EPIGENETIC REGULATION OF GENE EXPRESSION BY LONG NON-CODING RNAS

Aki Keean Braceros

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biochemistry in the University of North Carolina at Chapel Hill.

Chapel Hill
2023

Approved by:

J. Mauro Calabrese

Greg G. Wang

Brian D. Strahl

Jill M. Dowen

Daniel J. McKay

Douglas H. Phanstiel

ABSTRACT

Aki Keean Braceros: Epigenetic Regulation of Gene Expression by Long Non-coding RNAs
(Under the direction of J. Mauro Calabrese)

Genomic imprinting is an epigenetic regulatory process of gene expression that can be controlled by repressive long non-coding RNAs (lncRNAs). The expression of lncRNAs leads to the recruitment of Polycomb Repressive Complexes (PRCs) over specific regions of the genome that cause changes to the chromatin landscape resulting in gene silencing. The lncRNA *Xist* is the best characterized for its role in recruiting PRCs and silencing an entire X chromosome, and a handful of others, including the imprinted lncRNA *Airn*, function similarly over smaller autosomal regions albeit still over multiple megabases (Mbs) of chromatin. Still, the mechanisms of lncRNA-mediated repression are unclear, and questions remain whether lncRNA function is attributed to its act of transcription or lncRNA product.

This dissertation discusses novel insights that address this debate from evidence based on *Airn* and its repressed domain in F1-hybrid mouse cell lines, owing to its mono-allelic expression and *cis*-acting nature. First, we reveal how long-distance repression by *Airn* is dependent on specific DNA regulatory elements that facilitate three-dimensional proximity of the lncRNA's gene and RNA product to its target loci. Second, we introduce new understandings into the repressive function and potency of *Airn* through direct comparisons with that of *Xist* under ectopic chromosomal conditions. Finally, we summarize these gained perspectives, which argue in favor of a model whereby *Airn* is a functional lncRNA product that induces repression over long distances, and further highlight intriguing observations from this work worth investigating

going forward. The paradigms we established on the epigenetic mechanisms of *Airn*-mediated repression inspire future studies on repressed domains elsewhere in the genome that are likely regulated by lncRNAs or locus control elements.

ACKNOWLEDGEMENTS

The support I received over the last six years has played an integral role in the completion of this dissertation. Firstly, my dissertation advisor, Mauro Calabrese, provided me a lot of guidance and a work home to grow both scientifically and professionally. I am a better scientist today and tomorrow because of him. I would especially like to acknowledge Mauro's empathy; he was always accommodating to my overall health by frequently checking in and allowing me the necessary opportunities of self-care, especially during really tough periods, including my long-distance relationship, the COVID-19 pandemic, and writing grants/manuscripts. He also supported me in pursuing a 10-week internship that ultimately shaped my career path. Thank you, Mauro, for believing in me and for being a wonderful advisor.

My work home would not have felt like a home if it was not for my lab colleagues. Past members, Megan Schertzer and David Lee, were critical factors that influenced my decision to join Mauro's lab. Thank you to Megan for taking me under her wing and handing down a project I became passionate about. None of this work would have found its successes without her foundational contributions. I am happy that we continue to stay in touch and bond over Pizza Hut. Thank you to David for being like an older brother. His welcoming quality instantly made me feel a part of the lab. A huge acknowledgement to Rachel Cherney and Jackson Trotman. It brought me so much security to adventure through this Ph.D. with Rachel. We joined the lab together and now we are leaving the lab together. I am going to miss having her as my work bestie. Jackson is also like a big brother to me. I truly appreciate his willingness to listen to me vent and rant about everything, as well as discuss ideas and topics no matter how basic or

complex. He is truly every person's dream coworker. Thank you to the rest of the Calabrese lab members: Mickey Murvin, Elizabeth Abrash, Sam Boyson, Quinn Eberhard, and Ryan Salcido. I enjoyed our daily interactions of laughs and sharing "tea" but also watching them develop their scientific personalities. The lab is in good hands. Thank you also to the rotation students that overlapped with me. I am happy to see them all flourish from the other side. There will be no other work environment like the Calabrese lab, and I will truly miss everyone.

In addition to Mauro's mentorship, I am honored to have received guidance from some of the most terrific scientists and human beings. I owe a huge thank you to my thesis committee: Greg Wang, Brian Strahl, Jill Dowen, Dan McKay, and Doug Phanstiel. When I started grad school at UNC, I envisioned instantly who would be my committee members, and I am lucky to say that I was able to have my dream team. The incredible feedback and discussions on my dissertation projects over the years strongly influenced the impact of the work, and I am honored to be able to forever say I was mentored by these pioneers. My former advisors, Jeff Brown (Undergrad), Ed Valente (Undergrad), and Toshi Tsukiyama (Tech), have all inspired and directed me to this Ph.D. path: Jeff introduced the world of the cell to me, Ed taught me the value and excitement of computer data science, and Toshi presented my current passion in genomics. I was also heavily influenced by former colleagues during my time in Toshi's lab: Laura Hsieh, Sarah Swygert, Jeff McKnight, Laura McKnight, and Sam Cutler. I owe these incredible individuals so much thanks for encouraging me to pursue research science and helping me identify my potential. I would also like to acknowledge Pairwise (my internship company), especially Gina Pham, Justin Fear, and Graham Johnson. Gina is the reason I discovered Pairwise and made my internship possible. I am so grateful to have had the opportunity to work in her Genomics team and network with so many fun and talented professionals across the

company. Justin was just an absolute joy to work under. I admire his immense talents, and I truly appreciate all that he has taught me to become a better bioinformatician. Graham was my co-mentor alongside Justin and gave me a great project that allowed me to really flourish with my genomics and bioinformatics knowledge. Graham was not only my coworker but is also a friend, even before joining Pairwise, so I was super lucky to be able to work with him. Thank you to Pairwise for giving me a really valuable experience.

Outside of lab, my dearest friends promoted the "life" aspect of my work-life balance. Sarah Barr and I went to undergrad together and even worked in the same research lab, but it was not until entering the Ph.D. program together that birthed the beginning of our friendship, which ultimately flourished into an everlasting one. She is one of the most dynamic people I know and is a motivational go-getter. I am honored to call her one of my best friends and to have had her officiate my marriage ceremony. I will miss our frequent Girls Nights, late-night fast-food runs, and immense number of laughs. I would also like to highlight the following people, who, despite not living in North Carolina (anymore), provided me endless love and a home when I needed a quick escape: Ali Wrede, Carlos Patiño Descovich, Laura Pellicer, and Rebecca Parks.

The last 3 years of my Ph.D./life has been impacted drastically since meeting my love, Hendric Vogel. I remain speechless when it comes to expressing how much love and appreciation I have for him. We spent our first year long-distance and dedicated many hours, emotions, and activities to keep our relationship strong. Since moving to North Carolina, he provides me many acts of love that keep me marching forward: from cooking meals for me to giving me massages to running errands to telling me "I love you" every day. I am just so incredibly grateful to call him my husband. I look forward to our future of many adventures together.

Finally, I would like to thank my family: Mom, Dad, Ami, Dawn, and Obaachan. They have supported my goals and dreams from the get-go. My parents and sisters never questioned my ambitions despite being thousands of miles away and did whatever it took to give me the wings to fly. I am soaring now because of them. A special shoutout to my Obaachan (grandma). Even though it has been 10 years since she has passed, she continues to live brightly in my heart and mind every single day. She raised and taught me my stubbornness and fight for happiness. I miss her daily and cannot wait to continue showing her the rest of the world. I dedicate this dissertation to my family.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

3D	Three-dimensional
Airn	*Antisense Igf2r RNA* (lncRNA)
Bp	Base pair
B/C	Black 6 mother crossed with Castaneous father
C57BL/6J	Black 6 (mouse strain)
(d)Cas9	(Catalytically dead) CRISPR-associated protein 9
CAST/EiJ	Castaneous (mouse strain)
C/B	Castaneous mother crossed with Black 6 father
CGI	CpG island
ChIP	Chromatin immunoprecipitation (method)
CHART	Capture Hybridization Analysis of RNA Targets (method)
CRISPR	Clustered regularly interspaced short palindromic repeat
ESC	(Mouse) Embryonic stem cell
FISH	Fluorescence *in situ* hybridization (method)
H2AK119ub	Mono-ubiquitylated lysine 119 of histone H2A
H3K27ac	Acetylated lysine 27 of histone 3
H3K27me3	Tri-methylated lysine 27 of histone H3
H3K4me2	Di-methylated lysine 4 of histone H3
H3K9me1/2/3	Mono-/di-/tri-methylated lysine 9 of histone H3
H-E	Highly-expressing
Hi-C	Chromatin capture conformation sequencing (method)
irMEFs	Irradiated mouse embryonic fibroblasts
Kb	Kilobase (pair)

KD	Knockdown
lncRNA	Long non-coding RNA
MAT	Maternal
Mb	Megabase (pair)
mRNA	Messenger RNA
PAT	Paternal
PRC1	Polycomb Repressive Complex 1
PRC2	Polycomb Repressive Complex 2
PRCs	Polycomb Repressive Complexes
qPCR	Quantitative polymerase chain reaction
RBP	RNA-binding protein
RP(K)M	Reads per (kilobase per) million
RT	Reverse transcription
rtTA	Reverse-tetracycline transactivator
sgRNA	Single guide RNA
Trunc.	*Airn* truncation
TSC	(Mouse) Trophoblast stem cell
WT	Wildtype
Xa	Active X chromosome
XCI	X chromosome inactivation
Xi	Inactive X chromosome
Xist	*X Inactivate Specific Transcript* (lncRNA)

CHAPTER 1: Introduction

1.1 Genomic imprinting and lncRNAs

Regulation of gene expression is an essential process that promotes proper cell identity and tissue homeostasis. This is especially critical during early development of the embryo, which requires precise coordination of gene expression networks to either maintain pluripotency or drive differentiation [1,2]. Genomic imprinting is an epigenetic gene regulatory process in which the dosage and preferential expression of ~150 mammalian genes is determined by the parent-of-origin inheritance of chromosomes and is highly conserved in placental mammals [3–5]. The genes subject to genomic imprinting play important roles in development and adult tissue homeostasis. Thus, imprinting defects can lead to various developmental disorders, congenital syndromes, and even increased risks for cancer [3]. Given these severe risks, genomic imprinting must be properly controlled and executed. A notable feature of imprinted genes is that they tend to exist in clusters throughout the genome called imprinted domains, which allow the expression states of clustered genes to be synchronously regulated. Such a coordinated process are mediated by long non-coding RNAs (lncRNAs) [6–10].

LncRNAs are a class of RNAs that are greater than 200 nucleotides in length, do not code for protein, and make up a major portion of the transcribed genome [7,11]. Like their messenger RNA (mRNA) counterpart, lncRNAs are transcribed by RNA Polymerase II (Pol II) and usually processed with 5'-end capping and 3'-end polyadenylation. However, in contrast to mRNAs, lncRNAs are spliced less efficiently, are shorter lived, and found predominantly retained in the nucleus [7,11]. For these reasons, the identification and characterization of lncRNAs and their

functions remain largely ambiguous, and current knowledge is limited to only a handful of annotated lncRNAs. Nonetheless, from these studied RNAs, it is clear that lncRNAs have roles in the nucleus that involve regulation of gene expression, and genomic imprinting is a major model through which our understanding of lncRNA-mediated gene regulation is based on [3,7–11]. In this regard, a common imprinting mechanism used by lncRNAs to regulate gene expression is through the recruitment of the Polycomb Represses Complexes (PRCs) to repress their respective imprinted domains [3,8–10,12–14].

1.2 Polycomb Repressive Complexes

The Polycomb Repressive Complexes 1 and 2 (PRC1 and PRC2) are highly conserved multi-subunit enzymes whose general function is to cooperatively repress gene expression by creating repressive chromatin environments. The PRCs are prominently recognized for their critical roles in embryonic development and stress response across multiple kingdoms of life. PRC1 and PRC2 are canonically known to mono-ubiquitinate lysine 119 of histone H2A (H2AK119ub1) and tri-methylate lysine 27 of histone H3 (H3K27me3), respectively. However, their ability to physically organize both local and three-dimensional (3D) genome architecture has gained greater attention for its importance in PRCs' full repressive potential [15,16]. Importantly, these functions by PRCs can achieve both reversible and permanent forms of gene silencing, which give cells and tissues the dynamic ability to regulate gene expression in a physiological context-dependent manner [12,13,16–18].

While the core functions of PRCs are conserved in various cell types and across species, PRCs take on diverse forms of subcomplexes defined by auxiliary factors that associate with and modulate the activity of core subunits, rendering specific and context-dependent functions [12,13,16–

[18]. For PRC1, the core components comprise of a RING1A or RING1B subunit partnered with a specific PCGF homologue that partitions PRC1 into two major subclasses: canonical PRC1 (cPRC1) and variant PRC1 (vPRC1). cPRC1 contains either PCGF2 or PCGF4 as its core PCGF subunit and the accessory subunits comprise of one of five chromobox family proteins (CBX2, CBX4, CBX6, CBX7, and CBX8) and one of three polyhomeotic-like proteins (PHC1-3). The exact CBX and PHC homologues included are dependent on their cell type-specific expression. The main role of cPRC1 is to compact chromatin structure through its CBX and PHC subunits, which serve to populate PRC interactions over large distances, bringing together target genes for concentrated repression and preventing chromatin accessibility to transcriptional activating machinery [15,16,19–21]. On the other hand, vPRC1 is assembled by the core subunit PCGF1, PCGF3, PCGF5, or PCGF6 and is further defined by the accessory subunit RYBP or YAF2. In contrast to cPRC1, vPRC1 does not appear to have chromatin organization activity but instead possesses stronger catalytic activity in H2AK119ub modification as its primary function, which is stimulated by RYBP and YAF2 [20,22,23].

