iScience

Article

The Contribution of lincRNAs at the Interface between Cell Cycle Regulation and Cell State Maintenance

Adriano Biasini,
Adam Alexander
Thil Smith, Baroj
Abdulkarim, Maria
Ferreira da Silva,
Jennifer Yihong
Tan, Ana Claudia
Marques

anaclaudia.marques@unil.ch

HIGHLIGHTS

Genes differentially
expressed between mESC
cell cycle stages are
enriched in lincRNAs

CC-lincRNAs are co-
expressed with cell cycle
and pluripotency genes

CC-lincRNAs are often
mESC specific and their
promoters enriched in
pluripotency TFs

Upregulation of two CC-
lincRNAs results in
deregulated mESC cell
cycle progression

Article

The Contribution of lincRNAs at the Interface between Cell Cycle Regulation and Cell State Maintenance

Adriano Biasini,[1,3] Adam Alexander Thil Smith,[1,2,3] Baroj Abdulkarim,[1] Maria Ferreira da Silva,[1] Jennifer Yihong Tan,[1] and Ana Claudia Marques[1,4,*]

SUMMARY

Cell cycle progression is controlled by the interplay of established cell cycle regulators. Changes in these regulators' activity underpin differences in cell cycle kinetics between cell types. We investigated whether long intergenic noncoding RNAs (lincRNAs) contribute to embryonic stem cell cycle adaptations. Using single-cell RNA sequencing data for mouse embryonic stem cells (mESCs) staged as G1, S, or G2/M we found differentially expressed lincRNAs are enriched among cell cycle-regulated genes. These lincRNAs (CC-lincRNAs) are co-expressed with genes involved in cell cycle regulation. We tested the impact of two CC-lincRNA candidates and show using CRISPR activation that increasing their expression is associated with deregulated cell cycle progression. Interestingly, CC-lincRNAs are often differentially expressed between G1 and S, their promoters are enriched in pluripotency transcription factor (TF) binding sites, and their transcripts are frequently co-regulated with genes involved in the maintenance of pluripotency, suggesting a contribution of CC-lincRNAs to mESC cell cycle adaptations.

INTRODUCTION

The cell cycle is a dynamic sequence of events leading from one parent cell to two daughter cells. This requires replication of chromosomes during the Synthesis phase (S phase) and their segregation into two daughter cells during Mitosis (M phase). M and S phases are separated by two Gap phases, G1 and G2, that act as checkpoints to prevent cell division without genome replication and aberrant polyploidy (Murray and Hunt, 1993). Progression through cell cycle is regulated by cell cycle stage-specific activation and repression of numerous proteins including Cyclin Dependent Kinases (CDKs) and proteins from the Cyclin family (Morgan, 1995). In most somatic cells, the oscillatory expression or activity of distinct Cyclin-Cdk complexes allows the activation and repression of cell cycle regulators and promotes cell cycle transitions (Morgan, 1995). One key regulator of this process is the retinoblastoma protein (RB) that controls G1 and prevents entry into S phase. Upon entrance in G1, RB is unphosphorylated (active) and blocks the expression of genes required for G1/S transition. During G1, RB is phosphorylated and becomes inactive, allowing cells to progress to S phase (Weinberg, 1995). In embryonic stem cells (ESCs), RB is hyperphosphorylated, resulting in suppression of the G1-S checkpoint and thereby rapid shuttling between DNA synthesis and mitosis, decreasing the average duration of the ESC cell cycle (reviewed in Zaveri and Dhawan [2018]).

These significant adaptations in the ESC cell cycle are important for the maintenance of the embryonic stem cell state and cell fate decisions, as highlighted by the partial overlap between the gene regulatory networks that control the two processes (Soufi and Dalton, 2016). For example, both Oct4 and Nanog, two core pluripotency factors, control genes involved in cell cycle regulation: in mouse embryonic stem cells (mESCs), Oct4 represses the expression of *p21*, a Cyclin-dependent kinase inhibitor that is expressed in somatic cells but not in embryonic stem cells (Lee et al., 2010); *NANOG*, whose expression is cell cycle regulated, controls S-phase entry by regulating the expression of *Cdc25C* and *CDK6* in human ESCs (hESCs) (Zhang et al., 2009). Although the association between cell cycle dynamics and cell state is well established, the molecular mechanisms underlying this connection remain uncharacterized (Soufi and Dalton, 2016).

In addition to proteins, noncoding RNAs, including long intergenic noncoding RNAs (lincRNAs), have also been shown to contribute to cell cycle progression (Kitagawa et al., 2013). An example of this is MALAT1, a

[1]Department of Computational Biology, University of Lausanne, Lausanne, Switzerland

[2]Present address: Baker Heart & Diabetes Institute, 75 Commercial Road, Melbourne, VIC 3004, Australia

[3]These authors contributed equally

[4]Lead Contact

*Correspondence: anaclaudia.marques@unil.ch

https://doi.org/10.1016/j.isci.2020.101291

1

lincRNA that is frequently upregulated in multiple human cancers (Sun and Ma, 2019). In human fibroblasts, depletion of MALAT1 leads to decreased expression of several genes involved in cell cycle progression and results in G1 arrest. MALAT1 is also involved in splicing of *B-Myb*, a gene involved in the transcriptional regulation of several mitotic proteins (Tripathi et al., 2013). More recently, the cohesion regulator long non-coding RNA (CONCR) has been found to be necessary for cell cycle progression and DNA replication. CONCR expression is activated by the transcription factor MYC and is upregulated in multiple cancer types (Marchese et al., 2016). Silencing of CONCR leads to a significant decrease in DNA synthesis. At the molecular level, CONCR physically interacts with DDEA/H-box helicase 11, which ensures the proper separation of sister chromatids during the cell division process. The absence of CONCR leads to the loss of sister chromatid cohesion and affects metaphase (Marchese et al., 2016). Finally, lincRNA expression is often dysregulated in cancer and the characterization of subsets of cancer-associated lincRNAs highlights their potential roles as cell cycle progression modulators (Schmitt and Chang, 2016; Notzold et al., 2017).

LincRNAs are also part of the network controlling stem cell fate maintenance and differentiation (Yan et al., 2017). Because lincRNA expression is often tissue specific (Derrien et al., 2012; Tuck et al., 2018), in contrast to proteins, we hypothesized they can support cell type-specific activity of ubiquitously expressed genes and act at the intersection of cell cycle and mESC cell state regulation. The ability of tissues-specific non-coding RNAs to modulate the activity of ubiquitously expressed gene was already exemplified by lncSCA7. This lincRNA regulates ATXN7 levels in retinal and cerebellar neurons and contributes to specific degeneration of these cells in SCA7 patients (Tan et al., 2014). Additionally, the relative short half-lives of lincRNAs (Clark et al., 2012; Mukherjee et al., 2017) further underscores their potential as modulators of temporally resolved processes such as cell cycle progression.

Here, we investigate the contributions of lincRNAs to embryonic stem cell cycle adaptations.

RESULTS

LincRNA Expression Is Often Cell Cycle Regulated

We used publicly available single-cell RNA sequencing (scRNAseq) data for 279 mESCs with known cell cycle stage (Buettner et al., 2015) to assess the extent of cell cycle-regulated lincRNA expression in mESCs. We estimated the expression of protein-coding transcripts and lincRNAs in each of these cells. After excluding cells and genes that failed quality control (Figure S1) we identified 10,487 genes, including 781 lincRNAs, whose expression can be robustly detected in 246 cells. As previously shown (Buettner et al., 2015), gene expression patterns in this dataset reflect the cell cycle stages of the individual cells (Figure 1A), supporting its use to identify transcripts whose expression are cell cycle dependent. Using DEseq2, we identified 638 genes (6.1%) whose expression is significantly different between at least two cell cycle stages (Table S1). The proportion of differentially expressed lincRNAs (n = 70, 8.96%) is significantly higher than that found for protein-coding genes (n = 501, 5.51%) (proportions test p value = 7×10^{-10}, Figure 1B, Table S1), indicating that lincRNA expression is more dynamic throughout the mESC cell cycle than is the expression of protein-coding genes (Figure 1D).

The median expression of mRNAs is roughly 14 times higher than that of lincRNAs (Figure S2A) and steady-state abundance can impact the ability to detect differential gene expression, as highlighted by the significantly higher expression (two-tailed Wilcoxon test, p value = 3×10^{-43}) of genes identified as differentially expressed (Figure S2A). To assess whether lincRNAs' relatively low expression impacts our differential gene expression analysis results, we repeated the analysis by re-quantifying protein-coding gene expression using a subset (1/14) of randomly sampled reads from each library. At this sequencing depth, the median mRNA expression is comparable with that of lincRNAs in the full data (Figure S2B). We repeated expression quantification, quality control, normalization, filtering, and differential gene expression analyses using these randomly sampled reads and found that only 60% of the mRNAs found to be differentially expressed using the full data are also differentially expressed when using the down-sampled libraries. This result indicates that the higher proportion of differentially expressed lincRNAs is likely an underestimate owing to the limited power to measure their expression using single cell RNA sequencing data.

