# Extensive microRNA-mediated crosstalk between lncRNAs and mRNAs in mouse embryonic stem cells

Jennifer Y. Tan,[1,2,3] Tamara Sirey,[1,2] Frantisek Honti,[1,2,3] Bryony Graham,[4,5] Allison Piovesan,[6] Matthias Merkenschlager,[5] Caleb Webber,[1,2] Chris P. Ponting,[1,2] and Ana C. Marques[1,2,3]

*[1]MRC Functional Genomics Unit, University of Oxford, Oxford OX1 3QX, United Kingdom; [2]University of Oxford, Department of Physiology, Anatomy and Genetics, Oxford OX1 3QX, United Kingdom; [3]Department of Physiology, University of Lausanne, 1005 Lausanne, Switzerland; [4]MRC Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, Oxford University, Oxford OX3 9DS, United Kingdom; [5]Lymphocyte Development Group, MRC Clinical Sciences Centre, Imperial College London, London W12 0NN, United Kingdom; [6]Department of Experimental, Diagnostic and Specialty Medicine (DIMES),Unit of Histology, Embryology and Applied Biology, University of Bologna, 40126 Bologna, BO, Italy*

**Recently, a handful of intergenic long noncoding RNAs (lncRNAs) have been shown to compete with mRNAs for binding to miRNAs and to contribute to development and disease. Beyond these reports, little is yet known of the extent and functional consequences of miRNA-mediated regulation of mRNA levels by lncRNAs. To gain further insight into lncRNA-mRNA miRNA-mediated crosstalk, we reanalyzed transcriptome-wide changes induced by the targeted knockdown of over 100 lncRNA transcripts in mouse embryonic stem cells (mESCs). We predicted that, on average, almost one-fifth of the transcript level changes induced by lncRNAs are dependent on miRNAs that are highly abundant in mESCs. We validated these findings experimentally by temporally profiling transcriptome-wide changes in gene expression following the loss of miRNA biogenesis in mESCs. Following the depletion of miRNAs, we found that >50% of lncRNAs and their miRNA-dependent mRNA targets were up-regulated coordinately, consistent with their interaction being miRNA-mediated. These lncRNAs are preferentially located in the cytoplasm, and the response elements for miRNAs they share with their targets have been preserved in mammals by purifying selection. Lastly, miRNA-dependent mRNA targets of each lncRNA tended to share common biological functions. Post-transcriptional miRNA-mediated crosstalk between lncRNAs and mRNA, in mESCs, is thus surprisingly prevalent, conserved in mammals, and likely to contribute to critical developmental processes.**

**[Supplemental material is available for this article.]**

Transcript abundance for large numbers of eukaryotic genes is modulated post-transcriptionally by microRNAs (miRNAs, ~22-nt noncoding RNAs) (Stark et al. 2005). The recognition and binding of a mature miRNA to response elements (MREs) present within the target transcript lead to its degradation or translational repression (Ambros et al. 2003; Wienholds and Plasterk 2005; Bartel 2009). When a pair of transcripts is targeted by a particular miRNA, changes in the abundance of one can modulate the level of the other (Franco-Zorrilla et al. 2007; Marques et al. 2011; Salmena et al. 2011; Tay et al. 2014). Transcripts engaging in such crosstalk are referred to as competitive endogenous RNAs (ceRNAs) (Salmena et al. 2011). Intricate networks of crosstalking RNAs are proposed to regulate coordinately the relative abundance of functionally related transcripts (Sumazin et al. 2011; Ala et al. 2013; Han et al. 2013; Tan et al. 2014; Wehrspaun et al. 2014). This suggests a layer of post-transcriptional regulation that is overlaid upon transcriptional programs.

Changes in endogenous levels of ceRNAs can, for example, lead to changes in cell status (Wang et al. 2013) and have been associated with disease (Poliseno et al. 2011; Tan et al. 2014). Furthermore, some transcribed pseudogenes have preserved their ancestral parent genes' ceRNA function despite having lost their ability to encode functional proteins, which argues for their sustained biological roles (Marques et al. 2012). Nevertheless, the biological relevance of ceRNAs has recently been challenged (Broderick and Zamore 2014) because the level of one transcript, *Aldoa*, required to significantly alter the level of one highly abundant miRNA, miR-122, and its targets in adult hepatocytes was found to exceed the changes observed in vivo, even under extreme physiological or disease conditions (Denzler et al. 2014). In contrast, recent results from genome-wide quantitative analyses suggests ceRNA competition involving a number of miRNAs is relevant in a physiological context and determined by the endogenous miRNA:target pool ratio (Bosson et al. 2014). The differences between the results of these analyses, together with discrepancies in the conclusions of different attempts to mathematically model miRNA-mediated interactions (Ala et al. 2013; Bosia et al. 2013; Hausser and Zavolan 2014; Jens and Rajewsky 2015; Yuan et al. 2015), suggest that substantially more genetic and genomic evidence from many ceRNAs and miRNAs in diverse cell types will

**Corresponding author: anaclaudia.marques@unil.ch**
Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.181974.114.

be required to resolve this issue and to establish the general prevalence and physiological relevance of ceRNAs.

The miRNA-mediated crosstalk among transcripts can involve coding as well as noncoding transcripts, including intergenic long noncoding RNAs (lncRNAs) (Franco-Zorrilla et al. 2007; Cesana et al. 2011; Wang et al. 2013). Thousands of lncRNAs have been annotated in eukaryotic genomes (Derrien et al. 2012; Ulitsky and Bartel 2013), many of which are preferentially located in the cytoplasm (van Heesch et al. 2014), where they can engage in miRNA-mediated interactions with other transcripts. Both computational and experimental evidence support the extensive targeting of lncRNAs by miRNAs (Paraskevopoulou et al. 2013). While a small number of lncRNAs are currently known to function as ceRNAs (Cesana et al. 2011; Fan et al. 2013; Wang et al. 2013; Tan et al. 2014), the full extent of lncRNAs possessing miRNA-dependent regulatory roles remains to be determined (Ulitsky and Bartel 2013).

The relatively low transcript abundance of most lncRNAs (Cabili et al. 2011; Derrien et al. 2012) might be thought to limit their ability to effectively modulate, in a miRNA-dependent manner, mRNA abundance (Ebert and Sharp 2010; Ala et al. 2013; Figliuzzi et al. 2013; Denzler et al. 2014). Nevertheless, some established ceRNAs neither share an unusually high number of predicted MREs with their mRNA targets nor are especially abundant (Cesana et al. 2011; Wang et al. 2013), which suggests that other factors, such as miRNA-target affinity or miRNA turnover, might explain their efficient crosstalk. These ceRNAs are also no different from most other lncRNAs (Cabili et al. 2011; Derrien et al. 2012) with respect to their highly restricted spatial and temporal expression patterns. For example, *linc-MD1* is a muscle-specific ceRNA that regulates transcript abundance of two key myogenic transcription factors, *Maml1* and *Mef2c*, which are required for activating muscle-specific gene expression (Cesana et al. 2011); in addition, *linc-RoR* competes for miR-145 binding with key self-renewal transcription factor transcripts, namely *Nanog*, *Pou5f1*, and *Sox2*, and is expressed during induced pluripotent stem cell (iPSC) reprogramming and in undifferentiated embryonic stem cells (ESCs) (Loewer et al. 2010; Wang et al. 2013). The narrow expression profile of these ceRNAs might specify the cells or tissues in which their activity exerts the greatest effect. Cell-fate decisions, such as those involving *linc-RoR* or *linc-MD1*, often involve switch-like responses in the expression levels of key regulatory genes that result in coordinated changes in transcription profiles driven by one or more key transcription factors. Several miRNAs have been found to contribute to the regulation of such switches (Mukherji et al. 2011), and lncRNAs have often been implicated in the regulation of the circuitry underlying cell-fate decisions (Jia et al. 2010; Cesana et al. 2011; Guttman et al. 2011; Sun et al. 2013; Wang et al. 2013). These results suggest that relatively lowly abundant, yet specifically expressed, lncRNAs might function efficiently as ceRNAs regulating the transition between pluripotent and differentiated cell states. In such bistable states, small changes in miRNA levels induced by ceRNAs may have a greater impact on cellular homeostasis than in fully differentiated normal cells.

In this study we sought to determine the relative prevalence of miRNA-mediated changes induced by lncRNAs. We took advantage of publicly available and experimentally determined expression data on the impact of knockdown, using shRNAs, of over 140 lncRNAs in mouse embryonic stem cells (mESCs) (Guttman et al. 2011). These data previously provided support for the notion that some lncRNAs act as protein-binding scaffolds coordinating cell-type specific gene expression changes transcriptionally (Guttman et al. 2011). Our analysis demonstrates that lncRNAs can also contribute to mESC fate decisions via post-transcriptional miRNA-mediated mechanisms.

## Results

### Extensive miRNA-mediated crosstalk among lncRNAs and mRNAs

To investigate the extent of miRNA-dependent gene expression regulation of mRNAs by lncRNAs, we took advantage of a large set of experimentally determined expression profile changes in mouse embryonic stem cells (mESCs) induced by the targeted knockdown of 147 lncRNAs and 40 regulatory protein-coding gene controls (Guttman et al. 2011). Throughout the manuscript, we refer to individual lncRNAs using the names given to these transcripts in the original study. We refer to mRNAs whose expression is differentially up- or down-regulated following knockdown, using shRNAs, of these noncoding and coding RNAs, as their "targets" (Fig. 1A). Depletion of lncRNAs resulted in expression level changes of an average of 163 targets, a similar number of targets to those observed for protein-coding gene controls (on average, 197 targets) (Guttman et al. 2011).

We considered whether some of these gene expression changes (Guttman et al. 2011) were a consequence of increased post-transcriptional repression of transcripts sharing miRNA response elements (MREs) with the depleted lncRNAs. In contrast to transcriptional regulation by lncRNAs that can lead to either activation or repression of their targets' expression, the primary consequence of competition among lncRNAs and mRNA targets for binding to the same miRNAs is a positive correlation between their transcripts' levels. We applied this signature to predict miRNA-dependent lncRNA-mRNA interactions (Fig. 1A).