For PRC2, the core subunits are EZH1/2, EED, SUZ12, and RBBP4/6. Like PRC1, PRC2 can be partitioned into two mutually exclusive subgroups in mammals: PRC2.1 and PRC2.2. PRC2.1 is comprised of different accessory subunit combinations of one PCL protein (PHF1/PCL1, MTF2/PCL2, and PHF19/PCL3) and either EPOP or RALI, whereas PRC2.2 contains JARID2 and AEBP2.The major roles of the accessory subunits for PRC2 subcomplexes are to promote recruitment of both PRC2 and PRC1 forms and to enhance the catalytic activity of PRC2 at specific sites in the genome in both synergistic and independent manners [12,16,24–26]. PRC2 has been implicated in chromatin organization; however, it is likely in an indirect manner through its role in recruiting cPRC1 through its H3K27me3 mark [15,16,27].

The subunits within PRC1 and PRC2 are also involved in their recruitment to the genome. Specifically, they can recognize their own and each other's histone modification that, in turn, propagates their recruitment and spread [12,18,28]. While this is the case, in other specific regions of the genome, such as imprinted domains that are often thousands and millions of base pairs in size and require large-scale repression, PRC recruitment is dependent on lncRNAs [12]. The greatest and most characterized of these examples is the lncRNA *Xist* in X chromosome inactivation (XCI), which is an early development process that can occur in either random or imprinted fashion [8–10,12,29]. *Xist* orchestrates the chromosome-wide transcriptional silencing and recruitment of PRCs to stably repress the entire X chromosome (chrX) in female placental mammals, which equalizes the dosage of chrX gene expression between XX females and XY males [12,29–31]. Behind *Xist*, another lncRNA called *Airn* in mice has come into light in the last decade as a potent gene silencer with the capacity to recruit the PRCs to chromatin over a 15-Mb domain on chr17 in extraembryonic tissues, the largest autosomal region known to date that is repressed by a lncRNA [12,32]. To this regard, *Xist* is undoubtedly an outlier in its repressive potency, thus comparisons with *Airn* and its imprinted domain are critical to provide mechanistic insights that can be used to build a better generalizable model for lncRNA repressive function.

1.3 Recruitment of PRCs by *Xist* and *Airn* lncRNAs

Both PRC1 and PRC2 are essential for proper XCI to occur, and their widespread recruitment to the future inactive X (Xi) is dependent on the *Xist* lncRNA [12,31]. *Xist* (*X* Inactivate Specific Transcript) is the exemplar for repressive lncRNAs and has been a consistent pioneer model for understanding lncRNA function. The *Xist* gene produces an approximately 18-Kb transcript that is abundant, spliced, 5'-end capped, 3'-end polyadenylated, and relatively stable

for a lncRNA. Despite these properties, *Xist* harbors low overall sequence conservation, except for a series of unique tandem repeats (referred to as Repeats A-F) that render its silencing function through high-affinity interactions with RNA-binding proteins (RBPs) [12,33,34]. The complete identity of RBPs and their cognate Repeat-binding site(s) in *Xist* remains under heavy investigation. Nonetheless, among the *Xist* Repeat-RBP interactions known, both mouse *Xist* and human *XIST* interact with HNRNPK via their Repeats B/C and Repeats B/D, respectively, to recruit specific forms of vPRC1 to the chrX, which, in turn, drives the cooperative and hierarchical recruitment of other PRCs [12,31,33–39]. Notably, Repeat A is also implicated in the early recruitment and complete spreading of PRCs over the Xi, but the mechanisms appear to be independent of HNRNPK and are still unclear [40–42]. Future studies are needed to determine the mechanisms through which PRC2 is recruited to Xi. Overall, *Xist* is an outlier with its ability to recruit PRCs over a 165-Mb chromosome, and only a handful of other annotated lncRNAs have been implicated in PRC recruitment but to much smaller extents. Until recently, the lncRNA *Airn* has been shown to recruit PRCs and silence genes over a 15-Mb domain in mouse extraembryonic tissues, which is significantly more potent than originally determined [12,32,37,43,44].

Airn (antisense *Igf2r* RNA) is an imprinted lncRNA gene whose expression comes solely from the paternally inherited allele of chr17 in mice. The gene is predicted as 90 to 120 Kbs in length and is antisense transcribed via an unmethylated CpG island (CGI) element located within intron 2 of a protein-coding gene called *Igf2r* [4,12,45,46]. Indeed, the act of transcription over the *Airn* gene causes transcriptional interference of *Igf2r* rendering its imprinted, repressed state. This mechanism of repression sparked a debate if repression of other surrounding genes associated with *Airn* expression are a result of transcriptional interference as well (discussed more below).

The *Airn* gene and its associated repressive function is poorly conserved, and to our current knowledge, is specific to rodents only. Still, efforts in other species have been made to identify an *Airn* orthologue that confers repressive activity. Yotova *et al.* investigated the CGI that is also found in the intron 2 of human *IGF2R* gene, which is an important regulatory element that marks the *Airn* promoter of mouse *Igf2r*. While they identified promoter activity from this CGI that also produces an antisense lncRNA, the lncRNA does not extend over the *IGF2R* promoter for transcriptional interference, nor does it result in imprinted expression of surrounding genes [12,47]. In marsupials, *Igf2r* exhibits imprinted expression patterns yet lacks a CGI in its intron 2 and instead contains one in intron 12. While an antisense lncRNA is produced from this CGI, the transcript is less than 1 Kb in size, thus significantly smaller than mouse *Airn* and does not overlap the *Igf2r* promoter [12,48] More recently, new evidence for the *IGF2R* domain in cattle has been shown to exhibit similar imprinted expression patterns to that of mouse and might involve an *Airn*-like gene activity [49]. This opens the possibility of *Airn* being conserved in other kingdoms of life that have yet to be studied. Until then, mouse *Airn* is an intriguing model beyond *Xist* for understanding mechanisms of lncRNA-mediated repression and their relationships with PRCs.

Airn's RNA product is unusual: it is detected at low copies with variable lengths up to 90 Kbs, 5'-end capped but inconsistently 3'-end polyadenylated, short-lived, predominantly unspliced, and detected exclusively near its site of transcription in the nucleus [12,37,50,51]. Despite these unstable properties, expression of *Airn* is associated with an increase in PRC-modified chromatin and repression of surrounding genes [12,52]. How *Airn* recruits PRCs to its targeted domain remains unclear. Furthermore, the extent of *Airn*-induced gene silencing is tissue-specific and variable throughout developmental stages. In the mouse embryo, *Airn* transcription causes

the silencing of its overlapping antisense partner *Igf2r*, whereas in more differentiated somatic and extraembryonic tissues, its expression-induced repression extends to distal genes located 100 Kbs to 15 Mbs away [12,37,43–45,53]. Given that the silencing of *Igf2r* is clearly linked to transcriptional inference via *Airn*, questions arose whether the distal silencing associated with *Airn* expression was also due to its act of transcription or in fact due to a functional lncRNA product [12,54–57]. Furthermore, the large distances between the *Airn* locus and its target gene loci suggest that 3D genome architecture and DNA regulatory elements play critical roles in controlling the potency of *Airn*'s repression and recruitment of PRCs. Whether this mechanism applies to other domains regulated by lncRNAs is unclear, but evidence on *Xist*- and *Airn*-repressed regions suggest that is the case.

1.4 LncRNA-mediated repression through 3D genome architecture and DNA regulatory elements

There is no doubt that the spatial organization of the genome is essential for proper gene expression. Indeed, transcription and 3D chromatin architecture are strongly coupled, whereby the state of one influences the state of the other and vise-versa [58–61]. The genome is organized in a hierarchical fashion. From the highest order, chromosomes tend to occupy specific positions within the nucleus called chromosome territories. Next, chromosomal regions that share biochemical and functional properties, such as transcriptional states of gene expression and histone modifications, tend to cluster together in nuclear space, thereby forming the "A" and "B" compartments. Within these compartments, Mb-sized genomic regions that have a high propensity to exhibit retained internal chromatin interactions are referred to as "contact domains" (also known as topologically associated domains, or TADs). A subset of these interactions within

contact domains are Kb-sized DNA loops that bring distal sites into proximity, such as genes and their cognate regulatory elements [58,61–64]. Finally, the integrity of these higher-ordered structures is strongly dependent on local chromatin structure and nucleosome dynamics [60]. Overall, each level of genome architecture influences or is influenced by, to some extent, transcription, and lncRNAs have been implicated to influence this relationship, particularly at the level of 3D organization [10,65,66]. This is especially true for the Xi [65,67,68].

Given that repressive lncRNAs like *Xist* and *Airn* induce gene silencing and recruit the PRCs over multiple Mbs of chromatin, it is not surprising that 3D genome architecture, likely influenced in part by PRC activity, plays a key role. Critical works over the last ten years have profiled the conformation of the chrX and its temporal relationship with *Xist* expression, and it is clear that its structural organization is influential in the effective chromosome-wide silencing by *Xist* [67,69–71]. From both high-resolution microscopic and genomics observations, the Xi adopts a compact 3D bipartite confirmation that is distinct from its active X (Xa) counterpart and localized to the nuclear periphery [67]. The bipartite structure is comprised of two large "superdomains" that harbor high-density intramolecular interactions and are separated by a lncRNA gene called *Dxz4* that is partially required for their proper formation [67,69,72,73]. Furthermore, the *Dxz4* locus is strongly bound by the major architectural proteins CTCF and Cohesin, which appear to be responsible for mediating the observed DNA interactions between *Dxz4* and X-linked loci [67]. Still, the factors that promote the precise bipartite conformation alongside *Dxz4* are unknown. The chromatin scaffolding protein SMCHD1 has been implicated for the intermediate step of bipartite formation by attenuating the A/B compartments and local contact domains. Interestingly, SMCHD1 recruitment to the Xi is dependent on the initial *Xist*-HNRNPK-vPRC1 mechanism, and its chromatin re-organization activity is required for the

large-scale recruitment and spread of other PRC forms. This supports the notion that PRCs and chromatin organization are important for *Xist*-mediated silencing [67,74,75]. While this is the case for establishing the Xi, disrupting the integrity of the bipartite structure after Xi establishment does not cause loss of PRCs and reactivation of X-linked genes, suggesting that maintenance of repression is dependent on PRCs and independent of chromatin structure.

While the functional significance of the bipartite Xi structure is not yet fully understood, there are several aspects of 3D chromatin architecture during the establishment of this unique conformation that suggest its importance in facilitating *Xist* silencing over the entire chromosome. First, upon expression, *Xist* RNA can diffuse away from its site of transcription and preferentially localize to active genes within regions that form pre-existing DNA interactions with the *Xist* locus [71,76–78]. The initiation of silencing by *Xist* leads to the *Xist*-HNRNPK-vPRC1 pathway, which then recruits SMCHD1 that begins the chromatin reorganization process to form the bipartite structure. *Xist*'s diffusion capacity together with the formation of the highly condensed conformation makes it easier for the *Xist* to readily localize, spread, and maintain PRC recruitment and activity for stable repression of the entire Xi [67]. Such intricate mechanisms of Xi maintenance are essential for preventing inappropriate reactivation. The mechanistic insights gained on 3D chromatin architecture of the Xi can be applied to other lncRNA-repressed regions, such as the imprinted domain regulated by *Airn*.

