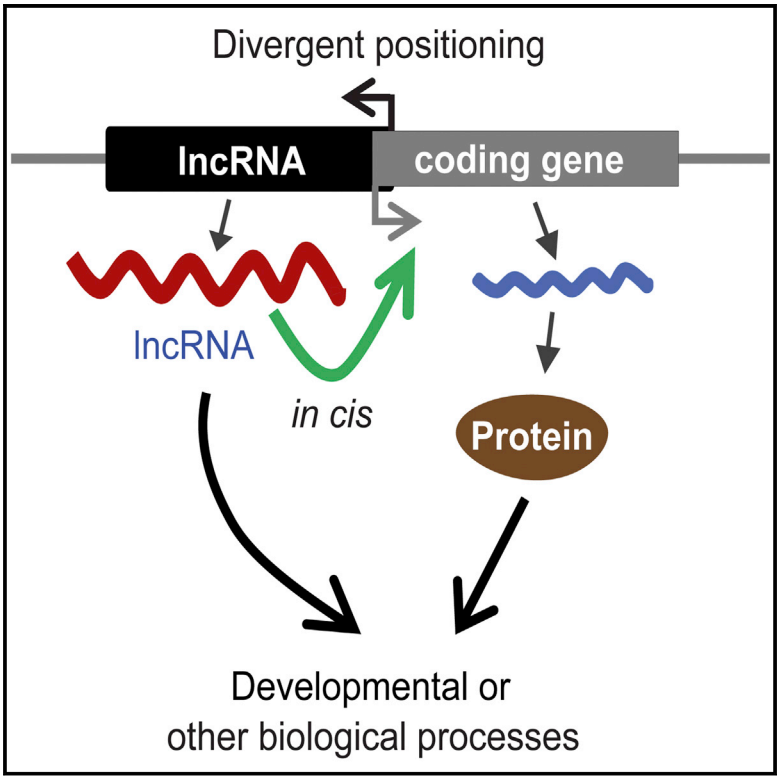
Article

Cell Stem Cell

Divergent lncRNAs Regulate Gene Expression and Lineage Differentiation in Pluripotent Cells

Graphical Abstract



Authors

Sai Luo, J. Yuyang Lu, Lichao Liu, ..., Jianlong Wang, Yong E. Zhang, Xiaohua Shen

Correspondence

xshen@tsinghua.edu.cn

In Brief

Based on broad genomic analysis and specific functional characterization in pluripotent cells, Luo and colleagues suggest that a major class of lncRNAs that is arranged divergently to nearby genes plays a role in transcriptional regulation to fine-tune gene expression and lineage differentiation.

Highlights

- Genomic organization of divergent lncRNAs correlates strongly with regulatory genes
- Divergent lncRNAs regulate the expression of adjacent genes in pluripotent cells
- The lncRNA *Evx1as* binds locally to chromatin and promotes *EVX1* transcription in *cis*
- *Evx1as* acts upstream of *EVX1* to mediate mesendodermal lineage differentiation

Accession Numbers

GSE70420
GSE70419
GSE62899
GSE58514

Luo et al., 2016, Cell Stem Cell *18*, 637–652
May 5, 2016
http://dx.doi.org/10.1016/j.stem.2016.01.024

CellPress

Cell Stem Cell
Article

Divergent lncRNAs Regulate Gene Expression and Lineage Differentiation in Pluripotent Cells

Sai Luo,[1] J. Yuyang Lu,[1] Lichao Liu,[1] Yafei Yin,[1] Chunyan Chen,[2] Xue Han,[1] Bohou Wu,[1] Ronggang Xu,[1] Wei Liu,[1] Pixi Yan,[1] Wen Shao,[1] Zhi Lu,[3] Haitao Li,[1] Jie Na,[1] Fuchou Tang,[4] Jianlong Wang,[5] Yong E. Zhang,[2] and Xiaohua Shen[1,*]
[1]Tsinghua-Peking Center for Life Sciences, School of Medicine, Tsinghua University, Beijing 100084, China
[2]Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China
[3]School of Life Sciences, Tsinghua University, Beijing 100084, China
[4]Biodynamic Optical Imaging Center, Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China
[5]The Black Family Stem Cell Institute, Department of Developmental and Regenerative Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
*Correspondence: xshen@tsinghua.edu.cn
http://dx.doi.org/10.1016/j.stem.2016.01.024

SUMMARY

Divergent lncRNAs that are transcribed in the opposite direction to nearby protein-coding genes comprise a significant proportion (~20%) of total lncRNAs in mammalian genomes. Through genome-wide analysis, we found that the distribution of this lncRNA class strongly correlates with essential developmental regulatory genes. In pluripotent cells, divergent lncRNAs regulate the transcription of nearby genes. As an example, the divergent lncRNA *Evx1as* promotes transcription of its neighbor gene, *EVX1*, and regulates mesendodermal differentiation. At a single-cell level, early broad expression of *Evx1as* is followed by a rapid, high-level transcription of *EVX1*, supporting the idea that *Evx1as* plays an upstream role to facilitate *EVX1* transcription. Mechanistically, *Evx1as* RNA binds to regulatory sites on chromatin, promotes an active chromatin state, and interacts with Mediator. Based on our analyses, we propose that the biological function of thousands of uncharacterized lncRNAs of this class may be inferred from the role of their neighboring adjacent genes.

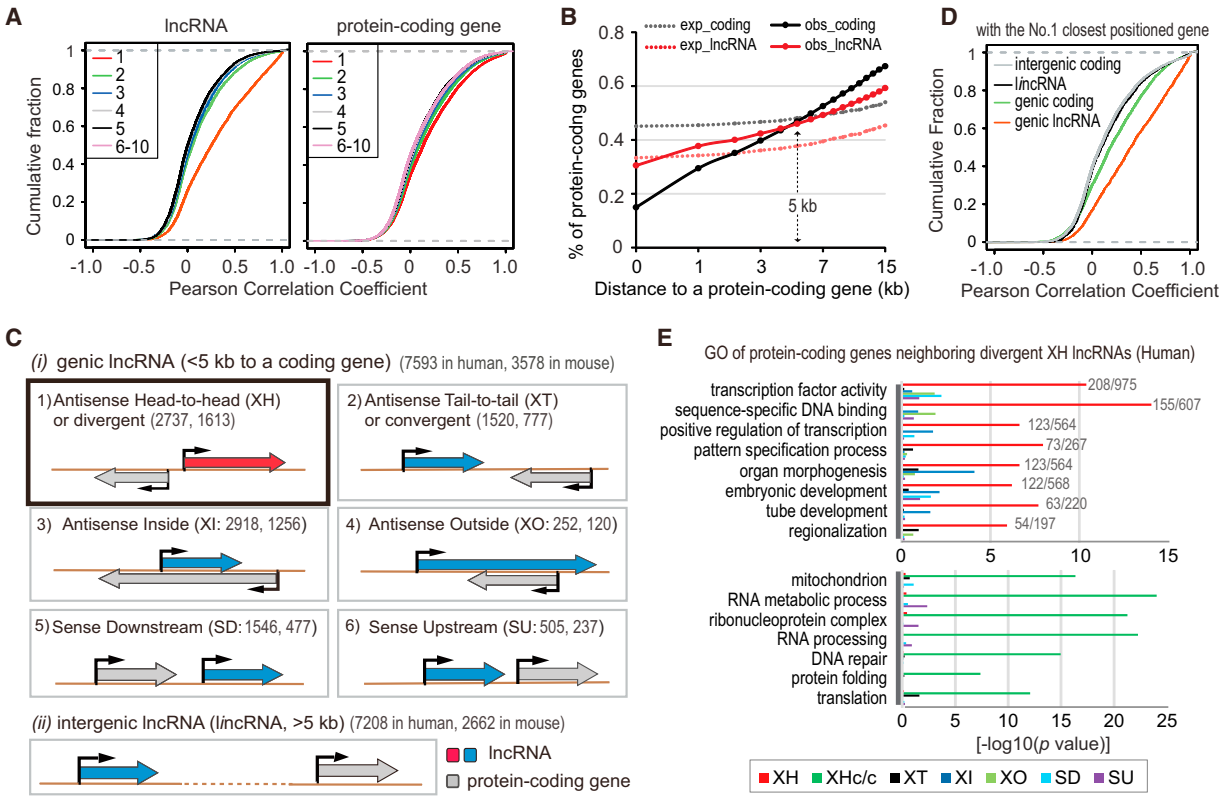
INTRODUCTION

Much of the developmental complexity of higher eukaryotes is thought to arise from gene regulation rather than from an increase in the number of protein-coding genes (Morris and Mattick, 2014). RNA may represent a hidden layer of regulatory information in complex organisms, as increasing amounts of genetic information are expressed as and transacted by RNA (Taft et al., 2007; Yin et al., 2015). Genome-wide transcriptome analyses have identified thousands of long noncoding RNAs (lncRNAs) (Derrien et al., 2012). It has been proposed that lncRNAs may serve as versatile regulators of diverse aspects of biology in physiological and pathological contexts (Batista and Chang, 2013; Sauvageau et al., 2013). However, the functionality of vast majorities of lncRNAs is unknown. Identifying functional lncRNAs and then inferring biological pathways in which they act in represent major challenges in understanding genome complexity and RNA-mediated gene regulation.

Various methods based on chromatin features, genome distribution, expression pattern, and subcellular localization have been used to categorize lncRNAs and to characterize their function (Cabili et al., 2011; Derrien et al., 2012; Mondal et al., 2010; Ponjavic et al., 2009). However, a unified approach to classify all lncRNA genes and link lncRNA biotypes with function is still lacking. Initial evidence of genomic juxtaposition and co-expression of tissue-specific lncRNAs and protein-coding genes was reported. For example, brain-expressed lncRNAs show regionally enriched expression profiles that are similar to those of adjacent protein-coding genes of neurological importance (Mercer et al., 2008; Ponjavic et al., 2009). A few lncRNAs expressed in the lung and foregut endoderm are positioned adjacent to transcription factors critical for lung development (Herriges et al., 2014). Studies of subsets of lncRNAs expressed in human or mouse embryonic stem cells (ESCs) showed coordinated expression with genomically associated developmental genes during differentiation (Dinger et al., 2008; Sigova et al., 2013).

Evidence suggests that antisense transcription may be associated with promoters of genes encoding transcriptional regulators (Cabili et al., 2011; Derrien et al., 2012; Hu et al., 2014; Lepoivre et al., 2013; Sigova et al., 2013). However, the biological significance of these observations is not understood. It has been a matter of debate whether lncRNA expression correlates with neighboring (*cis*) or distal (*trans*) protein-coding genes and whether lncRNAs can regulate their protein-coding neighbors in *cis* (Cabili et al., 2011; Derrien et al., 2012; Mondal et al., 2010; Ørom et al., 2010; Sigova et al., 2013).

Here we revealed a non-random distribution of lncRNAs in the genome through comprehensive locus categorization. Divergent lncRNAs that are transcribed on the opposite strand from their neighboring protein-coding genes represent an interesting class comprising ~20% of total lncRNAs in mammalian genomes. However, no clear function for this lncRNA class has yet been identified. An interesting hypothesis is that divergent gene organization may allow lncRNA transcripts to regulate their adjacent coding genes. However, the mechanisms by which divergent lncRNA transcripts regulate their nearby sense mRNAs are not fully understood.



