Supplemental Data

Supplementary Figures

Supplementary Figure 1 Additional PAR-CLIP quality controls. (a) Subcellular localization of wild-type and mutant FUS. C, cytoplasmic; N, nuclear; protein targets of antibodies are indicated. Quantification of subcellular distribution of FUS mutants is shown in the bar graph; dark grey, cytoplasmic localized protein; light grey, nuclear localized protein. Band intensities were measured using ImageJ and the sum of each set (C+N) was set to 100%. Tagged wild-type FUS is mostly nuclear whereas tagged mutant FUS is mostly cytoplasmic, similar to reports for untagged protein. (b) Crosslinking of RNA to FUS in PAR-CLIP experiments is dependent on the incorporation of 4SU and UV 365 nm irradiation. Four similar-sized cell pellets obtained from FLAGHA-tagged FUS expressing cell lines grown in presence or absence of 4SU and treated with or without UV 365 nm light were carried through the PAR-CLIP protocol up to the SDS-PAGE gel. 4SU, 4-thiouridine; UV 365 nm, samples crosslinked at 365 nm wavelength. (c) FLAGHA-tagged wild-type and mutant FUS expressing HEK293 T-Rex Flp-In cells show similar protein abundance. Cell lysates of obtained from the same numbers of HEK293 T-Rex Flp-In cells expressing the indicated
FLAGHA-tagged protein were separated by SDS-PAGE and Western blot analysis was performed using anti-HA and anti-beta actin antibody. Denaturing SDS-PAGE demonstrates that crosslinked RNA and FLAGHA-tagged RBP co-migrate. Aliquots of FUS and EWSR1 PAR-CLIP samples (see Figure 1a) were separated by SDS-PAGE. The blot was first analyzed by phosphorimaging visualizing radiolabeled RNA and subsequently by immunoblotting with anti-HA antibody. The lower of the double bands for EWSR1 was found by mass spectrometry analysis (data not shown) to additionally contain FUS and TAF15, suggesting that FET proteins may heterodimerize or bind in close proximity.
Supplementary Figure 2 Power analysis shows subsaturation in discovery of CCs. We performed simulations where sequence reads were randomly removed to emulate smaller sequence datasets. The number of CCs is plotted as a function of the number of uniquely mapped reads, in steps of 500,000. The plot shows the mean number of CCs obtained in three independent simulations. Error bars indicate the maximum and minimum obtained value.
Supplementary Figure 3 Binding frequencies across transcript regions for reference RBPs. (a) Distribution of CCs across transcript regions for PUM2, QKI and IGF2BP1. These are included for comparison with Figure 1e. (b) For comparison, we calculated the total number of nucleotides in 5’ UTRs, CDS and 3’ UTRs. The numbers were calculated by summarizing over all coding RefSeq transcripts.
**Supplementary Figure 4** Extended analysis of positional distribution of CCs at intron-exon junctions for various RBPs. (a) Enriched binding near splice acceptors was not observed for QKI, despite frequent binding to introns similar to FUS. CCs of ≥25% T-to-C changes were used. Due to the smaller size of the PUM2 and QKI datasets, a minimum of 5 reads per CC instead of 10 were required. The Y-axis indicates the number of observed CCs per 4 nt segment. (b) Positional distribution of CCs for FET-proteins based on size-normalized datasets. Each dataset was reduced to contain the same number of sites as QKI (n = 6822), by random sampling of CCs.
Crosslinked cluster of SON (exonic):

CAGTTTAGTTAATAATATGTGATTTAGAAAATTTATCTTTAATCTACTCAAGATCGATCGACATGACATGGTAAAT # reads error
-----------------------------------------------ATTTATCTTTAAGACTCAAG--------------------------- 55 1
-----------------------------------------------ATCTATCTTTAAGACTCAAG--------------------------- 40 1
-----------------------------------------------ATTTATCTTTAAGACTCAAG--------------------------- 36 1
-----------------------------------------------ATTTATCTTTAAGACTCAAG--------------------------- 27 1
-----------------------------------------------ATTTATCTTTAAGACTCAAG--------------------------- 25 0
-----------------------------------------------ATTTATCTTTAAGACTCAAG--------------------------- 4 1
-----------------------------------------------ATTTATCTTTAAGACTCAAG--------------------------- 3 1
-----------------------------------------------TTAATAATATGTGATTTAGAAAATCTTTAAGACTCAAG-------- 2 1
-----------------------------------------------ATCTATCTTTAAGACTCAAG--------------------------- 1 1

