Position-specific binding of FUS to nascent RNA regulates mRNA length

Akio Masuda[1], Jun-ichi Takeda[1], Tatsuya Okuno[1], Takaaki Okamoto[1], Bisei Ohkawara[1], Mikako Ito[1], Shinsuke Ishigaki[2], Gen Sobue[2], Kinji Ohno[1]

[1]Division of Neurogenetics, Center for Neurological Diseases and Cancer, and
[2]Department of Neurology, Nagoya University Graduate School of Medicine, Nagoya, Japan
65 Tsurumai, Showa-ku, Nagoya 466-8550, Japan

Supplemental Information

Contents

Supplemental Figure legends

Supplemental Fig. S1. Positional analysis of FUS-CLIP tags and knockdown of *Fus*.

(A) FUS-CLIP clusters normalized by the total number of nucleotides in each region.

(B) Average positional FUS-CLIP tag density in 648 introns longer than 100 kb. The mean and 95% confidence interval are indicated. *p*-value is calculated using Kruskal-Wallis test followed by Bonferroni post hoc test.

(C) FUS-dependent intron conservation for genes encoding RNA-binding proteins. Conservation scores, which are according to phyloP of 30 vertebrate species, were obtained from the UCSC Genome Bioinformatics. Average conservation scores of 6471 introns (> 200 nt) for genes encoding RNA-binding proteins are classified into three categories depending on the CLIP-tag density on the intron. Similar to the previous report (Nakaya et al. 2013), we excluded intronic regions that give rise to noncoding RNAs. In the phyloP plots, sites predicted to be conserved are given positive phyloP scores, while sites predicted to be fast-evolving are given negative scores. The mean and 95% confidence interval are indicated. *p*-value is calculated by Kruskal-Wallis test followed by Bonferroni post hoc test.

(D) Western blots showing efficiencies of two siRNAs (siFus-1 and siFus-2) in N2A cells.

Supplemental Fig. S2. FUS facilitates local accumulation of RNAP II and suppresses nascent transcripts.

(A) Shift of traveling ratio (TR) by *Fus*-knockdown. TR measures the ratio of the density of RNAP2 near the TSSs (-300 to +100 nt) over that for the rest of the gene bodies. In both siCont- and siFus-treated cells, 4,671 genes carry more than 200 RNAP II ChIP-tag coverages on both the TSS and the gene body. TRs are calculated using the 4,671 genes. The mean and 95% confidence interval are indicated. Upon loss of FUS, TR is significantly increased for the analyzed genes (p-value = 3.84 x 10^{-23} by paired Wilcoxon test).

(B) Real-time PCR quantification of ChIP-DNA precipitated with antibodies for the whole RNAP II (N20 antibody) and Ser2-phosphorylated CTD (H5 antibody) normalized for that with control antibody or input DNA. All experiments are triplicated, and the mean and SD are indicated. * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$ by two-way ANOVA and post-hoc Bonferroni test.

Supplemental Fig. S3. FUS binds to RNA segments around APA sites and regulates activation of APA in a position-dependent manner

(A) Pie chart showing distribution of reads in CAGE-seq (CAGE) and PolyA-seq (PolyA) on genomic features. CDS, coding sequence.

(B) Pie chart showing the frequencies of alternative polyA sites (APA), alternative splice sites (alt SS), alternative TSSs (alt TSS), and constitutive splice sites (cons SS) within (Cluster) or outside (non-Cluster) of FUS-CLIP-clustered regions. Each APA, alt SS, alt TSS, and cons SS is categorized into either 'Cluster' or 'non-Cluster' group. The values denote the numbers of each site in Cluster and non-Cluster regions. $p < 2.2e-16$ by Fisher's exact test (table 2 x 4).

