



This is a repository copy of *SRSF1-dependent inhibition of C9ORF72-repeat RNA nuclear export : genome-wide mechanisms for neuroprotection in amyotrophic lateral sclerosis*.

White Rose Research Online URL for this paper:  
<https://eprints.whiterose.ac.uk/177074/>

Version: Published Version

---

**Article:**

Castelli, L.M., Cutillo, L., Souza, C.D.S. et al. (14 more authors) (2021) SRSF1-dependent inhibition of C9ORF72-repeat RNA nuclear export : genome-wide mechanisms for neuroprotection in amyotrophic lateral sclerosis. *Molecular Neurodegeneration*, 16 (1). 53.

<https://doi.org/10.1186/s13024-021-00475-y>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:  
<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.




[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

RESEARCH ARTICLE

Open Access



# SRSF1-dependent inhibition of *C9ORF72*-repeat RNA nuclear export: genome-wide mechanisms for neuroprotection in amyotrophic lateral sclerosis

Lydia M. Castelli<sup>1†</sup>, Luisa Cutillo<sup>2†</sup>, Cleide Dos Santos Souza<sup>1</sup>, Alvaro Sanchez-Martinez<sup>3</sup>, Ilaria Granata<sup>4</sup>, Ya-Hui Lin<sup>1</sup>, Monika A. Myszczyńska<sup>1</sup>, Paul R. Heath<sup>1</sup>, Matthew R. Livesey<sup>1,5</sup>, Ke Ning<sup>1,5</sup>, Mimoun Azzouz<sup>1,5</sup>, Pamela J. Shaw<sup>1,5</sup>, Mario R. Guarracino<sup>4</sup>, Alexander J. Whitworth<sup>3</sup>, Laura Ferraiuolo<sup>1,5</sup>, Marta Milo<sup>6,7\*</sup> and Guillaume M. Hautbergue<sup>1,5\*</sup> 

## Abstract

**Background:** Loss of motor neurons in amyotrophic lateral sclerosis (ALS) leads to progressive paralysis and death. Dysregulation of thousands of RNA molecules with roles in multiple cellular pathways hinders the identification of ALS-causing alterations over downstream changes secondary to the neurodegenerative process. How many and which of these pathological gene expression changes require therapeutic normalisation remains a fundamental question.

**Methods:** Here, we investigated genome-wide RNA changes in *C9ORF72*-ALS patient-derived neurons and *Drosophila*, as well as upon neuroprotection taking advantage of our gene therapy approach which specifically inhibits the SRSF1-dependent nuclear export of pathological *C9ORF72*-repeat transcripts. This is a critical study to evaluate (i) the overall safety and efficacy of the partial depletion of SRSF1, a member of a protein family involved itself in gene expression, and (ii) a unique opportunity to identify neuroprotective RNA changes.

\* Correspondence: [marta.milo@astrazeneca.com](mailto:marta.milo@astrazeneca.com);  
[g.hautbergue@sheffield.ac.uk](mailto:g.hautbergue@sheffield.ac.uk)

<sup>†</sup>Lydia M. Castelli and Luisa Cutillo contributed equally to this work.

<sup>6</sup>Department of Biomedical Science, University of Sheffield, Western Bank, Sheffield S10 2TN, UK

<sup>1</sup>Sheffield Institute for Translational Neuroscience (SITraN), Department of Neuroscience, University of Sheffield, 385 Glossop Road, Sheffield S10 2HQ, UK

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

**Results:** Our study shows that manipulation of 362 transcripts out of 2257 pathological changes, in addition to inhibiting the nuclear export of repeat transcripts, is sufficient to confer neuroprotection in C9ORF72-ALS patient-derived neurons. In particular, expression of 90 disease-altered transcripts is fully reverted upon neuroprotection leading to the characterisation of a human C9ORF72-ALS disease-modifying gene expression signature. These findings were further investigated *in vivo* in diseased and neuroprotected *Drosophila* transcriptomes, highlighting a list of 21 neuroprotective changes conserved with 16 human orthologues in patient-derived neurons. We also functionally validated the high neuroprotective potential of one of these disease-modifying transcripts, demonstrating that inhibition of ALS-upregulated human KCNN1–3 (*Drosophila* SK) voltage-gated potassium channel orthologs mitigates degeneration of human motor neurons and *Drosophila* motor deficits.

**Conclusions:** Strikingly, the partial depletion of SRSF1 leads to expression changes in only a small proportion of disease-altered transcripts, indicating that not all RNA alterations need normalization and that the gene therapeutic approach is safe in the above preclinical models as it does not disrupt globally gene expression. The efficacy of this intervention is also validated at genome-wide level with transcripts modulated in the vast majority of biological processes affected in C9ORF72-ALS. Finally, the identification of a characteristic signature with key RNA changes modified in both the disease state and upon neuroprotection also provides potential new therapeutic targets and biomarkers.

**Keywords:** Amyotrophic lateral sclerosis, C9ORF72-repeat expansions, Pre-clinical models, Transcriptome, Genome-wide mechanisms of neuroprotection, SRSF1-dependent RNA nuclear export, Disease-modifying gene expression signature, Voltage-gated potassium ion channel

## Background

Polymorphic GGGGCC hexanucleotide-repeat expansions in the *C9ORF72* gene cause the most common forms of familial amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) [1, 2], a spectrum of fatal diseases which respectively lead to progressive death of motor neurons and of neurons in the frontal/temporal lobes of the brain. While ALS causes progressive paralysis and death usually within 2–5 years from symptom onset [3, 4], FTD patients present with altered cognitive features and psychological disinhibition [5]. No effective disease-modifying therapy is currently available for these diseases.

Alteration of multiple biological processes in ALS leads to pathophysiological changes including stress responses, mitochondrial dysfunction, alterations in axonal transport, autophagy and protein clearance, cell death, excitotoxicity, neuroinflammation and dysregulated astrocyte-motor neuron crosstalk among others [6] in association with widespread alteration of the RNA metabolism and gene expression [7]. Consistent with this, the expression levels, alternative splicing and polyadenylation site usage of thousands of transcripts are altered in *C9ORF72* repeat-expansion carriers [8, 9]. This raises serious challenges for the identification of altered transcripts causing neurodegeneration since an unknown proportion of RNA changes may occur as a secondary downstream process resulting from disease progression and dysregulation of RNA-processing factors. *C9ORF72* repeat-expansions contribute to

neuronal injury through three non-mutually exclusive mechanisms: haploinsufficiency, repeat-RNA sequestration of proteins and repeat-associated non-AUG (RAN) translation of neurotoxic dipeptide-repeat proteins (DPRs), the latter being considered one of the main drivers of neurodegeneration [10, 11].

We recently showed that repeat-RNA sequestration of the splicing factor and nuclear export adaptor SRSF1 (serine/arginine-rich splicing factor 1 previously known as ASF/SF2) triggers the nuclear export of intron-1-retaining *C9ORF72* repeat transcripts which lead to the cytoplasmic RAN translation of DPRs [12]. Conversely, depleting SRSF1 confers a selective disease-modifying approach for neuroprotection in patient-derived neurons and *Drosophila* models of disease [12]. Interaction of dephosphorylated SRSF1 with NXF1 (nuclear RNA export factor 1) upon completion of splicing [13, 14] induces handover of the mRNA through remodeling of NXF1 into a high RNA-binding conformation [15, 16] that licenses the nuclear export process via transient interactions of NXF1 with the protruding FG-repeats of nucleoporins which decorate the channel of the nuclear pore [17, 18]. The remodeling mechanism of NXF1 provides in turn a control mechanism to retain un-spliced transcripts within the nucleus [19–21]. *C9ORF72* intron-1 repeat-RNA sequestration of SRSF1 is thought to trigger inappropriate remodeling of NXF1 into the high RNA affinity mode that promotes the nuclear export of pathological *C9ORF72* pre-mRNAs which retain the intron-1 repeat expansions [12, 22].

The genome-wide functions of SRSF1 have not yet been investigated in neurons. In proliferative cells, it contributes to shaping the transcriptome through several RNA-processing functions involved in: (i) constitutive and alternative splicing [23–25]; (ii) nuclear export of mRNAs [13, 15, 26]; (iii) RNA stability/ surveillance [27] and (iv) release of paused RNA polymerase II from promoters [28]. The functions of SRSF1 have been extensively studied in immortalised and cancer cells, where overexpression leads to altered splicing functions linked to transformation and oncogenesis [29, 30]. SRSF1 plays an essential role in the tissue-specific splicing of the CaMKII $\delta$  (Ca<sup>2+</sup>/calmodulin-dependent kinase II $\delta$ ) transcript that occurs during embryonic development of the heart, while in contrast, SR-rich proteins are dispensable to the viability of mature cardiomyocytes [31]. Accordingly, individual depletion (>90%) of each of the conserved SRSF1–7 proteins only affected the SRSF1-dependent nuclear export of 225 transcripts in immortalised cells and approximately 100–400 transcripts for the other SRSF factors, indicating that the NXF1-dependent nuclear export adaptor function involves redundancy and/or cooperation [26]. Taken together, further investigation is required to understand the genome-wide contribution of SRSF1 function in a neuronal context.

In this study, we examined the genome-wide mechanisms of neuroprotection by which SRSF1 depletion confers neuroprotection through investigation of whole-cell and cytoplasmic transcriptomes from healthy control and C9ORF72-ALS *Drosophila* and patient-derived neurons. Strikingly, the nuclear export inhibition of C9ORF72 repeat transcripts led to expression changes for ~250 human mRNAs involved in multiple disease pathways out of ~2250 changes in C9ORF72-ALS patient-derived neurons, indicating that, while the neurodegenerative process is characterised by a large number of gene expression changes, a small proportion of transcript changes is sufficient to suppress the neurodegenerative process. The analysis of SRSF1-depleted C9ORF72-ALS *Drosophila* transcriptomes led to a similar observation with conserved manipulation of cellular pathways. Moreover, levels of approximately one third of SRSF1-RNAi induced neuroprotective changes are completely reversed compared to the disease state, providing small in vitro and in vivo disease-modifying gene expression signatures. Finally, based on the integration of human-*Drosophila* transcriptomes that identified 16 conserved human neuroprotective transcript changes, we used pharmacological inhibition and genetic manipulation to show that the expression levels of a conserved small conductance Ca<sup>2+</sup>-activated potassium channel, which is increased in the ALS models, can be manipulated to mitigate the death of human C9ORF72-ALS

motor neurons as well as the neurodegeneration-associated locomotor deficits in *Drosophila*.

## Methods

### Experimental cell and animal models

#### *Patient-derived iNPC lines and differentiation into iNeurons*

The human iNeuron cells were derived from iNPC lines of three healthy controls and three C9ORF72-ALS patients (Table 1). They have been used in multiple studies including the manuscript describing the neuroprotective role of the SRSF1-RNAi depletion [12]. The following Research Resource Identifiers have been registered for human iNPC lines 154 (RRID:CVCL\_XX76), 155 (RRID:CVCL\_UF81), 3050 (RRID: CVCL\_UH66), 78 (RRID:CVCL\_UF8), 183 (RRID:CVCL\_UF85) and 201 (RRID:CVCL\_UF86). Control and patient iNeurons were differentiated and cultured exactly as described previously [12] in 6-well plates (200,000 cells per well) and transduced with 5 MOI LV-Control-RNAi or LV-SRSF1-RNAi lentivirus for 5 days.

#### *Patient derived iPSC lines and differentiation into motor neurons*

The human motor neurons were derived from human induced pluripotent stem cells (iPSC) lines (Table 1). iPSC cells derived from 2 patients carrying the mutation in C9orf72 (CS28iALS-C9nxx; RRID:CVCL\_W558 and CS29iALS-C9nxx; RRID:CVCL\_W559) and 1 iPSC cell derived from an unaffected control (CS14iCTR-21nxx; RRID:CVCL\_JK54) were obtained from Cedars-Sinai, a nonprofit academic healthcare organization. The cell line control MIFF1 (RRID:CVCL\_1E69) [32] was kindly provided by Prof Peter Andrews and Dr. Ivana Barbaric (Centre for Stem Cell Biology, The University of Sheffield). iPSCs were maintained in Matrigel-coated plates (Corning®) according to the manufacturer's recommendations in complete mTeSR™-Plus™ Medium (StemCell Technologies). Cultures were replenished with fresh medium every day. Cells were passaged every 4 to 6 days as clumps using ReLeSR™ (StemCell Technologies) an enzyme-free reagent for dissociation according to the manufacturer's recommendations. For all the experiments in this study, iPSCs were used between passage 20 and 35, all iPSCs were cultured in 5% O<sub>2</sub>, 5% CO<sub>2</sub> at 37 °C. For Motor Neuron differentiation, neural differentiation of iPSCs was performed using the modified version dual SMAD inhibition protocol [33]. Briefly iPSC cells were transferred to Matrigel-coated plates. On the day after plating (day 1), after the cells have reached ~100% confluence, the cells were washed once with PBS and then the medium was replaced with neural medium (50% of KnockOut™ DMEM/F-12 (ThermoFisher Scientific), 50% of Neurobasal (ThermoFisher Scientific), 0.5× N2 supplement (ThermoFisher Scientific), 1× Gibco®

**Table 1** List and characteristics of patient-derived neurons used for the genome-wide investigation of SRSF1 depletion by RNA-seq

| Patient cell lines | Ethnicity | Gender | Cell type   | Age at biopsy sample collection (years) |
|--------------------|-----------|--------|---|---|
| H1 (3050)          | Caucasian | Male   | Healthy control                                   | 68                                      |
| H2 (155)           | Caucasian | Male   | Healthy control                                   | 40                                      |
| H3 (154)           | Caucasian | Female | Healthy control                                   | 55                                      |
| C9-1 (78)          | Caucasian | Male   | C9ORF72-ALS                                       | 66                                      |
| C9-2 (183)         | Caucasian | Male   | C9ORF72-ALS                                       | 49                                      |
| C9-3 (201)         | Caucasian | Female | C9ORF72-ALS                                       | 66                                      |
| CS14               | Caucasian | Female | Healthy control                                   | 30–35                                   |
| MIFF1              | Caucasian | Male   | Healthy control                                   | Unknown                                 |
| ALS-28             | Caucasian | Male   | C9ORF72-ALS                                       | 47                                      |
| ALS-29             | Caucasian | Male   | C9ORF72-ALS                                       | 47                                      |
| ALS-29 ISO         | Caucasian | Male   | Isogenic ALS-29 line lacking the repeat expansion | 47                                      |

GlutaMAX™ Supplement (ThermoFisher Scientific), 0.5x B-27 (ThermoFisher Scientific), 50 U/ml penicillin (Lonza) and 50 mg/ml streptomycin (Lonza), supplemented with SMAD inhibitors (DMH-1 2 μM; SB431542–10 μM and CHIR99021 3 μM [[Tocris]). The medium was changed every day for 6 days. On day 7, the medium was replaced for neural medium supplemented with DMH-1 2 μM, SB431542–10 μM and CHIR 1 μM, All-Trans Retinoic Acid 0.1 μM (RA; StemCell Technologies) and Purmorphamine 0.5 μM (PMN; Tocris). The cells were kept in this medium until day 12 when is possible to see a uniform neuroepithelial sheet. The cells were then split 1:6 with Accutase (Gibco™) onto matrigel substrate in the presence of 10 μM of rock inhibitor (Y-27632 dihydrochloride; Tocris), giving rise to a sheet of neural progenitor cells (NPC). After 24 h of incubation, the medium was changed to neural medium supplemented with RA 0.5 μM and PMN 0.1 μM, and changed every day for 6 more days. On day 19 the motor neuron progenitors were split with accutase onto matrigel-coated plates and the medium was replaced with neural medium supplemented with RA 0.5 μM, PMN 0.1 μM, compound E 0.1 μM (Cpd E; Tocris), BDNF 10 ng/mL (ThermoFisher Scientific), CNTF 10 ng/mL (ThermoFisher Scientific) and IGF 10 ng/mL (ThermoFisher Scientific). The cells were then fed alternate days with neuronal medium until day 40.

