Evolutionary divergence of *Firre* localization and expression

CHRISTIAN MUCH,[1,5] MICHAEL J. SMALLEGAN,[1,2,5] TAEYOUNG HWANG,[1] SKYLAR D. HANSON,[1] GABRIJELA DUMBOVIĆ,[1,3,6] and JOHN L. RINN[1,4,6]

[1]BioFrontiers Institute, University of Colorado Boulder, Boulder, Colorado 80303, USA
[2]Department of Molecular, Cellular and Developmental Biology, University of Colorado Boulder, Boulder, Colorado 80302, USA
[3]Institute for Cardiovascular Regeneration, Goethe University Frankfurt, 60596 Frankfurt am Main, Germany
[4]Department of Biochemistry, University of Colorado Boulder, Boulder, Colorado 80302, USA

ABSTRACT

Long noncoding RNAs (lncRNAs) are rapidly evolving and thus typically poorly conserved in their sequences. How these sequence differences affect the characteristics and potential functions of lncRNAs with shared synteny remains unclear. Here we show that the syntenically conserved lncRNA *Firre* displays distinct expression and localization patterns in human and mouse. Single molecule RNA FISH reveals that in a range of cell lines, mouse *Firre* (*mFirre*) is predominantly nuclear, while human *FIRRE* (*hFIRRE*) is distributed between the cytoplasm and nucleus. This localization pattern is maintained in human/mouse hybrid cells expressing both human and mouse *Firre*, implying that the localization of the lncRNA is species autonomous. We find that the majority of *hFIRRE* transcripts in the cytoplasm are comprised of isoforms that are enriched in RRD repeats. We furthermore determine that in various tissues, *mFirre* is more highly expressed than its human counterpart. Our data illustrate that the rapid evolution of syntenic lncRNAs can lead to variations in lncRNA localization and abundance, which in turn may result in disparate lncRNA functions even in closely related species.

Keywords: *Firre*; lncRNA; localization; conservation

INTRODUCTION

Pervasive transcription of mammalian genomes produces thousands of lncRNAs that are rapidly evolving and therefore usually poorly conserved at the sequence level (Pang et al. 2006). However, as an attestation to their functional relevance, evolutionary conservation of lncRNAs can be found at other levels, including structure and synteny (Diederichs 2014; Ulitsky 2016; Ross et al. 2021). LncRNAs with the same secondary or tertiary structure can exert identical molecular functions despite divergent nucleotide sequences (Smith et al. 2013). In contrast, lncRNAs expressed from syntenically conserved regions, in which the lncRNA gene locus is embedded in the same chromosomal context with identical neighboring genes, may differ in both sequence and structure. Hundreds of syntenic lncRNA loci have been identified across vertebrates, in which the respective transcript can be similar or entirely disparate in function (Hezroni et al. 2015).

The function of a lncRNA is commonly tied to its subcellular localization. In contrast to mRNAs that are typically enriched in the cytoplasm in order to get translated, lncRNAs can be found virtually throughout the cell, and therefore can act in a wide range of cellular processes. Nuclear lncRNAs, for instance, have been described as regulating transcription, chromatin architecture, chromatin remodeling, and nuclear bodies (Yao et al. 2019). Cytoplasmic lncRNAs, on the other hand, have been implicated in the control of translation, mRNA turnover, and post-translational modifications (Yao et al. 2019). Ultimately, the subcellular localization of a lncRNA will be defined by its sequence motifs and its interactions with various factors in *trans*. Several characteristics of lncRNAs have been linked to their increased nuclear compartmentalization, including their inefficient splicing, intron retention, facilitated association with nuclear RNA-binding proteins, and tethering to chromatin or the nuclear lamina (Derrien

[5]These authors contributed equally to this work.
[6]These authors contributed equally to this work.
Corresponding authors: john.rinn@colorado.edu, dumbovic@med.uni-frankfurt.de
Article is online at http://www.rnajournal.org/cgi/doi/10.1261/rna.079070.121.

et al. 2012; Miyagawa et al. 2012; Chen et al. 2016; Mele et al. 2017; Azam et al. 2019; Dumbovic et al. 2021; Statello et al. 2021). Short RNA motifs that guide nuclear retention, such as cytosine-rich elements, have been identified by mutagenesis analyses and high-throughput screening approaches (Zhang et al. 2014; Lubelsky and Ulitsky 2018; Shukla et al. 2018). The presence of such motifs alone or their combinatorial enrichment in a larger sequence context could potentially initiate the binding of nuclear factors—directly through their primary sequence, by harboring specific RNA modifications, or by forming higher-order structures—and thus enrich a lncRNA in the nucleus.

Another critical aspect for the role of a lncRNA in the cell is its abundance. As opposed to mRNAs, many lncRNAs are on average lowly expressed with a high degree of tissue-specificity, although notable exceptions exist (Cabili et al. 2011; Derrien et al. 2012). While some molecular processes may require only a few molecules (e.g., *linc-p21*, Dimitrova et al. 2014; Groff et al. 2016), others depend on several hundred transcripts per cell (e.g., *NORAD*, Lee et al. 2016). Importantly, lncRNA expression has to be fine-tuned to ensure the stoichiometry of molecular interactions in *cis* and in *trans*, for example when recruiting chromatin modifiers to different genomic loci (Rinn et al. 2007; Wang et al. 2011), sequestering proteins (Wu et al. 2016) or competing for miRNA binding (Cesana et al. 2011; Salmena et al. 2011).

The *Firre* locus is syntenically conserved between human and mouse, and its duplication has been associated with developmental disorders (Abe et al. 2014; Ha et al. 2019; Miolo et al. 2020). In *mFirre* loss- and gain-of-function mouse models, *mFirre* was shown to modulate lymphogenesis as well as the inflammatory response (Lewandowski et al. 2019). Interestingly, several lines of evidence point to exclusively nuclear roles of *mFirre*, including the regulation of H3K27me3 (Yang et al. 2015; Fang et al. 2020) and the organization of nuclear architecture (Hacisuleyman et al. 2014; Darrow et al. 2016; Barutcu et al. 2018; Bonora et al. 2018; Froberg et al. 2018; Kriz et al. 2021). Indeed, single-molecule RNA fluorescence in situ hybridization (smRNA FISH) demonstrates a nuclear localization of *mFirre* in mouse cells (Hacisuleyman et al. 2014; Bergmann et al. 2015; Lewandowski et al. 2019). However, the subcellular localization of *hFIRRE* has not conclusively been shown, and a direct comparison between human and mouse *Firre* in terms of localization and abundance as an approximation of the lncRNA's function is lacking.

In this study we focused on *Firre* to determine if syntenic lncRNA loci that lack sequence conservation have similar properties across an array of human and mouse cell lines and tissues. Using smRNA FISH and computational analyses, we observe a difference in localization and abundance between human and mouse *Firre*. Specifically, we detect that *mFirre* is localized in the nucleus, while *hFIRRE* is both nuclear and cytoplasmic, a property that we show is sequence intrinsic. We find that the cytoplasmic population of *hFIRRE* transcripts is enriched in local repeating RNA domains (RRDs). In all tissues examined, *mFirre* is more abundant than *h*FIRRE. Further promoter analysis revealed surprisingly similar transcription factor (TF) binding site (TFBS) instances between human and mouse *Firre* despite highly divergent promoter sequences. Collectively, our study highlights important differences in localization and expression of human and mouse *Firre*, suggesting the potential for distinct functionalities of a syntenically conserved lncRNA.