(legend on next page)

In-depth characterization of the *Evx1as/EVX1* locus reveals a coupled transcription activation of *Evx1as* and *EVX1* in an *Evx1as*-dependent manner during ESC differentiation. *Evx1as* RNA promotes *EVX1* transcription by coating its own locus on chromatin and modulating local chromatin state and configuration. An early, broad expression pattern of *Evx1as* prior to *EVX1* activation supports *Evx1as* as an upstream pilot factor regulating *EVX1* expression. Remarkably, knocking down 18 of 24 divergent lncRNAs led to downregulation of nearby genes. *Evx1as* depletion or deletion elicited global expression changes in lineage differentiation known to involve *EVX1*. Our work suggests that divergent lncRNAs, or at least a subset of them, can positively regulate the transcription of nearby genes in *cis* and participate in biological processes similar to those controlled by the nearby protein-coding genes.

RESULTS

LncRNA Locus Classification Reveals a Non-random Genomic Distribution

In an effort to classify all lncRNA genes and reveal their potential biological roles, we determined their genomic distribution patterns relative to protein-coding loci. Pairwise Pearson correlation analysis revealed that lncRNAs, but not protein-coding genes, exhibit a significantly higher expression correlation ($p < 10^{-38}$) with the closest positioned (#1) gene than with other distal nearby genes (#2–#10) (Figure 1A). Interestingly, within a 5-kb distance range, the proportion of protein-coding genes observed to neighbor a lncRNA rather than a coding gene is much higher than expected from a random distribution (t test, $p < 1.29^{-9}$) (Figure 1B). We therefore chose a distance cutoff of 5 kb from a protein-coding gene to define two classes of lncRNA genes: intergenic (henceforth called "*l*incRNAs") and genic lncRNAs (Figure 1C; Table S1). Genic lncRNAs exhibit significantly greater expression correlation with their nearest coding genes than *l*incRNA/coding pairs and genic coding/coding pairs (Figures 1D and S1A).

The set of genic lncRNAs was further classified into six locus biotypes (Figure 1C; Tables S1 and S2). We first considered lncRNAs transcribed in an antisense direction (designated "X") and designated those that are positioned head-to-head to protein-coding genes as the divergent or "XH" biotype. Antisense lncRNA/coding gene pairs in the tail-to-tail position are designated convergent or "XT." The gene body of an antisense lncRNA can be located within a protein-coding gene (antisense-inside, "XI") or can completely encompass a protein-coding gene (antisense-outside, "XO"). For lncRNAs transcribed in the same direction as the nearest gene (designated "S"), the transcription start site (TSS) of the lncRNA gene can be located (<5 kb) downstream ("SD") or upstream ("SU") of the TSS of the neighboring coding gene.

Divergent XH and antisense-inside XI lncRNAs comprise 19%–27% and 20%–21% of total lncRNAs, respectively, representing the two largest genic lncRNA biotypes in human and mouse genomes (Figure S1B; Tables S1, S2, and S3). Antisense lncRNAs are more likely to be co-expressed with nearby genes than control *l*incRNA/coding pairs and neighboring coding/coding pairs (Wilcoxon $p < 5 \times 10^{-6}$), whereas sense lncRNAs show no obvious difference (Figure S1C).

Divergent lncRNAs Associate with Transcription and Development

Gene Ontology (GO) analysis revealed that protein-coding genes associated with divergent XH lncRNAs are strongly enriched in regulatory functions, including transcription factor activity, pattern specification, and embryonic development (>1.5-fold, $p < 10^{-6}$) (Figures 1E and S1D). The 400 overlapping genes that neighbor divergent lncRNAs in both human and mouse exhibit higher enrichment in these functions than those in each species considered separately (>3-fold, $p < 3 \times 10^{-8}$) (Figure 1F). About 42% (168 genes) of them encode transcription factors and developmental regulators (Table S4). This suggests that many divergent lncRNAs may be conserved at the syntenic level across mammalian species. In comparison, divergent protein-coding/coding genes (XHc/c) are significantly enriched in housekeeping activities (Figure 1E; Table S3).

Mutations of genes neighboring divergent XH and XI lncRNAs are more likely to produce mouse developmental and survival phenotypes than other lncRNA biotypes and divergent coding/coding gene pairs (Figure 1G). The local chromatin environments of divergent lncRNAs in ESCs exhibit strong and specific enrichments of bivalent (H3K27me3 and H3K4me3) and enhancer (H3K27ac and H3K4me1) marks compared with the promoters of lncRNAs from other biotypes (Figure 1H). Strong enrichments of regulatory chromatin marks correlate with enriched development-related functions in nearby coding genes and imply that divergent lncRNAs may be developmentally regulated.

Interestingly, divergent lncRNAs originated relatively early and show a skewed distribution toward older evolutionary ages compared with *l*incRNAs (Figures 1I and S1E–S1G). The mean

Figure 1. Divergent lncRNAs Correlate with Genes That Have Essential Regulatory Functions in Transcription and Development

(A) Expression correlation analysis of genes with their ten nearest genes in 23 human tissues.

(B) A comparison of the observed (obs) versus expected (exp) distributions of genes in the genome. The y axis shows the percentage of coding genes that lies next to a lncRNA gene or a coding gene.

(C) LncRNA classification. Gene numbers in human and mouse are indicated sequentially in parenthesis.

(D) Genic lncRNAs exhibit significantly greater expression correlation with their nearest neighbors ($p < 4 \times 10^{-40}$).

(E and F) GO analysis of coding genes neighboring various biotypes of lncRNAs. Selected GO terms in XH lncRNAs (>1.5-fold and $p < 1 \times 10^{-6}$) in the upper panel of (E) or in the overlapping set of XH lncRNAs in both mouse and human ($p < 3 \times 10^{-8}$) in (F) are shown. Divergent coding/coding genes (XHc/c) serve as a control. The numbers of genes associated with a XH lncRNA or in a particular GO term in the genome are indicated sequentially.

(G) Mammalian phenotype analysis of coding genes neighboring lncRNA biotypes. The heatmap is plotted as (−log10(*p* value)). Darker colors indicate more significant p values.

(H) Chromatin marks in regions surrounding the TSS of various lncRNA biotypes (upper) and coding genes (lower) in human ESCs.

(I) Evolutionary age distributions of human lncRNAs. The x axis shows the age assignment at which a lncRNA first appears.

See also Figure S1 and Tables S1, S2, S3, and S4.

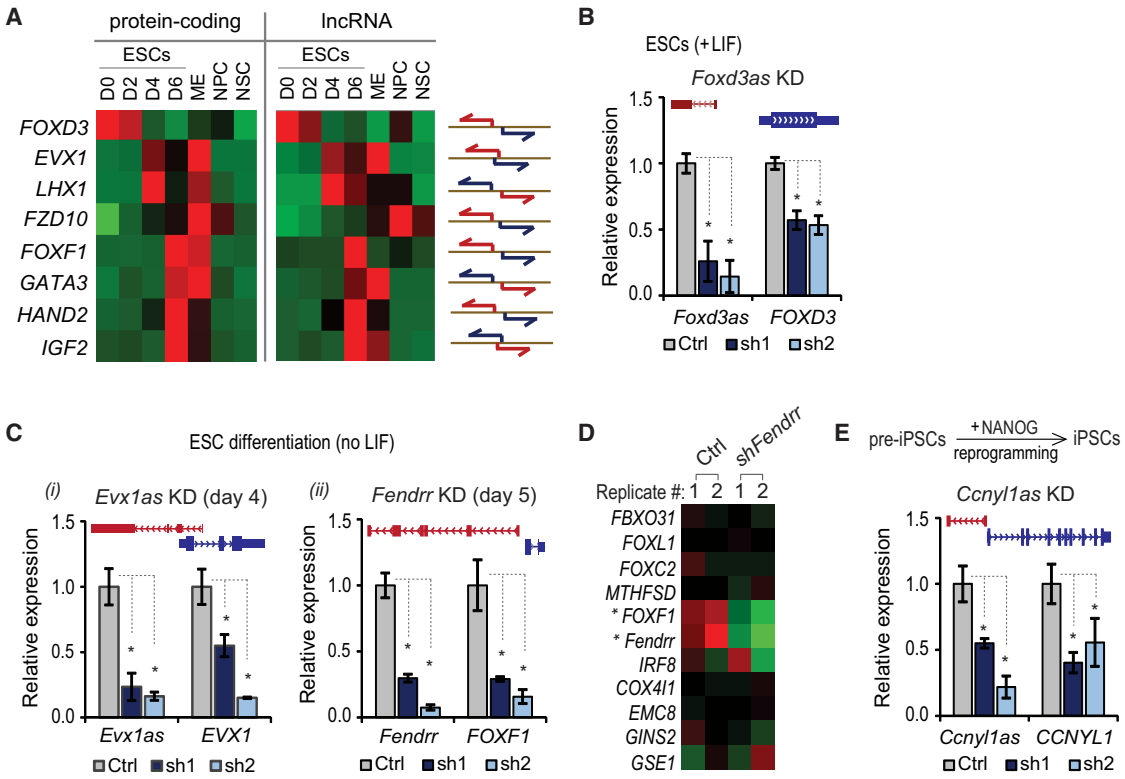


Figure 2. Divergent lncRNAs Regulate Nearby Transcription in Pluripotent Stem Cells

(A) Co-activation of selected divergent lncRNA/coding gene pairs in ESCs on days 0 to 6 (D0 to D6) of LIF withdrawal and in lineage-committed cells. NPCs, neural precursor cells; NSCs, neural stem cells.

(B, C, and E) RT-qPCR analysis of effects of lncRNA knockdown by RNAi on the divergent coding genes in various culture conditions as indicated.

(D) Heatmap of the expression of genes within ±500 kb of the *Fendrr*/*FOXF1* locus upon *Fendrr* RNAi.

Relative positioning of lncRNAs (in red) and protein-coding genes (in blue) is shown in (A)–(C) and (E). The y axis represents relative mean expression normalized to *GADPH* and the scramble shRNA control (Ctrl) cells. Data are shown as mean ± SD (n = 4, including two technical repeats for two independent knockdown). *$p < 0.05$. See also Figure S2 and Table S5.

evolutionary age of divergent lncRNAs is significantly older than lincRNAs (4.8 versus 5.7 in human [Wilcoxon $p < 2 \times 10^{-16}$]; 5.8 versus 6.3 in mouse [Wilcoxon $p < 3 \times 10^{-9}$]), implying that divergent lncRNAs might be functional and maintained by long-term selection.

Divergent lncRNAs Regulate Nearby Transcription in Pluripotent Stem Cells

Pluripotent ESCs have been established as an ideal system to study the fine details of transcriptional and epigenetic regulation during cell-fate switches (Shen et al., 2009). Withdrawal of leukemia inhibitory factor (LIF) induces ESC differentiation into three germ layers, namely mesoderm, endoderm, and neuroectoderm. Mesendoderm (ME) is a transient cell state prior to further differentiation into mesoderm and endoderm. By analyzing RNA-seq data from differentiating ESCs and lineage-committed cells (Yin et al., 2015), we found that several divergent pairs of lncRNAs and protein-coding genes are specifically co-expressed (Figure 2A). To test whether divergent juxtaposition manifests a regulatory interaction, we selected three divergent lncRNAs and studied the effects of lncRNA knockdown by RNAi on nearby transcription.