**Supplementary Figure 5** Representation of a PAR-CLIP CC from reads corresponding to the SON gene transcript. Reads were aligned to the genome and T-to-C changes are indicated in red. Green indicates the two complementary segments of a predicted stem-loop and orange indicates the additional nucleotides proposed to constitute the FET protein family RRE.
Supplementary Figure 6 siRNA-mediated silencing of FUS. HEK293 T-Rex Flp-In cells were transfected with three different siRNAs targeting FUS and two different control siRNAs (Applied Biosystems) using Lipofectamine RNAiMAX (all from Invitrogen) at 50 nM final concentration, each in duplicates (10 arrays in total). (a) Knockdown efficiency was assayed after 72 h by anti-FUS immunoblots. (b) Volcano plot of whole-transcriptome mRNA changes and P-values after FUS silencing. 200 ng of total RNA was assayed using U133 Plus 2.0 arrays (Affymetrix). Raw files were processed in R using the Bioconductor package and the RMA algorithm. Duplicates were averaged, redundant probe sets merged and genes were evaluated for differential expression using Student’s t-test followed by correction for multiple testing (Storey and Tibshirani’s method). No genes were significantly changed at 5% FDR, 1 gene at 10% FDR and 16 genes at 20% FDR. (c) The cumulative distribution of mRNA changes (at a zoomed-in scale compared to panel b) for transcripts with and without FUS binding sites (as determined by PAR-CLIP) reveals that FUS-bound transcripts were not notably different in their response to FUS silencing compared to unbound transcripts.
**Supplementary Tables**

**Supplementary Table 1** Annotation table of mapped PAR-CLIP sequence reads.

| Library               | Number of mapped reads (error distance 1) | Number of uniquely mapped reads | mRNA reads (%) | miRNA reads (%) | miscRNA reads (%) | piRNA reads (%) | rRNA reads (%) | snRNA reads (%) | tRNA reads (%) | other (%) |
|-----------------------|------------------------------------------|---------------------------------|----------------|----------------|------------------|----------------|---------------|----------------|---------------|-----------|
| FUS stable\(^1\)     | 4,894,941                                | 2,007,928                       | 38.77          | 0.21           | 2.6              | 0.04           | 28.51         | 0.78           | 0.25          | 28.84     |
| FUS inducible\(^2\)  | 5,688,260                                | 2,064,289                       | 40.49          | 0.02           | 2.46             | 0.05           | 6.47          | 0.24           | 0.11          | 50.18     |
| EWSR1 stable         | 4,348,695                                | 2,153,122                       | 31.88          | 0.05           | 2.99             | 0.03           | 54.29         | 0.72           | 0.46          | 9.59      |
| EWSR1 inducible      | 2,885,147                                | 1,308,001                       | 29.51          | 0.13           | 4.62             | 0.03           | 47.21         | 0.74           | 1.24          | 16.53     |
| TAF15 stable         | 3,780,613                                | 1,324,082                       | 33.25          | 0.21           | 2.94             | 0.04           | 26.43         | 1.83           | 1.49          | 33.8      |
| TAF15 inducible      | 4,008,522                                | 1,332,335                       | 46.84          | 0.09           | 2.37             | 0.03           | 10.33         | 1              | 0.78          | 38.55     |
| FUS-R521G            | 4,271,115                                | 1,792,475                       | 39.68          | 0.24           | 3.78             | 0.04           | 37.65         | 0.89           | 0.91          | 16.8      |
| FUS-R521H            | 1,826,791                                | 850,452                         | 34.84          | 0.26           | 4.51             | 0.03           | 47.68         | 1.43           | 1.42          | 9.83      |

\(^1\)refers to cell lines constitutively expressing the indicated protein

\(^2\)refers to cell lines which express the indicated protein after induction with doxycycline
Supplementary Table 2 Normalization of target gene counts. Datasets were reduced by random sampling of reads. Read numbers were individually tuned for each dataset to equalize the number of targeted genes. Two size normalizations were performed; one where the number of target genes was dictated by EWSR1 ("Resampled to EWSR1 gene count"), and a more radical size reduction dictated by TAF15 ("Resampled to TAF15 gene count"). Size reduction decreased overall overlaps, meaning that the reduced datasets had higher fractions of uniquely targeted genes.

