.9, and weight decay = 1×10^{-6} was adopted as the optimizer. Batch size was set to 256. We also applied early stopping to prevent overfitting, in which the AUPRC was selected as the monitoring metric for an imbalanced dataset.

Variant effect prediction using NeuroNet

After the model was trained, we utilized NeuroNet to infer the functional effects of individual variants on epigenomic features. Given a genetic variant, we construct a variant-centered

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

Transcription factor binding motif analysis

The TF binding motif analysis was performed using GimmeMotifs[75]. The short-range and long-range enhancers were defined as enhancers with the mean absolute E-P distance below 5% and above 95% of the distribution, respectively. The differential motifs between short- and long-range enhancers were identified using the “gimme maelstrom” command. This command combined the results from different methods using the Stouffer’s method, yielding an integrated *Z*-score. A higher *Z*-score means the motif better matches the sequence of the peak region in one condition than the other.

Network analysis

We downloaded the human PPIs from STRING version 12.0[28], including 19,622 proteins and 6,857,702 protein interactions. We extracted high-confidence (combined score >700) PPIs for downstream analysis, including 16,185 proteins and 236,000 protein interactions. To eliminate the bias caused by hub proteins[76], we ran the random walk with restart (RWR) algorithm over the PPI network, wherein the restart probability was set to 0.5, resulting in a smoothed network after preserving the top 5% predicted edges (N = 6,243,766). We then applied the Louvain method[77], a classic community detection algorithm searching for densely connected modules, to decompose the network into different modules. After the algorithm converged, we obtained 1,261 modules with an average size of 13 nodes.

The enrichment of genes of interest within each module was tested using the hypergeometric test. Modules with false discovery rate (FDR) < 0.1 after the BH procedure were reported.

Gene ontology analysis

The gene ontology (GO) analysis in this work was performed based on clusterProfiler[25] and enrichR[78]. All enrichment analyses were focused on the biological process and molecular function. The gene enrichment was evaluated based on a two-sided Fisher's exact test, and only terms with adjusted $P < 0.1$ (after BH correction) were reported. To remove the GO term redundancy in some cases, we applied the “simplify” function provided by clusterProfile.

Benchmarking of NeuroNet prediction against other variant scores

To benchmark NeuroNet against DeepSEA, the Pearson correlation between NeuroNet prediction for four MN epigenomic features and DeepSEA prediction for matched epigenomic features was calculated. For each MN epigenomic feature, all correlation

coefficients were first sorted and evenly divided into highly- and weakly-correlated sets, and then the rank of brain- and neuron-related correlations was examined using a two-sided Fisher's exact test within these two sets.

SNP2TFBS[35] was used to predict the variant effect on TF binding affinity. The binding affinities for different TFs were averaged for each studied variant to estimate its overall effect. Given a particular NeuroNet cutoff, the noncoding variants were split into two groups according to their NeuroNet predictions. We compared the SNP2TFBS scores between these two groups of variants and examined the difference using a two-sided *t*-test. The *t*-statistics was reported to indicate the SNP2TFBS score difference between two groups. The final NeuroNet score threshold was determined at the largest difference.

With the determined NeuroNet threshold, the correlation between NeuroNet prediction and other scores was inspected by comparing scores between NeuroNet-positive and NeuroNet-negative variants using a two-sided Fisher's exact test.

Whole-genome sequencing data in Project MinE

Project MinE WGS data (Data Freeze 2) consists of 6,715 sporadic ALS patients and 2,885 healthy controls recruited at specialized neuromuscular centers in the UK, Belgium, Germany, Ireland, Italy, Spain, Turkey, the United States and the Netherlands[79]. Patients were diagnosed with possible, probable or definite ALS according to the 1994 El-Escorial criteria[80]. All controls were free of neuromuscular diseases and matched for age, sex and geographical location. Sample-level and variant-level quality controls (QCs) have been described previously[79]. Briefly, variants with missingness > 5% were removed, as were variants out of Hardy-Weinberg equilibrium (HWE) in controls ($P < 1 \times 10^{-6}$) and monomorphic variants. All samples had missingness <10% across all 22 chromosomes and samples with an inbreeding coefficient >3 standard deviations from the mean of the distribution were excluded, as were unexpected related samples.

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

After QCs, single nucleotide variants (SNVs) meeting all the following criteria were reserved for the survival analysis: (1) minor allele frequency (estimated based on non-Finnish European samples in the gnomAD whole-genome collection 2.1.1) < 0.5%; (2) minor allele count ≥ 3 within the QCed cohort; (3) residing within MN CREs (promoters and enhancers) defined by the ABC model; (4) NeuroNet score > 0.06. These criteria together defined the "functional rare noncoding variants" in our study. Next, screened variants were collapsed to their target genes based on the ABC E-P links and the number of total mapped variants for each gene was counted (i.e., gene burden). To associate the gene burden to survival, we applied the Cox proportional-hazards model implemented by the "coxph" function in the R "survival" package, with 26 covariates including age, sex, age^2, sex x age, sequencing platform, *C9ORF72* status and the first 20 principal components. Because many genes shared the same group of regulatory variants resulting in duplicate tests, we collapsed those genes into unique sets and carried out the Bonferroni multiple testing correction. Tests with adjusted $P < 0.05$ were reported.

We also performed a case-control analysis by including the control samples in Project MinE. 2,264 controls were reserved for analysis after QCs. Functional rare noncoding variants were identified in the same way as that in the survival analysis, and then included in the

case-control study. Gene burdens were calculated similarly. To associate gene burden to the disease status, we applied the SKAT-O method[40] implemented in the RVTESTS[40], with 25 covariates including age, sex, age^2, sex x age, sequencing platform, and the first 20 principal components.

Benchmarking of survival analysis

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

Transcriptome analysis for the AnswerALS cohort

For AnswerALS dataset, RNA-seq data of iPSC-derived MNs and phenotype data were obtained for 245 ALS patients. Gene expression was normalized for gene length and then sequencing depth to produce transcripts per kilobase million (TPM) using Kallisto v0.46.0[81]. Read counts were trimmed Mean of M-values (TMM) normalized[82] to account for differences in library size. For each gene, based on the expression level in and outside the top 25%, patients were split into two groups – high-expression and low-expression groups. We used multivariable linear regression to determine the relationship between gene expression level and rate of change in the ALSFRS-R, and Cox proportional-hazards regression on two patient groups to determine the relationship between gene expression and survival time from symptom onset. In both cases, age of disease onset, sex, site of disease onset, and *C9ORF72* status were included in the model as covariates. Significance testing was performed for all genes expressed in iMNs as determined by >0 read counts in more than half of samples, and mean read counts across all samples >25th percentile. The ALSFRS-R was scored from a maximum of 48 where a reduced score indicates increasing disability[47]. We calculated the rate of change in the ALSFRS-R (i.e., Δ(ALSFRS-R)) assuming a linear decline from a score of 48 at symptom onset.

CRISPR-editing of survival-associated genetic variant

The variant chr7:76,009,472:C>T (hg19 genome build, ENSEMBL) was introduced to *C9ORF72*-ALS 183 iPSCs (passage 21) (RRID:CVCL_UF85) using CRISPR/Cas9-mediated homology directed repair (Synthego[83]). Donor sequence: GTGCCAGTCGCAGGCTCTCCAATCAGGAGGCTGCCGTGGGGCCAGGGGTGTGGCTCC GGTTCCCACGCGGGGCGAATTCCCTCTCCCGGTAGTCTCCCCAG. gRNA sequence: GGGAACCGGAGCCGCGCCCC. *C9ORF72*-ALS 183 iPSCs were derived from a 51 year old caucasian male with *C9ORF72* G4C2 repeat expansion and a diagnosis of familial ALS at the time of sampling[84,85]. Editing efficiency was confirmed via Sanger sequencing using the following primers: fwd (5'-3'): CAGAAGCCCTGAGGAGTGC, rvs (5'-3'): GCCTCGGAGGTTTTTCTGGA. Sequencing traces were uploaded to the Synthego ICE Analysis tool (v3) (https://ice.synthego.com), and the indel efficiency within the mixed pool was calculated at 91%. Editing efficiency remained stable throughout motor neuron differentiation. Sequence traces are available on request. CRISPR-edited cells were engineered by Synthego.

