nature structural & molecular biology

Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre

Ezgi Hacisuleyman[1–3,13], Loyal A Goff[2–4,13], Cole Trapnell[2,3], Adam Williams[5], Jorge Henao-Mejia[5], Lei Sun[6], Patrick McClanahan[7], David G Hendrickson[2,3], Martin Sauvageau[2,3], David R Kelley[2,3], Michael Morse[3], Jesse Engreitz[3], Eric S Lander[3], Mitch Guttman[8], Harvey F Lodish[6,9,10], Richard Flavell[5,11], Arjun Raj[7] & John L Rinn[2,3,12]

RNA, including long noncoding RNA (lncRNA), is known to be an abundant and important structural component of the nuclear matrix. However, the molecular identities, functional roles and localization dynamics of lncRNAs that influence nuclear architecture remain poorly understood. Here, we describe one lncRNA, Firre, that interacts with the nuclear-matrix factor hnRNPU through a 156-bp repeating sequence and localizes across an ~5-Mb domain on the X chromosome. We further observed Firre localization across five distinct *trans*-chromosomal loci, which reside in spatial proximity to the *Firre* genomic locus on the X chromosome. Both genetic deletion of the *Firre* locus and knockdown of hnRNPU resulted in loss of colocalization of these *trans*-chromosomal interacting loci. Thus, our data suggest a model in which lncRNAs such as Firre can interface with and modulate nuclear architecture across chromosomes.

It has become clear that mammalian genomes encode many lncRNAs with diverse functions in development and disease[1–6]. Recent work has begun to identify lncRNAs' myriad roles, including but not limited to formation of ribonucleoprotein complexes with epigenetic regulatory machinery, transcriptional and post-transcriptional regulation of gene expression and the formation of subcompartments in the nucleus to mediate higher-order chromosomal architecture[7–14]. Examples of these phenomena can be observed in X-chromosome dosage compensation in mammals. Several lncRNAs have been shown to recruit epigenetic regulatory complexes (for example, Polycomb)[15,16], some of which are brought to the future inactive X chromosome[17,18]. Specifically, the lncRNA Xist binds to multiple proteins as a 'scaffold' to mediate the silencing of genes on the X chromosome and to affect the higher-order chromosomal architecture needed to establish proper epigenetic silencing[14,19,20].

RNA molecules have long been known to be key constituents of the nuclear matrix[21–24], and several noncoding RNAs (ncRNAs)[24,25] have been demonstrated to be involved in the formation of nuclear subcompartments, such as the nucleolus and paraspeckles[26], and of higher-order chromosomal architecture[19,20,27]. However, lncRNAs' mechanistic diversity, their influence on nuclear architecture and their consequent cellular roles remain unresolved.

Here, we identify a long intergenic noncoding RNA (lincRNA), termed functional intergenic repeating RNA element (Firre), that localizes across a 5-Mb domain around its site of transcription. This domain of Firre localization is also in spatial proximity to at least five other *trans*-chromosomal loci within the nucleus. This cross-chromosomal colocalization requires Firre because genetic deletion of *Firre* results in a loss of spatial proximity between its *trans*-chromosomal binding sites. We further identified a unique 156-bp repeating RNA domain in the Firre sequence that is required for both interaction with the nuclear-matrix factor hnRNPU and localization of Firre transcripts in a punctate manner in the nucleus. Strikingly, knockdown of hnRNPU, similarly to deletion of the *Firre* locus, results in a loss of spatial proximity between the *Firre* locus and its *trans*-chromosomal binding sites. Collectively, these findings suggest a model in which lncRNAs, such as Firre, can function as nuclear-organization factors that interact with and influence higher-order nuclear architecture across chromosomes.

RESULTS

Firre is a nuclear-retained and chromatin-associated lncRNA

We previously identified Firre (previously referred to as linc-RAP-1) as being required for proper adipogenesis in a loss-of-function screen

---

[1]Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts, USA. [2]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts, USA. [3]Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, Massachusetts, USA. [4]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [5]Department of Immunobiology, Yale University School of Medicine, New Haven, Connecticut, USA. [6]Whitehead Institute for Biomedical Research, Cambridge, Massachusetts, USA. [7]Department of Bioengineering, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [8]Department of Biology, California Institute of Technology, Pasadena, California, USA. [9]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [10]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [11]Howard Hughes Medical Institute, Yale University School of Medicine, New Haven, Connecticut, USA. [12]Department of Pathology, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA. [13]These authors contributed equally to this work. Correspondence should be addressed to J.L.R. (johnrinn@fas.harvard.edu).

Received 17 December 2013; accepted 30 December 2013; published online 26 January 2014; doi:10.1038/nsmb.2764

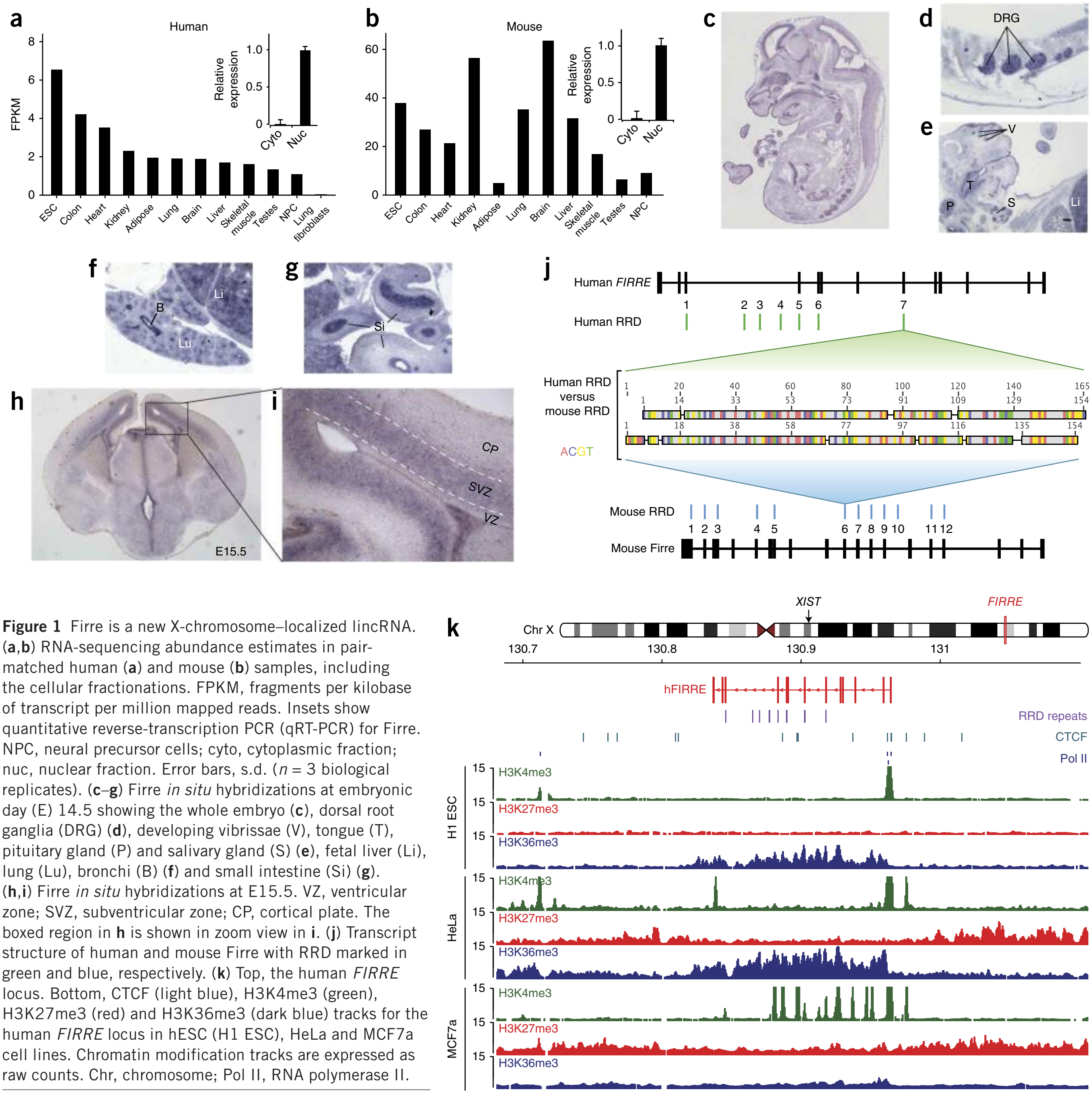


Figure 1 Firre is a new X-chromosome–localized lincRNA. (**a**,**b**) RNA-sequencing abundance estimates in pair-matched human (**a**) and mouse (**b**) samples, including the cellular fractionations. FPKM, fragments per kilobase of transcript per million mapped reads. Insets show quantitative reverse-transcription PCR (qRT-PCR) for Firre. NPC, neural precursor cells; cyto, cytoplasmic fraction; nuc, nuclear fraction. Error bars, s.d. ($n = 3$ biological replicates). (**c**–**g**) Firre *in situ* hybridizations at embryonic day (E) 14.5 showing the whole embryo (**c**), dorsal root ganglia (DRG) (**d**), developing vibrissae (V), tongue (T), pituitary gland (P) and salivary gland (S) (**e**), fetal liver (Li), lung (Lu), bronchi (B) (**f**) and small intestine (Si) (**g**). (**h**,**i**) Firre *in situ* hybridizations at E15.5. VZ, ventricular zone; SVZ, subventricular zone; CP, cortical plate. The boxed region in **h** is shown in zoom view in **i**. (**j**) Transcript structure of human and mouse Firre with RRD marked in green and blue, respectively. (**k**) Top, the human *FIRRE* locus. Bottom, CTCF (light blue), H3K4me3 (green), H3K27me3 (red) and H3K36me3 (dark blue) tracks for the human *FIRRE* locus in hESC (H1 ESC), HeLa and MCF7a cell lines. Chromatin modification tracks are expressed as raw counts. Chr, chromosome; Pol II, RNA polymerase II.

in mouse adipocyte precursors[28]. A detailed subsequent analysis of Firre revealed many interesting and distinguishing features: (i) a diverse expression pattern of Firre *in vivo*, with enrichment in neural-crest tissues as shown by *in situ* hybridization (**Fig. 1a**–**i**); (ii) a conserved intergenic human ortholog located on the X chromosome (**Fig. 1j**,**k**); (iii) a unique 156-bp repeating RNA domain (RRD) that occurs 16 and 8 times in *Mus musculus* (mouse) and *Homo sapiens* (human) transcripts, respectively, with 96% sequence identity within species and 68% across species (**Fig. 1j**); (iv) numerous alternatively spliced isoforms with differential inclusion or exclusion of RRD sequences (**Supplementary Fig. 1a**,**b**, **Supplementary Table 1**); (v) stability of Firre transcripts even after 6 h of actinomycin D treatment (**Supplementary Fig. 1c**,**d**).