Until recently, the 3D chromatin conformation of the *Airn*-repressed domain was poorly understood. However, given the scope of *Airn*'s repressive potency in extraembryonic tissues coupled with its unstable transcriptional properties, there is much more motive to determine how 3D chromatin structures might facilitate long-distance repression by *Airn*. From allele-specific fluorescence *in situ* hybridization (FISH) microscopy and first-generation chromatin capture

conformation (3C) methods in mouse trophoblast stem cells (TSCs) and visceral yolk sacs, we and others revealed pre-existing DNA interactions that frequently occur between the *Airn* locus and imprinted targets, such as the *Slc22a3* gene that is ~250 Kbs away. Indeed, these interacting regions persisted on the allele that expresses *Airn* and are subject to PRC2-catalyzed H3K27me3 modifications [43,52]. Furthermore, the frequency and shorter spatial distances to *Airn* correlated with extent of H3K27me3 intensity across FISH-probed regions of the *Airn*-expressing allele in TSCs, suggesting that proximity between *Airn* and its target loci predicts the extent of *Airn*-mediated repression [52]. Notably, the regions of pre-existing and *Airn*-induced DNA interactions coincide CGI regulatory elements that overlap the promoters of *Airn* target genes. In general, the precise role of CGIs is unclear, but much of their functional implication in transcriptional regulation comes from their high affinity for PRC components [79–81]. Indeed, the CGIs found near *Airn* in 3D space are unique from other CGIs within the targeted domain in that they bind RING1B and EZH2 with some of the highest affinity on both *Airn*- and non-expressing alleles. In addition, these sites are the most sensitive to *Airn*-induced PRC recruitment, appearing as nucleation sites for the strongest targeting and spread of *Airn* silencing [52]. Given the apparent relationship between *Airn*-mediated PRC recruitment and 3D DNA interactions through such DNA regulatory elements, we speculated that CGIs functionally serve as Polycomb Response Elements in flies, through which Polycomb-mediated chromatin organization and repressive domain formation is established over large distances [16,82]. In this same regard, we suspect that the PRCs at CGIs, even on the allele that does not express *Airn*, are mediating the long-range interactions with the *Airn* locus, creating a 3D interconnected network that promotes distal silencing by *Airn.* In mouse ESCs, similar scenarios are observed elsewhere in the genome involving PRC-bound poised enhancer and silencer elements that facilitate PRC-repressed

networks [15,16,27]. Consistent with this notion, we previously deleted the CGI at the *Slc22a3* gene that interacts with *Airn* in TSCs, which caused a multi-Mb reduction in H3K27me3 across the 15-Mb domain, suggesting a co-dependency between CGIs and *Airn* through which long-distance DNA interactions form and nucleate the spread of repression by *Airn* [52]. Still, it is unclear to what extent regions across *Airn*'s imprinted domain interact both in the absence and presence of *Airn* expression, to what extent these conformations control distal repression by *Airn*, and to what extent the PRC-bound CGIs control this process. Furthermore, since transcription and 3D chromatin organization are strongly coupled, it is still unclear whether the distal silencing associated with *Airn* expression is in fact due to its act of transcription or a functional lncRNA product.

1.5 The debate on *Airn*-mediated repression: act of transcription versus lncRNA product

While questions remain surrounding the mechanistic action of repression by *Airn* and other repressive lncRNAs, multiple lines of evidence suggest that *Xist* produces a functional lncRNA product that is responsible for executing XCI. For one, its RNA can be visualized microscopically, coating and forming a cloud-like appearance over the Xi [77,78]. At the molecular level, *Xist* is detected non-uniformly associating with Xi chromatin via capture hybridization methods [71,76]. These patterns of *Xist* RNA localization are conserved in its function, whereby ectopically expressing it from essentially any chromosomal location results in the same chromosome-wide effect. Notably, the underlying regions to which *Xist* localizes are also subject to transcriptional repression and PRC recruitment [12,83–86]. Overall, the collective evidence diminishes any doubt that *Xist* lncRNA is functional, but whether the same mechanisms are true for *Airn* is largely unclear.

The critical experiments demonstrating functionality from *Xist*'s lncRNA product are missing for *Airn*. Still, certain observations on *Airn* relative to its imprinted domain hint that *Airn*-mediated repression is indeed due to a functional lncRNA product. In TSCs, expression of *Airn* represses genes and induces the deposition of H2AK119ub and H3K27me3 in a highly non-uniform fashion, like *Xist*, across a 15-Mb window [52,53]. We previously demonstrated that the level of intensity for both gene repression and PRC-directed modifications could be modulated by altering levels of *Airn* transcription from its endogenous promoter [52]. These findings support a role for the *Airn* lncRNA product in PRC recruitment due to the chromatin within the *Airn*-targeted domain being highly sensitive to levels of *Airn* RNA. Also like *Xist*, we see by standard and single-molecule FISH microscopy that expression of *Airn* leads to a single RNA focus that stays retained to its site of transcription, suggesting that *Airn* lncRNA is present and can be associated with the silencing induced by its expression [52]. One hypothesis in support of the act of transcription model is that the *Airn* gene possesses enhancers that promote the expression of distal genes like *Slc22a3* through enhancer-promoter interactions. This suggests it is the consequence of *Airn* transcription that disrupts these interactions and causes imprinted repression of the distal genes [43]. However, deletion of the *Airn* gene, including the enhancers, on the maternal allele that does not express *Airn* did not cause repression of the distal genes, thus providing some of the first direct evidence against the act of transcription model [43]. While these findings make a starting good case that *Airn* is a functional lncRNA product, its unusual sequence and transcriptional properties, especially when compared to *Xist*, still raise the question of whether its act of *Airn* transcription is responsible for its regulatory effect. This same question has been posed for other putative regulatory lncRNAs given that much of our understanding regarding their regulatory mechanisms involves their act of transcription, and not the lncRNA

product [87]. Nonetheless, in the present day, we are more equipped with advanced technologies and sophisticated model systems to definitively address these unknowns by continuing our focus on *Airn* and its repressive mechanisms.

1.6 Using F1-hybrid models to study lncRNA-mediated repression

An attractive feature of repressive lncRNAs like *Xist* and *Airn* that make them useful models for studying lncRNA function is their *cis*-acting nature. In other words, their mono-allelic expression coupled with their repression of genes solely on the same chromosome makes both lncRNAs and their repressive events traceable model systems. As noted above, *Airn*-induced repression appears the most potent in extraembryonic tissues. This is especially true in the trophoblast, in which genomic imprinting, PRC activity, and gene regulation by lncRNAs are highly prevalent [12,52,53,88,89]. Therefore, given that both *Xist* and *Airn*, among others, have imprinted expression patterns (both from the paternal allele) and act in *cis*, their study *in vivo* necessitates the need for controlled F1-hybrid systems. Our group utilizes F1-hybrid mouse trophoblast stem cells (TSCs), which are precursors and an *ex vivo* model of the trophoblast [32,37,53,90–94]. F1-hybrid TSCs are a fantastic model system for studying allelic physical events associated with each parentally inherited chromosome due to the inherit differences in genetic content, such as single nucleotide polymorphisms (SNPs), between the parent strains of in-bred mice (i.e., in our system, between C57BL/6J and CAST/EiJ) [52,53,93]. In particular, with high throughput sequencing based assays, we can extrapolate the data and map each allele back to their parent-of-origin through genome-associated SNPs in the sequencing reads. A critical advantage to using F1-hybrid TSCs is that they harbor an internal control in all genomic profiling experiments, in the form of the allele on which *Airn* is not expressed. Additionally, we

often make it a point to perform each assay at least once with the reciprocal F1-hybrid line to avoid making any conclusions based on results derived from strain bias, and not parent-of-origin bias. In this dissertation, the work presented in Chapter 2 was predominantly performed in F1-hybrid TSCs to study *Airn* and its imprinted domain on chr17.

A disadvantage to our TSC model is that our studies are limited to assessing only steady-state events of genomic imprinting and lncRNA-mediated repression. Thus, in order to study the temporal aspects of these events, we recently engineered a new F1-hybrid inducible gene system in mouse embryonic stem cells (ESC). In our system, any lncRNA gene of interest can be cloned downstream of a doxycycline-inducible promoter in a DNA vector that enables its Cre-mediated insertion into the C57BL/6J allele of the *Rosa26* locus on chr6. We refer to these cells as "RMCEs," and we have had previous success in both the gene insertion and chromosome analysis of *Xist* and mutant forms of *Xist* [83]. In this dissertation, the work presented in Chapter 3 reveals our recent success in the insertion of the full 89-kb *Airn* gene in our RMCE system. For the first time, we were able to express *Airn* from an ectopic locus, thereby allowing us to investigate effects arising only from expression of the lncRNA itself, decoupled from any effects of endogenous *cis*-regulatory features or chromosomal structures that could have evolved in concert with the *Airn* gene. Furthermore, with this system, we also reveal for the first time direct comparisons of the transcriptional and repressive properties between *Xist* and *Airn* under the same expression and chromosomal contexts. Overall, our F1-hybrid TSC and RMCE model systems open the doors to new areas of studies on *Xist*, *Airn*, and other regulatory lncRNAs, increasing our capacity to shift paradigms within the lncRNAs and gene regulation fields.

REFERENCES

1. Theunissen, T. W. & Jaenisch, R. Mechanisms of gene regulation in human embryos and pluripotent stem cells. *Development* **144**, 4496–4509 (2017).

2. Latham, K. E. Preimplantation embryo gene expression: 56 years of discovery, and counting. *Mol. Reprod. Dev.* **90**, 169–200 (2023).

3. Monk, D., Mackay, D. J. G., Eggermann, T., Maher, E. R. & Riccio, A. Genomic imprinting disorders: lessons on how genome, epigenome and environment interact. *Nat. Rev. Genet.* **20**, 235–248 (2019).

4. Barlow, Denise P., Bartolomei, M. S. Genomic Imprinting in Mammals. *Cold Spring Harb. Perspect. Biol.* **6**, a018382 (2014).

5. Tucci, V. *et al.* Genomic Imprinting and Physiological Processes in Mammals. *Cell* **176**, 952–965 (2019).

6. Engreitz, J. M., Ollikainen, N. & Guttman, M. Long non-coding RNAs: Spatial amplifiers that control nuclear structure and gene expression. *Nature Reviews Molecular Cell Biology* **17**, 756–770 (2016).

7. Statello, L., Guo, C. J., Chen, L. L. & Huarte, M. Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* **22**, 96–118 (2021).

8. Autuoro, J. M., Pirnie, S. P. & Carmichael, G. G. Long noncoding RNAs in imprinting and X chromosome inactivation. *Biomolecules* **4**, 76–100 (2014).

9. MacDonald, W. A. & Mann, M. R. W. Long noncoding RNA functionality in imprinted domain regulation. *PLoS Genet.* **16**, 1–22 (2020).

10. Llères, D., Imaizumi, Y. & Feil, R. Exploring chromatin structural roles of non-coding RNAs at imprinted domains. *Biochem. Soc. Trans.* **49**, 1867–1879 (2021).

11. Yao, R. W., Wang, Y. & Chen, L. L. Cellular functions of long noncoding RNAs. *Nat. Cell Biol.* **21**, 542–551 (2019).

12. Trotman, J. B., Braceros, K. C. A., Cherney, R. E., Murvin, M. M. & Calabrese, J. M. The control of polycomb repressive complexes by long noncoding RNAs. *Wiley Interdiscip. Rev. RNA* **12**, 1–41 (2021).

13. Piunti, A. & Shilatifard, A. The roles of Polycomb repressive complexes in mammalian development and cancer. *Nat. Rev. Mol. Cell Biol.* **22**, 326–345 (2021).

14. Almeida, M., Bowness, J. S. & Brockdorff, N. The many faces of Polycomb regulation by RNA. *Current Opinion in Genetics and Development* **61**, 53–61 (2020).

15. Illingworth, R. S. Chromatin folding and nuclear architecture: PRC1 function in 3D.

Current Opinion in Genetics and Development **55**, 82–90 (2019).

16. Guo, Y. & Wang, G. G. Modulation of the high-order chromatin structure by Polycomb complexes. *Front. Cell Dev. Biol.* **10**, 1–10 (2022).

17. Vijayanathan, M., Trejo-Arellano, M. G. & Mozgová, I. Polycomb Repressive Complex 2 in Eukaryotes—An Evolutionary Perspective. *Epigenomes* **6**, 1–29 (2022).

18. Blackledge, N. P. & Klose, R. J. The molecular principles of gene regulation by Polycomb repressive complexes. *Nat. Rev. Mol. Cell Biol.* **22**, 815–833 (2021).

19. Simon, J. A. & Kingston, R. E. Occupying Chromatin: Polycomb Mechanisms for Getting to Genomic Targets, Stopping Transcriptional Traffic, and Staying Put. *Mol. Cell* **49**, 808–824 (2013).

20. Blackledge, N. P. *et al.* PRC1 Catalytic Activity Is Central to Polycomb System Function. *Mol. Cell* (2019). doi:10.1016/j.molcel.2019.12.001

21. Boyle, S. *et al.* A central role for canonical PRC1 in shaping the 3D nuclear landscape. *Genes Dev.* (2020). doi:10.1101/gad.336487.120

22. Blackledge, N. P. *et al.* Variant PRC1 Complex-Dependent H2A Ubiquitylation Drives PRC2 Recruitment and Polycomb Domain Formation. *Cell* **157**, 1445–1459 (2014).

23. Fursova, N. A. *et al.* Synergy between Variant PRC1 Complexes Defines Polycomb-Mediated Gene Repression. *Mol. Cell* (2019). doi:10.1016/j.molcel.2019.03.024

24. Healy, E. *et al.* PRC2.1 and PRC2.2 Synergize to Coordinate H3K27 Trimethylation. *Mol. Cell* (2019). doi:10.1016/J.MOLCEL.2019.08.012

25. Glancy, E. *et al.* PRC2.1- and PRC2.2-specific accessory proteins drive recruitment of different forms of canonical PRC1. *Mol. Cell* **83**, 1393-1411.e7 (2023).

26. Højfeldt, J. W. *et al.* Non-core Subunits of the PRC2 Complex Are Collectively Required for Its Target-Site Specificity. *Mol. Cell* (2019). doi:10.1016/J.MOLCEL.2019.07.031

27. Ngan, C. Y. *et al.* Chromatin interaction analyses elucidate the roles of PRC2-bound silencers in mouse development. *Nat. Genet.* **52**, 264–272 (2020).

