



Published in final edited form as:

*Mol Cell*. 2019 August 08; 75(3): 523–537.e10. doi:10.1016/j.molcel.2019.05.028.

## LncRNA-induced spread of Polycomb controlled by genome architecture, RNA abundance, and CpG island DNA

Megan D. Schertzer<sup>1,2</sup>, Kean C.A. Bracer<sup>1,3</sup>, Joshua Starmer<sup>4</sup>, Rachel E. Cherney<sup>1,2</sup>, David M. Lee<sup>1,2</sup>, Gabriela Salazar<sup>5</sup>, Megan Justice<sup>2,6</sup>, Steven R. Bischoff<sup>7</sup>, Dale O. Cowley<sup>7</sup>, Pablo Ariel<sup>8</sup>, Mark J. Zylka<sup>5</sup>, Jill M. Downen<sup>2,6,9,10,11</sup>, Terry Magnuson<sup>4,11</sup>, J. Mauro Calabrese<sup>1,11,12,\*</sup>

<sup>1</sup>Department of Pharmacology, University of North Carolina, Chapel Hill, NC, 27599

<sup>2</sup>Curriculum in Genetics and Molecular Biology, University of North Carolina, Chapel Hill, NC, 27599

<sup>3</sup>Curriculum in Mechanistic, Interdisciplinary Studies of Biological Systems, University of North Carolina, Chapel Hill, NC, 27599

<sup>4</sup>Department of Genetics, University of North Carolina, Chapel Hill, NC, 27599

<sup>5</sup>Department of Cell Biology and Physiology and UNC Neuroscience Center, University of North Carolina, Chapel Hill, NC, 27599

<sup>6</sup>Integrative Program for Biological and Genome Sciences, University of North Carolina, Chapel Hill, NC, 27599

<sup>7</sup>Animal Models Core, University of North Carolina, Chapel Hill, NC, 27599

<sup>8</sup>Microscopy Services Laboratory and Department of Pathology and Laboratory Medicine, University of North Carolina, Chapel Hill, NC, 27599

<sup>9</sup>Department of Biochemistry and Biophysics, University of North Carolina, Chapel Hill, NC, 27599

<sup>10</sup>Department of Biology, University of North Carolina, Chapel Hill, NC, 27599

<sup>11</sup>Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC, 27599

<sup>12</sup>Lead contact.

\*Correspondence: jmcablabr@med.unc.edu.

#### Author Contributions

M.D.S. and J.M.C. conceived the study, M.D.S., K.C.A.B., R.E.C., G.S., M.J., D.M.L., S.R.B. and J.M.C. performed the experiments, M.D.S., J.S., P.A., and J.M.C. performed the analysis, D.O.C., J.M.D., M.J.Z., and T.R.M. provided reagents and advice, and M.D.S. and J.M.C. wrote the paper.

#### Declaration of interests

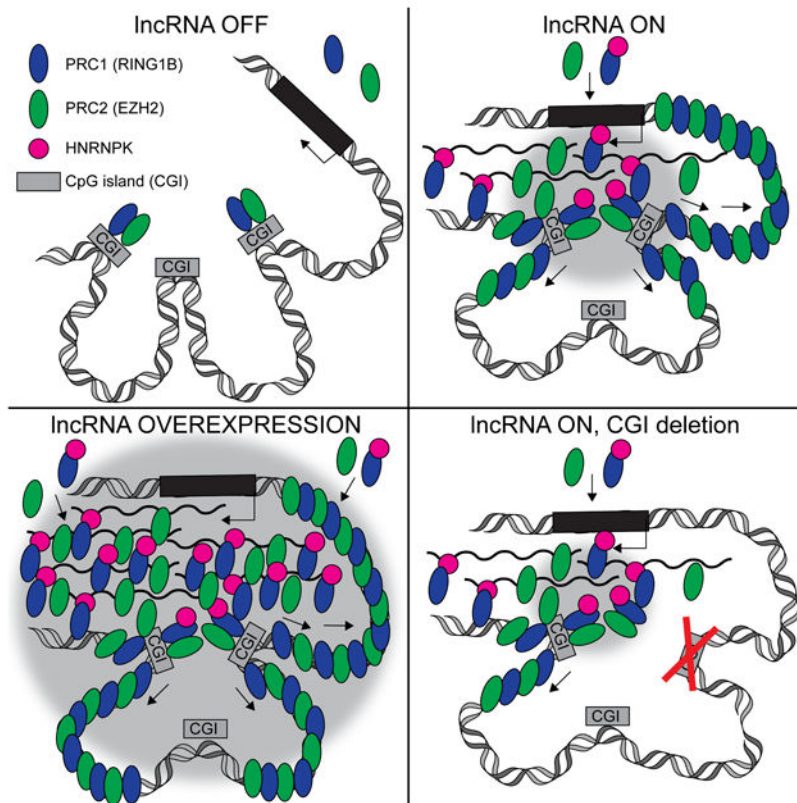
D.O.C. is employed by, has equity ownership in, and serves on the board of directors of TransViragen, the company contracted by UNC-Chapel Hill to manage its Animal Models Core Facility. The authors declare no other competing interests.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## Summary

Long noncoding RNAs (lncRNAs) cause Polycomb Repressive Complexes (PRCs) to spread over broad regions of the mammalian genome. We report that in mouse trophoblast stem cells, the *Airn* and *Kcnq1ot1* lncRNAs induce PRC-dependent chromatin modifications over multi-megabase domains. Throughout the *Airn*-targeted domain, extent of PRC-dependent modification correlated with intra-nuclear distance to the *Airn* locus, pre-existing genome architecture, and the abundance of *Airn* itself. Specific CpG islands (CGIs) displayed characteristics indicating that they nucleate the spread of PRCs upon exposure to *Airn*. Chromatin environments surrounding *Xist*, *Airn*, and *Kcnq1ot1* suggest common mechanisms of PRC engagement and spreading. Our data indicate that lncRNA potency can be tightly linked to lncRNA abundance, and that within lncRNA-targeted domains, PRCs are recruited to CGIs via lncRNA-independent mechanisms. We propose that CGIs that autonomously recruit PRCs interact with lncRNAs and their associated proteins through 3-dimensional space to nucleate the spread of PRCs in lncRNA-targeted domains.

## Graphical Abstract



## eTOC

Schertzer et al. studied relationships between long noncoding RNAs (lncRNAs) and Polycomb Repressive Complexes (PRCs) in mouse trophoblast stem cells. They found that genome architecture, lncRNA abundance, and CpG island DNA each play important roles in dictating the intensity of PRC-induced chromatin modifications within lncRNA target domains.

## Introduction

Long noncoding RNAs (lncRNAs) play essential roles in development by directing Polycomb Repressive Complexes (PRCs) to broad genomic regions. In the most extreme example, expression of the lncRNA *Xist* causes PRCs to modify chromatin over the entire inactive X chromosome. Several other lncRNAs also cause PRCs to engage with smaller genomic regions. In the mouse placenta, the lncRNA *Aim* silences 10 genes in 10 megabases (Mbs) on chr17 and the lncRNA *Kcnq1ot1* silences seven genes in ~800 kilobases (kb) on chr7. Like *Xist*, these lncRNAs act *in cis*, meaning that they only target regions located on the same chromosome from which they were transcribed (Andergassen et al., 2017; Lee and Bartolomei, 2013).

The two major PRCs, PRC1 and PRC2, catalyze the mono-ubiquitylation of lysine 119 on histone H2A (H2AK119ub) and the trimethylation of lysine 27 on histone H3 (H3K27me3), respectively. These modifications repress gene expression through parallel mechanisms that compact chromatin and antagonize transcriptional activators (Schwartz and Pirrotta, 2013; Simon and Kingston, 2013). The two PRCs are also interdependent throughout the genome. Regions of chromatin modified by one PRC are usually modified by the other, PRC1 can be recruited by PRC2 and vice versa, and loss of either PRC destabilizes most, or all, PRC1- and PRC2-silenced regions (Blackledge et al., 2014; Kalb et al., 2014; Schwartz and Pirrotta, 2013; Simon and Kingston, 2013). Indeed, *Xist*, *Aim*, and *Kcnq1ot1* require both PRC1 and 2 for full repressive activity (Almeida et al., 2017; Kalantry et al., 2006; Terranova et al., 2008).

Nevertheless, the mechanisms through which *Xist* and related lncRNAs induce the spread of PRCs over chromatin remain unclear. While local features of the genome correlate with levels of *Xist*-induced, PRC-dependent modification, the mechanisms that underlie the correlations are unclear (Calabrese et al., 2012; Cotton et al., 2014; Kelsey et al., 2015; Loda et al., 2017; Pinter et al., 2012). The RNA-binding protein HNRNPK bridges PRC1 with *Xist*, and this interaction is required to spread both PRC1 and 2 over the inactive X (Almeida et al., 2017; Pintacuda et al., 2017). However, whether *Xist* directly travels with PRCs, or if *Xist* causes PRCs to spread via secondary interactions, remains debated (Cerase et al., 2014; Smeets et al., 2014; Sunwoo et al., 2015). Moreover, while *Aim* and *Kcnq1ot1* also direct PRCs to chromatin (Pandey et al., 2008; Regha et al., 2007), it is unclear if they do so through mechanisms shared with *Xist*. Indeed, transcription over the *Aim* and *Kcnq1ot1* genes, and not necessarily the lncRNAs themselves, plays a role in local silencing by both lncRNAs; whether the lncRNA products are required for distal silencing or spread of PRCs remains unclear (Andergassen et al., 2017; Korostowski et al., 2012; Latos et al., 2012). Lastly, the mechanisms that give rise to specific patterns of PRC-dependent modifications within the *Aim* and *Kcnq1ot1* target domains remain unclear (Andergassen et al., 2017; Pandey et al., 2008; Regha et al., 2007).

Considering these unknowns, we set out to study molecular phenotypes associated with *Aim* and *Kcnq1ot1* in contexts that allowed direct comparison to *Xist*, using female, F1-hybrid, mouse trophoblast stem cells (TSCs) that naturally express all three lncRNAs (Calabrese et

al., 2015). We found that genome architecture, lncRNA abundance, and CpG island (CGI) DNA all play roles in coordinating the spread of PRCs induced by *Xist*, *Airn*, and *Kcnq1ot1*.

## Results

### Megabase-sized domains of H3K27me3 require continued expression of *Airn* and *Kcnq1ot1*

In TSCs, *Xist*, *Airn*, and *Kcnq1ot1* are monoallelically expressed from paternally inherited chromosomes (Lee and Bartolomei, 2013). Their monoallelism, coupled with their cis-acting nature, necessitates the use of F1-hybrid cells to study their effects on chromatin (Figure S1A). To this end, we previously derived F1-hybrid TSCs from reciprocal crosses between C57BL/6J and CAST/EiJ mice, and demonstrated that the TSCs can be used to study all three lncRNAs (Calabrese et al., 2015; Calabrese et al., 2012).

While *Xist* is known to recruit PRCs to gene-dense regions of the X (Calabrese et al., 2012; Chadwick, 2007; Marks et al., 2009), PRC targeting had not been examined around *Airn* and *Kcnq1ot1* using sequencing-based approaches. Thus, we used ChIP-Seq to measure the density of H2AK119ub and H3K27me3, covalent modifications catalyzed by PRC1 and PRC2, respectively, in reciprocal F1-hybrid TSC lines. As expected, the two modifications were highly correlated throughout the genome ( $r = 0.746$ ; Figure S1B). We used H3K27me3 as a surrogate for H2AK119ub in most of our work below owing to its higher signal in ChIP-Seq.

We observed that H3K27me3 density was high around known *Airn* and *Kcnq1ot1* target genes, and dropped sharply near non-targets, supporting previous views that chromatin-associated factors work in concert with lncRNAs to control the spread of PRCs in lncRNA domains (Figure 1A, B, upper panels; (Calabrese et al., 2012; Cotton et al., 2014; Kelsey et al., 2015; Loda et al., 2017; Pinter et al., 2012)). Unexpectedly, both *Airn* and *Kcnq1ot1* were centered in H3K27me3-enriched regions that extended for megabases beyond their originally defined target genes (Figure 1A, B, lower panels area between dotted lines). We used hiddenDomains to call peaks of H3K27me3 in TSCs (Starmer and Magnuson, 2016), and then used allelic data within those peaks to identify sites of significant parent-of-origin bias (Table S1). Strikingly, 83% of all paternally-biased autosomal peaks of H3K27me3 in TSCs were found in the regions surrounding *Airn* and *Kcnq1ot1* (Figure S1C; Table S1), where patterns of H2AK119ub mirrored those of H3K27me3 (Figure S1D, E; Table S1). Of the 118 and 91 paternally-biased H3K27me3 peaks surrounding *Airn* and *Kcnq1ot1*, 43 and 32 overlapped genes (starting 2kb upstream of transcription starts and extending to transcription ends) and 76 and 3 peaks were intergenic, respectively. H3K27me3 signal in these peaks rivaled or surpassed signal in H3K27me3 peaks on the X (Figure S1F). These data suggest that in TSCs, *Airn* and *Kcnq1ot1* each direct PRCs to regions that span megabases, potentially in a manner analogous to *Xist*.

To determine whether PRC-induced chromatin modifications around *Airn* and *Kcnq1ot1* were lncRNA-dependent, we truncated each lncRNA in a way that phenocopies lncRNA knockout in embryos (Figure S2A; (Mancini-Dinardo et al., 2006; Sleutels et al., 2002)). We derived clonal truncation lines for each lncRNA, and in two of them we profiled H3K27me3

via ChIP-Seq (Figure S2B-D). We then compared paternal H3K27me3 between lncRNA-truncated and wild-type TSCs. Upon truncation, we observed loss of H3K27me3 in the regions surrounding both lncRNAs (Figure 1C, D; Table S1). Consistent with activity *in cis*, *Airn*-truncated TSCs had wild-type H3K27me3 levels around *Kcnq1ot1*, and vice versa (Figure S2D). These data show that in TSCs, *Airn* and *Kcnq1ot1* direct PRCs to 13 and 2.3 Mb regions, respectively.

We performed RNA-Seq in three truncation clones to determine whether lncRNA loss coincided with gene derepression. Of 61 genes in the *Airn* domain whose allelic expression met our threshold for consideration (avg. of 10 allelic reads per dataset), 14 were derepressed upon *Airn* truncation, 6 of which are known *Airn* targets in the placenta (Figure 1E; Table S2; (Andergassen et al., 2017)). In the *Kcnq1ot1* domain, 27 genes met our threshold, and 6 were derepressed upon *Kcnq1ot1* truncation; 4 were known *Kcnq1ot1* targets (Figure 1F; Table S2). Non-impacted genes also trended towards higher expression upon lncRNA truncation (Figure 1E, F). Thus, lncRNA truncation leads to derepression of genes *in cis*. Locations of derepressed and non-impacted genes relative to H3K27me3 levels are shown in Figure S2E. 62 and 16 genes in the *Airn* and *Kcnq1ot1* domains, respectively, did not meet our threshold for allelic analysis.

### Spread of H3K27me3 in the *Airn* domain is influenced by pre-existing genomic architecture and additional features on the paternal allele

On the X, H3K27me3 levels correlate with 3D contacts in place prior to induction of *Xist*, and with specific regulatory elements (Calabrese et al., 2012; Cotton et al., 2014; Engreitz et al., 2013; Kelsey et al., 2015; Loda et al., 2017; Pinter et al., 2012). Whether similar trends are true in regions silenced by *Airn* and *Kcnq1ot1* is unknown.

Owing to its large size, we first focused on the region targeted by *Airn*. Based on the studies above, and data demonstrating DNA loops restrict signals that control gene expression (Downen et al., 2014; Rao et al., 2014), we hypothesized that variation in H3K27me3 around *Airn* was due to three factors: first, a pre-existing genomic architecture that might make certain regions more susceptible to targeting due to their ability to contact *Airn*; second, within susceptible regions, a greater affinity of *Airn* for specific sites over others; and third, DNA loops that restrict PRC spread around sites of *Airn* contact.

We tested the first two hypotheses using fluorescence in situ hybridization (FISH). We designed FISH probes to 9 regions surrounding *Airn*, each harboring different extents of *Airn*-dependent H3K27me3, including a probe in a region whose H3K27me3 was unaffected by *Airn* (“Neg control”, purple bar; Figure 2A). We used RNA/DNA FISH to measure spatial distance between each region of interest and the *Airn* locus, distinguishing paternal from maternal alleles by co-localization of *Airn* RNA and DNA FISH probes (Figure 2B).

For 8 of 9 loci, average spatial distance to *Airn* was less on the paternal allele (Figure 2C). We also examined two loci in *Airn*-truncation TSCs, *Park2* and *Arid1b*, and found that differences between maternal and paternal distributions were reduced (Figure 2D). The average difference in distance between the paternal and maternal alleles, i.e. the extent of genomic compaction at a locus, was a strong predictor of H3K27me3 (Figure 2E,



$R^2=0.738$ ). These data indicate that compaction in the *Airn* region depends on continued expression of *Airn* and correlates with underlying levels of H3K27me3. Distance in base pairs to the *Airn* locus was also a strong predictor of paternal H3K27me3 (Figure 2F panel (i);  $R^2 = 0.399$ ).

Next, we tested the hypothesis that, in addition to distance in base pairs from *Airn*, H3K27me3 levels were influenced by chromosomal conformations in place prior to the onset of *Airn* expression, which rendered certain regions more likely than others to come into proximity to *Airn*. In our experiment, the maternal allele served as a surrogate to approximate the conformation that the paternal allele would be in if *Airn* were not expressed. If extent of *Airn*-induced H3K27me3 was influenced by pre-existing chromosomal conformations in place prior to the onset of *Airn* expression, we would expect distance to *Airn* on the maternal allele, which provides a readout for those conformations, to be a better predictor of paternal H3K27me3 than the distance to the *Airn* locus in base pairs. Consistent with this notion, our expectation held true (Figure 2F panel (i) vs (ii);  $p=0.003$ ; empirically derived by bootstrapping).

Intriguingly, distance to *Airn* on the paternal allele was a better predictor of paternal H3K27me3 than distance on the maternal allele (Figure 2F panel (ii) vs (iii);  $p=0.037$ ; empirically derived by bootstrapping). This increase in predictive power supports the view that, within broader domains capable of contacting *Airn*, additional factors on the paternal allele cause the lncRNA to associate with certain sites more than others. Thus, in addition to distance in base pairs from the lncRNA expressing locus (Figure 2Fi) and pre-existing genomic architecture (Figure 2Fii), local features of chromatin (Figure 2Fiii) likely contribute to the control of PRCs by *Airn*.