Tissue culture of iPSCs and iPSC-derived MNs

iPSCs were cultured in mTesR Plus media (StemCell Technologies) in Matrigel-coated 6-well plates (Corning) and maintained at 37°C, 5% CO2. Cells were passaged when ~80%

782 confluent using ReLeSR (StemCell Technologies), according to the manufacturer's
783 instructions. iPSCs were differentiated into neural progenitor cells (NPCs) using a modified
784 version of the dual SMAD inhibition protocol[19,86]. On the day after plating (day 1), after the
785 cells had reached ~100% confluence, the cells were washed once with PBS and the medium
786 was replaced with neutralization medium (50% of KnockOut DMEM/F-12, 50% of
787 Neurobasal), 0.5× N2 supplement (ThermoFisher), 1x Gibco GlutaMAX Supplement
788 (ThermoFisher), 0.5x B-27 supplement (ThermoFisher), 50 U ml−1 penicillin and 50 mg
789 ml−1 streptomycin, supplemented with SMAD inhibitors (DMH-1 2 μm; SB431542 10 μm;
790 and CHIR99021 3 μm). This medium was changed every day for 6 days, and on day 7 the
791 medium was replaced for neural medium supplemented with DMH-1 2 μm, SB431542 10 μm
792 and CHIR 1 μm, All-Trans Retinoic Acid 0.1 μm (RA), and Purmorphamine 0.5 μm (PMN),
793 the cells were kept in this medium until day 12 when it was possible to observe a uniform
794 neuroepithelial sheet. At this point the cells were split 1:6 with Accutase (Gibco), onto
795 matrigel substrate in the presence of 10 μm of ROCK inhibitor (Y-27632 dihydrochloride,
796 Tocris), giving rise to a sheet of neural progenitor cells (NPC). After 24 hours of incubation
797 the medium was changed to neural medium supplemented with RA 0.5 μm and PMN 0.1
798 μm, and the medium was changed every day for 6 more days. On day 19 motor neuron
799 progenitors were split with accutase onto to matrigel-coated plates and the medium was
800 replaced with neural medium supplemented with RA 0.5 μm, PMN 0.1 μm, compound E 0.1
801 μm (Cpd E, Tocris), BDNF 10ng/mL, CNTF 10ng/mL and IGF 10ng/mL until day 28. On day
802 29, the media was replaced with neuronal media (neurobasal media supplemented with 1%
803 of B27, BDNF 10ng/mL, CNTF 10ng/mL and IGF 10ng/mL). The cells were then fed on
804 alternate days with neuronal medium until day 40.

805 ***Measurement of gene expression in iPSC-derived MNs***

806 Gene expression was determined via quantitative real-time polymerase chain reaction
807 (qRT-PCR). Primer sequences were as follows: *FGL2: Fwd:*
808 *TCAAAGTGTCCCAGCCAAGAA, Rvs: ACTCTGTAGGTCTCACTGCTTC. FAM185BP: Fwd:*
809 *TCAGGGTCTGGCTGTGTAGA, Rvs: TCCCAGACAGATCACTTTGCC. CCDC146: Fwd:*
810 *ACTCCACAAAGCTCATCAGAAAG, Rvs: TCCTGAAGTTTTGATACAATTTTGC.* RNA was
811 extracted using a Direct-zol RNA Purification Kits (Zymo) according to the manufacturer's
812 instructions. 2μg of total RNA was converted to cDNA in an initial reaction mixture containing
813 1μL 10mM dNTPs, 1μL 40μm random hexamer primer (ThermoFisher Scientific), and
814 DNAse/RNAse-free water to a total volume of 14μL. This mixture was heated for 5 minutes
815 at 70°C then incubated on ice for 5 minutes. 4μL of 5x FS buffer, 2μL 0.1M DTT, and 1μL
816 M-MLV reverse transcriptase (200U/μL) (ThermoFisher Scientific) were then added, and
817 cDNA conversion was performed in a PCR thermocycler (37°C for 60 minutes, 85°C for 10
818 minutes). cDNA was amplified using RT-PCR with Brilliant III SYBR Green (Agilent) as per
819 the manufacturer's instructions. Ct analysis was performed using CFX Maestro software
820 (BioRad). β-actin was chosen as a reference gene because expression is stable in
821 iPSC-derived neurons[87]. qPCR products were sequenced to confirm target specificity and
822 sequence traces are available on request. Statistical analysis was performed using
823 GraphPad Prism version 10.0.0.

824 ***Measurement of ALS phenotypes in iPSC-derived MNs***

825 Survival of iPSC-derived MN was assessed from day 28-40 using an Incucyte® Live Cell
826 Analysis System (Sartorius) at 20x magnification across 3 biological repeats. Neurite length

827 and number of neurite branch-points were used as readouts for cell survival. For each
828 biological repeat, 10 wells were imaged per group with 9 fields of view per well. The
829 neurotrack module (Sartorius) was used to determine neurite metrics from brightfield
830 images. In each case statistical significance was calculated by applying an unpaired *t*-test at
831 each timepoint. Statistical analysis was performed using GraphPad Prism version 10.0.0.

832 ***Immunocytochemistry***

833 For immunostaining, neural progenitor cells (NPC) and motor neurons (MN) were washed
834 with phosphate-buffered saline (PBS) and fixed with 4% paraformaldehyde for 10 min at
835 room temperature. After fixation samples were washed three times with PBS and
836 permeabilized with 0.3% Triton X-100 diluted in PBS for 5 min. The cells were subsequently
837 blocked in 5% Donkey serum (Millipore) for 1h at room temperature. After blocking, cell
838 cultures were incubated with the appropriate primary antibodies (guinea pig anti-MAP2
839 1:1000 [Synaptic Systems #188004]; chicken anti-TUJ1 1:1000 [Merck Millipore #AB9354];
840 mouse anti-SMI32 1:1000 [Biolegend #801701]; rabbit anti-TDP-43 1:100 [Cell Signaling
841 Technologies #89789S] diluted in PBS containing 5% of DS overnight. Cells were then
842 washed with PBS three times. Fluorescent secondary antibodies (Alexa Fluor 488, 555, 594
843 or 647, diluted 1:400 with DS) (ThermoFisher Scientific #A-21202, #A-21432, A-21450,
844 #A-32744, #A-21206) were subsequently added to the cells and incubated for 1h. The
845 samples were washed with PBS three more times and incubated with Hoechst 33342
846 (Invitrogen) for nuclear staining for 5 minutes. All experiments included cultures where the
847 primary antibodies were not added; non-specific staining was not observed in these negative
848 controls. Images were obtained using an Opera Phenix™ High Content Screening System at
849 40x magnification. Images were analyzed using the Harmony™ Image analysis system. 405,
850 488, 594, and 647 nm lasers along with the appropriate excitation and emission filters were
851 used. These settings were kept consistent while taking images from all separate
852 differentiations.

853 At day 40, cells were fixed in 4% paraformaldehyde (in PBS) for assessment of TDP-43
854 proteinopathy via immunocytochemistry. TDP-43 localization was determined by nuclear and
855 perinuclear TDP-43 intensity (Arbitrary Fluorescence Unit). Neurons were designated as
856 exhibiting mislocalized TDP-43 when nuclear TDP-43 staining in the lowest 10-percentile per
857 field-of-view was coincident with perinuclear TDP-43 staining in the highest 50-percentile for
858 the same field of view. Thresholds were designed in a blinded fashion to prevent overfitting.
859 For all analyses at least 25 fields were randomly selected and scanned per well of a 96-well
860 plate in triplicate. Statistics were calculated per field-of-view.

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

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

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

3D-iN longitudinal survival tracking

3D-iNs were allowed to mature until Day 30 after dox and were then transduced with a *SYN1*::GFP lentivirus to enable labeling of individual neurons as previously described[53,56]. 72-hours prior to the start of the survival assay (Day 37 after Dox), 3D-iNs were treated with ASO gapmers designed using an in-house strategy; 10µm of either a scrambled or *CCDC146* ASO were administered via gymnotic uptake. On Day 39 (24 hours prior to start of tracking), 3D-iNs were embedded in 100% Matrigel (Corning) for 30 minutes to allow gelling. MM was replenished after the 3D-iNs were fixed. Following a baseline image at Day 0 of the tracking (Day 40 of differentiation), glutamate (L-Glutamic acid monosodium salt hydrate; 5mM Milipore) was added to the medium and replenished every other day thereafter. Continuous z-stacks starting from the 3D-iN surface were captured every other day (from Day 0) using a Zeiss AxioZoom.v16 wide-field upright fluorescence microscope. Image processing and analysis were performed using ImageJ software (NIH).