(C) Normalized complexity map of FUS-CLIP tags of the mouse brain CLIP experiments (Lagier-Tourenne et al. 2012) around APA sites detected in *Fus*-knockout or *Fus*-knocked down mouse brain (Lagier-Tourenne et al. 2012).
(D) FUS-dependent RNAP II accumulation (ChIP) around all annotated APA sites. The RNAP II ChIP-seq data are divided into 4 groups based on the number of FUS-CLIP tags in regions spanning 500-nt upstream and 500-nt downstream of the target sites, and FUS-dependent RNAP II accumulation is plotted for each group.
(E) Two representative APA sites that are upregulated (*Zbtb24*) or downregulated (*Ewsr1*) by *Fus*-knockdown. Total RNAs are reverse-transcribed using oligo-dT primer and real-time RT-PCR quantification is performed (bottom graphs). $*p < 0.05$ and $**P < 0.01$ by t-test compared to siCont. Gene structures are shown on the top. Primer positions are indicated by blue arrowheads. Distribution of FUS-CLIP tags (red) and RNAP II-ChIP tags (green) are shown below the gene structures.
(F) FUS enhances polyadenylation and transcription termination of the APA site in *Ewsr1* minigene. *Fus* mRNA is knocked down in N2A cells (siFus-2) or overexpressed in HEK293 cells (Fus) along with overexpression of the *Ewsr1*-WT minigene shown in Fig. 3C. Ratios of polyA-tailed RNA to total RNA are calculated by real-time RT-PCR of nuclear RNA using oligo-dT primer (polyA-tailed) and gene-specific primer (total RNA). Ratios of the amount of downstream RNA to upstream RNA are quantified by real-time RT-PCR. Schematic of the plasmid construct harboring the APA site and the flanking regions of *Ewsr1* are indicated at the top. Primer positions are also indicated by grey arrowheads. Although these primer sets can amplify transcripts from both the minigene and the endogenous gene, the transcript from the minigene was at least 100 times more than that from the endogenous gene according to real time RT-PCR. $*p < 0.05$ and $**p < 0.01$ by t-test.
(G) 3' RACE using nested RT-PCR to detect polyadenylated mRNA in *Fus*-suppressed N2A cells and *Fus*-overexpressed HEK293 cells transfected with the *Zbtb24*-WT-minigene or the *Ewsr1*-WT-minigene shown in Fig. 3C. Schematic of the minigenes and location of the primers for 3' RACE analysis are shown in Fig. 3C.
(H) Sequencing chromatograms of the 3' RACE products for the APA transcripts of the *Zbtb24*-WT-minigene (left) and the *Ewsr1*-WT-minigene (right) are shown with the corresponding genomic sequences in black. The polyA tails are shaded in green and the alternative PASs are indicated by blue boxes.

Supplemental Fig. S4. Binding of FUS to nascent RNA downstream of PAS terminates transcription and promotes polyadenylation.

(A) Position-specific regulation of the luciferase activity of minigene with the *Ewsr1*-APA site by FUS-MS2 fusion protein (red bar). Mutations in PAS abolish the enhancing effect (lower panel). Renilla luciferase activities are normalized to that of co-transfected firefly luciferase activity. $**p < 0.01$ and $***p < 0.001$, compared to relative luciferase activity of EGFP+MS2 (dark green) by one-way ANOVA and post-hoc Tukey test. Color code for MS-coat fusion proteins are indicated in Fig. 4A.

(B) Tethered FUS (red bar) downstream of *Gapdh*-PAS terminates transcription and enhances polyadenylation, whereas upstream tethering of FUS only terminates transcription. The copy number of a downstream transcript is estimated with primer pair RG. Nuclear RNA was reverse transcribed using oligo-dT primer (polyA-tailed) and gene-specific primer (total RNA). Real-time RT-PC was performed with primer pair RLUC, and ratios of polyA-tailed RNA to total RNA are calculated. $*p < 0.05$ and $**p < 0.01$ by t-test.

(C) Tethered FUS (red bar) downstream of the *Ewsr1*-APA site enhances polyadenylation. Ratios of polyA-tailed RNA to total RNA are calculated as indicated in *(B)*. $**p < 0.01$ by t-test.

Supplemental Fig. S5. FUS interacts with CPSF160, and promotes binding of CPSF160 to PAS during stalling of RNAP II.

(A) Time course analysis of distribution of RNAP II, FUS and 3' end processing factors in N2A cells treated with an RNAP II inhibitor, DRB. Cells are harvested and fractionated at the indicated time points followed by immunoblotting with the indicated antibodies as in Fig. 5A.

(B) Distribution of RNAP II, CPSF160, and FUS in N2A cells treated with an RNAP II inhibitor, α-amanitin (10 μg/ml). Cells are harvested and fractionated at the indicated time points followed by immunoblotting with the indicated antibodies, as in Fig. 5A in which RNAP II is inhibited by DRB.

(C) Nascent RNA is required for the interaction between RNAP II and FUS. N2A cells treated with an RNAP II inhibitor, α-amanitin (10 μg/ml), for the indicated durations are co-immunoprecipitated using anti-RNAP II antibody followed by immunoblotting.