To validate our *in silico* differential gene expression predictions we used DNA content to sort mESC cells into G1, S, and G2/M stages and measured the cell cycle stage expression of twelve differentially expressed lincRNAs, including Neat1, and three mESC cell cycle protein-coding genes (*Ccnb1, Ccnd1* and *Ccne1*) by quantitative PCR (Figures 1C and S3). The cell cycle protein-coding gene and lincRNA expression patterns
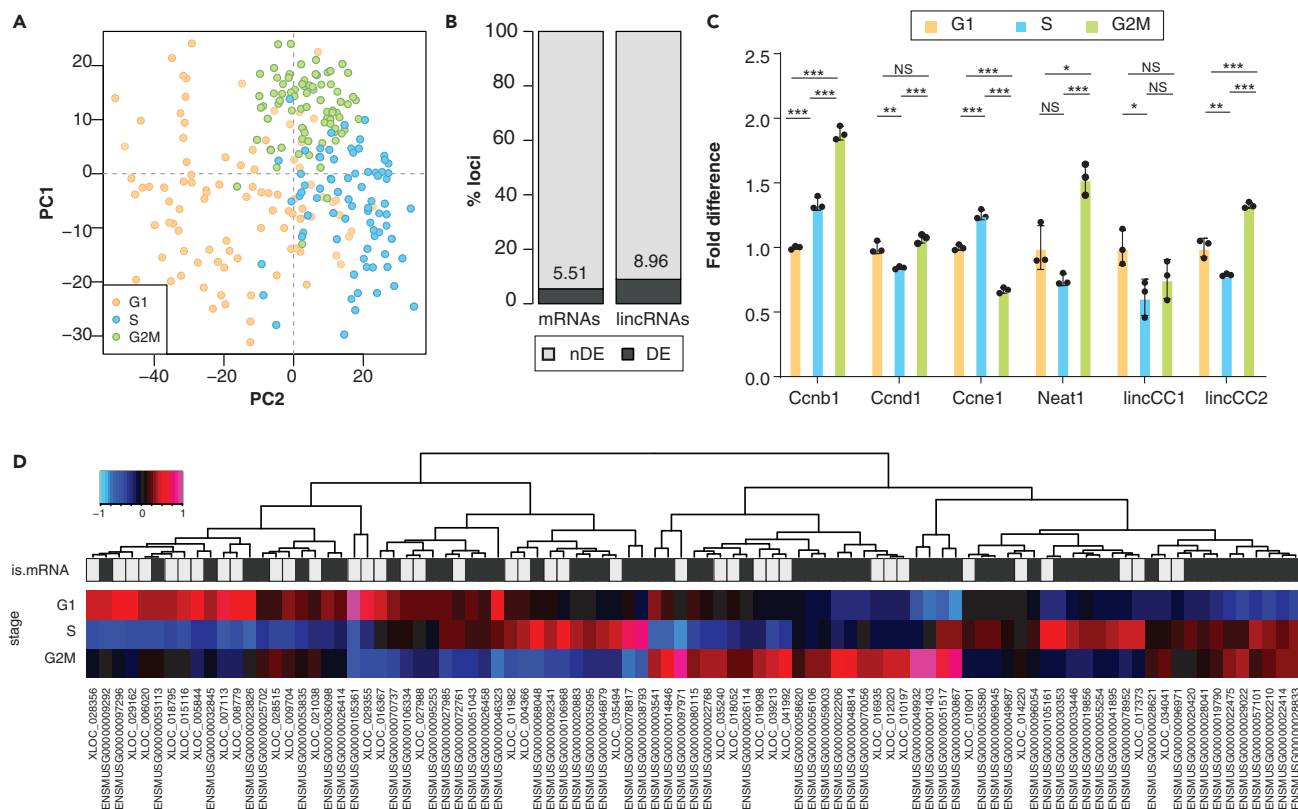
Figure 1. LincRNA Expression Is Often Cell Cycle Regulated

(A) Principal component analysis of gene expression for all loci passing the technical noise filter. The first two principal components (PC1 on the x axis and PC2 on the y axis) together explain 5.15% of the total variability, and separate cells in different cell cycle stages (G1: orange, S: blue; G2/M: green).

(B) Percentage of protein-coding genes (left) and lincRNAs (right) that are differentially expressed (dark gray, numbers indicate the percentage) between at least two cell cycle stages.

(C) Fold difference in normalized expression (relative to G1) of known cell cycle-regulated genes (Ccnb1, Ccnd1, Ccne1, Neat1) and two differentially expressed novel lincRNAs (lincCC1 and lincCC2) in mESCs at different cell cycle stages (G1: orange, S: blue, G2/M: green).

(D) Heatmap representing the median Z score of the shifted log expression level across all cells in each of the three stages (G1, S and G2M, rows) for each differentially expressed lincRNA and mRNA (columns). Rectangles in "is.mRNA" row are colored gray and black for lincRNA and mRNA, respectively. Color scale top right. Significance (two-tailed unpaired t test p value) between cell cycle stage expression is indicated as NS p > 0.05, *p < 0.05, **p < 0.01, ***p < 0.001.

measured by qPCR were generally consistent with what was estimated using scRNA sequencing data. The results of this analysis support that despite the relatively low expression of lincRNAs that complicates the accurate estimation of their expression, our differential expression predictions are generally robust.

Differentially Expressed lincRNAs Contribute to the Regulation of Cell Cycle Progression

Consistent with the role of differentially expressed protein-coding genes in the regulation of cell cycle progression, these genes are significantly enriched (hypergeometric test adjusted p value < 0,05) in annotations with gene ontology terms such as "cell division" and other related cell cycle processes (Figure 2A, Table S2), as previously described (Dominguez et al., 2016; Liu et al., 2017; Pena-Diaz et al., 2013). The lower sensitivity of scRNAseq together with the differences in embryonic and adult cell cycle dynamics and experimental design likely explain the relatively smaller enrichment in "cell division"-related terms among differentially expressed genes relative to previous studies (Dominguez et al., 2016; Pena-Diaz et al., 2013; Liu et al., 2017). Furthermore, 8.8% (44 of 501) of differentially expressed protein-coding loci are orthologous to genes previously shown to have a periodic expression throughout the human cell cycle (Santos et al., 2015), a significant 3.1-fold enrichment relative to all mESC expressed protein-coding loci (214, hypergeometric test p value = 1×10^{-11}), in cell cycle-related terms. These results support that genes differentially expressed between cell cycle stages are indeed enriched in cell cycle regulators.
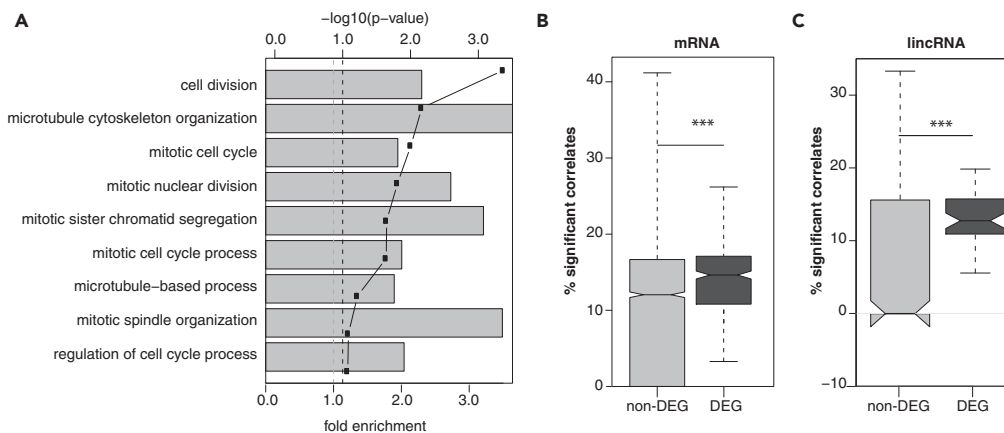
Figure 2. Genes Differentially Expressed between Cell Cycle Stages Have Cell Cycle-Related Functions

(A) Results of the GO term enrichment analysis for cell cycle stage differentially expressed genes. Gray bars indicate the fold enrichment, and points the associated -log10 of the p value. Dashed line indicates the 10% FDR cutoff. Distribution of the percentage of genes cell cycle regulators that are significantly correlated in expression with (B) protein-coding genes and (C) lincRNAs. Differentially or non-differentially expressed genes from each biotope are represented in dark and light gray, respectively. Significance of the distribution comparison (two-tailed Mann-Whitney-Wilcoxon rank-sum test) are indicated as *** p value < 0.001.

The enrichment of cell cycle-related genes among those differentially expressed between cell cycle stages suggests that some of the lincRNAs identified here as differentially expressed might also contribute to mESC cell cycle regulation. Most lincRNAs differentially expressed between cell cycle stages were annotated *de novo* (53 *de novo* lincRNAs) using mESC RNA sequencing data (Tan et al., 2015). We searched the literature for functions of the remaining seventeen differentially expressed lincR-NAs annotated by ENSEMBL (Table S3). Of these, only four have been previously characterized: Neat1, Malat1, the host transcript for snord93, and miRNA-24 pri-miRNA transcript. Malat1 is a cancer-associated lincRNA and has been previously implicated in cell cycle progression. For example, Malat1 is highly expressed during the S and M phases of human fibroblasts where it controls cell cycle-related processes (Tripathi et al., 2013). Neat1, whose knockdown was recently shown to impair S phase transition (Adriaens et al., 2016), has also been frequently associated with cancer and cell cycle progression (Yu et al., 2017). Relatively little is known about snord93, but changes in this RNA's levels impact cell proliferation (Patterson et al., 2017), which would be consistent with a role in cell cycle progression. Finally, miRNA-24 post-transcriptionally regulates MYC and E2F2, two known cell cycle genes, and changes in its levels impair G1 transition (Lal et al., 2009). In conclusion, all of the four annotated and characterized lincRNAs we identified as being differentially expressed either have established roles or have been associated with cell cycle progression.

Given the paucity of functional annotations for lincRNAs, and to assess broadly the contributions of differentially expressed lincRNAs to the cell cycle, we reasoned that genes that functionally participate in cell cycle regulation should genetically interact with known cell cycle regulators. To first validate this idea, we considered cell cycle phase differentially expressed protein-coding genes not annotated as cell cycle regulators and found their expression is more often significantly (p value = 1×10^{-8}, Wilcoxon rank-sum test) correlated with the expression of annotated cell cycle genes than non-differentially expressed protein-coding loci (Figure 2B). Similarly, we found that the expression of differentially expressed lincRNAs is also significantly more often correlated with the levels of cell cycle genes than non-differentially expressed lincRNAs (Figure 2C, p value = 2×10^{-4} Wilcoxon rank-sum test), supporting their contributions to mESC cell cycle progression. Hereafter, we refer to differentially expressed lincRNAs and protein-coding genes as cell cycle-associated lincRNAs and protein-coding genes, or CC-lincRNAs and CC-PCGs, respectively.

Cell Cycle lincRNAs Are Associated with Stem Cell Cycle Adaptations

The cell cycle is a ubiquitous process, yet most lincRNAs are expressed in cell type-specific manner (Ulitsky and Bartel, 2013). Do the CC-lincRNAs identified here contribute to cell cycle regulation ubiquitously, or

Figure 3. CC-lincRNAs Are Associated with Stem Cell Cycle Adaptations

(A–C) (A) Distribution of tissue specificity (measured as Tau) for CC-lincRNAs, CC-PCGs, and annotated cell cycle genes. Distribution of the fold difference in expression of (B) lincRNAs and (C) protein-coding genes during a 216-h neuronal commitment of mESCs. Fold difference in expression is relative to the median fold difference at time = 8 h.

(D) Fold enrichment in binding (bars, x axis) of pluripotency transcription factors (y axis) for CC-PCGs (dark gray) and CC-mRNAs (light gray).

(E) Cell cycle stage-average (G1: orange, S: blue, G2/M: green) of the expression Z scores of CC-PCGs (left panel) and CC-lincRNAs (right panel).

(F) Proportional Euler diagram of the number of lincRNAs (left) and protein-coding genes (right) differentially expressed between G1 and S (orange), G1 and G2M (blue), or S and G2M (green).