### **Intensity of H3K27me3 in lncRNA target domains correlates with TADs, DNA loops, and SMC1 and CTCF binding**

DNA loops and topologically associated domains (TADs) divide the genome into compartments with distinct chromatin and gene expression patterns that may influence targeting by *Xist* (Darrow et al., 2016; Dixon et al., 2012; Downen et al., 2014; Engreitz et al., 2013; Giorgetti et al., 2016; Rao et al., 2014). To determine whether DNA loops and TADs might also influence *Airn* and *Kcnq1ot1*, we examined Hi-C, ChIA-PET, and ChIP-Seq data in mouse embryonic stem cells (ESCs). TAD boundaries are often conserved between cell types (Dixon et al., 2012), and we reasoned that, as a first pass, inferring the location of DNA loops and TADs in TSCs using ESC data was a viable approach. Independently, we profiled, via ChIP-Seq, Cohesin (SMC1) and CTCF binding in F1-hybrid TSCs. Genome-wide and in the *Airn* and *Kcnq1ot1* domains, SMC1 and CTCF peak locations were concordant between ESCs and TSCs (Figure S3A). Moreover, SMC1 and CTCF binding in TSCs was detected at DNA loops anchored by SMC1 and CTCF in ESCs (Figure S3B, C), consistent with the notion that the two cell types harbor many of the same DNA loops.

Throughout the *Airn* and *Kcnq1ot1* target domains, inflections in TSC H3K27me3 density coincided with ESC TAD boundaries and SMC1-bound DNA loops (Figure 3A, B), supporting the notions that DNA loops influence spread of H3K27me3 in lncRNA target domains, and that *Airn* and *Kcnq1ot1* direct PRCs over multiple TADs. Moreover, SMC1

and CTCF showed reduced binding in lncRNA-silenced domains on paternal relative to maternal alleles. This reduction was stronger for SMC1 than CTCF and correlated with the range over which *Xist*, *Airn*, and *Kcnq1ot1* direct PRCs to chromatin (Figure 3C, D). Thus, the more potent the lncRNA, the more likely that its targeted regions lack DNA loops anchored by SMC1 and CTCF.

### **LncRNA repressive potency correlates with abundance, stability, and underlying features of the genome**

Our data show that in TSCs, *Airn* and *Kcnq1ot1* direct PRCs to megabase-sized regions in which H3K27me3 levels are influenced by genome architecture and underlying features of chromatin. We also found that *Airn* and *Kcnq1ot1* each control PRCs to different extents, and both to lesser extents than *Xist*.

We examined if differences in lncRNA abundance could account for differences in repressive potency, which we define here as the ability of a lncRNA to induce PRC-dependent chromatin modifications. We estimated copy number of *Xist*, *Airn*, and *Kcnq1ot1* in TSCs using RNA-Seq and found that the lncRNAs are expressed at an average of 232, 8.7, and 7.6 copies per TSC, respectively (Figure 4A, S4A). Using Actinomycin D, we found the half-life of *Xist* was ~6.2 hours, while the half-lives of *Airn* and *Kcnq1ot1* were each ~1.7 hours (Figure 4B). Thus, in TSCs, abundance and stability correlate with lncRNA potency, but they do not account for differences in potency between *Airn* and *Kcnq1ot1*.

Based on these data, we hypothesized that changes in *Airn* abundance would affect its potency. We created TSCs in which we could recruit a transcriptional activator (dCas9-VP160) or a repressor (dCas9-KRAB) to the endogenous *Airn* promoter in a doxycycline-inducible manner (Schertzer et al., 2018). RNA-Seq showed *Airn* increased to ~27.2 copies per cell by recruiting the activator and decreased to ~0.6 copies per cell by recruiting the repressor. Almost all of the boost in *Airn* expression occurred on the paternal allele, presumably because DNA methylation prevented the activator from accessing the maternal allele (Figure S4B, C). Overexpression increased the size of *Airn* foci detected by FISH but did not change *Airn* subcellular distribution (Figure S4D-F).

We observed a striking correlation between RNA abundance and repressive potency in the *Airn* domain (Figure 4C-E). Expression-induced changes in H3K27me3 were variable throughout the domain and were inversely proportional to changes in gene expression (Figure 4C-E; Figure S5A). The largest changes in H3K27me3 occurred on the centromeric side of *Airn*, centered around three regions that appeared to be sites from which H3K27me3 spread outwards, owing to their high levels of H3K27me3 in *Airn*-overexpression cells that dropped rapidly with increasing distance on either or both sides (Fig 4E, arrows). We draw two major conclusions: (1) that RNA abundance can affect lncRNA control over PRCs, but it is not the only factor to do so, and (2) that genomic features – likely a mix of 3D architecture, chromatin-bound factors, and the sequence of DNA itself – play important roles in the lncRNA-induced spread of PRCs on chromatin.

### CGIs bind PRCs autonomously and can nucleate spread of H3K27me3 by *Airn*

We examined regions in the *Airn* domain whose patterns of H3K27me3 suggested the presence of H3K27me3 nucleation sites (Figure 4E, arrows). Strikingly, all 6 regions coincided with CGIs (Figure 4E, blue ticks; S5B), which are known to recruit PRCs in mammals (Farcas et al., 2012; Li et al., 2017; Lynch et al., 2012; Mendenhall et al., 2010; Oksuz et al., 2018; Riising et al., 2014; Woo et al., 2010). The 8 CGIs were all found at lowly to moderately expressed genes and had the highest levels of H3K27me3 in *Airn*-overexpressing TSCs relative to all other CGIs in the *Airn* domain (Table S3; Figure S5B). Using non-allelic data as a search feature, 7 of the 8 CGIs co-localized with MACS-defined peaks of RING1B (catalytic subunit of PRC1), and 6 also co-localized with peaks of EZH2 (catalytic subunit of PRC2). In contrast, only 27 of the 83 remaining CGIs in the *Airn* domain co-localized with RING1B peaks, and none with EZH2 peaks (Table S3; Figure S5B). Moreover, consistent with trends elsewhere (Figure S3D, E), 5 of the 8 CGIs in question co-localized with SMC1 peaks and none with CTCF (Table S3; Figure S5B). Assuming that *Airn* targets PRCs *de novo* to chromatin, we hypothesized that the 8 CGIs should harbor higher RING1B and EZH2 signal relative to surrounding regions, and that CGIs on the *Airn*-targeted paternal allele would harbor more signal than CGIs on the untargeted maternal allele.

To test these hypotheses, we profiled RING1B in wild-type, *Airn*-overexpression, *Airn*-truncation, and *Kcnq1ot1*-truncation TSCs, and EZH2 in wild-type TSCs. We observed enrichment of RING1B and EZH2 at CGIs relative to surrounding DNA; however, near-equal levels of RING1B and EZH2 were found at CGIs on maternal and paternal alleles (Figure 5A, red lines overlap with blue lines). Outside of CGIs, we observed broad enrichment of RING1B and EZH2 on the paternal allele. This enrichment was responsive to *Airn* expression and mirrored enrichment of H3K27me3 and H2AK119-ub (Figures 5B, C vs. Figures 1A, S1D). Analogous patterns of RING1B and EZH2 surrounded *Kcnq1ot1* (Figure S6A-C).

To test our hypothesis that specific CGIs nucleate the spread of PRCs in the *Airn* target domain, we used CRISPR to delete the CGI at *Slc22a3*, which is located ~234 kb upstream of *Airn* yet harbors some of the highest density of H3K27me3 in the *Airn* domain and shows evidence of RING1B binding on both alleles (Figure S6D; Table S3). As a control, we deleted a size-matched region ~1,383 kb upstream of *Airn* that occurs within an H3K27me3 peak in the *Park2* intron but does not overlap a CGI (Figure S6E). We profiled H3K27me3 in two independent clones of each deletion. Strikingly, deletion of the *Slc22a3* CGI, but not the size-matched control, caused a ~4.6 Mb reduction in H3K27me3 in the *Airn* domain (Figures 5D, 5E, S6F). This loss could not be ascribed to reduced *Airn* RNA abundance upon *Slc22a3* CGI deletion (Figure S6G). Thus, specific CGIs can play outsized roles in nucleating the spread of H3K27me3 in the *Airn* target domain.

### *Xist*-induced H3K27me3 density is highest around CGIs that bind PRCs autonomously

The inactive X displayed patterns of H3K27me3 similar to those seen upon over-expression of *Airn*, where H3K27me3 levels culminated at single points, then decreased in intensity until inflecting or crossing into a H3K27me3-depleted region (Figure S7A vs. 4E;



(Calabrese et al., 2012)). In light of these similarities, we examined levels of H3K27me3, RING1B, and EZH2 at X-linked CGIs in TSCs. Analogous to the *Airn* region, the highest levels of H3K27me3 on the inactive X were found at CGIs over which we could detect peaks of RING1B and EZH2; the more PRC binding that could be detected, the greater the levels of surrounding H3K27me3, and the greater the difference in H3K27me3 between the inactive and active X (Figure 6A; Table S4). However, at CGIs, the binding of both RING1B and EZH2 was substantially higher on the active X relative to the inactive X, despite H3K27me3 levels showing the opposite enrichment in the majority of cases (Figure 6A). Most CGIs co-bound by RING1B and EZH2 on the active X coincided with lowly expressed genes, consistent with their PRC-mediated repression (Table S4). Outside of CGIs, RING1B and EZH2 were broadly enriched on the inactive X and their enrichment in H3K27me3 peaks was strongly co-correlated (Figure 6B;  $r=0.98$ ). Thus, in TSCs, regions on the inactive X that harbor the most H3K27me3 coincide with CGIs that bind the highest levels of RING1B and EZH2. Unexpectedly, at these CGIs, more RING1B and EZH2 signal is found on the active X than on the inactive X.

We next examined *Xist*-induced spread of PRCs in ESCs, a commonly used cell-based model in *Xist* research. As part of a separate study, we inserted a doxycycline-inducible *Xist* gene into the *Rosa26* locus on chr6 in mouse ESCs (D.M.L. and J.M.C., in press). We profiled RING1B and H3K27me3 in these ESCs before, 12 hours after, and 72 hours after induction of *Xist* (Figure S7B, C). Similar to what we observed in TSCs, the highest levels of *Xist*-induced H3K27me3 were found around CGIs on chr6 that bound PRCs prior to induction of *Xist* (Figure 6C). However, consistent with recent studies performed in ESC models (Fursova et al., 2019; Zyllicz et al., 2019), there was little relative difference in the change in *Xist*-induced H3K27me3 levels at PRC-bound versus unbound CGIs ( 1.08 vs 0.87, respectively; upper right corner of lower two panels in Figure 6C), whereas in TSCs, the difference was dramatic ( 1.29 vs 0.26, respectively; upper right corner of first and last panels in lowest row in Figure 6A). Moreover, unlike what was observed on the TSC inactive X, we observed no *Xist*-dependent depletion of PRCs at CGIs on ESC chr6 (Figure 6A vs 6C), and overall levels of *Xist*-induced H3K27me3 were lower on ESC chr6 than they were on the TSC inactive X, or even in the TSC *Airn* domain (Figure 6D vs S7A, 4E). Similarly, H3K27me3 levels around *Kcnq1ot1* were lower in ESCs than in TSCs and were lower in a third cell type, cortical neurons (Figure S7D, E). Lastly, we note that relative to the TSC X, there were ~10-fold more RING1B/EZH2-bound CGIs on ESC chr6 (Figure 6A vs 6C). These data highlight potential differences in interactions between PRCs and CGIs in ESCs versus TSCs, and suggest, along with data from (Andergassen et al., 2017; Lewis et al., 2006; Umlauf et al., 2004), that relative to other mouse cell types, TSCs are primed to respond to PRC-controlling lncRNAs.

### ***Xist*, *Airn*, and *Kcnq1ot1* require HNRNPK to spread H3K27me3**

*Xist* requires the RNA-binding protein HNRNPK to induce PRC spread (Pintacuda et al., 2017). Considering the similarities between *Xist*, *Airn*, and *Kcnq1ot1*, we examined whether *Airn* and *Kcnq1ot1* also required HNRNPK to induce PRC spread.

We first examined whether *Airn* and *Kcnq1ot1* showed evidence of HNRNPK association. We used a formaldehyde-based RNA immunoprecipitation (RNA IP) protocol followed by RNA-Seq in *Xist*-expressing SM33 ESCs (which also express *Kcnq1ot1* and low levels of *Airn*) and in TSCs (Raab et al., 2019). IP revealed strong enrichment of HNRNPK over all three lncRNAs relative to IgG in both cell types, and peaks of HNRNPK enrichment were identified by MACS (Figure 7A, Table S5). In contrast, IP of CTCF, a protein that binds RNA with high affinity in a sequence non-specific manner (Kung et al., 2015), yielded little enrichment over the lncRNAs (Figure 7A, Table S5). Moreover, in TSCs, relative to the set of 23366 UCSC Known Genes, *Xist*, *Airn*, and *Kcnq1ot1* harbored the 4<sup>th</sup>, 6<sup>th</sup>, and 7<sup>th</sup> most HNRNPK signal, respectively (Table S5, 'ts.hk.norm' column). In contrast, the lncRNAs harbored the 296<sup>th</sup>, 3048<sup>th</sup>, and 4244<sup>th</sup> most CTCF signal, and in terms of length-normalized expression, they were the 171<sup>st</sup>, 6550<sup>th</sup>, and 6798<sup>th</sup> most highly expressed transcripts, respectively (Table S5, 'ts.ctcf.norm' and 'TSC input' columns). HNRNPK enrichment was also observed over Repeat B and C in *Xist*, which are known HNRNPK-interacting regions (Pintacuda et al., 2017). Taken together, these data show that *Xist*, *Airn*, and *Kcnq1ot1* associate with HNRNPK, likely at levels above most other genes.

Next, we used CRISPR to knock-down HNRNPK in TSCs and profiled H3K27me3 and RNA expression after four days of Cas9 induction (Schertzer et al., 2018). By two days of Cas9 induction, HNRNPK knockdown caused TSC death (not shown). Nevertheless, four days after Cas9 induction, relative to non-targeting sgRNA controls, HNRNPK levels were substantially reduced in surviving TSCs (Figure S7F, G). Loss of HNRNPK coincided with a significant loss of H3K27me3 in *Xist*, *Airn*, and *Kcnq1ot1* target domains (Figure 7B), but did not coincide with changes in gene silencing, presumably because TSCs that lost the most HNRNPK died (Figure S7H). Moreover, upon HNRNPK knockdown, H3K27me3 was reduced around CGIs that bound PRCs even in the absence of lncRNA expression (Figure 7C). Thus, in lncRNA domains, H3K27me3 levels at PRC-bound CGIs are more dependent on HNRNPK than H3K27me3 levels at PRC-unbound CGIs. Also, similar to *Xist*, *Airn* and *Kcnq1ot1* rely on HNRNPK to spread PRCs.

## Discussion

Via orthogonal assays, we compared the genomic properties in the domains targeted by *Airn* and *Kcnq1ot1* to those on the *Xist*-targeted X chromosome. We gained several insights into mechanism which we enumerate below. While *Xist*, *Airn*, and *Kcnq1ot1* are all monoallelically expressed due to X-inactivation and imprinting (Lee and Bartolomei, 2013), we posit that these transcriptional regulatory mechanisms are unlikely to impact function of the lncRNA after its production. Thus, principles defined by our study are likely to be relevant to other lncRNAs, as well.

We found that variation in intensity of H3K27me3 in *Airn* and *Kcnq1ot1* domains mirrored variation on the inactive X, where H3K27me3-enriched regions are separated by regions that partially or fully escape *Xist*-induced silencing (Calabrese et al., 2012; Chadwick, 2007; Marks et al., 2009; Pinter et al., 2012). Within the *Airn* domain, variation in H3K27me3 could be partly explained by large-scale, pre-existing conformations of chromatin that rendered strongly silenced regions more likely to come in proximity to the *Airn* locus than

weakly-silenced ones. Our results support the view that long-distance contacts in place prior to the onset of lncRNA expression play roles in dictating the intensity of PRC-induced modification in lncRNA target domains (Engreitz et al., 2013; Kelsey et al., 2015; Marks et al., 2015).

However, more than *Xist*, *Airn* and *Kcnq1ot1* appeared to be influenced by genome architecture. Inflections in H3K27me3 density in the *Airn* and *Kcnq1ot1* domains tended to colocalize with DNA loops and TADs (Dixon et al., 2012; Downen et al., 2014), whereas such structures are largely absent on the inactive X (Rao et al., 2014). Accordingly, we found that the more potent the lncRNA, the more likely it was to disrupt binding of SMC1 and, to a lesser extent, CTCF. Thus, while DNA loops may influence the initial spread of H3K27me3 on the inactive X, they are more likely to be overridden, ultimately, by the repressive effect of *Xist*.

We observed a strong correlation between expression, stability, and potency of *Xist*, *Airn*, and *Kcnq1ot1*. In TSCs, *Xist*, the most potent of the three, was expressed most highly, at ~232 molecules per cell, and had the longest half-life, at ~6.2 hours. *Airn* and *Kcnq1ot1* were expressed at ~8 molecules per cell and had ~1.7-hour half-lives. Increasing or decreasing expression of *Airn* changed its potency over a 13 Mb domain. Thus, factors that control the balance between expression and stability likely play major roles in controlling the potency of *Airn* and other lncRNAs as well.

Within the *Airn* domain, specific CGIs appeared to nucleate the spread of H3K27me3 upon lncRNA exposure, owing to their high levels of H3K27me3 and the nearby decrease in H3K27me3 as distance from the CGIs increased. These CGIs bound RING1B and EZH2 at near equal levels on paternal and maternal alleles, but were centered in broad regions of H3K27me3 enrichment only on the lncRNA-expressing (paternal) allele. Similar relationships between CGIs, PRC binding, and H3K27me3 density were found surrounding *Kcnq1ot1* and on the TSC inactive X, although at X-linked CGIs, RING1B and EZH2 binding were higher on the active X than on the inactive one. Deletion of a lynchpin CGI at the *Slc22a3* promoter caused a multi-megabase loss of H3K27me3 in the *Airn* domain, whereas a control deletion did not. *Xist*, *Airn*, and *Kcnq1ot1* all required HNRNPK to induce the spread of PRCs, and HNRNPK loss reduced H3K27me3 at CGIs pre-loaded with PRCs in all three lncRNA target domains.