28. Van Kruijsbergen, I., Hontelez, S. & Veenstra, G. J. C. Recruiting polycomb to chromatin. *Int. J. Biochem. Cell Biol.* **67**, 177–187 (2015).

29. Lee, J. T. & Bartolomei, M. S. X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell* **152**, 1308–1323 (2013).

30. Brockdorff, N. Noncoding RNA and Polycomb recruitment. *RNA* **19**, 429–442 (2013).

31. Bousard, A. *et al.* The role of Xist -mediated Polycomb recruitment in the initiation of X-chromosome inactivation . *EMBO Rep.* **20**, 119 (2019).

32. Braceros, A. K. *et al.* Proximity-dependent recruitment of Polycomb repressive complexes by the lncRNA Airn. *Cell Rep.* **42**, 112803 (2023).

33. Chu, C. *et al.* Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404–416 (2015).

34. Raposo, A. C., Casanova, M., Gendrel, A. V. & da Rocha, S. T. The tandem repeat modules of Xist lncRNA: A swiss army knife for the control of X-chromosome inactivation. *Biochem. Soc. Trans.* **49**, 2549–2560 (2021).

35. Nakamoto, M. Y., Lammer, N. C., Batey, R. T. & Wuttke, D. S. HnRNPK recognition of the B motif of Xist and other biological RNAs. *Nucleic Acids Res.* **48**, 9320–9335 (2020).

36. Pintacuda, G. *et al.* hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing. *Mol. Cell* **68**, 955-969.e10 (2017).

37. Schertzer, M. D. *et al.* lncRNA-Induced Spread of Polycomb Controlled by Genome Architecture, RNA Abundance, and CpG Island DNA. *Mol. Cell* **75**, 523-537.e10 (2019).

38. Colognori, D., Sunwoo, H., Kriz, A. J., Wang, C. Y. & Lee, J. T. Xist Deletional Analysis Reveals an Interdependency between Xist RNA and Polycomb Complexes for Spreading along the Inactive X. *Mol. Cell* **74**, 101-117.e10 (2019).

39. Wei, G., Almeida, M., Bowness, J. S., Nesterova, T. B. & Brockdorff, N. Xist Repeats B and C, but not Repeat A, mediate de novo recruitment of the Polycomb system in X chromosome inactivation. *Dev. Cell* **56**, 1234–1235 (2021).

40. Bousard, A. *et al.* The role of *Xist* -mediated Polycomb recruitment in the initiation of X-chromosome inactivation. *EMBO Rep.* (2019). doi:10.15252/embr.201948019

41. Żylicz, J. J. *et al.* The Implication of Early Chromatin Changes in X Chromosome Inactivation. *Cell* **176**, 182-197.e23 (2019).

42. Colognori, D., Sunwoo, H., Wang, D., Wang, C. Y. & Lee, J. T. Xist Repeat A contributes to early recruitment of Polycomb complexes during X-chromosome inactivation. *Dev. Cell* **56**, 1236–1237 (2021).

43. Andergassen, D. *et al.* The Airn lncRNA does not require any DNA elements within its locus to silence distant imprinted genes. (2019). doi:10.1371/journal.pgen.1008268

44. Andergassen, D. *et al.* Mapping the mouse Allelome reveals tissue-specific regulation of allelic expression. *Elife* **6**, 1–29 (2017).

45. Wutz, A., Smrzka, O. W. & Schweifer, N. Imprinted expression of the Igf2r gene depends on an intronic CpG island. *Nature* **389**, 745–749 (1997).

46. Sleutels, F., Zwart, R. & Barlow, D. P. The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* **415**, 810–813 (2002).

47. Yotova, I. Y. *et al.* Identification of the human homolog of the imprinted mouse Air non-coding RNA. *Genomics* **92**, 464–473 (2008).

48. Suzuki, S., Shaw, G. & Renfree, M. B. Identification of a novel antisense noncoding RNA, ALID, transcribed from the putative imprinting control region of marsupial IGF2R 06 Biological Sciences 0604 Genetics. *Epigenetics and Chromatin* **11**, 1–11 (2018).

49. Liu, X. *et al.* Genomic imprinting of the IGF2R/AIR locus is conserved between bovines and mice. *Theriogenology* **180**, 121–129 (2022).

50. Braidotti, G, Baubec, T, Seidl, C, Smrzka, O, Stricker, S, Yotova, I, Barlow, D. *The Air Noncoding RNA: An Imprinted cis-silencing Transcript*. (2004).

51. Seidl, C. I. M., Stricker, S. H. & Barlow, D. P. The imprinted Air ncRNA is an atypical RNAPII transcript that evades splicing and escapes nuclear export. *EMBO J.* **25**, 3565–3575 (2006).

52. Schertzer, M. D. *et al.* lncRNA-Induced Spread of Polycomb Controlled by Genome Architecture, RNA Abundance, and CpG Island DNA. *Mol. Cell* **75**, 523-537.e10 (2019).

53. Calabrese, J. M., Starmer, J., Schertzer, M. D., Yee, D. & Magnuson, T. A survey of imprinted gene expression in mouse trophoblast stem cells. *G3 Genes, Genomes, Genet.* **5**, 751–759 (2015).

54. Latos, P. A. *et al.* Airn Transcriptional Overlap, But Not Its lncRNA Products, Induces Imprinted Igf2r Silencing. *Science (80-.).* **338**, 1469–1472 (2012).

55. Pauler, F. M., Barlow, D. P. & Hudson, Q. J. Mechanisms of long range silencing by imprinted macro non-coding RNAs. *Curr. Opin. Genet. Dev.* **22**, 283–289 (2012).

56. Sleutels, F., Tjon, G., Ludwig, T. & Barlow, D. P. Imprinted silencing of Slc22a2 and Slc22a3 does not need transcriptional overlap between Igf2r and Air. *EMBO J.* **22**, 3696–3704 (2003).

57. Stricker, S. H. *et al.* Silencing and transcriptional properties of the imprinted Airn ncRNA are independent of the endogenous promoter. *EMBO J.* **27**, 3116–3128 (2008).

58. Li, Y., Hu, M. & Shen, Y. Gene regulation in the 3D genome. *Hum. Mol. Genet.* **27**, R228–R233 (2018).

59. van Steensel, B. & Furlong, E. E. M. The role of transcription in shaping the spatial organization of the genome. *Nat. Rev. Mol. Cell Biol.* **20**, 327–337 (2019).

60. Venkatesh, S. & Workman, J. L. Histone exchange, chromatin structure and the regulation of transcription. *Nat. Rev. Mol. Cell Biol.* **16**, 178–189 (2015).

61. Dekker, J. & Mirny, L. The 3D Genome as Moderator of Chromosomal Communication. *Cell* **164**, 1110–1121 (2016).

62. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).

63. Rao, S. S. P. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. (2014). doi:10.1016/j.cell.2014.11.021

64. Gibcus, J. H. & Dekker, J. The Hierarchy of the 3D Genome. *Mol. Cell* **49**, 773–782 (2013).

65. Yan, F., Wang, X. & Zeng, Y. 3D genomic regulation of lncRNA and Xist in X chromosome. *Semin. Cell Dev. Biol.* (2019). doi:10.1016/j.semcdb.2018.07.013

66. Gil, N. & Ulitsky, I. Regulation of gene expression by cis-acting long non-coding RNAs. *Nat. Rev. Genet.* **21**, 102–117 (2020).

67. Fang, H., Disteche, C. M. & Berletch, J. B. X Inactivation and Escape: Epigenetic and Structural Features. *Front. Cell Dev. Biol.* **7**, 1–12 (2019).

68. Pandya-Jones, A. & Plath, K. The "lnc" between 3D chromatin structure and X chromosome inactivation. *Semin. Cell Dev. Biol.* **56**, 35–47 (2016).

69. Giorgetti, L. *et al.* Structural organization of the inactive X chromosome in the mouse. *Nature* **535**, 575–579 (2016).

70. Simon, M. D. *et al.* High-resolution Xist binding maps reveal 2-step spreading during X-inactivaiton. *Nature* **504**, 465–469 (2014).

71. Engreitz, J. M. *et al.* The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science (80-.).* **341**, 1–9 (2013).

72. Deng, X. *et al.* Bipartite structure of the inactive mouse X chromosome. *Genome Biol.* **16**, 1–21 (2015).

73. Froberg, J. E., Pinter, S. F., Kriz, A. J., Jégu, T. & Lee, J. T. Megadomains and superloops form dynamically but are dispensable for X-chromosome inactivation and gene escape. *Nat. Commun.* **9**, 1–19 (2018).

74. Wang, C. Y., Jégu, T., Chu, H. P., Oh, H. J. & Lee, J. T. SMCHD1 Merges Chromosome Compartments and Assists Formation of Super-Structures on the Inactive X. *Cell* 406–421 (2018). doi:10.1016/j.cell.2018.05.007

75. Jansz, N. *et al.* Smchd1 Targeting to the Inactive X Is Dependent on the Xist-HnrnpK-PRC1 Pathway. *Cell Rep.* **25**, (2018).

76. Simon, M. D. *et al.* High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. *Nature* **504**, 465–469 (2013).

77. Cerase, A., Pintacuda, G., Tattermusch, A. & Avner, P. Xist localization and function:

New insights from multiple levels. *Genome Biol.* **16**, 1–12 (2015).

78. Brockdorff, N. Localized accumulation of Xist RNA in X chromosome inactivation. *Open Biol.* **9**, (2019).

79. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **25**, 1010–22 (2011).

80. Blackledge, N. P. & Klose, R. J. CpG island chromatin: A platform for gene regulation. *Epigenetics* **6**, 147–152 (2011).

81. Simon, J. A. & Kingston, R. E. Occupying Chromatin: Polycomb Mechanisms for Getting to Genomic Targets, Stopping Transcriptional Traffic, and Staying Put. *Mol. Cell* **49**, 808–824 (2013).

82. Ogiyama, Y., Schuettengruber, B., Papadopoulos, G. L., Chang, J. M. & Cavalli, G. Polycomb-Dependent Chromatin Looping Contributes to Gene Silencing during Drosophila Development. *Mol. Cell* **71**, 73-88.e5 (2018).

83. Trotman, J. B. *et al.* Elements at the 5′ end of Xist harbor SPEN-independent transcriptional antiterminator activity. *Nucleic Acids Res.* **48**, 10500–10517 (2020).

84. Holding, C. Autosome inactivation by ectopic XISTin differentiated cells. *Genome Biol.* **3**, (2002).

85. Hall, L. L. *et al.* An ectopic human XIST gene can induce chromosome inactivation in postdifferentiation human HT-1080 cells. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 8677–8682 (2002).

86. Trotman, J. B. *et al.* Ectopically expressed Airn lncRNA deposits Polycomb with a potency that rivals Xist. *bioRxiv* (2023). doi:10.1101/2023.05.09.539960

87. Engreitz, J. M. *et al.* Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* **539**, 452–455 (2016).

88. McAninch, D., Roberts, C. T. & Bianco-Miotto, T. Mechanistic insight into long noncoding RNAs and the placenta. *International Journal of Molecular Sciences* **18**, (2017).

89. Maclary, E. *et al.* PRC2 represses transcribed genes on the imprinted inactive X chromosome in mice. *Genome Biol.* **18**, 1–17 (2017).

90. Quinn, J., Kunath, T. & Rossant, J. Mouse trophoblast stem cells. *Methods Mol. Med.* **121**, 125–148 (2006).

91. Oda, M., Shiota, K. & Tanaka, S. Trophoblast Stem Cells. *Methods Enzymol.* **419**, 387–400 (2006).

92. Roberts, R. M. & Fisher, S. J. Trophoblast Stem Cells. *Biol. Reprod.* **84**, 412–21 (2011).

93. Calabrese, J. M. *et al.* Site-specific silencing of regulatory elements as a mechanism of x inactivation. *Cell* **151**, 951–963 (2012).

94. Rossant, J. & Cross, J. C. Placental development: Lessons from mouse mutants. *Nature Reviews Genetics* **2**, 538–548 (2001).

95. Plath, K. *et al.* Role of histone H3 lysine 27 methylation in X inactivation. *Science (80-.).* **300**, 131–135 (2003).

96. Kohlmaier, A. *et al.* A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. *PLoS Biol.* **2**, (2004).

97. Dixon-McDougall, T. & Brown, C. J. Multiple distinct domains of human XIST are required to coordinate gene silencing and subsequent heterochromatin formation. *Epigenetics and Chromatin* **15**, 1–18 (2022).

98. Holoch, D. *et al. A cis-acting mechanism mediates transcriptional memory at Polycomb target genes in mammals*. *Nature Genetics* **53**, (Springer US, 2021).

99. Terranova, R. *et al.* Polycomb Group Proteins Ezh2 and Rnf2 Direct Genomic Contraction and Imprinted Repression in Early Mouse Embryos. *Dev. Cell* **15**, 668–679 (2008).

100. Hernández-Romero, I. A. & Valdes, V. J. De Novo Polycomb Recruitment and Repressive Domain Formation. *Epigenomes* **6**, (2022).

101. Plys, A. J. *et al.* Phase separation of polycomb-repressive complex 1 is governed by a charged disordered region of CBX2. *Genes Dev.* **33**, 799–813 (2019).