Figure 3. Continued

(G) Distribution of the correlations between pairs of pluripotency-associated genes (black) or known pluripotency genes, and differentially expressed (CC, gray) or non-differentially expressed (nonCC, white) protein-coding genes and lincRNAs. Significance (two-tailed Mann-Whitney-Wilcoxon rank-sum test) for the comparison between pluripotency-associated genes and the different gene classes is indicated on top of the boxplot as NS p > 0.05, **p < 0.01, ***p < 0.00.

are their functions restricted to mESCs? To gain initial insights into this question we used publicly available transcriptome-wide data (Yue et al., 2014) to estimate the tissue specificity of lincRNAs and protein-coding genes across 29 adult and developing mouse tissue and cell lines. We estimated Tau, a measure of tissue specificity, for CC-lincRNAs, CC-PCGs, and known cell cycle regulators. Tau varies between 0% for "ubiquitously expressed" to 100% for "tissue-specifically expressed" genes (Kryuchkova-Mostacci and Robinson-Rechavi, 2017). As expected, CC-PCGs, as well as established cell cycle genes, are broadly expressed (median tau = 24.8%). In contrast, CC-lincRNAs are significantly more tissue specific (median tau = 42%, Wilcox rank-sum test p value = 7×10^{-13}, Figure 3A) than their protein-coding counterparts and are as tissue specific as other mESC-expressed lincRNAs (data not shown).

To further test if CC-lincRNA expression is often restricted to mESCs, we investigated changes in their transcript abundance upon neuronal differentiation of mESCs. *In vitro* neural differentiation is a well-defined and highly efficient process (>80% of differentiated cells). We took advantage of publicly available transcriptome-wide expression for a time course of neuronal commitment of mESC (Sun et al., 2017) to investigate the changes in noncoding and coding gene expression. Consistent with their tissue-specific expression (Figure 3A), CC-lincRNA expression decreases rapidly upon differentiation (Figure 3B), supporting their contribution to cell cycle regulation being restricted to mESCs. In contrast, CC-PCGs are expressed at similar levels throughout neuronal commitment (Figure 3C).

The transcriptional network controlling mESC cell state and function is regulated by a set of mESC core transcription factors (Chambers and Tomlinson, 2009). The short G1 phase that characterizes the embryonic stem cell cycle, which is critical to ensure maintenance of pluripotency, is in part orchestrated by stem cell-specific factors (Boward et al., 2016). We took advantage of publicly available chromatin immunoprecipitation sequencing (ChIP-seq) data for pluripotency transcription factors in mESCs (Chen et al., 2008) to assess the extent of these factors' binding to cell cycle-regulated lincRNA promoters. We found that CC-lincRNA promoters are enriched in mESC core transcription factors (TFs) supporting their role in the network underlying mESC cell state. Specifically, promoters of CC-lincRNAs are significantly enriched, relative to all expressed lincRNAs, in the binding of pluripotency transcription factors, including Nanog, Oct4, and Sox2 (false discovery rate [FDR]<0.05, permutation test) (Figure 3D). We found no significant enrichment by most pluripotency TFs at the promotes of CC-PCGs (Figure 3D).

To assess what aspect of mESC cell cycle progression might be more often modulated by CC-lincRNAs, we investigated their relative expression across different cell cycle phases. Consistent with previous observations that mRNA expression peaks at G2/M phase in human cells (Cho et al., 2001; Whitfield et al., 2002) we found that CC-protein-coding genes were highly expressed in S and G2/M (Figure 3E) and were often differentially expressed between G1 and G2/M (Figure 3F). In contrast, the levels of CC-lincRNAs were higher in G1 relative to all other cell cycle stages (Figure 3E) and most were differentially expressed between G1 and S phases (Figure 3F).

Given the significantly elevated expression of CC-lincRNAs in G1, their tissue-specific expression and evidence that their transcription is regulated by stem cell-specific transcription factors, we hypothesized that CC-lincRNAs may contribute to the interplay between cell cycle and maintenance of pluripotency in mESCs. If CC-lincRNAs participate in the network controlling maintenance of mESC cell state, they should be co-expressed with genes involved in maintenance of pluripotency. To test this hypothesis, we investigated the correlation, during mESC neuronal commitment, between CC-genes and genes implicated in maintenance of pluripotency (Xu et al., 2013). First, we found that, consistent with the interplay between cell cycle control and maintenance of mESC cell state, the median pairwise correlation between CC-PCGs and pluripotency genes (Spearman's r = 0.06) is similar to what is found for pairs of genes implicated in pluripotency (r = 0.10, two-tailed Wilcoxon test p value = 0.3, Figure 3G). Consistent with our hypothesis that CC-lincRNAs contribute to maintenance of mESC cell state, we found that the extent of their association with pluripotency genes (r = 0.05) is also similar to what we estimate for pairs of genes implicated in pluripotency (two-tailed Wilcoxon test p value = 0.7, Figure 3G). For comparison,
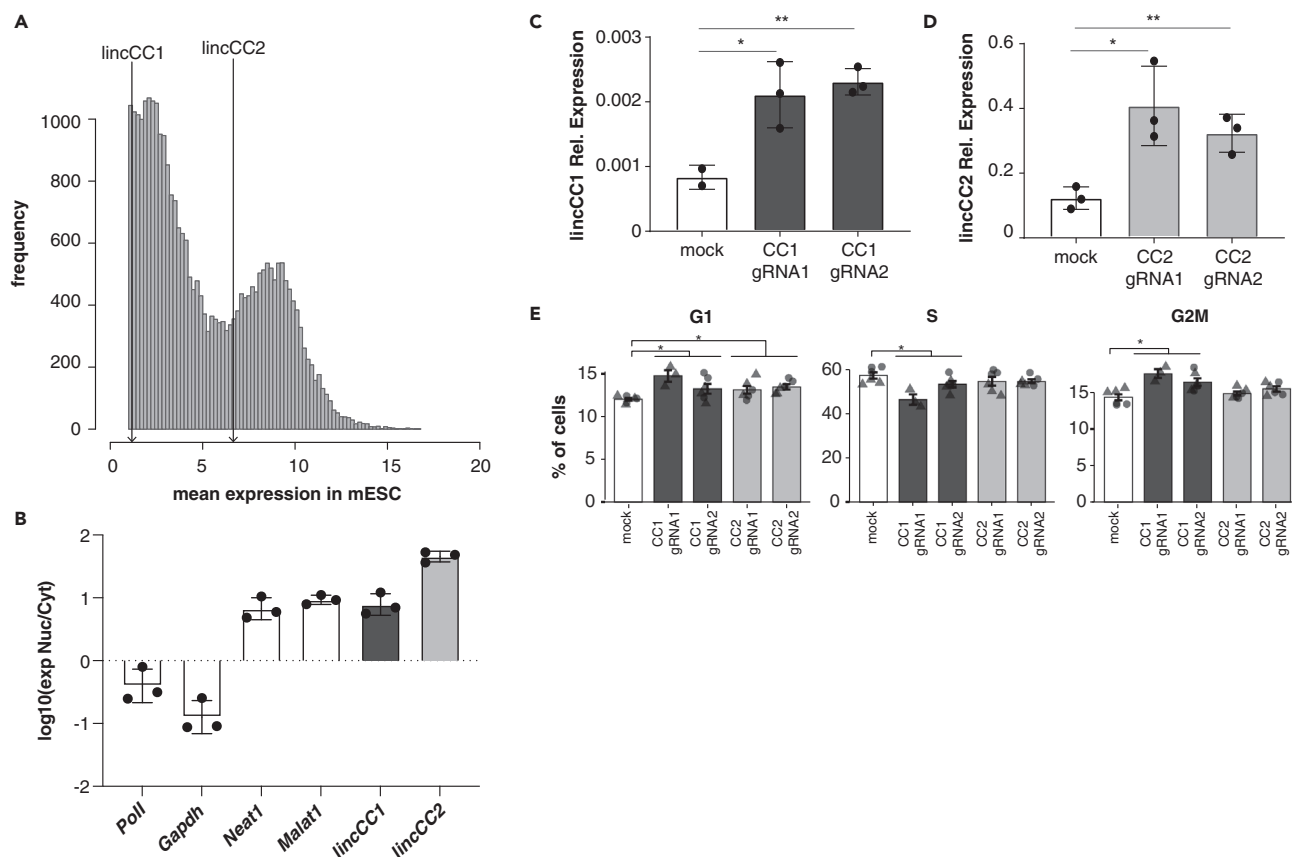
Figure 4. Candidate lincRNA Overexpression Impacts Cell Cycle Progression

(A) Histogram of the mean cell expression of all considered loci in mESCs. Lines indicate the expression of lincCC1 and lincCC2.

(B–D) (B) Ratio between the nuclear and cytoplasmic expression (log10) for two cytosolic mRNAs (*PolII* and *Gapdh*), two nuclearly retained lincRNAs (*Malat1* and *Neat1*), and lincCC1 and lincCC2. Each point corresponds to the results for one biological replicate. Relative (C) lincCC1 and (D) lincCC2 expression 72 h after co-transfection of Cas9-VP160 with scramble-, gRNA1-, or gRNA2-expressing constructs for three independent biological replicates.

(E) Percentage of mESCs at G1, S, and G2/M stages of the cell cycle 72 h post transfection of mESCs with scramble (white), lincCC1 (dark gray), or lincCC2 (light gray) targeting gRNAs. Technical replicates for two different biological replicates represented as either triangles or circles. Significance of comparisons is indicated as NS p > 0.05, *p < 0.05, **p < 0.01.

we also estimated the strength of the association with genes involved in pluripotency for non-cell cycle differentially expressed mESC coding (r = 0.03) and non-coding (r = −0.01) genes and found this to be significantly lower than that of CC-lincRNAs or CC-mRNAs (two-tailed Wilcoxon test p values: nonCC-mRNA = 0.002, nonCC-lncRNA = 0.0001, Figure 3G). Similar results are obtained when considering only significant correlations (correlation test p < 0.05) between pluripotency gene and coding or noncoding transcripts.

These results support the hypothesis that some CC-lincRNAs modulate specific aspects of mESC cell cycle and may contribute to the regulation of stem cell cycle adaptions.

Candidate CC-lincRNA Analysis Supports lincRNA Roles in Cell Cycle Regulation

Next, we aimed to experimentally validate the contributions of previously uncharacterized CC-lincRNAs to mESC cell cycle regulation. Considering that relatively low transcript abundance may limit individual lincRNA's ability to regulate cell cycle progression, we selected one modestly expressed CC-lincRNA (XLOC_018328, hereafter lincCC1, Figure S4A) and one CC-lincRNA that is relatively highly expressed (Gm26853, hereafter lincCC2, Figure S4B) in mESCs (Figure 4A). We determined subcellular localization of these transcripts in mESCs by subcellular fractionation followed by qPCR and found that both tested CC-lincRNAs are predominantly nuclear (Figure 4B).

We employed CRISPR activation (CRISPRa) to increase endogenous lincRNAs' transcription. For each lincRNA, we designed two guide RNAs (gRNA) to target VP160-fused dead-Cas9 (dCas9-VP160) (Cheng et al., 2013) to the vicinity of lincRNA promoters. As a control, we designed a non-targeting scrambled gRNA. We transiently co-transfected each of the gRNA-expressing constructs with a dCas9-VP160 expressing vector. Seventy-two hours post transfection we observed an average 2-fold upregulation (two-tailed unpaired t test p value CC1 gRNA1 = 0.02, CC1 gRNA2 = 0.0002, CC2 gRNA1 = 0.0001, CC2 gRNA2 = 0.0001) of candidate lincRNAs expression (Figures 4C and 4D) in mESCs treated with targeting gRNA, relative to control.