#### ***Drosophila* stocks and husbandry**

Flies were raised under standard conditions in a humidified, temperature-controlled incubator with a 12 h: 12 h light:dark cycle at 25 °C, on food consisting of agar, cornmeal, molasses, propionic acid and yeast. Transgene expression was induced using the tissue specific D42-GAL4 or nSyb-GAL4 driver. The following strains were obtained from the Bloomington *Drosophila* Stock Center (RRIDs indicated): UAS-Luciferase RNAi (BDSC\_31603),

D42-GAL4 (BDSC\_8816), nSyb-GAL4 (BDSC\_51635), SK<sup>Mi10378</sup> (BDSC\_55475), SK<sup>Mi09020</sup> (BDSC\_51246). The following strains were obtained from the Vienna *Drosophila* Resource Center: UAS-SF2-RNAi (VDR v27776). UAS-G4C2x3 and UAS-G4C2x36 lines [34] were a gift from Prof Adrian M. Issacs (Department of Neurodegenerative Disease, UCL Institute of Neurology, London). Unless otherwise stated, all experiments were conducted using male flies.

#### **Pharmacological treatments**

In order to evaluate the neuroprotective potential of apamin (Sigma Aldrich, 178,270), a potent antagonist of calcium-activated potassium channels KCNN1 and KCNN3 [35], motor neurons derived from iPSC cells from unaffected controls and C9ORF72-ALS patients were exposed to apamin (0.1–10 μM) diluted in neuronal medium for 72 h. Cells treated with dimethyl sulfoxide (DMSO; Sigma Aldrich), the vehicle for dilution of apamin, were used as a control.

#### **Total, nuclear and cytoplasmic RNAs fractionation from patient-derived iNeurons**

Three wells of a 6-well plate were lysed in Reporter lysis buffer (*Promega*) for 10 min on ice before centrifugation at 17,000 g, 5 min, 4 °C for subsequent western blot and total RNA extractions. Six wells of a 6-well plate were used for nuclear/cytoplasmic fractionations that were performed as described previously [12]. Briefly, cells were lifted from the plates in DEPC PBS, pelleted by centrifugation at 400 g for 5 min and quickly washed with hypotonic lysis buffer (10 mM HEPES pH 7.9, 1.5 mM MgCl<sub>2</sub>, 10 mM KCl, and 0.5 mM DTT). Cells were then lysed in hypotonic lysis buffer containing 0.16 U μl<sup>-1</sup> Ribosafe RNase inhibitors (Bioline), 2 mM PMSF and SIGMAFAST Protease Inhibitor Cocktail tablets, EDTA free (Sigma-Aldrich) for 15 min on ice. All



lysates underwent differential centrifugation (1500 g, 3 min, 4 °C then 3500 g, 8 min, 4 °C and then 17,000 g, 1 min, 4 °C) transferring the supernatants to fresh tubes after each centrifugation. The pellets from centrifugation at 1500 g for 3 min were lysed in Reporter lysis buffer (Promega) containing 0.16 U  $\mu\text{l}^{-1}$  Ribosafe RNase inhibitors, 2 mM PMSF and Protease Inhibitors for 10 min on ice and subsequently centrifuged at 17,000 g, 5 min, 4 °C to obtain the nuclear fractions. RNA was extracted from total, nuclear and cytoplasmic fractions as detailed below. Equal volumes of total, nuclear and cytoplasmic lysates were resolved by SDS-PAGE, electroblotted onto nitrocellulose membrane and probed using mouse anti-SSRP1 [1/500 dilution; Abcam #ab26212] and chicken anti-class III  $\beta$ -tubulin (TUJ1) [1/500; Millipore #AB9354] antibodies. SSRP1 was detected using HRP-conjugated mouse secondary antibody [1/10000; Promega #W4021] and TUJ1 was detected using HRP-conjugated chicken secondary antibody [1/10000; Promega #G1351].

#### RNA extraction

250  $\mu\text{l}$  iNeuron total, nuclear or cytoplasmic extracts were added to 750  $\mu\text{l}$  PureZOL™ (Bio-Rad) to extract the RNA. Total *Drosophila* RNA was extracted from *Drosophila* heads with the specific transgenes driven by D42-GAL4 which had been ground to a powder under liquid nitrogen before addition of 250  $\mu\text{l}$  reporter lysis buffer (Promega) and addition of 750  $\mu\text{l}$  PureZOL™ to the frozen lysates to extract the RNA. Briefly, lysates were cleared by centrifugation for 10 min at 12,000 g at 4 °C. One fifth the volume of chloroform was added and tubes were vigorously shaken for 15 s. After 10 min incubation at room temperature, tubes were centrifuged 12,000 g, 10 min, 4 °C and supernatants collected. RNA was precipitated for 30 min at room temperature with an equal volume of isopropanol and 2  $\mu\text{l}$  glycogen and subsequently pelleted at 12,000 g, 20 min, 4 °C. Pellets were washed with 70% DEPC ethanol and re-suspended in DEPC water. All PureZol™ extracted RNA samples were treated with DNaseI (Sigma Aldrich) and quantified using a Nanodrop (NanoDropTechnologies). Fractionated extracts were subjected to RNA extraction using Direct-zol™ RNA microprep kits (Zymo Research) following the manufacturer's protocol, including the recommended in-column DNase I treatment and quantified using a Nanodrop. RNA quality was then assessed using a eukaryote total RNA Nano 6000 Kit (Agilent Technologies) prior to high depth RNA sequencing or microarray analysis and qRT-PCR.

#### Quantitative RT-PCR (qRT-PCR)

Following quantification, 1  $\mu\text{g}$  RNA (iNeurons samples) or 2  $\mu\text{g}$  RNA (*Drosophila* samples) was converted to

cDNA using BioScript Reverse Transcriptase (Bioline). qRT-PCR primers were designed using Origene or Primer-BLAST and validated using a 1 in 4 serial template dilution series (standard curve with  $R^2 > 0.97$ ). qRT-PCR reactions were performed in duplicate using the Brilliant III Ultra-Fast SYBR Green QPCR Master Mix (Agilent Technologies) on a C1000 Touch™ thermos Cyclers CFX96™ Real-Time System (BioRAD) using an initial denaturation step, 45 cycles of amplification (95 °C for 30 s; 60 °C for 30 s; 72 °C for 1 min) prior to recording melting curves. qRT-PCR data was analysed using CFX Manager™ software (Version 3.1) (BioRAD) and GraphPad Prism (Version 7). The sequences of qPCR primers are provided in Additional file 1.

#### Immunofluorescence microscopy

For immunostaining, cells were washed with phosphate-buffered saline (PBS) and fixed with 4% paraformaldehyde (Sigma Aldrich) for 10 min at room temperature. After fixation samples were washed three times with PBS and permeabilized with 0.3% Triton X-100 diluted in PBS for 5 min. The cells were subsequently blocked in 5% Donkey serum (DS; Millipore) for 1 h. After blocking, cell cultures were incubated with the appropriate primary antibodies (rabbit anti-Nestin, 1:500 [Biolegend #841901]; mouse anti-Pax6, 1:200 [Millipore #MAB5552]; mouse anti-Tuj1, 1:1000 [Biolegend #801201]; mouse anti-NeuN, 1:1000 [Biolegend #SIG-39860]; goat anti-Chat, 1:100 [Millipore #AB144P]; mouse anti-SM132 [Biolegend #801701], 1:500; guinea pig anti-Map2 [Synaptic Systems #188004], 1:1000; rabbit anti-Caspase 3, 1:200 [Millipore #AB3623]) diluted in PBS containing 1% of DS overnight. Cells were then washed with PBS three times. Fluorescent secondary antibodies (Alexa Fluor 488, 555, 594 or 647, diluted 1:400 with DS [ThermoFisher Scientific #A-21202, #A-21432, A-21450, #A-32744, #A-21206]) were subsequently added to the cells and incubated for 1 h. The samples were washed with PBS three more times and incubated with 1.0 mg/mL 4,6-diamidino-2-phenylindole (DAPI; Sigma Aldrich) for nuclear staining. As a specificity control, all experiments included cultures where the primary antibodies were not added. Non-specific staining was not observed in such negative control conditions. RNA foci were visualized using RNA fluorescence in situ hybridization (FISH) as described previously [36]. Images were taken with the Opera Phenix™ High Content Screening System at  $\times 40$  magnification using the Harmony™ Image analysis system. We used 405, 488 and 594 nm and 647 lasers, along with the appropriate excitation and emission filters. These settings were kept consistent while taking images from all cultures.

### High-content automated imaging microscopy

To investigate whether apamin protects motor neurons (MNs) from cell death, MNs were treated with apamin for 72 h (0.1–10  $\mu$ M) and then were stained for active caspase 3, a typical apoptotic marker. MNs were plated at  $2 \times 10^4$  cells per well on matrigel-coated 96-well plates. After treatment, the cells were fixed and stained for active caspase 3 and MAP2, which was used as a marker to define the boundary of cells and DAPI for nuclear staining. A quantitative imaging analysis of the MN was conducted through the The Opera Phenix™ High Content Screening System at  $\times 40$  magnification using the Harmony™ Image analysis system. The following morphological features were assessed for both treated and control: percentage caspase-3 positive cells and the number of fragmented nuclei. At least 25 fields were randomly selected and scanned per well of a 96-well plate in triplicate. To identify and remove any false readings generated by the system, three random treated and untreated wells were selected and counted manually (blinded to groups).

### *Drosophila* locomotor and lifespan assays

The startle induced negative geotaxis (climbing) assay was performed using a counter-current apparatus and D42-GAL4 driver. Briefly, 20–50 flies were placed into the first chamber, tapped to the bottom, and given 10 s to climb a 10 cm distance. This procedure was repeated five times (five chambers), and the number of flies that remained within each chamber counted. The weighted performance of several groups of flies for each genotype was normalized to the maximum possible score and expressed as Climbing index [37]. For larval crawling assays, nSyb-GAL4 was used and wandering third instar larvae were placed in the centre of a 1% agar plate and left to acclimatise for 30 s, after which the number of peristalsis waves that occurred in the following minute were recorded.

### Bioinformatics analysis

#### Next generation RNA sequencing (RNA-seq)

Total RNA samples with RNA integrity numbers (RIN) comprised between 9.3 and 10.0 were sent to the Centre of Genomic Research at the University of Liverpool for RNA-seq (project LIMS14705). Dual-indexed strand-specific RNA-seq libraries were prepared from the submitted total RNA samples using RiboZero rRNA depletion and the NEBNext Ultra Directional RNA library preparation kits (*New England Biolabs*). Paired-end 2x150bp sequencing was performed on an *Illumina HiSeq 4000* platform. RNA-seq reads were quality-checked and trimmed by the Centre of Genomic Research at the University of Liverpool. Fastq files were then aligned to the Human Genome GRCh38 using STAR 2.7 aligner [38]

and the Ensembl built on a University of Sheffield Unix cluster. The RNA-seq data have been deposited in *Gene Expression Omnibus* (GEO) under accession number GSE139900. An average of 104 million 150 bp paired-end sequencing reads were obtained (Supplementary Table 1). Approximately 5–10% of the genome is stably transcribed in human cell lines [39, 40]. The size of the human genome is  $3 \times 10^9$  bp. Therefore according to the Lander/Waterman equation (Coverage = read length  $\times$  number of reads / size of dataset): Transcriptome coverage =  $(2 \times 150 \text{ bp} \times 104 \times 10^6 \text{ reads}) / (3 \times 10^9 \text{ bp} \times 10/100) = 104$  fold.

### Comparison of transcription profiles from fibroblasts and fibroblast-derived cells

Microarray data previously obtained (GEO Accession number GSE87385) from human fibroblasts, induced astrocytes and induced oligodendrocytes were compared to the transcriptome of induced neurons and post-mortem laser captured motor neurons (GEO Accession number GSE29652). After merging all expression data based on Gene ID (Official Gene Symbol), we obtained 10,688 gene transcripts that were annotated on both the microarray platform used to define the transcriptome of fibroblasts, induced astrocytes, oligodendrocytes and post-mortem motor neurons, as well as the RNA-seq data relative to induced human neurons. Expression data of these 10,688 transcripts were normalised across platforms based on expression of housekeeping genes and groups visualised using the *Qlucore* visualisation software. Principal component analysis (PCA) plots were obtained by performing F-test ANOVA multi-group comparison analysis applying  $p$ -value  $< 0.01$ . No fold change parameter was applied.

### RNA-seq analysis: generation of differentially-expressed transcript lists

Over 80% of trimmed RNA-Seq reads provided by the Centre for Genomics Research at the University of Liverpool were aligned to the human GRCh38.79 genome build using STAR2.7 [38]. Supplementary Table S1 provides numbers and proportion of aligned reads. Quantification of transcripts abundance was done using the aligned .bam files with RSEM, a method to accurately quantify transcripts from RNA-Seq data with or without a reference genome [41]. Transcripts counts were then analysed with EdgeR [42] to normalise the data and quantify differential expression. Data normalisation was performed using the relative log expression (RLE) implemented in EdgeR. After normalisation, transcripts abundance was filtered for low and no reads. We retained transcripts for which counts per million (cpm) were greater or equal than 2 in at least two samples. Transcripts sequenced in each group were compared to one

another to evaluate transcriptome coverage across our experimental conditions. For each comparison, we evaluated the biological variation using the maximisation of the negative binomial dispersion using the empirical Bayes likelihood function, as implemented in EdgeR. Differentially-expressed transcript isoforms were computed for fold change  $FC > 2$  and  $p$ -value  $p < 0.05$ , which were evaluated using the quasi-likelihood (QL) methods with empirical-Bayes Test in EdgeR. Differentially-expressed transcripts were annotated using BioMart [43] based on their Ensembl transcript Id to recover gene symbol, gene description and biotype.

#### **RNA-seq analysis: splicing**

For the splicing analysis, sequencing reads were aligned to the GRCh38.79 genome build through STAR two-pass mode (v2.5.4b) [38]. The gtf file was used as a guide during the first pass to find the superset of all novel splice junctions that were then used in the second-pass to improve the consistency of alignment and quantification across these spliced transcripts. The DEXSeq module of Bioconductor was used to identify differential exon usage [44]. We used the python scripts provided by the package to annotate the genome and to count the reads overlapping the exons. The significance thresholds for differential exon usage were set at a Benjamini–Hochberg false discovery rate of 5%.

#### **Microarray analysis**

*Drosophila\_2* gene expression arrays (*Affymetrix*) were used in this study. Total RNA samples were prepared according to the manufacturers' protocol. Briefly, 200 ng of total RNA was converted into cDNA using an oligo(dT) which also carries the binding site for T7 RNA polymerase. Following first strand synthesis, residual RNA was degraded by addition of RNaseH and a double stranded cDNA molecule was generated using DNA Polymerase I and DNA ligase. These cDNA molecules were used as a substrate for the T7 RNA polymerase to produce multiple copies of antisense RNA using an IVT labelling system. The cRNA molecules produced incorporated biotin labelled ribonucleotides, which acted as a target for the subsequent detection of hybridization, using fluorescently labelled streptavidin. 12.5  $\mu$ g of cRNA molecules were heat fragmented and applied to the GeneChips in a hybridization solution according to the Affymetrix protocol. Hybridization took place overnight in a rotating hybridization oven at 60 rpm, 45 °C for 16 h. The GeneChip arrays were washed using the ThermoFisher Fluidics Station. After washing and development of the fluorescent signal, the GeneChip arrays were scanned using GC30007 scanner. Gene level differential expression analysis was carried out using Transcriptome Analysis Console 3.1 (*Affymetrix*). The CEL files were

loaded into the software and sorted into the appropriate groups. Data were normalised using the RMA algorithm and the resultant grouped. CHP files compared for differential analysis. The software used a one way between subject ANOVA (unpaired) to calculate expression level differences with default values fold change  $FC > 2$  and  $p$  value  $p < 0.05$ . The microarray data have been deposited in *Gene Expression Omnibus* (GEO) under accession number GSE138592.

#### **Statistical analysis of data**

For qRT-PCR data, either one-way or two-way ANOVA (analysis of variance) with Tukey's correction for multiple comparisons was used. For *D. melanogaster* climbing assays, a Kruskal–Wallis nonparametric test with Dunn's post-hoc correction for multiple comparisons was used and data reported as mean  $\pm$  95% CI. For *D. melanogaster* crawling assays One-way ANOVA parametric with Bonferroni's multiple comparison test was used and data reported as mean  $\pm$  SEM. Data were plotted using GraphPad Prism 7. Significance is indicated as follows; NS: non-significant,  $P \geq 0.05$ ; \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ ; \*\*\*\* $P < 0.0001$ .

