Published paper

Cooper-Knock, J., Bury, J.J., Heath, P.R., Wyles, M., Higginbottom, A., Gelsthorpe, C., Highley, J.R., Hautbergue, G., Rattray, M., Kirby, J. and Shaw, P.J. (2015) *C9ORF72 GGGGCC Expanded Repeats Produce Splicing Dysregulation which Correlates with Disease Severity in Amyotrophic Lateral Sclerosis.* PLoS One, 10 (5). e0127376

10.1371/journal.pone.0127376

PLOS | ONE

RESEARCH ARTICLE

C9ORF72 GGGGCC Expanded Repeats Produce Splicing Dysregulation which Correlates with Disease Severity in Amyotrophic Lateral Sclerosis

Johnathan Cooper-Knock[1], Joanna J. Bury[1], Paul R Heath[1], Matthew Wyles[1], Adrian Higginbottom[1], Catherine Gelsthorpe[1], J. Robin Highley[1], Guillaume Hautbergue[1], Magnus Rattray[2], Janine Kirby[1], Pamela J. Shaw[1]*

1 Sheffield Institute for Translational Neuroscience (SITraN), University of Sheffield, 385A Glossop Road, Sheffield, S10 2HQ, United Kingdom, **2** Life Sciences, The University of Manchester, Michael Smith Building, Oxford Road, Manchester, M13 9PT, United Kingdom

* pamela.shaw@sheffield.ac.uk

OPEN ACCESS

Citation: Cooper-Knock J, Bury JJ, Heath PR, Wyles M, Higginbottom A, Gelsthorpe C, et al. (2015) *C9ORF72* GGGGCC Expanded Repeats Produce Splicing Dysregulation which Correlates with Disease Severity in Amyotrophic Lateral Sclerosis. PLoS ONE 10(5): e0127376. doi:10.1371/journal.pone.0127376

Academic Editor: Huaibin Cai, National Institute of Health, UNITED STATES

Received: January 4, 2015

Accepted: April 15, 2015

Published: May 27, 2015

Data Availability Statement: The microarray dataset is available on GEO, accession number GSE68608. All relevant data are within the paper and its Supporting Information files.

Funding: The authors acknowledge grants from the 'EU Framework 7' (Euromotor No259867) and the SOPHIA project (funded by 'EU Joint Programme - Neurodegenerative Disease Research' and 'Medical Research Council') to PJS and JK. PJS is an NIHR Senior Investigator. JCK and JRH are supported by 'Motor Neurone Disease Association' / 'Medical Research Council' Lady Edith Wolfson Fellowship

Abstract

Objective

An intronic GGGGCC-repeat expansion of *C9ORF72* is the most common genetic variant of amyotrophic lateral sclerosis (ALS) and frontotemporal dementia. The mechanism of neurodegeneration is unknown, but a direct effect on RNA processing mediated by RNA foci transcribed from the repeat sequence has been proposed.

Methods

Gene expression profiling utilised total RNA extracted from motor neurons and lymphoblastoid cell lines derived from human ALS patients, including those with an expansion of *C9ORF72*, and controls. In lymphoblastoid cell lines, expansion length and the frequency of sense and antisense RNA foci was also examined.

Results

Gene level analysis revealed a number of differentially expressed networks and both cell types exhibited dysregulation of a network functionally enriched for genes encoding 'RNA splicing' proteins. There was a significant overlap of these genes with an independently generated list of GGGGCC-repeat protein binding partners. At the exon level, in lymphoblastoid cells derived from *C9ORF72*-ALS patients splicing consistency was lower than in lines derived from non-C9ORF72 ALS patients or controls; furthermore splicing consistency was lower in samples derived from patients with faster disease progression. Frequency of sense RNA foci showed a trend towards being higher in lymphoblastoid cells derived from patients with shorter survival, but there was no detectable correlation between disease severity and DNA expansion length.

awards ([MR/K003771/1] and [G0 800380] respectively). Samples used in this research were in part obtained from the UK MND DNA Bank for MND Research, funded by the 'Motor Neurone Disease Association' and the 'Wellcome Trust'. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Significance

Up-regulation of genes encoding predicted binding partners of the *C9ORF72* expansion is consistent with an attempted compensation for sequestration of these proteins. A number of studies have analysed changes in the transcriptome caused by *C9ORF72* expansion, but to date findings have been inconsistent. As a potential explanation we suggest that dynamic sequestration of RNA processing proteins by RNA foci might lead to a loss of splicing consistency; indeed in our samples measurement of splicing consistency correlates with disease severity.

Introduction

GGGGCC repeat expansions within intron 1 of the *C9ORF72* gene are the most common cause of familial amyotrophic lateral sclerosis (ALS) and familial frontotemporal degeneration (FTD) [1,2], though how this genetic change results in neuronal injury is not yet understood. Evidence is being gathered for a gain-of-function toxicity mediated by either sequestration of RNA binding proteins (RBPs) by RNA foci transcribed from the repeat sequence [3–8], or via repeat associated non-ATG (RAN) translation of the repeat sequence to produce a dipeptide repeat protein [9–11], or a combination of both mechanisms.

Gene expression profiling has the potential to identify biological pathways aberrantly affected by the *C9ORF72* expansion. In addition, if toxicity is mediated by nuclear RNA foci developed from an intronic expansion, then transcriptome changes may be relatively upstream in disease pathogenesis [12]. On this basis we have studied gene expression changes in motor neurons and lymphoblastoid cell lines derived from individuals with *C9ORF72*-ALS.

We have previously suggested that dynamic sequestration by RNA foci of a number of RBPs might affect nuclear speckle function and thus disrupt mRNA splicing [8]. It has been proposed that splicing errors are a normal occurrence for which the cell is able to compensate [13]. Therefore an excessive splicing error rate may not immediately result in disease; however in time compensatory mechanisms might be overwhelmed in vulnerable cells. This is more consistent with the variable phenotype and late age of onset seen in *C9ORF72*-ALS than a model of binary toxicity resulting from a small number of specific splicing errors. Therefore we aimed to derive a measure of the overall splicing error rate in biosamples containing the *C9ORF72* repeat expansion. Additionally we used Southern hybridisation and RNA fluorescence in-situ hybridisation (FISH) to examine the relationship between the changes in the splicing error rate, disease severity, the length of the GGGGCC repeat expansion and the abundance of RNA foci.

Results

Transcriptome analysis

Motor neurons. Network analysis using WGCNA identified six significant networks within 5,000 genes considered (**Fig 1**) all of which were differentially expressed between *C9ORF72*-ALS and control groups, and showed significant functional enrichment (**Table 1**). Based on the median fold change, three networks were down-regulated and three networks were up-regulated in *C9ORF72*-ALS (**Table 1**). Specifically, within the brown network which was significantly enriched for transcripts related to the Gene Ontology (GO) term 'RNA splicing', 58.2% of transcripts were up-regulated. The yellow and green networks were also up-regulated and functionally enriched for 'male sex differentiation' and 'erythrocyte homeostasis' respectively. The

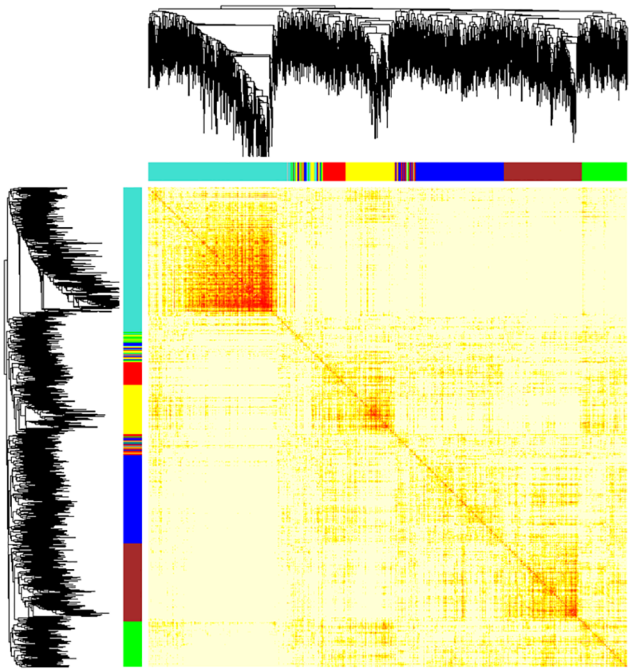


Fig 1. Gene level network analysis of transcriptome changes in motor neurons from *C9ORF72*-ALS cases. WGCNA analysis identified six gene networks which were dysregulated between *C9ORF72*-ALS and control samples. A clustering tree and heat map are shown illustrating separation of the gene networks, a lower branch height or darker colour denotes a greater Pearson correlation coefficient between pairs of genes.

doi:10.1371/journal.pone.0127376.g001

turquoise, blue and red networks were all down-regulated and functionally enriched for 'cholesterol biosynthetic process', 'regulation of glucose metabolic process' and 'regulation of nuclear division' respectively.

Lymphoblastoid cell lines. Two samples failed the Affymetrix quality control (QC) assessment and were excluded from the analysis based upon a low % presence call and/or AUC value. Network analysis using WGCNA identified nine significant networks which were differentially expressed between *C9ORF72*-ALS and control groups (Fig 2), and showed significant functional enrichment (Table 2). Based on the median fold change, five networks were down-

Table 1. Gene level network analysis of transcriptome changes in motor neurons from *C9ORF72*-ALS cases.

