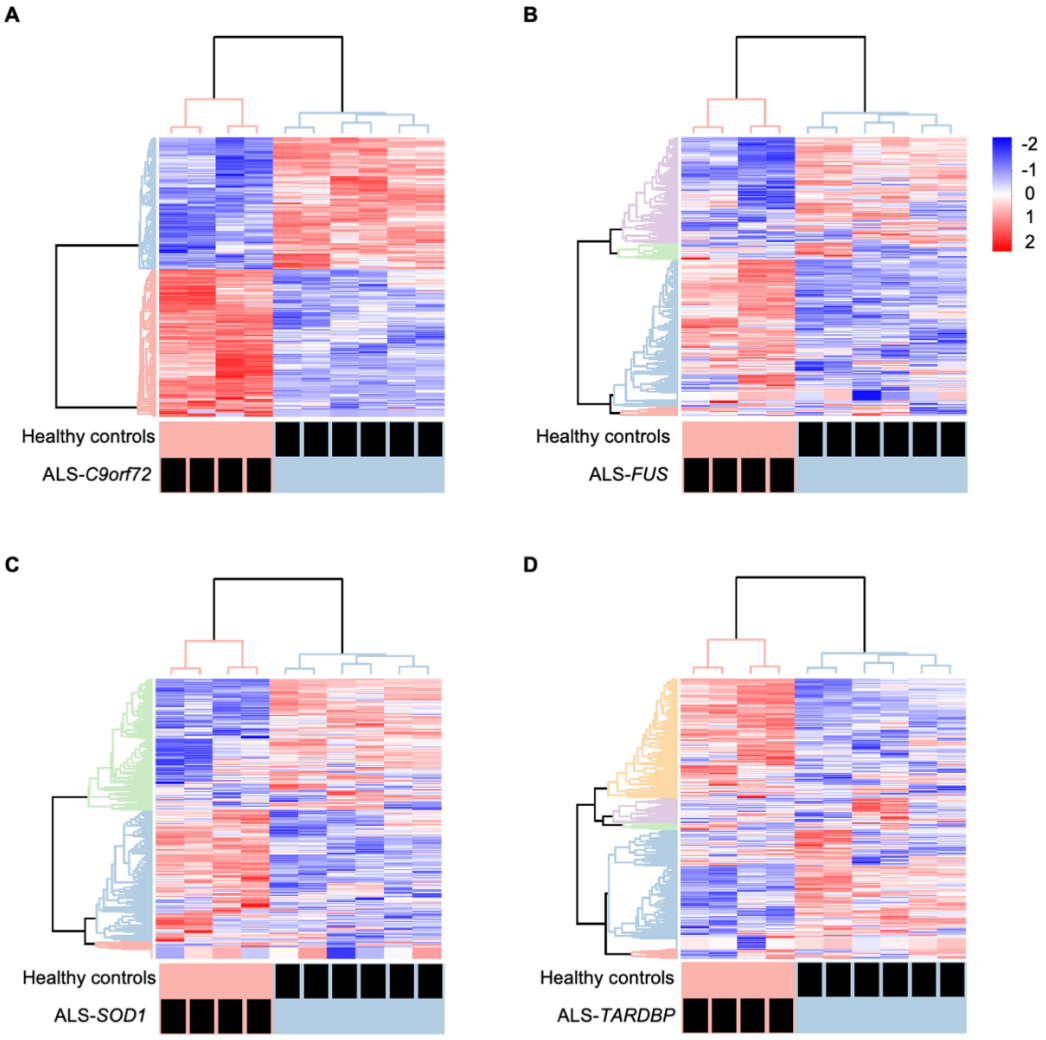
Multiomics and machine-learning identify novel transcriptional and mutational signatures in amyotrophic lateral sclerosis

Alberto Catanese[1,2,†], Sandeep Rajkumar[1], Daniel Sommer[1], Pegah Masrori[3,4,5], Nicole Hersmus[3,4,5], Philip Van Damme[3,4,5], Simon Witzel[6], Albert Ludolph[2,6], Ritchie Ho[7,8,9,10], Tobias M Boeckers[1,2], Medhanie Mulaw[11,†]

[1] Institute of Anatomy and Cell Biology, Ulm University School of Medicine, Ulm, Germany
[2] German Center for Neurodegenerative Diseases (DZNE), Ulm site, Ulm, Germany
[3] Laboratory of Neurobiology, Center for Brain & Disease Research, VIB, Leuven, Belgium.
[4] Department of Neurology, University Hospitals Leuven, Leuven, Belgium.
[5] Experimental Neurology, Department of Neurosciences, Leuven Brain Institute, KU Leuven, Leuven, Belgium.
[6] Department. of Neurology, Ulm University School of Medicine, Ulm, Germany
[7] Center for Neural Science and Medicine, Cedars-Sinai Medical Center, Los Angeles, CA, USA
[8] Board of Governors Regenerative Medicine Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA
[9] Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA, USA
[10] Department of Neurology, Cedars-Sinai Medical Center, Los Angeles, CA, USA
[11] Internal Medicine I and Institute of Molecular Medicine and Stem Cell Aging, Medical Faculty, University Hospital Ulm, University of Ulm University, Ulm, Germany.
[*] Co-senior authors

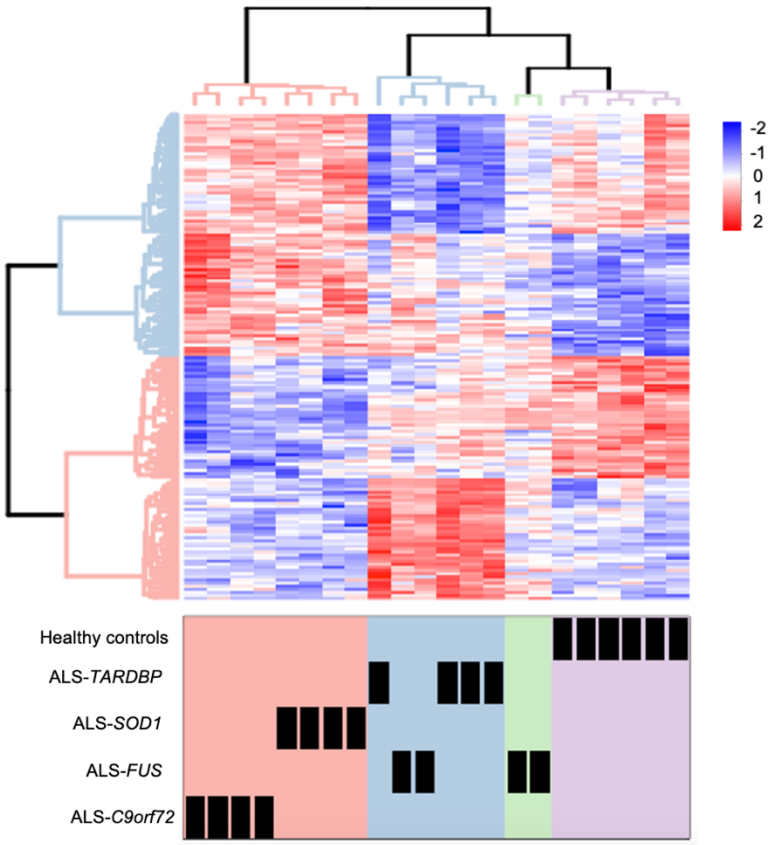
This file contains 13 Supplementary Figures and 5 Supplementary Tables



