The lncRNA *Firre* functions as a transcriptional activator from a distance

Michael J. Smallegan[1,2], Christian Much[3], Erika Lasda[4], Jordan Lewandowski[5], John L. Rinn[1,6]

[1]BioFrontiers Institute, University of Colorado Boulder, Boulder, CO, USA
[2]Department of Molecular, Cellular and Developmental Biology, University of Colorado Boulder, Boulder, CO, USA
[3]Department of Biochemistry, Weill Cornell Medicine, New York, NY, USA
[4]Department of Biochemistry and Molecular Genetics, University of Colorado Denver Anschutz Medical Campus, Aurora, CO, USA
[5]Flagship Pioneering, Cambridge, MA, USA
[6]Department of Biochemistry, University of Colorado Boulder, Boulder, CO, USA

***Correspondence: john.rinn@colorado.edu**

Abstract

There are currently over 16,000 long-noncoding RNAs (lncRNAs) annotated in mammalian genomes, a number on par with protein-coding genes. Numerous studies have demonstrated that lncRNAs can influence gene expression programs but lack the temporal resolution to identify the immediate regulatory events. Thus, even for the most well-studied, the early and primary targets of lncRNAs remain elusive. Here we describe a multifaceted effort to determine the primary targets of a lncRNA and the underlying temporal dynamics. Our approach combines several genetically defined loss-of and gain-of-function, rescue, and inducible lncRNA models. We further resolved gene regulatory events temporally (30 min to four days). Applying this approach, we find that the lncRNA *Firre* is an RNA-based transcriptional activator. These transcriptional activation events are not mediated by the epigenetic regulators PRC2, WDR5, or G9A, nor are they due to the proximal binding of the *Firre* lncRNA. Instead, *Firre* activates transcription by increasing chromatin accessibility (approximately 30 min after induction) and activating primary target genes (about an hour later). Overall, this study suggests *Firre* functions as an RNA-based transcriptional activator.

Introduction

Mammalian genomes are replete with novel regulatory elements, including DNA enhancers, newly discovered small proteins, and long noncoding RNAs (lncRNAs) (Abascal et al. 2020; Consortium 2012; Derrien et al. 2012). These findings have inspired efforts to define the targets of these newfound noncoding regulatory elements. For example, there are more than 16,000 lncRNAs in the human genome (Derrien et al. 2012; Cabili et al. 2011; Consortium 2012; Abascal et al. 2020), a number on par with protein-coding genes – yet their primary regulatory sites remain poorly understood. Numerous studies have demonstrated that lncRNAs can influence gene expression programs (Rinn and Chang 2012, 2020; Wang and Chang 2011; Morris and Mattick 2014). However, only a few lncRNAs have been identified to play important roles *in vivo,* and even fewer have been mechanistically characterized (Andergassen and Rinn 2021).

A major limitation in understanding lncRNAs' mechanistic function has been the lack of knowing the direct regulatory targets. Determining primary targets for lncRNAs that function *in cis* has been facile and is usually due to a DNA-based enhancer encoded in the locus or low abundance lncRNA that facilitates promoter activation *in cis* (Tong and Yin 2021; Tan et al. 2017; Groff et al. 2016; Maamar et al. 2013; Dimitrova et al. 2014; Paralkar et al. 2016; Guttman and Rinn 2012; Maass et al. 2012; Elling et al. 2018; Miyagawa et al. 2012; Engreitz et al. 2016; Rinn and Chang 2020; Carpenter et al. 2013). In contrast, *trans*-acting RNA-based gene regulatory events are less well understood (Guttman and Rinn 2012). This raises the question: do lncRNAs act directly to regulate transcription, akin to transcription factors (TFs), or are they affecting other cellular processes resulting in gene regulation from a distance?

Transcription factors bind, in *trans*, to DNA and act by recruiting transcriptional machinery (Lee and Young 2000; Mitchell and Tjian 1989). TF binding in proximity to the transcription start site (TSS) is an orthogonal piece of information that permits the inference of direct action independent of the timing of gene expression changes (Lammers et al. 2020; Wang and Deng 2022; Charoensawan et al. 2015; Stewart-Ornstein et al. 2013; Weikum et al. 2017). We currently lack a parallel mechanistic understanding of lncRNA-based gene regulation.

The time scale at which lncRNAs affect gene expression is also relatively unexplored. Most loss- or gain-of-function studies of lncRNAs measure the transcriptome at a single time point on

the order of multiple hours or days(Guttman et al. 2011; Huarte et al. 2010; Loewer et al. 2010; Guttman and Rinn 2012). Thus, the earliest or primary effects of the lncRNA are obscured by layers of gene regulatory cascades and compensation mechanisms. In contrast, TFs can regulate primary gene targets within minutes. For example, p53 and NFKB *trans*-activate their primary target sites within minutes (Allen et al. 2014; Mahat et al. 2016; Sasse et al. 2019). For understanding TF regulation, the *temporal dynamics* of gene expression changes became essential to disentangling primary (and likely direct) effects from secondary and tertiary effects.

Here we set out to identify the primary targets and underlying temporal dynamics of lncRNA-based gene regulation *in trans*. This effort required choosing a lncRNA that is known to be functional and that functions *in trans*. To this end we chose to first focus on the lncRNA *Firre* (Functional intergenic repeating RNA element), based on the following properties: (i) *Firre* is a *trans*-acting RNA that does not regulate gene expression *in cis*; (ii) gain and loss-of *Firre* RNA function *in vivo* phenotypes ranging from lethality under LPS exposure to defects in hematopoiesis, respectively; (iii) *Firre* is syntenically conserved across vertebrates; (iv) *Firre* is genetically associated with human diseases (Lewandowski et al. 2019; Abe et al. 2014; Hacisuleyman et al. 2016, 2014; Barutcu et al. 2018). Although many studies have attempted to understand the mechanism of *Firre*, these efforts have resulted in several conflicting models. A conundrum that the knowledge of the primary regulatory sites of *Firre* would help to sort out.

To determine the primary regulatory sites of *Firre*, we used a combination of genetically defined mouse embryonic stem cell (mESC) lines. Briefly, we derived *Firre* knockout (*Firre*^{KO}) lines, KO lines with an inducible *Firre* transgene, and a wild-type (WT) line that contains an inducible *Firre* transgene. Importantly, we have a control cell line in each genetic background to account for the effect of doxycycline (dox) – thus controlling for non-specific dox effects, which can be significant.

We monitored gene regulation events by RNA-seq across 16 time points (from 30 min to four days) to identify the temporal order of *Firre*-based regulatory events. We also performed a time course of ATAC-seq to decipher the earliest changes in chromatin accessibility that lead to changes in target gene expression. Finally, we combined this temporal information across genotypes to identify *Firre* targets that occur in all conditions tested and are not due to dox effects. We identified 30 genes commonly regulated in WT and *Firre*^{KO} genetic backgrounds upon induction of *Firre*.

Interestingly, all but one of these genes are upregulated, suggesting that *Firre* has transcriptional activation activity. Using loss-of-function studies, we determined that this gene activation is not due to key epigenetic factors (WDR5, G9a, or PRC2), as previously suggested. We further determined where *Firre* was localized relative to these primary gene-regulation events using RNA Affinity Purification followed by DNA sequencing (RAP-DNA-seq) and single molecule RNA FISH (smRNA FISH). We found that *Firre* RNA does not appear to localize in proximity to these sites, despite the rapid induction. Instead, *Firre* functions to increase chromatin accessibility as measured by ATAC-seq in less than 30 minutes – that in turn leads to mature transcript production approximately 2 hours later. Collectively, our study identifies *Firre* function as an RNA-based transcriptional activator via increasing chromatin accessibility and from a distance.

Results

Firre's gene regulatory effects across static WT and KO samples

We first took the approach of analyzing RNA-seq data from WT and *Firre*^{KO} mice across multiple tissues. We reason that the possible regulatory targets of *Firre* could be commonly misregulated across tissues and cells. To this end, we analyzed nine primary tissue RNA-seq experiments comparing WT and *Firre*^{KO} genotypes (methods) (Fig 1A). We observed a total of 842 genes that were differentially expressed (KO vs WT for a given tissue; $P < 0.05$, |fold-change| >= 1.5) in at least one tissue (Fig 1B). However, there are very few differentially expressed genes in common in multiple tissue contexts (69 genes in two or more tissues and four genes in three or more tissues). The only gene commonly differentially expressed in all nine tissues is *Firre*. Together, these results suggest either that *Firre* regulation may be tissue-specific or that the heterogeneity of tissues confounds identifying regulatory targets of *Firre*.

To control for cellular heterogeneity in tissues, we derived mESCs from WT and *Firre*^{KO} mice (methods). After establishing primary WT and *Firre*^{KO} mESC lines, we performed RNA-seq on six biological replicates for WT and *Firre*^{KO}. We found 921 genes that were differentially expressed (314 downregulated and 607 upregulated, $P < 0.05$, |fold-change| >= 1.5; Fig. 1C). We note that this number of possible regulatory targets far exceeds the number of molecules of *Firre* in mESCs approximately 30 molecules per cell(Much et al. 2022). Due to compensatory

mechanisms and the stoichiometry of Firre, these expression changes are not likely to be due to the direct action of Firre.

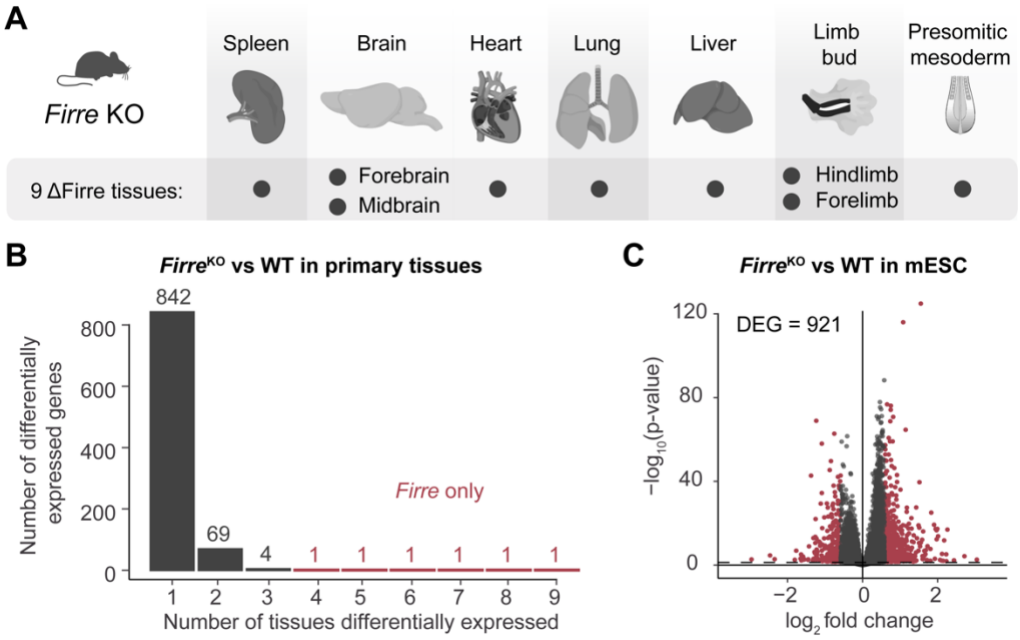


Figure 1: Firre knockout gene expression changes *in vivo* and in mESCs. (A) Schematic of tissues examined by RNA-seq in a *Firre* knockout mouse model. Each black dot indicates if one or two independent studies examined the represented tissue. **(B)** Bar plot showing the number of genes differentially regulated in a *Firre* knockout per tissue (number above each bar). **(C)** Volcano plot showing a total of 921 genes significantly differentially expressed in *Firre* knockout mESC lines relative to WT background. *Firre* not shown: $\log_2$ fold-change -6.9, $-\log_{10}$(p-value) = 196.6.

Temporal gene regulation by *Firre* in a rescue model

To dispense with the confounding effects of compensation, we need to be able to control the expression of *Firre* and measure the dynamics of gene expression changes before the cell returns to homeostasis. However, the rate at which lncRNAs mediate gene expression changes is unknown. Therefore, we started by measuring the transcriptome at widely spaced time intervals up to four days after the induction of *Firre*.