Quantitative real-time PCR

Total RNA from iN neuronal cultures was extracted using the RNA Extraction Kit (Qiagen) and reverse transcribed using NEB ProtoScript II First Strand Synthesis Kit (NEB). RNA concentration and quality was checked via NanoDrop 1000 (Thermofisher). Real-time PCR was conducted using the iTaq Universal SYBR Green Supermix (Bio-Rad) with primers (*Fwd: CTGCTCTTGAAAGTGTCCTGGG*, *Rvs: CTTCCTTGTGAAGCTCTCGCCT*; sequences from Origene). For ASO testing , N = 4 independent cultures of control iNs were used.

Small-molecule drug search

We searched chemical activity databases ChEMBL[88] and PubChem[89] for CCDC146 and its similar proteins, but did not obtain any active compounds for these proteins. Here, similar proteins were defined as those with at least 50% sequence identities to the canonical sequence of human CCDC146 protein, provided by the UniProt database[90]. We also searched patent databases including WIPO (https://patentscope.wipo.int/), EPO (https://www.epo.org/en/searching-for-patents) and USPTO

(https://www.uspto.gov/patents/search), using search terms “CCDC146” and “coiled-coil domain-containing protein 146”, to confirm that there is currently no available drugs for this target.

Protein structure and binding site prediction

The predicted protein structure of CCDC146 protein was obtained from the AlphaFold Protein Structure Database[91] (https://alphafold.ebi.ac.uk/). To evaluate the druggability of CCDC146, we predicted potential ligand binding sites on its surface and obtained negative prediction results using P2Rank[74], which is an accurate ligand binding site prediction tool.

Extended Data Figures

Extended Data Fig. 1 | Additional results for E-P links and NeuroNet prediction. a, Gene ontology (GO) enrichment for genes with short-range enhancers in iMNs. GO terms with adjusted $P < 0.1$ (Benjamini-Hochberg correction) are shown. *P*-value by two-sided Fisher’s exact test. **b-d**, Pearson correlation between NeuroNet and DeepSEA predictions for H3K4me3 (**b**), H3K27ac (**c**), and ATAC (**d**), respectively. *P*-value by two-sided Fisher’s exact test. Neuron- and brain-related tissues and cell types are highlighted. PCC, Pearson correlation coefficient.

Extended Data Fig. 2 | Additional results for rare variant burden analysis. a, Regulatory relationships between chr2:51,259,489:G>C and target genes. Peak signals from MNs derived from the donor CS14 were visualized. Significant genes in NeuroNet-powered survival analysis are highlighted in red. **b**, Quantile-quantile (Q-Q) plot of a standard rare variant burden-based survival analysis without NeuroNet-based variant filtering. Genes sharing the same set of variants were collapsed into groups. The diagonal line represents the expected $-\log_{10}$(*P*-value) under the null hypothesis. The gray shaded areas represent the 95% confidence intervals under the null hypothesis. **c**, Overlap of genes between survival and risk analysis with different *P*-value cutoffs. Overlap was examined using two-sided Fisher’s exact test. The dashed gray line represents odds ratio = 1. CI, confidence interval.

Extended Data Fig. 3 | CRISPR/Cas9-mediated modeling of chr7:76,009,472:C>T exacerbates ALS-associated phenotypes in iPSC-derived *C9ORF72*-ALS MNs. a, Representative images of iPSC-derived MNs (top panels: isogenic control; bottom panels: edited) studied at day 40 post-iPSC-stage when they expressed a battery of neuronal markers, including SMI-32, MAP2 and TUJ1. Scale bar: 50μm. **b**, Quantification of number of neurite branch-points in an age- and sex-matched normal control, an isogenic control carrying a G4C2-repeat expansion of *C9ORF72*, and chr7:76,009,472:C>T-edited iMNs across 12 days of tracking using an Incucyte live-cell analysis system (Sartorius). Error bars represent standard error. **c**, Representative images from day 1 and day 11 for isogenic control iMNs carrying a G4C2-repeat expansion of *C9ORF72* (top panels) and chr7:76,009,472:C>T-edited iMNs (bottom panels). Cell soma are coloured yellow and neurites are coloured purple. Scale bar: 1.6mm.

Extended Data Fig. 4 | Suppression of *CCDC146* in ALS patient-derived neurons via antisense oligonucleotide (ASO) rescues survival deficits irrespective of baseline *CCDC146* expression and genetic backgrounds. a-e, Kaplan-Meier (KM) estimator plots of 3D-iN survival starting (day 0) from day 40 after initiation of iPSC differentiation. $N = 2$ donors; $N = 200$ neurons, $N = 4$ spheroids per condition, per donor. Comparisons were

performed between ALS (*C9ORF72* and sporadic) and control iNs with scrambled ASO (**a**, **b**), and between *CCDC146*-targeting ASO and scrambled ASO for ALS (*C9ORF72* and sporadic) iNs (**c**, **d**) and control iNs (**e**), respectively. *P*-value by Mantel-Cox log-rank test. **, $P < 0.01$; ***, $P < 0.001$; NS, not significant. **f**, Representative images of control iNs treated with the scrambled ASO (top panels) or the *CCDC146*-targeting ASO (bottom panels). Images were taken from approximately the same location on the plat at Day 0 (left panels) and Day 8 (right panels) showing no obvious toxicity. Scale bar: 200μm. Ctrl, control. **g**, Quantification of *CCDC146* mRNA expression via q-RT-PCR in $N = 1$ control line and $N = 8$ ALS patient-derived iN lines. $N = 3$ technical replicates per cell line. The red dashed line indicates control expression level. *P*-value by the Dunnett's test. **, adjusted $P < 0.01$; ****, adjusted $P < 0.0001$; NS, not significant; #, patient cell lines used in ASO experiments.

Extended Data Fig. 5 | The AlphaFold-predicted three-dimensional structure of CCDC146 protein. **a**, The carton illustration of the predicted protein structure of CCDC146, obtained from https://alphafold.ebi.ac.uk/entry/Q8IYE0. pLDDT, predicted local distance difference test. **b**, Protein surface of the predicted structure, illustrated using PyMol (https://pymol.org/).

Supplementary Tables

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

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

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

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

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

References

1. Feldman, E. L. *et al.* Amyotrophic lateral sclerosis. *Lancet* **400**, 1363–1380 (2022).
2. Chiò, A. *et al.* Prognostic factors in ALS: A critical review. *Amyotroph. Lateral Scler.* **10**, 310–323 (2009).
3. Cooper-Knock, J. *et al.* Clinico-pathological features in amyotrophic lateral sclerosis with expansions in C9ORF72. *Brain* **135**, 751–764 (2012).
4. Al-Chalabi, A., van den Berg, L. H. & Veldink, J. Gene discovery in amyotrophic lateral sclerosis: implications for clinical management. *Nat. Rev. Neurol.* **13**, 96–104 (2017).
5. Cooper-Knock, J. *et al.* Advances in the genetic classification of amyotrophic lateral sclerosis. *Curr. Opin. Neurol.* **34**, 756–764 (2021).
6. Li, X. *et al.* Dynamic incorporation of multiple in silico functional annotations empowers rare variant association analysis of large whole-genome sequencing studies at scale. *Nat. Genet.* **52**, 969–983 (2020).
7. Zhang, F. & Lupski, J. R. Non-coding genetic variants in human disease. *Hum. Mol. Genet.* **24**, R102–10 (2015).

