



Multimomics 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} and Medhanie Mulaw^{11,†}

[†]These authors contributed equally to this work.

Amyotrophic lateral sclerosis is a fatal and incurable neurodegenerative disease that mainly affects the neurons of the motor system. Despite the increasing understanding of its genetic components, their biological meanings are still poorly understood. Indeed, it is still not clear to which extent the pathological features associated with amyotrophic lateral sclerosis are commonly shared by the different genes causally linked to this disorder. To address this point, we combined multimomics analysis covering the transcriptional, epigenetic and mutational aspects of heterogeneous human induced pluripotent stem cell-derived C9orf72-, TARDBP-, SOD1- and FUS-mutant motor neurons as well as datasets from patients' biopsies. We identified a common signature, converging towards increased stress and synaptic abnormalities, which reflects a unifying transcriptional program in amyotrophic lateral sclerosis despite the specific profiles due to the underlying pathogenic gene. In addition, whole genome bisulphite sequencing linked the altered gene expression observed in mutant cells to their methylation profile, highlighting deep epigenetic alterations as part of the abnormal transcriptional signatures linked to amyotrophic lateral sclerosis. We then applied multi-layer deep machine-learning to integrate publicly available blood and spinal cord transcriptomes and found a statistically significant correlation between their top predictor gene sets, which were significantly enriched in toll-like receptor signalling. Notably, the overrepresentation of this biological term also correlated with the transcriptional signature identified in mutant human induced pluripotent stem cell-derived motor neurons, highlighting novel insights into amyotrophic lateral sclerosis marker genes in a tissue-independent manner. Finally, using whole genome sequencing in combination with deep learning, we generated the first mutational signature for amyotrophic lateral sclerosis and defined a specific genomic profile for this disease, which is significantly correlated to ageing signatures, hinting at age as a major player in amyotrophic lateral sclerosis. This work describes innovative methodological approaches for the identification of disease signatures through the combination of multimomics analysis and provides novel knowledge on the pathological convergencies defining amyotrophic lateral sclerosis.

- 1 Institute of Anatomy and Cell Biology, Ulm University School of Medicine, 89081 Ulm, Germany
- 2 Translational Protein Biochemistry, German Center for Neurodegenerative Diseases (DZNE), Ulm site, 89081 Ulm, Germany
- 3 Laboratory of Neurobiology, Center for Brain & Disease Research, VIB, 3000 Leuven, Belgium
- 4 Department of Neurology, University Hospitals Leuven, 3000 Leuven, Belgium
- 5 Experimental Neurology, Department of Neurosciences, Leuven Brain Institute, KU Leuven, 3000 Leuven, Belgium
- 6 Department of Neurology, Ulm University School of Medicine, 89081 Ulm, Germany
- 7 Center for Neural Science and Medicine, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA

Received November 07, 2022. Revised February 15, 2023. Accepted February 25, 2023. Advance access publication March 8, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of the Guarantors of Brain.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

8 Board of Governors Regenerative Medicine Institute, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA
 9 Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA
 10 Department of Neurology, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA
 11 Unit for Single-Cell Genomics, Medical Faculty, Ulm University, 89081 Ulm, Germany

Correspondence to: Alberto Catanese
 Institute of Anatomy and Cell Biology
 Ulm University Albert-Einstein-Allee 11, 89069 Ulm, Germany
 E-mail: alberto.catanese@uni-ulm.de

Keywords: ALS; omics; deep learning; motor neurons

Introduction

Amyotrophic lateral sclerosis (ALS) is the most prevalent motor neuron disease and is characterized by a devastating progression leading to death within 1–5 years following diagnosis.¹ The recent advances in techniques and mathematical methods have consistently deepened our understanding of the genetic basis of ALS, helping in gaining more knowledge into the extremely heterogeneous pathological landscape of this fatal disease.² Nevertheless, despite the importance of these findings, an efficacious treatment for this major neurodegenerative disorder is still missing. In fact, the genes that have been causally linked to ALS are involved in broadly different cellular pathways and functions.³ This suggests that the molecular pathomechanisms characterizing the different ALS cases might largely depend on the mutated genes. Still, the clinical presentation of ALS patients cannot entirely be explained on the basis of the underlying genetic cause, thus leaving a crucial question unanswered: what are the core commonalities characterizing the different ALS cases that, together with the genetic background, contribute to the onset and progression of this pathology? In this study, we sought to identify convergent (and divergent) ALS-related alterations at the transcriptomic, epigenetic and genetic levels with the aim of providing a portrait biologically defining this neurodegenerative disease.

Materials and methods

Human iPSCs and differentiation into motor neurons

Human induced pluripotent stem cell (hiPSC) lines (Table 1) were either generated at Ulm University or purchased from The Induced Pluripotent Stem Cell (iPSC) Core (David and Janet Polak Foundation Stem Cell Core Laboratory) at the Board of Governors Regenerative Medicine Institute (Cedars-Sinai Medical Center, Los Angeles, CA, USA) and BioCat GmbH. hiPSCs were cultured on Matrigel®-coated (Corning, Cat. No. 354277) six-well plates (BD Falcon, 351146) in mTeSR™1 medium (STEMCELL Technologies, Cat. No. 85850) under ideal incubation conditions (37°C, 5% CO₂, 5% O₂). At 80% confluency, the cells were passaged 1:3 or 1:6 as required, using dispase (STEMCELL Technologies, Cat. No. 07923). hiPSCs were differentiated into motor neurons (MNs) following the protocol previously described in Catanese et al.⁴ Briefly, hiPSC colonies were detached and cultivated in suspension in ultra-low attachment flasks T75 for 3 days to induce the formation of embryoid bodies in human embryonic stem cell medium (Dulbecco's modified Eagle medium/F12 + 20% knockout serum replacement + 1% NEAA + 1% β-mercaptoethanol + 1% antibiotic-antimycotic + SB-431542 10 μM + Dorsomorphin 1 μM + CHIR 99021 3 μM + Purmorphamine 1 μM + ascorbic acid 200 ng/μl + cAMP 10 μM + 1% B27 + 0.5% N2). On

the fourth day, the medium was switched to MN Medium (Dulbecco's modified Eagle medium/F12 + 24 nM sodium selenite + 16 nM progesterone + 0.08 mg/ml apotransferrin + 0.02 mg/ml insulin + 7.72 μg/ml putrescine + 1% NEAA, 1% antibiotic-antimycotic + 50 mg/ml heparin + 10 μg/ml of the neurotrophic factors BDNF, GDNF and IGF-1, SB-431542 10 μM, Dorsomorphin 1 μM, CHIR 99021 3 μM, Purmorphamine 1 μM, ascorbic acid 200 ng/μl, retinoic acid 1 μM, cAMP 1 μM, 1% B27, 0.5% N2). Ultimately, after five further days of cultivation, the embryoid bodies were dissociated into single cells with Accutase (Sigma Aldrich, Cat. No. A6964) and plated onto six-well plates (Corning, Cat. No. 3516) precoated with Growth Factor Reduced Matrigel (Corning, Cat. No. 356230).

RNA isolation from human MN

Day *in vitro* 28 (DIV28) hiPSC-MN from two wells of a six-well plate were washed with phosphate buffered saline (PBS), the cells were then scraped gently, collected in 1× PBS, centrifuged for 5 min at 300g and RNA was isolated from the pellet using the RNeasy Mini Kit (Qiagen, Cat. No. 74106) based on instructions from the manufacturer.

Isolation and preservation of peripheral blood mononuclear cells

Peripheral blood was collected by venepuncture into heparin tubes. All blood was processed within 1 h after collection. The heparin tubes were centrifuged at 300g for 20 min at 15°C (all centrifugation steps were done in a swing-out bucket rotor). The upper plasma layer was removed to transfer the buffy coat layer in a 15 ml tube containing 4 ml PBS, 0.1% BSA and 2 mM EDTA. The diluted cells were carefully transferred to 15 ml tube with 3 ml density medium Lymphoprep™ (Stemcell Technologies, Cat. No. 07801). To obtain a gradient, we centrifuged the tubes at 160g for 30 min at 15°C. The supernatant was removed to eliminate the platelets followed by a third centrifugation at 350g for 20 min at 15°C, the remaining supernatant was removed, peripheral blood mononuclear cells (PBMCs) were recovered from the plasma/lymphoprep interface and were transferred to a 15 ml tube with 5 ml of PBS. To precipitate the PBMCs, a final centrifugation at 400g for 8 min at 15°C was carried out. The cells were then resuspended in 1 ml of FBS, 10% dimethyl sulphoxide and aliquoted in cryovials. The samples were placed in Mr. Frosty™ (Thermo Fisher Scientific, Cat. No. 5100-0001) at –80°C to enable controlled rate freezing in preparation for final storage in liquid nitrogen vapours at –180°C.