To further determine the subcellular localization of Firre, we used single-molecule RNA fluorescence *in situ* hybridization (RNA FISH) targeting Firre (as described in ref. 29). We adopted a dual labeling strategy to independently target the introns and exons of Firre, thus marking the site of transcription on the X chromosome (intronic probes) and the location of the mature transcripts (exonic probes) separately (**Fig. 2a** and **Supplementary Table 2**). RNA FISH analysis revealed an exclusively nuclear and focal distribution for Firre (**Fig. 2a** and **Supplementary Fig. 2**) in all cells tested. Notably, Firre exhibited strong expression foci near its site of transcription in both male and female mouse and human embryonic stem cells (mESCs and hESCs, respectively) (**Fig. 2a**). We also note that the Firre RNA was localized around its site of transcription but extended slightly beyond this site

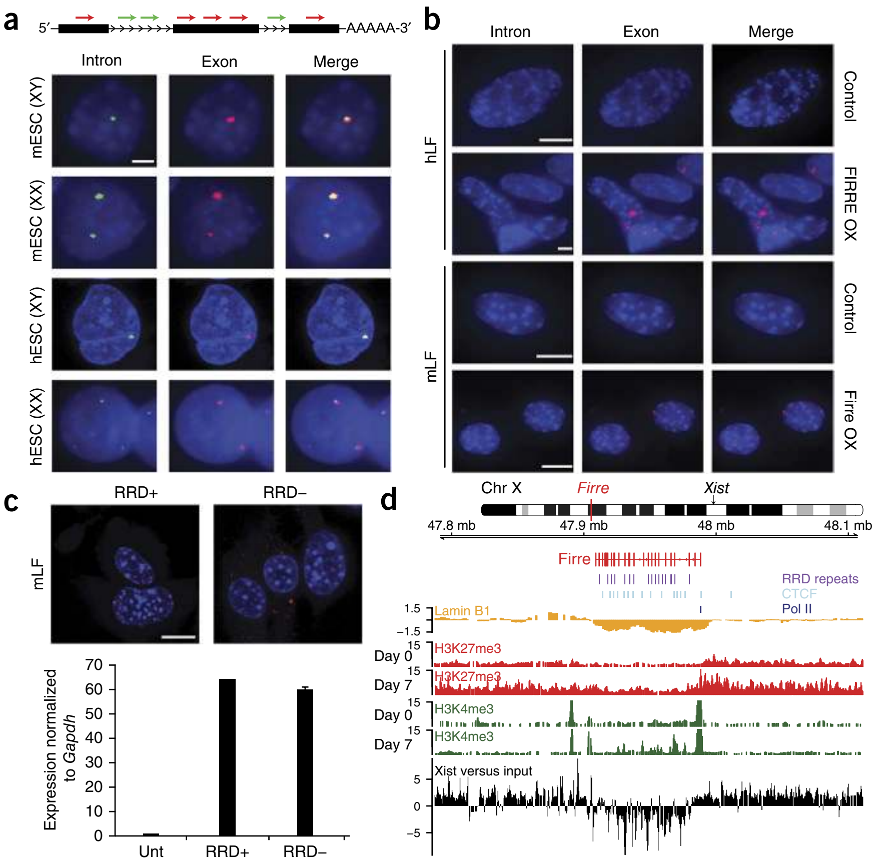


Figure 2 Firre is a new, strictly nuclear lincRNA that escapes X-chromosome inactivation. (**a**–**c**) Single-molecule RNA FISH of introns in green (A594), exons in red (Cy3) and nuclei (4′,6-diamidino-2-phenylindole (DAPI)). (**a**) Male and female hESCs and mESCs. Scale bar, 20 µm. (**b**) Viral overexpression of Firre in human and mouse lung fibroblasts (hLF and mLF) that do not express Firre. Scale bars, 15 µm. OX, overexpression. (**c**) Viral overexpression of Firre isoforms with (+) or without (–) RRD in mLFs. Scale bar, 15 µm. Bottom, quantification of expression. Unt, untransfected control. Error bars, s.d. ($n = 3$ biological replicates). (**d**) The mouse *Firre* locus. The unique repetitive domain RRD (purple), CTCF (blue), lamin B1 (orange), H3K27me3 (red) and H3K4me3 (green) and Xist RAP (black) are shown. Lamin B1 and Xist are plotted as log fold change on the *y* axis relative to input, and chromatin modifications are shown as raw counts.

in all six human and mouse cell lines tested. Subcellular localization and expression of Firre in cell lines with and without inactive X chromosomes was similar to those observed in ESCs (**Supplementary Fig. 2**). Thus, Firre is nuclear localized and forms expression foci on both X chromosomes before and after X-chromosome inactivation.

We tested whether overexpression of Firre was sufficient to form the observed endogenous local expression foci. Briefly, we ectopically expressed Firre via retrovirus-mediated integration in human and mouse lung fibroblasts, which do not express Firre (**Fig. 2b** and **Supplementary Fig. 3a**). We observed many sites of focal accumulation of Firre (**Fig. 1b**). We obtained similar results upon ectopic expression of Firre in human cells that endogenously express this lncRNA (HEK293 and HeLa) (**Supplementary Fig. 3**). We repeated the experiments above, using isoforms of Firre either with (isoform 4; one repeat) or without (isoform 6) the RRD. Strikingly, in the absence of the RRD, the nuclear localization of Firre was disrupted, and we detected Firre RNA in the cytoplasm (**Fig. 2c**). Thus, RRD is required for the focal nuclear localization of Firre.

The *Firre* locus escapes X-chromosome inactivation

Our observation from RNA FISH in female mouse and human ESCs and HEK293 cells led us to hypothesize that *Firre* might escape X-chromosome inactivation. To test this, we analyzed the local chromatin environment within the *Firre* locus, using existing chromatin immunoprecipitation (ChIP) data for numerous histone modifications and transcription factors. Several of these data are consistent with *Firre* escaping X-chromosome inactivation: First, we observed an appreciable depletion of lamin B1 across the mouse *Firre* locus and across the human *FIRRE* locus in various cell lines (**Figs. 1k** and **2d**). Lamin B1 is a matrix protein involved in nuclear stability and chromatin organization, and it is known to mark heterochromatin. The domain of lamin B1 depletion extends precisely across the body of the *Firre* gene but not into the upstream or downstream regions. Second, the *Firre* locus was specifically and significantly ($P < 1.0 \times 10^{-8}$) depleted of trimethylated histone H3 K27 (H3K27me3) in differentiated mESCs and in human cells before and after X-chromosome inactivation (**Fig. 2d**). Third, the *Firre* locus was enriched for trimethylated histone H3 K4 (H3K4me3) with and without Firre transcription (**Fig. 2d**). Finally, we observed a striking localization pattern of CCCTC-binding factor (CTCF) (**Fig. 2d**)—which can function as an insulator between chromatin domains and facilitate interchromosomal interactions—adjacent to almost every exon of Firre[30].

To further test the hypothesis that *Firre* escapes X-chromosome inactivation, we investigated whether Xist RNA itself localizes on the *Firre* locus upon X-chromosome inactivation. Specifically, we examined the localization of Xist on DNA by using data generated by RNA antisense purification (RAP) in mouse lung fibroblasts (mLFs). In contrast to the enrichment of Xist previously observed across most of the X chromosome[19], a strong and focal depletion was present in Xist binding at the *Firre* locus; this was similar to what was observed at genes known to escape X-chromosome inactivation (**Fig. 2d**). Interestingly, the Xist-depleted boundaries are consistent with the previously identified boundaries for the lamin-depleted regions. Collectively, these data indicate that the *Firre* locus escapes X-chromosome inactivation and has a notable enrichment for CTCF and H3K4me3 and depletion for H3K27me3 and lamin B1.

Firre localizes to chromatin in *cis* and in *trans*

The focal nuclear localization of Firre near its site of transcription led us to identify the direct interactions between Firre and chromatin in the nucleus. To resolve the DNA-binding sites of Firre genome wide, we performed RAP[19]. RAP, along with chromatin isolation by RNA purification (CHiRP)[31], capture hybridization analysis of RNA targets (CHART)[32] and chromatin oligoaffinity precipitation (ChOP)[33], identifies genome-wide DNA-binding locations for RNAs

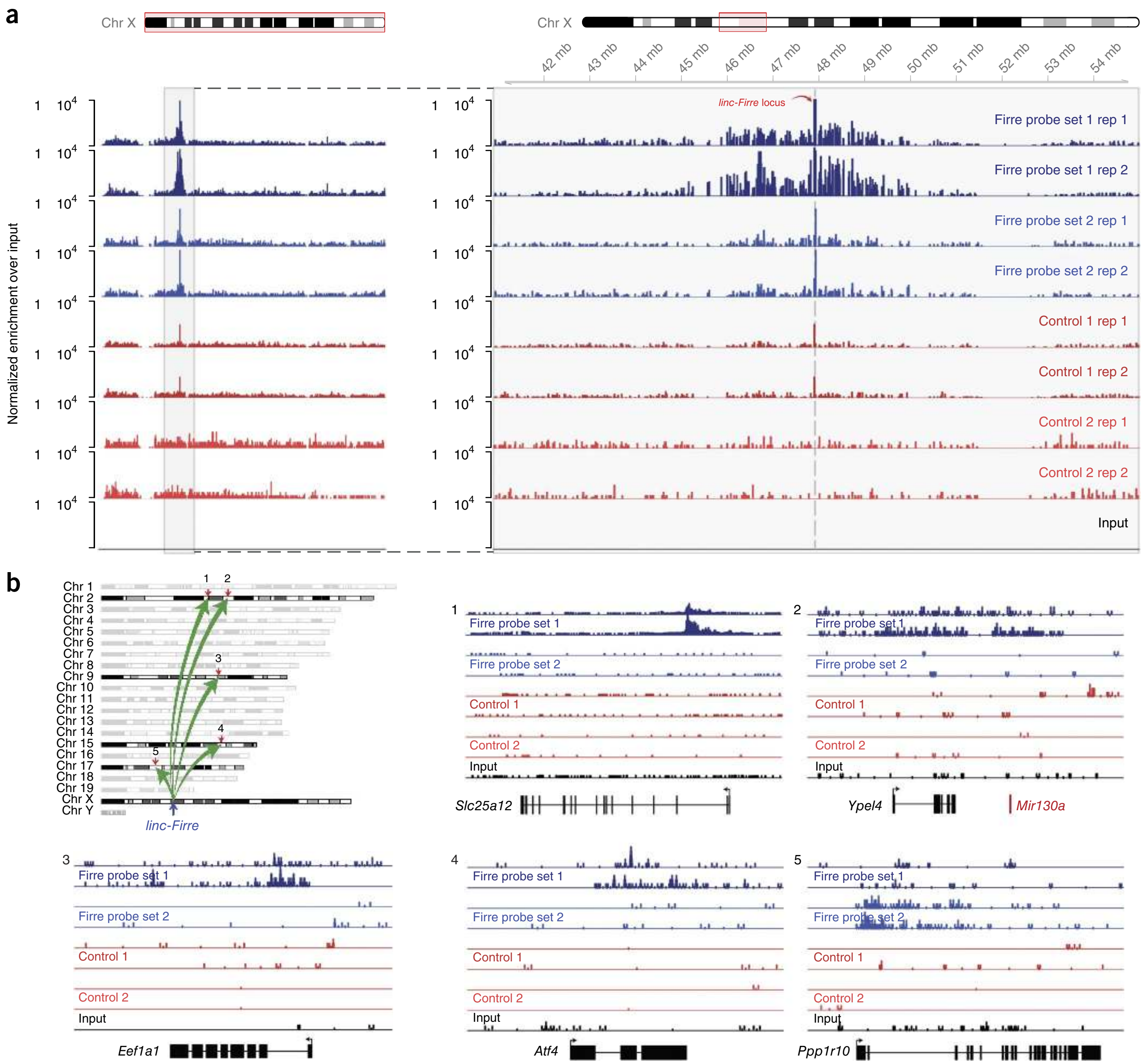


Figure 3 Firre forms *cis* and *trans* chromosomal contacts. (**a**) RAP by Firre shown along the X chromosome in male mESCs as fold enrichment relative to input after normalization for library depth. The genomic ~5-Mb region flanking the Firre locus is highlighted. Replicate enrichment tracks normalized to the input (black) are presented for each of two Firre-targeting probes (blue tracks) and two sets of sense probes (negative controls) (red tracks). (**b**) RAP by Firre shown for five distinct interchromosomal genomic loci (1–5). Local genomic contexts are presented for each of the *trans*-interacting sites and their adjacent genes: *Slc25a12*, *Ypel4*, *Eef1a1*, *Atf4* and *Ppp1r10*. Counts for the *trans*-chromosomal contacts of Firre are shown after normalization for sequencing depth. Chr, chromosome.