To predict the extent of miRNA-mediated regulation by lncRNAs, we first identified mESC-expressed miRNAs and then predicted which transcripts they bind and regulate. miRNA levels were quantified, in quadruplicate, using NanoString Technology (Methods; Supplemental Table S1), and subsequent analysis was performed considering only the 25% most highly expressed miRNAs (160 from 117 miRNA families) (Garcia et al. 2011), except where otherwise stated. MREs were predicted using TargetScan (version 6.2) (Garcia et al. 2011) across the entire sequence of the lncRNAs (Guttman et al. 2011) and within the longest annotated 3′ UTRs of mouse protein-coding genes (Ensembl build 70) (Supplemental Table S2; Methods; Flicek et al. 2012).

We first, and as a negative control, considered the mRNA targets for each of the 40 regulatory protein-coding gene controls (Guttman et al. 2011). For each target, we calculated the density (number per kilobase [kb] of 3′ UTR sequence) of predicted response elements for mESC-expressed miRNAs it shared with the transcription factor that had been identified in the original study as significantly altering its expression (Guttman et al. 2011). These transcription factors' targets were expected to be modulated transcriptionally, not post-transcriptionally (Guttman et al. 2011), and as predicted, no significant differences were found in the densities of shared MREs between transcription factor and up- or down-regulated targets (Supplemental Fig. S1A). We created a set of control miRNA seeds based on shuffled dinucleotides of mESC-expressed miRNAs that were not homologous to the seeds of mouse mESC miRNAs (80 shuffled miRNAs) (Methods). As expected, no significant difference was found between the densities of MREs ($MRE_{shuffled}$) for these shuffled miRNAs shared between

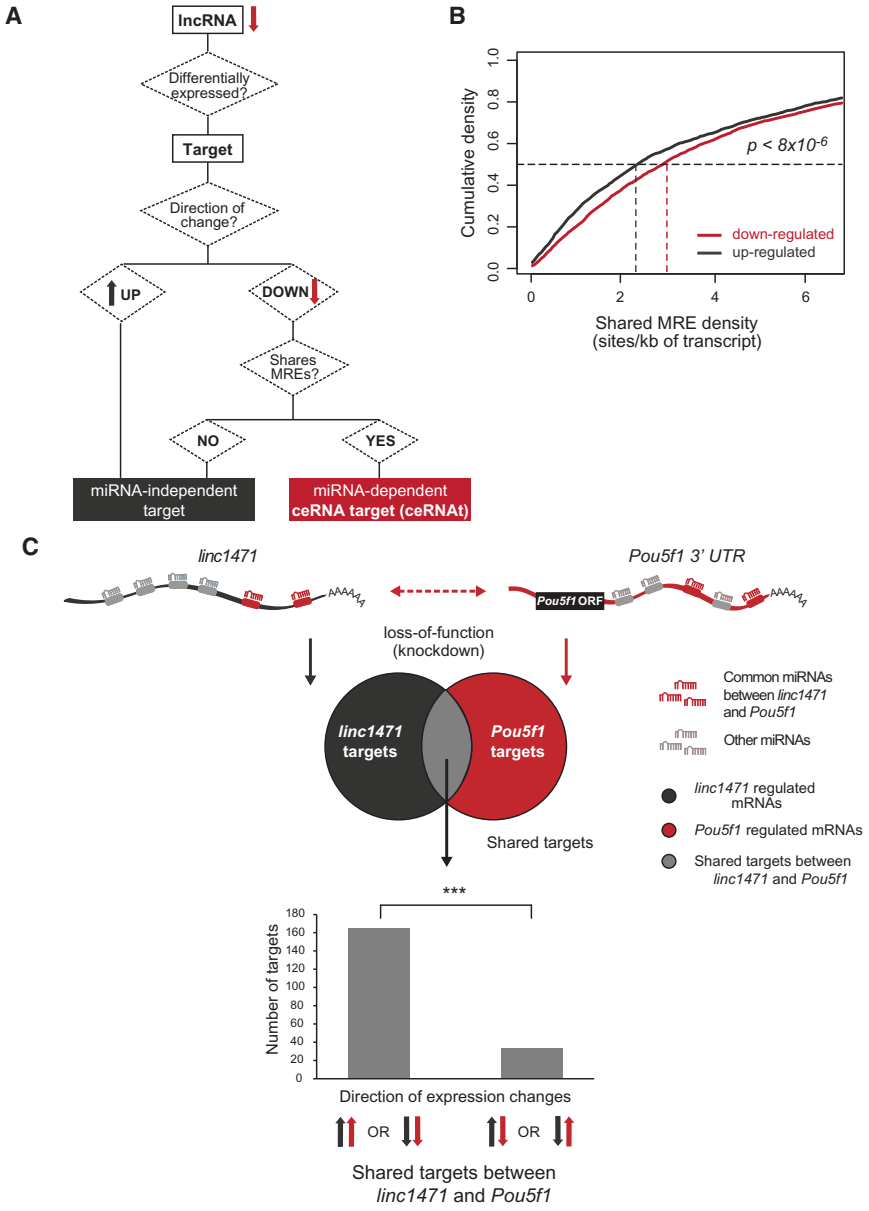



**Figure 1.** MiRNA-dependent regulation of mRNA abundance by lncRNAs. (*A*) Diagram illustrating the classification of an individual lncRNA's mRNA target as either a competitive endogenous RNA target (ceRNAt) (red) or an miRNA-independent target (dark gray). Red and dark gray arrows represent down- and up-regulation, respectively. (*B*) Cumulative distribution plots for the density of predicted miRNA response elements (MREs) for the top 25% most highly expressed miRNA families in mESCs shared between lncRNAs and their respective down-regulated targets (median = 2.9 sites/kb of transcript, red) and up-regulated targets (median = 2.3 sites/kb of transcript, black). (*C*) Transcription factor *Pou5f1* (red) is predicted to compete (dotted red arrow) for binding to miR-421 and miR-762 MREs (red oblongs within transcript) with *linc1471* (dark gray). MREs for miRNAs not shared between the two genes are represented in light gray. Bar chart represents the number of miRNA-independent targets of *linc1471* that are shared with *Pou5f1*, whose levels changed in the same or opposite direction upon *linc1471* (dark gray) or *Pou5f1* (red) knockdown. Arrows indicate the direction of the observed expression changes following *linc1471* and *Pou5f1* knockdown. (***) $P < 0.001$.

down-regulated (median of 1.1 MREs/kb) or up-regulated targets (median of 0.90 MREs/kb) (Supplemental Fig. S1B) and lncRNA.

Next, we considered the mRNA targets for each of the 147 lncRNAs and the density of MREs for mESC-expressed miRNAs they shared. In contrast to results for regulatory transcription factor and $MRE_{shuffled}$ controls, mRNAs that were down-regulated upon lncRNA knockdown shared a significantly higher number of predicted MREs with these lncRNAs (median of 2.9 MREs/kb of 3′ UTR) than up-regulated targets (median of 2.3 MREs/kb; $P < 8 \times 10^{-6}$, two-tailed Mann-Whitney *U* test) (Fig. 1B). Similar results were obtained when MRE predictions were considered for all miRNAs expressed in mESCs or the 75% or 50% most highly expressed miRNAs in mESCs (Supplemental Fig. S1C–E). These results were obtained using computationally predicted MREs, which are known to have relatively high false-positive and -negative rates (Maziere and Enright 2007). Consequently, we next considered a stringent set of MREs that overlap experimentally derived Argonaute-bound regions in mESCs (Leung et al. 2011). With these, we found that down-regulated mRNA targets contain over twofold higher densities of MREs shared with their lncRNA (mean of $5.0 \times 10^{-3}$ MREs/kb) than up-regulated mRNAs (mean of $2.1 \times 10^{-3}$ MREs/kb; $P < 0.05$, two-tailed Mann-Whitney *U* test).

These results are consistent with the abundance of some lncRNAs' mRNA targets being modulated through their competition for binding mESC-expressed miRNAs. On average, 19.3% (Supplemental Table S2) of lncRNAs' targets exhibited the two signatures of miRNA-dependent crosstalk: (1) Depletion of the lncRNA is associated with down-regulation of its mRNA target, and (2) coding and noncoding transcripts contain predicted MREs for the same miRNAs (Fig. 1A). Hereafter, we refer to mRNA targets of lncRNAs with these signatures as competitive endogenous RNA targets (ceRNAts) (Supplemental Table S3).

What then of mRNAs whose changes in transcript abundance could not be explained by a ceRNA mechanism (Fig. 1A)? We hypothesized that some of these changes might be a consequence of secondary effects of transcriptional regulation mediated by transcription factors whose transcripts are primary miRNA-dependent targets of lncRNAs (i.e., TF ceRNAts). To test this hypothesis, we considered 10 lncRNAs that have TF ceRNAts whose effects on gene expression upon knockdown were also experimentally determined by Guttman et al. (2011). Consistent with this indirect mode of mRNA regulation, the levels of twice as many miRNA-independent targets of these lncRNAs were affected in the same direction rather than in opposite directions upon knockdown of either the lncRNA or its respective TF ceRNAt (averages 15.8% versus 7.8%, respectively; $P < 0.05$, two-tailed Mann-Whitney *U* test). This indicates that a proportion

of gene expression changes that are not primarily miRNA-dependent may be explained through secondary transcriptional activation by TFs whose transcripts are ceRNAts.