## **Results**

### **Identifying transcriptomes from healthy and C9ORF72-ALS human-derived neurons**

Depleting *SRSF1* mRNA levels by approximately 60% confers neuroprotection through inhibition of the nuclear export and RAN translation of pathological *C9ORF72*-repeat transcripts [12]. To investigate the genome-wide mechanisms by which *SRSF1* depletion confers neuroprotection and evaluate the safety of this manipulation, we identified whole-cell and cytoplasmic transcriptomes to investigate RNA changes at expression, splicing and nuclear export levels. Neurons derived from induced neural progenitor cells reprogrammed from the fibroblasts of three different ALS patients harbouring *C9ORF72*-repeat expansion mutations (*C9ORF72*-ALS) and three healthy controls (Table 1) were treated with lentivirus expressing either scrambled control-RNAi (C-RNAi) or *SRSF1*-RNAi ( $\Delta$ *SRSF1*) prior to nuclear and cytoplasmic fractionation in the same conditions and cell lines previously used to show that partial depletion of *SRSF1* confers neuroprotection [12]. Western blot analysis using antibodies directed against the SSRP1 chromatin remodelling factor confirmed the absence of nuclear contamination in the cytoplasmic fractions (Fig. 1a). Similarly to our previous study, treatment of patient-derived neurons with *SRSF1*-RNAi lentivirus led to partial depletion of *SRSF1* at mRNA (~60%) and protein levels (Fig. 1b-c respectively). We also confirmed that the depletion of *SRSF1* did not affect the splicing of *C9ORF72* intron-1 (Fig. 1d) and specifically



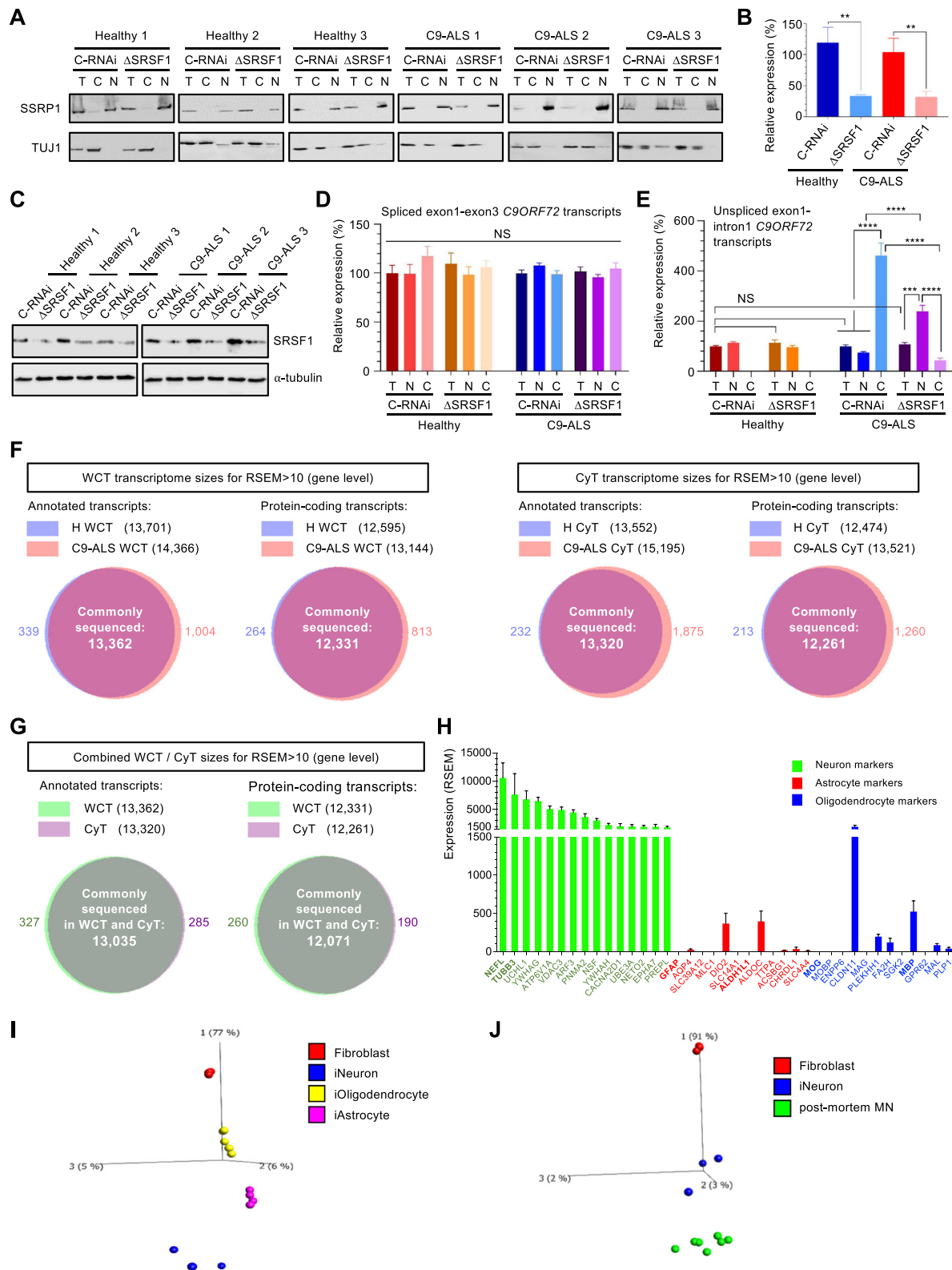


Fig. 1 (See legend on next page.)

(See figure on previous page.)

**Fig. 1** Generation of whole-cell and cytoplasmic transcriptomes from healthy and C9ORF72-ALS patient-derived neurons. **A** Three healthy control and three C9ORF72-ALS (C9-ALS) lines of patient-derived neurons were treated with Ctrl-RNAi (C-RNAi) or SRSF1-RNAi ( $\Delta$ SRSF1) prior to whole-cell (T) lysis or nuclear (N) and cytoplasmic (C) fractionation. Western blots were probed for the nuclear chromatin remodelling SSRP1 factor and the neuronal cytoplasmic marker TUJ1. **B** Relative expression levels of *SRSF1* mRNA in whole-cell patient-derived neurons prepared in A were quantified using qRT-PCR in biological triplicates following normalization to U1 snRNA levels and to 100% for healthy neurons treated with C-RNAi (mean  $\pm$  SEM; one-way ANOVA with Tukey's correction for multiple comparisons, \*\*:  $p < 0.01$ ; N (qRT-PCR reactions) = 3). **C** Western blots analysis of SRSF1 protein expression in the three healthy and three C9-ALS neuron lines treated with either C-RNAi or SRSF1-RNAi. **D** Total, nuclear and cytoplasmic levels of intron1-spliced C9ORF72 transcripts (as measured by the exon1-exon3 junction) were quantified using qRT-PCR in biological triplicates following normalization to U1 snRNA levels and to 100% for whole-cell healthy neurons treated with C-RNAi (mean  $\pm$  SEM; one-way ANOVA with Tukey's correction for multiple comparisons, NS: non-significant; N (qRT-PCR reactions) = 3). **E** Total, nuclear and cytoplasmic levels of unspliced C9ORF72 transcripts retaining intron1 (as measured by the exon1-intron1 junction) were quantified using qRT-PCR in biological triplicates following normalization to U1 snRNA levels and to 100% for whole-cell healthy neurons treated with C-RNAi (mean  $\pm$  SEM; one-way ANOVA with Tukey's correction for multiple comparisons, NS: non-significant; \*\*\*:  $p < 0.001$ ; \*\*\*\*:  $p < 0.0001$ ; N (qRT-PCR reactions) = 3). **F** Whole cell transcriptome (WCT; left) and cytoplasmic transcriptome (CyT, right) sizes for quantified transcripts at gene level. **G** Combined WCT and CyT sizes for quantified transcripts at gene level. **H** Abundance of neuron (green), astrocyte (red) and oligodendrocyte (blue) markers in the human derived neurons (combined from all 24 protein-coding transcriptomes shown in Fig. 1F). **I** Principal component analysis (PCA) plot representing transcriptomes from patient-derived neurons, astrocytes, oligodendrocytes and fibroblasts of origin. **J** Principal component analysis (PCA) plot representing transcriptomes from patient-derived neurons, human post-mortem motor neurons (MNI) and fibroblasts of origin

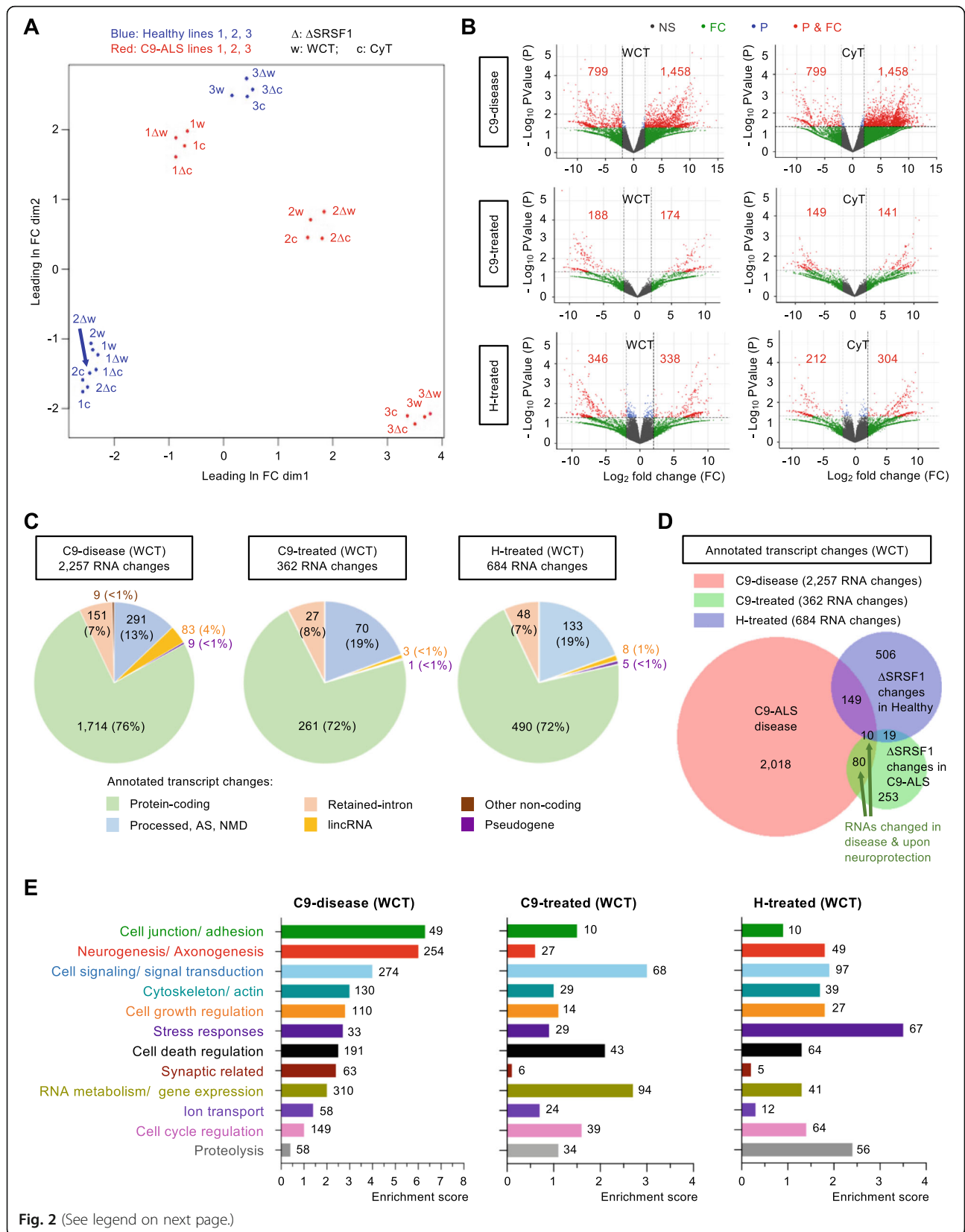
inhibited the nuclear export of *C9ORF72* transcripts retaining the pathological repeat-expansions in intron-1, showing nuclear accumulation and concomitant cytoplasmic decrease of these repeat transcripts (Fig. 1e).

We next proceeded to extract total RNA from the whole-cell and cytoplasmic fractions to analyse genome-wide RNA expression changes using next generation RNA sequencing (RNA-seq). Independent triplicates of rRNA-depleted RNA sequencing libraries were subjected to high depth RNA sequencing (averaging 104 million reads per sample, >100-fold transcriptome coverage; Table S1) to investigate differential expression and splicing with high confidence in unrelated patient samples, which present a high level of genetic variability (Material and Methods). The lists of over 40,000 quantified transcript isoforms are presented in Table S2 under 4 tabs corresponding to each of whole-cell transcriptomes (WCT) and cytoplasmic transcriptomes (CyT) for either healthy control (H) or C9ORF72-ALS (C9) patient-derived neurons treated with C-RNAi or SRSF1-RNAi lentivirus. Over 13,000 annotated transcripts and 12,000 protein-coding mRNAs were commonly sequenced at gene level across all conditions reflecting the generation of datasets with high transcriptome coverage without notable RNA sequencing bias between transcriptomes (Fig. 1f-g). Based on cell type-specific transcriptome databases from human-derived brain cells [45] and mouse brains [46], we identified the top 20-expressed transcripts in our datasets that are present within (i) the top 1000-specific human-derived neurons and 2034-enriched mouse brain neurons; (ii) the top 354-specific human-derived astrocytes and 2616-enriched mouse brain astrocytes and (iii) the top 260-specific human-derived oligodendrocytes and 2227-enriched mouse brain oligodendrocytes (Supplementary Table 3). This

analysis showed that our differentiation protocol successfully yields cells with high expression of neuronal markers (such as *NEFL* and *TUBB3/TUJ1*) and low expression levels of transcripts known to be specifically enriched in astrocytes (such as *GFAP* or *ALDH1L1*) and oligodendrocytes (such as *MOG* or *MBP*) (Fig. 1h). Moreover, a principal component analysis (PCA) comparing induced patient-derived neurons, astrocytes and oligodendrocytes to their parental fibroblasts clearly showed specific segregation of the 4 different cell types away from each other and from the fibroblasts of origin; with glial cells, i.e. astrocytes and oligodendrocytes clustering closer to each other than to neurons (Fig. 1i). Moreover, our protocol of differentiation leads to patient-derived neuron transcriptomes which cluster close to human post-mortem motor neurons on the main component of a PCA plot comparing fibroblasts to motor neurons (Fig. 1j). Overall, high depth RNA-seq transcriptomes have been generated from cells successfully differentiated into patient-derived neurons.

#### Neuroprotective depletion of SRSF1 preserves the transcriptomes of human-derived neurons

A multidimensional scale analysis of the 24 identified transcriptomes (whole-cell and cytoplasmic triplicates from healthy and C9ORF72-ALS neurons treated with control- or SRSF1-RNAi lentivirus) shows that the partial depletion of SRSF1 does not overall disrupt the transcriptomes, with maintenance of the genetic variability between individuals (Fig. 2a). Differentially-expressed transcript isoforms were selected for fold change  $FC > 2$  and  $p$ -value  $p < 0.05$ . Despite a marked reduction in the abundance of sequenced SRSF1 isoforms in the matched pair of neurons treated with control- or SRSF1-RNAi, the differential expression of SRSF1 transcripts, which



(See figure on previous page.)