Total RNA isolation of peripheral blood mononuclear cells and real-time PCR

Total RNA was extracted from PBMCs using TRI Reagent® (Sigma Aldrich/Merck) according to the manufacturer's instructions.

Table 1 hiPSC donors information

Cell line	Gene	Mutation	Age	Sex	Source	Catalogue number
Healthy controls						
Healthy I	–	–	45	Female	Ulm University ⁵	N/A
Healthy II	–	–	64	Male	BioCat GmbH	SC600A-WT
Healthy III	–	–	52	Female	Cedars-Sinai	CS14iCTR-21
ALS patients						
ALS-C9orf72 I	C9orf72	(G ₄ C ₂)1.8kb	60	Male	Ulm University ⁵	N/A
ALS-C9orf72 III	C9orf72	(G ₄ C ₂)2.7kb	50	Female	Cedars-Sinai	CS30iALS-C9
ALS-FUS II	FUS	c.1484delG	27	Male	Ulm University ⁶	N/A
ALS-FUS III	FUS	c.1504delG	19	Male	Ulm University ⁷	N/A
ALS-TARDBP I	TARDBP	p.Gly298Ser	62	Male	Cedars-Sinai	CS47iALS-TDP
ALS-TARDBP II	TARDBP	p.N390D	26	Male	Cedars-Sinai	CS5ZLDiALS
ALS-SOD1 I	SOD1	p.A5V	40	Female	Cedars-Sinai	CS07iALS-SOD1A4
ALS-SOD1 II	SOD1	p.G94A	57	Male	Cedars-Sinai	CS2RJViALS

Reverse transcription was performed on 1 µg of total RNA with Superscript™ III first-strand synthesis system (Invitrogen/Thermo Fisher) also according to the manufacturer's instructions. All samples were processed at the same time, the resulting cDNA diluted 1:10 in nuclease-free water. Real-time reactions were run in triplicate using 5 µl/reaction with Fast™ SYBR Green Master Mix (Applied Biosystems/Thermo Fisher). Real-time PCRs were performed on a 96-well format using the QuantStudio3™ (Applied Biosystems/Thermo Fisher). The experiment was performed blinded, according to standards of the MIQE guidelines.⁸ These experiments were performed and analysed under double-blind conditions.

RNA-sequencing

For RNA-sequencing (RNA-seq), messenger RNA was purified from total RNA using poly-T oligo-attached magnetic beads. After fragmentation, the first-strand cDNA was synthesized using random hexamer primers, followed by the second strand cDNA synthesis using either dUTP for directional library or dTTP for the non-directional library. For the non-directional library, it was ready after end repair, A-tailing, adapter ligation, size selection, amplification and purification. The directional library was ready after end repair, A-tailing, adapter ligation, size selection, USER enzyme digestion, amplification and purification. The library was checked with Qubit and real-time PCR for quantification and a bioanalyser for size distribution detection. Quantified libraries will be pooled and sequenced on Illumina platforms, according to effective library concentration and data amount. The clustering of the index-coded samples was performed according to the manufacturer's instructions. After cluster generation, the library preparations were sequenced on an Illumina platform and paired-end reads were generated.

DNA isolation

DNA from DIV28 hiPSC-MNs grown on six-well plates was isolated using the QIAamp DNA Mini Kit (Qiagen, Cat. No. 51306) as per the manufacturer's instructions with slight modifications. Briefly, two wells were first washed with 1× PBS (Gibco), and cells were scraped gently, collected in 1× DPBS, centrifuged for 5 min at 300g and the pellet was resuspended in 200 µl of 1× DPBS. Next, 20 µl of proteinase K and 200 µl of buffer AL were added, briefly vortexed, heated at 56°C for 10 min with mixing at 350 rpm and centrifuged briefly. Then, 200 µl of 100% ethanol was added, briefly vortexed and centrifuged. The entire contents were transferred to QIAamp mini

spin columns, and centrifuged for 1 min at 8000 rpm. Next, 500 µl of buffer AW1 was added and centrifuged for 1 min at 8000 rpm, then 500 µl of buffer AW2 was added and centrifuged at 14 000 rpm for 3 min first and then for 1 min. The contents of the column were eluted using 20 µl nuclease-free DEPC-treated distilled water (Roth), by incubating for 5 min at room temperature and centrifuging for 1 min at 8000 rpm. This step was repeated once more after adding the eluate back into the column to further concentrate the DNA in the samples. The DNA concentration and purity were estimated using a spectrophotometer (NanoDrop, ThermoFisher), and the samples were snap frozen in liquid N₂ and stored at –80°C until shipment.

Whole genome bisulphite sequencing

After sample DNA testing by agarose gel electrophoresis, positive control DNAs were added into the DNAs and they were fragmented into 200–400 bp using Covaris S220. Next, terminal repairing, A-ligation, methylation sequencing and adapter ligation were performed to the DNA fragments. The final DNA library was ready after bisulphite treatment (EZ DNA Methylation Gold Kit, Zymo Research; after bisulphite treatment, unmethylated cytosine will change into uracil, whereas methylated cytosine will stay unchanged), size selection and PCR amplification steps. Library concentration was first quantified by Qubit2.0, and then was diluted to 1 ng/µl before checking the insert size on an Agilent 2100 and being quantified with more accuracy by quantitative PCR (effective concentration of library >2 nM). After passing library testing, different libraries were pooled together and then fed into Illumina devices according to effective concentration and expected data volume. The sequencing strategy was paired-end sequencing.

RNA-seq analysis

Raw reads were mapped using HISAT2^{9–11} mapped reads (concordant) were kept for downstream analysis and gene counting was performed using featureCounts.¹² Data normalization and differential expression analysis were performed using the R package limma.¹³ Heatmaps, principal component analysis and radar plots were also generated using R. Gene Set Enrichment Analysis was performed using the GSEA application.^{14,15} Statistical significance was set at adjusted P-value cut-off of 0.05 and false discovery rate of 0.25. Self-organizing maps (SOMs) were conducted using the Bioconductor package oposSOM.¹⁶

For the published RNA-seq datasets, the raw gene expression count matrix was downloaded from the Gene Expression Omnibus repository. GSE112681 (397 and 645 ALS and control samples; referred to in the paper as 'blood') and GSE137810 (963 and 280 ALS and control samples under subseries GSE153960, respectively; filtered for ALS motor neuron disease and control set only, referred in the manuscript as 'spinal'). Downstream analyses were performed as described in the sections on RNA-seq and Deep-learning analysis.

Whole genome bisulphite sequencing analysis

Paired-end whole genome bisulphite sequencing (WGBS) read mapping, methylation and single nucleotide polymorphisms (SNP) calling, extraction and reporting were performed using gemBS.¹⁷ Differential methylation analysis and overlap analyses were conducted using R and Bioconductor packages. HOMER was used for annotation of methylation sites and motif and ontology analyses.¹⁸ Additional enrichment analyses were performed using Gene Set Enrichment Analysis.^{13–15} Single nucleotide variants were filtered using Bcftools (<http://samtools.github.io/bcftools/bcftools.html>) to exclude known SNPs and low-quality variants. Circos plots were generated using the online resource described by Krzywinski and colleagues (<http://circos.ca/>).¹⁹

Analysis of single nucleotide variants

Analysis of single nucleotide variants and mutational signatures of WGBS and published whole genome sequencing (AnswerALS)²⁰ was performed using R and the Bioconductor package SomaticSignatures.²¹ Somatic signatures were generated according to previously described methodologies^{6–22} by looking at six different types of mutations and their 16 (4 × 4) possible flanking base combinations at the 5' and 3'. This led to a maximum of 96 (16 × 6) tri-nucleotide motifs that were used to generate the ALS mutational signatures.

Monte Carlo simulation

After generating empirical information of overlap between differential analyses of the four familial ALS groups against the control samples, Monte Carlo Simulation was performed to quantify the chance of randomly encountering such overlaps. We then generated a randomly reshuffled gene expression matrix of our dataset, and performed differential analysis and looked at overlaps between the different comparisons. This step was iterated 1000 times to calculate the average rate of random overlaps. *P*-values of the significance of the number of empirical versus random overlaps (in at least two, three or four comparisons as shown in [Supplementary Fig. 4](#)) were obtained by using the formula

$$P = (r + 1)/(n + 1)$$

where *r* is the number of iterations in the random overlap analysis that had higher overlap values than the empirical overlap count whereas *n* is the total number of iterations (1000).