Our data suggest that lncRNAs preferentially induce the spread of PRCs from CGIs that autonomously bind PRCs. In our model (Figure 7D), individual lncRNA foci associate with high levels of PRCs owing to RNA-binding proteins such as HNRNPK, which bind both lncRNAs and PRCs and may aggregate with themselves and other proteins (Hentze et al., 2018; Pintacuda et al., 2017). Relative to sites on chromatin that lack bound PRCs, these same RNA-binding proteins may stabilize lncRNA foci at PRC-bound CGIs. The stabilization of a lncRNA carrying a payload of PRCs at a CGI would initiate PRC spread in a domain of contact, beyond the spread that was nucleated by the CGI prior to lncRNA exposure. Network-like interactions between PRC-bound CGIs, which may exist even in the absence of lncRNA expression (Isono et al., 2013), could explain how lowly expressed lncRNAs like *Airn* induce multi-megabase effects (one lncRNA focus could contact multiple

CGIs simultaneously), and why deletion of the *Slc22a3* CGI caused such a strong loss of H3K27me3 (disrupting a key CGI in a network might disrupt lncRNA access to other CGIs, as well). The PRCs and HNRNPK are now known to tether *Xist* to the X both during and after X-inactivation is complete (Colognori et al., 2019), providing precedent for the notion that interactions between lncRNAs, HNRNPK, and possibly other proteins tether lncRNAs to PRC-bound CGIs in domains including the X.

Importantly, our proposed CGI-mediated nucleation model appears to be more relevant for *Airn* and *Kcnq1ot1* than it is for *Xist*, owing to the greater non-uniformity of H3K27me3 centered around CGIs in the target domains of the former two lncRNAs, and the fact that deletion of a single PRC-bound CGI in the *Airn* domain caused a multi-megabase loss of H3K27me3. Relative to *Xist*, which is stable and can diffuse away from its site of transcription to form hundreds of nuclear foci (Cerese et al., 2014; Smeets et al., 2014; Sunwoo et al., 2015), lncRNAs such as *Airn* or *Kcnq1ot1* are unstable, may be less diffusible, and likely access far fewer regions on chromatin before being degraded or otherwise turned over. These differences rationalize how *Airn* or *Kcnq1ot1* might exhibit a greater reliance on PRC-bound CGIs to spread PRCs within their target domains. The stability of *Xist*, its affinity for actively transcribed regions of chromatin, and the large number of PRC-bound CGIs that *Xist* has access to on the X almost certainly lessen its dependence on any one CGI, particularly during the early stages of X-inactivation in the embryo or in ESCs, where many regions on the X are still transcribed and many CGIs (likely many more than in TSCs) are PRC-bound (Engreitz et al., 2013; Fursova et al., 2019; Loda et al., 2017; Pinter et al., 2012; Zylicz et al., 2019). However, the sum of PRCs bound to chromatin at the onset of X-inactivation may still play important roles in tethering *Xist* to chromatin; indeed, recent work by Colognori, Sunwoo, and colleagues suggest this to be true (Colognori et al., 2019).

Moreover, in TSCs, a subset of CGIs appear to seed high levels of H3K27me3 in broad windows on the inactive X well after their initial exposure to *Xist* (see Figure 6A). In *Drosophila*, deposition of H3K27me3 on newly incorporated nucleosomes requires the presence of DNA elements that recruit PRCs (Laprell et al., 2017). The TSC lines used in our study have maintained their H3K27me3 levels for months in culture after their initial exposure to *Xist* in the blastocyst. Therefore, rather than being related to an event occurring at the onset of X-inactivation, it seems likely that on the TSC inactive X, the increased levels of H3K27me3 surrounding PRC-bound CGIs are due to ongoing synergy between CGI- and *Xist*-dependent PRC recruitment. Thus, in certain cell types, subsets of CGIs on the X may control the intensity of PRC-induced chromatin modifications locally, long after initial exposure to *Xist*.

CGIs are known to nucleate the spread of PRCs throughout the genome (Farcas et al., 2012; Li et al., 2017; Lynch et al., 2012; Mendenhall et al., 2010; Oksuz et al., 2018; Riising et al., 2014; Woo et al., 2010), and, in prior studies of *Xist*, it has been noted that the presence of PRCs bound to chromatin correlates with the intensity of PRC-induced modifications precipitated by expression of *Xist* (Cotton et al., 2014; Kelsey et al., 2015; Loda et al., 2017; Pinter et al., 2012). However, to our knowledge, our work is the first to directly demonstrate that a single CGI is required to maintain wild-type levels of H3K27me3 in a lncRNA target

domain. In our view, this result and others we describe above imply that the proteins that cause PRCs to engage with CGIs elsewhere in the genome likely nucleate the spread of PRCs even within lncRNA target domains.

A key question that remains is, how do lncRNAs induce the spread of PRCs from CGIs? It is possible that the act of transcription, and not the lncRNA *per se*, plays a role (Kornienko et al., 2013). Indeed, transcription of *Airn* over the *Igf2r* promoter silences the latter gene, and transcription of *Kcnq1ot1* blocks access to enhancers in its gene body (Korostowski et al., 2012; Latos et al., 2012). Still, the hundreds of *Xist* RNA foci that surround the inactive X must harbor function after being transcribed (Cerase et al., 2014; Smeets et al., 2014; Sunwoo et al., 2015). Considering this in relation to data we describe above, we posit that like *Xist*, the *Airn* and *Kcnq1ot1* lncRNAs encode function by recruiting RNA-binding proteins such as HNRNPK, which may nucleate super-stoichiometric interactions with themselves, other RNA-binding proteins, and the PRCs around a given lncRNA focus (Hentze et al., 2018; Pintacuda et al., 2017).

In concert with lncRNA-induced effects, transcription may alter nuclear architecture in a way that facilitates PRC spread over short and long genomic spans (Engreitz et al., 2013; Mele and Rinn, 2016; Nozawa et al., 2017). Within TADs, transcription may promote PRC spread in a process related to DNA loop extrusion (Fudenberg et al., 2016). Between TADs, affinity between transcribed regions may help PRC-bound CGIs co-localize with lncRNA foci in 3D space.

Within lncRNA target domains and between cell types, altered activity of factors that lncRNAs require to interface with PRCs and CGIs may cause lncRNAs to vary in potency. Similar alterations, induced pharmacologically, may offer new avenues to exogenously control lncRNA silencing function.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### TSC derivation and culture

The C/B and B/C TSC lines used in this work are female and correspond to the C/B and B/C TSCs used in (Calabrese et al., 2012), and are referred to as CB.1 and BC.1 TSCs in (Calabrese et al., 2015). TSCs were cultured at 37°C on pre-plated irradiated MEF feeder cells (irMEFs) in TSC media [RPMI (Invitrogen), 20% Qualified FBS (Invitrogen), 1mM sodium pyruvate (Invitrogen), 100µM β-mercaptoethanol (Sigma), and 2mM L-glutamine] supplemented with Fgf4 (25ng/ml; Invitrogen) and Heparin (1µg/ml; Sigma) just before use. At passage, TSCs were trypsinized with 0.125% Trypsin (Invitrogen) for 3 minutes at room temperature and gently dislodged from their plate with a sterile, cotton-plugged Pasteur pipette (ThermoFisher). To deplete MEF feeder cells from TSCs prior to RNA isolation or crosslinking, TSCs were preplated for 40 minutes and cultured for three days in 70% MEF-conditioned TSC media supplemented with Fgf4 (25ng/ml; Invitrogen) and Heparin (1µg/ml; Sigma). This was done once for harvesting chromatin and twice before RNA isolation.



## Cortical neuron derivation and culture

Reciprocal F1-hybrid cortical neurons were derived from crosses between C57BL/6J and CAST/EiJ mice, cultured, and fixed with 1% formaldehyde. Embryonic day 13.5 (E13.5) to E16.5 mouse cortices were dissected and trypsinized with TrypLE express at 37 °C for 10 min. Dissociated neurons were cultured with Neurobasal medium with 5% fetal bovine serum, GlutaMAX (Invitrogen), B27 (Invitrogen) and Antibiotic-Antimycotic (Invitrogen) and changed into Neurobasal medium supplemented with 4.84 µg/ml-1 uridine 5'-triphosphate (Sigma), 2.46 µg/ml-1 5-fluoro-2'-deoxyuridine (Sigma), GlutaMAX (Invitrogen), B27 (Invitrogen), and Antibiotic-Antimycotic (Invitrogen) at DIV1 and DIV3. Neurons were fixed in 10cm plates at DIV5.

## ESC culture

ESCs (both the *Rosa26*RMCE and SM33 lines; both male) were maintained at 37°C in a humidified incubator at 5% CO<sub>2</sub>. Media was changed daily and consisted of DMEM high glucose plus sodium pyruvate, 0.1 mM non-essential AA, 100µM β-mercaptoethanol, 2 mM L-glutamine, 1000 U/ml LIF (ESG1107, Millipore Sigma, St. Louis, MO) and 15% ES-qualified FBS. RMCE cells were maintained on approximately 1.5 million irMEFs per 10-cm plate. At the passage prior to harvesting cells for ChIP, ESCs were pre-plated for 40 minutes and cultured for two days in 70% MEF-conditioned media supplemented as above.

## METHOD DETAILS

### Generation of stable cell lines

**Generation of *Airn* and *Kcnq1ot1* truncation TSCs**—To create targeting vectors to truncate the *Airn* and *Kcnq1ot1* lncRNAs, a triple-polyA sequence (from (Meng et al., 2013); kind gift of L. Meng and A. Beaudet) was cloned into the NotI and XhoI restriction sites of the PGK-neo vector (Addgene #51422; (Luo et al., 2014)). 5' prime and 3' homology arms of about 800bp each were amplified from the RP23-81B3 and RP23-101N20 BACs (BACPAC Resources), to target the *Airn* and *Kcnq1ot1* loci, respectively, and subsequently cloned upstream and downstream of the triple-polyA-neo cassette. These arms flanked sgRNA recognition sites that were designed to cut at approximately the same genomic coordinates of previous triple-polyA-mediated truncations of *Airn* and *Kcnq1ot1* in the mouse (Mancini-Dinardo et al., 2006; Sleutels et al., 2002). sgRNAs were cloned into pX330 (Addgene #42230; (Cong et al., 2013)). Oligonucleotides used to amplify homology arms and to clone sgRNAs are listed in Table S7.

Targeting vectors were linearized with HindIII and co-electroporated into C/B TSCs with pX330 at a 1:1 ratio using the Neon® Instrument (electroporation program: 950V, 30 ms, 2 pulses; Invitrogen). G418 selection (200µg/ml; Gibco) was started two days after electroporation. Individual colonies were picked on day 8 of G418 selection, and selection was continued for 7 additional days before cells were harvested for RNA expression analysis.

**Generation of *Airn* OE and *Airn* KD TSCs**—(Schertzer et al., 2018) describes our rationale and construction of the piggyBac-based vectors. To create doxycycline-inducible

Cas9, dCas9-VP160, and dCas9-KRAB vectors, a parent vector was created in which a bGH-polyA signal and an EF1 $\alpha$  promoter driving expression of a hygromycin resistance gene was ligated into the cumate-inducible piggyBac transposon vector from System Biosciences after its digestion with HpaI and SpeI, which cut just downstream of each chicken  $\beta$ -globin insulator sequence and removed all other internal components of the original vector. The TRE promoter from pTRE-Tight (Clontech) was then cloned upstream of the bGH-polyA site, and Cas9, dCas9-VP160, and dCas9-KRAB were then each cloned behind the TRE promoter by digestion with AgeI and SalI (NEB) followed by Gibson Assembly (NEB), to generate piggyBac cargo vectors capable of inducibly expressing Cas9, dCas9-VP160, and dCas9-KRAB, respectively, upon addition of doxycycline (Addgene #126029, #126031, #126030; Schertzer et al).

In parallel, an sgRNA targeting the *Airm* promoter region was cloned into pX330 (Table S7). Subsequently, the entire U6-sgRNA expression cassette, as well as a U6-sgRNA expression cassette that lacked an sgRNA targeting sequence, was cloned into the PacI site upstream of the rtTA3-IRES-Neo cassette in the rtTA-piggyBac-Cargo vector described in ((Kirk et al., 2018); Addgene #126028; (Schertzer et al., 2018)). The *Airm*-targeting-sgRNA-rtTA-Cargo vector and the non-targeting-sgRNA-rtTA-Cargo vector were each co-electroporated into wild-type C/B TSCs along with the dCas9-VP160-Cargo and dCas9-KRAB-Cargo vectors, respectively, and with the pUC19-piggyBac transposase from (Kirk et al., 2018), using the Neon® Instrument (electroporation program: 1300V, 40 ms, 1 pulse; Invitrogen). C/B TSCs were then selected on G418 (200 $\mu$ g/ml; Gibco) and Hygromycin B (150 $\mu$ g/ml; Gibco) for 9 days. Electroporation of all four vector combinations was performed a second time, and piggyBac-expressing TSCs from both series of electroporation were expanded and treated with 1 $\mu$ g/ml of doxycycline (Sigma) for four days prior to crosslinking for H3K27me3 ChIP and RNA preparation for RNA-seq, as described above.

**Generation of CGI and non-CGI deletion TSCs**—To delete the *Slc22a3* CGI and the non-CGI control region, 4 unique sgRNAs were designed that flanked each region to be excised (Table S7; Figure S6D, E). Each sgRNA was cloned into the rtTA-BsmBI piggyBac vector from (Schertzer et al., 2018), and starter cultures for each sgRNA were pooled together in equal amounts prior to liquid culture expansion and plasmid preparation using the Invitrogen HiPure Midiprep kit. The pooled vectors were co-electroporated into 500,000 TSCs with Cas9-Cargo and piggyBac transposase vectors at an 8:2:1 ratio using the Neon® Instrument (electroporation program: 1300V, 40 ms, 1 pulse; Invitrogen). Two days after electroporation, cells were selected with G418 (200 $\mu$ g/ml; Gibco) and Hygromycin B (150 $\mu$ g/ml; Gibco) for 13 days, followed by 4 days of dox treatment (1 $\mu$ g/ml). 2,000 dox-induced cells were then plated on a 10cm plate with pre-plated irMEFs. After 7 days, individual colonies were picked and plated on irMEFs for expansion. Clonal lines were passaged once off of irMEFs prior to harvests for genotyping and *Airm* RNA expression analysis.

Genotyping PCR reactions were performed with genomic DNA using Apex Taq DNA Polymerase (Genesee Scientific) and custom primers. The first set of primers flanked the deletion and identified clonal lines with at least one allele deleted. The second set only amplified a wildtype allele, with one primer sitting outside the deletion and the other inside.

Primers used are listed in Table S7 and their locations relative to the sgRNAs are shown in Figure S6D, E.

**Generation of ESCs with *Xist* inserted into the *Rosa26* locus**—A recombinase-mediated cassette exchange (RMCE) approach was used to insert a doxycycline inducible *Xist* gene into the *Rosa26* locus in an ESC line. Briefly, a male F1-hybrid mouse ESC line derived from a cross between C57BL/6J (B6) and CAST/EiJ (Cast) mice was made competent for RMCE by insertion of a custom homing cassette into the *Rosa26* locus via homologous recombination. *Xist* transgenes were cloned via PCR or recombineering into a custom RMCE-cargo vector and then electroporated along with a plasmid expressing Cre-recombinase into RMCE-competent cells using a Neon® Transfection System (Invitrogen). After selection on hygromycin (150µg/mL) and ganciclovir (3µM), individual colonies were picked and genotyped, then selected on G418 (200 µg/mL) after transfection with a pUC19-piggyBAC transposase and a piggyBac-based cargo vector containing an rtTA-expression. Creation of the ESC line is described in its entirety in a manuscript currently under revision by D.M.L., S.R.B, D.O.C, and J.M.C.

**Generation of HNRNPK knockdown TSCs**—Two sgRNAs targeting HNRNPK were designed using Desktop Genetics and cloned into the rtTA-BsmbI piggyBac vector from (Schertzer et al., 2018). Starter cultures for each sgRNA were maxi prepped using the Qiagen kit. Both sgRNA vectors were co-electroporated into 500,000 TSCs with Cas9-Cargo and (piggyBac) transposase vectors at a 1:1:1 ratio using the Neon® Instrument (electroporation program: 1300V, 40 ms, 1 pulse; Invitrogen). Two days after electroporation, cells were selected with G418 (200ug/ml; Gibco) and Hygromycin B (150ug/ml; Gibco) for 10 and 8 days (first and second experimental replicate, respectively), followed by 4 days of dox treatment (1ug/ml) prior to crosslinking for H3K27me3 CHIP and RNA preparation for RNA-seq, as described above.

### RNA Isolation, qPCR, and RNA-Seq

TSCs were passaged twice off of irMEFs with 40 minutes of pre-plating prior to RNA preparation. RNA was isolated using Trizol (Invitrogen). For RT-qPCR assays in Figure S2B, 2µg of RNA was reverse transcribed using Superscript III (Invitrogen). For assays in Figures 4B and S7B, MultiScribe RT (Applied Biosystems) was used with 2.5 µg RNA. qPCR was performed using iTaq Universal SYBR Green (Biorad) and custom primers (Table S7). RNA-Seq libraries were prepared from 1µg of RNA using Stranded mRNA-Seq Kits (Kapa Biosciences) and RNA HyperPrep Kits with RiboErase (Kapa Biosciences; Table S6).

### CHIP-Seq

Prior to crosslinking for ChIP, TSCs were passaged one time off of irMEFs with 40 minutes of pre-plating. For all ChIPs except EZH2, cells were crosslinked in RPMI media and 10% FBS with 0.6% formaldehyde for 10 minutes at room temperature. 125mM glycine was used to quench for 5 minutes at room temperature. For Ezh2 ChIPs, cells were crosslinked in 1.5mM EGS (ethylene glycol bis(succinimidyl succinate); ThermoFisher Pierce) in PBS for 30 minutes and then in 0.6% formaldehyde for 10 minutes at room temperature, and 50mM

glycine was used to quench for 10 minutes at room temperature. ChIPs were performed using 10 to 20 million cells, 5 to 10 $\mu$ l of antibody, and 25 $\mu$ l of Protein A/G agarose beads (Santa Cruz). Sonication was performed on a Vibracell VCX130 (Sonics) with cycles of 30% intensity for 30 seconds with 1 minute of rest on ice between cycles. TSCs crosslinked in 0.6% formaldehyde and EGS/formaldehyde were sonicated for 10 and 12 cycles, respectively. Crosslinked cortical neurons and ESCs were sonicated for 10 cycles. Antibodies used were: H3K27me3 (Abcam ab6002), H2AK119ub (Cell Signaling #8240), CTCF (kind gift from V. Lobanenkov), EZH2 (Cell Signaling #5246), RING1B (Cell Signaling #5694), and SMC1 (Bethyl A300-055A).