by cross-linking chromatin and RNA. These procedures are followed by the targeting and pulldown of a specific RNA with antisense oligonucleotides[19]. We performed RAP in male mESCs by using two sets of 120-bp antisense probes targeting Firre and two sets of sense probes as negative controls (**Supplementary Table 3**). This was followed by sequencing to identify genomic regions directly bound by Firre. We observed an ~5-Mb domain of Firre localization around the *Firre* locus (**Fig. 3a**). Strikingly, we also observed five significantly enriched peaks (Cuffdiff2; 1% false discovery rate (FDR)) of Firre located on chromosomes 2, 9, 15 and 17 that overlap known genes including *Slc25a12*, *Ypel4*, *Eef1a1*, *Atf4* and *Ppp1r10* (**Fig. 3b**). Notably, four out of five of these genes have previously described regulatory roles during adipogenesis[34–37], results consistent with our previous study showing the role of Firre in adipogenesis[28]. Expanding this search to regions not overlapping mRNAs, we observed a total of 34 additional significant (Cuffdiff2; 1% FDR) localization sites for Firre.

Collectively these data suggest that Firre is localized on multiple chromosomes yet has only one predominant nuclear localization site in male and two in female cells around its site of transcription. These observations suggest two possible models. One possibility is that Firre could be shuttled from its site of transcription to these

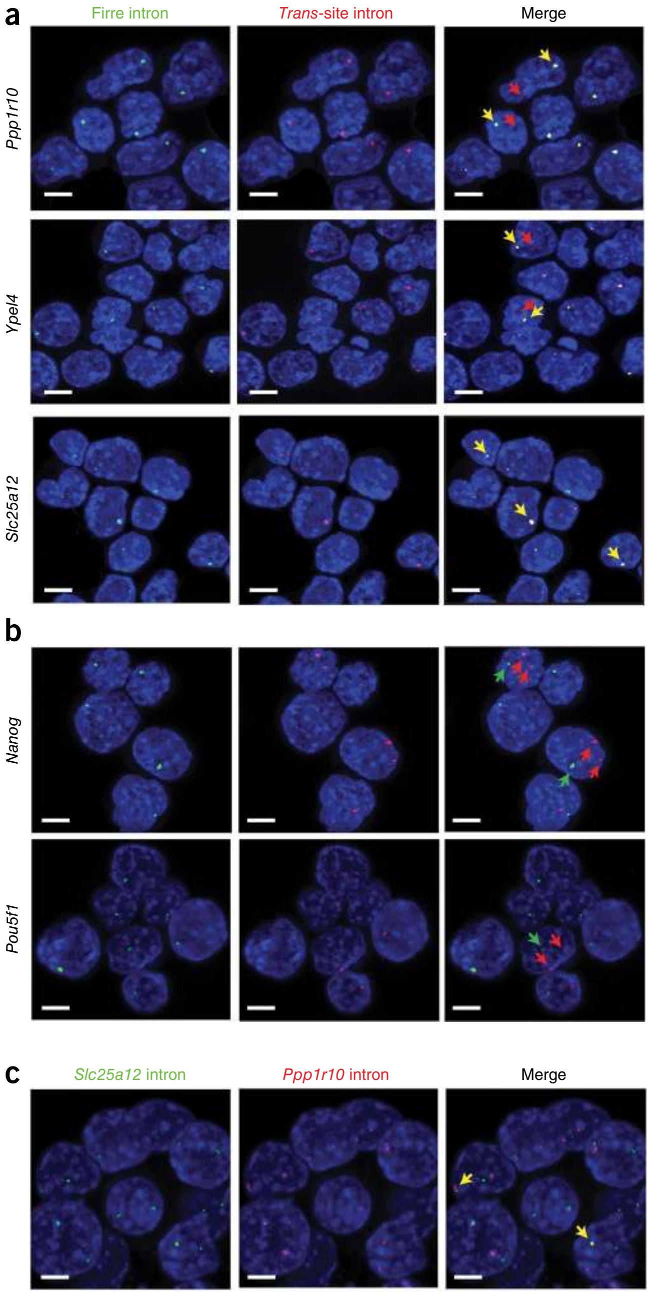


Figure 4 The *trans*-chromosomal contacts of Firre are colocalized within the nucleus. (**a**) Co-RNA FISH. Introns of Firre are labeled in green and the introns of its *trans* targets in red. Nuclei are labeled by DAPI. Colocalization of Firre with *Ppp1r10*, *Ypel4* and *Slc25a12*. (**b**) Co-RNA FISH of *Nanog* and *Pou5f1*. Green and red arrows indicate nonoverlapping genomic loci, and yellow arrows indicate colocalization of both loci. (**c**) RNA FISH colocalization of the *trans*-interacting loci *Ppp1r10* and *Slc25a12*. Scale bars, 40 µm.

sites on other chromosomes. Alternatively, the focal localization of Firre to its own genomic locus could mean that it serves as a regional organizing factor to bring the *trans*-interacting sites into the three-dimensional proximity of the *Firre* locus on the X chromosome.

Firre *trans*-chromosomal sites are in spatial proximity

In order to determine the nature of the *trans*-chromosomal interactions for Firre, we performed single-molecule RNA FISH in mESCs on the transcription sites of Firre and three of the *trans*-interacting genes (*Slc25a12*, *Ypel4* and *Ppp1r10*) (**Fig. 4a**). For negative controls, we performed similar RNA FISH for Firre and several genes with high expression in mESCs that were not detected by RAP as *trans* targets (for example, *Pou5f1* (also known as *Oct4*), *Nanog* and *Sox2*) (**Fig. 4b** and **Supplementary Fig. 4a**). Remarkably, we observed colocalization between Firre and all three *trans* sites tested (*Slc25a12*, 73.9% of cells; *Ypel4*, 79.4% of cells; and *Ppp1r10*, 78.1% of cells) (**Fig. 4a**) and between these *trans* sites (**Fig. 4c** and **Supplementary Fig. 4b**). Conversely, we did not observe any colocalization of Firre and unbound targets *Pou5f1*, *Nanog* or *Sox2* (**Fig. 4b** and **Supplementary Fig. 4a**). Thus, these results are consistent with the latter model, in which the *Firre* locus resides in three-dimensional proximity to these *trans*-chromosomal binding sites.

Firre regulates key pluripotency pathways

To determine the functional role of Firre, we generated a Firre-knockout male mESC line by deleting the entire *Firre* locus on the X chromosome (Δ*Firre*). Briefly, we introduced loxP sites 5′ and 3′ of the *Firre* locus by a two-step targeting strategy. Then we infected the cells that contained this locus with a Cre plasmid and clonally selected the cells with the proper deletion. Comparison of wild-type and Δ*Firre* growth rates revealed a marked retardation in growth rate and colony formation in the knockout (**Fig. 5a**,**b**). We also note an intermediate growth defect when the cells were grown on a mouse embryonic fibroblast feeder layer (**Fig. 5a**). The Δ*Firre* cells on feeders were able to form bigger and more colonies in the same amount of time as compared to the Δ*Firre* cells grown without feeders (**Fig. 5a**). To identify the gene pathways and molecular signatures that are altered upon deletion of Firre, we conducted massively parallel RNA sequencing (RNA-seq) comparisons between wild-type and Δ*Firre*. Briefly, we isolated RNA from three wild-type and two Δ*Firre* replicate cultures and subjected it to paired-end Illumina sequencing to a mean depth of $\sim 9 \times 10^6$ fragments aligned per replicate. We identified 1,077 genes with significant differential expression (Cuffdiff2; 5% FDR) between the wild-type and Δ*Firre* mESCs (**Fig. 5c**). Preranked gene-set enrichment analysis (GSEA) demonstrated that Δ*Firre* cells were significantly enriched ($P < 0.01$) for genes involved in extracellular-matrix organization and cell-surface receptor-ligand interactions (**Fig. 5d**). Conversely, Δ*Firre* mESCs, relative to wild type, were depleted for genes involved in mRNA processing, nuclear export and electron transport chain–mediated energy metabolism (**Fig. 5d**). Notably, we observed an increase in transforming growth factor β (TGF-β) signaling in the Δ*Firre* mESCs (**Fig. 5e**). Interestingly TGF-β signaling is known to be a potent inhibitor of adipogenesis[38]. This is consistent with our previous observation that knockdown of Firre strongly inhibits adipogenesis in mouse preadipocytes[28] and growth defects observed in mESC cultures (**Fig. 5a**,**b**).

We next tested whether the Firre *trans* localization sites were affected by the absence of Firre. We did not observe a global enrichment for the five *trans*-site genes in the list of significantly differentially expressed genes ($P < 1.0$; hypergeometric test). We did observe one key exception, *Ppp1r10*, one of the three validated *trans* sites, that was significantly decreased (Cuffdiff2; 1% FDR) in the Δ*Firre* cells relative to wild type. However, we cannot preclude the possibility of perturbations to mRNA stability, translation or processing at any of the remaining *trans* sites.