To temporally resolve gene regulatory effects by *Firre*, we generated *Firre*^{KO} mESCs that also contain a dox-inducible *Firre* transgene; we term this line "*Firre*^{RESCUE}" (Fig. 2A; Supplemental

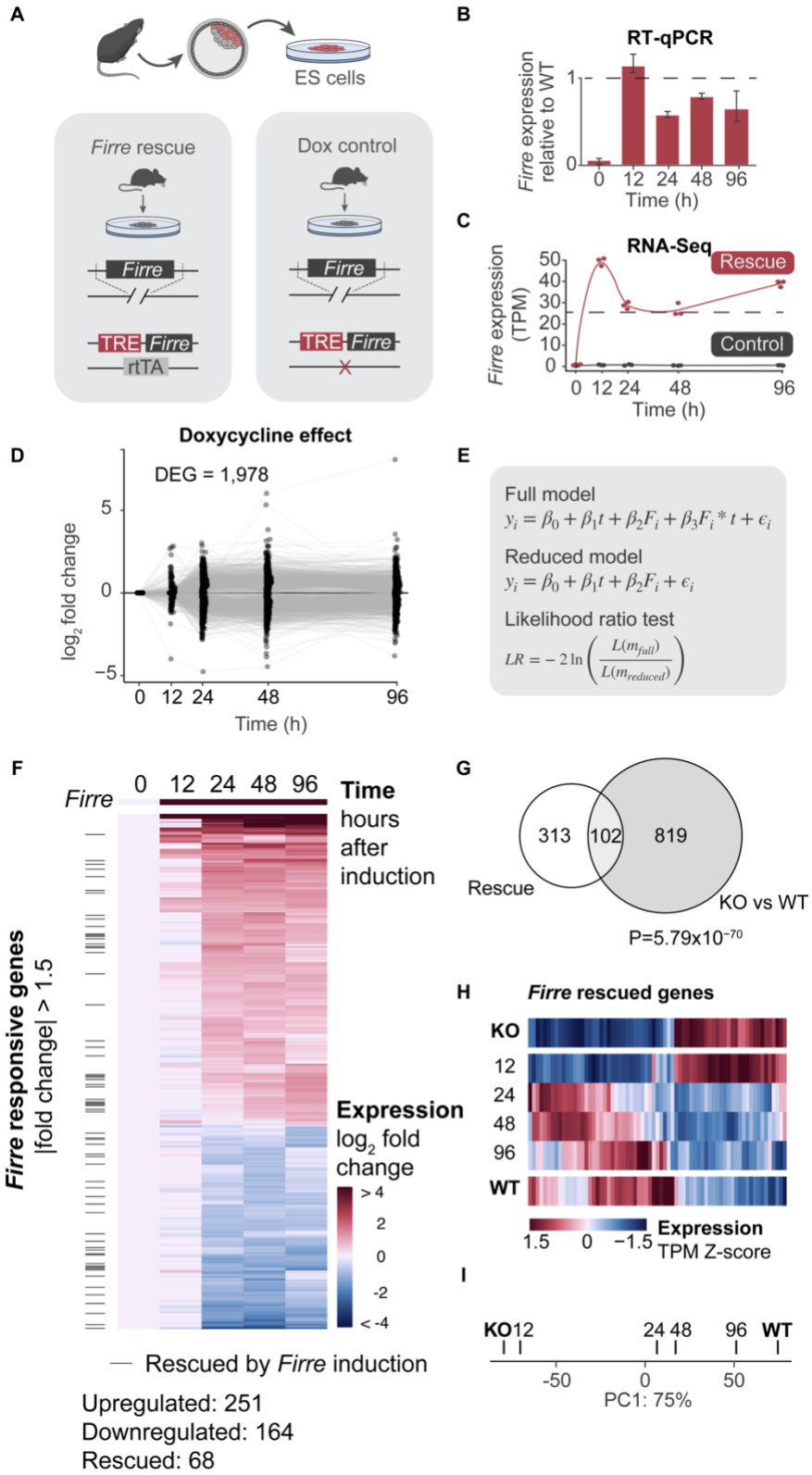
Fig. S1A). Since dox itself causes gene-expression changes, we derived a “dox control” line (Fig. 2A). This cell line is genetically identical to the *Firre*^{RESCUE} line but the reverse tetracycline-controlled transactivator (rtTA) element is present but not expressed (Supplemental Fig. S1B). Thus, the *Firre* transgene is not inducible in the dox control line.

We first characterized the *Firre* transgene in the *Firre*^{RESCUE} and dox control lines using a time course of 0, 12, 24, 48, and 96 hours. Specifically, we treated both rescue and control lines with dox for four days, monitoring samples at each time point by quantitative reverse-transcription PCR (RT-qPCR) with primers targeting *Firre*. In the *Firre*^{RESCUE} line, *Firre* expression is induced to WT levels by 12 hours and remains expressed through four days (Fig. 2B). We did not observe any expression of transgenic *Firre* in the dox control line.

To identify temporally regulated genes upon the introduction of *Firre*, we generated a transcriptome-wide time course in the *Firre*^{RESCUE} and dox control cell lines. We treated cells with dox for 12, 24, 48, and 96 hours and performed RNA-seq with three biological replicates at each timepoint. We observed that *Firre*’s expression measured by RNA-seq follows the induction pattern detected by RT-qPCR (Fig. 2C; Supplemental Fig. S1C; Supplemental Fig. S2A).

Next, we characterized the dox gene-regulatory effect in the dox control line. We determined differentially expressed genes due to dox by comparing each gene’s expression at a given time point to the 0-time point. Surprisingly, treatment with dox dynamically regulated 1,978 genes ($P < 0.05$; |fold-change| >= 1.5; Fig. 2D), indicating that it is essential to control for the gene-regulatory effects of dox.

To control for the effect of dox and identify genes regulated by *Firre*, we used a likelihood ratio test between two regression models (Fig. 2E). The first model contains terms for time (t), the baseline difference between cell lines (F_i), and the interaction of the two ($t*F_i$). The second model lacks the interaction term. Genes regulated similarly in the *Firre*^{RESCUE} and control lines fit well in both models. However, if the interaction term has an explanatory value, the likelihood ratio test will have a small p-value. Using this approach, we found 414 genes to be significantly


A
ES cells
Firre rescue
Dox control
Firre
TRE Firre
rtTA
B
RT-qPCR
Firre expression relative to WT
Time (h)
C
RNA-Seq
Rescue
Control
Firre expression (TPM)
Time (h)
D
Doxycycline effect
DEG = 1,978
log2 fold change
Time (h)
E
Full model
$y_i = \beta_0 + \beta_1 t + \beta_2 F_i + \beta_3 F_i * t + \epsilon_i$
Reduced model
$y_i = \beta_0 + \beta_1 t + \beta_2 F_i + \epsilon_i$
Likelihood ratio test
$LR = -2\ln\left(\frac{L(m_{full})}{L(m_{reduced})}\right)$
F
Firre
Time hours after induction
Firre responsive genes |fold change| > 1.5
Expression log2 fold change
Rescued by Firre induction
Upregulated: 251
Downregulated: 164
Rescued: 68
G
313
102
819
Rescue
KO vs WT
P=5.79x10-70
H
Firre rescued genes
KO
12
24
48
96
WT
Expression TPM Z-score
I
KO 12 24 48 96 WT
PC1: 75%

186

187

Figure 2: Temporal gene regulation in a *Firre* transgene rescue model. (A) Schematic showing the *Firre* knockout mESC line containing a dox-inducible *Firre* transgene and dox-control model that does not express rtTA. **(B)** Measuring induction of *Firre* transgene in rescue cells by RT-qPCR. **(C)** Expression levels of *Firre* in rescue (red) and dox-control mESCs (black) as measured by RNA-seq. **(D)** $\log_2$ fold change of significantly differentially expressed genes that are affected by dox treatment in the dox control cell line. Fold-changes relative to zero timepoint. **(E)** Linear model to derive genes that are significantly differentially regulated upon *Firre* transgene induction and not due to dox treatment (likelihood ratio test). **(F)** Heatmap of significantly differentially expressed genes upon *Firre* induction and not dox treatment. Shades of red and blue represent $\log_2$ fold changes relative to the zero-time point of up- and down-regulation, respectively, upon *Firre* transgene induction. Black tick marks represent genes that were down-regulated in then knockout and up-regulated upon transgene induction, or "rescued" gene expression. **(G)** Venn diagram of number of significantly differentially expressed genes (overlap) in WT vs KO and KO vs rescue cells. **(H)** Heatmap of the TPM of 68 rescued genes across each time point. Top and bottom heatmap represent the KO and WT expression levels, respectively. The heatmaps in between represent the temporal timing of when genes are rescued. **(I)** Principal component analysis of the primary component (PC1 that represent 75% of the variation in the data) at each time point.

regulated by *Firre* across time ($P < 0.05$; |fold-change| >= 1.5; Fig. 2F; Supplemental Fig. S2E-F).

To assess the extent to which the reintroduction of *Firre* in a *Firre*^{KO} background restored the WT expression pattern, we identified genes that were significantly differentially expressed in response to *Firre*'s induction and differentially expressed in the static *Firre*^{KO} vs WT analysis. The genes responding to *Firre*'s induction significantly overlapped the differentially expressed genes in the *Firre*^{KO} (102 genes; hypergeometric $P=5.79\times10^{-70}$; Fig. 2G). The overall pattern of expression changes indicated a reciprocal relationship (Pearson correlation = -0.81; $P<2.2\times10^{-16}$; Supplemental Fig. S2G). We considered genes reciprocally regulated to within 20% of the *Firre*^{KO} vs. WT fold-change to be rescued (68 genes; Fig. 2H; Supplemental Fig. S2H-I). We next examined the temporal dynamics underlying how gene-expression programs return to WT levels (Fig. 2H). Notably, there appears to be a transition between 12 and 24 hours. To test this, we performed a principal component analysis on the expression values of the rescued genes.

The principal component values showed a high similarity in cell lines not exposed to *Firre* and the 12-hour time point. We observed a shift towards rescued gene-expression levels at 24 hours. Interestingly, after this larger shift at 24 hours, the remaining changes are more gradual towards the WT state (Fig. 2H-I).

Gene regulation by *Firre* before 12 hours of induction in rescue model

The time course above established that gene expression changes occur by 12 hours, however, expression changes in a shorter time window would give more confidence that these are direct effects of *Firre*. As a first test, we measured the levels of *Firre* induction prior to 12 hours by RT-qPCR at 0, 3, 6, and 12 hours. Interestingly, *Firre* reached maximal expression at three hours (Fig. 3A). Based on the early induction of *Firre* at three hours we performed RNA-seq at a much more fine-grained temporal resolution prior to 12 hours. Specifically, we sampled every 30 minutes until 5.5 hours post addition of dox (Fig. 3B). RNA was isolated at each time point in duplicate and subjected to RNA-seq. We observed induction of *Firre* expression within 30 minutes by RNA-seq (Fig. 3C), with a maximal induction around two hours (18.5-fold above WT levels; Supplemental Fig. S3A). We also observed that mESCs have minimal expression changes resulting from dox exposure in this short time window relative to the response at 12 hours (12 differentially expressed genes; $P < 0.05$; |fold-change| >= 1.5; Fig. 3D).

We next analyzed this short time course to home in on early *Firre* regulatory events. We used the same likelihood ratio test approach (Fig. 2E) that determines genes significantly regulated by *Firre* induction across time while controlling for dox-responsive genes. We identified 18 genes (12 upregulated and six downregulated) that respond to *Firre* (Fig. 3E; Supplemental Fig. S3B-C). We note that over the 5.5-hour time course, the upregulated genes continue their trajectory to higher expression levels, while the downregulated genes transiently decrease expression and then return to their previous expression levels (Fig. 3E; Supplemental Fig. 3C-D). Together these results suggest that *Firre*'s rapid expression mediates robust upregulation of some genes and transient downregulation of others.

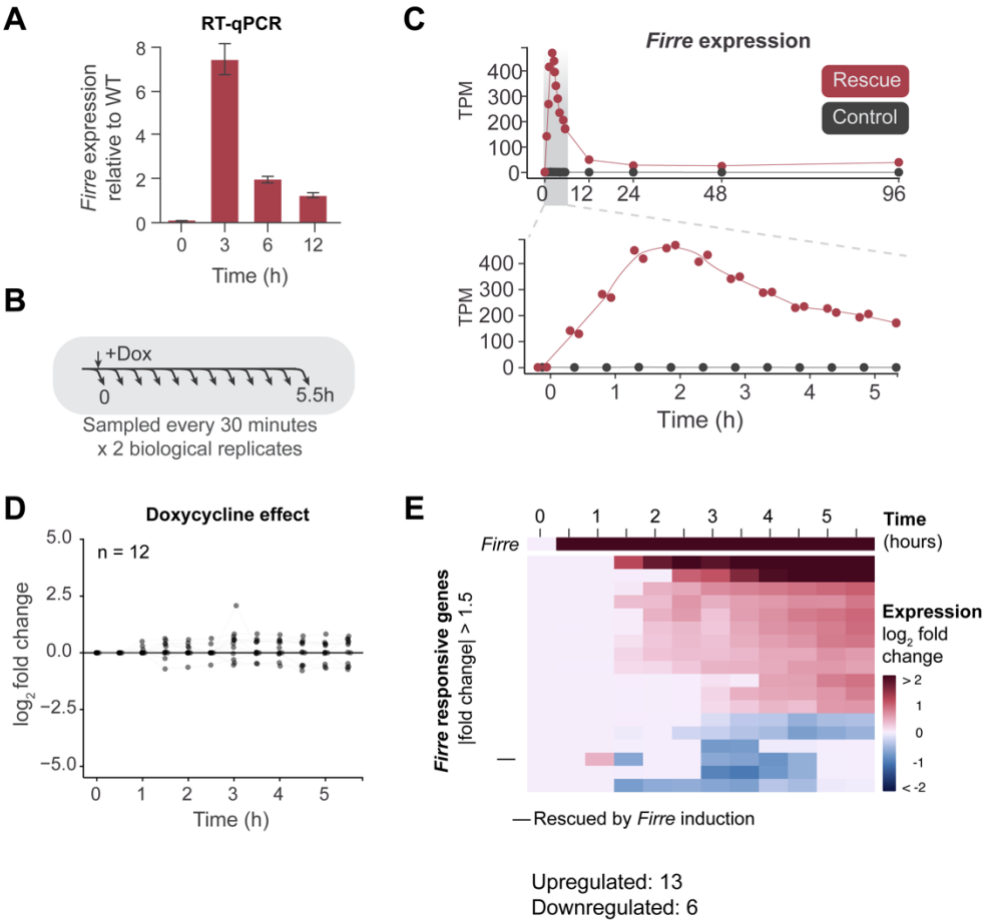


Figure 3: Early temporal regulation in a *Firre* transgene rescue model. (A) Schematic showing sampling of cells that were treated with dox for 0 to 5.5 hours at intervals of 30 min. **(B)** Gene expression levels of *Firre* transgene expression over time as measured by RT-qPCR. *C*t values were normalized against levels of *Gapdh* and *ActB*. Data are mean range of two biological replicates. **(C)** Measurement of *Firre* expression levels in *Firre*^{RESCUE} (red) and dox control (black) cell lines over time as measured by RNA-seq. Bottom, zoom in of *Firre* expression at early time points. **(D)** $\log_2$ fold change of significantly differentially expressed genes that are affected by dox treatment in the dox control cell line. Fold-changes relative to zero timepoint. **(E)** Heatmap of genes differentially expressed due to *Firre* transgene induction and not dox treatment at each early time point.

