Published in final edited form as:

Nat Struct Mol Biol. 2014 July ; 21(7): 585–590. doi:10.1038/nsmb.2842.

Dicer-microRNA-Myc circuit promotes transcription of hundreds of long noncoding RNAs

Grace X Y Zheng¹, Brian T Do¹, Dan E Webster¹, Paul A Khavari¹, and Howard Y Chang^{1,2}

¹Program in Epithelial Biology, Stanford University School of Medicine, Stanford, California, USA

²Howard Hughes Medical Institute, Stanford, California, USA

Abstract

Long noncoding RNAs (lncRNAs) are important regulators of cell fate, yet little is known about mechanisms controlling lncRNA expression. Here we show that transcription is quantitatively different for lncRNAs and mRNAs—as revealed by deficiency of Dicer (Dcr), a key RNase that generates microRNAs (miRNAs). Dcr loss in mouse embryonic stem cells led unexpectedly to decreased levels of hundreds of lncRNAs. The canonical Dgcr8-Dcr-miRNA pathway is required for robust lncRNA transcriptional initiation and elongation. Computational and genetic epistasis analyses demonstrated that Dcr activation of the oncogenic transcription factor cMyc is partly responsible for lncRNA expression. A quantitative metric of mRNA-lncRNA decoupling revealed that Dcr and cMyc differentially regulate lncRNAs versus mRNAs in diverse cell types and *in vivo*. Thus, numerous lncRNAs may be modulated as a class in development and disease, notably where Dcr and cMyc act.

Mammalian genomes encode thousands of noncoding RNAs. Two classes of noncoding RNAs, lncRNA and miRNAs, have important roles in regulating gene expression. lncRNAs are defined as transcripts of >200 nt that lack a canonical open reading frame and do not function by encoding proteins. lncRNAs are synthesized by RNA polymerase II, spliced and sometimes polyadenylated¹. Many lncRNAs can regulate chromatin states by interacting with chromatin- modification complexes, which control diverse developmental processes and human diseases, including cancer².

miRNAs are endogenous ~22-nt RNAs that are processed by the RNase III enzymes Drosha and Dcr. miRNAs post-transcriptionally regulate the expression of thousands of mRNAs through interaction between the 5' end of miRNAs (miRNA seeds) and the 3' untranslated

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

Correspondence should be addressed to H.Y.C. (howchang@stanford.edu).

Accession codes. Data have been deposited in the Gene Expression Omnibus (GEO) under accession number GSE55338.

AUTHOR CONTRIBUTIONS

G.X.Y.Z. and H.Y.C. initiated the project. G.X.Y.Z. and H.Y.C. designed the experiments. G.X.Y.Z. performed the experiments and the computational analysis. B.T.D., D.E.W. and P.A.K. designed and implemented bioinformatics and microarray screens. The manuscript was prepared by G.X.Y.Z. and H.Y.C. with input from all authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

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

region (UTR)³. Although lncRNAs lack 3' UTRs, lncRNAs have been reported to be downregulated by miRNAs^{4–6}. In addition, lncRNAs can titrate away the effect of miRNAs on endogenous mRNA targets by acting as a 'sponge'⁷. Nonetheless, the transcriptional-control mechanism of lncRNAs is not well understood, and how the biogenesis of lncRNAs and miRNAs may intersect has not been addressed on a genome-wide scale.

Here, we took a systematic approach to study the global effect of miRNAs on lncRNA expression in mouse embryonic stem cells (mESCs), in which individual miRNAs and lncRNAs have been extensively studied. The miR-295 family constitutes the dominant miRNA population in mESCs⁸, and it has been linked to a number of functions in embryonic stem (ES) cells, including maintenance of pluripotency and proliferation⁹. Similarly, thousands of lncRNAs are expressed in mESCs and have been implicated in both pluripotency maintenance and regulation of differentiation potential^{10,11}.

We used extensive RNA deep sequencing (RNA-seq) to study the mESC transcriptome before and after the loss of Dcr and identified an unexpected transcriptional-regulatory circuit for lncRNAs. Experiments with loss-of-function mutants of other components of the miRNA-biogenesis machinery further validated the role of Dcr and cMyc in this circuit. Our study opens up the possibility that lncRNAs are regulated as a class, in a genome-wide phenomenon with important developmental and clinical implications.

RESULTS

lncRNAs are globally downregulated in Dcr-knockout mESCs

To assess the effect of loss of Dcr and miRNAs on the expression of lncRNAs, we performed paired-end deep sequencing of polyadenylated RNA species from wild-type (WT) and Dcr-knockout (KO) mESCs. We used two biological-replicate cultures of WT mESC lines and three independently derived Dcr-KO mESC lines to generate over 800 million RNA-seq reads. We aligned them with TopHat¹², assembled transcripts with Cufflinks¹³ and classified the transcripts by using existing databases from the University of California, Santa Cruz, and published lncRNA annotations^{10,11,14,15}. Because RNA content per cell was not significantly different between WT and Dcr KO, we normalized the data to total RNA level. In total, we detected 14,348 coding RNAs and 2,229 known lncRNAs in WT and Dcr-KO mESCs (Fig. 1a).