Supplementary Figure 1

Gene-centric transcriptome analysis of ALS-related MN

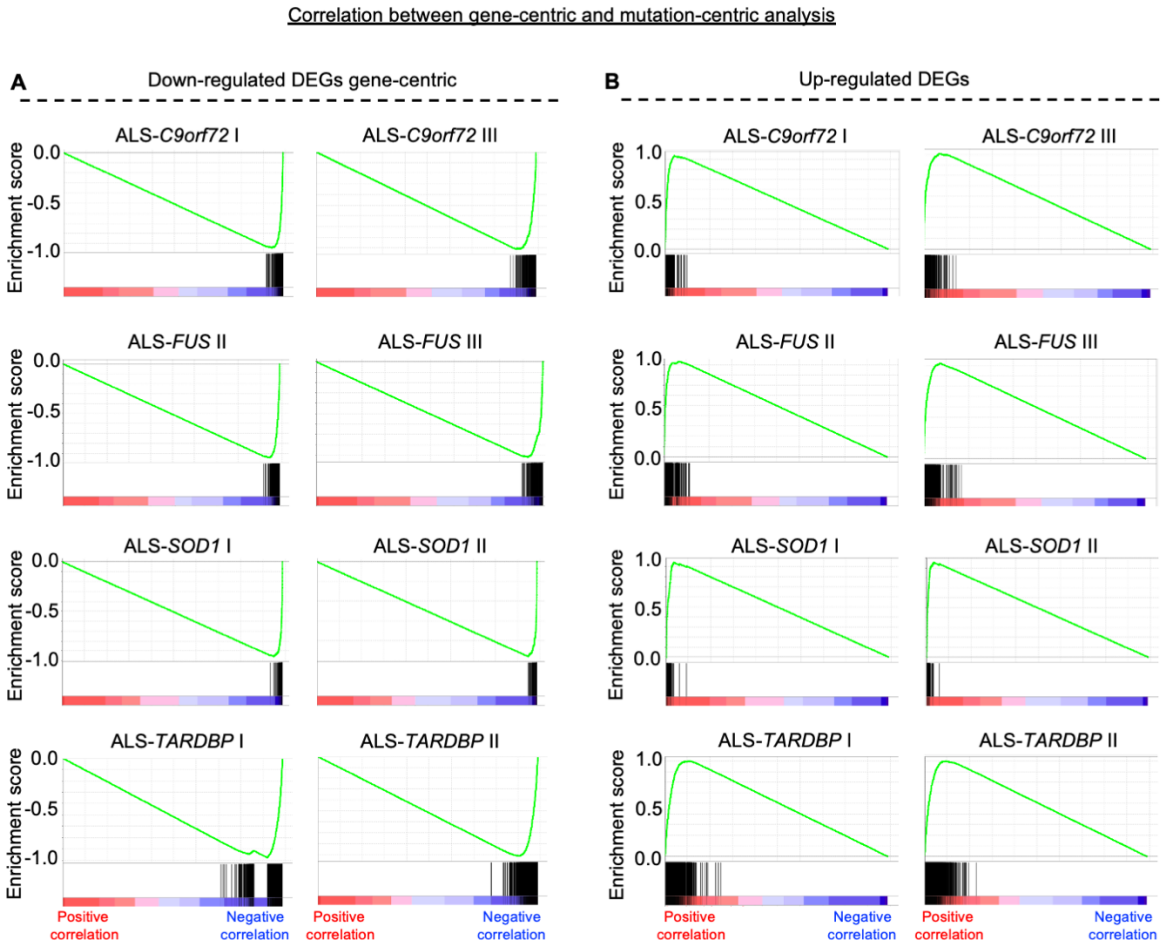
Heatmaps showing the hierarchical clustering of the DEGs identified in the transcriptomes of ALS-*C9orf72* (**A**), ALS-*FUS* (**B**), ALS-*SOD1* (**C**) and ALS-*TARDBP* (**D**) MN when compared to MN from healthy controls.



Supplementary Figure 2

Gene-centric transcriptome analysis separates mutants from control MN

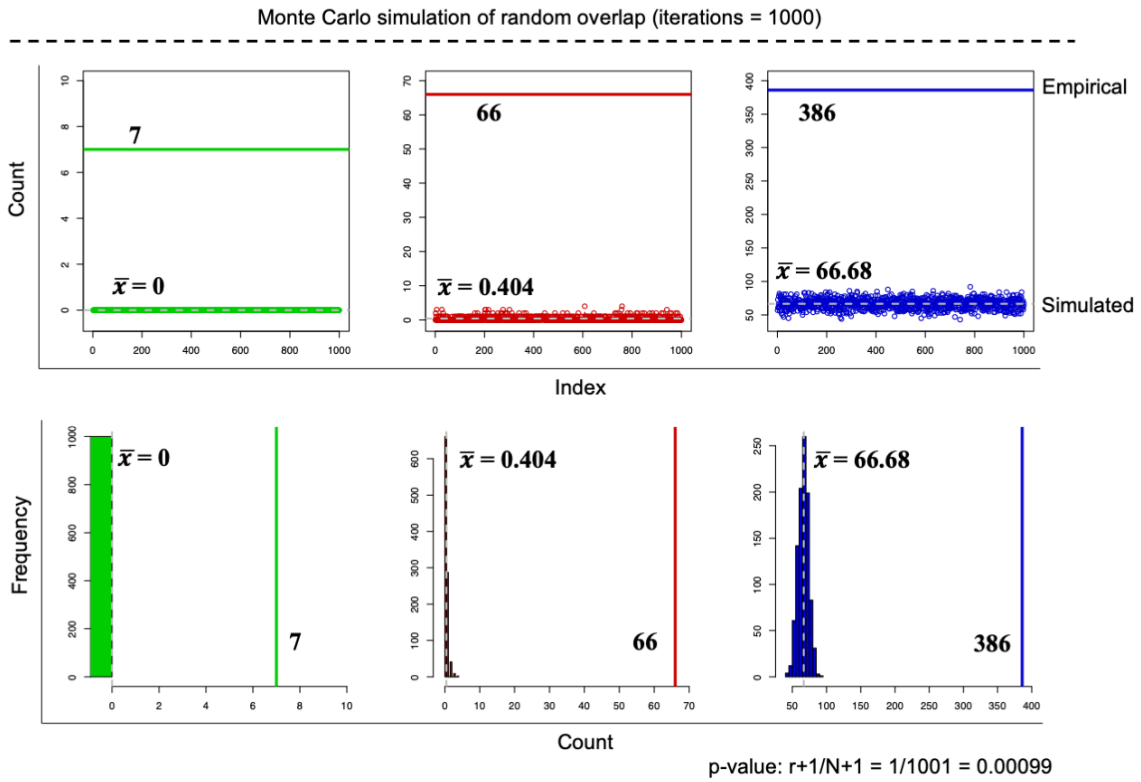
Heatmaps showing the hierarchical clustering based on the determinants of the PCA shown in Fig. 1A. Healthy controls (violet cluster) are separate from all the ALS samples.