8. Almeida, N. *et al.* Employing core regulatory circuits to define cell identity. *EMBO J.* **40**, e106785 (2021).
9. GTEx Consortium *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
10. Li, X. *et al.* The impact of rare variation on gene expression across tissues. *Nature* **550**, 239–243 (2017).
11. Zhou, J. & Troyanskaya, O. G. Predicting effects of noncoding variants with deep learning–based sequence model. *Nat. Methods* **12**, 931–934 (2015).
12. Zhou, J. *et al.* Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. *Nat. Genet.* **50**, 1171–1179 (2018).
13. Kelley, D. R. *et al.* Sequential regulatory activity prediction across chromosomes with convolutional neural networks. *Genome Res.* **28**, 739–750 (2018).
14. Avsec, Ž. *et al.* Effective gene expression prediction from sequence by integrating long-range interactions. *Nat. Methods* **18**, 1196–1203 (2021).
15. Yuan, H. & Kelley, D. R. scBasset: sequence-based modeling of single-cell ATAC-seq using convolutional neural networks. *Nat. Methods* **19**, 1088–1096 (2022).
16. Agarwal, V. & Shendure, J. Predicting mRNA Abundance Directly from Genomic Sequence Using Deep Convolutional Neural Networks. *Cell Rep.* **31**, 107663 (2020).
17. Muroňová, J. *et al.* Lack of CCDC146, a ubiquitous centriole and microtubule-associated protein, leads to non-syndromic male infertility in human and mouse. *bioRxiv* 2023.02.27.530236 (2023) doi:10.1101/2023.02.27.530236.
18. Fulco, C. P. *et al.* Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019).
19. Zhang, S. *et al.* Genome-wide identification of the genetic basis of amyotrophic lateral sclerosis. *Neuron* **110**, 992–1008.e11 (2022).
20. Project MinE: study design and pilot analyses of a large-scale whole-genome sequencing study in amyotrophic lateral sclerosis. *Eur. J. Hum. Genet.* **26**, 1537–1546 (2018).
21. Bendl, J. *et al.* The three-dimensional landscape of cortical chromatin accessibility in Alzheimer's disease. *Nat. Neurosci.* **25**, 1366–1378 (2022).
22. Ivansson, E. L. *et al.* Variants within the SP110 nuclear body protein modify risk of canine degenerative myelopathy. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E3091–100 (2016).
23. Chen, M. *et al.* Emerging roles of activating transcription factor (ATF) family members in tumourigenesis and immunity: Implications in cancer immunotherapy. *Genes Dis* **9**, 981–999 (2022).
24. Ryu, H. *et al.* Sp1 and Sp3 are oxidative stress-inducible, antideath transcription factors in cortical neurons. *J. Neurosci.* **23**, 3597–3606 (2003).
25. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovations* **2**, (2021).
26. Milioto, C. *et al.* PolyGR and polyPR knock-in mice reveal a conserved neuroprotective extracellular matrix signature in C9orf72 ALS/FTD neurons. *Nat. Neurosci.* (2024) doi:10.1038/s41593-024-01589-4.
27. Ohno, M. *et al.* Nardilysin regulates axonal maturation and myelination in the central and peripheral nervous system. *Nat. Neurosci.* **12**, 1506–1513 (2009).
28. Szklarczyk, D. *et al.* The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**, D638–D646 (2023).
29. Lee, J. K., Shin, J. H., Lee, J. E. & Choi, E.-J. Role of autophagy in the pathogenesis of

amyotrophic lateral sclerosis. *Biochim. Biophys. Acta* **1852**, 2517–2524 (2015).
30. Moll, T. *et al.* Membrane lipid raft homeostasis is directly linked to neurodegeneration. *Essays Biochem.* **65**, 999–1011 (2021).
31. Richter, F. *et al.* Genomic analyses implicate noncoding de novo variants in congenital heart disease. *Nat. Genet.* **52**, 769–777 (2020).
32. Chen, K. M., Cofer, E. M., Zhou, J. & Troyanskaya, O. G. Selene: a PyTorch-based deep learning library for sequence data. *Nat. Methods* **16**, 315–318 (2019).
33. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
34. Currin, K. W. *et al.* Genetic effects on liver chromatin accessibility identify disease regulatory variants. *Am. J. Hum. Genet.* **108**, 1169–1189 (2021).
35. Kumar, S., Ambrosini, G. & Bucher, P. SNP2TFBS - a database of regulatory SNPs affecting predicted transcription factor binding site affinity. *Nucleic Acids Res.* **45**, D139–D144 (2017).
36. Davydov, E. V. *et al.* Identifying a High Fraction of the Human Genome to be under Selective Constraint Using GERP++. *PLoS Comput. Biol.* **6**, e1001025 (2010).
37. Ritchie, G. R. S., Dunham, I., Zeggini, E. & Flicek, P. Functional annotation of noncoding sequence variants. *Nat. Methods* **11**, 294–296 (2014).
38. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
39. Huang, Y.-F., Gulko, B. & Siepel, A. Fast, scalable prediction of deleterious noncoding variants from functional and population genomic data. *Nat. Genet.* **49**, 618–624 (2017).
40. Lee, S. *et al.* Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am. J. Hum. Genet.* **91**, 224–237 (2012).
41. Zhan, X., Hu, Y., Li, B., Abecasis, G. R. & Liu, D. J. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics* **32**, 1423–1426 (2016).
42. Nicolas, A. *et al.* Genome-wide Analyses Identify KIF5A as a Novel ALS Gene. *Neuron* **97**, 1268–1283.e6 (2018).
43. Baron, D. M. *et al.* ALS-associated KIF5A mutations abolish autoinhibition resulting in a toxic gain of function. *Cell Rep.* **39**, 110598 (2022).
44. Li, C. *et al.* Genome-wide analyses identify NEAT1 as genetic modifier of age at onset of amyotrophic lateral sclerosis. *Mol. Neurodegener.* **18**, 77 (2023).
45. Baxi, E. G. *et al.* Answer ALS, a large-scale resource for sporadic and familial ALS combining clinical and multi-omics data from induced pluripotent cell lines. *Nat. Neurosci.* **25**, 226–237 (2022).
46. Cedarbaum, J. M. *et al.* The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. *J. Neurol. Sci.* **169**, 13–21 (1999).
47. Proudfoot, M., Jones, A., Talbot, K., Al-Chalabi, A. & Turner, M. R. The ALSFRS as an outcome measure in therapeutic trials and its relationship to symptom onset. *Amyotroph. Lateral Scler. Frontotemporal Degener.* **17**, 414–425 (2016).
48. Kuriyama, S., Tamiya, Y. & Tanaka, M. Spatiotemporal expression of UPK3B and its promoter activity during embryogenesis and spermatogenesis. *Histochem. Cell Biol.* **147**, 17–26 (2017).
49. Shi, Y. *et al.* Haploinsufficiency leads to neurodegeneration in C9ORF72 ALS/FTD human induced motor neurons. *Nat. Med.* **24**, 313–325 (2018).
50. Fiesel, F. C., Schurr, C., Weber, S. S. & Kahle, P. J. TDP-43 knockdown impairs neurite

outgrowth dependent on its target histone deacetylase 6. *Mol. Neurodegener.* **6**, 64 (2011).
51. Klim, J. R. *et al.* ALS-implicated protein TDP-43 sustains levels of STMN2, a mediator of motor neuron growth and repair. *Nature Neuroscience* vol. 22 167–179 Preprint at https://doi.org/10.1038/s41593-018-0300-4 (2019).
52. Neumann, M. *et al.* Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314**, 130–133 (2006).
53. Linares, G. R. *et al.* SYF2 suppression mitigates neurodegeneration in models of diverse forms of ALS. *Cell Stem Cell* **30**, (2023).
54. Shaw, P. J. & Ince, P. G. Glutamate, excitotoxicity and amyotrophic lateral sclerosis. *J. Neurol.* **244 Suppl 2**, S3–14 (1997).
55. Van Den Bosch, L., Van Damme, P., Bogaert, E. & Robberecht, W. The role of excitotoxicity in the pathogenesis of amyotrophic lateral sclerosis. *Biochim. Biophys. Acta* **1762**, 1068–1082 (2006).
56. Bowles, K. R. *et al.* ELAVL4, splicing, and glutamatergic dysfunction precede neuron loss in MAPT mutation cerebral organoids. *Cell* **184**, (2021).
57. Baysoy, A., Bai, Z., Satija, R. & Fan, R. The technological landscape and applications of single-cell multi-omics. *Nat. Rev. Mol. Cell Biol.* **24**, 695–713 (2023).
58. Yadav, A. *et al.* A cellular taxonomy of the adult human spinal cord. *Neuron* **111**, 328–344.e7 (2023).
59. Gautier, O. *et al.* Human motor neurons are rare and can be transcriptomically divided into known subtypes. *bioRxiv* (2023) doi:10.1101/2023.04.05.535689.
60. Workman, M. J. *et al.* Large-scale differentiation of iPSC-derived motor neurons from ALS and control subjects. *Neuron* **111**, 1191–1204.e5 (2023).
61. Diekstra, F. P. *et al.* UNC13A is a modifier of survival in amyotrophic lateral sclerosis. *Neurobiol. Aging* **33**, 630.e3–8 (2012).
62. Li, M. J. *et al.* GWASdb v2: an update database for human genetic variants identified by genome-wide association studies. *Nucleic Acids Res.* **44**, D869–76 (2016).
63. Cuttler, K., Hassan, M., Carr, J., Cloete, R. & Bardien, S. Emerging evidence implicating a role for neurexins in neurodegenerative and neuropsychiatric disorders. *Open Biol.* **11**, 210091 (2021).
64. Miller, T. M. *et al.* Trial of Antisense Oligonucleotide Tofersen for SOD1 ALS. *N. Engl. J. Med.* **387**, 1099–1110 (2022).
65. Smith, B. N. *et al.* Exome-wide rare variant analysis identifies TUBA4A mutations associated with familial ALS. *Neuron* **84**, 324–331 (2014).
66. Yamashita, S. *et al.* Amyotrophic lateral sclerosis in a patient with Kartagener syndrome. *Amyotroph. Lateral Scler.* **11**, 402–404 (2010).
67. De Decker, M. *et al.* C21orf2 mutations found in ALS disrupt primary cilia function. *bioRxiv* 2022.02.28.482239 (2022) doi:10.1101/2022.02.28.482239.
68. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
69. Uhlén, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
70. Humphrey, J. *et al.* Integrative transcriptomic analysis of the amyotrophic lateral sclerosis spinal cord implicates glial activation and suggests new risk genes. *Nat. Neurosci.* **26**, 150–162 (2023).
71. Nguyen, D.-T. *et al.* Pharos: Collating protein information to shed light on the druggable genome. *Nucleic Acids Res.* **45**, D995–D1002 (2017).