The divergent lncRNA *Foxd3as* and its neighbor gene *FOXD3*, which encodes a transcription factor critical for early embryonic development and ESC self-renewal, are co-expressed in ESCs but are downregulated upon differentiation (Figure S2A). Interestingly, *Foxd3as* depletion led to a decrease of *FOXD3* mRNA in ESCs (Figure 2B). In addition, during ESC differentiation, knockdown of the divergent lncRNA *Evx1as* attenuated activation of its nearby even-skipped gene, *EVX1* (Figures 2C and S2B). Consistently, knockdown of another divergent lncRNA *Fendrr* led to attenuated activation of its nearest neighbor gene *FOXF1* upon ESC differentiation (Figures 2C and S2A). Importantly, *Fendrr* RNAi failed to affect other genes within ±500 kb of the *Fendrr* locus (Figure 2D; Table S5). In addition, *Fendrr* is specifically correlated with *FOXF1* in expression, but not with other nearby genes across diverse tissues (Figure S2C), arguing against a consequence of chromosome proximity in dictating nearby gene expression.

Moreover, an opposite experimental setting to ESC differentiation is cellular reprogramming to the pluripotent state. Knockdown of the divergent lncRNA *Ccnyl1as* led to decreased expression of *CCNYL1* in induced pluripotent stem cells (iPSCs) during reprogramming (Figure 2E).

Prevalent Transcriptional Regulation of Nearby Genes by Divergent lncRNAs

We then wondered how many randomly selected lncRNAs would have an effect on neighbor genes when knocked down. Many divergent lncRNAs are expressed in ESCs treated with retinoic acid (RA), providing enough candidates for a small-scale RNAi screen. Of 41 randomly selected lncRNAs (Table S5), we successfully knocked down 16 with >50% efficiency (Figures 3A, 3B, S2D, and S2E). Of these 16 lncRNAs, depletion of 10 led to decreased expression of the divergent coding gene, while 6 had no effect, suggesting context-dependent lncRNA regulation.

To ask whether this phenomenon is conserved across species, we studied three divergent lncRNAs in the human breast cancer cell line MCF7, a differentiated mammary cell line distinct from the mouse embryonic cells we analyzed previously. RNAi knockdown of *Gata3as*, *Nbr2*, and *Igf1ras* downregulated their corresponding divergent protein-coding genes, *GATA3*, *BRCA1*, and *IGF1R*, which are known to be involved in tumorigenesis (Figures 3C and S2F).

Finally, we investigated divergent regulation in an in vivo setting during early embryonic development. *Gata6as* and *GATA6* are co-activated at the morula and blastocyst stages of mouse embryos (Figure S2G). Microinjection of siRNAs against *Gata6as* into one-cell or two-cell embryos significantly attenuated transcriptional activation of *GATA6* and decreased the number of cells expressing *GATA6* in blastocysts by RT-qPCR and immunostaining (Figures 3D–3F and S2H).

In summary, among 24 divergent lncRNAs that we successfully knocked down, depletion of 18 (75%) led to downregulation of nearby protein-coding genes. Interestingly, knockdown of all 12 lncRNAs neighboring a transcription factor gene had a negative effect on their coding partners (Figure 3G). These results suggest that positive regulation on nearby transcription by divergent lncRNAs is a prevalent phenomenon in mouse and human. In comparison, among 20 genes in 12 divergent coding/coding pairs tested, only 4 (in three pairs) appeared to regulate nearby mRNA expression upon depletion (Figures 3G, S2I, and S2J).

Evx1as Is Required for Proper Activation of *EVX1*

To determine a true regulatory role of divergent lncRNAs, we characterized the *Evx1as*/*EVX1* locus in depth. Their expression is highly correlated during development (Figures S3A and S3B). In ESCs, both genes are repressed but are activated to peak expression on day 4 of LIF withdrawal and are specifically enriched in ME cells (Figures 4A–4C). RACE analysis revealed two isoforms of *Evx1as*. Both initiate within the first exon of *EVX1* and overlap by 8 nt with the *EVX1* mRNA in day 4 differentiated ESCs, while only the long isoform (~2,788 nt) can be detected by northern blot (Figure 4C).

We showed that *Evx1as* RNAi by seven shRNAs attenuates *EVX1* activation (Figure S2B). To control for possible off-target effects of RNAi, we used an alternative knockdown approach by antisense oligonucleotides (ASOs), which induce RNA degradation by recruiting RNase H to their target RNAs (Wheeler et al., 2012). We assayed the effect of *Evx1as* ASOs while we artificially activated *Evx1as* and *EVX1* by CRISPR-on (Konermann et al., 2015). Co-expression of sgRNAs targeting to the promoter region of *Evx1as*/*EVX1* with the transcription activator dCas9-VP64 increased levels of both transcripts by ~30- to 60-fold in undifferentiated ESCs (Figures 4D and 4E). Consistent with RNAi, *Evx1as* knockdown by ASO attenuated *EVX1* activation during CRISPR-on (Figure 4E). In addition, inhibition of transcription elongation of *Evx1as* by CRISPR interference (CRISPRi) significantly downregulated both pre-mRNA and mRNA levels of *EVX1* (Figures S3C and S3D).

We next inactivated *Evx1as* or *EVX1* through genomic deletion (Figures 4D and S3E–S3I). *EVX1* activation during ESC differentiation or CRISPR-on is significantly blocked or attenuated in *Evx1as* null ESCs (Figures 4F and S3J). Importantly, nuclear run-on revealed that *Evx1as* ASOs, which degrade *Evx1as* RNA transcripts, attenuate the transcription rate of *EVX1* during CRISPR-on (Figures 4G). In addition, the decreased level of *Evx1as* nascent transcripts upon its depletion suggests a positive feedback of *Evx1as* RNA to regulate its own transcription. These results suggest that besides its DNA sequences, *Evx1as* RNA and/or transcription regulate transcriptional activation of *EVX1*. In contrast, *EVX1* knockout or RNAi failed to affect *Evx1as* expression (Figures 4F, S3I, and S3K).

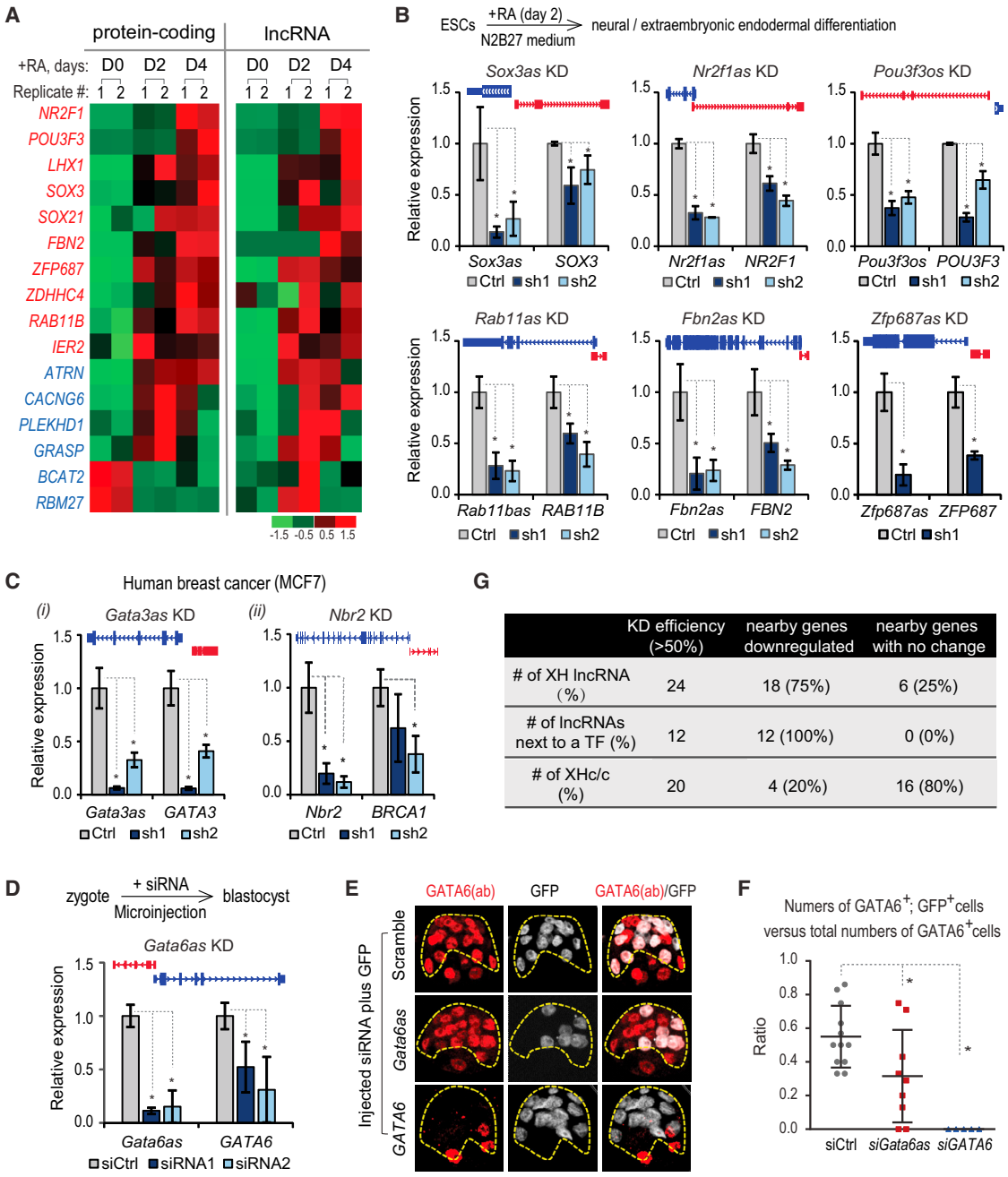
Evx1as Promotes *EVX1* Transcription in *cis*

Next, we studied whether overexpression of *Evx1as* or *EVX1* in ESCs affects the transcription of the other gene. Ectopic expression of *Evx1as* failed to alter *EVX1* mRNA levels (Figure S4A), arguing against *trans* regulation of *EVX1* transcription. To assess the *cis* effect, we knocked in a strong constitutive promoter (CMV early enhancer/chicken β actin [CAG] promoter) upstream of the *Evx1as* or *EVX1* TSS (Figures 4H, S4B, and S4C). Interestingly, CAG-knockin ESCs that overexpress *Evx1as* exhibited a 5-fold upregulation of *EVX1* mRNA (Figures 4I and S4D), suggesting a role of *Evx1as* in *cis* promoting the basal transcription of *EVX1* in ESCs. In comparison, neither *trans* nor *cis* overexpression of *EVX1* affected *Evx1as* transcript levels (Figures 4I and S4A), indicating that *EVX1* is dispensable for *Evx1as* expression.

Considering that genomic alterations may disrupt regulatory DNA in the *Evx1as*/*EVX1* promoters, we sought to guide and tether the *Evx1as* RNA to the *Evx1as*/*EVX1* locus by a CRISPR-mediated strategy (Shechner et al., 2015). We co-transfected a catalytically dead dCas9 with an RNA transcript fused with an sgRNA (Figure 4H). Compared with controls targeting a non-related locus (the TSS of *REX1*), tethering *Evx1as* RNA (both short and long isoforms) to the promoter of *Evx1as*/*EVX1* significantly increased the levels of *EVX1* pre-mRNAs and mRNAs (Figures 4J and S4E). In comparison, the levels of *EVX1* mRNA, but not pre-mRNA, were elevated to a less extent by tethering a reverse sequence of *Evx1as* RNA (Figure 4J). Tethering *GFP*, *HOTTIP*, or *EVX1* RNA to the *Evx1as*/*EVX1* locus failed to increase *EVX1* transcription (Figures S4E and S4F). Tethering *Evx1as* RNA to the *REX1* promoter had no effect on the level of *REX1* mRNA (Figure S4G), indicating context- or sequence-dependent lncRNA regulation. Thus, RNA tethering demonstrates a direct role for *Evx1as* RNA in *cis* regulation of *EVX1* transcription.