We evaluated the quality of the sequencing data by comparing the expression of our transcripts against existing transcriptome data from mESCs. The comparison of transcript expression between the WT and another ES-cell line from ref. 11 revealed a high degree of similarity (Pearson correlation coefficient $r = 0.91$). In addition, the expression of protein-coding genes in WT and Dcr-KO mESCs agreed well with published microarray data¹⁶ ($r = 0.80$ and 0.81 respectively). We also found that the expression of a few well-known targets of miR-295, such as *Casp2*, showed a similar level of derepression in Dcr-KO mESCs, as previously reported⁹ (Supplementary Fig. 1a,b). We calculated a cumulative density function (cdf) plot comparing expression differences for the set of miR-295 targets as determined by seed complementarity (7-mer M8 or T1A). Relative to a control set of genes matched for 3'-UTR length, dinucleotide composition and expression level, the miR-295-target set was

more derepressed upon Dcr loss (Supplementary Fig. 1f,g). This result shows that our sequencing data were able to faithfully show a transcriptome-wide miR-295 signature associated with mESCs.

We next compared lncRNA expression levels in WT versus Dcr-KO mESCs. Similarly to mRNAs, lncRNAs with a complementary seed match to miR-295 were also derepressed upon Dcr loss (Supplementary Fig. 1h). However, to our surprise, the expression of hundreds of lncRNAs decreased in Dcr-KO mESCs ($P = 3.95 \times 10^{-13}$; Fig. 1b,c), in contrast to the expression of mRNAs, whose log₂ fold change between Dcr KO and WT was around 0. We also observed this trend when comparing mRNAs with expression level matched to that of lncRNAs in WT mESCs (Supplementary Fig. 1e). Of 2,229 lncRNAs, 714 were significantly changed in level in Dcr KO (false discovery rate (FDR) = 0.05), of which 64% were downregulated (Fig. 1d). Over 50% (375 out of 714) of significantly regulated lncRNAs showed at least a two-fold downregulation, and over 15% of lncRNAs (113 out of 714) showed at least a four-fold downregulation. Quantitative PCR with reverse transcription (RT-qPCR) of nine lncRNAs (Fig. 1e) and NanoString analysis of 102 lncRNAs (Fig. 2b) validated the decrease of lncRNA levels in Dcr-KO mESCs ($r = 0.99$ and 0.76 respectively; Supplementary Fig. 2a). Interestingly, many of the downregulated lncRNAs are required to maintain the pluri-potency and differentiation potential of mESCs¹⁰, thus suggesting that reduction of lncRNA levels may have important consequences in the biology of mESCs.

Canonical miRNA pathway is required for lncRNA expression

We next examined whether the canonical Dgcr8-Dcr-miRNA pathway was required for lncRNA expression in mESCs (Fig. 2a). First, to test whether downregulated lncRNAs could be rescued by Dcr, we transfected Dcr or a catalytically dead Dcr point mutant into Dcr-KO mESCs and assayed the expression of lncRNAs by RT-qPCR and NanoString analysis. Dcr successfully rescued the expression of nearly all of the downregulated lncRNAs examined (and, as a positive control, Dcr overexpression also rescued the expression of miR-295), whereas the Dcr mutant did not, thus suggesting that the Dcr RNase III domain essential for miRNA maturation is important for lncRNA expression (Fig. 2b and Supplementary Fig. 2b,d). Canonical miRNA biogenesis requires processing of primary miRNAs by DGCR8 to generate substrate hairpin RNAs for DCR (Fig. 2a). Consistently with this model, Dgcr8-KO mESCs also showed significant downregulation of lncRNAs ($r = 0.72$; Fig. 2b and Supplementary Fig. 2c,f,g). Because miR-295 is the major miRNA in mESCs, we hypothesized that loss of miR-295 led to loss of lncRNAs. Indeed, overexpression of miR-295 in Dcr-KO mESCs, but not miR-96 (the second most abundant miRNA in mESCs¹⁶) or a control miRNA that is not expressed in mESCs, rescued the expression of lncRNAs (Fig. 2b and Supplementary Fig. 2b,e). In addition, miR-295 rescued the expression of downregulated lncRNAs in Dgcr8-KO mESCs (Supplementary Fig. 2c,e). Taken together, these data suggest that a stepwise Dgcr8-Dcr-miR295 pathway is responsible for the expression of many lncRNAs in mESCs.

Dicer-miRNA pathway controls lncRNA transcription

To better understand the mechanism of lncRNA downregulation in Dcr-KO mESCs, we performed metabolic-labeling experiments to determine whether the decrease in lncRNA

expression was due to decreased transcription or increased degradation. We treated WT and Dcr-KO mESCs with 4-thio-uridine (4sU) for 15 min (assuming negligible degradation among newly synthesized RNAs within this time frame) and assayed the expression of newly synthesized and total RNAs to calculate the rate of transcription and degradation for each lncRNA¹⁷ (Fig. 3a). We performed the same calculation on several coding mRNAs and on positive controls and used published data¹⁸ to check that our method could accurately measure their half-lives ($r = 0.83$, Supplementary Fig. 3a). Although the half-lives of lncRNAs and mRNAs did not change significantly (and some even increased) in Dcr-KO mESCs, the transcription rates of downregulated lncRNAs were significantly lower in Dcr-KO mESCs (Fig. 3b,c). In contrast, mRNAs with constant expression between WT and Dcr-KO mESCs showed constant rates of transcription (Supplementary Fig. 3b,c), thus suggesting that Dcr loss specifically decreased the transcription of tested lncRNAs.