Gene regulation by *Firre* in an overexpression model

We hypothesized that the robust primary targets of *Firre* would be affected similarly in a WT background with a *Firre* transgene. To test this, we generated a *Firre* WT mESC line with the

same dox-inducible *Firre* transgene (*Firre* overexpression line / *Firre*^{OE}). A WT background served as a dox control (Fig. 4A). We first used RT-qPCR to characterize the induction of the *Firre* transgene in the WT and dox control backgrounds. We found that the *Firre* transgene does not induce in the dox control line. In contrast, in the *Firre*^{OE} line, we observed a rapid induction of the *Firre* transgene with a maximum expression at three hours and taper around 6 hours (Fig. 4B).

To mirror our previous experiments in the rescue model, we generated a long (96-hour; three biological replicates) and a short (5.5-hour; two biological replicates) RNA-seq time course (Fig. 4C). By RNA-seq, we observed a maximal induction of *Firre* (46.2-fold WT levels) at two hours, and *Firre* expression levels returned to 4.5-fold WT levels by 12 hours (Fig. 4D-E; Supplemental Fig. S4A-C). Overall, these are very similar patterns of temporal induction of *Firre* in the rescue model that also had a maximum induction of *Firre* at two hours and returned to near WT levels at 12 hours. We next analyzed our RNA-seq data for dox-related false-positive response genes. As we did previously, we characterized the effects of dox alone by comparing gene expression to time zero. We see 63 genes changing within the 5.5-hour window and 762 genes changing in the longer time course (Supplemental Fig. S4D-E).

To identify *Firre*-responsive genes, we used the same regression models and likelihood ratio test described above (Fig. 2E), which identifies genes temporally regulated by *Firre* but not due to dox effects. In the 96-hour time course, we identified 176 differentially expressed genes ($P < 0.05$; |fold-change| > 1.5; Fig. 4F; Supplemental Fig. S4F-G) that exhibit coherent regulation by *Firre* across time. We note that this is a smaller number of regulated genes than in the rescue time course. Nonetheless, a significant fraction of these genes are also regulated in the *Firre*^{RESCUE} line (hypergeometric $P=2.2\times10^{-107}$; Fig. 4G). We next used the same analyses on our early time course of monitoring every 30 minutes for 5.5 hours. We identified 26 (25 upregulated and one downregulated) potential early *Firre* regulatory targets (Fig. 4H; Supplemental Fig. S4H-J). We also observe a significant (hypergeometric $P=4.21\times10^{-42}$; Fig. 4I) overlap between immediate *Firre* responding genes in rescue and overexpression models. Collectively, we observe good correspondence in genes that are regulated in both the shorter and longer time course across overexpression and rescue models.

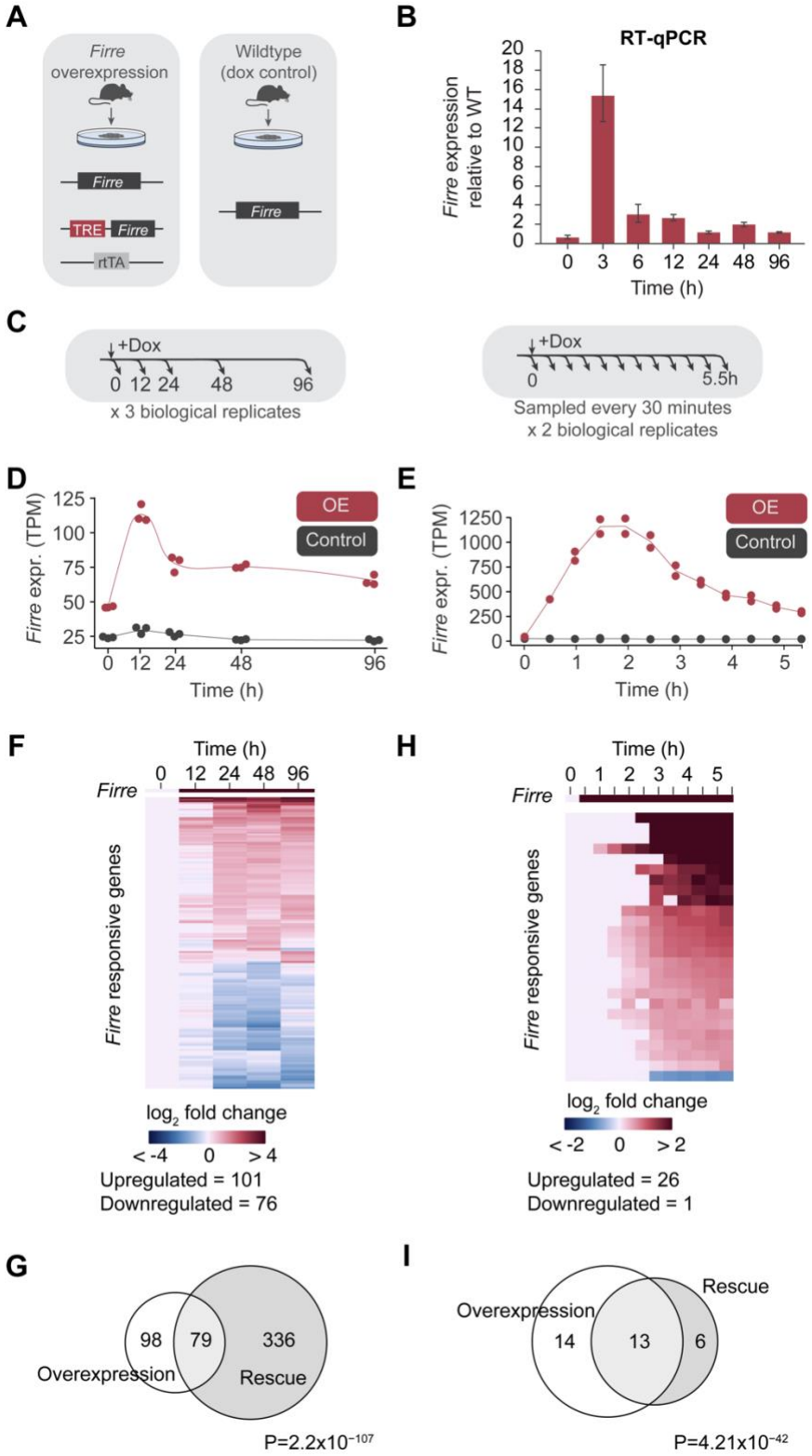


Figure 4: Temporal regulation of *Firre* transgene in a WT background. (A) Schematic of *Firre* transgene in a WT background and dox control of *Firre* transgene in WT background. **(B)** Schematic of time points sampled in WT plus *Firre* transgene model **(C)** RNA-seq measurement

of *Firre* transgene expression (red) and in dox control (black) over the long time course of every 12 hours. (**D)** Measurement of *Firre* transgene (red) in WT background and in dox control (black). **(E)** Heatmap of genes that are significantly differentially expressed upon exposure to *Firre* transgene and not dox treatment over the long time course. **(F)** Venn diagram of genes in the "rescue" and WT background models in the long time course. **(G)** Heatmap of genes significantly differentially expressed upon exposure to the *Firre* transgene across shorter time course. **(H)** Venn diagram representing the overlap of *Firre* response genes in WT and KO backgrounds in the short time course.

Firre **regulatory targets across time and genotype**

We next wanted to determine the genes that respond to *Firre* in both the rescue and the overexpression model and not due to dox in either condition. Moreover, we combined the data from the short and longer time courses: 156 RNA-seq experiments across 16-time points and four distinct genotypes (Fig. 5A). We used the same likelihood ratio test as described previously and added a term describing the genetic background (F_{KO}) (Fig. 5B).

Using this model, we identified 29 significantly differentially expressed ($P < 0.05$, |fold-change| > 1.5) early *Firre*-responsive genes (Fig. 5C; Supplemental Fig. S5B-C). All but one of these genes was upregulated. These genes started increasing expression in 1.5-4.5 hours, and 22/30 genes remained at elevated levels through 96 hours (Fig. 5C-D; Supplemental Fig. S5A-B). One gene, *Prdm14*, decreased expression. Unlike the transient downregulation in the *Firre*^{RESCUE} line, *Prdm14* maintained its lower expression through 96 hours. Since these 30 genes are rapidly regulated concordantly in two genetic backgrounds in response to *Firre*, they represent high confidence candidates to be *Firre* target genes.

We next wanted to independently verify an early *Firre* responsive gene by smRNA FISH. We chose *Shf* and designed a complementary probe set to monitor its abundance and localization across time (methods). We tested 0, 2, 4, and 6-hour time points after induction with dox in both *Firre*^{OE} and rescue lines. In both cell lines, we observed a continuous increase in the expression of *Shf* after the induction of *Firre*, verifying our RNA-seq results (Supplemental Fig. S6). In summary, we identified a cohort of high-confidence *Firre* target genes that are predominantly upregulated upon *Firre* induction.

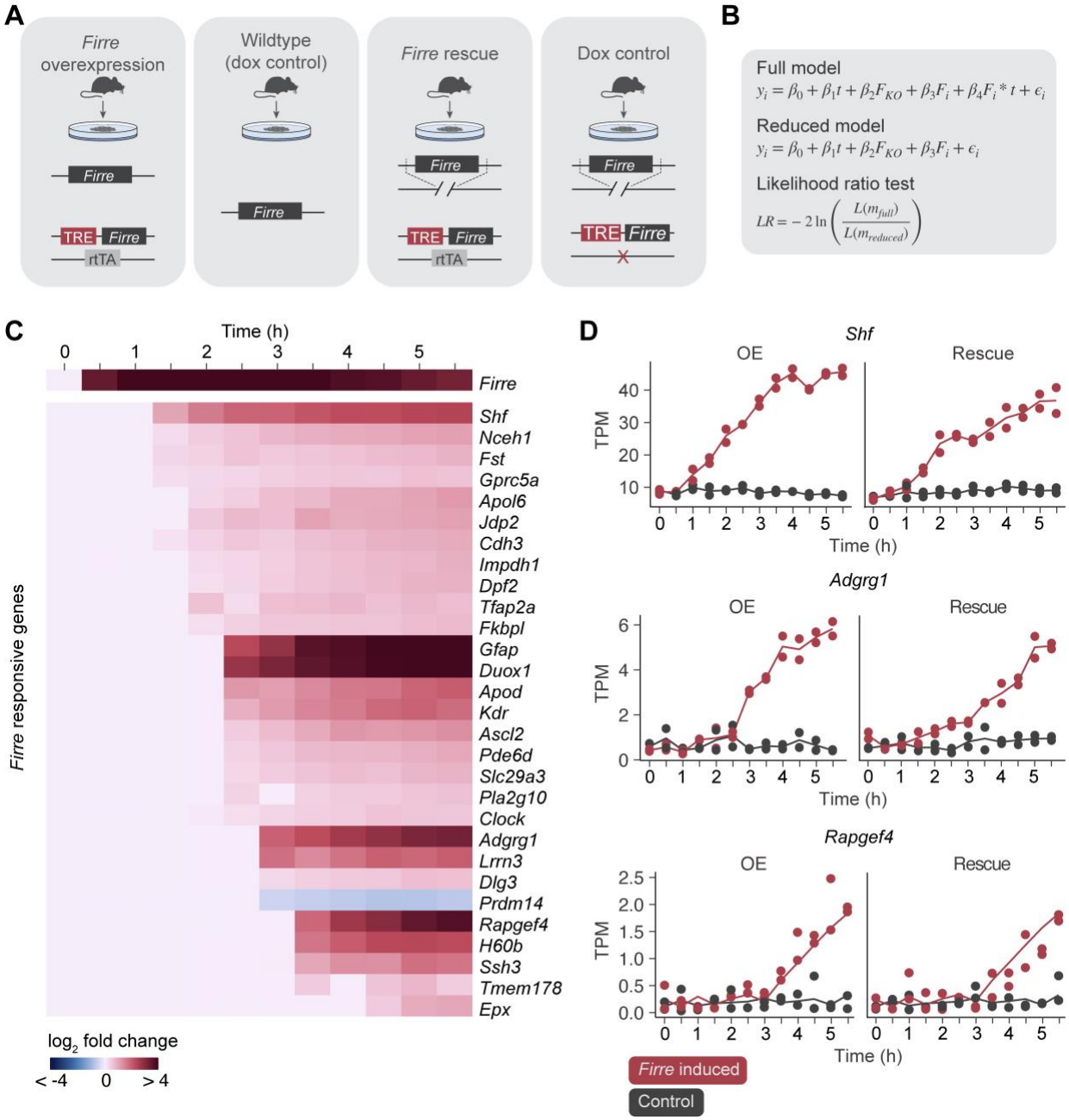


Figure 5: *Firre* transgene targets across genotype and time. (A) Schematic of all Firre mESC lines tested. **(B)** Linear model to determine genes that are regulated by *Firre* transgene independent of dox and across WT and KO backgrounds. **(C)** Heatmap of gene expression changes of Firre-responsive genes over time as determined by the linear model in B. **(D)** Gene expression changes of representative genes due to *Firre* induction and not dox treatment.