**Fig. 2** Genome-wide effects of the SRSF1 depletion in human healthy and C9ORF72-ALS neurons. **A** Multidimensional scale analysis of the WCT and CyT transcriptomes for healthy (blue) and C9-ALS (red) lines of patient-derived neurons treated with C-RNAi or SRSF1-RNAi ( $\Delta$ ). **B** Volcano plots representing the genome-wide distribution of differentially-expressed transcripts according to  $p$ -values and fold changes for C9-disease (H\_C-RNAi versus C9\_C-RNAi), C9-treated (C9\_C-RNAi versus C9\_SRSF1-RNAi) and H-treated (H\_C-RNAi versus H\_SRSF1-RNAi) neuron groups. NS: non-significant; FC: fold changes < 2; P:  $p$ -values > 0.05; P & FC: significant  $p$ -values (< 0.05) & fold changes (> 2). Red labels indicate the numbers of significantly down- or up-regulated annotated transcripts. **C, D** Venn diagrams representing differentially-expressed transcripts at WCT level for C9-disease, C9-treated and H-treated groups. **E** Bar charts representing enrichment scores for the 12 top biological processes identified via functional annotation clustering for the C9-disease group (1470 protein coding DEGs). The enrichment scores corresponding to the 12 altered biological pathways are also reported for the C9-treated (254 protein-coding DEGs) and H-treated groups (470 protein-coding DEGs). The GO terms are provided with gene IDs and statistics in Table S6. Numbers at bar extremities indicate the numbers of genes altered in each pathways. The neuroprotective depletion of SRSF1 mitigates most of the disease-altered pathways

was experimentally validated (Fig. 1b), provided  $p$ -values overall not statistically significant due to the high variability of SRSF1 expression between individual subjects (2 to 5 fold; Table S4).

Lists of differentially expressed transcript isoforms (total RNA changes) and differentially expressed genes (DEGs) are provided in Table S5 for annotated and protein-coding transcripts in WCT and CyT transcriptomes. Changes were investigated in: (i) C9ORF72-ALS disease (H\_C-RNAi versus C9\_C-RNAi; tabs 1 & 4); C9ORF72-ALS neurons treated with SRSF1-RNAi (C9\_C-RNAi versus C9\_SRSF1-RNAi; tabs 2 & 5); healthy control neurons treated with SRSF1-RNAi (H\_C-RNAi versus H\_SRSF1-RNAi; tabs 3 & 6). We identified that the total expression levels of 2257 transcripts (corresponding to 1804 DEGs) are altered in C9ORF72-ALS (C9-disease group) by comparing the expression of transcripts quantified in healthy and C9-ALS neurons treated with control RNAi (Fig. 2b-c, Table S5). On the other hand, the neuroprotective effects of the depletion of SRSF1 were investigated in either C9ORF72-ALS or healthy patient-derived neurons by comparing transcripts levels in each cell type treated with control-RNAi and SRSF1-RNAi (C9-treated and H-treated groups respectively). Thus, comparing transcript levels in C9ORF72-ALS neurons treated with either control-RNAi or SRSF1-RNAi lentivirus (C9-treated group) identified that a total of 362 RNA changes (351 DEGs; < 1% of the transcriptome comprising transcripts from ~42,500 coding and non-coding genes) is implicated in the neuroprotection conferred by the partial depletion of SRSF1 (Fig. 2b-c, Table S5). We also characterised that the depletion of SRSF1 leads to 684 transcript changes (642 DEGs; ~1.5% transcriptome) in healthy neurons treated (H-treated group) (Fig. 2b-c, Table S5). The very small proportion of quantified transcript changes induced by the depletion of SRSF1 is in full agreement with the multidimensional scale analysis that did not show global alteration of SRSF1-depleted transcriptomes (Fig. 2a). Consistent with the mRNA splicing and nuclear export functions of SRSF1, 72% of manipulated RNAs in SRSF1-RNAi-treated neurons are protein

coding transcripts (Fig. 2c). However, the SRSF1-RNAi-induced transcript changes in healthy and C9ORF72-ALS neurons do not overlap (Fig. 2d) reinforcing the concept that healthy control and C9-ALS transcriptomes are very diverse at a global level due to widespread alteration of RNA metabolism in the C9ORF72-ALS disease state. Interestingly, the expression levels of approximately a quarter of transcripts altered in disease (80 + 10 or 90 out of 362) are also reciprocally changed upon neuroprotection (Fig. 2d).

#### Manipulation of SRSF1 ameliorates multiple dysregulated pathways in C9ORF72-ALS neurons

A gene ontology (GO) analysis was performed with the protein-coding lists of WCT DEGs provided in Supplementary Table 5 using DAVID 6.8 [47, 48]. The results are presented in Table S6 under 3 tabs for the investigation of biological processes modulated in C9-disease, C9-treated and H-treated groups. Transcripts altered in C9-disease encode proteins involved in cell junction/ adhesion, neurogenesis/ axonogenesis, cytoskeleton, cell signaling, regulation of cell growth and cell death, stress responses, synaptic signaling, RNA metabolism/ gene expression, ion transport and cell cycle regulation (Fig. 2e, C9-disease). Interestingly, despite the fact that the depletion of SRSF1 in healthy control and C9ORF72-ALS neurons led to different transcript changes (Fig. 2d), the same pathways are manipulated upon SRSF1 depletion (Fig. 2e, C9-treated versus H-treated). In agreement with the proliferative and oncogenic functions of SRSF1 [29, 30], the depletion of SRSF1 in healthy neurons leads to down-regulation of markers of the G1/S cell cycle transition and mitosis (Supplementary Figure 1A) as well as of cell proliferation and cancer markers [49, 50] (Supplementary Figure 1B). Interestingly, reducing the expression level of SRSF1 also promotes expression of transcripts involved in neuron differentiation, axonogenesis and synaptic transmission (Supplementary Figure 1C), suggesting additional roles of SRSF1 that may provide neuroprotective benefits beyond inhibiting the nuclear export of C9ORF72 repeat transcripts and the production of DPRs.



As shown above and reported in the literature, a broad range of cellular processes are affected in C9ORF72-ALS. Remarkably, the neuroprotection conferred by the depletion of SRSF1 appears to act upon the vast majority of biological pathways dysregulated in disease with the exception of synaptic-related signaling which is poorly manipulated and protein degradation which is up-regulated upon neuroprotection (Fig. 2e, C9-disease versus C9-treated). Out of 2257 RNA changes which occur in 1804 genes in C9ORF72-ALS, manipulating the expression levels of 362 transcripts only (~16%, including 261 protein-coding genes) is sufficient to confer neuroprotection in vitro (Fig. 2c) suggesting that the large majority of RNA alterations are not directly related to pathogenesis but are likely to be downstream consequences of the neurodegenerative process. Importantly, this shows that neuroprotection can be achieved without mitigating most of the disease-altered transcript changes. While the precise gene expression changes causing neurodegeneration still remain to be determined, the identification of 261 neuroprotective mRNAs offers new promising perspectives to understand disease pathophysiology and identify novel potential therapeutic targets for C9ORF72-ALS.

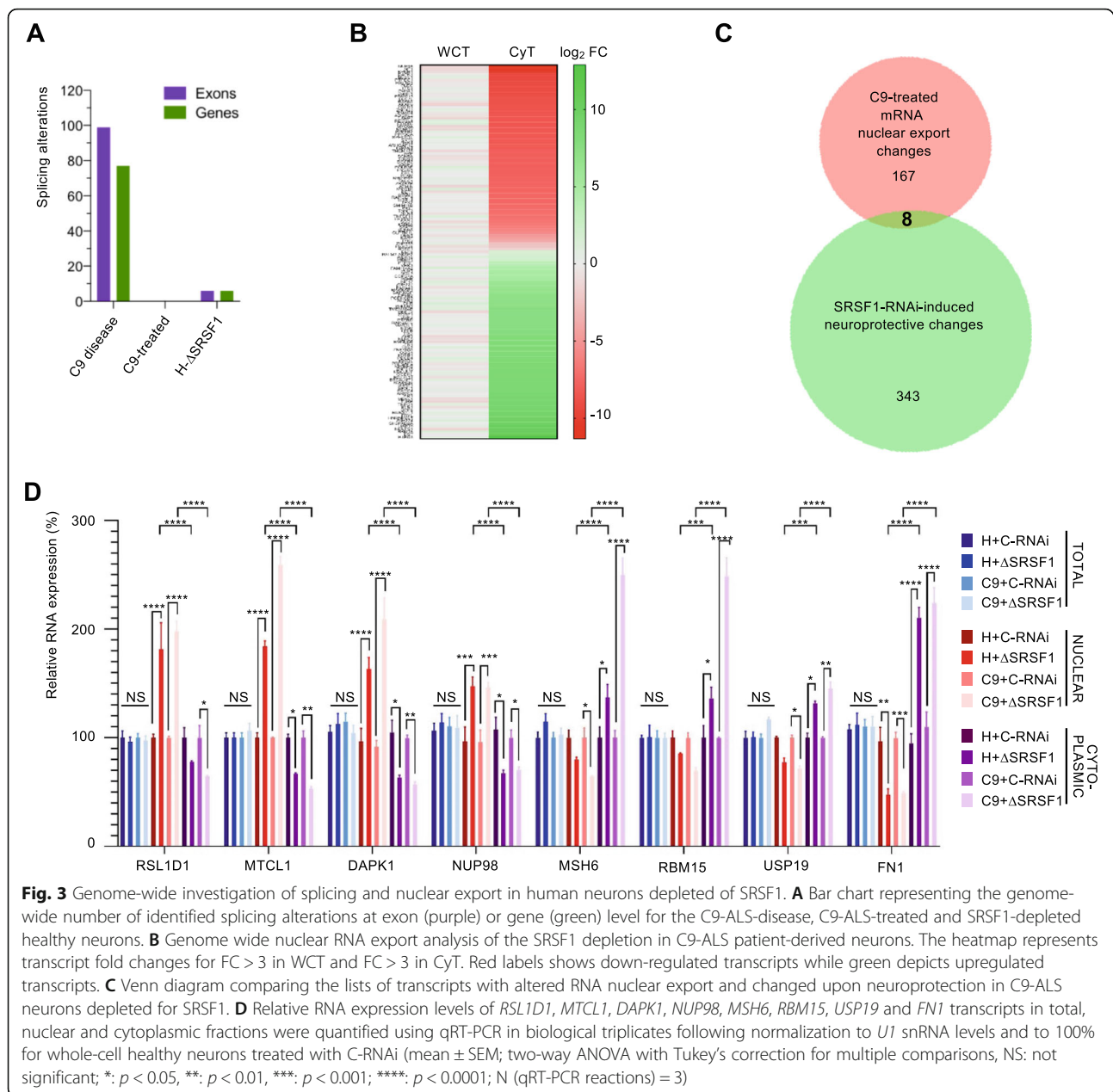
#### **Depletion of SRSF1 confers neuroprotection independently of genome-wide modulation of splicing and mRNA nuclear export**

The genome wide effects of the neuroprotective depletion of SRSF1 were next investigated at splicing and mRNA nuclear export levels. For the splicing analysis, exon reads were extracted from bam files and differences in exon usage were computed for C9-disease, C9-treated and H-treated neurons (Methods). This methodology allows the detection of differential splicing in an exon-centric manner, through the analysis of differential exon usage between the conditions under study, directly related to the diversity of genes and, hence, to alternative splicing. This analysis is independent of the knowledge of the exact transcript isoform(s) that can often generate misleading results since one exon can be shared between several assembled transcripts [51–53]. Table S7 reports all statistically differentially expressed exons identified at a false discovery rate of 5% under 6 tabs for either WCT or CyT: C9-disease (tabs 1 & 4), C9-treated (tabs 2 & 5) and H-treated (tabs 3 & 6). Table S8 provides a summary of the changes. Cytoplasmic exon usage changes are less noisy than the whole-cell samples that contain processing pre-mRNAs. Altered transcript isoforms that have been exported into the cytoplasm are also more likely to have functional consequences at the protein level. Ninety-nine differentially expressed cytoplasmic exons were identified in 77 genes in C9-disease, while no splicing changes were detected in neuroprotected

C9ORF72-ALS neurons and only 6 were found in SRSF1-depleted healthy control neurons (Fig. 3a). In particular, no changes were detected in the splicing of the *C9ORF72* transcripts in either C9ORF72-ALS or healthy control neurons in agreement with qRT-PCR assays in Fig. 1d. These data indicate that the partial depletion of SRSF1, although neuroprotective, has no significant effect on the genome-wide splicing of transcripts.

We also investigated the potential impact of partial depletion of SRSF1 on the genome-wide nuclear export of transcripts. Transcript abundance with FC >3 in the cytoplasmic transcriptomes were intersected with transcripts not significantly changing in the whole cell transcriptomes (FC <3) in a similar approach used for measuring the mRNA nuclear export dependence of SRSF1–7 proteins in a murine cell line [26]. This analysis allowed identifying transcripts with altered cytoplasmic expression but unchanged steady-state levels e.g. transcripts with specific nuclear export defects. The nuclear export of 177 annotated transcripts, which include 137 mRNAs, is altered in C9ORF72-ALS neurons treated with SRSF1-RNAi representing 0.4% of the transcriptome while 202 mRNAs have altered nuclear export in healthy neurons depleted of SRSF1 (Table S9). Approximately half of the transcripts either showed nuclear export inhibition or stimulation (Fig. 3b; extended heat map with transcript IDs in Supplementary Figure 2), in full agreement with a previous study which identified nuclear export alterations of 225 transcripts upon depletion of over 90% of SRSF1 in a proliferating cell context [26]. The absence of significant overlap between the 177 transcripts with altered RNA nuclear export and the 362 transcript changes implicated in the SRSF1-RNAi-induced neuroprotection (Fig. 3c) suggests that neuroprotection is conferred independently of the nuclear export modulation of cellular transcripts, but rather through inhibition of the nuclear export of pathological *C9ORF72*-repeat transcripts which cannot be detected by RNA-seq due to the pure and repeat GC-rich nature of the expansions. It was however quantified and confirmed by qRT-PCR in Fig. 1e. In a therapeutic view, it is noteworthy that partial depletion of SRSF1 does not alter the CaMKII $\delta$  transcript essential to the developing heart [31] at expression, splicing or nuclear export level.

To experimentally validate the outputs of the bioinformatics analysis, we investigated some of the 4 most down-regulated and 4 most up-regulated transcripts identified in the SRSF1-dependent nuclear export list (Table S9; tab C9-treated ann. transcripts) which represents a compilation of predicted data from both whole-cell and cytoplasmic transcriptomes. We used qRT-PCR to quantify the mRNA expression levels in total, nuclear and cytoplasmic fractions isolated from healthy and



C9ORF72-ALS patient-derived neurons treated with either control-RNAi or SRSF1-RNAi lentivirus. As predicted from the bioinformatic investigation, we showed that the expression levels of *RSL1D1*, *MTCL1*, *DAPK1*, *NUP98* and *MSH6*, *RBM15*, *USP19*, *FN1* are indeed respectively decreased and increased in the cytoplasm while the total levels remain unchanged and the nuclear expression levels are altered in the opposite direction (Fig. 3d). Interestingly, fold changes for transcripts with the highest altered nuclear export were no more than 2.5 fold, indicating that the genome-wide effects of the SRSF1 depletion on nuclear export is limited (both in

the numbers and FC of affected transcripts) reminiscent of the study which showed that the family of SRSF1–7 proteins play redundant and/or cooperative roles in the NXF1-dependent nuclear export adaptor function [26].