72. Kelleher, K. J. *et al.* Pharos 2023: an integrated resource for the understudied human proteome. *Nucleic Acids Res.* **51**, D1405–D1416 (2023).
73. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
74. Krivák, R. & Hoksza, D. P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. *J. Cheminform.* **10**, 39 (2018).
75. Bruse, N. & van Heeringen, S. J. GimmeMotifs: an analysis framework for transcription factor motif analysis. *bioRxiv* 474403 (2018) doi:10.1101/474403.
76. Krishnan, A. *et al.* Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder. *Nat. Neurosci.* **19**, 1454–1462 (2016).
77. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, P10008 (2008).
78. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 1–14 (2013).
79. Project MinE ALS Sequencing Consortium. Project MinE: study design and pilot analyses of a large-scale whole-genome sequencing study in amyotrophic lateral sclerosis. *Eur. J. Hum. Genet.* **26**, 1537–1546 (2018).
80. Brooks, B. R. El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. Subcommittee on Motor Neuron Diseases/Amyotrophic Lateral Sclerosis of the World Federation of Neurology Research Group on Neuromuscular Diseases and the El Escorial 'Clinical limits of amyotrophic lateral sclerosis' workshop contributors. *J. Neurol. Sci.* **124 Suppl**, 96–107 (1994).
81. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
82. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).
83. Ludwik, K. A., Telugu, N., Schommer, S., Stachelscheid, H. & Diecke, S. ASSURED-optimized CRISPR protocol for knockout/SNP knockin in hiPSCs. *STAR Protoc* **4**, 102406 (2023).
84. Varcianna, A. *et al.* Micro-RNAs secreted through astrocyte-derived extracellular vesicles cause neuronal network degeneration in C9orf72 ALS. *EBioMedicine* vol. 40 626–635 Preprint at https://doi.org/10.1016/j.ebiom.2018.11.067 (2019).
85. Castelli, L. M. *et al.* SRSF1-dependent inhibition of C9ORF72-repeat RNA nuclear export: genome-wide mechanisms for neuroprotection in amyotrophic lateral sclerosis. *Mol. Neurodegener.* **16**, 53 (2021).
86. Du, Z.-W. *et al.* Generation and expansion of highly pure motor neuron progenitors from human pluripotent stem cells. *Nat. Commun.* **6**, 6626 (2015).
87. Artyukhov, A. S. *et al.* New genes for accurate normalization of qRT-PCR results in study of iPS and iPS-derived cells. *Gene* **626**, 234–240 (2017).
88. Zdrazil, B. *et al.* The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Res.* **52**, D1180–D1192 (2024).
89. Kim, S. *et al.* PubChem 2023 update. *Nucleic Acids Res.* **51**, D1373–D1380 (2023).
90. UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **51**, D523–D531 (2023).
91. Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).

Figure 1

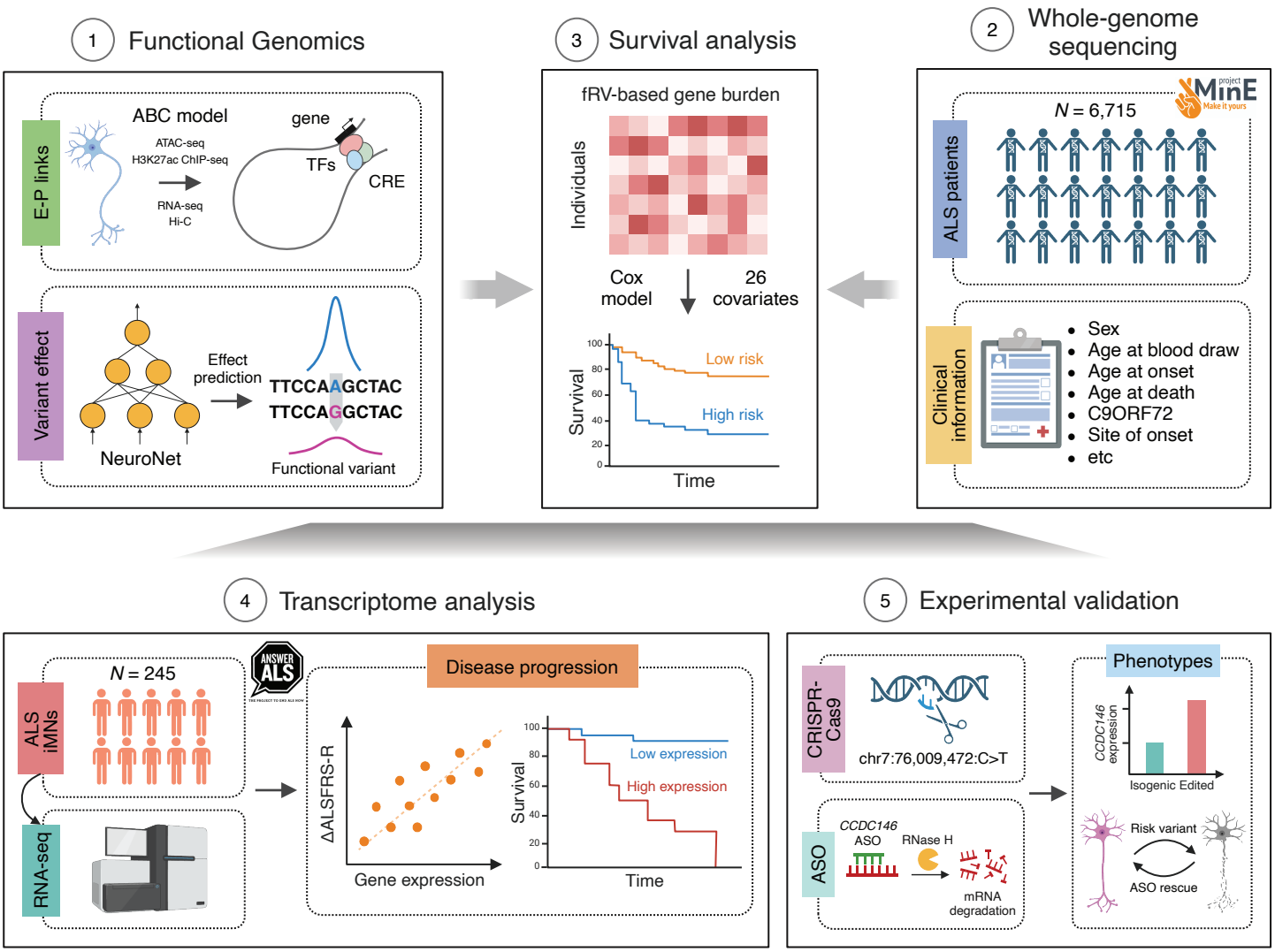

1 Functional Genomics
E-P links
ABC model
ATAC-seq
H3K27ac ChIP-seq
RNA-seq
Hi-C
gene
TFs
CRE
Variant effect
NeuroNet
Effect prediction
TTCCAAGCTAC
TTCCAGGCTAC
Functional variant
3 Survival analysis
fRV-based gene burden
Individuals
Cox model
26 covariates
Survival
Low risk
High risk
Time
2 Whole-genome sequencing
ALS patients
N = 6,715
Clinical information
Sex
Age at blood draw
Age at onset
Age at death
C9ORF72
Site of onset
etc
4 Transcriptome analysis
ALS iMNs
N = 245
RNA-seq
Disease progression
ΔALSFRS-R
Gene expression
Survival
Low expression
High expression
Time
5 Experimental validation
CRISPR-Cas9
chr7:76,009,472:C>T
ASO
CCDC146
ASO
RNase H
mRNA degradation
Phenotypes
CCDC146 expression
Isogenic Edited
Risk variant
ASO rescue