For H3K27me3 and SMC1 ChIPs, 10 million crosslinked TSCs were re-suspended in lysis buffer 1 (50mM HEPES pH 7.3, 140mM NaCl, 1mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, and 1x PIC (protease inhibitor cocktail; Sigma) and incubated for 10 min at 4C, and then incubated with lysis buffer 2 (10mM Tris-HCl pH 8.0, 200mM NaCl, 1mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, and 1x PIC) for 10 min at RT. For H3K27me3 ChIPs, cells were resuspended in lysis buffer 3 (10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA pH8.0, 0.5 mM EGTA pH 8.0, 0.1% Na-deoxycholate, 0.5% N-lauroylsarcosine, and 1x PIC) and then sonicated. For SMC1 ChIPs, cells were resuspended in a sonication buffer (20mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA pH 8.0, 0.1% SDS, 1% Triton X-100, and 1x PIC) and then sonicated. ChIPs were performed by incubating sonicated cell lysates at a concentration of 20 million cells/1ml of lysis buffer 3 containing 1% Triton X-100 with pre-conjugated antibody/agarose beads overnight at 4°C. After over night H3K27me3 ChIP, beads were washed 5x in RIPA buffer (50 mM HEPES pH 7.3, 500 mM LiCl, 1 mM EDTA, 1% NP-40 and 0.7% Na-Deoxycholate) for 5 minutes each and then once in TE. After overnight SMC1 ChIP, beads were washed once with lysis buffer 3 (20mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA pH 8.0, 0.1% SDS, 1% Triton X-100, and 1x PIC), once with High Salt Buffer C (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 2 mM EDTA pH 8.0, 0.1% SDS, 1% Triton X-100, and 1x PIC), once with Buffer D (10mM Tris-HCl pH 8.0, 250 mM LiCl, 1mM EDTA pH 8.0, 1% NP-40, and 1xPIC), and once with TE + 50 mM NaCl. To elute the DNA, beads were re-suspended in Elution buffer (50mM Tris pH 8.0, 10mM EDTA, and 1% SDS) and placed on a 65°C heat block for 17 minutes to 1 hour with frequent vortexing. Crosslinks were reversed overnight at 65°C, eluates were incubated with Proteinase K and RNase A, and DNA was extracted with phenol/chloroform and precipitated with ethanol.

For CTCF, EZH2, and RING1B ChIPs, 10 million crosslinked cells were resuspended in buffer 4 (50 mM Tris-HCl pH 7.5, 140 mM NaCl, 1 mM EDTA, 1 mM EGTA, 0.1% Na-deoxycholate, 0.1% SDS), sonicated to generate 200-500 bp DNA fragments, cleared via centrifugation, and diluted to 20 million cells equivalents per ml of buffer 4 containing 1% Triton-X100. Post-ChIP, beads were washed 3x with buffer 4 containing 1% Triton-X100, once with buffer 4 containing 1% Triton-X100 and 500mM NaCl, once with buffer 5 (20 mM Tris pH 8.0, 1 mM EDTA, 250 mM LiCl, 0.5% NP-40, 0.5% Na-deoxycholate), and once with TE before eluting and reversing crosslinks as above.

DNA was prepared for sequencing on the Illumina platform using Next Reagents (NEB) and Agencourt AMPure XP beads (Beckman Coulter).

## DNA/RNA FISH

BACs and fosmids (Key Resources Table) were ordered from the BACPAC resource center and fingerprinted with restriction digestion prior to use to verify inserted DNA. Fluorescent labeling was performed using BioPrime (Invitrogen). For RNA/DNA FISH in wild-type TSCs, BAC RP23-309H2O was used to mark the *Airn* DNA locus, and fosmid W11-2156F18 was used to detect *Airn* RNA. For RNA/DNA FISH in *Airn* truncation TSCs, fosmid W11-662A5 was used to detect expression of *Igf2r*, which marks the maternal allele owing to its monoallelic expression. *Igf2r* remained monoallelic even in *Airn* truncation cells, because the polyA signal for the G418 resistance gene that we used to select for *Airn* truncations is in the same orientation as *Igf2r* transcription. The G418 polyA signal therefore causes early termination of *Igf2r* on the paternal allele, effectively suppressing its transcription even if paternal *Igf2r* might have been reactivated by truncation of *Airn*.

RNA/DNA FISH was performed essentially as in (Byron et al., 2013). TSCs were fixed on coverslips for 10 minutes in 4% paraformaldehyde/PBS, followed by a 10-minute permeabilization on ice in 0.5% TritonX-100 in PBS and 1:200 Ribonucleoside Vanadyl Complex (NEB). Coverslips were stored at -20C in 70% ethanol until use. To initiate the RNA/DNA FISH protocol, coverslips were dehydrated by serial 3-minute incubations with 75%, 85%, 95%, and 100% ethanol, and air-dried for 5 minutes. Biotinylated RNA FISH probe was added and coverslips were placed cell-side down in a chamber humidified with 50% formamide/2xSSC overnight at 37°C. After overnight incubation, coverslips were washed 3x with 50% formamide/2xSSC at 42C, 3x with 1xSSC at 50C, 1x with 1xSSC at room temperature, and 1x with 4xSSC. Each wash was 5 minutes long. After the final wash, streptavidin AlexaFluor 555 (Invitrogen) was diluted 1 to 2000 in 4xSSC 1ug/ml BSA, and added dropwise to each coverslip. Coverslips were incubated cell-side down at 37°C for one hour in a chamber humidified with 4xSSC. Coverslips were then washed for 10 minutes each with 4xSSC, 4xSSC plus 0.1% TritonX-100 (Fisher Scientific), and then 4xSSC again in a chamber humidified with 4xSSC at 37C. Coverslips were rinsed with 1xSSC then 1xPBS, then fixed with 2% paraformaldehyde in PBS for 5 minutes at room temperature. Coverslips were rinsed twice with PBS, then incubated for 5 minutes at room temperature in 200mM NaOH to degrade RNA. Cells were then rinsed with 70% ethanol and denatured at 80°C for 20 minutes in preheated 70% formamide/2xSSC. Coverslips were then washed with ice cold 70% ethanol, and with 100% ethanol, then allowed to air dry. DNA FISH probes were then added and coverslips were placed face down in a chamber humidified with 50% formamide/2xSSC overnight at 37°C The next day, coverslips were washed 3x with 50% formamide/2xSSC at 42°C, and 3x with 1xSSC at 55°C. Each wash was 5 minutes long. Coverslips were then rinsed 1x with PBS before a 2-minute incubation in DAPI stock diluted 1:1000 in water. Coverslips were rinsed twice more and affixed to glass slides using Vectashield (VectorLabs), then sealed with nail polish.

Four dimensional datasets were acquired by taking multi-channel Z-stacks on an Olympus BX61 widefield fluorescence microscope using a Plan-Aprochromat 63X/1.4 oil objective and a Hamamatsu ORCA R2 camera, controlled by Volocity 6.3 software. Excitation was provided by a mercury lamp and the following filters were used for the four fluorescent channels that were imaged: 377/25 ex, 447/30 em for DAPI (DAPI-5060B Semrock filter);



482/17 ex, 536/20 em for AlexaFluor488, (Semrock FITC-3540B filter); 562/20 ex, 642/20 em for AlexaFluor 555 (Semrock TXRED-4040B filter); 628/20 ex, 692/20 em for Cy5 (Semrock Cy5 4040A filter). Pixel size was 0.103  $\mu\text{m}$ , Z spacing was 0.2  $\mu\text{m}$ , and images had 1344X1024 pixels. Approximately 40 Z slices were acquired for each Z-stack. Z-stacks were deconvolved using the iterative-constrained algorithm (Mediacy AutoQuantX3) with default algorithm settings. Sample settings for the deconvolution were: peak emissions for dyes (670 nm, 565 nm, 519 nm, 461 nm for Cy5, AlexaFluor 555, AlexaFluor 488 and DAPI respectively), widefield microscopy mode, NA = 1.4, RI of oil = 1.518, and RI of sample = 1.45. After deconvolution, DNA and RNA FISH signals were located using the “Spots” function in Imaris software (version 8.3.1, Bitplane) and marked with equal sized spheres. To initially call spots on all images, spot detection values were set at 0.5 $\mu\text{m}$  for xy and 1.5 $\mu\text{m}$  for z, and background subtraction and auto quality settings were used. We manually optimized the quality/sensitivity setting to match the expected 1 RNA spot or 2 DNA spots per cell. Only nuclei in which we could observe the expected number of DNA and RNA FISH signals (four and one, respectively), were counted. Images are shown as maximum intensity projections that were made using ImageJ.

To correct for chromatic aberrations that distort the relative positions of spots (in XYZ) labeled with different fluorophores, we systematically corrected all spot positions based on calibrations performed with 0.2 $\mu\text{m}$  diameter Tetraspeck beads (ThermoFisher). These were diluted in ethanol, dried on coverslips, mounted with Vectashield, and imaged with the same filters as the sample slides, and deconvolved and analyzed with the same settings as the samples. Because each Tetraspeck bead is labeled with four fluorophores, the spots in each fluorescent channel for a given bead should localize to the same position in XYZ. Thus, by analyzing the deviations of the detected spot positions from the actual spot positions (where all channels should be colocalized) across the field of view we were able to determine the corrections that had to be applied to compensate for the system’s chromatic aberrations. These corrections were a function of the particular pair of fluorophores that needed to be compared, as well as the position in X and Y in the field of view. We found that the Z position required a shift dependent on the fluorophore pair, independent of XY position in the field (mean shift = -0.462  $\mu\text{m}$ ). In contrast, the XY coordinates of detected spots required a shift dependent on the position in the field of view (the difference between actual and detected spot positions between channels varied linearly across X and Y). Once we had constructed the chromatic correction model for the entire field of view in XYZ and the necessary fluorescent channel pairs, we applied the corrections to the beads and found that the distance between a detected spot in the Cy5 and AlexaFluor488 channels was smaller than 136 nm in 95% of cases. This gives an upper bound on the precision of our measurement and analysis scheme, i.e. we could confidently make statements about distances larger than 136 nm, but not smaller. We then applied the chromatic correction model to each imaged spot, prior to measuring distances between DNA FISH probes.

### **ERCC spike-ins to measure lncRNA copy # per cell**

We took the following approach to calculate transcript copy number per TSC. First, we purified and quantified total RNA from known numbers of TSCs, in triplicate, to determine that the average TSC contains 30pg of RNA (not shown). Second, prior to preparing RNA-

seq libraries, we added 2 $\mu$ l of a 1:100 dilution of ERCC RNA Spike-In Mix #1 (Invitrogen, 4456740) to 1 $\mu$ g of the TSC RNA sample of interest according to Invitrogen's recommendation. We then proceeded with RNA-seq library preparation using RNA HyperPrep Kits with RiboErase (Kapa Biosciences). For lncRNA copy number per cell measurements (Figure 4), RNA-seq libraries were sequenced from the following RNA preparations: single preparations of RNA from replicate derivations of dCas9-VP160/non-targeting-sgRNA-rtTA-expressing C/B TSCs, single preparations of RNA from replicate derivations of dCas9-VP160/*Aim*-targeting-sgRNA-rtTA-expressing C/B TSCs, single preparations of RNA from replicate derivations of dCas9-KRAB/non-targeting-sgRNA-rtTA-expressing C/B TSCs, and single preparations of RNA from replicate derivations of dCas9-KRAB/*Aim*-targeting-sgRNA-rtTA-expressing C/B TSCs. Counts from wild-type TSCs and dCas9-VP160/ and dCas9-KRAB/non-targeting sgRNA TSCs were collectively considered "wild-type".

To calculate transcript copy number per cell, reads were aligned to a version of the mm9 genome with ERCC.fa sequences doped in, and a standard curve was created to link RPKM values for each of the ERCC-Spike-In RNAs to their molecular abundance (see Figure S4A for an example). These RPKM-to-abundance ratios were used to calculate molecular abundance of our lncRNAs of interest in 1 $\mu$ g of RNA, and this abundance was divided by 33,333 (the approximate number of TSCs that would give rise to 1 $\mu$ g of RNA) to determine the lncRNA copy number per TSC reported in Figure 4A.

### Measurement of lncRNA half-life

TSCs were treated with a final concentration of 5 $\mu$ g/ml Actinomycin D (Sigma) for 10 minutes, 20 minutes, and 30 minutes, and 1, 2, 4, and 8 hours prior to lysis with Trizol (Invitrogen) and RT-qPCR to measure expression of *Xist*, *Aim*, and *Kcnq1ot1* relative to GAPDH (Table S7). Actinomycin D treatment and RNA preparation was performed twice in total, once each on separate days.

To model lncRNA half-life, lncRNA levels were measured by qRT-PCR relative to *Gapdh* mRNA at each time point. Levels were then represented as a percentage relative to 0 hr and transformed by the natural log. Linear models using time as a predictor of RNA percentage were fit to the data for each biological replicate and then used to find the time at which the percent of RNA remaining relative to the 0 hr time point was 50%. This value was reported as the half-life.

### RNA fractionation

To isolate RNA from cytosolic, soluble nuclear, and chromatin-bound fractions, cells were passaged once off of irMEFs with 40 minutes of pre-plating and cultured in conditioned media on a 10cm plate. *Aim*-overexpression TSCs were induced with doxycycline at 1 $\mu$ g/mL for 3 days. For RNA harvest, cells were washed twice with 1mL cold PBS and scraped in 1mL PBS + 1mM PMSF + 1:100 protease inhibitor cocktail (PIC, Sigma P8340). 200 $\mu$ L was removed at this step and cells were resuspended in 1mL Trizol (total RNA). The remaining cells were centrifuged at 1500 $\times$ rcf for 5min, and resuspended in 250 $\mu$ L low salt solution (10mM KCl, 1.5mM MgCl<sub>2</sub>, 20mM Tris-HCl pH 7.5) supplemented with 1mM

PMSF, 1mM DTT, and 1x PIC. Triton X-100 was added to a final concentration of 0.1% and cells were rotated for 10min at 4°C, then centrifuged for 5min at 1500x*rcf*. 200uL of supernatant was added to 1mL Trizol (cytosolic fraction). The remaining supernatant was discarded and the nuclear pellet was washed by rotating for 2min at 4°C in low salt solution with Triton X-100 and centrifuged at 1300rpm for 10min. Nuclei were resuspended in 100uL Buffer B (3mM EDTA, 0.2mM EGTA) supplemented with 1mM DTT, 1mM PMSF, and 1x PIC, rotated for 30min at 4°C and centrifuged at 1700x *rcf* for 10min. 80uL of supernatant was added to 1mL Trizol (soluble nuclear fraction). The chromatin pellet was washed by rotating for 2-5min in Buffer B and centrifuged at 1700x*rcf* for 10min. The pellet was resuspended in 1mL Trizol (chromatin-bound fraction). Isolation of RNA from Trizol was performed according to manufacturer protocol. Equal amounts of RNA (1µg) were reverse transcribed using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) with random primers. qPCR was performed using iTaq Universal SYBR Green (Bio-Rad) and custom primers (Table S7). For a given RNA species, log-transformed Cq values for each fraction were added together, and the percentage of total signal coming from each fraction was plotted in R.

### Stellaris RNA FISH

Custom Stellaris® FISH probes were designed against the first 40kb of *Airm* and the first 2kb of *Xist* by utilizing the Stellaris® RNA FISH Probe Designer (Biosearch Technologies, Inc., Petaluma, CA) available online at [www.biosearchtech.com/stellarisdesigner](http://www.biosearchtech.com/stellarisdesigner) and labeled with Quasar 570 and 670 dyes, respectively. The ShipReady Quasar 670 probe set was used for *Gapdh* (Cat# SMF-3140-1). Cells were grown on MEFs on glass coverslips for 2 days in the presence or absence of 1µg/mL doxycycline before being washed once with 1x PBS, fixed for 10min at room temperature with 4% formaldehyde in 1x PBS, washed twice with 1x PBS, and permeabilized overnight with cold 75% ethanol at 4°C. 1uL of 2.5µM probe was added to 100uL of hybridization solution (10% dextran sulfate, 2x SSC, 10% formamide) and pre-warmed to 37°C. Coverslips were washed at 37°C for 2-5min in pre-warmed wash buffer (2x SSC, 10% formamide). Coverslips were incubated with diluted probes overnight at 37°C in a humidified chamber, then washed twice with wash buffer at 37°C for 30min, adding DAPI to 5ng/mL for the second wash. Coverslips were rinsed with 2x SSC, mounted using Prolong Gold and allowed to cure overnight at room temperature. Images were acquired and deconvolved similarly to DNA/RNA FISH images and maximum intensity projections were made using ImageJ (Schindelin et al., 2012).

### RNA immunoprecipitation

RNA IP experiments were performed using the protocol outlined in (Raab et al., 2019). Prior to fixation, TSCs were passaged once off of irMEFs with 40 minutes of pre-plating and cultured in conditioned media. Prior to fixation of SM33 ESCs, *Xist* expression was induced with 1µg/mL doxycycline for 24 hours. TSCs and SM33 ESCs were trypsinized, washed twice with PBS, then fixed in 0.3% methanol-free formaldehyde for 30 min at 4 °C. Formaldehyde was quenched with 125 mM glycine for 5 min at room temperature. Cells were snap frozen in liquid nitrogen and stored at -80 °C.