| Dataset | Protein | FUS | EWSR1 | TAF15 | Mut. FUS | FUS | EWSR1 | TAF15 | Mut. FUS | FUS | EWSR1 | TAF15 | Mut. FUS |
|---------|---------|-----|-------|-------|----------|-----|-------|-------|----------|-----|-------|-------|----------|
| Uniquely mapped reads | | 4069214 | 3458882 | 3443557 | 2642445 | 1900000 | 3458882 | 3443557 | 2500000 | 1400000 | 2500000 | 3443556 | 1500000 |
| Targeted genes | | 6845 | 4488 | 3113 | 4732 | 4563 | 4488 | 3113 | 4536 | 3083 | 3159 | 3113 | 3084 |
| % unique targets (not targeted by FUS) | | 14.5% | 13.4% | 19.4% | 32.1% | 26.8% | 37.7% | 43.3% | 44.7% | 47.2% |
**Supplementary Table 3** Gene ontology (GO) analysis of genes uniquely bound by mutant FUS (not bound by the wild-type protein), as depicted in Figure 1d.

| GO Term             | Unique mut. FUS targets | Genome | Fold enrichment | P-value     | Description                                                        |
|---------------------|-------------------------|--------|-----------------|-------------|-------------------------------------------------------------------|
| GO:0005515          | 43.4%                   | 34.6%  | 1.3             | 7.48E-08    | protein binding                                                  |
| GO:0005739          | 10.0%                   | 5.6%   | 1.8             | 3.65E-07    | mitochondrion                                                    |
| GO:0005829          | 10.1%                   | 6.0%   | 1.7             | 2.96E-06    | cytosol                                                           |
| GO:0005634          | 34.5%                   | 27.4%  | 1.3             | 3.27E-06    | nucleus                                                           |
| GO:0005737          | 29.4%                   | 22.9%  | 1.3             | 8.39E-06    | cytoplasm                                                        |
| GO:0005783          | 8.1%                    | 4.7%   | 1.7             | 1.23E-05    | endoplasmic reticulum                                            |
| GO:0000502          | 1.0%                    | 0.1%   | 6.8             | 1.36E-05    | proteasome complex                                               |
| GO:0031145          | 1.5%                    | 0.4%   | 4.1             | 2.67E-05    | anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process |
| GO:0051436          | 1.5%                    | 0.4%   | 4.1             | 2.67E-05    | negative regulation of ubiquitin-protein ligase activity during mitotic cell cycle |
| GO:0005743          | 3.3%                    | 1.4%   | 2.4             | 3.43E-05    | mitochondrial inner membrane                                      |
| GO:00051437         | 1.5%                    | 0.4%   | 4               | 4.15E-05    | positive regulation of ubiquitin-protein ligase activity during mitotic cell cycle |
| GO:0009313          | 0.4%                    | 0.0%   | 23.8            | 7.36E-05    | oligosaccharide catabolic process                                 |

1 fraction of genes uniquely bound by mutant FUS (916) annotated with a specific GO term

2 background fraction (all ENSEMBL genes) for a specific GO term
Supplementary Methods

Oligonucleotides and siRNA duplexes

The following oligodeoxynucleotides were used for plasmid preparation, mutagenesis reactions and sequencing:

FLAGHA_BamHI_for,
5′GATCGACCGGTGACTACAAGGACGACGATGACAAGTACCCTTATGACGTGCCC
GATTACGCTG;

FLAGHA_BamHI_rev,
5′GATCCAGCGTAATCGGGCACGTCATAAGGGTACTTGTCATCGTCGTCTTTGTAG
TCACCGGTG;

FUS_pET23a_SalI_for, 5′ACGCGTCGACCCATGGCCTCAAACGATTATACCC;

FUS_pET23a_NotI_rev, 5′ATAAGTTTAGCGGCGCATAACGGCCCTCTCCCTGCGATC;

FUS_NcoI_for, 5′CAATCCCATGGACTACAAGGACGACGATGAC

FUS_EcoRV_rev, 5′GCAATCGATATCTCAGTGGTGGTGGTGTTGTTG;

FUS_PCR_for, 5′ACGCGTCGACATGGCCTCAAACGATTATACCC;

FUS_PCR_rev, 5′ATAAGTTTAGCGGCGCATAACGGCCCTCTCCCTGCGATC;