(D) DRB increases the amount of CPSF160 bound to FUS. N2A cells treated with indicated concentrations of DRB for 2 hrs are co-immunoprecipitated using anti-FUS antibody followed by immunoblotting with the indicated antibodies.

(E) Slow-processing C4 RNAP II (C4) increases CPSF160 bound to FUS. N2A cells are transfected with plasmids expressing α-amanitin-sensitive wild-type (WT_S), α-amanitin-resistant slow C4, and α-amanitin-resistant wild-type (WT_R) RNAP II. α-Amanitin (10 μg/ml) is added to the medium to inhibit endogenous RNAP II at 8 hrs after transfection, and cells are incubated for additional 40 hrs. Total cell lysates are co-immunoprecipitated with anti-FUS antibody.

(F) Primer positions (RIP) and an RNA-probe used in Figs. 5D and E. Schematic of the *Ewsr1*-WT minigene (Fig. 3C) and the flanking regions of *Ewsr1* are shown.

(G) Tethered FUS downstream (right panel) but not upstream (left panel) of an APA site enhances binding of CPSF160 to PAS in HEK293 cells. MS2-tethering system shown in Fig. 4A was transfected in HEK293 cells. Ratios of co-immunoprecipitated RNA to input RNA are quantified by real-time RT-PCR. *$p < 0.05$ by t-test.

Supplemental Fig. S6. FUS regulates mRNA length.

(A) FUS-RNA interaction (right graph) correlates with difference of fractional mRNA length by *Fus*-knockdown (left graph). Normalized CLIP-coverage, which was calculated as in Fig. 6C, are plotted against three categories of Refseq genes that are evenly divided according to difference of fractional mRNA length. Box denotes 25th and 75th percentiles; line within box denotes 50th percentile; whiskers denote values of the 75th percentile + 1.5 times the vertical distance covered by the box. *p*-value is calculated using Kruskal-Wallis test followed by Steel-Dwass post hoc test.

(B-D) GSEA analysis of one of the ten gene sets (Figs. 6D, E), for which FUS-CLIP clusters are located from positions -100 to 0 nt. upstream of the APA site. The TSSs and polyA sites are detected by CAGE-seq and PolyA-seq, respectively, in 7377 RefSeq gene. Assuming that the length of the entire gene region of a specific gene is 100%, the fractional mRNA lengths of the 7377 transcripts after siCont and siFus treatments were calculated. The difference of the fractional mRNA lengths between siFus-treated and siCont-treated transcripts was calculated for the 7377 genes. The 7377 genes are plotted in descending order of the difference of the fractional mRNA lengths *(B)*. The total number of FUS-CLIP tags normalized by FPKM of each gene in siCont-treated cells is plotted in *(C)* on the same ordinate. The individual genes are named as "Gene 1" to "Gene 7377". Thus, the shaded areas in *(B)* and *(C)* are identical to the green and orange areas in Figs. 6B and 6C, respectively. Blue lines in *(B)* and *(C)* indicate 47 genes harboring FUS-CLIP clusters in a 100-nt window ending at the APA site. Note that the 47 genes labeled in blue are enriched in the lower region of panels *(B)* and *(C)*. The GSEA curve (blue line) of the 47 genes in *(D)* shows a quantitative measure showing that the 47 genes are enriched in the lower region of panels *(B)* and *(C)*, where the mRNA lengths are shortened by *Fus*-knockdown.

(E) Relationship between a ratio of fractional mRNA lengths and a shift of TSSs (left graph) or polyA sites (right graph) by *Fus*-knockdown in N2A cells.

(F) Shift of TSSs and polyA sites by *Fus*-knockdown in N2A cells. Shifts of TSSs and PolyA sites are calculated as in Fig. 6A.