Network	Number of Genes	P-value C9ORF72-ALS Vs Control	Top Gene Ontology Enrichment	P-value for Enrichment Analysis	Median Fold Change
Turquoise	1555	0.008	Cholesterol biosynthetic process	0.001	0.62
Blue	1020	0.003	Regulation of glucose metabolic process	0.01	0.47
Brown	901	0.008	RNA splicing	7.45E-04	1.49
Yellow	635	0.003	Male sex differentiation	0.02	1.91
Green	579	0.0005	Erythrocyte homeostasis	0.01	1.75
Red	321	0.006	Regulation of nuclear division	0.01	0.49

WGCNA analysis identified six gene networks which were dysregulated between *C9ORF72*-ALS and control samples. The median fold change of genes within each network and the functional enrichment of each of the gene networks is tabulated. A fold change of >1 equates to up-regulation and a fold change of <1 equates to down-regulation.

doi:10.1371/journal.pone.0127376.t001

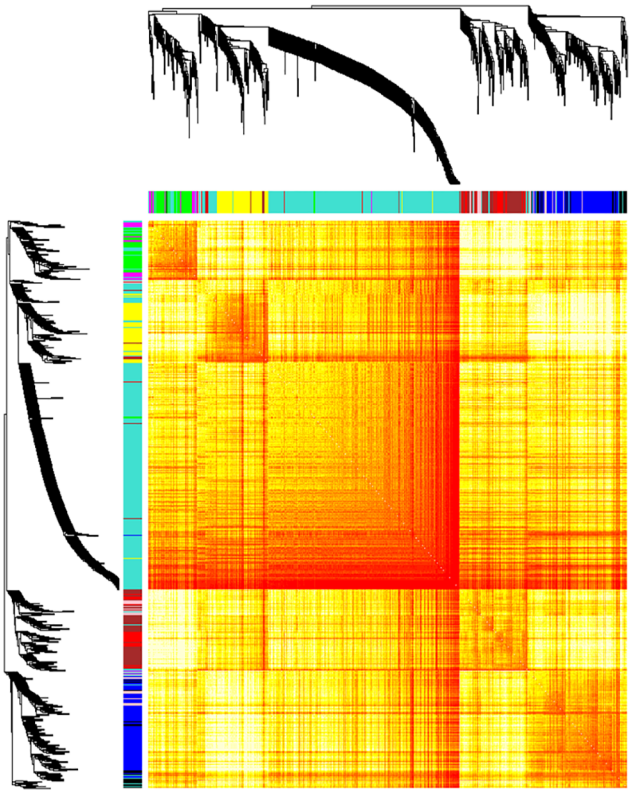


Fig 2. Gene level network analysis of transcriptome changes in lymphoblastoid cell lines derived from *C9ORF72*-ALS cases. WGCNA analysis identified nine gene networks which were dysregulated between *C9ORF72*-ALS and control samples. A clustering tree and heat map are shown illustrating separation of the gene networks, a lower branch height or darker colour denotes a greater Pearson correlation coefficient between pairs of genes. The median fold change of genes within each network and the functional enrichment of each of the gene networks is tabulated (B). A fold change of >1 equates to up-regulation and a fold change of <1 equates to down-regulation.

doi:10.1371/journal.pone.0127376.g002

Table 2. Gene level network analysis of transcriptome changes in lymphoblastoid cell lines derived from C9ORF72-ALS cases.

Network	Number of Genes	P-value C9ORF72-ALS Vs Control	Top Gene Ontology Enrichment	P-value for Enrichment Analysis	Median Fold Change
Turquoise	4653	8.11E-64	Positive regulation of apoptosis	0.01	1.19
Blue	1403	0.0000001	Regulation of action potential in neuron	0.02	0.88
Brown	1038	4.69E-09	Protein catabolic process	0.002	0.79
Yellow	854	2.48E-09	Synaptic transmission	0.004	1.16
Green	537	0.0000002	RNA splicing	1.50E-05	1.27
Red	427	9.17E-08	Positive regulation of apoptosis	0.02	0.74
Black	391	3.43E-09	Striated muscle tissue development	0.02	0.86
Pink	367	3.54E-08	Inflammatory response	0.004	0.86
Magenta	336	0.0000001	Protein catabolic process	1.43E-05	1.53

WGCNA analysis identified nine gene networks which were dysregulated between C9ORF72-ALS and control samples. The median fold change of genes within each network and the functional enrichment of each of the gene networks is tabulated. A fold change of >1 equates to up-regulation and a fold change of <1 equates to down-regulation.

doi:10.1371/journal.pone.0127376.t002

regulated and four networks were up-regulated in *C9ORF72*-ALS (Table 2). Specifically 92% of transcripts were up-regulated within the green network, which was significantly enriched for transcripts related to the GO term 'RNA splicing'. The turquoise and red networks were both functionally enriched for genes related to 'positive regulation of apoptosis', the blue and yellow networks were functionally enriched for categories directly related to neuronal function, and the black network was functionally enriched for 'striated muscle tissue development'; the pink network was functionally enriched for 'inflammatory response'; and the brown and magenta networks were functionally enriched for 'protein catabolic process.'

Analysis of networks enriched for the GO term 'RNA Splicing'. The brown network in the motor neurons and the green network in the lymphoblastoid cell lines were both up-regulated in *C9ORF72*-ALS samples and significantly enriched for transcripts related to the GO term 'RNA splicing.' We set out to determine whether the two networks contained similar transcripts or only transcripts with similar functional enrichment.

To make this comparison we reverted to the original lists of transcripts not filtered by Pearson correlation coefficient because computational burden was no longer an issue, and we were interested in all transcripts associated with the network signal and not just the most correlated. Examination of all genes significantly correlated (as quantified by Pearson correlation coefficient) ($p<0.05$) with the brown network signal and associated with the GO term 'RNA Processing' in motor neurons derived from *C9ORF72*-ALS patients revealed 88 transcripts encoding 74 unique genes (S1 Table). Examination of all genes significantly correlated ($p<0.05$) with the green network signal and associated with the GO term 'RNA Processing' in lymphoblastoid cells derived from *C9ORF72*-ALS patients revealed 459 transcripts encoding 236 unique genes (S1 Table). Given the difference in cell types and microarray platforms, there was evidence for significant similarity between the lists: 54% of the motor neuron list was also present within the lymphoblastoid cell list. Previously we have identified candidate binding partners of the GGGGCC repeat expansion by RNA pulldown and mass spectroscopy [8]. 20% of the unique hits identified in this way were present within the lymphoblastoid cell line list ($p<0.0001$) of which 89% were up-regulated in the *C9ORF72*-ALS samples compared to controls (S1 Table); and 10% of the unique hits were present within the motor neuron list ($p<0.0001$) of which 77% were up-regulated in the *C9ORF72*-ALS samples compared to controls (S1 Table).

Analysis of splicing. There was no significant difference in the total number of splicing events observed in lymphoblastoid cell lines derived from *C9ORF72*-ALS patients, non-*C9ORF72*-ALS patients and controls (Fig 3). However, the nature of those splicing events was significantly different. It is expected that functionally appropriate splicing would be similar in samples of a particular group and therefore we propose that splicing consistency is a marker of the error rate in RNA splicing. Splicing consistency was significantly reduced in the *C9ORF72*-ALS group compared to non-*C9ORF72*-ALS patients and controls (Fig 4A, S1 Fig). It is noteworthy that control cases do not share a common disease process and therefore might be expected to have quite different patterns of splicing. This is good evidence that splicing in *C9ORF72-ALS* is actively disrupted. In addition, splicing was less consistent in *C9ORF72-ALS* patients who lived <2 years following diagnosis compared to those that lived >4 years suggesting a link with the aggressiveness of the disease course (Fig 4B).

qPCR based validation of transcriptome changes

Candidates for qPCR validation were chosen from those genes which were up-regulated in the lymphoblastoid cells derived from *C9ORF72*-ALS cases compared to controls, and also identified as candidate binding partners of the GGGGCC repeat expansion by RNA pulldown and mass

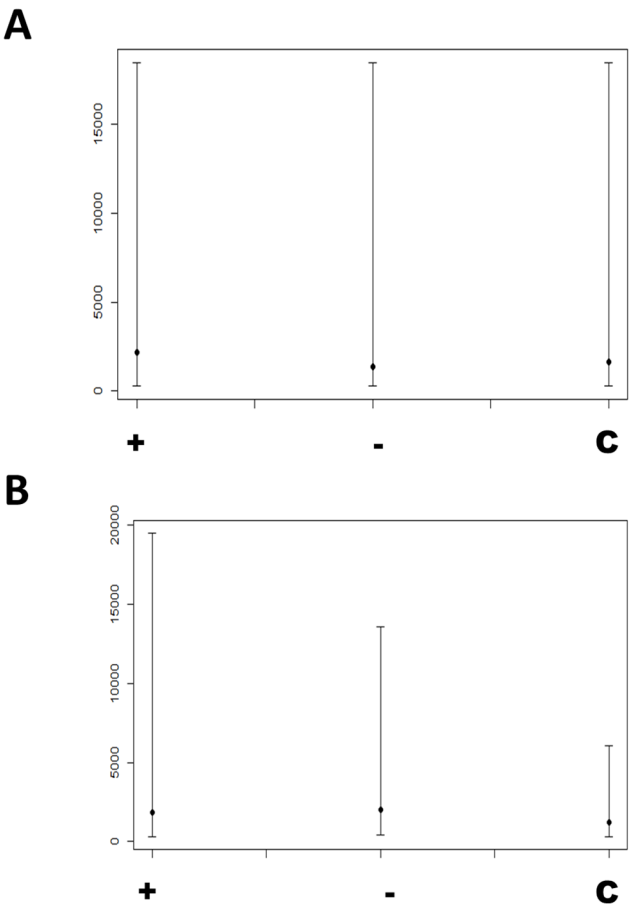


Fig 3. Frequency of exon inclusion and exclusion events. Plots of median and 95% CI for numbers of (A) exon inclusion and (B) exon exclusion events in *C9ORF72*-ALS (+), non-*C9ORF72* ALS (-) and control (C) derived lymphoblastoid cell lines, as determined by FIRMA score. There was no significant difference between sample groups.

doi:10.1371/journal.pone.0127376.g003

spectroscopy [8]. qPCR confirmed up-regulation of *HNRNPF* (1.43 fold, t-test, p = 0.001), *RBM3* (1.18 fold, t-test, p = 0.03) and *FUS* (1.35 fold, t-test, p = 0.005) but not *HNRNPH2*.

Estimation of expansion size and quantification of the abundance of RNA foci in lymphoblastoid cell lines

GGGGCC repeat expansion size and abundance of sense and antisense RNA foci was determined in lymphoblastoid cell lines derived from 17 patients with short (<2 years) disease duration and 7 patients with long (>4 years) disease duration. No difference in minimum (t-test, p = 0.10), modal (t-test, p = 0.41) or maximum (t-test, p = 0.57) repeat size was detectable between groups by Southern blotting (data not shown).

The frequency of lymphoblastoid cells containing sense RNA foci was higher in lines derived from 3 patients with short (<2 years) disease duration compared to lines derived from 3 patients with long (>4 years) disease duration, however this trend did not reach significance (average frequency of sense foci+ cells was 0.35 versus 0.12, t-test, p = 0.099). There was no such trend in the frequency of lymphoblastoid cells containing antisense RNA foci (average frequency of antisense foci+ cells 0.20 versus 0.32, t-test, p = 0.29). Example cells are shown in Fig 5.