Epigenetic and proximal influence of *Firre* trans-activation

Although many lncRNAs have been identified to bind to epigenetic regulatory factors, how these interactions affect gene expression remains poorly understood. One hypothesis is that *Firre* could bind to epigenetic complexes that regulate histone modifications and alter these modifications proximal to regulatory sites. To test this hypothesis, we designed a strategy to examine the influence of the Polycomb repressive complex 2 (PRC2), *G9a* (EHMT2), and WDR5 that influence histone H3 lysine 27 trimethylation (H3K27me3), H3 lysine 9 trimethylation (H3K9me3), and H3 lysine 4 trimethylation (H3K4me3) states, respectively.

Specifically, we used GapmeRs to knockdown the PRC2 components *Ezh2*, *Suz12*, *Jarid2*, and *Aebp2* as well as *G9a* and *Wdr5* in *Firre*^{OE} cells. After 24 hours of GapmeR exposure, the *Firre* transgene was induced by dox for six hours (Fig. 6A). We reasoned that if any of the three epigenetic complexes are required for *Firre*-mediated gene expression regulation, their knockdown would abolish the upregulation of the *Firre*-responsive genes *Adgrg1* and *Shf* (Fig. 6B).

We validated the effective depletion of PRC2, G9a, and WDR5 complexes at the RNA and protein level (Supplemental Fig. S6A). Using a non-targeting control (NTC) GapmeR confirmed the induction of *Firre* and the concomitant upregulation of the *Firre* responders *Adgrg1* and *Shf* after the addition of dox, while the non-*Firre* target *Tug1* remained unchanged (Fig. 6C and Supplemental Fig. S6B). Surprisingly, we found that the depletion of any of these three epigenetic regulatory complexes did not decrease the upregulation of the tested *Firre*-responsive genes (Fig. 6C). Collectively, these results demonstrate that *Firre*-mediated activation of target genes is not dependent on the epigenetic complexes PRC2, G9a, and WDR5.

To probe for another possible mechanism of how *Firre* regulates its targets, we next determined if the *Firre* RNA is in proximity to its target genes. We first leveraged previously published RNA affinity purification (RAP) data that identified 34 *Firre*-bound genomic locations in mESCs (Hacisuleyman et al. 2014) where *Firre* was bound and searched for overlaps with *Firre* responder genes. The 34 RAP peaks did not overlap within 50 kb of the gene bodies of any of the *Firre* responder genes identified in any experiment (520 genes). Thus, under homeostatic

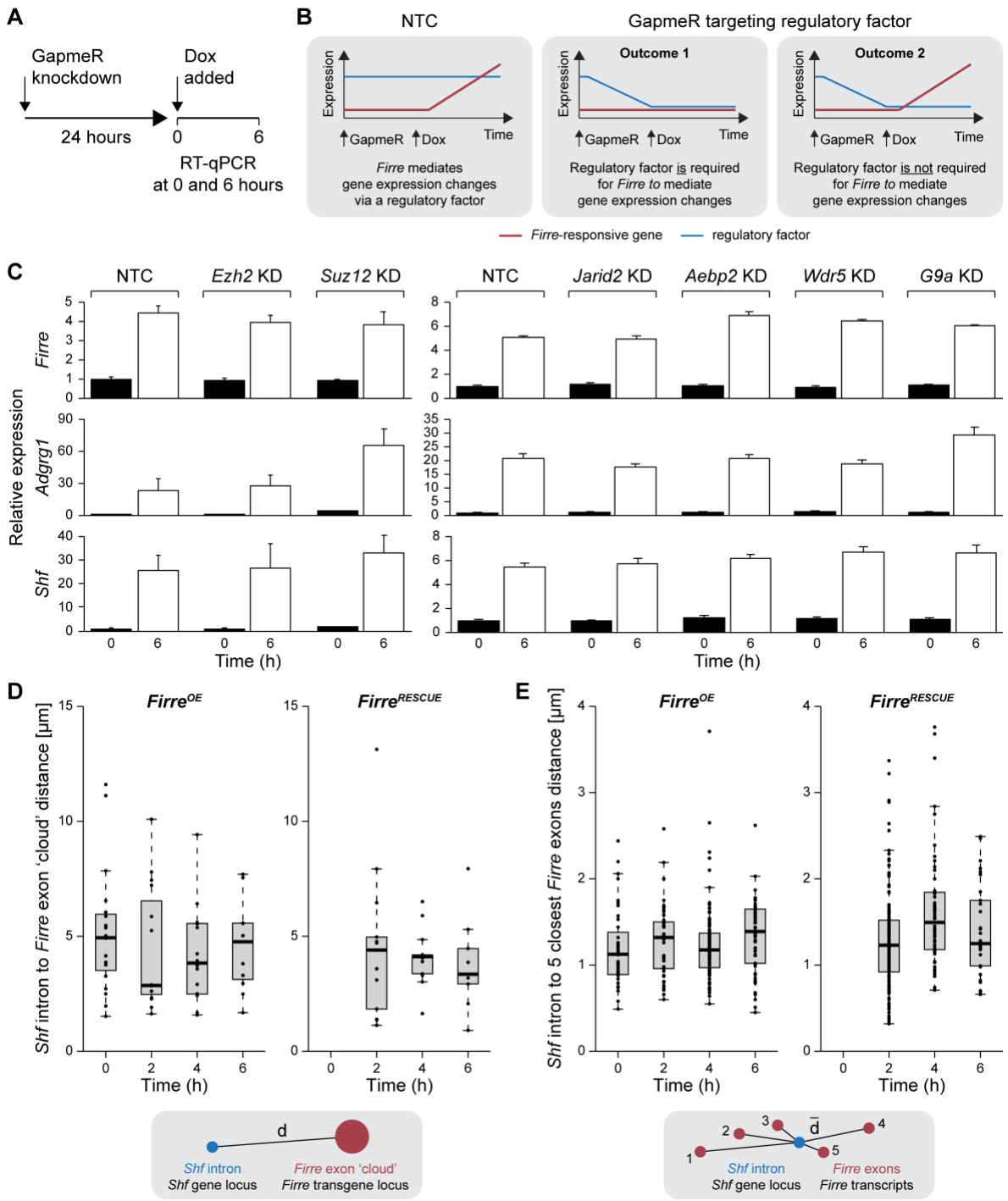


Figure 6: Probing epigenetic and proximal mechanisms of *Firre*-mediated gene regulation. (A) Timeline of GapmeR-mediated gene knockdown and dox-induced *Firre* transgene expression. (B) Depiction of possible outcomes for *Firre*-responsive gene expression (red line) after knockdown of a regulatory factor (blue line) and induction of *Firre* transgene expression. **(C)** Relative gene expression of *Firre* and its targets *Adgrg1* and *Shf* in *Firre*^{OE} cells transfected with non-targeting control- (NTC), *Ezh2*-, *Suz12, Jarid2*-, *Aebp2*-, *Wdr5*-, or *G9a*-

targeting GapmeRs, and treated without (zero hours) or with (six hours) dox as determined by RT-qPCR. **(D** and **E)** smRNA FISH-based quantification of the distance between *Shf* intron and *Firre* transgene (D) and *Shf* intron and the five closest *Firre* exons (distance averaged) (E) in *Firre*-overexpression and *Firre*-rescue cells after 0, 2, 4, and 6 hours of dox treatment. The images below depict the rationale of the distance measurements.

conditions, *Firre* does not appear to be in proximity to genes with early or late expression changes.

As a second approach, we aimed at determining the spatial relationship between *Firre* and its target genes by smRNA FISH. To this end, we designed differentially labeled probe sets complementary to *Firre* exons as well as either the exons or the first intron of the early response gene *Shf*. We performed smRNA FISH at 0, 2, 4, and 6 hours post-Firre induction in the $Firre^{OE}$ and $Firre^{RESCUE}$ cell lines and confirmed the upregulation of *Shf* RNA by *Firre* (Supplemental Fig. S6C-D).

We then measured the distance of the site of transcription of *Shf*, as indicated by its intron signal, from the site of transcription of *Firre*, as approximated by the typical *Firre* exon-stained "expression cloud" amongst the many individual *Firre* molecules in the nucleoplasm. We found that the *Firre* transgene was not in close proximity to the *Shf* gene nor increased in proximity upon induction of *Firre*, indicating no direct interaction of the *Firre* and *Shf* loci (Fig. 6D). We next determined if individual *Firre* RNA molecules were in proximity to the *Shf* locus. We found that *Firre* exons on average were more than one micron away from the *Shf* locus and remained so despite increased expression of *Firre* and *Shf* (Fig. 6E, Supplemental Fig. S6C-D). Thus, both RAP and smRNA FISH demonstrate that *Firre* does not form direct contacts with its most immediate and primary gene targets.

Firre **regulates chromatin accessibility proximal to regulatory targets**

Another possibility for how *Firre* could be activating genes is by increasing chromatin accessibility in *cis*-regulatory regions of the target genes. To this end, we performed ATAC-seq across time in the four different mESC lines. We focused on a short time window and generated

ATAC-seq libraries from cells harvested every 30 min for 2.5 hours to identify the earliest changes in chromatin accessibility.

Using MACS2 to call peaks separately in each sequenced sample, we merged the peak sets to create a consensus peak set with 145,143 peaks. Peak signals were quantified by counting the number of reads in each peak across all samples. To detect peaks changing accessibility due to the effect of dox, we compared the peak signals of each time point to those of the zero-time point in the control cell lines. We found no peaks significantly changing in time due to dox (0 dox associated peak changes; $P < 0.05$). We further determined peaks that changed in both rescue and overexpression models. Briefly, we tested for differential accessibility by comparing all conditions in which *Firre* is induced (timepoint > 0 in *Firre*^{RESCUE} and overexpression lines) to conditions in which the *Firre* transgene is not induced (dox control lines; zero-time point in *Firre*^{RESCUE} and overexpression lines). We identified a total of 55 ATAC peaks that change ($P < 0.05$) in response to *Firre*'s induction (Fig. 7A; Supplemental Fig. S6A). Of these 55 peaks, the vast majority (51 peaks) increase in accessibility, while only four decrease accessibility over the time course (Fig. 7A; Supplemental Fig. S7A).

Next, we wanted to determine how many of the differentially accessible ATAC peaks were in proximity to genes changing in expression upon *Firre* induction. We identified ten differentially accessible ATAC peaks within +/- 50 kb of the TSS of early *Firre* responsive genes. One of these peaks was in proximity to two neighboring genes which both increased in expression (*Duox1* and *Shf*). This revealed eleven genes, including *Firre* itself, that showed proximal increased chromatin accessibility and concordant temporal regulation by *Firre* (Fig. 7B). Analysis of the ATAC-seq peak center location relative to the TSS for the ten non-*Firre* genes revealed that all but one ATAC-peaks are downstream of the TSS (Fig. 7C). The gene *Rapgef4*, for example, rapidly gained ATAC peak signals within its first intron at a proximal enhancer element (Fig. 7C; Supplemental Fig. S7B).

We next wanted to define the timing of chromatin accessibility and gene expression changes upon *Firre* induction. To this end, we compared the temporal dynamics of ATAC- and RNA-seq changes side by side. We found that for all eleven genes, chromatin accessibility was increasing already within 30 min and before most gene expression changes occurred (Fig. 7E).

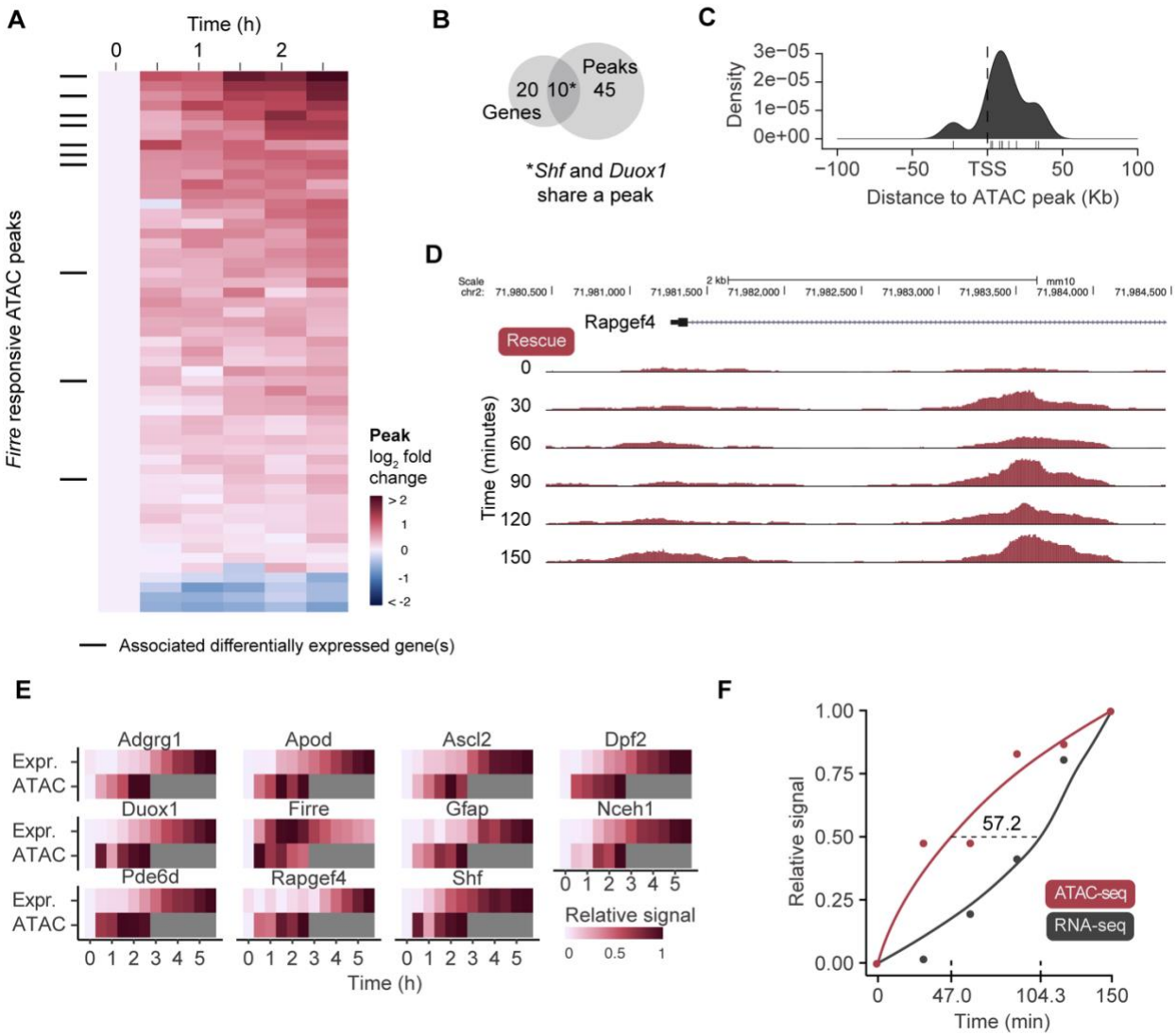


Figure 7: Changes in chromatin accessibility upon induction of *Firre* transgene. (A) Heatmap showing chromatin accessibility $\log_2$-fold changes across time. Increased chromatin accessibility relative to the zero-time point is depicted in shades of red, decreased chromatin accessibility in blue. Black tick marks indicate differentially expressed genes. **(B)** Venn diagram of *Firre*-responsive genes and chromatin accessibility sites. Bottom, TSS metaplot of the density of ATAC-seq peak centers relative to *Firre* responsive gene TSSs. Tick marks indicate the peak centers. **(B)** Genome browser tracks of chromatin accessibility change across time within *Rapgef4* in *Firre*^{RESCUE} cells. **(D)** Heatmap of chromatin accessibility and gene expression changes at each time point after induction of *Firre.* Time points where chromatin accessibility was not measured are indicated in grey. **(E)** Plot of chromatin accessibility (red) and gene expression (black) signal relative to the maximum measured over time. The dashed line represents the difference in half-maximum values.