Supplemental Materials and Methods

ChIP and ChIP-seq

N2A cells in a 15-cm dish were subjected to crosslinking in 1% formaldehyde for 10 min at RT, then quenched with 125 mM glycine for 5 min. Cells were washed with 8 ml cold PBS, then transferred to a 15-ml tube in 1 ml PBS containing 0.5 mM PMSF. The cell pellets were suspended in 1 ml lysis buffer (50 mM Tris-HCl pH 7.4, 100 mM NaCl, 1 mM $MgCl_2$, 0.1 mM $CaCl_2$, 1% NP-40, 0.5% sodium deoxycholate, and 0.1% SDS) containing 1 mM phenylmethanesulfonyl fluoride, 10 μg/ml aprotinin and 10 μg/ml leupeptin. The lysate was sonicated on ice using a sonicator (Astrason S4000, Misonix, Inc.) at a power setting of 40 units with 10 pulses of 20-sec on and 30-sec off. Samples were centrifuged at 12,000 x *g* at 4 ˚C for 10 min to remove insoluble material, and the supernatant containing DNA-protein complexes was collected. Sheared chromatin yielded fragments of 300-1500 bp.

Specific antibody (2 μg) and 20 μl of Dynabeads proteinG or 30 μl of Dynabeads M-450 Rat anti-Mouse IgM were added to the chromatin lysate, and were incubated on a rocking platform overnight at 4 ˚C. As a negative control, isotype-matched antibodies purified from wild-type mouse serum were used. Immune-complexes were washed twice each in the following buffers: wash buffer (1x PBS, 0.1% SDS, 0.5% deoxycholate, 0.5% NP-40), high-salt wash buffer (5x PBS, 0.1% SDS, 0.5% deoxycholate, 0.5% NP-40), and PNK buffer (50 mM Tris-Cl pH 7.4, 10 mM $MgCl_2$, 0.5% NP-40). Immune-complexes were eluted twice in 50 μl of buffer containing 1% SDS and 100 mM $NaHCO_3$ at RT for 15 min. The samples were added with 10 μl of 5 M NaCl and 5 μl of 5 mg/ml proteinase K, and incubated at 65 ˚C for 4 hrs to reverse the DNA-protein crosslinks. The final DNA fragments were purified with QIAquick PCR Purification Kit (Qiagen) and stored at -20 ˚C. ChIP-fragemtns were quantified by real-time PCR essentially as described in real-time RT-PCR above. The PCR primers are listed in Supplemental Table S4.

For ChIP-seq analysis, chromatin was sheared to yield fragments of 100-300 bp and chromatin-immunoprecipitation was performed as described above using 8WG16 antibody. The ChIP'ed-DNAs and input DNAs were barcode-tagged, pooled, and sequenced in 100-base paired-end run by an Illumina HiSeq2000 at Otogenetics Corp. All ChIP-seq data were registered in DDBJ DRA Sequence Read Archive with an accession of DRA001194.

Nascent-seq

Nasent-seq was performed as described elsewhere (Menet et al. 2012). Nascent RNA, which was derived from chromatin fraction of N2A cells, was DNase-treated with TURBO DNase (Life Technologies). Polyadenylated RNA and ribosomal RNA were removed from the nascent RNA with oligo-dT beads (Truseq RNA sample Prep kit, Illumina) and Ribo-Zero magnetic kit (Epicentre), respectively.

Nascent-seq libraries were made using Truseq RNA Sample Prep kit (Illumina) with some modifications for strand-specific sequencing. dUTP was added instead of dTTP in the second strand cDNA synthesis. Uracil N-glycosylase was added to digest these dUTP containing DNAs after the adapter ligation. High-throughput 100-bp paired-end sequencing was performed with the Illumina HiSeq2000 sequencer (Beijing

Genomics Institute).

Construction of plasmids

The mouse *Fus* cDNA was obtained from Open Biosystems. To construct mammalian expression vectors for FUS, the mouse *Fus* cDNA was subcloned into the mammalian expression vector p3xFLAG-CMV-10 (Sigma) at BamHI and KpnI sites to generate p3xFLAG-*Fus*.

Mammalian expression vectors for wild-type (WTS), α-amanitin-resistant wild-type (WTR) and α-amanitin-resistant slow C4 mutant (C4) RNAP II (Schor et al, 2009) were kind gifts of Dr. Alberto R. Kornblihtt at University of Buenos Aires.

The minigene carrying the APA site of *Zbtb24* and *Ewsr1* gene (Fig. 3C) were constructed by insertion of PCR-amplified mouse genomic fragment into XbaI and EcoRV sites, and EcoRV and HindIII sites of pRBG4 vector, respectively (Masuda et al, 2008). We introduce artificial mutations in CLIP-tag clustered regions using QuikChange site-directed mutagenesis kit (Agilent). These plasmids were cut with HindIII or EcoRV before transfection to remove SV40 polyadenylation signal from the minigenes.