To assess the impact of changes in candidate lincRNA expression upon cell cycle progression in mESC, we determined the proportions of cells in G1, S, and G2/M phases using fluorescence-activated cell sorting (FACS). We used EdU incorporation and DNA content to identify the proportion of single cells in each cell cycle phase following CC-lincRNA overexpression (Figure S4C). We found that, consistently with the proposed role of lincCC1 and lincCC2 in the modulation of cell cycle progression, upregulation of these lincRNAs is associated with significant changes in the proportion of cells within different cell cycle stages. LincCC1 upregulation is associated with a significant increase (p value = 0.0002) in the number of cells in G2/M and small, yet significant changes in the proportion of cells in G1 (p = 0.005) and S phases (p value = 0.002) (Figure 4E). LincCC2 mostly affects the proportion of cells in G1 (p value = 0.02).

These observations are consistent with the role of these CC-lincRNAs in the regulation of cell cycle progression.

DISCUSSION

The cell cycle is central to the development of a multicellular organism from a single cell zygote and is required for cell renewal. Dysregulation of this process can cause disease, most notably cancer. Progression through the cell cycle requires dynamic and tightly regulated transitions between well-defined cell cycle stages, which are controlled by changes in cell cycle regulators' activity. Although control of cell cycle progression has been primarily assigned to protein-coding genes, several lincRNAs have recently been shown to contribute to this process (Kitagawa et al., 2013). For example, a recent siRNA screen in HeLa cells revealed that knockdown of 26 lincRNAs, whose expression is deregulated in cancer, impacts different aspects of cell cycle progression (Notzold et al., 2017). The role of lincRNAs in this process is further supported by a more recent comprehensive high-content RNAi screen in the same cells (Stojic et al., 2020). In parallel, characterization of individual lincRNAs has also illustrated how these noncoding RNAs contribute to the regulation of different aspects of cell cycle (Kitagawa et al., 2013).

In contrast to most cell cycle protein-coding regulators, lincRNA expression is often restricted in time and space, as supported by the analysis of their expression across adult and embryonic tissues and in single cells (for example, Derrien et al., 2012; Tuck et al., 2018; Ulitsky and Bartel, 2013). This suggests that most lincRNA-encoded functions are tissue specific. Although lincRNAs have been previously implicated in the regulation of cell cycle progression (Adriaens et al., 2016; Tripathi et al., 2013; Marchese et al., 2016), their role, if any, in cell type-specific modulations of this central biological process remains understudied. Here, we used mESC cycle, which is characterized by a truncated G1 phase, as a model system to assess the contributions of lincRNAs expressed in mESCs to the modulation of this stem cell adaptation.

To identify lincRNAs that are putative cell cycle regulators, we reasoned that for noncoding genes the functional moiety is the transcript; thus, differential noncoding gene expression between cell cycle stages would result in differential activity and enrich for lincRNAs with roles in cell cycle regulation. Consistent with this hypothesis, our genome-wide analyses and experimental validations support the role of lincRNAs differentially expressed between mESC cell cycle stages as modulators of cell cycle progression, likely through interaction with other cell cycle regulators.

Interestingly and relative to protein-coding genes, whose activity during cell cycle is often modulated by post-translational modifications, lincRNAs are enriched among differentially expressed transcripts, supporting that, in mESCs, their expression is more frequently dynamic throughout cell cycle progression. The fraction of cell cycle-regulated lincRNAs is similar to a previous estimate based on HeLa cells bulk RNA sequencing (~9% [Dominguez et al., 2016]). However, the dynamics of lincRNA expression throughout the cell cycle differs between cell types, as in contrast with this earlier study in HeLa (Dominguez et al., 2016) that revealed no preferential cell stage-specific expression in mESCs, we found most lincRNAs are highly

expressed in G1 and differentially expressed between G1 and S phases. Given the critical importance of the G1-S transition in maintenance of embryonic stem cell state (Boward et al., 2016), we hypothesize that a subset of lincRNAs with roles in mESC cell cycle progression also have roles in maintenance of stem cell state. Supporting this hypothesis is the observation that these lincRNAs' expression is often restricted to pluripotent cells and rapidly decreases upon exit from pluripotency and entry into neural commitment. Furthermore, the expression of these lincRNAs is often regulated by core pluripotency transcription factors, including Nanog, Sox2, and Oct4. Finally, we provide preliminary evidence that cell cycle lincRNAs are part of the network underlying pluripotency.

Although further work is now required to disentangle how individual lincRNAs modulate cell cycle progression in mESCs, our results suggest that tissue-specific regulation by lincRNAs contributes to cell type-specific adaptations of ubiquitous processes.

Limitations of the Study

The present study aimed to identify the contributions of lincRNAs, which are generally tissue specifically expressed, to embryonic stem cell type-specific adaptations. Toward this end we explored publicly available cell cycle-staged single-cell RNA sequencing data and analyzed the impact of candidate lincRNA expression on cell cycle progression.

Our computational analysis was limited by the absence of an alternative dataset that could be used to replicate our observations. One consequence of this limitation is that analysis presented in Figure 2 uses the same data as that used to identify differential expressed genes and is therefore not independent.

The candidate analysis we reported is limited by the absence of efficient approaches to deplete nuclear lincRNA expression in mESCs. As a consequence and despite several attempts, using either shRNAs or CRISRPi, we were unable to transiently or constitutively deplete lincCC1 or CC2. Establishing and characterizing mESCs where candidate lincRNA function is lost would have allowed us to understand the biological and molecular contributions of lincRNAs to mESC cell cycle adaptations.

Resource Availability

Lead Contact

Ana Claudia Marques, anaclaudia.marques@unil.ch.

Data and Code Availability

All code used is available upon requestMaterials availability Is not relevant because no new materials were generated in the context of this project.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101291.

REFERENCES

Adriaens, C., Standaert, L., Barra, J., Latil, M., Verfaillie, A., Kalev, P., Boeckx, B., Wijnhoven, P.W., Radaelli, E., Vermi, W., et al. (2016). p53 induces formation of NEAT1 lncRNA-containing paraspeckles that modulate replication stress response and chemosensitivity. Nat. Med. *22*, 861–868.

Boward, B., Wu, T., and Dalton, S. (2016). Concise review: control of cell fate through cell cycle and pluripotency networks. Stem Cells *34*, 1427–1436.

Buettner, F., Natarajan, K.N., Casale, F.P., Proserpio, V., Scialdone, A., Theis, F.J., Teichmann, S.A., Marioni, J.C., and Stegle, O. (2015). Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. Nat. Biotechnol. *33*, 155–160.

Chambers, I., and Tomlinson, S.R. (2009). The transcriptional foundation of pluripotency. Development *136*, 2311–2322.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106–1117.

Cheng, A.W., Wang, H., Yang, H., Shi, L., Katz, Y., Theunissen, T.W., Rangarajan, S., Shivalila, C.S., Dadon, D.B., and Jaenisch, R. (2013). Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. Cell Res. *23*, 1163–1171.

Cho, R.J., Huang, M., Campbell, M.J., Dong, H., Steinmetz, L., Sapinoso, L., Hampton, G., Elledge, S.J., Davis, R.W., and Lockhart, D.J. (2001). Transcriptional regulation and function during the human cell cycle. Nat. Genet. *27*, 48–54.

Clark, M.B., Johnston, R.L., Inostroza-Ponta, M., Fox, A.H., Fortini, E., Moscato, P., Dinger, M.E., and Mattick, J.S. (2012). Genome-wide analysis of long noncoding RNA stability. Genome Res. *22*, 885–898.

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

Dominguez, D., Tsai, Y.H., Gomez, N., Jha, D.K., Davis, I., and Wang, Z. (2016). A high-resolution transcriptome map of cell cycle reveals novel connections between periodic genes and cancer. Cell Res. *26*, 946–962.

Kitagawa, M., Kitagawa, K., Kotake, Y., Niida, H., and Ohhata, T. (2013). Cell cycle regulation by long non-coding RNAs. Cell Mol. Life Sci. *70*, 4785–4794.

Kryuchkova-Mostacci, N., and Robinson-Rechavi, M. (2017). A benchmark of gene expression tissue-specificity metrics. Brief Bioinform. *18*, 205–214.

Lal, A., Navarro, F., Maher, C.A., Maliszewski, L.E., Yan, N., O'day, E., Chowdhury, D., Dykxhoorn, D.M., Tsai, P., Hofmann, O., et al. (2009). miR-24 Inhibits cell proliferation by targeting E2F2, MYC, and other cell-cycle genes via binding to "seedless" 3'UTR microRNA recognition elements. Mol. Cell *35*, 610–625.

Lee, J., Go, Y., Kang, I., Han, Y.M., and Kim, J. (2010). Oct-4 controls cell-cycle progression of embryonic stem cells. Biochem. J. *426*, 171–181.

Liu, Y., Chen, S., Wang, S., Soares, F., Fischer, M., Meng, F., Du, Z., Lin, C., Meyer, C., Decaprio, J.A., et al. (2017). Transcriptional landscape of the human cell cycle. Proc. Natl. Acad. Sci. U S A *114*, 3473–3478.

Marchese, F.P., Grossi, E., Marin-Bejar, O., Bharti, S.K., Raimondi, I., Gonzalez, J., Martinez-Herrera, D.J., Athie, A., Amadoz, A., Brosh, R.M., Jr., and Huarte, M. (2016). A long noncoding RNA regulates sister chromatid cohesion. Mol. Cell *63*, 397–407.

Morgan, D.O. (1995). Principles of CDK regulation. Nature *374*, 131–134.

Mukherjee, N., Calviello, L., Hirsekorn, A., De Pretis, S., Pelizzola, M., and Ohler, U. (2017). Integrative classification of human coding and noncoding genes through RNA metabolism profiles. Nat. Struct. Mol. Biol. *24*, 86–96.

Murray, A., and Hunt, T. (1993). The Cell Cycle, an Introduction (Oxford University Press).

Notzold, L., Frank, L., Gandhi, M., Polycarpou-Schwarz, M., Gross, M., Gunkel, M., Beil, N., Erfle, H., Harder, N., Rohr, K., et al. (2017). The long non-coding RNA LINC00152 is essential for cell cycle progression through mitosis in HeLa cells. Sci. Rep. *7*, 2265.

Patterson, D.G., Roberts, J.T., King, V.M., Houserova, D., Barnhill, E.C., Crucello, A., Polska, C.J., Brantley, L.W., Kaufman, G.C., Nguyen, M., et al. (2017). Human snoRNA-93 is processed into a microRNA-like RNA that promotes breast cancer cell invasion. NPJ Breast Cancer *3*, 25.

