

ARTICLE

Received 15 Sep 2015 | Accepted 3 Jun 2016 | Published 5 Jul 2016

DOI: 10.1038/ncomms12143

OPEN

# Distinct and shared functions of ALS-associated proteins TDP-43, FUS and TAF15 revealed by multisystem analyses

Katannya Kapeli<sup>1,2,3,\*</sup>, Gabriel A. Pratt<sup>1,2,4,\*</sup>, Anthony Q. Vu<sup>1,2</sup>, Kasey R. Hutt<sup>1,2</sup>, Fernando J. Martinez<sup>1,2</sup>, Balaji Sundararaman<sup>1,2</sup>, Ranjan Batra<sup>1,2,5</sup>, Peter Freese<sup>6</sup>, Nicole J. Lambert<sup>6</sup>, Stephanie C. Huelga<sup>1,2,4</sup>, Seung J. Chun<sup>7</sup>, Tiffany Y. Liang<sup>1,2</sup>, Jeremy Chang<sup>1,2</sup>, John P. Donohue<sup>8</sup>, Lily Shiue<sup>8</sup>, Jiayu Zhang<sup>9</sup>, Haining Zhu<sup>9</sup>, Franca Cambi<sup>10</sup>, Edward Kasarskis<sup>10</sup>, Shawn Hoon<sup>11</sup>, Manuel Ares Jr.<sup>8</sup>, Christopher B. Burge<sup>6</sup>, John Ravits<sup>5</sup>, Frank Rigo<sup>7</sup> & Gene W. Yeo<sup>1,2,3,4,11</sup>

The RNA-binding protein (RBP) TAF15 is implicated in amyotrophic lateral sclerosis (ALS). To compare TAF15 function to that of two ALS-associated RBPs, FUS and TDP-43, we integrate CLIP-seq and RNA Bind-N-Seq technologies, and show that TAF15 binds to ~4,900 RNAs enriched for GGUA motifs in adult mouse brains. TAF15 and FUS exhibit similar binding patterns in introns, are enriched in 3' untranslated regions and alter genes distinct from TDP-43. However, unlike FUS and TDP-43, TAF15 has a minimal role in alternative splicing. In human neural progenitors, TAF15 and FUS affect turnover of their RNA targets. In human stem cell-derived motor neurons, the RNA profile associated with concomitant loss of both TAF15 and FUS resembles that observed in the presence of the ALS-associated mutation FUS R521G, but contrasts with late-stage sporadic ALS patients. Taken together, our findings reveal convergent and divergent roles for FUS, TAF15 and TDP-43 in RNA metabolism.

<sup>1</sup> Department of Cellular and Molecular Medicine, University of California at San Diego, La Jolla, California 92093, USA. <sup>2</sup> Stem Cell Program and Institute for Genomic Medicine, University of California at San Diego, La Jolla, California 92093, USA. <sup>3</sup> Department of Physiology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117549, Singapore. <sup>4</sup> Department of Bioinformatics and Systems Biology, University of California at San Diego, La Jolla, California 92093, USA. <sup>5</sup> Department of Neurosciences, University of California at San Diego, La Jolla, California 92093, USA. <sup>6</sup> Department of Biology, MIT, Cambridge, Massachusetts 02142, USA. <sup>7</sup> Ionis Pharmaceuticals, Carlsbad, California 92010, USA. <sup>8</sup> Department of Molecular, Cell and Developmental Biology, Sinsheimer Labs, University of California, Santa Cruz, California 95064, USA. <sup>9</sup> Department of Molecular and Cellular Biochemistry, College of Medicine, University of Kentucky, Lexington, Kentucky 40536, USA. <sup>10</sup> Department of Neurology, University of Kentucky, Lexington, Kentucky 40536, USA. <sup>11</sup> Molecular Engineering Laboratory, A\*STAR, Singapore 138673, Singapore. \* These authors contributed equally to this work. Correspondence and requests for materials should be addressed to G.W.Y. (email: geneyeo@ucsd.edu).

Amiotrophic lateral sclerosis (ALS) is a fatal disease characterized by progressive degeneration of motor neurons (MNs) in the motor cortex, brainstem and spinal cord. Although the precise pathogenesis of ALS remains unknown, aberrant RNA processing appears to be an important contributing factor. The RNA-binding protein (RBP) TAR DNA-binding protein 43 (TDP-43) was initially recognized as a major constituent of pathological ubiquitinated protein aggregates in the brain and spinal cord tissue of patients with sporadic ALS (sALS)<sup>1,2</sup>. Dominant mutations in *TDP-43* were subsequently identified in ALS patients<sup>3–8</sup> with evidence that these mutations were indeed causative of ALS pathogenesis<sup>9</sup>. Shortly thereafter, mutations in the gene encoding another RBP, fused in sarcoma (*FUS*, also known as translocated in liposarcoma or TLS), were identified in a subset of patients with familial ALS and sALS<sup>10,11</sup>. Although mutations in *FUS* and *TDP-43* are present in only a small fraction of ALS cases, abnormal activity of *FUS* and *TDP-43* is observed in a large fraction of ALS cases.

The discovery of mutations in the genes encoding *TDP-43* and *FUS* received much attention as these proteins have strikingly similar protein domain architectures<sup>12</sup>. This motivated a search for more structurally similar RBPs as candidate ALS genes and, as a result, mutations in TATA box-binding protein (TBP)-associated factor 15 (*TAF15*) were identified in patients with sALS and familial ALS<sup>13,14</sup>. *FUS* and *TAF15* belong to the FET family of heterogeneous nuclear ribonucleoprotein (hnRNP) proteins, which includes Ewing sarcoma breakpoint region 1. As the protein structure of *TAF15* is similar to those of *FUS* and *TDP-43* (ref. 13), it was predicted that *TAF15* would be functionally similar to these RBPs. Similar to *FUS* and *TDP-43*, *TAF15* is predominantly localized to the nucleus but shuttles to and from the cytoplasm, participates in transcription, is thought to affect alternative splicing (AS) and has been found to form cytoplasmic inclusions in all *FUS*-FTLD subtypes and in some sALS patient tissues<sup>13,15,16</sup>.

Another commonality among *TDP-43*, *FUS* and *TAF15* is that the vast majority of ALS-associated mutations identified in the genes encoding these RBPs are found in their C-terminal Gly-rich domains. An emerging hypothesis is that mutations within the Gly-rich region of these RBPs promote their pathological aggregation<sup>17,18</sup>. Aggregation of *FUS*, *TDP-43* and *TAF15* proteins is often accompanied by loss of their nuclear localization; yet it is unclear whether protein aggregation or mislocalization to the cytoplasm is the initiating pathogenic event<sup>19</sup>. In efforts to investigate the normal nuclear function of these RBPs, comprehensive RNA-binding maps of *TDP-43*, *FUS* and *TAF15* in the normal mouse<sup>20–22</sup> or human<sup>15</sup> central nervous system (CNS) have been determined. These studies revealed global roles for *TDP-43*, *FUS* and *TAF15* in AS and motif specificities for *TDP-43* and *FUS* in the CNS. Furthermore, loss of *TDP-43* or *FUS* expression affects the RNA levels of genes containing long introns<sup>20–22</sup>. Our understanding of *FUS*, *TDP-43* and *TAF15* function in RNA processing has primarily come from examining these proteins individually and under different conditions, making comparisons difficult. This approach has limited our understanding of how the activities of these RBPs may converge on common pathways or act in parallel. A systematic comparison of *FUS*, *TDP-43* and *TAF15* to determine their shared and unique functions in mature and developing neurons would be valuable in understanding their contribution to development and ultimately disease.

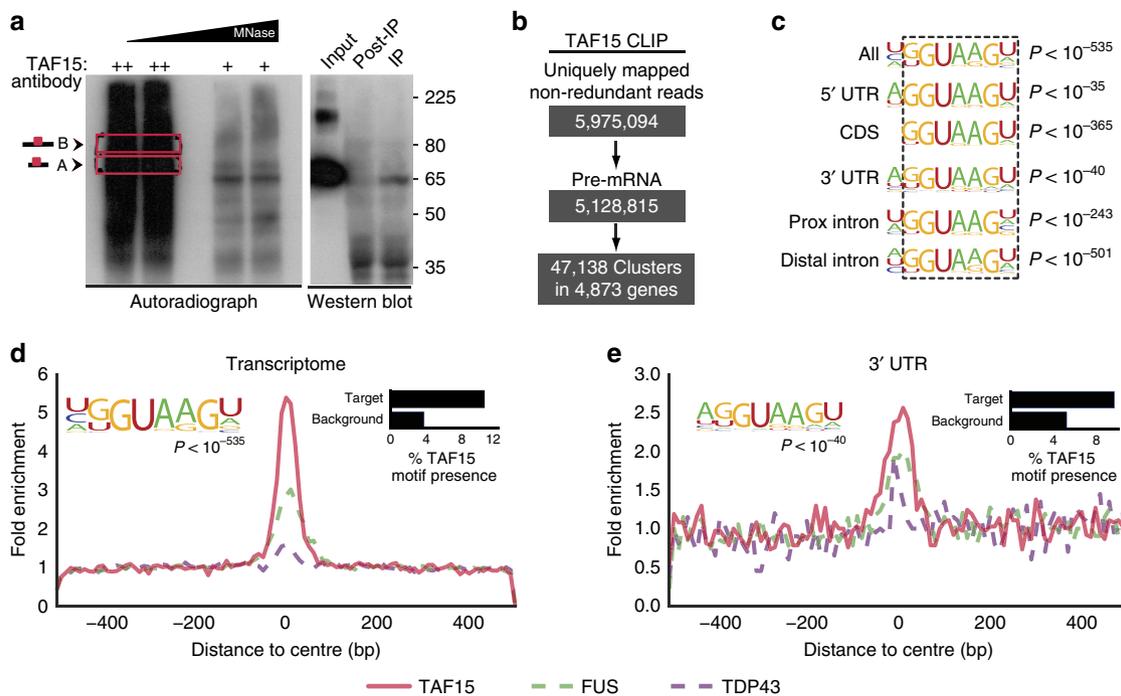
Here we identify 4,873 RNA targets of *TAF15* in the mouse brain that reveal a *TAF15*-binding motif. Expanding on our previous studies<sup>21,22</sup>, we find that *FUS* and *TAF15* exhibit similar global RNA interaction profiles *in vivo*, but affect a strikingly small subset of common genes. Unexpectedly, *TAF15* influences a

small fraction of AS events compared with *TDP-43* and *FUS* in the mouse CNS. In human neural progenitor cells (NPCs), we find that *TAF15* and *FUS* affect the stability of distinct mRNA populations, many of which are bound by *TAF15* and *FUS*. Depletion of *TAF15*, *FUS* and *TDP-43* in human-induced pluripotent stem cell (iPSC)-derived MNs also affects different genes. Subsets of *TAF15* and *FUS*-regulated mRNAs, including ALS-associated genes, are also differentially expressed in spinal cord MNs dissected from sALS patients and iPSC-derived MNs from ALS patients harbouring a *R521G* mutation in *FUS*. Taken together, these findings uncover points of functional convergence and divergence of *FUS*, *TAF15* and *TDP-43*.