Evx1as Binds to Its Own Locus and Promotes Chromatin Looping

Subcellular fractionation detected *Evx1as* RNA present in both cytoplasm and nucleus, and most of the nuclear *Evx1as* is bound



G

	KD efficiency (>50%)	nearby genes downregulated	nearby genes with no change
# of XH lncRNA (%)	24	18 (75%)	6 (25%)
# of lncRNAs next to a TF (%)	12	12 (100%)	0 (0%)
# of XHc/c (%)	20	4 (20%)	16 (80%)

Figure 3. Prevalent Transcriptional Regulation of Nearby Genes by Divergent lncRNAs

(A) Heatmap of the expression of 16 randomly selected divergent lncRNA/coding gene pairs during RA-induced ESC differentiation. Gene names in red or blue indicate that lncRNA knockdown affected or failed to affect the nearby mRNA, respectively.

(B and C) Representative knockdowns for genes (B) shown in (A) or in MCF7 cells (C). The y axis represents relative mean expression normalized to *GADPH* and the scramble shRNA (Ctrl) cells. Data are shown as mean ± SD (n = 4, including two technical repeats for two independent knockdown). *p < 0.05.

(D) *Gata6as* RNAi in mouse zygotes attenuates the activation of *GATA6* in blastocysts (E4.25). The y axis represents relative expression normalized to *TUB4* (n = 4 independent injection). *p < 0.05.

(E and F) *Gata6as* RNAi decreases the number of cells expressing *GATA6* in blastocysts (E3.75). siRNAs were co-injected with *H2B-GFP* mRNA into one cell of two-cell embryos. GATA6 staining is in red and GFP in white; the dotted yellow lines indicate the inner cell mass of blastocyst embryos. The y axis (F) shows the ratio of $GATA6^+$, GFP^+ double-positive cells versus the total number of $GATA6^+$ cells. Each symbol represents one injected embryo. *p < 0.05.

(G) Statistical summary of the lncRNAs and protein-coding genes analyzed by RNAi.

See also Figure S2 and Table S5.

to chromatin (Figure S5A). Consistent with its association on chromatin, RNA fluorescence in situ hybridization (FISH) typically detected one or two nuclear signals from *Evx1as* transcripts (Figures 5A, S5B, and S5C). RNA FISH failed to detect cytosolic *Evx1as*, likely resulting from their diffused presence and low cytosolic concentrations.

To reveal the DNA targets of *Evx1as* transcripts, we performed chromatin isolation by RNA purification (ChIRP) using antisense oligos tiling along the entire *Evx1as* transcript sequence that does not overlap with *EVX1* (Chu et al., 2011) (Figure S5D). Undifferentiated ESCs that do not express *Evx1as* failed to yield any significant DNA peaks, demonstrating the specificity of RNA affinity capture in our ChIRP-seq assay (Figures 5B and S5E).

ChIRP-seq analysis of day 4 differentiated ESCs revealed that *Evx1as* RNA transcripts coat their own gene locus and extend >50 kb downstream of its annotated transcript end site (TES) (Figure S5E). In contrast, polyA and ribominus total RNA-seq show no or few signal reads beyond the annotated TES of *Evx1as* (Figures 4A and S5F). In addition, analysis of chromatin-bound RNA in macrophages failed to detect significant signals beyond cleavage and polyadenylation sites of genes (Bhatt et al., 2012). Thus, it is less likely that *Evx1as* ChIRP-seq captures a rare nuclear, chromatin-bound transcript that goes beyond the annotated boundary of *Evx1as*.

Evx1as ChIRP-seq show signals concentrated around the last exon and the 3′ downstream region that overlaps with a potential enhancer site, which is enriched in DNase I hypersensitivity signals, H3K4me3, H3K4me1, and H3K27ac marks, as well as multiple chromatin and transcription regulators in ESCs (Figures 5B and S5F). While moderate binding observed at the promoters, ChIRP-qPCR confirmed that *Evx1as* RNA, but not *EVX1* mRNA, binds strongly to this 3′ regulatory site (Figure 5C). Its deletion from the genome resulted in attenuated activation of *Evx1as* and *EVX1* for ~2-fold during ESC differentiation (Figures S5G and S5H), suggesting its role as an enhancer in regulating the expression of *Evx1as/EVX1*.

CRISPR-on by sgRNAs targeting to this potential enhancer site increased the levels of both *Evx1as* and *EVX1* transcripts by ~10- to 17-fold in WT ESCs, whereas no obvious increase was observed in ESCs treated with *Evx1as* ASO (Figure 5D). Tethering *Evx1as* RNA to this site moderately increased the level of *EVX1* mRNA (Figure S5I). Thus, chromatin association of *Evx1as* RNA may promote enhancer activity. Interestingly, this potential enhancer site interacts with the promoter of *Evx1as/EVX1* by SMC1 ChIA-PET analysis in ESCs (Dowen et al., 2014) (Figures 5B and S5F). We performed chromatin conformation capture (3C) and found that differentiation further enhances their interaction by 2-fold (Figure 5E). However, this increase was not detected in *Evx1as* null ESCs, suggesting a role of *Evx1as* in modulating the enhancer-promoter interaction (Figure 5E).

Evx1as Facilitates Mediator Binding and an Active Chromatin State

LncRNA-mediated recruitment of epigenetic regulators on chromatin has been described previously (Lai et al., 2013; Trimarchi et al., 2014). MED1 and MED12 are the core and kinase components of Mediator, a multiprotein complex that functions as a transcriptional coactivator, respectively (Malik and Roeder, 2010). Chromatin immunoprecipitation (ChIP) showed that both MED1 and MED12 bind to the promoter and potential enhancer of *Evx1as/EVX1* in ESCs, and differentiation further enhances their chromatin association (Figures 5F and S5J). This increase correlates with activation of *Evx1as/EVX1* at day 4 differentiation; however, it is completely blocked in *Evx1as* null ESCs (Figure 5G). In comparison, CTCF binds to the potential enhancer but not the promoter in an *Evx1as*-independent manner (Figure S5K).

We then wondered whether *Evx1as* might interact with regulatory proteins bound locally on chromatin. We performed RNA immunoprecipitation (RIP) of Mediator, CTCF, WDR5, and EZH2 as well as FLAG-tagged EVX1. Interestingly, only MED1 and MED12 captured *Evx1as* RNA transcripts but not *EVX1* or *T* mRNA in differentiated ESCs (Figure 5H; data not shown). Consistently, biotin-labeled *Evx1as* transcripts but not the control *GFP* RNA captured MED1 and MED12 in vitro (Figure S5L). Moreover, *MED12* knockdown by RNAi resulted in attenuated activation of *Evx1as* and *EVX1* during ESC differentiation (Figure S5M).

Activation of *Evx1as/EVX1* in day 4 differentiation is accompanied by an increased binding of H3K4me3 and H3K27ac at the promoter region (Figures 5I and 5J). However, differentiated *Evx1as* null ESCs failed to increase these active histone modifications despite normal levels of H3K27me3 at the promoter (Figures 5I and 5J; data not shown). In addition, the level of H3K27ac at the potential enhancer site is also increased during differentiation in WT ESCs, but not in *Evx1as* null ESCs (Figure 5I). Based on these results, we propose that chromatin association of *Evx1as* transcripts may facilitate the binding of the Mediator complex to shape a local, active chromatin environment required for activation of *EVX1*.

Evx1as and *EVX1* Show Distinct Expression Dynamics in Single Cells

Next, to distinguish between the cause and consequence of lncRNA expression, we sought to investigate whether *Evx1as* and *EVX1* are differentially regulated at the single-cell level during the time course of ESC differentiation (Figures 6A, S6A, and S6B). During the early stage of ESC differentiation, *Evx1as* exhibits low-level expression in 17%–26% of cells with a median of 17 molecules per cell on day 0 and 23 molecules on day 2 (Figures 6B–6D). On days 3 and 4, 55% to 62% of cells express *Evx1as* with 56 and 65 transcripts per cell, respectively. In comparison, *EVX1* RNA transcripts are not detected on days 0 and 2 but reach a median of 326 molecules per cell in ~14% cells on day 3, suggesting a rapid and synchronous activation of *EVX1*. On day 4, 19% of cells express *EVX1*, with 411 transcripts per cell. On day 6, only 4% of cells express *EVX1* with 614 transcripts per cell, whereas 59% of cells still express *Evx1as* with 44 transcripts per cell. Thus, *Evx1as* exhibits a gradual increase in transcript abundance and an early, broad, yet low-level expression pattern during ESC differentiation, whereas *EVX1* exhibits a burst of high-level transcription in a relatively confined population of cells.

Interestingly, most of the *EVX1*-expressing cells co-express *Evx1as* (Figure 6C). *Evx1as* expression accompanies the time course of *EVX1* activation and deactivation, suggesting that *Evx1as* modulates both the extent and kinetics of *EVX1* expression. On the other hand, many *Evx1as*-expressing cells do not

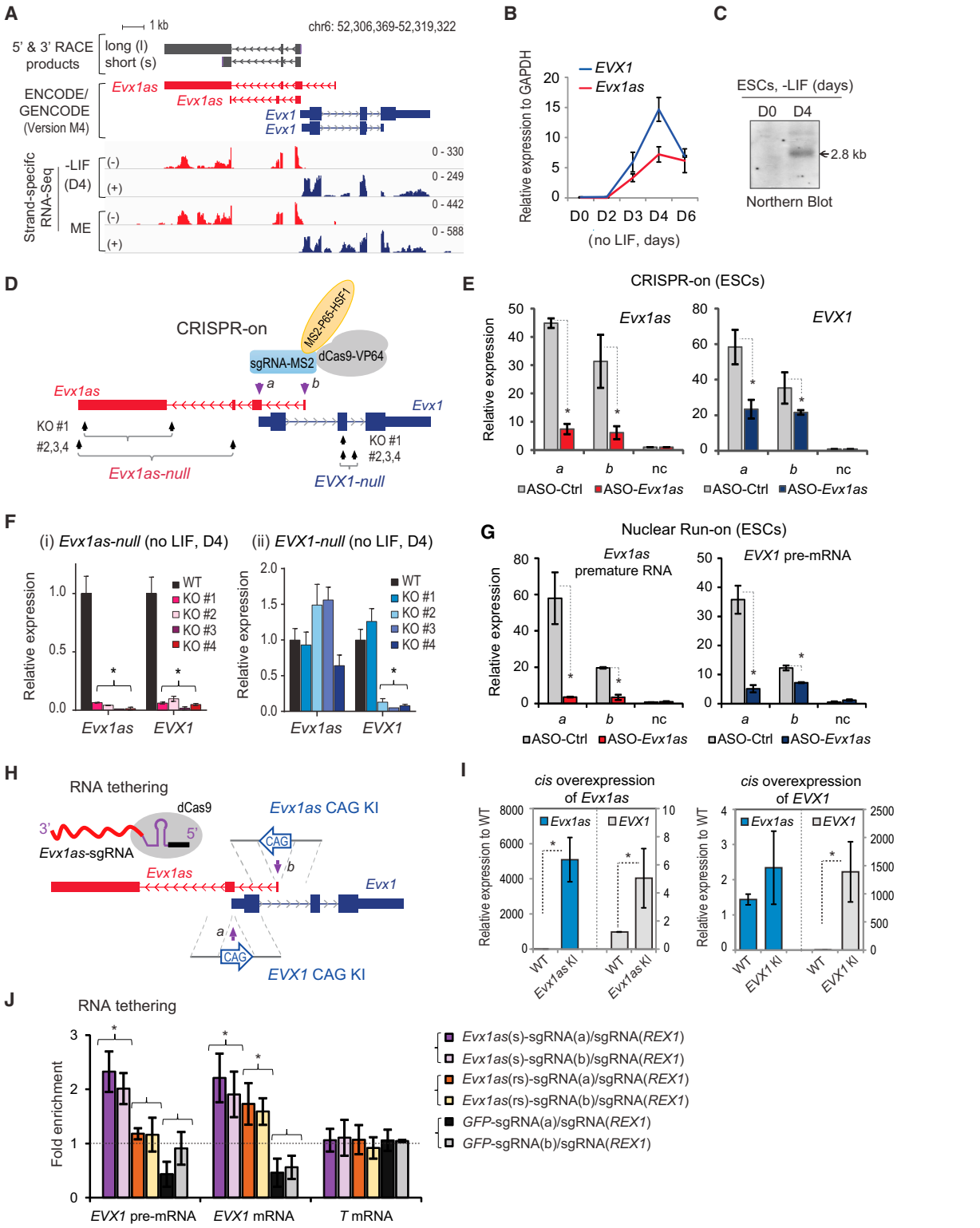


Figure 4. *Evx1as* Promotes *EVX1* Transcription in *cis*

(A) The *Evx1as/EVX1* gene locus. Two *Evx1as* isoforms are detected by rapid amplification of cDNA ends (RACE).