Chromatin-state analysis further supported the role of Dcr-miRNA pathway in lncRNA synthesis. We performed chromatin immuno-precipitation followed by deep sequencing (ChIP-seq) of trimethylated histone H3 K4 (H3K4me3), which is associated with transcriptional initiation, and trimethylated H3 K36 (H3K36me3), which is associated with transcriptional elongation in both WT and Dcr-KO mESCs. Both modifications have been used to identify and characterize transcription of lncRNA genes¹¹. Although the levels of H3K4me3 and H3K36me3 were not significantly different for lncRNAs whose expression remained constant, Dcr KO led to significant decreases in both H3K4me3 and H3K36me3 levels at downregulated lncRNA loci ($P = 0.025$ and 2.2×10^{-16} respectively, Fig. 3d–g and Supplementary Figs. 1d and 3d,e). This result suggests that both transcriptional initiation and consequent elongation of lncRNA genes are inhibited in Dcr-KO mESCs, with elongation apparently being more severely affected.

cMyc links Dcr pathway to transcription of lncRNA genes

Because lncRNA downregulation in Dcr KO seems to occur at the level of transcription, we hypothesized that these lncRNAs might be regulated by a common transcription factor which itself was downregulated upon Dcr loss. To test this hypothesis, we performed a bioinformatic screen for DNA-binding proteins whose binding sites are enriched near downregulated lncRNA loci relative to control (upregulated lncRNA) loci (Fig. 4a and Online Methods). In addition, we focused on DNA-binding proteins whose expression was downregulated in Dcr-KO mESCs. Notably, cMyc is the only factor that met both criteria. cMyc ChIP-seq data in mESCs showed that downregulated lncRNA genes have almost four times as many cMyc-binding sites as those of upregulated lncRNA genes (Fig. 4b and Supplementary Fig. 4b). Furthermore, cMyc expression decreased by ten-fold upon Dcr loss (Fig. 4a and Supplementary Figs. 1b and 4a), an effect due to the loss of miR-295 expression, as documented in previous studies¹⁹. RNA-seq of cMyc^{-/-} mESCs¹⁹ showed a downregulation of lncRNAs over mRNAs similar to that in Dcr KO (Fig. 4c). Conversely, enforced expression of cMyc in Dcr-KO cells rescued the expression of most lncRNAs examined (Fig. 4d). The identification of an enriched transcription factor further suggests that Dcr regulation of lncRNAs occurs indirectly through transcription.

To further ascertain the role of cMyc in regulating lncRNA expression, we compared the expression changes of lncRNA genes with and without cMyc-binding sites. Indeed, lncRNA genes with cMyc-binding sites showed greater decreases in expression compared to those without cMyc-binding sites in cMyc^{-/-} and Dcr-KO mESCs (Fig. 4c and Supplementary Fig. 4c). Intriguingly, about 40% of the protein-coding genes also had cMyc-binding sites, but they were down-regulated by only 4% in Dcr-KO mESCs (Supplementary Fig. 4d), thus suggesting that lncRNA genes are much more sensitive than protein-coding genes to cMyc expression.

To check whether cMyc affects transcription initiation and elongation of lncRNAs, we compared H3K4me3 and H3K36me3 levels at lncRNA genes with cMyc-binding sites in WT and Dcr-KO cells. Indeed, 99% of the H3K4me3 loss and 95% of the H3K36me3 loss at lncRNA gene loci in Dcr-KO cells occurred at cMyc-target genes but not at non-cMyc targets (Fig. 4e). In contrast to lncRNA genes, levels of only H3K36me3, but not H3K4me3, at protein-coding genes with cMyc-binding sites were affected in Dcr-KO mESCs (Supplementary Fig. 4e,f). In addition, Pol2 ChIP-seq data with and without cMyc inhibition also showed that cMyc inhibition reduced Pol2 elongation at lncRNA genes with cMyc-binding sites (Supplementary Fig. 4g). These results reinforce the idea that cMyc is a major determinant of transcription of lncRNA genes that affects both the transcriptional initiation and elongation.

Genomic organization affects lncRNA sensitivity to Dcr loss

To understand how transcription of lncRNA genes could be different from that of protein-coding genes, we compared the genomic contexts of both classes of transcripts (Supplementary Fig. 5a,b). lncRNA genes were preferentially located adjacent to protein-coding genes (33% lncRNAs relative to 13% mRNAs), and these lncRNA genes were downregulated in Dcr-KO mESCs by 30% more than intergenic lncRNA genes that were not near a protein-coding gene ($P < 2.2 \times 10^{-16}$). Interestingly, although the lncRNAs in the divergent pairs of lncRNA and protein-coding genes were downregulated in Dcr-KO mESCs, their protein-coding-gene counterparts were upregulated (although to a smaller extent), thus suggesting that lncRNA genes and protein-coding genes compete for transcription when they are divergent from each other (Supplementary Fig. 5c).