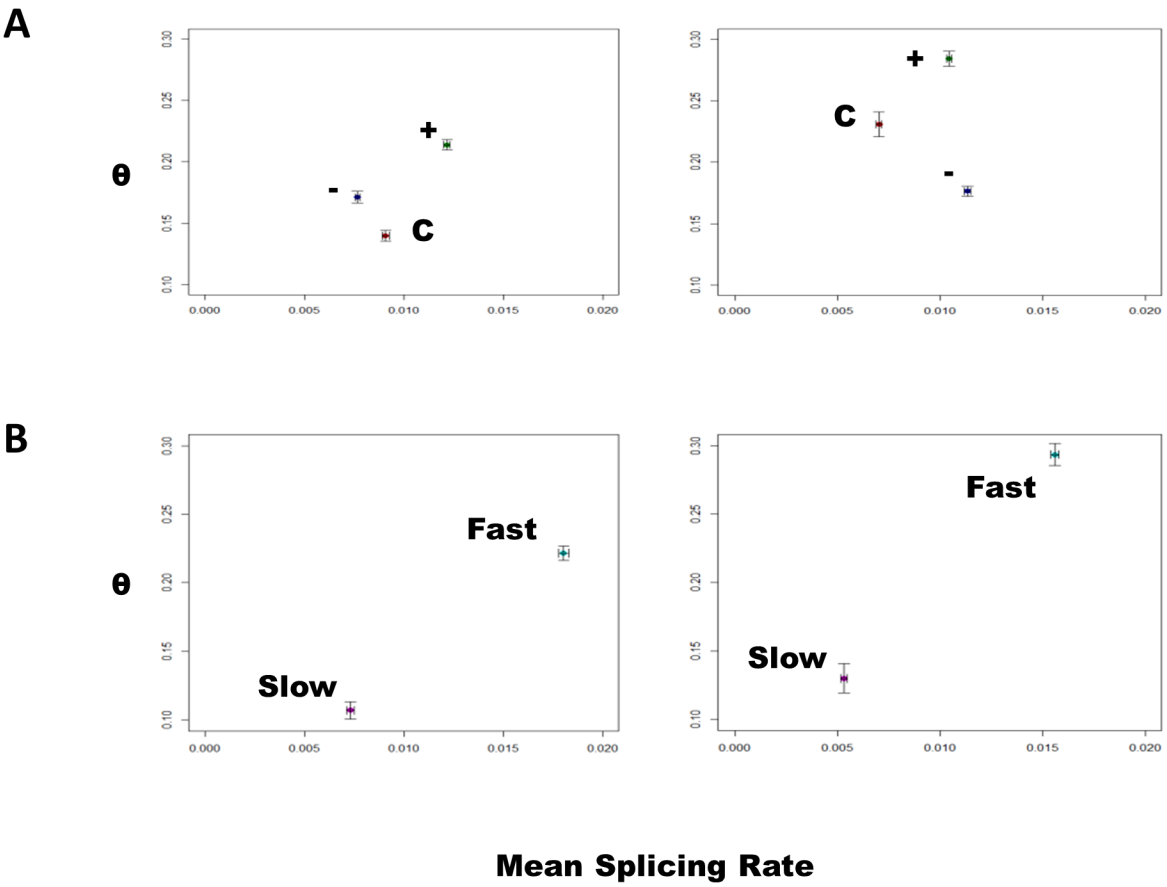


Fig 4. Plots of θ against the mean splicing rate with 95% confidence intervals. Exon inclusion events are shown in the left panel and exclusion inclusion events are shown in the right panel. θ is higher indicating reduced consistency of splicing in (A) *C9ORF72-ALS* (+) compared to non-*C9ORF72* ALS (-) and control (C) derived lymphoblastoid cell lines; and (B) in cell lines derived from patients with rapid (length <2 years, Fast) compared to slowly (length >4 years, Slow) progressive *C9ORF72*-ALS.

doi:10.1371/journal.pone.0127376.g004

Discussion

There is an urgent need to understand the mechanisms of neuronal injury in *C9ORF72*-disease. In order to establish the biological pathways altered by the presence of the GGGGCC repeat expansion we carried out gene expression profiling of isolated motor neurons from spinal cord and lymphoblastoid cell lines derived from human ALS patients and controls. Moreover, it has been suggested that the *C9ORF72* expansion has a direct effect on the transcriptome, possibly via the formation of RNA foci [1,8]; if this is the case then transcriptome changes may represent a relatively upstream component of pathogenesis and a suitable therapeutic target.

Transcriptome analysis in *C9ORF72*-ALS motor neurons

Six gene networks were identified as differentially expressed between *C9ORF72-ALS* and control motor neurons (**Fig 1**, **Table 1**). Several networks were significantly enriched for GO categories previously implicated in ALS including 'cholesterol biosynthetic process' which occurs primarily in the endoplasmic reticulum (ER) [14], 'regulation of glucose metabolic process' [15], 'regulation of nuclear division' [16] and 'RNA splicing' [17] (**Table 1**). The 'RNA splicing' network overlapped with a similar network in the lymphoblastoid cells and will be discussed further below.

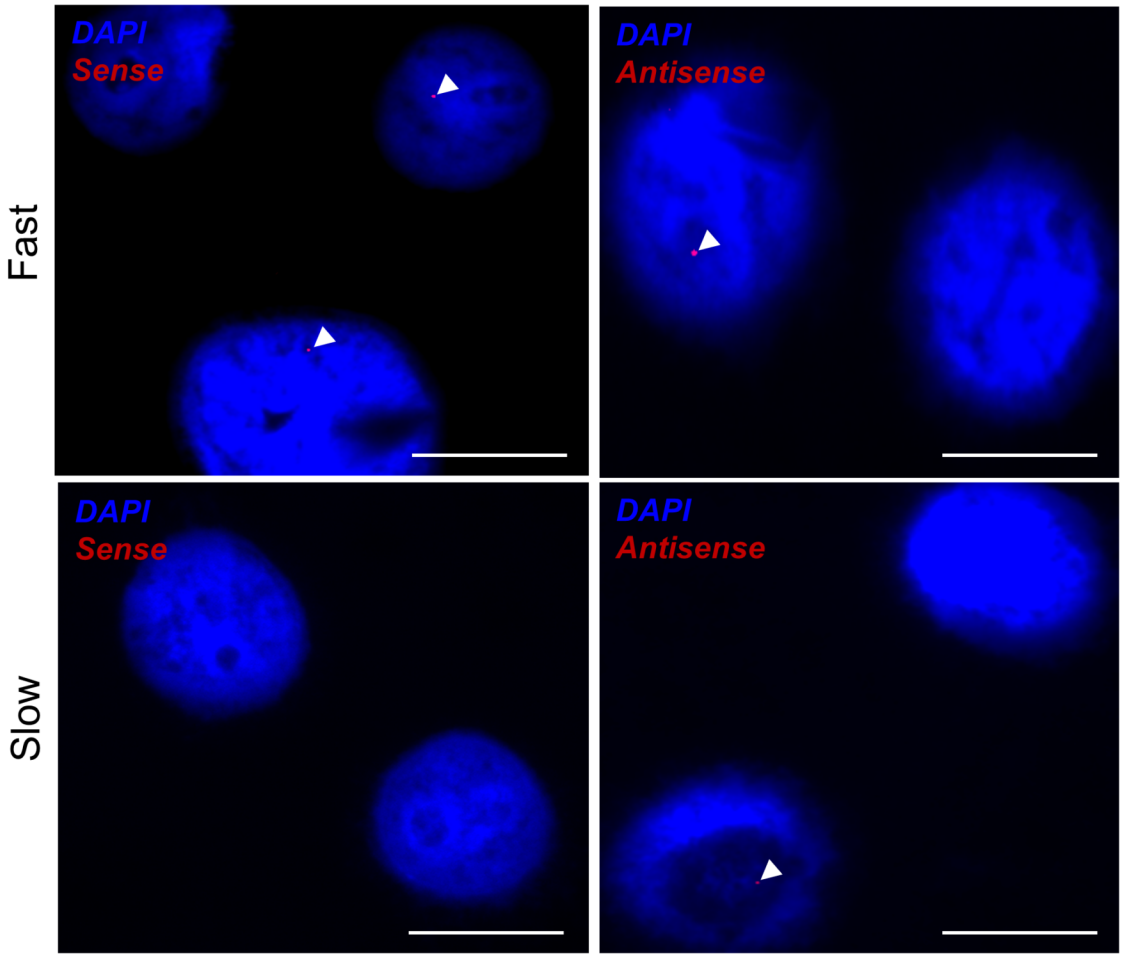


Fig 5. RNA foci in lymphoblastoid cell lines derived from patients with short or long survival. RNA FISH was performed for sense and antisense RNA foci in lymphoblastoid cells. Example cells are shown derived from patients with rapid (length <2 years, Fast, upper panels) compared to slowly (length >4 years, Slow, lower panels) progressive *C9ORF72*-ALS. GGGGCC-repeat sense RNA foci are visualised (arrowheads) in the left panels whereas GGCCCC-repeat antisense RNA foci are visualised (arrowheads) in the right panels. Scale bar 10 μm.

doi:10.1371/journal.pone.0127376.g005

Dysregulation of gene networks related to glucose and cholesterol metabolism, both of which were down-regulated in *C9ORF72*-ALS motor neurons, is interesting. Increasingly ER stress is implicated in ALS. As well as a site of lipid synthesis, the ER is responsible for correct protein folding [18]. Protein aggregates are a prominent feature of all forms of ALS; ER stress activates the unfolded protein response (UPR) and chronically can lead to apoptosis. Indeed activation of the UPR has been observed in sporadic ALS patients [19]. ER stress has been observed to impact negatively on cholesterol synthesis [20]. In a different cell model we have previously demonstrated a deficit in glucose metabolism associated with ALS [21].

Transcriptome analysis in *C9ORF72*-ALS lymphoblastoid cells

Nine transcript networks were identified as differentially expressed between *C9ORF72*-ALS and control lymphoblastoid cells (**Fig 2**, **Table 2**). With the exception of the gene network enriched for 'RNA splicing', the functional enrichment of differentially expressed networks in the *C9ORF72*-ALS lymphoblastoid cells and motor neurons was distinct. This is not unexpected given the use of non-overlapping cases in the sample sets and different analysis platforms;

indeed this makes the identification of a common network all the more robust. Moreover, there are differences in physiology between the cell types: motor neurons are post-mitotic whereas lymphoblastoid cells are actively dividing which may explain why 'regulation of nuclear division' did not appear in the lymphoblastoid cell networks.

Some common themes arise in the enrichment of the differentially expressed networks in the lymphoblastoid cells: three networks were enriched for functional categories related to nerve or muscle function (**Table 2**). This suggests that the presence of the expansion in the lymphoblastoid cells has an effect on genes important for neuromuscular function. Even if not deleterious to the lymphoblastoid cells, these same changes may be toxic if they occur in the cell types vulnerable to the neurodegenerative pathology in ALS. Two networks were enriched for functional categories related to protein catabolism (**Table 2**). Failure of protein catabolism has been implicated previously in ALS [22], indeed several genetic variants of ALS are caused by mutations in genes with roles in protein degradation e.g. VCP [23] and UBQLN2 [24]. Two networks were enriched for functional categories related to regulation of apoptosis (**Table 2**). Dysregulation of pathways related to apoptosis has also been previously implicated in ALS [25]. Both protein processing and regulation of apoptosis have been linked to ALS in the context of ER stress [18]; misfolded protein accumulation can induce ER stress, and chronic ER stress can lead to apoptosis.