RESULTS

Firre lncRNA subcellular localization differs in human and mouse cells

At the sequence level, the lncRNA *Firre* is poorly conserved in mammals. Multiple-sequence-alignment analyses of the gene locus and its repeats suggest an ectopic evolutionary event in the primate lineage that conserves the sequence of *Firre* only in primates. Conversely, *Firre* is syntenically conserved across the mammalian clade—yet lacks sequence conservation (Hacisuleyman et al. 2016). For example, the *Firre* locus is syntenically conserved on the X-chromosome in human and mouse (Fig. 1A). This level of conservation opens the possibility that *Firre* adopted distinct functions in these species. To investigate this hypothesis, we set out to study the localization of *Firre* as an indicator of its function in human and mouse cells. For this purpose, we performed smRNA FISH using probes targeting *Firre* exons. To ensure the specificity of the FISH signal, we applied a very stringent probe design algorithm that dismisses any probe sequences with a potential to substantially cross-hybridize with non-*Firre* transcripts. The specificity of mouse and human probes was confirmed in *Firre* knockout mouse ES cells and in human foreskin fibroblast cells that do not express *hFIRRE*, respectively (Supplemental Fig. S1A,B). We then quantified the number of nuclear and cytoplasmic *Firre* FISH signals in human iPS and mouse ES cells. We found *mFirre* to be exclusively enriched in the nucleus of mouse ES cells; however, *hFIRRE* RNA was detected in both the nucleus and cytoplasm of human iPS cells (Fig. 1B,C; Supplemental Fig. S1C,D). In both human and mouse stem cells, *Firre* showed cell-to-cell variability in expression levels, with cells expressing between five to over a hundred molecules. On average, *Firre* displayed a higher expression in mouse ES cells (on average, 34 molecules) than in human iPS cells (on average, 19 molecules, Fig. 1B,C; Supplemental Fig. S1C,D).

We next wanted to investigate if this disparate localization and expression pattern between human and mouse was consistent in other cell types. For this purpose, we selected a set of human and mouse cell lines and performed

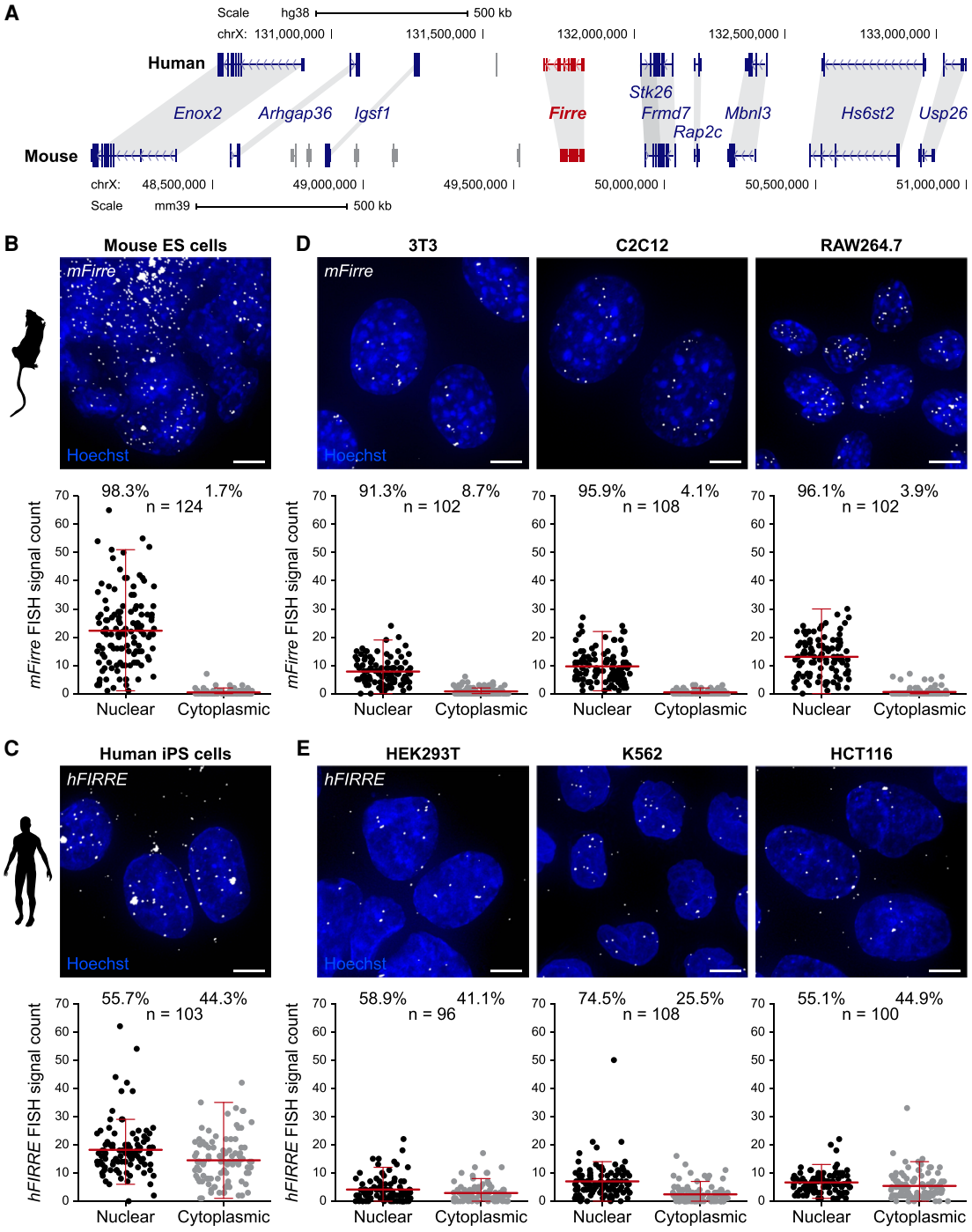

FIGURE 1. *Firre* is predominantly nuclear in mouse but not in human cells. (*A*) Schematic depicting the human and mouse *Firre* locus (red) and their neighboring coding genes that are shared (blue) and species-specific (gray). (*B–E*) Maximum intensity projections of *Firre* smRNA FISH on mouse ES cells (*B*), human iPS cells (*C*), mouse 3T3, C2C12, and RAW264.7 cell lines (*D*), and human HEK293T, K562, and HCT116 cell lines (*E*). *Below*, quantification of *Firre* FISH signals in the nucleus and cytoplasm of each cell line. Nuclei are stained with Hoechst (blue), *Firre* exon in gray. Scale bar is 5 µm. Data are mean and non-outlier range of quantified cells.

smRNA FISH staining for *Firre*. Quantitative analysis revealed that similarly to mouse ES cells, *mFirre* was predominantly to exclusively nuclear in all investigated mouse cell lines (3T3, C2C12, and RAW264.7, Fig. 1D; Supplemental Fig. S1C). Conversely, in human cell lines (HEK293T, K562, and HCT116), *hFIRRE* was localized to both the nucleus and the cytoplasm (Fig. 1E; Supplemental Fig. S1D), in accordance with our observations in human iPS cells. Although

the total number of *Firre* molecules in human (on average, 6 nuclear and 4 cytoplasmic) and mouse cell lines (on average, 10 nuclear and 0.6 cytoplasmic) was comparable, considering the well described nuclear functions of *mFirre*, it is important to note that the mouse cell lines on average possessed more nuclear *Firre* than the human cell lines (Supplemental Fig. S1C,D). As the cytoplasmic FISH signal can be observed in four human cell lines of diverse cellular contexts but not in human foreskin fibroblasts (Supplemental Fig. S1B), it can be concluded that cross-hybridization to an off-target RNA with the quantified abundance of <10 molecules per cell is unlikely and that cytoplasmic *hFIRRE* is specifically detected. Abundance levels as determined by smRNA FISH are largely in good agreement with publicly available RNA-seq data of human and mouse *Firre* expression levels in the investigated cell lines (Supplemental Fig. S1E,F). Of note, as stem cells tend to express more genes and more isoforms of each gene compared to differentiated cells, their TPM values might be underestimating *Firre* expression levels in relation to the other cell lines (Ramalho-Santos et al. 2002; Pritsker et al. 2005; Wu et al. 2010). In summary, we show that the *Firre* lncRNA exhibits a different subcellular localization in human and mouse cells, with *hFIRRE* being both nuclear and cytoplasmic and *mFirre* being predominantly nuclear.