Because cMyc has an important role in regulating synthesis of both lncRNAs and mRNAs, we wondered whether cMyc-regulated lncRNA genes overlapped with lncRNA genes that were divergent from protein-coding genes. 123 of the downregulated lncRNA genes are both cMyc targets and divergent from an adjacent protein-coding gene (Supplementary Fig. 5d). This 46% overlap is highly significant ($P = 4.4 \times 10^{-43}$, hypergeometric distribution), thus suggesting that cMyc regulation and the genomic organization of genes encoding downregulated lncRNAs may functionally intersect. Moreover, the effect of having both features in a lncRNA gene is not additive, further supporting the idea that the two pathways are not independent from each other (Supplementary Fig. 5e). Thus, cMyc regulation of transcription of lncRNA genes occurs most potently in the specific genomic context of divergent transcription.

Dcr and cMyc regulation of lncRNA expression is widespread

We wondered whether the effect of Dcr and cMyc on the regulation of lncRNA expression applied to additional cell types beyond mESCs. We surveyed microarray data in Gene Expression Omnibus (GEO) and Cancer Cell Line Encyclopedia (CCLE)²⁰ for differential regulation of lncRNAs and mRNAs (Fig. 5a). For the effect of Dcr, we selected microarray experiments from GEO that studied Dcr KO in mouse tissues or knockdown in human cells. To study the effect of cMyc, we selected experiments from CCLE that resulted in gain or loss of cMyc expression. We mapped probes to annotated lncRNAs and mRNAs and calculated the differential expression of lncRNAs and mRNAs between perturbations versus controls (Fig. 5a). We defined the outcomes as positive or negative decoupling, depending on the relative magnitude of expression changes between all lncRNAs and mRNAs in each experiment. Positive decoupling indicates that lncRNA expression change is significantly higher than mRNA expression change, whereas negative decoupling means relative depletion of lncRNAs ($P < 0.05$, FDR < 0.05 ; Online Methods).

Indeed, Dcr loss preferentially depleted lncRNAs expressed in multiple mouse tissues *in vivo* and in human cells (12 significant experiments out of 19, $P < 0.05$; Fig. 5b). Dcr KO led to negative decoupling of lncRNAs in adrenal-gland cells, granulocytes, liver cells, fibroblasts, inner-ear cells, T regulatory cells and HeLa cells (Fig. 5b). Dcr KO in neurons and oocytes showed inconsistent results, and we detected positive decoupling in a single preadipocyte experiment, which may reflect experimental noise or unique regulation in this cell type. In addition, over 400 microarray experiments surveyed in CCLE had perturbed cMyc levels. These experiments covered over 35 different cell types, and not all cMyc perturbation is associated with changes in Dcr expression. We found that significant positive decoupling (lncRNA induction over mRNAs) is associated with cMyc overexpression (185/295 experiments, 66%), and significant negative decoupling is linked to cMyc depletion (100/134 experiments, 75%), consistently with our observation in mESCs (Fig. 5b). This result suggests that regulation of lncRNAs by Dcr and cMyc is conserved between mice and humans and is widespread in diverse types of tissues *in vivo* and in cancer cells.

DISCUSSION

Here we provided the first demonstration, to our knowledge, that DGCR8-Dcr-miRNA circuitry can influence the expression of hundreds of lncRNAs in mESCs through transcriptional regulation (Fig. 5c). Initially suggested by a bioinformatics screen, and later validated by gain-of-function and loss-of-function studies, the link between miRNA circuitry and lncRNA transcription acts in part through a transcriptional factor, cMyc (Fig. 5c). In addition, cMyc seems to have the strongest effect on lncRNAs that are divergently transcribed from protein-coding genes. The quantitative difference between transcription for lncRNA and mRNA occurs in many different cell and tissue types in mice and humans, thus suggesting that the phenomenon is both widespread and important.

Tissue-specific and cell type-specific expression of lncRNAs have been highlighted by many studies^{21,22}. Our results introduce the idea that hundreds of lncRNAs can also be regulated by the same factors. Dcr emerges as a central biogenesis factor for both small and long regulatory RNAs in ES cells—via post-transcriptional RNA processing for small RNAs

and as an indirect regulator of transcriptional activation of lncRNAs. Dcr requirement has been used to implicate small regulatory RNAs in diverse biological processes, and our results expand this interpretation to potentially implicate lncRNAs.

cMyc, a master regulator of proliferation and pluripotency of mESCs²³, affects transcriptional-pause release of thousands of protein-coding genes²⁴. Our study shows that cMyc also affects expression of hundreds of lncRNA genes. In addition, lncRNA genes, especially the ones that are divergently transcribed from neighboring protein-coding genes, seem more sensitive to cMyc regulation than their protein-coding gene counterparts are. Although the exact mechanism that renders divergent lncRNA genes more sensitive to cMyc regulation is still unclear, unique features and functions of lncRNA genes may offer some explanations. For example, divergent lncRNA genes and protein-coding genes may compete for cMyc occupancy when there is limited cMyc expression. Protein-coding genes, many of which are critical in maintaining normal cellular programs, may be able to outcompete lncRNA genes for cMyc binding at their promoters. Second, divergent lncRNA genes tend to have fewer U1-binding sites and more polyadenylation sites than their protein-coding gene counterparts²⁵. cMyc may be able to interact with U1 and other elongation factors that allow preferential elongation of protein-coding genes over lncRNA genes.