EWSR1_PCR_for, 5′ACGCGTCGACATGGCCTCAAACGATTATACCC;

EWSR1_PCR_rev, 5′ATAAGTTTAGCGGCGCATAACGGCCCTCTCCCTGCGATC;

TAF15_PCR_for, 5′ACGCGTCGACATGGCCTCAAACGATTATACCC;

TAF15_PCR_rev, 5′ATAAGTTTAGCGGCGCATAACGGCCCTCTCCCTGCGATC;

FUS_C521G_for, 5′GAGCACAGACAGGATGGCAGGGAGAGG;

FUS_C521G_rev, 5′CCTCTCCCTGCGCATACTCCTGCTGGTCTG;

FUS_G521A_for, 5′CACAGACAGACAGGAGGGAGGGAGGCG;
The following oligoribonucleotides were used for FUS protein in vitro binding studies:

\[ \text{AUU_trinucleotide_repeat}, \]
5′AUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU;

\[ \text{GGU_trinucleotide_repeat}, \]
5′GGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGU;

\[ \text{SON_AB}, 5′\text{GAUUUAUCUUAACUCAAGAUACUGAACAUGACA}; \]
\[ \text{SON_AA}, 5′\text{GAUUUAUCUUAACUCAAGAUACUGAACAUGACA}; \]
\[ \text{SON_BB}, 5′\text{GAUUAAGAUAUAACUCAAGAUACUGAACAUGACA}; \]
\[ \text{SON_BA}, 5′\text{GAUUAAGAUAUAACUCAAGAUACUGAACAUGACA}; \]
\[ \text{SON_UA_shifted}, 5′\text{GAUUUAUCUUCUUAACUCAAGAUACUGAACAUGACA}; \]
\[ \text{SON_no_UA_in_loop}, 5′\text{GAUUUAUCUUCUCAACUCAAGAUACUGAACAUGACA}; \]
\[ \text{SON_1st_U_deleted}, 5′\text{GAUUUAUCUUCUACUCAAGAUACUGAACAUGACA}; \]

Pre-annealed siRNA duplexes were purchased from Applied Biosystems: FUS (s5402, s5403, s5401), Silencer Select Negative Control #1 siRNA (#4390843).

**Plasmids for the creation of stable cell lines**

Plasmids pENTR4_FUS, _EWSR1 and _TAF15 were generated by PCR amplification of the respective coding sequences (CDS) followed by restriction digestion with SalI and NotI and ligation into pENTR4 (Invitrogen) (primers: FUS_PCR_for, FUS_PCR_rev, EWSR1_PCR_for, EWSR1_PCR_rev, TAF15_PCR_for, TAF15_PCR_rev). pENTR4_FUS, _EWSR1 and _TAF15 were recombined into both pFRT_TO_DESTFLAGHA and
pFRT_DESTFLAGHA modified destination vectors (Invitrogen) using Gateway LR Clonase II enzyme mix according to the manufacturer's instructions (Invitrogen) to allow for both doxycycline dependent and independent (constitutive) expression of FLAGHA-tagged protein in stably transfected Flp-In T-REx HEK293 cells (Invitrogen). The mutagenesis reactions to create pENTR4_FUS_R521H and pENTR4_FUS_R521G were performed using the QuikChange II XL Site-Directed Mutagenesis kit according to the manufacturer’s instructions (Stratagene). The plasmids described in this study can be obtained from Addgene (www.addgene.org).

**Plasmids for protein expression**

The pET23(a) vector (Novagen, #69745) was modified to contain an N-terminal FLAGHA tag aside from its encoded C-terminal His6 tag yielding pET23(a)_mod. To achieve this, the pET23(a) plasmid was first digested with BamHI followed by the ligation of the pre-annealed oligodeoxynucleotides FLAGHA_BamHI_for and FLAGHA_BamHI_rev. Next, PCR amplification using FUS_pET23a_SalI_for and FUS_pET23a_NotI_rev from pENTR4_FUS yielded the coding sequence (CDS) without the stop codon of FUS. The PCR product was SalI and NotI digested and ligated into the SalI and NotI digested pET23(a)_mod vector. Another PCR was performed (primers FUS_NcoI_for and FUS_EcoRV_rev) and the FLAGHA_FUS_His6 cDNA was amplified. A regular pENTR4 vector (Invitrogen) was NcoI and EcoRV digested. A fill-in reaction with T4 DNA polymerase was performed to create blunt ended products of both the FLAGHA_FUS_His6 cDNA and the pENTR4 vector. Then, the FLAGHA_FUS_His6 cDNA was ligated into the pENTR4 vector. The pENTR4_FLAGHA_FUS_His6 construct was recombined into pDEST8 destination vector using GATEWAY LR recombinase according to manufacturer’s protocol (Invitrogen).
Mammalian cell culture and creation of stable cell lines