#### Manipulation of SRSF1 modulates multiple dysregulated pathways in C9ORF72-ALS *Drosophila*

Partial depletion of SRSF1 suppressed the C9ORF72-repeat neurodegeneration-associated locomotor deficits of G4C2x36 *Drosophila* [34] through the transgenic expression of SRSF1-specific RNAi sequences [12]. Here, we investigated the transcriptomes of *Drosophila* heads

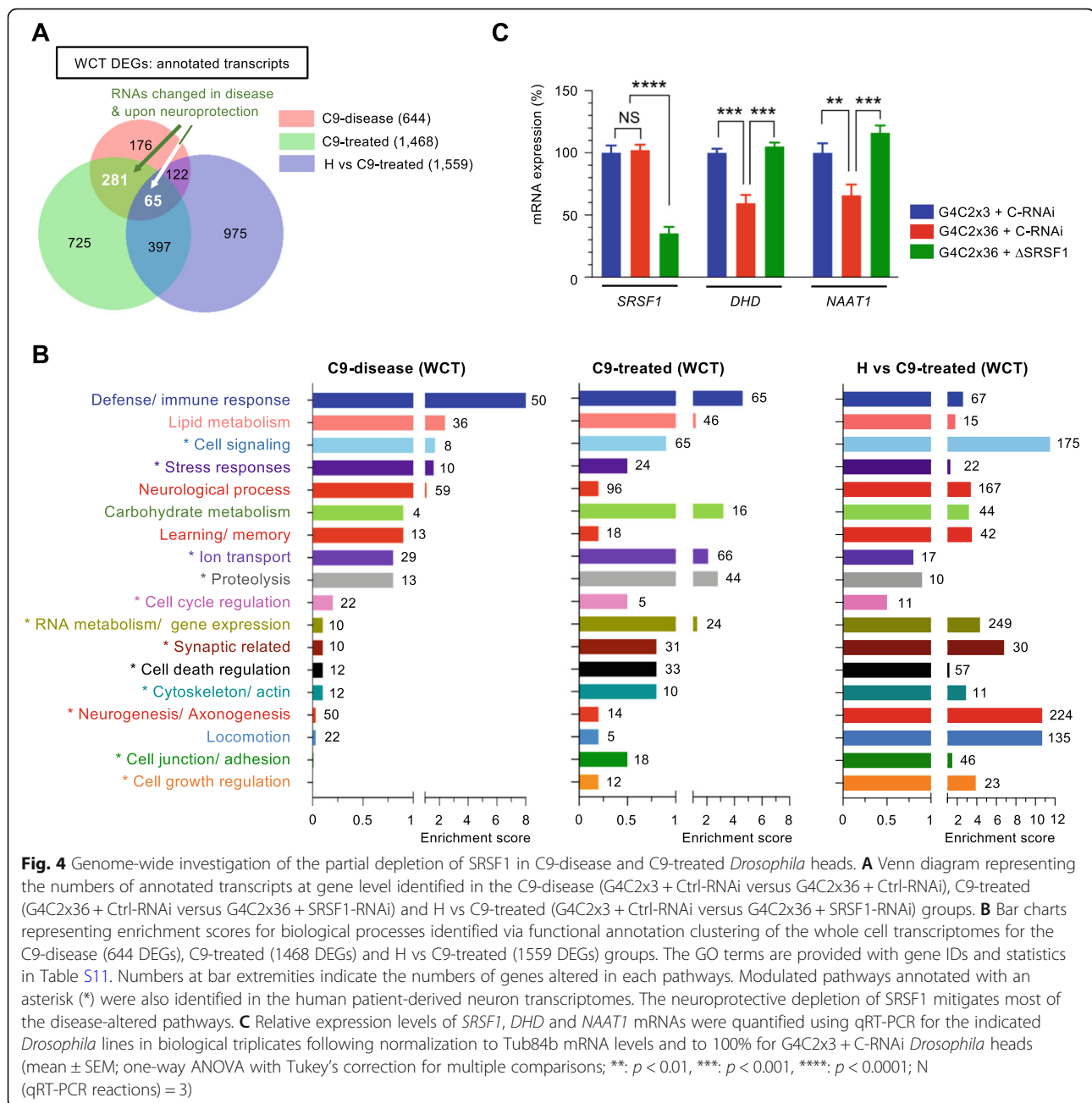
from the same lines driving G4C2x36 expression by D42-GAL4: (i) healthy control flies expressing 3 G4C2 repeats and a luciferase-RNAi control (G4C2x3\_C-RNAi); (ii) C9ORF72-ALS model expressing 36 G4C2 repeats and the RNAi control (G4C2x36\_C-RNAi); (iii) C9ORF72-ALS-neuroprotected flies expressing 36 G4C2 repeats and the disrupted SRSF1 allele (G4C2x36\_SRSF1-RNAi).

Lists of differentially expressed transcripts were identified for  $FC > 2$  and  $p < 0.05$  (Methods). Table S10 reports changes in: (i) C9-disease (tab 1; G4C2x3\_C-RNAi versus G4C2x36\_C-RNAi), (ii) C9-treated (tab 2; G4C2x36\_C-RNAi versus G4C2x36\_SRSF1-RNAi), (iii) H vs C9-treated (tab 3; G4C2x3\_C-RNAi versus G4C2x36\_SRSF1-RNAi). Six hundred forty-four DEGs were identified in C9-disease flies while expression of 1468 and 1559 genes was respectively changed in the C9-treated and H vs C9-treated groups. Venn diagrams show that SRSF1-RNAi-induced neuroprotection is achieved without normalizing all of the transcripts that are altered in disease and that expression levels of 346 transcripts are altered both in disease and upon neuroprotection (Fig. 4a). The DAVID ontology analysis is available for all groups in Table S11 (tabs 1, 2, 3). Transcripts altered in the C9ORF72-ALS *Drosophila* heads include biological processes previously identified in patient-derived neurons (such as cell signalling, stress responses, and ion transport) but also changes in defence/ immune response, lipid/ carbohydrate metabolisms, neurological process and learning/ memory (Fig. 4b, C9-disease). On the other hand, the neuroprotective effects conferred by the depletion of SRSF1 act again remarkably upon the vast majority of pathways dysregulated in disease (Fig. 4b, C9-treated). Interestingly, other cellular processes that were found altered in both C9-disease and C9-treated patient-derived neurons (such as protein degradation, gene expression, synaptic-related, cell death regulation, cytoskeleton and cell junction/adhesion) are also enriched in neuroprotected *Drosophila* heads (Fig. 4b, C9-treated). This indicates that the inhibition of the RAN translation of C9ORF72-repeat transcripts mediated by the depletion of SRSF1 confers neuroprotection by inducing changes in conserved cellular pathways. This is amplified in the H vs C9-treated group which highlights the manipulated pathways that have been modified in C9-treated flies compared to the healthy animals and which confer neuroprotection in C9-ALS flies (Fig. 4b). Highly enriched changes include transcripts involved in cell signaling, synaptic-related processes, neurogenesis/ axonogenesis and locomotion. Overall, the modulation of these biological processes appear particularly relevant to a *Drosophila* model of C9ORF72-ALS and the mitigation of the neurodegeneration associated locomotor deficits. Using qRT-PCR assays, we further

experimentally validated that thioredoxin (*DHD*), the most downregulated transcript in C9ORF72-ALS *Drosophila*, and the sodium-dependent nutrient amino acid transporter *NAAT1* in the ion transport pathway, exhibit reduced mRNA expression levels in disease while they are upregulated to normal levels upon SRSF1 depletion (Fig. 4c) in agreement with the computed DEG lists (Table S10, C9-disease and C9-treated).

#### Identifying conserved transcripts reciprocally altered in disease and upon neuroprotection

The depletion of SRSF1 provides a unique opportunity to investigate manipulated disease-modifying mechanisms through the identification of gene expression changes that occur in both C9ORF72-ALS and upon neuroprotection. The expression levels of only 90 transcripts affected in disease (~4% out of 2257 RNA changes) are also altered following SRSF1 depletion (Fig. 5a). A clustered heat map represents how transcripts are modulated in diseased and neuroprotected neurons for each individual subject (Supplementary Figure 3). It is noteworthy that, despite genetic variability between individuals, the depletion of SRSF1 reverses altered expression levels for almost all disease-associated transcripts across the cases. This analysis was further validated by investigating the fold change values calculated for each of the 90 transcripts which have modified expression levels in the C9-disease and C9-treated groups (Fig. 5b, enlarged heatmap in Supplementary Figure 4, transcript IDs and FC values in Table S12, tab 1). Remarkably, the altered expression levels of 88 transcripts, which include 68 mRNAs, are completely reversed upon SRSF1 depletion. We next sought to perform the same analysis using the *Drosophila* transcriptomes and investigated the 346 transcript changes that are both affected in disease and manipulated upon SRSF1-RNAi-dependent neuroprotection (Fig. 5c) showing again, as for the human-derived neuron data, that the transcripts which are downregulated or upregulated in C9-disease are largely completely reversed in the neuroprotected C9-treated flies (Fig. 5d, transcript IDs and FC values in Table S12, tab 2). The identification of these disease-modifying gene expression signatures with transcripts involved in different biological functions that show reversed expression levels upon neuroprotection in both in vitro and in vivo models is entirely consistent with our previous conclusions indicating that the depletion of SRSF1 counteracts multiple disease-altered biological processes (Figs. 2e and 4b). This reflects the neuroprotective effects of SRSF1 manipulation at genome-wide level through inhibition of the nuclear export of pathological C9ORF72-repeat transcripts and mitigation of the DPR-associated neurotoxicity which are known to broadly alter gene expression and multiple



cellular pathways [54–59]. Changes in RNA expression levels however do not necessarily mirror protein levels and further separate studies will be required to confirm the altered expression of genes identified here at the functional level.

We next seek to integrate the data obtained in our patient-derived neuron and *Drosophila* transcriptomes by identifying potential RNA expression changes conserved in disease and upon SRSF1-RNAi-induced neuroprotection. To this purpose, we used BioMart [43] which allows identifying orthologous genes and therefore

conserved gene changes in the human and fly disease-modifying gene expression signatures. Forty differentially-expressed human genes out of 90 C9-disease/treated transcript changes have 49 fly homologues while 99 differentially-expressed fly genes out of 346 C9-disease/treated transcript changes have 78 human homologues (Table S12, tabs 3 and 4). Only 2 orthologous gene changes are commonly predicted to be up-regulated in the human and fly disease groups while their expression levels are down-regulated upon neuroprotection in both the human and fly C9-treated groups:



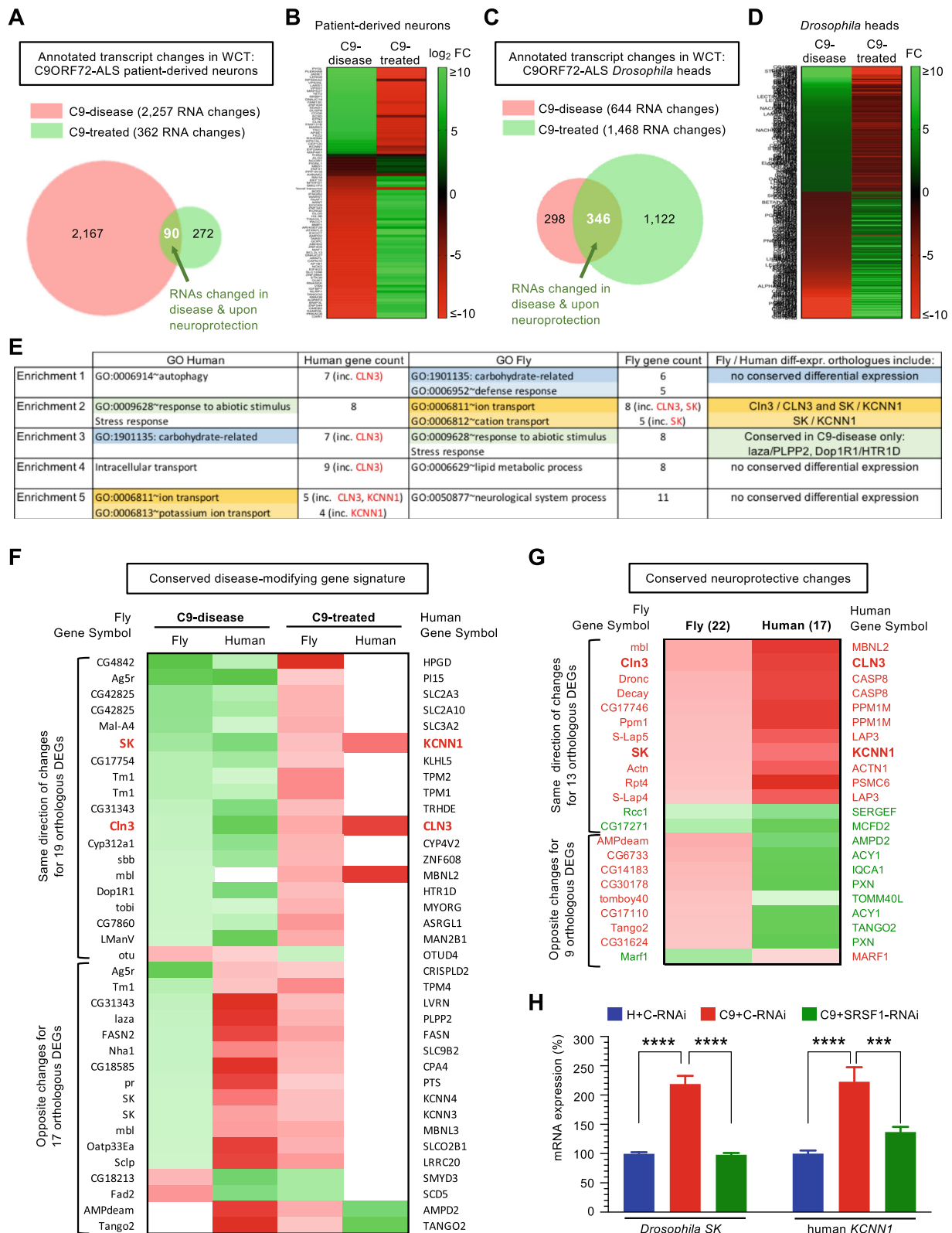


Fig. 5 (See legend on next page.)

(See figure on previous page.)

**Fig. 5** Identifying conserved disease-modifying gene expression signatures and neuroprotective transcripts. **A** Venn diagram representing differentially expressed genes (DEGs) identified in the human C9-disease and C9-treated groups. **B** Identification of a human disease-modifying gene expression signature. Heatmap representing the computed fold changes (FC) for the common human transcripts modulated in both C9-disease and C9-treated groups. Red labels show down-regulated transcripts while green depicts upregulated transcripts. **C** Venn diagram representing differentially expressed genes (DEGs) identified in the *Drosophila* C9-disease and C9-treated groups. **D** Identification of a *Drosophila* disease-modifying gene expression signature. Heatmap representing the FC for the common *Drosophila* transcripts modulated in both C9-disease and C9-treated groups. Red labels show down-regulated transcripts while green depicts upregulated transcripts. **E** Gene ontology analysis of conserved differentially-expressed transcripts identified in the human and *Drosophila* disease-modifying gene expression signatures carbohydrate metabolism, ion transport and response to stress as commonly altered in disease and manipulated upon neuroprotection. **F** Heatmap representing the computed fold changes for the orthologous genes in the disease-modifying signatures. Red labels show down-regulated transcripts while green depicts upregulated transcripts. **G** Heatmap representing the computed fold neuroprotective changes for the orthologous genes in the C9-treated groups. Red labels show down-regulated transcripts while green depicts upregulated transcripts. **H** qRT-PCR quantification of *Drosophila SK* and human *KCNN1* orthologous transcripts. Relative expression levels of *SK* mRNA was quantified for the indicated *Drosophila* lines in biological triplicates following normalization to Tub84b mRNA levels and to 100% for G4C2x3 + C-RNAi *Drosophila* heads (mean  $\pm$  SEM; one-way ANOVA with Tukey's correction for multiple comparisons; \*\*\*\*:  $p < 0.0001$ ; N (qRT-PCR reactions) = 3). Relative expression levels of *KCNN1* mRNA was quantified in whole-cell patient-derived neurons in biological triplicates following normalization to U1 snRNA levels and to 100% for healthy neurons treated with C-RNAi (mean  $\pm$  SEM; one-way ANOVA with Tukey's correction for multiple comparisons, \*\*\*:  $p < 0.001$ , \*\*\*\*:  $p < 0.0001$ ; N (qRT-PCR reactions) = 3)

human small conductance calcium-activated potassium channel protein 1 (*KCNN1*)/ fly small conductance calcium-activated potassium channel protein (*SK*) and the human and fly lysosomal/endosomal transmembrane proteins *CLN3/Cln3*. Gene ontology analysis was further performed on the 40 and 99 differentially-expressed and conserved genes respectively identified in the in vitro and in vivo disease-modifying gene expression signatures (Table S12, tabs 5 and 6). Top enriched categories in both human and fly include response to abiotic stimulus/ stress response, carbohydrate metabolism and cation/potassium transport (Fig. 5e) indicating that if only 2 orthologous gene changes are commonly differentially-expressed, neuroprotection is achieved through manipulation of different genes which are involved in the same biological processes. Interestingly, the fully conserved C9-disease/treated changes in *CLN3/Cln3* and *KCNN1/SK* expression levels are both involved in the ion transport pathway. A heatmap of conserved genes showing at least one differentially-expressed change in the C9-disease or the C9-treated groups highlights 19 similar and 17 opposite direction of orthologous changes (Fig. 5f; Table S12, tab 7). Most orthologous changes are identified in the C9-disease groups indicating that neuroprotection largely involves manipulation of different genes in C9ORF72-ALS patient-derived neurons and *Drosophila*.