Deep-learning analysis

An R implementation of the Keras/TensorFlow²³ binary classifier was used for the deep-learning analysis. Glorot uniform initializer and hyperbolic tangent activation were used for kernel initialization and kernel activation, respectively, with stochastic gradient

descent as a model optimizer and binary cross-entropy for loss measurement. An initial layer of 16 units with an additional three hidden layers of 16 units was used, with a final output layer of one binary unit.

Ethical approval

All procedures with hiPSCs have been performed in compliance with the guidelines of the Federal Government of Germany within the context of the German Network for Motor Neuron Diseases (MND-NET) and have been approved by the ethical committee of Ulm University (19/12). All participants gave informed consent for the study. The use of human material was approved by the Declaration of Helsinki concerning Ethical Principles for Medical Research Involving Human Subjects, and experiments were performed according to the principles set out in the Department of Health and Human Services Belmont Report. All experiments with human blood were conducted in accordance with approval by the Research Ethics Committee of the academic hospital Leuven with informed and signed consent from individuals (S65097, S59292 and S60803).

Data availability

The in-house RNA-seq and WGBS datasets are available on request. We also used publicly available datasets, which are accessible under the Gene Expression Omnibus accession numbers GSE112681, GSE137810, GSE106382 and from the AnswerALS consortium.

Results

Gene-centric RNA-seq analysis reveals transcriptional similarities in familial ALS

First, we investigated the presence of a transcriptional signature shared by different ALS cases using hiPSCs from eight familial (fALS) patients and three healthy controls without known pathologies. The patients carried mutations within the C9orf72 (1.8 and 2.7 kb GGGGCC expansion), FUS (c.1484delG and c.1504delG), SOD1 (p.A5V and p.G94A) and TARDBP (p.G298S and p.N390D) mutations ([Table 1](#)).^{24–26} HiPSCs from all genotypes were differentiated into spinal MN⁴ and their transcriptomes analysed after 28 DIV, thus anticipating the later time points at which MN loss has been observed in ALS cultures.^{4,27} The analysis was first performed by pooling the patients according to the mutated genes (gene-centric approach) and identified differentially expressed genes (DEGs) in all the comparisons to healthy controls ([Supplementary Fig. 1](#)). Principal component analysis ([Fig. 1A](#)) separated the control from the ALS genotype and highlighted strongly divergent pathologic transcriptional programs characterizing the four mutant groups, which were mainly clustered according to their mutated gene. This separation was also confirmed by hierarchical clustering performed with the top factor loadings for PC1 and PC2 ([Supplementary Fig. 2](#)), where control samples were grouped in a distinct cluster (purple). ALS-C9orf72 and ALS-SOD1 were part of the same cluster (pink) but they also showed segregation according to their gene-centric mutational group. ALS-TARDBP clustered along with one FUS patient (light-blue cluster), whereas the remaining FUS samples formed a separate cluster (light-green). Even though the different mutant samples mainly clustered according to their genetic background, we observed also some degree of variability between patients carrying different mutations in the same gene, as in the case of FUS. For this reason, we controlled whether our gene-centric

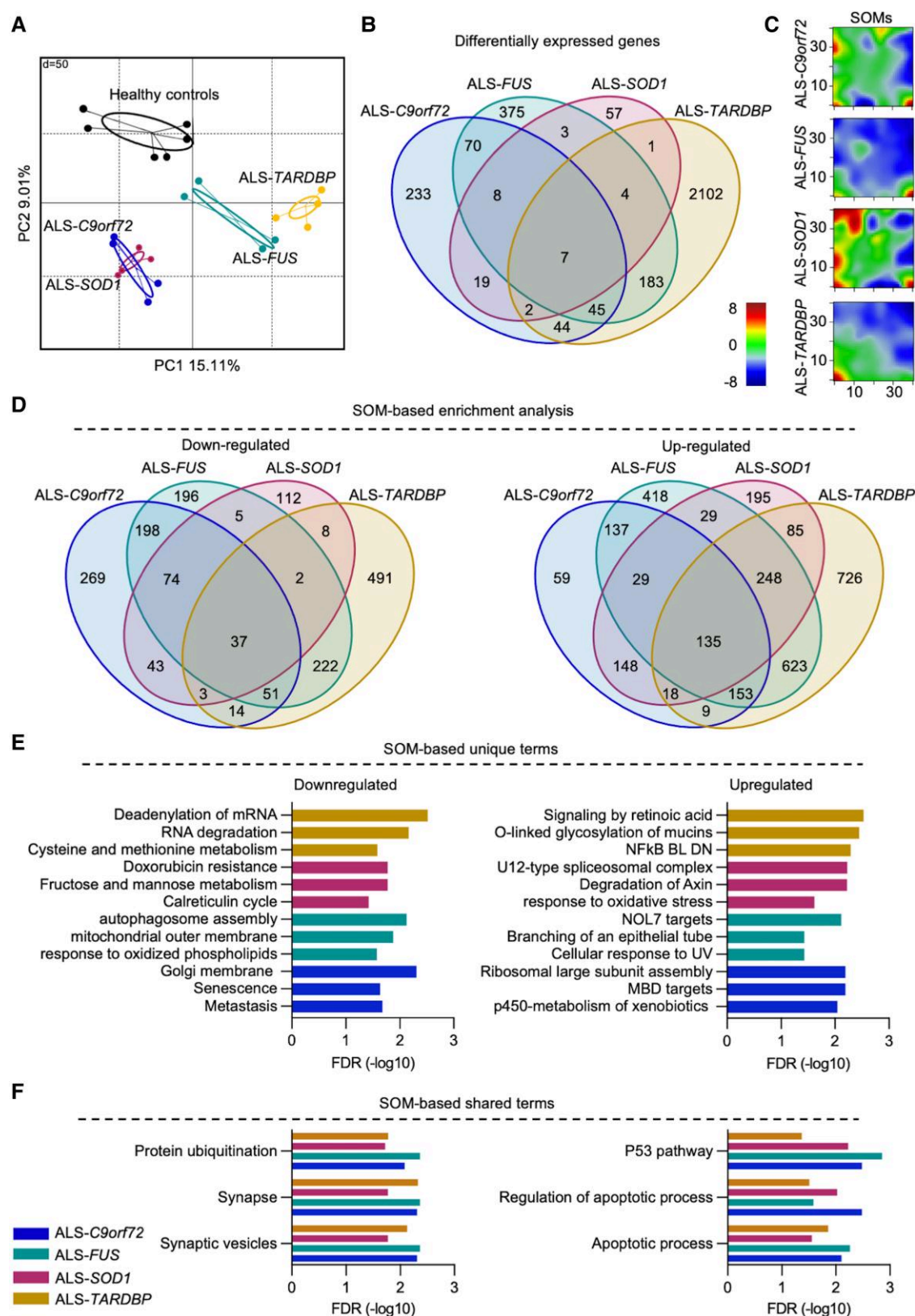


Figure 1 RNA-seq reveals divergent and convergent transcriptional features in ALS. (A) Transcriptome-based principal component analysis grouping of the different hiPSC-derived MN (two independent replicates for each cell line) according to their mutated gene (two patients for each mutation and three controls). (B) Venn diagram highlighting the shared DEGs among the ALS mutations: the only seven genes commonly altered in all the ALS genotypes did not significantly enrich for any biological term. (C) SOMs representation of the transcriptional landscape of the different ALS groups. The colour scale in the maps represents the gene expression level expressed as relative to the mean of the portrait (0 value, green colour) after normalization. (D) Venn diagrams showing the number of SOM-based significantly down- and upregulated terms in the different ALS groups compared to healthy controls. (E) List of the top three unique and (F) shared significantly down- and upregulated terms based on the SOMs.

approach might fail in capturing some important patient-specific alterations, resulting in a simplification of the pathological transcriptome characterizing the different individuals. To do so, we compared the transcriptome of hiPSC-derived MN from each patient line to the healthy controls to individually identify differentially down- or upregulated transcripts. We then compared these DEGs to those highlighted by the gene-centric approach and found a strongly significant correlation between the dysregulated genes identified by the two approaches in all the comparisons (Supplementary Fig. 3A and B). This confirmed that patients with different mutations within the same gene are characterized by strongly convergent pathological transcriptomes (as compared to healthy controls), which are efficiently captured by our gene-centric approach. Based on this evidence, we confirmed the gene-centric statistical strategy as the best fit for the scope of our investigations.