To formalize this observation, we created a meta-profile of ATAC-seq signal changes and RNA-seq signal changes for these genes (Fig. 7F). We observe that, on average, ATAC peaks reach half-maximal at 47 minutes and RNA abundance at 104 minutes. Collectively, these data provide evidence that *Firre* functions as an RNA-based transcriptional activator by increasing chromatin accessibility that in turn activates increased RNA transcription.

Discussion

The mammalian genome encodes thousands of recently discovered noncoding regulatory elements – ranging from newly identified enhancers to lncRNAs (Consortium 2012; Abascal et al. 2020; Birney et al. 2007; Consortium 2004). Many recent efforts have focused on identifying where, when, and how these sites regulate transcription. This has led to many new insights into how enhancers and lncRNAs work *in cis* to activate or repress target genes (Anderson et al. 2016; Fanucchi et al. 2018; Krawczyk and Emerson 2014; Gil and Ulitsky 2019). However, our understanding of *trans*-acting elements has remained poorly understood. Arguably the least is known about how lncRNAs work to regulate gene expression.

We found that *Firre*-mediated regulatory events occurred rapidly and rescued loss-of-function defects over much longer time scales. Specifically, we found that induction of *Firre* increases chromatin accessibility within 30 minutes, initial gene expression changes occurred in a few hours, and reach homeostasis (in a KO background) after 24 hours. These findings demonstrate the importance of combining very short and longer time frames of lncRNA-mediated gene regulation. Short time courses provide insight into the primary targets, and longer time courses identify stable changes to the gene expression program that are more likely to potentiate disease states associated with *Firre*.

We further investigated the possible mechanisms of the observed gene activation by *Firre*. We started with testing a common theme that lncRNAs bind and influence epigenetic processes (e.g., histone modifications). We found that the loss-of-function for three key epigenetic pathways (PRC2, WDR5, and G9a) did not influence gene activation by *Firre*. We observed similar *Firre* mediated gene activation rates under all conditions tested – further underscoring the robustness of these regulatory events. Thus, these early gene regulatory events are unlikely

due to a commonly proposed mechanistic model of recruitment or eviction of epigenetic complexes to derepress or activate *Firre* target genes.

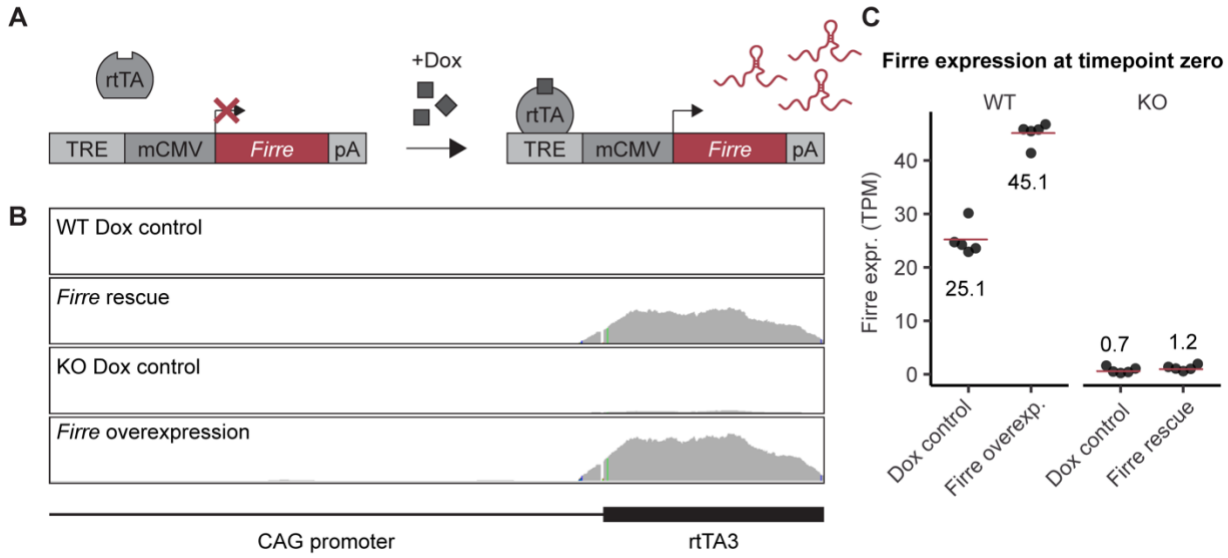
We also tested if the *Firre* RNA localized to these sites by two independent methods: RAP and smRNA FISH. In neither case did we observe that *Firre* binds in proximity to its regulatory sites. Using smRNA FISH we homed in on time points that corresponded to near and after activation events occurred and did not find colocalization of *Firre* exons at sites of gene activation. A possible caveat to both approaches is the targeted nature requiring probes to hybridize to *Firre in situ*. It is possible, although unlikely, that *Firre* cannot be targeted by either approach if it was bound by chromatin complexes making it inaccessible to probes. In contrast, we observed a swift increase in chromatin accessibility within minutes of inducing *Firre*. These findings are surprising in that *Firre* does not need to be in proximity to corresponding sites of gene regulation despite such rapid activation of these genes.

Together these results raise the question of how does a lncRNA regulate gene expression so rapidly? Based on our observation that chromatin accessibility changes precede any gene expression changes, it is unlikely that these effects are the result of changes in the abundance of a protein. That would require a gene to be activated, spliced, translated, and shuttled back to the nucleus, which seems unlikely to occur within 30 minutes. Moreover, it is unlikely that *Firre* directly binds to these sites and activates gene expression akin to a transcription factor. Finally, since *Firre* is not in proximity to these sites it is unlikely able to evict or recruit epigenetic complexes to derepress or activate chromatin.

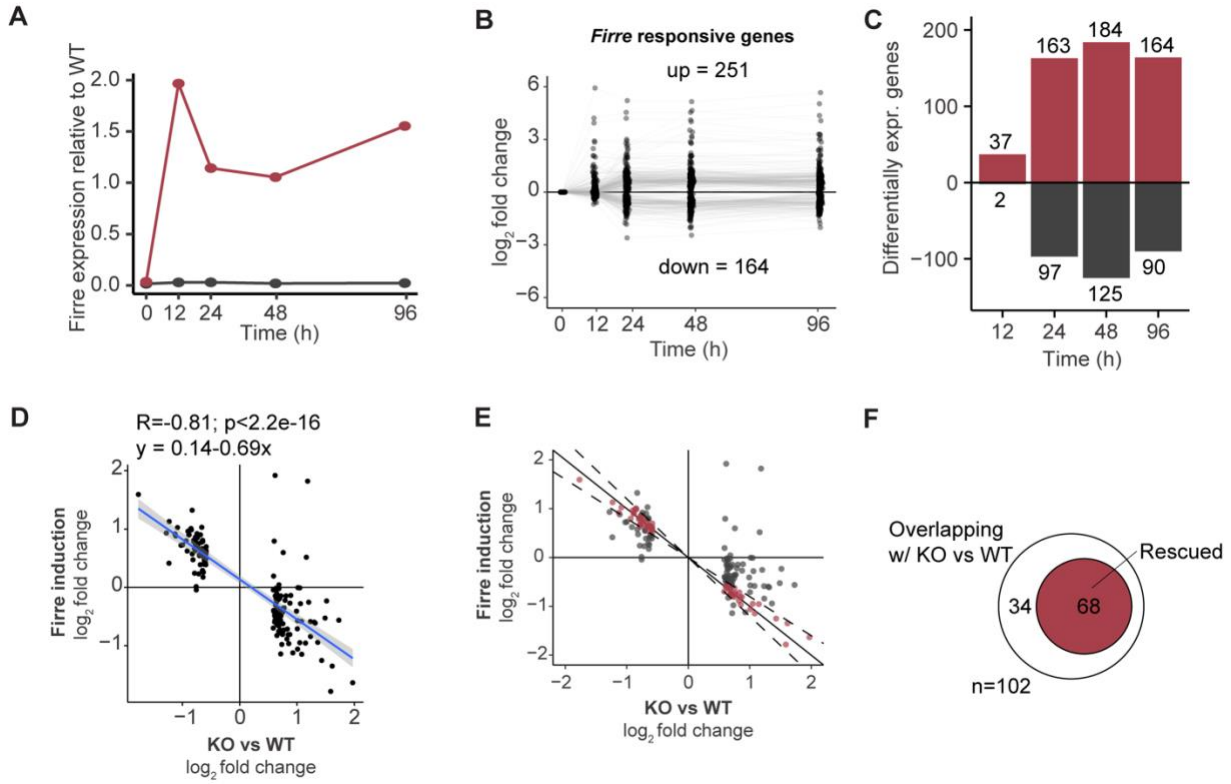
Yet, the chromatin accessibility sites hold a clue of how *Firre* could activate transcription. Based on this information about what is unlikely we can envision a few possible mechanistic insights. First, most of the earliest regulatory events occur downstream of promoter regions, suggesting that *Firre* does not affect transcription at the promoter. Rather these data suggest *Firre* is activating promoter-proximal enhancers to induce target gene expression. Second, since *Firre* does not appear to localize to these sites it suggests that *Firre* must be functioning in the nucleoplasm (as *Firre* localization is almost exclusively nuclear). It is likely that this nucleoplasmic function involves a ribonucleic-based molecular reaction that rapidly leads to either repressing a repressor or activating an activator of transcription, respectively (Brivanlou and Darnell 2002). This would also hold true for interactions with chromatin-remodeling complexes prior to their localization at these regulatory sites.

We can now test these, and other possible RNA-based gene activation models owing to the identification of the early and primary regulatory events. These sites can be assayed to determine what molecular interactions occur upon induction of *Firre*. Thus, we argue the best first step in determining the mechanism of a lncRNA is knowing the primary targets. It is much more difficult to determine the mechanism before knowing the primary regulatory sites. Collectively, these results represent a way forward in dissecting the mechanism by which a model lncRNA can activate gene expression from a distance.