## Results

**TAF15 binds RNAs enriched for GGUAAGU motifs *in vivo*.** To identify *in vivo* RNA substrates recognized by *TAF15*, we performed CLIP (crosslinking immunoprecipitation)-seq in whole-brain tissue from adult mice using a commercially available antibody that specifically recognizes the N terminus of the *TAF15* protein. We isolated RNA from low and high molecular weight *TAF15* protein–RNA complexes (Fig. 1a, bands A and B, respectively) and converted the RNA into sequencing libraries for transcript identification. No protein–RNA complexes were immunoprecipitated when using nonspecific IgG or in the absence of ultraviolet crosslinking (Supplementary Fig. 1a). Interactions between *TAF15* and *FUS* have previously been detected<sup>23,24</sup>. Therefore, we tested whether *TAF15* and *FUS* interact post-cell lysis. Uniquely tagged versions of *TAF15* and *FUS* proteins were expressed separately in HEK293T cells, and upon mixing lysates from these cell lines we found that V5-tagged *TAF15* immunoprecipitates Myc-tagged *FUS* (Supplementary Fig. 1b, lane 10) and *vice versa* (Supplementary Fig. 1b, lane 14). This demonstrated that *TAF15* and *FUS* can physically associate post-cell lysis. For our *TAF15* CLIP-seq experiments, the use of ultraviolet-crosslinked cells and highly stringent lysis and wash conditions prevented co-immunoprecipitation of *FUS* (Supplementary Fig. 1a) and *TDP-43* (Supplementary Fig. 1c) with *TAF15*, ensuring that *FUS*-RNA and *TDP-43*-RNA complexes were not inadvertently recovered. Given the high overlap in sequence similarity between the *TAF15* target RNAs isolated from the low (band A) and high (band B) molecular weight complexes (Supplementary Fig. 1d), the libraries were combined (Fig. 1b and Supplementary Fig. 1e), resulting in 5.9 million non-redundant sequenced reads that mapped to 13,633 annotated protein-coding pre-mRNAs having more than 10 reads (5,128,815 reads, 85.8%), noncoding genes (139,382 reads, 2.3%) and intergenic regions (706,897 reads, 11.8%) in the mouse genome (mm9).

Using a published cluster-finding algorithm<sup>22</sup>, we identified 47,138 *TAF15*-binding clusters in 4,873 genes. We applied the HOMER algorithm to these clusters to discover *in vivo* *TAF15*-binding motifs. The consensus motif GGUAAGU was statistically significantly enriched in *TAF15* clusters (Fig. 1c,  $P < 10^{-535}$ ). Interestingly, this motif is similar to the 5' splice site sequence, GURAGU (ref. 25); however, enrichment of this motif in both the coding sequence and the 3' untranslated repeat (UTR) provided evidence that we did not inadvertently extract the 5' splice site sequence within introns. Distribution analysis also illustrated that the *TAF15* motif is enriched within the centre of the CLIP clusters in the transcriptome (Fig. 1d) and also within 3' UTRs (Fig. 1e). We searched for the *TAF15* motif in clustered reads from published *FUS* (ref. 21) and *TDP-43* (ref. 22) CLIP-seq experiments and observed that the *TAF15* motif was also enriched in transcriptome-wide *FUS* CLIP clusters and, to a lesser extent, in *TDP-43* CLIP clusters residing in 3' UTRs



**Figure 1 | CLIP-seq reveals that TAF15 binds GGUAAGU motifs in the mouse brain.** (a) Autoradiograph of TAF15 protein-RNA complexes from the mouse brain immunoprecipitated with an antibody against TAF15 (left panel). RNA residing in the regions outlined by the red boxes was recovered for sequencing. TAF15-RNA complexes migrated at the expected size and were efficiently recovered because little protein remained in post-immunoprecipitation lysate (right panel, middle lane). (b) Flow chart illustrating CLIP-seq reads analysed to define TAF15 clusters. (c) *De novo* sequence motifs enriched above background within the transcriptome or specific genetic regions with associated binomial *P* values. (d) Positional distribution of the TAF15 motif GGUAAGU within TAF15 (red), FUS (green) or TDP-43 (purple) CLIP clusters. Inset graph shows the per cent enrichment of the TAF15 motif GGUAAGU within TAF15 targets ('Target') or within the transcriptome ('Background'). (e) Positional distribution analysis of the TAF15 motif GGUAAGU as in d but specifically within CLIP clusters residing in 3' UTRs. Inset graph shows the per cent enrichment of the TAF15 motif GGUAAGU within the 3' UTRs of TAF15 targets ('Target') or within the 3' UTRs of the transcriptome ('Background').

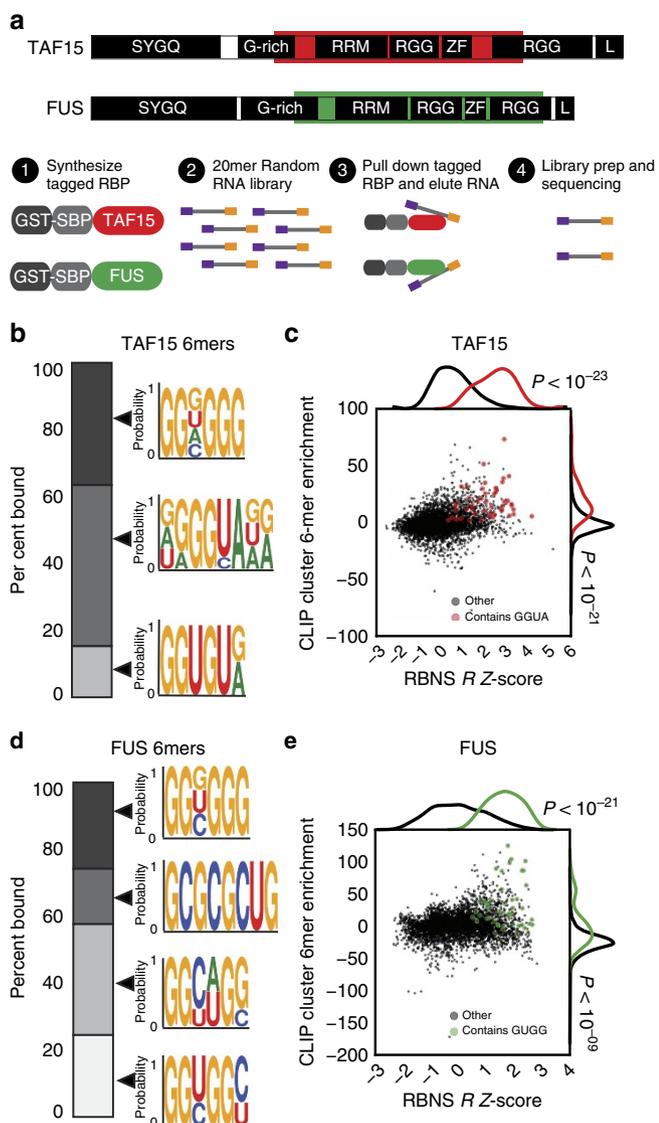
(Fig. 1d,e). We conclude that TAF15 interacts with binding sites enriched for a GGUAAGU motif within thousands of genes *in vivo*.

#### RNA Bind-n-Seq reveals TAF15 binding to GGUA motif *in vitro*.

To characterize the *in vitro* sequence specificity of TAF15, we applied RNA Bind-n-Seq (RBNS)<sup>26</sup> to recombinant TAF15 and, as a comparison, to recombinant FUS protein. Briefly, truncated forms of recombinant TAF15 or FUS containing both the RNA recognition motif and zinc-finger domain (amino acids 204–415 for TAF15 and amino acids 235–481 for FUS) were incubated with an RNA pool consisting of random 20mer RNAs flanked by short primers used to add adapters for high-throughput sequencing (Fig. 2a). For FUS, this truncated region was previously shown to exhibit high affinity for RNA<sup>27</sup>. Complementary to *in vivo* interactions identified by CLIP-seq, this method evaluates TAF15 and FUS independently of its *in vivo* complex interaction with RNA. For TAF15, RBNS discovered degenerate G-rich and GU-rich motifs and notably an (A/G)GGUA motif that resembled the GGUAAGU motif that was identified *in vivo* by CLIP (Fig. 2b). In fact, the shared GGUA 4mer was significantly enriched in hexamers that were over-represented in both RBNS and TAF15 CLIP-derived clusters relative to the appropriate control backgrounds (Fig. 2c). In addition, the same GGUA motif was enriched in the TAF15 clusters located within 3' UTRs of target genes (Supplementary Fig. 2a). RBNS applied to FUS domains identified a similar degenerate G-rich motif, a GC-rich motif, and a GGUGG motif (bottom motif in Fig. 2d) that resembled motifs identified in published *in vivo* CLIP

studies<sup>20,21</sup>. A similar evaluation of the GUGG 4mer (or GGUG, not shown) confirmed enrichment within FUS *in vivo* CLIP-seq-derived clusters in the transcriptome (Fig. 2e) and in 3' UTRs (Supplementary Fig. 2b). Interestingly, we found that the distribution of the GUGG 4mer was also enriched in the hexamers derived from the TAF15 RBNS experiment (Supplementary Fig. 2c). Similarly, the GGUA 4mer was enriched in the FUS hexamers (Supplementary Fig. 2d). Although both of these motifs were found at a lower level of significance in the hexamers derived from experiments interrogating the other protein, our results suggest that TAF15 and FUS share some affinity with each other's motifs. It is noteworthy that the affinities of TAF15 and FUS to k-mers containing GGUA and GGUG, although significantly different from background, is relatively weak compared with previously studied RBPs such as RBFOX2, MBNL1 and CELF1 (ref. 26). We conclude that TAF15 interacts with a previously undiscovered GGUA core motif within significantly enriched clusters *in vivo*. Importantly, our results demonstrate that the interactions of FUS and TAF15 with their RNA-binding sites can occur independently of cofactor associations.

**TAF15 interacts with many FUS-RNA targets.** Similar to FUS and TDP-43, TAF15 clusters were predominantly found within introns (Supplementary Fig. 3a), consistent with previously published results in HEK293 cells<sup>28</sup>, mouse neurons and human brain tissue<sup>15</sup>. As intronic regions account for a substantial proportion of nucleotides in transcribed RNA, this distribution



**Figure 2 | RNA Bind-n-Seq confirms enrichment for GGUA motifs in RNAs that bind TAF15 *in vitro*.** (a) Experimental overview of RBNS. Truncated versions of TAF15 and FUS (highlighted in red and green, respectively) were tagged and incubated at different concentrations with a diverse pool of RNA oligonucleotides flanked by adapters. The tagged proteins were retrieved with streptavidin-coated beads and bound RNAs were sequenced. Input RNA was sequenced in parallel to quantify the proportions of bound RNA molecules. (b) RNA-binding preferences for truncated TAF15 shown as motif logos made from aligning RBNS 6mers weighted by their enrichments. Motif proportions were determined by summing the enrichments of each motif's aligned 6mers. (c) Scatter plot correlating the per cent enrichment above background of 6mers in TAF15 mouse brain CLIP-seq versus TAF15 RBNS *R* Z-scores. Red dots represent all significant 6mers containing the GGUA motif. Histograms show normalized distributions of 6mers containing (red) or not containing (black) the GGUA motif in CLIP-seq (right) or RBNS (top). *P* values shown are computed by a Kolmogorov-Smirnov statistic. (d) RNA-binding preferences for truncated FUS as in b. (e) Scatter plot and histogram analyses are as described in c using FUS mouse brain CLIP-seq versus FUS RBNS *R* Z-scores. Green dots represent all significant 6mers containing the GUGG motif.