In agreement with the gene-centric heterogeneity within the ALS genotype, we found that only seven transcripts (the genes *AGAP7P*, *HOXC8*, *MIR939*, *PDLIM1* and the novel transcripts *ENSG00000228613*, *ENSG00000272428*, *ENSG00000275216*) were significantly altered in all the ALS subtypes when compared to control MN, whereas most of the DEGs were either gene-specific or shared among a reduced number of ALS subgroups (Fig. 1B). Nevertheless, Monte Carlo simulation could not reproduce the degree of shared DEGs' expression across the different mutations when performing 1000 iterations with random transcripts (Supplementary Fig. 4). This ruled out randomness as the cause of the overlaps observed in the different transcriptomes of the ALS mutants.

On the basis of the transcriptional heterogeneity within the ALS genotype, we asked whether the DEGs observed in the ALS subgroups might converge into common biological pathways and created SOMs to identify a common expression pattern. This confirmed a significantly different transcriptional landscape among the genes considered (Fig. 1C), as most of the significantly down- and upregulated terms identified using SOMs were either exclusive to or only partially shared among the ALS subgroups (Fig. 1D). In fact, the top three enriched terms identified in each mutant genotype by the SOMs appeared to be specific for the respective gene. When looking at the downregulated terms, we found that metastasis, senescence and Golgi membrane were enriched in ALS-C9orf72 whereas autophagosome assembly, mitochondrial outer membrane and response to oxidized phospholipids were specific for ALS-FUS. SOD1-mutant cultures showed reduced expression of the genes involved in the calreticulin cycle and doxorubicin resistance as well as monosaccharide metabolism, whereas terms related to RNA processing and amino acids metabolism were downregulated in ALS-TARDBP. In contrast, upregulated genes associated with p450, ribosomal assembly and MBD targets were specific for ALS-C9orf72; NOL7 targets, epithelial branching and cellular response to UV were detected in the transcriptomes of FUS mutants. In ALS-SOD1 we detected high expression of genes involved in axin degradation and oxidative stress as well as in spliceosomal complex, whereas ALS-TARDBP MN were characterized by upregulation of mucins' glycosylation, NFkB and retinoic acid signalling (Fig. 1E).

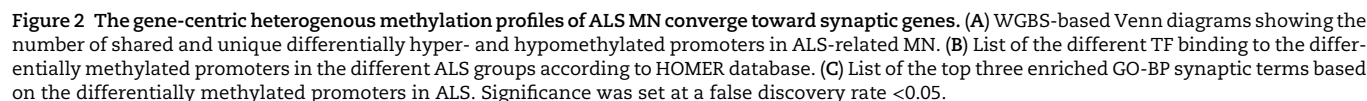
Notably, when we specifically looked at the significantly altered terms shared by all the subgroups (37 down- and 135 upregulated), we found that the genes showing reduced expression in ALS were mainly involved in synaptic processes and ubiquitination. In contrast, most of the upregulated terms were related to apoptosis, stress and DNA damage (Fig. 1F). Thus, despite the different ALS subgroups being characterized by strongly heterogeneous

transcriptomes, we could identify a specific set of biological alterations mainly related to cell death and to the synaptic microenvironment, highlighting a potential crucial role of the synapse in ALS pathology independently from the specific underlying mutation.^{4,28}

The ALS transcriptional signatures correlate with epigenetic alterations

We then aimed to uncover whether the shared transcriptional alterations in ALS might originate from epigenetic abnormalities by performing WGBS with samples obtained from the same hiPSC lines after differentiation into MN. In line with the transcriptome data, we noticed a different methylation pattern characterizing the different ALS mutations (Supplementary Fig. 5A) and most of the differentially methylated regions (DMRs) were either unique or partially shared among the disease subgroups (Supplementary Fig. 5B). Since the promoters were the most significantly enriched regions identified by WGBS after the CpG islands (Supplementary Table 1), we focused on these specific DNA sequences. When we looked at the DMRs within the promoter regions, we identified a very restricted number of shared terms among the different ALS cases. In particular, there was no hypermethylated and only one hypomethylated DMR shared by all four subgroups (Fig. 2A), again highlighting a deep heterogeneity within the different ALS subtypes also on the epigenetic level. This was further confirmed when we predicted the transcription factors (TFs) binding to the promoters identified by DMRs analysis. Among the significantly enriched TFs, Slug was the only one predicted to bind differentially hypermethylated promoters in all the ALS MN cultures. By contrast, we could not predict any shared TF whose hypomethylated binding motif was shared by all the ALS subtypes (Fig. 2B). Performing gene enrichment analysis by feeding the targets of the TFs significantly linked to the DMRs into the Reactome database also highlighted a high degree of variability within the ALS genotype: in fact, we did not identify any communality in the top enriched terms associated with hyper- or hypomethylated motifs in the ALS subtypes (Supplementary Fig. 6). Interestingly, when we looked at the enriched gene ontology biological processes (GO-BP) we found that terms associated with synaptic morphology and function were significantly enriched in the hyper- as well as hypomethylated promoters of all ALS-related MNs (Fig. 2C). Thus, epigenetic abnormalities contribute to the synaptic alterations and appear to be shared among ALS cases with different genetic backgrounds.

We next aligned the RNA-seq and WGBS data to evaluate, at a single-gene level, the degree of correlation between transcriptional and epigenetic alterations defying ALS. As expected, we found a significant correlation between hypermethylated promoters and downregulated transcripts, as well as between hypomethylation and higher RNA levels, in all the ALS subtypes analysed (Supplementary Fig. 7A). By looking at the leading edges of these correlations, we identified three hypermethylated/downregulated and 23 hypomethylated/upregulated genes shared by all the ALS mutants (Supplementary Fig. 7B). Interestingly, by performing enrichment analysis with the leading edges genes on the basis of the RNA-seq-WGBS correlation we found reduced synaptic terms in FUS, SOD1 and TARDBP cases, but not in ALS-C9orf72 (Supplementary Fig. 8 and Supplementary Table 2). This suggests that the altered expression of synaptic transcripts in presence of GGGGCC expansion might originate from the reduced levels of C9orf72 protein observed in ALS,^{29–32} rather than from their methylation pattern. By contrast, we found that the hypomethylation of the *LY6E*, *LY6H*, *LYNX1* and *PSCA* promoters (all located in



chromosome 7)³³ significantly enriched the acetylcholine receptor binding term in all the subtypes considered, in agreement with their role in controlling the trafficking, assembly and function of nicotinic acetylcholine receptors (Supplementary Fig. 9 and Supplementary Table 3).³⁴

Deep machine-learning defines an ALS transcriptional portrait converging on toll-like receptor activation

We then reasoned to identify an ALS signature by overcoming the reduced degree of transcriptional similarities highlighted by the analysis of hiPSC-derived MN. We considered a whole blood transcriptome dataset from 397 ALS patients and 645 healthy controls (GSE112681)³⁵ to train a deep-learning algorithm and identify an ALS-specific transcriptional profile. Here, 70% of the data was used as training input, while the remaining 30% validated the prediction accuracy of the model (Fig. 3A). Afterwards, we generated a correlation matrix for all the mapped genes and scored their degree of correlation to the prediction model. Enrichment analysis based on the terms identified by the deep-learning approach highlighted that terms related to toll-like receptor (TLR) cascade, immune response and autophagy were significantly associated with ALS (Fig. 3B). In a second layer of analysis, we applied our deep-learning method to an independent transcriptome originating from the spinal cord of ALS patients (GSE137810)³⁶ (Fig. 3C), which also highlighted a downregulation of 141 synaptic transcripts that, according to the SynGO database,³⁷ were significantly associated with pre- and postsynaptic terms, including synaptic vesicles and neurotransmitter release (Supplementary Fig. 10, Supplementary Table 4). In addition, we detected a significant association of TLR signalling and immune response, as well as clathrin-mediated endocytosis, vesicle budding and TP53 regulation with the spinal transcriptome of ALS patients (Fig. 3D). On the basis of these striking similarities, we investigated to which extent the blood and spinal cord RNA-seq datasets share common transcriptional alterations linked to ALS. To our surprise, we found a strongly significant degree of correlation between these two datasets (Fig. 3E), indicating that the transcriptional profile defining ALS is recapitulated in both blood and spinal cord tissue. We next filtered the top 5% ALS predictor genes belonging to the blood transcriptome, as well as the top 5% belonging to the spinal cord one (Supplementary Fig. 11), and overlapped these gene lists: we identified 14 statistically significant transcripts positively (RGS18, LY96, SKAP2, AQP9, TLR2, TLR6, TLR8, CTSS, EVI2B, CD58, FAM126B, NECAP1, PP3CB and OXR1) and only one negatively (PDE4C) associated with both tissues in ALS (Fig. 3F). Of note, IL18 displayed positive correlation in the spinal cord but negative in the blood of samples derived from patients. We then performed enrichment analysis with the 14 genes positively correlating to ALS and found that, despite their low number, they significantly enriched in pathways mainly related to the pro-inflammatory TLR signalling (Fig. 3G). This data strengthens the relevance of inflammatory biomarkers detected in the blood of ALS patients,³⁸ as we identified a list of novel transcripts systemically altered in individuals suffering from motor neuron disease.