A network of genes enriched for 'RNA splicing' as was up-regulated in both cell types under examination suggesting that it may represent an upstream effect of the expansion. Analysis of the 'RNA splicing' network signal in both models showed that the similarity extended beyond the functional enrichment to the actual genes dysregulated. Moreover the dysregulated gene lists were significantly enriched with independently generated candidate protein binding partners of the GGGGCC-repeat expansion from our own work [8] and that of others [7,26,27]. We have previously proposed that *C9ORF72*-disease involves dynamic sequestration of a significant number of RBPs involved in mRNA splicing, by RNA foci transcribed from the GGGGCC repeat [8]. The observed up-regulation of genes encoding these proteins in both *C9ORF72*-ALS motor neurons and lymphoblastoid cells is consistent with attempted compensation by the cell for a sequestration process.

Exon splicing in *C9ORF72*-ALS lymphoblastoid cells

In view of these findings we attempted to examine global splicing function within the cell. Despite their being non-neuronal, we utilised lymphoblastoid cells in this analysis because of the large number of samples available and the accessibility of high quality RNA. It is reasonable to expect that a molecular phenotype observed in lymphoblastoid cells might also be present in the central nervous system (CNS). We have previously shown that detectable *C9ORF72* expansion length [28] and transcription of RNA foci [8] are comparable between lymphoblastoid cell lines and the CNS.

Splicing errors are likely to be a normal occurrence for which the cell is able to compensate [13]. However, if the load of these errors is increased then the compensatory mechanism may be overcome and the probability of this occurring might be expected to increase with time. This is consistent with the late age of onset and markedly variable phenotype found in *C9ORF72*-disease. In order to quantify splicing errors, we defined functionally appropriate splicing as likely to be consistent between members of a particular group: *C9ORF72*-ALS, non-*C9ORF72* ALS or controls. We identified a reduction in splicing consistency, or an increase in the splicing error rate, in *C9ORF72*-ALS samples compared to non-*C9ORF72* ALS samples and controls; moreover the splicing error rate was higher in samples derived from *C9ORF72*-ALS patients with shorter survival compared to samples derived from *C9ORF72*-ALS patients with

a longer survival, suggesting a link with CNS toxicity. Consistent with our hypothesis there was a trend for the frequency of sense RNA foci to be higher in lymphoblastoid cell lines derived from patients with a shorter survival; this might be expected to increase the sequestration of RNA splicing proteins and thus exacerbate the production of splicing errors.

A number of studies have previously examined the transcriptome in the presence of expanded *C9ORF72* [3–5,29]; the findings of these studies have so far been inconsistent. An increase in the number of splicing errors is a potential explanation for this finding. We await further validation of our results using new technologies such as RNA sequencing and utilising newly emerging disease models including iPS derived motor neurons from patients with *C9ORF72* mutations.

Materials and Methods

Transcriptome analysis

Laser captured motor neurones. Brain and spinal cord tissue from eight *C9ORF72*-ALS patients and three neurologically normal human control subjects was obtained from the Sheffield Brain Tissue Bank (**Table 3**). *C9ORF72*-ALS samples were identified by repeat-primed PCR of the *C9ORF72* gene [1,2]. Clinically these patients resembled the full clinical spectrum of *C9ORF72*-ALS: Mean age of onset was 61 years (range 56 to 66 years) and mean disease duration was approximately 2 years (range 7 months to 43 months). Tissue donated for research was obtained with written informed consent from the next of kin, and in accordance with the UK Human Tissue Authority guidelines on tissue donation. The work was approved by the South Yorkshire Ethics Committee.

Spinal cord sections from the limb enlargements were collected postmortem, processed according to standard protocols [30], and stored at −80ºC until required. Cervical spinal cord sections were prepared, between 800 and 1200 motor neurons were isolated and RNA was extracted using methods described previously [31]. RNA quantity and quality was assessed on the Nanodrop spectrophotometer and Agilent Bioanalyser, respectively, to ensure all samples were of comparable and sufficient quality to proceed. RNA (20–25ng) was linearly amplified using the Affymetrix Two Cycle cDNA synthesis protocol to produce biotin-labelled copy RNA. Copy RNA (15μg) was fragmented for 15min and hybridized to the Human Genome U133 Plus 2.0 GeneChips, according to Affymetrix protocols. Array washing and staining was

Table 3. Clinical information relating to motor neurons laser captured from ALS patients and controls, utilised in gene level microarray analysis.

Sample Type	Gender	Age	Duration	Diagnosis	Presentation	C9orf72
Control1	F	52	-	-	-	-
Control2	M	63	-	-	-	-
Control3	F	65	-	-	-	-
Patient1	F	62	2.00	Familial	Bulbar	+
Patient2	F	61	3.33	Sporadic	Bulbar	+
Patient3	M	66	1.17	Familial	Bulbar	+
Patient4	F	56	3.58	Familial	Limb	+
Patient5	M	62	1.67	Sporadic	Bulbar	+
Patient6	F	61	3.50	Sporadic	Limb	+
Patient7	M	70	2.17	Familial	Limb	+
Patient8	F	58	0.58	Sporadic	Limb	+

Age at symptom onset and disease duration is provided in years. Abbreviations: M = male, F = female.

doi:10.1371/journal.pone.0127376.t003

performed in the GeneChip fluidics station 400 and arrays were scanned on the GeneChip 3000 scanner. GeneChip Operating Software (GCOS) was used to generate signal intensities for each transcript.

Data Analysis. Data were normalised using the Puma package which quantifies technical variability to improve the estimation of gene expression [32, 33]. The next step was to identify networks of genes with correlated expression which are likely to represent functional groups. To reduce the computational burden and enhance the signal strength in the data, genes were ranked by t-statistic in a disease versus control comparison; the top 10,000 genes were then taken forward. For network detection, genes were further filtered to find the 5000 most connected (as quantified by Pearson correlation coefficient) genes; by definition, networked genes are strongly connected and therefore this should not lead to loss of information [34]. Network detection was performed using the weighted gene coexpression network analysis (WGCNA) package [35]. The correlation between expression of a given network of genes and whether a sample was a *C9ORF72*-case or a control was quantified and a Student's asymptotic p-value calculated; p-values <0.05 were taken to be significant. Differentially expressed networks were examined and an enrichment analysis performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) [36,37]. Enrichment was calculated by functional annotation clustering using the 'high' i.e. specific, Gene Ontology 'biological processes' terms.

Lymphoblastoid cell lines. Lymphoblastoid cell lines derived from Caucasian ALS patients (n = 56) and neurologically normal controls (n = 15), all of Northern European descent, were obtained from the UK Motor Neurone Disease Association (MNDA) DNA Bank (**Table 4**). *C9ORF72*-ALS samples were identified by repeat-primed PCR of the *C9ORF72* gene [1,2]. Clinically these patients resembled the full clinical spectrum of *C9ORF72*-ALS: Mean age of onset was 58 years (range 28 to 75 years) and mean disease duration was approximately 2 years (range 2 months to 83 months). All samples were collected with written informed consent from the donor, and the work was approved by the South Yorkshire Ethics Committee.

Total RNA was extracted from ALS patient and control-derived lymphoblastoid cell lines using QIAGEN's RNeasy Mini Kit following the manufacturer's recommendations. A 75μL LCL suspension, containing approximately $5x10^6$ cells, typically yields between 1.9 and 13.6μg total RNA with a mean concentration of approximately 170ng/μl as assessed the by the NanoDrop 1000 spectrophotometer (Thermo Scientific). The quality of the isolated material was analysed using the 2100 bioanalyzer with an RNA 6000 Nano LabChip Kit (Agilent Technologies, Inc.). Linear amplification of RNA with an input of approximately 300ng of starting material was performed using the Ambion Whole Transcript (WT) Expression Assay (*Applied Biosystems*) and Affymetrix GeneChip WT Terminal Labelling Kit. This procedure generated fragments of biotin-labelled sense-stranded copy DNA (6–10μg) between 40 and 70 nucleotides in length that were hybridized onto Human Exon 1.0ST GeneChip Arrays according to Affymetrix protocols. Array washing, staining and visualisation were performed as described for motor neuron derived RNA.

Data analysis. Network analysis of gene expression in the lymphoblastoid cell lines was identical to that in the motor neurons. However, because of the exon level probing, after Puma normalisation, approximately twice as many transcripts were quantified. This was taken into consideration in the filtering steps and the same proportion of transcripts were analysed at each stage rather than the exact same number i.e. the top 20,000 genes ranked by t-statistic were filtered to the 10,000 most connected for the network analysis.

Exon level data were analysed using the 'finding isoforms using robust multichip analysis' (FIRMA) package [38] which itself is part of the Aroma Affymetrix package [39]. The FIRMA step was then applied to detect alternative splicing; a FIRMA score is calculated for each exon. The score represents the result of fitting a transcript-level model to the observed data and

Table 4. Clinical information relating to lymphoblastoid cell lines derived from ALS patients and controls, utilised in exon level microarray analysis.

Sample Type	Gender	Age	Duration	Diagnosis	Presentation	C9orf72
Control1	F	52	-	-	-	-
Control2	M	69	-	-	-	-
Control3	F	65	-	-	-	-
Control4	F	84	-	-	-	-
Control5	M	56	-	-	-	-
Control6	F	59	-	-	-	-
Control7	M	73	-	-	-	-
Control8	F	67	-	-	-	-
Control9	M	47	-	-	-	-
Control10	M	64	-	-	-	-
Control11	F	41	-	-	-	-
Control12	M	36	-	-	-	-
Control13	M	61	-	-	-	-
Control14	M	54	-	-	-	-
Control15	F	63	-	-	-	-
Patient1	F	69	>4.00	Familial	Limb	+
Patient2	F	61	2.96	Familial	Limb	-
Patient3	F	28	1.10	Familial	Bulbar	+
Patient4	M	44	2.11	Familial	Respiratory	-
Patient5	F	46	Unknown	Familial	Bulbar	-
Patient6	M	69	1.76	Familial	Limb	+
Patient7	M	48	Unknown	Familial	Mixed	-
Patient8	M	57	5.71	Familial	Mixed	-
Patient9	F	57	1.21	Familial	Mixed	+
Patient10	M	63	>5.00	Familial	Limb	+
Patient11	F	62	0.17	Familial	Bulbar	+
Patient12	F	64	6.92	Familial	Limb	+
Patient13	M	59	<1.00	Familial	Unknown	+
Patient14	M	63	1.71	Familial	Mixed	+
Patient15	F	56	4.14	Familial	Limb	+
Patient16	M	47	1.63	Familial	Limb	+
Patient17	F	51	0.97	Familial	Bulbar	+
Patient18	F	61	Unknown	Familial	Bulbar	-
Patient19	M	73	1.88	Sporadic	Respiratory	-
Patient20	M	60	1.15	Sporadic	Bulbar	+
Patient21	M	64	2.36	Sporadic	Bulbar	-
Patient22	F	68	3.31	Sporadic	Bulbar	-
Patient23	M	68	1.56	Sporadic	Limb	+
Patient24	F	72	4.66	Sporadic	Limb	+
Patient25	M	58	1.40	Sporadic	Bulbar	-
Patient26	M	54	2.89	Sporadic	Bulbar	-
Patient27	M	53	3.28	Sporadic	Limb	-
Patient28	F	52	2.25	Sporadic	Limb	+
Patient29	M	72	2.58	Sporadic	Limb	-
Patient30	M	60	1.08	Sporadic	Bulbar	-
Patient31	F	67	1.47	Sporadic	Bulbar	+