Although we found that cMyc and divergent genomic organization can explain a majority (over 60%) of Dcr-dependent lncRNAs in mESCs, additional mechanisms that connect Dcr to transcription may have roles (Fig. 5c). This is partly supported by results of our Dcr-KO microarray screen, in which lncRNA downregulation was observed in many Dcr-deficient systems that do not involve cMyc (Fig. 5b). In these systems, miRNAs other than miR-295 may be involved in regulating expression of other transcription factors or elongation factors.

Our microarray screen also shows that cMyc activity is perturbed in over 400 microarray experiments with varying degrees of Dcr expression. This suggests that cMyc itself may be able to regulate lncRNA expression independently of Dcr. cMyc has potent oncogenic activity in a broad spectrum of human cancers. Its regulation on lncRNA expression broadens the potential regulatory roles of lncRNAs in many types of cancer.

ONLINE METHODS

ES-cell culture

Feeder-free WT, three lines of Dcr-KO mESCs, cMyc^{f/f} and cMyc^{-/-} were generated and maintained on gelatin as described previously⁸. DGCR8 WT and KO mESCs were obtained from the Blelloch laboratory and were cultured on gelatin in the same manner.

Oligonucleotides and miRNAs used in all the experiments

All the oligos in Supplementary Table 1 are from Elim Biopharm, except miRIDIAN miR mimics (Thermo Scientific).

Paired-end RNA sequencing

TRIzol (Qiagen) was used to extract RNA from cells (WT and Dcr-KO mESCs, cMyc^{f/f} and cMyc^{-/-} mESCs). Polyadenylated RNA was isolated from 1–3 µg total RNA with the

MicroPoly(A) Purist Kit (Ambion). Samples were processed according to the manufacturer's protocol. Subsequently, libraries were prepared according to the dUTP protocol²⁶. Over 800 million sequencing reads (100 nt) were generated on Hi-Seq 2000 Illumina platforms.

RT-qPCR of lncRNAs and miRNAs

TRIzol (Qiagen) was used to extract RNA from cells. Turbo-DNA Free Kit (Ambion) was used to treat RNA with DNase. Stratagene Brilliant II SYBR Green qRT-PCR (Agilent) master mix was used for lncRNAs. For miRNAs, a Superscript III kit (Invitrogen) was used to reverse transcribe 1 µg RNA, and LightCycler 480 SYBR Green I Master (Roche) was used for qPCR. β-actin and GAPDH were usually used for normalization.

NanoString analysis

TRIzol (Qiagen) was used to extract RNA from cells. 100 ng of RNA was used for each NanoString assay, which was performed according to the manufacturer's protocols. The lincRNA probe set was selected from the custom probe set used in ref. 10, and housekeeping genes (Actb, B2m and Gapdh) were added for data normalization.

Metabolic labeling

Cells were seeded in 15-cm plates for 24 h. 4-thiouridine (4sU, Sigma) was added to the cell-culture medium to a 500-µM final concentration 15 min before RNA extraction. We used this time point, assuming that there was negligible degradation of RNA. Biotinylation and purification of 4sU-labeled RNA was based on the protocol used by Rabani *et al.*¹⁷. A certain amount (usually 10%) of RNA was set aside as total RNAs before biotinylated RNA was captured with Dynabeads.

ChIP-seq

ChIP-seq was carried out according to the Farnham protocol²⁷. Approximately 50×10^6 – 100×10^6 cells were used for each ChIP-seq experiment. Chromatin was isolated, sheared and incubated with 5 µg of antibody (anti-histone H3K4me3, Abcam, ab8580, lot GR87168-1 and anti-histone H3K36me3, Abcam, ab9050, lot GR66372-1, with validation provided on the manufacturer's website) overnight at 4 °C. Staph A cells were blocked overnight at 4 °C with 10 mg/mL BSA and were incubated with chromatin for 15 min at room temperature. Sequencing libraries were prepared with Illumina's library preparation protocols, and barcodes were added to pool the libraries for Hi-Seq 2000.

Transfection and FACS sorting

24 h before transfection 5×10^5 Dcr-KO mESCs were plated per well of gelatinized six-well plates. Cells were transfected with 10 µl Lipofectamine 2000 (Invitrogen), 1 µg pcDNA3.1-GFP plasmid, 3 µg of pFRT/TO/FLAG/HA-DEST Dicer (Addgene) or pDcr-dead mutant (with catalytic residues E110a and E110b modified to alanine by QuikChange Lightning Site-Directed Mutagenesis Kit, Stratagene) in 1,500 µl of Opti-MEM (Invitrogen). In the case of miR-295 overexpression, 20 nM miR-295 mimic or control mimic was transfected along with 4 µg of pcDNA3.1-GFP plasmid. In cMyc rescue experiments, Dcr-KO mESCs

were transfected with 10 μ l Lipofectamine 2000, 1 μ g pcDNA3.1-GFP plasmid, 3 μ g of pCX-cMyc (Addgene) or pCX-backbone in 1,500 μ l of Opti-MEM. 4 h after transfection, transfection mix was removed from cells and replaced with ES-cell medium. 48 h after transfection, cells were collected for RT-qPCR analysis. For NanoString analysis, GFP-positive cells were isolated by FACS.