We then overlapped our ALS transcriptional signature to the SOM-based profiles generated from the transcriptomes of hiPSC-derived MN, to resolve the mutation-specific discrepancies observed in ALS. We observed a strong correlation tendency in the case of ALS-C9orf72, ALS-FUS and ALS-SOD1, which reached significance in ALS-TARDBP (Fig. 3H). To rule out possible biases

arising from our culture setup, we analysed an independent transcriptome (GSE106382),²⁶ which included hiPSCs from SOD1, FUS, TARDBP and sporadic patients. Again, we could observe a clear separation between controls and all the ALS lines (with the exception of a minimal overlap between the control and sporadic groups because of one healthy individual), as well as a strongly divergent transcriptional program in the different ALS mutations (Supplementary Fig. 12A). SOMs from the different mutant subgroups based on these datasets (Supplementary Fig. 12B) confirmed the heterogeneous landscape characterizing the different mutations at transcriptional level. In agreement with these findings, this microarray-based transcriptome highlighted strong differences between the ALS subgroups that did not share any downregulated term, whereas there were 128 commonly shared upregulated ones (Supplementary Fig. 12C). We then checked the correlation between the deep-learning data and the different transcriptomes of this independent dataset. In line with the data from the in-house cell lines, we observed a strong correlation between all the ALS subtypes and the deep-learning signature, with the SOD1 and TARDBP cases showing the best performance and reaching statistical significance (Supplementary Fig. 12D). This strengthened the idea of common transcriptional alterations underlying the different ALS subtypes independently from the specific pathogenic mutation. Indeed, we identified a significant correlation between pathways associated with TLR and the transcriptomes of fALS hiPSC-derived MN, indicating that this specific biological process significantly contributes to ALS manifestation (Fig. 3I). To better confirm this evidence, we performed single-tube quantitative PCR using blood from an independent cohort of 10 sALS, 10 fALS patients and 10 healthy controls without known pathologies (Supplementary Table 5) to analyse eight transcripts selected from the group of genes identified by overlapping the two deep-learning datasets (Fig. 3F): LY96, TLR2, TLR6, TLR8, EVI2B, CD58, FAM126B, PP3CB. The expression levels of these genes were then integrated, under double-blind conditions, into the deep-learning algorithm to evaluate the accuracy of genotype prediction. In agreement with the significant correlation between this gene set and fALS transcriptome, the expression of these transcripts was sufficient to significantly recognize the familial patients and separate them from the other two groups (Supplementary Fig. 13). Thus, these genes might represent novel transcriptional biomarkers for the identification of fALS cases from blood samples.

The ALS mutational profile correlates with ‘ageing’ signatures

Finally, we asked whether the ALS portrait hereby described could be explained by a specific pattern of genomic mutations commonly shared in ALS and independent from the pathogenic ones. To identify the mutational processes underlying ALS and the probable biological factors associated with them, we generated somatic mutational signatures of 866 samples (773 ALS patients and 93 healthy controls) from the whole genome sequencing dataset of AnswerALS.²⁰ Briefly, a somatic mutational signature analysis looks at not only the base substitutions but also the 5' and 3' flanking bases to generate specific tri-nucleotide motifs whose frequencies can be mathematically analysed to deduce mutational signatures that inform the likely origin of the mutation.^{6–22}

We first generated 96 non-negative matrix factorization-based somatic signatures (the maximum possible number of signatures) (Fig. 4A and a representative plot is shown in Fig. 4B). We then

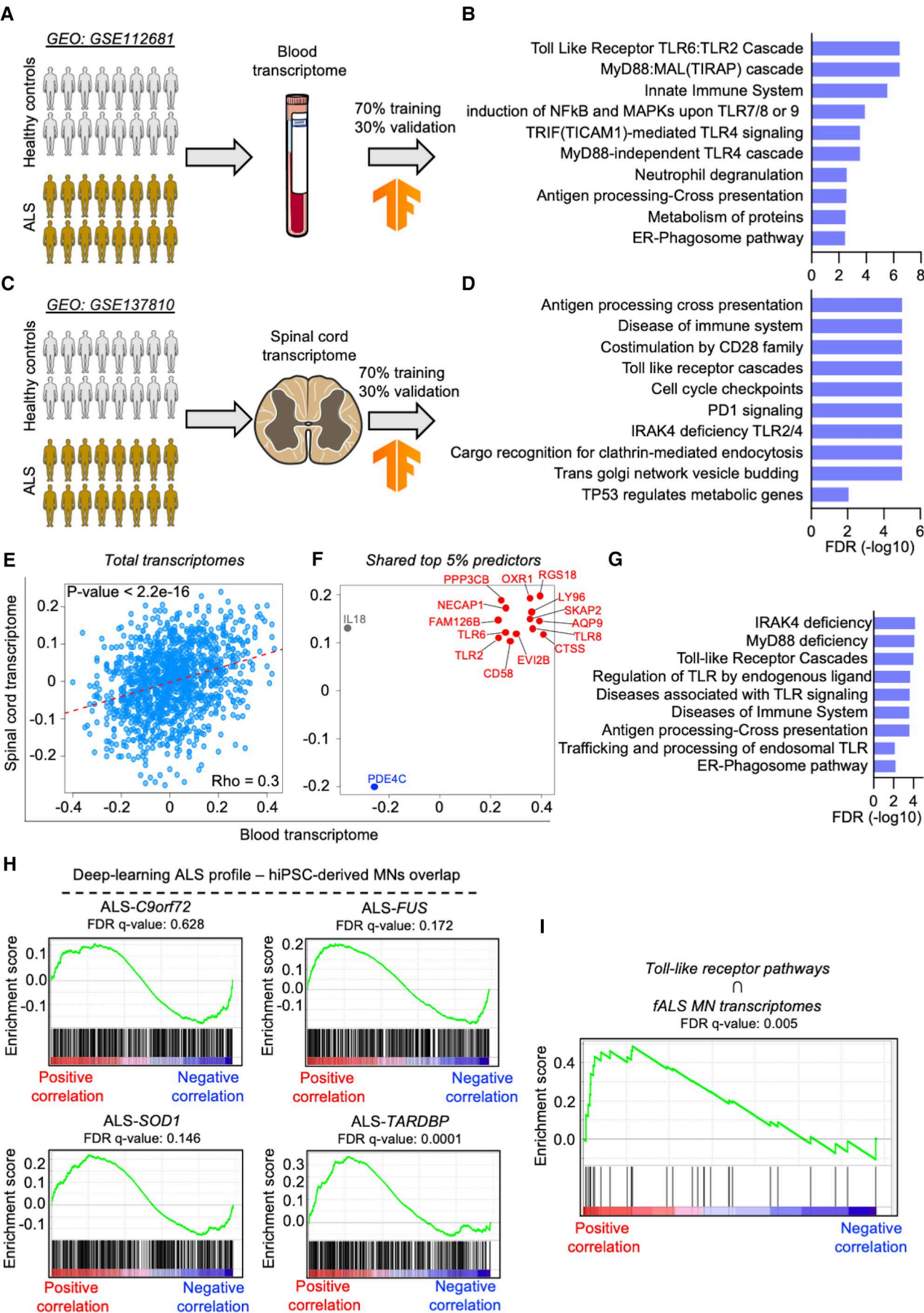


Figure 3

(continued)