To construct luciferase reporter vectors with the APA site of *Gapdh /Ewsr1*, the APA and adjacent region was PCR-amplified from mouse genomic DNA, and ligated into the XbaI and BamHI sites of the pRL-SV40 vector (Promega) to be substituted for the 3' UTR of the luciferase gene. We cloned the synthesized MS2 RNA sequence of 5'-CGTACACCATCAGGGTACG-3' into XbaI (upstream of the APA site) or BamHI site (downstream of the APA site). We introduce artificial mutations in PAS using QuikChange site-directed mutagenesis kit (Agilent).

The luciferase reporter vectors harboring the APA site of *Ewsr1* upstream of luciferase gene was constructed by inserting the APA and the flanking regions into NheI site of the pRL-SV40 vector.

The EGFP expression vector, pEGFP-N1, was obtained from Clontech.

The expression vector of a fusion cDNA encoding FUS and the MS2 coat protein was constructed by insertion of mouse *Fus* cDNA into HindIII and BamHI sites and insertion of MS2 coat protein's cDNA into BamHI and NotI sites of pEGFP-N1 vector. The expression vector of a fusion cDNA encoding EGFP and the MS2 coat protein with NLS was described previously (Masuda et al. 2008).

RNA affinity purification assay

RNA affinity purification assay was performed as previously described (Masuda et al. 2008). Briefly, the biotinylated RNA-probes were synthesized using the RiboMAX System (Promega) from a PCR-amplified fragment prepared as follows. The ~200-bp fragment of *Ewsr1* intron 8 (Supplemental Fig. S5F; RNA-probe) was PCR-amplified from pRBG4-int8 vector and pRBG4-int8mutPAS using the following primers. The forward primer was 5'-*TAATACGACTCACTATAGGG*AATCCTTCTGTAACTTTGGTAACTG-3', where the italicized is T7 promoter. The reverse primer was 5'-GTTGCATGCAGCTTTACTCCTTGTT-3'.

Biotinylated RNAs were mixed with nuclear extract of N2A cells in the binding buffer (20 mM HEPES, pH 7.8, 50 mM KCl, 3 mM $MgCl_2$,500 mM EDTA and 0.05% Triton X) and the bound proteins were purified with streptavidin-sepharose beads.

Depletion of FUS from nuclear extract was performed using Protein G HP spin trap (GE Healthcare) and anti-FUS antibody according to the manufacturer's instructions. The purified proteins were analyzed by Western blotting.

RNA immunoprecipitation (RIP) assay

To eliminate the effects of cytoplasmic RNA, nuclear fraction was isolated and RIP assays were performed. N2A cells transfected with siRNA or overexpression vectors were resuspended in Buffer A (10 mM HEPES-KOH pH 7.8, 10 mM KCl, 0.1 mM EDTA, 0.1 % NP-40 and proteinase inhibitors). The fraction of crude nuclei was pelleted by centrifugation and the pellet was washed once in Buffer A without NP-40. RNA was extracted from one-twentieth of nuclei and save as "Input" RNA. The rest of nuclei were suspended in NETN buffer and insoluble elements were cleared by centrifugation. Immunoprecipitation was performed as described above. Immune complexes were washed four times by NETN buffer and bound RNAs were isolated using Quickgene-total RNA purification kit (Wako) with DNase I treatment.

Bioinformatic analysis

CLIP-seq and ChIP-seq reads were mapped to the mouse genome (NCBI build 37.1/mm9) with default parameters using the BioScope v1.3.1 (Life Technologies). RNA-seq reads were mapped to the mouse genome with default parameters using the TopHat v2.0.9 (Trapnell et al. 2013). CAGE-seq and Nascent-seq reads were mapped to the mouse genome with default parameters using BWA v0.7.10 (Li and Durbin 2010). PolyA-seq reads were mapped to the mouse genome with default parameters using the Bowtie2 v2.1.0 (Langmead and Salzberg 2012). We removed multiply aligned reads, unreliable reads, and PCR duplicates with Avadis NGS software v1.3 (Strand). Adaptor sequences were trimmed with Trimmomatic v0.32 (Bolger et al. 2014). Files were converted with Samtools v0.1.19 (Li et al. 2009). BED files were generated with Bedtools v2.18.2 (Quinlan and Hall 2010). We also made Perl scripts, if necessary.