Pena-Diaz, J., Hegre, S.A., Anderssen, E., Aas, P.A., Mjelle, R., Gilfillan, G.D., Lyle, R., Drablos, F., Krokan, H.E., and Saetrom, P. (2013). Transcription profiling during the cell cycle shows that a subset of Polycomb-targeted genes is upregulated during DNA replication. Nucleic Acids Res. *41*, 2846–2856.

Santos, A., Wernersson, R., and Jensen, L.J. (2015). Cyclebase 3.0: a multi-organism database on cell-cycle regulation and phenotypes. Nucleic Acids Res. *43*, D1140–D1144.

Schmitt, A.M., and Chang, H.Y. (2016). Long noncoding RNAs in cancer pathways. Cancer Cell *29*, 452–463.

Soufi, A., and Dalton, S. (2016). Cycling through developmental decisions: how cell cycle dynamics control pluripotency, differentiation and reprogramming. Development *143*, 4301–4311.

Stojic, L., Lun, A.T.L., Mascalchi, P., Ernst, C., Redmond, A.M., Mangei, J., Barr, A.R., Bousgouni, V., Bakal, C., Marioni, J.C., et al. (2020). A high-content RNAi screen reveals multiple roles for long noncoding RNAs in cell division. Nat. Commun. *11*, 1851.

Sun, Y., and Ma, L. (2019). New insights into long non-coding RNA MALAT1 in cancer and metastasis. Cancers (Basel) *11*, 216.

Sun, N., Yu, X., Li, F., Liu, D., Suo, S., Chen, W., Chen, S., Song, L., Green, C.D., McDermott, J., et al. (2017). Inference of differentiation time for single cell transcriptomes using cell population reference data. Nat. Commun. *8*, 1856.

Tan, J.Y., Vance, K.W., Varela, M.A., Sirey, T., Watson, L.M., Curtis, H.J., Marinello, M., Alves, S., Steinkraus, B., Cooper, S., et al. (2014). Cross-talking noncoding RNAs contribute to

cell-specific neurodegeneration in SCA7. Nat. Struct. Mol. Biol. *21*, 955–961.

Tan, J.Y., Sirey, T., Honti, F., Graham, B., Piovesan, A., Merkenschlager, M., Webber, C., Ponting, C.P., and Marques, A.C. (2015). Extensive microRNA-mediated crosstalk between lncRNAs and mRNAs in mouse embryonic stem cells. Genome Res. *25*, 655–666.

Tripathi, V., Shen, Z., Chakraborty, A., Giri, S., Freier, S.M., Wu, X., Zhang, Y., Gorospe, M., Prasanth, S.G., Lal, A., and Prasanth, K.V. (2013). Long noncoding RNA MALAT1 controls cell cycle progression by regulating the expression of oncogenic transcription factor B-MYB. PLoS Genet. *9*, e1003368.

Tuck, A.C., Natarajan, K.N., Rice, G.M., Borawski, J., Mohn, F., Rankova, A., Flemr, M., Wenger, A., Nutiu, R., Teichmann, S., and Buhler, M. (2018). Distinctive features of lincRNA gene expression suggest widespread RNA-independent functions. Life Sci. Alliance *1*, e201800124.

Ulitsky, I., and Bartel, D.P. (2013). lincRNAs: genomics, evolution, and mechanisms. Cell *154*, 26–46.

Weinberg, R.A. (1995). The retinoblastoma protein and cell cycle control. Cell *81*, 323–330.

Whitfield, M.L., Sherlock, G., Saldanha, A.J., Murray, J.I., Ball, C.A., Alexander, K.E., Matese, J.C., Perou, C.M., Hurt, M.M., Brown, P.O., and Botstein, D. (2002). Identification of genes periodically expressed in the human cell cycle and their expression in tumors. Mol. Biol. Cell *13*, 1977–2000.

Xu, H., Baroukh, C., Dannenfelser, R., Chen, E.Y., Tan, C.M., Kou, Y., Kim, Y.E., Lemischka, I.R., and Ma'ayan, A. (2013). ESCAPE: database for integrating high-content published data collected from human and mouse embryonic stem cells. Database (Oxford) *2013*, bat045.

Yan, P., Luo, S., Lu, J.Y., and Shen, X. (2017). Cis- and trans-acting lncRNAs in pluripotency and reprogramming. Curr. Opin. Genet. Dev. *46*, 170–178.

Yu, X., Li, Z., Zheng, H., Chan, M.T., and Wu, W.K. (2017). NEAT1: a novel cancer-related long non-coding RNA. Cell Prolif. *50*, e12329.

Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D., et al. (2014). A comparative encyclopedia of DNA elements in the mouse genome. Nature *515*, 355–364.

Zaveri, L., and Dhawan, J. (2018). Cycling to meet fate: connecting pluripotency to the cell cycle. Front. Cell Dev. Biol. *6*, 57.

Zhang, X., Neganova, I., Przyborski, S., Yang, C., Cooke, M., Atkinson, S.P., Anyfantis, G., Fenyk, S., Keith, W.N., Hoare, S.F., et al. (2009). A role for NANOG in G1 to S transition in human embryonic stem cells through direct binding of CDK6 and CDC25A. J. Cell Biol. *184*, 67–82.

Supplemental Information

The Contribution of lincRNAs

at the Interface between Cell Cycle

Regulation and Cell State Maintenance

Adriano Biasini, Adam Alexander Thil Smith, Baroj Abdulkarim, Maria Ferreira da Silva, Jennifer Yihong Tan, and Ana Claudia Marques

SUPPLEMENTARY INFORMATION

TRANSPARENT METHODS

Processing of cell cycle staged single cell RNA sequencing data

Mouse genome sequence and annotation files were downloaded from ENSEMBL (GRCm38.82 i.e. mm10). LincRNAs expressed in mESCs (Tan et al., 2014) were added to the transcript annotations using custom Python scripts in order to appropriately deal with novel transcripts overlapping existing loci, leading to a total of 55,596 loci. The FASTA sequences of the 96 ERCC spike-ins used in the experiment were downloaded from UCSC GoldenPath. We used CGAT ((Heger et al., 2013), v 0.2.3) to extract transcript sequences longer than 200 nucleotides, and Kallisto ((Bray et al., 2016)**,** v 042.4, default parameters) to build a transcriptome index that included 114,842 transcripts across 49,285 loci and ERCCs. The FASTQ files for staged mESC single-cell RNA-sequencing were downloaded from ArrayExpress (accession number E-MTAB-2805). For each cell, we estimated transcript expression using Kallisto (v. 0.42.4) (Bray et al., 2016) without bootstraps. Hit counts were imported into R and summarized to the gene level using tximport (Soneson et al., 2015). After removing genes without any hits in any cells, only 38,016 genes remained.

Basic cell-level quality control was performed as previously described (Buettner et al., 2015).   Briefly, we excluded cells: 1) with a cell-wise gene detection rate below 20%; 2) cells with <1.6M total hits; 3) cells with proportions of hits to ERCCs lower than 15% or higher than 65%; 4) cells with a proportion of hits to mitochondrial genes lower than 1% or higher than 10%; and 5) cells with an estimated cell size higher or lower than 2 median absolute deviations from the stage-median cell size. In total, after these filtering steps, 94/95 G1 cells, 76/88 S cells, and 74/96 G2/M cells were retained for further analysis. We then performed gene-level quality control. We considered only genes that had at least one hit count in at least 4% of samples. Hit counts were normalized using the scran package (v1.0.4, (Lun et al., 2016)). Size factors were calculated using pools of samples of sizes 10 to 35 cells from each stage. We used scLVM (Brennecke et al., 2013, Buettner et al., 2015) to filter out genes whose total variance is not significantly higher than that expected from pure technical variability (fitTechnicalNoise

parameters: mincv2=0.01, quan=0.10 ; getVariableGenes parameters: threshold = 0.10, minBiolDisp = 0.30), and subsequently also removed ERCCs & mitochondrial genes. After this final filtering step, 10,487 genes were kept.

Except for differential expression analysis, where scran-normalised counts were used, all subsequent analyses were performed on scran-normalised shifted log10 counts (sl10 = log10(normalised counts + 1)).

Processing of bulk RNA sequencing for a time-course of mESC to neurons differentiation

We used STAR (Dobin et al., 2013) (v 2.5.3a) to map bulk RNA sequencing data for a 216-hour time course of mESC differentiation into neurons (Sun et al., 2017) to the Mouse transcriptome. Transcript quantification was performed using RSEM (Li and Dewey, 2011) (v 1.3.0) and for each time point we considered the average of expression across time points. We estimated the Spearman correlation between genes of interest and pluripotency associated genes (Xu et al., 2013) using R.

Differential expression analysis

Kallisto pseudo-count data was imported into R (v 3.3.1) and prepared for DESeq2 (v 1.12.4) (Love et al., 2014) using the *tximport* (Soneson et al., 2015) package (v 1.0.3). Principal Components Analysis (PCA) was applied using R package FactoMineR (Le et al., 2008) to an expression matrix consisting of all cells passing QC in rows, and all genes passing QC in columns. The cell cycle signal typically appeared most strongly across the first two principal components, with the third component correlating strongly with the gene dropout rate (not shown). Differentially expressed genes (FDR< 5%, log2 fold change > 0.1) were called between pairs of cell cycle stages using DESeq2 (Love et al., 2014, Jaakkola et al., 2017) using appropriate contrasts.

In order to assess whether the overall lower expression of lincRNAs affected our ability to call them as DE, we performed a down-sampling experiment to bring mRNA expression down to levels comparable to lincRNAs. We first estimated the average scaling factor between mRNA and lincRNA expression in the original data. We estimated the difference between the median Transcripts Per Million reads (TPM) in quality-controlled cells between lincRNAs and mRNAs to be 14.196, i.e. on average

mRNAs are 14x more highly expressed than lincRNAs. We randomly sub-sampled the original FASTQ files to 1/14th of their original size (i.e. using a factor of 0.0714) using seqtk (v 2015.10.15). We quantified transcript expression using Kallisto and filtered cells and genes as described above prior to differential gene expression analysis. For the sub-sampled data, the number of tested genes was similar to the full data (9 923 vs 10 391). We then counted the number of mRNA genes called as DE in the full analysis that passed the quality control filters for the sub-sampled data (388), and the number of those called as DE in the sub-sampled analysis (233, i.e. 60.1%).

Co-expression analysis.

We considered a gene to be a cell cycle regulator if: 1) was annotated as a cell cycle gene (Buettner et al., 2015); 2) it was annotated with a GO term containing the string "cell cycle" or 3) had a Human ortholog annotated in CycleBase (Santos et al., 2015). Pluripotency-associated genes were extracted from the ESCAPE database (Xu et al., 2013).