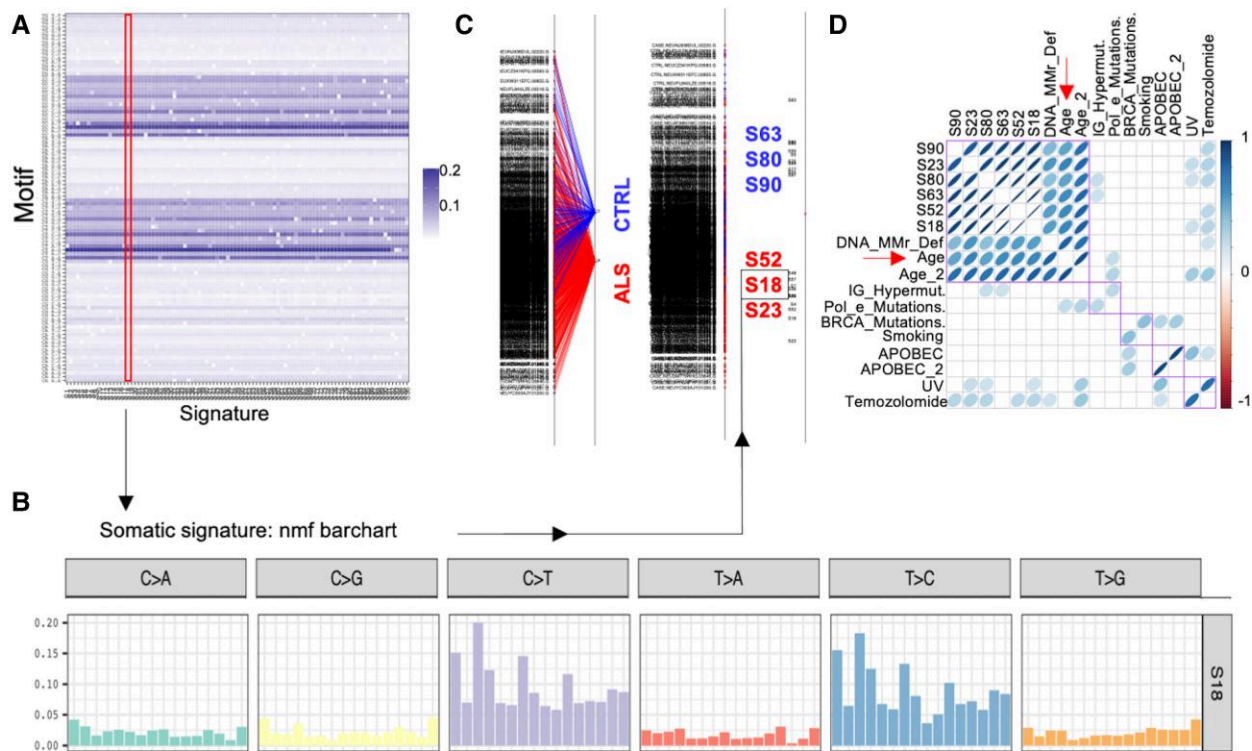


Figure 4 The mutational load of ALS genomes correlates with DNA damage and ageing signatures. (A) Mutational signatures were generated from single nucleotide variants of the AnswerALS whole genome dataset. We generated the maximum number of signatures that can be generated (96) using non-negative matrix factorization (nmf). We show here the contribution of the 96 tri-nucleotide motif combinations (y-axis) and the calculated signatures (x-axis). Signature 18 (which will be discussed in subsequent panels) is highlighted (red area). (B) A representative signature (S18) mutational spectrum profile is shown here, where a T>C and C>T substitution composed motifs (flanked by different combinations of 5' and 3' bases) predominantly contribute to the signature. (C) We performed a between groups analysis classifier to assess whether we can distinguish between ALS (red) and control (CTRL; blue) samples. All signatures are shown on the left, followed by centroids of ALS and CTRL samples (middle), and a highlight of six signatures that had the highest loading (weight) in classifying ALS and CTRL samples (right). (D) Correlation analysis between the six top signatures (S63, S80, S90, S52, S18 and S23; displayed in C) and publicly available ones with a known biological association. The blue spheres are drawn only for the correlations with a significant P-value at a cut-off of 0.05. The density of the colour and shape of the sphere indicate the strength of a positive correlation. Of note, the top 6 signatures are not only significantly correlated to each other but also to signatures associated with DNA mismatch repair deficiency (DNA MMR def.) and ageing signatures (Age and Age_2). The purple squares represent the ordering of terms based on unsupervised hierarchical clustering based on the correlation matrix.

used between group analysis and support vector machines to identify signatures that could discriminate between ALS and controls samples. As a result, we identified six signatures that showed the strongest association with either ALS or control centroids (Signatures 18, 23, 52, 63, 80 and 90; Fig. 4C). In the final analysis, we correlated the tri-nucleotide motifs of these top six significant signatures to publicly available ones with defined underlying biological association.⁷ Interestingly, we observed that all the six signatures showed statistically significant correlation to ageing ('Age' and 'Age2') and DNA mismatch and repair deficiency

('DNA_MMR_Def.') (Fig. 4D; $P < 0.05$). Additionally, we also observed that some signatures, to a lesser extent, showed significant correlation to Immunoglobulin hypermutation ('IG_Hypermut.'), Ultraviolet ('UV') and Temozolomide ('Temozolomide'). We thus conclude here that most of the mutations that are acquired by ALS patients are largely attributable to the ageing phenotype and associated biological factors like the ageing-dependent decline in DNA repair efficiency, in agreement with the increased p53 signalling characterizing the transcriptome of the fALS cases considered (Fig. 1F).

Figure 3 (Continued)
Machine-learning identifies defines an ALS transcriptional signature converging on TLR signalling. (A) Schematic representation of the deep-learning analysis based on the whole blood transcriptome from ALS patients and healthy controls. The analysis was performed by using 70% of the data as training dataset and the remaining 30% for the validation. (B) Significant enrichment of pathways associated with inflammation and TLR signalling. (C) A second layer of deep-learning analysis based on spinal cord transcriptome from ALS patients was performed following the same procedure as used for the blood transcriptome. (D) This also showed higher activation of pro-inflammatory and TLR pathways in ALS. (E) Correlation plot showing a significant overlap of the blood and spinal cord total transcriptomes after deep-learning machine analysis. Model correlation coefficients were obtained for each dataset to assess significant correlation in gene ranks using Spearman correlation coefficient. After sorting the top 5% predictor genes of ALS, we identified 14 transcripts (highlighted in red) positively associated with ALS in both transcriptomics (F), which clustered in TLR, inflammation and autophagy pathways (G). (H) Enrichment plots showing the positive correlation degree between the transcriptional signature based on the double-layer (blood and spinal cord) deep-learning analysis and the different ALS genotypes. (I) The enrichment in TLR-related transcripts highlighted by the deep-learning approach positively and significantly correlated with the transcriptional program of familial ALS MN.

Discussion

The genetic heterogeneity of ALS reflects the vast number of biochemical alterations that have been associated to this disease and represents a major obstacle for the development of novel effective treatments. Indeed, a large portion of the newly designed therapeutic strategies undergoing clinical trials, such as antisense oligonucleotides,^{39,40} aim to target pathomechanisms linked to specific genetic causes. Still, the genetic and pathobiological complexity characterizing this disease cannot *a priori* exclude the presence of pathological features commonly shared across the ALS spectrum. The identification of common alterations would not only improve our understanding of the mechanisms underlying this neurodegenerative disease, but might also open new therapeutic scenarios for a large portion of patients. Such an approach is made possible by the development of novel mathematical and machine-learning tools that, together with the increased accessibility to deep-sequencing analysis, offer the possibility of integrating different layers of biological information with the final aim of identifying druggable targets, as well as diagnostic and prognostic markers for human diseases.^{41–43} To apply this strategy to ALS, we performed the first integrative, machine-learning-based analysis of transcriptomic, epigenetic and genetic data obtained from human cultured MN, blood and post-mortem neuronal samples. In the first step of our analysis, we looked at the transcriptional similarities between *C9orf72*, *FUS*, *SOD1* and *TARDBP*-mutant MN. By performing independent gene-centric analysis we highlighted that the ALS subgroups, when compared to the healthy controls, were mainly characterized by discrepant alterations. This confirmed previous descriptions of divergent transcriptome changes in the presence of diverse pathogenic backgrounds,^{44,45} which have been observed even when comparing single hiPSC lines with different mutations within the same gene.⁴⁶ The inter-patient variability represents indeed an ongoing challenge for hiPSC-based studies, where the efficacy of neuronal differentiation, cellular composition and donor sex represent a source of additional experimental variabilities.⁴⁷ Still, despite the large transcriptional discrepancies documented when investigating ALS genes or even single variants using hiPSC-based models, some studies have already identified a certain degree of transcriptional overlap across the disease spectrum. For example, previous transcriptomic analysis including MN from *FUS* and *TARDBP*-patients,²⁷ as well as sporadic, *C9orf72*, *SOD1* and *TARDBP*-mutant cells,⁴⁸ revealed altered monoamine and lipid metabolism in ALS compared to controls, while single-cell RNA-seq identified common signatures in sporadic and *C9orf72* familial cases.⁴⁹ In addition, the phenotypical manifestation of disease features in ALS MN, such as mitochondrial dysfunction and electrophysiological impairment, reveals a higher degree of coherence across different cases than the underlying transcriptional diversities observed in traditional comparative analysis.^{46,50} Accordingly, even though we used a restricted number of iPSC lines, we showed that the major determinants of a pathologic transcriptional landscape are conserved across patients even in the presence of different variants in the same gene. These observations indicate that, because of the high-dimensional nature of RNA-seq data, a significant amount of relevant information might be kept hidden if the maximum potential of computational and statistical methods is not fully exploited. Indeed, by performing unsupervised machine-learning we could achieve a significantly higher coverage of the transcriptional commonalities characterizing the different ALS cases included in our study. Importantly, the increased representation of pro-apoptotic pathways and p53 signalling in all