(*Continued*)

Table 4. (*Continued*)

Sample Type	Gender	Age	Duration	Diagnosis	Presentation	C9orf72
Patient32	F	37	1.74	Sporadic	Limb	+
Patient33	M	56	2.20	Sporadic	Limb	+
Patient34	M	59	1.84	Sporadic	Limb	-
Patient35	F	70	2.13	Sporadic	Limb	-
Patient36	M	38	2.83	Sporadic	Mixed	-
Patient37	M	45	1.47	Sporadic	Limb	+
Patient38	F	48	~4.00	Sporadic	Bulbar	+
Patient39	F	72	1.87	Sporadic	Bulbar	-
Patient40	M	72	0.52	Sporadic	Limb	+
Patient41	F	75	1.05	Sporadic	Limb	-
Patient42	F	52	2.18	Sporadic	Limb	-
Patient43	F	58	1.33	Sporadic	Mixed	+
Patient44	M	47	1.57	Sporadic	Limb	+
Patient45	F	48	5.95	Sporadic	Limb	+
Patient46	M	64	0.66	Sporadic	Limb	+
Patient47	F	37	4.50	Sporadic	Bulbar	+
Patient48	M	70	1.24	Sporadic	Limb	-
Patient49	F	70	3.04	Sporadic	Limb	-
Patient50	M	61	2.57	Sporadic	Bulbar	-
Patient51	M	62	1.96	Sporadic	Limb	+
Patient52	F	58	<1.00	Sporadic	Bulbar	+
Patient53	M	61	~4.00	Sporadic	Mixed	+
Patient54	M	65	1.40	Sporadic	Limb	+

Age at symptom onset and disease duration is provided in years. Abbreviations: M = male, F = female.

doi:10.1371/journal.pone.0127376.t004

observing the disparity between the model and the exon-level intensity of each individual exon. Thus exons with a different level of expression to their parent transcript i.e. those which are spliced in or out, are identified. Utilising all probes specific for an entire transcript results in a significant improvement on the estimation of exon expression compared to using the relatively small number of probes specific to a given exon in isolation. FIRMA scores were log transformed prior to analysis. Highly negative or positive values of the FIRMA score are indicative of alternative exon skipping or inclusion respectively. The 1st and 99th percentiles of the FIRMA score for all exons in all samples were used to identify exons with the most evidence of alternative splicing, as used previously [40].

Consistency of splicing within a sample group such as patients or controls was evaluated by comparing the number of splicing events which occurred in 1, 2, 3. . .n samples within the group. To allow comparison between groups, comparison was made with the situation in which exons are spliced in or out at random. In each case the random situation was modelled with a Poisson distribution and the observed data were fitted to a negative binomial distribution (**S1 Fig**). θ is a quantification of the overdispersion in the negative binomial distribution with respect to an equivalent Poisson distribution, which is therefore a measure of non-random choice, i.e. consistency, in the splicing observed in each sample group. The variance of a negative binomial distribution is given by $\mu+\mu 2/\theta$ where μ = mean. In contrast the variance of a Poisson distribution is equal to μ. Therefore a higher level of θ corresponds to a variance closer to the Poisson distribution and reduced consistency of splicing.

qPCR based validation of transcriptome changes

Total RNA from lymphoblastoid cells was amplified using High Capacity RNA-to-cDNA kit (Applied Biosystems). Quantitative PCR (QPCR) primers for HNRNPF, FUS, HNRNPH2 and RBM3 transcripts were designed using Eurofins online primer design software (http://www.eurofinsdna.com). QPCR of 75 *C9ORF72*-ALS cases and 35 controls was performed using Brilliant II SYBR Green QPCR Master Mix (Stratagene) on the Stratagene 3000, as described previously [41]. RNA from groups of five *C9ORF72*-ALS or five control samples was pooled. These samples were obtained from the UK MNDA DNA Bank and included those samples utilised in the microarray analysis, as well as additional samples. T-tests were used to determine if the relative differences in transcript expression in lymphoblastoid cells between *C9ORF72*-ALS samples and controls were statistically significant.

Estimation of expansion size in lymphoblastoid cell lines

GGGGCC expansion size was estimated using a Southern hybridisation based protocol as previously described [28] using DNA derived from patients with rapid (<2 years, n = 17) or slowly (>4 years, n = 7) progressive disease.

Quantification of abundance of RNA foci in lymphoblastoid cell lines

A 5' TYE-563-labeled LNA (16-mer fluorescent)-incorporated DNA probe was used against the sense (Exiqon, Inc.; batch number 607323) and the antisense RNA hexanucleotide repeat (Exiqon, Inc.; batch number 610331). Slides were prepared and RNA foci were visualised as described previously [8]. More than fifty lymphoblastoid cells derived from patients with rapid (<2 years, n = 3) or slowly (>4 years, n = 3) progressive disease were imaged.

Supporting Information

S1 Fig. Quantification of splicing events shared between lymphoblastoid cell lines of each sample group. Plots of the number of splicing events (y-axis) which were present in a given number of lymphoblastoid cell lines (x-axis) within a particular sample group. Sample groups from top to bottom are: normal controls, *C9ORF72*-ALS patients, non-*C9ORF72* ALS patients, *C9ORF72*-ALS patients with survival <2 years and *C9ORF72*-ALS patients with survival >4 years. In each plot the left-hand line represents a Poisson fit to the observed data i.e. the random case. The right-hand line is the observed data and the dotted line represents the negative binomial distribution fit to the observed data. In each case the negative binomial provides a relatively good fit to the observed data. θ as shown in Fig 4, is a quantification of the overdispersion in the negative binomial compared to the Poisson fit to the observed data i.e. the degree of consistency in the splicing observed in each sample group.
(TIF)

S1 Table. Genes associated with the 'RNA splicing' network signal and within the GO term 'RNA Processing.' Genes listed are within the GO category 'RNA Processing' and are significantly associated ($p<0.05$) with the 'RNA splicing' network signal in either *C9ORF72*+ lymphoblastoid cell lines or motor neurons. A fold change of >1 equates to up-regulation and a fold change of <1 equates to down-regulation.
(XLSX)

Acknowledgments

Samples used in this research were in part obtained from the UK MND DNA Bank for MND Research, funded by the MND Association and the Wellcome Trust. We are grateful to all of the patients with ALS and their family members who donated biosamples for research.

Author Contributions

Conceived and designed the experiments: JCK JRH GH MR JK PJS. Performed the experiments: JCK JJB PRH MW AH CG. Analyzed the data: JCK JJB PRH JRH MR PJS. Contributed reagents/materials/analysis tools: MR JK PRH PJS. Wrote the paper: JCK JJB PRH MW AH CG JRH GH MR JK PJS.

References

1. DeJesus-Hernandez M, Mackenzie I, Boeve B, Boxer A, 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: 245–256. doi: 10.1016/j.neuron.2011.09.011 PMID: 21944778
2. Renton Alan E, 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: 257–268. doi: 10.1016/j.neuron.2011.09.010 PMID: 21944779
3. Donnelly CJ, Zhang PW, Pham JT, Heusler AR, Mistry NA, Vidensky S, et al. RNA Toxicity from the ALS/FTD C9ORF72 Expansion Is Mitigated by Antisense Intervention. Neuron 2013; 80: 415–428. doi: 10.1016/j.neuron.2013.10.015 PMID: 24139042
4. 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. Science Translational Medicine 2013; 5: 208ra149.
5. Lagier-Tourenne C, Baughn M, Rigo F, Sun S, Liu P, Li HR, et al. Targeted degradation of sense and antisense C9orf72 RNA foci as therapy for ALS and frontotemporal degeneration. Proc Natl Acad Sci U S A 2013; 110:E4530–E4539. doi: 10.1073/pnas.1318835110 PMID: 24170860
6. Mizielinska S, Lashley T, Norona FE, Clayton EL, Ridler CE, Fratta P, et al. C9orf72 frontotemporal lobar degeneration is characterised by frequent neuronal sense and antisense RNA foci. Acta Neuropathol. 2013; 126: 845–857. doi: 10.1007/s00401-013-1200-z PMID: 24170096
7. Lee YB, Chen HJ, Peres JN, Gomez-Deza J, Attig J, Stalekar M, et al. Hexanucleotide Repeats in ALS/FTD Form Length-Dependent RNA Foci, Sequester RNA Binding Proteins, and Are Neurotoxic. Cell Rep. 2013; 5: 1178–1186. doi: 10.1016/j.celrep.2013.10.049 PMID: 24290757
8. Cooper-Knock J, Walsh MJ, Higginbottom A, Highley JR, Dickman MJ, Edbauer D, et al. Sequestration of multiple RNA Recognition Motif-containing proteins by C9ORF72 repeat expansions. Brain 2014; 137:2040–51. doi: 10.1093/brain/awu120 PMID: 24866055
9. Ash PE, Bieniek KF, Gendron TF, Caulfield T, Lin WL, Dejesus-Hernandez M, et al. Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. Neuron 2013; 77: 639–646. doi: 10.1016/j.neuron.2013.02.004 PMID: 23415312
10. Mackenzie IR, Arzberger T, Kremmer E, Troost D, Lorenzl S, Mori K, et al. Dipeptide repeat protein pathology in C9ORF72 mutation cases: clinico-pathological correlations. Acta Neuropathol. 2013; 126: 859–879. doi: 10.1007/s00401-013-1181-y PMID: 24096617
11. Mori K, Weng SM, Arzberger T, May S, Rentzsch K, Kremmer E, et al. The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. Science 2013; 339: 1335–1338. doi: 10.1126/science.1232927 PMID: 23393093
12. Cooper-Knock J, Shaw PJ, Kirby J The widening spectrum of C9ORF72-related disease; genotype/phenotype correlations and potential modifiers of clinical phenotype. Acta Neuropathol. 2014; 127:333–45. doi: 10.1007/s00401-014-1251-9 PMID: 24493408
13. Pickrell JK, Pai AA, Gilad Y, Pritchard JK Noisy splicing drives mRNA isoform diversity in human cells. PLoS Genet. 2014; 6: e1001236.
14. Parakh S, Spencer DM, Halloran MA, Soo KY, Atkin JD Redox regulation in amyotrophic lateral sclerosis. Oxid Med Cell Longev 2013; 2013: 408681. doi: 10.1155/2013/408681 PMID: 23533690
15. Jawaid A, Salamone AR, Strutt AM, Murthy SB, Wheaton M, McDowell EJ, et al. ALS disease onset may occur later in patients with pre-morbid diabetes mellitus. Eur J Neurol. 2010; 17: 733–739. doi: 10.1111/j.1468-1331.2009.02923.x PMID: 20074230