To illustrate this phenomenon, consider the proposed miRNA-mediated crosstalk between *linc1471* and the mRNA encoding the transcription factor POU5F1 (also known as OCT4). Two of the five recognition elements predicted in the 3′ UTR of *Pou5f1* are for miR-421 and miR-762, for which MREs are also predicted in *linc1471*, whose knockdown (Fig. 1C), in turn, leads to a significant 21-fold decrease in *Pou5f1* mRNA abundance (Guttman et al. 2011). If some of the *linc1471*'s targets whose levels cannot be explained by a ceRNA mechanism are a secondary effect of miRNA-mediated decrease in *Pou5f1* expression, then one would expect that the levels of these genes would be affected upon knockdown of either *linc1471* or *Pou5f1* in the same direction. Indeed, and as expected, miRNA-independent targets of *linc1471* that are also differentially expressed upon *Pou5f1* knockdown (by 14.8-fold) (Guttman et al. 2011) are five times more likely to change in the same ($n$ = 165) rather than in opposing directions ($n$ = 33) ($P < 10^{-4}$, Fisher's exact test) (Fig. 1C). Furthermore, of the *linc1471* targets that we hypothesized might be a secondary effect of *Pou5f1* changes, 21.2% show evidence that POU5F1 protein binds at their promoter in mESCs (Karwacki-Neisius et al. 2013), a significantly ($P < 0.05$, Fisher's exact test) higher proportion than found when considering all other *linc1471* targets (13.7%).

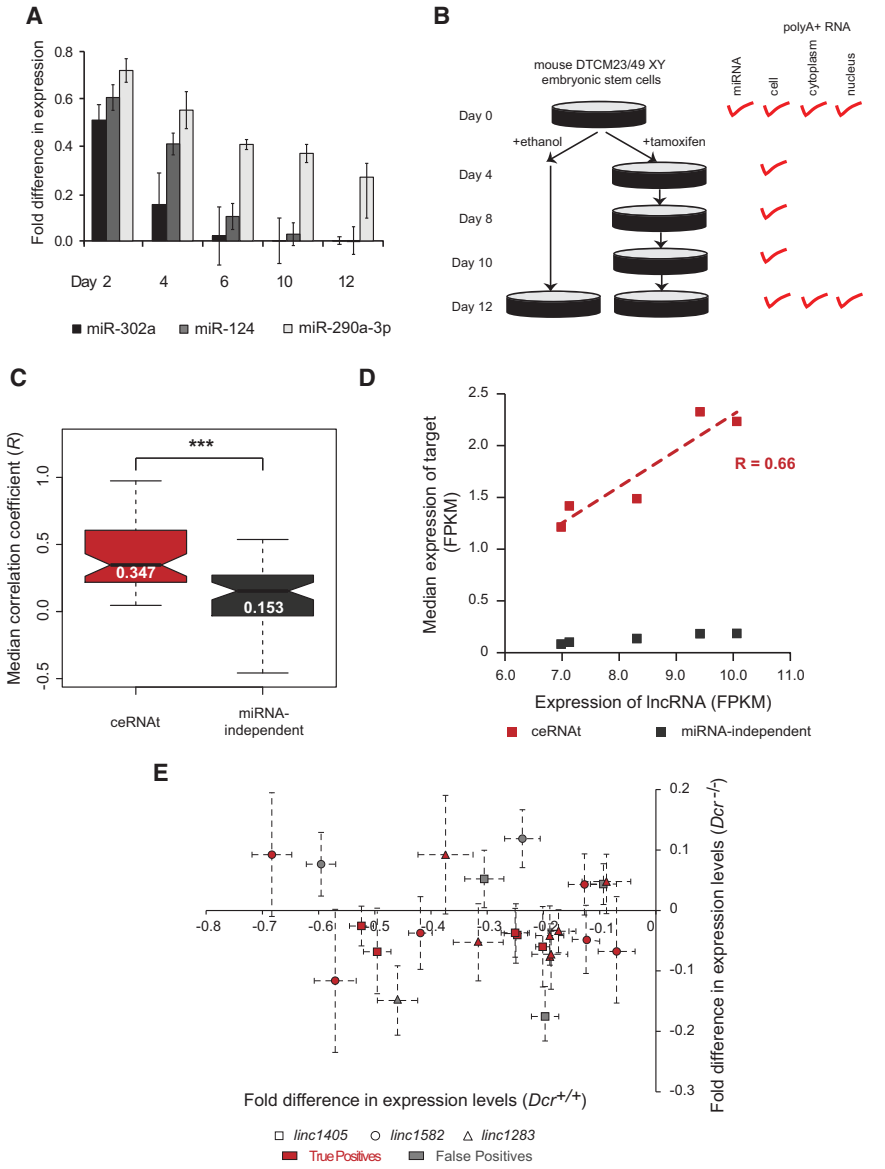



**Figure 2.** The expression levels of lncRNAs and their ceRNAts are positively correlated upon loss of miRNA biogenesis. (*A*) Loss of miRNA biogenesis ($Dcr^{-/-}$) is associated with decreased levels, relative to the wild-type control, of mature miR-302a (black), miR-124 (dark gray), and miR-290a-3p (light gray) abundance over a 12-d time course. Fold difference in expression relative to controls was determined using qRT-PCR in triplicate. (*B*) Long poly(A)-selected RNA from total cellular extracts of DTCM23/49 XY mouse embryonic stem cells (mESCs) was collected on days 0, 4, 8, 10, and 12 following exposure to tamoxifen. Long poly(A)-selected RNA was also collected from the nucleus and cytoplasm of these cells before (0) and 12 d after treatment with tamoxifen. Total RNA used to quantify miRNA expression was extracted before tamoxifen treatment. (*C*) Median correlation coefficients between lncRNA expression and their respective ceRNAts (red, median $R = 0.347$) and miRNA-independent targets (dark gray, $R = 0.153$) over the 12-d time course following loss of miRNA biogenesis. (*D*) Pearson's correlation between *linc1510*'s expression (average across replicates of the gene expression measured as fragment per kilobase of exon per million reads mapped [FPKM], *x*-axis) and the median expression of all of its targets annotated as either ceRNAts ($R = 0.66$, red) or miRNA-independent targets ($R = 0.19$, dark gray) at each time point (*y*-axis). (*E*) Fold difference in normalized (using *Gadph*) gene expression following knockdown of *linc1405* (square symbols), *linc1582* (circles), or *linc1283* (triangles) and siRNA transfection control in $Dcr^{+/+}$ (*x*-axis) or $Dcr^{-/-}$ (*y*-axis) mESCs. True and false ceRNA positives are highlighted in red or gray, respectively. Error bars represent S.E.M. across replicates. (***) $P < 0.001$.

## lncRNAs and their respective ceRNAts are coordinately up-regulated upon loss of miRNA biogenesis

We then undertook a more direct approach to experimentally validate lncRNA crosstalk with mRNAs via miRNAs. For this, we took advantage of mESCs in which a conditional mutation of a key gene in miRNA biogenesis, *Dicer1*, has been introduced. These cells contain a tamoxifen-inducible *Cre* recombinase that drives recombination between *loxP* sites flanking the *Dicer1* RNase III domain (Nesterova et al. 2008). Loss of this domain (Supplemental Fig. S2A) and ablation of miRNA biogenesis thus occur following tamoxifen addition. The time for complete miRNA removal, following *Dicer1* loss-of-function, varies considerably owing to dependences on miRNA initial abundance and stability, as illustrated in Figure 2A for three miRNAs, miR-302a, miR-124, and miR-290a-3p. This temporal variation of miRNA abundance allowed us to investigate the miRNA-dependency of interactions

between transcripts. In particular, we expected the expression levels over time for a lncRNA and its ceRNAts to increase coordinately as the levels of miRNAs mediating their crosstalk diminish. No expression correlation was expected following loss of miRNA biogenesis between lncRNAs and mRNA targets that we predicted to be primarily regulated in a miRNA-independent manner, in particular for lncRNAs and mRNAs that do not share predicted MREs for mESC-expressed miRNAs (Fig. 1A).

We determined the temporal variation of mRNA levels by collecting, in triplicate, and then sequencing poly-adenylated RNA from mESCs prior to tamoxifen addition (day 0) and at days 4, 8, 10, and 12 thereafter (Fig. 2B). On average, 88.0% (range 87.0%–89.6%) of RNA sequencing paired-end reads was mapped uniquely to the mouse genome (assembly version mm9). As expected, mRNA and lncRNA expression levels, in general, clustered by time point (Supplemental Fig. S2B). No significant decrease in *Myc* expression levels was observed (fold-change = 0.68 ± 0.17, FDR = 0.58) in our experiment, in contrast to homozygous *Dgcr8*$^{-/-}$ (Melton et al. 2010) or *Dicer1*$^{-/-}$ (Zheng et al. 2014) mESCs, suggesting that, as opposed to a recent analysis of mESC lncRNAs (Zheng et al. 2014), transcriptional regulation by *Myc* does not contribute significantly to the changes in lncRNA or mRNA levels observed in our study.

For a subset of five miRNAs, we estimated the correlation between their abundance (measured by qPCR) and the levels of their predicted mRNA/lncRNA targets (measured by RNA-seq). As expected, miRNA and target abundance were significantly anti-correlated ($R = -0.46$, $P < 0.001$, Pearson's correlation) (Supplemental Fig. S2C).

We removed from our analysis 24 of the initially considered lncRNAs because their loci are now seen to overlap annotated protein-coding genes (Ensembl build 70). Of the remaining lncRNAs, 104 show evidence of expression in our experiment and had multiple ceRNAts and miRNA-independent targets expressed in these cells (Supplemental Fig. S2D). As expected, lncRNA-ceRNAt pairs exhibited greater correlation (median $R = 0.35$) in expression levels over time than lncRNA and miRNA-independent targets (median $R = 0.15$, $P < 3.1 \times 10^{-8}$, two-tailed Mann-Whitney *U* test) (Fig. 2C). This difference remained when we considered down- and up-regulated miRNA-independent targets separately (Supplemental Fig. S2E). As a control, we considered the difference in correlation between lncRNAs and down-regulated targets sharing MREs for shuffled miRNAs (ceRNAt for $MRE_{shuffled}$) and the remaining down- or up-regulated targets. As expected, we found no significant difference in median correlation between these groups (Supplemental Fig. 2F). These observations are consistent with the decreased abundance of miRNAs leading to coordinated changes in the abundance of transcripts that can regulate each other's levels via miRNA-mediated crosstalk. For example, the expression following loss of miRNA biogenesis of *linc1510* was much better correlated with its ceRNAts (25 targets, $R = 0.66$) than with its miRNA-independent targets (86 targets, $R = 0.19$) (Fig. 2D).

