bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Single cell transcriptomics identifies master regulators of dysfunctional pathways in SOD1 ALS motor neurons

Seema C. Namboori[1#], Patricia Thomas[2#], Ryan Ames[1,3], Lawrence O. Garrett[1], Craig R. G. Willis[4], , Lawrence W. Stanton[5,6], Akshay Bhinge[1,2*]

1. Living Systems Institute, University of Exeter, EX4 4QD, UK.

2. College of Medicine and Health, University of Exeter, EX1 2LU, UK.

3. Centre for Biomathematical Modelling and Analysis, University of Exeter, EX4 4QD, UK.

4. Department of Sport and Health Sciences, College of Life and Environmental Sciences, University of Exeter, EX1 2LU, UK.

5. Genome Institute of Singapore, 138672, Singapore.

6. Current address: Humacyte, NC 27713, USA.

# Contributed equally

* To whom correspondence should be addressed

Abstract

Background

Bulk RNA-Seq has been extensively utilized to investigate the molecular changes accompanying motor neuron degeneration in Amyotrophic Lateral Sclerosis (ALS). However, due to the heterogeneity and degenerating phenotype of the neurons, it has proved difficult to assign specific changes to neuronal subtypes and identify which factors drive these changes. Consequently, we have utilized single cell transcriptomics of degenerating motor neurons derived from ALS patients to uncover key transcriptional drivers of dysfunctional pathways.

Results

Single cell analysis of spinal neuronal cultures derived from ALS and isogenic iPSC allowed us to classify cells into neural subtypes including motor neurons and

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

interneurons. Differential expression analysis between disease and control motor neurons revealed downregulation of genes involved in synaptic structure, neuromuscular junction, neuronal cytoskeleton and mitochondrial function. Interestingly, interneurons did not show similar suppression of these homeostatic functions. Single cell expression data enabled us to derive a context-specific transcriptional network relevant to ALS neurons. Master regulator analysis on this network identified core transcriptional factors driving the ALS disease signature. Specifically, we were able to correlate suppression of HOXA1 and HOXA5 to synaptic dysfunction in ALS motor neurons. Our results suggest that suppression of HOX genes may be a general phenomenon in SOD1 ALS.

Conclusions

Our results demonstrate the utility of single cell transcriptomics in mapping disease-relevant gene regulatory networks driving neurodegeneration in ALS motor neurons. We propose that ALS-associated mutant SOD1 leads to inhibition of transcriptional networks driving homeostatic programs specific to motor neurons, thereby providing a possible explanation for the relative resistance of spinal interneurons to degeneration in ALS.

Keywords

ALS, spinal MN, single cell RNA-seq, HOXA1, HOXA5, synapse

Introduction

Amyotrophic Lateral Sclerosis (ALS), also known as Lou Gehrig's disease, is an age-onset fatal neurodegenerative disorder that affects motor neurons in the brain and spinal cord(1). Patients display progressive paralysis and eventually die due to failure of the respiratory muscles, commonly within 3-5 years of diagnosis(2). Despite extensive research, the causes underlying the observed degeneration are

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

incompletely understood. Hence, so far, there is no cure for ALS. Understanding the molecular drivers of neurodegeneration in ALS can potentially lead to the development of life-saving therapies. Approximately 20% of ALS cases are familial with mutations identified in genes spanning diverse cellular functions including SOD1(3). Animal models have been extensively used to study ALS and have revealed important insights into disease mechanisms(4). However, species-specific differences and the need to overexpress the mutant proteins to generate phenotypes have necessitated developing human models of the disease that can complement existing animal models(5).

Induced pluripotent stem cells (iPSC) derived from patients suffering with ALS provide a powerful model to study ALS. ALS patient-derived iPSC bear the disease-causing mutations in a physiologically relevant background and can be readily differentiated into clinically relevant neurons(6). Such diseased neurons can now be compared to healthy neurons to model key aspects of the disease such as neuron survival, morphometric defects, electrophysiological dysfunctions and protein/RNA aggregation foci *in vitro*(7-12).

Molecular characterisation of these neurons using "omics" tools has uncovered important insights into disease pathophysiology(8,9,13,14). However, application of genomic tools such as RNA-sequencing to ALS neurons in bulk has serious drawbacks. Current differentiation protocols generate motor neurons at efficiencies ranging from 50-80% with the efficiency varying depending upon the iPSC line used. The differentiated neurons are usually a mix of motor neurons (MN) and "non-motor neurons", typically spinal interneurons (IN). Additionally, long term cultures of these neurons that are required for phenotypic characterization commonly display some degree of glial cell proliferation. Importantly, ALS motor neurons display progressive degeneration. This suggests that at any given time point, neurons can be expected to

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

be in different stages of degeneration and hence may display a differential expression of key drivers of disease pathology. Bulk RNA analysis of these cultures would not only average cell type expression but also the motor neuron specific degeneration expression signatures. To address these issues, we performed single cell transcriptomic analysis of degenerating ALS SOD1 mutant neurons. Our analysis reveals MN specific transcriptional networks regulating synaptic function and neuronal cytoskeleton to be downregulated in ALS MN. Importantly, single cell data enabled us to build a disease relevant transcriptional network that was used to identify key transcription factors driving the disease signature.

Materials and Methods

Human iPSC culture of iPSC

ALS patient-derived iPSC bearing SOD1 E100G/+ (ND35662) mutation were obtained from the Coriell Institute for Medical Research. ALS iPSC and the genome edited isogenic control iPSC were maintained as colonies on human ES qualified matrigel (Corning) in mTeSR (StemCell Technologies). Colonies were routinely passaged in a 1:6 split using Dispase. Mycoplasma testing was conducted regularly to rule out mycoplasma contamination of cultures.

Differentiation of iPSC into spinal motor neurons

iPSC were plated as colonies onto matrigel and differentiated by treatment with neuronal differentiation media (N2B27: DMEM/F12, Neurobasal, N2 supplement 1%, B27 supplement 1%, L-glutamine 1%, ascorbic acid 5uM, insulin 20ug/ml) supplemented with SB431542 (80uM), CHIR9921 (3uM) and LDN8312 (0.2uM) from day 0 till day 4. Cells were caudalized by treatment with 0.1uM retinoic acid starting at day 2 and ventralized with 1uM purmorphamine starting at day 4 and continued till day 10. At day 10, OLIG2 positive motor neuron progenitors were re-

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

plated onto poly-D-lysine/laminin coated wells and differentiated by treating the cells with N2B27 media supplemented with BDNF 20ug/ml, GDNF 10ug/ml and DAPT 10uM. DAPT treatment was stopped at day 14 and neuronal cultures were pulsed with mitomycin at a dose of 10ug/ml for 1 hour to prevent further proliferation of any undifferentiated progenitors. Neuronal cultures were maintained till day 44 by changing media every other day. Survival of MN was assessed using 3 independent differentiations.

Immunofluorescence

Cells were fixed with 4% paraformaldehyde, permeabilized with ice-cold methanol for 5 minutes and washed with PBS containing 10% serum for 1 hour at room temperature. Cells were incubated with primary antibodies (Table S1) diluted into PBS containing 10% serum and incubated overnight at 4°C. Next day, cells were washed and incubated with Alexa-fluor conjugated secondary antibodies (Molecular probes) for 45 minutes at room temperature and nuclei were stained with Hoechst 33542 (Molecular probes).

Quantitative RT-PCR

Total RNA was extracted with the miRNeasy kit (Qiagen) and reverse transcribed using random hexamers and the High Capacity reverse transcription system from Applied Biosystems. Quantitative PCR was performed using the SYBR GREEN PCR Master Mix from Applied Biosystems. The target gene mRNA expression was normalized to the expression of RPL13 and relative mRNA fold changes were calculated by the ΔΔCt method. Primer sequenced are included in Table S2.

Single-cell capture and library preparation

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Single cells were captured using standard protocol of C1 single-cell auto prep system (Fluidigm). Briefly, differentiated neuronal cultures at day 44 were dissociated into single cells by Accutase and loaded onto the C1 chip. Post cell capture, each well of the chip was manually inspected to identify wells bearing single cells. Next, lysis, reverse transcription and PCR amplification of the cDNA was perform in an automated fashion within the C1 instrument. To prepare single-cell libraries, cDNA products from each single cell were harvested from C1 chip followed by concentration and quantification using PicoGreen dsDNA Assay kit. Sequencing libraries were generated using Illumina Nextera XT library preparation kit.

Read processing, mapping and quality control

Sequence data were processed and mapped to the human reference genome (hg19) using TopHat (v2.0.11)(15) with Bowtie2 (v2.2.1)(16). Gene expression levels were quantified with HTSeq-count (v0.6.1p1)(17) and converted to fragments per kilobase million (FPKM) for human gene annotation (GENCODE release 19)(18). This yielded 57,241 transcripts from 365 libraries. Only libraries deemed to be single cells were retained for further analysis (332 cells). Additionally, libraries were qualified as low quality if: 1) the mapping ratio (mapped reads/total sequenced reads) was < 0.5; 2) total number of genes expressed were < 4000; 3) total number of mapped reads < 0.5 million. A gene was deemed to be poorly expressed if it was found to be present in less than 20 cells at an FPKM expression threshold set at 0.1. The filtering process yielded 303 single cells and 14774 genes for further analysis.

Weighted gene coexpression network analysis

To identify functional modules of genes associated with ALS we used weighted gene coexpression network analysis (WGCNA) implemented in the R statistical language. FPKM values for filtered libraries and features (as described above) were log transformed and used to build a coexpression network. The coexpression network

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

was constructed with WGCNA using a soft thresholding power of 6 using the signed-hybrid approach. Modules in the network were identified using the *cutTreeDynamic* function with a minimum module size of 20. Modules with similar expression profiles were merged if their eigengene correlation coefficients were >=0.9. To define modules related with ALS we used Pearson's correlation coefficient to assess associations between module eigengenes and disease state. *P-values* were corrected using the method of Benjamini and Hochberg and correlations with an adjusted p-value < 0.01 were deemed significant. Gene ontology enrichment analysis of disease associated modules was carried out using the anRichment R package.

Master regulator analysis

The master regulator analysis was performed using the RTN R package. Log transformed FPKM values for 211 neurons were used as input to build a transcriptional network for 845 TFs present in the dataset. TF annotations were obtained from AnimalTFDB. P-values for network edges were estimated from a pooled null distribution using 1000 permutations. Edges with an adjusted p-value < 0.05 were retained for DPI filtering. Post-DPI filtering, the MRA generates a list of TF with predicted targets that is termed as the regulon. The regulon for each TF can be classed as positive or negative based on the Pearson correlations. To identify master regulators, the differential gene expression between ALS and control MN was used as a phenotype and sorted from most upregulated to most downregulated. The RTN package was used to conduct a GSEA like analysis to identify whether a TF regulon (positive or negative) was enriched towards one end of the sorted list of differentially expressed genes. P-values were estimated based on 1000 permutations of the dataset and adjusted using the Benjamini Hochberg method. At a p-value cutoff of 0.05, we identified 52 TFs as significant. Next, we assessed whether the identified TFs were differentially expressed and whether the direction of fold change

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

was concordant with the regulon of that TF. Applying this criterion generated a list of 13 TF that were deemed as master regulators of the ALS signature.

Overexpression of HOXA1 and HOXA5 in SH-S5Y5 cells

HOXA1 and HOXA5 cDNA expressing plasmids driven by a CMV promoter were purchased from origene. A CMV driven GFP plasmid was used as control. SH-S5Y5 cells were plated on poly-D-lysine coated 24-well plates at 1 million cells per well. Plasmid transfections were performed using lipofectamine 3000 according to manufacturer's instruictions. Cells were induced to differentiate the next day by changing media to N2B27 + 10uM retinoic acid. This resulted in the cells exiting the cell cycle preventing plasmid loss. Cells were harvested at day 6 of differentiation.