T-REx HEK293 Flp-In cells (Invitrogen) were grown in D-MEM high glucose (1x) with 10% fetal bovine serum, 100 U per ml of penicillin, 100 µg per ml streptomycin, 100 µg per ml zeocin and 15 µg per ml blasticidin. Cell lines constitutively or inducibly expressing FLAGHA-tagged proteins were generated by co-transfection of pFRT_TO_FLAGHA_GOI (gene of interest) or pFRT_FLAGHA_GOI constructs with pOG44 (Invitrogen) using lipofectamine 2000 (Invitrogen). Cells constitutively or inducibly expressing FUS, EWSR1, TAF15, FUS-R521G or FUS-R521H were cultivated in D-MEM high glucose (1x, Invitrogen) with 10% fetal bovine serum, 100 U per ml of penicillin, 100 µg per ml streptomycin and 100 µg per ml hygromycin (Invivogen). In the case of inducible expression 15 µg per ml blasticidin (Invivogen) was also added to this medium. Induction was achieved by adding 1 µg per ml doxycycline to the growth medium 15 to 20 h before crosslinking.

Insect cell culture, recombinant protein expression in Spodoptera frugiperda (Sf9) cells and protein purification

Sf9 cells were grown in Grace’s Insect Medium (Invitrogen, #11605-094), supplemented with 10% fetal bovine serum, 1% Pluronic F-68 (Invitrogen, #24040-032), 100 U per ml of penicillin, 100 µg per ml streptomycin and maintained in room air at 26°C in spinner culture (80 rpm). The pDEST8_FLAGHA_FUS_His6 was transformed into MAX Efficiency DH10Bac competent *E. coli* (Invitrogen). Bacmid DNA was isolated using PureLink HQ Mini Plasmid Purification and transfected into Sf9 cells using Cellfectin II Reagent kit (all from Invitrogen). Three rounds of viral amplifications yielded 250 ml of cell supernatant containing 2 x 10^8 plaque forming units per ml virus. 25 ml of this solution were used for infection of one liter of Sf9 culture maintained at a density of 1 x 10^6 cells per ml. Four days after infection, Sf9 cells were washed by centrifugation (500 x g) in 1x PBS, and pellets were
suspended in 5-times the pellet volume of ice-cold lysis buffer (50 mM Tris-HCl, pH 8.0, 5 mM MgCl₂, 1 M KCl, 10% glycerol, 5 mM imidazol, 0.1% triton-X-100, 1 mM beta-mercaptoethanol and Complete EDTA-free protease inhibitor cocktail (Roche)). The suspension was incubated on ice for 10 min and then additionally suspended by 20 strokes with a Dounce homogenizer. Insoluble material was removed by centrifugation for 20 min at 20,000 x g and the supernatant was further cleared by passing through a 5µm Supor membrane syringe filter (Pall Acrodisc). Sf9-expressed FUS was purified using the AektaExplorer (Amersham Bioscience). 10 ml of Co²⁺ TALON beads were washed 3 times with de-ionized water and packed into a XK 16 column (Amersham Bioscience). The column was equilibrated with 4 column volumes (CV) lysis buffer supplemented with 5 mM imidazole and 1 mM DTT. The cell lysate was loaded onto the equilibrated Co²⁺ TALON column (GE Healthcare Life Sciences XK 16/20 column with an AK16 adapter) using a flow rate of 1 ml per min. The column was washed with 2 CV lysis buffer supplemented with 5 mM imidazole and 2 CV supplemented with 13 mM imidazole. The protein was eluted from the column in 4 CV elution buffer (50 mM Tris-HCl, pH 8.0, 5 mM MgCl₂, 1 M KCl, 10% glycerol, 400 mM imidazol, 0.1% triton-X-100, 1 mM beta-mercaptoethanol) running a gradient with a final concentration of 400 mM imidazole. During elution, 1 ml fractions were collected and analyzed on SDS-gels. Fractions containing FUS were pooled and dialyzed overnight in 2 times 1 l of dialysis buffer (20 mM Tris-HCl, pH 8.0, 300 mM KCl, 5 mM MgCl₂, 0.1% v per v Triton X-100, 50% v per v glycerol, 1 mM DTT) using dialysis bags (Spectrum, SpectraPor, 08-667E) with a molecular cutoff of 12 to 14 kDa. Protein concentrations were estimated by comparing Coomassie stain intensity against a BSA standard on a 10% SDS-gels.
Electrophoretic mobility shift assay (EMSA)