The differential localization of human and mouse *Firre* lncRNA is sequence-driven

We next wanted to understand whether the distinct subcellular localization of human and mouse *Firre* is dependent on their RNA sequence or cellular factors. To this end, we analyzed the subcellular distribution of human and mouse *Firre* within the same cellular context. We performed the analysis on a human/mouse hybrid cell line that carries a single human X-chromosome in an otherwise mouse cell background (Fig. 2A). We reasoned that if the differences in the subcellular localization of human and mouse *Firre* are driven by species-specific cellular environments and not by the sequence, human and mouse *Firre* would adopt equivalent subcellular distributions in the same cellular context. On the other hand, both RNAs will adopt their species-specific distribution if the difference is driven by their sequence. RNA-seq confirmed that the hybrid cell line expressed both *hFIRRE* and *mFirre* and that within the same cellular context *mFirre* is expressed more abundantly than *hFIRRE* (Fig. 2B). Of note, since ~90% of the cells retain the human X-chromosome, bulk RNA-seq will underestimate the expression levels of *hFIRRE* compared to *mFirre*. In order to precisely dissect the subcellular localization of human and mouse *Firre* in the same cellular context, we

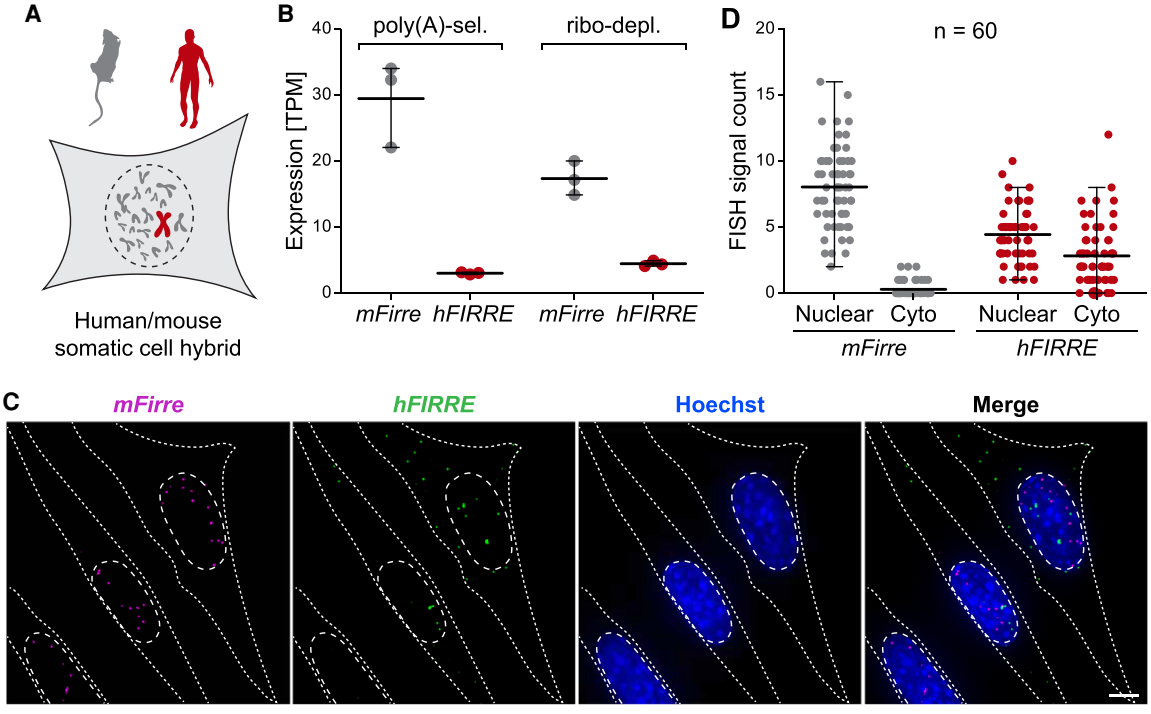

FIGURE 2. *Firre* lncRNA localization is driven by its sequence. (*A*) Schematic showing a human/mouse somatic cell hybrid that harbors a human X-chromosome in a mouse cell background. (*B*) RNA-seq analysis of *mFirre* and *hFIRRE* expression in hybrid cells. Libraries were prepared using poly(A) enrichment or ribosomal RNA depletion. Data are mean and non-outlier range of three biological replicates. (*C*) Maximum intensity projections of smRNA-FISH for *hFIRRE* and *mFirre* on hybrid cells. Nuclei are stained with Hoechst (blue), *mFirre* exon in magenta, *hFIRRE* exon in green. Scale bar is 5 µm. (*D*) Quantification of nuclear and cytoplasmic FISH signals corresponding to *mFirre* and *hFIRRE* in hybrid cells. Data are mean and non-outlier range of quantified cells.

performed human and mouse *Firre* smRNA FISH and quantified the number of *hFIRRE* and *mFirre* signals in the nucleus and cytoplasm of cells endogenously coexpressing both transcripts. Interestingly, while *mFirre* was exclusively nuclear as observed in other mouse cells, *hFIRRE* localized to both nuclear and cytoplasmic compartments (on average, 61% nuclear and 39% cytoplasmic) (Fig. 2C,D). This shows that in a mouse background, where all *mFirre* transcripts are retained in the nucleus, *hFIRRE* can be exported to the cytoplasm, and adopt its subcellular localization that was prevalent in the human cell lines. Collectively, these results show that the localization of human and mouse *Firre* is dependent on their sequence.

hFIRRE encodes an RRD-enriched, cytoplasmic isoform in human cells

Although the *Firre* locus is poorly conserved at the sequence level in mammals, a commonality across species is the enrichment of the local repeat RRD (Hacisuleyman et al. 2014). In both species, the RRD is enriched in *Firre* exons. While mouse RRDs are highly conserved (on average, 91% to 99% conservation between individual repeats), human RRDs share lower similarity and have higher sequence variability between repeat units (62% to 89%) (Supplemental Fig. S2A,B). These repeats are transcribed into the mature *Firre* lncRNA (Fig. 3A,B). We designed probes specific to human and mouse RRD and costained with exon probes. In mouse ES cells, RRD and exon signal colocalize (on average, 91%), showing that the vast majority of *mFirre* RNA molecules contain RRDs (Fig. 3C). Interestingly, smRNA FISH on other mouse cell types revealed the existence of an isoform that is enriched in RRD (on average, 35% and 25% in 3T3 and C2C12 cells, respectively). In both 3T3 and C2C12 cells, the RRD-enriched isoform is retained in the nucleus. Contrary to mouse, *hFIRRE* is transcribed into an RRD-enriched transcript that is mostly localized in the cytoplasm of human iPS cells (on average, 54% of all cytoplasmic *hFIRRE*) (Fig. 3D). This is also observed in HEK293T and K562 cells, where the RRD-enriched isoform is abundantly expressed in the cytoplasm (on average, 65% and 81%, respectively, of all cytoplasmic *hFIRRE*) (Fig. 3D). Of note, we confirm that the RRD probes specifically detect nuclear *mFirre* as well as nuclear-cytoplasmic *hFIRRE*, thus validating the differential localization of human and mouse *Firre* with an independent FISH probe set (Supplemental Fig. S1A,B).

To understand the transcript isoforms that constitute nuclear and cytoplasmic *hFIRRE*, we analyzed publicly available RNA-seq data for nuclear and cytoplasmic K562 cell fractions. The *hFIRRE* locus is annotated in GENCODE v38 to encode 17 isoforms that are comprised of either zero, four, five, or six RRD-containing exons (Fig. 3E). We quantified reads over the 17 *hFIRRE* isoforms using Salmon (Patro et al. 2017), obtaining abundance estimates for all 17 isoforms in the combined cell fractions (TPM > 0). We found that isoforms containing RRDs display higher expression levels in cytoplasmic fractions than isoforms devoid of RRDs (Fig. 3F). Interestingly, a small isoform (753 bp) is the most abundant *hFIRRE* transcript (Fig. 3F). This small isoform is comprised mostly of RRD-containing exons (87%) and—consistent with our smRNA FISH results—is enriched in the cytoplasm. Finally, we noticed that in the nucleus there are RRD-free isoforms that are more abundant than RRD-containing isoforms (Fig. 3F).

Compared to mouse, the *hFIRRE* locus is enriched in other repeat types, mostly transposons (Fig. 3A,B; Supplemental Fig. S2C). One of those repeats is a full-length, primate-specific HERV-H within intron 10. We thus aimed to determine whether HERV-H is transcribed into a compartment-specific *hFIRRE* isoform. To test this, we designed smRNA FISH probes against HERV-H and analyzed its localization relative to the *hFIRRE* transcription site (colocalization with intron signal) and spliced *hFIRRE* (exon only). Results show that HERV-H RNA is highly expressed and predominantly nuclear, but does not form a part of unspliced or mature *hFIRRE* (on average, 6% of all *hFIRRE* RNA molecules colocalize with HERV-H RNA) (Fig. 3G,H; Supplemental Fig. S2D). Some of the HERV-H RNA is in close proximity to the *hFIRRE* active transcription site, which cannot be explained by random positioning (on average, 0.78 µm for the *hFIRRE* locus, compared to 1.15 µm for the non-HERV-H-containing *hTUG1* locus) (Fig. 3I; Supplemental Fig. S2E). In summary, our data show that the *hFIRRE* locus produces isoforms enriched in the RRD motif that are predominant in the cytoplasm. Furthermore, the *hFIRRE* locus harbors an HERV-H element that is transcribed but not part of the *hFIRRE* transcript.