Analysis of HOXA1 and HOXA5 binding sites

ChIP-Seq derived peaks for HOXA1, HOXA5, H3K4Me1 and H3K4Me3 were downloaded from the gene expression omnibus for datasets GSM1208634, GSM1239461, GSM1208810, GSM1208811 repectively. Fold enrichments were used as is from the peak files. Peaks were classified into 3 classes: strongly enriched (fold enrichment > 2.0), weakly enriched (fold enrichment >= 1.5 and <= 2.0) and not enriched (fold enrichment < 1.5). A gene was deemed as expressed if a H3K4Me3 or H3K4Me1 peak was found within 5kb of the transcriptional start site or if a H3K4Me1 peak was found within the gene body. A gene was deemed a HOXA1 or HOXA5 target if a HOXA1 or HOXA5 binding site was found within 10kb of the transcriptional start site, within the gene body or 2 kb from the transcriptional termination site.

Analysis of mouse SOD1 G93A ALS MN

Microarray expression values for mouse SOD1 G93A MN were downloaded from the gene expression omnibus for dataset GSE46298. Expression values were background subtracted, normlaized and log transformed using the RMA R package. Threshold for

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

low expression was set to 30 percentile of the entire dataset. P-values for individual genes were estimated based on a 2-tailed Student's t-test and corrected for multiple hypothesis using the Benjamini Hochberg procedure. Only genes with median expression values above background were included in the analysis.

Results

Single cell RNA-Seq analysis of ALS and control neurons

We previously developed an *in vitro* model of ALS MN degeneration where MN differentiated from mutant SOD1 iPSC display disease-specific phenotypes such as reduced cell survival and morphometric defects compared to their isogenic control counterparts(19). We differentiated iPSC derived from patients bearing the SOD1 E100G mutation, as well as the corresponding CRISPR edited isogenic control into MN (Fig.1a) (19). Mature MN appeared by day 30 and expressed MN markers such as ISL1, CHAT and NF-H in addition to the pan-neuronal markers TUJ1 and MAP2 (Fig. 1b). ALS and control iPSC could be differentiated into MN at similar efficiencies (~72% ISL1+ TUJ1+ neurons). When cultured for a further 2 weeks (day 44), ALS iPSC derived neurons displayed a 40% loss in survival compared to the isogenic control MN (Fig. 1c). To gain deeper insights into the mechanisms driving neurodegeneration, we performed single cell RNA-sequencing on the ALS and isogenic control MN at day 44 of our differentiation protocol (Fig. 1a). Neuronal cultures were dissociated into single cells, captured and lysed in a fully automated fashion using a microfluidic platform followed by library preparation and deep sequencing of transcripts in the individual cells (Methods). We captured a total of 332 individual cells that included 165 cells from the ALS culture and 167 cells from the isogenic control.

Single cell transcriptomes were sequenced to an average depth of 1.5e6 reads per cell and the overall sequencing depth was similar across the ALS and control datasets

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(Figure S1a). The proportion of total reads mapped to the genome is an indicator of library quality(20). Both ALS and control cells displayed similar mapping statistics indicating there was no bias towards poor quality cells in any one dataset (Figure S1b). At a FPKM threshold of 0.1, both datasets expressed 8000 to 9000 genes (Figure S1c). To remove poorly amplified RNA libraries (indicated by low mapping ratios and low numbers of expressed genes) and lowly expressed genes, single cell libraries were subjected to a set of quality control criteria (Methods). After quality filtering, we retained a total of 303 high quality cells (153 cells for ALS and 150 cells for the control). These 303 cells expressed 14774 genes in total with ALS cells expressing on average 7170 genes while the control dataset expressed 7745 genes (Figure S1d) with the overall distribution of the number of genes being similar between the two datasets. Each gene was classified based on whether it was protein coding, long non-coding, pseudogene or small nuclear/nucleolar RNA. We did not find any systematic difference in distribution of the gene classes expressed between the ALS and controls datasets (Figure S1e). In summary, our single cell transcriptomes for the ALS and controls sets were similar in quality and character on a genome-wide level.

Our differentiation protocol was designed to generate spinal MN(21). *In vivo*, spinal MN at different rostro-caudal levels of the spinal cord are demarcated by specific combinations of HOX gene expression (known as the HOX code)(22). To ascertain the rostrocaudal address of our *in vitro* differentiated neurons, we estimated the percentage of cells expressing each of the 39 HOX genes and plotted the data as a heatmap (Fig 1d). The heatmap showed most of the cells expressed HOXA5 and HOXB5 with very few cells expressing HOXB8 and HOXD8 and none expressing HOX genes from paralog groups 9 and higher (Fig. 1d). This clearly indicated that all of our cells were largely restricted to the hindbrain or brachial spinal cord identity as

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

would be expected from our differentiation protocol that employed 0.1uM retinoic acid without any GDF11(23). Next, we assessed expression of the classical markers of neuronal subtypes for motor neurons (ISL1, CHAT, VAChT or SLC18A3), interneurons (GAD1, GAD2, ARX) and non-neuronal cells (S100B, SOX9, MKI67) across all cells. As we have shown previously, our data indicates that iPSC-derived neuronal cultures display wide variation in expression across individual single cells that typically gets averaged in bulk analysis(24) (Fig.1e). To resolve this heterogeneity and enable differential expression between relevant classes of neurons, we sought to classify cells into specific neural lineages.

Classification of single cells into neural subtypes

Gene expression of selective markers in single cells is routinely used to classify neurons into distinct lineages(25,26). However, using one or two markers to classify single cells can lead to erroneous classification as single cell data typically has a high rate of dropouts, especially for lowly expressed transcripts. On the other hand, clustering of single cell data based on the expression of all detected transcripts might lead to sub-optimal clustering due to the inclusion of genes that are irrelevant to the classification. We sought to circumvent these problems by first identifying genes that can be used to classify cells into relevant cell types (neurons vs glia and MN vs IN). These lists of genes were termed as classifier gene sets. Cells were classified into distinct neural subtypes by sequentially clustering single cells based on the relevant classifier gene set as described below (Fig. 2a).

Classification of single cells into neurons and glia

We first identified genes differentially present between neurons and glia using a recently published gene expression dataset on purified human neurons, astrocytes and oligodendrocytes from frozen brain tissue(27). Genes that displayed a fold change of at least 20 between neurons and astrocytes or oligodendrocytes were

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

included for future analysis. Differential gene expression analysis identified 522 genes as upregulated and 433 genes as downregulated in neurons versus glial cells. Gene ontology using DAVID(28) on the differentially expressed gene set showed enrichment of specific functional categories related to neuronal physiology. Categories related to neuron development such as GABAergic synapse and postsynaptic cell membrane were found to be enriched in the neuron-activated genes while cell cycle and glial differentiation terms were deemed to be enriched in the downregulated genes confirming that our identified gene set was able to distinguish neurons and glia (Figure S2). Out of the 955 genes, 755 genes were expressed in our filtered single cell dataset and were further used to cluster cells into neurons and glia. We identified 63 cells out of the starting pool of 332 cells as mitotic as these expressed high levels of the proliferating cell marker MKI67 and removed them from further analysis. The remaining 269 single cells were clustered based on the neurons vs glia classifier gene set. Consensus k-means clustering using the R package SC3(29) identified three distinct clusters (Fig.2b). We calculated the mean expression of classical neuronal markers (SYP, RBFOX1, SYN1, CAMKV), neuronal lineage specific markers (CHAT, GAD1 and GAD2) and glial markers (SOX9, S100B, REST, PAX6, MKI67) in each of these 3 clusters. Expression profiles of classical markers indicated that clusters 2 and 3 were neuronal i.e. these cells expressed pan-neuronal and neuronal subtype specific markers but did not express glial markers (total 211 cells) while cluster 1 included neuronal progenitors and glial cells i.e. marker expression profiles were opposite to that of cluster 1 (58 cells) (Fig. 2c).

Classification of neurons into MN and IN

We sought to further classify the neuronal cells in clusters 2 and 3 into MN and IN by clustering the cells based on established MN and IN markers that were expressed in our dataset (MN markers: CHAT, SLC18A3 or VACHT, ISL1, ONECUT1, PRPH, IN

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

markers: GAD1, GAD2, ARX, DLX1, DLX6). To avoid confounding the classification due to the SOD1 mutation, the ALS and isogenic control datasets were clustered separately. Consensus clustering of the control neuronal cells using the MN/IN marker gene set identified three clusters (Figure S3a). Analysis of mean expression of hallmark MN and IN markers indicated that cluster c1 comprised of IN i.e. these neurons expressed high levels of IN markers but not MN markers (49 cells) while clusters c2 and c3 were MN i.e. these cells expressed high levels of MN markers but not IN markers (total 74 cells) (Figure S3b). However, we observed a subset of IN that expressed ISL1 (Figure S3b). ISL1 expression in interneurons has been observed in the mouse spinal cord within lamine V, VI and VII(30). However, to avoid the possibility that these were cell doublets that had escaped detection, this subset (28 cells) was removed. Re-clustering the remaining neurons identified three clusters with expected marker expression: cluster c1 (21 cells) displayed IN markers and clusters c2 and c3 (total 74 cells) displayed MN markers (Fig. 2d and 2e). We used a similar clustering workflow on the SOD1 neuronal cells to identify MN and IN clusters (Fig. 2a, 2f, 2g). For the SOD1 ALS dataset, marker expression levels clearly indicated that cluster a1 comprised of IN while cluster a2 comprised of MN. Cluster a3 displayed expression of both MN and IN markers (possible doublets) and was removed from further analysis. In summary, our clustering approach identified 74 MN (c2 plus c3 clusters in Fig. 2d,e) and 21 IN (c1 cluster in Fig. 2d,e) in the control dataset, while 24 MN (a2 cluster in Fig. 2f,g) and 33 IN (a1 cluster in Fig. 2f,g) were identified in the SOD1 ALS dataset.

Differential expression analysis of ALS and control neurons

To gain deeper insights into the transcriptional changes within ALS neurons, we compared gene expression in SOD1 ALS MN with the isogenic control MN using Monocle(31). Differential expression analysis identified 84 upregulated genes and

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

332 downregulated genes in ALS MN at a FDR < 0.05. Gene ontology (GO) enrichment analysis of the differentially expressed genes identified several pathways dysregulated in ALS MN. The downregulated gene ontology terms were distributed across 2 main categories, 1) synaptic function and 2) mitochondrial structure and function. Terms enriched related to synaptic function included "synapse organization", "synapse assembly", "axon", "trans-synaptic signalling", and "neuromuscular junction" while mitochondria related terms included "mitochondrial outer membrane ", "NADH:ubiquinone oxidoreductase, mitochondria" and "oxidative phosphorylation" (Fig. 2h). On the other hand, GO terms enriched in the upregulated genes included terms related to the cell cycle and protein targeting to the endoplasmic reticulum (ER) (Fig. 2i). Deficiency in the mitochondrial oxidative phosphorylation pathway as well as suppression of translation secondary to ER stress with concomitant increase in cell cycle related genes has been identified previously in bulk analysis of ALS SOD1 MN(19), which was recapitulated in our single cell expression analysis. Importantly, dysregulation of synaptic signalling and structure in our ALS model supports the dying back hypothesis of ALS that posits neuronal degeneration is secondary to pathology initiated at the distal end of the axon and neuromuscular junction(32). Next, we asked whether mutant SOD1 induces similar changes in the IN gene expression program. Strikingly, comparative analysis between ALS and control IN revealed far fewer genes dysregulated in ALS IN compared to ALS MN (24 genes upregulated while 48 genes downregulated). Additionally, there was minimal, though significant, overlap in the gene expression programs perturbed between the two classes of neurons with 4 genes shared in the upregulated set and 9 genes shared in the downregulated set (Figure S4). Given the small number of genes identified as dysregulated in ALS IN, GO enrichment analysis was unable to find significant enrichment of any functional categories. Overall, our

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

data indicates that mutant SOD1 causes significant changes in the transcriptional program of MN compared to IN.