10 pmol oligonucleotide were labeled with 5 pmol [γ-32P]-ATP in a 10 µl reaction containing 70 mM Tris-HCl (pH 7.6 @ 25), 10 mM MgCl2, 5 mM DTT, and 5 U T4 PNK. The reaction was denatured (95°C, 30 sec), and placed on ice. After 1 min, 5 U T4 PNK was added and the reaction was incubated at 37°C for 15 min. After 15 min, regular ATP was added to a final concentration of 1 mM and the reaction was incubated for an additional 5 min at 37°C. The reaction volume was increased to 50 µL and denatured at 95°C for 30 sec. The unincorporated [γ-32P]-ATP was removed by passing the reaction mixture through a G25 column (GE Life Science). The eluate volume was increased to 100 µl and the 100 nM oligonucleotide was stored at -20°C. In a 20 µl reaction, 1 nM labeled RNA was incubated with protein concentrations varying from 0-1 µM in buffer containing 20 mM Tris-HCl (pH 7.65 @ 25°C), 300 mM KCl, 5 mM MgCl2, 1 mM DTT, 0.1U per µl RNasin (Promega), 100 ng yeast tRNA, and 0.1 mg per ml acetylated BSA at 30°C for 1 hour in 1.5 ml passivated (50 µL 1 mg per ml acetylated BSA, 27°C, 1 hour), siliconized eppendorf tubes. After 1 hour, 5 µl of buffer with 50% glycerol and bromophenol blue was added. Following a 30 min pre-run, the reaction was separated by native page (25 mM Tris, 0.2 M Glycine, 6% 49:1 acrylamide:bisacrylamide, ammoniumpersulfat, TEMED), at 4°C at 300V in 25 mM Tris 0.2 M Glycine containing buffer. The reaction was loaded onto running gels, and the species were separated for 2 hours at 300 V at 4°C. The [γ-32P] radioactive signal was detected using phosphorimager screens and the signal was quantified using ImageGauge software. Curves and binding constants were calculated using Kaleidagraph software. Contrast and brightness of images in Figure 2 was adjusted, equally across the frame, in Photoshop (Adobe).
Preparation of whole cell extracts and Western blotting

For whole cell mammalian lysates, cells were washed with ice-cold PBS and lysed in 3 pellet volumes 10 mM HEPES-KOH, pH 7.5, 150 mM KCl, 2 mM EDTA, 1 mM NaF, 0.5% NP-40, 0.5 mM DTT and complete EDTA-free protease inhibitor cocktail (Roche). The lysate was incubated for 10 min on ice and cleared for 10 min at 20,000 x g and 4°C. Total protein concentration was measured by Bradford protein assay (Biorad). Whole cell lysates were analyzed on a 10% SDS-gel. Protein samples were then transferred to nitrocellulose membrane (BioRAD, Trans-Blot; 1.5 mAmp per cm² membrane for 1.5 hrs) and probed with the indicated antibodies. Signals were developed using the ECL kit (Amersham) under standard conditions. The luminescence signal was recorded with a Fujifilm Image Reader LAS-3000. Contrast and brightness of presented images was adjusted in Photoshop (Adobe) (adjustments were applied equally across the frames).