Supplementary Figure 3

Mutation-centric transcriptome analysis correlates significantly with the gene-centric approach.

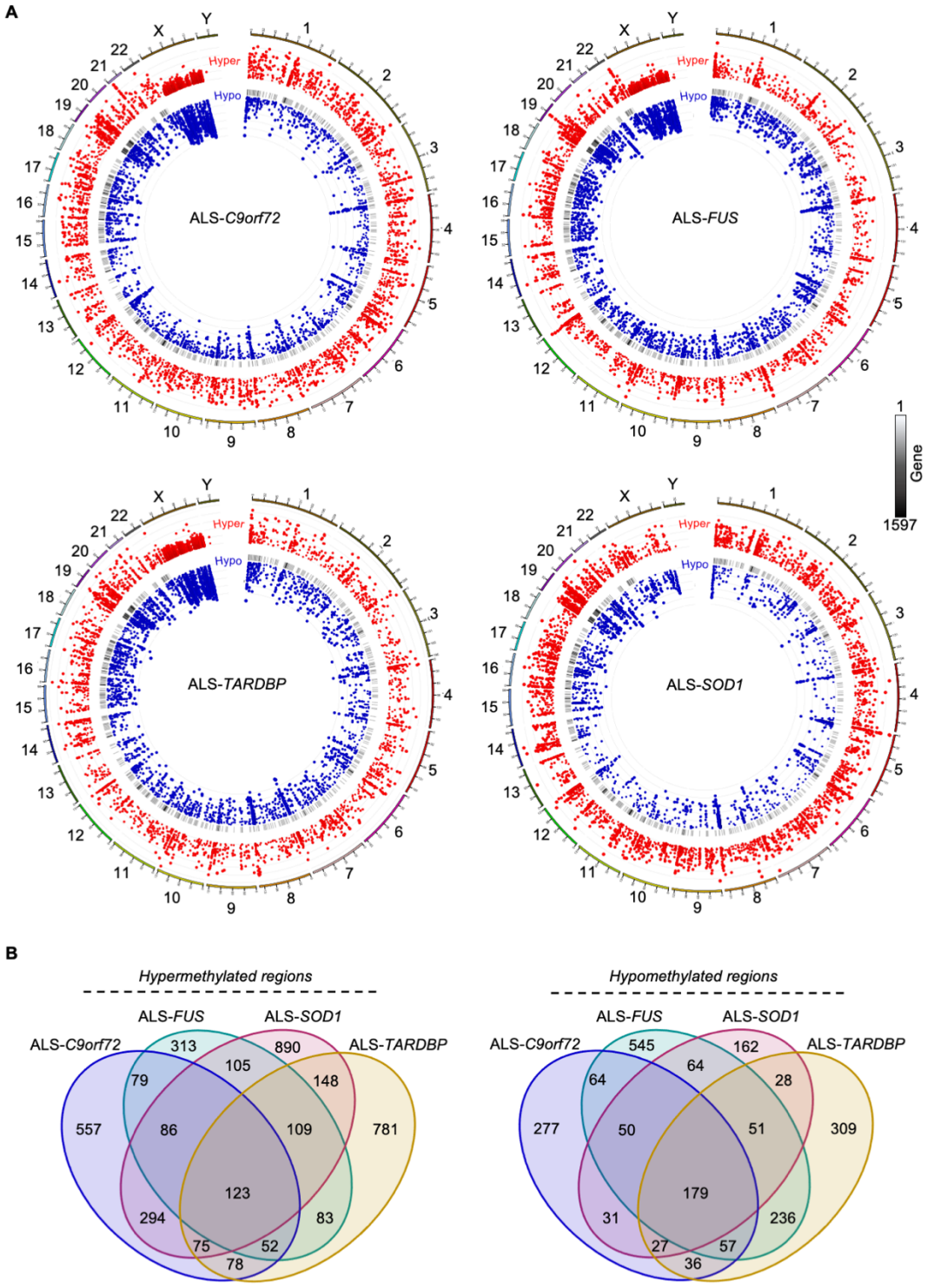
(**A**) Correlation analysis between the downregulated and (**B**) upregulated DEGs identified in each single mutant line and the respective ALS group (gene-centric, obtained by pooling the lines with mutations within the same gene). As displayed, there is a strongly significant (p-value < 0.001; FDR < 0.001 in each comparison) correlation between both approaches, indicating that the mutation-specific alterations account for a minimal variability within the pathological transcriptome of ALS MN.



Supplementary Figure 4

Shared expression of DEGs in ALS MN does not occur randomly

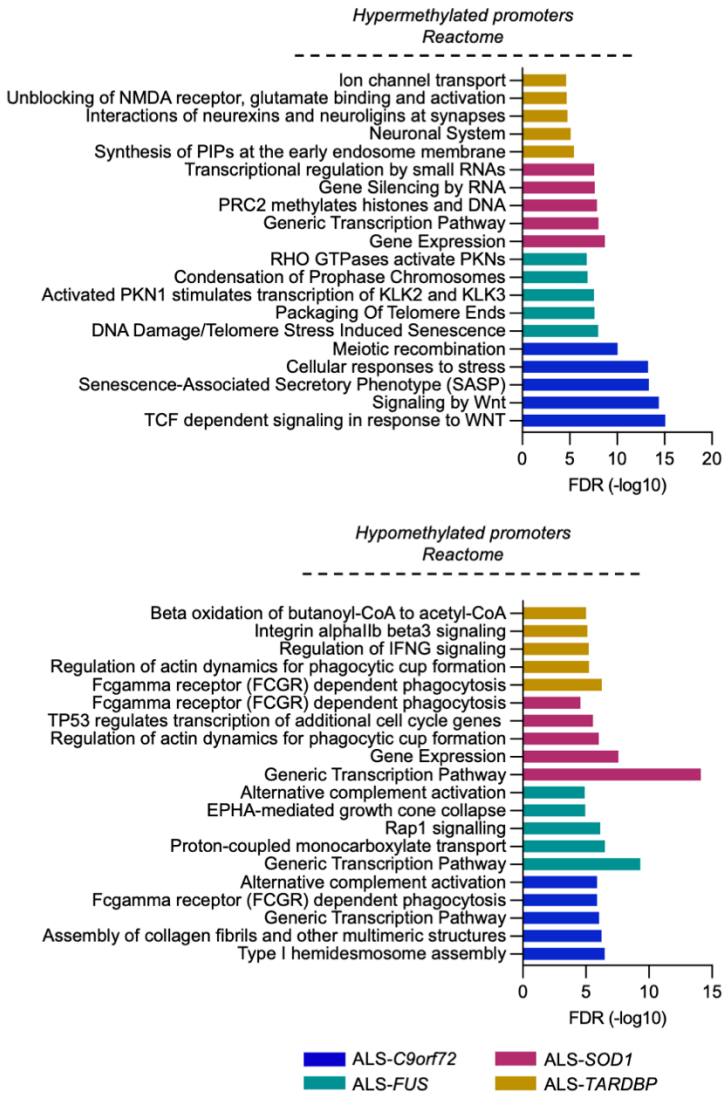
Monte Carlo simulation of the DEGs´ sharing degree expression across the different ALS groups performed by 1000 random iterations. This approach failed in reproducing the number of DEGs shared by all (green), three (red) or two (blue) ALS subgroups revealed in the RNAseq analysis. Dashed lines represent the number of shared genes computed by the simulation, while the continuous lines indicate the number of shared genes identified. Empirical numbers were obtained by calculating the number of genes shared in all 4 (green graphs), at least 3 (red graphs) or at least 2 (blue graphs) ALS groups according to our RNAseq data shown in Fig. 1.