Network analysis of single neurons using WGCNA

To gain a systems level understanding of the observed transcriptional changes, we performed weighted gene co-expression network analysis (WGCNA)(33,34) to identify modules of co-regulated genes dysregulated in ALS MN and IN. Since co-expression analysis works on identifying co-regulated sets of genes, we postulated that a large sample set with high variability per gene across the samples would improve performance of the network analysis. Hence, we included the entire set of neurons for both SOD1 and isogenic control datasets (a total of 211 cells). WGCNA identified a total of thirteen co-regulated modules of which three (modules blue, purple and black) were significantly downregulated in ALS MN compared to isogenic control MN (adjusted p-value < 0.01) (Fig. 3a). GO enrichment analysis of these modules revealed association of each module with specific functional categories (Fig. 3b). Synaptic function and structure terms were enriched in the blue module, microtubule and cytoskeleton terms were enriched in the purple module, while oxidative phosphorylation and mitochondrial membrane terms were enriched in the black module. These observations were in accordance with the downregulation of these functional categories observed in the differential gene expression analysis of ALS MN. Interestingly, the blue module (enriched for synaptic terms) associated significantly

with MN compared to IN in both the ALS and control samples but was observed to be strongly downregulated in ALS MN (Fig. 3a). This indicates that genes mediating synaptic signalling are highly expressed in MN compared to IN. Downregulation of these genes specifically in ALS MN suggests a possible mechanism that can explain the relative susceptibility of ALS MN to degeneration.

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We hypothesized that downregulation of the observed gene modules may be driven by concomitant dysregulation of specific transcription factors (TFs). To test this hypothesis, the expression of ~1000 TFs in our dataset was correlated to the module eigengenes estimated by WGCNA. Correlation analysis identified sequence-specific TFs that correlated both positively and negatively with the modules. Given a module that is downregulated in ALS MN, a TF that correlates positively with this module is expected to be inhibited in ALS MN while a TF that correlates negatively is expected to be activated in ALS MN. TFs that matched these expectations were deemed as concordant. Application of the criterion of concordance to our correlation analysis identified TFs associated with specific modules (Fig. 3c). For example, HOXA5 displayed greater than 4-fold suppression in ALS MN compared to isogenic controls while its expression correlated positively with the blue module that was identified as downregulated in ALS MN (Fig. 3c). This strongly indicates that HOXA5 is involved in regulating synaptic functions in MN. In addition, we observed SOX4 and HMGB2 activated in ALS MN and negatively correlated with the blue module (Fig. 3c), indicating that these TFs act in an opposite fashion to HOXA5. Additionally, we observed positive correlations between HOXA5 and ATF2 with the purple module (cytoskeleton organization) while ATF2 also correlated positively with the black module (oxidative phosphorylation) (Fig. 3c). The WGCNA analysis strongly suggested that downregulation of specific pathways by mutant SOD1 is accompanied by corresponding de-regulation of TFs and such factors are readily identified from our single cell dataset.

Master regulator analysis to identify TFs driving ALS disease signature

Genes differentially expressed in ALS MN can be considered as a molecular phenotype that drives the cellular phenotype of neuronal degeneration. We wanted to identify TFs that regulated this disease signature. Identification of such master

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

regulators requires a context-specific transcriptional network(35). Our single cell RNA-Seq data offered a unique opportunity to build such a network relevant to ALS and MN.

The gene interactions detected by WGCNA include both direct and indirect interactions. We postulated that removal of indirect interactions would significantly enhance our power to identify dysregulated TFs driving ALS MN degeneration. To this end, we deployed the ARACNE algorithm that uses an information theoretic approach to prune indirect regulations in transcriptional networks(36). Though the pruning procedure does not necessarily remove all indirect interactions, it significantly enriches the network for direct interactions.

ARACNE was implemented on the gene expression matrix derived from all 211 neurons using the RTN package(37), and downstream targets were identified for 845 TFs expressed in the neuronal cells. Key TF drivers were identified using master regulator analysis (Figure S5). The predicted targets of a TF (termed as the regulon of that TF) were classified as activated or repressed based on the direction of the correlation between the TF and the target gene. Next, an ALS signature was defined as genes that were differentially expressed between ALS MN and isogenic control MN. Master regulators were identified based on whether there was a statistically significant overlap between the activated or repressed regulon of a TF and the ALS disease signature (Figure S5, Methods). The master regulator analysis (MRA) identified 51 TFs at a FDR < 0.05. We filtered out TFs that were not found to be differentially expressed in the ALS MN compared to control. Next, we checked for concordance between the expression change of a TF and its regulon and filtered out non-concordant TFs (Figure S5). This identified a core set of 13 TFs that satisfied the following criteria: 1) the TFs were differentially expressed in ALS MN compared to control, 2) the regulons of these TFs showed a significant overlap with the ALS gene

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

expression signature, 3) the TF and its regulon expression was concordant (Fig. 4a). Interestingly, HOXA5 was amongst the TFs identified by this analysis, where the positive regulon of HOXA5 was downregulated in ALS MN while the negative regulon was upregulated in ALS MN (Figure S6). Additionally, HOXA5 itself was downregulated in ALS MN (Fig. 4a). To gain further insights into the functions of the identified master regulators, we performed gene ontology enrichment analysis on the regulons of these TFs (Figure. S7). Strikingly, HOXA1, HOXA5 and HOXD1 regulons were enriched in genes related to synaptic structure and signalling as well as maintaining the axon cytoskeleton organization (Figure S7). This indicated a direct link between inhibition of synaptic functions in ALS MN and HOXA1/HOXA5/HOXD1 downregulation. Additionally, the GO enrichment analysis linked downregulation of ATF2 to autophagy and PRDM2 to mitochondrial structural defects. Thus, our master regulator analysis linked functional deficits in ALS MN to downregulation of specific TFs. We asked whether the identified TFs regulated each other. Extracting a sub-network comprising only of the master regulators revealed that HOX TFs closely interacted within themselves as compared to the remaining TFs with HOXA1, HOXA5 and HOXD1 displaying the maximum number of connections to other TFs (Fig. 4b). This suggested that synaptic dysregulation driven by altered HOX gene expression is a major driver of the disease signature in ALS.

Next, we asked whether these TFs were dysregulated in ALS IN as well. Comparison of the gene expression profiles of the identified TFs across MN and IN in both the ALS and isogenic cells revealed distinct patterns of expression across the 4 subsets (Fig. 4c). Group 1 TFs were expressed at higher levels in control MN compared to IN and were significantly downregulated in ALS MN. HOXA1, HOXA5, HOXD1 and MAFG belonged to this group where the average expression levels of these TFs were

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2-10 fold higher in MN but were strongly downregulated (> 5 fold) in ALS MN. Group 2 and 3 TF expression levels were similar in control MN and IN, however Group 2 TFs were dysregulated specifically in ALS MN (ATF2, PRDM2, TSC22D2, ZMIZ1, ZNF134) while Group 3 TFs were dysregulated in both ALS MN and IN (HMGB2, HOXD8, ZBTB6) (Fig. 4c). Finally, Group 4 included just SOX4 where the expression was lower in control MN compared to IN but significantly elevated in both ALS MN and IN (Fig. 4c). The expression pattern of Group 1 TFs suggested that these genes could be involved in MN homeostatic networks and downregulation of these TFs in ALS MN could, at least in part, explain the relative susceptibility of MN compared to IN in ALS. Group 2 TFs regulated core cellular functions such as autophagy and mitochondrial respiration, and downregulation of these TFs could result in cellular stress due to dysregulation of core cellular pathways. Group 3 and 4 TFs expression dysregulation was observed in both MN and IN. This suggested that these TFs may play an ancillary role in accelerating the degeneration process but may not be major drivers of degeneration on their own.

Dysregulation of TFs in the SOD1 mouse model of ALS

Since our observations regarding dysregulated TFs in SOD1 ALS were based on analyzing iPSC derived from a single patient, we decided to confirm our findings in a mouse model of SOD1 ALS. The SOD1 G93A mouse is a well-established animal model of ALS that displays degeneration of MN over time accompanied by disruption of synaptic functions. A recent study had analyzed gene expression changes in MN isolated from the lumbar spinal cord of G93A mice via laser capture microdissection and analysed gene expression changes via microarrays (38). The study had used two different mouse strains, 129Sv and C57, each expressing the mutant G93A hSOD1 gene and analysed gene expression changes at various stages over the course of MN degeneration. The 129Sv strain displays fast disease progression while the C57 strain

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

displays slow disease progression. We first analysed gene expression of all candidate TFs in this dataset except HOXA1 and HOXA5 as these were not expressed in lumbar MN. We did not observe significant differences in TF expression at the pre-symptomatic stages in either strain (Fig. 5a,5c). However, both Sox4 and Hmgb2 displayed strong activation in ALS MN at later stages in both strains, in concordance to our data (Fig. 5a, 5c). Further, Atf2 dislayed downregulation in the 129Sv strain at the endstage of the disease (Fig. 5a). Surprisingly, Prdm2, Zbtb6 and Zmiz1 displayed activation in ALS MN in contrast to our observation in iPSC derived MN. Since HOXA1 and HOXA5 were not expressed in lumbar MN, we analysed expression of other Hox genes relevant to the mouse lumbar spinal region (Fig. 5b, 5d). We observed significant downregulation of multiple Hox genes in ALS MN derived from both the 129Sv and the C57 strains (Fig. 5b, 5d). The expression changes in the slow progressing C57 strain, though milder, were largely in concordance with changes observed in 129Sv (Fig. 5b). Overall, our analysis indicated the involvement of ATF2, SOX4, HMGB2 as well as HOX genes in SOD1 ALS.

Early drivers of MN degeneration

We next asked whether any of these TFs displayed dysregulation in the early stages of the degeneration process using our iPSC-based model. The expression levels of the 13 master regulators were assessed in MN differentiated from ALS and control iPSC at D30 of differentiation. At this stage, ALS MN did not display any significant differences in survival compared to the control MN. Strikingly, HOXA1 and HOXA5 but not HOXD1 showed significant downregulation in ALS neuronal cultures in spite of the assay being performed in bulk (Fig. 6a). HOXD8, PRDM2 and ZBTB6 were also downregulated in ALS albeit to a lesser extent (Fig. 6a). We did not observe significant upregulation of SOX4 or HMGB2 at this stage. This indicated that HOXA1 and HOXA5 could be early drivers of neurodegeneration in ALS. Since HOX genes

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

were also observed to be downregulated in the mouse model dataset and our network analysis had linked HOX genes to synaptic function, we decided to investigate the role of HOX genes in ALS. We next assessed whether the targets of HOXA1 and HOXA5, as predicted by the MRA, were downregulated in ALS. In concordance with our expectation, several HOXA1/HOXA5 synaptic targets were significantly downregulated, although modestly, in ALS neurons at D30 compared to the control (Fig. 6b). To confirm a direct role of HOXA1/HOXA5 in regulating synaptic gene expression, we exogenously expressed HOXA1 and HOXA5 in the neuronal cell line SH-SY5Y and induced these cells to differentiate. Overexpression of HOXA1 and HOXA5 significantly upregulated several synaptic genes confirming that HOXA1 and HOXA5 activate synaptic gene expression in neurons (Fig. 6c). Next, we asked whether synaptic genes are direct targets of these TFs. To answer this question, we analysed ChIP-seq data mapping HOXA1 and HOXA5 binding sites in colon carcinoma cells(39). For most synaptic genes expressed in the carcinoma cells (assessed by the proximal localization of H3K4Me3 or H3K4Me1 signal), the ChIP-Seq analysis identified either a HOXA1 or HOXA5 site proximal to the promoter or within the gene body, thereby indicating that at least a subset of synaptic genes are direct targets of these TFs (Fig. 6d). Interestingly, the ChIP-Seq data revealed HOXA1 binding sites in close proximity to the HOXA5 promoter and vice versa (Figure S8a) suggesting that HOXA1 and HOXA5 display reciprocal regulation. Further, we found HOXA1 binding sites proximal to the promoters in 9 out of 13 TFs identified by the MRA (Figure S8b). This observation indicated that HOXA1 might regulate other master regulators including HOXA5 and suggests the possibility that downregulation of HOXA1 could be an early event triggering synaptic breakdown in SOD1 ALS MN.