Figure 2

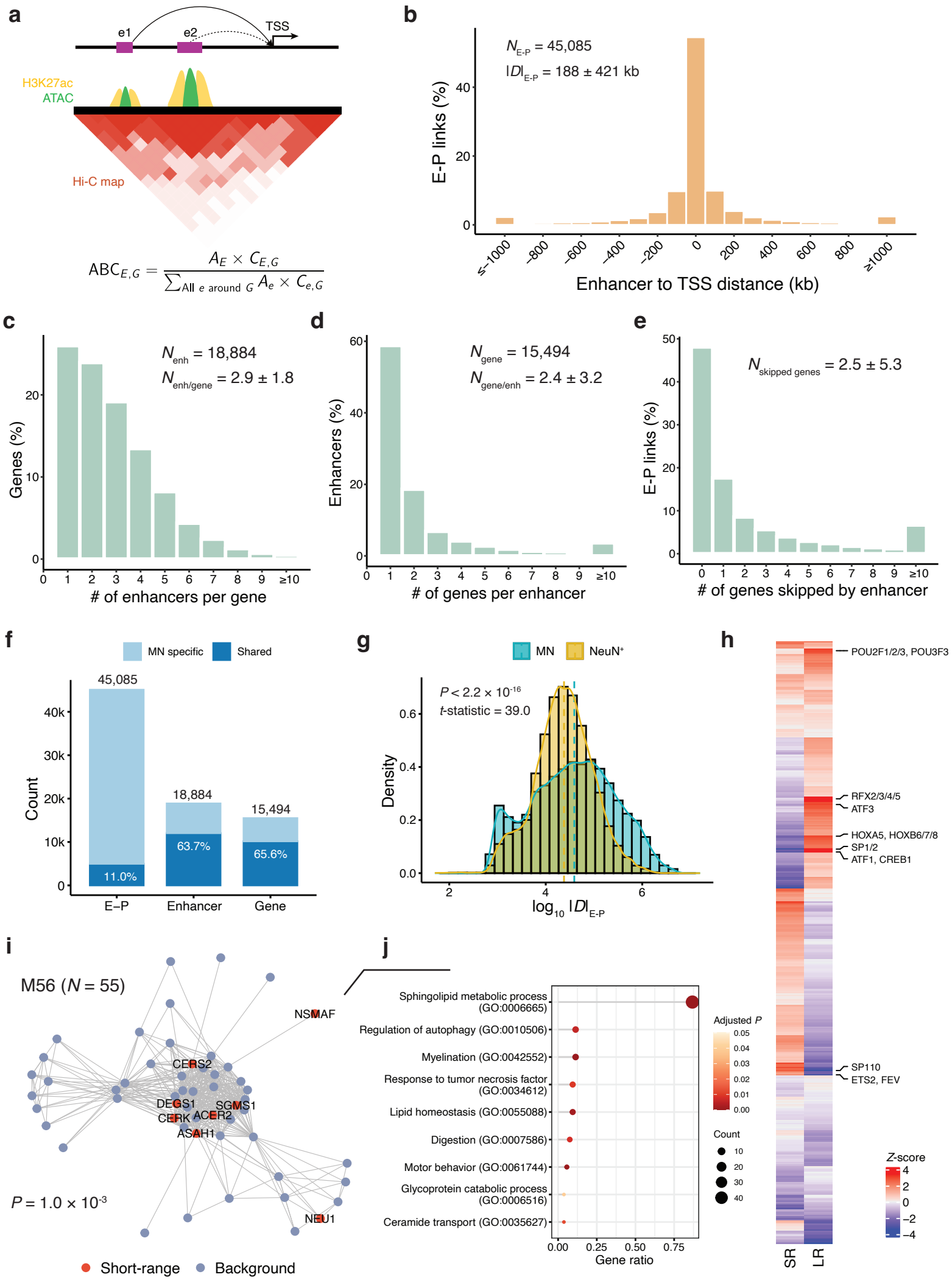

a
e1
e2
TSS
H3K27ac
ATAC
Hi-C map
$ABC_{E,G} = \frac{A_E \times C_{E,G}}{\sum_{\text{All } e \text{ around } G} A_e \times C_{e,G}}$
b
$N_{E\text{-}P}$ = 45,085
$|D|_{E\text{-}P}$ = 188 ± 421 kb
E-P links (%)
Enhancer to TSS distance (kb)
c
N_{enh} = 18,884
$N_{enh/gene}$ = 2.9 ± 1.8
Genes (%)
# of enhancers per gene
d
N_{gene} = 15,494
$N_{gene/enh}$ = 2.4 ± 3.2
Enhancers (%)
# of genes per enhancer
e
$N_{skipped\ genes}$ = 2.5 ± 5.3
E-P links (%)
# of genes skipped by enhancer
f
MN specific
Shared
45,085
18,884
15,494
11.0%
63.7%
65.6%
Count
E–P
Enhancer
Gene
g
MN
NeuN+
$P < 2.2 \times 10^{-16}$
t-statistic = 39.0
Density
$\log_{10} |D|_{E\text{-}P}$
h
POU2F1/2/3, POU3F3
RFX2/3/4/5
ATF3
HOXA5, HOXB6/7/8
SP1/2
ATF1, CREB1
SP110
ETS2, FEV
Z-score
SR
LR
i
M56 (N = 55)
NSMAF
CERS2
DEGS1
SGMS1
CERK
ACER2
ASAH1
NEU1
$P = 1.0 \times 10^{-3}$
Short-range
Background
j
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