102. Isono, K. *et al.* SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. *Dev. Cell* **26**, 565–577 (2013).

103. Shukla, S. *et al.* Small-molecule inhibitors targeting Polycomb repressive complex 1 RING domain. *Nat. Chem. Biol.* **17**, 784–793 (2021).

104. Danishuddin, Subbarao, N., Faheem, M. & Khan, S. N. Polycomb repressive complex 2 inhibitors: emerging epigenetic modulators. *Drug Discov. Today* **24**, 179–188 (2019).

105. Bonev, B. *et al.* Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171**, 557-572.e24 (2017).

106. Du, Z. *et al.* Polycomb Group Proteins Regulate Chromatin Architecture in Mouse Oocytes and Early Embryos. *Mol. Cell* **77**, 825-839.e7 (2020).

107. Oksuz, O. *et al.* Capturing the Onset of PRC2-Mediated Repressive Domain Formation. *Mol. Cell* **70**, 1149-1162.e5 (2018).

108. Crispatzu, G. *et al.* The chromatin, topological and regulatory properties of pluripotency-

associated poised enhancers are conserved in vivo. *Nat. Commun.* **12**, 1–17 (2021).

109. Kraft, K. *et al.* Polycomb-mediated genome architecture enables long-range spreading of H3K27 methylation. *Proc. Natl. Acad. Sci. U. S. A.* **119**, 1–10 (2022).

110. Bloom, K. S. Centromeric Heterochromatin: The Primordial Segregation Machine. *Annu. Rev. Genet.* **48**, 457–484 (2014).

111. Allshire, R. C. & Madhani, H. D. Ten principles of heterochromatin formation and function. *Nat. Rev. Mol. Cell Biol.* **19**, 229–244 (2018).

112. Cutter DiPiazza, A. R. *et al.* Spreading and epigenetic inheritance of heterochromatin require a critical density of histone H3 lysine 9 tri-methylation. *Proc. Natl. Acad. Sci. U. S. A.* **118**, 1–10 (2021).

113. Nicetto, D. & Zaret, K. Role of H3K9me3 Heterochromatin in Cell Identity Establishment and Maintenance. *Curr. Opin. Genet. Dev.* **55**, 1–10 (2019).

114. Shankar, S. R. *et al.* G9a, a multipotent regulator of gene expression. *Epigenetics* **8**, 16–22 (2013).

115. Nagano, T. *et al.* The Air Noncoding RNA Epigenetically Silences Transcription by Targeting G9a to Chromatin. *Science (80-.).* **322**, 1717–1720 (2008).

116. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, (2012).

117. Haws, S. A., Simandi, Z., Barnett, R. J. & Phillips-Cremins, J. E. 3D genome, on repeat: Higher-order folding principles of the heterochromatinized repetitive genome. *Cell* **185**, 2690–2707 (2022).

118. Navarro-Cobos, M. J., Balaton, B. P. & Brown, C. J. Genes that escape from X-chromosome inactivation: Potential contributors to Klinefelter syndrome. *Am. J. Med. Genet. Part C Semin. Med. Genet.* **184**, 226–238 (2020).

119. Zhang, Y. *et al.* Genes that escape X-inactivation in humans have high intraspecific variability in expression, are associated with mental impairment but are not slow evolving. *Mol. Biol. Evol.* **30**, 2588–2601 (2013).

120. Balaton, B. P., Dixon-Mcdougall, T., Peeters, S. B. & Brown, C. J. The eXceptional nature of the X chromosome. *Hum. Mol. Genet.* **27**, R242–R249 (2018).

121. Berletch, J. B., F, Y. & Disteche, C. M. Escape from X inactivation in mice and humans. *Genome Biol.* **11**, (2010).

122. Wainer Katsir, K. & Linial, M. Human genes escaping X-inactivation revealed by single cell expression data. *BMC Genomics* **20**, 1–17 (2019).

123. Carrel, L. & Brown, C. J. When the lyon(Ized chromosome) roars: Ongoing expression from an inactive X chromosome. *Philos. Trans. R. Soc. B Biol. Sci.* **372**, (2017).

CHAPTER 2: Proximity-dependent recruitment of Polycomb repressive complexes by the lncRNA *Airn*[1]

2.1 Introduction

Genomic imprinting is a process known to occur for some ~150 mammalian genes, resulting in their preferential expression from one parentally inherited allele over the other. Dysregulated imprinting can lead to developmental disorders, cancer, and changes in metabolism. Studies of genomic imprinting have also yielded important insights into fundamental mechanisms of gene regulation, including the recognition that long noncoding RNAs (lncRNAs) control the expression of genes that are essential for proper human development [1-4].

Airn (*Antisense of Igf2r Non-Protein Coding RNA*) is a gene that in mice is imprinted and produces a lncRNA specifically from the paternal allele of chromosome 17. The *Airn* locus produces lncRNA transcripts that are upwards of ~90kb in length but have heterogeneous 3′ ends [5]. In addition, *Airn* transcripts are lowly abundant, short-lived, retained near their site of transcription, predominantly unspliced, and not obviously conserved outside of rodents [6-12]. Nevertheless, *Airn* expression results in transcriptional repression over a domain that spans ~15 megabases (Mb) in extraembryonic tissues of the mouse, the largest autosomal region known to be under the control of a repressive lncRNA [13,14]. Repression by *Airn* occurs predominantly if not exclusively *in cis*, on the same chromosome from which the lncRNA is transcribed [13,14]. The

[1] This chapter previously appeared as an article in the Journal of Cell Reports. The original citation is as follows: Braceros et al. "Proximity-dependent recruitment of Polycomb repressive complexes by the lncRNA *Airn*," *Cell Reports* 42 no. 7(July 2023): 112803.

mechanism by which *Airn* mediates repression over its 15Mb target domain is not clear. *Airn*'s lack of conservation, lack of splicing, and the instability and variable length of its RNA product have raised questions about whether it is the *Airn* lncRNA itself or merely the act of its transcription that mediates repression [15,16]. Accumulating data are supportive of a role for the *Airn* lncRNA product in mediating long-range repression [14,17,18], yet it remains unclear what properties of the RNA may enable it to do so. Moreover, the intensity of repression across the *Airn* target domain is non-uniform [14], implying that features of the genome influence repression by *Airn* in ways that are not yet clear.

For its full repressive effect, *Airn* requires several histone-modifying enzymes, including the Polycomb Repressive Complexes (PRCs), which are known to be recruited to chromatin by the expression of a handful of other lncRNAs, including *Xist* during the process of X chromosome Inactivation [19-21]). There are two major PRCs, PRC1 and PRC2, each of which can be classified into different sub-complexes that contain core and auxiliary components [22]. PRC1 monoubiquitinates histone H2A at lysine 119 (H2AK119ub) and is comprised of canonical and variant complexes termed cPRC1 and vPRC1, respectively. PRC2 tri-methylates histone H3 at lysine 27 (H3K27me3) and is comprised of sub-complexes called PRC2.1 and PRC2.2. The different auxiliary factors that distinguish PRC sub-complexes modulate their enzymatic activities, interaction partners, effects on 3-dimensional (3D) genome organization, and ultimately, effects on gene expression [22]. For example, specific forms of vPRC1 have been shown to interface the most closely with the lncRNA *Xist* [23,24].

We previously found that in mouse trophoblast stem cells (TSCs), expression of *Airn* represses genes and induces the deposition of H2AK119ub and H3K27me3 in a highly non-uniform fashion across a 15Mb target domain [14]. Intensity of gene repression and PRC-directed

modifications could be modulated by altering levels of *Airn* transcription from its endogenous promoter, supporting a role for the *Airn* lncRNA product in PRC recruitment and indicating that chromatin within the *Airn* target domain is highly sensitive to levels of *Airn*. Within the domain, the regions most highly decorated in PRC-deposited modifications centered around a subset of CpG island (CGI) promoters bound by the catalytic components of PRC1 and PRC2, even on the maternal allele, which does not express *Airn*. These and other data led us to hypothesize that the non-uniform repression across the *Airn* target domain was mediated by DNA regulatory elements that preferentially contact the *Airn* locus through 3D space and focus *Airn*'s repressive activity over certain genomic regions.

Herein, we set out to test that hypothesis and examine in greater detail the extent to which repression by *Airn* is influenced by chromatin architecture and underlying features of the genome. In TSCs and mouse embryonic stem cells (ESCs), we found that variation in repression across the domain could be partly explained by 3D DNA contacts that exist in the absence of *Airn* expression, which appear to bring certain genomic regions in closer proximity to the *Airn* locus over others. Regions within the domain that associated the most robustly with the *Airn* lncRNA product associated the most robustly with PRC1 and PRC2. Seemingly similar DNA regulatory elements located in different regions within the target domain had different effects on *Airn*-dependent repression, possibly by modulating local PRC recruitment or frequency of contact between *Airn* and target DNA. Our data support the notion that the *Airn* lncRNA product recruits several forms of PRC1 and PRC2 to chromatin and demonstrates that DNA regulatory elements can control the regional intensity with which it does so.

2.2 Results

2.2.1 Airn expression induces large-scale changes to chromatin architecture

Because *Airn* is monoallelically expressed and *cis*-acting, its target domain exists in different states on each allele; the paternal allele being repressed by *Airn*, and the maternal allele existing in the non-repressed state. For this reason, we and others have studied *Airn* in F1-hybrid cell lines or animals derived from different strains of inbred mice [13,14,17]. In F1-hybrids, physical events associated with maternal and paternal alleles can be distinguished by single-nucleotide polymorphisms (SNPs) in sequencing reads [25].

To determine how *Airn* expression alters chromatin architecture, we performed *in situ* Hi-C [26] in three F1-hybrid TSC lines: one line derived from a cross between a CAST/EiJ mother and C57BL/6J father (C/B TSCs), one derived from the reciprocal cross, a C57BL/6J mother and CAST/EiJ father (B/C TSCs), and a third TSC line in which we previously used CRISPR to insert a triple-polyadenylation cassette ~3kb downstream of the *Airn* transcription start site in C/B TSCs, generating a mutant with an expected null phenotype (C/B *Airn* truncation TSCs; [8,14]). Hi-C libraries were generated in biological triplicate from each TSC line and sequenced to an aggregate depth per genotype of at least 1.6 billion paired-end 150 nucleotide (nt) reads (Table S1).

We constructed allele-specific contact maps to examine how *Airn* expression alters chromatin architecture across its target domain [27,28]. In C/B WT TSCs, on the paternal B6 allele, which expresses *Airn*, relative to the maternal CAST allele, which does not, we observed a reduction in short-range DNA contacts concomitant with an increase in long-range contacts, which were largely contained within the 15Mb target domain previously shown to be repressed by *Airn* in TSCs (Figures 1A and B, panel (i); [14]). The greatest increases in contact occurred

within a 4.5Mb interval that extended from the *Airn* locus and terminated at the genes *Prr18*, *T*, and *Pde10a* (Figures 1A and B, panel (i)). In *Airn* truncation TSCs, the increased contacts were not detectable, demonstrating their dependence on *Airn* expression (Figures 1A and B, panel (ii)). Moreover, the overall trends that were observed in C/B WT TSCs were also observed in the reciprocal F1-hybrid wildtype cell line -- B/C TSCs -- in which the paternal, *Airn*-expressing allele is of CAST origin, confirming that differences in chromatin architecture between paternal and maternal alleles are due to parent-of-origin and not strain-specific effects (Figures 1A and B, panel (iii)). The regions that underwent the strongest *Airn*-dependent changes in contact frequency were the ones that most clearly shifted from the "A" to the "B" chromosome compartment specifically on the alleles that expressed *Airn* (Figure 1C, panels (i-iii); Figure S1; [26,29]). Thus, in TSCs, *Airn* expression induces large-scale changes to chromatin architecture that are largely contained within a 15Mb genomic interval previously shown to be subject to *Airn*-dependent repression [13,14].

2.2.2 Airn-dependent changes in chromatin architecture coincide with the presence of PRC-deposited modifications

To understand how DNA contacts with the *Airn* locus correlate with *Airn*-dependent repression, we created a series of "viewpoint" plots, in which contacts between *Airn* and surrounding regions were extracted and visualized in two dimensions. Consistent with the heatmaps of Figure 1, we observed a dramatic increase in *Airn*-dependent contacts with the *Airn* locus on the paternal allele towards the centromeric end of chr17, including a pronounced shoulder of increased contacts surrounding the genes *Prr18*, *T*, and *Pde10a* (Figure 2A). We also observed augmented contacts between *Airn* and the gene *Qk* on both alleles in all TSC lines

profiled (Figure 2A). In C/B TSCs, the intensity of allele-specific contacts with *Airn* as measured by Hi-C correlated remarkably well with allele-specific distances to the *Airn* locus that we had previously measured by DNA FISH, corroborating both forms of measurement (Figure 2B; Spearman's ρ of -0.82 and -0.95 and p = 0.007 and 0.001 on maternal and paternal alleles, respectively; FISH probe locations from [14] shown under panel A(i)).