Co-expression analysis with cell cycle genes was performed using the single cell RNA sequencing data with staged mESCs (Buettner et al., 2015). We considered all loci expressed (normalized shifted log10>0.14) in at least 50 cells (out of a possible 244). The expression cut-off was defined based on inspection of the shifted log10 expression values. We estimated the Spearman correlation between all these loci and considered a gene pair to be significantly correlated if the absolute Spearman correlation between their expression was > 0.10 and the correlation test p-value was < 0.10. The correlation between any given pair was only calculated in cells where both genes were expressed. Correlations for gene pairs where there were less than 20 such value pairs were set to 0.

Co-expression analysis with pluripotency genes was performed using the time course of mESC to neurons differentiation data, as described above.

Enrichment analyses

Enrichment analyses for Gene Ontology Biological Process, KEGG and REAC terms in *Mus musculus* were carried out using R package "gProfileR" (v. 0.6.1) (Reimand et al., 2016). Terms were considered to be significantly enriched if the associated

corrected p-value was below 10% relative to the specified background of mESC expressed genes.

Enrichment in binding of mESC transcription factors (TFs) (Chen et al., 2008) at CC-lincRNA promoters (defined as +/-1 kb from their annotated TSS) was estimated using the Genome Association Tester (GAT) (Heger et al., 2013). GAT tests whether the enrichment of TF binding at CC-lincRNA promoters is different from what would be expected based on 10,000 random samplings (with replacement) of intergenic segments with the same length and GC content as the CC-lincRNA promoters. The gat-compare tool was used to test the significance of the enrichment of TF binding at CC-lincRNA promoters relative to that at non CC-lincRNA promoters. Core TF binding sites were obtained from (Chen et al., 2008).

Loci tissue specificity metrics in Mouse

We manually selected 150 Mouse paired-end total RNA bulk RNA-sequencing datasets from the ENCODE project (Consortium, 2012), covering a range of tissues, sexes and ages, and added 3 publicly-available Mouse embryonic stem cell datasets (Tan et al., 2015). We defined a "tissue.simple" grouping of samples, based on a high-level description of the tissue of origin (e.g. central nervous system, or heart) and of the developmental stage (adult or embryonic). Gene expression was estimated using Kallisto and imported into R using tximport as described above. Library size normalization was performed with DESeq2, and counts were transformed into shifted log10 normalised values. Values below a background level of 0.1 were set to 0. This cut-off was obtained based on the distribution of shifted log10 expression values. Genes with no expression values above 0 in any sample were discarded. Expression values were averaged across technical and biological replicates using the median. Tau was calculated as described in (Kryuchkova-Mostacci and Robinson-Rechavi, 2017), briefly as Tau = sum(1-(v/max(v)))/(length(v)-1), where v is a vector of average expression values for different tissues.

Cell Culture

ES-E14-Tg2a (E14) cells were grown on 0.1% gelatin-coated tissue culture dishes, in DMEM (Thermo Fischer, 41965-039) supplemented with 1x Non-Essential Amino

Acids (Thermo Fischer 11140-035), 50 uM β-mercaptoethanol (Thermo Fischer 31350-10), 15% Fetal Bovine Serum (Thermo Fischer 10499-044), 500 Units/ml of Pennicilin/Streptomycin (Thermo Fischer, 15140122), and 100 Units/ml of Recombinant mouse LIF Protein (Merck ESG1107). Culture were seeded at an average density of ~ 3.8×10^4 cells/cm^2 and passaged every 48 h.

Analysis of cell cycle stage and subcellular gene expression

We resuspended 10^6 cells in 1 ml of PBS, containing 1 uL/mL of fluorescent reactive dye (LIVE/DEAD) and incubated in the dark for 30 minutes at room temperature. Cells were then washed with PBS, spun, and resuspended in fresh medium. We add 2 uL of DNA Hoechst 33324 dye (20 mM) and cells were incubated at 37°C for 20 minutes. Following incubation, cells were centrifuged at 4°C for 4 minutes at 400 g and the resulting cell pellet was resuspended in 500 ul of 3% FBS in PBS with 0.1% EDTA and kept on ice. Unstained and single dye controls prepared in parallel were used for FACS calibration.

Subcellular fractionation and RNA extraction of mESCs was carried out using the PARIS kit (Thermo Fisher, AM1921) according to the manufacturer's instructions. Genomic DNA removal from the cytosolic and nuclear RNA fractions was performed through RNeasy (Qiagen, 74104) on-column DNAse digestion (Qiagen, 79254) was performed according to manufacturers instructions.

Cells were sorted in an AstriosEQ (Beckmann Coulter) cell sorter and collected in 500 uL of RTL buffer from the RNeasy extraction kit (QIAGEN). Forward and side scatter (FSC & SSC, 488nm DPSS laser) were used as is common for size and doublet exclusion. We excluded dead cells based on LIVE/DEAD fluorescence (488nm DPSS laser) and used DNA content (355nm DPSS laser) to define three gates: G1, S and G2/M (Supplementary Figure S5).

RNA was extracted using RNeasy mini kit (Qiagen) according to manufacturer's instructions. RNA was reverse transcribed using Quatitect Reverse Transcription Kit (Qiagen 205313). Quantitative PCR analysis was performed using SYBR Green Quantitative PCR Master Mix (Roche 0692404001) in a Light Cycler 96 Real Time PCR system (Roche). Briefly, the RTqPCR reaction was assembled in 10uL with

forward and reverse primers at a final concentration of .5 uM each, and SYBR Green
Quantitative PCR Master Mix at a final concentration of 1X.

Candidate lincRNA overexpression and cell cycle analysis.

We used E-CRISPR (Heigwer et al., 2014) to design 2 guide RNA (gRNA) sequences
located within 1 kb of lincCC1 and lincCC2 promoters respectively. We synthetized
oligos containing these sequences and BbsI restriction sites. Annealed oligos were
inserted into the pKLV-U6gRNA(BbsI)-PGKpuro2ABFP vector (Ochiai et al., 2015).
We seeded $1.5*10^5$ E14 cells/well in 6-well plates and allowed cells to grow overnight.
We used lipofectamine 2000 (Thermo Fischer) to transfect 1000 ng of gRNA
expression vector and 1000 ng of pAC95-pmax-dCas9-Vp160-2A-Neo (Cheng et al.,
2013) vector into mESC cultures overnight according to manufacturer's instructions.
We replaced growth medium 8 hours after transfection to minimize toxicity. Seventy-
two hours post transfection one sample was collected for RNA to ensure the efficiency
of the over-expression by qPCR; in parallel the other cells were pulsed with 10 uM
Edu for 30 minutes in the dark at 37ºC. Cells were washed with 1% BSA in PBS and
trypsinized cells. Following a second wash with 1% BSA in PBS, cells were fixed at
room temperature, protected from the light for 15 minutes in 100 uL of Click-iT fixative
solution. After fixation, cells were washed in 1% BSA in PBS, resuspended in 100 uL
of 1x Click-iT saponin-based permeabilization and wash reagent, and incubated at
room temperature in the dark for 15 minutes. Following incubation, 500 uL of Click-iT
reaction cocktail containing Alexa Fluor 488 Fluorescent Dye Azide were added and
samples were incubated at room temperature in the dark for 30 minutes. Reaction was
quenched by resuspending cells in 1 mL of 1x Click-iT saponin-based permeabilization
reagent. To stain the samples for DNA content, 1 uL of Fx Cycle FarRed staining
solution (Thermo Fischer F10348) was added to each sample, and Pure Link RNase
A (Thermo Fischer 12091039) was added at a final concentration of 100 ng/mL. The
samples were incubated at 4ºC in the dark for 30 minutes, analyzed by flow cytometry
analysis. Cells stained for DNA incorporation and content were analyzed using a 10
color/3 laser Beckman Coulter Gallios Analyzer. Blue (405 nm) and red (640 nm)
excitation lasers were used for excitation of the Alexa Fluor 488 Fluorescent Dye Azide
and Fx FarCycle Red respectively. Emission was detected using channels FL-09 and

FL-06 for Alexa Fluor 488 and Fx FarCycle Red respectively. Two biological replicates, with 3 technical replicates each, were tested for both candidate lincRNAs using two different guide RNAs (gRNA1 and gRNA2). We excluded flow cytometry data for one of CC1's gRNA1 experiments due to issues with RNA collection preventing us from confirming candidate overexpression. 30-60k events were flowed per sample, except for the first experiment where 15-30k events were flowed.

Flow data was analyzed in FlowJo (v10.2). Flow events were gated on Cells and Singlets, and spillover was compensated. Low incorporation 2n and 4n populations, as well as a medium-to-high incorporation population, were easily identifiable in the DNA incorporation vs content scatterplot, and gated respectively as G1, G2/M and S (and ungated). This stage-level gating was performed by two gaters, blind to the direction and amplitude of target lincRNA expression modulation. Stage population fractions were comparable between gaters (Pearson correlations between stage fractions 0.65-0.97 depending on stage & experiment), and thus all results presented here are those of one gater.