Firre binds hnRNPU in an RRD-dependent manner

We next turned to identify proteins that interact with Firre that might mediate its 5-Mb X-chromosome localization. As a first approach

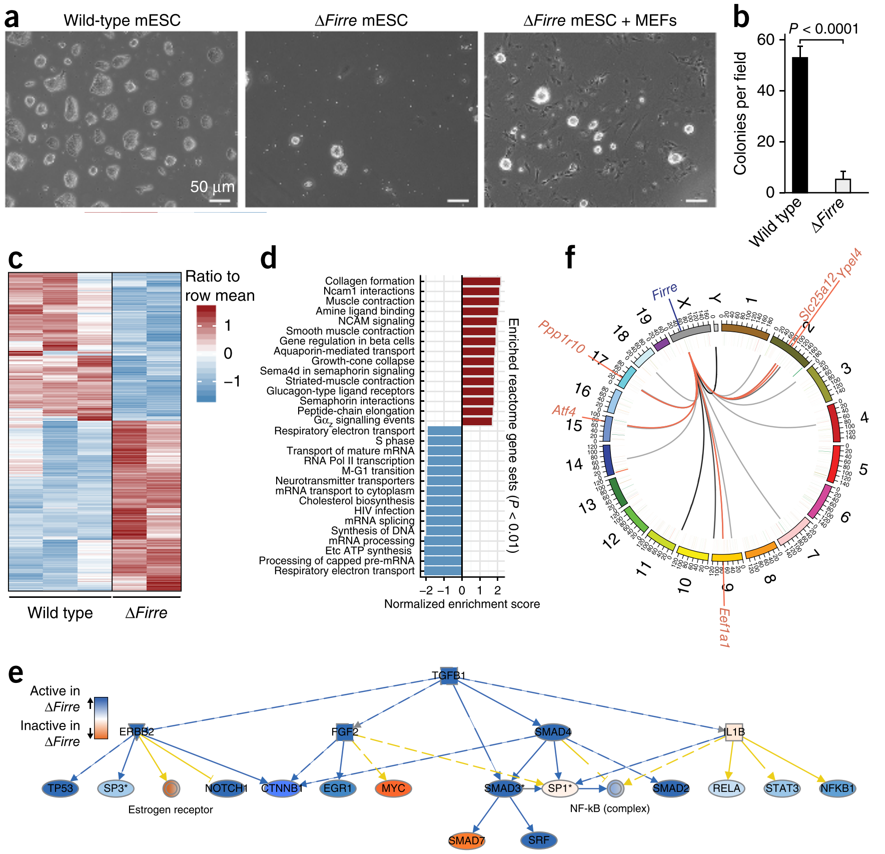
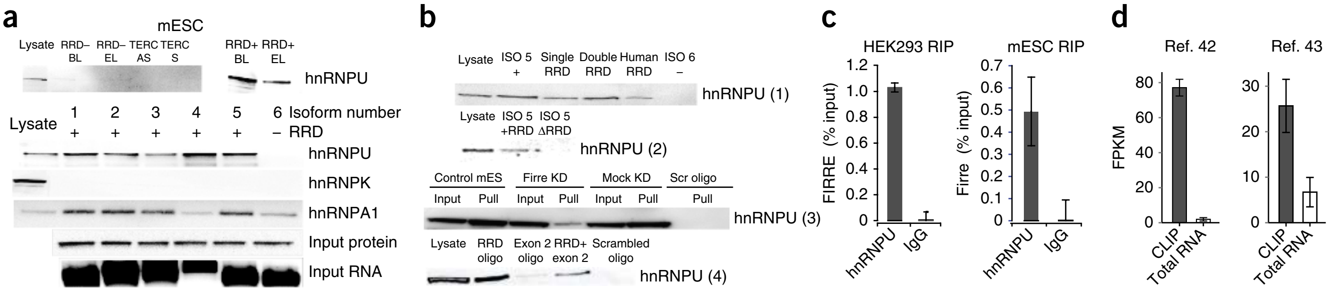


Figure 5 Characterization of Firre knockout in male mESCs. (**a**,**b**) Genetic deletion of Firre in male mESCs (Δ*Firre)* with (+MEF) and without feeder cells. Images (**a**) and quantification (**b**) of Δ*Firre* in mESCs are shown. Error bars indicate s.e.m. for replicate fields ($n = 7$ for wild type; $n = 6$ for Δ*Firre*); *P* value calculated with two-tailed Student's *t* test. (**c**) Heat map of 892 significantly differentially expressed genes (Cuffdiff2; 1% FDR) between wild-type (WT) and Δ*Firre* male mESCs. (**d**) Top 15 enriched and depleted significant ($P < 0.01$; Mann–Whitney *U* test). Reactome gene sets from a preranked GSEA analysis on Cuffdiff2 test statistics. (**e**) Ingenuity pathway analysis mechanistic-network diagram of significant ($P < 6.31 \times 10^{-44}$ by Fisher's exact test) increases in predicted downstream TGF-β signaling activity in the Δ*Firre* male mESCs relative to wild type. Color scale indicates activation *z* score of predicted downstream targets for a given gene. (**f**) Circos diagram of significant Firre RAP peaks (links) interacting with the *Firre* genomic locus (blue) in male mESCs. Peaks intersecting genic regions are highlighted in red and specifically labeled. Log_2 fold changes for significant (Cuffdiff2; 1% FDR) differentially expressed genes (Δ*Firre*/wild type) are inscribed at corresponding genomic locations within the circle. Analyses in **c**–**f** are of the same RNA-seq data set containing two replicate cultures of Δ*Firre* and three wild-type cultures.

to identify the candidate protein partners of Firre, we performed RNA pulldown assays in mESCs and mouse adipocyte lysates by biotinylating the RNA, either by body-labeling (*in vitro* transcription) or 3′-end–labeling (pCp-biotin) and subsequent MS. Unrelated ncRNAs (sense and antisense telomerase RNA TERC) served as negative controls (**Supplementary Fig. 5a**). To identify proteins that preferentially coprecipitated with Firre in an RRD-dependent manner, we used five different RRD-positive isoforms and one RRD-negative isoform and took the difference between the peptide counts



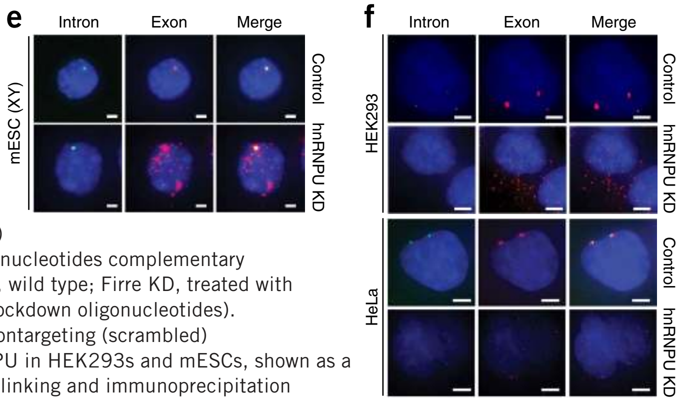


Figure 6 hnRNPU binds to the RRD of Firre and regulates its nuclear localization. (**a**) Western blots for RNA pulldowns*:* five with RRD and one without. Biotin end-labeling (EL) versus body-labeling (BL) for hnRNPU interaction is shown. TERC antisense (AS) and sense (S) are negative controls. hnRNPK and A1 are shown for RRD specificity. 20% of lysate was loaded. Input protein lysate and RNA are shown as loading controls. (**b**) Western blots for RNA pulldowns. (1,2) Pulldowns performed with *in vitro*–biotinylated synthetic RRD constructs (1) or with an endogenous isoform with RRD or the same isoform with RRD deleted (2) in mESCs. (3) RNA pulldowns (pull) with targeting desthiobiotin-DNA oligonucleotides complementary to Firre or nontargeting oligonucleotides (scr oligo) in mESCs (control mES, wild type; Firre KD, treated with oligonucleotides to knock down Firre; mock KD, treated with scrambled knockdown oligonucleotides). (4) Pulldowns with targeting (RRD, exon 2 or a combination of both) and nontargeting (scrambled) desthiobiotin oligonucleotides (oligo) in HEK293 cells. (**c**) RIP with hnRNPU in HEK293s and mESCs, shown as a percentage of input. Error bars, s.d. ($n = 3$ biological replicates). (**d**) Cross-linking and immunoprecipitation (CLIP) data as gene-level FPKM values. Error bars, s.d. ($n = 3$ biological replicates). (**e**,**f**) RNA FISH targeting Firre in mESCs, HEK293s and HeLa cells in the absence of hnRNPU. Green, introns; red, exons; blue, nuclei. Scale bars, 20 μm (**e**) or 5 μm (**f**). **Supplementary Figure 7** shows original images for **a** and **b**.

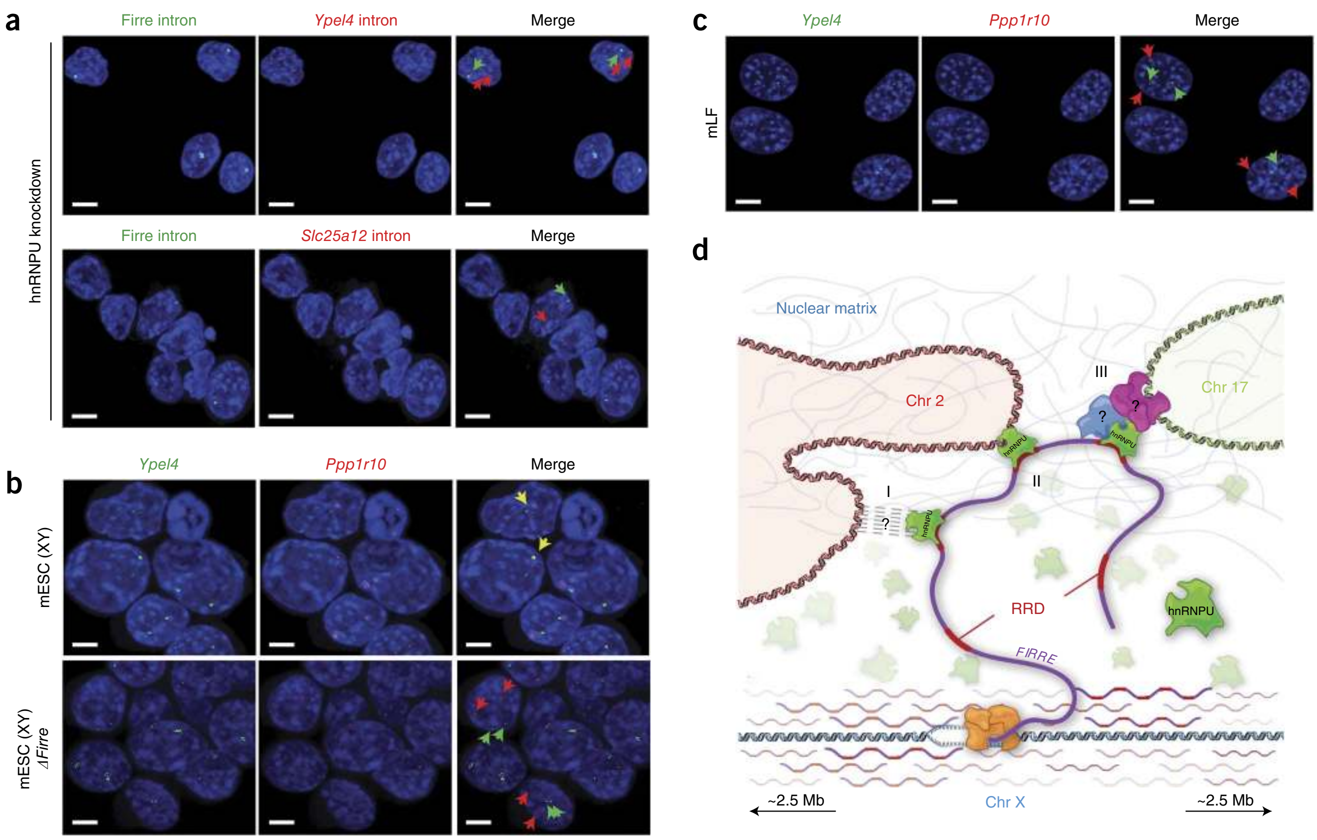


Figure 7 The *trans*-chromosomal contacts of Firre are mediated via its interaction with the nuclear-matrix protein hnRNPU. (**a**) RNA FISH colocalization of Firre introns (green) with *Ypel4* or *Slc25a12* introns (red) in the absence of hnRNPU. (**b**) RNA FISH colocalization of the *trans*-interacting loci *Ppp1r10* and *Ypel4* in the absence of Firre expression in Δ*Firre* male mESCs compared to wild-type mESCs. (**c**) RNA FISH colocalization of *trans* sites in mLFs. Scale bars in **a**–**c**, 40 µm. (**d**) A model for Firre as a regional organization factor. Firre transcripts accumulate at the site of their transcription. hnRNPU binds to the RRD of Firre and facilitates interactions with *trans*-chromosomal regions through one of several possible mechanisms: tertiary interactions with nuclear-matrix components (I), direct binding of hnRNPU to matrix attachment regions *in trans* (II) or as-yet-undetermined interactions with other protein complexes to facilitate indirect binding to DNA (III).

of the RRD-positive and RRD-negative isoforms (**Supplementary Fig. 5b**). We repeated the differential analysis for each of the five isoforms and took the top 10% of the differential peptide-count scores for each isoform identified in both mESC and adipocyte lysates. We identified eight candidate proteins that physically associate with Firre in an RRD-dependent manner, on the basis of highest unique peptide counts (**Supplementary Fig. 5c**).