To further investigate mechanisms of pathogenesis and neuroprotection, we identified the conserved gene changes in the C9-disease and C9-treated transcriptomes. Eight hundred twelve human differentially-expressed genes out of 1804 DEGs found in C9-disease (Supplementary Table 5) have 1251 fly homologues and 65 differentially-expressed fly orthologues (Table S13, tab 1). Two hundred one differentially-expressed fly genes out of 644 DEGs (Supplementary Table 10) have

143 human homologues and 47 differentially-expressed human orthologues (Table S13, tab 2). A heatmap of the conserved changes identified in C9-disease is presented in Supplementary Figure 5, showing that only a small proportion of disease-altered gene expression changes have differentially-expressed orthologues. In addition to the challenge posed by integrating transcriptomes from human-derived neurons and *Drosophila* heads which contain multiple cell types, this might also be explained by the fact that only 17.8 and 24.4% of genes have annotated orthologues in the *Homo sapiens* and *Drosophila melanogaster Ensembl* genomes respectively (Supplementary Figure 6A-B). Likewise, 188 human differentially-expressed genes out of 351 neuroprotective gene changes have 261 fly homologues and 22 differentially-expressed fly orthologues (Table S13, tab 3) while 510 differentially-expressed fly genes out of 1468 neuroprotective DEGs have 953 human homologues and 17 differentially-expressed human orthologues (Table S13, tab 4). Only 17 human and 22 fly orthologous genes are involved in conferring the SRSF1-RNAi dependent neuroprotection with 13 having similar changes of direction (Fig. 5g). Conserved manipulation of these transcripts suggests that manipulation of other cellular pathways including various metabolic enzymes and apoptosis likely play important neuroprotective roles beyond modulation of *KCNN1/SK*, *CLN3/Cln3* and the ion transport pathway.

Alterations of voltage-gated ion channels have been reported in C9ORF72-ALS [60–63] and previous reports including patents for pharmacological modulation of SK channels in ALS indicate a potential role of small conductance Ca<sup>2+</sup>-activated potassium (SK) channels in ALS. We therefore experimentally validated our RNA-seq data using qRT-PCR assays to show that *KCNN1/SK* transcripts were indeed up-regulated in the C9-disease

groups and down-regulated to normal expression levels upon SRSF1-RNAi-induced neuroprotection in the C9-treated groups (Fig. 5h).

#### Manipulating SK/KCCN ion channel activity alleviates C9ORF72-ALS motor neuron death and *Drosophila* locomotor deficits

We next wanted to functionally validate that our gene expression signature is enriched in genes with disease-modifying potential by testing the effects of manipulating the conserved *Drosophila* SK and human SK channel subunits in C9ORF72-ALS patient-derived neurons and *Drosophila*.

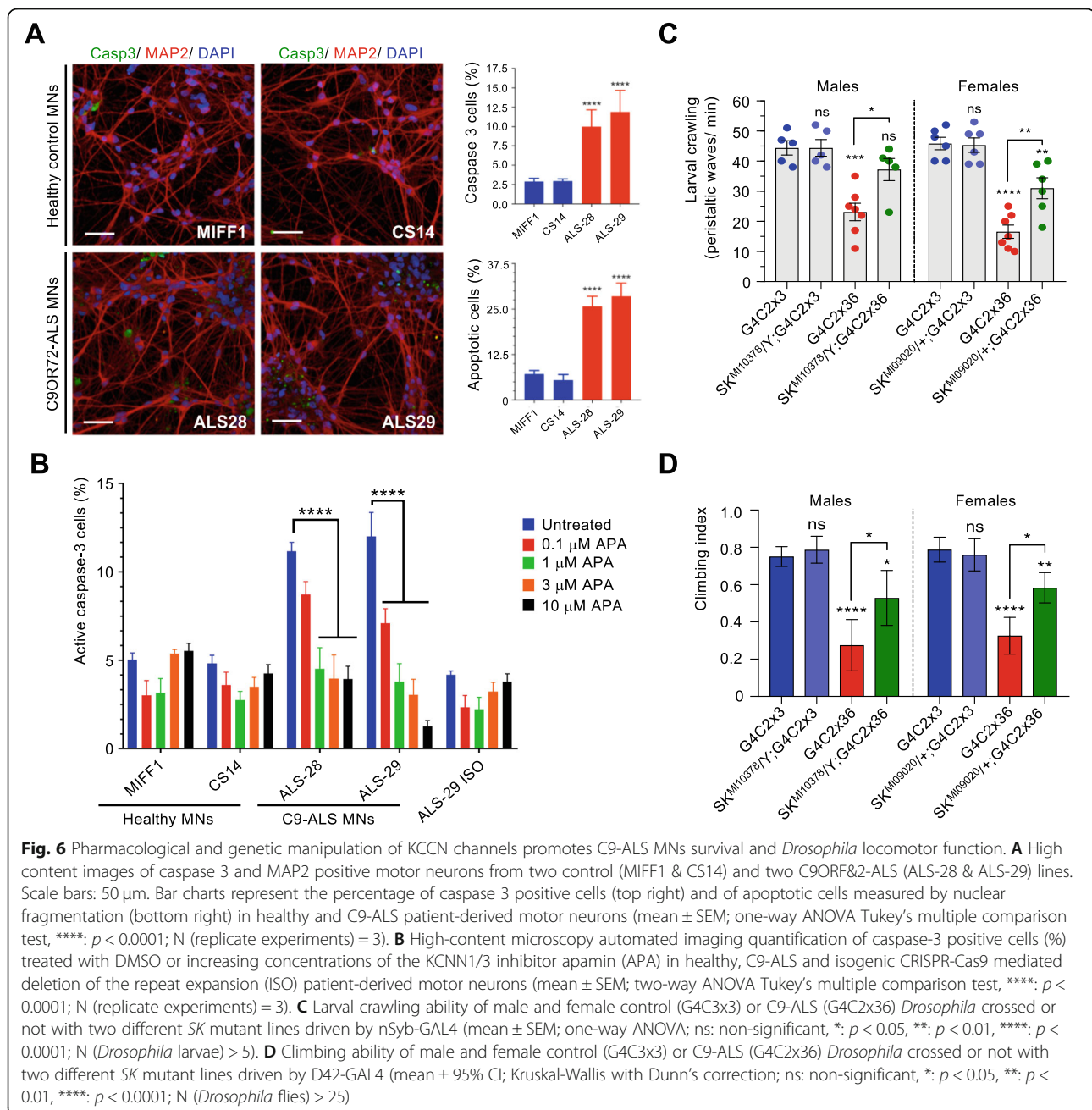
Four KCNN1–4 SK channel isoforms are found in the human genome. Using qRT-PCR quantification, we observed that *KCCN1* and *KCNN3* transcripts are significantly up-regulated in C9ORF72-ALS models while they are down-regulated upon neuroprotection (Supplementary Figure 7). On the other hand, *KCNN4* is not affected by the depletion of SRSF1, but is drastically reduced at mRNA level in C9ORF72-ALS patient-derived neurons. We next sought to functionally investigate the effects of inhibiting the up-regulated KCNN1 (encoding Kca2.1) and KCNN3 (encoding Kca2.3) transcripts in C9ORF72-ALS motor neurons. Motor neurons were generated from 2 lines each of healthy control (MIFF1, CS14) and C9ORF72-ALS patients (ALS-28, ALS-29; Table 1). Motor neurons were differentiated from neural progenitor cells, which express the characteristic Nestin and Pax6 markers (Supplementary Figure 8), through a 40-day differentiation protocol that leads to mature motor neurons expressing ChAT (Supplementary Figure 9). We further validated that C9ORF72-ALS motor neurons recapitulate RNA foci, a characteristic pathological hallmark of disease (Supplementary Figure 10). Interestingly, compared to healthy controls, C9ORF72-ALS motor neurons displayed significantly higher levels of caspase-3 positive cells, a marker of apoptosis (Fig. 6a, top right bar chart), which correlated with a significant increase in nuclear fragmentation and apoptosis (Fig. 6a, bottom right bar chart). This offers a unique opportunity to test the effects of the pharmacological inhibition of KCCN1/3 channels on C9ORF72-ALS motor neuron survival. The addition of increasing concentrations of apamin, an antagonist of KCNN1 and KCNN3 channels [35], to the motor neuron cultures resulted in a dose-dependent reduction of caspase-3 positive cells specific to C9ORF72-ALS motor neurons (Fig. 6b). These data show that, consistent with the RNA-seq investigation, C9ORF72-ALS motor neurons have an enhanced expression of KCNN subunits and that inhibiting the activity of these Ca<sup>2+</sup>-activated potassium channels, as in the SRSF1-RNAi intervention, decreases apoptosis and cell death, promoting in turn neuronal survival.

To further evaluate the neuroprotective potential of inhibiting the SK channel function in C9-*Drosophila*, Ctrl and C9 flies were crossed with two different loss-of-function SK mutant fly lines (Methods) and locomotor ability was analysed in hemizygous (male) or heterozygous (female) SK mutants. Strikingly, partial loss-of-function of SK channel activity significantly restored both the crawling activity of C9 larvae (Fig. 6c) and the climbing ability of C9 adult flies (Fig. 6d). Taken together, beyond highlighting the involvement of SK channels in C9ORF72-ALS, these data highlight the powerful approach of our study by targeting SRSF1-mediated transcriptomic changes to delineate the most C9ORF72-disease relevant mechanisms and neuroprotective strategies.

#### Discussion

Multiple studies have reported thousands of transcriptome changes in C9ORF72-ALS neurons from cell and animal models as well as from post-mortem human brains, raising challenges for the identification of altered expression of transcripts that cause disease. Here, we combined for the first time in vitro and in vivo transcriptome investigations of C9ORF72-ALS human-derived neurons and *Drosophila* with a disease-modifying strategy of neuroprotection which leads to specific inhibition of the SRSF1-dependent nuclear export of pathological C9ORF72 repeat transcripts [12]. In perfect agreement with published transcriptome studies, we identified over 2000 transcript changes in human C9ORF72-ALS affecting cellular pathways involved in neuronal-related processes, signalling, RNA metabolism, cell junction/ adhesion, cytoskeleton, cell death regulation and responses to stress [64–66]. Dysregulation in the expression of genes involved in synaptic-related processes, neuron differentiation and membrane hypoexcitability were also reported in C9ORF72-ALS human-derived neurons [64] while immune response and RNA processing alterations were also identified in a human post-mortem ALS brain study which investigated genome-wide expression changes using samples stratified by disease severity [67].

Strikingly, we discovered that neuroprotection is conferred by manipulating the expression of a small proportion of C9ORF72-ALS-altered transcripts (362 transcripts genome-wide) in addition to inhibiting the nuclear export of pathological C9ORF72-repeat transcripts and DPR-associated neurotoxicity, known to cause widespread alteration of gene expression through disruption of the nucleolus [57], splicing [54, 55] and nucleocytoplasmic transport of proteins [56, 59] although potentially indirectly [58]. This also implies that, out of thousands of RNA changes occurring in C9ORF72-ALS, the vast majority of diseased gene



expression changes occur secondary to the neurodegenerative process and a complete reversal is not required to achieve a neuroprotective effect. In both C9ORF72-ALS human-derived neurons and G4C2x36 *Drosophila*, the expression of approximately one third of the SRSF1-RNAi-manipulated transcripts altered in disease is completely reversed upon neuroprotective depletion of SRSF1, revealing in turn in vitro and in vivo C9ORF72-ALS disease-modifying small gene expression signatures. Identifying these out of thousands of RNA changes occurring in C9ORF72-ALS provides a novel exciting basis

to evaluate disease-modifying treatments, discover potential new biomarkers and importantly pinpoint a small number of new targets with therapeutic potential.

To demonstrate the power identification of our disease-modifying transcript approach and emphasizing that our data integration leads to the identification of neuroprotective changes, we further showed that decreasing the activity of conserved SK potassium channels, which are upregulated in disease, mitigates the death of human-derived motor neurons as well as locomotor deficits in *Drosophila*. Our genome-wide



investigation indicated that the nuclear export of *KCNN1-3* transcripts is not dependent on SRSF1 (Table S9) however the disease-altered overexpression of *KCNN1-3/SK* mRNAs is down-regulated to physiological levels under SRSF1-RNAi-induced neuroprotection in both C9ORF72-ALS patient-derived neurons and *Drosophila* (Figs. 5f-h, S7). This interestingly suggests that the SRSF1-dependent inhibition of the nuclear export of *C9ORF72*-repeat transcripts and RAN translation of DPRs, which cause broad dysregulation of the RNA metabolism, restores the normal expression of *KCNN1-3/SK* and other disease-altered transcripts which exhibit reversed expression upon neuroprotection.

Taken together, this investigation provides further validation that the partial depletion of SRSF1 is a promising gene therapy approach for neuroprotection, showing minimal effects of SRSF1 depletion on the splicing and nuclear export of cellular transcripts in either healthy or C9ORF72-ALS neurons at a global genome-wide level. Our results are in perfect agreement with an independent study showing redundant/ cooperative roles of SRSF1-7 in the NXF1-dependent nuclear mRNA export adaptor function [26]. Consistent with this, the bulk nuclear export of human mRNAs is driven by the Transcription-Export (TREX) complex [68–70], independently of SRSF1 which is not associated to TREX [71], as simultaneous depletion of two TREX subunits (including the general mRNA nuclear export adaptor Alyref) leads to a drastic reduction in the genome-wide recruitment of NXF1 onto polyadenylated mRNAs [72].

Finally, our data demonstrate genome-wide efficacy of the neuroprotective depletion of SRSF1 with a remarkable mitigation of multiple unrelated cellular pathways altered in C9ORF72-ALS. In conclusion, the mechanisms of neuroprotection conferred by the partial depletion of SRSF1 involve both changes in the expression of a small number of neuroprotective transcripts encoding proteins involved in various functions (ion transport, RNA metabolism, synaptic transmission, etc.) as well as the inhibition of the nuclear export of *C9ORF72* repeat transcripts and the subsequent translation of dipeptide repeat proteins that also trigger widespread gene expression alterations and neurodegeneration [54–59, 73]. Additional neuroprotective benefits provided by the depletion of SRSF1, which leads to the stimulation of transcripts encoding proteins involved in neuron differentiation, axonogenesis and synaptic transmission (Figure S1C), may explain why the partial suppression of one fly *ALYREF* gene, which was initially reported as a suppressor of *C9ORF72*-repeat mediated neurotoxicity in a *Drosophila* loss-of-function screen [74] and also identified as a binding partner of *C9ORF72*-repeat RNAs [36, 75], provide poor mitigating effects on the locomotor deficits of C9ORF72-ALS *Drosophila* [12].

Interestingly, altered cytoplasmic distribution of SRSF1 has been reported within the infarcted area of stroke victims when compared to the contralateral area, implicating a role for the cytoplasmic localisation of SRSF1 in tissue repair after ischaemia [76]. Several cytoplasmic related functions have also been reported for SRSF1, including nonsense-mediated decay (NMD) through the RS domain [27], mRNA degradation of specific targets [77], autoregulation of its mRNA through binding to its 3'UTR to prevent translation initiation [78], targeting of mRNAs to stress granules following stress [79, 80] and translation regulation [81, 82]. It remains unknown whether this potential role in neuronal tissue repair is due to an impairment in the nuclear functions of SRSF1 or an enhancement of its cytoplasmic roles. Interestingly, our study which pinpoints neuroprotective benefits following partial depletion of SRSF1 would suggest that the nuclear reduction of SRSF1 may be the factor contributing to neuronal tissue repair after ischaemia through stimulation of the expression of transcripts involved in neuron differentiation, axonogenesis and synaptic transmission. Further investigation is now required to validate the potential safety and efficacy of the partial depletion of SRSF1 in the brain and spinal cords of pre-clinical mammalian models of C9ORF72-ALS/FTD.