For each IP, 10 million cells were resuspended in 0.5 ml RIPA buffer (50 mM Tris-HCl pH 8, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS, 5 mM EDTA, 150 mM KCl) supplemented with 0.5 mM DTT, 1x protease inhibitor cocktail (Sigma), and 2.5  $\mu$ l RNAsin (Ambion), and incubated on ice for 10 min prior to lysing using a Vibracell VCX130 (Sonic) with two cycles of 30% intensity for 30 seconds with 1 minute of rest on ice between cycles, followed by centrifugation at 4  $^{\circ}$ C for 20 min at maximum speed. Subsequently, extracts were diluted with 0.5ml fRIP buffer (25 mM Tris-HCl pH 7.5, 5 mM EDTA, 0.5% Igepal CA-630, 150 mM KCl) supplemented with 0.5 mM DTT, 1x protease inhibitor cocktail (Sigma), and 2.5  $\mu$ l RNAsin (Ambion). In parallel, per IP, 25ul of protein A/G agarose beads (Santa Cruz) were pre-conjugated with antibody overnight in PBS and 0.5% BSA at 4  $^{\circ}$ C. 10  $\mu$ l of HNRNPK (Abcam, ab39975) antibody was used for HNRNPK IP in SM33 cells, and 10  $\mu$ l of HNRNPK (Santa Cruz, sc28380) antibody was used for HNRNPK IP in TSCs. 10ul of CTCF antibody (a kind gift from V. Lobanenko) was used for CTCF IP in TSCs. 10ug of mouse IgG (Invitrogen, 02-6502) was used as the "IgG control". After sonication, clarification, and dilution in fRIP buffer, extracts were united with antibody/bead mixtures and incubated overnight at 4  $^{\circ}$ C with end-over-end rotation. Beads were washed consecutively with fRIP buffer (25 mM Tris-HCl pH 7.5, 5 mM EDTA, 0.5% Igepal CA-630, 150 mM KCl), three times in ChIP buffer (50 mM Tris-HCl pH 7.5, 140 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS), once in high salt buffer (ChIP buffer, but with 500 mM NaCl) and once in (20 mM Tris pH 8.0, 1 mM EDTA, 250 mM LiCl, 0.5% NP-40, 0.5% Na-deoxycholate). All washes were performed for 5 min at 4  $^{\circ}$ C. After the first and final wash, solutions were transferred to clean tubes. After the final wash, beads were resuspended in 100ul of 1x reverse crosslinking buffer (1x PBS, 2% N-lauroyl sarcosine, 10 mM EDTA) supplemented with 5 mM DTT. Per IP, 20  $\mu$ l proteinase K and 1  $\mu$ l RNAsin were added and samples were incubated 1 h at 42  $^{\circ}$ C, 1 h at 55  $^{\circ}$ C, and 30 min at 65  $^{\circ}$ C and were mixed by pipetting every 15 minutes. Following, samples were mixed with 1ml of Trizol. 200ul of chloroform was added, the aqueous phase was extracted and mixed with 1 volume of ethanol, vortexed, then purified using a Zymo-spin IC column, using the on-column DNase digestion as per the manufacturer's instruction. RNA was eluted in 15  $\mu$ l of deionized water. RNA-seq libraries were prepared using 9ul of immunoprecipitated RNA from each condition mixed with 1  $\mu$ l of 1:250  $\mu$ l dilution of ERCC spike-in mix 1 (Invitrogen, 4456740). The SM33 input library was prepared from 200ng of RNA and 1ul of a 1:100 dilution of ERCC spike-ins, and the TSC input library was prepared from 100ng of RNA and 1ul of a 1:250 dilution of ERCC spike-ins. Libraries were prepared using the Kapa RiboErase kit following the manufacturer's instructions, pooled, and sequenced using single-end 75-bp reads on an Illumina Nextseq 500.

### **HNRNPK and H3K27me3 Immunofluorescence**

TSCs were fixed on coverslips as described above in preparation for DNA/RNA FISH. To initiate the IF protocol, coverslips were washed twice in PBS and blocked for 30 minutes at room temperature in blocking solution (1x PBS with 0.2% Triton X-100, 1% goat serum, and 6 mg/mL IgG-free BSA). Then, coverslips were washed in 0.2% triton/1x PBS and incubated with HNRNPK antibody (Santa Cruz 28380) and H3K27me3 antibody (Cell Signaling #9733) diluted 1:200 in block solution for 1 hour at RT. Coverslips were washed

3x in 0.2% triton/1x PBS for 4 minutes each and incubated with secondary antibody (AlexaFluor 488 goat anti-mouse, A-11029 and AlexaFluor 594 goat anti-rabbit, A-11037) diluted 1:400 in block solution for 30 minutes at RT. After incubation, coverslips were washed 3x in 0.2% triton/1x PBS for 4 minutes each and rinsed 1x with PBS before a 2-minute incubation in DAPI stock diluted to 1 $\mu$ g/ml in water. Coverslips were rinsed twice more and mounted to glass slides using Prolong Gold (Thermo Fisher Scientific P10144). Images were acquired and deconvolved similarly to DNA/RNA FISH images and maximum intensity projections were made using Image J.

### Protein isolation and western blotting

To isolate protein for western blotting, cells were washed with PBS, and then lysed with RIPA buffer (10 mM Tris-Cl (pH 7.5), 1 mM EDTA, 0.5 mM EGTA, 1% NP40, 0.1% sodium deoxycholate, 0.1% SDS, 140 mM NaCl) supplemented with 1 mM PMSF (Fisher Scientific) and 1x protease inhibitor cocktail (Sigma) for 15 minutes at 4°C. Prior to western blotting, protein levels were quantified using the DC assay from Biorad. For western blotting, primary and secondary antibody incubations were done for 1hr at room temperature. Antibodies used were HNRNPK (Santa Cruz sc-28380, 1:5000 dilution), TBP (Abcam ab818, 1:2000 dilution), donkey anti-mouse IgG-HRP secondary (Santa Cruz; sc-2314; 1:2500), and donkey anti-rabbit IgG-HRP secondary (Santa Cruz; sc-2313; 1:2500).

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Sequence alignment and processing

RNA sequence reads were aligned to genomic sequence using Star (version 2.6.0a; (Dobin et al., 2013)) and ChIP sequencing reads were aligned using bowtie2 (Langmead et al., 2009), using default parameters. All mm9 genome annotations were obtained from the UCSC genome browser. Variant sequence data was obtained from the Sanger Institute (<http://www.sanger.ac.uk/resources/mouse/genomes/>). Samtools was used to filter for reads that had a mapping quality greater than or equal to 30 (Li et al., 2009). CAST/EiJ pseudogenome creation and allele-specific read retention was performed as in (Calabrese et al., 2015; Calabrese et al., 2012). All genome-related plots were generated using R.

### ChIP-Seq Peak Calling

ChIP-Seq data between wild-type C/B and B/C TSCs were pooled to identify peaks in H3K27me3, CTCF, SMC1, and EZH2 datasets. Peaks of RING1B in TSCs were identified by pooling ChIP-Seq data from *Airn* OE, *Airn* WT, *Airn* KO, and *Kcnq1ot1* KO TSCs. H3K27me3 peaks in ESCs were called separately for the 0hr and 72hr datasets. RING1B peaks in ESCs were identified by pooling 0hr, 12hr, and 72hr datasets. hiddenDomains 3.0 was used to call peaks from H3K27me3 ChIP-Seq data, using input DNA sequenced from formaldehyde crosslinked TSCs or ESCs as a control (Starmer and Magnuson, 2016). Default parameters plus max.read.count=10 were used in the main hiddenDomains.R script, and neighboring enriched bins were merged to generate the final set of H3K27me3 peaks. MACS2 was used to call peaks from all other ChIP-Seq datasets, with the same input used in hiddenDomains and parameters -broad -broad-cutoff 0.01 (Zhang et al., 2008). All peaks



were called using allele-nonspecific data, owing to its significantly higher coverage relative to allele-specific data.

### Parent-of-origin bias in H3K27me3 peaks

Allele-specific reads falling within peaks of H3K27me3 were counted in the C/B and B/C TSC datasets, and counts were imported into edgeR and normalized using edgeR's counts per million (CPM) metric (Robinson et al., 2010). H3K27me3 ChIP-seq data in C/B TSCs was from (Calabrese et al., 2012), and B/C data was generated as part of this study. Allelic data from individual H3K27me3 replicates within C/B and B/C TSC datasets were merged before importing into edgeR. Autosomal peaks with  $\geq 1$  allelic cpm in each dataset were analyzed. X-linked peaks were excluded. Differential enrichment between Cast and B6 alleles within each individual peak in each F1-hybrid TSC line was tested via edgeR's generalized linear model likelihood ratio test, and p-values from both tests were adjusted for false discovery using the Benjamini-Hochberg correction method. Peaks exhibiting PO biases with false discovery rates scores of  $\geq 0.1$  in both C/B and B/C TSCs were considered to be significantly biased (Table S1).

### Allelic changes upon lncRNA truncation

Allelic reads were retained that fell within exons as well as introns for all UCSC Known Genes. Counts were then used to calculate the proportion of paternal expression for each gene in each sample, and this proportion was then arcsine transformed. To detect differentially expressed genes, two tailed t-tests were performed on the transformed data comparing three knockout TSC lines to five wild-type TSC lines. For genes in the *Airn* region, the three *Kcnq1ot1* knockout TSC lines plus the two wildtype C/B TSC lines were used as *Airn* wild-type. Similarly, for genes in the *Kcnq1ot1* region, the three *Airn* knockouts plus the two wildtype C/B TSC lines were used as *Kcnq1ot1* wild-type. The arcsine transformation was used to eliminate the bounds of '0' and '1' in proportions and to spread the data out at the extremes and only the extremes, thereby validating the assumptions inherent in a two-sided t-test. See Table S2.

### Genome alignability

The proportion of mm9 that could be uniquely mapped using 45 or 75bp sequence tags, depending on the length of read from the dataset in question, was defined as genome alignability.

### Genome-wide correlations in ChIP-Seq datasets

To derive Pearson's r values between H3K27me3 and H2AK119ub pooled datasets, reads were counted in 10kb bins genome wide using bedtools coverage (Quinlan and Hall, 2010). Counts were normalized for dataset size. For H3K27me3 and H2AK119ub comparisons, only bins with total H3 rpm  $\geq 1$  were used (H3 data from (Calabrese et al., 2012)). To derive Pearson's r values between RING1B and EZH2 density with H3K27me3 peaks on the X chromosome, paternal reads were counted per H3K27me3 peak. Only peaks with an average of  $>10$  paternal reads per dataset were used.

### Bootstrap approach for FISH measurements

We used a bootstrap approach to determine if the correlations between paternal H3K27me3 density in the nine regions probed for DNA FISH and distance in base pairs to *Airn*, maternal distance to *Airn*, and paternal distance to *Airn* (Figure 2F) were significantly different. For each iteration of the bootstrap, we calculated the mean values for the bootstrap-dataset on the maternal and paternal allele, and used a linear regression fit to determine the R-squared value between the maternal and paternal distance measurements and paternal H3K27me3. We repeated this process 10,000 times and used the distributions of R-squared values under each condition and either a one-sample (Figure 2Fii) or two-sample (Figure 2Fiii) test to calculate empirical p-values that assess whether the differences in R-squared values between our three comparisons were significantly different.

### Hi-C and ChIA-PET data

Hi-C data was downloaded from and processed using Juicebox (Durand et al., 2016). ESC SMC1 ChIP-PET contact calls were from (Downen et al., 2014), and ESC SMC1 and CTCF ChIP-Seq data were from (Kagey et al., 2010; Stadler et al., 2011).

### Chromosome tiling plots using bedtools

Chromosome-scale H3K27me3 tiling density plots in Figures 4, 5, 6, 7, S6, and S7 were created by summing total H3K27me3 counts in 40kb bins across each chromosome, moving in 4kb increments (using bedtools coverage on sorted bam files). All counts per bin were normalized for alignability, where total reads per bin were divided by the proportion of alignable bases per bin. Bins with alignability of less than 0.5 (i.e. less than 50% alignable at 75bp (Figures 4, 5, 6, 7, and S6) and 45bp resolution (Figure S7) were excluded from tiling density plots to avoid potential uncertainty that would be introduced by normalizing highly non-unique regions.

In Figures 4, 5, 6, S6, and S7, read counts were normalized between datasets by multiplying by 1 million then dividing by the total number of reads per dataset. Read normalization for HNRNPK knockdown experiments (Figure 7) is described in a separate subheaded section below.

### H3K27me3 ChIP normalization using *Drosophila* DNA

It was possible that HNRNPK knockdown would cause a global reduction in H3K27me3 that if left unaccounted for during library preparations, would have obscured a reduction in H3K27me3 in lncRNA target domains. To circumvent this possibility, after H3K27me3 ChIP in non-targeting and HNRNPK knockdown TSCs and prior to the preparation of sequencing libraries, an amount of sonicated *Drosophila melanogaster* DNA (kind gift of D. McKay) equal to 1% of the total amount of DNA in the lowest yielding ChIP sample (74.2 picograms) was added to 10ul of each ChIP (one third of the total volume of eluted ChIP'd DNA for each sample). After sequencing, reads were aligned to the mm9 and dm6 genomes. Normalization factors were created by dividing the total number of aligned *Drosophila* reads in each sample by the lowest *Drosophila* read count amongst samples, giving a factor of 1 for the lowest sample and values greater than 1 for all other samples. Binned read counts (from mm9) were divided by these normalization factors and then divided by the input DNA

amount in ng for each IP. To be able to directly compare Y-axes displayed in Figure 7B and C, read counts were then divided by bin size in kb, which was 40kb for Figure 7B and 4kb for Figure 7C.

### Determining feature overlap using bedtools

To determine feature overlap for Venn diagrams in Figure S3 and for CGI classification in Figures 5, 6, 7, and S6, bedtools intersect was used on MACS2 “broadpeak” files and UCSC-annotated CGIs.

### Metagenes

Allele specific metagene plots in Figures 5, 6, and S6 were constructed using the following approach. For each dataset, counts of sequence read starts were recorded in 500bp bins surrounding the annotated TSS or center of the feature for the gene/feature class in question. In addition to normalization for gene number, allelic counts were normalized for the number of uniquely alignable SNPs present in each bin for the specific gene/feature set being analyzed. Non-allelic metagenes in Figures 6 and S3 were generated with HOMER (Heinz et al., 2010). To create tag directories of aligned reads, “makeTagDirectory” was used. Then, “annotatePeaks.pl” was used to generate metagenes with 500bp bins in a 100kb window for Figure 6 and 50bp bins in a 4kb window for Figure S3. The “Coverage” column was used for plotting.

### Measurement of Signal over IgG in HNRNPK RNA IP data

RNA-IP reads were aligned to a version of the mm9 genome with ERCC.fa sequences doped in. Samtools was used to filter aligned reads for  $q > 30$ . (Li et al., 2009). Reads were overlaid with UCSC known gene annotations using featureCounts to determine the read count per transcript (Liao et al., 2014). For normalization, counts per ERCC spike-in transcript were generated for each dataset using featureCounts on the ERCC92.gtf file. The upper quartile values from the set of ERCC spike-in transcripts quantified for each dataset were used to normalize all datasets relative to their respective total RNA input dataset (either SM33 or TSC). Wiggle tracks in Figure 7A and transcript read counts in Table S5 were scaled using these factors. HNRNPK and CTCF normalized counts were divided by IgG normalized counts to give the signal relative to IgG values that are reported in Figure 7A.

## DATA AND SOFTWARE AVAILABILITY

All sequencing data generated as a part of this study have been deposited to NCBI GEO under the accession number GSE118402. Raw image and western data can be accessed through Mendeley, under the links <http://dx.doi.org/10.17632/bv9y5rcpzz.1> and <http://dx.doi.org/10.17632/nk84zzwjkh.1>.

## CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Mauro Calabrese (jmcablabr@med.unc.edu).

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank V. Lobanenkov for the CTCF antibody, and apologize to colleagues we did not cite for space. This work was supported by NIH Grant GM121806 and Basil O'Connor Award #5100683 from the March of Dimes Foundation (J.M.C.), NIH Grant DP1ES024088 (M.J.Z), NIH Grant R35GM124764 (J.M.D.), and NIH Grant R01GM101974 (T.R.M). D.M.L, R.E.C., and M.J. were supported in part by the NIGMS training award T32 GM007092, and K.C.A.B. by the NIGMS award T32 GM119999. Microscopy was supported in part by NCI grant P30 CA016086.