Firre is more highly expressed in mouse than in human tissues

To determine whether *Firre* expression levels differ globally between human and mouse, we analyzed *Firre* expression in 102 human and 286 mouse RNA-seq data sets across 123 unique tissues. Retrieving RNA-seq expression levels for *Firre* from the EMBL-EBI Expression Atlas (Papatheodorou et al. 2020), we found that *mFirre* is more abundant than *hFIRRE*. This is the case when considering the average expression levels across all tissues (19.6 TPM for *mFirre* and 2.8 TPM for *hFIRRE*) as well as when comparing matching tissues between mouse and human (Fig. 4A,B).

We hypothesized that the observed abundance differences could be driven by cross-species changes in the regulatory potential of *Firre*'s promoter. To test this hypothesis, we aimed to analyze the similarity of TFBS occurrences in *Firre*'s promoter relative to other syntenic lncRNA promoters across the two species. To this end, we identified all syntenically conserved lncRNAs (n = 3705 pairs) as those that are flanked by the same

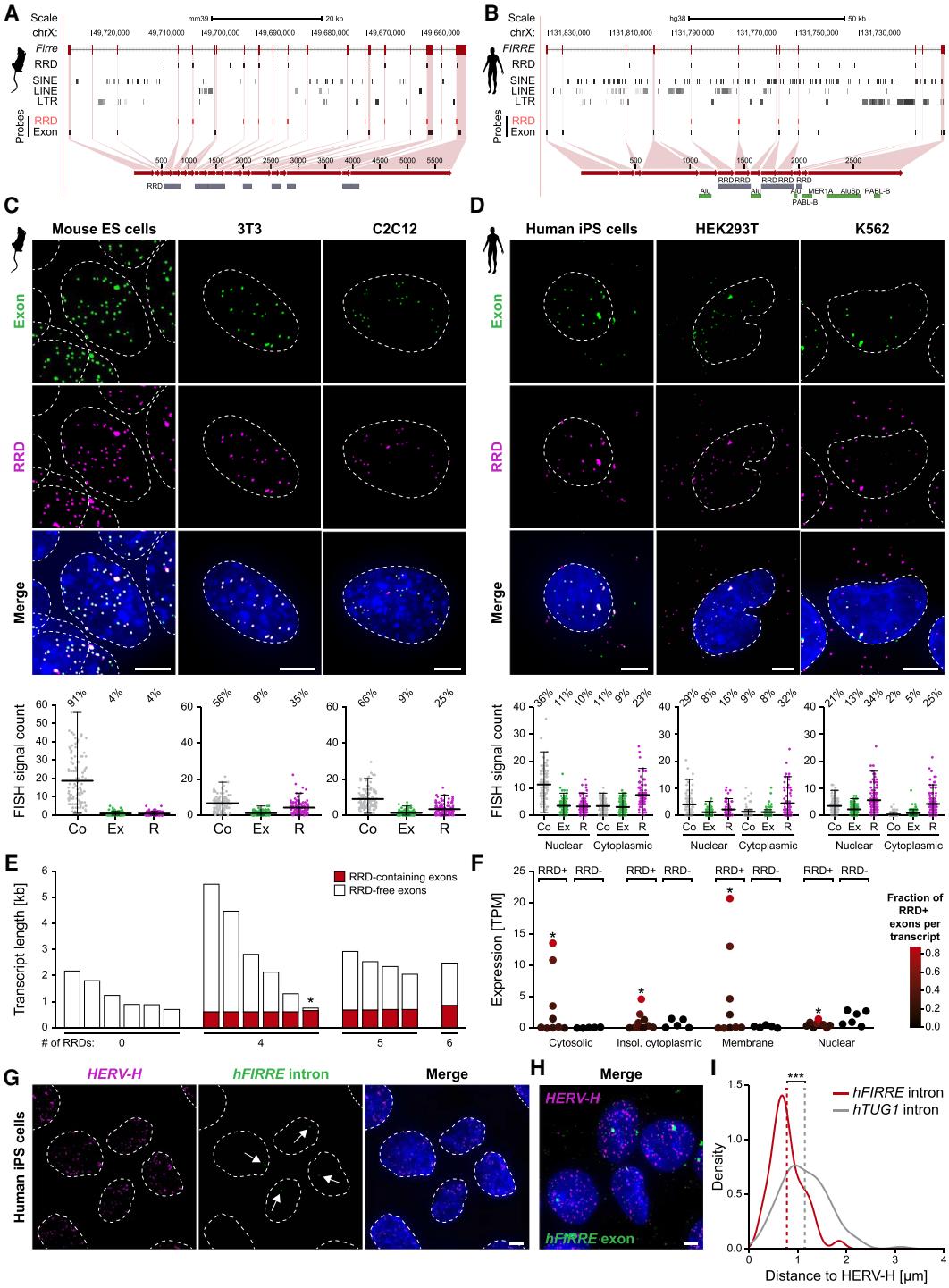

FIGURE 3. The cytoplasmic human *FIRRE* transcript is an RRD-enriched isoform. (*A*,*B*) Schematic showing the *hFIRRE* (*A*) and *mFirre* locus (*B*) as well as the location of local (RRD) and global (SINE, LINE, LTR) repeats. The positions of *Firre* exon and RRD FISH probes are indicated in relation to the gene and its main transcript isoform. (*C*,*D*) Maximum intensity projections of smRNA-FISH for *Firre* exon and RRD on mouse ES cells, 3T3 and C2C12 (*C*) as well as human iPS cells, HEK293T and K562 (*D*). *Below*, quantifications of FISH signals show the number of exon-only (Ex), RRD-only (R), and colocalized exon-RRD spots (Co). Nuclei are stained with Hoechst (blue), *Firre* exon in green, RRD in magenta. Scale bar is 5 µm. Data are mean and non-outlier range of quantified cells. (*E*) Transcript lengths of *hFIRRE* isoforms with different numbers of RRD-containing exons. (*F*) RNA-seq analysis of *hFIRRE* isoform expression levels in different K562 fractions based on their fraction of RRD-free (RRD-) and RRD-containing exons (RRD+). The asterisk denotes the short, RRD-enriched isoform shown in *E*. (*G*,*H*) Maximum intensity projections of smRNA FISH of HERV-H transcripts and *hFIRRE* intron (highlighted by arrows, *G*) or *hFIRRE* exon (*H*) on human iPS cells. Nuclei are stained with Hoechst (blue), HERV-H in magenta, *hFIRRE* intron and exon in green. Scale bar is 5 µm. (*I*) Density plot of the distance between *hFIRRE* and *hTUG1* intron to their closest HERV-H transcript. The dashed line indicates the mean distance. (***) $P = 3.9 \times 10^{-9}$, two-tailed Student's *t*-test.