The highest-ranked candidate from this analysis was hnRNPU (**Supplementary Fig. 5c**). hnRNPU was of particular interest because it is required for the proper localization of Xist, for interaction with the scaffold-attachment regions on DNA and for the formation of highly structured chromatin territories[39–41]. To confirm the interaction between hnRNPU and Firre, we repeated the RNA pulldown experiments described above and assayed for hnRNPU via western blotting. In both mESC and adipocyte lysates, hnRNPU coprecipitated with Firre but not with the negative controls (**Fig. 6a** and **Supplementary Fig. 5d**). We further tested additional hnRNP family members and found either no association or association independent of RRD (**Fig. 6a** and **Supplementary Fig. 5d**). Together these data suggest that Firre associates with hnRNPU.

To test whether the Firre RRD element is required and sufficient to interact with hnRNPU, we performed RNA pulldowns by using single and double copies of the mouse and human RRD sequences in mESC lysates. We observed binding of hnRNPU to both mouse and human synthetic RRD constructs (**Fig. 6b**). Finally, western blot analysis confirmed that hnRNPU binds to the Firre isoform containing RRD but not to the ΔRRD isoform (**Fig. 6b**). To determine whether the hnRNPU-Firre interaction is biologically relevant at endogenous levels, we captured endogenous Firre by using complementary DNA oligonucleotides and confirmed that hnRNPU copurifies specifically with the lncRNA (**Fig. 6b** and **Supplementary Fig. 6**).

To further validate the hnRNPU-Firre interaction, we performed the reciprocal experiment: RNA immunoprecipitation (RIP) targeting hnRNPU. Consistently with the RNA pulldowns, there was strong enrichment of Firre relative to IgG controls after normalization to total input (**Fig. 6c**). Finally, analysis of publicly available data from UV-cross-linked RIP targeting hnRNPU verified that Firre directly and specifically binds to hnRNPU (**Fig. 6d**)[42,43]. Collectively, these results suggest that Firre interacts with hnRNPU and that the RRD is both required and sufficient for this interaction.

hnRNPU is required for focal localization of Firre

We investigated whether hnRNPU, in turn, regulates the spatial expression of Firre, on the basis of the requirement for the RRD in establishing the proper localization of Firre and its interaction with hnRNPU. Briefly, we transfected short interfering RNAs (siRNAs) targeting hnRNPU in mESCs and human (HEK293 and HeLa) cell lines and observed a >90% decrease in hnRNPU expression

(**Supplementary Fig. 6c**). We confirmed the previously described role of hnRNPU in Xist localization in HEK293 cells (**Supplementary Fig. 6d**). After the knockdown of hnRNPU, we repeated RNA FISH targeting of Firre as above. In all cell lines tested, we observed a strong delocalization of Firre and in several instances even translocation into the cytoplasm (**Fig. 6e**,**f**). Thus, both the RRD and hnRNPU are required for the proper focal and nuclear localization of Firre and its retention in the nucleus.

hnRNPU is required for proximal *trans* localization of Firre

Having found that hnRNPU regulates the specific localization of Firre, we next tested whether hnRNPU is required to maintain the proximal localization of the *Firre* locus and its *trans*-chromosomal localization sites. To this end, we repeated the RNA FISH between Firre and either *Slc25a12* or *Ypel4* upon siRNA-mediated depletion of hnRNPU in mESCs. In both cases, we observed a considerable decrease in colocalization of each *trans* site with Firre in the absence of hnRNPU (**Fig. 7a**).

Firre is required for *trans*-chromosomal colocalization

To test the functional contribution of Firre to *trans*-chromosomal colocalization, we repeated the co-FISH experiments between the *trans* sites in male Δ*Firre* mESCs. Strikingly, the *Ppp1r10* and *Ypel4* gene loci no longer colocalized in the absence of Firre (**Fig. 7b**) (15% colocalization in Δ*Firre* relative to 72% in wild type), thus suggesting a requirement for the *Firre* gene locus in facilitating the formation of cross-chromosomal interactions in mESCs. Furthermore, we did not observe colocalization of *trans* sites in mLFs that did not express Firre (**Fig. 7c**). Collectively, the above results suggest a potential role for Firre, along with hnRNPU, in either maintaining or establishing higher-order nuclear architecture (**Fig. 7d**).

DISCUSSION

Here, we have identified and characterized a new ncRNA-encoding gene, *Firre*, that has important roles in both cell physiology and nuclear architecture. The numerous properties of Firre provide new insights into how RNA sequences can result in both *cis* and *trans* localization on chromatin, how these interactions between specific RNA sequences and nuclear-matrix factors influence nuclear organization and how these RNA-protein interaction properties summate to modulate properties of higher-order nuclear architecture and consequently affect cell physiology. Together, these observations suggest an intriguing model in which Firre, and potentially many other lncRNAs[8,13,22,27,44–47], function as nuclear-organization factors (**Fig. 4b**). Specifically, Firre may serve to interface with and to modulate the topological organization of multiple chromosomes. Consistently with this model, genetic deletion of *Firre* results in a loss of nuclear proximity of several *trans*-chromosomal loci to the *Firre* locus. Moreover, the proper localization of Firre requires both a specific 156-bp sequence and a physical interaction with hnRNPU to maintain the multichromosomal nuclear interactions. Thus, we propose that lncRNAs, through the interaction with nuclear-matrix proteins, such as hnRNPU, might impart specificity in organizing a proper 'zip code' of chromosomal territories within the nucleus. For example, either the chromosomal binding of Firre or sequence-specific interactions may serve as a *cis* localization signal to initiate subcompartment formation or maintain the subcompartments within the nucleus. Consistently with our model, two recent studies[19,20] have demonstrated that the Xist RNA uses a 'local proximity search' to guide its localization across large expanses of the X chromosome during X-chromosome inactivation. Here, we broaden this phenomenon and show that these interactions are not merely restricted to a single chromosome and can extend across multiple chromosomes in regional proximity. Several other observations in this study highlight potential gene-regulatory roles of Firre, mediated by *trans*-chromosomal interactions. Intriguingly, we observe an array of CTCF-binding sites across the *Firre* locus. CTCF has previously been shown to have a critical role in X-chromosome pairing and counting[48]. Similarly, this array of CTCF-binding sites across the *Firre* locus might further facilitate interchromosomal interactions with the 5-Mb X-chromosome localization domain. Finally, our study demonstrates that the formation of these cross-chromosomal interactions is altered upon genetic depletion of the *Firre* locus. This model has several implications for potential new roles for lncRNA-mediated gene regulation. For example, lncRNAs could bring genes involved in a similar biological process into proximity, thus allowing for co-regulation in space and time and serving as nuclear-organization factors[19]. This appears to be the case for Firre; here, several genes involved in energy metabolism and adipogenesis are organized together in spatial proximity and are typically coexpressed, consistently with the previously described role of Firre in adipogenesis[28]. Underscoring the physiological relevance of such a model, either genetic deletion or transcriptional depletion of Firre in mESCs and adipocytes, respectively, results in the perturbation of cell physiology[28]. Future studies will require genetic studies in mouse models to further illuminate the role of Firre in mammalian development and disease.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. All sequencing and related data have been deposited in the Gene Expression Omnibus database under accession code GSE45157.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank Biosearch, especially M. Beal, H. Johansson, A. Orjalo, S. Coassin and R. Cook, for materials, advice and help with FISH experiments. We are grateful to the entire Rinn and Raj laboratories for generous help with experiments, bioinformatics and preparation of the manuscript. This work was supported by US National Institutes of Health grants 1DP2OD00667 (J.L.R.), P01GM099117 (J.L.R.), 1DP20D008514 (A.R.) and P50HG006193-01 (J.L.R.) and by the Howard Hughes Medical Institute (R.F.).

AUTHOR CONTRIBUTIONS

E.H. designed and performed experiments. L.A.G. performed computational analysis. J.L.R. directed research. E.H., L.A.G. and J.L.R. wrote the manuscript. C.T. and D.R.K. helped with computational analysis. M.S., D.G.H., P.M., J.E., M.G., A.R., M.M. and E.S.L. contributed experiments. A.W., J.H.-M. and R.F. generated knockout mESCs. L.S. and H.F.L. helped with experimental techniques.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Huarte, M. & Rinn, J.L. Large non-coding RNAs: missing links in cancer? *Hum. Mol. Genet.* **19**, R152–R161 (2010).
2. Rinn, J.L. & Chang, H.Y. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81**, 145–166 (2012).
3. Carninci, P. *et al.* The transcriptional landscape of the mammalian genome. *Science* **309**, 1559–1563 (2005).

4. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775–1789 (2012).
5. Batista, P.J. & Chang, H.Y. Cytotopic localization by long noncoding RNAs. *Curr. Opin. Cell Biol.* **25**, 195–199 (2013).
6. Guttman, M. & Rinn, J.L. Modular regulatory principles of large non-coding RNAs. *Nature* **482**, 339–346 (2012).
7. Clemson, C.M. *et al.* An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol. Cell* **33**, 717–726 (2009).
8. Maass, P.G. *et al.* A misplaced lncRNA causes brachydactyly in humans. *J. Clin. Invest.* **122**, 3990–4002 (2012).
9. Nie, L. *et al.* Long non-coding RNAs: versatile master regulators of gene expression and crucial players in cancer. *Am. J. Transl. Res.* **4**, 127–150 (2012).
10. Pandey, R.R. *et al.* Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol. Cell* **32**, 232–246 (2008).
11. Rinn, J.L. *et al.* Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311–1323 (2007).
12. Tsai, M.C. *et al.* Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**, 689–693 (2010).
13. Vallot, C. *et al.* XACT, a long noncoding transcript coating the active X chromosome in human pluripotent cells. *Nat. Genet.* **45**, 239–241 (2013).
14. Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J. & Lee, J.T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750–756 (2008).
15. Kaneko, S., Son, J., Shen, S.S., Reinberg, D. & Bonasio, R. PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1258–1264 (2013).
16. Davidovich, C., Zheng, L., Goodrich, K.J. & Cech, T.R. Promiscuous RNA binding by Polycomb repressive complex 2. *Nat. Struct. Mol. Biol.* **20**, 1250–1257 (2013).
17. Jeon, Y., Sarma, K. & Lee, J.T. New and Xisting regulatory mechanisms of X chromosome inactivation. *Curr. Opin. Genet. Dev.* **22**, 62–71 (2012).
18. Plath, K., Mlynarczyk-Evans, S., Nusinow, D.A. & Panning, B. Xist RNA and the mechanism of X chromosome inactivation. *Annu. Rev. Genet.* **36**, 233–278 (2002).
19. Engreitz, J.M. *et al.* The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973 (2013).
20. Simon, M.D. *et al.* High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. *Nature* **504**, 465–469 (2013).
21. Bouvier, D., Hubert, J., Seve, A.P. & Bouteille, M. Nuclear RNA-associated proteins and their relationship to the nuclear matrix and related structures in HeLa-cells. *Can. J. Biochem. Cell Biol.* **63**, 631–643 (1985).
22. Nickerson, J.A., Krochmalnic, G., Wan, K.M. & Penman, S. Chromatin architecture and nuclear RNA. *Proc. Natl. Acad. Sci. USA* **86**, 177–181 (1989).
23. Pederson, T. & Bhorjee, J.S. Evidence for a role of RNA in eukaryotic chromosome structure: metabolically stable, small nuclear RNA species are covalently linked to chromosomal DNA in HeLa cells. *J. Mol. Biol.* **128**, 451–480 (1979).
24. Umlauf, D., Fraser, P. & Nagano, T. The role of long non-coding RNAs in chromatin structure and gene regulation: variations on a theme. *Biol. Chem.* **389**, 323–331 (2008).
25. Wilusz, J.E., Sunwoo, H. & Spector, D.L. Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev.* **23**, 1494–1504 (2009).
26. Mao, Y.S., Sunwoo, H., Zhang, B. & Spector, D.L. Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. *Nat. Cell Biol.* **13**, 95–101 (2011).
27. Delpretti, S. *et al.* Multiple enhancers regulate *Hoxd* genes and the *Hotdog* lncRNA during cecum budding. *Cell Rep.* **5**, 137–150 (2013).
28. Sun, L. *et al.* Long noncoding RNAs regulate adipogenesis. *Proc. Natl. Acad. Sci. USA* **110**, 3387–3392 (2013).
29. Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008).
30. Cuddapah, S. *et al.* Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res.* **19**, 24–32 (2009).
31. Chu, C., Quinn, J. & Chang, H.Y. Chromatin isolation by RNA purification (ChIRP). *J. Vis. Exp.* **61**, 3912 (2012).
32. Simon, M.D. *et al.* The genomic binding sites of a noncoding RNA. *Proc. Natl. Acad. Sci. USA* **108**, 20497–20502 (2011).
33. Mariner, P.D. *et al.* Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. *Mol. Cell* **29**, 499–509 (2008).
34. Lee, E.K. *et al.* miR-130 suppresses adipogenesis by inhibiting peroxisome proliferator-activated receptor γ expression. *Mol. Cell. Biol.* **31**, 626–638 (2011).
35. Nukitrangsan, N. *et al.* Effect of *Peucedanum japonicum* Thunb on the expression of obesity-related genes in mice on a high-fat diet. *J. Oleo Sci.* **60**, 527–536 (2011).
36. Rubi, B., del Arco, A., Bartley, C., Satrustegui, J. & Maechler, P. The malate-aspartate NADH shuttle member Aralar1 determines glucose metabolic fate, mitochondrial activity, and insulin secretion in beta cells. *J. Biol. Chem.* **279**, 55659–55666 (2004).
37. Seo, J. *et al.* Atf4 regulates obesity, glucose homeostasis, and energy expenditure. *Diabetes* **58**, 2565–2573 (2009).
38. Choy, L. & Derynck, R. Transforming growth factor-β inhibits adipocyte differentiation by Smad3 interacting with CCAAT/enhancer-binding protein (C/EBP) and repressing C/EBP transactivation function. *J. Biol. Chem.* **278**, 9609–9619 (2003).
39. Göhring, F. & Fackelmayer, F.O. The scaffold/matrix attachment region binding protein hnRNP-U (SAF-A) is directly bound to chromosomal DNA *in vivo*: a chemical cross-linking study. *Biochemistry* **36**, 8276–8283 (1997).
40. Hasegawa, Y. *et al.* The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. *Dev. Cell* **19**, 469–476 (2010).
41. Lobov, I.B., Tsutsui, K., Mitchell, A.R. & Podgornaya, O.I. Specificity of SAF-A and lamin B binding *in vitro* correlates with the satellite DNA bending state. *J. Cell. Biochem.* **83**, 218–229 (2001).
42. Huelga, S.C. *et al.* Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. *Cell Rep.* **1**, 167–178 (2012).
43. Xiao, R. *et al.* Nuclear matrix factor hnRNP U/SAF-A exerts a global control of alternative splicing by regulating U2 snRNP maturation. *Mol. Cell* **45**, 656–668 (2012).
44. Wong, L.H. *et al.* Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. *Genome Res.* **17**, 1146–1160 (2007).
45. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G.D. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS ONE* **5**, e13984 (2010).
46. Lee, B.K. & Iyer, V.R. Genome-wide studies of CCCTC-binding factor (CTCF) and cohesin provide insight into chromatin structure and regulation. *J. Biol. Chem.* **287**, 30906–30913 (2012).
47. Price, A.L., Jones, N.C. & Pevzner, P.A. *De novo* identification of repeat families in large genomes. *Bioinformatics* **21** (suppl. 1), i351–i358 (2005).
48. Donohoe, M.E., Silva, S.S., Pinter, S.F., Xu, N. & Lee, J.T. The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. *Nature* **460**, 128–132 (2009).

ONLINE METHODS

Repetitive sequence analysis (FSA). The sequence for Firre was scanned for repetitive elements with the *ab initio* repeat-detection algorithm RepeatScout[47]. This sequence was then aligned back to the genome (mm9 or hg19) with BLAT[49] with the following parameters: -stepSize=5 -repMatch=2253 -minScore=50 -minIdentity=0. Genomic DNA for the hits was extracted and multiply aligned with the Fast Statistical Aligner, a probabilistic multiple alignment tool specifically engineered to accommodate multiple alignment of sequences with potentially nonuniform evolutionary constraint[50]. Fast Statistical Alignment uses pair hidden Markov models to estimate gap and substitution parameters for the multiple alignment–scoring function, improving alignment robustness.

Cloning *Firre*. Total RNA (1 µg) was reverse transcribed according to the instructions in the Superscript III kit (Life Technologies, cat. no. 18080-051). The thermocycling conditions were 25 °C for 10 min, 55 °C for 1 h, 70 °C for 15 min and 4 °C final. 2 µL of the cDNA was mixed with 21 µL of water, 2 µL of 10 µM primers and 25 µL 2× Phusion Mastermix (New England BioLabs, cat. no M0531S). The PCR conditions were: (i) 98 °C for 30 s, (ii) 98 °C for 10 s, (iii) 66 °C for 30 s, (iv) 72 °C for 3 min, (v) 72 °C for 5 min and (vi) 4 °C final, with 45 cycles repeating steps 2–4. The extension time varied with the length of the lincRNA. The products were checked on 1% agarose gel. Nested PCR was performed when necessary with purified PCR products instead of cDNA. Longer isoforms were gel-purified and then subjected to the following cleaning steps. The PCR products were purified with SPRI beads (Beckman Coulter, cat. no. A63880), according to the instructions in the manual. SPRI beads were added to the PCR product and incubated at room temperature for 2 min. The mix was put on a magnet for 4 min, and the supernatant was removed. The beads were washed with 100 µL of 70% EtOH for 30 s twice and placed at 37 °C for 5 min until the beads appeared dry. The PCR product immobilized on the beads was eluted with 30 µL of water on the magnet for 5 min.

The purified PCR product was quantified and used in BP reactions. The amount of DNA to be added was calculated as described in the Gateway cloning manual (Invitrogen). The BP reaction was set up according to the BP Clonase II instructions (Life Technologies, cat. no. 11789020).

For transformations, 1 vial of Omnimax 2T1R (Life Technologies, cat. no 8540-03) cells was used for four BP reactions. The steps outlined in the Omnimax 2T1R manual were followed. The transformation plates were incubated at 37 °C overnight, and the colonies were sequenced through Genewiz. When the inserts were verified, the plasmids were prepared with the Qiagen miniprep kit (Qiagen, cat. no. 27104).

***In situ* hybridization.** The Firre probe was generated by PCR from adult-brain cDNA and subcloned in pCRII-TOPO (Life Technologies, cat. no. K4610-20). The antisense riboprobe was generated by *in vitro* transcription with SP6 polymerase (Roche Applied Science, cat. no. 10810274001) as previously described[51]. For nonradioactive *in situ* hybridizations, staged embryos were dissected in 1× PBS (Invitrogen) and fixed in 4% paraformaldehyde overnight at 4 °C. For E14.5 cross-sections, embryos were washed overnight at 4 °C in 30% sucrose/PBS and then by a 1:1 ratio of 30% sucrose/OCT Clear Frozen Section Compound (VWR, cat. no. 95057-838) for 1 h. Embryos were then placed in fresh OCT, frozen and stored at −80 °C until sectioning. Frozen serial sections, 20 µm thick, were prepared with a HM550 cryostat (Thermo Scientific) and mounted onto Superfrost Plus slides (VWR, cat. no. 48311-703). Sections were permeabilized with 10 µg/ml proteinase K (Roche Diagnostics) for 10 min, washed with 1× PBS, treated 10 min in RIPA buffer and cross-linked again for 5 min in cold 4% paraformaldehyde. Sections were then prehybridized for 1 h at room temperature at 70 °C (50% formamide, 5× SSC, 5× Denhardts, 500 µg/ml salmon sperm DNA and 250 µg/ml yeast RNA) and then incubated overnight at 70 °C in the same solution containing 2 µg/ml of DIG-labeled riboprobe. Sections were washed, blocked 1 h with 10% sheep serum and incubated overnight at 4 °C with 0.375 U/ml alkaline phosphatase–labeled anti-DIG antibody (Roche Diagnostics). Signal was detected by exposure of sections to NBT-BCIP (Sigma, cat. no. B1911), 0.1% Tween-20. Reactions were stopped with washes in 1× PBS supplemented with 0.1% Tween-20, and sections were mounted in Fluoromount-G (Southern Biotech, cat. no. 0100-01). The E14.5 whole-embryo cross-section image (**Fig. 1c**) was taken from the Eurexpress Database (assay ID euxassay_013928, http://www.eurexpress.org/). Nonradioactive *in situ* hybridizations of E15.5 embryo brains were performed on 40-µm vibratome sections (Leica) mounted on Superfrost Plus slides (VWR) with reported methods[52]. Sense probes were used as negative controls in all experiments. The PCR primers used to generate the Firre probe and the Firre riboprobe sequence used for the hybridizations are listed in **Supplementary Table 4**.