Sample-level summaries of the flow cytometry results were generated using FlowJo (v10.2). Summaries were imported into R. Samples were quality-controlled based on their percentages of cells/singlets (indicative of cell health), and MFIs (indicative of staining strength). Experiments for which the associated qPCRs did not validate over-expression of the targets were excluded. Ultimately, 27 samples from 2 experiments and 5 treatment groups (scrambled, two different guides for lincCC1 and two for lincCC2) were retained. For each stage, we ran a linear mixed model (restricted log-likelihood maximisation, R package nlme version 3.1-137) to investigate the impact of treatment on the percentage of cells in that stage. We used the log2-transformed percentage as outcome, target lincRNA (scrambled, lincCC1, lincCC2) as a fixed effect, and biological replicates as a random intercept.
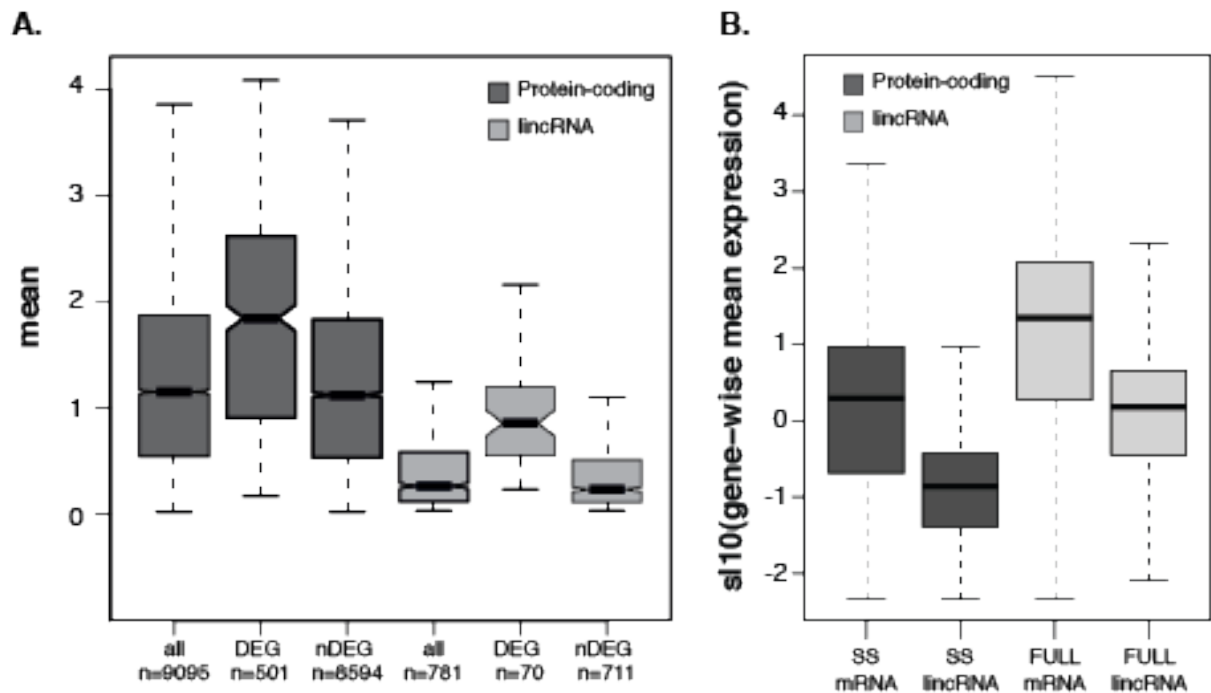
SUPPLEMENTARY FIGURES
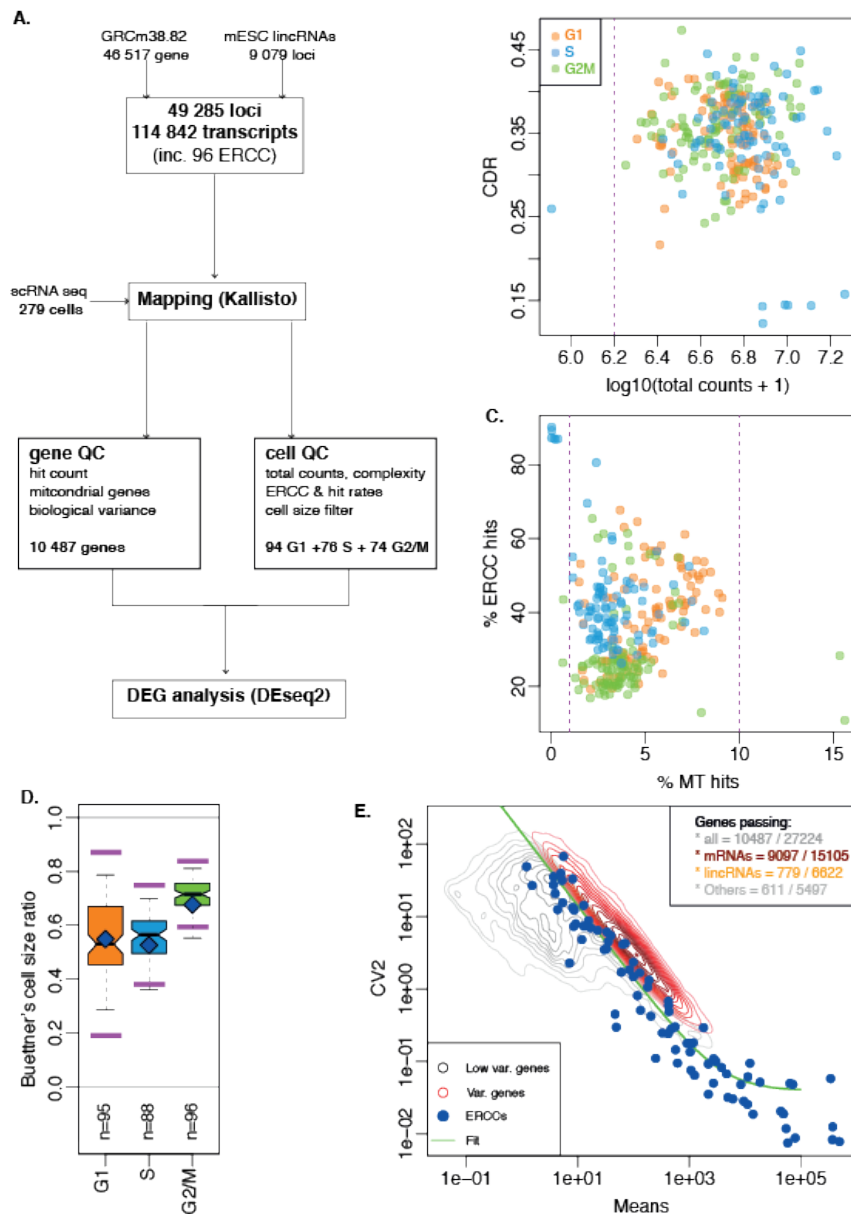
Supplementary Figure S1 . Schematics of the scRNA sequencing analysis, related to Figure 1.

A) Schematic representation of the differential expression analysis pipeline. **B)** Scatterplot showing for each sample (cell) the total hit (mapped read) counts & "Cell Detection Rate", i.e. fraction of genes detected by at least 1 hit in each cell. Low library quality exclusion thresholds are indicated by dashed magenta lines. Cells are coloured by cell cycle stage. **C)** Scatterplot showing for each cell the percent of hits to mitochondrial (MT) genes & to ERCC spike-ins. Low library quality thresholds are indicated by dashed magenta lines. **D)** Distributions of Buettner's cell size ratio (Buettner et al., 2015) for cells in each stage. Magenta lines show medians +/- 2 median absolute deviations; cells falling outside of these windows are excluded. **E)**

Technical noise filter (Brennecke et al., 2013) as implemented in scLVM (Buettner et al., 2015): each gene is represented by its cross-sample mean expression (x-axis) and coefficient of variation (squared: CV2, y-axis). Blue dots correspond to ERCC spike-ins, while the green line represents the non-linear CV2 ~ mean model that was fit to the spike-ins, and then used to test which genes show significantly higher-than-technical variability. Genes in black do not pass the significance test, those in red do. Counts of genes passing the test are indicated in the top-right hand corner.

Supplementary Figure S2. Cell-cycle stage specific expression, related to Figure 1.

A) boxplots of cross-sample mean expression (normalised shifted log10 counts) for genes of various classes: all=tested for differential expression. DE=differentially-expressed. NDE: non-differentially expressed. mRNA: protein-coding genes. lincRNA: lincRNA loci. **B)** Boxplots of shifted log10 mean loci expression for protein-coding genes (mRNA, dark grey) and lincRNAs (light grey), in both the sub-sampled (1/14th) data (SS) and FULL data from cell cycle staged single-cell RNAseq dataset (Buettner et al., 2015).
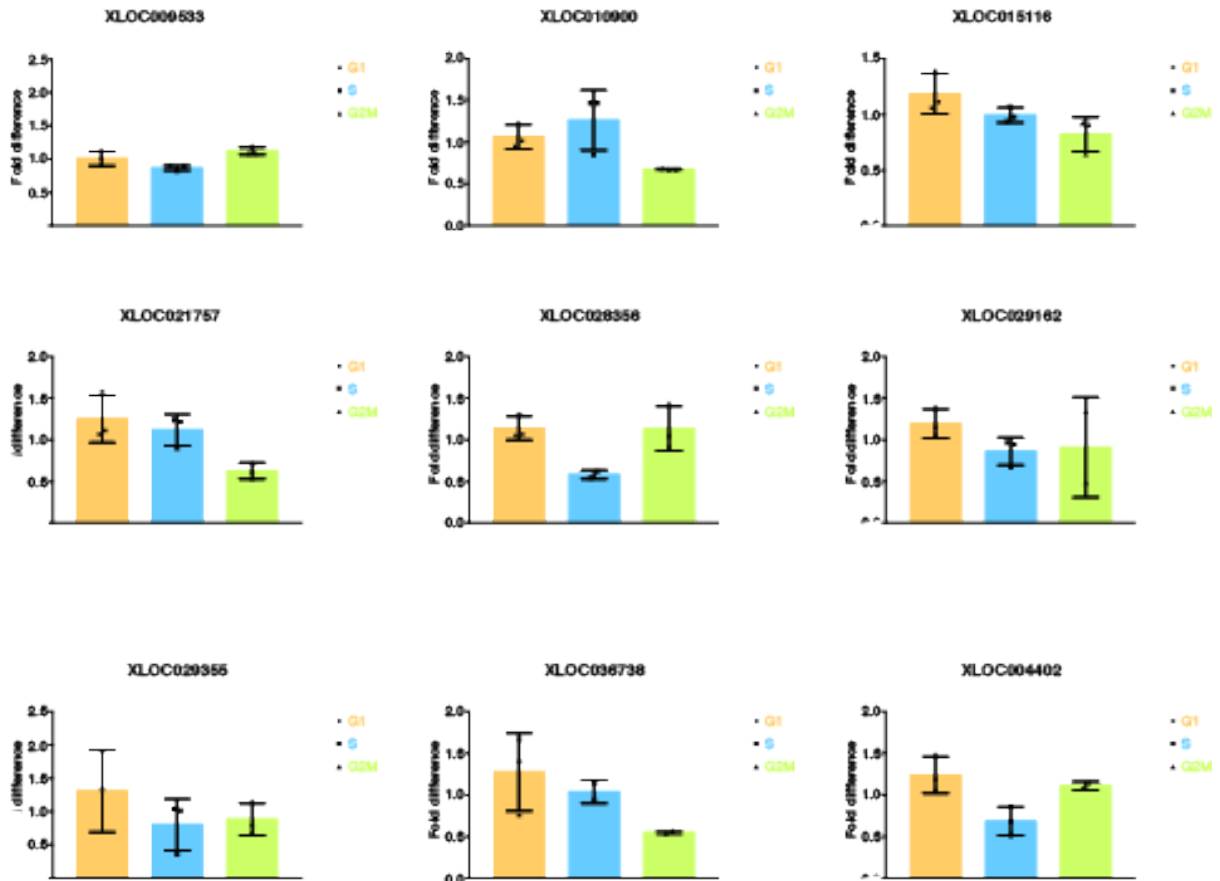
Supplementary Figure S3. RT-qPCR analysis of selected lincRNA's expression in mESC sorted by cell cycle stage based on DNA content, related to Figure 1. Fold difference in relative expression for 9 predicted cell-cycle differentially expressed lincRNAs.
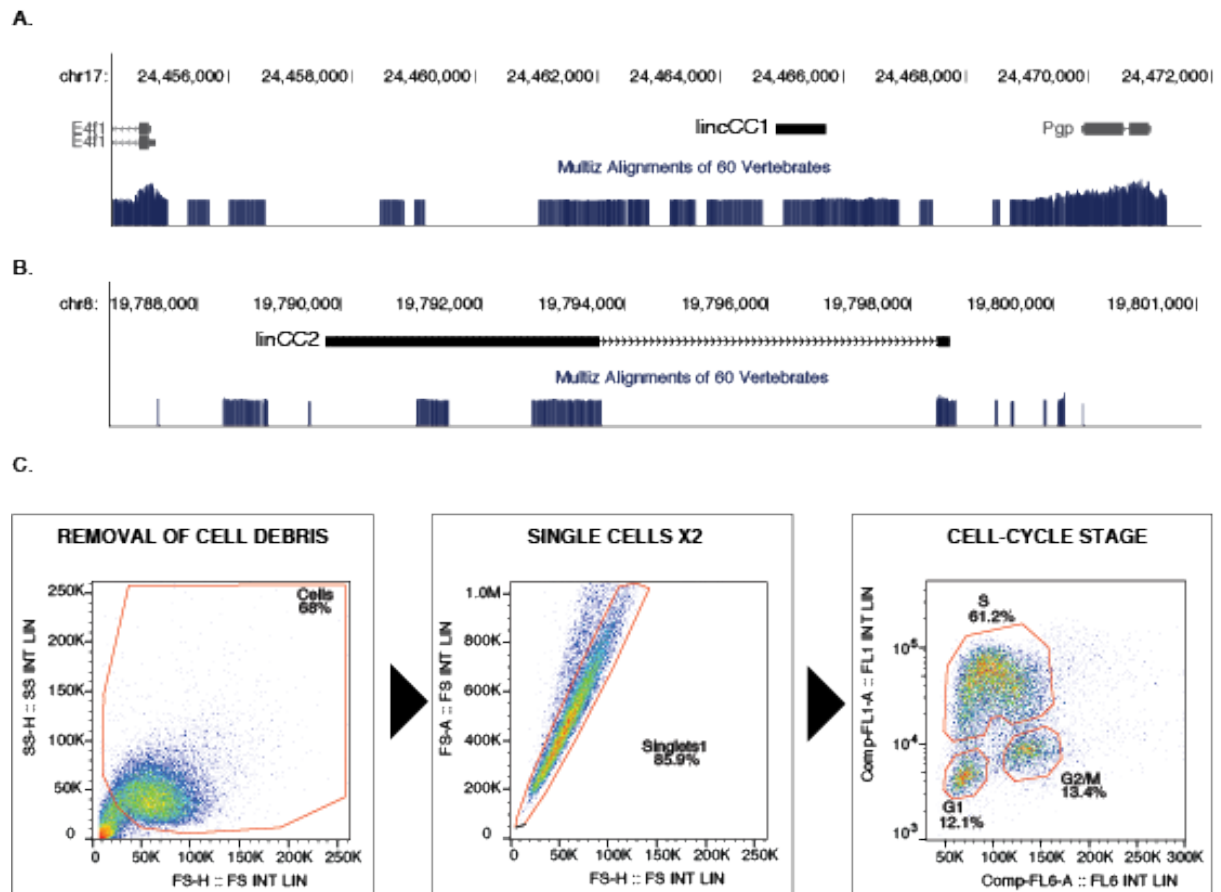
Supplementary Figure S4. Characteristics of candidate CC-lincRNA, related to Figure 4.