RNA-seq analysis

Reads were aligned to the mouse genome (mm9) with Tophat¹², and uniquely mapped reads were assembled into transcripts with Cufflinks¹³. Transcripts with <50 reads were removed from the analysis. Isoforms were collapsed into one transcript, and collapsed transcripts were annotated by comparison to databases of known protein-coding genes, lncRNA genes, pseudo-genes, rRNA genes and repeats²⁸. lncRNA genes were annotated by comparison to existing annotations outlined in three studies^{11,14,15}. Transcripts that did not match anything known were termed unannotated. Because our sequencing and analysis protocols selected for polyadenylated and nonrepetitive reads, we discarded reads that mapped to rRNAs and repeats. DESeq²⁹ was used to normalize reads between samples, and log fold change of each transcript between WT and Dcr-KO cells was calculated. FDR 0.05 was used to decide whether a transcript was significantly up- or downregulated. RNA-seq data from Guttman *et al.*¹¹ was reanalyzed with the same methods as described above.

miRNA-target analysis

miR-295 mRNA targets were predicted by identification of mRNAs that have at least one T1A or M2-8 7-mer in their 3' UTRs⁹. Because lncRNAs do not have 3' UTRs, we used the entire sequence to search for the 7-mer. Control gene sets were generated to match for sequence length (3' UTR for mRNAs, and entire sequence for lncRNAs), dinucleotide composition, and expression level in WT mESCs.

ChIP-seq analysis

ChIP-seq reads were mapped to the mouse genome (mm9) with Bowtie³⁰, allowing no mismatch. We used the MACS³¹ algorithm for peak detection of cMyc ChIP-seq data³², with FDR 0.05 as the cutoff. Enrichment over background was calculated from the input sequenced reads. A transcript was classified as a cMyc target if it had a cMyc binding peak within -5 kb and +2 kb of its transcription start site.

Identification of divergent gene pairs

Divergent gene pairs were identified according to the methods listed in Sigova *et al.*³³. A gene pair was classified as divergent if the two transcription start sites were within 2 kb from each other and the directions of transcription did not overlap. A gene pair was classified as 'sense overlap' if the directions of transcription were the same. A gene pair was classified as 'body overlap' if the directions of transcription overlapped with each other. The remaining genes were classified as intergenic.

Bioinformatics screen

Transcription factor-binding sites were counted in the test and control sets, and an enrichment score for each transcription factor was calculated as the ratio of the number of binding sites in the test set over those in the control set. Detailed methodology is described in Webster *et al.*³⁴. To identify transcription factors enriched in the downregulated lncRNA set, we compared downregulated lncRNAs to either upregulated lncRNAs, or lncRNAs with no significant expression change. Only the first comparison yielded transcription factors with high enrichment scores. We also performed similar analysis on upregulated lncRNAs relative to lncRNAs with no significant expression change in order to identify any transcription factors that may regulate upregulated lncRNAs. However, we were not able to identify any with a high enrichment score.

Microarray analysis (GEO and CCLE data)

All Affymetrix mouse and human gene-expression microarrays were obtained from Gene Expression Omnibus (GEO) and the Cancer Cell Line Encyclopedia (CCLE)²⁰. For GEO, series matrix files were downloaded and quantile-normalized with R's limma package, and within each experiment, biological replicates were identified and grouped automatically by provided annotations. For CCLE, raw CEL files were downloaded and quantile-normalized with R's affy package. To identify Dicer-KO studies, a text search was done within GEO annotations for studies involving Dicer, and to identify cMyc perturbation studies, a search was done within CCLE for cell lines in which cMyc was misregulated at least two-fold relative to the average cMyc expression value across all cell lines. Probes for all array platforms were then mapped to either lncRNAs or protein-coding genes according to databases outlined in RNA-seq analysis (Online Methods), and for each study, each lncRNA and mRNA was assigned a log fold change relative to its associated control. The control for each GEO study was determined on the basis of text annotations, and the control for all CCLE data was defined as the average expression across all cell lines. Within each study, differential regulation of lncRNAs and mRNAs was determined from the distribution of lncRNA log fold changes versus those of mRNAs. Direction of differential regulation was determined from the mean of the distributions, and statistical significance was calculated with the Kolmogorov-Smirnov test.

Statistical analyses

All test statistics were calculated with R (<http://www.r-project.org/>). These include the Wilcoxon rank-sum test, K-S test, *t* test, hyper-geometric distribution and DESeq.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank members of the Chang laboratory, and P. Sharp (Massachusetts Institute of Technology) and A. Giraldez (Yale) for discussion, and for sharing unpublished data. We thank R. Blueloch (University of California, San Francisco) and A. Bradley (Wellcome Trust Sanger Institute) for sharing DGCR8 WT and KO mESCs, and cMyc KO mESCs respectively. G.X.Y.Z. was supported by the Leukemia and Lymphoma Society (grant 5549-13 to G.X.Y.Z.) and a Dean's Fellowship from Stanford University. The study was supported by the US National

Institutes of Health (grant R01-CA118750 to H.Y.C.) and California Institute for Regenerative Medicine (grant RB4-05763 to H.Y.C.). H.Y.C. is supported as an Early Career Scientist of the Howard Hughes Medical Institute.