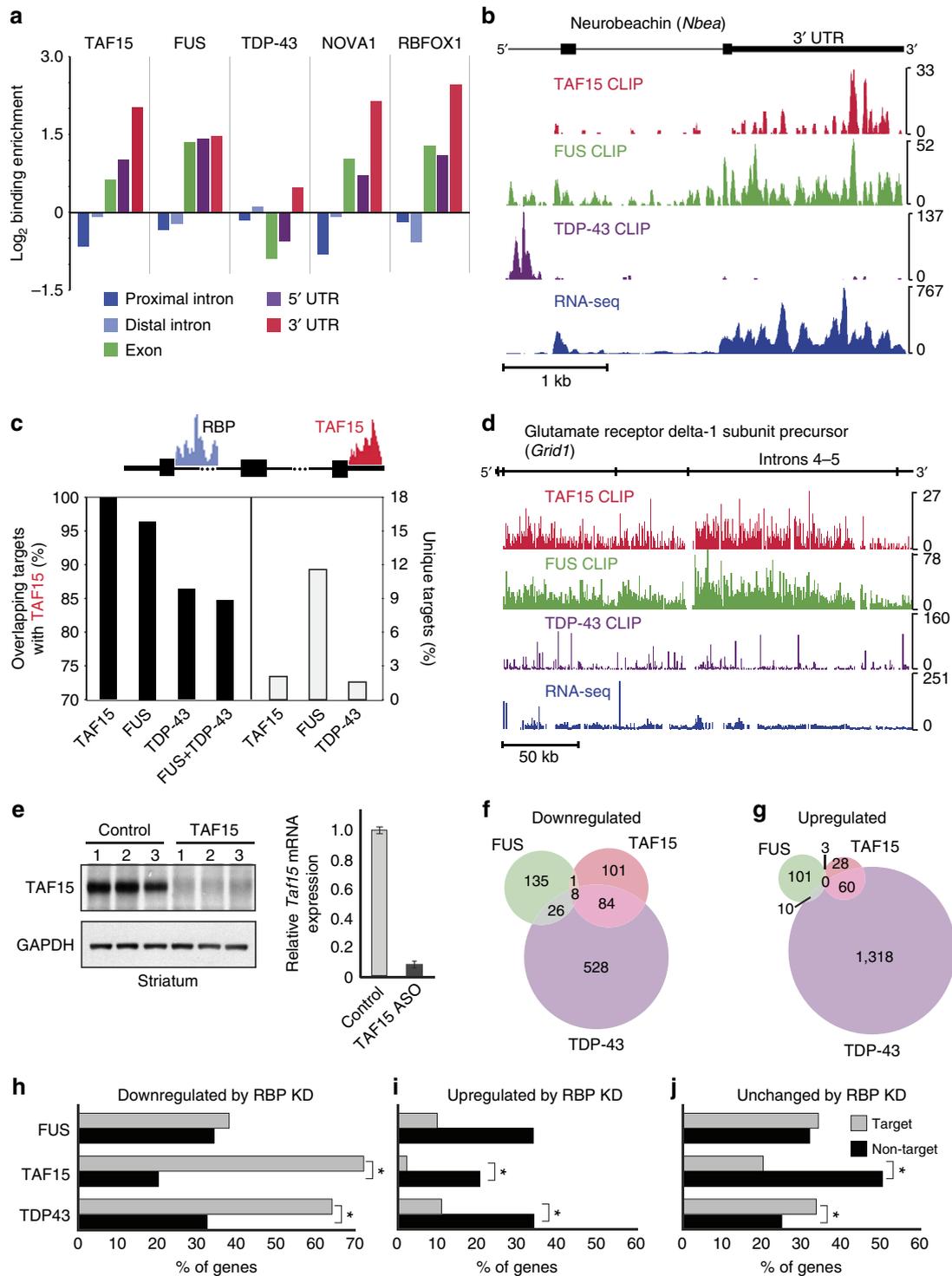
was expected and similar to other predominantly nuclear-localized RBPs such as RBFOX1 and NOVA1 (Supplementary Fig. 3a). Unlike TDP-43, we found that TAF15 and FUS binding

was significantly enriched in the 3' UTR, akin to RBFOX1 and NOVA1 that also have proposed 3' end formation roles<sup>29,30</sup> (Fig. 3a). To illustrate this, the 3' UTR of the neurobeachin (*Nbea*) transcript, encoding a protein involved in synaptic function and autism<sup>31</sup>, is enriched for TAF15 and FUS binding (Fig. 3b). FUS had a similar binding profile as TAF15, whereas TDP-43 bound an intronic region upstream of the penultimate exon, but with no cluster in the 3' UTR (Fig. 3b).

We found that targets of RBFOX1 and NOVA1 do not overlap with TAF15 target genes (Supplementary Fig. 3b). In contrast, the majority of FUS (98%) and TDP-43 (86%) target RNAs were also TAF15 targets (Fig. 3c). For genes that were targets of both TAF15 and FUS, 38% had at least one binding site that overlapped between TAF15 and FUS (Supplementary Fig. 3c). Our results indicate that TAF15 and FUS bind to the same genes with close proximity, consistent with our findings that the GUGG motif preferred by FUS was enriched in TAF15 CLIP clusters and the GGUA motif preferred by TAF15 was also enriched in FUS CLIP clusters (Supplementary Fig. 2c,d). TAF15 also exhibited a 'saw-tooth'-like pattern of deposition within genes containing long introns, such as the glutamate receptor delta-1 subunit precursor gene (*Grid1*), similar to FUS, but dissimilar to TDP-43 (ref. 21; Fig. 3d). We conclude that TAF15 and FUS binding are enriched in the 3' UTRs of target genes and both harbour the same 'saw-tooth'-like profiles in long introns.

#### Distinct roles of TAF15, FUS and TDP-43 on gene expression.

To identify TAF15-regulated RNAs, single-stranded antisense oligonucleotides (ASOs) complementary to TAF15 RNA or non-targeting control ASOs (Control) were delivered into the adult mouse striatum. TAF15 mRNA and protein were depleted by at least 90% in mice treated with TAF15-targeting ASOs (Fig. 3e). RNA extracted from striata of three mice was subjected to strand-specific RNA sequencing (RNA-seq) library generation and sequencing. On average, 24.8 million reads were obtained for each library, of which 86% mapped to the mouse genome (mm9). We identified 194 and 91 genes (Supplementary Data 1) that were significantly ( $P < 0.05$ ) downregulated (Fig. 3f) and upregulated (Fig. 3g), respectively, when TAF15 protein was depleted. To examine overlapping and unique effects of TAF15, FUS and TDP-43 on RNA expression, we re-analysed RNA-seq data sets in which FUS and TDP-43 were depleted from the mouse striatum in the same manner as TAF15 (refs 21,22). Although *Fus* and *Tdp-43* expression remained unchanged upon TAF15 depletion, the *Taf15* mRNA level appears to be slightly increased upon FUS depletion (Supplementary Fig. 3d). Similar to FUS and TDP-43 (ref. 21), we found that genes downregulated by loss of TAF15 exhibited exceptionally long introns (Supplementary Fig. 3e). Despite this similar trend in regulation, there was a poor overlap between the differentially regulated genes such that by our conservative re-analysis, only eight genes (including *Park2*, *Nrxn1* and *Kcnp4*) were commonly downregulated (Fig. 3f) and no genes were commonly upregulated (Fig. 3g). To distinguish between direct and indirect effects of RBP binding on gene expression, we measured the fraction of affected genes that were directly bound as determined by CLIP-seq. A significantly higher proportion of genes downregulated upon TAF15 or TDP-43 loss were direct targets of that RBP (Fig. 3h). Closer examination revealed that this association remains significant for the subset of downregulated genes that exhibited TAF15 (and to some extent TDP-43) binding in the 3' UTR (Supplementary Fig. 3f). In support of this result, genes that were downregulated upon TAF15 loss were more likely to contain the TAF15 'GGUAA' motif in their 3' UTRs or introns (Supplementary Fig. 3g). Genes that were upregulated or were unaffected upon loss of FUS,