Genome browser view of the genomic location (mm10) of **A)** lincCC1 and **B)** lincCC2. Blue bars indicate conservation in multiZ alignments (60 vertebrates). **C)** gating strategy to identify cells from distinct cell cycle stages. Panel 1 shows the size exclusion gate based on FSC-H & SSC-H for removing debris. Panels 2 shows the FSC-H vs FSC-A doublet exclusion gate; a similar gate was implemented for SSC-H vs SSC-A. Panel 3 shows the partitioning of cells into 3 stages based on DNA content (measured on FL6-A) and EdU incorporation (measured on FL1-A).
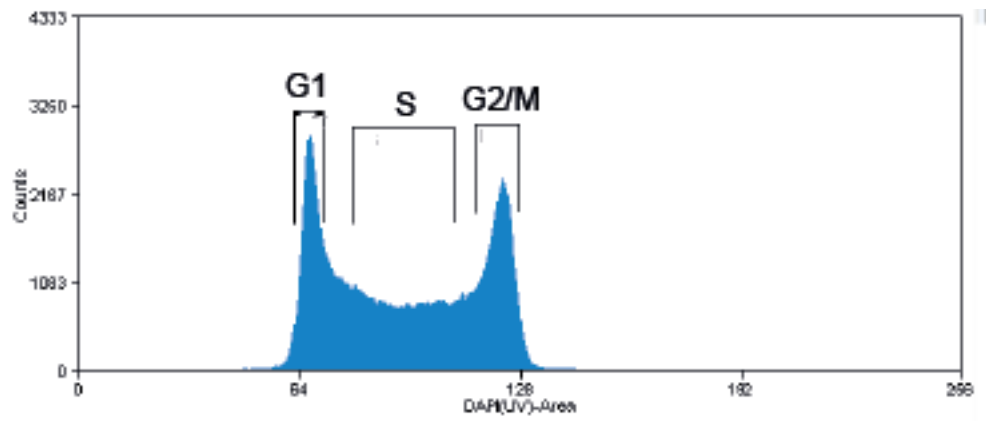
Supplementary Figure S5. Representative DNA content histogram for mESC, related to Figure 4.

Histogram of cellular DNA content measures as stained by DAPI, showing gates for delineating clearly-defined G1, S or G2/M stages.

REFERENCES

BRAY, N. L., PIMENTEL, H., MELSTED, P. & PACHTER, L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol,* 34**,** 525-7.

BRENNECKE, P., ANDERS, S., KIM, J. K., KOLODZIEJCZYK, A. A., ZHANG, X., PROSERPIO, V., BAYING, B., BENES, V., TEICHMANN, S. A., MARIONI, J. C. & HEISLER, M. G. 2013. Accounting for technical noise in single-cell RNA-seq experiments. *Nat Methods,* 10**,** 1093-5.

BUETTNER, F., NATARAJAN, K. N., CASALE, F. P., PROSERPIO, V., SCIALDONE, A., THEIS, F. J., TEICHMANN, S. A., MARIONI, J. C. & STEGLE, O. 2015. Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat Biotechnol,* 33**,** 155-60.

CHEN, X., XU, H., YUAN, P., FANG, F., HUSS, M., VEGA, V. B., WONG, E., ORLOV, Y. L., ZHANG, W., JIANG, J., LOH, Y. H., YEO, H. C., YEO, Z. X., NARANG, V., GOVINDARAJAN, K. R., LEONG, B., SHAHAB, A., RUAN, Y., BOURQUE, G., SUNG, W. K., CLARKE, N. D., WEI, C. L. & NG, H. H. 2008. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell,* 133**,** 1106-17.

CHENG, A. W., WANG, H., YANG, H., SHI, L., KATZ, Y., THEUNISSEN, T. W., RANGARAJAN, S., SHIVALILA, C. S., DADON, D. B. & JAENISCH, R. 2013. Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res,* 23**,** 1163-71.

CONSORTIUM, E. P. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature,* 489**,** 57-74.

DOBIN, A., DAVIS, C. A., SCHLESINGER, F., DRENKOW, J., ZALESKI, C., JHA, S., BATUT, P., CHAISSON, M. & GINGERAS, T. R. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics,* 29**,** 15-21.

HEGER, A., WEBBER, C., GOODSON, M., PONTING, C. P. & LUNTER, G. 2013. GAT: a simulation framework for testing the association of genomic intervals. *Bioinformatics,* 29**,** 2046-8.

HEIGWER, F., KERR, G. & BOUTROS, M. 2014. E-CRISP: fast CRISPR target site identification. *Nat Methods,* 11**,** 122-3.

JAAKKOLA, M. K., SEYEDNASROLLAH, F., MEHMOOD, A. & ELO, L. L. 2017. Comparison of methods to detect differentially expressed genes between single-cell populations. *Brief Bioinform,* 18**,** 735-743.

KRYUCHKOVA-MOSTACCI, N. & ROBINSON-RECHAVI, M. 2017. A benchmark of gene expression tissue-specificity metrics. *Brief Bioinform,* 18**,** 205-214.

LE, S., JOSSE, J. & HUSSON, F. 2008. FactoMineR: An R package for multivariate analysis. *Journal of Statistical Software,* 25**,** 1-18.

LI, B. & DEWEY, C. N. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics,* 12**,** 323.

LOVE, M. I., HUBER, W. & ANDERS, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol,* 15**,** 550.

LUN, A. T., MCCARTHY, D. J. & MARIONI, J. C. 2016. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res,* 5**,** 2122.

OCHIAI, H., SUGAWARA, T. & YAMAMOTO, T. 2015. Simultaneous live imaging of the transcription and nuclear position of specific genes. *Nucleic Acids Res,* 43**,** e127.

REIMAND, J., ARAK, T., ADLER, P., KOLBERG, L., REISBERG, S., PETERSON, H. & VILO, J. 2016. g:Profiler-a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res,* 44**,** W83-9.

SANTOS, A., WERNERSSON, R. & JENSEN, L. J. 2015. Cyclebase 3.0: a multi-organism database on cell-cycle regulation and phenotypes. *Nucleic Acids Res,* 43**,** D1140-4.

SONESON, C., LOVE, M. I. & ROBINSON, M. D. 2015. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res,* 4**,** 1521.

SUN, N., YU, X., LI, F., LIU, D., SUO, S., CHEN, W., CHEN, S., SONG, L., GREEN, C. D., MCDERMOTT, J., SHEN, Q., JING, N. & HAN, J. J. 2017. Inference of differentiation time for single cell transcriptomes using cell population reference data. *Nat Commun,* 8**,** 1856.

TAN, J. Y., SIREY, T., HONTI, F., GRAHAM, B., PIOVESAN, A., MERKENSCHLAGER, M., WEBBER, C., PONTING, C. P. & MARQUES, A. C. 2015. Extensive microRNA-mediated crosstalk between lncRNAs and mRNAs in mouse embryonic stem cells. *Genome Res,* 25**,** 655-66.

TAN, J. Y., VANCE, K. W., VARELA, M. A., SIREY, T., WATSON, L. M., CURTIS, H. J., MARINELLO, M., ALVES, S., STEINKRAUS, B., COOPER, S., NESTEROVA, T., BROCKDORFF, N., FULGA, T., BRICE, A., SITTLER, A., OLIVER, P. L., WOOD, M. J., PONTING, C. P. & MARQUES, A. C. 2014. Cross-talking noncoding RNAs contribute to cell-specific neurodegeneration in SCA7. *Nat Struct Mol Biol,* 21**,** 955-961.

XU, H., BAROUKH, C., DANNENFELSER, R., CHEN, E. Y., TAN, C. M., KOU, Y., KIM, Y. E., LEMISCHKA, I. R. & MA'AYAN, A. 2013. ESCAPE: database for integrating high-content published data collected from human and mouse embryonic stem cells. *Database (Oxford),* 2013**,** bat045.