16. Wang W, Bu B, Xie M, Zhang M, Yu Z, Tao D. Neural cell cycle dysregulation and central nervous system diseases. Prog Neurobiol. 2009; 89: 1–17. doi: 10.1016/j.pneurobio.2009.01.007 PMID: 19619927
17. 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: 518–530. doi: 10.1038/nrneurol.2012.156 PMID: 22890216
18. Matus S, Valenzuela V, Medinas DB, Hetz C ER Dysfunction and Protein Folding Stress in ALS. Int J Cell Biol 2013; 2013: 674751. doi: 10.1155/2013/674751 PMID: 24324498
19. Vijayalakshmi K, Alladi PA, Ghosh S, Prasanna VK, Sagar BC, Nalini A, et al. Evidence of endoplasmic reticular stress in the spinal motor neurons exposed to CSF from sporadic amyotrophic lateral sclerosis patients. Neurobiol Dis. 2011; 41: 695–705. doi: 10.1016/j.nbd.2010.12.005 PMID: 21168498
20. Rohrl C, Eigner K, Winter K, Korbelius M, Obrowsky S, Kratky D, et al. Endoplasmic reticulum stress impairs cholesterol efflux and synthesis in hepatic cells. J Lipid Res. 2014; 55: 94–103. doi: 10.1194/jlr.M043299 PMID: 24179149
21. Raman R, Allen SP, Goodall EF, Kramer S, Ponger LL, Heath PR, et al. Gene expression signatures in motor neuron disease fibroblasts reveal dysregulation of metabolism, hypoxia-response and RNA processing functions. Neuropathol Appl Neurobiol. 2015; 41:201–226 doi: 10.1111/nan.12147 PMID: 24750211
22. Blokhuis AM, Groen EJ, Koppers M, van den Berg LH, Pasterkamp RJ Protein aggregation in amyotrophic lateral sclerosis. Acta Neuropathol. 2013; 125: 777–794. doi: 10.1007/s00401-013-1125-6 PMID: 23673820
23. Johnson JO, Mandrioli J, Benatar M, Abramzon Y, Van Deerlin VM, Trojanowski JQ, et al.Exome sequencing reveals VCP mutations as a cause of familial ALS. Neuron 2010; 68: 857–864. doi: 10.1016/j.neuron.2010.11.036 PMID: 21145000
24. Deng HX, Chen W, Hong ST, Boycott KM, Gorrie GH, Siddique N, et al. Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia. Nature 2011; 477: 211–215. doi: 10.1038/nature10353 PMID: 21857683
25. Sathasivam S, Shaw PJ Apoptosis in amyotrophic lateral sclerosis—what is the evidence? Lancet Neurol. 2005; 4: 500–509. PMID: 16033692
26. Mori K, Lammich S, Mackenzie IR, Forne I, Zilow S, Kretzschmar H, et al. hnRNP A3 binds to GGGGCC repeats and is a constituent of p62-positive/TDP43-negative inclusions in the hippocampus of patients with C9orf72 mutations. Acta Neuropathol. 2013; 125: 413–423. doi: 10.1007/s00401-013-1088-7 PMID: 23381195
27. Haeusler AR, Donnelly CJ, Periz G, Simko EA, Shaw PG, Kim MS, et al. C9orf72 nucleotide repeat structures initiate molecular cascades of disease. Nature 2014; 507: 195–200. doi: 10.1038/nature13124 PMID: 24598541
28. Buchman VL, Cooper-Knock J, Connor-Robson N, Higginbottom A, Kirby J, Razinskaya OD, et al. Simultaneous and independent detection of C9ORF72 alleles with low and high number of GGGGCC repeats using an optimised protocol of Southern blot hybridisation. Mol Neurodegener. 2013; 8: 12. doi: 10.1186/1750-1326-8-12 PMID: 23566336
29. Ismail A, Cooper-Knock J, Highley JR, Milano A, Kirby J, Goodall E, et al. Concurrence of multiple sclerosis and amyotrophic lateral sclerosis in patients with hexanucleotide repeat expansions of C9ORF72. J Neurol Neurosurg Psychiatry 2013; 84: 79–87. doi: 10.1136/jnnp-2012-303326 PMID: 23085936
30. Ince PG, McArthur FK, Bjertness E, Torvik A, Candy JM, Edwardson JA. Neuropathological diagnoses in elderly patients in Oslo: Alzheimer's disease, Lewy body disease, vascular lesions. Dementia 1995; 6: 162–168. PMID: 7620529
31. Ferraiuolo L, Heath PR, Holden H, Kasher P, Kirby J, Shaw PJ. Microarray analysis of the cellular pathways involved in the adaptation to and progression of motor neuron injury in the SOD1 G93A mouse model of familial ALS. J Neurosci. 2007; 27: 9201–9219. PMID: 17715356
32. Pearson RD, Liu X, Sanguinetti G, Milo M, Lawrence ND, Rattray M. puma: a Bioconductor package for propagating uncertainty in microarray analysis. BMC Bioinformatics 2009; 10: 211. doi: 10.1186/1471-2105-10-211 PMID: 19589155
33. Rattray M, Liu X, Sanguinetti G, Milo M, Lawrence ND Propagating uncertainty in microarray data analysis. Brief Bioinform. 2006; 7: 37–47. PMID: 16761363
34. Ghazalpour A, Doss S, Zhang B, Wang S, Plaisier C, Castellanos R, et al. Integrating genetic and network analysis to characterize genes related to mouse weight. PLoS Genet. 2006; 2: e130. PMID: 16934000
35. Langfelder P, Horvath S WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 2008; 9: 559. doi: 10.1186/1471-2105-9-559 PMID: 19114008

36. Huang DW, Sherman BT, Lempicki RA Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protocols 2008; 4: 44–57.

37. Huang DW, Sherman BT, Lempicki RA Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Research 2009; 37: 1–13. doi: 10.1093/nar/gkn923 PMID: 19033363

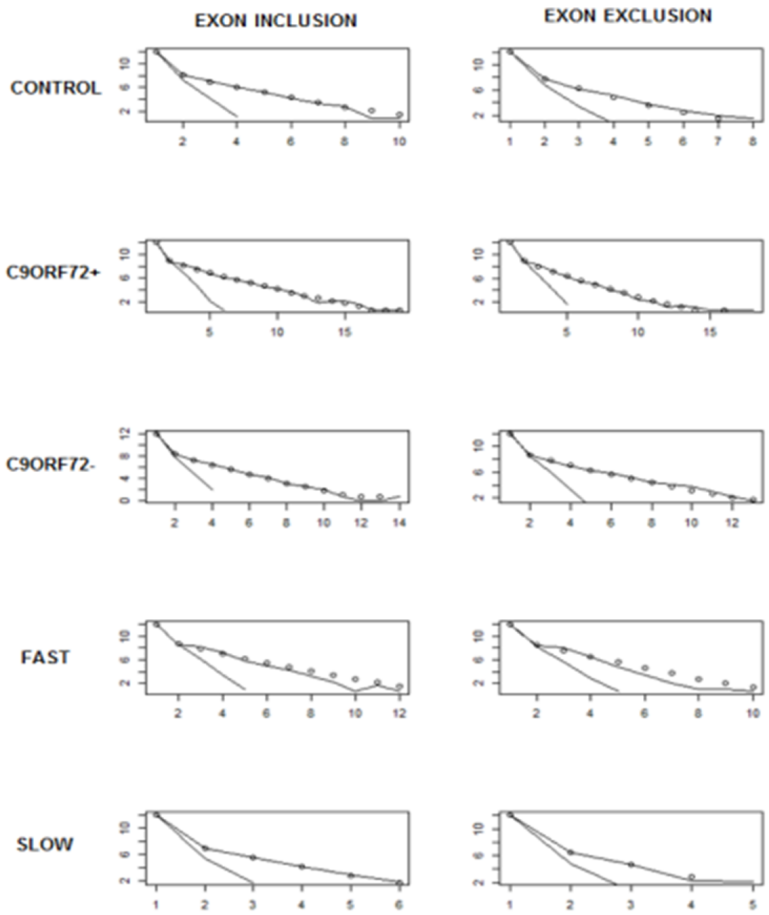
38. Purdom E, Simpson KM, Robinson MD, Conboy JG, Lapuk AV, Speed TP. FIRMA: a method for detection of alternative splicing from exon array data. Bioinformatics 2008; 24: 1707–1714. doi: 10.1093/bioinformatics/btn284 PMID: 18573797

39. Bengtsson H, Simpson K, Bullard J, Hansen K. aroma.affymetrix: A generic framework in R for analyzing small to very large Affymetrix data sets in bounded memory. 2008; Tech Report #745, Department of Statistics, University of California, Berkeley.

40. Sveen A, Agesen TH, Nesbakken A, Rognum TO, Lothe RA, Skotheim RI. Transcriptome instability in colorectal cancer identified by exon microarray analyses: Associations with splicing factor expression levels and patient survival. Genome Med. 2011; 3: 32. doi: 10.1186/gm248 PMID: 21619627

41. Kirby J, Halligan E, Baptista MJ, Allen S, Heath PR, Holden H, et al. Mutant SOD1 alters the motor neuronal transcriptome: implications for familial ALS. Brain 2005; 128: 1686–1706. PMID: 15872021

S1_Supplementary Figure



S1_Fig.tif
Plots of the number of splicing events (y-axis) which were present in a given number of lymphoblastoid cell lines (x-axis) within a particular sample group. Sample groups from top to bottom are: normal controls, C9ORF72-ALS patients, non-C9ORF72 ALS patients, C9ORF72-ALS patients with survival C9ORF72-ALS patients with survival > 4 years. In each plot the left-hand line represents a Poisson fit to the observed data i.e. the random case. The right-hand line is the observed data and the dotted line represents the negative binomial distribution fit to the observed data. In each case the negative binomial provides a relatively good fit to the observed data. θ as shown in Fig 4, is a quantification of the overdispersion in the negative binomial compared to the Poisson fit to the observed data i.e. the degree of consistency in the splicing observed in each sample group.