the ALS groups confirmed the major feature of this disease, namely a high MN vulnerability leading to the selective loss of this neuronal population. On the other hand, we could identify a significant loss of synaptic transcripts as a novel commonly shared feature of ALS. Synaptic disruption is observed in several neurodegenerative conditions, but has often been considered as a consequence of dying neurons rather than a pathological mechanism actively contributing to neuronal death in ALS. Thanks to our top-down integrative approach, we could show that synaptic abnormalities are not only occurring at the transcriptional level, but also at the epigenetic one, indicating that these alterations are unlikely to result from the apoptotic processes ongoing in mutant cultures. In fact, even if the presence of heterogenous pathogenic mutations differentially affects the methylation pattern of the ALS genome, these epigenetic changes still significantly converge towards synaptic terms. Thus, our results are in support of a central role played by the synaptic microenvironment in disease progression, as also shown by previous works demonstrating neuroprotection through re-establishment of synaptic structure and physiological properties.^{4,51} In addition, a significant loss of synaptic transcripts was observed also in the ALS post-mortem samples, confirming the translational relevance of these findings and highlighting the potential of integrative analysis including patients' biopsies.⁴⁹ Even though a misbalanced contribution of different cell types might have a significant impact on the bulk transcriptome of ALS spinal cords due to MN loss and glial activation,³⁶ our multi-tissue approach revealed a subset of genes biologically associated with TLR signalling. This subset of genes significantly resembled an ALS portrait in multiple samples, including blood cells and confirmed the biological relevance of this pathway in ALS progression.⁵² The striking correlation observed between the blood and spinal transcriptomes represents an important novelty for the neurodegeneration field, as we could highlight disease-specific alterations even in cells not directly linked to the pathology. This opens the fascinating possibility of using transcriptional biomarkers obtainable through low-invasive biopsies for diagnostic and prognostic scopes. Indeed, the advantage of blood cell-based transcriptomes has also been described in an elegant integrative study aimed at identifying signatures of vulnerability across different neurodegenerative disorders.⁴² This evidence supports the need for producing more large-scale omics datasets for ALS,^{20,53} which at the moment are very limited and often include a reduced number of healthy controls. This represents a limitation that has to be kept in mind even in regards to this study, since the inter-individual variability characterizing omics experiments is a source of larger inconsistency within healthy controls groups than in patients, which still share the underlying traits of the disease considered. Indeed, the markers linked to TLR identified here failed in differentiating sporadic patients from healthy controls. This could be, partly, due to the fact that the genes we used were those with high expression and predict likelihood of ALS class/phenotype in the deep-learning model of both datasets (Fig. 3F; genes shown in the top right quadrant, labelled in red). In contrast, we did not identify genes (except for *PDE4C*) whose expression is highly correlated to the control group. Additionally, controls are often highly heterogenous in their transcriptional profile due to many confounding factors and are usually selected merely due to the absence of a given pathology. A well-controlled and stratified sampling approach, combined with a sufficiently large sample size, can overcome this problem and improve the deep-learning model development.

Our data still provide important proof-of-principle evidence on the possibility of circumventing the heterogeneity that characterizes ALS,

with the goal of identifying common alterations defining this disease. We demonstrate the presence of an ALS-related pathological portrait not only at the transcriptional and epigenetic levels, but we also identified a mutational signature specific for this neurodegenerative disorder. Even though our experimental design included a restricted number of ALS patients and a limited cohort of healthy controls, we could show that ALS significantly correlates with ageing and DNA damage signatures by mapping the mutational load within their genome.⁷ Ageing represents the major risk factor for neurodegenerative diseases⁵⁴ and it has been shown that ALS disrupts the expression and network of ageing-associated genes.⁴⁹ Thus, we demonstrate that alterations observed on physiological ageing in ALS⁵⁵ occur even at the genetic level, where the genomic mutational load of patients mirrors the prognostication of other detrimental features, such as accumulation of DNA damage, observed at the cellular levels.²⁵

In conclusion, our integrated multiomics approach identified multiple layers of commonalities that provide a crucial insight into how different mutations and divergent pathomechanisms can still converge over time into a singular presentation of ALS disease. This study represents a novel and important step towards the identification of common molecular alterations and, with the help of larger cohorts and deeper omics datasets, of a unifying signature across the heterogenous spectrum of this neurodegenerative disease.

Acknowledgements

The authors are grateful to Sabine Seltenheim for the excellent technical support. The authors want also to thank all the donors who made this study possible.

Funding

This project was financed by the Else Kröner-Fresenius-Stiftung (project number 2019_A111 to A.C.) and by the Deutsche Forschungsgemeinschaft (German Research Foundation) - SFB1506 'Ageing at interfaces' (project A01 to A.C. and T.M.B.; project B01 to M.M.). D.S. received financial support from the 'Experimental Medicine' graduate program of the Medical Faculty of Ulm University. P.M. has a research Fellowship of the European Academy of Neurology (no award/grant number). P.V.D. holds a senior clinical investigatorship of FWO-Vlaanderen (G077121N) and is supported by the E. von Behring Chair for Neuromuscular and Neurodegenerative Disorders, the ALS Liga België and the KU Leuven funds 'Een Hart voor ALS', 'Laeversfonds voor ALS Onderzoek' and the 'Valéry Perrier Race against ALS Fund'. Several authors of this publication are members of the European Reference Network for Rare Neuromuscular Diseases (ERN-NMD). This work was supported in part by the European Union's ERA-Net for Research Programmes on Rare Diseases (INTEGRALS).

Supplementary material

Supplementary material is available at *Brain* online.

Competing interests

P.V.D. reports to have served on advisory boards for Biogen, CSL Behring, Alexion Pharmaceuticals, Ferrer, QurAlis, Cytokinetics,

Argenx, UCB, Muna Therapeutics, Alector, Augustine Therapeutics and VectorY (paid to institution). All the other authors declare no competing interests.