Cellular fractionation. The cells grown in 15-cm dishes were washed with 5 mL of 1× PBS and trypsinized with 3 mL of TrypLE (Invitrogen) at 37 °C for 3–5 min. The trypsin was quenched with five volumes of ice-cold growth medium (DMEM (Invitrogen), 10% FBS (Invitrogen), 1% pen-strep (Invitrogen), and 1% L-glutamine (Invitrogen)), and the cells were pelleted at 200*g* for 3 min and resuspended in 1 ml of ice-cold 1× PBS. The resuspension was centrifuged at 200*g* for 10 min at 4 °C. The supernatant was carefully removed without disturbing the pellet; the remaining packed pellet volume was estimated for the next steps. The pellet was resuspended in five packed pellet volumes of ice-cold cytoplasmic extraction buffer (20 mM Tris, pH 7.6 (Ambion), 0.1 mM EDTA (Ambion), 2 mM $MgCl_2$ (Ambion), 1× protease inhibitors (VWR) and 0.5 U/µL RNaseOUT (Invitrogen)). The cells were incubated first at room temperature for 2 min, then on ice for 10 min. The cells were lysed by addition of CHAPS to a final concentration of 0.6%. The sample was then homogenized by passage through a 1-ml syringe and was centrifuged at 500*g* for 5 min at 4 °C. The 70–80% of the supernatant was taken and saved at −80 °C; this was the cytoplasmic fraction. The remaining supernatant was carefully removed, and the pellet was washed with cytoplasmic extraction buffer supplemented with 0.6% (w/v) CHAPS. The sample was centrifuged at 500*g* for 5 min at 4 °C, and the entire supernatant was discarded. The wash step was repeated one more time. The pellet was then resuspended in two packed pellet volumes of nuclei suspension buffer (10 mM Tris, pH 7.5, 150 mM NaCl, 0.15% (v/v) NP–40, 1× protease inhibitors and 0.5 U/µL RNaseOUT). The nuclear suspension was layered on five packed pellet volumes of sucrose cushion (10 mM Tris, pH 7.5, 150 mM NaCl, 24% (w/v) sucrose, 1× protease inhibitors and 0.5 U/µL RNaseOUT) and pelleted at 14,000 r.p.m. for 10 min at 4 °C. The supernatant was discarded, and the pellet was washed with ten packed pellet volumes of ice-cold 1× PBS supplemented with 1 mM EDTA. The sample was then centrifuged at 500*g* for 5 min at 4 °C. The pellet constituted the nuclear fraction.

Fluorescence *in situ* hybridization (FISH). The FISH protocol was performed as described previously[29]. Briefly, oligonucleotide probes targeting and tiling the intron of Firre were conjugated to Alexa594 fluorophores, and the probes targeting and tiling the exon were conjugated to tetramethylrhodamine (TMR) and purified by HPLC. Before the hybridization, the adherent cells were fixed (10 min with 4% formaldehyde) and permeabilized with 70% EtOH in two-chamber cover glasses. mESCs and hESCs were fixed in solution after they were collected from the plate: the cells were incubated at room temperature in 2% formaldehyde solution for 10 min, and this was followed by centrifugation at 1,000*g* for 3 min. The cells were washed with 1× PBS twice with centrifugation at 1,000*g* for 3 min in between. The cells were permeabilized with 70% EtOH. The ESCs were then plated on gelatinized cover glasses. Prior to the hybridization, the cells were rehydrated with wash buffer containing 10% formamide (Ambion, cat. no. AM9342) and 2× SSC (Ambion, cat. no. AM9765) for 5 min. Then the probes (0.5 ng/µL final) were hybridized in 10% dextran sulfate (Sigma, cat. no. D8906), 10% formamide and 2× SSC at 37 °C overnight. After hybridization, the cells were washed in wash buffer at 37 °C for 30 min twice (with the addition of DAPI in the second wash) and then in 2× SSC twice. The imaging was done immediately after with 2× SSC as the mounting medium.

The same protocol was followed for the co-FISH experiments. The probes targeting and tiling the introns of the *trans* sites (*Slc25a12*, *Ypel4*, and *Ppp1r10*) were conjugated to Quasar570. Co-FISH assays were conducted as indicated in wild-type male mESCs, Δ*Firre* male mESCs or mLFs. Quasar670 was used as an additional fluorophore when working with three colors and *trans* sites; Quasar570 and Quasar670 were used together when staining for *trans* sites.

Actinomycin-D treatment. Actinomycin-D (Act-D) (Sigma, cat. no. A9415-2MG) was resuspended in DMSO with a final concentration of 2 mg/mL. Act-D was thoroughly mixed with the 2I medium (2 µg/ml) and added on the male mESCs at 0, 1.5, 3 and 6 h.

RNA antisense purification (RAP) analysis. RAP was performed as described[19]. Briefly, the RNA of interest was tiled with 120-bp antisense nucleotides that had been biotinylated. Two distinct pools of antisense probes targeting Firre and one pool containing sense probes (negative control) were generated. The hybridization was done in duplicate and cross-linked, and 20 ng (350 fmol) of oligos were added to the precleared lysates and the mix was rotated at 4 °C. The oligos were then captured by streptavidin beads, and the elutions for RNA and DNA were performed. Consistently with standard ChIP-seq assays, duplicate pulldowns were performed and sequenced to control for technical variability.

For X-chromosome enrichment analysis (**Fig. 3a**), the X chromosome was divided into 10-Kb bins, and a linear regression of counts per bin was performed against each replicate and the input control. The slope of the linear regression was used as a normalization factor (alpha) between the two libraries. Enrichment levels relative to input were calculated by division of the experimental counts for each bin by the input counts times alpha.

To identify regions bound by the Firre RNA in *trans*, we used the Scripture peak-calling algorithm[53] to call significant peaks across each of the replicate sequencing .bam files, including the input control and antisense control. All peaks were merged with Bedtools mergeBed[54] to obtain the universe of significant peaks across all samples. A .gtf file of significant peaks, along with the replicate .bam files for each of the samples was used as input for Cuffdiff2 (ref. 26) for quantification and differential testing. Cuffdiff2 was run with default parameters with the addition of the '–no-length-correction' argument to disable length correction. Significant peaks were called with the Cuffdiff2 test statistic with $P < 0.1$.

Retroviral overexpression of Firre. The overexpression vector for Firre was made by modification of the pLenti6.3/TO/V5-DEST (Snap Gene) destination vector. We modified it by removing the WPRE, the SV40 promoter and the blasticidin-resistance gene, keeping the gateway tails the same, to prevent any interference with the lincRNA structure and function. All the transductions were done as follows: the cells were split into 12-well dishes and resuspended in medium with 4 μg/mL polybrene. Immediately after, 100 μL of virus (of the same titer or with volume adjustment if not the same titer) was added to each well. The untransduced control was used to measure the overexpression levels by qRT-PCR. The sequences of the isoforms 1.1 (+RRD), 1.4 (+RRD) and 1.6 (–RRD) used for mouse transductions are in **Supplementary Table 1** (.bed file), and the sequence of the human isoform used for HEK293 and HeLa transductions is given in **Supplementary Table 4**.

Cell culture. HEK293 (ATCC, CRL-1573), HeLa (ATCC, CCL-2), mLF (ATCC, CCL-206) and hLF (ATCC, IMR90) cells were grown in growth medium (described above) at 37 °C at 5% CO_2. Male (Novus, NBP1-41162) and female (RIKEN, AES0010) mouse ESCs were grown in previously gelatinized (0.2%) dishes with 2I medium containing 125 mL DMEM/F12 (Invitrogen), 83.5 μL BSA fraction V (50 μg/ml relative to DMEM) (Invitrogen, 15260-037, 75 mg/mL), 125 mL Neurobasal medium (Invitrogen, 21103-049), 625 μL of the Ndiff Neuro2 (200×, relative to Neurobasal medium) (Millipore, SCM012), 2.5 mL B27 minus vitamin A (50×, relative to Neurobasal medium) (Invitrogen, 12587-010), 2 μL β-mercaptoethanol, 1 μM PD0325901 (Stemgent, 04-0006), 3 μM CHIR99021 (Stemgent 04-0004), 25 μL LIF ESGRO (from Chemicon, ESG1106), 1% pen-strep (Invitrogen, 15140-163), 1% nonessential amino acids (Invitrogen, 11140-076) and 1% L-glutamine (Invitrogen, 25030-164). The plating density of mES cells was chosen to be 30,000–50,000/cm^2. Adipocytes were grown as described previously[28].

Pulldown with *in vitro*–biotinylated RNA. The cloned-in transcript in pdest14 plasmid vector was linearized with Nhe1. Phenol-chloroform–extracted and ethanol-precipitated template was then used in *in vitro* transcription, which included 20 μg/ml DNA template, 40 mM Tris, pH 7.9, 2.5 mM spermidine, 26 mM $MgCl_2$, 0.01% Triton X-100, 8 mM GTP, 5 mM ATP, 5 mM CTP, 1.3 mM UTP, 0.7 mM Bio-16-UTP (Epicentre), 5 mM DTT, 20 mM $MgCl_2$, 80 U/ml RNAseOUT and 20 U T7 RNA polymerase (Life Technologies, cat. no. 18033-019). The mix was incubated at 37 °C until a white precipitate formed. After the reaction reached completion, EDTA to 60 mM final concentration was added to dissolve the precipitate. Owing to the biotin, RNA will partition into the organic layer if it is phenol extracted. Therefore, the *in vitro* transcriptions were first treated with DNase (Worthington, cat. no. LS006353) (37 °C for 10 min followed by EDTA addition and 75 °C for 10 min), then cleaned with Bio-Spin 30 columns (Bio-Rad, cat. no. 732-6231).

For end-labeling, the *in vitro*–transcribed RNA (without labeled UTPs) was treated with the 3′-end–biotinylation kit of Thermo Scientific Pierce (cat. no. 20160).

The lysate was prepared by lysis of 15 cm dishes in 1 ml of lysis buffer (150 mM KCl, 25 mM Tris-HCl, pH 7.4, 5 mM EDTA, 5 mM $MgCl_2$, 1% NP-40, 1× protease inhibitor, 0.5 mM DTT, 100 U/ml RNAseOUT) for 30 min at 4 °C. The lysate was centrifuged at 13,000 r.p.m. for 30 min and filtered with a 0.45-μm filter. The concentration of the lysate was measured by BCA protein assay (Thermo Scientific, cat. no. 23225).

For the pulldown, 1.5 mg of the lysate was initially precleared with Magnetic MyOne Streptavidin T1 beads (Life Technologies, cat. no. 65601) for 30 min at 4 °C. The beads were prepared as described in the manual. The precleared lysate was diluted 2× and supplied with 0.1 μg/μl tRNA, to which 30 pmoles of biotinylated RNA was added. The RNA was incubated in the lysate for 2 h at 4 °C rocking, after which 40 μl of MyOne Streptavidin T1 beads was added to the mix. The mix was incubated for another hour at 4 °C. The beads were washed three times (10 min each) with 1 ml of the wash buffer (lysis buffer but with 300 mM KCl) on a magnetic rack. Finally, the beads were resuspended in 30 μl of sample buffer (4×, Bio-Rad) and reducing agent (20×, Bio-Rad) and boiled for 5 min at 95 °C. The samples were then run on 4–12% gradient Bis-Tris gels and stained with Sypro Red as described (Life Technologies, cat. no. S-12000) for protein detection and MS or were transferred to a PVDF membrane for western blotting.