## References

- Almeida M, Pintacuda G, Masui O, Koseki Y, Gdula M, Cerase A, Brown D, Mould A, Innocent C, Nakayama M, et al. (2017). PCGF3/5-PRC1 initiates Polycomb recruitment in X chromosome inactivation. *Science* 356, 1081–1084. [PubMed: 28596365]
- Andergassen D, Dotter CP, Wenzel D, Sigl V, Bammer PC, Muckenhuber M, Mayer D, Kulinski TM, Theussl HC, Penninger JM, et al. (2017). Mapping the mouse Allelome reveals tissue-specific regulation of allelic expression. *eLife* 6.
- Blackledge NP, Farcas AM, Kondo T, King HW, McGouran JF, Hanssen LL, Ito S, Cooper S, Kondo K, Koseki Y, et al. (2014). Variant PRC1 Complex-Dependent H2A Ubiquitylation Drives PRC2 Recruitment and Polycomb Domain Formation. *Cell* 157, 1445–1459. [PubMed: 24856970]
- Byron M, Hall LL, and Lawrence JB (2013). A multifaceted FISH approach to study endogenous RNAs and DNAs in native nuclear and cell structures. *Curr Protoc Hum Genet Chapter* 4, Unit 4 15.
- Calabrese JM, Starmer J, Schertzer MD, Yee D, and Magnuson T (2015). A survey of imprinted gene expression in mouse trophoblast stem cells. *G3 (Bethesda)* 5, 751–759. [PubMed: 25711832]
- Calabrese JM, Sun W, Song L, Mugford JW, Williams L, Yee D, Starmer J, Mieczkowski P, Crawford GE, and Magnuson T (2012). Site-specific silencing of regulatory elements as a mechanism of X inactivation. *Cell* 151, 951–963. [PubMed: 23178118]
- Cerase A, Smeets D, Tang YA, Gdula M, Kraus F, Spivakov M, Moindrot B, Leleu M, Tattermusch A, Demmerle J, et al. (2014). Spatial separation of Xist RNA and polycomb proteins revealed by superresolution microscopy. *Proc Natl Acad Sci U S A* 111, 2235–2240. [PubMed: 24469834]
- Chadwick BP (2007). Variation in Xi chromatin organization and correlation of the H3K27me3 chromatin territories to transcribed sequences by microarray analysis. *Chromosoma* 116, 147–157. [PubMed: 17103221]
- Colognori D, Sunwoo H, Kriz AJ, Wang CY, and Lee JT (2019). Xist Deletional Analysis Reveals an Interdependency between Xist RNA and Polycomb Complexes for Spreading along the Inactive X. *Mol Cell*.
- Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823. [PubMed: 23287718]
- Cotton AM, Chen CY, Lam LL, Wasserman WW, Kobor MS, and Brown CJ (2014). Spread of X-chromosome inactivation into autosomal sequences: role for DNA elements, chromatin features and chromosomal domains. *Hum Mol Genet* 23, 1211–1223. [PubMed: 24158853]
- Darrow EM, Huntley MH, Dudchenko O, Stamenova EK, Durand NC, Sun Z, Huang SC, Sanborn AL, Machol I, Shamim M, et al. (2016). Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. *Proc Natl Acad Sci U S A* 113, E4504–4512. [PubMed: 27432957]
- Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, and Ren B (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380. [PubMed: 22495300]

- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, and Gingeras TR (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. [PubMed: 23104886]
- Dowen JM, Fan ZP, Hnisz D, Ren G, Abraham BJ, Zhang LN, Weintraub AS, Schuijers J, Lee TI, Zhao K, et al. (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* 159, 374–387. [PubMed: 25303531]
- Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, and Aiden EL (2016). Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst* 3, 99–101. [PubMed: 27467250]
- Engreitz JM, Pandya-Jones A, McDonel P, Shishkin A, Sirokman K, Surka C, Kadri S, Xing J, Goren A, Lander ES, et al. (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* 341, 1237973. [PubMed: 23828888]
- Farcas AM, Blackledge NP, Sudbery I, Long HK, McGouran JF, Rose NR, Lee S, Sims D, Cerase A, Sheahan TW, et al. (2012). KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *eLife* 1, e00205. [PubMed: 23256043]
- Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, and Mirny LA (2016). Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep* 15, 2038–2049. [PubMed: 27210764]
- Fursova NA, Blackledge NP, Nakayama M, Ito S, Koseki Y, Farcas AM, King HW, Koseki H, and Klose RJ (2019). Synergy between Variant PRC1 Complexes Defines Polycomb-Mediated Gene Repression. *Mol Cell*.
- Giorgetti L, Lajoie BR, Carter AC, Attia M, Zhan Y, Xu J, Chen CJ, Kaplan N, Chang HY, Heard E, et al. (2016). Structural organization of the inactive X chromosome in the mouse. *Nature* 535, 575–579. [PubMed: 27437574]
- Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, and Glass CK (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38, 576–589. [PubMed: 20513432]
- Hentze MW, Castello A, Schwarzl T, and Preiss T (2018). A brave new world of RNA-binding proteins. *Nat Rev Mol Cell Biol* 19, 327–341. [PubMed: 29339797]
- Isono K, Endo TA, Ku M, Yamada D, Suzuki R, Sharif J, Ishikura T, Toyoda T, Bernstein BE, and Koseki H (2013). SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. *Dev Cell* 26, 565–577. [PubMed: 24091011]
- Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 467, 430–435. [PubMed: 20720539]
- Kalantry S, Mills KC, Yee D, Otte AP, Panning B, and Magnuson T (2006). The Polycomb group protein Eed protects the inactive X-chromosome from differentiation-induced reactivation. *Nat Cell Biol* 8, 195–202. [PubMed: 16415857]
- Kalb R, Latwiel S, Baymaz HI, Jansen PW, Muller CW, Vermeulen M, and Muller J (2014). Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nat Struct Mol Biol* 21, 569–571. [PubMed: 24837194]
- Kelsey AD, Yang C, Leung D, Minks J, Dixon-McDougall T, Baldry SE, Bogutz AB, Lefebvre L, and Brown CJ (2015). Impact of flanking chromosomal sequences on localization and silencing by the human non-coding RNA XIST. *Genome Biol* 16, 208. [PubMed: 26429547]
- Kirk JM, Kim SO, Inoue K, Smola MJ, Lee DM, Schertzer MD, Wooten JS, Baker AR, Sprague D, Collins DW, et al. (2018). Functional classification of long non-coding RNAs by k-mer content. *Nat Genet*.
- Kornienko AE, Guenzl PM, Barlow DP, and Pauler FM (2013). Gene regulation by the act of long non-coding RNA transcription. *BMC Biol* 11, 59. [PubMed: 23721193]
- Korostowski L, Sedlak N, and Engel N (2012). The Kcnq1ot1 long non-coding RNA affects chromatin conformation and expression of Kcnq1, but does not regulate its imprinting in the developing heart. *PLoS Genet* 8, e1002956. [PubMed: 23028363]

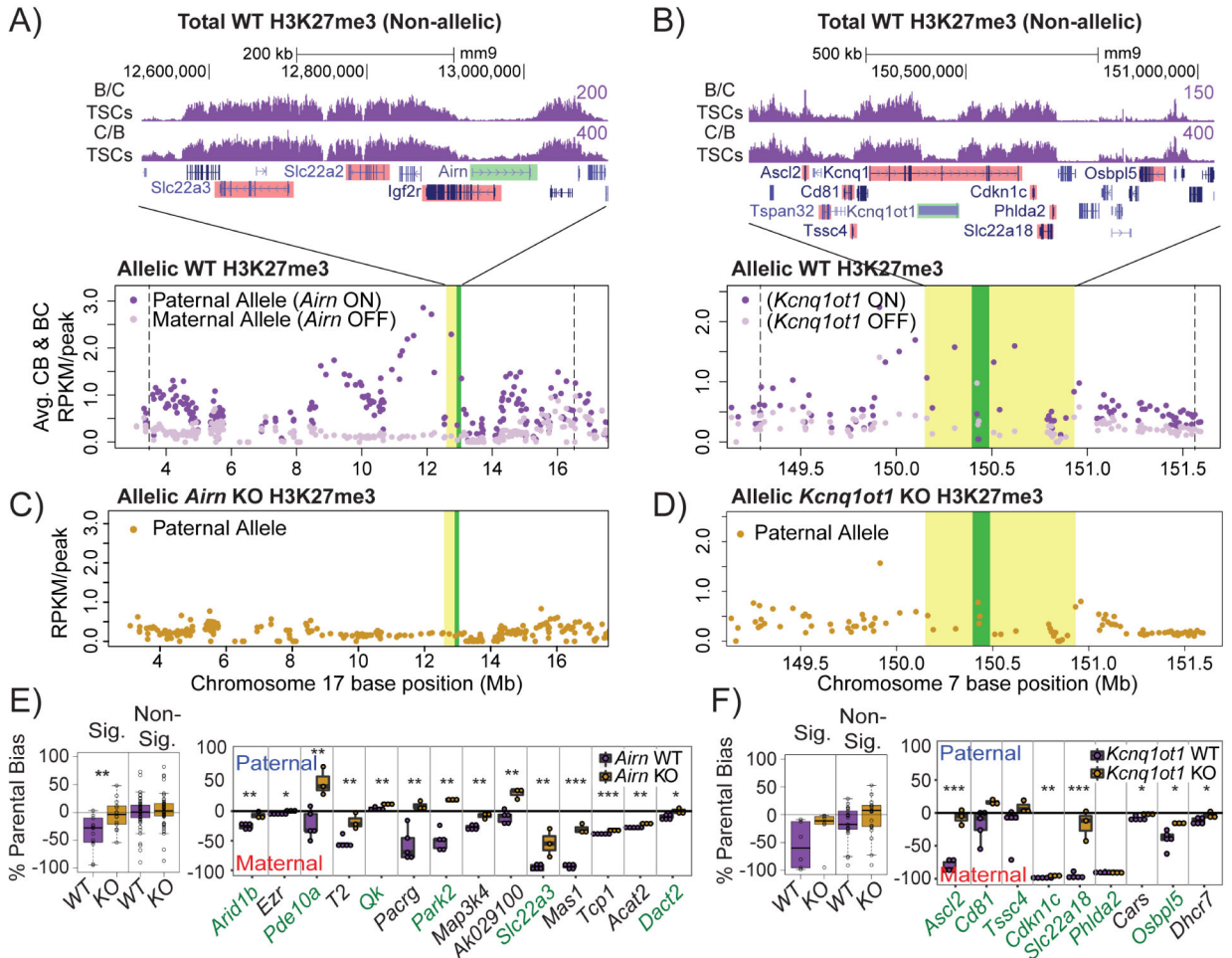


- Kung JT, Kesner B, An JY, Ahn JY, Cifuentes-Rojas C, Colognori D, Jeon Y, Szanto A, del Rosario BC, Pinter SF, et al. (2015). Locus-specific targeting to the X chromosome revealed by the RNA interactome of CTCF. *Mol Cell* 57, 361–375. [PubMed: 25578877]
- Langmead B, Trapnell C, Pop M, and Salzberg SL (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, R25. [PubMed: 19261174]
- Laprell F, Finkl K, and Muller J (2017). Propagation of Polycomb-repressed chromatin requires sequence-specific recruitment to DNA. *Science* 356, 85–88. [PubMed: 28302792]
- Latos PA, Pauler FM, Koerner MV, Senergin HB, Hudson QJ, Stocsits RR, Allhoff W, Stricker SH, Klement RM, Warczok KE, et al. (2012). Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science* 338, 1469–1472. [PubMed: 23239737]
- Lee JT, and Bartolomei MS (2013). X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell* 152, 1308–1323. [PubMed: 23498939]
- Lewis A, Green K, Dawson C, Redrup L, Huynh KD, Lee JT, Hemberger M, and Reik W (2006). Epigenetic dynamics of the Kcnq1 imprinted domain in the early embryo. *Development* 133, 4203–4210. [PubMed: 17021040]
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. [PubMed: 19505943]
- Li H, Liefke R, Jiang J, Kurland JV, Tian W, Deng P, Zhang W, He Q, Patel DJ, Bulyk ML, et al. (2017). Polycomb-like proteins link the PRC2 complex to CpG islands. *Nature* 549, 287–291. [PubMed: 28869966]
- Liao Y, Smyth GK, and Shi W (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. [PubMed: 24227677]
- Loda A, Brandsma JH, Vassilev I, Servant N, Loos F, Amiras A, Splinter E, Barillot E, Poot RA, Heard E, et al. (2017). Genetic and epigenetic features direct differential efficiency of Xist-mediated silencing at X-chromosomal and autosomal locations. *Nature communications* 8, 690.
- Luo Y, Lin L, Bolund L, and Sorensen CB (2014). Efficient construction of rAAV-based gene targeting vectors by Golden Gate cloning. *BioTechniques* 56, 263–268. [PubMed: 24806227]
- Lynch MD, Smith AJ, De Gobbi M, Flenley M, Hughes JR, Vernimmen D, Ayyub H, Sharpe JA, Sloane-Stanley JA, Sutherland L, et al. (2012). An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. *EMBO J* 31, 317–329. [PubMed: 22056776]
- Mancini-Dinardo D, Steele SJ, Levorse JM, Ingram RS, and Tilghman SM (2006). Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. *Genes Dev* 20, 1268–1282. [PubMed: 16702402]
- Marks H, Chow JC, Denissov S, Francoijs KJ, Brockdorff N, Heard E, and Stunnenberg HG (2009). High-resolution analysis of epigenetic changes associated with X inactivation. *Genome Res* 19, 1361–1373. [PubMed: 19581487]
- Marks H, Kerstens HHD, Barakat TS, Splinter E, Dirks RAM, van Mierlo G, Joshi O, Wang SY, Babak T, Albers CA, et al. (2015). Dynamics of gene silencing during X inactivation using allele-specific RNA-seq. *Genome Biology* 16.
- Mele M, and Rinn JL (2016). "Cat's Cradling" the 3D Genome by the Act of LncRNA Transcription. *Mol Cell* 62, 657–664. [PubMed: 27259198]
- Mendenhall EM, Koche RP, Truong T, Zhou VW, Issac B, Chi AS, Ku M, and Bernstein BE (2010). GC-rich sequence elements recruit PRC2 in mammalian ES cells. *PLoS Genet* 6, e1001244. [PubMed: 21170310]
- Meng L, Person RE, Huang W, Zhu PJ, Costa-Mattioli M, and Beaudet AL (2013). Truncation of Ube3a-ATS unsilences paternal Ube3a and ameliorates behavioral defects in the Angelman syndrome mouse model. *PLoS Genet* 9, e1004039. [PubMed: 24385930]
- Nozawa RS, Boteva L, Soares DC, Naughton C, Dun AR, Buckle A, Ramsahoye B, Bruton PC, Saleeb RS, Arnedo M, et al. (2017). SAF-A Regulates Interphase Chromosome Structure through Oligomerization with Chromatin-Associated RNAs. *Cell* 169, 1214–1227 e1218. [PubMed: 28622508]

- Oksuz O, Narendra V, Lee CH, Descostes N, LeRoy G, Raviram R, Blumenberg L, Karch K, Rocha PP, Garcia BA, et al. (2018). Capturing the Onset of PRC2-Mediated Repressive Domain Formation. *Mol Cell* 70, 1149–1162 e1145. [PubMed: 29932905]
- Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, Nagano T, Mancini-Dinardo D, and Kanduri C (2008). Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell* 32, 232–246. [PubMed: 18951091]
- Pintacuda G, Wei G, Roustan C, Kirmizitas BA, Solcan N, Cerase A, Castello A, Mohammed S, Moindrot B, Nesterova TB, et al. (2017). hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing. *Mol Cell* 68, 955–969 e910. [PubMed: 29220657]
- Pinter SF, Sadreyev RI, Yildirim E, Jeon Y, Ohsumi TK, Borowsky M, and Lee JT (2012). Spreading of X chromosome inactivation via a hierarchy of defined Polycomb stations. *Genome Res* 22, 1864–1876. [PubMed: 22948768]
- Quinlan AR, and Hall IM (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. [PubMed: 20110278]
- Raab JR, Smith KN, Spear CC, Manner CJ, Calabrese JM, and Magnuson T (2019). SWI/SNF remains localized to chromatin in the presence of SCHLAP1. *Nat Genet* 51, 26–29. [PubMed: 30510238]
- Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680. [PubMed: 25497547]
- Regha K, Sloane MA, Huang R, Pauler FM, Warczok KE, Melikant B, Radolf M, Martens JH, Schotta G, Jenuwein T, et al. (2007). Active and repressive chromatin are interspersed without spreading in an imprinted gene cluster in the mammalian genome. *Mol Cell* 27, 353–366. [PubMed: 17679087]
- Riising EM, Comet I, Leblanc B, Wu X, Johansen JV, and Helin K (2014). Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. *Mol Cell* 55, 347–360. [PubMed: 24999238]
- Robinson MD, McCarthy DJ, and Smyth GK (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. [PubMed: 19910308]
- Schertzer MD, Thulson E, Bracerros KCA, Lee DM, Hinkle ER, Murphy RM, Kim SO, Vitucci ECM, and J.M. C (2018). A piggyBac-based toolkit for inducible genome editing in mammalian cells. *bioRxiv*.
- Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat Methods* 9, 676–682. [PubMed: 22743772]
- Schwartz YB, and Pirrotta V (2013). A new world of Polycombs: unexpected partnerships and emerging functions. *Nat Rev Genet* 14, 853–864. [PubMed: 24217316]
- Simon JA, and Kingston RE (2013). Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. *Mol Cell* 49, 808–824. [PubMed: 23473600]
- Sleutels F, Zwart R, and Barlow DP (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* 415, 810–813. [PubMed: 11845212]
- Smeets D, Markaki Y, Schmid VJ, Kraus F, Tattermusch A, Cerase A, Sterr M, Fiedler S, Demmerle J, Popken J, et al. (2014). Three-dimensional super-resolution microscopy of the inactive X chromosome territory reveals a collapse of its active nuclear compartment harboring distinct Xist RNA foci. *Epigenetics Chromatin* 7, 8. [PubMed: 25057298]
- Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Scholer A, van Nimwegen E, Wirbelauer C, Oakeley EJ, Gaidatzis D, et al. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* 480, 490–495. [PubMed: 22170606]
- Starmer J, and Magnuson T (2016). Detecting broad domains and narrow peaks in ChIP-seq data with hiddenDomains. *BMC Bioinformatics* 17, 144. [PubMed: 27009150]

- Sunwoo H, Wu JY, and Lee JT (2015). The Xist RNA-PRC2 complex at 20-nm resolution reveals a low Xist stoichiometry and suggests a hit-and-run mechanism in mouse cells. *Proc Natl Acad Sci U S A* 112, E4216–4225. [PubMed: 26195790]
- Terranova R, Yokobayashi S, Stadler MB, Otte AP, van Lohuizen M, Orkin SH, and Peters AH (2008). Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. *Dev Cell* 15, 668–679. [PubMed: 18848501]
- Umlauf D, Goto Y, Cao R, Cerqueira F, Wagschal A, Zhang Y, and Feil R (2004). Imprinting along the Kcnq1 domain on mouse chromosome 7 involves repressive histone methylation and recruitment of Polycomb group complexes. *Nat Genet* 36, 1296–1300. [PubMed: 15516932]
- Woo CJ, Kharchenko PV, Daheron L, Park PJ, and Kingston RE (2010). A region of the human HOXD cluster that confers polycomb-group responsiveness. *Cell* 140, 99–110. [PubMed: 20085705]
- Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9, R137. [PubMed: 18798982]
- Zylicz JJ, Bousard A, Zumer K, Dossin F, Mohammad E, da Rocha ST, Schwalb B, Syx L, Dingli F, Loew D, et al. (2019). The Implication of Early Chromatin Changes in X Chromosome Inactivation. *Cell* 176, 182–197 e123. [PubMed: 30595450]

- *Airn* and *Kcnq1ot1* direct PRCs to multi-megabase domains in trophoblast stem cells
- *Airn*-induced H3K27me3 correlates w/DNA structure, RNA abundance, and PRC-bound CGIs
- Deletion of a single PRC-bound CGI caused a 4.5Mb loss of H3K27me3 in the *Airn* domain
- Like *Xist*, *Airn* and *Kcnq1ot1* require HNRNPK to spread H3K27me3 in domains