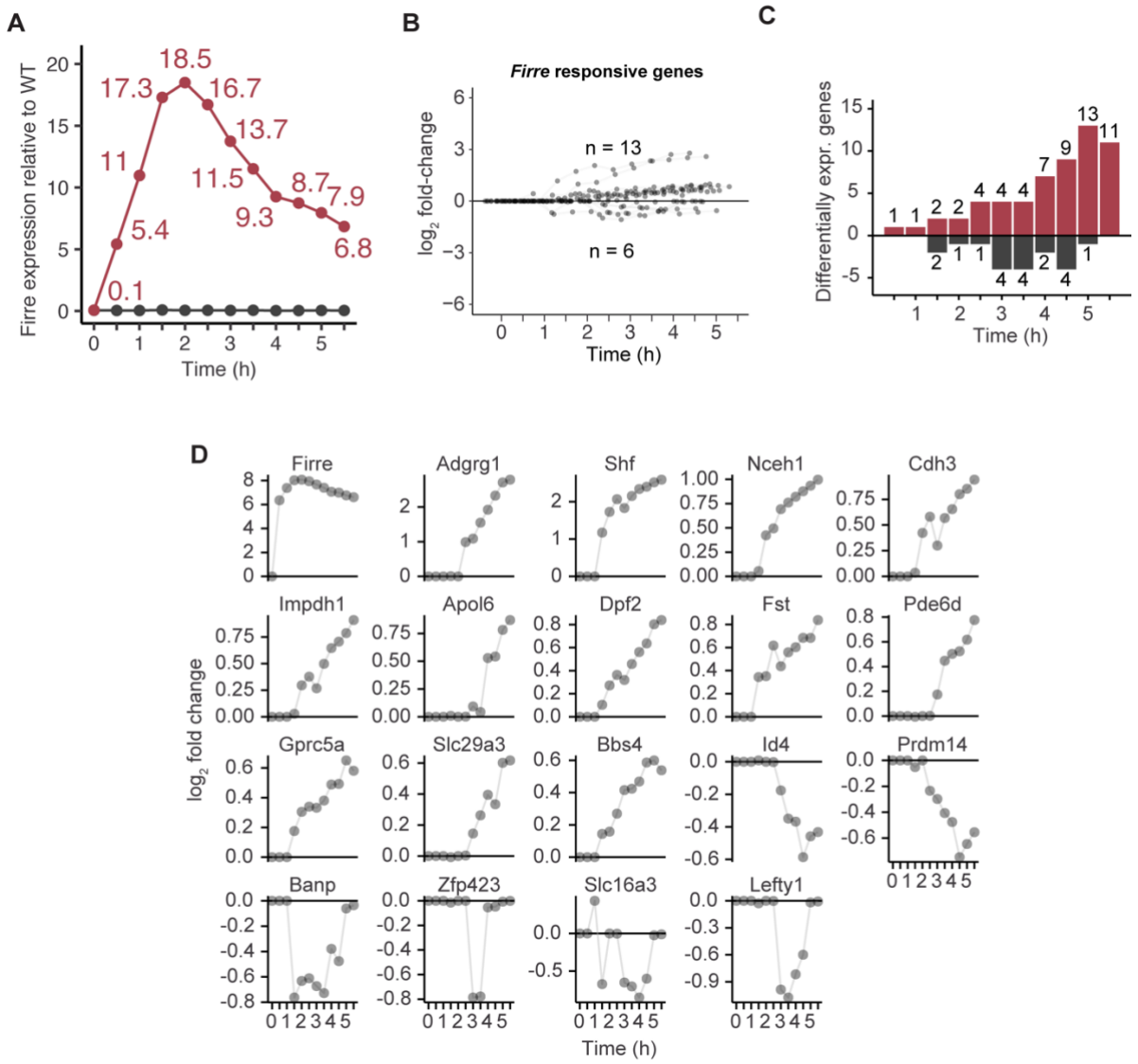
Supplemental Figures



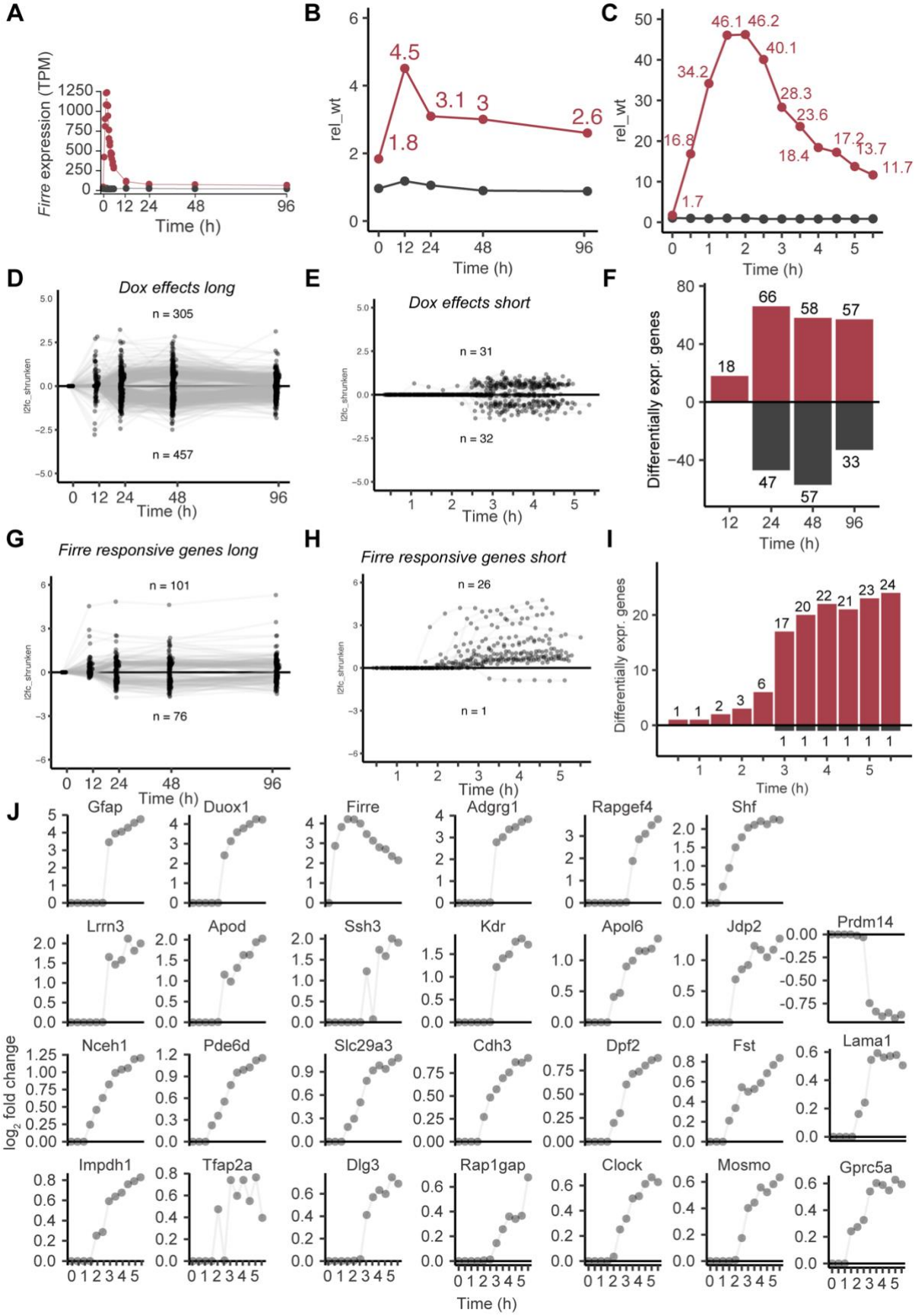
Supplemental Fig S1: Characterization of the mESC cell lines used. (A) Diagram showing the *Firre* transgene construct. **(B)** Genome browser alignment tracks of RNA-seq reads over the rtTA3 element. Dox control lines show no expression of rtTA3. **(C)** *Firre*'s expression measured by RNA-seq at time point zero in the four mESC cell lines. Red line indicates the mean expression.



Supplemental Fig S2: Temporal gene regulation in the *Firre*^{RESCUE} line. (A) *Firre*'s expression measured by RNA-seq relative to *Firre*'s wildtype expression level (~25 TPM). **(B)** Time course plot of the shrunken $\log_2$ fold-changes relative to the zero time point of *Firre* responsive genes with the numbers of up and down-regulated genes indicated. **(C)** The number of genes in each timepoint exceeding the |fold-change| > 1.5 threshold. The negative axis indicates downregulated genes. (D) Scatter plot of the significantly differentially expressed genes overlapping in the *Firre* KO vs WT and in the *Firre*^{RESCUE} time course. The time course fold-change (y-axis) is represented by the maximum fold-change in any one time point. A linear regression is plotted in blue. (E) Scatter plot of *Firre* KO vs WT fold-changes compared the fold-changes in the *Firre*^{RESCUE} time course. The dashed lines indicate the 20% reciprocal fold-change cutoff for genes to be considered rescued. Rescued genes are plotted in red and genes that didn't make the cut-off are plotted in gray. (F) Venn diagram showing the number of rescued genes compared to the total number of overlapping differentially expressed genes.

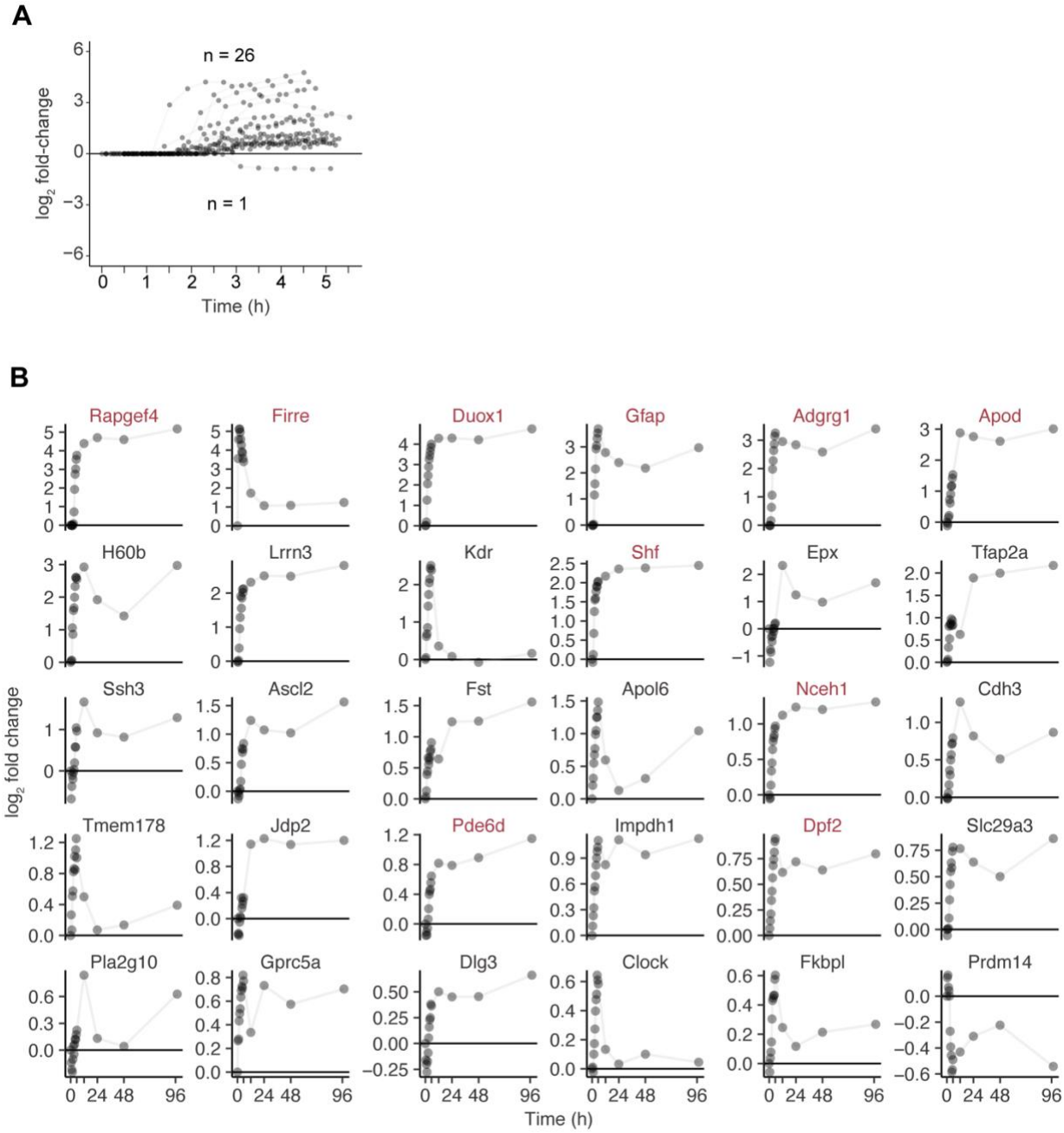


Supplemental Fig S3: Dynamic gene regulation in the *Firre*^{RESCUE} short time course. (A) *Firre*'s expression profile by RNA-seq plotted relative to *Firre*'s wildtype expression level. The red points are the *Firre*^{RESCUE} line and the dox control line is plotted in gray. (B) Time course plot of the shrunken log2 fold-changes for the genes significantly differentially expressed. (C) Bar plot showing the number of genes with a |fold-change| > 1.5 at each time point. Red bars indicate the number of genes up-regulated while black indicates the number of genes down-regulated. (D) Individual profile plots of the log2 fold-change compared to the 0-time point for the significantly differentially expressed genes.