Discussion

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ALS patient-derived iPSC have provided unprecedented access to human diseased motor neurons enabling researchers to follow the course of degeneration in dish(6). Investigation of these models using genomics has uncovered key pathways dysregulated in ALS neurons. However, so far, iPSC derived neuronal cultures have been typically analysed in bulk. Hence, it has been difficult to assign observed pathway aberrations specifically to MN or other neuronal subsets present in the *in vitro* culture. The advent of single cell genomics has now allowed the analysis of individual neurons in mixed cultures(24,40). This game-changing technology now allows identification of genome-wide gene expression in individual neurons thereby enabling neuronal classification into subtypes. We have applied this technology to analyse RNA expression in individual neurons derived from ALS iPSC. This not only allowed us to distinguish gene expression changes in MN and IN, but also enabled us to construct context-specific gene regulatory networks. Network analysis is a powerful way to understand how genes interact with each other to bring about cellular phenotypes(41,42), in the current context case: neuronal degeneration. By building a transcriptional network specific to the ALS landscape, we were able to uncover network modules specifically downregulated in ALS MN. Further, by using master regulator analysis, we were able to identify TFs that drive these disease-associated network modules. Significantly, our analysis uncovered a hitherto unknown link between HOXA1/HOXA5 downregulation and synaptic dysfunction in ALS MN.

The dying back hypothesis posits that neuropathology is initiated in the distal axons and synapses of motor neurons subsequently leading to axon retraction and degeneration of the soma proximally(32). In support of this hypothesis, defects in the neuromuscular junction and distal axons were identified in mouse models of ALS before onset of symptoms(43). Similarly, defects in synaptic activity have been

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

observed in MN derived from ALS associated C9ORF72 repeat expansion and mutant TDP43 patient derived iPSC(9,10). Interestingly, these defects were observed on prolonged *in vitro* culture of the MN correlating with onset of degenerative phenotypes(9). These observations strongly indicate that synaptic dysfunction is a major driver of neurodegeneration in ALS. Synaptic collapse can be a downstream effect of either impaired delivery or production of synaptic proteins and mRNAs. Impaired delivery can occur secondary to defects in the cytoskeleton leading to inefficient axonal transport. Accordingly, neurofilament protein inclusions have been observed in mutant SOD1 MN that correlated with onset of neurite collapse(7). On the other hand, our data reveals inhibition of the synaptic genes at the transcriptional level thereby impairing production. Further, we find direct links between transcriptional suppression of the synaptic program and downregulation of HOXA1 and HOXA5 transcription factors.

HOX genes encode homeobox containing transcription factors that play a role in spinal cord development(44). Specifically, HOX genes are involved in spinal cord patterning along the rostro-caudal axis(22,45). Mutational analyses on mouse Hox genes have revealed roles of these factors in defining regional boundaries, MN development and axon guidance(22). However, little is known about the expression and role of these genes subsequent to MN-specification. Our data indicates that mature MN display high levels of HOX gene expression under normal physiological conditions. We considered the possibility that the observed expression of HOX genes may be due to the fact that iPSC derived MN display a foetal transcriptome. However, this is not the case as HOX genes including HOXA5 were found to be expressed in adult MN micro-dissected from human spinal cords(46). Additionally, both HOXA1 and HOXA5 were found to be expressed in micro-dissected postnatal mouse and rat cervical spinal MN(46,47). This strongly suggests that HOX genes

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

have a role beyond MN specification. Our data indicates that HOXA1 and HOXA5 activate genes involved in synaptic structure and assembly. Additionally, ChIP-Seq mapping of binding sites indicates that most of the identified synaptic genes are direct downstream targets of these transcription factors. Accordingly, deletion of Hoxa5 in post-natal mice was shown to downregulate genes involved in synaptic functions(48). Strikingly, we find that expression of these two transcription factors is specifically high in MN compared to IN but is strongly downregulated in ALS. This suggests one possible mechanism via which mutant SOD1 targets MN homeostatic TFs thereby explaining the relative susceptibility of MN to degeneration. The fact that these factors are also downregulated in D30 ALS MN cultures before the onset of any degenerative phenotypes suggests that these may be the early drivers of degeneration. We also find downregulation of ATF2 and PRDM2, TFs that were linked to core cellular processes such as oxidative phosphorylation, mitochondrial structure and autophagy. It must be noted that inhibition of mitochondrial function and autophagy have been directly linked to defects in axonal transport and neuromuscular synaptic function(49,50). On the other hand, SOX4 and HMGB2 were found to be upregulated in our dataset as well as in ALS mouse models. However, whether these factors drive degeneration in parallel to or downstream of HOXA1/A5, is unclear. The observation that these TFs were not as strongly dysregulated in D30 MN indicates their subsequent downregulation or upregulation may cause exacerbation of the degenerative cycle set in motion earlier by HOXA1 and HOXA5 suppression. Accordingly, knockdown of Hmgb2 has been shown to increase viability while replating mouse MN suggesting that Hmgb2 may be a generic stress-related factor(35). Is HOX gene suppression observed in other regions of the spinal cord? We analysed Hox gene expression in microarray expression data from adult lumbar MN that were microdissected from the SOD1 G93A mouse model of ALS and

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

observed downregulation of Hox genes relevant to the lumbar rostro-caudal address of the spinal cord. This strongly suggests that mutant SOD1 suppresses HOX gene expression across other spinal segments.

The underlying cause of MN degeneration in ALS is very likely to be multi-factorial with multiple drivers collaborating to cause MN demise. Our data indicates that one of these drivers could be the inhibition of HOXA1 and HOXA5 that leads to suppression of genes required for axonal cytoskeleton and synaptic function, thereby resulting in lower availability of synaptic proteins at the neuromuscular junction.

Conclusions

Using in-depth analysis of degenerating MN, we have demostrated that TFs regulating MN homeostasis are affected in ALS, which may explain why MN are highly susceptible to degeneration. Our results display the power of applying single cell analysis to iPSC based neurodegenerative models to uncover core transcriptional drivers of specific pathways involved in motor neuron degeneration. We expect that wider use of single cell genomics especially multi-omics technologies to measure different molecular entities from the same cell (51) combined with network biology will help uncover novel regulators that can be targeted using small molecules or gene therapy.

Conflict of Interests

The authors declare that they have no conflict of interest.

Acknowledgements

We thank the sequencing centre at the Genome Institute of Singapore for help with Illumina sequencing and mapping.

Funding

This work was generously supported by the Wellcome Trust Institutional Strategic Support Award (WT204909MA), the Joint Council Office ASTAR, Singapore and

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

start-up funds provided to the lead author by the University of Exeter, U.K. R.A. is supported by a UKRI EPSRC Innovation Fellowship. C.R.G.W is supported by the Biotechnology and Biological Sciences Research Council-funded South West Biosciences Doctoral Training Partnership [BB/J014400/1; BB/M009122/1].

Authors' contributions

A.B and L.W.S provided funding for the project. A.B. S.C.N and L.O.G designed and conducted the experiments. A.B., P.T, R.A, C.R.G.W analysed the data. All authors contributed towards interpreting the data and writing the manuscript.

References

1. Brown, R.H., Jr. and Al-Chalabi, A. (2017) Amyotrophic Lateral Sclerosis. *N Engl J Med*, **377**, 1602.
2. Taylor, J.P., Brown, R.H., Jr. and Cleveland, D.W. (2016) Decoding ALS: from genes to mechanism. *Nature*, **539**, 197-206.
3. Hardiman, O., Al-Chalabi, A., Chio, A., Corr, E.M., Logroscino, G., Robberecht, W., Shaw, P.J., Simmons, Z. and van den Berg, L.H. (2017) Amyotrophic lateral sclerosis. *Nat Rev Dis Primers*, **3**, 17085.
4. Philips, T. and Rothstein, J.D. (2015) Rodent Models of Amyotrophic Lateral Sclerosis. *Curr Protoc Pharmacol*, **69**, 5 67 61-21.
5. Van Damme, P., Robberecht, W. and Van Den Bosch, L. (2017) Modelling amyotrophic lateral sclerosis: progress and possibilities. *Dis Model Mech*, **10**, 537-549.
6. Sances, S., Bruijn, L.I., Chandran, S., Eggan, K., Ho, R., Klim, J.R., Livesey, M.R., Lowry, E., Macklis, J.D., Rushton, D. *et al.* (2016) Modeling ALS with motor neurons derived from human induced pluripotent stem cells. *Nat Neurosci*, **19**, 542-553.
7. Chen, H., Qian, K., Du, Z., Cao, J., Petersen, A., Liu, H., Blackbourn, L.W.t., Huang, C.L., Errigo, A., Yin, Y. *et al.* (2014) Modeling ALS with iPSCs reveals that mutant SOD1 misregulates neurofilament balance in motor neurons. *Cell Stem Cell*, **14**, 796-809.
8. Kiskinis, E., Sandoe, J., Williams, L.A., Boulting, G.L., Moccia, R., Wainger, B.J., Han, S., Peng, T., Thams, S., Mikkilineni, S. *et al.* (2014) Pathways disrupted in human ALS motor neurons identified through genetic correction of mutant SOD1. *Cell Stem Cell*, **14**, 781-795.
9. Selvaraj, B.T., Livesey, M.R., Zhao, C., Gregory, J.M., James, O.T., Cleary, E.M., Chouhan, A.K., Gane, A.B., Perkins, E.M., Dando, O. *et al.* (2018) C9ORF72 repeat expansion causes vulnerability of motor neurons to Ca(2+)-permeable AMPA receptor-mediated excitotoxicity. *Nat Commun*, **9**, 347.
10. Devlin, A.C., Burr, K., Borooah, S., Foster, J.D., Cleary, E.M., Geti, I., Vallier, L., Shaw, C.E., Chandran, S. and Miles, G.B. (2015) Human iPSC-derived motoneurons harbouring TARDBP or C9ORF72 ALS mutations are dysfunctional despite maintaining viability. *Nat Commun*, **6**, 5999.
11. Simone, R., Balendra, R., Moens, T.G., Preza, E., Wilson, K.M., Heslegrave, A., Woodling, N.S., Niccoli, T., Gilbert-Jaramillo, J., Abdelkarim, S. *et al.* (2018) G-quadruplex-binding small molecules ameliorate C9orf72 FTD/ALS pathology in vitro and in vivo. *EMBO Mol Med*, **10**, 22-31.