References

1. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. *Annu Rev Biochem.* 2012; 81:145–166. [PubMed: 22663078]
2. Batista PJ, Chang HY. Long noncoding RNAs: cellular address codes in development and disease. *Cell.* 2013; 152:1298–1307. [PubMed: 23498938]
3. Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell.* 2009; 136:215–233. [PubMed: 19167326]
4. Jeggari A, Marks DS, Larsson E. miRcode: a map of putative microRNA target sites in the long non-coding transcriptome. *Bioinformatics.* 2012; 28:2062–2063. [PubMed: 22718787]
5. Cesana M, et al. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell.* 2011; 147:358–369. [PubMed: 22000014]
6. Braconi C, et al. microRNA-29 can regulate expression of the long non-coding RNA gene MEG3 in hepatocellular cancer. *Oncogene.* 2011; 30:4750–4756. [PubMed: 21625215]
7. Tay Y, Rinn J, Pandolfi PP. The multilayered complexity of ceRNA crosstalk and competition. *Nature.* 2014; 505:344–352. [PubMed: 24429633]
8. Calabrese JM, Seila AC, Yeo GW, Sharp PA. RNA sequence analysis defines Dicer's role in mouse embryonic stem cells. *Proc Natl Acad Sci USA.* 2007; 104:18097–18102. [PubMed: 17989215]
9. Zheng GX, et al. A latent pro-survival function for the mir-290–295 cluster in mouse embryonic stem cells. *PLoS Genet.* 2011; 7:e1002054. [PubMed: 21573140]
10. Guttman M, et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature.* 2011; 477:295–300. [PubMed: 21874018]
11. Guttman M, et al. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat Biotechnol.* 2010; 28:503–510. [PubMed: 20436462]
12. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics.* 2009; 25:1105–1111. [PubMed: 19289445]
13. Trapnell C, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol.* 2010; 28:511–515. [PubMed: 20436464]
14. Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. *Cell.* 2009; 136:629–641. [PubMed: 19239885]
15. Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. *Cell.* 2013; 154:26–46. [PubMed: 23827673]
16. Leung AK, et al. Genome-wide identification of Ago2 binding sites from mouse embryonic stem cells with and without mature microRNAs. *Nat Struct Mol Biol.* 2011; 18:237–244. [PubMed: 21258322]
17. Rabani M, et al. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat Biotechnol.* 2011; 29:436–442. [PubMed: 21516085]
18. Sharova LV, et al. Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. *DNA Res.* 2009; 16:45–58. [PubMed: 19001483]
19. Melton C, Judson RL, Bluelloch R. Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. *Nature.* 2010; 463:621–626. [PubMed: 20054295]
20. Barretina J, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature.* 2012; 483:603–607. [PubMed: 22460905]
21. Cabili MN, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 2011; 25:1915–1927. [PubMed: 21890647]

22. Harrow J, et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* 2012; 22:1760–1774. [PubMed: 22955987]
23. Davis AC, Wims M, Spotts GD, Hann SR, Bradley A. A null c-myc mutation causes lethality before 10.5 days of gestation in homozygotes and reduced fertility in heterozygous female mice. *Genes Dev.* 1993; 7:671–682. [PubMed: 8458579]
24. Rahl PB, et al. c-Myc regulates transcriptional pause release. *Cell.* 2010; 141:432–445. [PubMed: 20434984]
25. Almada AE, Wu X, Kriz AJ, Burge CB, Sharp PA. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature.* 2013; 499:360–363. [PubMed: 23792564]
26. Parkhomchuk D, et al. Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res.* 2009; 37:e123. [PubMed: 19620212]
27. O’Geen H, Echipare L, Farnham PJ. Using ChIP-seq technology to generate high-resolution profiles of histone modifications. *Methods Mol Biol.* 2011; 791:265–286. [PubMed: 21913086]
28. Karolchik D, et al. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* 2004; 32:D493–D496. [PubMed: 14681465]
29. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol.* 2010; 11:R106. [PubMed: 20979621]
30. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009; 10:R25. [PubMed: 19261174]
31. Zhang Y, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 2008; 9:R137. [PubMed: 18798982]
32. Chen X, Vega VB, Ng HH. Transcriptional regulatory networks in embryonic stem cells. *Cold Spring Harb Symp Quant Biol.* 2008; 73:203–209. [PubMed: 19022762]
33. Sigova AA, et al. Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proc Natl Acad Sci USA.* 2013; 110:2876–2881. [PubMed: 23382218]
34. Webster DE, et al. Enhancer-targeted genome editing selectively blocks innate resistance to oncoprotein inhibition. *Genome Res.* 2014; 24:751–760. [PubMed: 24443471]

Figure 1.

Global lncRNA downregulation in Dcr-KO mESCs. **(a)** Breakdown of classes of transcripts and their abundance in WT and Dcr-KO mESCs. RPKM, reads per kilobase per million. **(b)** Cumulative distribution functions (cdfs) of $\log_2$ fold change (LFC) of lncRNAs (red) and coding transcripts (blue) between Dcr WT and KO mESCs ($n = 3$ KO lines; $P = 3.95 \times 10^{-13}$ by two-sided Wilcoxon rank-sum test). **(c)** Sequencing tracks of long intergenic noncoding (linc) RNA linc1296 (nomenclature from ref. 10). The y axis represents number of reads per million. The gene structure of linc1296 is shown at the bottom in black, with the exons as black boxes. **(d)** Bar plots of $\log_2$ fold change of lncRNAs and protein-coding transcripts with significantly changed expression between Dcr WT and KO mESCs. **(e)** Expression of nine lncRNAs and one known miR-295-target, *Casp2*, validated by RT-qPCR ($n = 3$ experiments, mean $\pm$ s.e.m., * $P \leq 0.05$; ** $P \leq 0.01$ by two-sided t test).