S1_Table.xlsx
Genes listed are within the GO category ‘RNA Processing’ and are significantly associated (pC9ORF72+ lymphoblastoid cell lines or motor neurons. A fold change of >1 equates to up-regulation and a fold change of

S1 Table

Gene Symbol	Lymphoblastoid Cell Lines		Motor Neurons	
	Transcript ID	Fold Change (C9ORF72+ Vs Control)	Transcript ID	Fold Change (C9ORF72+ Vs Control)
A1CF	ENST00000373995	1.24		
A2BP1	ENST00000340209	0.41		
ABCB5	ENST00000441315	1.19	1555371_at	0.43
	ENST00000406935	1.22		
ADAD2	ENST00000268624	0.93		
	ENST00000315906	2.78		
ADAR	ENST00000492630	1.34		
ADARB1	ENST00000481022	0.74		
ADAT2	ENST00000342031	1.73		
	ENST00000367594	1.16		
ADAT3	ENST00000454697	1.84		
APP	ENST00000348990	5.45		
	ENST00000359726	5.54		
	ENST00000466453	0.38		
	ENST00000464867	0.73		
	ENST00000491395	0.69		
ARL6IP4	ENST00000315580	1.18		
ATXN1	ENST00000483591	1.12		
	ENST00000483954	1.43		
BAT1	ENST00000462421	0.81		
	ENST00000458215	0.67		
C1D	ENST00000407324	1.07		
	ENST00000470189	0.86		
C1DP2	ENST00000407324	1.07		
	ENST00000470189	0.86		
C1DP3	ENST00000407324	1.07		
	ENST00000470189	0.86		
C1ORF66	ENST00000462397	0.81		
CCAR1	ENST00000480627	2.43		
CDK5RAP1	ENST00000473791	0.82		
	ENST00000375351	1.99		
CDKN2A	ENST00000304494	1.12		
	ENST00000470819	1.09		
	ENST00000380150	0.88		
CELF1	ENST00000395292	1.16		
CELF6	ENST00000287202	0.86		
CPSF3	ENST00000489403	1.22		
	ENST00000458452	1.45		
	ENST00000430786	0.4		
	ENST00000421495	5.21		
	ENST00000435064	4.64		

CPSF3L	ENST00000496353	0.85		
	ENST00000450926	4.03		
	ENST00000467408	0.62		
	ENST00000419704	4.08		
	ENST00000478641	0.79		
	ENST00000294579	4.4		
CPSF4	ENST00000465132	1.31		
	ENST00000412686	1.1		
CPSF7	ENST00000413232	2.14	217866_at	1.53
CRNKL1	ENST00000490258	0.82		
CSDC2	ENST00000460790	1.22		
CSTF1	ENST00000428552	2.01		
CSTF2	ENST00000415585	0.8		
DDX1	ENST00000478695	1.95		
DDX17	ENST00000479734	1.15	208718_at	1.32
	ENST00000475004	1.77		
	ENST00000477112	1.13		
DDX56	ENST00000448192	1.13		
	ENST00000473924	2.05		
DHX16	ENST00000480966	1.09		
DHX35	ENST00000449559	1.77		
DHX9	ENST00000483416	1.17		
DIS3	ENST00000475871	1.17		
DKC1	ENST00000492372	2.03		
	ENST00000475423	1.22		
	ENST00000413910	2.08		
DUS3L	ENST00000428664	2.01		
ELAC2	ENST00000426905	9.67		
	ENST00000476042	0.48		
	ENST00000484122	1.15		
	ENST00000407336	0.35		
	ENST00000427703	4.86		
	ENST00000395962	9.49		
	ENST00000491478	0.75		
	ENST00000338034	20.7		
ELAVL4	ENST00000474675	1.2		
EXOSC1	ENST00000489158	1.15		
EXOSC10	ENST00000469634	0.89		
EXOSC3	ENST00000482614	0.83		
EXOSC7	ENST00000459856	1.79		
	ENST00000461361	0.88		
EXOSC8	ENST00000389704	1.18		
	ENST00000488779	2.55		
	ENST00000488108	0.63		
	ENST00000379809	2.71		
	ENST00000495092	4.32		
FARS2	ENST00000274680	0.81		
FTSJ1	ENST00000485486	2.26		
	ENST00000466371	1.86		

	ENST00000487353	0.89		
FTSJ2	ENST00000486040	1.1		
FUS	ENST00000491029	2.76	215744_at	0.39
	ENST00000483853	1.18		
	ENST00000474990	1.31		
GEMIN6	ENST00000281950	1.23		
GTF2F2	ENST00000461904	1.15		
HNRNPA2B1	ENST00000495810	1.21		
	ENST00000490912	1.13		
HNRNPA3	ENST00000432457	1.99		
HNRNPC	ENST00000336053	2.5		
	ENST00000320084	2.5		
HNRNPF	ENST00000437529	3.39		
HNRNPH2	ENST00000457902	2.19	201406_at	1.88
HNRNPH3	ENST00000467249	1.92	208990_s_at	2.16
	ENST00000480987	2.73		
	ENST00000469172	1.17		
HNRNPK	ENST00000481820	1.24		
HNRNPL	ENST00000423415	1.19		
HNRNPM	ENST00000348943	3.89	233605_x_at	0.73
HNRNPR	ENST00000476660	0.84		
	ENST00000464516	1.46		
	ENST00000478691	1.15		
HNRNPU	ENST00000283179	4.3		
	ENST00000465881	1.22		
	ENST00000413625	0.66		
	ENST00000483966	1.32		
HNRNPUL1	ENST00000270069	1.13		
	ENST00000392006	1.21		
	ENST00000263367	1.21		
HNRPLL	ENST00000410076	1.14		
HSD17B10	ENST00000495986	1.2		
HSD3B7	ENST00000454009	0.93		
IMP4	ENST00000462392	2.96		
INTS12	ENST00000493425	1.26		
INTS4	ENST00000433818	1.12		
INTS6	ENST00000440606	1.11		
IVNS1ABP	ENST00000468217	1.18		
	ENST00000392007	0.5		
	ENST00000494880	1.23		
KRR1	ENST00000229214	1.14	203203_s_at	2.13
LARP6	ENST00000344870	1.1		
LARP7	ENST00000344442	1.2		
LOC100130109	ENST00000469451	1.79		
LOC643167	ENST00000492779	0.74	227223_at	2.44
	ENST00000427743	1.93		
LOC644063	ENST00000481820	1.24		
LOC644390	ENST00000423415	1.19		
LOC644422	ENST00000328748	1.31		

LOC652147	ENST00000429650	0.72		
	ENST00000480835	4.3		
LOC653155	ENST00000490653	0.61	202126_at	1.87
	ENST00000489943	0.31	237093_at	0.4
LOC653884	ENST00000341154	1.22		
MAGOH	ENST00000495868	1.25		
	ENST00000371470	1.16		
MBNL1	ENST00000493459	2.16	1558111_at	0.5
	ENST00000477171	1.43	1556658_a_at	0.41
	ENST00000478535	0.32		
	ENST00000460166	0.3		
MLH1	ENST00000476172	0.75		
	ENST00000435176	3.79		
	ENST00000432299	3.05		
MOV10	ENST00000471160	2.69		
	ENST00000486416	1.13		
MPHOSPH10	ENST00000468427	0.5		
	ENST00000493360	1.16		
	ENST00000476969	1.37		
	ENST00000244230	1.1		
MTO1	ENST00000462039	0.73	224430_s_at	2.26
	ENST00000445187	1.22		
	ENST00000487960	2.92		
MTPAP	ENST00000488290	1.08		
NCBP1	ENST00000375147	1.15		
NCBP2	ENST00000447325	2.3		
	ENST00000482976	3.18		
NHP2L1	ENST00000488571	1.14		
NOLC1	ENST00000476468	2.07		
NONO	ENST00000472185	0.17		
	ENST00000490044	1.27		
	ENST00000450092	1.23		
	ENST00000373856	2.4		
NOP14	ENST00000398071	1.29		
NOP56	ENST00000462630	1.14		
	ENST00000470143	1.12		
NOP58	ENST00000478508	1.31		
NPM3	ENST00000468544	1.91		
PA2G4	ENST00000303305	1.12		
PA2G4P4	ENST00000303305	1.12		
PABPC4	ENST00000492468	2.63		
	ENST00000372862	3.25		
	ENST00000451091	0.5		
	ENST00000492519	0.81		
	ENST00000470443	0.56		
	ENST00000461578	0.3		
PABPN1	ENST00000397276	1.58		
PCBP2	ENST00000359282	1.43		
	ENST00000359462	1.43		

	ENST00000439930	1.43		
PDCD11	ENST00000490787	1.91		
	ENST00000466959	0.89		
PDCD7	ENST00000380204	2.28		
PES1	ENST00000433575	2.42		
PIWIL2	ENST00000454009	0.93		
POLR2B	ENST00000438333	3.73		
	ENST00000484821	1.68		
POLR2D	ENST00000487079	1.37	203664_s_at	0.75
POLR2H	ENST00000476003	1.11		
	ENST00000490958	1.13		
	ENST00000460083	1.34		
PPAN	ENST00000486482	0.85		
	ENST00000430370	1.32		
	ENST00000446223	1.32		
PPIE	ENST00000427454	1.39		
	ENST00000467741	0.42		
	ENST00000482751	1.19		
PPIG	ENST00000418888	1.16		
PPIL1	ENST00000483552	1.27		
PPP1R8	ENST00000486634	0.65		
	ENST00000236412	1.19		
PPP2R1A	ENST00000473455	2.09		
	ENST00000454220	1.32		
	ENST00000473820	1.47		
PPP4R2	ENST00000482242	1.98	226317_at	1.5
	ENST00000460360	1.19		
PRKRA	ENST00000490501	0.77		
	ENST00000463882	1.13		
PRMT5	ENST00000421938	1.31		
PRPF18	ENST00000298451	1.75		
	ENST00000417658	1.24		
PRPF3	ENST00000476970	0.68		
	ENST00000470824	0.86		
PRPF31	ENST00000321030	1.19		
	ENST00000467851	1.75		
PRPF38B	ENST00000467302	1.26		
PRPF4	ENST00000488937	1.17		
PRPF40A	ENST00000359961	11.19	218053_at	1.26
	ENST00000450135	5.78		
	ENST00000493468	0.38		
	ENST00000486100	0.4		
	ENST00000440252	23.92		
PRPF4B	ENST00000490653	0.61	202126_at	1.87
	ENST00000489943	0.31	237093_at	0.4
PTBP2	ENST00000482253	1.22		
PUS10	ENST00000421319	1.21		
PUS7	ENST00000487277	1.2		
QTRTD1	ENST00000488665	1.6		