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

12. Lopez-Gonzalez, R., Lu, Y., Gendron, T.F., Karydas, A., Tran, H., Yang, D., Petrucelli, L., Miller, B.L., Almeida, S. and Gao, F.B. (2016) Poly(GR) in C9ORF72-Related ALS/FTD Compromises Mitochondrial Function and Increases Oxidative Stress and DNA Damage in iPSC-Derived Motor Neurons. *Neuron*, **92**, 383-391.
13. Krach, F., Batra, R., Wheeler, E.C., Vu, A.Q., Wang, R., Hutt, K., Rabin, S.J., Baughn, M.W., Libby, R.T., Diaz-Garcia, S. *et al.* (2018) Transcriptome-pathology correlation identifies interplay between TDP-43 and the expression of its kinase CK1E in sporadic ALS. *Acta Neuropathol*, **136**, 405-423.
14. De Santis, R., Santini, L., Colantoni, A., Peruzzi, G., de Turris, V., Alfano, V., Bozzoni, I. and Rosa, A. (2017) FUS Mutant Human Motoneurons Display Altered Transcriptome and microRNA Pathways with Implications for ALS Pathogenesis. *Stem Cell Reports*, **9**, 1450-1462.
15. Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S.L. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*, **14**, R36.
16. Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods*, **9**, 357-359.
17. Anders, S., Pyl, P.T. and Huber, W. (2015) HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166-169.
18. Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S. *et al.* (2012) GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res*, **22**, 1760-1774.
19. Bhinge, A., Namboori, S.C., Zhang, X., VanDongen, A.M.J. and Stanton, L.W. (2017) Genetic Correction of SOD1 Mutant iPSCs Reveals ERK and JNK Activated AP1 as a Driver of Neurodegeneration in Amyotrophic Lateral Sclerosis. *Stem Cell Reports*, **8**, 856-869.
20. Ilicic, T., Kim, J.K., Kolodziejczyk, A.A., Bagger, F.O., McCarthy, D.J., Marioni, J.C. and Teichmann, S.A. (2016) Classification of low quality cells from single-cell RNA-seq data. *Genome Biol*, **17**, 29.
21. Maury, Y., Come, J., Piskorowski, R.A., Salah-Mohellibi, N., Chevaleyre, V., Peschanski, M., Martinat, C. and Nedelec, S. (2015) Combinatorial analysis of developmental cues efficiently converts human pluripotent stem cells into multiple neuronal subtypes. *Nat Biotechnol*, **33**, 89-96.
22. Philippidou, P. and Dasen, J.S. (2013) Hox genes: choreographers in neural development, architects of circuit organization. *Neuron*, **80**, 12-34.
23. Lippmann, E.S., Williams, C.E., Ruhl, D.A., Estevez-Silva, M.C., Chapman, E.R., Coon, J.J. and Ashton, R.S. (2015) Deterministic HOX patterning in human pluripotent stem cell-derived neuroectoderm. *Stem Cell Reports*, **4**, 632-644.
24. Wang, J., Jenjaroenpun, P., Bhinge, A., Angarica, V.E., Del Sol, A., Nookaew, I., Kuznetsov, V.A. and Stanton, L.W. (2017) Single-cell gene expression analysis reveals regulators of distinct cell subpopulations among developing human neurons. *Genome Res*, **27**, 1783-1794.
25. Zeisel, A., Munoz-Manchado, A.B., Codeluppi, S., Lonnerberg, P., La Manno, G., Jureus, A., Marques, S., Munguba, H., He, L., Betsholtz, C. *et al.* (2015) Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science*, **347**, 1138-1142.
26. Zhong, S., Zhang, S., Fan, X., Wu, Q., Yan, L., Dong, J., Zhang, H., Li, L., Sun, L., Pan, N. *et al.* (2018) A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature*, **555**, 524-528.
27. Zhang, Y., Sloan, S.A., Clarke, L.E., Caneda, C., Plaza, C.A., Blumenthal, P.D., Vogel, H., Steinberg, G.K., Edwards, M.S., Li, G. *et al.* (2016) Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. *Neuron*, **89**, 37-53.

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

28. Huang da, W., Sherman, B.T. and Lempicki, R.A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*, **37**, 1-13.
29. Kiselev, V.Y., Kirschner, K., Schaub, M.T., Andrews, T., Yiu, A., Chandra, T., Natarajan, K.N., Reik, W., Barahona, M., Green, A.R. *et al.* (2017) SC3: consensus clustering of single-cell RNA-seq data. *Nat Methods*, **14**, 483-486.
30. Bui, T.V., Akay, T., Loubani, O., Hnasko, T.S., Jessell, T.M. and Brownstone, R.M. (2013) Circuits for grasping: spinal dI3 interneurons mediate cutaneous control of motor behavior. *Neuron*, **78**, 191-204.
31. Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S. and Rinn, J.L. (2014) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol*, **32**, 381-386.
32. Dadon-Nachum, M., Melamed, E. and Offen, D. (2011) The "dying-back" phenomenon of motor neurons in ALS. *J Mol Neurosci*, **43**, 470-477.
33. Zhang, B. and Horvath, S. (2005) A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*, **4**, Article17.
34. Langfelder, P. and Horvath, S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**, 559.
35. Ikiz, B., Alvarez, M.J., Re, D.B., Le Verche, V., Politi, K., Lotti, F., Phani, S., Pradhan, R., Yu, C., Croft, G.F. *et al.* (2015) The Regulatory Machinery of Neurodegeneration in In Vitro Models of Amyotrophic Lateral Sclerosis. *Cell Rep*, **12**, 335-345.
36. Basso, K., Margolin, A.A., Stolovitzky, G., Klein, U., Dalla-Favera, R. and Califano, A. (2005) Reverse engineering of regulatory networks in human B cells. *Nat Genet*, **37**, 382-390.
37. Fletcher, M.N., Castro, M.A., Wang, X., de Santiago, I., O'Reilly, M., Chin, S.F., Rueda, O.M., Caldas, C., Ponder, B.A., Markowetz, F. *et al.* (2013) Master regulators of FGFR2 signalling and breast cancer risk. *Nat Commun*, **4**, 2464.
38. Nardo, G., Iennaco, R., Fusi, N., Heath, P.R., Marino, M., Trolese, M.C., Ferraiuolo, L., Lawrence, N., Shaw, P.J. and Bendotti, C. (2013) Transcriptomic indices of fast and slow disease progression in two mouse models of amyotrophic lateral sclerosis. *Brain*, **136**, 3305-3332.
39. Yan, J., Enge, M., Whitington, T., Dave, K., Liu, J., Sur, I., Schmierer, B., Jolma, A., Kivioja, T., Taipale, M. *et al.* (2013) Transcription factor binding in human cells occurs in dense clusters formed around cohesin anchor sites. *Cell*, **154**, 801-813.
40. Yao, Z., Mich, J.K., Ku, S., Menon, V., Krostag, A.R., Martinez, R.A., Furchtgott, L., Mulholland, H., Bort, S., Fuqua, M.A. *et al.* (2017) A Single-Cell Roadmap of Lineage Bifurcation in Human ESC Models of Embryonic Brain Development. *Cell Stem Cell*, **20**, 120-134.
41. Li, Y., McGrail, D.J., Latysheva, N., Yi, S., Babu, M.M. and Sahni, N. (2018) Pathway perturbations in signaling networks: Linking genotype to phenotype. *Semin Cell Dev Biol*.
42. Lefebvre, C., Rieckhof, G. and Califano, A. (2012) Reverse-engineering human regulatory networks. *Wiley Interdiscip Rev Syst Biol Med*, **4**, 311-325.
43. Fischer, L.R., Culver, D.G., Tennant, P., Davis, A.A., Wang, M., Castellano-Sanchez, A., Khan, J., Polak, M.A. and Glass, J.D. (2004) Amyotrophic lateral sclerosis is a distal axonopathy: evidence in mice and man. *Exp Neurol*, **185**, 232-240.
44. Carpenter, E.M. (2002) Hox genes and spinal cord development. *Dev Neurosci*, **24**, 24-34.
45. Mazzoni, E.O., Mahony, S., Peljto, M., Patel, T., Thornton, S.R., McCuine, S., Reeder, C., Boyer, L.A., Young, R.A., Gifford, D.K. *et al.* (2013) Saltatory remodeling of Hox chromatin in response to rostrocaudal patterning signals. *Nat Neurosci*, **16**, 1191-1198.
46. Nichterwitz, S., Chen, G., Aguila Benitez, J., Yilmaz, M., Storvall, H., Cao, M., Sandberg, R., Deng, Q. and Hedlund, E. (2016) Laser capture microscopy coupled with Smart-seq2 for precise spatial transcriptomic profiling. *Nat Commun*, **7**, 12139.

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

47. Hedlund, E., Karlsson, M., Osborn, T., Ludwig, W. and Isacson, O. (2010) Global gene expression profiling of somatic motor neuron populations with different vulnerability identify molecules and pathways of degeneration and protection. *Brain*, **133**, 2313-2330.
48. Lizen, B., Moens, C., Mouheiche, J., Sacre, T., Ahn, M.T., Jeannotte, L., Salti, A. and Gofflot, F. (2017) Conditional Loss of Hoxa5 Function Early after Birth Impacts on Expression of Genes with Synaptic Function. *Front Mol Neurosci*, **10**, 369.
49. Miller, K.E. and Sheetz, M.P. (2004) Axonal mitochondrial transport and potential are correlated. *J Cell Sci*, **117**, 2791-2804.
50. Rudnick, N.D., Griffey, C.J., Guarnieri, P., Gerbino, V., Wang, X., Piersaint, J.A., Tapia, J.C., Rich, M.M. and Maniatis, T. (2017) Distinct roles for motor neuron autophagy early and late in the SOD1(G93A) mouse model of ALS. *Proc Natl Acad Sci U S A*, **114**, E8294-E8303.
51. Macaulay, I.C., Ponting, C.P. and Voet, T. (2017) Single-Cell Multiomics: Multiple Measurements from Single Cells. *Trends Genet*, **33**, 155-168.

Figure 1. Single cell RNA-seq of iPSC-derived MN.

a) Schematic depicting the differentiation protocol used to derive MN from human iPSC. The horizontal line depicts the timeline of differentiation with the numbers indicating days. Day 0 indicates the first media change to induce differentiation. RA: retinoic acid, Pur: Purmorphamine. **b)** MN at day 30 stain positive for the MN markers ISL1, CHAT and NF-H as well as the pan-neuronal markers TUJ1 and MAP2. Representative images were derived from differentiation of the isogenic control iPSC. **c)** MN (ISL1+, TUJ1+) were counted at d30 and d44 of the differentiation protocol. D44 MN counts were normalized to d30 counts. SOD1 E100G indicates the MN derived from mutant SOD1 iPSC. SOD1 E100E indicates the isogenic corrected control MN. Error bars shown are s.e.m. n=3 independent biological replicates. * indicates p-value < 0.01. n.s. indicates not significant. **d)** Heatmap displaying % of cells expressing specific HOX genes. FPKM threshold was set at 0.1. White space indicates that the corresponding HOX paralog is not expressed in humans. Coloured solid arrows indicate the HOX code for specific spinal segments along the rostro-caudal axis. **e)** Boxplots displaying distribution of expression levels of MN markers (ISL1, CHAT, SLC18A3 or VaCHT), interneuron

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

markers (GAD1, GAD2, ARX) and non-neuronal markers (S100B, SOX9, MKI67) across all cells.

Figure 2. Clustering and classification of single cells into neural subtypes.

a) Flow chart displaying the steps taken to classify single cells into neurons and glia, and further classify the neurons into MN and IN. Numbers in brackets denote the number of cells retained after each filtering step. b) Heat map displaying clustering of all cells based on neurons vs glia classifier gene set. SC3 clustered cells into three distinct clusters. The heatmap colours indicate a similarity score between the cells based on how frequently they were clustered together using various SC3 clustering parameters. Red indicates high, while blue indicates low similarity score. c) Expression profiles for neuronal markers (SYP, RBFOX1, SYN1, CHAT, GAD1, GAD2, CAMKV) and non-neuronal markers (S100B, SOX9, REST, PAX6 and MKI67) across the 3 clusters. We estimated the median expression of each marker gene for each cluster and then scaled the values by the highest expression such that they range from 0 (not expressed) to 1 (highest expression). Comparison of the neuronal and non-neuronal marker profiles indicates that cluster 1 comprises of non-neuronal cells while clusters 2 and 3 comprise of neurons. Red represents a value of 1 while blue represents 0. d) Heatmap displaying clustering of neurons in the control dataset using MN and IN markers into 3 clusters. Red indicates high, while blue indicates low cell similarity score as described for panel b. e) Expression profiles for MN markers (CHAT, ISL1, SLC18A3, ONECUT1 and PRPH) and IN markers (GAD1, GAD2, ARX, DLX1 and DLX6) across the identified clusters in panel d. Scaled FPKM values were estimated as described for panel c. Cluster c1 expresses IN markers while clusters c2 and c3 express MN markers. f) Heatmap displaying clustering of neurons in the ALS dataset using MN and IN markers into 3 clusters. Red indicates high while

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

blue indicates low cell similarity score as described for panel b. g) Expression profiles for MN and IN markers across ALS neuronal clusters in panel f indicate that cluster a1 comprises of IN while cluster a2 is comprised of MN. Cluster a3 expressed both PRPH and GAD2 but clustered with MN. Hence, this cluster was removed from further analysis as it likely represented doublets. h) GO enrichment analysis of genes downregulated in ALS MN compared to control MN. i) GO enrichment analysis of genes upregulated in ALS MN compared to control MN.