The magnitude of *Airn*-dependent contacts (i.e., paternal contacts subtracted from maternal contacts) correlated remarkably well with the underlying intensity of *Airn*-dependent, PRC-directed chromatin modifications (Figures 2C-F; Spearman's ρ between *Airn*-dependent contacts and H3K27me3 and H2AK119ub in C/B TSCs, 0.69 and 0.67, respectively, p < 2e-16 for both comparisons). Moreover, we used ChIP-Seq to profile seven individual components of PRC1 and PRC2 (RING1B, RYBP, CBX7, KDM2B, EZH2, MTF2, and JARID2). Within the target domain, six out of seven PRC components exhibited a signature of *Airn*-responsiveness, defined as a broad shoulder of paternal enrichment on the centromeric side of the *Airn* locus (Figure S2A). The same six PRC components appeared responsive to *Xist* (Figure S3). The sole exception was the PRC1.1 component KDM2B (Figures S2A and S3). Because the PRCs and their modifications to chromatin can induce DNA compaction [7,22,26], and *Airn* expression correlates with the presence of PRCs within its target domain, our data suggest that the large-scale changes in genome architecture induced by expression of *Airn* depend, at least in part, on the PRCs and their modifications to chromatin.

2.2.3 Airn-dependent repression centers around regions that form pre-existing contacts with the Airn locus and harbor CGIs bound by vPRC1

DNA contacts detected by Hi-C can occur by chance with a frequency that increases with decreasing distance from the locus in question [26]. To quantify the relative intensity of contacts that occur with the *Airn* locus after correcting for distance-dependent effects, we created a series of Observed-over-Expected (O/E) plots with *Airn* as the viewpoint, in which detected contacts were normalized by those expected from a distance-dependent decay model [26].

These O/E viewpoint plots revealed three local maxima of contact with *Airn* that fell within the 4.5Mb genomic interval most intensely repressed by *Airn*, extending from the *Airn* locus and terminating at *Prr18*, *T*, and *Pde10a* (Figures 2, 3A, S4). Specifically, maxima were detected surrounding the genes *Prr18*/*T*/*Pde10a*, the gene *Qk*, and the gene *Slc22a3* (Figures 3A and S4). While intensity of O/E contact with *Prr18*/*T*/*Pde10a* increased dramatically upon expression of *Airn*, intensity of O/E contact with *Qk* and *Slc22a3* changed to lesser extents or not at all (Figures 3A and S4). All three maxima were present in *Airn* truncation TSCs, highlighting contact with the *Airn* locus even in the absence of *Airn* expression (Figures 3A, panel (ii) and S4). Each gene within these maxima are driven by CGI promoters that we found in previous work to either bind high levels of RING1B and EZH2 (in the cases of *Prr18*/*T*/*Pde10a* and *Qk*), or are present in the region of the *Airn* target domain that accumulates the highest levels of *Airn*-dependent, PRC-directed chromatin modifications (in the case of *Slc22a3*; Figure 2; [14]).

We were intrigued that points of contact with *Airn* centered around PRC-bound CGIs. CGIs often mark high-density sites of PRC binding, and PRCs and the modifications that they deposit on chromatin can mediate long-range 3D contacts independently of CTCF and Cohesin [30-39]. Moreover, the genes within the points of contact -- *Pde10a*, *Qk*, and *Slc22a3* -- are all

repressed by *Airn* in TSCs ([14]; Table S2). Also, the *Airn* gene body itself harbors two CGIs that bind RING1B/PRC1 [14], and the *Airn* lncRNA has previously been found to associate with the *Slc22a3* CGI [18]. Lastly, although the reasons remain unclear, we previously found that deletion of the *Slc22a3* CGI resulted in a dramatic loss of *Airn*-induced accumulation of H3K27me3, most notably in the 4.5Mb interval beginning at *Airn* and terminating at *Prr18*, *T*, and *Pde10a* [14]. These data raise the possibility that features associated with the CGIs in regions that form augmented DNA contacts with *Airn* play roles in modulating the local intensity of *Airn*-dependent repression.

With that possibility in mind, we used ChIP-Seq to examine what factors and chromatin modifications were enriched over CGIs contained within points of 3D contact with *Airn*. These included CGIs at *Pde10a*, *Qk*, *Slc22a3*, and *Airn* itself, as well as the CGI promoter of the gene *Map3k4*. Like *Qk*, *Map3k4* is repressed by *Airn* and forms a detectable contact with the *Airn* locus by Hi-C (Figures 3A and S4B), but partially escapes silencing and sits within a region that resists the local accumulation of *Airn*-induced, PRC-deposited chromatin modifications (Figures 2C and D; Table S2). In total, we examined ChIP-Seq data for the same seven PRC components whose allelic tiling density profiles are shown in Figures S2 and S3, four chromatin modifications (H3K27me3, H2K119ub, H3K4me2, and H3K27ac), and two architectural factors (SMC1A/Cohesin and CTCF; data from this study and [14,40]). A summary of the allele-specific enrichment of each factor is found in Table S3, and the total, non-allelic genome browser density tracks are shown in (Figure 3B). The asterisks above each CGI indicate whether the factor was detected on the maternal allele, paternal allele, or both (Figure 3B; Table S3).

While there was not one singular pattern of enrichment, notable similarities emerged. Each CGI except for the one found at the promoter of *Airn* showed some level of peak-like

enrichment for vPRC1 on the maternal allele (Figure 3B). Likewise, all of the CGIs examined except for the one at the *Airn* promoter showed peak-like enrichment of at least one of the two chromatin modifications associated with transcriptional activation (H3K4me2 or H3K27ac), also on the maternal allele. SMC1A/Cohesin was also detected on the maternal allele of these same set of CGIs, although its intensity of enrichment was low relative to intergenic peaks (Figure 3B). Thus, the regions within the target domain that form 3D contacts with the *Airn* locus on the maternal allele all harbor CGI promoters that are enriched in vPRC1, Cohesin, and signatures of transcriptional activity.

In contrast, CGIs in the region that underwent the strongest *Airn*-dependent changes in chromatin architecture, surrounding the genes *Prr18*, *T*, and *Pde10a*, were associated with sharp peaks of cPRC1 and PRC2 as well as vPRC1 (Figures 3B and S2A). Of those CGIs, the CGI at *Pde10a* associated with the highest levels of both PRC1 and PRC2 (Table S3). Likewise, the intergenic regions that underwent the strongest *Airn*-dependent changes in chromatin architecture were similarly enriched in cPRC1 and PRC2 as well as vPRC1 (Figure S2A). Thus, while the presence of vPRC1, Cohesin, and transcription can identify regions that contact the *Airn* locus in the absence of *Airn* expression, the presence of cPRC1 and PRC2 at both CGIs and intergenic regions correlate more strongly with the presence of *Airn*-dependent, PRC-deposited modifications and long-range changes to architecture.

2.2.4 Presence of Airn lncRNA on chromatin correlates with presence of PRC1 and PRC2 and centers around pre-existing contacts with the Airn locus

We next sought to determine whether the *Airn* lncRNA itself preferentially associated with specific DNA regions. To address this question, we used CHART-Seq, an approach to

identify genomic regions located proximal to lncRNAs [41]. We performed CHART-Seq for *Airn* (Figure S5A) in C/B wildtype TSCs, in *Airn* truncation TSCs, and in a C/B TSC line from [14], in which we used CRISPR-Cas9 to over-express *Airn* from its endogenous promoter (*Airn* Highly-Expressing or H-E TSCs). RT-qPCR confirmed expected levels of *Airn* expression in each TSC line (Figure S5B).

In C/B wildtype TSCs, *Airn* CHART-Seq revealed enrichment of DNA on the paternal but not maternal allele across the *Airn* target domain, beginning near the centromere and ending ~3Mb downstream of the *Airn* locus, the same region where the last *Airn*-induced PRC-dependent modifications are visible (Figures 4A and B, panels (i), S5C and S5D). In H-E TSCs, the enrichment increased with over-expression of *Airn* (Figures 4A, panel (ii), S5C and S5D). Conversely, in *Airn* truncation cells, DNA enrichment on from the paternal allele was lost (Figure 4A, panel (iii), S5C and S5D). These data indicate that DNA recovered by CHART is sensitive to overall levels of *Airn* expression. Moreover, we observed a strong correlation between paternal *Airn* CHART-Seq and H3K27me3, RING1B, and EZH2 ChIP-Seq signal throughout the target domain (lowest Spearman's ρ across C/B wildtype and H-E TSC comparisons is 0.53, $p<2e-9$ for all comparisons; Figures 4A vs 4B-D). Calibrated and non-calibrated ChIP-Seq for H3K27me3 and EZH2 showed the same patterns of enrichment (Figures 4B, 4D vs S2B, S2C). We were especially struck by the lower CHART-Seq signal in H-E TSCs that began just upstream of *Map3k4* and *Qk*, genes that are repressed by *Airn* but located at inflection points where the intensity of PRCs and PRC-directed modifications drops precipitously (Figures 4A-D). Likewise, particularly in H-E TSCs, local maxima of *Airn* lncRNA association coincide with genomic regions that form local maxima of Hi-C O/E contacts with the *Airn* locus even in the absence of *Airn* expression (Figures 4A and 3A, panel (ii)).

We also performed *Airn* CHART-Seq in mouse ESCs, in which we used CRISPR-Cas9 to force expression of *Airn* from its endogenous promoter (H-E ESCs). In H-E ESCs, *Airn* was expressed at a level approximately equal to WT TSCs (Figure S5B). As a negative control, we performed CHART-Seq in uninduced ESCs, which do not express meaningful levels of *Airn*. We also performed ChIP-Seq to determine the extent to which DNA retrieved by CHART correlated with intensity of H3K27me3. We observed associations between the *Airn* lncRNA and DNA in H-E ESCs across a domain that was remarkably similar in size and contour to the domain associated with *Airn* in TSCs (Figure 4F vs 4A). Also as in TSCs, *Airn* associations were significantly correlated with underlying H3K27me3 (Figure 4F vs 4G; Spearman's $\rho = 0.64$, $p = 7.9e\text{-}14$), and *Map3k4* and *Qk* resisted *Airn* and PRC-directed modifications in H-E ESCs (Figures 4E and F). Furthermore, we observed that local maxima of Hi-C O/E contacts with the *Airn* locus in ESCs coincided or were proximal to local maxima of *Airn* lncRNA association and H3K27me3, most notably at *Prr18*/*T*/*Pde10a*, *Slc22a3*, and on the telomeric side of *Airn*, in a region surrounding the genes *Dact2 and Wdr27* (Figure 4G vs 4E, F; Hi-C, [42]). The only exception to this pattern was at *Qk*, which despite forming DNA contacts with the *Airn* locus, resisted contacting the *Airn* lncRNA and accumulating PRC-directed modifications in both TSCs and ESCs.

Altogether, our data support the notion that a major function of the *Airn* lncRNA product is to recruit the PRCs to chromatin over a 15Mb domain, and that proximity to the *Airn* lncRNA product dictates the intensity with which the recruitment occurs. Moreover, pre-existing DNA contacts – those that occur with the *Airn* locus in the absence of *Airn* expression – center within regions of chromatin that become decorated in PRC-directed modifications upon *Airn*

expression. Lastly, *Map3k4* and *Qk* resist associations with the *Airn* lncRNA despite forming DNA contacts with the *Airn* locus.

2.2.5 DNA regulatory element deletions alter levels of PRC-directed modifications and gene repression throughout the Airn target domain

Our data indicate that regional proximity to the *Airn* lncRNA product correlates with local intensity of gene repression and PRC recruitment. To study the extent to which DNA regulatory elements might control proximity to *Airn*, we focused on the region surrounding the genes *Prr18*/*T*/*Pde10a*, which harbors several CGIs that bind PRCs and undergoes increased frequency of 3D contact with the *Airn* locus upon *Airn* expression (Figures 2 and 3). We used CRISPR to individually delete the CGI promoters of *T* and *Pde10a*, as well as a 190kb cluster of intergenic CTCF and SMC1A/Cohesin peaks located between *T* and *Pde10a* (pink rectangles in Figure 3B; Figures S6A-C). We derived heterozygous clonal TSC lines harboring deletions for each element on their paternal alleles (Δ*T*, four lines; Δ*Pde10a*, two lines; Δ*Cluster*, two lines; Figures S5A-C). As controls, we derived four clonal non-targeting (NTG) TSC lines that harbor the same doxycycline-inducible *Cas9* transgene and underwent the same process of electroporation, clonal selection, and induction as above, but that express a non-targeting sgRNA that does not match the mouse genome (Figures S6A-D; Table S5). Additionally, we revived the two TSC lines in which we previously deleted the *Slc22a3* CGI on the paternal allele (Δ*Slc22a3* TSCs from [14]; Figure S6D). In that study, we found that *Slc22a3* CGI deletion caused a ~4.5Mb reduction in the intensity of H3K27me3, beginning essentially at the *Slc22a3* gene and extending through the distal cluster of PRC-bound CGIs at *Prr18*/*T*/*Pde10a* [14]. RT-qPCR and RNA-Seq

showed that *Airn* expression levels varied by no more than two-fold across our panel of lines (Figures S6E).