References

- Ryan M, Heverin M, McLaughlin RL, Hardiman O. Lifetime risk and heritability of amyotrophic lateral sclerosis. *JAMA Neurol.* 2019;76:1367-1374.
- Zhang S, Cooper-Knock J, Weimer AK, et al. Genome-wide identification of the genetic basis of amyotrophic lateral sclerosis. *Neuron.* 2022;110:992-1008.
- Renton AE, Chió A, Traynor BJ. State of play in amyotrophic lateral sclerosis genetics. *Nat Neurosci.* 2014;17:17-23.
- Catanese A, Rajkumar S, Sommer D, et al. Synaptic disruption and CREB-regulated transcription are restored by K⁺ channel blockers in ALS. *EMBO Mol Med.* 2021;13:e13131.
- Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.* 2013;3:246-259.
- Nik-Zainal S, Alexandrov LB, Wedge DC, et al. Mutational processes molding the genomes of 21 breast cancers. *Cell.* 2012;149:979-993.
- Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. *Nature.* 2013;500:415-421.
- Bustin SA, Benes V, Garson JA, et al. The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clin Chem.* 2009;55:611-622.
- Kim D, Langmead B, Salzberg SL. HISAT: A fast spliced aligner with low memory requirements. *Nat Methods.* 2015;12:357-360.
- Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* 2019;37:907-915.
- Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc.* 2016;11:1650-1667.
- Liao Y, Smyth GK, Shi W. Featurecounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 2013;30:923-930.
- Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43:e47.
- Mootha VK, Lindgren CM, Eriksson KF, et al. PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet.* 2003;34:267-273.
- Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102:15545-15550.
- Löffler-Wirth H, Kalcher M, Binder H. oposSOM: R-package for high-dimensional portraying of genome-wide expression landscapes on Bioconductor. *Bioinformatics.* 2015;31:3225-3227.
- Merkel A, Fernández-Callejo M, Casals E, et al. gemBS: High throughput processing for DNA methylation data from bisulfite sequencing. *Bioinformatics.* 2018;35:737-742.
- Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell.* 2010;38:576-589.
- Krzywinski M, Schein J, Birol I, et al. Circos: An information aesthetic for comparative genomics. *Genome Res.* 2009;19:1639-1645.
- Baxi EG, Thompson T, Li J, 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.* 2022;25:226–237.
21. Gehring JS, Fischer B, Lawrence M, Huber W. Somatic signatures: Inferring mutational signatures from single-nucleotide variants. *Bioinformatics.* 2015;31:3673–3675.
 22. Alexandrov LB, Stratton MR. Mutational signatures: The patterns of somatic mutations hidden in cancer genomes. *Curr Opin Genet Dev.* 2014;24:52–60.
 23. Abadi M, Agarwal A, Barham P, et al. TensorFlow: Large-scale machine learning on heterogeneous systems. Preprint at arXiv. 2015;1603.04467.
 24. Japtok J, Lojewski X, Naumann M, et al. Stepwise acquirement of hallmark neuropathology in FUS-ALS iPSC models depends on mutation type and neuronal aging. *Neurobiol Dis.* 2015;82:420–429.
 25. Higelin J, Demestre M, Putz S, et al. FUS mislocalization and vulnerability to DNA damage in ALS patients derived hiPSCs and aging motoneurons. *Front Cell Neurosci.* 2016;10:290.
 26. Catanese A, Olde Heuvel F, Mulaw M, et al. Retinoic acid worsens ATG10-dependent autophagy impairment in TBK1-mutant hiPSC-derived motoneurons through SQSTM1/p62 accumulation. *Autophagy.* 2019;15:1719–1737.
 27. Fujimori K, Ishikawa M, Otomo A, et al. Modeling sporadic ALS in iPSC-derived motor neurons identifies a potential therapeutic agent. *Nat Med.* 2018;24:1579–1589.
 28. Hall CE, Yao Z, Choi M, et al. Progressive motor neuron pathology and the role of astrocytes in a human stem cell model of VCP-related ALS. *Cell Rep.* 2017;19:1739–1749.
 29. Shi Y, Lin S, Staats KA, et al. Haploinsufficiency leads to neurodegeneration in C9ORF72 ALS/FTD human induced motor neurons. *Nat Med.* 2018;4:313–325.
 30. Sareen D, O'Rourke JG, Meera P, et al. Targeting RNA foci in iPSC-derived motor neurons from ALS patients with a C9ORF72 repeat expansion. *Sci Transl Med.* 2013;5:208ra149.
 31. Henstridge CM, Sideris DI, Carroll E, et al. Synapse loss in the prefrontal cortex is associated with cognitive decline in amyotrophic lateral sclerosis. *Acta Neuropathol.* 2018;135:213–226.
 32. Bauer CS, Cohen RN, Sironi F, et al. An interaction between synapsin and C9orf72 regulates excitatory synapses and is impaired in ALS/FTD. *Acta Neuropathol.* 2022;144:437–464.
 33. Upadhyay G. Emerging role of lymphocyte antigen-6 family of genes in cancer and immune cells. *Front Immunol.* 2019;10:819.
 34. Miwa JM, Anderson KR, Hoffman KM. Lynx prototoxins: Roles of endogenous mammalian neurotoxin-like proteins in modulating nicotinic acetylcholine receptor function to influence complex biological processes. *Front Pharmacol.* 2019;10:343.
 35. van Rheenen W, Diekstra FP, Harschnitz O, et al. Whole blood transcriptome analysis in amyotrophic lateral sclerosis: A biomarker study. *PLoS ONE.* 2018;3:e0198874.
 36. Humphrey J, Venkatesh S, Hasan R, et al. Integrative transcriptomic analysis of the amyotrophic lateral sclerosis spinal cord implicates glial activation and suggests new risk genes. *Nat Neurosci.* 2023;26:150–162.
 37. Koopmans F, van Nierop P, Andres-Alonso M, et al. SynGO: An evidence-based, expert-curated knowledge base for the synapse. *Neuron.* 2019;103:217–234.
 38. Staats KA, Borchelt DR, Tansey MG, Wymer J. Blood-based biomarkers of inflammation in amyotrophic lateral sclerosis. *Mol Neurodegener.* 2022;17:11.
 39. Miller T, Cudkowicz M, Shaw PJ, et al. Phase 1-2 trial of antisense oligonucleotide Tofersen for SOD1 ALS. *N Engl J Med.* 2020;383:109–119.
 40. Tran H, Moazami MP, Yang H, et al. Suppression of mutant C9orf72 expression by a potent mixed backbone antisense oligonucleotide. *Nat Med.* 2022;28:117–124.
 41. Sardar R, Sharma A, Gupta D. Machine learning assisted prediction of prognostic biomarkers associated with COVID-19, using clinical and proteomics data. *Front Genet.* 2021;12:636441.
 42. Huseby CJ, Delvaux E, Brokaw DL, et al. Blood RNA transcripts reveal similar and differential alterations in fundamental cellular processes in Alzheimer's disease and other neurodegenerative diseases. *Alzheimers Dement.* Published online 21 December 2022. doi:10.1002/alz.12880
 43. Eckardt JN, Middeke JM, Riechert S, et al. Deep learning detects acute myeloid leukemia and predicts NPM1 mutation status from bone marrow smears. *Leukemia.* 2022;36:111–118.
 44. Namboori SC, Thomas P, Ames R, et al. Single-cell transcriptomics identifies master regulators of neurodegeneration in SOD1 ALS iPSC-derived motor neurons. *Stem Cell Rep.* 2021;16:3020–3035.
 45. Dash BP, Freischmidt A, Weishaupt JH, et al. Downstream effects of mutations in SOD1 and TARDBP converge on gene expression impairment in patient-derived motor neurons. *Int J Mol Sci.* 2022;23:9652.
 46. Smith AST, Chun C, Hesson J, et al. Human induced pluripotent stem cell-derived TDP-43 mutant neurons exhibit consistent functional phenotypes across multiple gene edited lines despite transcriptomic and splicing discrepancies. *Front Cell Dev Biol.* 2021;9:728707.
 47. Workman MJ, Lim RG, Wu J, et al. Large-scale differentiation of iPSC-derived motor neurons from ALS and control subjects. *Neuron.* Published online 2 February 2023. doi:10.1016/j.neuron.2023.01.010
 48. Lee H, Lee JJ, Park NY, et al. Multi-omic analysis of selectively vulnerable motor neuron subtypes implicates altered lipid metabolism in ALS. *Nat Neurosci.* 2021;24:1673–1685.
 49. Ho R, Workman MJ, Mathkar P, et al. Cross-comparison of human iPSC motor neuron models of familial and sporadic ALS reveals early and convergent transcriptomic disease signatures. *Cell Syst.* 2021;12:159–175.e9.
 50. Devlin AC, Burr K, Borooah S, et al. Human iPSC-derived motoneurons harbouring TARDBP or C9ORF72 ALS mutations are dysfunctional despite maintaining viability. *Nat Commun.* 2015;6:5999.
 51. Saxena S, Roselli F, Singh K, et al. Neuroprotection through excitability and mTOR required in ALS motoneurons to delay disease and extend survival. *Neuron.* 2013;80:80–96.
 52. Lee JY, Lee JD, Phipps S, et al. Absence of toll-like receptor 4 (TLR4) extends survival in the hSOD1 G93A mouse model of amyotrophic lateral sclerosis. *J Neuroinflammation.* 2015;12:90.
 53. van der Spek RAA, van Rheenen W, Pulit SL, et al. The project MinE databrowser: Bringing large-scale whole-genome sequencing in ALS to researchers and the public. *Amyotroph Lateral Scler Frontotemporal Degener.* 2019;20:432–440.
 54. Hou Y, Dan X, Babbar M, et al. Ageing as a risk factor for neurodegenerative disease. *Nat Rev Neurol.* 2019;15:565–581.
 55. Valdez G, Tapia JC, Lichtman JW, et al. Shared resistance to ageing and ALS in neuromuscular junctions of specific muscles. *PLoS ONE.* 2012;7:e34640.