Figure 3

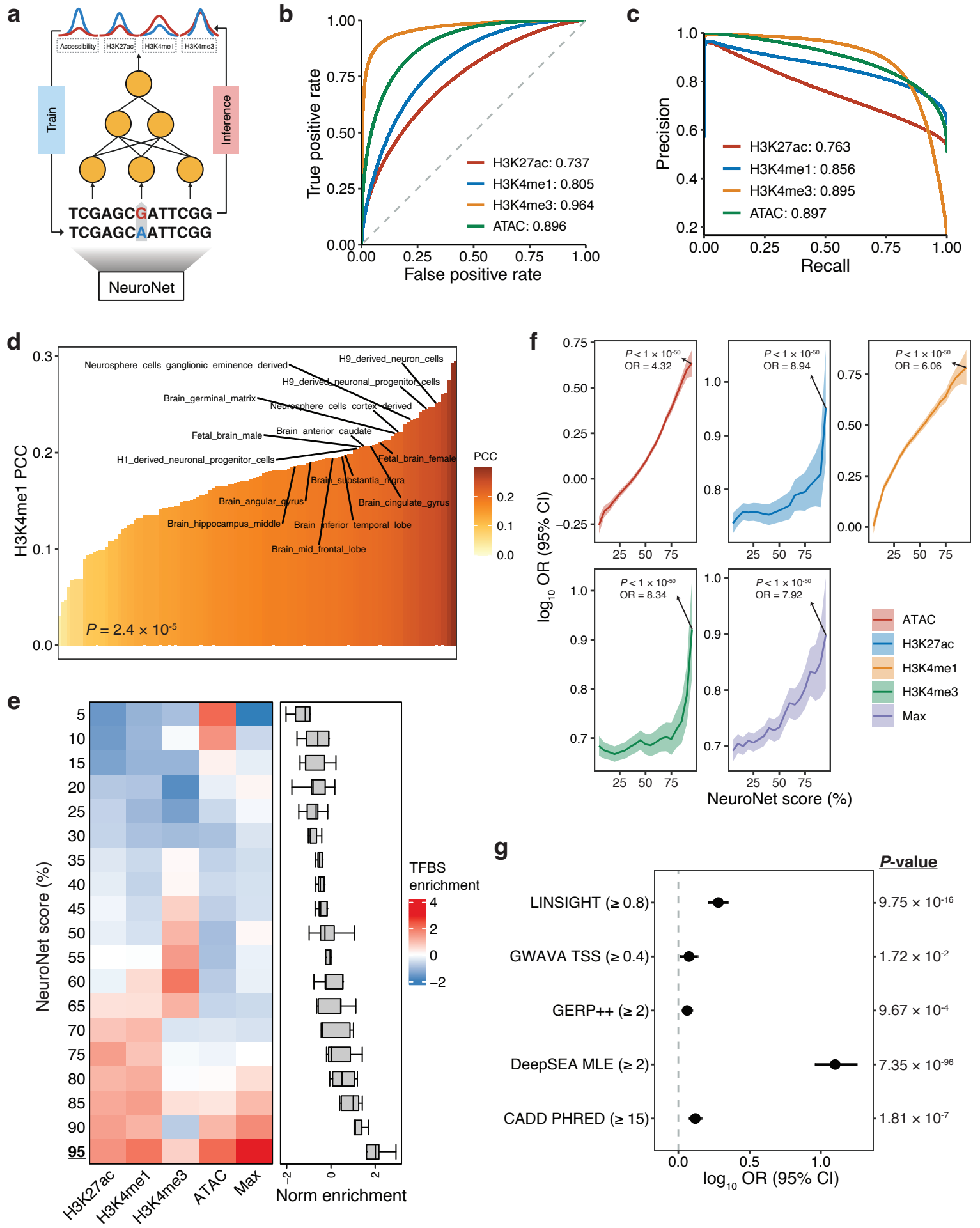

a
Accessibility
H3K27ac
H3K4me1
H3K4me3
Train
Inference
TCGAGCGATTCGG
TCGAGCAATTCGG
NeuroNet
b
True positive rate
False positive rate
H3K27ac: 0.737
H3K4me1: 0.805
H3K4me3: 0.964
ATAC: 0.896
c
Precision
Recall
H3K27ac: 0.763
H3K4me1: 0.856
H3K4me3: 0.895
ATAC: 0.897
d
H3K4me1 PCC
PCC
Neurosphere_cells_ganglionic_eminence_derived
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
P = 2.4 × 10-5
e
NeuroNet score (%)
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 score (%)
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)

Figure 4

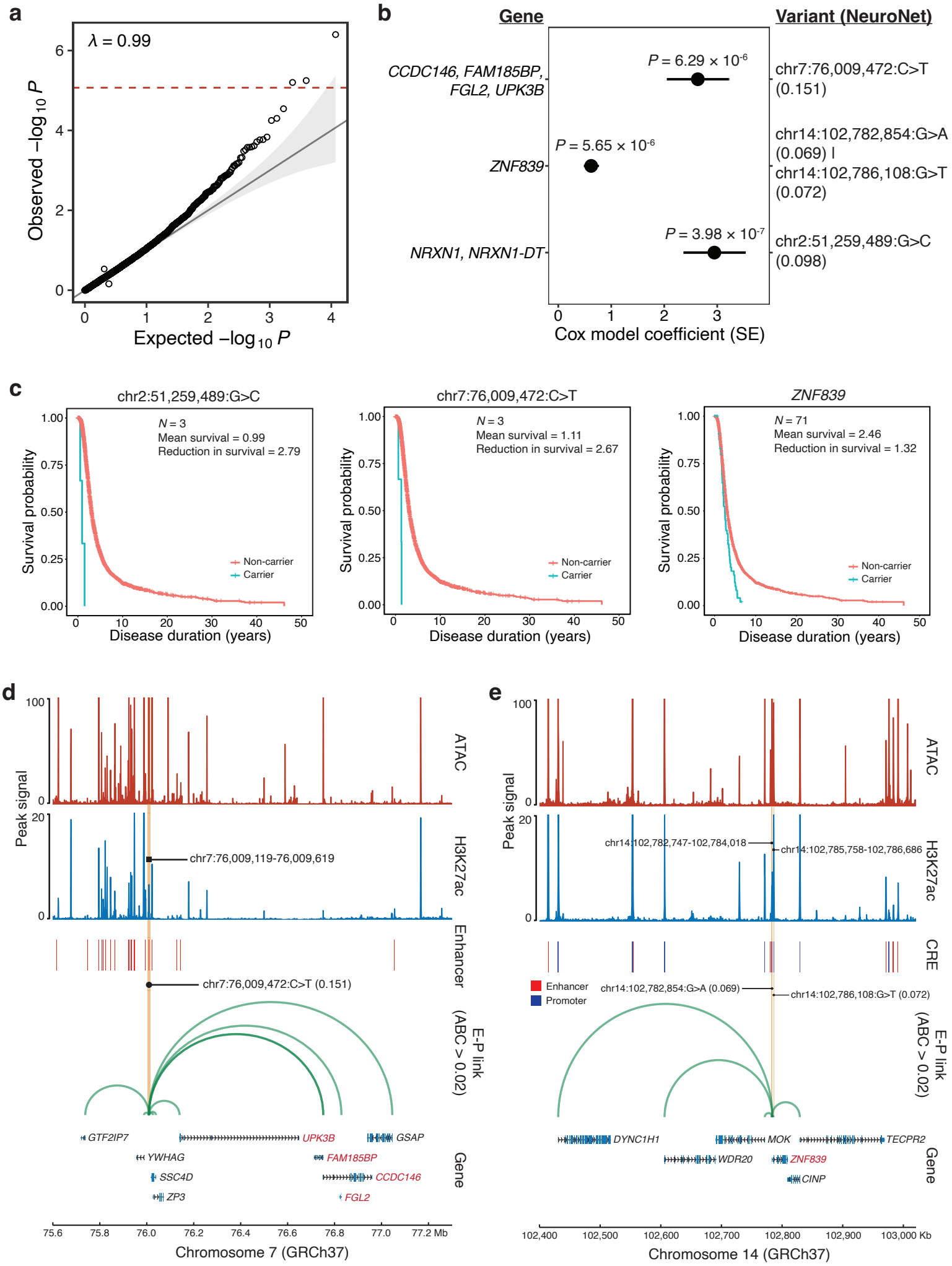

a
λ = 0.99
Observed −log10 P
Expected −log10 P
b
Gene
Variant (NeuroNet)
CCDC146, FAM185BP, FGL2, UPK3B
P = 6.29 × 10-6
chr7:76,009,472:C>T (0.151)
ZNF839
P = 5.65 × 10-6
chr14:102,782,854:G>A (0.069) |
chr14:102,786,108:G>T (0.072)
NRXN1, NRXN1-DT
P = 3.98 × 10-7
chr2:51,259,489:G>C (0.098)
Cox model 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)
GTF2IP7
UPK3B
GSAP
YWHAG
FAM185BP
SSC4D
CCDC146
ZP3
FGL2
Gene
Chromosome 7 (GRCh37)
e
chr14:102,782,747-102,784,018
chr14:102,785,758-102,786,686
CRE
Enhancer
Promoter
chr14:102,782,854:G>A (0.069)
chr14:102,786,108:G>T (0.072)
DYNC1H1
MOK
TECPR2
WDR20
ZNF839
CINP
Chromosome 14 (GRCh37)