The expression level of 51% (53 of 104) of lncRNAs was significantly (empirical $P < 0.05$) better correlated with the levels of their ceRNAts than expected, following *Dicer1* knockout, based on estimates for 1000 sets (of the same size) of randomly selected mESC-expressed transcripts pairs (Methods). These 53 lncRNAs, whose miRNA dependency of ceRNAt interactions was validated, are hereafter referred to as long noncoding competitive endogenous RNAs or lnceRNAs (Supplemental Table S5). Only eight of 104 (8%) lncRNAs were significantly correlated following *Dicer1* conditional excision with its miRNA-independent targets.

Next, we sought to experimentally test our predictions. Using siRNA, we knocked down three lnceRNAs in wild-type (*Dcr*$^{+/+}$) and *Dicer1*-null (*Dcr*$^{-/-}$) mESCs and measured, using qPCR, the level of 10 of their putative ceRNAts. Genes found to be differentially expressed upon lncRNA knockdown by Guttman et al. (2011) but that could not be replicated in our experiment, likely due to differences in experimental design, were excluded from the remainder of the analysis (maximum two targets per lnceRNA) (Supplemental Fig. S3). Consistent with the miRNA dependency of their regulatory interactions, 62.5%–87.5% ceRNAts for the three lnceRNAs tested were significantly down-regulated in *Dcr*$^{+/+}$ but not in *Dcr*$^{-/-}$ mESCs after multiple test corrections (Bonferroni corrected $P < 0.005$, two-tailed *t*-test) (Fig. 2E; Supplemental Fig. S3).

The ability of a transcript to modulate another's abundance in a miRNA-dependent manner depends on its relative abundance. The median expression of lnceRNAs (1.7 FPKM) was over 10-fold higher than that of other mESC lncRNAs (0.17 FPKM) (Supplemental Fig. S4A). On average, lnceRNA expression was higher than 48% of ceRNAts with 91% of lnceRNAs being more highly expressed than at least one of their predicted ceRNAts (Supplemental Fig. S4B).

In summary, our results are consistent with over half of the mESC-expressed lncRNAs investigated here interacting with their mRNA targets in a miRNA-dependent manner.

## lnceRNAs are enriched in the cytoplasm

We expected lnceRNAs to be enriched in the cytoplasm because post-transcriptional regulation of gene expression by miRNAs occurs preferentially in this subcellular compartment (Bartel 2004). Gene expression in cytoplasmic and nuclear mESC fractions was determined by extracting and sequencing poly(A)-selected RNA in triplicate (Methods) before (day 0) and after (day 12) loss of *Dicer1* function. Expression data from the different experimental conditions were clearly separated using multidimensional scaling analysis (Supplemental Fig. S5A). We estimated the expression in the cytosol relative to the nucleus as $r = \text{expression}_{\text{cytosol}}/\text{expression}_{\text{nucleus}}$ for each mESC expressed locus.

The 104 mESC-expressed lncRNAs considered were less abundant in the cytoplasmic fraction (median $r = 0.529$) than mRNAs (median $r = 0.917$, $P < 2.2 \times 10^{-16}$, two-tailed Mann-Whitney *U* test) (Supplemental Fig. S5B) as was seen previously for a large set of human lncRNAs (Derrien et al. 2012). However, relative to the set of all Ensembl-annotated lncRNAs expressed in mESCs, they were more abundant in the cytoplasm (median $r = 0.438$, $P < 0.04$, two-tailed Mann-Whitney *U* test) (Supplemental Fig. S5B). This difference may reflect, at least in part, the increased efficiency of RNAi targeting of transcripts in the cytosol which, in the original study (Guttman et al. 2011), would have favored selection of cytosol-enriched lncRNAs for transcriptome-wide profiling. The subset of lncRNAs previously reported to physically interact with chromatin (26 considered in our analysis) (Guttman et al. 2011) were found to be more abundant in the nucleus of mESC (median $r = 0.406$) than the remainder of the lncRNAs tested (median $r = 0.563$), which is consistent with their chromatin association and proposed transcriptional roles that take place in the nucleus. Of these, 50% (13 out of 26) were predicted to be lnceRNAs. A subset of these may be bifunctional lncRNAs that shuttle between the nucleus and the cytoplasm and regulate gene expression at both transcriptional and post-transcriptional levels. Further work will be required to establish the biological or otherwise relevance of this observation.

As expected, and relative to the remaining 51 lncRNAs with no evidence for miRNA-mediated regulatory roles (median $r$ = 0.397), the 53 lnceRNAs were significantly enriched in the cytosol (median $r$ = 0.555, $P$ < 0.03, two-tailed Mann-Whitney $U$ test) (Fig. 3A; Supplemental Table S6).

Loss of miRNA biogenesis and release from miRNA-mediated regulation are expected to lead to increases in the cytoplasmic abundance of ceRNAts. Indeed, while their median nuclear abundances decreased (by 4.7% over 12 d), their cytoplasmic levels increased by 7.2% ($P < 2.0 \times 10^{-6}$, two-tailed Mann-Whitney $U$ test) (Fig. 3B; Methods). A small yet not statistically significant change was also observed between the levels in the cytoplasmic (median $r$ = 1.03) and nuclear (median $r$ = 0.98) subcellular fractions ($P$ = 0.06, two-tailed Mann-Whitney $U$ test) of putative ceRNAts of the 53 non-lnceRNAs following loss of miRNA biogenesis. No significant change in relative levels, following *Dicer1* loss-of-function, in the cytoplasm or the nucleus was detected for miRNA-independent targets of both lnceRNAs and non-lnceRNAs following loss of miRNA biogenesis (Supplemental Fig. S5C,D).

## Recognition elements for miRNAs shared between lnceRNAs and ceRNAts have evolved under constraint in mammals

We reasoned that if the proposed miRNA-mediated regulation by lnceRNAs adds an important layer of gene expression regulation, then their MRE sequences, in particular those shared between lnceRNAs and their ceRNAts, would show signatures of purifying selection consistent with the action of natural selection to preserve this regulatory layer. Using publicly available poly(A)-selected RNA sequencing data for human H1 embryonic stem cells (Bernstein et al. 2012), we found evidence of conserved transcription (Methods) for 29 of the 104 lncRNAs (28.0%). Of these, 66% (19) are lnceRNAs that we hereafter refer to as conserved lnceRNAs (Supplemental Table S7). We estimated the nucleotide substitution rate ($d_{MRE}$), between mouse and human, across response elements for mESC expressed and mammalian conserved miRNAs (57 miRNA families) predicted within the sequence of conserved lnceRNAs ($d_{MRE}$ = 0.376) (Fig. 4A; Methods). This rate was significantly and substantially suppressed compared with random samples of putatively neutrally evolving sequence ($P < 10^{-4}$, empirical $P$-value) (Fig. 4A). In contrast, no significant difference ($P$ = 0.247, empirical test) in nucleotide substitution rate was observed between $MRE_{shuffled}$ ($d_{MRE\text{-}shuffled}$ = 0.429) and putatively neutral sequence (Supplemental Fig. S6). Next, we compared the rate estimated for MREs to what would be expected for non-MRE in conserved lnceRNA sequence that has been matched in length (Methods). We also accounted for the observed difference in G + C content in predicted MREs (%G + C = 42.0%) and non-MRE sequence (%G + C = 48.6%, two-tailed Mann-Whitney $U$ test, $P < 6 \times 10^{-4}$) (Supplemental Fig. S7). The MRE nucleotide substitution rate normalized to a neutral rate ($d_{MRE}/d_{AR}$ = 0.872) was significantly lower ($P < 2.2 \times 10^{-16}$, two-tailed Mann-Whitney $U$ test) than the rate for non-MRE sequence ($d_{nonMRE}/d_{AR}$ = 0.977) (Fig. 4B). This implies that MREs within conserved lnceRNAs evolved under stronger selective constraint than other lnceRNA regions. Consequently, despite the known low sensitivity of MRE prediction algorithms (Maziere and Enright 2007), this signature of purifying selection suggests that a fraction of the predicted MRE sequences are functional and conserved between mouse lnceRNAs and their human orthologs.

Finally, the nucleotide substitution rates of lnceRNA MREs for miRNAs that are shared (median $d_{MRE\text{-}shared}/d_{AR}$ = 0.831) with their respective ceRNAts evolved under significantly greater constraint ($P < 2.2 \times 10^{-16}$, two-tailed Mann-Whitney $U$ test) than MREs within lnceRNAs that were not shared with their ceRNAts (median $d_{MRE\text{-}nonshared}/d_{AR}$ = 0.900) (Fig. 4C).

## ceRNAts of individual lnceRNAs tend to be functionally related

Finally, we investigated whether miRNA-dependent regulation by each one of the 53 lnceRNAs preferentially affects transcripts of functionally related genes. For this, we took advantage of an integrative phenotypic-linkage network of mouse protein-coding genes (Honti et al. 2014). This network integrates gene–gene linkage information from diverse and complementary sources including gene coexpression, protein physical interaction, co-citation, and gene functional annotation data and, relative to networks built using individual data types, exhibits improved coverage and accuracy (Honti et al. 2014). For each lnceRNA, we estimated the average link weight between genes classified as either ceRNAts or miRNA-independent targets; a higher link weight reflects the increasing likelihood of two genes in the network being functionally related (Honti et al. 2014).