(B) Co-activation of *Evx1as* and *EVX1* during ESC differentiation induced by LIF withdrawal.

(C) Northern blot detected the long isoform of *Evx1as*.

(D) Schematic diagram of CRISPR/Cas9-mediated activation (CRISPR-on) and knockouts (KO). Arrowheads indicate relative locations of sgRNAs.

(legend continued on next page)

express *EVX1*, indicating that *Evx1as* alone may not suffice for *EVX1* activation. Alternatively, *Evx1as* might inhibit *EVX1* transcription in some contexts that do not need *EVX1* expression. Nevertheless, the distinct expression profiles revealed by single-cell time-course analysis support the notion that *Evx1as* functions upstream to facilitate rapid, high-level transcription of *EVX1* when required.

Evx1as and *EVX1* Help to Regulate Mesendodermal Differentiation

In vitro differentiation of ESCs induced by LIF withdrawal resembles gastrulation of the early post-implantation embryo (Keller, 2005). ME cells in vivo are transiently present in the primitive streak of early embryo prior to further differentiation into mesoderm/posterior streak and endoderm/anterior streak (Tam and Loebel, 2007). EVX1, a homeodomain transcription factor, is a concentration-dependent repressor (Briata et al., 1995; Dush and Martin, 1992). It promotes mesoderm differentiation by inhibiting *GSC*, an endoderm/anterior streak gene (Kalisz et al., 2012).

Interestingly, *Evx1as* RNAi blocked activation of all ME-related marker genes analyzed, including *T*, *GSC*, *EOMES*, and *SOX17* (Figure 7A). Consistently, RNA-seq and gene set enrichment analysis (GSEA) revealed global underrepresentation of ME genes in day 4 differentiated *Evx1as* null and knockdown ESCs compared with the WT control (Figures 7B–7D and S6C–S6E; Table S6). This result indicates a role of *Evx1as* RNA and its DNA locus in regulating ESC differentiation. GO analysis showed that downregulated genes are significantly enriched in terms related to mesendodermal development (Figure 7C). In comparison, *EVX1* null and knockdown cells exhibit moderate decreases in expression of a subset of ME and mesoderm/posterior streak genes, including *T* and *WNT5a*, but aberrant upregulation of markers known to be expressed in cells toward the anterior of the streak (*GSC*, *LHX1*, *CXCR4*, *EOMES*) and definitive endoderm (*SOX17* and *FOXA2*) (Figures 7D and S6E–S6G; Table S6). Expression changes revealed by RNA-seq were confirmed by analysis of marker genes (Figures 7E and 7F).

Moreover, consistent with the failure to fully activate ME differentiation, *Evx1as* null ESCs show increased expression of a subset of neural genes and pluripotency genes, whereas only moderate increases were observed in *EVX1* null ESCs (Figures 7B and 7D). In comparison, overexpression of *Evx1as in trans* had no effect on the expression of differentiation genes (Figure S6H), suggesting that the role of *Evx1as* in lineage differentiation likely results from its *cis*-regulatory function and may be partly mediated through *EVX1*. As mesoderm and endoderm development are tightly coupled, the divergent *Evx1as/EVX1* gene locus helps to regulate mesoderm or endoderm fate choices through coordinated expression of *Evx1as* and *EVX1*, leading to orchestrated lineage differentiation of ESCs (Figure 7G).

DISCUSSION

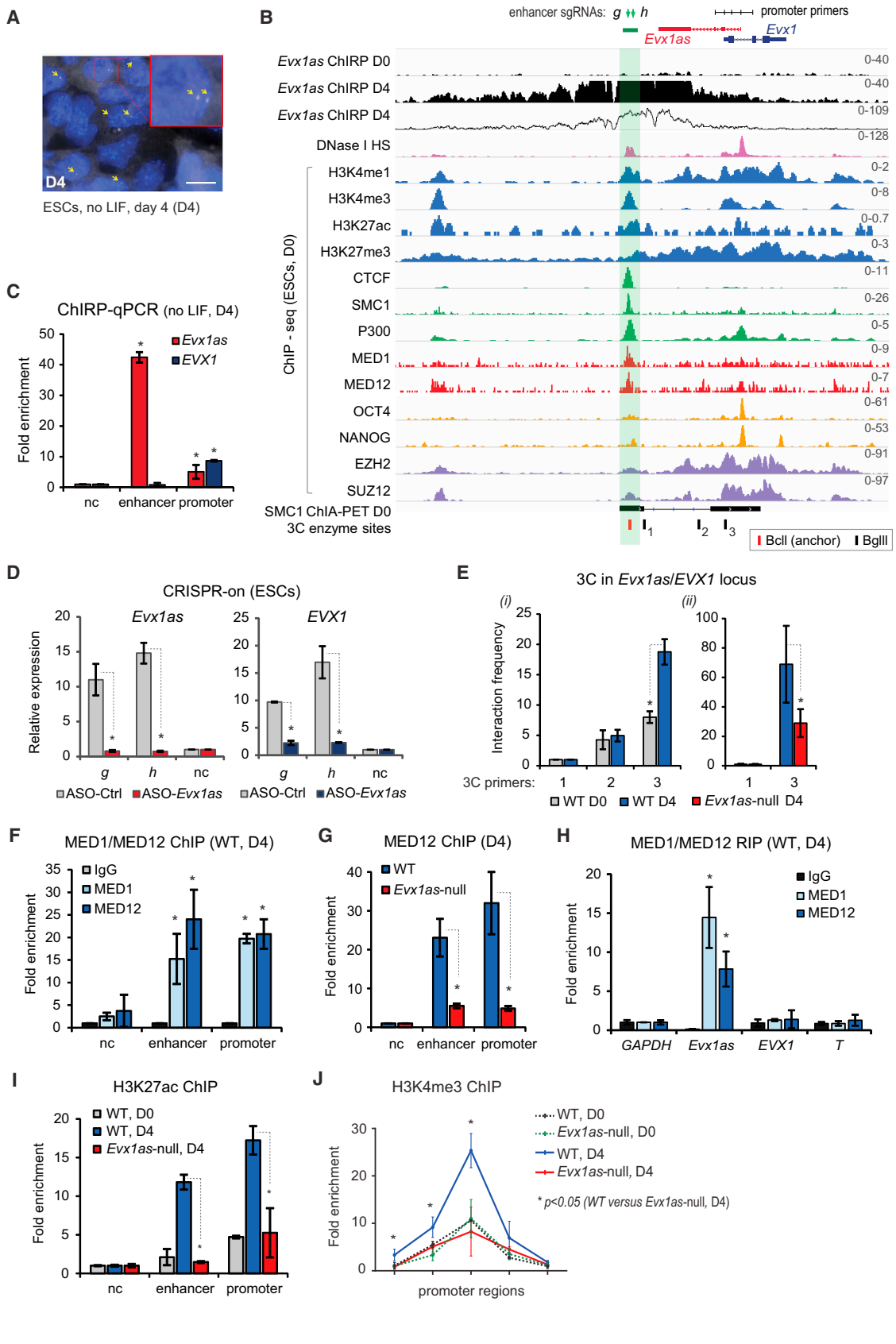
Proteins are believed to be the major functional executors in cells and organisms. It is intriguing how their expression might be regulated by the noncoding portions of the genome, including thousands of lncRNA transcripts, resulting in greater morphological diversity in higher eukaryotes. A non-random distribution of lncRNAs in the genome suggests that locus classification is an effective first step toward a genome-wide understanding of RNA-mediated gene regulation. To unravel the functional linkage between lncRNAs and nearby coding genes, we focused on one class of lncRNA—the divergent biotype. These lncRNAs are particularly interesting because (1) they comprise a significant proportion of all lncRNA genes in mammalian genomes, (2) they tend to co-localize and co-express with developmental and transcription regulator genes, (3) they associate with regulatory epigenetic marks, and (4) they have deeper evolutionary origins than intergenic lincRNAs. However, the prevalence of divergent lncRNA-mediated transcription regulation has been underappreciated, and the functions and mechanisms of action of divergent lncRNAs are not fully understood.

Divergent lncRNAs Mediate Genuine *cis* Regulation of Nearby Transcription

The evolutionary origins and maturation of divergent lncRNAs suggest that they are advantageous to organisms and have thus become fixed in populations. In yeast and bacteria, genes that must respond in a switch-like manner, such as stress-response and environment-specific genes, are enriched for antisense expression (Qi and Arkin, 2014; Xu et al., 2011). In metazoans with compact genomes, such as *C. elegans* and *Drosophila*, 64% to 71% of lncRNAs are positioned divergently to protein-coding genes (Table S1). These neighbor genes are enriched in functions related to morphogenesis, transcription, chromatin organization, and locomotion (data not shown). Thus, divergent lncRNA/coding gene organization tends to be ancient and conserved, reflecting selection to preserve its functionality. It has been hypothesized that divergent lncRNAs might represent an evolutionary intermediate between upstream antisense RNAs and protein-coding genes (Wu and Sharp, 2013).

The genomic loci and flanking regions of developmental and transcription factor genes tend to be replete with conserved noncoding sequences (Ponjavic et al., 2009; Woolfe et al., 2005). Some divergent lncRNA loci, including *Fendrr* and *Mdgt*, appear to be essential for animal development and survival (Grote et al.,

(E and G) Effects of *Evx1as* knockdown by ASO on steady-state RNA levels (E) or nascent transcripts in nuclear run-on (G) during CRISPR-on. The sgRNAs *a* and *b* target the promoters of *Evx1as/EVX1*, and "nc" targets to a random sequence.
(F) Analysis of knockout ESCs. Four independent clones are shown for each gene.
(H) Schematic diagram of CAG knockin (KI) and RNA tethering.
(I) Analysis of CAG knockin ESCs. The y axis represents the fold change compared with WT ESCs. n = 4 independent knockin ESC clones.
(J) The effect of tethering *Evx1as* transcripts to the *Evx1as/EVX1* promoters. *Evx1as* (a short isoform [s]; or a reverse short isoform [rs]) or *GFP* RNA fused with either of sgRNA *a* and *b* or with an sgRNA targeting the TSS of *REX1*. The y axis represents the fold enrichment of RNA tethering to the *Evx1as/EVX1* promoters normalized to the corresponding RNA tethering to the *REX1* TSS (*$p < 0.05$).
In (B), (E)–(G), (I), and (J), RT-qPCR data are shown as mean ± SD (n = 3 biological replicates unless otherwise indicated). *$p < 0.05$ compared with the controls. See also Figures S3 and S4.