Supplementary Figure 5

Heterogenous distribution of differentially methylated regions within the whole genome of ALS-related MN

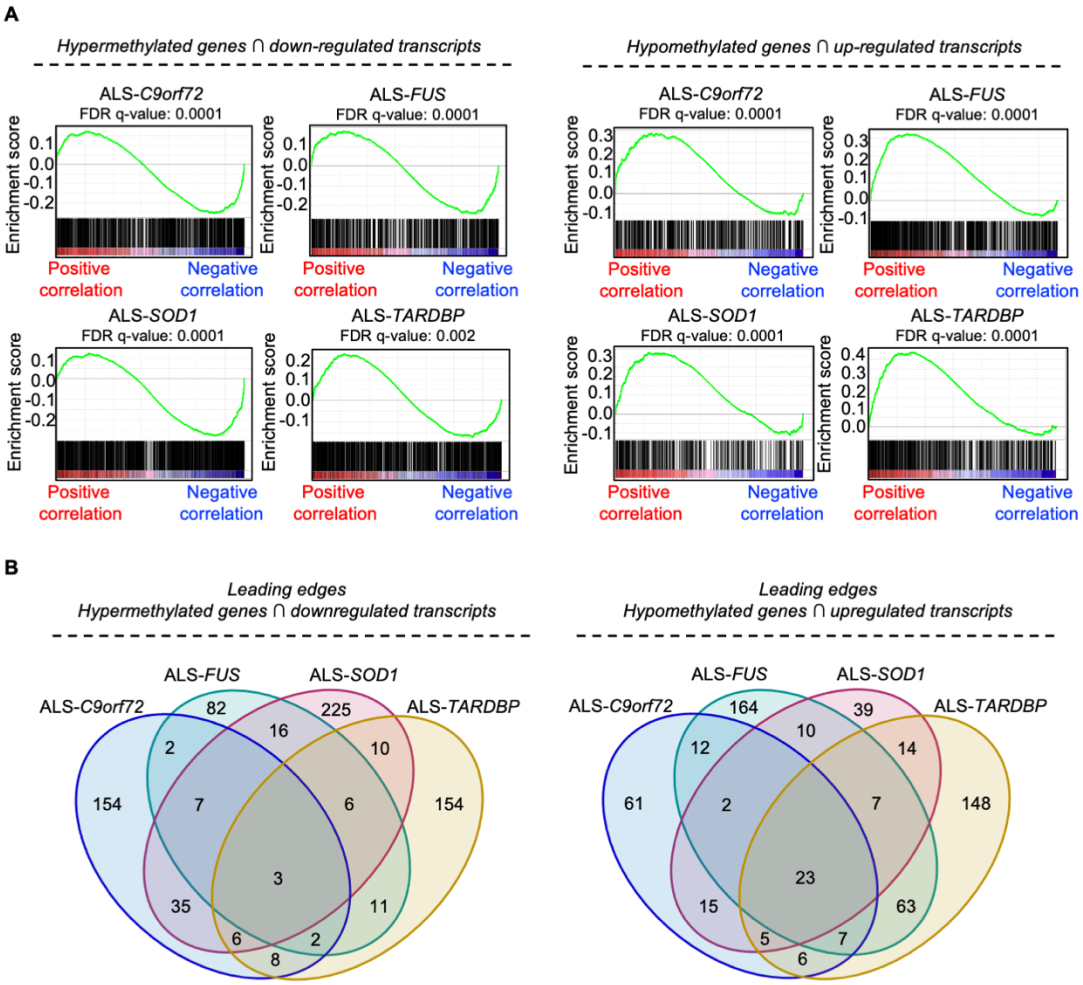
(**A**) Chromosome-based circos plots showing the differential methylation pattern of ALS-*C9orf72*, ALS-*FUS*, ALS-*SOD1* and ALS-*TARDBP* MN when compared to controls. (**B**) Venn diagrams representing the unique and shared DMRs in the different ALS groups.



Supplementary Figure 6

Top unique pathways associate with the DMRs of ALS MN

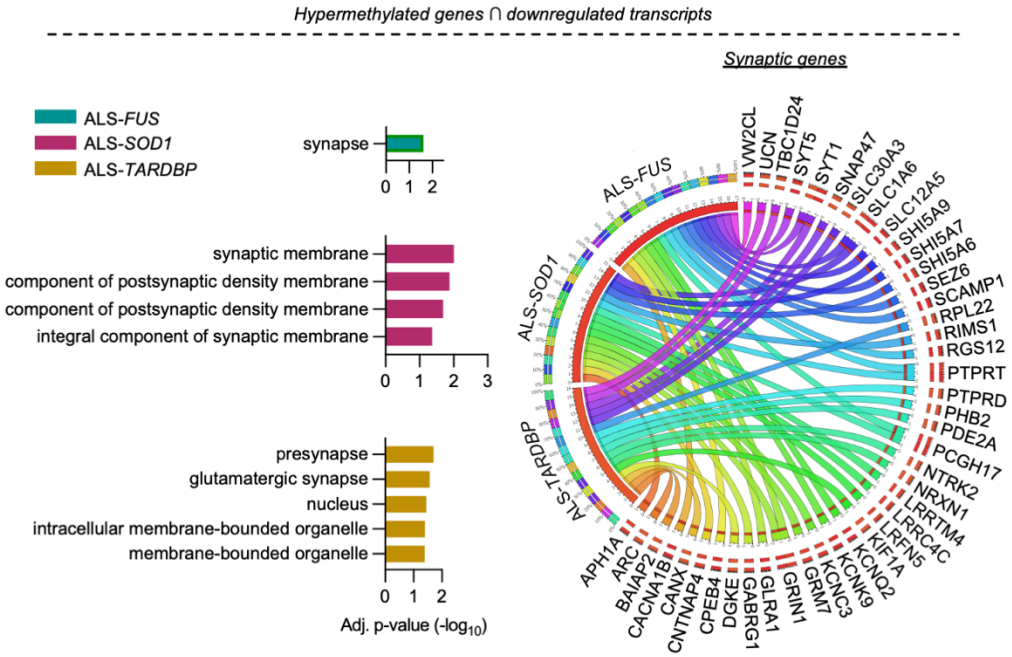
Reactome analysis of the top terms associated with the significantly hyper- and hypomethylated promoters identified with HOMER (see Figure 2) in the different ALS subgroups.