Strikingly, lnceRNAs' ceRNAts were found to be substantially more closely related to each other (median of average link weights = 0.534) than were miRNA-independent transcripts (median of average link weights = 0.205, two-tailed Mann-Whitney $U$ test, $P$ < 0.026) (Fig. 5A). This result was unaltered when we considered separately miRNA-independent targets that were either down- (median of average link weights = 0.236, two-tailed Mann-Whitney $U$ test, $P$ < 0.046) or up-regulated (median of average link weights = 0.189, two-tailed Mann-Whitney $U$ test, $P$ < 0.012) upon lncRNA knockdown (Supplemental Fig. S8A). Similar results were obtained using another measure of functional relatedness, the sum of the linkage weights, after subsampling to the same target group size (Supplemental Fig. S8B; Methods). As a control, we tested differences in functional clustering between down-regulated genes sharing shuffled MREs (median

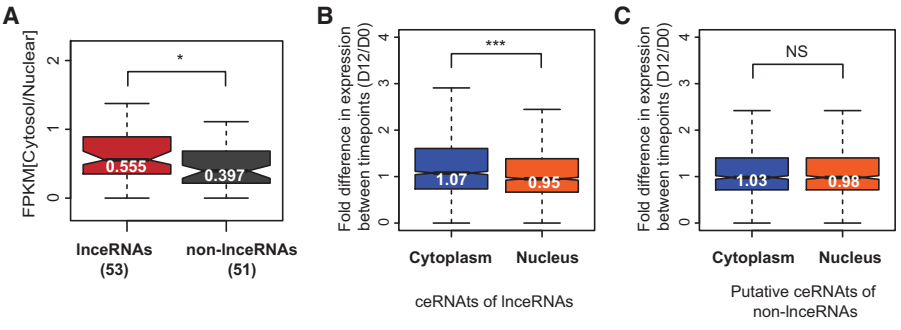



**Figure 3.** lnceRNAs are enriched in the cytoplasm. (*A*) Ratio between gene expression (in FPKM) in the cytoplasmic and nuclear fraction for lnceRNAs (red, $r$ = 0.555) and lncRNAs that were not annotated as lnceRNAs (gray, $r$ = 0.397). Relative fold difference in expression measured in the cytoplasm (blue) and nucleus (orange) of cells before (day 0) and after (day 12) *Dicer1* loss-of-function for ceRNAts and putative ceRNAts of (*B*) lnceRNAs and (*C*) non-lnceRNAs. Median fold differences are shown in the corresponding box plot. (NS) Not significant, (*) $P$ < 0.05, (***) $P$ < 0.001.

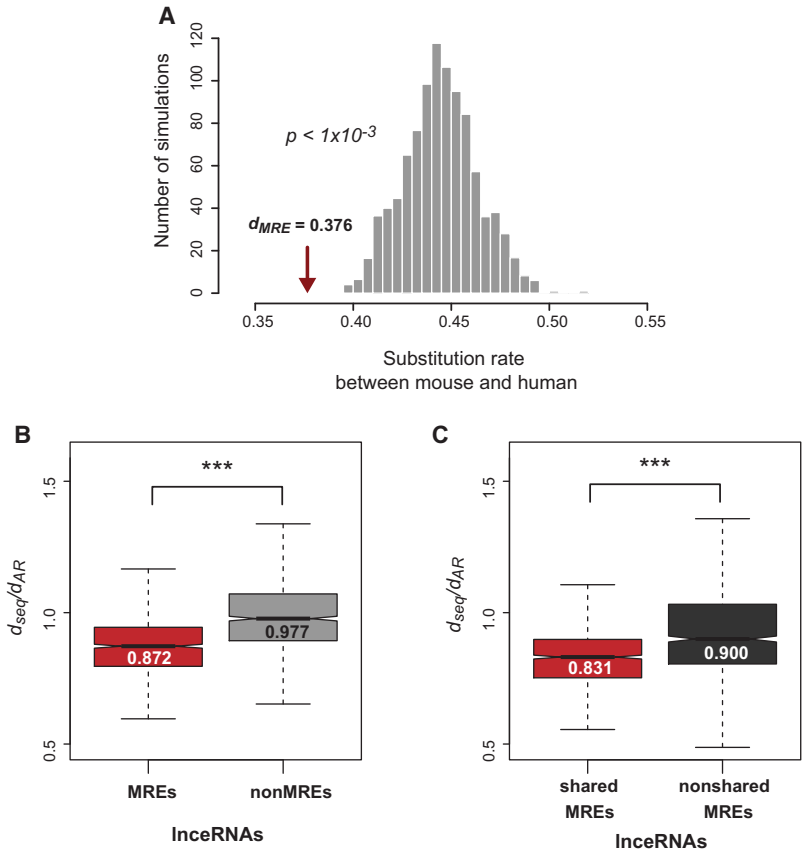



**Figure 4.** MiRNA response elements shared between lnceRNAs and ceRNAts are conserved through mammalian evolution. (*A*) Histogram representing the substitution rate between mouse and human for a neutrally evolving sequence in the vicinity of lnceRNA MREs with the same length (ancestral repeats, ARs). Vertical arrow represents the substitution rate estimated for lnceRNA MREs ($d_{MRE} = 0.376$). (*B*) Distribution of nucleotide substitution rate normalized by neutral local rate, between mouse and human, for predicted MREs ($d_{MRE}/d_{AR} = 0.872$, red) and non-MREs ($d_{nonMRE}/d_{AR} = 0.977$, gray) sequence within lnceRNAs. (*C*) Distribution of mouse-human substitution rate normalized by neutral local rate, measured at MREs shared between lnceRNAs and their ceRNAts (red, $d_{MRE\text{-}shared}/d_{AR} = 0.831$) and MREs for miRNAs not shared between lnceRNA and their targets (gray, $d_{MRE\text{-}nonshared}/d_{AR} = 0.900$). (***) $P < 0.001$.

of average link weights = 0.151) and the remaining lnceRNA targets (median of average link weights = 0.183 and 0.177 for down- and up-regulated $miRNA_{shuffled}$-independent targets) and as expected found no significant difference ($P = 0.10$ and $P = 0.28$ for down- and up-regulated miRNA-independent targets, two-tailed Mann-Whitney *U* test) (Supplemental Fig. S8C). These results argue that coordinated miRNA-mediated modulation of gene expression levels by lnceRNAs tends to affect predominantly functionally related protein-coding genes. This finding, together with the evolutionary constraint observed for shared MREs, argues that miRNA-mediated crosstalk between lncRNAs and mRNAs is required for normal biological processes.

On average, mESC-expressed lnceRNAs have 18.9 predicted MREs per kb of transcript that are specific to 12 different mESC-expressed miRNAs. This MRE density is over 1.5-fold higher than the density in 3′ UTRs of ceRNAts (11.3 MREs predicted per kb; $P < 5.9 \times 10^{-12}$, two-tailed Mann-Whitney *U* test) (Supplemental Fig. S9; Supplemental Table S8). A single lncRNA might, therefore, be more likely than an mRNA to post-transcriptionally regulate the abundance of its targets via crosstalk with different miRNAs.

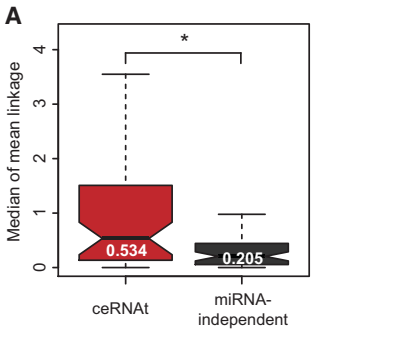
We next sought to identify lnceRNAs that post-transcriptionally regulate the abundance of functionally related mRNAs. To do so, for each lnceRNA we compared the mean of link weights for all its target transcripts to that expected from simulated data (Methods). For 28 of 53 (53%) lnceRNAs, we found that their ceRNAts were significantly more functionally related (empirical $P < 0.05$) than expected.

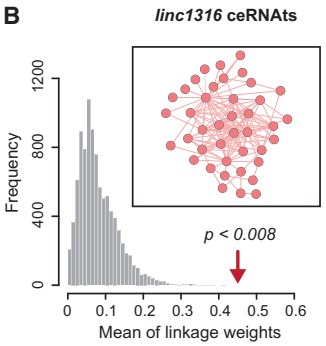
These mRNA targets were enriched in genes involved in the regulation of the cell cycle, developmental process, cell signaling/communication or regulation of differentiation (Supplemental Table S9) which is consistent with the originally proposed roles of these lncRNAs in regulating the circuitry underlying cell-state decision in mESCs (Guttman et al. 2011). For example, depletion of *linc1316* led to changes in the levels of 62 ceRNAts and 57 miRNA-independent targets. These lnceRNAs' ceRNAts are strongly functionally interrelated (Fig. 5B) and are enriched in genes involved in the regulation of cellular differentiation and developmental process (Fig. 5C). In contrast, miRNA-independent targets of *linc1316* are neither functionally related nor are their annotations enriched in any particular GO biological process (Supplemental Fig. S10; Ashburner et al. 2000). Interestingly, *linc1316* harbors predicted binding sites for well-known miRNAs involved in the maintenance of ESC pluripotency, including miR-290-295 and miR-200 families (Peter 2009; Melton et al. 2010; Lichner et al. 2011), and is likely to post-transcriptionally regulate transcriptional factors that promote stem cell self-renewal, such as *Hmga2* (Nishino et al. 2008) and *Myc* (Singh and Dalton 2009) that are down-regulated upon its knockdown.

## Discussion

Here we used experimental and computational genomics approaches to investigate the prevalence and properties of lnceRNAs. We focused our analysis on lncRNAs with proposed roles in the circuitry underlying pluripotent and differentiated cell states. Not only have most lnceRNAs been described in the context of this critical cellular transition but their impact on gene expression profiles, when pools of miRNAs are limited and changes in their repertoires can lead to repression or activation of transcriptional programs, is likely to be greater than in fully differentiated cells or in homeostasis. Environmental or cellular stress, for example upon starvation or infection (Ebert et al. 2007; Franco-Zorrilla et al. 2007), may also offer similar opportunities for strong effects of small changes in target gene expression.