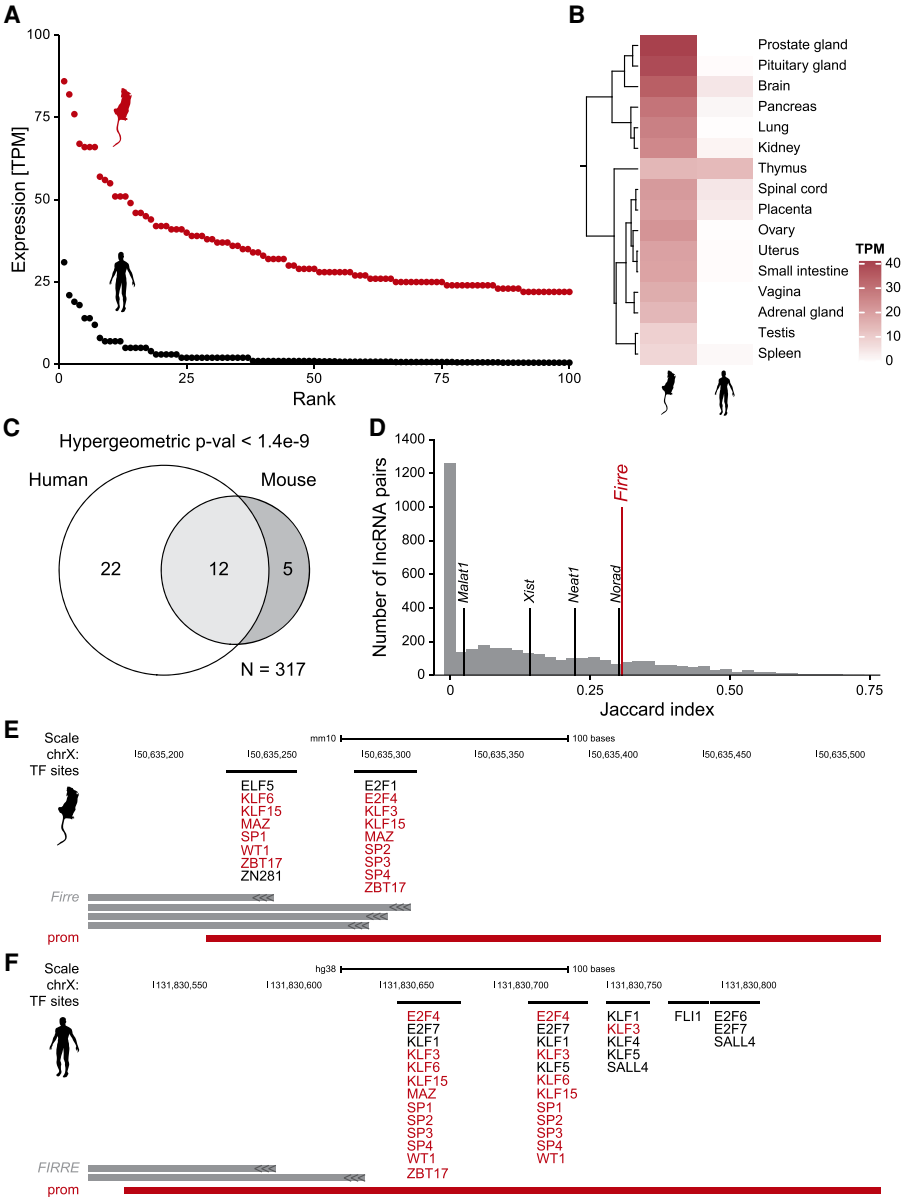

FIGURE 4. *Firre* displays higher expression levels in mouse than in human tissues. (*A*) Ranked expression levels of *Firre* in mouse (red) and human (black) tissues. (*B*) Heat map showing expression of *Firre* in a range of human and mouse tissues. (*C*) Venn diagram indicating common transcription factor binding sites (TFBSs) in human and mouse *Firre* promoters. (*D*) Histogram representing the similarity of TFBSs for all pairs of syntenically conserved lncRNAs in human and mouse. The position of *Firre* (83rd percentile) and other syntenic lncRNA pairs is highlighted. (*E*,*F*) Schematic showing transcription factors with motif instances at each site for mouse (*E*) and human *Firre* promoter regions (*F*). Shared transcription factors are highlighted in red. Note that the *mFirre* promoter is annotated as proximal enhancer in the ENCODE SCREEN cCRE annotation.

orthologous protein coding genes in the human and mouse genomes. We defined promoter regions by overlapping the candidate *cis*-regulatory elements (cCREs) defined by the ENCODE SCREEN project with the GENCODE-defined transcription start sites (TSSs) (Frankish et al. 2019; The ENCODE Project Consortium et al. 2020). For each promoter sequence, we predicted unique TFBS instances via FIMO using motifs from the HOCOMOCO v11 motif database with the subset of orthologous TFs that have known motifs in both human and mouse (n = 317). With a unique set of TFBS instances for all mouse and human syntenic lncRNA promoters at hand, we next computed the extent to which TFBSs were in common between syntenic lncRNA promoter pairs. We observed that while most promoter pairs have zero TFBSs in common (n = 1246; 34% of the pairs), *Firre*'s promoter exhibits a significant overlap in TFBS instances (hypergeometric P-value $<1.4 \times 10^{-9}$; Fig. 4C). More globally, *Firre* exhibits stronger preservation of TFBS instances than most syntenic promoters (83rd percentile; Wilcoxon P-value $< 2.2 \times 10^{-16}$).

Considering the caveat that TSSs are poorly annotated for many lncRNAs, we note that *Firre* has a larger TFBS overlap than *Malat1*, *Xist*, *Neat1*, *Norad*, and other well studied and annotated lncRNAs (Fig. 4D–F). This is surprising considering that the *Firre* promoter has very low primary sequence conservation between human and mouse (46% sequence identity). Curiously, in human the promoter region is annotated as such, yet in mouse the promoter region is annotated as a proximal enhancer (ENCODE SCREEN cCRE annotation; human cCRE: EH38E2768830, mouse cCRE: EM10E0926021). Together, we show that human and mouse *Firre* are expressed at vastly different levels, even though the TFBS instances are largely preserved between human and mouse *Firre* promoters.

DISCUSSION

The molecular function of a lncRNA, or for that matter that of any coding or noncoding RNA, strongly depends on its subcellular localization and expression level. As lncRNA sequences are rapidly evolving, it becomes prudent to consider diverging localizations and abundances even in closely related species when examining lncRNA function.

In this study, we analyzed the localization and expression of the lncRNA *Firre* in mouse and human. With *Firre*, we selected a lncRNA that is conserved at the synteny but not at the sequence level, reasoning that this could be a premise for a potential functional divergence between the investigated species. Indeed, we observed that the localization of *Firre* differs in mouse and human cells, where *mFirre* was highly enriched in the nucleus, and *hFIRRE* was both nuclear and cytoplasmic. Whereas the nuclear functions of *mFirre* have been well characterized (Hacisuleyman et al. 2014; Yang et al. 2015; Darrow et al. 2016; Barutcu et al. 2018; Bonora et al. 2018; Froberg et al. 2018; Fang et al. 2020; Kriz et al. 2021), the molecular functions of *hFIRRE*, especially with respect to its cytoplasmic localization, are unknown. Translational profiling has not identified *hFIRRE* to associate with ribosomes (van Heesch et al. 2014, 2019; Carlevaro-Fita et al. 2016), rendering it unlikely of being translated or regulating the translation of other RNAs. Moreover, there is no evidence of an open reading frame that contains compensatory mutations throughout evolution by PhyloCSF. It therefore remains tempting to speculate about the functions of *hFIRRE* outside of the nucleus, particularly regarding its implications in disease, where it is frequently duplicated or overexpressed (Abe et al. 2014; Ha et al. 2019; Shi et al. 2019; Miolo et al. 2020; Wang et al. 2021).

The differential localization of *mFirre* and *hFIRRE* was maintained in human/mouse hybrid cells. In this cell line, *hFIRRE* is ectopically expressed from a human X-chromosome, while *mFirre* is expressed endogenously from the mouse genome. A fraction of *hFIRRE* translocated to the cytoplasm, but *mFirre* remained in the nucleus. This observation argues for a sequence-intrinsic mechanism of *Firre* localization that is, for instance, different from the one noted for the syntenically conserved lncRNA *FAST* (Guo et al. 2020). In human and mouse stem cells, *FAST* was described to localize differently depending on the expression of the splicing factor PPIE. In human stem cells that express PPIE at low levels, *hFAST* is spliced and exported to the cytoplasm. *mFAST*, on the other hand, is retained in the nucleus of mouse stem cells, because PPIE inhibits *mFAST* splicing and thus prevents its nuclear export (Guo et al. 2020). A *trans*-factor such as PPIE did not influence the human-specific localization of *hFIRRE* in the human/mouse hybrid cells, and *hFIRRE* and *mFirre* localized differently in the same mouse cell background as they would in their native environment.

In search of the sequence determinant for *Firre*'s differential localization, we observed that the cytoplasmic isoforms of *hFIRRE* in human cells were enriched in RRDs. We recently reported on the evolution of local repeats within the *Firre* locus, including the exonic RRD (Hacisuleyman et al. 2016). We observed that a single RRD acts as a nuclear retention signal (Hacisuleyman et al. 2016). However, the existence of cytoplasmic *hFIRRE* isoforms enriched in RRDs in our smRNA FISH experiments highlights the importance of considering the full-length context of an RNA. Whether one or more RRD copies serve as a cytoplasmic localization signal, confer more stability to the transcript in the cytoplasm, or the cytoplasmic *hFIRRE* isoforms lack a nuclear retention signal remains to be determined.