(legend on next page)

2013; Sauvageau et al., 2013). One might propose that the function of divergent lncRNAs may result from its sequence overlap with DNA motifs shared with nearby genes or reflect coupled transcription across neighboring loci because they are subject to common regulatory sequences and local chromatin features (Ebisuya et al., 2008). However, we discounted this notion for the following reasons.

First, the fact that genic lncRNA/coding gene pairs exhibited significantly higher expression correlation than genic coding/coding pairs argues against a simple proximity effect. Second, *Fendrr* knockdown specifically affected its divergently positioned gene *FOXF1*, but not other nearby genes. Third, *EVX1* appears to be dispensable for *Evx1as* expression. Fourth, depletion of *Evx1as* transcripts by loss-of-function approaches without altering genomic sequences, including CRISPRi, RNAi, and ASOs, led to downregulation of *EVX1*. On the other hand, tethering of *Evx1as* RNA alone to the *Evx1as*/*EVX1* locus increased the basal transcription of *EVX1.* Fifth, *Evx1as* transcripts bind to their own locus on chromatin and interact with Mediator, providing a mechanistic evidence for lncRNA-mediated *cis* regulation. Sixth, differences in expression levels and activation kinetics between *Evx1as* and *EVX1* in single cells support the hypothesis that lncRNAs transcriptionally regulate the divergent locus. Finally, compared with 75% of lncRNAs exhibiting a *cis* regulatory effect, only ~20%–25% of divergent coding/coding genes analyzed appear to have an effect on nearby transcription upon depletion. Although we cannot rule out the idea that divergent coding transcripts have a regulatory role, it appears more common for the divergent lncRNA/coding pairs than the coding/coding pairs, at least in the subset that we studied. Rare cases have been reported that the mRNA of a divergent protein-coding gene, *Wrap53*, functions as a noncoding transcript to regulate nearby p53 gene transcription (Mahmoudi et al., 2009; Saldaña-Meyer et al., 2014). Thus, functional interaction between nearby lncRNAs and protein-coding genes likely reflects genuine *cis* regulation by lncRNAs rather than being a simple consequence of transcriptional coupling due to proximity.

Mechanistic Investigation of *Evx1as* Function on *EVX1* Transcription

While coating chromatin, *Evx1as* RNA appears to simultaneously assemble relevant chromatin complexes and promote chromatin looping. Our results support a hypothesized role for divergent transcripts in signaling or guiding chromatin complexes to shape local chromatin state and structure. Despite its low-level expression, the *cis* tethering of *Evx1as* RNA to transcription sites suggests that local, effective concentration of divergent RNA transcripts may be sufficient to modulate expression of both the lncRNA and its nearest neighbor. Thus, divergent lncRNAs may provide another layer of transcription regulation in addition to promoters, enhancers, and terminators. Compared with *cis*-regulatory DNA elements, lncRNA transcripts that lack coding potential are more flexible, mobile, and transient, thus providing a convenient means to precisely regulate nearby gene expression in a site-specific manner.

Early, broad expression of *Evx1as* in single cells suggests that *Evx1as* may have a "window of opportunity" in which to integrate multiple regulatory signals and to prime a permissive yet poised chromatin and/or transcription state, allowing for rapid activation of nearby *EVX1* in response to induction signals. Interestingly, *EVX1* exhibits an abrupt, synchronous activation in day 3 differentiated ESCs, which correlates with the early induction of ME. Synchronous activation of *EVX1* has also been identified in mouse pre-gastrulation embryos (Dush and Martin, 1992). *EVX1* transcripts are not detected at embryonic day E6.0 but are present a short time later at approximately E6.25 in a localized region of epiblast cells that will soon be found in the primitive streak. The importance of transcription synchrony has been demonstrated in fly (Lagha et al., 2013). Replacement of the strongly paused *snail* promoter with non-paused promoters causes stochastic activation of *snail* expression and disrupted mesoderm invagination during fly morphogenesis. We speculate that synchronous activation of *EVX1* mediated by *Evx1as* may be similarly required for coordinated cell behavior during ME induction in vitro and in vivo.

Evx1as Modulates ME Differentiation

The fact that both *Evx1as* and *EVX1* promote ME differentiation toward mesoderm/posterior streak fates is consistent with a regulatory role of *Evx1as* on *EVX1* transcription. Intriguingly, *Evx1as* null ESCs show severe downregulation of ME, mesodermal, and endodermal genes, whereas *EVX1* null ESCs show a modest decrease in few ME genes but significant upregulation of endodermal and anterior streak genes. A convenient explanation is that *Evx1as* might have functions beyond controlling *EVX1* expression, as supported by the fact that a significant proportion of *Evx1as* is found present in cytoplasm by cell fractionation. However, other protein substrates of *Evx1as* await further analysis if they exist. In addition, ChIRP-seq did not reveal additional

Figure 5. *Evx1as* Modulates Local Chromatin State and Configuration

(A) *Evx1as* RNA FISH in day 4 differentiated ESCs. Scale bar represents 10 μm.

(B) The *Evx1as*/*EVX1* locus in genome browser. The upper tracks show normalized read densities of *Evx1as* ChIRP-seq in day 0 and day 4 differentiated ESCs. Tracks below show signals of DNase I hypersensitivity, ChIP-seq, and SMC1 ChIA-PET in ESCs. The potential enhancer is boxed in light green.

(C) Chromatin association of *Evx1as* or *EVX1* RNA. The y axis shows fold enrichment in day 4 ESCs to an unrelated region ("nc," primers *CSa*) and ChIRP signals in day 0 ESCs.

(D) Effects of *Evx1as* knockdown by ASO during CRISPR-on at the potential enhancer in ESCs. Relative locations of sgRNAs *g* and *h* are shown as arrowheads in (B).

(E) 3C analysis in WT and *Evx1as* null ESCs on days 0 and 4 of LIF withdrawal. Relative locations of 3C primers upstream of each *BclI* or BglII site (vertical lines) are shown at the bottom of (B).

(F, G, I, and J) ChIP-qPCR of Mediator (F), MED12 (G), H3K27ac (I), and H3K4me3 (J). The y axis shows fold enrichment normalized to the "nc" (*CSa*) and the input. In (J), promoter primers are shown sequentially on the top of (B).

(H) RIP-qPCR of Mediator. The y axis shows fold enrichment relative to the input and *GAPDH* mRNA.

In (C)–(J), data are shown as mean ± SD (n = 3 biological replicates unless otherwise indicated). *$p < 0.05$. See also Figure S5.

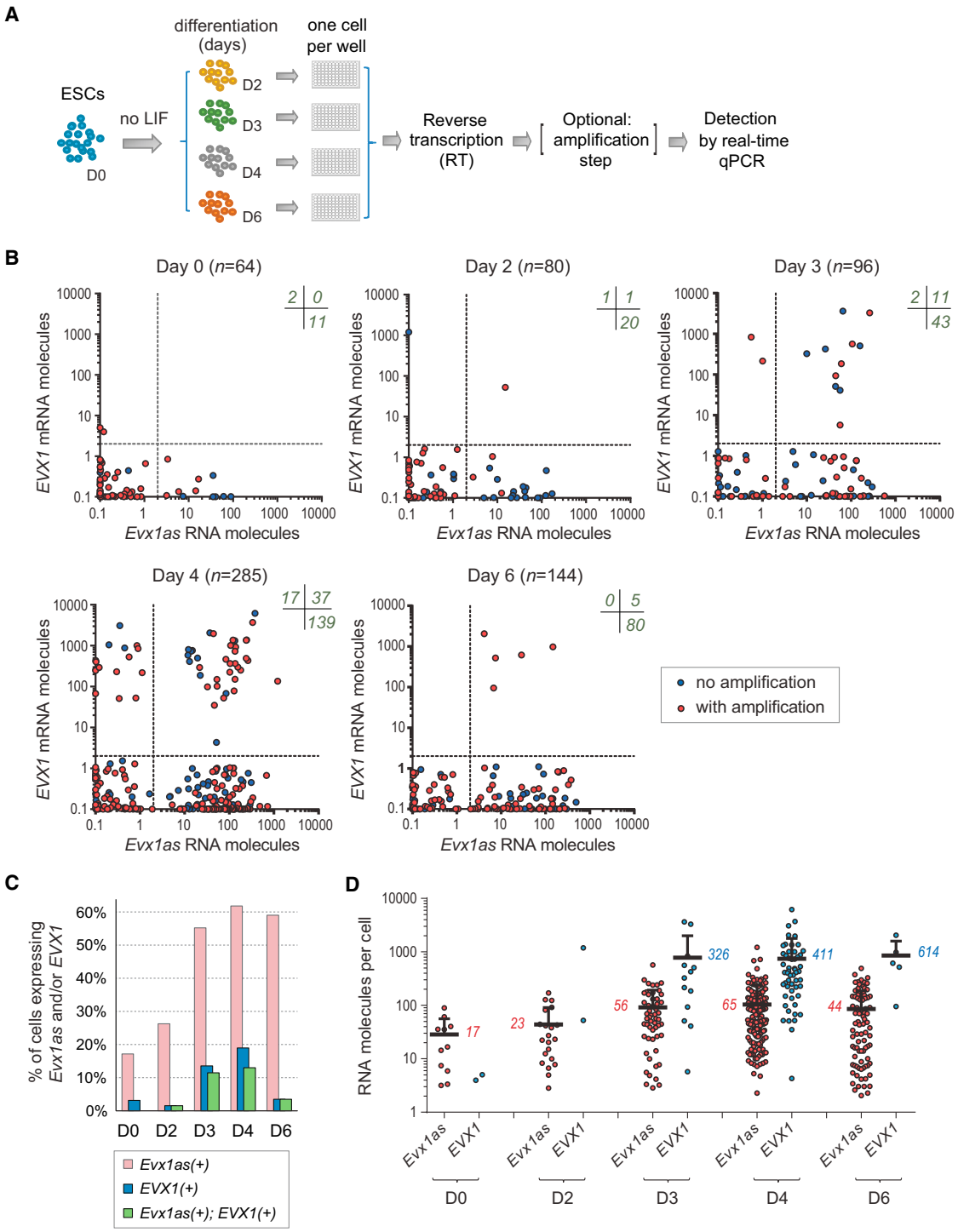


Figure 6. Single-Cell Time-Course Analysis of *Evx1as* and *EVX1* Expression during ESC Differentiation