Figure 3. Network analysis of all neurons using WGCNA

a) WGCNA identified 13 modules of co-regulated genes across the neuronal dataset. Module eigengenes for each module were compared across 4 groups: ALS MN vs Control MN, ALS IN vs Control IN, ALS MN vs ALS IN and Control MN vs Control IN. Association of a given module with a specific group was estimated using Pearson correlations. For each A vs B comparison, a correlation of 1 (red) indicates that the module eigengenes were higher in group A compared to group B. A correlation of -1 (blue) indicates that the module eigengenes were lower in group A compared to group B. Correlations with an adjusted p-value < 0.01 were considered significant (indicated by red *). **b)** GO enrichment analysis of modules significantly associated with ALS MN. The blue, purple and black modules were found to be negatively associated with ALS MN. **c)** Correlations between TF expression and module eigengenes for the blue, purple and black modules. Correlation with an adjusted p-value of < 0.01 were considered significant.

Figure 4. Master regulator analysis (MRA) of ALS disease signature.

a) TFs identified by the MRA as the most significant drivers of the ALS disease signature. The red/green heatmap indicates the fold upregulation or downregulation of the TF in ALS MN compared to control MN. Red indicates upregulation while green indicates downregulation. The grey panel shows the distribution of the targets

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of each TF (regulon of the TF) amongst the top 500 upregulated and downregulated genes in ALS MN. Though the analysis was performed across the entire dataset, only the top 500 genes are shown for visualization purposes. TF targets were further classified as activated or repressed based on correlation. Yellow vertical ticks indicate genes activated by the TF while blue vertical ticks indicate genes repressed by the TF. If a TF is downregulated in ALS MN, then it follows that genes activated by the TF (yellow vertical ticks) are downregulated in ALS MN (right side of the grey panel). Similar logic applies for TFs that are upregulated. Such TFs were deemed to be concordant with their regulons. **b)** Sub-network displaying the interactions between the master regulators identified in panel 4a. Only interactions with a FDR < 0.1 were used to build this network. The edges denote the significance of the interaction: thicker the edge, higher the significance or strength of the interaction between the TFs. Node colour indicates whether the TF was upregulated (red) or downregulated (green) in ALS MN compared to control MN. Legend displays log2(fold change). **c)** Boxplots displaying the expression profiles of the identified master regulators across ALS MN, ALS IN, Control MN and Control IN. TFs were grouped into 4 groups (colour coded red, blue, yellow and green) based on their expression profiles. Group red showed high expression in control MN compared to control IN and was downregulated in ALS MN. Group blue showed similar expression in control MN and IN but was lower in ALS MN. Group yellow showed similar expression in control MN and IN, but was dysregulated in both ALS MN and IN. Group green was lower in control MN compared to IN but increased in both ALS MN and IN compared to control.

Figure 5. Analysis of master regulator TFs in the mouse model of ALS.

a) Candidate TF and **c)** Lumbar Hox gene expression changes in ALS vs control MN in the 129Sv strain. **b)** Candidate TF and **d)** Lumbar Hox gene expression changes in

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ALS vs control MN in the C57 strain. preSym: pre-symptomatic, Sym: symptomatic. Yellow indicates upregulation while cyan indicates downregulation in ALS MN compared to control MN. P-values were estimates using Student's two-tailed t-test. P-values were corrected by the Benjamini Hochberg procedure to yield false discover rates (FDR). The heatmap displays the FDR for each comparison.

Figure 6. Inhibition of synaptic genes in ALS MN maybe due to inhibition of HOXA1 and HOXA5.

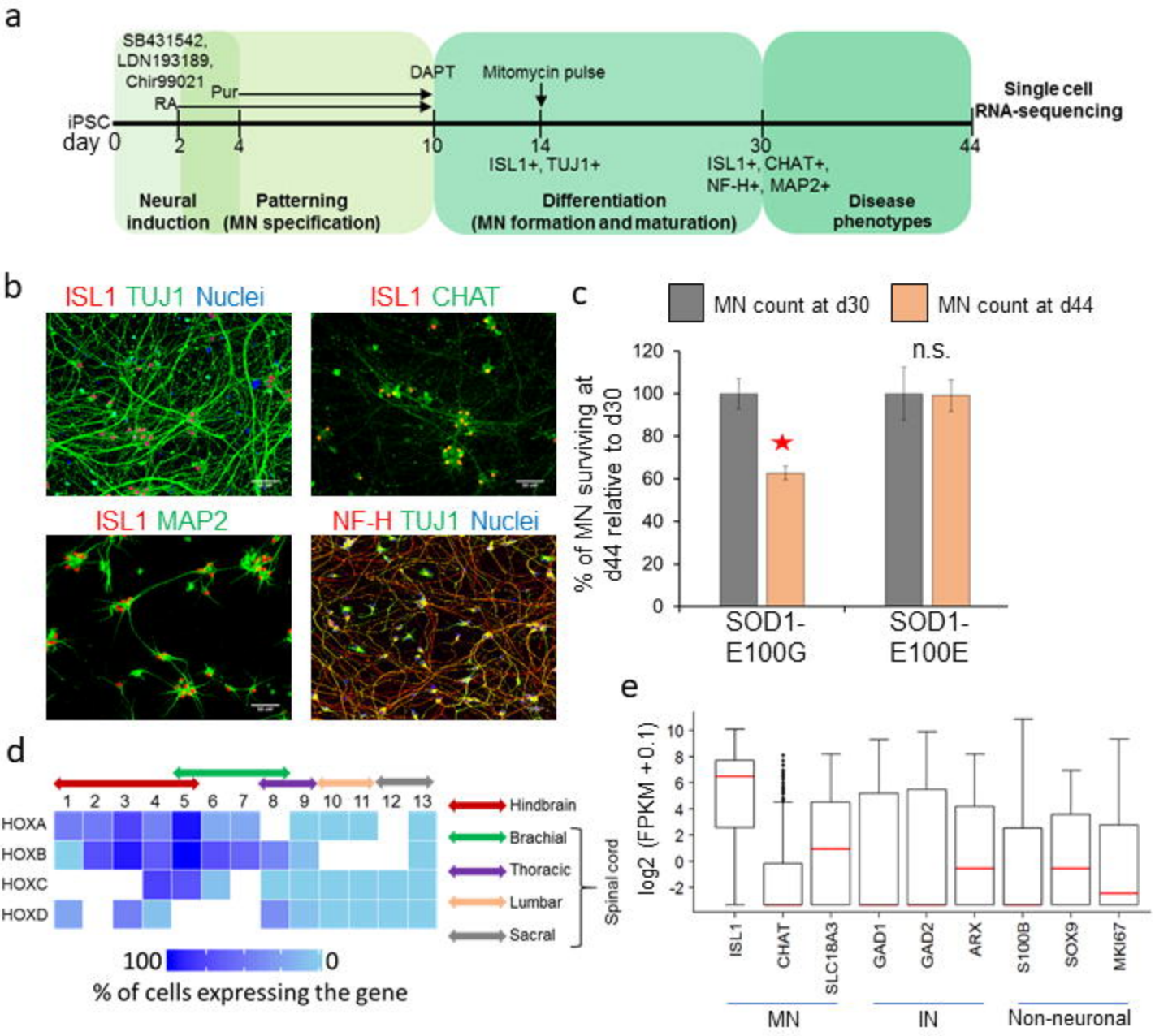
a) Quantitative RT-PCR to measure expression levels of master regulators in bulk ALS MN cultures compared to control at day 30 of differentiation i.e. before onset of any degenerative phenotypes. HOXA1, HOXA5, HOXD8, PRDM2 and ZBTB6 were downregulated in ALS compared to control. ** indicates p-value < 0.01, * indicates p-value < 0.05. Error bars shown are s.e.m. n=3 independent differentiations. P-values were estimated using Student's 2-tailed t-test. **b)** Quantitative RT-PCR to measure expression levels of synaptic targets of HOXA1 and HOXA5 in bulk ALS MN cultures compared to control at day 30 of differentiation. * indicates p-value < 0.05. Error bars shown are s.e.m. n=3 independent differentiations. P-values were estimated using Student's 1-tailed t-test as we expected the synaptic genes to be downregulated in ALS MN. **c)** Quantitative RT-PCR to measure expression levels of synaptic targets of HOXA1 and HOXA5 after exogenously expressing HOXA1 and HOXA5 in SH-Sy5Y neuronal cells. Exogenous expression of GFP was used as a control and fold changes were estimated relative to control. * indicates p-value < 0.05. Error bars shown are s.e.m. n=3 independent transfections. P-values were estimated using Student's 1-tailed t-test as we expected the synaptic genes to be upregulated after HOXA1/A5 expression. **d)** Heatmap displays HOXA1 and HOXA5 binding sites on the genome proximal to synaptic targets as measured by ChIP-Seq analysis in colon carcinoma cell lines. Binding sites were classified as not enriched

bioRxiv preprint doi: https://doi.org/10.1101/593129; this version posted March 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(white), weakly enriched (light blue) and strongly enriched (sky blue). A gene was considered a target if an enriched TF binding site was observed within the gene body or 10kb upstream from the transcriptional start site or 2.0 kb downstream from the transcriptional stop site. A gene was considered expressed if a H3K4Me3 signal was observed in the promoter (5 kb from the start site) of that gene or H3K4Me1 signal was observed either in the promoter or within the gene body.

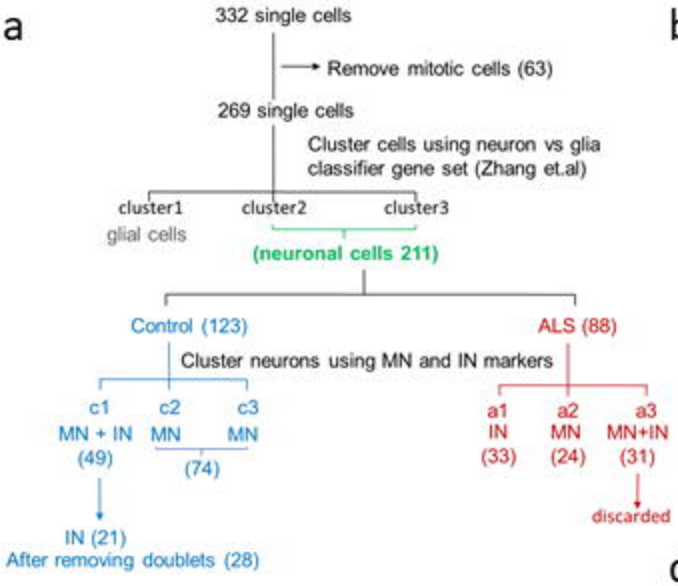
certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available u
aCC-BY-NC-ND 4.0 International license.