Supplementary Figure 7

Significant correlation between the transcriptome and methylome profiles in ALS MN

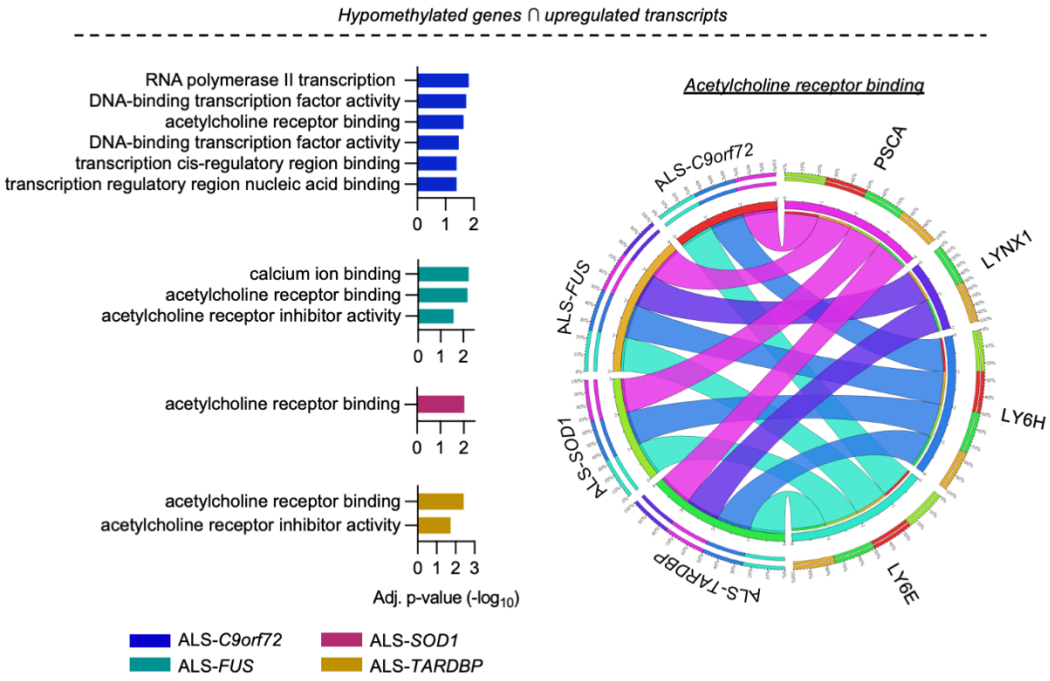
(**A**) We identified a statistically significant correlation between the expression levels (down- or up-regulation) of transcripts with the methylation status (hyper- or hypomethylated) of their respective promoters. (**B**) When we looked at the shared DEGs whose promoters where also significantly hyper- (assuming down-regulation) or hypomethylated (assuming up-regulation), we observed that the shared downregulated genes were only 3, while the upregulated were 23.



Supplementary figure 8

Significantly downregulated and hypermethylated synaptic genes in ALS

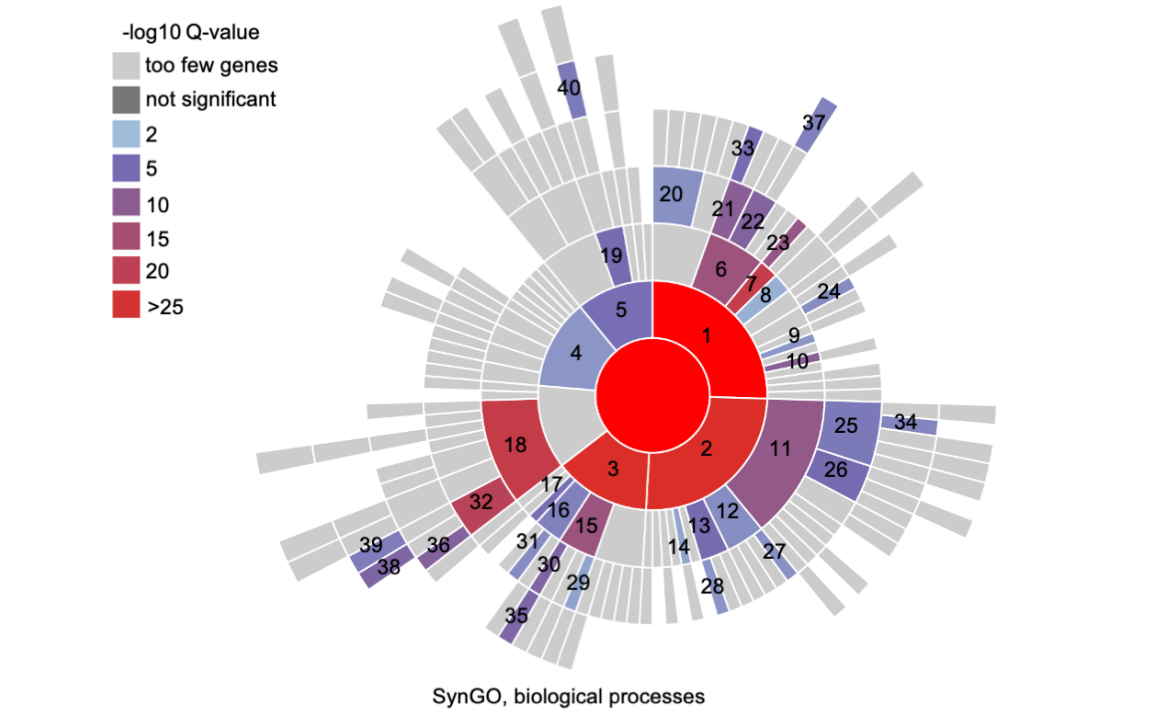
Significantly enriched synaptic terms were obtained by analyzing the downregulated synaptic transcripts whose promoter is also hypermethylated displayed in the circos plot (also reported in Supplementary Table 2). This suggests epigenetic silencing of these genes in ALS-*FUS,* -*TARDBP* and -*SOD1*. The different ALS groups and genes are color-coded by the different lines, which represent the reciprocal interconnection and association.



Supplementary Figure 9

Significantly upregulated and hypomethylated genes in ALS

All the ALS subgroups display up-regulation of *Acetylcholine receptor binding* genes (also reported in Supplementary Table 3), whose promoters are hypo-methylated as well. The different ALS groups and genes are color-coded by the different lines, which represent the reciprocal interconnection and association.