A
Firre expression (TPM)
Time (h)
B
rel_wt
4.5
3.1
3
2.6
1.8
Time (h)
C
rel_wt
46.1
46.2
40.1
34.2
28.3
23.6
16.8
17.2
13.7
18.4
11.7
1.7
Time (h)
D
Dox effects long
n = 305
n = 457
l2fc_shrunken
Time (h)
E
Dox effects short
n = 31
n = 32
l2fc_shrunken
Time (h)
F
Differentially expr. genes
18
66
58
57
47
57
33
Time (h)
G
Firre responsive genes long
n = 101
n = 76
l2fc_shrunken
Time (h)
H
Firre responsive genes short
n = 26
n = 1
l2fc_shrunken
Time (h)
I
Differentially expr. genes
1
1
2
3
6
17
20
22
21
23
24
Time (h)
J
Gfap
Duox1
Firre
Adgrg1
Rapgef4
Shf
Lrrn3
Apod
Ssh3
Kdr
Apol6
Jdp2
Prdm14
Nceh1
Pde6d
Slc29a3
Cdh3
Dpf2
Fst
Lama1
Impdh1
Tfap2a
Dlg3
Rap1gap
Clock
Mosmo
Gprc5a
log2 fold change
Time (h)

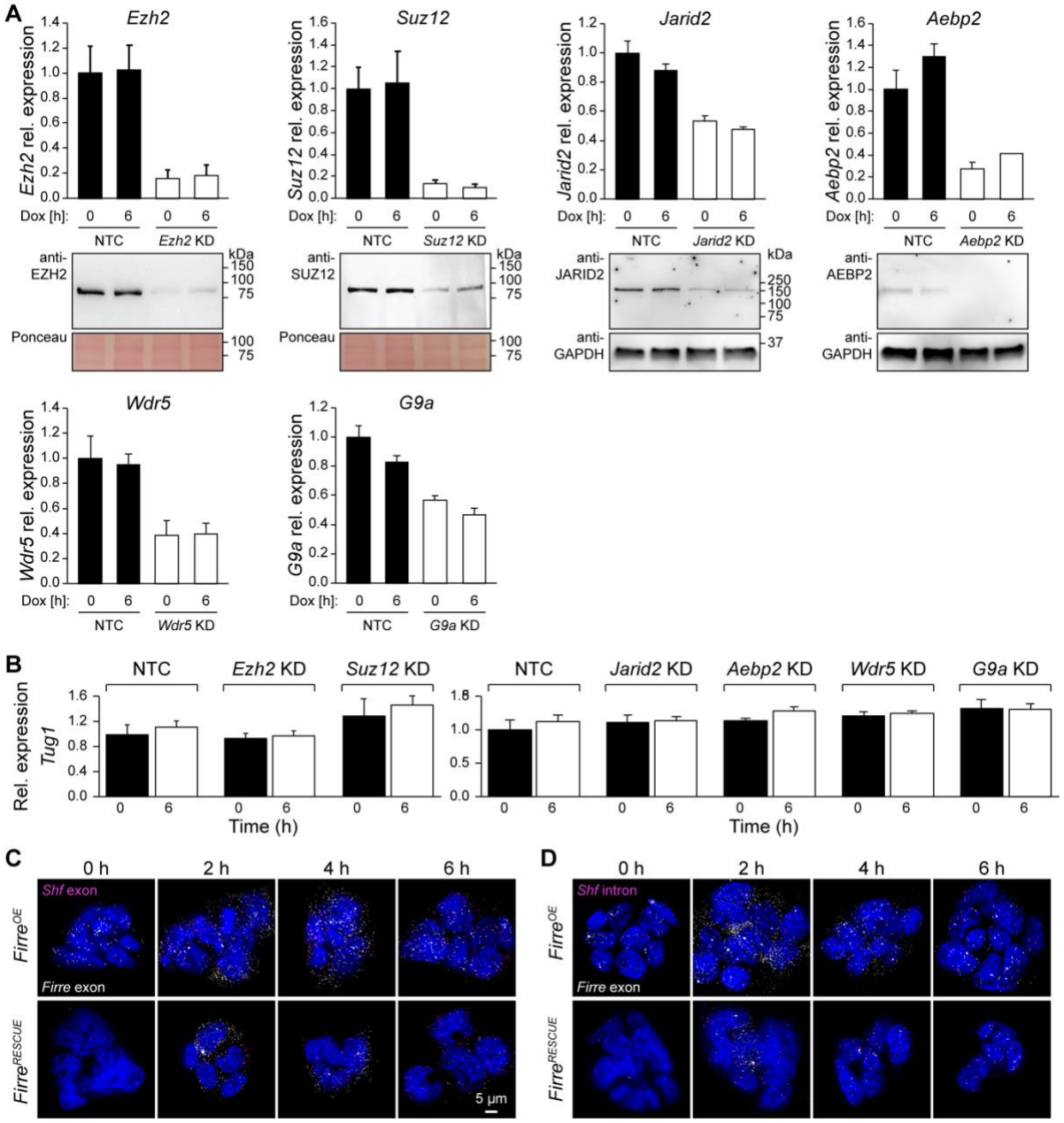
582

Supplemental Fig S4: Gene regulation dynamics in the *Firre*^{OE} time course. (A) *Firre*'s expression profile measured by RNA-seq at all measured time points. The *Firre*^{OE} line is plotted in red, while the dox control line is plotted in black. (B) *Firre'*s expression relative to wildtype *Firre* expression in the 96-hour time course. (C) *Firre'*s expression relative to wildtype in the 5.5-hour time course. (D) Time course plot of the shrunken log2 fold-changes for the significantly differentially expressed genes in the dox control line in the 96-hour time course. (E) Time course plot for the dox control line short time course. (F) The number of differentially expressed genes with |fold-change| > 1.5 at each time point in the *Firre*^{OE} line. (G) Time course plot of the differentially expressed genes in the *Firre*^{OE} long time course. (H) Time course plot for the *Firre*^{OE} short time course. (I) The number of genes at each time point with |fold-change| > 1.5 for the *Firre*^{OE} short time course. (J) Individual gene profile plots showing the fold-change vs the 0-timepoint for the significantly differentially expressed genes in the *Firre*^{OE} short time course.

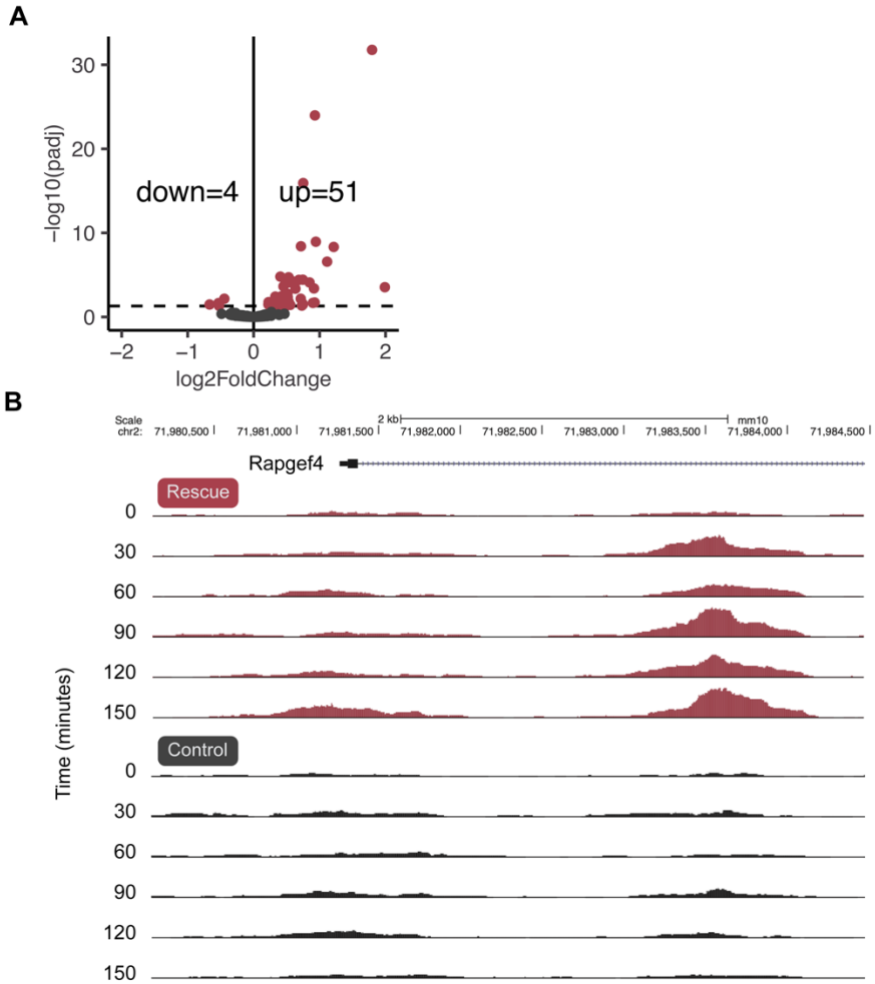


Red indicates genes with nearby differentially accessible ATAC-seq peaks.

Supplemental Fig S5: Gene expression profile plots for genes detected as differentially expressed in the combined model. (A) Time course plot of the shrunken log2 fold-change for the genes significantly differentially expressed in the short time course combined model (*Firre*^{OE} and *Firre*^{RESCUE}). (B) Individual gene plots show the fold-change compared to the zero-timepoint for all measured timepoints. Genes with nearby differentially accessible ATAC-seq peaks are colored in red.



Supplemental Fig S6: Chromatin modifier knockdown validation. (A) Relative expression of the genes targeted for GapmeR knockdown measured by RT-qPCR and western blot. (B) Relative expression of a control gene (Tug1) where the expression level should not be affected by the induction of Firre. smRNA FISH images of the *Shf* exon and *Firre* transgene (C) and *Shf* intron and *Firre* transgene (D) in *Firre*-overexpression and *Firre*-rescue cells after 0, 2, 4, and 6 hours of dox treatment.



Supplemental Fig S7: Differentially accessible ATAC-seq peaks. (A) Volcano plot showing the fold-change (x-axis) vs -log10(p-value) (y-axis). Significantly differentially accessible peaks are plotted in red with a cutoff of (padj < 0.05). (B) Example browser tracks of ATAC-seq peak signal through time in the *Firre*^{RESCUE} and dox control line.

Methods

Derivation of genetically modified mESCs

mESC derivations were performed with the assistance of Harvard's Genome Modification Facility and performed as described in (Behringer et al. 2014). Briefly, 2-5 IU of pregnant mare serum gonadotropin (PMSG) and 2-5 IU of human chorionic gonadotropin HCG were administered 48 hours a part by intraperitoneal (i.p.) injection to female Firre-deficient mice. PMSG/HCG treated mice were setup for timed matings (1-2 per cage) with either male mice deficient for Firre (Firre-/y) or with male mice that contain an inducible Firre transgene in the Firre-deficient background (Firre-/y; tg-Firre; CAGs-rtTA3) (generation of mice previously described in Lewandowski et al (Lewandowski et al. 2019).Female mice were checked for copulation plugs the following morning. Approximately 72 hours later, blastocysts were flushed out of the uterine horns from female mice with successful copulation plugs. Blastocysts with a detected cavity were selected for culture and expansion. Individual clones were screened by PCR genotyping for Firre wild-type, knockout, transgene, and CAGs-rtTA3 alleles as described in Lewandowski et al. (Lewandowski et al. 2019).

Cells were maintained on 0.1% gelatin in KnockOut DMEM (Thermo Fisher, 10829018) supplemented with 12.5% FCS (MilliporeSigma, ES-009-B), 1X GlutaMAX supplement (Thermo Fisher, 35050061), 1X non-essential amino acids (Thermo Fisher, 11140050), 100 U/mL penicillin/streptomycin (Thermo Fisher, 15140122), 100 μM 2-mercaptoethanol (Thermo Fisher, 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. Cells were routinely tested for mycoplasma contamination by PCR.

Cells were treated with 1 ng/μL dox (MilliporeSigma, D9891-5G) the day after they were seeded and one hour after the medium was changed to reduce cell stress.

RNA isolation, quantification, and quality control

Total RNA was extracted using TRIzol lysis reagent (Thermo Fisher) and purified using the RNeasy Plus Mini Kit (Qiagen) according to the manufacturer's instructions. RNA concentration

was measured with a Qubit fluorometer (Invitrogen), and RNA quality was determined using a Bioanalyzer or TapeStation (Agilent Technologies).

RT-qPCR

Total RNA was reverse-transcribed using the SuperScript IV First-Strand Synthesis System with random hexamers (Thermo Fisher) according to the manufacturer's instructions. qRT-PCR samples were prepared using TaqMan probes and the TaqMan Fast Advanced Master Mix (Thermo Fisher) according to the manufacturer's instructions and run in technical triplicates on a Bio-Rad C1000 Touch Thermal Cycler. C_t values were normalized against the internal controls *Gapdh* and *ActB*. Fold differences in expression levels were calculated according to the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

RNA-Seq

RNA-seq libraries were prepared from 1 µg of total RNA using the TruSeq RNA Library Prep Kit v2 (Illumina) with poly(A) selection according to the manufacturer's instructions. Multiples of 12 randomized samples were each multiplexed, pooled, and run in one HiSeq lane.

ATAC-seq

ATAC-seq libraries were prepared from 25,000 cells following the Omni-ATAC protocol (Corces et al., 2017) and using the Tagment DNA Enzyme and Buffer Kit (Illumina). Samples were pooled and run in one HiSeq lane.

Knockdowns

For knockdown assays, triplicates of $3x10^5$ mESCs were seeded per well of a 6-well plate. The next day, cells were reverse transfected with 25 nM (non-targeting control, *Jarid2*, *Aebp2*, *Suz12*, *Ezh2*) or 50 nM (*Wdr5*, *G9a*) GapmeRs using Lipofectamine RNAiMAX Transfection Reagent (Thermo Fisher) according to manufacturer's instructions. After 24 hours, either no dox (0-hour time point) or 1 ng/µL dox (6-hour time point) was added without performing a media change, and cells were harvested after 6 hours for RNA and protein extraction.

Protein extraction and western blotting

Cells were collected by trypsinization, washed with cold PBS, and incubated in lysis buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1% NP-40, 3 mM MgCl2, 5% glycerol, 1X protease inhibitors) for 20 min on ice. Protein extracts were collected as supernatant after centrifugation at 14,000xg for 10 min at 4°C, resolved on an Any kD Mini PROTEAN TGX Precast Protein Gel (Bio-Rad) and transferred on a nitrocellulose membrane (Millipore) via wet transfer in the cold room overnight. Membranes were blocked in 5% milk in TBS-T (TBS with 0.1% Tween 20) and probed with anti-JARID2 (D6M9X), anti-AEBP2 (D7C6X), anti-EZH2 (D2C9), anti-SUZ12 (Cell Signaling Technology, D39F6), or anti-GAPDH antibody (14C10, all Cell Signaling Technology) at a dilution of 1:1000 in blocking buffer overnight at 4°C or for one hour at room temperature. The membrane was washed in TBS-T, incubated with a rabbit HRP-coupled secondary antibody (Agilent) in TBS-T for 1 hour, and washed again in TBS-T. Proteins were detected with the SuperSignal West Pico PLUS Chemiluminescent Substrate (Thermo Fisher) and acquired using an ImageQuant LAS 4000 (GE Healthcare).

smRNA FISH

Oligonucleotides tiling *Firre* exons (including its RRDs), *Shf* exons, and *Shf* intron 1 were designed with the Stellaris RNA FISH probe designer (LGC Biosearch Technologies, version 4.2), labeled with Quasar 570 or 670, and produced by LGC Biosearch Technologies.