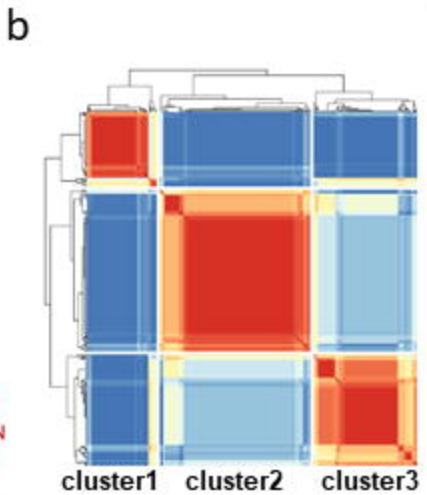
Fig.1

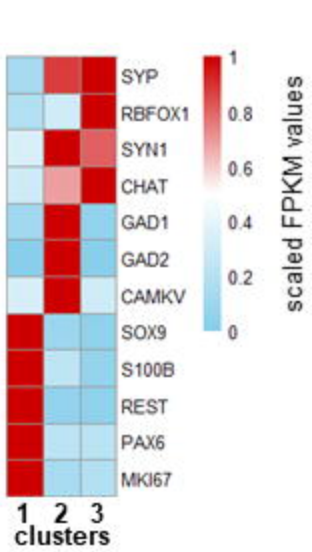


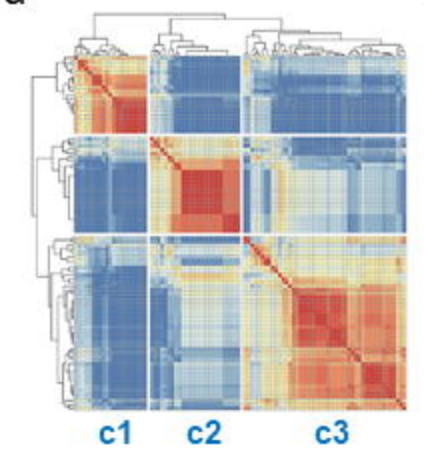
certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available u
aCC-BY-NC-ND 4.0 International license.

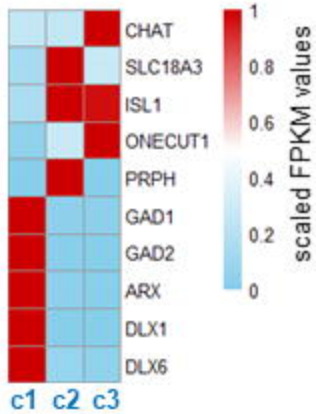
Fig.2

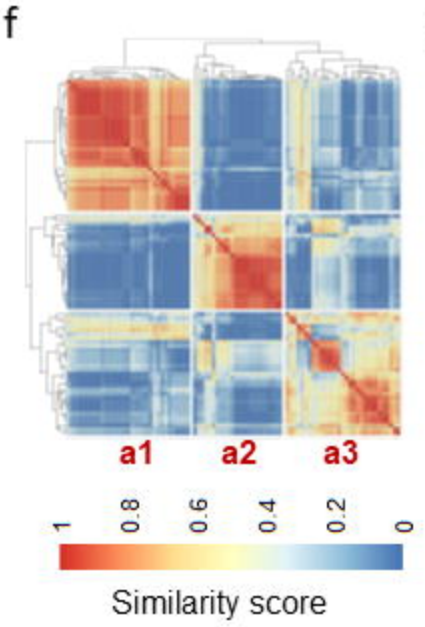


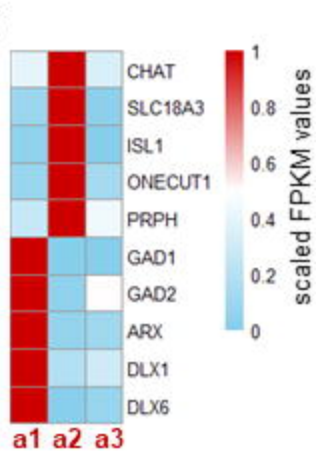












h

Downregulated in ALS MN

GO term	FDR
synapse organization	5.90E-08
synaptic signaling	1.33E-06
axon	1.04E-05
synapse assembly	8.46E-05
synaptic membrane	0.00018
neuromuscular junction	0.00092
mitochondrial outer membrane	0.02244
NADH:ubiquinone oxidoreductase, mitochondria	0.02285
Oxidative phosphorylation	0.04570

i

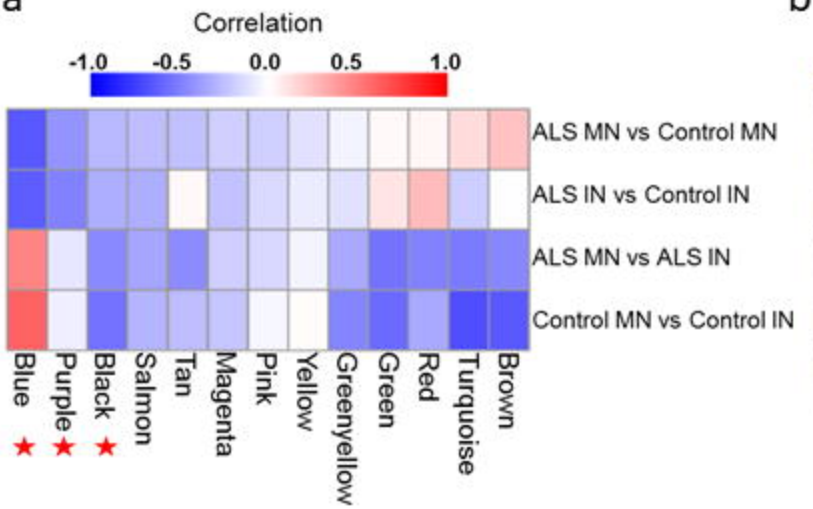
Upregulated in ALS MN

GO term	FDR
protein targeting to ER	7.92E-11
Cell Cycle, Mitotic	0.00044

certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available u
aCC-BY-NC-ND 4.0 International license.

Fig.3

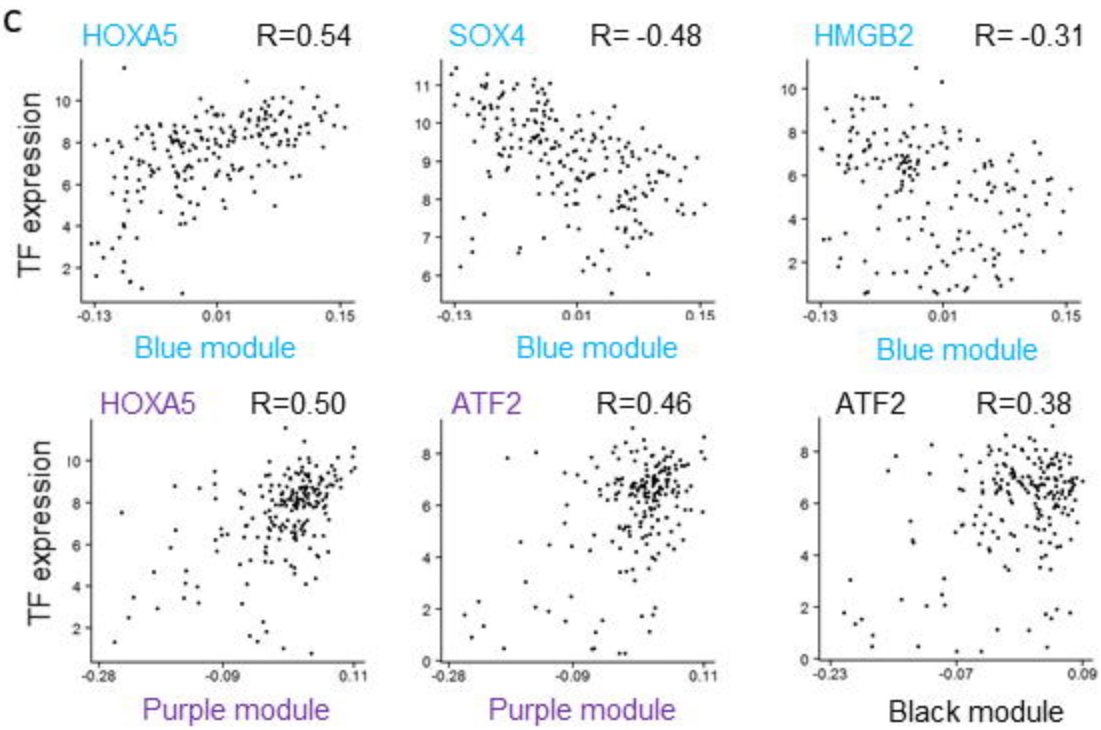
a



b

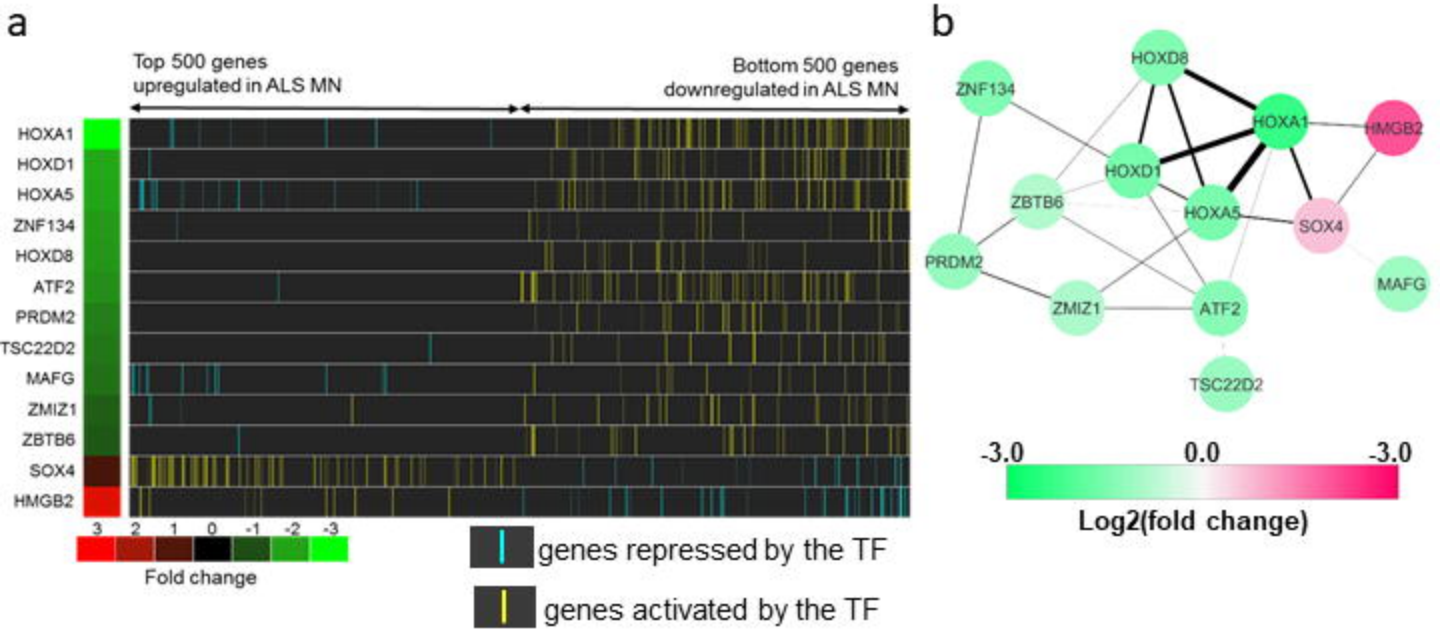
Module	GO term	FDR
Blue	synapse	1.03E-21
	anterograde trans-synaptic signaling	2.31E-17
	axon development	1.05E-09
	neurotransmitter transport	1.69E-09
Purple	regulation of cytoskeleton organization	7.15E-05
	microtubule-based process	0.000164
	cytoskeletal protein binding	0.002501
Black	membrane protein complex	1.13E-16
	mitochondrial inner membrane	5.95E-12
	Oxidative phosphorylation (KEGG)	8.44E-11