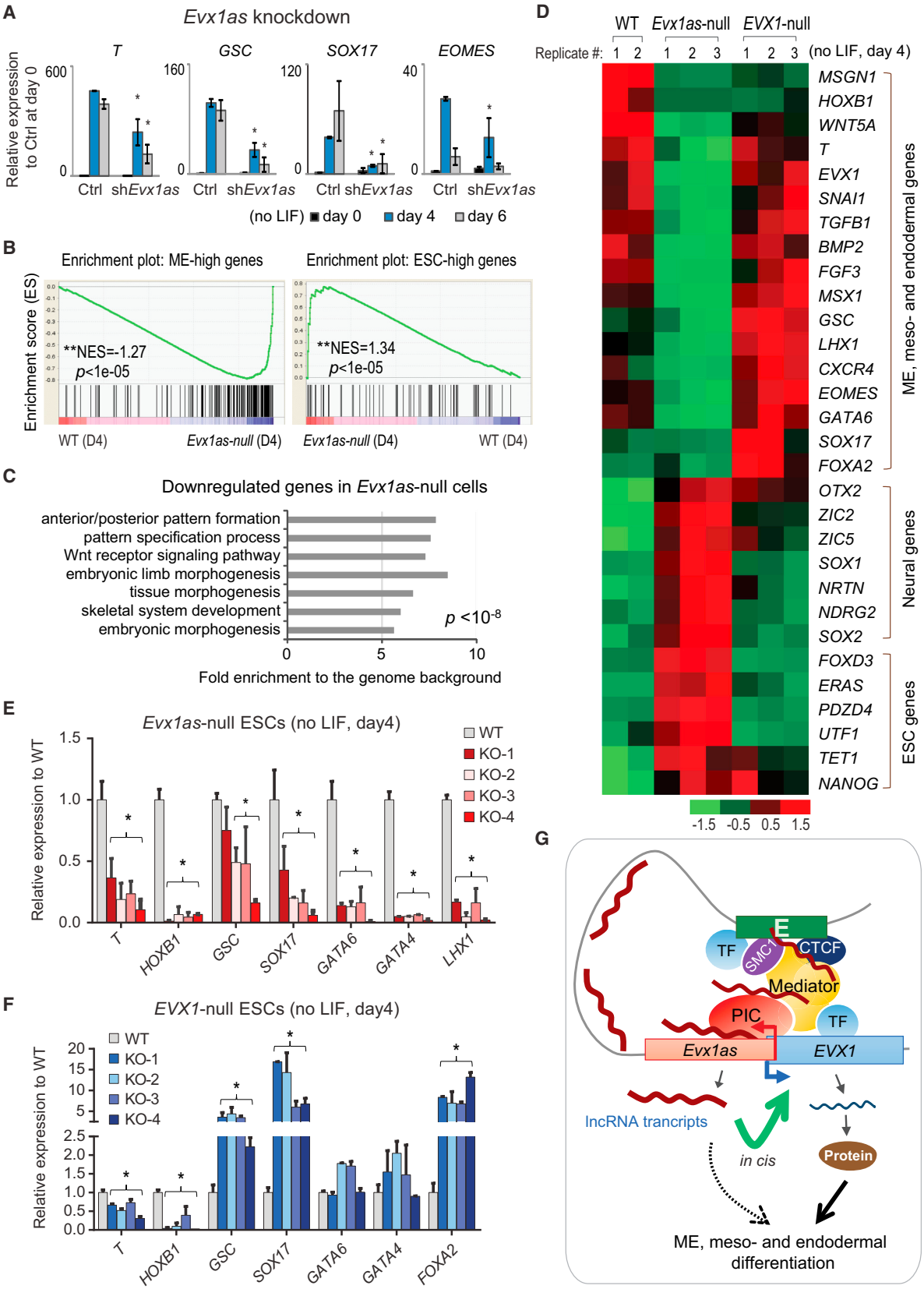
(A) The scheme.

(B) Single-cell scatter plots showing the abundance of *Evx1as* and *EVX1* RNA. Dotted lines indicate thresholds set at more than two RNA molecules per cell. The total number (*n*) of cells analyzed is shown on the top. The inset tables show the number of cells in three of the four quadrants delineated by the thresholds. Cells subjected to a preamplification or without amplification before qPCR detection are represented by red or blue dots, respectively.

(C) The percentages of cells expressing *Evx1as* and/or *EVX1* during differentiation.

(D) Plot of the number of *Evx1as* and *EVX1* molecules per cell during differentiation. Cells expressing *Evx1as* and *EVX1* are shown by red and blue dots, respectively. Median numbers of RNA molecules per cell are indicated only if a cell population contained five or more cells at each time point.

See also Figure S6.



(legend on next page)

genomic targets of *Evx1as*, ruling out its *trans* activity in regulating different target gene(s) on chromatin.

Alternatively, the precise spatiotemporal expression of *EVX1* may be required for ME-related differentiation. In *Drosophila*, the pair-rule segmentation gene *even skipped* (*eve*), the fly homolog of *EVX1*, acts as a concentration-dependent morphogen repressing different genes at different concentrations in different locations (Fujioka et al., 1995; Jaynes and Fujioka, 2004). Although it is specifically expressed in odd-numbered stripes, *eve* null alleles completely abolish all segmentation, while weak *eve* mutations cause deletions of alternate segments (Manoukian and Krause, 1992).

Knockout of *EVX1* in ESCs likely affects only a small subset of cells co-expressing *Evx1as* and *EVX1*. In contrast, deletion of *Evx1as*, the upstream regulator of *EVX1*, may affect a broader population of cells, resulting in stochastic and low-level expression of *EVX1* in cells that should either express *EVX1* at high levels or not express *EVX1*. Dysregulated expression of *EVX1* in *Evx1as* null ESCs may elicit a severe, pleiotropic effect on mesendodermal cells and the subsequent differentiation of mesoderm and endoderm lineages than a simple, complete loss of the *EVX1* gene in *EVX1* null ESCs. Consistent with the *cis*-regulatory role of *Evx1as* in controlling *EVX1* transcription, this model supports the notion that the overall level and extent of *EVX1* expression must be tightly regulated by *Evx1as* and reinforces the importance of lncRNA-mediated transcriptional control to ensure that genes are expressed in the right amounts at the correct times in cell populations.

Prevalence and Functional Inference of *cis*-Regulatory lncRNAs

In diverse in vitro and in vivo contexts we tested, including pluripotency maintenance, lineage differentiation, reprogramming, human cancer, and zygotic development, lncRNA depletion caused decreased expression of nearby genes in most cases. Cases of transcriptional inhibition by divergent lncRNAs have been reported (Ariel et al., 2014; Han et al., 2014; Latos et al., 2012). The outcome of lncRNA-mediated control can be activation or silencing dependent on the biological context required for the function of a lncRNA. We propose that divergent lncRNA-mediated transcription regulation of nearby genes may represent a common mechanism that is utilized to finely tune the spatiotemporal expression of pleiotropic developmental loci, thus contributing in part to the increased phenotypic complexity of higher eukaryotes (Figure 7G).

Evidence shows that lncRNAs as a class are preferentially located in the chromatin and nucleus of the cell when compared with protein-coding mRNAs (Derrien et al., 2012). Analysis of chromatin-associated RNAs suggested that lncRNAs, as an integral component of chromatin, may regulate various biological functions through fine-tuning the chromatin architecture (Mondal et al., 2010). We suspect that lncRNA-mediated *cis* regulation is unlikely to be limited to the divergent lncRNA biotype and might be prevalent among protein-coding genes with a nearby lncRNA. Lack of *Evx1as* causes global defects in activating ME-related differentiation programs known to involve nearby EVX1, suggesting that divergent lncRNAs may participate in biological processes similar to those controlled by the nearby protein-coding genes. From this point of view, the functionality of many uncharacterized lncRNAs can be rapidly predicted from the function of their neighboring protein-coding genes. We believe that this functional inference will help to generate meaningful hypothesis and better experimental designs when investigating lncRNA transcripts whose functions are largely unknown.

EXPERIMENTAL PROCEDURES

Additional experimental procedures and details are provided in the Supplemental Experimental Procedures. DNA sequences for primers, shRNAs, siRNAs, ASOs, sgRNA, and ChIRP probes are listed in Table S7.

LncRNA Classification

LncRNAs were classified into locus biotypes based on their transcription orientation and the positions of their transcription start and end sites with respect to nearby protein-coding loci (5-kb distance cutoff).

Bioinformatics Analysis

The observed fraction of protein-coding genes that are located in a defined genomic distance from a neighboring lncRNA or coding gene was compared with simulated distributions by random positioning. Pairwise Pearson correlation analysis of each lncRNA with its nearest ten genes was performed across 23 human tissues. GO and phenotype terms with $p < 1 \times 10^{-6}$ were considered significant. In evolutionary age analysis, we dated lncRNA genes on the vertebrate phylogenetic tree by following a previous strategy (Zhang et al., 2010). We filtered out sequences overlapping with protein-coding exons to avoid bias caused by neighboring genes. All RNA-seq and ChIP-seq datasets used in this study are listed in Table S8.

Cell Culture and Knockdown Analysis

ESC differentiation, reprogramming, and knockdown are described in the Supplemental Experimental Procedures. *Evx1as* ASOs were transfected to cells carrying CRISPR-on transfection. siRNAs were injected into mouse one-cell

Figure 7. *Evx1as* and *EVX1* Are Required for Mesendodermal Differentiation

(A) Marker gene analysis of *Evx1as* knockdown by RNAi during ESC differentiation.

(B) *Evx1as* null ESCs show global downregulation of ME genes and upregulation of ESC genes by GSEA. Normalized enrichment score (NES) and nominal p values are shown.

(C) GO analysis of downregulated genes in day 4 differentiated *Evx1as* null cells.

(D) Heatmap of the expression of representative genes on day 4 ESC differentiation.

(E and F) RT-qPCR analysis of *Evx1as* null (E) and *EVX1*-null (F) ESCs on day 4 differentiation. The y axis shows expression relative to *GADPH* and the control. Data are shown as mean ± SD (n = 3 biological replicates). *$p < 0.05$ compared with the control.

(G) A *cis*-regulatory model. *Evx1as* RNA stays attached on chromatin to its site of transcription and the downstream regulatory region and locally facilitates the binding of the transcription coactivator Mediator, which may help assembly of relevant chromatin and transcription machineries, thereby promoting an open chromatin configuration and contacts at the shared promoter and enhancer ("E") that are required for efficient transcription of *EVX1*. The role of *Evx1as* in ESC differentiation may be partly mediated through EVX1 and/or through functions beyond controlling *EVX1* expression (indicated by the dotted arrow). Nevertheless, the entire *Evx1as*/*EVX1* gene locus is required for orchestrated lineage differentiation of ESCs. TF, transcription factor; PIC, the preinitiation complex.

See also Figure S6 and Table S6.

or two-cell embryos. Mouse work follows the animal ethics rules at Tsinghua University.

CRISPR/Cas9-Mediated Genome Editing, Activation, and RNA Tethering

CRISPR/Cas9-mediated knockout and knockin, CRISPR-on, and RNA tethering were performed as previously described (Konermann et al., 2015; Shechner et al., 2015; Yin et al., 2015).

Nuclear Run-On, ChIRP, and 3C

Nuclear run-on, ChIRP, and the 3C analysis were performed as previously described (Patrone et al., 2000; Yin et al., 2015).

Single-Cell Analysis, Northern Blotting, and RNA FISH

Single-cell analysis was performed as described previously (Tang et al., 2010). A threshold of two RNA molecules per cell and five or more cells expressing more than two molecules of the corresponding RNA was chosen for calculating the median expression levels. About 1 μg of enriched polyA+ RNA was loaded per lane in Northern blot analysis. In FISH, a total of 48 probes labeled with Quasar570 were used to target *Evx1as* RNA (Stellaris).

ACCESSION NUMBERS

The accession numbers for ChIRP-seq and RNA-seq reported in this paper are GEO: GSE70420, GEO: GSE70419, GEO: GSE62899, and GEO: GSE58514.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, and eight tables and can be found with this article online at http://dx.doi.org/10.1016/j.stem.2016.01.024.

AUTHOR CONTRIBUTIONS

X.S. conceived of and supervised the study. X.S. and S.L. designed the experiments. S.L., J.Y.L., L.L., Y.Y., X.H., B.W., R.X., W.L., P.Y., and W.S. performed experiments. J.Y.L. performed bioinformatics analyses. C.C. and Y.E.Z. performed evolutionary analysis. Z.L., H.L., J.N., F.T., and J.W. contributed reagents/technical assistance/suggestions. X.S. wrote the manuscript with help from S.L. and J.Y.L.