Figure 5

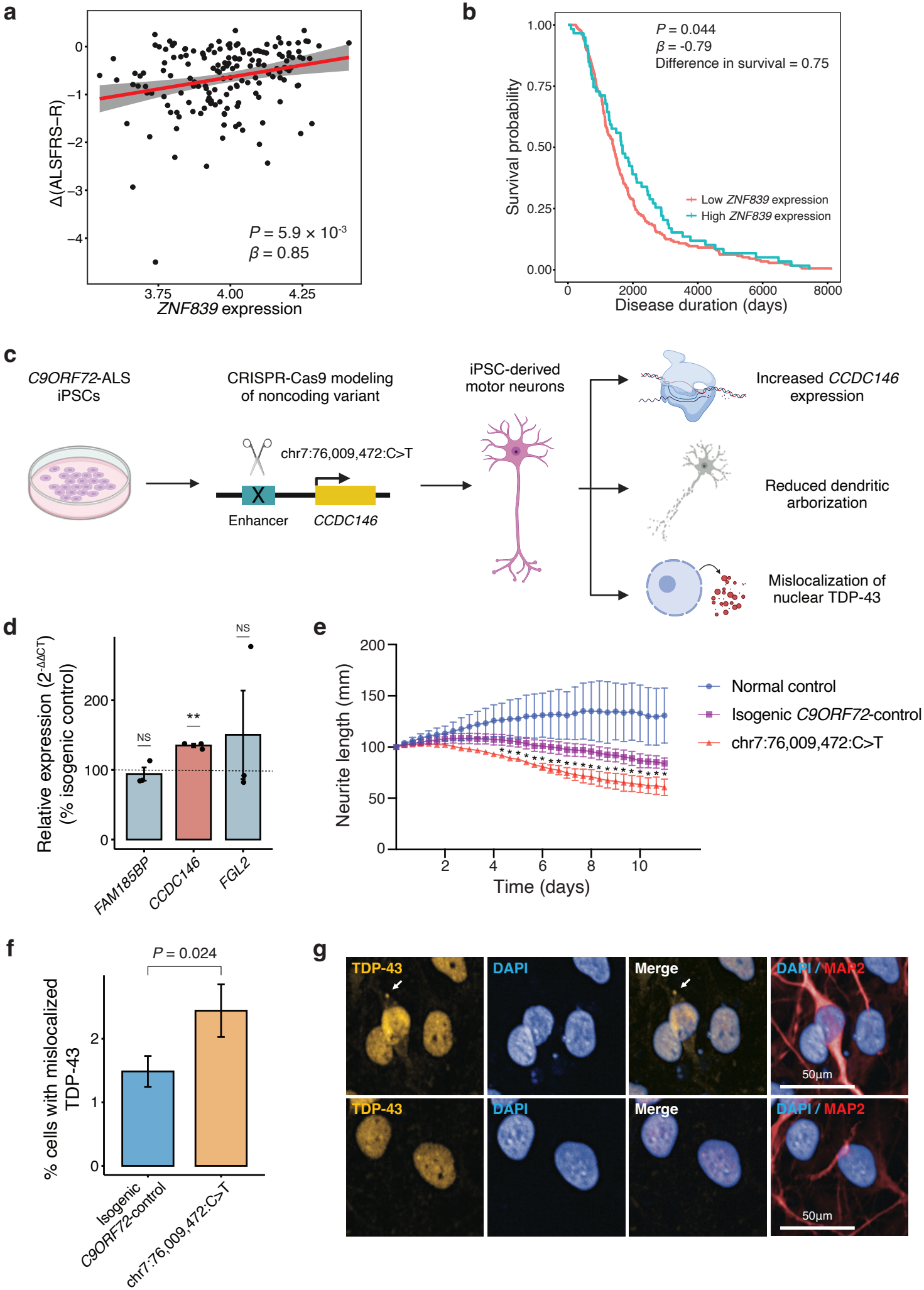
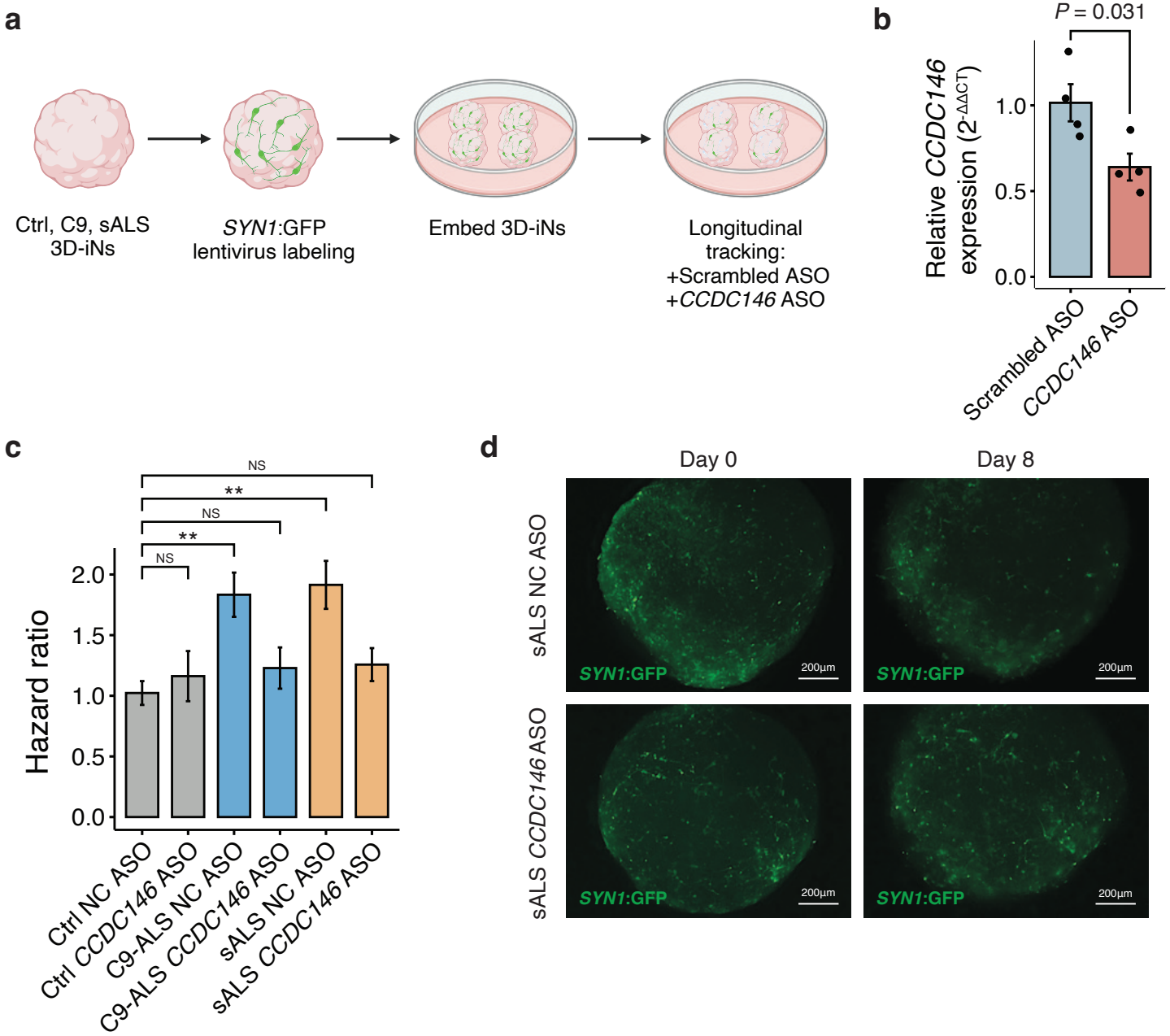

a
Δ(ALSFRS−R)
0
−1
−2
−3
−4
P = 5.9 × 10-3
β = 0.85
3.75
4.00
4.25
ZNF839 expression
b
Survival probability
1.00
0.75
0.50
0.25
0.00
P = 0.044
β = -0.79
Difference in survival = 0.75
Low ZNF839 expression
High ZNF839 expression
0
2000
4000
6000
8000
Disease duration (days)
c
C9ORF72-ALS iPSCs
CRISPR-Cas9 modeling of noncoding variant
chr7:76,009,472:C>T
X
Enhancer
CCDC146
iPSC-derived motor neurons
Increased CCDC146 expression
Reduced dendritic arborization
Mislocalization of nuclear TDP-43
d
Relative expression (2-ΔΔCT) (% isogenic control)
200
100
0
NS
**
NS
FAM185BP
CCDC146
FGL2
e
Neurite length (mm)
200
150
100
50
0
2
4
6
8
10
Time (days)
Normal control
Isogenic C9ORF72-control
chr7:76,009,472:C>T
f
% cells with mislocalized TDP-43
P = 0.024
2
1
0
Isogenic C9ORF72-control
chr7:76,009,472:C>T
g
TDP-43
DAPI
Merge
DAPI / MAP2
50µm
TDP-43
DAPI
Merge
DAPI / MAP2
50µm

Figure 6


a
Ctrl, C9, sALS 3D-iNs
SYN1:GFP lentivirus labeling
Embed 3D-iNs
Longitudinal tracking:
+Scrambled ASO
+CCDC146 ASO
b
P = 0.031
Relative CCDC146 expression (2-ΔΔCT)
1.0
0.5
0.0
Scrambled ASO
CCDC146 ASO
c
NS
**
NS
**
NS
Hazard ratio
2.0
1.5
1.0
0.5
0.0
Ctrl NC ASO
Ctrl CCDC146 ASO
C9-ALS NC ASO
C9-ALS CCDC146 ASO
sALS NC ASO
sALS CCDC146 ASO
d
Day 0
Day 8
sALS NC ASO
sALS CCDC146 ASO
SYN1:GFP
200µm