Even though *Firre* expression levels were comparable between mouse and human cell lines, they vastly differed in tissues, with *mFirre* being more abundant than *hFIRRE*. Surprisingly, these abundance differences cannot be readily explained by differences in promoter motifs, as *Firre* has one of the most similar TFBS overlaps across syntenic lncRNA promoters in human and mouse. The relationship between TF binding to the magnitude and context of transcription is not understood well enough to conclude that a high TFBS similarity would lead to a similar expression pattern. Alternative explanations for *Firre*'s different expression patterns between the two species include higher chromatin accessibility at the mouse promoter for TFs to bind, the existence of a stronger enhancer region for *mFirre*, or a higher stability of *mFirre* RNA.

This study provides insights into the functional divergence of the syntenically conserved lncRNA *Firre* by examining its localization and expression in human and mouse. The possibility of a conserved lncRNA adopting distinct characteristics in different species warrants caution when applying findings from mouse models to human physiology. *Firre* thus represents an example that highlights the rapidly evolving nature of lncRNAs that can generate different functional outputs in closely related species.

MATERIALS AND METHODS

Cells

HEK293T, K562, HCT116, BJ, 3T3, C2C12, and RAW264.7 cells were obtained from ATCC and cultured according to ATCC guidelines at 37°C and 5% CO_2. The human/mouse somatic cell hybrid (Coriell, GM10324) was cultured in DMEM (Thermo Fisher Scientific, 11965092) supplemented with 10% FCS (MilliporeSigma, F8067), 1× GlutaMAX supplement (Thermo Fisher Scientific, 35050061), 4×10^{-7} M aminopterin (MilliporeSigma, A3411-10MG), 1×10^{-4} M hypoxanthine (MilliporeSigma, H9636-1G), and 1.6×10^{-5} M thymidine (MilliporeSigma, T1895-1G) at 37°C and 7.5% CO_2. Human iPS cells were cultured on vitronectin (Thermo Fisher Scientific, A14700) in Essential 8 Flex medium with E8 supplement (Thermo Fisher Scientific, A2858501) at 37°C and 5% CO_2. Mouse ES cells were maintained on 0.1% gelatin (MilliporeSigma, ES-006-B) in KnockOut DMEM (Thermo Fisher Scientific, 10829018) supplemented with 12.5% FCS (MilliporeSigma, ES-009-B), 1× GlutaMAX supplement (Thermo Fisher Scientific, 35050061), 1× nonessential amino acids (Thermo Fisher Scientific, 11140050), 100 U/mL penicillin/streptomycin (Thermo Fisher Scientific, 15140122), 100 µM 2-mercaptoethanol (Thermo Fisher Scientific, 31350010), 1 µM PD0325901 (MilliporeSigma, PZ0162-5MG), 3 µM CHIR99021 (MilliporeSigma, SML1046-5MG), and 100 U/mL LIF (MilliporeSigma, ESG1107) at 37°C and 7.5% CO_2. All cells were routinely tested for mycoplasma contamination by PCR.

smRNA FISH

Oligonucleotides targeting human and mouse *Firre* exon, intron, RRD, human TUG1 intron (Dumbovic et al. 2021), and HERV-H were designed with the Stellaris RNA FISH probe designer (LGC Biosearch Technologies, version 4.2), labeled with Quasar 570 or Quasar 670, and produced by LGC Biosearch Technologies. Probe sequences are provided in Supplemental Table S1.

Cells were seeded on glass coverslips coated with 0.01% poly-L-lysine (MilliporeSigma, P4707-50ML), vitronectin (human iPS cells) or gelatin (mouse ES cells). For the suspension cell line K562, all the steps were performed in an Eppendorf tube with brief centrifugation at 300*g* for 5 min between the steps. Cover glasses were washed twice with PBS, fixed in 3.7% formaldehyde in PBS for 10 min at room temperature, and washed again twice with PBS. Cover glasses were immersed in 70% ethanol at 4°C for at least 1 h and then washed with 1 ml of wash buffer A (LGC Biosearch Technologies, SMF-WA1-60) supplemented with 10% deionized formamide (Thermo Fisher Scientific, 4440753) at room temperature for 5 min. Cells were hybridized with 80 µL of hybridization buffer (LGC Biosearch Technologies, SMF-HB1-10) supplemented with 10% deionized formamide containing the FISH probes at a 1:100 dilution (stock 12.5 µM in TE buffer) in a humid chamber at 37°C overnight for up to 16 h. The next day, cells were washed with 1 mL of wash buffer A supplemented with 10% deionized formamide at 37°C for 30 min and stained with wash buffer A supplemented with 10% deionized formamide containing 10 µg/mL Hoechst 33342 (Thermo Fisher Scientific, H3570) at 37°C for 30 min. Cover glasses were washed with 1 mL of wash buffer B (LGC Biosearch Technologies, SMF-WB1-20) at room temperature for 5 min, equilibrated 5 min in base glucose buffer (2× SSC, 0.4% glucose solution, 20 mM Tris pH 8.0 in RNase-free H_2O), and then incubated 5 min in base glucose buffer supplemented with a 1:100 dilution of glucose oxidase (stock 3.7 mg/mL) and catalase (stock 4 mg/mL). Coverslips were mounted with ProLong Glass Antifade Mountant (Thermo Fisher Scientific, P36984) on a glass slide, and left to curate at 4°C overnight before proceeding to image acquisition.

Microscopy and image analysis

Image acquisition was performed using a DeltaVision Elite widefield microscope with an Olympus UPlanSApo 100×/1.40-numerical aperture oil objective lens and a PCO Edge sCMOS camera. *Z*-stacks of 200 nm step size were acquired. Images were deconvolved with the built-in DeltaVision SoftWoRx Imaging software and maximum intensity projections were created using ImageJ/Fiji. FISH spots were quantified manually using ImageJ/Fiji, wherein the brightness and contrast of each channel were adjusted. Imaris image analysis software (Bitplane) was used to calculate the distance between FISH spots.

RNA-seq library preparation

Total RNA from human/mouse hybrid cells was extracted with the Maxwell LEV SimplyRNA Purification kit (Promega, AS1280) with DNase I treatment. RNA quality was assessed with the Bioanalyzer 2100 (Agilent). One microgram of total RNA was subjected to poly(A)-RNA enrichment and library preparation with the KAPA mRNA HyperPrep kit (Roche) or ribosomal RNA depletion with the KAPA RNA HyperPrep with RiboErase kit (Roche), and sequenced on Illumina's NextSeq platform using a 75-bp paired-end strategy.

RNA-seq analysis

RNA-seq expression abundance measurements were retrieved from the EBI Expression Atlas (https://www.ebi.ac.uk/gxa/home; date of retrieval: April 12, 2021) for 102 human experiments and 286 mouse experiments. *Firre* RNA-seq abundance measurements for cell lines that were used for smRNA FISH were retrieved from RefineBio for the experimental accession numbers listed in Supplemental Table S2 (Lee et al. 2020). RNA-seq from K562 subfractions was retrieved from ENCODE (The ENCODE Project Consortium 2012) (accession numbers: ENCSR040YBR, ENCSR038WEK, ENCSR885DVH, ENCSR860DWK, ENCSR696YIB), aligned to hg38 and quantified using GENCODE v38 annotations using nf-core/rnaseq v3.2 with the "star_salmon" aligner (Ewels et al. 2020). RRD transcript analysis was performed by overlapping RRD positions with the *hFIRRE* exons annotated in GENCODE v38 and plotting the transcript abundance estimations calculated by salmon v1.4 (Patro et al. 2017). RNA-seq for the human/mouse hybrid cell line was aligned and quantified with nf-core/rnaseq v1.4.2. The genomic sequence for human chrX and corresponding gene annotations (hg38; GENCODE v38) were appended to the mouse genome and annotation file, respectively (mm39; GENCODE M27), and gene level abundance estimations from Salmon v0.14.1 were plotted for *Firre*.