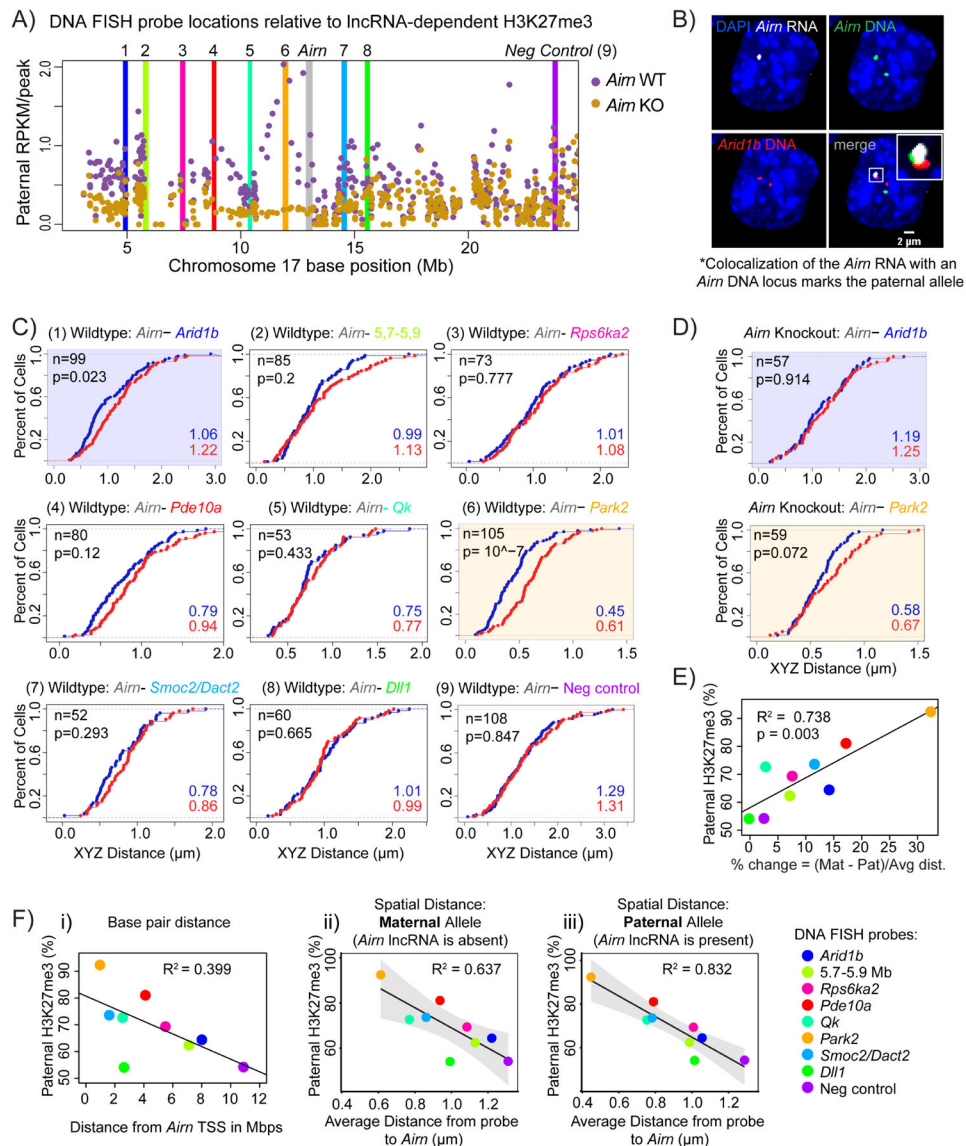


**Figure 1. Megabase-sized domains of H3K27me3 require continued expression of *Airn* and *Kcnq1ot1*.**

(A, B) Paternally-biased H3K27me3 surrounds *Airn* and *Kcnq1ot1* in TSCs. Dark/light purple dots, paternal/maternal signal in H3K27me3 peaks avgd. from C/B + B/C TSCs. Green bars, *Airn/Kcnq1ot1* loci. Dotted lines, first/last biased H3K27me3 peak used to define domain size (13 Mb for *Airn* and 2.3 Mb for *Kcnq1ot1*). Yellow shading/upper insets, non-allelic H3K27me3 data in previously-defined *Airn/Kcnq1ot1* domains relative to maternally/paternally-biased genes (pink/green, respectively). (C, D) Paternal signal in H3K27me3 peaks in truncation C/B TSCs (gold). Peaks are as in A, B. (E, F) Parent-of-origin expression in wild-type (purple) and truncation (gold) C/B TSCs. Paternal/maternal biases represented from 0 to 100; maternal values multiplied by -1. Value of 0, equal expression from both alleles. Value of 100, 100% expression from one allele. Green name, known lncRNA target. Asterisks, genes de-repressed in lncRNA truncation (\*\*\*,  $p < 0.001$ ; \*\*,  $p < 0.01$ ; \*,  $p < 0.05$ , two-tailed t-test). Sig./Non-Sig, avg. bias of de-repressed/non-target gene.

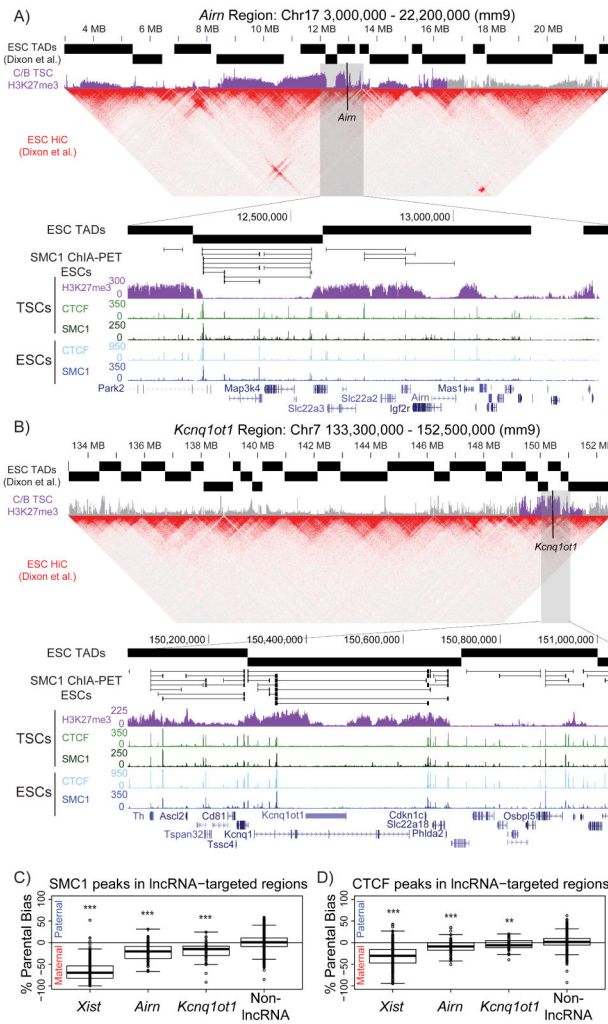
See also Figures S1, S2. Tables S1, S2, S6.





**Figure 2. Spread of H3K27me3 in the *Airn* domain is influenced by pre-existing genomic architecture and additional features on the paternal allele.**

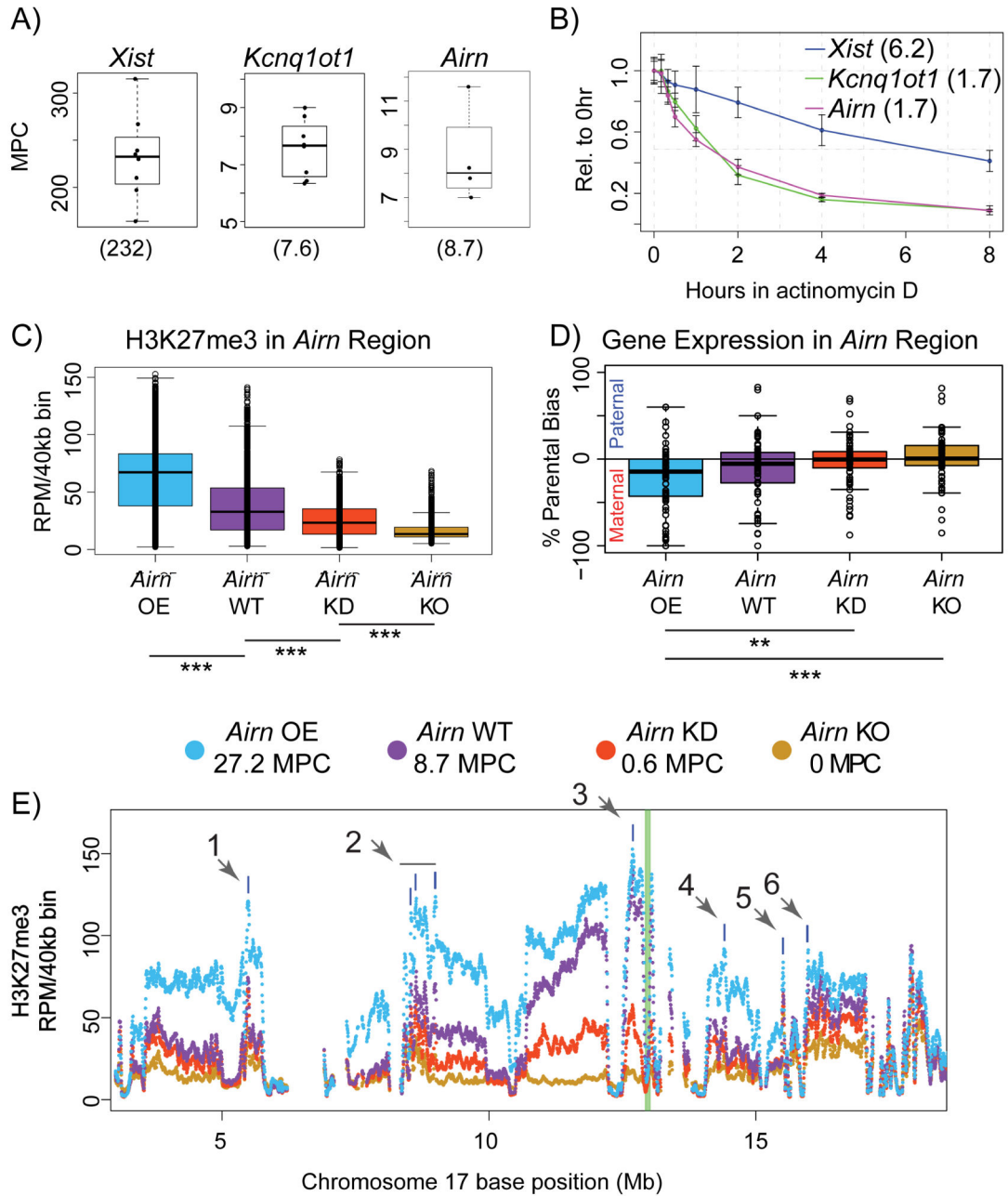
(A) DNA FISH probe location vs. paternal H3K27me3. (B) Representative FISH image. White box, signal overlap on paternal allele. Scale bar, 2  $\mu$ m. (C, D) Cumulative distribution plots for probes in (A). Spatial distance to *Airn* shown on paternal (blue) and maternal (red) alleles. Blue/red numbers, avg. distance on paternal/maternal alleles. n, # of cells. p-values from two-sample KS-tests. Shaded plots in (C; wild-type TSCs) correspond to those in (D; *Airn*-truncation TSCs). (E, F) Correlation between paternal H3K27me3 in regions probed for DNA FISH and (E) % avg. difference in distance between maternal and paternal alleles and (F): (i) distance in base pairs to *Airn* TSS, and avg. distance measured via FISH from probe to *Airn* on (ii) maternal and (iii) paternal alleles from (C). Grey in (ii) and (iii), 95% confidence intervals. See also Table S6.



**Figure 3. Intensity of H3K27me3 in lncRNA target domains correlates with TADs, DNA loops, and SMC1 and CTCF binding.**

(A, B) Hi-C data/TADs (Dixon et al., 2012), SMC1 loops (Downen et al., 2014), SMC1/CTCF binding (Kagey et al., 2010; Stadler et al., 2011) in ESCs, SMC1/CTCF binding in TSCs around *Airn* (A) and *Kcnq1ot1* (B). Purple, non-allelic H3K27me3 signal (C/B TSCs). H3K27me3 shading turns gray at last detected peak of paternally-biased H3K27me3. (C, D) Avg. parent-of-origin bias from C/B and B/C TSCs of (C) SMC1 and (D) CTCF peaks in *Xist*, *Airn*, and *Kcnq1ot1* target regions. \*\*\*,  $p < 0.001$ ; \*\*,  $p < 0.01$  relative to “Non-lncRNA” (all other autosomal peaks in genome); Tukey’s HSD test. Scales as in Figure 1E, F.

See also Figure S3, Table S6.



**Figure 4. LncRNA repressive potency correlates with abundance, stability, and underlying features of the genome.**

(A) Molecules per cell (MPC) of *Xist*, *Airn*, *Kcnq1ot1* by RNA-Seq. (B) Stability of *Xist*, *Airn*, *Kcnq1ot1* in TSCs after 5 $\mu$ g/ml of Actinomycin D. Mean  $\pm$ SD half-lives in parentheses. (C) Boxplots of H3K27me3 density in 40kb sliding bins across *Airn* domain, and (D) parent-of-origin expression for 61 considered genes (Figure 1) in *Airn*-overexpression (OE), -wild-type (WT), knockdown (KD), and truncation (KO) TSCs. \*\*\*, p < 0.001; \*\*, p < 0.01, Tukey's HSD test. Y-axis as in Figure 1E, F. (E) H3K27me3 density in 40kb sliding bins across *Airn* domain in OE, WT, KD, and KO TSCs. *Airn* MPC is above

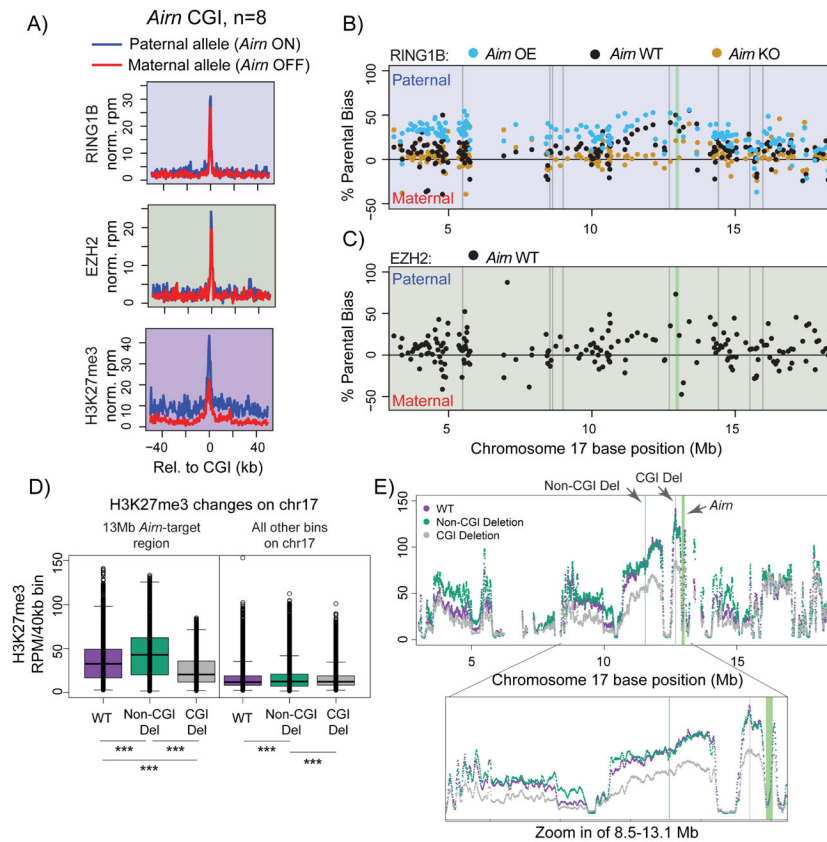
the density plot. Blue ticks + grey arrows, CGIs in Figure 5A. Green bar, *Airn* locus. “WT”, TSCs expressing dCas9-VP160 and a non-targeting sgRNA.  
See also Figures S4 and S5, Tables S3, S6, and S7.