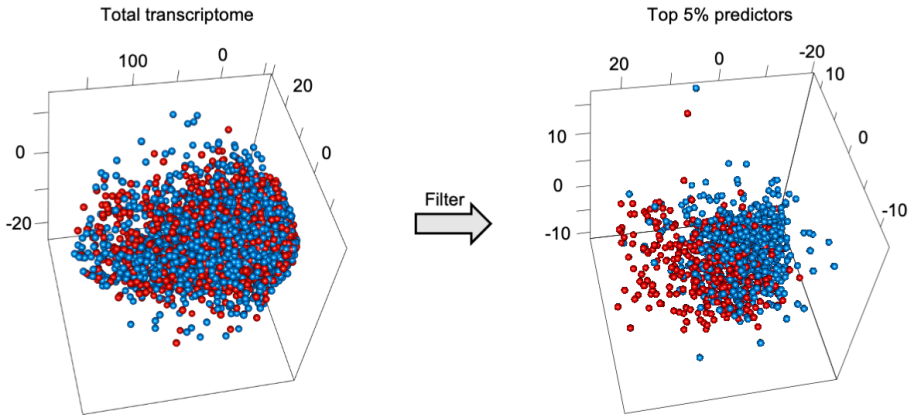


1 Synapse organisation
2 Process in the presynapse
3 Process in the postsynapse
4 Metabolism
5 Transport
6 Synapse assembly
7 Postsynapse organization
8 Modification of synapse structure
9 Regulation of synapse organization
10 Synapse adhesion
11 Synaptic vesicle cycle
12 Regulation of presynaptic membrane potential
13 Regulation of presynaptic calcium levels
14 Presynaptic dense core vesicle release
15 Regulation of postsynaptic receptor levels
16 Regulation of postsynaptic membrane potential
17 Regulation of postsynaptic receptor activity
18 Trans-synaptic signaling
19 Neurotransmitter receptor transport
20 Structural constituent of postsynapse
21 Regulation of synapse assembly
22 Postsynaptic specialization assembly
23 Regulation of postsynaptic organization
24 Postsynaptic cytoskeleton organization
25 Synaptic vesicle exocytosis
26 Synaptic vesicle endocytosis
27 Ligand-gated ion channel activity
28 Voltage-gated calcium channel activity
29 Neurotransmitter receptor localization
30 Postsynaptic receptor endocytosis
31 Transmitter-gated ion channel activity
32 Chemical synapse transmission
33 Regulation of presynapse assembly
34 Regulation of synaptic vesicle exocytosis
35 Regulation of postsynaptic receptor endocytosis
36 Modulation of chemical synaptic transmission
37 Regulation of postsynaptic density assembly
38 Presynaptic modulation of chemical synaptic
39 Postsynaptic modulation of chemical synaptic
40 neurotransmitter receptor endosomal transport

Supplementary Figure 10

Synaptic transcripts are downregulated in the ALS spinal transcriptome

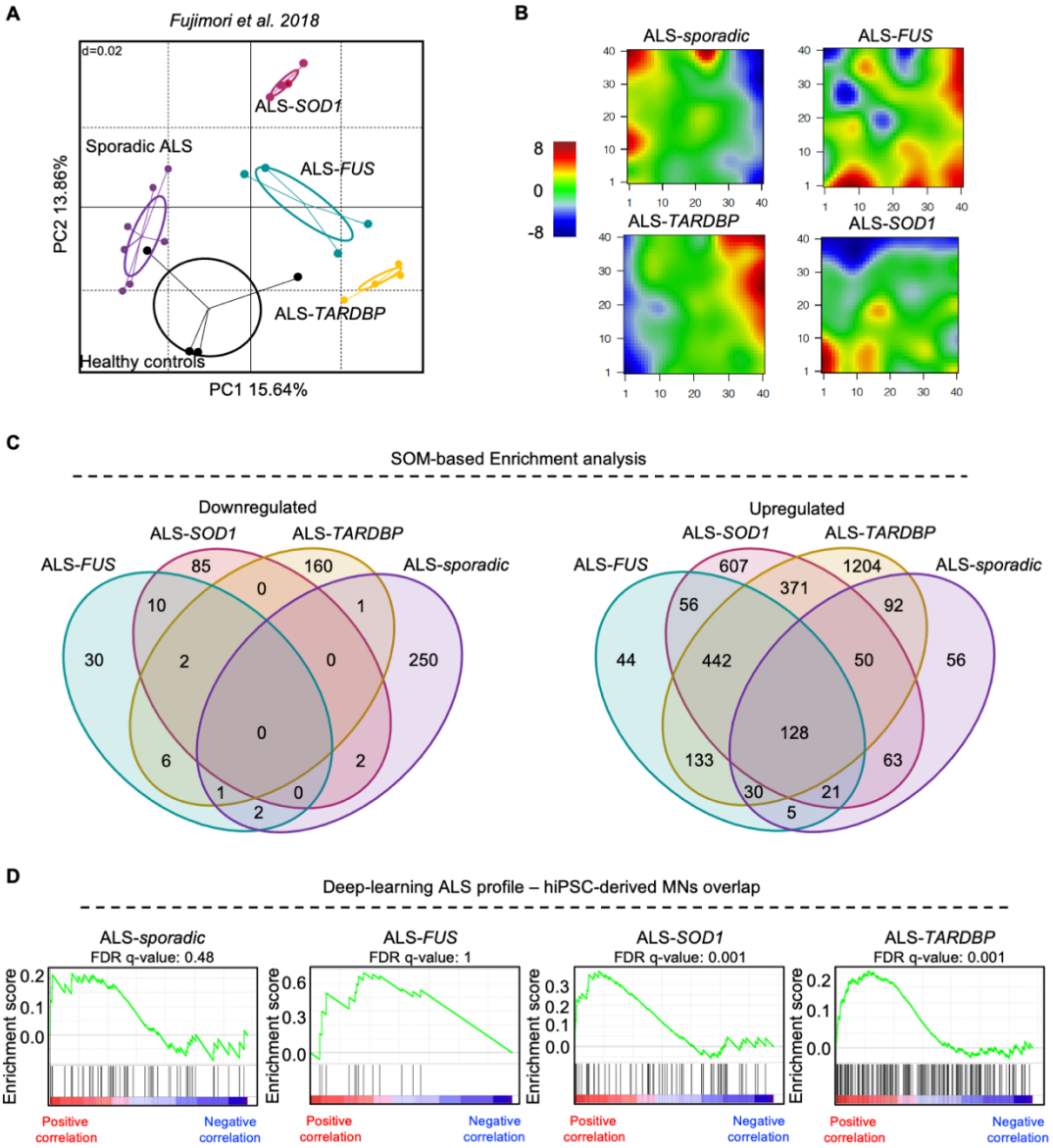
Graphical representation of the significantly enriched (colored according to-log10 Q-value) SynGO biological processes. The tiles represent, from the center to the border, the different main and child SynGO terms, which are numbered and listed in case of statistical signifance. The analysis was performed with the downregulated synaptic genes in ALS *post mortem* samples (the full gene list can be found in the Supplementary Table 4).



Supplementary Figure 11

Top 5% predictor genes separated ALS from control transcriptomes

Filtering the top 5% ALS and Healthy predictor genes from the blood and spinal cord transcriptomes efficiently separates the two genotypes after deep-learning analysis.