Preparation of nuclear extracts

Cells were harvested by trypsinization and centrifuged for 5 min at 2,000 x g. All following steps were performed at 4°C. The cell pellet was resuspended in 1x PBS and centrifuged for 5 min at 2,000 x g. Next, the packed cell volume (PCV) was recorded and the cells were resuspended in 5x PCV Hypotonic Lysis Buffer (10 mM HEPES pH 7.9 (KOH), 10 mM KCl, 1.5 mM MgCl2, 0.5 mM DTT, Complete EDTA-free protease inhibitor cocktail (Roche)) and incubated on ice for 10 min. After a further centrifugation step (5 min at 2,000 x g) the pellet was resuspended in 2x PCV Hypotonic Lysis Buffer. The suspension was homogenized with 5 strokes in a Dounce glass homogenizer (type B pestle) and cell lysis was ensured microscopically. The lysate was centrifuged for 10 min at 2,000 x g, the supernatant was saved as the cytoplasmic extract (further centrifuged at 13,000 x g for 30 min) and the exact volume was recorded. To wash the nuclei, they were resuspended in 2x PCV Hypotonic
Lysis Buffer and further centrifuged for 10 min at 2,000 x g. After discarding the supernatant, the nuclei were resuspended in 1x SDS sample buffer so that the final volume equaled that of the cytoplasmic extract, sonicated for 15 sec and boiled for 3 min. Purity of the fractions was tested by probing with anti-lamin and anti-tubulin antibodies.

**Antibodies**

Monoclonal anti-HA.11 (clone 16B12, Covance), polyclonal anti-FUS (Abcam, AB23439), beta-tubulin (Sigma, T4026), beta-actin (Sigma, SAB3500350) and lamin C (Abcam, AB16048) were used as primary antibodies at a 1:1000 dilution. Anti-Flag M2 (Sigma, F3165) was used for PAR-CLIP. HRP-conjugated anti-rabbit Ig and anti-mouse Ig (both from DAKO) were used as secondary antibodies for Western blot analysis.

**PAR-CLIP**

PAR-CLIP was performed as described before. Briefly, the growth medium of HEK293 T-REx Flp-In cells expressing FLAGHA-tagged FET proteins was supplemented with 100 μM 4SU for 12 hours prior to crosslinking. After decanting the growth medium, cells were irradiated uncovered with 0.15 J per cm2 total energy of 365 nm UV light in a Stratalinker 2400. Cells were harvested at 500 x g and lysed in 3 cell pellet volumes of NP40 lysis buffer (50 mM HEPES, pH 7.5, 150 mM KCl, 2 mM EDTA, 1 mM NaF, 0.5% (v per v) NP40, 0.5 mM DTT, complete EDTA-free protease inhibitor cocktail (Roche)). After centrifugation at 13,000 x g the cleared cell lysate was treated with RNase T1. FLAGHA-tagged FET proteins were immunoprecipitated with an anti-FLAG antibody conjugated to Protein G Dynabeads (Invitrogen). After a second RNase T1 digestion, beads were washed in high-salt wash buffer (50 mM HEPES-KOH, pH 7.5, 500 mM KCl, 0.05% (v per v) NP40, 0.5 mM DTT, complete EDTA-free protease inhibitor cocktail (Roche)) and resuspended in dephosphorylation buffer (50 mM Tris-HCl, pH 7.9, 100 mM NaCl, 10 mM MgCl2, 1 mM DTT). Following
incubation with calf intestinal alkaline phosphatase beads were washed again. Next, the crosslinked RNA was radiolabelled using T4 polynucleotide kinase. After a final wash step, RNA–FET complexes were released from the beads by incubating at 90°C and subsequently separated on SDS-gels. The excision of the bands corresponding to the expected masses of the proteins was followed by the electroelution of the RNA–FET complexes. Following proteinase K digestion, RNA was recovered from the eluate by acidic phenol-chloroform extraction and ethanol precipitation. The 5′-32P-phosphorylated RNA was then carried through a standard cDNA library preparation protocol4. Both the 3′ and the 5′ Solexa adapters were ligated to the RNA followed by its reverse transcription. The resulting cDNA was amplified by PCR and the final PCR product was Solexa sequenced.

Processing of PAR-CLIP reads

The raw sequencing data was processed as previously described3. Briefly, after the removal of the Solexa adapters sequences that were either too short (less than 20 nucleotides) or too repetitive (for exact scoring parameters please refer to5) were discarded. The remaining sequences were mapped against the human genome (NCBI36 hg18 assembly) while allowing at most one error. Non-uniquely mapping reads were discarded and clusters were built from overlapping reads, requiring at least ten reads per cluster and at least one T to C change ("crosslinked clusters" or "CCs").