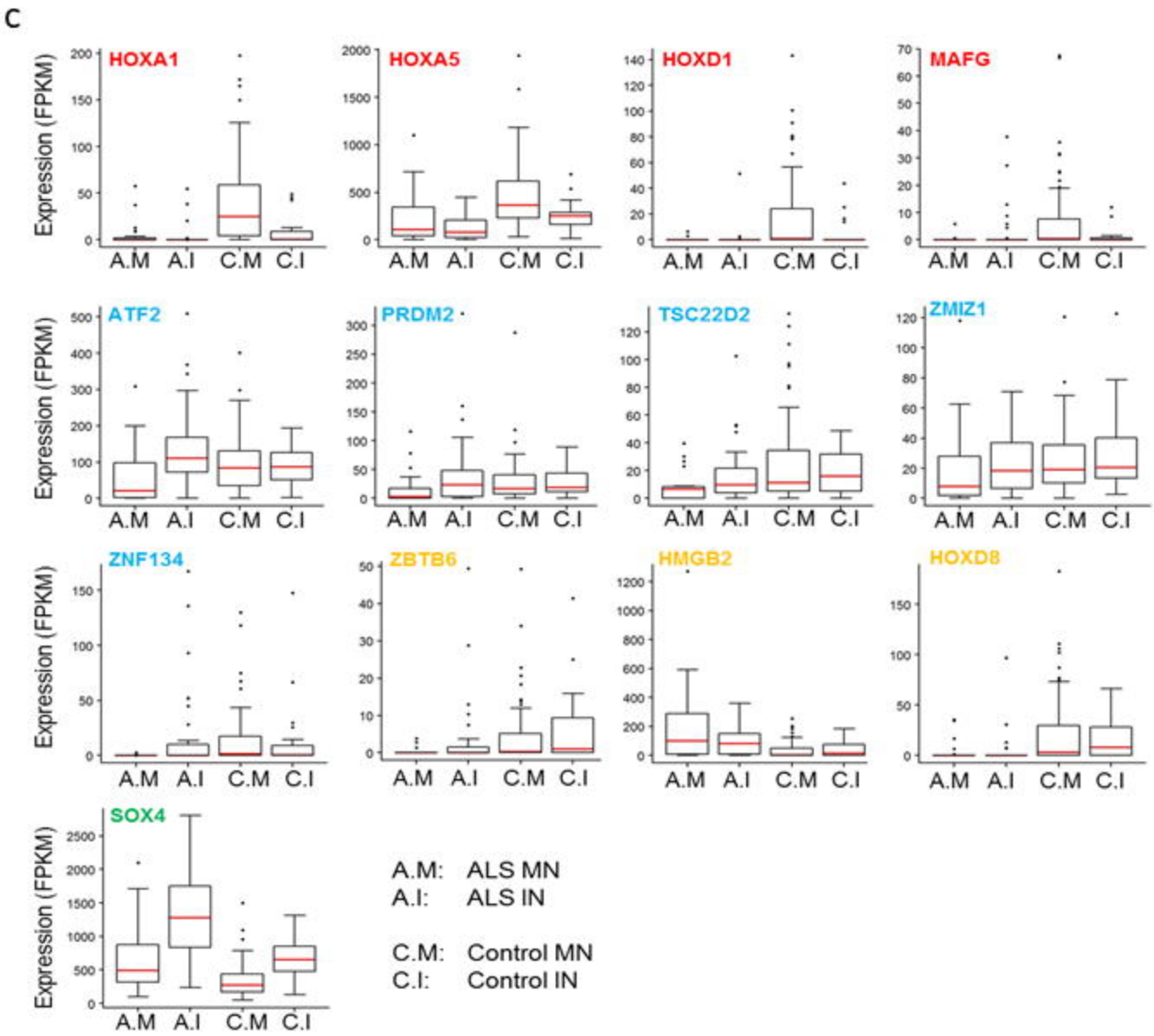
c



certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available u
aCC-BY-NC-ND 4.0 International license.

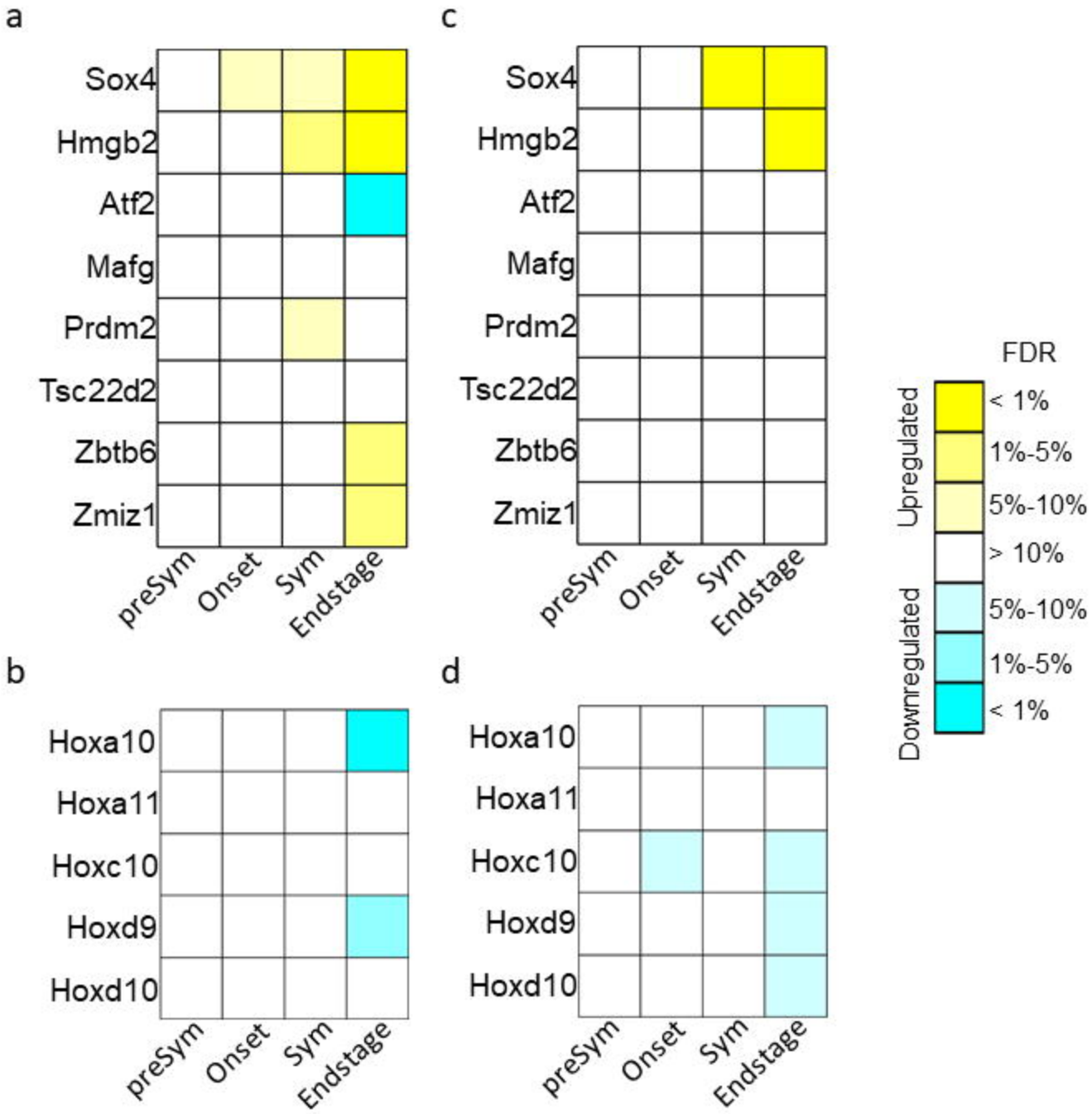
Fig.4





certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available u
aCC-BY-NC-ND 4.0 International license.

Fig.5



certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available u
aCC-BY-NC-ND 4.0 International license.

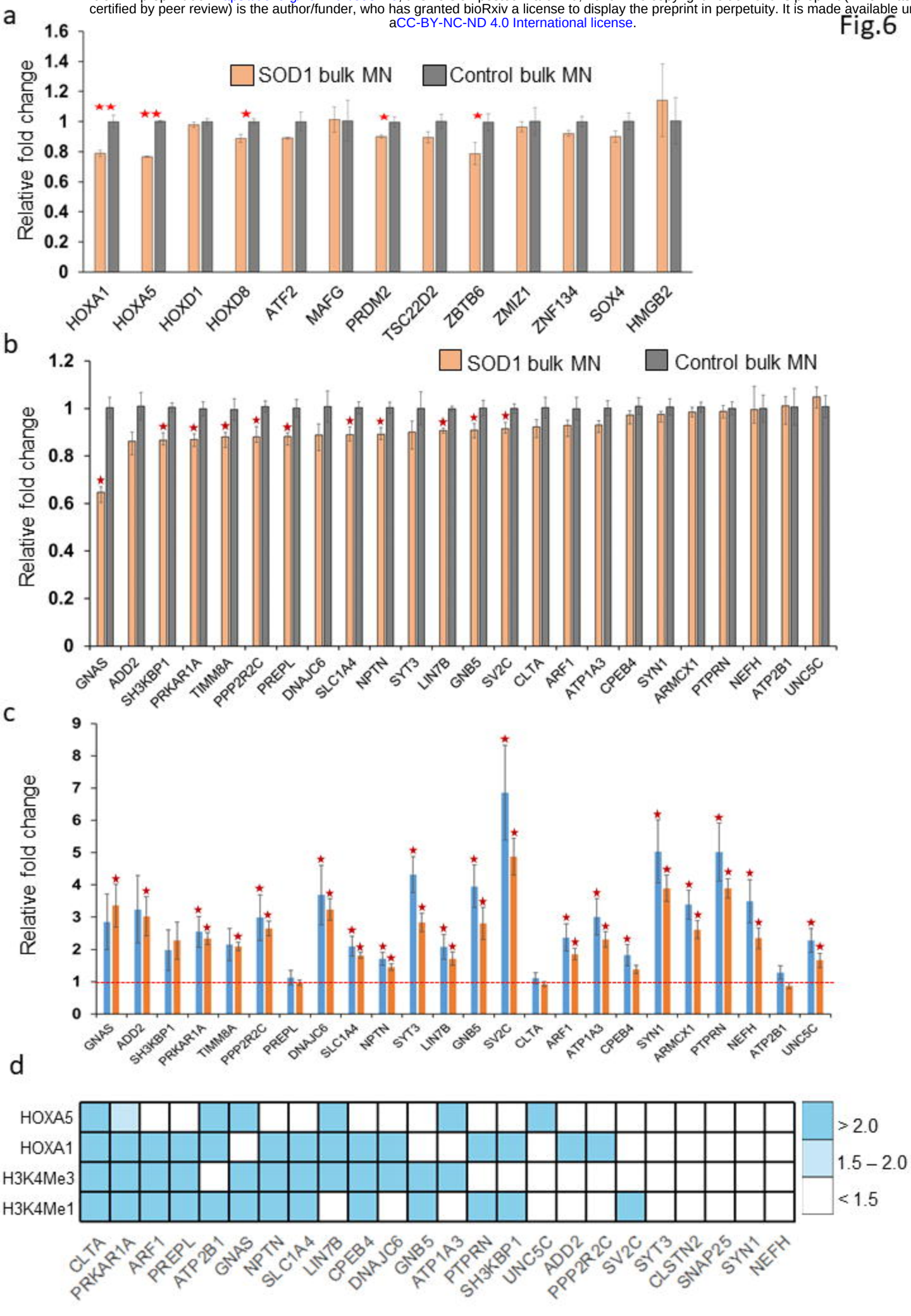



Fig.6