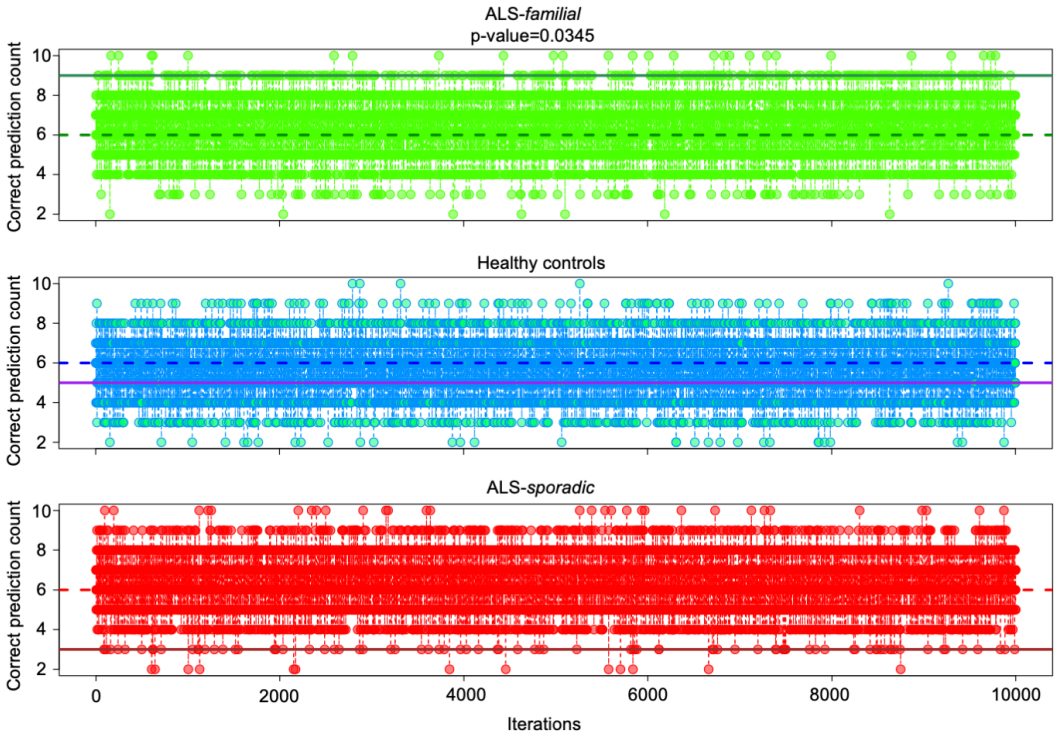


Supplementary Figure 12

Confirmation of the ALS transcriptional signature in independent hiPSC-derived MN

(**A**) We considered the microarray-based transcriptome analysis of hiPSC MN published by Fujimori and colleagues[5], which again revealed a high heterogeneity between the different ALS subgroups included. (**B**) Also in this case, SOMs analysis confirmed the reduced degree of shared altered terms across the ALS cases when compared to healthy control, as displayed in (**C**) the Venn diagrams. (**D**) Nevertheless, in line with the *in-house* results, the transcriptional profile of all the ALS subgroups showed a positive correlation with the transcriptional signature identified by deep learning, which reached statistical significance in the case of ALS-*SOD1* and ALS-*TARDBP*.

The color scale in the SOMs represents the gene expression level expressed as relative to the mean of the portrait (0 value, green color) after normalization.



Supplementary Figure 13

Genes associated with toll-like receptor signaling are predictive of fALS genotype

We performed, in double blind conditions, single-tube qPCRs against the LY96, TLR2, TLR6, TLR8, EVI2B, CD58, FAM126B, PP3CB genes using RNA obtained from the blood of 20 ALS patients (n=10 familial and n=10 sporadic) and 10 healthy controls. The expression levels of these transcripts were then integrated into the deep-learning algorithm shown in Figure 3 to evaluate the accuracy of these markers in predicting the correct genotype. We then used Monte Carlo simulation to disclose whether the association computed by the deep learning approach on the basis of only the expression of these 8 transcripts might yield enough robustness to recognize an ALS patient. Notably, the number of familial patients (green graph) correctly recognized was significantly higher (p=0.0345) than the one simulated, indicating that these genes might represent valid transcriptional biomarkers at least for this portion of ALS patients. Dashed lines represent the simulated number of individuals associated correctly to their genotype, while the continuous lines indicate the number of patients correctly associated based on the expression levels of LY96, TLR2, TLR6, TLR8, EVI2B, CD58, FAM126B, PP3CB.

Supplementary Table 1: Ranking of the differentially methylated genomic regions in fALS MN

Annotation	P-value			
	ALS-*C9orf72*	ALS-*FUS*	ALS-*SOD1*	ALS-*TARDBP*
CpG islands	1e-687	1e-1261	1e-515	1e-855
Promoters	1e-224	1e-445	1e-175	1e-341
Exons	1e-131	1e-297	1e-100	1e-170
Protein-coding	1e-103	1e-247	1e-84	1e-137
Coding	1e-102	1e-208	1e-65	1e-131

Supplementary Table 2: Synaptic genes hypermethylated and downregulated in fALS MN

Gene	Enrichment Score			
	ALS-*C9orf72*	ALS-*FUS*	ALS-*SOD1*	ALS-*TARDBP*
SYT1	-	-0.2251	-	-0.1091
KCNK9	-	-0.229	-	-
RGS12	-	-0.2202	-	-
KCNQ2	-	-0.1685	-	-
PTPRT	-	-0.1553	-0.1698	-
SCAMP1	-	-0.1419	-	-
PCDH17	-	-0.1207	-0.2670	-
CACNA1B	-	-0.1126	-	-
SLC1A6	-	-0.0984	-0.2671	-
SEZ6	-	-0.0950	-	-
SLC12A5	-	-0.0824	-	-
VWC2L	-	-0.0704	-	-
SHISA7	-	-0.0598	-	-
CPEB4	-	-0.0560	-	-
GLRA1	-	-0.0435	-	-
GRIN1	-	-0.0368	-0.2747	-
SYT5	-	-0.0297	-	-
LRRC4C	-	-	-0.2728	-
LRRTM4	-	-	-0.1988	-
SHISA9	-	-	-0.1944	-
LRFN5	-	-	-0.1374	-
SHISA6	-	-	-0.1082	-
GABRG1	-	-	-0.2446	-
APH1A	-	-	-0.2237	-
KCNC3	-	-	-0.1896	-
GRM7	-	-	-0.1199	-
RIMS1	-	-	-0.1063	-
CNTNAP4	-	-	-0.0566	-
NRXN1	-	-	-	-0.1621
UCN	-	-	-	-0.1611
SNAP47	-	-	-	-0.1539
CANX	-	-	-	-0.1521
PDE2A	-	-	-	-0.1463
NTRK2	-	-	-	-0.1442
PTPRD	-	-	-	-0.1404
TBC1D24	-	-	-	-0.1219
KIF1A	-	-	-	-0.1131
SLC30A3	-	-	-	-0.0922
PHB2	-	-	-	-0.0804
BAIAP2	-	-	-	-0.0409
RPL22	-	-	-	-0.0414
ARC	-	-	-	-0.1592
DGKE	-	-	-	-0.0023