Cells were seeded on cover glasses coated with 0.1% gelatin. 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 incubated in 70% ethanol at 4°C for at least 1 hour and then washed with 1 mL of wash buffer A (LGC Biosearch Technologies) at room temperature for 5 min. Cells were hybridized with 100 µL of hybridization buffer (LGC Biosearch Technologies) containing the FISH probes at a 1:100 dilution in a humid chamber at 37 °C overnight up to 16 hours. The next day, cells were washed with 1 mL of wash buffer A at 37 °C for 30 min and stained with wash buffer A containing 10 µg/mL Hoechst 33342 (Thermo Fisher) at 37 °C for 30 min. Cover glasses were washed with 1 mL of wash buffer B (LGC Biosearch Technologies) at room temperature for 5 min, mounted with ProLong Gold (Thermo Fisher) on a glass slide, and left to curate at 4°C overnight before proceeding to image acquisition.

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 capturing the entire cell colony were acquired. Images were deconvolved with the built-in DeltaVision SoftWoRx Imaging software and maximum intensity projections were created using ImageJ/Fiji.

Validation

Firre distance quantification

Imaris image analysis software (Bitplane) was used to quantify the distances between FISH spots. Images were prepared and FISH spots detected as previously described (Gil and Ulitsky 2019; Khong and Parker 2018)Coordinates of each *Shf* intron spot and the closest '*Firre* cloud' or the five closest *Firre* exon spots within the same cell were recorded and used to calculate their Euclidean distance from each other.

<u>Data analysis</u>

Tissue and static Firre^{KO} vs WT analysis

FASTQs for Firre^{KO} and WT mouse tissue RNA-seq experiments were retrieved from GEO GSE83631 and GSE127554 corresponding to Lewandowski et al. (Lewandowski et al. 2019)and Andergassen et al. (Andergassen et al. 2019)The static Firre^{KO} vs WT comparison was carried out using the Firre WT and KO mESCs before adding dox (n = 12). Quality control, read mapping, and quantification was performed using nf-core/rnaseq v1.4.2 (Ewels et al. 2020)Reads were mapped to mm10 and quantified with Salmon v1.5.2 (Patro et al. 2017) using the Gencode M25 annotation. Significantly differentially expressed genes were calculated using DESeq2 (Love et al. 2014) (comparing Firre^{KO} vs WT) with an adjusted p-value cutoff of 0.05 and a fold-change cutoff of 1.5 on the shrunken $\log_2$ fold change values from DESeq2.

Firre induction time course quantification

Quality control, read mapping, and quantification was performed using nf-core/rnaseq v1.4.2. Reads were mapped to mm10 and quantified with Salmon using the Gencode M25 annotation.

Doxycycline control RNA-seq

Differential expression was calculated using DEseq2 by using the Wald test to compare each time point back to the zero-time point. The rescue and overexpression dox control lines and the corresponding short and long-time courses were analyzed separately. Cutoffs to call differentially expressed genes were made using an adjusted p-value cutoff of 0.05 and whether a gene achieved a fold-change greater than 1.5 at any time point (calculated using the shrunken $\log_2$ fold change).

Firre induction time course analysis

Differential expression was calculated with DEseq2 using the likelihood ratio test to compare a full model which contains the Firre induction/time point interaction term to a reduced model lacking that term. Calling differentially expressed genes proceeded in two steps: 1. A cutoff was made on the adjusted p-values from the likelihood ratio test ($p < 0.05$), filtering to genes that changed significantly in time as compared to the dox control. 2. A fold change cutoff was made by comparing the expression through time back to the zero time point (without any comparison to the dox control line). A gene was called differentially expressed (responding to Firre) if it achieved a fold change greater than 1.5 in any timepoint as measured by the shrunken $\log_2$ fold change from DEseq2.

The combined model using all four cell lines was analyzed in the same manner with the same thresholding strategy, but the full and reduced models contained an additional indicator variable for the Firre genotype (WT or KO).

The $\log_2$ fold changes shown in the heatmaps are the shrunken $\log_2$ fold changes reported by DEseq2.

Identification of rescued genes

A gene was considered rescued if 1. It was significantly differentially expressed in the static Firre^{KO} vs WT comparison and in the Firre^{RESCUE} induction experiment, and 2. If the fold change was reciprocal to within 20% of the fold change value in the Firre^{KO} vs WT.

ATAC-seq time course analysis

Quality control, read mapping, and peak calling were carried out using nf-core/atacseq v1.2.0. Reads were mapped to mm10 and the mm10 v2 blacklist from ENCODE was used. Within the nf-core/atacseq pipeline MACS2 v2.2.7.1 (Feng et al. 2012; Zhang et al. 2008) was used to call peaks in each experiment separately. Consensus peaks were called using the peak regions present in any single experiment. Reads within peaks were calculated using featureCounts. Differentially accessible peaks were calculated using DEseq2 to compare the read counts in the samples in which Firre was present (>0 timepoint in the Firre^{RESCUE} and overexpression lines) and the samples in which Firre was absent (dox control and 0-time points). A cutoff was made using the adjusted p-value ($p<0.05$).

References

Abascal F, Acosta R, Addleman NJ, Adrian J, Afzal V, Aken B, Akiyama JA, Jammal OA, Amrhein H, Anderson SM, et al. 2020. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* 583: 699–710.

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. *Developmental Medicine & Child Neurology* n/a-n/a.

Allen MA, Andrysik Z, Dengler VL, Mellert HS, Guarnieri A, Freeman JA, Sullivan KD, Galbraith MD, Luo X, Kraus WL, et al. 2014. Global analysis of p53-regulated transcription identifies its direct targets and unexpected regulatory mechanisms. *Elife* 3: e02200.

Andergassen D, Rinn JL. 2021. From genotype to phenotype: genetics of mammalian long non-coding RNAs in vivo. *Nat Rev Genet* 23: 229–243.

Andergassen D, Smith ZD, Lewandowski JP, Gerhardinger C, Meissner A, Rinn JL. 2019. In vivo Firre and Dxz4 deletion elucidates roles for autosomal gene regulation. *Elife* 8: e47214.

Anderson KM, Anderson DM, McAnally JR, Shelton JM, Bassel-Duby R, Olson EN. 2016. Transcription of the non-coding RNA upperhand controls Hand2 expression and heart development. *Nature* 539: 433–436.

Barutcu AR, Maass PG, Lewandowski JP, Weiner CL, Rinn JL. 2018. A TAD boundary is preserved upon deletion of the CTCF-rich Firre locus. *Nature communications* 9: 1444.

Behringer R, Gertsenstein M, Nagy KV, Nagy A. 2014. Manipulating the Mouse Embryo: A Laboratory Manual, 4th ed. Cold Spring Harbor Press, New York.

Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Stamatoyannopoulos JA, et al. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447: 799–816.

Brivanlou AH, Darnell JE. 2002. Signal Transduction and the Control of Gene Expression. *Science* 295: 813–818.

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 & Development* 25: 1915–1927.

Carpenter S, Aiello D, Atianand MK, Ricci EP, Gandhi P, Hall LL, Byron M, Monks B, Henry-Bezy M, Lawrence JB, et al. 2013. A Long Noncoding RNA Mediates Both Activation and Repression of Immune Response Genes. *Science* 341: 789–792.

Charoensawan V, Martinho C, Wigge PA. 2015. “Hit-and-run”: Transcription factors get caught in the act. *Bioessays* 37: 748–754.

Consortium EP. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature Publishing Group* 489: 57–74.

Consortium TEP. 2004. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* 306: 636–640.

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 Research* 22: 1775–1789.

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. *Molecular Cell* 1–14.

Elling R, Robinson EK, Shapleigh B, Liapis SC, Covarrubias S, Katzman S, Groff AF, Jiang Z, Agarwal S, Motwani M, et al. 2018. Genetic Models Reveal cis and trans Immune-Regulatory Activities for lincRNA-Cox2. *CellReports* 25: 1511-1524.e6.

Engreitz JM, Haines JE, Perez EM, Munson G, Chen J, Kane M, McDonel PE, Guttman M, Lander ES. 2016. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature Publishing Group* 539: 452–455.

Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, Garcia MU, Tommaso PD, Nahnsen S. 2020. The nf-core framework for community-curated bioinformatics pipelines. *Nat Biotechnol* 38: 276–278.

Fanucchi S, Fok ET, Dalla E, Shibayama Y, Börner K, Chang EY, Stoychev S, Imakaev M, Grimm D, Wang KC, et al. 2018. Immune genes are primed for robust transcription by proximal long noncoding RNAs located in nuclear compartments. *Nat Genet* 51: 138–150.

Feng J, Liu T, Qin B, Zhang Y, Liu XS. 2012. Identifying ChIP-seq enrichment using MACS. *Nat Protoc* 7: 1728–1740.

Gil N, Ulitsky I. 2019. Regulation of gene expression by cis-acting long non-coding RNAs. *Nat Rev Genet* 21: 102–117.

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. *CellReports* 16: 2178–2186.

Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L, et al. 2011. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* 477: 295–300.

Guttman M, Rinn JL. 2012. Modular regulatory principles of large non-coding RNAs. *Nature* 482: 339–346.

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. *Nature Structural & Molecular Biology* 21: 198–206.

Hacisuleyman E, Shukla CJ, Weiner CL, Rinn JL. 2016. Function and evolution of local repeats in the Firre locus. *Nature communications* 7: 11021.

Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, Khalil AM, Zuk O, Amit I, Rabani M, et al. 2010. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* 142: 409–419.

Khong A, Parker R. 2018. mRNP architecture in translating and stress conditions reveals an ordered pathway of mRNP compaction. *J Cell Biol* 217: 4124–4140.

Krawczyk M, Emerson BM. 2014. p50-associated COX-2 extragenic RNA (PACER) activates COX-2 gene expression by occluding repressive NF-κB complexes. *Elife* 3: e01776.

Lammers NC, Kim YJ, Zhao J, Garcia HG. 2020. A matter of time: Using dynamics and theory to uncover mechanisms of transcriptional bursting. *Curr Opin Cell Biol* 67: 147–157.

Lee TI, Young RA. 2000. TRANSCRIPTION OF EUKARYOTIC PROTEIN-CODING GENES. *Genetics* 34: 77–137.

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.

Loewer S, Cabili MN, Guttman M, Loh Y-H, Thomas K, Park IH, Garber M, Curran M, Onder T, Agarwal S, et al. 2010. Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nature Genetics* 42: 1113–1117.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15: 550.

Maamar H, Cabili MN, Rinn J, Raj A. 2013. linc-HOXA1 is a noncoding RNA that represses Hoxa1 transcription in cis. *Genes & Development* 27: 1260–1271.

Maass PG, Rump A, Schulz H, Stricker S, Schulze L, Platzer K, Aydin A, Tinschert S, Goldring MB, Luft FC, et al. 2012. A misplaced lncRNA causes brachydactyly in humans. *The Journal of clinical investigation* 122: 3990–4002.

Mahat DB, Salamanca HH, Duarte FM, Danko CG, Lis JT. 2016. Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation. *Mol Cell* 62: 63–78.

Mitchell PJ, Tjian R. 1989. Transcriptional Regulation in Mammalian Cells by Sequence-Specific DNA Binding Proteins. *Science* 245: 371–378.

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.

Morris KV, Mattick JS. 2014. The rise of regulatory RNA. *Nature Reviews Genetics* 15: 423–437.

Much C, Smallegan MJ, Hwang T, Hanson SD, Dumbovic G, Rinn JL. 2022. Evolutionary divergence of Firre localization and expression. *Rna* rna.079070.121.

Paralkar VR, Taborda CC, Huang P, Yao Y, Kossenkov AV, Prasad R, Luan J, Davies JOJ, Hughes JR, Hardison RC, et al. 2016. Unlinking an lncRNA from Its Associated cis Element. *Molecular Cell* 62: 104–110.

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.

Rinn JL, Chang HY. 2012. Genome regulation by long noncoding RNAs. *Annual review of biochemistry* 81: 145–166.

Rinn JL, Chang HY. 2020. Long Noncoding RNAs: Molecular Modalities to Organismal Functions. *Annu Rev Biochem* 89: 283–308.

Sasse SK, Gruca M, Allen MA, Kadiyala V, Song T, Gally F, Gupta A, Pufall MA, Dowell RD, Gerber AN. 2019. Nascent transcript analysis of glucocorticoid crosstalk with TNF defines primary and cooperative inflammatory repression. *Genome Res* 29: 1753–1765.

Stewart-Ornstein J, Nelson C, DeRisi J, Weissman JS, El-Samad H. 2013. Msn2 Coordinates a Stoichiometric Gene Expression Program. *Curr Biol* 23: 2336–2345.

Tan JY, Smith AAT, Silva MF da, Matthey-Doret C, Rueedi R, Sönmez R, Ding D, Kutalik Z, Bergmann S, Marques AC. 2017. cis-Acting Complex-Trait-Associated lincRNA Expression Correlates with Modulation of Chromosomal Architecture. *CellReports* 18: 2280–2288.

Tong C, Yin Y. 2021. Localization of RNAs in the nucleus: cis- and trans- regulation. *Rna Biol* 18: 1–14.

Wang KC, Chang HY. 2011. Molecular mechanisms of long noncoding RNAs. *Molecular Cell* 43: 904–914.

Wang Z, Deng W. 2022. Dynamic transcription regulation at the single-molecule level. *Dev Biol* 482: 67–81.

Weikum ER, Knuesel MT, Ortlund EA, Yamamoto KR. 2017. Glucocorticoid receptor control of transcription: precision and plasticity via allostery. *Nat Rev Mol Cell Bio* 18: 159–174.

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