The MS analysis was done as follows. The eluates from RNA pulldowns done in three different cellular contexts (mouse adipose tissue, mouse adipocyte and mESC lysates) with five different RRD+ isoforms and one RRD– isoform were run on a gel as described. The bands that were differential between RRD+ and RRD– isoforms were cut and processed for MS. To identify proteins that preferentially coprecipitated with Firre in an RRD-dependent manner, we took the difference between the peptide counts of the RRD+ and RRD– isoforms. We repeated the differential analysis for each of the RRD+ isoforms and took the top 10% of the differential peptide-count scores for each isoform identified in both mESC and adipocyte lysates.

The transfer for western blotting was done in transfer buffer that was prepared in the following ratios: 100 ml of 10× TG (Bio-Rad), 200 ml methanol and 700 ml ddH_2O. The membrane was activated in methanol first and equilibrated in transfer buffer before transfer. The transfer was done at 70 W for 1 h. After the transfer, the membrane was washed with methanol for blocking and incubated with the primary antibody diluted 1:1,000 in 0.1% Tween, 1% nonfat milk and 1× PBS for 4–5 h at RT or 4 °C O/N. After the primary-antibody incubation (human hnRNPU (3G6), Santa Cruz sc-32315 (validation of reactivity shown on the Santa Cruz website) and mouse hnRNPU, Abcam ab20666 (validated in ref. 55)), the membrane was washed three times with wash buffer (1× PBS supplemented with 0.1% Tween) and then incubated with the secondary antibody for 1 h at RT. The membrane was washed again three times with wash buffer and then developed with SuperSignal West Chemiluminescent Thermo Scientific reagents. 20% of the input lysate was used for all the western blots (**Fig. 6a** and **Supplementary Fig. 5d**).

Endogenous RNA pulldown. The 23-bp to 25-bp oligos (sequences in **Supplementary Table 4**) were synthesized with 18S linker and desthiobiotin at the 5′ end through IDT. The same protocol for RNA pulldowns above was followed, except instead of addition of biotinylated RNA, the DNA oligos were added to the lysate, and an annealing step was followed. Annealing was done by incubation of the lysate and the oligos at 37 °C for 15 min, at room temperature for 15 min and and at 4 °C for 6 h to O/N. The rest of the steps were the same as above, except (i) during the incubation with the beads, heparin was spiked in at 0.5 μg/μL in the last half hour; (ii) in the first wash, 0.5 μg/μL heparin was spiked in again; and (iii) instead of boiling of the beads, the RNA and protein were eluted from the beads with 12.5 mM biotin for 30 min at RT and 3 h at 4 °C. 60% of the elution was used for RNA extraction and 40% to run a protein gel.

RNA immunoprecipitation. The protein lysate (1.5 mg) was incubated with 6–8 μg of the hnRNPU or IgG antibody (hnRNPU (3G6), Santa Cruz sc-32315; hnRNPU (H-94), Santa Cruz sc-25374 (validation on the Santa Cruz website); Anti-hnRNPU, Abcam ab20666; Mouse (G3A1) IgG1, Cell Signaling 5415) at

4 °C for 2–3 h (antibody validation is available on the manufacturers' websites). Then 45 µL of protein G Dynabeads (Life Technologies, cat. no. 10003D) that were previously washed twice in 500 µL lysis buffer were added to the lysate and antibody mix. The lysate, antibody and beads were incubated at 4 °C for another 2 h. The beads were washed three times (10 min each) with 1 mL of lysis buffer. The RNA was extracted by addition of 1 mL of TRIzol (Life Technologies, cat. no. 15596-018) to the beads. For the total input RNA, 10% of the input lysate was mixed with 1 mL of TRIzol. For 1 ml of TRIzol, 200 µL of chloroform was added, and the mix was centrifuged at 4 °C at 13,000 r.p.m. for 15 min. The aqueous layer was then added to 1 volume of isopropanol, 1/10 volume KOAc and 1 µL of glycoblue and kept at −20 °C for at least one hour. The samples were then centrifuged at 13,000 r.p.m. at 4 °C for 30 min. The supernatant was removed, and the pellet was washed with 1 mL of ice-cold 70% EtOH twice (centrifuging 2 min each time at 4 °C). The pellet was then resuspended in 15 µL of RNase-free water.

CLIP-seq analysis. We analyzed two human hnRNPU CLIP-seq data sets generated independently by Huelga *et al.* (GSE34993)[42] and Xiao *et al.* (GSE34491)[43]. We downloaded fastq files and aligned with TopHat[56] to hg19 and a custom transcriptome GTF consisting of UCSC coding genes, a recently published lncRNA catalog[57] and cloned Firre isoforms. We used the RNA-seq differential expression software Cuffdiff to estimate read counts to Firre in all CLIP and total RNA data sets[26]. We performed a Poisson-based statistical test for enrichment of aligned reads in the CLIP versus total RNA.

RNAi-mediated knockdown of hnRNPU. mESCs were transfected by the reverse-transfection method in six-well-plate format. Briefly, the Lipofectamine RNAiMAX (6 µl/well, Life Technologies, cat. no. 13778030) and siRNA (50 nM final, Dharmacon On-Targetplus smart pool for mouse (L-051574-01-0005) and for human (L-013501-00-0005)) complexes were prepared in 400 µl of Opti-MEM and incubated at room temperature for 20–30 min, during which the mESCs were prepared for splitting. The split was done as follows: the cells were washed with 1× PBS and trypsinized for 3 min. The trypsin was quenched by 2I medium, and the cells were then centrifuged for 5 min at 850 r.p.m. at 4 °C, resuspended in new 2I medium and counted. Approximately 280,000 cells were plated for each well of a six-well plate. Immediately after, the Lipofectamine–oligo complexes were added to the wells. The medium was changed after 24 h, and the cells were harvested after 96 h for the complete knockdown of the protein (checked by qRT-PCR and by western blot). The same protocol was used for HEK293s and HeLa cells; however, the cells were plated at a density of ~21,000 cells/cm^2.

RNA extraction. RNA extraction was performed by addition of 1 ml of TRIzol to each well of a six-well plate. 200 µL of chloroform was added, and the mix was centrifuged at 4 °C at 13,000 r.p.m. for 15 min. The aqueous layer was processed on the RNeasy Mini columns (Qiagen, cat. no. 74104). The RNA was reverse transcribed with the SuperScriptIII First-Strand Synthesis kit. The cDNA synthesis was performed at 25 °C for 5 min, 50 °C for 1 h and 70 °C for 15 min. Then the cDNA (15 ng per well of the 384-well qPCR plate) was added 1:1 to the SYBR and primer mix (100 nM) for qRT-PCR. All the primers used in qRT-PCR are shown in **Supplementary Table 4**.

RNA-seq library preparation, sequencing and analysis. 200 ng of extracted RNA from each of two Δ*Firre* and three wild-type JM8A male mouse ES-cell cultures was used as input for the Illumina TruSeq library preparation kit, according to the manufacturer's guidelines. Libraries were individually barcoded, and library size distribution and quality were assayed with a DNA High-sensitivity Chip on the Agilent Bioanaylzer 2100. Libraries were pooled, and paired-end 35-bp fragments were generated on an Illumina MiSeq sequencer to an average depth of 9 million fragments per sample. Fragments were aligned to the mouse genome (mm9) with Tophat2 (ref. 58) with default options and the UCSC transcriptome as a reference. Aligned reads were quantified against all mouse UCSC genes with Cuffdiff2 (ref. 59) with default options. Significantly differentially expressed (DE) genes were selected with an FDR of 5%. A circular representation of DE gene projections onto mm9 (**Fig. 5f**) was generated with the Circos utility (http://circos.ca/). Because RNA-seq data are heteroscedastic and not normally distributed, we chose to evaluate pathway enrichment by using a preranked GSEA analysis. This was conducted by the GSEA tool[60] with a list of all genes ranked by cuffdiff2 test statistic (KO/WT) against the c2.cp.reactome.v4.0.symbols.gmt gene-set collection (MSigDB, Broad). Gene sets were selected as significantly enriched if the nominal *P* value was less than 0.01.

Targeting and generation of conditional Δ*Firre* mESCs. To generate ESCs specifically deficient in Firre, a two-step targeting strategy was used to introduce loxP sites in the 5′ and 3′ ends of the *Firre* locus. Targeting of only one allele was needed to obtain *Firre*-deficient ESCs, because *Firre* resides on the X chromosome, and male ESCs (JM8) were used. To generate the *Firre* 3′ targeting construct, we amplified the homology arms (primer sequences in **Supplementary Table 4**), which were then cloned into the pEASY-FLIRT vector.

In this targeting construct, the neo cassette that serves as a selection marker during the targeting process was flanked by two Frt sites. To generate the *Firre* 5′ targeting construct, we amplified the homology arms (primer sequences in **Supplementary Table 4**), which were then cloned into the newly generated pEASY-Hygro vector. In this construct, the hygro cassette that serves as a selection marker during the targeting process was flanked by two loxP sites.

ES cells were maintained under standard conditions and targeted as previously described[30]. In brief, the vector targeting the 3′ end of *Firre* was electroporated into C57BL/6 ES cells (JM8) and grown under selection with neomycin. Homologous-recombinant ES cells identified by PCR analysis were subsequently electroporated with the vector targeting the 5′ end of *Firre* and grown under selection with hygromycin. Double-targeted ES cells were identified by PCR analysis. To delete *Firre*, double-targeted ES cells were electroporated with a Cre recombinase–expressing plasmid (pGK-Cre-bPA). PCR genotyping was used to identify clones in which *Firre* had been deleted.

49. Kent, W.J. BLAT: the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
50. Bradley, R.K. *et al.* Fast statistical alignment. *PloS Comput. Biol.* **5**, e1000392 (2009).
51. Arlotta, P. *et al.* Neuronal subtype-specific genes that control corticospinal motor neuron development *in vivo. Neuron* **45**, 207–221 (2005).
52. Tiveron, M.C., Hirsch, M.R. & Brunet, J.F. The expression pattern of the transcription factor Phox2 delineates synaptic pathways of the autonomic nervous system. *J. Neurosci.* **16**, 7649–7660 (1996).
53. Guttman, M. *et al. Ab initio* reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat. Biotechnol.* **28**, 503–510 (2010).
54. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
55. Vizlin-Hodzic, D., Johansson, H., Ryme, J., Simonsson, T. & Simonsson, S. SAF-A has a role in transcriptional regulation of Oct4 in ES cells through promoter binding. *Cell. Reprogram.* **13**, 13–27 (2011).
56. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
57. Kelley, D. & Rinn, J. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biol.* **13**, R107 (2012).
58. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
59. Trapnell, C. *et al.* Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **31**, 46–53 (2013).
60. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).