ACKNOWLEDGMENTS

We thank J. Silva for the pre-iPSCs. We thank Y. Chen, L. Yu, Y. Qi, Q. Xi, D. Li, X. Fu, T. Xie, R.G. Roeder, and Shen Laboratory members for insightful discussion and critical reading. Grant support is from the National Basic Research Program of China (2012CB966703), the National Natural Science Foundation of China (31471219, 8141101062, and 31428010), and the Center for Life Sciences (CLS) at Tsinghua University.

Received: October 20, 2015
Revised: December 17, 2015
Accepted: January 29, 2016
Published: March 17, 2016

REFERENCES

Ariel, F., Jegu, T., Latrasse, D., Romero-Barrios, N., Christ, A., Benhamed, M., and Crespi, M. (2014). Noncoding transcription by alternative RNA polymerases dynamically regulates an auxin-driven chromatin loop. Mol. Cell *55*, 383–396.

Batista, P.J., and Chang, H.Y. (2013). Long noncoding RNAs: cellular address codes in development and disease. Cell *152*, 1298–1307.

Bhatt, D.M., Pandya-Jones, A., Tong, A.J., Barozzi, I., Lissner, M.M., Natoli, G., Black, D.L., and Smale, S.T. (2012). Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. Cell *150*, 279–290.

Briata, P., Van De Werken, R., Airoldi, I., Ilengo, C., Di Blas, E., Boncinelli, E., and Corte, G. (1995). Transcriptional repression by the human homeobox protein EVX1 in transfected mammalian cells. J. Biol. Chem. *270*, 27695–27701.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. *25*, 1915–1927.

Chu, C., Qu, K., Zhong, F.L., Artandi, S.E., and Chang, H.Y. (2011). Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol. Cell *44*, 667–678.

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., 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.

Dinger, M.E., Amaral, P.P., Mercer, T.R., Pang, K.C., Bruce, S.J., Gardiner, B.B., Askarian-Amiri, M.E., Ru, K., Soldà, G., Simons, C., et al. (2008). Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. Genome Res. *18*, 1433–1445.

Dowen, J.M., Fan, Z.P., Hnisz, D., Ren, G., Abraham, B.J., Zhang, L.N., Weintraub, A.S., Schuijers, J., Lee, T.I., Zhao, K., and Young, R.A. (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell *159*, 374–387.

Dush, M.K., and Martin, G.R. (1992). Analysis of mouse Evx genes: Evx-1 displays graded expression in the primitive streak. Dev. Biol. *151*, 273–287.

Ebisuya, M., Yamamoto, T., Nakajima, M., and Nishida, E. (2008). Ripples from neighbouring transcription. Nat. Cell Biol. *10*, 1106–1113.

Fujioka, M., Jaynes, J.B., and Goto, T. (1995). Early even-skipped stripes act as morphogenetic gradients at the single cell level to establish engrailed expression. Development *121*, 4371–4382.

Grote, P., Wittler, L., Hendrix, D., Koch, F., Währisch, S., Beisaw, A., Macura, K., Bläss, G., Kellis, M., Werber, M., and Herrmann, B.G. (2013). The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. Dev. Cell *24*, 206–214.

Han, P., Li, W., Lin, C.H., Yang, J., Shang, C., Nurnberg, S.T., Jin, K.K., Xu, W., Lin, C.Y., Lin, C.J., et al. (2014). A long noncoding RNA protects the heart from pathological hypertrophy. Nature *514*, 102–106.

Herriges, M.J., Swarr, D.T., Morley, M.P., Rathi, K.S., Peng, T., Stewart, K.M., and Morrisey, E.E. (2014). Long noncoding RNAs are spatially correlated with transcription factors and regulate lung development. Genes Dev. *28*, 1363–1379.

Hu, H.Y., He, L., and Khaitovich, P. (2014). Deep sequencing reveals a novel class of bidirectional promoters associated with neuronal genes. BMC Genomics *15*, 457.

Jaynes, J.B., and Fujioka, M. (2004). Drawing lines in the sand: even skipped et al. and parasegment boundaries. Dev. Biol. *269*, 609–622.

Kalisz, M., Winzi, M., Bisgaard, H.C., and Serup, P. (2012). EVEN-SKIPPED HOMEOBOX 1 controls human ES cell differentiation by directly repressing GOOSECOID expression. Dev. Biol. *362*, 94–103.

Keller, G. (2005). Embryonic stem cell differentiation: emergence of a new era in biology and medicine. Genes Dev. *19*, 1129–1155.

Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 583–588.

Lagha, M., Bothma, J.P., Esposito, E., Ng, S., Stefanik, L., Tsui, C., Johnston, J., Chen, K., Gilmour, D.S., Zeitlinger, J., and Levine, M.S. (2013). Paused Pol II coordinates tissue morphogenesis in the Drosophila embryo. Cell *153*, 976–987.

Lai, F., Orom, U.A., Cesaroni, M., Beringer, M., Taatjes, D.J., Blobel, G.A., and Shiekhattar, R. (2013). Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature *494*, 497–501.

Latos, P.A., Pauler, F.M., Koerner, M.V., Şenergin, H.B., Hudson, Q.J., Stocsits, R.R., Allhoff, W., Stricker, S.H., Klement, R.M., Warczok, K.E.,

et al. (2012). Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. Science *338*, 1469–1472.

Lepoivre, C., Belhocine, M., Bergon, A., Griffon, A., Yammine, M., Vanhille, L., Zacarias-Cabeza, J., Garibal, M.A., Koch, F., Maqbool, M.A., et al. (2013). Divergent transcription is associated with promoters of transcriptional regulators. BMC Genomics *14*, 914.

Mahmoudi, S., Henriksson, S., Corcoran, M., Méndez-Vidal, C., Wiman, K.G., and Farnebo, M. (2009). Wrap53, a natural p53 antisense transcript required for p53 induction upon DNA damage. Mol. Cell *33*, 462–471.

Malik, S., and Roeder, R.G. (2010). The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. Nat. Rev. Genet. *11*, 761–772.

Manoukian, A.S., and Krause, H.M. (1992). Concentration-dependent activities of the even-skipped protein in Drosophila embryos. Genes Dev. *6*, 1740–1751.

Mercer, T.R., Dinger, M.E., Sunkin, S.M., Mehler, M.F., and Mattick, J.S. (2008). Specific expression of long noncoding RNAs in the mouse brain. Proc. Natl. Acad. Sci. USA *105*, 716–721.

Mondal, T., Rasmussen, M., Pandey, G.K., Isaksson, A., and Kanduri, C. (2010). Characterization of the RNA content of chromatin. Genome Res. *20*, 899–907.

Morris, K.V., and Mattick, J.S. (2014). The rise of regulatory RNA. Nat. Rev. Genet. *15*, 423–437.

Ørom, U.A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytnicki, M., Notredame, C., Huang, Q., et al. (2010). Long noncoding RNAs with enhancer-like function in human cells. Cell *143*, 46–58.

Patrone, G., Puppo, F., Cusano, R., Scaranari, M., Ceccherini, I., Puliti, A., and Ravazzolo, R. (2000). Nuclear run-on assay using biotin labeling, magnetic bead capture and analysis by fluorescence-based RT-PCR. Biotechniques *29*, 1012–1014, 1016–1017.

Ponjavic, J., Oliver, P.L., Lunter, G., and Ponting, C.P. (2009). Genomic and transcriptional co-localization of protein-coding and long non-coding RNA pairs in the developing brain. PLoS Genet. *5*, e1000617.

Qi, L.S., and Arkin, A.P. (2014). A versatile framework for microbial engineering using synthetic non-coding RNAs. Nat. Rev. Microbiol. *12*, 341–354.

Saldaña-Meyer, R., González-Buendía, E., Guerrero, G., Narendra, V., Bonasio, R., Recillas-Targa, F., and Reinberg, D. (2014). CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. Genes Dev. *28*, 723–734.

Sauvageau, M., Goff, L.A., Lodato, S., Bonev, B., Groff, A.F., Gerhardinger, C., Sanchez-Gomez, D.B., Hacisuleyman, E., Li, E., Spence, M., et al. (2013). Multiple knockout mouse models reveal lincRNAs are required for life and brain development. eLife *2*, e01749.

Shechner, D.M., Hacisuleyman, E., Younger, S.T., and Rinn, J.L. (2015). Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. Nat. Methods *12*, 664–670.

Shen, X., Kim, W., Fujiwara, Y., Simon, M.D., Liu, Y., Mysliwiec, M.R., Yuan, G.C., Lee, Y., and Orkin, S.H. (2009). Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. Cell *139*, 1303–1314.

Sigova, A.A., Mullen, A.C., Molinie, B., Gupta, S., Orlando, D.A., Guenther, M.G., Almada, A.E., Lin, C., Sharp, P.A., Giallourakis, C.C., and Young, R.A. (2013). Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. Proc. Natl. Acad. Sci. USA *110*, 2876–2881.

Taft, R.J., Pheasant, M., and Mattick, J.S. (2007). The relationship between non-protein-coding DNA and eukaryotic complexity. BioEssays *29*, 288–299.

Tam, P.P., and Loebel, D.A. (2007). Gene function in mouse embryogenesis: get set for gastrulation. Nat. Rev. Genet. *8*, 368–381.

Tang, F., Barbacioru, C., Nordman, E., Li, B., Xu, N., Bashkirov, V.I., Lao, K., and Surani, M.A. (2010). RNA-Seq analysis to capture the transcriptome landscape of a single cell. Nat. Protoc. *5*, 516–535.

Trimarchi, T., Bilal, E., Ntziachristos, P., Fabbri, G., Dalla-Favera, R., Tsirigos, A., and Aifantis, I. (2014). Genome-wide mapping and characterization of Notch-regulated long noncoding RNAs in acute leukemia. Cell *158*, 593–606.

Wheeler, T.M., Leger, A.J., Pandey, S.K., MacLeod, A.R., Nakamori, M., Cheng, S.H., Wentworth, B.M., Bennett, C.F., and Thornton, C.A. (2012). Targeting nuclear RNA for in vivo correction of myotonic dystrophy. Nature *488*, 111–115.

Woolfe, A., Goodson, M., Goode, D.K., Snell, P., McEwen, G.K., Vavouri, T., Smith, S.F., North, P., Callaway, H., Kelly, K., et al. (2005). Highly conserved non-coding sequences are associated with vertebrate development. PLoS Biol. *3*, e7.

Wu, X., and Sharp, P.A. (2013). Divergent transcription: a driving force for new gene origination? Cell *155*, 990–996.

Xu, Z., Wei, W., Gagneur, J., Clauder-Münster, S., Smolik, M., Huber, W., and Steinmetz, L.M. (2011). Antisense expression increases gene expression variability and locus interdependency. Mol. Syst. Biol. *7*, 468.

Yin, Y., Yan, P., Lu, J., Song, G., Zhu, Y., Li, Z., Zhao, Y., Shen, B., Huang, X., Zhu, H., et al. (2015). Opposing Roles for the lncRNA Haunt and Its Genomic Locus in Regulating HOXA Gene Activation during Embryonic Stem Cell Differentiation. Cell Stem Cell *16*, 504–516.

Zhang, Y.E., Vibranovski, M.D., Landback, P., Marais, G.A.B., and Long, M. (2010). Chromosomal Redistribution of Male-Biased Genes in Mammalian Evolution with Two Bursts of Gene Gain on the X Chromosome. PLoS Biol. *8*, e1000494.