Author Manuscript

Author Manuscript

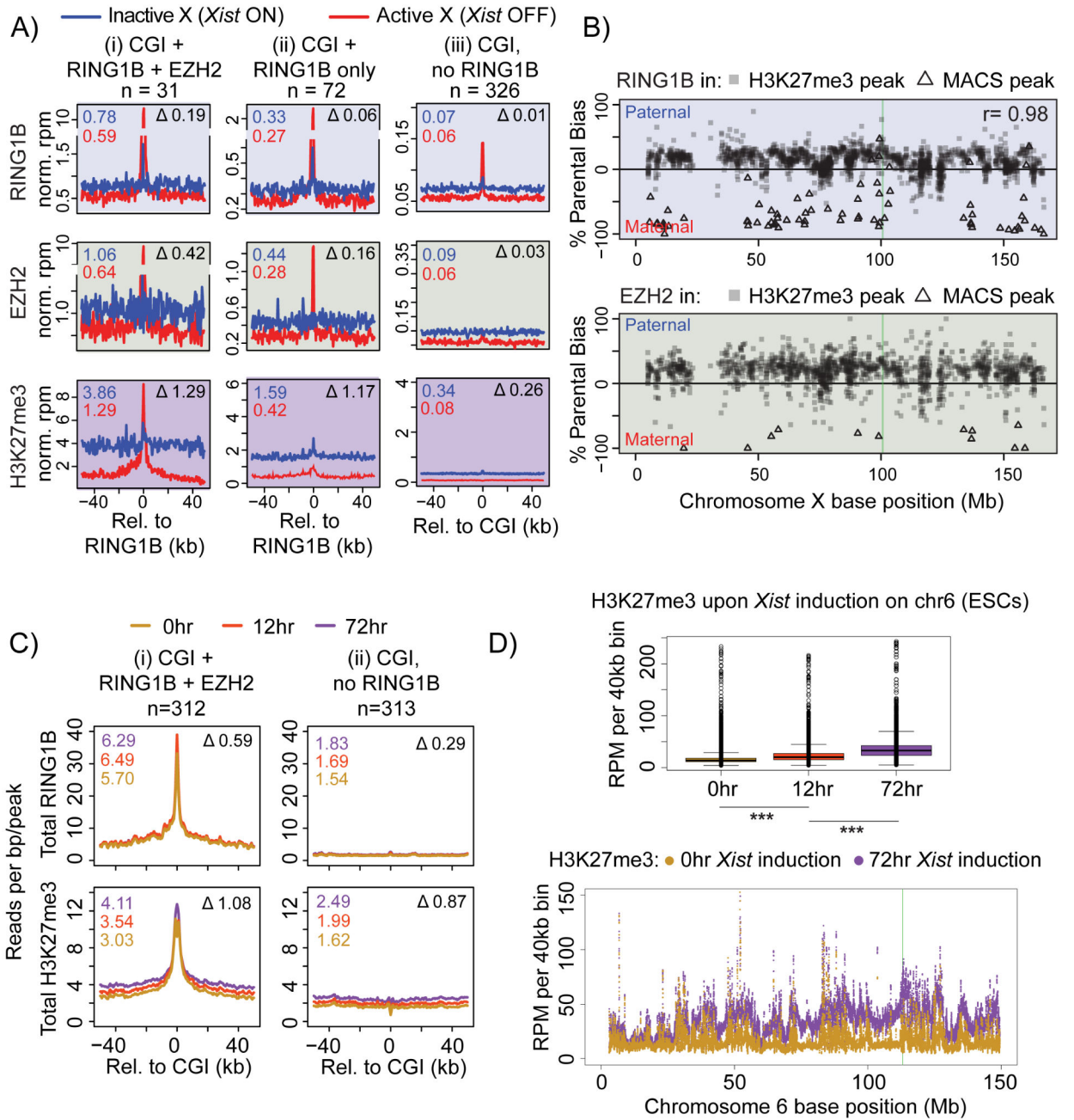
Author Manuscript

Author Manuscript



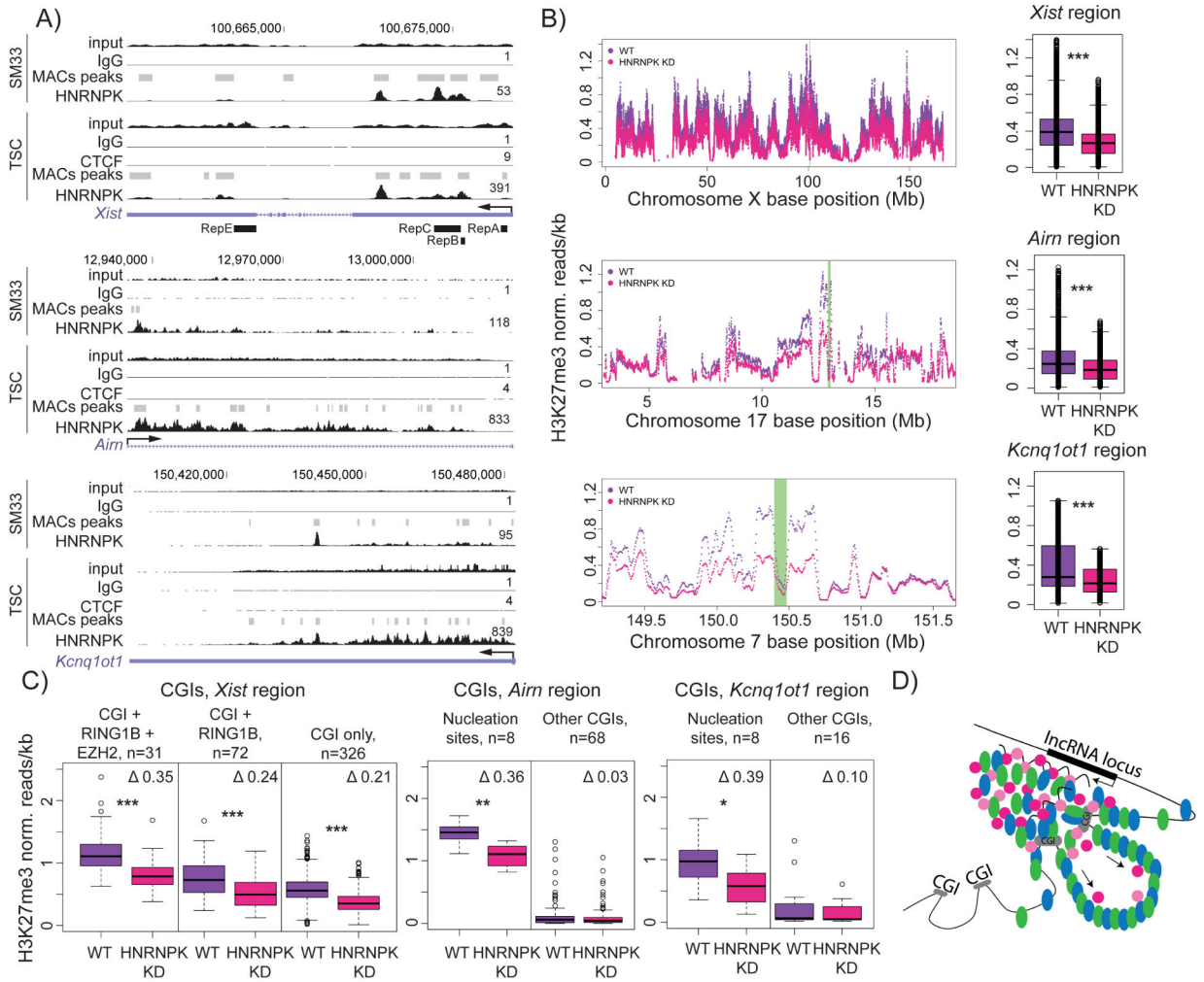
**Figure 5. CGIs bind PRCs autonomously and can nucleate spread of H3K27me3 by *Airn*.** (A) Allelic metagenes plots of RING1B, EZH2, H3K27me3 relative to 8 CGIs in H3K27me3 nucleation centers in *Airn* domain (i.e. blue ticks in Figure 4E). (B, C) Parent-of-origin bias in RING1B and EZH2 in peaks of H3K27me3 in *Airn* domain. Data from *Airn* wild-type (WT), overexpression (OE), and truncation (KO) TSCs shown. Green line, *Airn* locus. Grey lines, CGIs from in (A). Y-axis as in Figure 1E, F. Panels shaded in A-C for clarity. (D) Boxplot and (E) tiling plot of H3K27me3 density in 40kb bins sliding across *Airn* domain in WT, Non-CGI deletion, and CGI-deletion TSCs. (D) also shows non-*Airn* bins on chr17; note marginal increase in non-CGI relative to WT and CGI. \*\*\*,  $p < 0.001$ , Tukey's HSD test. Vertical lines in (E), *Airn*, Non-CGI, and CGI deletion location. For Non-CGI and CGI, data shown is avg. of two clones. See also Figure S6, Tables S3, S6.





**Figure 6. *Xist*-induced H3K27me3 density is highest around CGIs that bind PRCs autonomously.** (A) Metagenes of allelic RING1B, EZH2, and H3K27me3 density at CGIs that coincide w/ peaks of (i) RING1B and EZH2, (ii) RING1B only, or (iii) neither. Median signal on the active (red) and inactive X (blue) in the metagene window is shown in upper left, and difference in medians in upper right. Y-axes are broken in select plots to visualize trends on both X's. (B) Parent-of-origin bias of RING1B (top) and EZH2 (bottom) in peaks of H3K27me3 (squares) and peaks of RING1B and EZH2 (triangles). Green line, *Xist* locus. Y-axis as in Figure 1E, F. Pearson correlation between RING1B and EZH2 density in H3K27me3 peaks in upper right. Panels shaded in A+B for clarity. (C,D) H3K27me3

spreads from PRC-bound CGIs upon *Xist* induction in mouse ESCs. (C) Non-allelic RING1B and H3K27me3 density centered at chr6 CGIs bound by (i) RING1B+EZH2 or (ii) neither. Density is shown for three timepoints of *Xist* induction: 0hr, 12hr, and 72hr. Upper left, median at each timepoint. Upper right, difference in 72hr and 0hr medians. (D) Boxplot and tiling density of H3K27me3 across chr6 at 0hr (no *Xist* expression), 12hr, and 72hr *Xist* induction. 12hr tiling plot not shown for clarity. Green line, *Xist* insertion on chr6. \*\*\*,  $p < 0.001$ , Tukey's HSD test. See also Figures S7, Tables S4, S6.



**Figure 7. *Xist*, *Airn*, and *Kcnq1ot1* require HNRNPK to spread H3K27me3.**

(A) Wiggle tracks of RNA-IP data for input, IgG, and HNRNPK in SM33 ESCs + TSCs and CTCF in TSCs across *Xist*, *Airn*, and *Kcnq1ot1*. Blocks above HNRNPK tracks, MACS peaks. Right-justified numbers, signal over IgG. *Xist* repeats are below *Xist* diagram. (B) H3K27me3 tiling density and boxplots in *Xist*, *Airn*, and *Kcnq1ot1* domains in WT and HNRNPK knockdown TSCs. Green bars, lncRNA loci. \*\*\*,  $p < 0.001$ , two-tailed t-test. (C) Boxplots of H3K27me3 density  $\pm 2$  kb from CGI centers in *Xist*, *Airn*, and *Kcnq1ot1* domains. Difference in means between WT and knockdown, upper right corner. \*\*\*,  $p < 0.001$ ; \*\*,  $p < 0.01$ ; \*,  $p < 0.05$ , two-tailed t-test. “Nucleation sites” in *Airn/Kcnq1ot1* plots, CGIs from Figures 5A/S6A. (D) Model: Super-stoichiometric interactions between proteins such as HNRNPK (pink circles) that bind lncRNAs (squiggles) and PRC1/2 (blue/green ovals) concentrate PRCs in lncRNA foci. These same interactions tether lncRNA foci to CGIs (grey ovals) pre-bound by PRCs, nucleating PRC spread in contacted regions. See also Figure S7, Tables S5, S6.

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal H3K27me3	Abcam	Cat#ab6002
Rabbit monoclonal H3K27me3	Cell Signaling	Cat#9733
CTCF	gift from V.Lobanekov	N/A
Rabbit monoclonal EZH2	Cell Signaling	Cat#5246
Rabbit monoclonal RING1B	Cell Signaling	Cat#5694
Rabbit polyclonal SMC1	Bethyl Laboratories	Cat#A300-055A
Rabbit monoclonal H2AK119ub	Cell Signaling	Cat#8240
Mouse monoclonal HNRNPK	Abcam	Cat#ab39975
Mouse monoclonal HNRNPK	Santa Cruz	Cat#sc28380
Mouse IgG isotype control	Invitrogen	Cat#02-6502
Mouse monoclonal TBP	Abcam	Cat#ab818
Goat anti-mouse secondary, Alexa Fluor 488	Invitrogen	Cat#A-11029
Goat anti-rabbit secondary, Alexa Fluor 594	Invitrogen	Cat#A-11037
Donkey anti-mouse IgG-HRP secondary	Santa Cruz	Cat#sc-2314
Donkey anti-rabbit IgG-HRP secondary	Santa Cruz	Cat#sc-2313
Bacterial and Virus Strains		
Biological Samples		
Chemicals, Peptides, and Recombinant Proteins		
Heparin Sodium- Cell culture tested	Sigma	Cat#H3149
FGF4 Recombinant Human Protein	Invitrogen	Cat#PHG0154
ESGRO® Recombinant Mouse LIF Protein	Millipore Sigma	Cat#ESG1107
Critical Commercial Assays		
KAPA RNA HyperPrep Kits with RiboErase	Kapa Biosystems	Cat#KK8560
KAPA Stranded mRNA-seq Kits	Kapa Biosystems	Cat#KK8420
iTaq Universal SYBR Green	Biorad	Cat#1725121
BioPrime DNA Labeling System	Invitrogen	Cat#18094011
Superscript III Reverse Transcriptase	Invitrogen	Cat#18080044
High-Capacity cDNA Reverse Transcription Kit	Applied Biosystems	Cat#4368813
Apex Taq DNA Polymerase	Genesee Scientific	Cat#42-801B2
ERCC RNA Spike-in mix	Invitrogen	Cat#4456740
Deposited Data		
Raw and analyzed sequencing data	This study; See Table S6	GSE118402
Mouse reference genome NCBI build 37, NCBI37/mm9	Genome Reference Consortium	<a href="https://www.ncbi.nlm.nih.gov/grc/mouse">https://www.ncbi.nlm.nih.gov/grc/mouse</a>
Variant sequence data	Sanger Institute	<a href="http://www.sanger.ac.uk/resources/mouse/genomes/">http://www.sanger.ac.uk/resources/mouse/genomes/</a>
Microscopy images, FISH and IF	This study; Mendeley Data	<a href="http://dx.doi.org/10.17632/bv9y5rcpzz.1">http://dx.doi.org/10.17632/bv9y5rcpzz.1</a> <a href="http://dx.doi.org/10.17632/nk84zzwjkh.1">http://dx.doi.org/10.17632/nk84zzwjkh.1</a>

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Western blot	This study; Mendeley Data	<a href="http://dx.doi.org/10.17632/bv9y5rcpz.1">http://dx.doi.org/10.17632/bv9y5rcpz.1</a> <a href="http://dx.doi.org/10.17632/nk84zzwjkh.1">http://dx.doi.org/10.17632/nk84zzwjkh.1</a>
Experimental Models: Cell Lines		
mouse: C/B and B/C trophoblast stem cells	Calabrese lab	NA
mouse: C/B and B/C cortical neurons	Zylka lab	NA
mouse: SM33 embryonic stem cells	Plath lab	PMID: 23828888
mouse: <i>Xist</i> RMCE embryonic stem cells	Calabrese lab	D.M.L. and J.M.C. in press
Experimental Models: Organisms/Strains		
Oligonucleotides		
Primers	This study	See Table S7
Recombinant DNA		
PGK-neo vector	Addgene	Cat#51422; PMID: 24806227
pX330 vector	Addgene	Cat#42230; PMID: 23287718
PB_tre_dCas9_VP160 vector	Calabrese lab	Cat#126031; <a href="https://doi.org/10.1101/448803">https://doi.org/10.1101/448803</a>
PB_tre_dCas9_KRAB vector	Calabrese lab	Cat#126030; <a href="https://doi.org/10.1101/448803">https://doi.org/10.1101/448803</a>
PB_tre_Cas9 vector	Calabrese lab	Cat#126029; <a href="https://doi.org/10.1101/448803">https://doi.org/10.1101/448803</a>
PB_rtTA_BsmBI vector	Calabrese lab	Cat#126028; <a href="https://doi.org/10.1101/448803">https://doi.org/10.1101/448803</a>
pUC19-piggyBac transposase	Calabrese lab	PMID: 30224646
BAC for <i>Arid1b</i> probe	BACPAC Resources	Cat#RP23-223M13
BAC for 5.7-5.9Mb probe	BACPAC Resources	Cat#RP23-90J22
BAC for <i>Rps6ka2</i> probe	BACPAC Resources	Cat#RP23-457G11
BAC for <i>Pde10a</i> probe	BACPAC Resources	Cat#RP23-291O1
BAC for <i>Qk</i> probe	BACPAC Resources	Cat#RP23-338H10
BAC for <i>Park2</i> probe	BACPAC Resources	Cat#RP23-136N13
BAC for <i>Smoc2/Dact2</i> probe	BACPAC Resources	Cat#RP23-104G1
BAC for <i>Dll1</i> probe	BACPAC Resources	Cat#RP23-460P10
BAC for Neg Control probe	BACPAC Resources	Cat#RP23-343J15
BAC for <i>Airn</i> DNA probe	BACPAC Resources	Cat#RP23-309H20
Fosmid for <i>Airn</i> RNA	BACPAC Resources	Cat#W11-2156F18
Fosmid for <i>Igf2r</i> RNA	BACPAC Resources	Cat#W11-662A5
Software and Algorithms		
AutoQuantX3 deconvolution algorithm	Mediacy	<a href="http://www.mediacy.com/autoquantx3">http://www.mediacy.com/autoquantx3</a>
Imaris (8.3.1)	Bitplane	<a href="http://www.bitplane.com/imaris">http://www.bitplane.com/imaris</a>
FIJI (ImageJ)	Schindelin et al., 2012	<a href="https://imagej.net/Downloads">https://imagej.net/Downloads</a>
STAR (2.6.0a)	Dobin et al., 2013	<a href="https://github.com/alexdobin/STAR">https://github.com/alexdobin/STAR</a>
Bowtie2 (2.3.4)	Langmead and Salzberg, 2012	<a href="http://bowtie-bio.sourceforge.net/bowtie2/index.shtml">http://bowtie-bio.sourceforge.net/bowtie2/index.shtml</a>
Samtools (1.9)	Li et al., 2009	<a href="https://github.com/samtools/samtools">https://github.com/samtools/samtools</a>
R (3.4.3)	R Core Team	<a href="https://www.r-project.org/">https://www.r-project.org/</a>
edgeR (R package)	Robinson et al., 2010	<a href="https://bioconductor.org/packages/release/bioc/html/edgeR.html">https://bioconductor.org/packages/release/bioc/html/edgeR.html</a>



REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bedtools (2.26)	Quinlan and Hall, 2010	<a href="https://bedtools.readthedocs.io/en/latest/">https://bedtools.readthedocs.io/en/latest/</a>
hiddenDomains (3.0)	Starmer and Magnuson, 2016	<a href="https://sourceforge.net/projects/hiddendomains/">https://sourceforge.net/projects/hiddendomains/</a>
MACS2 (2.1.2)	Zhang et al., 2008	<a href="https://github.com/taoliu/MACS/">https://github.com/taoliu/MACS/</a>
Juicebox	Durand et al., 2016	<a href="https://github.com/aidenlab/Juicebox/">https://github.com/aidenlab/Juicebox/</a>
HOMER (4.10)	Heinz et al., 2010	<a href="http://homer.ucsd.edu/homer/">http://homer.ucsd.edu/homer/</a>
Subread- featureCounts (1.6.3)	Liao et al., 2014	<a href="http://subread.sourceforge.net/">http://subread.sourceforge.net/</a>
Other		
Vibracell ultrasonic processor	Sonics	Model#VCX130
Neon Transfection System	Invitrogen	Cat#MPK5000
Trizol Reagent	Invitrogen	Cat#15596018
Protein A/G PLUS agarose beads	Santa Cruz	Cat#sc-2003
Agencourt AMPure XP beads	Beckman Coulter	Cat#A63880
Stellaris FISH probes, Mouse <i>Gapdh</i> , Quasar 670	Biosearch Technologies	Cat#SMF-3140-1
Stellaris FISH probes, Custom Mouse <i>Aim</i> , Quasar 570	Biosearch Technologies	Stellaris RNA FISH Probe Designer
Stellaris FISH probes, Custom Mouse <i>Xist</i> , Quasar 670	Biosearch Technologies	Stellaris RNA FISH Probe Designer
Streptavidin Alexa Fluor 555 Conjugate	Invitrogen	Cat#S32355
Prolong Gold Antifade Mountant	Thermo Fisher	Cat# P10144
Vectashield Antifade Mounting Media	Vector Laboratories	Cat#H-1000
TetraSpeck™ Microspheres, 0.2 μm beads	Thermofisher	Cat#T7280