## Conclusions

Thousands of gene expression changes, involving altered expression and splicing of transcripts, are typically identified in neurodegenerative diseases such as ALS. Here, we show for the first time that modulating 16% of human disease-altered transcripts only (362 out of 2257 total pathological changes) is sufficient to confer the SRSF1-RNAi dependent neuroprotection previously identified as a promising novel gene therapy approach in C9ORF72-ALS patient-derived neurons and *Drosophila* [12]. Importantly, we found that the partial depletion of SRSF1 does not significantly alter transcriptomes (<1%) preserving the intrinsic variability of gene expression across individuals. It further leads to the neuroprotective manipulation of key transcripts involved in the vast majority of C9ORF72-ALS-altered biological processes without significantly affecting genome-wide splicing (none) or mRNA nuclear export (0.4% transcriptome with modulations <2.5 fold). Overall, our genome-wide investigation provide a solid rationale for the efficacy and safety of the partial depletion of SRSF1 in vitro and in vivo in preclinical patient-derived neurons and *Drosophila* models of C9ORF72-ALS. Integrating data between RNA-seq and microarrays as well as between diseased/neuroprotected human-derived neurons and *Drosophila* heads allowed identification of a disease-modifying signature with few but conserved neuroprotective targets with high therapeutic potential.

## Abbreviations

ALS: Amyotrophic lateral sclerosis; ANOVA: Analysis of variance; C9ORF72: Chromosome 9 open reading frame 72; DEGs: Differentially-expressed transcripts at gene levels; KCNN: Human small conductance calcium-activated potassium channel protein; MN: Motor neuron; PCA: Principal component analysis; SK: *Drosophila* small conductance calcium-activated potassium channel protein; SEM: Standard error of the mean; SRSF1: Serine/arginine-rich splicing factor 1

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13024-021-00475-y>.

**Additional file 1.** : List of qPCR primers used in this study.

**Additional file 2 : Table S1.** RNA-seq statistics.

**Additional file 3 : Table S2.** Annotated quantified transcripts for RSEM > 10.

**Additional file 4 : Table S3.** Common cell marker counts.

**Additional file 5 : Table S4.** SRSF1 transcript counts.

**Additional file 6 : Table S5.** Differentially-expressed transcripts in patient-derived neurons.

**Additional file 7 : Table S6.** Gene ontology analysis in patient-derived neurons.

**Additional file 8 : Supplementary Figure 1.** Downregulation of mitosis/ cell proliferation/ cancer makers and upregulation of neuronal-related functions in healthy neurons with partial depletion of SRSF1. (A) Heatmap representing the computed fold changes for the healthy-treated neurons group. Red labels down-regulated transcripts while green depicts upregulated transcripts. (B) Heatmap representing the computed fold changes for the healthy-treated neurons group. Red labels down-regulated transcripts while green depicts upregulated transcripts. (C) Heatmap representing the computed fold changes for the healthy-treated neurons group. Red labels down-regulated transcripts while green depicts upregulated transcripts.

**Additional file 9 : Table S7.** Splicing analysis in patient-derived neurons.

**Additional file 10 : Table S8.** Exon usage changes in patient-derived neurons.

**Additional file 11 : Table S9.** mRNA nuclear export analysis in patient-derived neurons.

**Additional file 12 : Supplementary Figure 2.** Enlargement of Fig. 3b. mRNA nuclear export analysis upon SRSF1 depletion at mRNA and FC levels.

**Additional file 13 : Table S10.** Differentially-expressed transcripts in *Drosophila*

**Additional file 14 : Table S11.** Gene ontology analysis in *Drosophila*

**Additional file 15 : Supplementary Figure 3.** Heatmap representing a clustered analysis of the C9-ALS disease-modifying transcript signature in patient-derived neurons. H: healthy lines 1, 2, 3; C9: C9ORF72-ALS lines 1, 2, 3; Δ: SRSF1 depletion.

**Additional file 16 : Supplementary Figure 4.** Enlargement of Fig. 5b. C9-ALS disease-modifying signature at transcript ID and FC levels.

**Additional file 17 : Table S12.** C9ORF72-ALS disease-modifying gene expression signature.

**Additional file 18 : Table S13.** Conserved human-fly gene expression changes.

**Additional file 19 : Supplementary Figure 5.** Orthologous gene changes in the human and fly C9-disease groups. 48 orthologues show conserved direction of differential expression at gene level (DEGs) while 33 others exhibit opposite direction of changes. Green and red labels respectively correspond to up-regulation and down-regulation of the expression levels of the corresponding transcripts.

**Additional file 20 : Supplementary Figure 6.** Orthologous genes in the human and fly genomes. BioMart (<http://www.ensembl.org/biomart/martview>) Ensembl Genes 99 was used with the *Homo sapiens* GRCh38.p13 and the *Drosophila melanogaster* BDGP6.28 genomes. A multispecies comparison was performed on 25.04.2020 based on (i) Homologue filters, Orthologous genes only or (ii) Homologue filters, Orthologous human Genes only. (A) 3681 human coding genes out of 20,449 have conserved fly orthologous genes. (B) 3411 fly coding genes of 13,947 have conserved human orthologous genes. List of orthologous genes are available upon request.

**Additional file 21 : Supplementary Figure 7.** Quantification of KCNN1-4 mRNA expression levels in healthy and C9ORF72-ALS patient-derived neurons treated with control or SRSF1-RNAi lentivirus. Relative expression levels of *KCNN1*, *KCNN2*, *KCNN3* and *KCNN4* mRNAs were quantified using qRT-PCR in biological triplicates following normalization to *U1* snRNA levels and to 100% for healthy neurons treated with Ctrl-RNAi (mean ± SEM; two-way ANOVA with Tukey's correction for multiple comparisons; NS: not significant; \*\*\*:  $p < 0.001$ , \*\*\*\*:  $p < 0.0001$ ; N (qRT-PCR reactions)=3).

**Additional file 22 : Supplementary Figure 8.** Characterisation of neural precursors cells (NPCs) derived from induced pluripotent stem cells (iPSCs) from healthy control and C9ORF72-ALS patients. (A) NPCs display positive staining for progenitor markers Pax6 and Nestin. Scale bars: 50 μm. (B) Quantification of cells positive for Nestin or Pax6 ( $n=3$ ).

**Additional file 23 : Supplementary Figure 9.** Differentiation of iPSCs from healthy and C9ORF72-ALS patients into functional motor neurons. (A) Schematic representation of the motor neuron differentiation protocol. (B) Representative images of motor neurons stained for the neuronal markers Tuj1, SMI-32, NeuN, MAP2 and the motor neuron specific marker ChAT. Scale bars: 50 μm.

**Additional file 24 : Supplementary Figure 10.** Motor neurons derived from C9ORF72-ALS patients develop characteristic sense and antisense RNA foci. Scale bars: 50 μm.

## Acknowledgements

We acknowledge patients and their families for participating in our research and donating biological samples to the NIHR Sheffield Biomedical Research Centre: Translational Neuroscience for chronic neurological disorders (IS-BRC-1215-20017). Micro-array analysis was undertaken by the Genomics Core Facility at the Sheffield Institute for Translational Neuroscience at the University of Sheffield. Data generation and processing of raw RNA sequencing reads were carried out by the Centre for Genomic Research which is based at the University of Liverpool (project LIMS14705).

## Authors' contributions

GMH and MM designed the transcriptomics studies. GMH, LF and AJW designed the experimental assays. PJS provided genetically characterised human biosamples. Computational analysis of the RNA-seq data was performed at gene expression level by LC and MM and at splicing level by IG and MRG. The *Drosophila* microarrays were run by PRH. LMC, CDSS, YHL and MAM performed the patient-derived neuron studies. ASM performed the *Drosophila* studies. GMH wrote the manuscript. LMC, PJS, AJW, LF and MM edited the manuscript. All authors approved the manuscript.

## Funding

This work was initiated with the Medical Research Council (MRC) grant MR/M010864/1 (KN, GMH, PJS) and the MND Association grant Hautbergue/ Apr16/846–791 (GMH, LF, AJW, PJS, LMC). This research was further supported by the MRC New Investigator research grant MR/R024162/1 (GMH) and the Biotechnology and Biological Sciences Research Council (BBSRC) grant BB/S005277/1 (GMH). LC was supported by H2020-EU Marie Curie fellowship CONTESSA (ID: 660388). CDSS is funded by an AstraZeneca Post-Doctoral award. LF was funded by the Thierry Latran Foundation (FLA AP2016/ Astrocyte secretome) and is currently supported by the MND Association grant Apr16/848–791 and the Academy of Medical Sciences Springboard Award. AJW was supported by MRC core funding (MC\_UU\_00015/6) and ERC Starting grant (DYNAMITO; 309742). GMH also reports grants Apr17/854–791 from the MND Association, Thierry Latran FTLAAP2016/ Astrocyte secretome and Royal Society International Exchanges grant IEC/R317010

during the course of this study. MA acknowledge grants from Alzheimer's Research UK (ARUK-PG2018B-005), European Research Council (ERC Advanced Award 294745) and MRC DPFS (129016). PJS is supported as an NIHR Senior Investigator Investigator (NF-SI-0617-10077) and by the MND Association (AMBRoSIA 972-797) and MRC grant MR/S004920/1.

#### Availability of data and materials

The RNA-seq and microarray data have respectively been deposited in *Gene Expression Omnibus* (GEO) under accession number GSE139900 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE139900>) and GSE138592 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE138592>). The R code for data analysis is available on request. All other data associated with this study are presented in the main text or Supplementary Materials.

#### Declarations

##### Ethics approval and consent to participate

Informed consent was obtained from all patients before collection of fibroblasts under study STH16573 and Research Ethics Committee reference 12/YH/0330 (Prof Dame Pamela Shaw, University of Sheffield, UK). Induced pluripotent stem cell (iPSC) lines were obtained from Cedars-Sinai (USA), a nonprofit academic healthcare organization. Patient-derived cell lines are described in Table 1.

##### Consent for publication

Not applicable.

##### Competing interests

GMH, MA, AJW and PJS have patents granted in the USA (US10/801027) and in Europe (EP3430143) for the use of inhibitors of SRSF1 to treat neurodegenerative disorders (WO2017207979A1). The authors declare no other relationships, conditions or circumstances that present a potential conflict of interest.

##### Author details

<sup>1</sup>Sheffield Institute for Translational Neuroscience (SITraN), Department of Neuroscience, University of Sheffield, 385 Glossop Road, Sheffield S10 2HQ, UK. <sup>2</sup>School of Mathematics, University of Leeds, Leeds LS2 9JT, UK. <sup>3</sup>MRC Mitochondrial Biology Unit, University of Cambridge, Cambridge Biomedical Campus, Hills Road, Cambridge CB2 0XY, UK. <sup>4</sup>National Research Council of Italy, High Performance Computing and Networking Institute (ICAR-CNR), 111 Via Pietro Castellino, 80131 Naples, Italy. <sup>5</sup>Neuroscience Institute, University of Sheffield, Western Bank, Sheffield S10 2TN, UK. <sup>6</sup>Department of Biomedical Science, University of Sheffield, Western Bank, Sheffield S10 2TN, UK. <sup>7</sup>Present Address: AstraZeneca, Academy House, 136 Hills Road, Cambridge CB2 8PA, UK.

Received: 25 March 2021 Accepted: 16 July 2021

Published online: 10 August 2021

#### References

- DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, et al. Expanded GGGGCC Hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron*. 2011;72(2):245–56. <https://doi.org/10.1016/j.neuron.2011.09.011>.
- Renton AE, Majounie E, Waite A, Simón-Sánchez J, Rollinson S, Gibbs JR, et al. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron*. 2011;72(2):257–68. <https://doi.org/10.1016/j.neuron.2011.09.010>.
- Brown RH, Al-Chalabi A. Amyotrophic lateral sclerosis. *N Engl J Med*. 2017;377(2):162–72. <https://doi.org/10.1056/NEJMra1603471>.
- van Es MDMA, MD POH, MD PAC, MD PAA-C, PhD PRJP, MD PJHV, et al. Amyotrophic lateral sclerosis. *Lancet*. 2017;390(10107):2084–98. [https://doi.org/10.1016/S0140-6736\(17\)31287-4](https://doi.org/10.1016/S0140-6736(17)31287-4).
- Olszewska DA, Lonergan R, Fallon EM, Lynch T. Genetics of frontotemporal dementia. *Curr Neurol Neurosci Rep*. 2016;16(12):107. <https://doi.org/10.1007/s11910-016-0707-9>.
- Ferraiuolo L, Kirby J, Grierson AJ, Sendtner M, Shaw PJ. Molecular pathways of motor neuron injury in amyotrophic lateral sclerosis. *Nat Rev Neurol*. 2011;7(11):616–30. <https://doi.org/10.1038/nrneuro.2011.152>.
- Cooper-Knock J, Kirby J, Ferraiuolo L, Heath PR, Rattray M, Shaw PJ. Gene expression profiling in human neurodegenerative disease. *Nat Rev Neurol*. 2012;8(9):518–30. <https://doi.org/10.1038/nrneuro.2012.156>.
- Cooper-Knock J, Bury JJ, Heath PR, Wyles M, Higginbottom A, Gelsthorpe C, et al. C9ORF72 GGGGCC expanded repeats produce splicing dysregulation which correlates with disease severity in amyotrophic lateral sclerosis. *PLoS One*. 2015;10(5):e0127376. <https://doi.org/10.1371/journal.pone.0127376>.
- Prudencio M, Belzil VV, Batra R, Ross CA, Gendron TF, Prgent LJ, et al. Distinct brain transcriptome profiles in C9orf72-associated and sporadic ALS. *Nat Neurosci*. 2015;18(8):1175–82. <https://doi.org/10.1038/nn.4065>.
- Gao F-B, Almeida S, Lopez-Gonzalez R. Dysregulated molecular pathways in amyotrophic lateral sclerosis-frontotemporal dementia spectrum disorder. *EMBO J*. 2017;36(20):2931–50. <https://doi.org/10.15252/embj.201797568>.
- Balendra R, Isaacs AM. C9orf72-mediated ALS and FTD: multiple pathways to disease. *Nat Rev Neurol*. 2018;14(9):544–58. <https://doi.org/10.1038/s41582-018-0047-2>.
- Hautbergue GM, Castelli LM, Ferraiuolo L, Sanchez-Martinez A, Cooper-Knock J, Higginbottom A, et al. SRSF1-dependent nuclear export inhibition of C9ORF72 repeat transcripts prevents neurodegeneration and associated motor deficits. *Nat Commun*. 2017;8(1):16063. <https://doi.org/10.1038/ncomms16063>.
- Huang YQ, Gattoni R, Stevenin J, Steitz JA. SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. *Mol Cell*. 2003;11(3):837–43. [https://doi.org/10.1016/S1097-2765\(03\)00089-3](https://doi.org/10.1016/S1097-2765(03)00089-3).
- Huang Y, Yario TA, Steitz JA. A molecular link between SR protein dephosphorylation and mRNA export. *Proc Natl Acad Sci U S A*. 2004;101(26):9666–70. <https://doi.org/10.1073/pnas.0403533101>.
- Tintaru AM, Hautbergue GM, Hounslow AM, Hung M-L, Lian L-Y, Craven CJ, et al. Structural and functional analysis of RNA and TAP binding to SF2/ASF. *EMBO Rep*. 2007;8(8):756–62. <https://doi.org/10.1038/sj.embor.7401031>.
- Hautbergue GM, Hung M-L, Golovanov AP, Lian L-Y, Wilson SA. Mutually exclusive interactions drive handover of mRNA from export adaptors to TAP. *Proc Natl Acad Sci U S A*. 2008;105(13):5154–9. <https://doi.org/10.1073/pnas.0709167105>.
- Hough LE, Dutta K, Sparks S, Temel DB, Kamal A, Tetenbaum-Novatt J, et al. The molecular mechanism of nuclear transport revealed by atomic-scale measurements. *Elife*. 2015;4. <https://doi.org/10.7554/eLife.10027>.
- Milles S, Mercadante D, Aramburu IV, Jensen MR, Banterle N, Koehler C, et al. Plasticity of an ultrafast interaction between nucleoporins and nuclear transport receptors. *Cell*. 2015;163(3):734–45. <https://doi.org/10.1016/j.cell.2015.09.047>.
- Walsh MJ, Hautbergue GM, Wilson SA. Structure and function of mRNA export adaptors. *Biochem Soc Trans*. 2010;38(1):232–6. <https://doi.org/10.1042/BST0380232>.
- Heath CG, Viphakone N, Wilson SA. The role of TREX in gene expression and disease. *Biochem J*. 2016;473(19):2911–35. <https://doi.org/10.1042/BCJ20160010>.
- Hautbergue GM. RNA nuclear export: from neurological disorders to Cancer. *Adv Exp Med Biol*. 2017;1007:89–109. [https://doi.org/10.1007/978-3-319-60733-7\\_6](https://doi.org/10.1007/978-3-319-60733-7_6).
- Castelli LM, Lin Y-H, Ferraiuolo L, Sanchez-Martinez A, Ning K, Azzouz M, et al. SRSF1-dependent nuclear export of C9ORF72 repeat-transcripts: targeting toxic gain-of-functions induced by protein sequestration. *Ther Targets Neurol Dis*. 2018;4:e1619.
- Caceres JF, Krainer AR. Functional analysis of pre-mRNA splicing factor SF2/ASF structural domains. *EMBO J*. 1993;12(12):4715–26. <https://doi.org/10.1002/j.1460-2075.1993.tb06160.x>.
- Tacke R, Manley JL. Determinants of SR protein specificity. *Curr Opin Cell Biol*. 1999;11(3):358–62. [https://doi.org/10.1016/S0955-0674\(99\)80050-7](https://doi.org/10.1016/S0955-0674(99)80050-7).
- Long JC, Cáceres JF. The SR protein family of splicing factors: master regulators of gene expression. *Biochem J*. 2009;417(1):15–27. <https://doi.org/10.1042/BJ20081501>.
- Müller-McNicoll M, Botti V, de Jesus Domingues AM, Brandl H, Schwich OD, Steiner MC, et al. SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. *Genes Dev*. 2016;30(5):553–66. <https://doi.org/10.1101/gad.276477.115>.
- Zhang Z, Krainer AR. Involvement of SR proteins in mRNA surveillance. *Mol Cell*. 2004;16(4):597–607. <https://doi.org/10.1016/j.molcel.2004.10.031>.
- Ji X, Zhou Y, Pandit S, Huang J, Li H, Lin CY, et al. SR proteins collaborate with 75K and promoter-associated nascent RNA to release paused