We performed H3K27me3 and H2AK119ub ChIP-Seq as well as RNA-Seq to examine how the deletions affected PRC activity and gene repression. Relative to NTG controls, ChIP-Seq in Δ*Slc22a3* TSCs revealed a dramatic loss of H3K27me3 and H2AK119ub throughout the *Airn* target domain (Figure 5A vs S6F, G), consistent with and extending results from [14]. However, whereas deletion of the *T* CGI had little to no effect, deletion of the *Pde10a* CGI unexpectedly caused a dramatic increase in the levels of H3K27me3 and H2AK119ub throughout the target domain, opposite to that observed in Δ*Slc22a3* TSCs (Figures 5B, 5C, S6G). Deletion of the cluster of CTCF and SMC1A/Cohesin peaks similarly increased H3K27me3 and H2AK119ub (Figure 5D). Cross-genotype comparisons within the target domain and across the remainder of chr17 are shown in Figure S6G.

RNA-Seq from deletion clones showed changes in gene expression consistent with changes in PRC-deposited modifications (Figure 5E vs 5A-D). In our previous study, we identified 27 genes within the target domain that were subject to repression by *Airn* [14]. In Δ*Slc22a3* TSCs, the relative paternal expression of these 27 genes increased significantly compared to their baseline in NTG, up to an average level that was slightly less than that observed in *Airn* truncation TSCs, which are effectively null mutants ([14]; Figure 5E; NTG vs Δ*Slc22a3* (A12 and A13 clones), $p = 0.038$ and 0.025, respectively, Welch two sample t-test). Conversely, in Δ*Pde10a* and Δ*Cluster* TSCs, paternal expression of these genes decreased relative to NTG TSCs, albeit with variability between clones (Figure 5E; NTG vs Δ*Pde10a* and NTG vs Δ*Cluster*). The decreased expression in Δ*Pde10a* and Δ*Cluster* TSCs was not as strong as that observed in *Airn* highly-expressing (H-E) TSCs (Figure 5E), but was nevertheless

consistent with the increase in PRC-deposited modifications (Figures 5C and D; [14]). Thus, seemingly similar DNA regulatory elements play critical but different roles in dictating the regional intensity of gene repression and PRC recruitment induced by *Airn* within its 15Mb target domain.

2.2.6 Changes in DNA contacts with Airn mirror changes in PRC activity caused by regulatory element deletion

To gain insight into the effects caused by regulatory element deletion, we used *in situ* Hi-C to examine DNA contacts in NTG, Δ*Slc22a3*, and Δ*Pde10a* TSCs (Table S1). Consistent with changes in gene repression and PRC-deposited modifications, deletion of the *Slc22a3* and *Pde10a* CGIs alternately diminished and increased *Airn*-dependent contacts across the entire target domain, coincident with corresponding changes in the intensities of compartmentalization (Figures 6A and B; Figure S7).

Examining contacts from the *Airn* viewpoint provided additional insights (Figures 6C and D). Deletion of the *Slc22a3* CGI was coincident with reduced levels of *Airn*-dependent contacts throughout the domain, with a possible exception at *Qk* (Figures 6C and D, panels (ii)). Proportionally, the greatest decreases in *Airn*-dependent contacts surrounded *Prr18*/*T*/*Pde10a* (Figures 6C and D, panels (ii)). In contrast, in Δ*Pde10a* cells, *Airn*-dependent contacts increased uniformly except at *Prr18*/*T*/*Pde10a* (Figures 6C and D, panels (iii)). Thus, deletion of the *Slc22a3* CGI reduced the interaction between *Airn* and DNA throughout its target domain, whereas deletion of the *Pde10a* CGI increased the interaction between *Airn* and all other regions in the domain save *Prr18*/*T*/*Pde10a*.

2.2.7 Airn expression is coincident with dissolution of DNA loops encasing Slc22a3 and a local increase in PRC-directed modifications

We next examined our data to determine how deletion of the *Slc22a3* CGI might restrict repression by *Airn*. In our major Hi-C datasets (described in Figure 1), we noted that the *Slc22a3* and *Airn* genes are located within the same contact domain (a contiguous region exhibiting high levels of inter-locus interactions), where they sit within nested DNA loops anchored by CTCF and Cohesin (Figures 7A and B). By Hi-C, the loops that surround *Slc22a3* were reduced by *Airn* expression, to the extent that they are no longer detectable by the SIP algorithm on *Airn*-expressing alleles (Figure 7A; [43]). Likewise, we observed a relative reduction in SMC1A and CTCF binding at those same loop anchors, again on *Airn*-expressing alleles, consistent with the reduced loop intensity (Figure 7B). We also searched for these DNA loops in NTG, Δ*Slc22a3*, and Δ*Pde10a* Hi-C data, but lower sequencing depth (~600 million read pairs per genotype) precluded a high-confidence analysis. Nonetheless, visual inspection of loops identified in our high-depth datasets was consistent with the notion that the *Slc22a3* CGI is required for *Airn*-dependent dissolution of the nested loops encasing *Slc22a3* (Figure 7C; note relative increase in intensity at loop anchor regions in the *ΔSlc22a3* genotype, denoted by the two grey arrows). We also observed that deletion of the *Slc22a3* CGI led to a local drop in H3K27me3 and had a lesser yet significant effect on H2AK119ub (Figure 7D). Thus, in contrast to the *Pde10a* CGI, which may effectively restrict surrounding regions from contacting *Airn* through mechanisms that remain to be determined, it is conceivable that prior to *Airn* expression, nested loops that encase *Slc22a3* and *Airn* may reduce the latter's ability to interact with distal DNA. Upon *Airn* expression, recruitment of PRCs locally, promoted by the *Slc22a3* CGI, could antagonize

Cohesin [37-39] and disrupt loops that then enable *Airn* to contact distal regions of chromatin more efficiently.

2.3 Discussion

We report a series of intriguing relationships between 3D DNA contacts, DNA regulatory elements, and PRC recruitment within the largest autosomal region known to be repressed by a mammalian lncRNA. Our results support the view that *Airn* is a potent *cis*-acting lncRNA that functions to maintain gene repression and recruit the PRCs within a 15Mb domain on mouse chr17. We show that extent of repression maintained by *Airn* can be modulated by discrete DNA regulatory elements that control the proximity of *Airn* to its genomic targets, a paradigm likely relevant to other domains governed by strong locus control regions, including the inactive X chromosome.

Using *in situ* Hi-C, we observed that expression of *Airn* is accompanied by changes in 3D contacts and compartmentalization on the centromeric side of the *Airn* locus. These changes correlated in-step with the intensity of *Airn*-induced H3K27me3 and H2AK119ub, and centered around three regions that contact the *Airn* locus even in the absence of *Airn* expression. Each of these regions harbor CGI promoters that bind components of vPRC1, exhibit signatures of transcriptional activity, and are located proximal to peaks of Cohesin on both the maternal and paternal alleles. Extent of association with the *Airn* lncRNA, as assessed by CHART-Seq, also correlated in-step with the intensity of PRCs and PRC-directed modifications, and centered around pre-existing DNA contacts, in both TSCs and ESCs. Of the seven different PRC subunits we profiled by ChIP-Seq, all except KDM2B were responsive to *Airn*. Two genes – *Map3k4* and *Qk* – resisted accumulating PRC-directed modifications, despite both genes being repressed by

Airn and forming contacts with the *Airn* locus. Moreover, the intensity of DNA contacts between *Map3k4*, *Qk*, and the *Airn* locus were relatively unchanged by *Airn* expression, and relative to surrounding regions, *Map3k4* and *Qk* resisted association with the *Airn* lncRNA.

Together, our data support the notion that spatial proximity to the *Airn* lncRNA product induces gene repression, and in most cases, also induces the accumulation of multiple forms of PRC1, PRC2, and PRC-directed chromatin modifications. *Airn* is a short-lived RNA that does not diffuse away from its site of transcription [12,14]. Thus, it follows that the regions that are most sensitive to repression by *Airn* are predisposed to contacting the *Airn* locus even in the absence of *Airn* expression. In turn, considering our findings along with prior data showing that the PRCs and their modifications to chromatin can induce DNA compaction [7,22,26], it seems likely that *Airn*-recruited PRCs and the modifications they deposit on chromatin are responsible for the major changes in chromatin architecture dependent on *Airn* expression. Such changes would presumably potentiate *Airn*-dependent repression, stabilizing the process by positive feedback.

Three notable regions within the *Airn* target domain – encompassing the genes *Slc22a3*, *Qk*, *Prr18/T/Pde10a* – exhibited augmented contacts with the *Airn* locus even in the absence of *Airn* expression. Of these regions, only *Slc22a3* formed a detectable DNA loop anchored at the *Airn* locus [43]. However, each of the regions harbored CGI promoters which themselves were associated with vPRC1, signatures of transcriptional activity, and nearby peaks of Cohesin. Prior works would suggest that any or all of these features could facilitate interactions between the regions and *Airn* DNA [26,31,44]. Upon expression of *Airn*, the region encompassing *Prr18/T/Pde10a* exhibited peak-like increases in contact with the *Airn* locus, while the intensity of detectable contacts with *Qk* and *Slc22a3* remained relatively unchanged. Deletion of specific DNA regulatory elements within the *Prr18/T/Pde10a* and *Slc22a3* regions affected the ability of

Airn to repress genes, induce PRC-directed modifications, and induce changes to chromatin architecture over megabases. Thus, DNA elements shape long-range contacts within the *Airn* target domain in ways that extend beyond single loop-based models of regulation [44].

Considered together, our data support the view that the extent of repression across the *Airn* target domain is governed by an equilibratory network of DNA regulatory elements that through direct or indirect means, control spatial proximity to the *Airn* lncRNA product (Figure 7E). Shifting the equilibrium in either direction has consequences on gene expression, chromatin modifications, and chromatin architecture. Indeed, we identified one CGI that appears to promote certain long-range contacts while restricting others (*Pde10a*), and another CGI that may promote long-range contacts by serving as a local Polycomb Response Element (*Slc22a3*; Figure 7E; [45]). To the latter point, a prior study demonstrated that *Airn* DNA preferentially interacts with the *Slc22a3* locus when the *Airn* lncRNA is not expressed [17], consistent with a model whereby *Airn*-induced, PRC-directed chromatin modifications at *Slc22a3* help to disengage the *Airn* locus from local DNA interactions and promote distal ones. Meanwhile, the CGI promoters of *Map3k4* and *Qk* gave the appearance of serving as boundary elements that attenuate local spread of repression by *Airn*. Our data suggest that variation in the genetic or epigenetic content of a DNA regulatory element has the potential to control gene expression by altering spatial equilibria between genes and locus control regions, be those repressors or enhancers [46-51]. By extension, unrecognized alterations to spatial equilibria that modulate contact with locus control regions may contribute to the challenge of assigning target genes to disease-associated SNPs [52].

Repression of the *Airn*-target gene *Igf2r* is due to the act of *Airn* transcription and does not depend on the *Airn* lncRNA product [15]. The mechanisms responsible for long-range repression by *Airn* remain unclear. Our observation that *Airn* lncRNA association and PRC-

directed chromatin modifications correlate in lock-step over a 15Mb domain in both TSCs and ESCs, together with data from [14,17,18], support the idea that *Airn* is a potent *cis*-acting repressive lncRNA that recruits the PRCs and possibly other repressive enzymes to chromatin. Future studies are needed to definitively prove this notion and demonstrate the mechanism by which the *Airn* lncRNA might mediate its long-range repressive effects.

2.4 Limitations of the Study

Considering that *Airn* acts largely if not exclusively *in cis*, F1-hybrid TSCs harbor an internal control in all genomic profiling experiments, in the form of the allele on which *Airn* is not expressed. Throughout our study, we interpret local differences in signal between *Airn*-expressing and non-expressing alleles to reflect *Airn*-dependent effects. Within individual TSC lines, these differences can be interpreted as absolute. However, when comparing signal between different TSC lines, particularly on *Airn*-expressing alleles, our Hi-C, CHART, and almost all ChIPs were performed under conditions that enable us to comment on relative but not absolute differences. Also, while the model in Figure 7 provides a parsimonious explanation of our data, our observations regarding 3D DNA contacts are correlative and we have not proven causation. It is also unclear to what extent *Airn* expression may induce changes in chromatin conformation independent of the PRCs. Lastly, the mechanisms by which *Airn* might recruit PRCs to chromatin are unclear. It is also unclear what molecular constituents occupy the interface between the *Airn* lncRNA and its genomic targets. If studies of *Xist* are any guide [53,54], it is likely that many hundreds of molecules of protein surround each molecule of *Airn*.

2.5 Methods

2.5.1 TSC culture

The mouse C/B TSC and B/C TSC lines used in this work correspond to the CAST/EiJ maternal/C57BL/6J paternal (C/B) and C57BL/6J maternal/CAST/EiJ paternal (B/C) TSCs used in [40,55] and are referred to as CB.1 and BC.1 TSCs in [55]. TSCs were cultured as in [56]. Briefly, TSCs were cultured on gelatin-coated, pre-plated irradiated mouse embryonic fibroblast (irMEF) feeder cells in TSC media (RPMI [Gibco, cat #: 11875093], 20% qualified FBS [Gibco, cat #: 26140079], 0.1mM penicillin-streptomycin [Gibco, cat #: 15140122], 1mM sodium pyruvate [Gibco, cat #: 11360070], 2mM L-glutamine [Gibco, cat #: 25030081], 100μM β-mercaptoethanol [Sigma-Aldrich, cat #: 63689]) supplemented with 25ng/mL FGF4 (Gibco, cat #: PHG0154) and 1μg/mL Heparin (Sigma-Aldrich, cat #: H3149) just before use, at 37°C in a humidified incubator at 5% CO_2. At passage, TSCs were trypsinized with 0.125% Trypsin-EDTA in PBS solution (Gibco, cat #: 25200-072) for ~4 minutes at room temperature and gently dislodged from the plate with a sterile, cotton-plugged Pasteur pipette. To deplete irMEFs from TSCs prior to all harvests, TSCs were pre-plated for 45 minutes at 37°C, transferred to a fresh culture plate, and then cultured for three days in 70% irMEF-conditioned TSC media supplemented with growth factors as above.