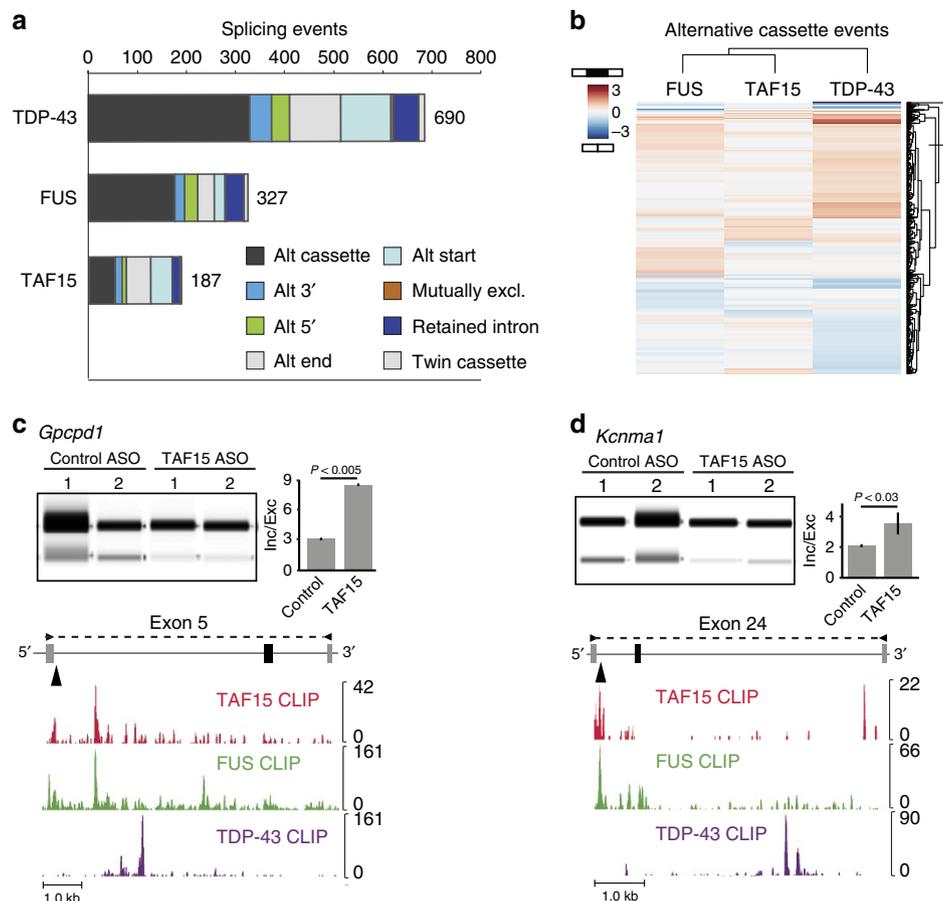
**Figure 3 | TAF15 and FUS exhibit similar RNA interaction profiles in the mouse brain.** (a) Fold change in binding enrichment of TAF15, FUS, TDP-43, NOVA1 and RBFOX1 after normalization to the average length of proximal introns (dark blue), distal introns (light blue), exons (green), 5' UTRs (purple) or 3' UTRs (red). (b) An example of 3' UTR binding by TAF15 (red) and FUS (green) but not by TDP-43 (purple) to Neurobeachin (*Nbea*) mRNA (chr3:55,428,730-55,433,169) in the mouse brain. RNA-seq results showing expression of *Nbea* is shown in blue. (c) Bar graph showing the per cent of gene targets that are common (black bars) or unique (white bars) to TAF15 and FUS, TDP-43 or both FUS and TDP-43. (d) An example of intronic 'saw-tooth' binding by TAF15 (red) and FUS (green) but not by TDP-43 (purple) to the glutamate receptor delta-1 subunit precursor (*Grid1*) mRNA (chr14:35,634,350-36,071,292) in the mouse brain. RNA-seq results showing expression of *Grid1* is shown in blue. (e) Confirmation of reduced TAF15 expression in the mouse striatum by western blot analysis (left) and qPCR (right). Knockdown was achieved by intrastriatal injection of ASOs complementary to TAF15 or a control ASO. Error bars represent s.d. (f, g) Venn diagrams showing overlap of genes downregulated (f) and upregulated (g) upon loss of TAF15, FUS or TDP-43 in the mouse striatum. (h-j) Per cent of genes that are downregulated (h), upregulated (i) or unchanged (j) upon ASO-mediated knockdown of the indicated RBP that has at least one binding site (target, grey) or no binding sites (non-target, black) by that RBP. Asterisks denote significant difference between target and non-target genes by Fisher's exact test at  $P < 0.05$ .

TAF15 or TDP-43 were generally not binding targets (Fig. 3i,j). Thus, we conclude that, although FUS, TAF15 and TDP-43 bind many of the same targets, only a small fraction of genes are similarly affected by loss of each of the three RBPs.

**TAF15 has a marginal role in AS.** Using splicing-sensitive microarrays, we detected 182 AS events that were altered upon TAF15 depletion in the mouse striatum (Fig. 4a). Although TAF15, TDP-43 and FUS proteins were reduced to similar levels, we observed fewer TAF15-dependent splicing events ( $n = 187$ ) compared with the number of splicing events altered by loss of FUS ( $n = 327$ ) or TDP-43 ( $n = 690$ ; Supplementary Data 2). There was little overlap in AS events altered by loss of TAF15, FUS or TDP-43 (Fig. 4b and Supplementary Fig. 4a). This suggests that, despite their high similarity in domain architecture and documented interactions with splicing factors<sup>32,33</sup>, FUS and TAF15 have distinct impacts on AS.

An AS event altered by TAF15 is exon 5 of the glycerophosphocholine phosphodiesterase 1 gene (*GPCPD1*; Fig. 4c), which was included upon TAF15 knockdown and harbours binding sites for TAF15 and FUS downstream of the 5' flanking exon

(indicated by an arrow). Another example, exon 24 in the calcium-activated potassium channel subunit alpha-1 gene (*Kcnma1*; Fig. 4d), was also included upon TAF15 loss and contained nearby TAF15- and FUS-binding sites (indicated by an arrow). Although TDP-43-binding sites were present near these exons, they were distinct from TAF15- and FUS-binding sites (Fig. 4c,d). A previous study reported that knockdown of TAF15 promoted the exclusion of exon 19 of the N-Methyl-D-Aspartate Receptor Subunit NR1 (*Grin1*) gene in mouse neurons<sup>15</sup>. TAF15- and FUS (but not TDP-43)-binding sites were present proximal to this exon, but depletion of TAF15 did not cause a statistically significant change ( $P = 0.106$ ) in the exclusion of exon 19 of *Grin1* in the mouse striatum (Supplementary Fig. 4b). This does not appear to be because of differences in tissue specificity as depletion of TAF15 in the mouse brain or spinal cord (Supplementary Fig. 4c) also had no significant effect on *Grin1* exon 19 splicing (Supplementary Fig. 4d). We conclude that in contrast to our and others' previous findings with FUS and TDP-43 (refs 20–22,34) and a study regarding TAF15 (ref. 15), TAF15 alters the splicing of a relatively small subset of genes, of which the majority (70%) are distinct from those regulated by either FUS or TDP-43.



**Figure 4 | TAF15 influences alternative splicing for a small subset of transcripts.** (a) Bar graph showing the number of alternative splicing events altered upon ASO-mediated depletion of TDP-43, FUS or TAF15 in the mouse striatum, as detected by splicing-sensitive microarray analyses. (b) Heatmap of alternative cassette events in a altered by FUS, TAF15 or TDP-43 depletion. Hierarchical clustering analysis was performed using separation (Sep) scores. Higher Sep scores (red) indicate inclusion events and lower Sep scores (blue) indicate exclusion events. (c) RT-PCR for exon 5 of glycerophosphocholine phosphodiesterase 1 (*Gpcpd1*; chr2:132,382,646–132,390,412) to assess alternative splicing in TAF15 knockdown samples compared with controls. Quantification of biological replicates is shown. Error bars represent s.d. Binding of TAF15 (red), FUS (green) and TDP-43 (purple) in the mouse brain is shown below. (d) RT-PCR of exon 24 in the potassium channel, calcium-activated large conductance subfamily M alpha, member 1 (*Kcnma1*) (chr14:24,149,961–24,156,401) to assess alternative splicing in TAF15 knockdown samples compared with controls. Quantification of biological replicates is shown. Error bars represent s.d. Binding of TAF15 (red), FUS (green) and TDP-43 (purple) in the mouse brain is shown below.

**TAF15 and FUS affect mRNA stability in neural progenitors.**

To evaluate the role of TAF15 in early neuronal development, we used human NPCs differentiated from iPSCs in which TAF15 protein levels and, for comparison, FUS protein levels, were individually depleted by lentiviral short hairpin RNAs (shRNAs; Fig. 5a,b). As TAF15 has a relatively minor role in AS, we investigated a potential role for TAF15 in RNA stability. NPCs were treated with the transcriptional inhibitor Actinomycin D for varying times, after which total RNA was collected and prepared for RNA-seq libraries (Fig. 5a). Half-life measurements were determined from a regression-based analysis of gene expression, assuming first-order decay kinetics. For each shRNA treatment, the median value for the coefficient of determination ( $R^2$ ) describing the log-linear fit across the time course for each gene, across all genes, was 0.54; this value was significantly higher ( $P \sim 0$ , by Kolmogorov–Smirnov two-tailed test) than the value (0.12) obtained by randomly shuffling the expression values for each time point within each gene (Fig. 5c and Supplementary Fig. 5a,b). To minimize false positives, we evaluated genes for which the  $R^2$  value was greater than 0.6. We identified 299 and 330 genes that were highly stabilized (increased half-life), as well as 132 and 44 genes that were highly destabilized (decreased half-life) upon loss of TAF15 and FUS, respectively (Fig. 5d, Supplementary Data 3).

We arbitrarily selected mRNAs that were most stabilized or destabilized by loss of TAF15 (marked by asterisks in Fig. 5e) and performed RNA immunoprecipitation followed by quantitative reverse transcriptase PCR (qRT-PCR) to determine whether these mRNAs were directly bound by TAF15 and FUS. TAF15 bound to most mRNAs (four of five) that were stabilized upon loss of TAF15 (*URB1*, *SNX9*, *CLN8* and *SMURF2*; Fig. 5f), of which *URB1*, *CLN8* and *SMURF2* also exhibited FUS interactions. In addition, TAF15 bound to most mRNAs (four of five) that were destabilized upon loss of TAF15 (*ATXN7L3B*, *PRKRIR*, *RAPGEF1* and *CGGBP1*; Fig. 5g). Notably, FUS bound to all these mRNAs (including *TCERG1*); however, FUS depletion did not appear to have an effect on mRNA stability of these transcripts (Fig. 5e and Supplementary Data 4). *ANAX2* and *TIAL1*, whose mRNAs were unaltered by TAF15 loss, were also examined for TAF15 and FUS binding (Supplementary Fig. 5c). A gene ontology (GO) analysis of genes affected at the mRNA stability level upon TAF15 knockdown revealed statistical enrichment for genes implicated in DNA-dependent transcription control ( $P < 10^{-26}$ ; Supplementary Data 5). An example of TAF15 mRNA turnover target involved in transcriptional control and also neurological diseases is the CGG-binding protein 1 (*CGGBP1*), which binds to CGG repeats in the promoter of the fragile X mental retardation 1 (*FMR1*) gene resulting in reduced expression<sup>35</sup>. We conclude that TAF15 and FUS control mRNA turnover in NPCs of distinct mRNA substrates.

**TAF15 and FUS affect different genes in human MNs.**

To discover the molecular events modulated by loss of TAF15, FUS and TDP-43 in an ALS-relevant cell type, we generated MNs from wild-type human iPSCs using a directed differentiation protocol<sup>36</sup>. Briefly, a combination of SMAD signalling inhibitors, Noggin, and the ALK5 inhibitor, SB431542, was used to yield a population of cells enriched for HB9, ISLET1 and TUJ1 (neuron-specific class III)-positive MNs with a minor fraction of OLIG2-positive oligodendrocytes (Fig. 6a,b). We subjected the MNs to lentivirus-packaged shRNAs targeting TAF15, FUS or TDP-43. As our *in vivo* findings indicated that TAF15 and FUS bind to similar RNA substrates, we also simultaneously depleted FUS and TAF15. Mature RNA and protein levels (Fig. 6c,d and Supplementary Fig. 6a) of the targeted RBPs were significantly

reduced and TAF15 and FUS protein levels did not exhibit reproducible changes (either up or down) in FUS and TAF15 depletions, respectively (Supplementary Fig. 6a). Reduction of TAF15, FUS or TDP-43 alone or in combination (TAF15 and FUS) in iPSC-derived MNs did not cause noticeable changes in cell morphology or death. We generated RNA-seq data from these cells, obtaining an average of 32.4 million uniquely mapped reads.