Supplementary Table 3: Acetylcholine receptor binding genes hypomethylated and upregulated in fALS MN

Gene	Enrichment Score			
	ALS-*C9orf72*	ALS-*FUS*	ALS-*SOD1*	ALS-*TARDBP*
PSCA	0.1494	0.1617	0.1080	0.2894
LY6H	0.2112	0.3539	0.2317	0.3182
LY6E	0.2242	0.2530	0.2381	0.2962
LYNX1	-	0.3606	-	0.3684

Supplementary Table 4: List of significantly downregulated synaptic genes in the ALS spinal cord

Gene	Fold change	Gene	Fold change	Gene	Fold change
ABHD17A	-0,509	*DNM2*	-0,345	*P2RX2*	-0,603
ABL1	-0,379	*DOC2B*	-0,406	*P2RX6*	-0,369
ADGRB1	-0,794	*DRD2*	-0,569	*P2RY4*	-0,544
ADGRL1	-0,536	*DRD4*	-0,813	*PICK1*	-0,410
ADORA1	-0,279	*DVL1*	-0,470	*PIP5K1C*	-0,615
ADORA2A	-0,786	*EEF2*	-0,137	*PLAT*	-0,476
ADRA2A	-0,429	*EPHB2*	-0,341	*PLEKHG5*	-0,803
ADRA2C	-0,452	*EPN1*	-0,386	*PRR7*	-0,538
AGAP3	-0,376	*EPS15L1*	-0,213	*PTPN23*	-0,370
AGRN	-0,738	*FZD9*	-0,517	*PTPRF*	-0,365
AKT1	-0,263	*GHRL*	-0,572	*PTPRS*	-0,361
ANO1	-0,529	*GIPC1*	-0,271	*RAB11B*	-0,293
ANO2	-0,499	*GIT1*	-0,457	*RAB11FIP3*	-0,473
AP3D1	-0,258	*GPC2*	-0,423	*RNF216*	-0,210
APBA2	-0,240	*GPER1*	-0,636	*SCRIB*	-0,492
ARC	-0,713	*GRID1*	-0,264	*SEMA3F*	-0,631
ARHGAP22	-0,375	*GRIK4*	-0,226	*SEMA4B*	-0,374
ARHGAP39	-0,530	*GRIK5*	-0,414	*SGTA*	-0,275
ARHGEF15	-0,726	*GRIN2D*	-0,378	*SH3GL1*	-0,293
ARHGEF2	-0,281	*GRIN3B*	-0,615	*SH3GLB2*	-0,470
ATG9A	-0,172	*GRIP2*	-0,544	*SHANK1*	-0,816
BCR	-0,398	*GRIPAP1*	-0,377	*SHANK3*	-0,842
BEGAIN	-0,657	*HIP1R*	-0,550	*SHISA7*	-0,606
CACNA1A	-0,602	*HTT*	-0,176	*SLC1A7*	-0,114
CACNA1B	-0,721	*IGF1R*	-0,273	*SLC6A11*	-0,338
CACNA1C	-0,467	*IGSF9B*	-0,682	*SLC6A6*	-0,241
CACNA1H	-0,999	*IQSEC1*	-0,341	*SLC6A8*	-0,249
CACNG4	-0,321	*IQSEC2*	-0,363	*SRC*	-0,220
CACNG7	-0,328	*IQSEC3*	-0,674	*SRCIN1*	-0,513
CHRNA10	-0,657	*ITGA3*	-0,197	*STX4*	-0,236
CHRNA4	-0,563	*ITPR3*	-0,643	*SYNGAP1*	-0,519
CHRNE	-0,345	*KCNC3*	-0,464	*TAMALIN*	-0,635
CLPTM1	-0,171	*KCNC4*	-0,358	*TAOK2*	-0,343
CPT1C	-0,523	*KLHL17*	-0,457	*TBC1D2*	-0,351
CRTC1	-0,526	*LAMA5*	-0,794	*TBC1D24*	-0,376
CTBP1	-0,223	*LHFPL4*	-0,483	*TRIM3*	-0,233
CTTN	-0,248	*LPAR2*	-0,367	*TRIO*	-0,204
CTTNBP2	-0,190	*LRFN1*	-0,468	*TSC2*	-0,544
CYTH1	-0,251	*LRFN3*	-0,489	*TSPOAP1*	-0,687
CYTH2	-0,252	*LRFN4*	-0,457	*VAC14*	-0,308
DAG1	-0,164	*LRRC4B*	-0,486	*VPS18*	-0,125
DAGLA	-0,446	*MARK2*	-0,339		
DBN1	-0,313	*MPP2*	-0,237		
DBNL	-0,430	*MTOR*	-0,116		
DGKQ	-0,587	*NECTIN1*	-0,492		
DGKZ	-0,552	*NLGN2*	-0,369		
DLG5	-0,333	*NLGN3*	-0,241		
DLGAP3	-0,756	*NRXN2*	-0,382		
DLGAP4	-0,354	*NTNG2*	-0,615		
DNAJC5	-0,219	*NUMBL*	-0,328		

Supplementary Table 5: Clinical data of the blood donors

	Diagnosis	Mutation	Age	Sex
Healthy controls	-	-	74	♀
	-	-	40	♀
	-	-	67	♀
	-	-	74	♂
	-	-	71	♂
	-	-	48	♀
	-	-	60	♂
	-	-	50	♂
	-	-	55	♂
	-	-	66	♂
ALS patients	fALS	*C9orf72*	60	♂
	fALS	*C9orf72*	63	♀
	fALS	*C9orf72*	52	♂
	fALS	*C9orf72*	52	♂
	fALS	*C9orf72*	59	♀
	fALS	*C9orf72*	55	♀
	fALS	*C9orf72*	43	♂
	fALS	*C9orf72*	64	♂
	fALS	*C9orf72*	63	♂
	fALS	*C9orf72*	67	♂
	sALS	negative for *SOD1, TARDBP, C9ORF72, FUS*	76	♂
	sALS	negative for *SOD1, TARDBP, C9ORF72, FUS*	55	♀
	sALS	negative for *SOD1, TARDBP, C9ORF72, FUS*	72	♂
	sALS	negative for *SOD1, TARDBP, C9ORF72, FUS*	85	♀
	sALS	negative for *SOD1, TARDBP, C9ORF72, FUS*	82	♀
	sALS	negative for *SOD1, TARDBP, C9ORF72, FUS*	78	♀
	sALS	negative for *SOD1, TARDBP, C9ORF72, FUS*	74	♂
	sALS	negative for *SOD1, TARDBP, C9ORF72, FUS*	75	♂
	sALS	negative for *SOD1, TARDBP, C9ORF72, FUS*	73	♂
	sALS	negative for *SOD1, TARDBP, C9ORF72, FUS*	52	♀
	sALS	negative for *SOD1, TARDBP, C9ORF72, FUS*	76	♀