Figure 2. Canonical microRNA pathway and miR-295 are required for lncRNA expression. **(a)** Schematic of the canonical miRNA pathway. **(b)** NanoString measurement of expression of 102 lncRNAs in three lines of Dcr-KO mESCs (normalized to WT), Dcr rescue in Dcr KO (normalized to the catalytically dead mutants), Dgcr8 KO (normalized to Dgcr8 WT) and miR-295 rescue in Dcr KO (normalized to a control miRNA mimic). The box on the left shows the ability of lncRNAs to maintain pluripotency (column 1) and to prevent differentiation (column 2) as well as their binding to chromatin (column 3), as cross-referenced with data from ref. 10. The blue and yellow colors in the heat map indicate down- and upregulated lncRNAs respectively ($n = 2$ cultures each).

Figure 3.

Dicer promotes lncRNA expression at the level of lncRNA transcription. **(a)** Schematic of 4sU experiment in WT and Dcr-KO mESCs. **(b,c)** Synthesis rate and half-life of lncRNAs, measured in Dcr WT and KO mESCs and displayed as a ratio of KO and WT ($n = 3$ trials for lncRNAs and $n = 4$ for Gapdh; * $P < 0.05$; ** $P < 0.01$ by two-sided t test). Error bars, s.e.m. **(d,e)** Sequencing signals of H3K4me3 and H3K36me3 in WT (purple) and Dcr KO (green) mESCs at linc543 (nomenclature from ref. 15). The y axis represents number of reads per million. The gene structure of linc543 is shown at the bottom in black, with exons shown as black boxes. **(f,g)** Average signal intensity of H3K4me3 and H3K36me3 across downregulated lncRNA genes in WT (purple) and Dcr KO (green) mESCs. y axis represents average log₂ value of normalized reads. P values indicate the significance of the area difference between WT and KO signals within the dotted line (two-sided Wilcoxon rank-sum test, $n = 100$ data points). The gray box at bottom represents the beginning and the end of an average lncRNA gene. The signals were also plotted for 50% up- and downstream of each lncRNA gene.

Figure 4.

cMyc-binding sites of lncRNAs affect their sensitivity to Dcr loss. **(a)** cMyc-binding-site enrichment in Dcr-dependent lncRNA genes. Top inset, schematic of bioinformatics strategy. TF, transcription factor; DB, database. Bottom inset, cMyc downregulation in Dcr-KO mESCs. RPKM, reads per kilobase per million. **(b)** cMyc ChIP-seq binding data, plotted alongside the expression change of lncRNAs in Dcr KO. **(c)** Cumulative distribution functions (cdfs) of log₂ fold change (LFC) for lncRNA genes (dark red) and protein-coding genes (blue) between cMyc^{flox/flox} (denoted f/f) and cMyc^{-/-} mESCs. lncRNA genes with cMyc-binding sites are shown in red, and lncRNA genes without cMyc-binding sites are shown in gray. lncRNAs (especially those encoded by lncRNA genes with cMyc-binding sites) are significantly downregulated in cMyc^{-/-} mESCs relative to protein-coding genes ($P < 2.2 \times 10^{-16}$ by two-sided Wilcoxon rank-sum test, $n = 2,229$ known lncRNA transcripts). **(d)** NanoString measurement of expression of 102 lncRNAs in three lines of Dcr-KO mESCs (KO1–KO3, normalized to WT, as in Fig. 2b) and cMyc rescue in Dcr KO (normalized to the control plasmid). The blue and yellow colors in the heat map indicate down- and upregulated lncRNAs respectively ($n = 2$ cultures). **(e)** Changes in intensity of H3K4me3 (left) and H3K36me3 (right) at lncRNA genes with (red) or without (gray) cMyc-binding sites between WT and Dcr-KO mESCs.

Figure 5.

Loss of Dcr and cMyc leads to a more severe change in expression of lncRNAs than of mRNAs in diverse cellular contexts. **(a)** Flowchart depicting the search for experiments that exhibit differential expression change of lncRNAs and mRNAs. GEO was mined for Dicer KO experiments, and CCLE was mined for cMyc perturbation experiments. Examples of negative and positive decoupling are shown in graphs. **(b)** lncRNA and mRNA regulation compared in 19 mouse and human tissues after Dicer knockout (top) and 415 human cell lines after cMyc perturbation (bottom). Blue represents downregulation of lncRNAs relative to protein-coding transcripts, and red represents upregulation of lncRNAs relative to protein-coding transcripts. GOF, gain of function; LOF, loss of function. The intensity of the color reflects the significance, determined by the Kolmogorov-Smirnov (KS) test ($*P < 0.05$ at top; numbers at bottom indicate significant experiments with $P < 0.05$). **(c)** A general model of lncRNA regulation by Dicer and cMyc. Solid arrows, known interactions; question mark, unknown mechanism; dashed arrows, new interactions shown in this work.