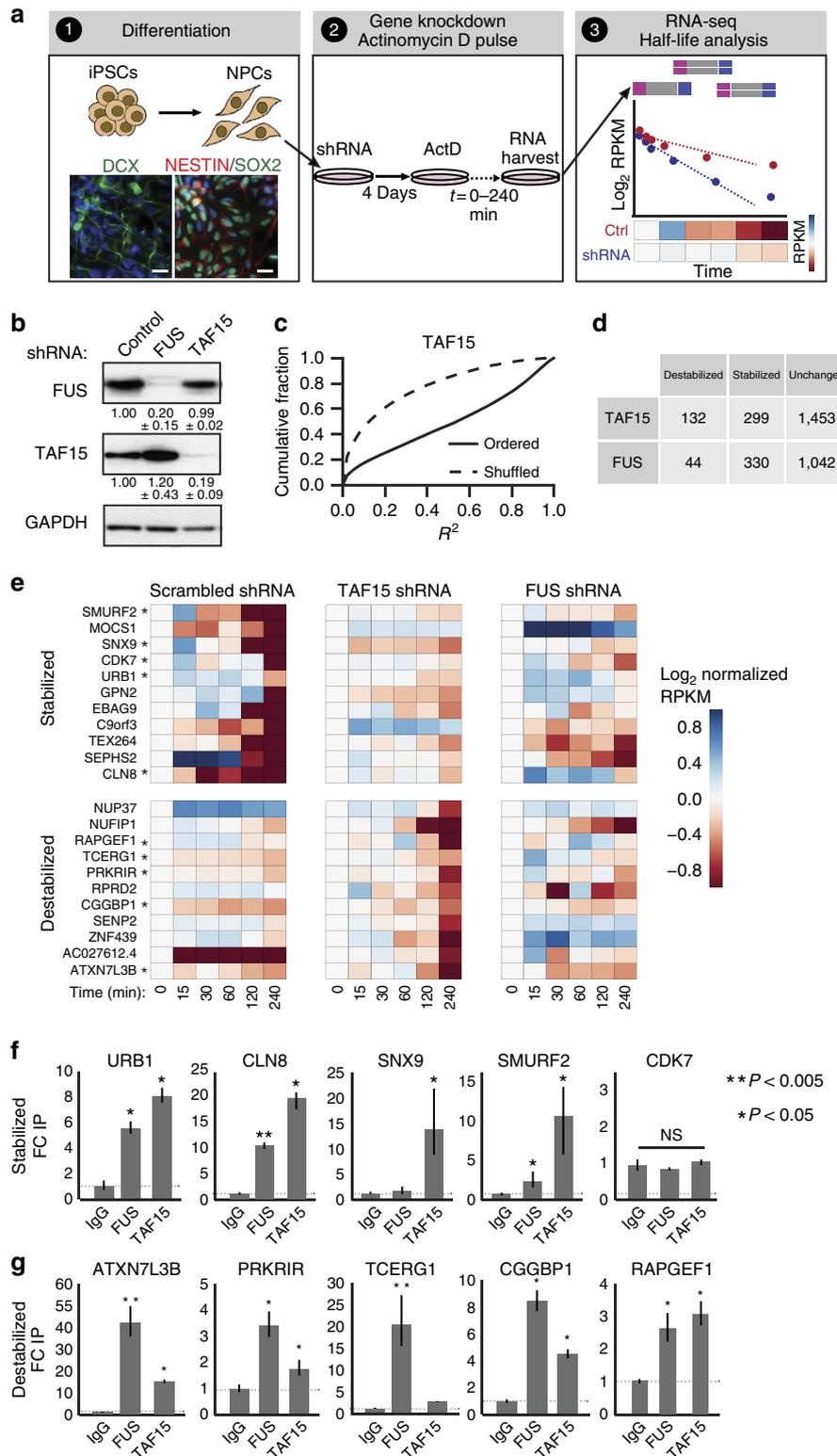
Similar to our *in vivo* depletion studies, we observed a minor overlap in the genes downregulated (61 genes) or upregulated (6 genes) upon loss of all three RBPs (Fig. 6e and Supplementary Data 6). In contrast to our findings in the adult mouse striatum, introns within downregulated genes affected by loss of TAF15 and FUS in MNs were not significantly longer than upregulated or unaffected genes (data not shown). Expectedly, ~76 and 85% of the genes in the FUS-only and TAF15-only knockdown experiments were also downregulated in the double knockdown. However, we found that a subset of genes ( $n = 144$ ) were downregulated only upon combined loss of TAF15 and FUS in human MNs (Fig. 6f), indicating a potential redundancy between TAF15 and FUS in controlling gene expression. These genes that were downregulated upon combined TAF15 and FUS loss were enriched for GO terms, reflecting extracellular cellular matrix composition, cell proliferation, wound healing and cytokine activity.

**Genes altered by RBP loss are akin to ALS-linked FUS mutant.**

To investigate whether the molecular changes observed upon loss of FUS, TAF15 or both proteins were relevant to ALS pathogenesis, we obtained fibroblasts from two ALS patients with the causative R521G mutation in *FUS*. The fibroblasts were reprogrammed into iPSCs and subjected to cellular, molecular and genetic characterization to confirm that they are pluripotent (Supplementary Fig. 6b), exhibit a normal karyotype and harbour the presence or absence of the mutation at nucleotide position 1,561 (Supplementary Fig. 6c). Three individual clones from two *FUS* R521G patient-derived iPSC lines (two clones were from one line) and two control iPSCs (from healthy, age-matched non-mutant individuals) were directly differentiated to MNs. RNA isolated from these cells were subjected to RNA-seq library preparation and sequencing to obtain an average of 20 million reads, of which 90% mapped uniquely to the human genome (hg19). To ensure that the differentiation process yielded MNs at similar stages of differentiation and similar subtypes of cells, we compared expression of a panel consisting of genes representing housekeeping, astrocyte, oligodendrocytes, neural precursor and neuronal subtypes. The similarities in expression profiles among the MN cell lines confirmed that differentiation of the iPSC lines was consistent and hence enabled downstream comparative analysis (Supplementary Fig. 6d). We identified 901 downregulated and 805 upregulated (Supplementary Data 7) genes that were differentially affected in the *FUS* R521G MNs compared with wild-type control MNs. Interestingly, although the majority of mutant-dependent gene expression changes were unique, there existed statistically significant overlaps in the genes downregulated in the *FUS* R521G MNs (relative to control) with genes downregulated upon loss of FUS ( $P < 10^{-9}$ ) or TAF15 ( $P < 10^{-3}$ ; Fig. 6g). Importantly, this overlap increased in number when we compared the genes affected by simultaneous depletion of both FUS and TAF15 ( $P < 10^{-22}$ ; Fig. 6g). In contrast, we observed no significant overlap in genes upregulated by any condition (Fig. 6g). Overall, these findings are consistent with our observations that FUS and TAF15 are redundant in their effects on molecular targets and implies a partial loss of molecular function by the *FUS* R521G mutation.

**Downregulated genes correlate with a sALS RNA signature.** To obtain insights into whether the genes affected by loss of ALS-associated RBPs resemble disease-specific RNA signatures, we turned to a RNA-seq data set generated from laser-capture microdissected spinal cord samples from sALS patients who had bulbar or arm onset of disease that was caudally progressing and thus had abundant residual MNs in the lumbar region at the time of death<sup>37</sup>. The RNA-seq data set consisted of samples from 13

sALS and 9 control patients. In all, 3,876 genes were significantly differentially regulated, of which 71% and 29% were upregulated and downregulated, respectively, in the sALS patient compared with normal samples (Supplementary Data 8). The differentially expressed genes were effectively able to separate the diseased patients from the control patients (Supplementary Fig. 6e). Next, we tested the hypothesis that ALS RBP-mediated RNA changes resemble the RNA signature that distinguishes sALS and normal



samples. For all the comparisons performed, we observed a significant overlap ( $P < 0.05$ , hypergeometric test) in genes that were upregulated in sALS samples and downregulated in FUS, TAF15, TDP-43, or FUS and TAF15 double knockdowns (Supplementary Fig. 6f). We also observed a significant inverse correlation of significantly changing genes between sALS samples and FUS, TDP-43 or FUS and TAF15 double knockdowns (Fig. 6h,  $R^2$  between  $-0.14$  and  $-0.32$ ,  $P < 0.05$ ), but not between sALS samples and the FUS R521G mutant MNs (Supplementary Fig. 6g). Despite the divergent sets of regulated genes whose mRNA levels are dependent on ALS-associated RBPs, we found that 2,747 genes were upregulated in sALS patient samples. Unlike *in vitro* differentiated MNs, the sALS patient samples represent more mature MNs at a late stage of disease progression. Our findings indicate that in late-stage sALS patient samples with TDP-43 pathology<sup>37</sup>, a subset of genes that are separable from those found in ALS iPSC-derived FUS R521G MNs, are abnormally higher compared with control patients. Among the commonly differentially regulated genes between knockdown and sALS samples, GO terms for extracellular space and matrix organization were statistically enriched ( $P < 0.01$ ).

## Discussion

Genetic and clinical evidence strongly supports causative roles for FUS, TDP-43 and TAF15 in ALS. Here we identify common and unique pathways normally controlled by these proteins utilizing diverse *in vitro* and *in vivo* neuronal systems (Supplementary Fig. 7). In the adult mouse brain, we identified TAF15-binding sites within  $\sim 4,900$  RNA substrates, and a GGUAAGU TAF15-binding motif not reported in previous studies<sup>15,28</sup>. We used RBNS technology to confirm a GGUA motif that was enriched within *in vivo* TAF15-binding sites. Together, we conclude that TAF15 and FUS can interact with their RNA motifs within *in vivo* RNA substrates without requiring complex cofactor associations. Overall, the RNA-binding pattern of TAF15 resembled that of FUS, but was distinct from TDP-43, even when all three RBPs targeted the same genes. TAF15 and FUS exhibited saw-tooth-like binding patterns on long introns, a pattern reminiscent of co-transcriptional splicing<sup>38</sup>. Genes downregulated upon loss of either TAF15 or FUS contained exceptionally longer introns. In addition, TAF15- and FUS-binding sites were also over-represented within 3' UTRs, possibly reflecting 3' end processing functions such as RNA turnover, transport and translation. Upregulation of genes upon FUS and TAF15 loss is likely a secondary effect as these genes are generally not targets. Lastly, unlike TDP-43 and FUS, loss of TAF15 appeared to have a minor impact on AS in the adult mouse brain.