Promoter analysis

Syntenic pairs of lncRNAs were calculated according to Hezroni et al. (2015). Promoter regions were defined by overlapping the candidate *cis*-regulatory elements (cCREs) defined by the ENCODE SCREEN project with the GENCODE-defined TSSs (versions M25 for mouse and v38 for human). Sequences for these promoter regions were extracted from mm10 and hg38, respectively. Motifs were retrieved from HOCOMOCO Core v11 and filtered to contain only TFs that have an ortholog with a motif in human and mouse ($n = 317$ motifs). TFBS instances were determined using the FIMO tool from the MEME suite (v5.1.1). To find unique instances of TFBS within each promoter, significant motif matches (Q-value <0.05) were de-duplicated by dropping motif matches corresponding to the same TF that largely overlapped (>50% overlap). The extent of TFBS instance overlap between a pair of syntenic lncRNAs was calculated using the multiset Jaccard index/Ruzicka similarity index to account for multiple motif instances within each promoter.

DATA DEPOSITION

RNA-seq files acquired for the hybrid cell line expression analysis have been deposited at Gene Expression Omnibus under accession number GSE186876.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We are grateful to Roy Parker and Carolyn Decker for access to, and training on, the DeltaVision Elite microscope. We acknowledge the support of Theresa Nahreini and Nicole Kethley in the Biochemistry Cell Culture Facility, Joe Dragavon in the BioFrontiers Advanced Light Microscopy Core, and Amber L. Scott in the BioFrontiers Sequencing Core. J.L.R. is a Howard Hughes Medical Institute (HHMI) Faculty Scholar. C.M. was supported by the Deutsche Forschungsgemeinschaft (DFG, MU 4462/1-1). T.H. was supported by a K99 grant from the National Institutes of Health under award number K99GM137072. This research was supported by NIH/NIGMS grant P01 GM099117 to J.L.R.

Author contributions: G.D. and J.L.R. conceived the study. G.D. and C.M. performed cell culture and smRNA FISH. S.D.H., G.D., and C.M. quantified the smRNA FISH experiments. M.J.S. and T.H. conducted the computational analyses. C.M., G.D., and M. J.S. wrote the manuscript with input from all authors.

Received November 29, 2021; accepted March 6, 2022.

REFERENCES

Abe Y, Kikuchi A, Kobayashi S, Wakusawa K, Tanaka S, Inui T, Kunishima S, Kure S, Haginoya K. 2014. Xq26.1-26.2 gain identified on array comparative genomic hybridization in bilateral periventricular nodular heterotopia with overlying polymicrogyria. *Dev Med Child Neurol* **56:** 1221–1224. doi:10.1111/dmcn.12553

Azam S, Hou S, Zhu B, Wang W, Hao T, Bu X, Khan M, Lei H. 2019. Nuclear retention element recruits U1 snRNP components to restrain spliced lncRNAs in the nucleus. *RNA Biol* **16:** 1001–1009. doi:10.1080/15476286.2019.1620061

Barutcu AR, Maass PG, Lewandowski JP, Weiner CL, Rinn JL. 2018. A TAD boundary is preserved upon deletion of the CTCF-rich *Firre* locus. *Nat Commun* **9:** 1444. doi:10.1038/s41467-018-03614-0

Bergmann JH, Li J, Eckersley-Maslin MA, Rigo F, Freier SM, Spector DL. 2015. Regulation of the ESC transcriptome by nuclear long noncoding RNAs. *Genome Res* **25:** 1336–1346. doi:10.1101/gr.189027.114

Bonora G, Deng X, Fang H, Ramani V, Qiu R, Berletch JB, Filippova GN, Duan Z, Shendure J, Noble WS, et al. 2018. Orientation-dependent Dxz4 contacts shape the 3D structure of the inactive X chromosome. *Nat Commun* **9:** 1445. doi:10.1038/s41467-018-03694-y

Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* **25:** 1915–1927. doi:10.1101/gad.17446611

Carlevaro-Fita J, Rahim A, Guigo R, Vardy LA, Johnson R. 2016. Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. *RNA* **22:** 867–882. doi:10.1261/rna.053561.115

Cesana M, Cacchiarelli D, Legnini I, Santini T, Sthandier O, Chinappi M, Tramontano A, Bozzoni I. 2011. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* **147:** 358–369. doi:10.1016/j.cell.2011.09.028

Chen CK, Blanco M, Jackson C, Aznauryan E, Ollikainen N, Surka C, Chow A, Cerase A, McDonel P, Guttman M. 2016. Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. *Science* **354:** 468–472. doi:10.1126/science.aae0047

Darrow EM, Huntley MH, Dudchenko O, Stamenova EK, Durand NC, Sun Z, Huang SC, Sanborn AL, Machol I, Shamim M, et al. 2016. Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. *Proc Natl Acad Sci* **113:** E4504–E4512. doi:10.1073/pnas.1609643113

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, et al. 2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* **22:** 1775–1789. doi:10.1101/gr.132159.111

Diederichs S. 2014. The four dimensions of noncoding RNA conservation. *Trends Genet* **30:** 121–123. doi:10.1016/j.tig.2014.01.004

Dimitrova N, Zamudio JR, Jong RM, Soukup D, Resnick R, Sarma K, Ward AJ, Raj A, Lee JT, Sharp PA, et al. 2014. *LincRNA-p21* activates *p21* in *cis* to promote Polycomb target gene expression and to enforce the G1/S checkpoint. *Mol Cell* **54:** 777–790. doi:10.1016/j.molcel.2014.04.025

Dumbovic G, Braunschweig U, Langner HK, Smallegan M, Biayna J, Hass EP, Jastrzebska K, Blencowe B, Cech TR, Caruthers MH, et al. 2021. Nuclear compartmentalization of TERT mRNA and TUG1 lncRNA is driven by intron retention. *Nat Commun* **12:** 3308. doi:10.1038/s41467-021-23221-w

The ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489:** 57–74. doi:10.1038/nature11247

The ENCODE Project Consortium, Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shoresh N, Adrian J, Kawli T, Davis CA, Dobin A et al. 2020. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583:** 699-710. doi:10.1038/s41586-020-2493-4

Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, Garcia MU, Di Tommaso P, Nahnsen S. 2020. The nf-core framework for community-curated bioinformatics pipelines. *Nat Biotechnol* **38:** 276–278. doi:10.1038/s41587-020-0439-x

Fang H, Bonora G, Lewandowski JP, Thakur J, Filippova GN, Henikoff S, Shendure J, Duan Z, Rinn JL, Deng X, et al. 2020. *Trans*- and *cis*-acting effects of *Firre* on epigenetic features of the inactive X chromosome. *Nat Commun* **11:** 6053. doi:10.1038/s41467-020-19879-3

Frankish A, Diekhans M, Ferreira AM, Johnson R, Jungreis I, Loveland J, Mudge JM, Sisu C, Wright J, Armstrong J, et al. 2019. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res* **47:** D766–D773. doi:10.1093/nar/gky955

Froberg JE, Pinter SF, Kriz AJ, Jegu T, Lee JT. 2018. Megadomains and superloops form dynamically but are dispensable for X-chromosome inactivation and gene escape. *Nat Commun* **9:** 5004. doi:10.1038/s41467-018-07446-w

Groff AF, Sanchez-Gomez DB, Soruco MML, Gerhardinger C, Barutcu AR, Li E, Elcavage L, Plana O, Sanchez LV, Lee JC, et al. 2016. In vivo characterization of Linc-p21 reveals functional *cis*-regulatory DNA elements. *Cell Rep* **16:** 2178–2186. doi:10.1016/j.celrep.2016.07.050

Guo CJ, Ma XK, Xing YH, Zheng CC, Xu YF, Shan L, Zhang J, Wang S, Wang Y, Carmichael GG, et al. 2020. Distinct processing of lncRNAs contributes to non-conserved functions in stem cells. *Cell* **181:** 621–636.e622. doi:10.1016/j.cell.2020.03.006

Ha TK, Mardy AH, Beleford D, Spanier A, Wayman BV, Penon-Portmann M, Wiita AP, Shieh JT. 2019. X-linked duplication copy number variation in a familial overgrowth condition. *Am J Med Genet C Semin Med Genet* **181:** 644–649. doi:10.1002/ajmg.c.31756

Hacisuleyman E, Goff LA, Trapnell C, Williams A, Henao-Mejia J, Sun L, McClanahan P, Hendrickson DG, Sauvageau M, Kelley DR, et al. 2014. Topological organization of multichromosomal regions by the long intergenic noncoding RNA *Firre*. *Nat Struct Mol Biol* **21:** 198–206. doi:10.1038/nsmb.2764