- polymerase. *Cell*. 2013;153(4):855–68. <https://doi.org/10.1016/j.cell.2013.04.028>.
29. Karni R, de Stanchina E, Lowe SW, Sinha R, Mu D, Krainer AR. The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat Struct Mol Biol*. 2007;14(3):185–93. <https://doi.org/10.1038/nsmb1209>.
  30. Anczuków O, Akerman M, Cléry A, Wu J, Shen C, Shirole NH, et al. SRSF1-regulated alternative splicing in breast Cancer. *Mol Cell*. 2015;60(1):105–17. <https://doi.org/10.1016/j.molcel.2015.09.005>.
  31. Xu X, Yang D, Ding J-H, Wang W, Chu P-H, Dalton ND, et al. ASF/SF2-regulated CaMKII $\delta$  alternative splicing temporally reprograms excitation-contraction coupling in cardiac muscle. *Cell*. 2005;120(1):59–72. <https://doi.org/10.1016/j.cell.2004.11.036>.
  32. Desmarais JA, Unger C, Damjanov I, Meuth M, Andrews P. Apoptosis and failure of checkpoint kinase 1 activation in human induced pluripotent stem cells under replication stress. *Stem Cell Res Ther*. 2016;7(1):17. <https://doi.org/10.1186/s13287-016-0279-2>.
  33. Du Z-W, Chen H, Liu H, Lu J, Qian K, Huang C-L, et al. Generation and expansion of highly pure motor neuron progenitors from human pluripotent stem cells. *Nat Commun*. 2015;6(1):6626–9. <https://doi.org/10.1038/ncomms7626>.
  34. Mizielińska S, Gronke S, Niccoli T, Ridler CE, Clayton EL, Devoy A, et al. C9orf72 repeat expansions cause neurodegeneration in *Drosophila* through arginine-rich proteins. *Science*. 2014;345(6201):1192–4. <https://doi.org/10.1126/science.1256800>.
  35. Aldrich R, Chandy KG, Grissmer S, Gutman GA, Kaczmarek LK, Wei AD, et al. Calcium- and sodium-activated potassium channels (version 2019.4) in the IUPHAR/BPS Guide to Pharmacology Database. GtoPdb CITE. 2019;2019:1–18.
  36. Cooper-Knock J, Walsh MJ, Higginbottom A, Robin Highley J, Dickman MJ, Edbauer D, et al. Sequestration of multiple RNA recognition motif-containing proteins by C9orf72 repeat expansions. *Brain*. 2014;137(7):2040–51. <https://doi.org/10.1093/brain/awu120>.
  37. Greene JC, Whitworth AJ, Kuo I, Andrews LA, Feany MB, Pallanck LJ. Mitochondrial pathology and apoptotic muscle degeneration in *Drosophila* parkin mutants. *Proc Natl Acad Sci U S A*. 2003;100(7):4078–83. <https://doi.org/10.1073/pnas.0737556100>.
  38. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21. <https://doi.org/10.1093/bioinformatics/bts635>.
  39. Cheng J, Kapranov P, Drenkow J, Dike S, Brubaker S, Patel S, et al. Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science*. 2005;308(5725):1149–54. <https://doi.org/10.1126/science.1108625>.
  40. Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science*. 2007;316(5830):1484–8. <https://doi.org/10.1126/science.1138341>.
  41. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011;12(1):323. <https://doi.org/10.1186/1471-2105-12-323>.
  42. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2009;26:139–40.
  43. Kasprzyk A. BioMart: driving a paradigm change in biological data management. *Database*. 2011;2011:bar049. <https://doi.org/10.1093/database/bar049>.
  44. Anders S, Reyes A, Huber W. Detecting differential usage of exons from RNA-seq data. *Genome Res*. 2012;22(10):2008–17. <https://doi.org/10.1101/gr.133744.111>.
  45. McKenzie AT, Wang M, Hauberg ME, Fullard JF, Kozlenkov A, Keenan A, et al. Brain cell type specific gene expression and co-expression network architectures. *Sci Rep*. 2018;8(1):8868. <https://doi.org/10.1038/s41598-018-27293-5>.
  46. Cahoy JD, Emery B, Kaushal A, Foo LC, Zamanian JL, Christopherson KS, et al. A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J Neurosci*. 2008;28(1):264–78. <https://doi.org/10.1523/JNEUROSCI.4178-07.2008>.
  47. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44–57. <https://doi.org/10.1038/nprot.2008.211>.
  48. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2009;37(1):1–13. <https://doi.org/10.1093/nar/gkn923>.
  49. Whitfield ML, George LK, Grant GD, Perou CM. Common markers of proliferation. *Nat Rev Cancer*. 2006;6(2):99–106. <https://doi.org/10.1038/nrc1802>.
  50. Waldman YY, Geiger T, Ruppin E. A genome-wide systematic analysis reveals different and predictive proliferation expression signatures of cancerous vs non-cancerous cells. *PLoS Genet*. 2013;9:e1003806.
  51. Han S, Miller JE, Byun S, Kim D, Risacher SL, Saykin AJ, et al. Identification of exon skipping events associated with Alzheimer's disease in the human hippocampus. *BMC Med Genet*. 2019;12(S1):13. <https://doi.org/10.1186/s12920-018-0453-8>.
  52. Björklund SS, Panda A, Kumar S, Seiler M, Robinson D, Gheeya J, et al. Widespread alternative exon usage in clinically distinct subtypes of invasive ductal carcinoma. *Sci Rep*. 2017;7(1):5568. <https://doi.org/10.1038/s41598-017-05537-0>.
  53. Feng H, Li T, Zhang X. Characterization of kinase gene expression and splicing profile in prostate cancer with RNA-Seq data. *BMC Genomics*. 2018;19(S6):564. <https://doi.org/10.1186/s12864-018-4925-1>.
  54. Kwon I, Xiang S, Kato M, Wu L, Theodoropoulos P, Wang T, et al. Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science*. 2014;345(6201):1139–45. <https://doi.org/10.1126/science.1254917>.
  55. Yin S, Lopez-Gonzalez R, Kunz RC, Gangopadhyay J, Borufka C, Gygi SP, et al. Evidence that C9ORF72 dipeptide repeat proteins associate with U2 snRNP to cause Mis-splicing in ALS/FTD patients. *Cell Rep*. 2017;19(11):2244–56. <https://doi.org/10.1016/j.celrep.2017.05.056>.
  56. Shi KY, Mori E, Nizami ZF, Lin Y, Kato M, Xiang S, et al. Toxic PRn poly-dipeptides encoded by the C9orf72 repeat expansion block nuclear import and export. *Proc Natl Acad Sci U S A*. 2017;114(7):E1111–7. <https://doi.org/10.1073/pnas.1620293114>.
  57. White MR, Mitrea DM, Zhang P, Stanley CB, Cassidy DE, Nourse A, et al. C9orf72 poly(PR) dipeptide repeats disturb biomolecular phase separation and disrupt nucleolar function. *Mol Cell*. 2019;74(4):713–6. <https://doi.org/10.1016/j.molcel.2019.03.019>.
  58. Vanneste J, Vercauteren T, Boeynaems S, Sicart A, Van Damme P, Daelemans D, et al. C9orf72-generated poly-GR and poly-PR do not directly interfere with nucleocytoplasmic transport. *Sci Rep*. 2019;9(1):15728–10. <https://doi.org/10.1038/s41598-019-52035-6>.
  59. Hayes LR, Duan L, Bowen K, Kalab P, Rothstein JD. C9orf72 arginine-rich dipeptide repeat proteins disrupt karyopherin-mediated nuclear import. *Elife*. 2020;9. <https://doi.org/10.7554/eLife.51685>.
  60. Wainger BJ, Kiskinis E, Mellin C, Wiskow O, Han SSW, Sandoe J, et al. Intrinsic membrane hyperexcitability of amyotrophic lateral sclerosis patient-derived motor neurons. *Cell Rep*. 2014;7(1):1–11. <https://doi.org/10.1016/j.celrep.2014.03.019>.
  61. Devlin A-C, Burr K, Borooah S, Foster JD, Cleary EM, Geti I, et al. Human iPSC-derived motoneurons harbouring TARDBP or C9ORF72 ALS mutations are dysfunctional despite maintaining viability. *Nat Commun*. 2015;6(1):5999–12. <https://doi.org/10.1038/ncomms6999>.
  62. Westergard T, McAvoy K, Russell K, Wen X, Pang Y, Morris B, et al. Repeat-associated non-AUG translation in C9orf72-ALS/FTD is driven by neuronal excitation and stress. *EMBO Mol Med*. 2019;11:1–14.
  63. Zhao C, Devlin A-C, Chouhan AK, Selvaraj BT, Stavrou M, Burr K, et al. Mutant C9orf72 human iPSC-derived astrocytes cause non-cell autonomous motor neuron pathophysiology. *Glia*. 2020;68(5):1046–64. <https://doi.org/10.1002/glia.23761>.
  64. Sareen D, O'Rourke JG, Meera P, Muhammad AKMG, Grant S, Simpkinson M, et al. Targeting RNA foci in iPSC-derived motor neurons from ALS patients with a C9ORF72 repeat expansion. *Sci Transl Med*. 2013;5:208ra149.
  65. Satoh J-I, Yamamoto Y, Kitano S, Takitani M, Asahina N, Kino Y. Molecular network analysis suggests a logical hypothesis for the pathological role of C9orf72 in amyotrophic lateral sclerosis/frontotemporal dementia. *J Cent Nerv Syst Dis*. 2014;6:69–78. <https://doi.org/10.4137/JCNSD.S18103>.
  66. Conlon EG, Fagegaltier D, Agius P, Davis-Parada J, Gregory J, Hubbard I, et al. Unexpected similarities between C9ORF72 and sporadic forms of ALS/FTD suggest a common disease mechanism. *Elife*. 2018;7:959.
  67. Jones AR, Troakes C, King A, Sahni V, De Jong S, Bossers K, et al. Stratified gene expression analysis identifies major amyotrophic lateral sclerosis genes. *Neurobiol Aging*. 2015;36:2006.e1–9.
  68. Strässer K, Masuda S, Mason P, Pfannstiel J, Oppizzi M, Rodriguez-Navarro S, et al. TREX is a conserved complex coupling transcription with messenger



- RNA export. *Nature*. 2002;417(6886):304–8. <https://doi.org/10.1038/nature746>.
69. Masuda S, Das R, Cheng H, Hurt E, Dorman N, Reed R. Recruitment of the human TREX complex to mRNA during splicing. *Genes Dev*. 2005;19(13):1512–7. <https://doi.org/10.1101/gad.1302205>.
  70. Viphakone N, Sudbery I, Griffith L, Heath CG, Sims D, Wilson SA. Co-transcriptional loading of RNA export factors shapes the human transcriptome. *Mol Cell*. 2019;75(2):310–8. <https://doi.org/10.1016/j.molcel.2019.04.034>.
  71. Dufu K, Livingstone MJ, Seebacher J, Gygi SP, Wilson SA, Reed R. ATP is required for interactions between UAP56 and two conserved mRNA export proteins, Aly and CIP29, to assemble the TREX complex. *Genes Dev*. 2010;24(18):2043–53. <https://doi.org/10.1101/gad.1898610>.
  72. Viphakone N, Hautbergue GM, Walsh M, Chang C-T, Holland A, Folco EG, et al. TREX exposes the RNA-binding domain of Nxf1 to enable mRNA export. *Nat Commun*. 2012;3(1):1006. <https://doi.org/10.1038/ncomms2005>.
  73. Hao Z, Liu L, Tao Z, Wang R, Ren H, Sun H, et al. Motor dysfunction and neurodegeneration in a C9orf72 mouse line expressing poly-PR. *Nat Commun*. 2019;10(1):2906–11. <https://doi.org/10.1038/s41467-019-10956-w> Nature Publishing Group.
  74. Freibaum BD, Lu Y, Lopez-Gonzalez R, Kim NC, Almeida S, Lee K-H, et al. GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. *Nature*. 2015;525(7567):129–33. <https://doi.org/10.1038/nature14974>.
  75. Lee Y-B, Chen H-J, Peres JN, Gomez-Deza J, Attig J, Štalekar M, et al. Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic. *Cell Rep*. 2013;5(5):1178–86. <https://doi.org/10.1016/j.celrep.2013.10.049>.
  76. García-Berrocó T, Llombart V, Colàs-Campàs L, Hainard A, Licker V, Penalba A, et al. Single cell Immuno-laser microdissection coupled to label-free proteomics to reveal the Proteotypes of human brain cells after ischemia. *Mol Cell Proteomics*. 2018;17(1):175–89. <https://doi.org/10.1074/mcp.RA117.000419>.
  77. Lemaire R, Prasad J, Kashima T, Gustafson J, Manley JL, Lafyatis R. Stability of a PKCI-1-related mRNA is controlled by the splicing factor ASF/SF2: a novel function for SR proteins. *Genes Dev*. 2002;16(5):594–607. <https://doi.org/10.1101/gad.939502>.
  78. Sun S, Zhang Z, Sinha R, Karni R, Krainer AR. SF2/ASF autoregulation involves multiple layers of post-transcriptional and translational control. *Nat Struct Mol Biol*. 2010;17(3):306–12. <https://doi.org/10.1038/nsmb.1750>.
  79. Delestienne N, Wauquier C, Soin R, Dierick J-F, Gueydan C, Krusys V. The splicing factor ASF/SF2 is associated with TIA-1-related/TIA-1-containing ribonucleoprotein complexes and contributes to post-transcriptional repression of gene expression. *FEBS J*. 2010;277(11):2496–514. <https://doi.org/10.1111/j.1742-4658.2010.07664.x>.
  80. Kulkarni M, Ozgur S, Stoecklin G. On track with P-bodies. *Biochem Soc Trans*. 2010;38(1):242–51. <https://doi.org/10.1042/BST0380242>.
  81. Sanford JR, Gray NK, Beckmann K, Cáceres JF. A novel role for shuttling SR proteins in mRNA translation. *Genes Dev*. 2004;18(7):755–68. <https://doi.org/10.1101/gad.286404>.
  82. Sanford JR, Ellis JD, Cazalla D, Cáceres JF. Reversible phosphorylation differentially affects nuclear and cytoplasmic functions of splicing factor 2/ alternative splicing factor. *Proc Natl Acad Sci U S A*. 2005;102(42):15042–7. <https://doi.org/10.1073/pnas.0507827102>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