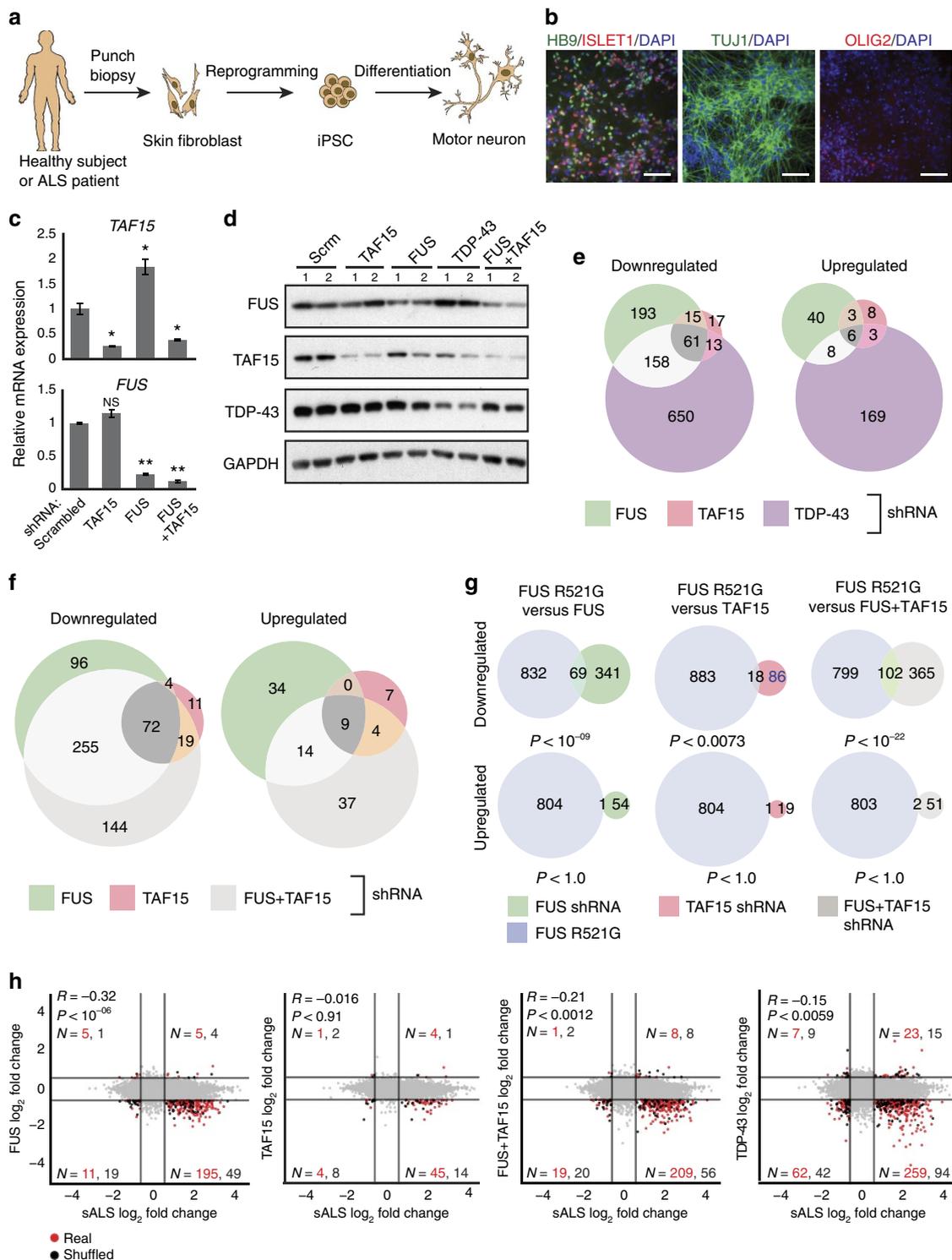
In models of early human neuronal development, we identified that TAF15 and FUS affected the mRNA turnover of distinct

subsets of RNA targets in human neuronal progenitor cells. Furthermore, loss of FUS, TAF15 or TDP-43 in human MNs derived from the same cells resulted in distinct changes in gene expression for each RBP. In addition, simultaneous depletion of FUS and TAF15 resulted in the downregulation of hundreds of additional genes. FUS and TAF15 have been shown to interact with each other<sup>23,24</sup> (Supplementary Fig. 1b) as well as other common proteins such as RNA Pol II (refs 39,40), spliceosome machinery<sup>24,32,33</sup> and transcription factors<sup>41</sup>. One possibility is that if FUS is unable to recruit regulatory factors to an RNA target, this function may be compensated for by TAF15.

To gain insight into disease, we compared the results of our loss-of-function studies to two models of ALS. The first model is MNs from ALS patients carrying the pathogenic FUS R521G mutation. Expression of FUS R521G from the mouse *MAPT* locus has been recently reported to cause neuronal toxicity in neurons of mice<sup>42</sup>. Previously, FUS R521G was also associated with a partial loss-of-function in RNA regulation in mouse spinal cords<sup>43</sup>. We did observe a small yet significant overlap in genes downregulated upon loss of TAF15, FUS or both proteins, and genes altered by FUS R521G. This overlapping set of genes may reflect the partial loss-of-function properties of FUS R521G (ref. 43). As mRNAs downregulated upon loss of these RBPs in the mouse brain are often direct binding targets of those RBPs, we speculate that the FUS R521G mutation, which causes cytoplasmic FUS mislocalization, resembles a partial loss-of-function of the RBPs in a model of early development. Nevertheless, the majority of expression changes caused by FUS R521G were mutant-specific such that they did not overlap with genes altered by loss of TAF15, FUS or both proteins. One interpretation is that these FUS R521G-specific gene changes may contribute the pathological, gain-of-function activities of mutant FUS that was observed to cause MN dysfunction in mice<sup>42</sup>.

To model late-stage ALS disease we utilized RNA-seq data obtained from spinal cord samples collected postmortem by laser-capture microdissection from sALS patients. These samples harboured ubiquitinated TDP-43 cytoplasmic inclusions and were from patients with no mutations in known ALS-causative genes, including FUS, TDP-43 or TAF15. Intriguingly, our comparisons of RNA signatures revealed an inverse correlation in a separate set of genes that were upregulated in the sALS samples but were downregulated upon loss of ALS-associated RBPs in *in vitro*-derived MNs. This indicates that these genes, whose levels are normally dependent and maintained by FUS, TAF15 and TDP-43, are aberrantly higher in late-stage ALS. A possible mechanism for gene upregulation is the breakdown of negative feedback loops as is observed for the effect of TDP-43 on its own expression<sup>44,45</sup>. We did not, however, observe a difference in TDP-43 mRNA levels between sALS and control neurons.

**Figure 5 | Loss of TAF15 or FUS affects mRNA stability in human neural precursor cells.** (a) Schematic of workflow. (1) Human iPSCs were differentiated into NPCs and stained for neuronal lineage markers DCX (left, green), NESTIN (right, red) and SOX2 (right, green) to confirm differentiation. 4,6-Diamidino-2-phenylindole (DAPI; blue) was used to locate nuclei. Scale bar, 25  $\mu\text{m}$ . (2) NPCs were infected with virus expressing shRNAs against TAF15 or FUS and then treated with Actinomycin D for indicated duration. (3) Poly(A)-selected RNA was converted into libraries for sequencing, and sequencing reads were used to calculate mRNA half-lives. (b) Validation of shRNA-mediated knockdown of FUS or TAF15 in NPCs by western blot analysis. Representative western blot is shown from one replicate with quantifications from biological triplicate knockdown experiments. (c) For each gene, the coefficient of determination ( $R^2$ ) reflecting the fit of the RPKM values to a log-linear regression was computed. The cumulative distribution functions of the  $R^2$  values for all genes in the TAF15 depletion experiment are depicted for real and shuffled values. (d) Table displaying the number of mRNAs whose half-lives were destabilized, stabilized or unchanged by depletion of TAF15 or FUS. Half-life changes, measured as  $\log_2$  (knockdown/control), that were greater than 1 were considered. (e) Heatmap of normalized RPKMs for stabilized and destabilized mRNAs upon shRNA-mediated knockdown of TAF15 or FUS. RPKMs are normalized for each gene to its RPKM at time 0. An asterisk indicates that the gene was examined for binding in f. (f) RNA immunoprecipitation was performed using antibodies against IgG (Control), TAF15 and FUS in NPCs. The relative fold change compared with the IgG control for genes that are stabilized by TAF15 loss was determined with qPCR. Values are means  $\pm$  s.d. for biological duplicates. Asterisk denotes a significant difference compared with IgG by Student's *t*-test where  $**P < 0.005$  and  $*P < 0.05$ . (g) RNA immunoprecipitation analysis as in f for mRNAs that were destabilized upon TAF15 loss.



**Figure 6 | Comparison of MN RNA signatures upon TAF15, FUS or TDP-43 loss to two models of ALS.** (a) Schematic of workflow to reprogramme iPSCs and to differentiate into MNs. (b) Immunofluorescence of human iPSC-derived MNs for MN marker HB9 (green), post-mitotic neuronal marker TUJ1 (green), neural stem cell marker ISLET1 (red) and oligodendrocyte marker OLIG2 (red). DAPI stain marks cell nuclei (blue). Scale bar, 25  $\mu$ m. (c) qRT-PCR and (d) western blot validation of shRNA-mediated depletion of TAF15, FUS and TDP-43 in MNs. Error bars represent s.d. from biological duplicate experiments. (e, f) Venn diagrams showing overlap of up- and downregulated genes in MNs upon depletion of TAF15, TDP-43, FUS or, simultaneously, FUS and TAF15 (FUS + TAF15). (g) Venn diagrams showing overlap of up- and downregulated genes between MNs with the FUS R521G mutation and knockdown of TAF15, FUS or FUS + TAF15. Statistical significance was determined by a hypergeometric test using genes expressed in MNs as background. (h) Scatter plots comparing gene expression changes ( $\log_2$  RPKM) in MNs from sALS patient samples compared with loss of TAF15, FUS, TDP-43 or FUS + TAF15. Each quadrant of a scatter plot shows genes (red dots) and gene counts ( $N$ , in red) that are significantly changing in sALS and RBP depletion experiments. Genes from a randomly ordered comparison are also shown (black dots) along with gene counts ( $N$ , in black).  $R^2$  and  $P$  values from linear regression analyses of genes significantly changing in both sALS and RBP knockdown experiments are shown.

Another plausible scenario is that in late stages of the disease, cytoplasmic inclusions of TDP-43 lead to stabilization of trapped, cytoplasmic RNA targets. Future studies to identify the mislocalized RNA targets in cytoplasmic bodies that are protected from degradation, such as stress granules, may yield further insight into disease-relevant targets at late stages in the disease.

In summary, our study delineates convergent and divergent RNA-processing functions of ALS-associated FUS, TAF15 and TDP-43 in normal and disease settings. Our comprehensive results shed light on multiple and distinct pathways by which these RBPs regulate gene expression in diverse neuronal systems and provide a framework for how they relate to ALS and other neurodegenerative diseases.