We integrated publicly available data on the transcriptome-wide impact of depletion of over 140 lncRNAs, in mESCs




C

| GO Term | Description | Enrichment | FDR q-value |
|---|---|---|---|
| GO:0007389 | pattern specification process | 7.9 | 8.9x10-4 |
| GO:0009888 | tissue development | 5.0 | 1.7x10-3 |
| GO:0048523 | negative regulation of cellular process | 3.2 | 2.0x10-3 |
| GO:0003002 | regionalization | 8.4 | 2.0x10-3 |
| GO:0065007 | biological regulation | 1.5 | 5.2x10-3 |
| GO:0019222 | regulation of metabolic process | 2.0 | 7.7x10-3 |
| GO:0006357 | regulation of transcription from RNA polymerase II promoter | 3.9 | 1.0x10-2 |
| GO:0030154 | cell differentiation | 2.5 | 1.7x10-2 |
| GO:0032502 | developmental process | 1.9 | 2.5x10-2 |
| GO:0045893 | positive regulation of transcription, DNA-dependent | 4.2 | 4.2x10-2 |

**Figure 5.** ceRNAts of lnceRNAs are functionally related. (*A*) Distribution of the mean linkage in an integrative functional network (Honti et al. 2014), for ceRNAts (median of mean linkage = 0.534, red) and miRNA-independent targets (median of mean linkage = 0.205, dark gray) of lnceRNAs. (*B*) Distribution of mean linkage weights for 1000 sets of randomly selected mESC-expressed genes. The red arrow indicates the mean linkage weight for ceRNAts of *linc1316*. The *inset* illustrates the connectivity of functional similarities (red edges) within miRNA-dependent target genes (red nodes). (*C*) Table illustrating Biological Processes Gene Ontology (GO) annotations that are significantly enriched within ceRNAts of *linc1316* relative to a background of mESC-expressed genes. (*) $P < 0.05$.

(Guttman et al. 2011) with in-house RNA sequencing profiles of mESCs following conditional loss of *Dicer1*, a key component of the miRNA biogenesis pathway. Our analysis indicates that over 50% of these lncRNAs induce changes in their targets in a miRNA-dependent manner. Furthermore, we predict that 88% of lncRNAs share miRNA binding sites with transcription factor encoding mRNAs (Zhang et al. 2012) that are also down-regulated upon lncRNA knockdown, suggesting that the changes induced via this miRNA-mediated mechanism can lead to secondary transcriptionally regulated effects.

The properties of lnceRNAs and their interactions with their mRNA targets are consistent with established rules of post-transcriptional regulation by miRNAs. More specifically, lnceRNAs are enriched in the cytoplasm and their interactions with their functionally related mRNA targets are dependent on the presence of miRNAs. The almost twofold higher density, relative to 3′ UTRs, of predicted MREs within lnceRNAs, together with their relative high expression, argues for these transcripts' enhanced ability to modulate their target levels post-transcriptionally. The increased evolutionary sequence constraint within MREs, in particular those shared between lnceRNAs and their ceRNAts, supports the conservation of these transcripts' regulatory roles, suggesting their biological relevance.

Considered together, our results are consistent with a high prevalence of miRNA-mediated interactions between lncRNAs, particularly those enriched in the cytoplasm, and their mRNA targets and furthermore suggest that this mechanism of lncRNA function, which hitherto has been relatively poorly studied, deserves further scrutiny, particularly in the context of the regulation of cell-fate decisions.

# Methods

## Mouse embryonic stem cell tissue culture

Feeder-free mouse DTCM23/49 XY embryonic stem cells were obtained as follows: mESC cultured in a feeder layer of irradiated MEFs as described by Nesterova et al. (2008) were passaged every 2 d (for 20 d) onto 0.1% gelatin-coated tissue culture dishes. The feeder-independent line was characterized by analyzing cells, mRNA expression levels of *Dicer1*, and key pluripotency markers and levels of MEF-specific and ES-specific miRNAs (data not shown but available upon request).

Mouse DTCM23/49 XY embryonic stem cells (Nesterova et al. 2008) were grown on 0.1% gelatin-coated plates in a culture medium at 37°C in a humidified atmosphere supplemented with 5% $CO_2$, in Knockout Dulbecco's Modified Eagle Medium (D-MEM, Invitrogen, #10829-018) containing 1% antibiotic penicillin/streptomycin (Invitrogen, #15140-122), supplemented with 10% fetal bovine serum (FBS, Invitrogen, #10108-165), 100 µM nonessential amino acids (Invitrogen, #11140-076), 100 µM L-glutamine (Invitrogen, #25030-024), 60 µM β-mercaptoethanol (Invitrogen, #31350-010), and 0.01% leukemia inhibitory factor (LIF, Millipore #ESG1106).

mESC cells were seeded at a density of $8.0 \times 10^5$ cells/dish in 10-cm$^2$ dishes and grown for 24 h prior to tamoxifen treatment. In triplicate, deletion of *Dicer1*'s RNAse III domain was induced by culturing the cells in the presence of 800 nM tamoxifen ([Z]-4-Hydroxytamoxifen [4-OHT], Sigma, #H7904). Cells treated with 0.1% ethanol were used as a control (three replicates). *Dicer1*-deficient colonies were selected and expanded from 10-cm$^2$ to 75-cm$^2$ tissue culture flasks (T-75). Cells were passaged at 70%–80% confluence (every 2–3 d) for 12 d.

## RNA extraction

Subcellular fractionation of mESCs was carried out using the PARIS kit (Ambion, #AM1921) following the manufacturer's instructions before (day 0) and 12 d after treatment with tamoxifen. RNA from total cell extracts and subcellular fractions was extracted using the miRNeasy kit (Qiagen, #217004), and genomic DNA was removed from all RNA extractions using the DNA-free kit (Ambion, #AM1906).

## Quantification of miRNA abundance

Mouse ESCs were harvested and total RNA was extracted using the miRNeasy kit (Qiagen, #217004) in quadruplicate. A total of 611 mouse and murine virus-associated miRNAs were quantified using the nCounter miRNA Expression Assay (NanoString Technologies) (Geiss et al. 2008) according to the manufacturer's instructions. Briefly, input RNA (105 ng) and miRtag linkers were ligated prior to hybridization with barcoded reporter and biotinylated capture probes at 65°C for 16 h. Samples were prepared for analysis on the nCounter Prep Station before data were collected at 555 FOV

on the nCounter Digital Analyzer. Data were analyzed using the NanoString Differential Expression (NanoStriDE) interface (Brumbaugh et al. 2011). Genome-wide miRNA abundance was normalized using a set of housekeeping mRNAs (Supplemental Table S10). Unique miRNAs were grouped into miRNA families (as annotated by TargetScan, v6.2) (Garcia et al. 2011) and their expression levels (normalized counts) combined.

### RNA sequencing, mapping, and quantification of gene expression

Directional poly(A)-selected RNA sequencing libraries were prepared and sequenced (Illumina HiSeq 2000) by BGI Tech Solutions. Total cellular poly(A)-selected RNA samples at day 0 and day 12 after tamoxifen treatment were sequenced to a depth of ~100 million (minimum 93 million; maximum 123 million) 100-bp paired-end reads per sample. Approximately 33 million (minimum 27 million; maximum 41 million) 50-bp paired-end reads per total cellular RNA extracts at days 4, 8, and 10 were sequenced. Cytosolic and nuclear RNA extracts were multiplexed and sequenced on one lane, yielding, on average, ~51 million (minimum 45 million; maximum 69 million) 50-bp paired-end reads.

Reads were aligned to the mouse reference genome (mm9) using TopHat (version 2.0.9) (Trapnell et al. 2009). Splice junctions from Ensembl build 70 (Flicek et al. 2012) were provided to facilitate read mapping across known splice junctions. Reads with paired mates mapping to distinct chromosomes were discarded. On average, 91.0% (minimum 78%; maximum 99%) of RNA sequencing reads was successfully mapped to the mouse genome (mm9). To account for differences in RNA sequencing depth across the five time points following *Dicer1* loss-of-function (day 0, 4, 8, 10, and 12), we considered the smallest number of mapped reads (27 million, day 10), and randomly sampled the same number of mapped reads from the remaining samples collected at the five time points. The number of subsampled RNA sequencing reads covering constitutively expressed nucleotides of lncRNAs (Guttman et al. 2011) and Ensembl build 70 protein-coding gene and lncRNA annotations (Flicek et al. 2012) were estimated using HTSeq (version 0.6.1) (Quinlan and Hall 2010) using default parameters with a minimum 1 aligned read to the respective genes. Expression levels were estimated as total fragments per kilobase of exon per million fragments mapped (FPKM) across the different libraries at each of the five time points and for each of the replicates.

To compare the abundance in the nuclear and cytoplasmic fraction of lncRNAs and their mRNAs targets before and after *Dicer1* loss-of-function (day 0 and day 12), locus expression level was determined in each of the compartments independently as previously described, and this was used to calculate the ratio between expression levels in the cytoplasm and nucleus.

Multidimensional scaling (MDS) analysis was performed using the edgeR package (Robinson et al. 2010).

### Quantitative PCR

RNA (1 µg per 20 µL sample reaction) was reverse-transcribed into cDNA using the QuantiTect Reverse Transcription kit (Qiagen, #205313) according to the manufacturer's instructions. For miRNA quantification, RNA was reverse-transcribed using the NCode VILO miRNA cDNA Synthesis Kit (Invitrogen, #A11193-050). Expression levels were estimated by real-time quantitative PCR (qRT-PCR) on a StepOneReal-Time PCR thermocycler (ABI) using SYBR green Master PCR mix (ABI, #4309155) and loci-specific primers for mRNAs of interest and for the following miRNAs: miR-124, miR-16, miR-200a, miR-290a-3p, miR-302a in triplicate (Supplemental Table S10). The expression levels of quantified mature miRNAs, lncRNAs, and mRNAs were normalized to that of *Gapdh*. Non-reverse-transcribed RNA was used as negative amplification control.