QTRTD1	ENST00000466870	0.79		
RALY	ENST00000413297	1.22		
	ENST00000481580	0.87		
	ENST00000489384	1.08		
	ENST00000246194	3.7		
RBM14	ENST00000409406	0.86	236157_at	0.59
RBM16	ENST00000479234	1.13		
RBM17	ENST00000465906	2.35	224780_at	1.37
	ENST00000418631	1.12		
	ENST00000467214	0.67		
	ENST00000418869	0.62		
	ENST00000481147	0.68		
	ENST00000446108	1.09		
RBM28	ENST00000223073	1.25		
	ENST00000415472	0.79		
RBM3	ENST00000490127	2.48		
	ENST00000472897	1.25		
RBM39	ENST00000492779	0.74	227223_at	2.44
	ENST00000427743	1.93		
RBM4	ENST00000409406	0.86	236157_at	0.59
RBM5	ENST00000395174	3.76		
	ENST00000471995	0.65		
	ENST00000441812	2.96		
	ENST00000404526	1.8		
	ENST00000489437	0.64		
RBM6	ENST00000493652	1.15	201967_at	1.4
	ENST00000483350	1.15		
RBM8A	ENST00000369307	2.61		
RBM9	ENST00000473487	0.88		
RBMS1	ENST00000492283	0.71	225265_at	1.75
	ENST00000494308	0.2	238185_at	2.49
	ENST00000474820	5.03		
	ENST00000490637	0.88		
	ENST00000464200	0.45		
	ENST00000477965	8.28		
ROD1	ENST00000450374	1.99		
RP9	ENST00000492391	1.36		
RPL11	ENST00000374550	1.98		
RPL14	ENST00000465280	11.32		
	ENST00000477056	1.14		
	ENST00000461368	2.28		
RPL35A	ENST00000448864	0.61		
	ENST00000464167	1.45		
	ENST00000442341	1.28		
	ENST00000496582	3.6		
RPL36A	ENST00000457902	2.19	201406_at	1.88
RPL5	ENST00000497519	1.29		
RPL7	ENST00000396466	6.07		
	ENST00000396465	1.14		

	ENST00000431653	1.67		
RPP14	ENST00000462046	0.42		
	ENST00000477305	2.84		
	ENST00000461393	3.35		
	ENST00000481972	2.34		
	ENST00000474660	2.53		
RPP30	ENST00000480635	0.87		
RPS16	ENST00000339471	1.17	213890_x_at	1.79
			226131_s_at	1.6
RPS24	ENST00000435275	0.48		
RPS6	ENST00000380381	1.27		
	ENST00000380384	1.09		
RPS7	ENST00000304921	1.15		
	ENST00000479123	1.81		
RRAGC	ENST00000474456	1.24		
RRP1	ENST00000473988	0.64		
	ENST00000475534	2.48		
RSRC1	ENST00000494002	1.13		
RTCD1	ENST00000260563	0.83		
SBDS	ENST00000496763	2.02		
SCNM1	ENST00000461862	1.35		
	ENST00000471039	1.35		
SETX	ENST00000419901	1.17	232229_at	0.61
SF1	ENST00000448404	6.38		
	ENST00000496969	0.41		
	ENST00000477596	0.85		
	ENST00000416674	0.63		
	ENST00000413951	0.35		
	ENST00000377387	1.22		
	ENST00000377394	5.41		
	ENST00000422298	4.67		
	ENST00000463343	2.29		
	ENST00000334944	12.45		
	ENST00000489544	0.11		
SF3A1	ENST00000485618	1.51	201356_at	1.39
	ENST00000444440	1.39		
SF3A3	ENST00000488934	1.14		
	ENST00000487062	1.21		
	ENST00000470585	1.19		
SF3B1	ENST00000462613	1.89		
	ENST00000479532	0.83		
	ENST00000409915	1.28		
SF3B2	ENST00000322535	1.12		
SF3B4	ENST00000271628	1.08	209044_x_at	1.54
SFPQ	ENST00000485365	1.73	226898_s_at	0.43
	ENST00000485454	1.3	214016_s_at	1.21
	ENST00000470472	1.98		
	ENST00000471991	1.22		
	ENST00000466745	0.67		

SFRS1	ENST00000423765	1.21	211784_s_at	1.27
SFRS11	ENST00000454435	0.4		
	ENST00000429640	3.05		
	ENST00000475204	0.65		
	ENST00000463116	0.37		
	ENST00000370950	1.3		
	ENST00000469170	1.94		
	ENST00000395136	0.71		
SFRS13A	ENST00000341154	1.22		
SFRS13B	ENST00000488604	1.45		
	ENST00000475068	1.12		
	ENST00000452027	0.87		
SFRS15	ENST00000467731	1.13	222310_at	2.84
SFRS17A	ENST00000313871	3.18		
	ENST00000440218	0.63		
SFRS4	ENST00000497015	1.62		
	ENST00000477193	1.08		
SFRS5	ENST00000394366	1.15		
SFRS6	ENST00000328748	1.31		
SFRS7	ENST00000378869	0.46		
	ENST00000415527	1.23		
SLBP	ENST00000489418	1.16		
SMC1A	ENST00000322213	1.29	201589_at	4.17
SMNDC1	ENST00000471297	1.12		
SNRNP200	ENST00000429650	0.72		
	ENST00000480835	4.3		
SNRNP35	ENST00000441315	1.19	1555371_at	0.43
	ENST00000406935	1.22		
SNRNP48	ENST00000397475	2.58		
	ENST00000496946	1.13		
SNRPB	ENST00000381342	2.86	208821_at	4.18
SNRPE	ENST00000469451	1.79		
SNRPEL1	ENST00000469451	1.79		
SNRPN	ENST00000346403	1.25		
SPOP	ENST00000347630	1.1		
SR140	ENST00000485374	3.31		
	ENST00000461941	87.15		
	ENST00000461844	1.7		
	ENST00000493598	4.14		
	ENST00000496100	4.65		
SRPK2	ENST00000465112	1.27		
	ENST00000462282	1.14		
	ENST00000493638	1.21		
SRRM1	ENST00000471693	0.9		
	ENST00000485541	1.29		
	ENST00000474843	0.75		
SRRT	ENST00000445337	2.15	214127_s_at	0.33
	ENST00000487311	1.35		
	ENST00000466432	0.84		

	ENST00000449389	1.54		
	ENST00000474896	0.86		
SSB	ENST00000474273	1.41		
	ENST00000461708	0.81		
	ENST00000470621	1.15		
SYF2	ENST00000476231	0.85		
SYNCRIP	ENST00000449619	1.41	236146_at	2.42
			217834_s_at	1.32
TARBP2	ENST00000266987	1.16	200020_at	1.31
TARDBP	ENST00000480464	1.51		
TFIP11	ENST00000422283	2.16		
	ENST00000450493	0.21		
	ENST00000492137	0.87		
	ENST00000481357	0.39		
	ENST00000407431	1.11		
THOC2	ENST00000441692	1.2		
	ENST00000432353	3.09		
	ENST00000496830	0.6		
	ENST00000419633	0.58		
	ENST00000492203	0.21		
	ENST00000464604	0.68		
TRA2A	ENST00000482395	0.91		
TRA2B	ENST00000453386	1.19		
	ENST00000342294	1.62		
	ENST00000485530	2.63		
TRMT2B	ENST00000372939	1.31		
	ENST00000490509	0.49		
TRMT6	ENST00000473131	1.09		
TRMT61B	ENST00000484060	0.54		
TRMU	ENST00000424260	1.55		
TRPT1	ENST00000394547	1.31		
TRUB2	ENST00000461180	1.23		
	ENST00000372890	1.13		
TSEN15	ENST00000462677	1.54		
	ENST00000367518	0.8		
	ENST00000485209	1.08		
TSEN54	ENST00000434205	3.02		
TTF2	ENST00000463696	3.79		
TYW1	ENST00000491969	16.5		
TYW3	ENST00000486467	1.1		
U2AF1	ENST00000478282	1.07		
UPF3B	ENST00000276201	1.59		
URM1	ENST00000372847	1.3		
USP39	ENST00000409470	5.52		
	ENST00000409025	3.77		
	ENST00000323701	3.65		
	ENST00000481409	0.03		
	ENST00000451004	2.59		
	ENST00000450066	3.24		

	ENST00000474572	0.1		
	ENST00000459775	8.41		
	ENST00000467885	0.33		
UTP11L	ENST00000486563	0.74		
UTP14A	ENST00000498179	1.1	221098_x_at	0.53
UTP6	ENST00000490218	1.3		
WDR12	ENST00000478869	2.4		
WDR4	ENST00000330317	1.18		
WDR77	ENST00000459665	1.16		
XRN2	ENST00000430571	1.35	208627_s_at	0.53
YBX1	ENST00000436427	1.12		
	ENST00000318612	1.79		
	ENST00000332220	1.22		
YTHDC1	ENST00000344157	1.18		
ZCCHC11	ENST00000466440	0.94		
	ENST00000470212	2.55		
ZCCHC6	ENST00000375947	1.15	244499_at	0.45
ZFC3H1	ENST00000308101	1.29		
ZMAT5	ENST00000397781	0.44		
ZNF638	ENST00000487638	1.38		
	ENST00000492262	1.37		
	ENST00000411930	0.39		
	ENST00000494621	1.49		
	ENST00000460310	1.35		
	ENST00000417778	1.31		
ZRANB2	ENST00000370920	1.19	213876_x_at	2.22
ZRSR2	ENST00000380308	1.19		
	ENST00000468028	1.48		
BICD1			242052_at	2.67
CELF2			1565599_at	0.29
			214405_at	1.52
CELF3			206817_x_at	0.34
CLP1			204370_at	1.92
CNOT6L			226153_s_at	2.23
			238438_at	2.15
DDX20			224315_at	1.52
DDX46			229098_s_at	2.03
DEDD2			225434_at	1.28
HNRNPA1			214280_x_at	1.28
			200016_x_at	1.43
INTS2			224308_s_at	1.68
JMJD6			215233_at	2.64
			212723_at	1.58
KIAA1429			223110_at	2.2
LOC643446			207939_x_at	2.68
LOC644037			214280_x_at	1.28
			200016_x_at	1.43
LSM7			204559_s_at	2.31
LUC7L3			220044_x_at	1.37

NOVA1			207437_at	1.98
NSUN2			223076_s_at	1.36
PABPC1			215157_x_at	1.35
PAPOLA			228569_at	1.6
PPARGC1A			219195_at	2.49
PTBP1			211271_x_at	3.97
RBM15B			226987_at	2.3
RBM22			236872_at	0.47
RBM25			212030_at	1.82
RBMS2			235558_at	0.37
RNPS1			207939_x_at	2.68
RPS17			211487_x_at	1.92
			212578_x_at	1.76
			201665_x_at	2.61
RPS19			202649_x_at	1.66
RRP1B			212844_at	0.47
SFRS3			202899_s_at	2.35
SMAD1			208015_at	0.34
TGS1			236371_s_at	0.41
THOC1			204064_at	2.71