Annotation

Human genomic coordinates of RefSeq transcripts, including exon-intron boundaries and CDS start or stop positions, were obtained from the UCSC browser6. Transcripts in unfinished genomic segments or segments with unknown location (‘_random’ chromosomes) were disregarded, as were transcripts that were not uniquely mapped so a single locus. Center positions of CCs were matched to transcript coordinates for the purpose of gene annotation,
and to determine whether binding took place in the 5′ UTR, CDS, 3′ UTR or intronic regions. Due to multiple overlapping transcript isoforms, a minor fraction could be assigned a single such region and was thus classified as ‘ambiguous’. Likewise, CCs that could not be assigned to a single specific gene were annotated as ‘ambiguous’ and disregarded in subsequent gene-centric analyses, as were clusters that matched to ribosomal RNA (‘miscRNA’ track in the UCSC browser). The annotation procedure was initially performed on individual reads, for the purpose of the clustering analysis (see below), and later on the final CCs.

Hierarchical clustering

Unfiltered binding profiles for each dataset were calculated by summarizing the total number of reads in each gene. These profiles were normalized based on the mean intensity in each dataset. Correlations between binding profiles were calculated based on the 5000 most variable genes as defined by the standard deviation across all datasets and using Spearman correlation as distance metric. Hierarchical clustering was performed using average linkage. After establishing reproducibility of replicates and uniqueness of the different proteins, replicate reads were pooled and clusters built and annotated as described above.

Site-level overlap analysis

We considered the top 1000 CCs in each datasets, as defined by the number of uniquely mapped reads and with the usual requirement of 25% T-to-C-containing reads. This approach simplifies downstream interpretation, as datasets that were originally of different sizes are thereby reduced to smaller and identically sized sets of high-ranking CCs. The methodology is similar to a previously described approach. We screened for pairs of such CCs were each CC was from a different datasets and where center positions were within 10 nt of each other. These were considered to represent the same binding site. A small number of CCs with very high read counts (>10,000, maximum 3 cases in each pair wise comparison) were disregarded.
in this analysis as they contributed positively to correlation values while potentially being the products of erroneous short read mapping. The analysis was also repeated, with a similar outcome, while only considering exonic sites (not shown).

**Cluster density plots**

To determine the frequency of binding near intron-exon and exon-intron junctions, all individual exons were extracted from RefSeq transcripts. For each unique junction, all CCs binding with ± 200 bp were identified and the frequency of binding for each 4 bp segment in this 400 bp region was determined.

**Gene ontology analysis**

Gene Ontology (GO) annotations were obtained from ENSEMBL using the BioMart tool\(^8\), and enrichment of GO terms was evaluated statistically using Fisher’s exact test. Genes that could not be mapped to the ENSEMBL GO annotation were excluded from the analysis. To account for multiple testing (30164 GO terms were evaluated), the observed P-values were compared to a simulated null-distribution (repeated scrambling of gene identities). A P-value of 1e-4 was found to be useful as a conservative threshold, as no false positives were observed at this level.

**Motif discovery**

To elucidate the RRE of FET proteins, we analyzed CCs as well as CCs with ≥2 T-to-C positional changes, as these could potentially show a stronger motif signal. The CCs were enriched for A and T as compared to random intronic regions, but use of standard bioinformatic tools\(^9,10\) as well as screening for short (4 nt) overrepresented sequence patterns (relative to mono- or dinucleotide shuffled sequences) did not return a significant motif.
Stem-loop structures were mapped using the Matlab Bioinformatics Toolbox, and were required to have perfectly complementary stems of at least 3 bp and loops of 3 bp or longer.
Supplemental References

1. Kwiatkowski, T.J., Jr. et al. Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science* **323**, 1205-8 (2009).
2. Vance, C. et al. Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. *Science* **323**, 1208-11 (2009).
3. Hafner, M. et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129-41 (2010).
4. Hafner, M. et al. Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. *Methods* **44**, 3-12 (2008).
5. Berninger, P., Gaidatzis, D., van Nimwegen, E. & Zavolan, M. Computational analysis of small RNA cloning data. *Methods* **44**, 13-21 (2008).
6. Kent, W.J. et al. The human genome browser at UCSC. *Genome Res* **12**, 996-1006 (2002).
7. Kishore, S. et al. A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nat Methods*.
8. Smedley, D. et al. BioMart--biological queries made easy. *BMC Genomics* **10**, 22 (2009).
9. Bailey, T.L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2**, 28-36 (1994).
10. Larsson, E., Lindahl, P. & Mostad, P. HeliCis: a DNA motif discovery tool for colocalized motif pairs with periodic spacing. *BMC Bioinformatics* **8**, 418 (2007).