### Prediction of miRNA response elements

TargetScan (version 6.2) (Garcia et al. 2011) was used to predict miRNA response elements in sequences of mouse lncRNAs (Guttman et al. 2011) and the longest 3′ UTRs of protein-coding mRNAs expressed in mouse ESCs (Ensembl build 70) (Flicek et al. 2012). As controls, a set of shuffled miRNA seed sequences was generated using uShuffle (Jiang et al. 2008). For each seed of the top 25% most highly expressed miRNAs, all possible shuffled combinations were generated (153 random seeds) using default parameters and with a shuffling size of 2 to maintain the dinucleotide frequencies of the shuffled sequences. After removing those homologous to all mESC-expressed miRNAs, a set of 80 shuffled MREs ($MRE_{shuffle}$) were obtained.

A conservative set of experimentally validated MREs was obtained by considering computationally predicted MREs overlapping (100% coverage) regions of the mouse genome enriched in Argonaute binding according to high-throughput CLIP-sequencing analysis in mESCs (Leung et al. 2011). We considered the peaks as annotated in the original study (Leung et al. 2011).

### Coexpression between lncRNAs and mRNA targets

For each lncRNA, we calculated its pairwise Pearson's correlation in expression across 12 d following *Dicer1* loss-of-function, with all its mRNAs targets, defined as genes that are differentially expressed in the lncRNA loss-of-function experiment conducted by Guttman et al. (2011). We considered only lncRNAs with evidence of expression across the 12-d time point (110 lncRNAs) and with more than two ceRNAts or miRNA-independent targets. For these lncRNAs (104), we calculated the median correlation coefficient for their ceRNAts and for their miRNA-independent targets and compared these values to what would be expected based on the median pairwise correlation between the lncRNAs and 1000 randomly selected groups of mESC-expressed genes sampled to the same size as ceRNAts or miRNA-independent sets. As a control, we repeated the analysis above for sets of putative ceRNAts and up- and down-regulated targets, defined using shuffled MREs.

### Experimental validation of miRNA-dependent lncRNA and ceRNAt interaction

We selected the most efficient small interference RNA (siRNA) oligo targeting *linc1582*, *linc1405*, and *linc1283* reported by Guttman et al. (2011) (Supplemental Table S9). One day prior to transfection, mESCs ($1.0 \times 10^5$ cells/mL) were seeded in six-well dishes. Knockdown siRNA and negative control oligos (Qiagen, #1027280) were transfected in triplicate into $Dcr^{+/+}$ (5 nM/well) and in $Dcr^{-/-}$ (10 nM/well) mESCs using Lipofectamine RNAiMAX Reagent (Invitrogen) according to the manufacturer's guidelines. Cells were grown under standard conditions for 48 h post-transfection and RNA extracted and reverse-transcribed as described above.

For each of the following three lncRNAs, *linc1253, linc1405*, and *linc1582*, we selected 10 putative ceRNAts and measured their expression by qPCR following lncRNA siRNA in $Dcr^{-/-}$ and $Dcr^{+/+}$ mESCs. Gene expression estimates were normalized to *Gadph*. We excluded from this analysis five targets whose levels were not significantly changed following lncRNA knockdown in $Dcr^{+/+}$. For each putative ceRNAt, we estimated its fold-change in expression following lncRNA depletion in $Dcr^{-/-}$ relative to $Dcr^{+/+}$ as follows:

fold difference in expression = (expression of ceRNAt $Dcr^{-/-}$ – expression of ceRNAt $Dcr^{+/+}$)/(expression of ceRNAt $Dcr^{+/+}$). Targets whose levels were significantly depleted (after Bonferroni multiple test correction, $P < 0.005$) following lncRNA knockdown in $Dcr^{+/+}$ mESC but remained unchanged in $Dcr^{-/-}$ mESC were classified as true positives. The remaining targets whose levels changed significantly upon lncRNA knockdown in $Dcr^{+/+}$ and $Dcr^{-/-}$ mESC were classified as false positives.

### Transcription factor analysis

We considered the 179 genes represented in the microarray used by Guttman and colleagues annotated as transcription factors (TFs) in AnimalTFDB (Zhang et al. 2012). We found that for 88% (91 of 104) of lncRNAs, their knockdown was associated with significant down-regulation of mESC-expressed transcription factors with whom they share MREs.

### Conservation of mouse lncRNA expression in humans

Poly(A)-selected RNA sequencing data from human embryonic stem cell (H1 hESC) (Bernstein et al. 2012) were mapped to the syntenic regions, in humans, of the 104 mouse lncRNAs that did not overlap annotated mouse protein-coding and showed expression evidence in our data set (obtained using liftOver [Meyer et al. 2013] with parameters: -minMatch = 0.2 -minBlocks = 0.01). A mouse lncRNA with at least five sequencing reads covering 20% or more of the syntenic region in humans was considered to be conserved in expression. The number of aligned reads was estimated using HTSeq (version 0.6.1) (Quinlan and Hall 2010) using default parameters, and read coverage was assessed using coverageBed (BEDTools version 2.17.0) (Quinlan and Hall 2010) with the –split option. At this cutoff, both the median depth (0%) and coverage (0 reads) of the human syntenic regions by hESCs RNA sequencing reads for 10,000 randomly selected sets of intervals in the mouse genome with the same length as the lncRNAs considered but with no evidence of transcriptional activity in mESCs (no reads across the entire region) are zero.

### Nucleotide substitution rates

Pairwise alignments of the different sequence features (all, shared, and nonshared MREs and non-MREs), between mouse (mm9) and human (hg19), were concatenated: shared MREs = 553 (3318 bp), nonshared MREs = 95 (570 bp), non-MREs = (23.346 bp), and $MRE_{shuffled}$ = 207 (1242 bp). Mouse and human alignments between neighboring (within 1 Mb) and nonoverlapping ancestral repeats (ARs), a good proxy for neutrally evolving sequence (Lunter et al. 2006), were used to simulate (1000 times) sequence alignments with similar G + C content and size, in mouse, of each of the considered sequences. Nucleotide substitution rates were estimated using the *REV* substitution model in *baseml* from the PAML package (Yang 1997).

To obtain empirical *P*-values, the estimated nucleotide substitution rate across the concatenated alignment of the sequences of interest was compared to the estimates obtained for the respective simulated putatively neutral sequence alignments.

### Integrated functional linkage network analysis

Functional similarity between ceRNAts or miRNA-independent target sets for each lnceRNA was estimated using an integrative phenotypic-linkage network of mouse protein-coding genes (Honti et al. 2014). For each lnceRNA, the median of functional linkages (measure of functional similarity) between lnceRNA targets in each group was calculated. These were then compared to a distribution of the same measures obtained from 1000 random bootstrapped gene sets, which were gene length-matched and mESC-expressed, containing the same number of genes as that in the gene set of interest. Functional linkages between lnceRNA targets (nodes) are represented as edges connecting the nodes using Cytoscape (Shannon et al. 2003).

Gene Ontology (GO) enrichment analysis was performed using the functional classification tool Database for Annotation, Visualization, and Integrated Discovered (DAVID, default parameters) (Huang et al. 2009) using default parameters and all mESC expressed genes (FPKM > 0) as background. The list of significantly enriched GO terms (after Benjamini-Hochberg correction) was summarized using REVIGO with default parameters (Supek et al. 2011), and only nonredundant common ancestral terms were reported.

### Statistics

All statistical analyses were done using the R package (R Core Team 2011). Asterisks indicate significance in the level of the comparison between the expressions of target transcripts ([*] $P < 0.05$; [**] $P < 0.01$; [***] $P < 0.001$; NS [not significant] $P > 0.05$). For each experimental analysis, statistical values were calculated using data collected from three or more independent experiments.

## Data access

The raw sequencing data and estimated transcript expression for the temporal profiling of mouse mESCs following loss of *Dicer1* function have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE58757.

## Acknowledgments

We thank Tudor Fulga and Bruno Steinkraus for help with miRNA quantification and members of the ACM and CPP groups for insightful comments and suggestions. This work was funded by the SNF (A.C.M., J.Y.T.); Medical Research Council (C.P.P., C.W., B.G., M.M.); Dorothy Hodgkin's Fellowship from the Royal Society (A.C.M.); European Research Council Advanced Grant (C.P.P., A.C.M., T.S.); and GENCODYS FP7 program grant (F.H., C.W., 251995); Clarendon Fund (J.Y.T.); and the Natural Sciences Engineering Research Council of Canada (J.Y.T.).

## References

Ala U, Karreth FA, Bosia C, Pagnani A, Taulli R, Leopold V, Tay Y, Provero P, Zecchina R, Pandolfi PP. 2013. Integrated transcriptional and competitive endogenous RNA networks are cross-regulated in permissive molecular environments. *Proc Natl Acad Sci* **110:** 7154–7159.

Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, Dreyfuss G, Eddy SR, Griffiths-Jones S, Marshall M, et al. 2003. A uniform system for microRNA annotation. *RNA* **9:** 277–279.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. 2000. Gene Ontology: tool for the unification of biology. *Nat Genet* **25:** 25–29.

Bartel DP. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116:** 281–297.

Bartel DP. 2009. MicroRNAs: target recognition and regulatory functions. *Cell* **136:** 215–233.

Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489:** 57–74.

Bosia C, Pagnani A, Zecchina R. 2013. Modelling competing endogenous RNA networks. *PLoS One* **8:** e66609.

Bosson AD, Zamudio JR, Sharp PA. 2014. Endogenous miRNA and target concentrations determine susceptibility to potential ceRNA competition. *Mol Cell* **56:** 347–359.