## Methods

**Injections of ASO in mice.** Stereotaxic injections of ASO complementary to TAF15 were performed in 8-week-old female C57Bl/6 mice to deplete TAF15. ASOs were delivered specifically to the striatum or brain/spinal cord by intrastriatal (12.5  $\mu$ g) or intracerebroventricular injection (300  $\mu$ g), respectively, as described previously<sup>21,46</sup>. Female mice were regularly monitored for 14 days until being killed and the tissues were harvested and frozen in TRIzol (Invitrogen). Control mice received a control ASO without any known target in the mouse genome under the same conditions. The ASOs were phosphorothioate 'gapmers' with sequences as follows (capitalized nucleotides containing 2  $\pm$  -O-(2-methoxy)ethyl modifications): GGTCCTctccatagTCGCT (TAF15; brain and striatum), TGGCAatattttacaACGCA (TAF15; spinal cord) and CTCAGTAACATTGA CACCAC (Control). All procedures were performed using a protocol approved by the Institutional Animal Care and Use Committee of Ionis Pharmaceuticals and the University of California at San Diego.

**Generation of neural precursor cells and MNs.** Human iPSCs derived from dermal fibroblast cells of a healthy individual (RRN08) were induced into neural precursor cells using a pan-neuronal protocol as previously described<sup>21</sup>. Briefly, stem cells were grown on Matrigel-coated plates (BD) in mTeSR1 growth media (Stem Cell Technologies). Stem cell colonies were grown on ultra low-attachment plates in DMEM/F12 + GlutaMAX supplemented with N2 and FGF-2 (20 ng ml<sup>-1</sup>). After 1 week, neural rosettes were manually picked, replated and maintained in DMEM/F12 + GlutaMAX supplemented with N2, B27 and FGF-2 (20 ng ml<sup>-1</sup>).

**Generation of human MNs.** Human MNs used in the shRNA knockdown experiments were differentiated from iPSCs (CVB) using a protocol modified from (ref. 36). Briefly, human iPSCs were maintained in hEB Media (Knockout D-MEM + 10% Knockout Serum Replacement (Life Technologies) + 10% Plasmanate (Biocare) + GlutaMAX + NEAA (Life Technologies) and supplemented with 10  $\mu$ M SB431542 and 1  $\mu$ M Dorsomorphin dihydrochloride (Tocris) on feeder-free dishes. Cells were maintained in SB431542 and Dorsomorphin until day 18 of differentiation. On days 4, 5 and 6 of differentiation, hEB media were mixed with N2 Base media (D-MEM/F12 + GlutaMAX, 1% N2 Supplement + 4.5 mM D-Glucose, 0.05 mM Ascorbic Acid (Sigma)) at a ratio of 70:30, 50:50 and 50:50, respectively. On days 7 and 8 of differentiation, cells were maintained in 50:50 combination of hEB media and maturation media (D-MEM/F12 + GlutaMAX, 2% N2 Supplement, 4% B27 serum-free supplement (Invitrogen), 9.0 mM D-Glucose and 0.1 mM ascorbic acid (Sigma)) supplemented with 2 ng ml<sup>-1</sup> each of ciliary neurotrophic factor, brain-derived neurotrophic factor and glial cell-derived neurotrophic factor (PeproTech). From days 7 to 22 of differentiation, cells were treated with 200 nM Smoothed Agonist (SAG; EMD Biosciences) and 1.5  $\mu$ M retinoic acid (Sigma). On day 18, cells were dissociated using Accutase and transferred to dishes coated with Poly-D-Lysine (Sigma) + Laminin (Life Technologies) and maintained in maturation media supplemented with retinoic acid and SAG. On day 22, cells were maintained in maturation media containing 2  $\mu$ M DAPT (Tocris). On day 26, cells were maintained in maturation media only. Throughout the differentiation protocol media was changed daily. The identity and purity of MNs were analysed by immunofluorescence for markers of stem cells, MNs, astrocytes and glial cells.

**Generation of MNs from fibroblast-derived iPSCs.** Adult human primary fibroblasts were obtained by Franca Cambia, Edward Kasarskis and Haining Zhu (University of Kentucky). Informed consent was obtained from all subjects before sample collection. The use of patient fibroblasts for research was approved by the University of Kentucky Institutional Review Board (IRB 05-0265). Briefly, adult human primary fibroblasts were cultured at 37 °C and 5% CO<sub>2</sub> in DMEM supplemented with 10% fetal bovine serum (FBS), NEAA and L-glutamine. To generate iPSC cells, control and ALS patient fibroblasts were transduced with the CytoTune iPSC Sendai Reprogramming Kit, as described in the manufacturer's protocol (Invitrogen). Colonies were manually passaged on Matrigel-coated plates and

grown in mTeSR1 growth media. After several passages, colonies were expanded using Accutase (Innovative Cell Technologies) and grown as a monolayer before differentiation. MN differentiation was performed as described above with the following modifications: CHIR99021 (Tocris) was added at 4  $\mu$ M until day 7 and the cells were either fixed for immunostaining or harvested for RNA in TRIzol (Life Technologies) 35 days post-neural induction. Three ALS patient lines GY6.2, GY7.3 and GY7.6 are referred to as FUS R521G Line 1, Line 2 and Line 3, respectively, and two wild-type sibling control lines KINI1ALS17.3 and KINI1ALS17.4 are referred to as wild-type sibling control Line 1 and Line 2, respectively.

**Lentiviral infections and transfections.** Lentiviral shRNA constructs (Open Biosystems) complementary to human TAF15 (TRCN0000020140, TRCN0000020141 or TRCN0000020143), human FUS/TLS (TRCN0000010450, TRCN0000039824 or TRCN0000039825) and human TDP-43 (TRCN0000016038) in the pLKO.1 vector system were used to produce lentivirus as previously described<sup>47</sup>. Virus produced from a pLKO.1 construct containing a control sequence was used as the control. At 60–70% confluency, NPCs were infected with virus (multiplicity of infection = 3) for 24 h, followed by a complete media change and further incubation for 72 h until cells were either collected and frozen in TRIzol (Invitrogen) or pelleted and frozen in liquid nitrogen for RNA and protein analyses, respectively. For lentiviral infection of MNs, media containing virus (multiplicity of infection = 5) were added to cells on day 28 of the MN differentiation protocol. After 24 h, a complete media change was performed and cells incubated for an additional 48 h. A second round of infection, similar to the first, began on day 31. On day 34 of MN differentiation, corresponding to a 6-day exposure period to shRNA expression, cells were either collected and frozen in TRIzol (Invitrogen) or pelleted and frozen in liquid nitrogen for RNA and protein analyses, respectively. For transfection of HEK293T cells, cells were plated in DMEM high-glucose media (Life Technologies) supplemented with 10% FBS. Cells were transfected with plasmid expressing human FUS-Myc cloned into pcDNA5 or TAF15-V5 cloned into pEF5-DEST using FuGENE 6 (Promega) according to the manufacturer's protocol for 24 h and then harvested for protein analysis.

**CLIP-seq library preparation and sequencing.** Brains from 8-week-old female C57Bl/6 mice were rapidly dissociated by forcing the tissue through a cell strainer with a pore size of 100  $\mu$ m (BD Falcon) before ultraviolet crosslinking. CLIP-seq libraries were constructed as previously described<sup>47</sup> using 10  $\mu$ g of a polyclonal antibody against TAF15 (300A-308, Bethyl Laboratories). Libraries were subjected to sequencing on a HiSeq2000 platform for 50 cycles. For each CLIP-seq library, the brain of one mouse was used.

**Computational analysis of CLIP-seq experiments.** CLIP-seq alignment and peak calling were performed as previously described<sup>21</sup>. Briefly, reads with the sequencing adapter or homopolymeric runs were trimmed and then mapped to the repeat-masked mouse genome (mm9) using Bowtie (version 0.12.2) with parameters  $-q -p 4 -e 100 -a -m 10 -best-strata$ . Reads that were flagged as PCR duplicates were removed. Significant clusters of reads were identified using a Poisson distribution with two different frequencies to determine a *P* value. First, a transcriptome-wide frequency was calculated by dividing the total length of all pre-mRNAs by the total number of CLIP reads mapping to the whole pre-mRNA transcriptome. Second, a gene-specific frequency was calculated by dividing the size of the gene-specific pre-mRNA by the total number of CLIP reads mapping to that gene-specific pre-mRNA. A significant cluster was annotated if it had sufficient reads to exceed a Bonferroni-corrected  $P < 10e^{-4}$  using both frequencies against the Poisson distribution.

**De novo motif analysis.** Motif analysis was performed as previously described<sup>48</sup>. Briefly, HOMER<sup>49</sup> was used to call *de novo* motifs using the command 'findMotifs.pl <foreground> fasta <outloc> -nofacts -p 4 -rna -S 20 -len 5,6,7,8,9 -noconvert -nogo -fasta <background>'. Foreground is defined as a fasta file of sequences taken from all called clusters, or all called clusters in a specific transcriptome region and background was randomly located clusters within the same gene regions as predicted TAF15 clusters.

**Peak annotations.** Transcriptome regions and gene classes were defined using annotations found in GENCODE version 17 (ref. 50). Depending on the analysis, clusters were either associated by the GENCODE-annotated 5' UTR, 3' UTR, exon or intronic regions. If a cluster overlapped multiple regions or a single part of a transcript was annotated as multiple regions, clusters were iteratively assigned first as exon, then 3' UTR, 5' UTR and finally as proximal or distal introns (as defined as 500 bp or greater from an exon-intron boundary). Overlapping peaks were calculated using bedtools<sup>51,52</sup>.

**Enrichment of peaks relative to region size.** To compute the fold enrichment of peaks in a given region, the fraction of peaks in that region was calculated as described above. The fractional region size was derived by dividing the total



RPKM at time  $t$ . Half-lives were derived from the decay rate using the equation  $t_{1/2} = \ln(2)/\lambda$ . Genes were included in the analysis if their decay rate was positive (that is, RPKMs decreased over time) and the linear regression line had a  $R^2$  fit greater than 0.6.

**Correlation of gene expression to CLIP binding and motifs.** Mouse brain CLIP-seq data for FUS and TDP-43 were previously described<sup>21,22</sup>. The binding location of each peak was assigned using the peak annotation method as described above. For each RBP, mouse brain CLIP-seq data and mouse striatum knockdown RNA-seq data were used to classify genes into the following categories: target and regulated, non-target and regulated, target and not regulated, and non-target and not regulated. A Fisher's exact test was performed to determine whether binding and regulation were significantly correlated. Motif analysis was performed similarly by determining whether a TAF15 'GGUAA' or FUS 'GUGG' motif was present in the 3' UTRs or introns of genes.

**Splicing-sensitive microarray analysis.** Total RNA from three individual control and TAF15 ASO-treated mice were prepared for hybridization to splicing-sensitive microarrays (Affymetrix). Separation scores (Sep scores) were generated as previously described<sup>61</sup>. For clustering of splicing events, a splicing event was included in clustering if, for any of the three experiments, TAF15, FUS or TDP-43 knockdown was significantly ( $|\text{Sep score}| > 0.5$ ,  $q$  value  $< 0.05$ ) differentially expressed. Hierarchical clustering was performed using Seaborn/SciPy on the Sep scores for each splicing event. Overlap analysis of splicing-sensitive microarray results and TAF15 mouse CLIP-seq data was performed as previously described<sup>61</sup>.