To identify FUS-CLIP tag clusters, we used MACS (version 1.4.2) (Zhang et al. 2008) with the following parameter "-f BAM -g mm --nomodel --shiftsize=25 -w -S --call-subpeaks" to identify FUS-CLIP tag clusters from three independent CLIP experiments. MACS clusters of three independent CLIP-seq data were added. The normalized FUS-CLIP signals were calculated by dividing FUS-CLIP coverage by FPKM of each gene. We used the regions of FUS- CLIP clusters to identify strand-specific motifs using the software HOMER (version 4.7) with default parameters (Heinz et al. 2010).

In the analysis of RNA-seq, expression levels of transcripts are calculated using the Cufflinks (version 2.0.2) (Trapnell et al. 2013) with a parameter "--library-type=fr-unstranded".

In the analysis of PolyA-seq, we defined polyA sites with the PolyA Score above 3.0, which was calculated as described elsewhere (Derti et al. 2012), to generate maps of genuine polyA sites and filter out internal priming events. The specificity of identification of polyA sites this filtering method has been reported to be 97.5% (Derti et al. 2012). Coverage of each polyA site was normalized to total number of mapped tags. To minimize variability between PolyA-seq procedures, we limited our analysis to polyA sites detected in both siCont-treated cells and siFus-treated cells.

We detected CAGE peaks and their CPM (count per million) with Paraclu and

edgeR, respectively, using nAnT-iCAGE (no-amplification non-tagging illuminaCAGE) data, which was generated by MOIRAI (Hasegawa et al. 2014). For mRNA length analysis, CAGE peaks with CPM > 1 were used.

To estimate the alternatively spliced strength of isoforms derived from RNA-seq, the Percent Spliced In (PSI) was calculated by MISO (version 0.5.2) with the following parameter "--run --prefilter --read-len 100 --paired-end 170 50" (Katz et al. 2010). All alternative splicing patterns predicted by MISO, such as alternative 3' splices site (A3SS), alternative 5' splices site (A5SS), mutually exclusive (MXE), retained intron (RI), and skipped exon (SE), were used.

To estimate mRNA lengths, the most 5' end of RefSeq transcripts was set to 0, and the most 3' end of RefSeq transcripts was set to 1 to normalize the length of mRNA. The weighed average positions (WAP) of TSSs in CAGE-seq and polyA sites in PolyA-seq of each gene were calculated by averaging detected positions while weighing the number of tags. Normalized mRNA length of each gene was calculated by subtracting WAP of polyA sites from WAP of TSSs.

To identify enriched Gene Ontology terms, we used the Database for Annotation, Visualization and Integrated Discovery (DAVID 6.7) (Dennis et al. 2003; Huang da et al. 2009).

Supplemental References

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114-2120.

Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4**: P3.

Derti A, Garrett-Engele P, Macisaac KD, Stevens RC, Sriram S, Chen R, Rohl CA, Johnson JM, Babak T. 2012. A quantitative atlas of polyadenylation in five mammals. *Genome Research* **22**: 1173-1183.

Hasegawa A, Daub C, Carninci P, Hayashizaki Y, Lassmann T. 2014. MOIRAI: a compact workflow system for CAGE analysis. *BMC Bioinformatics* **15**: 144.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**: 576-589.

Huang da W, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**: 44-57.

Katz Y, Wang ET, Airoldi EM, Burge CB. 2010. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods* **7**: 1009-1015.

Lagier-Tourenne C, Polymenidou M, Hutt KR, Vu AQ, Baughn M, Huelga SC, Clutario KM, Ling SC, Liang TY, Mazur C et al. 2012. Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. *Nature Neuroscience*.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357-359.

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**: 589-595.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.

Masuda A, Shen XM, Ito M, Matsuura T, Engel AG, Ohno K. 2008. hnRNP H enhances skipping of a nonfunctional exon P3A in CHRNA1 and a mutation disrupting its binding causes congenital myasthenic syndrome. *Hum Mol Genet* **17**: 4022-4035.

Menet JS, Rodriguez J, Abruzzi KC, Rosbash M. 2012. Nascent-Seq reveals novel

features of mouse circadian transcriptional regulation. *eLife* **1**: e00011.

Nakaya T, Alexiou P, Maragkakis M, Chang A, Mourelatos Z. 2013. FUS regulates genes coding for RNA-binding proteins in neurons by binding to their highly conserved introns. *RNA* **19**: 498-509.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.

Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature biotechnology* **31**: 46-53.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137.