Broderick JA, Zamore PD. 2014. Competitive endogenous RNAs cannot alter microRNA function in vivo. *Mol Cell* **54:** 711–713.
Brumbaugh CD, Kim HJ, Giovacchini M, Pourmand N. 2011. NanoStriDE: normalization and differential expression analysis of NanoString nCounter data. *BMC Bioinformatics* **12:** 479.
Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* **25:** 1915–1927.
Cesana M, Cacchiarelli D, Legnini I, Santini T, Sthandier O, Chinappi M, Tramontano A, Bozzoni I. 2011. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* **147:** 358–369.
Denzler R, Agarwal V, Stefano J, Bartel DP, Stoffel M. 2014. Assessing the ceRNA hypothesis with quantitative measurements of miRNA and target abundance. *Mol Cell* **54:** 766–776.
Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, et al. 2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* **22:** 1775–1789.
Ebert MS, Sharp PA. 2010. Emerging roles for natural microRNA sponges. *Curr Biol* **20:** R858–R861.
Ebert MS, Neilson JR, Sharp PA. 2007. MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nat Methods* **4:** 721–726.
Fan M, Li X, Jiang W, Huang Y, Li J, Wang Z. 2013. A long non-coding RNA, PTCSC3, as a tumor suppressor and a target of miRNAs in thyroid cancer cells. *Exp Ther Med* **5:** 1143–1146.
Figliuzzi M, Marinari E, De Martino A. 2013. MicroRNAs as a selective channel of communication between competing RNAs: a steady-state theory. *Biophys J* **104:** 1203–1213.
Flicek P, Amode MR, Barrell D, Beal K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fairley S, Fitzgerald S, et al. 2012. Ensembl 2012. *Nucleic Acids Res* **40:** D84–D90.
Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI, Rubio-Somoza I, Leyva A, Weigel D, Garcia JA, Paz-Ares J. 2007. Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat Genet* **39:** 1033–1037.
Garcia DM, Baek D, Shin C, Bell GW, Grimson A, Bartel DP. 2011. Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. *Nat Struct Mol Biol* **18:** 1139–1146.
Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, Fell HP, Ferree S, George RD, Grogan T, et al. 2008. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol* **26:** 317–325.
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.
Han K, Gennarino VA, Lee Y, Pang K, Hashimoto-Torii K, Choufani S, Raju CS, Oldham MC, Weksberg R, Rakic P, et al. 2013. Human-specific regulation of MeCP2 levels in fetal brains by microRNA miR-483-5p. *Genes Dev* **27:** 485–490.
Hausser J, Zavolan M. 2014. Identification and consequences of miRNA-target interactions—beyond repression of gene expression. *Nat Rev Genet* **15:** 599–612.
Honti F, Meader S, Webber C. 2014. Unbiased functional clustering of gene variants with a phenotypic-linkage network. *PLoS Comput Biol* **10:** e1003815.
Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4:** 44–57.
Jens M, Rajewsky N. 2015. Competition between target sites of regulators shapes post-transcriptional gene regulation. *Nat Rev Genet* **16:** 113–126.
Jia H, Osak M, Bogu GK, Stanton LW, Johnson R, Lipovich L. 2010. Genome-wide computational identification and manual annotation of human long noncoding RNA genes. *RNA* **16:** 1478–1487.
Jiang M, Anderson J, Gillespie J, Mayne M. 2008. uShuffle: a useful tool for shuffling biological sequences while preserving the k-let counts. *BMC Bioinformatics* **9:** 192.
Karwacki-Neisius V, Goke J, Osorno R, Halbritter F, Ng JH, Weisse AY, Wong FC, Gagliardi A, Mullin NP, Festuccia N, et al. 2013. Reduced Oct4 expression directs a robust pluripotent state with distinct signaling activity and increased enhancer occupancy by Oct4 and Nanog. *Cell Stem Cell* **12:** 531–545.
Leung AK, Young AG, Bhutkar A, Zheng GX, Bosson AD, Nielsen CB, Sharp PA. 2011. Genome-wide identification of Ago2 binding sites from mouse embryonic stem cells with and without mature microRNAs. *Nat Struct Mol Biol* **18:** 237–244.
Lichner Z, Pall E, Kerekes A, Pallinger E, Maraghechi P, Bosze Z, Gocza E. 2011. The miR-290–295 cluster promotes pluripotency maintenance by regulating cell cycle phase distribution in mouse embryonic stem cells. *Differentiation* **81:** 11–24.
Loewer S, Cabili MN, Guttman M, Loh YH, 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. *Nat Genet* **42:** 1113–1117.
Lunter G, Ponting CP, Hein J. 2006. Genome-wide identification of human functional DNA using a neutral indel model. *PLoS Comput Biol* **2:** 2–12.
Marques AC, Tan J, Ponting CP. 2011. Wrangling for microRNAs provokes much crosstalk. *Genome Biol* **12:** 132.
Marques AC, Tan J, Lee S, Kong L, Heger A, Ponting CP. 2012. Evidence for conserved post-transcriptional roles of unitary pseudogenes and for frequent bifunctionality of mRNAs. *Genome Biol* **13:** R102.
Maziere P, Enright AJ. 2007. Prediction of microRNA targets. *Drug Discov Today* **12:** 452–458.
Melton C, Judson RL, Blelloch R. 2010. Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. *Nature* **463:** 621–626.
Meyer LR, Zweig AS, Hinrichs AS, Karolchik D, Kuhn RM, Wong M, Sloan CA, Rosenbloom KR, Roe G, Rhead B, et al. 2013. The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res* **41:** D64–D69.
Mukherji S, Ebert MS, Zheng GXY, Tsang JS, Sharp PA, van Oudenaarden A. 2011. MicroRNAs can generate thresholds in target gene expression. *Nat Genet* **43:** 854–859.
Nesterova TB, Popova BC, Cobb BS, Norton S, Senner CE, Tang YA, Spruce T, Rodriguez TA, Sado T, Merkenschlager M, et al. 2008. Dicer regulates Xist promoter methylation in ES cells indirectly through transcriptional control of Dnmt3a. *Epigenetics Chromatin* **1:** 2.
Nishino J, Kim I, Chada K, Morrison SJ. 2008. Hmga2 promotes neural stem cell self-renewal in young but not old mice by reducing p16Ink4a and p19Arf Expression. *Cell* **135:** 227–239.
Paraskevopoulou MD, Georgakilas G, Kostoulas N, Reczko M, Maragkakis M, Dalamagas TM, Hatzigeorgiou AG. 2013. DIANA-LncBase: experimentally verified and computationally predicted microRNA targets on long non-coding RNAs. *Nucleic Acids Res* **41:** D239–D245.
Peter ME. 2009. Let-7 and miR-200 microRNAs: guardians against pluripotency and cancer progression. *Cell Cycle* **8:** 843–852.
Poliseno L, Haimovic A, Christos PJ, Vega Y, Saenz de Miera EC, Shapiro R, Pavlick A, Berman RS, Darvishian F, Osman I. 2011. Deletion of PTENP1 pseudogene in human melanoma. *J Invest Dermatol* **131:** 2497–2500.
Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26:** 841–842.
R Core Team. 2011. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.
Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26:** 139–140.
Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. 2011. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* **146:** 353–358.
Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13:** 2498–2504.
Singh AM, Dalton S. 2009. The cell cycle and Myc intersect with mechanisms that regulate pluripotency and reprogramming. *Cell Stem Cell* **5:** 141–149.
Stark A, Brennecke J, Bushati N, Russell RB, Cohen SM. 2005. Animal microRNAs confer robustness to gene expression and have a significant impact on 3′UTR evolution. *Cell* **123:** 1133–1146.
Sumazin P, Yang X, Chiu HS, Chung WJ, Iyer A, Llobet-Navas D, Rajbhandari P, Bansal M, Guarnieri P, Silva J, et al. 2011. An extensive microRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma. *Cell* **147:** 370–381.
Sun L, Goff LA, Trapnell C, Alexander R, Lo KA, Hacisuleyman E, Sauvageau M, Tazon-Vega B, Kelley DR, Hendrickson DG, et al. 2013. Long noncoding RNAs regulate adipogenesis. *Proc Natl Acad Sci* **110:** 3387–3392.
Supek F, Bosnjak M, Skunca N, Smuc T. 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6:** e21800.
Tan JY, Vance KW, Varela MA, Sirey T, Watson LM, Curtis HJ, Marinello M, Alves S, Steinkraus BR, Cooper S, et al. 2014. Cross-talking noncoding RNAs contribute to cell-specific neurodegeneration in SCA7. *Nat Struct Mol Biol* **21:** 955–961.
Tay Y, Rinn J, Pandolfi PP. 2014. The multilayered complexity of ceRNA crosstalk and competition. *Nature* **505:** 344–352.
Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25:** 1105–1111.
Ulitsky I, Bartel DP. 2013. lincRNAs: genomics, evolution, and mechanisms. *Cell* **154:** 26–46.
van Heesch S, van Iterson M, Jacobi J, Boymans S, Essers PB, de Bruijn E, Hao W, Macinnes AW, Cuppen E, Simonis M. 2014. Extensive localization of

long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. *Genome Biol* **15:** R6.
Wang Y, Xu Z, Jiang J, Xu C, Kang J, Xiao L, Wu M, Xiong J, Guo X, Liu H. 2013. Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. *Dev Cell* **25:** 69–80.
Wehrspaun CC, Ponting CP, Marques AC. 2014. Brain-expressed 3′UTR extensions strengthen miRNA cross-talk between ion channel/transporter encoding mRNAs. *Front Genet* **5:** 41.
Wienholds E, Plasterk RH. 2005. MicroRNA function in animal development. *FEBS Lett* **579:** 5911–5922.
Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* **13:** 555–556.
Yuan Y, Liu B, Xie P, Zhang MQ, Li Y, Xie Z, Wang X. 2015. Model-guided quantitative analysis of microRNA-mediated regulation on competing endogenous RNAs using a synthetic gene circuit. *Proc Natl Acad Sci* **112:** 3158–3163.
Zhang HM, Chen H, Liu W, Liu H, Gong J, Wang H, Guo AY. 2012. AnimalTFDB: a comprehensive animal transcription factor database. *Nucleic Acids Res* **40:** D144–D149.
Zheng GX, Do BT, Webster DE, Khavari PA, Chang HY. 2014. Dicer-microRNA-Myc circuit promotes transcription of hundreds of long noncoding RNAs. *Nat Struct Mol Biol* **21:** 585–590.

*Received July 22, 2014; accepted in revised form March 17, 2015.*