Hacisuleyman E, Shukla CJ, Weiner CL, Rinn JL. 2016. Function and evolution of local repeats in the *Firre* locus. *Nat Commun* **7:** 11021. doi:10.1038/ncomms11021

Hezroni H, Koppstein D, Schwartz MG, Avrutin A, Bartel DP, Ulitsky I. 2015. Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep* **11:** 1110–1122. doi:10.1016/j.celrep.2015.04.023

Kriz AJ, Colognori D, Sunwoo H, Nabet B, Lee JT. 2021. Balancing cohesin eviction and retention prevents aberrant chromosomal interactions, Polycomb-mediated repression, and X-inactivation. *Mol Cell* **81:** 1970–1987.e1979. doi:10.1016/j.molcel.2021.02.031

Lee S, Kopp F, Chang TC, Sataluri A, Chen B, Sivakumar S, Yu H, Xie Y, Mendell JT. 2016. Noncoding RNA NORAD regulates genomic stability by sequestering PUMILIO proteins. *Cell* **164:** 69–80. doi:10.1016/j.cell.2015.12.017

Lee AJ, Park Y, Doing G, Hogan DA, Greene CS. 2020. Correcting for experiment-specific variability in expression compendia can remove underlying signals. *Gigascience* **9:** giaa117. doi:10.1093/gigascience/giaa117

Lewandowski JP, Lee JC, Hwang T, Sunwoo H, Goldstein JM, Groff AF, Chang NP, Mallard W, Williams A, Henao-Meija J, et al. 2019. The *Firre* locus produces a *trans*-acting RNA molecule that functions in hematopoiesis. *Nat Commun* **10:** 5137. doi:10.1038/s41467-019-12970-4

Lubelsky Y, Ulitsky I. 2018. Sequences enriched in Alu repeats drive nuclear localization of long RNAs in human cells. *Nature* **555:** 107–111. doi:10.1038/nature25757

Mele M, Mattioli K, Mallard W, Shechner DM, Gerhardinger C, Rinn JL. 2017. Chromatin environment, transcriptional regulation, and splicing distinguish lincRNAs and mRNAs. *Genome Res* **27:** 27–37. doi:10.1101/gr.214205.116

Miolo G, Bernardini L, Capalbo A, Favia A, Goldoni M, Pivetta B, Tessitori G, Corona G. 2020. Identification of a de novo Xq26.2 microduplication encompassing *FIRRE* gene in a child with intellectual disability. *Diagnostics (Basel)* **10:** 1009. doi:10.3390/diagnostics10121009

Miyagawa R, Tano K, Mizuno R, Nakamura Y, Ijiri K, Rakwal R, Shibato J, Masuo Y, Mayeda A, Hirose T, et al. 2012. Identification of *cis*- and *trans*-acting factors involved in the localization of MALAT-1 noncoding RNA to nuclear speckles. *RNA* **18:** 738–751. doi:10.1261/rna.028639.111

Pang KC, Frith MC, Mattick JS. 2006. Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends Genet* **22:** 1–5. doi:10.1016/j.tig.2005.10.003

Papatheodorou I, Moreno P, Manning J, Fuentes AM, George N, Fexova S, Fonseca NA, Fullgrabe A, Green M, Huang N, et al. 2020. Expression Atlas update: from tissues to single cells. *Nucleic Acids Res* **48:** D77–D83. doi:10.1093/nar/gkaa339

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14:** 417–419. doi:10.1038/nmeth.4197

Pritsker M, Doniger TT, Kramer LC, Westcot SE, Lemischka IR. 2005. Diversification of stem cell molecular repertoire by alternative splicing. *Proc Natl Acad Sci* **102:** 14290–14295. doi:10.1073/pnas.0502132102

Ramalho-Santos M, Yoon S, Matsuzaki Y, Mulligan RC, Melton DA. 2002. "Stemness": transcriptional profiling of embryonic and adult stem cells. *Science* **298:** 597–600. doi:10.1126/science.1072530

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, et al. 2007. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129:** 1311–1323. doi:10.1016/j.cell.2007.05.022

Ross CJ, Rom A, Spinrad A, Gelbard-Solodkin D, Degani N, Ulitsky I. 2021. Uncovering deeply conserved motif combinations in rapidly evolving noncoding sequences. *Genome Biol* **22:** 29. doi:10.1186/s13059-020-02247-1

Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. 2011. A *ceRNA* hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* **146:** 353–358. doi:10.1016/j.cell.2011.07.014

Shi X, Cui Z, Liu X, Wu S, Wu Y, Fang F, Zhao H. 2019. LncRNA FIRRE is activated by MYC and promotes the development of diffuse large B-cell lymphoma via Wnt/ß-catenin signaling pathway. *Biochem Biophys Res Commun* **510:** 594–600. doi:10.1016/j.bbrc.2019.01.105

Shukla CJ, McCorkindale AL, Gerhardinger C, Korthauer KD, Cabili MN, Shechner DM, Irizarry RA, Maass PG, Rinn JL. 2018. High-throughput identification of RNA nuclear enrichment sequences. *EMBO J* **37:** e98452. doi:10.15252/embj.201798452

Smith MA, Gesell T, Stadler PF, Mattick JS. 2013. Widespread purifying selection on RNA structure in mammals. *Nucleic Acids Res* **41:** 8220–8236. doi:10.1093/nar/gkt596

Statello L, Guo CJ, Chen LL, Huarte M. 2021. Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol* **22:** 96–118. doi:10.1038/s41580-020-00315-9

Ulitsky I. 2016. Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nat Rev Genet* **17:** 601–614. doi:10.1038/nrg.2016.85

van Heesch S, van Iterson M, Jacobi J, Boymans S, Essers PB, de Bruijn E, Hao W, MacInnes AW, Cuppen E, Simonis M. 2014. Extensive localization of long noncoding RNAs to the cytosol

and mono- and polyribosomal complexes. *Genome Biol* **15:** R6. doi:10.1186/gb-2014-15-1-r6

van Heesch S, Witte F, Schneider-Lunitz V, Schulz JF, Adami E, Faber AB, Kirchner M, Maatz H, Blachut S, Sandmann CL, et al. 2019. The translational landscape of the human heart. *Cell* **178:** 242–260.e229. doi:10.1016/j.cell.2019.05.010

Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA, et al. 2011. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472:** 120–124. doi:10.1038/nature09819

Wang S, Wang Y, Wang S, Tong H, Tang Z, Wang J, Zhang Y, Ou J, Quan Z. 2021. Long non-coding RNA FIRRE acts as a miR-520a-3p sponge to promote gallbladder cancer progression via mediating YOD1 expression. *Front Genet* **12:** 674653. doi:10.3389/fgene.2021.674653

Wu JQ, Habegger L, Noisa P, Szekely A, Qiu C, Hutchison S, Raha D, Egholm M, Lin H, Weissman S, et al. 2010. Dynamic transcriptomes during neural differentiation of human embryonic stem cells revealed by short, long, and paired-end sequencing. *Proc Natl Acad Sci* **107:** 5254–5259. doi:10.1073/pnas.0914114107

Wu H, Yin QF, Luo Z, Yao RW, Zheng CC, Zhang J, Xiang JF, Yang L, Chen LL. 2016. Unusual processing generates SPA lncRNAs that sequester multiple RNA binding proteins. *Mol Cell* **64:** 534–548. doi:10.1016/j.molcel.2016.10.007

Yang F, Deng X, Ma W, Berletch JB, Rabaia N, Wei G, Moore JM, Filippova GN, Xu J, Liu Y, et al. 2015. The lncRNA *Firre* anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation. *Genome Biol* **16:** 52. doi:10.1186/s13059-015-0618-0

Yao RW, Wang Y, Chen LL. 2019. Cellular functions of long noncoding RNAs. *Nat Cell Biol* **21:** 542–551. doi:10.1038/s41556-019-0311-8

Zhang B, Gunawardane L, Niazi F, Jahanbani F, Chen X, Valadkhan S. 2014. A novel RNA motif mediates the strict nuclear localization of a long noncoding RNA. *Mol Cell Biol* **34:** 2318–2329. doi:10.1128/MCB.01673-13