**GO analysis.** Significantly enriched GO terms were identified using a hypergeometric test that compared the number of genes that were either regulated (RNA-seq data) or bound (CLIP-seq data) in each GO term to genes expressed (background) in each GO term. The background gene set was defined as genes that were expressed (RPKM  $> 1$ ) in the corresponding control experiment.

**Data availability.** The accession number for the sequencing data deposited in GEO for this paper is GSE77707.

## References

- Arai, T. *et al.* TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Biochem. Biophys. Res. Commun.* **351**, 602–611 (2006).
- Neumann, M. *et al.* Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314**, 130–133 (2006).
- Corrado, L. *et al.* High frequency of TARDBP gene mutations in Italian patients with amyotrophic lateral sclerosis. *Hum. Mutat.* **30**, 688–694 (2009).
- Daoud, H. *et al.* Contribution of TARDBP mutations to sporadic amyotrophic lateral sclerosis. *J. Med. Genet.* **46**, 112–114 (2009).
- Del Bo, R. *et al.* TARDBP (TDP-43) sequence analysis in patients with familial and sporadic ALS: identification of two novel mutations. *Eur. J. Neurol.* **16**, 727–732 (2009).
- Kühnlein, P. *et al.* Two German kindreds with familial amyotrophic lateral sclerosis due to TARDBP mutations. *Arch. Neurol.* **65**, 1185–1189 (2008).
- Lemmens, R. *et al.* TDP-43 M311V mutation in familial amyotrophic lateral sclerosis. *J. Neurol. Neurosurg. Psychiatry* **80**, 354–355 (2009).
- Rutherford, N. J. *et al.* Novel mutations in TARDBP (TDP-43) in patients with familial amyotrophic lateral sclerosis. *PLoS Genet.* **4**, e1000193 (2008).
- Sreedharan, J. *et al.* TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science* **319**, 1668–1672 (2008).
- Vance, C. *et al.* Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. *Science* **323**, 1208–1211 (2009).
- Kwiatkowski, T. J. J. *et al.* Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science* **323**, 1205–1208 (2009).
- Lagier-Tourenne, C., Polymenidou, M. & Cleveland, D. W. TDP-43 and FUS/TLS: emerging roles in RNA processing and neurodegeneration. *Hum. Mol. Genet.* **19**, R46–R64 (2010).
- Couthouis, J. *et al.* A yeast functional screen predicts new candidate ALS disease genes. *Proc. Natl Acad. Sci. USA* **108**, 20881–20890 (2011).
- Ticozzi, N. *et al.* Mutational analysis reveals the FUS homolog TAF15 as a candidate gene for familial amyotrophic lateral sclerosis. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **156B**, 285–290 (2011).
- Ibrahim, F. *et al.* Identification of *in vivo*, conserved, TAF15 RNA binding sites reveals the impact of TAF15 on the neuronal transcriptome. *Cell Rep.* **3**, 301–308 (2013).
- Neumann, M. *et al.* FET proteins TAF15 and EWS are selective markers that distinguish FTLD with FUS pathology from amyotrophic lateral sclerosis with FUS mutations. *Brain* **134**, 2595–2609 (2011).
- Vance, C. *et al.* ALS mutant FUS disrupts nuclear localization and sequesters wild-type FUS within cytoplasmic stress granules. *Hum. Mol. Genet.* **22**, 2676–2688 (2013).
- Guo, W. *et al.* An ALS-associated mutation affecting TDP-43 enhances protein aggregation, fibril formation and neurotoxicity. *Nat. Struct. Mol. Biol.* **18**, 822–830 (2011).
- Schwartz, J. C., Cech, T. R. & Parker, R. R. Biochemical properties and biological functions of FET proteins. *Annu. Rev. Biochem.* **84**, 141210135300003 (2014).
- Rogelj, B. *et al.* Widespread binding of FUS along nascent RNA regulates alternative splicing in the brain. *Sci. Rep.* **2**, 603 (2012).
- Lagier-Tourenne, C. *et al.* Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. *Nat. Neurosci.* **15**, 1488–1497 (2012).
- Polymenidou, M. *et al.* Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. *Nat. Neurosci.* **14**, 459–468 (2011).
- Thomsen, C., Grundevik, P., Elias, P., Ståhlberg, A. & Aman, P. A conserved N-terminal motif is required for complex formation between FUS, EWSR1, TAF15 and their oncogenic fusion proteins. *FASEB J.* **27**, 4965–4974 (2013).
- Sun, S. *et al.* ALS-causative mutations in FUS/TLS confer gain and loss of function by altered association with SMN and U1-snRNP. *Nat. Commun.* **6**, 6171 (2015).
- Shapiro, M. B. & Senapathy, P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* **15**, 7155–7174 (1987).
- Lambert, N. *et al.* RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. *Mol. Cell* **54**, 887–900 (2014).
- Yang, L. *et al.* Subcellular localization and RNAs determine FUS architecture in different cellular compartments. *Hum. Mol. Genet.* **24**, 5174–5183 (2015).
- Hoell, J. I. *et al.* RNA targets of wild-type and mutant FET family proteins. *Nat. Struct. Mol. Biol.* **18**, 1428–1431 (2011).
- Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
- Licalosi, D. D. *et al.* HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464–469 (2008).
- Olszewski, P. K. *et al.* Neurobeachin, a regulator of synaptic protein targeting, is associated with body fat mass and feeding behavior in mice and body-mass index in humans. *PLoS Genet.* **8**, e1002568 (2012).
- Jobert, L. *et al.* Human U1 snRNA forms a new chromatin-associated snRNP with TAF15. *EMBO Rep.* **10**, 494–500 (2009).
- Yamazaki, T. *et al.* FUS-SMN protein interactions link the motor neuron diseases ALS and SMA. *Cell Rep.* **2**, 799–806 (2012).
- Tollervy, J. R. *et al.* Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nat. Neurosci.* **14**, 452–458 (2011).
- Müller-Hartmann, H. *et al.* The human 20-kDa 5'-(CGG)(n)-3'-binding protein is targeted to the nucleus and affects the activity of the FMR1 promoter. *J. Biol. Chem.* **275**, 6447–6452 (2000).
- Chambers, S. M. *et al.* Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nat. Biotechnol.* **27**, 275–280 (2009).
- Batra, R. *et al.* Gene Expression Signatures of Sporadic ALS Motor Neuron Populations. *bioRxiv* (Cold Spring Harbor Labs Journals, 2016).
- Ameur, A. *et al.* Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nat. Struct. Mol. Biol.* **18**, 1435–1440 (2011).
- Schwartz, J. C. *et al.* FUS binds the CTD of RNA polymerase II and regulates its phosphorylation at Ser2. *Genes Dev.* **26**, 2690–2695 (2012).
- Kwon, I. *et al.* Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell* **155**, 1049–1060 (2013).
- Bertolotti, A., Lutz, Y., Heard, D. J., Chambon, P. & Tora, L. hTAF(II)68, a novel RNA/ssDNA-binding protein with homology to the pro-oncoproteins TLS/FUS and EWS is associated with both TFIID and RNA polymerase II. *EMBO J.* **15**, 5022–5031 (1996).
- Sharma, A. *et al.* ALS-associated mutant FUS induces selective motor neuron degeneration through toxic gain of function. *Nat. Commun.* **7**, 10465 (2016).
- Sephton, C. F. *et al.* Activity-dependent FUS dysregulation disrupts synaptic homeostasis. *Proc. Natl Acad. Sci. USA* **111**, E4769–E4778 (2014).
- Budini, M. & Buratti, E. TDP-43 autoregulation: implications for disease. *J. Mol. Neurosci.* **45**, 473–479 (2011).
- Polymenidou, M. *et al.* Misregulated RNA processing in amyotrophic lateral sclerosis. *Brain Res.* **1462**, 3–15 (2012).
- Rigo, F. *et al.* Pharmacology of a central nervous system delivered 2'-O-methoxyethyl-modified survival of motor neuron splicing oligonucleotide in mice and nonhuman primates. *J. Pharmacol. Exp. Ther.* **350**, 46–55 (2014).
- Yeo, G. W. *et al.* An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat. Struct. Mol. Biol.* **16**, 130–137 (2009).
- Lovci, M. T. *et al.* Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nat. Struct. Mol. Biol.* **20**, 1434–1442 (2013).

49. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
50. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).
51. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
52. Dale, R. K., Pedersen, B. S. & Quinlan, A. R. Pybedtools: a flexible Python library for manipulating genomic datasets and annotations. *Bioinformatics* **27**, 3423–3424 (2011).
53. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
54. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
55. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
56. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
57. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
58. Untergasser, A. *et al.* Primer3—new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115 (2012).
59. Wang, X., Spandidos, A., Wang, H. & Seed, B. PrimerBank: a PCR primer database for quantitative gene expression analysis, 2012 update. *Nucleic Acids Res.* **40**, D1144–D1149 (2012).
60. Conway, A. E. *et al.* Enhanced CLIP uncovers IMP protein-RNA targets in human pluripotent stem cells important for cell adhesion and survival. *Cell Rep.* **15**, 666–679 (2016).
61. Huelga, S. C. *et al.* Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. *Cell Rep.* **1**, 167–178 (2012).

## Acknowledgements

We thank members of the Yeo lab, especially Stefan Aigner, for critical reading of the manuscript. S.C.H. and G.A.P. were funded by National Science Foundation Graduate Research Fellowships. G.A.P. was also partially supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number T32GM008666. This work was supported by grants from the National Institutes of

Health (HG004659, NS075449 and HG007005 to G.W.Y.; NS077284 to H.Z.), the California Institute of Regenerative Medicine (RB3-05009 and RB4-06045 to G.W.Y.) and ALS Association (VC8K27 to G.W.Y.; 6SE340 to H.Z.). This work was partially supported by NIH grant HG007005 to C.B.B. We thank Ionis Pharmaceuticals for sharing reagents and unpublished results. G.W.Y. is an Alfred P. Sloan Research Fellow.

## Author contributions

G.W.Y. conceived the study. S.C.H., J.P.D., L.S. and M.A. performed splicing microarray analyses. P.F., N.J.L. and C.B.B. performed the RBNS experiments. T.Y.L. performed the CLIP experiments. A.Q.V., F.J.M., J.C. and K.K. generated iPSC lines and performed neural differentiation. B.S. assisted with the RNA stability experiments. K.R.H. and G.P. performed the bioinformatics analyses. F.R. and S.C.H. performed the antisense-oligonucleotide experiments. H.Z., J.Z., F.C. and E.K. provided the *FUS* R521G fibroblasts. R.B. and J.R. contributed sporadic ALS and control patient RNA-seq data and analyses. K.K., G.A.P. and G.W.Y. analysed data and wrote the manuscript.

## Additional information

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** F.R. is a paid employee of Ionis Pharmaceuticals. The remaining authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Kapeli, K. *et al.* Distinct and shared functions of ALS-associated proteins TDP-43, FUS and TAF15 revealed by multisystem analyses. *Nat. Commun.* **7**:12143 doi: 10.1038/ncomms12143 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016