2.5.2 ESC culture

Mouse E14 ESCs were cultured on gelatin-coated plates in ESC media (DMEM high glucose and sodium pyruvate [Gibco, cat #: 11995073], 15% qualified FBS, 0.1mM MEM non-essential AA [Gibco, cat #: 11140050], 0.1mM penicillin-streptomycin, 2mM L-glutamine, 100μM β-mercaptoethanol, 1:500 LIF) at 37°C in a humidified incubator with 5% CO_2. At

passage, ESCs were trypsinized with 0.125% Trypsin-EDTA in PBS solution for ~5 minutes at room temperature and dislodged from the plate at single-cell suspension. ESCs were passaged every other day and provided fresh media daily.

2.5.3 Generation of regulatory element deletions

Per regulatory element deletion, four unique sgRNAs were designed using CRISPOR [57], with two sgRNAs flanking the target site (Figure S2A-D, Table S5). As a negative control, an sgRNA using a non-targeting (NTG) sequence from (Invitrogen, cat #: A35526) was designed. Each sgRNA was cloned into the *BsmbI* site of the piggyBac-cargo rtTA vector (PB_rtTA_BsmBI) from [58] and transformed in DH5-alpha competent bacterial cells. Starter transformant cultures for each sequence-verified sgRNA were pooled together in equal volume amounts prior to liquid culture expansion and plasmid purification using the PureLink HiPure Plasmid Midiprep kit (Invitrogen, cat #: K2100004). The pooled sgRNAs were then co-electroporated with doxycycline-inducible Cas9-cargo (PB_tre_Cas9) and pUC19-piggyBac transposase vectors from [58] at an 8:2:1 plasmid ratio of 2.5µg total DNA into 1 million C/B TSCs on irradiated drug-resistant MEFs (irDR4-MEFs; ATCC, cat #: SCRC-1045) in a single well of a 6-well plate. The electroporations were performed using a Neon® Instrument (electroporation settings: 950V, 30ms, 2 pulses). Two days after electroporation, TSCs were selected with 150µg/mL hygromycin B (Corning, cat #: MT30240CR)) and 200µg/mL G418 (Gibco, cat #: 10131035) in irMEF-conditioned TSC media with growth factors for 11 days, followed by four days of 1µg/mL doxycycline treatment (Sigma-Aldrich, cat #: D9891) to induce Cas9 expression. 2,000 doxycycline-induced TSCs were then plated onto a pre-plated irMEFs 100-mm

dish for clonal selection and expansion. Prior to harvesting for genotyping assays, clonal TSC lines were passaged once off of irMEFs as above.

For genotyping assays, PCR was used to detect the presence or absence of target deletion DNA. "Wildtype" primers were designed to amplify either the flanking end or internal region of the deletion site. "Deletion" primers were designed to externally flank both ends of the deletion sites that would efficiently amplify a sizeable PCR product if the deletion occurred (Table S5). Sanger sequencing (Eton) was then used to detect the presence of informative B6/CAST SNPs in the PCR products for allelic identification.

2.5.4 Generation of Airn-overexpressing ESCs

750,000 ESCs were seeded in a single gelatin-coated well of a six-well plate, and the next day transfected with 2.5µg of an 8:2:1 plasmid ratio of piggyBac-cargo rtTA-*Airn* sgRNA, doxycycline-inducible piggyBac-cargo dCas9-VP160 (PB_tre_dCas9_VP160), and pUC19-piggyBac transposase from [58] using Lipofectamine 3000 (Invitrogen, cat #: L3000015) according to manufacturer instructions. The next day, transfected ESCs were selected with ESC media containing 150µg/mL hygromycin B and 200µg/mL G418 for 10 days, followed by 4 days of doxycycline treatment to induce *Airn* expression via dCas9-VP160 prior to harvests.

2.5.5 In situ Hi-C

Prior to crosslinking for Hi-C, TSCs were passaged once off of irMEFs as described above. TSCs were then trypsinized and washed once with PBS. 5-10 million cells were crosslinked in resuspension with 10mL of 1% formaldehyde (Thermo Scientific, cat #: 28906) in PBS solution for 10 minutes at room temperature, quenched with 200mM glycine for 5 minutes

at room temperature, and then washed twice with ice-cold PBS. Cells were then divided into 5 aliquots (1-2 million cells/aliquot), where one aliquot was used for each Hi-C experiment. Importantly, for the removal of all PBS washes and crosslinking solution, cells were spun at 160 x g for 5 minutes.

Hi-C libraries from C/B wildtype, B/C wildtype, and C/B *Airn* truncation TSCs were generated and sequenced as in the detailed protocol from [26], including DNA fragmentation with *MboI* and *MseI* restriction enzymes. Hi-C libraries from regulatory element deletion TSCs were generated using the Arima-HiC+ kit (Arima Genomics, cat #: A510008) according to the manufacturer instructions. Paired-end, 150-bp sequencing was performed using Illumina NovaSeq 6000 System.

2.5.6 (Calibrated) ChIP-Seq

Prior to crosslinking for ChIP, TSCs were passaged once off of irMEFs as above. For all ChIP experiments, except those for PRC components, adhered cells were crosslinked with 0.6% formaldehyde (Fisher Scientific, cat #: BP531-500) in RPMI media with 10% FBS for 10 minutes at room temperature, then quenched with 125mM glycine for 5 minutes at room temperature. Crosslinked cells were then washed twice with ice-cold PBS and scraped with ice-cold PBS with 0.05% Tween (Fisher Scientific, cat #: EW-88065-31) and PIC (Sigma Aldrich, cat #: P8340). The cells were then spun at 1,200 x g at 4°C to remove PBS, followed by resuspension in ice-cold PBS with PIC and divided into 10-million cell aliquots. For PRC component ChIPs, adhered C/B TSCs were crosslinked in PBS with 2mM DSG (disuccinimidyl glutarate; Thermo Scientific, cat #: 20593) for 45 minutes at room temperature and then in PBS with 1% formaldehyde (Thermo Scientific, cat #: 28906) for 15 minutes at room temperature.

Crosslinking was quenched with 200mM glycine for 5 minutes at room temperature. Cells were then washed, scraped, and aliquoted as above. All ChIPs were performed using 10 million cells, 10μL of antibody, and 30μL of Protein A/G agarose beads (Santa Cruz, cat #: sc-2003). Input chromatin was isolated accordingly to each antibody (see below) and sonicated to 100-500bp fragments using a Vibra-Cell VX130 (Sonics) with the following parameters: 8-10 cycles of 30% intensity for 30 seconds with 1 minute of rest on ice between cycles. Antibody-conjugated beads were prepared by incubating antibody with beads in 300μL Blocking Buffer (PBS, 0.5% BSA [Invitrogen, cat #: AM2616]) overnight at 4°C with rotation.

For H3K27me3 and H2AK119ub ChIPs, crosslinked TSCs were resuspended in 1mL Lysis Buffer 1 (50mM HEPES pH 7.5, 140mM NaCl, 1mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, PIC) and incubated with rotation for 10 minutes at 4°C. Cells were then resuspended in 1mL Lysis Buffer 2 (10mM Tris-HCl pH 8.0, 200mM NaCl, 1mM EDTA, 0.5 mM EGTA, PIC) for 10 minutes at room temperature. All buffer removal steps were performed with 5-minute 1,200 x g spins at 4°C. The extracted nuclei pellet was then resuspended and sonicated in 500μL Lysis Buffer 3 (10mM Tris-HCl pH 8.0, 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.1% sodium-deoxycholate, 0.5% N- lauroylsarcosine, PIC). Soluble chromatin was obtained with a 30-minute max speed spin at 4°C, mixed with 1% Triton X-100, and then incubated with pre-conjugated antibody beads overnight at 4°C with rotation. The ChIP beads were then washed five times in 1mL RIPA Buffer (50mM HEPES pH 7.5, 500mM LiCl, 1mM EDTA, 1% NP-40, 0.7% sodium-deoxycholate, PIC) and once with 1mL TE, each for 5 minutes at 4°C with rotation and spun at 2,000 x g for 2 minutes for buffer removal.

For PRC component ChIPs, crosslinked C/B TSCs were resuspended and sonicated in 500μL Low Salt Pol II ChIP Buffer (50mM Tris-HCl pH 7.5, 140mM NaCl, 1mM EDTA, 1mM

EGTA, 0.1% sodium-deoxycholate, 0.1% SDS, PIC). Soluble chromatin was obtained with a 30-minute max speed spin at 4°C, mixed with 1% Triton X-100, and then incubated with pre-conjugated antibody beads overnight at 4°C with rotation. The ChIP beads were then washed three times with 1mL Low Salt Pol II ChIP Buffer with 1% Triton X-100 and PIC, once with 1mL High Salt Pol II ChIP Buffer (50mM Tris-HCl pH 7.5, 500mM NaCl, 1mM EDTA, 1mM EGTA, 0.1% sodium-deoxycholate, 0.1% SDS, PIC), once with 1mL LiCl Wash Buffer (20mM Tris-HCl pH 8.0, 250mM LiCl, 1mM EDTA, 0.5% Na-deoxycholate, 0.5% NP-40, PIC), and once with 1mL TE, each for 5 minutes at 4°C with rotation and spun at 2,000 x g for 2 minutes for buffer removal.

For all ChIP DNA elution steps, washed beads were resuspended in Elution buffer (50mM Tris-HCl pH 8.0, 10mM EDTA, 1% SDS) and placed on a 65°C heat block for 17 minutes with frequent vortexing. ChIP DNA was then reverse crosslinked in 0.5% SDS and 100mM NaCl overnight at 65°C, followed by a 1-hour RNaseA (3µL; Thermo Scientific, cat #: EN0531) treatment at 37°C and a 2.5-hour Proteinase K (10µL; Invitrogen, cat #: 25530015) treatment at 56°C. DNA was then extracted with 1 volume of phenol:chloroform:isoamyl alcohol (Sigma-Aldrich, cat #: P3803) and precipitated with 2 volumes 100% ethanol, 1/10 volume 3M sodium-acetate pH 5.4, and 1/1000 volume linear acrylamide (Invitrogen, cat #: AM9520) overnight at -20°C. Precipitated DNA was then extracted with a 30-minute max speed spin at 4°C, washed once with ice-cold 80% ethanol, and resuspended in TE.

Calibrated ChIPs were performed to validate *Airn*-induced changes across genotypes. At the IP step, 5% sonicated chromatin from HEK293T cells was added to a standardized protein amount of input TSC chromatin across samples. Bradford protein assays (Bio-Rad, cat #: 5000006) with BSA protein standard were performed to determine protein quantity as a proxy for

input chromatin amount. HEK293T chromatin was sonicated as above to achieve 100-500bp fragments. All other steps were performed as normal.

ChIP-Seq libraries were prepared with NEBNext End Repair Module (NEB, cat #: E6050S), A-tailing by Klenow Fragment (3'→5' exo-; NEB, cat #: M0212S), and TruSeq 6-bp index adaptor ligation by Quick ligase (NEB, cat #: M2200S), and NEBNext High-Fidelity 2X PCR Master Mix (NEB, cat #: M0541S). All DNA size-selection purification steps were performed with AMPure XP beads (Beckman Coulter, cat #: A63880). Single-end, 75-bp sequencing was performed using an Illumina NextSeq 500/550 High Output v2.5 kit (Illumina, cat #: 20024906) on a NextSeq 500 System.

2.5.7 CHART-Seq

CHART was performed in duplicate as in the detailed protocol from [59]. TSCs were passaged once off of irMEFs as above. *Airn* Highly-Expressing TSCs from [14] and ESCs were induced with 1μg/mL doxycycline 4 days prior to crosslinking. Adhered TSCs and ESCs were crosslinked with 1% formaldehyde (Fisher Scientific, cat #: BP531-500) in PBS solution for 10 minutes at room temperature. On ice, cells were washed twice with ice-cold PBS and twice with ice-cold PBS + 0.05% Tween before being scraped, spun at 1,000 x g for 5 minutes at 4°C, and divided into 25-million cell aliquots.

Per aliquot, nuclei was extracted with two rounds of douncing in 4mL sucrose buffer (300mM sucrose, 10mM HEPES pH 7.5, 100mM KOAc, 1% Triton X-100, 0.1mM EGTA, 0.5mM spermidine, 0.15mM spermine, cOmplete EDTA-free protease inhibitor cocktail [Millipore, cat #: 11873580001], 1mM DTT, 80U SUPERase-in [Invitrogen, cat #: AM2696]), mixing 1:1 with glycerol buffer (25% glycerol, 10mM HEPES pH 7.5, 100mM KOAc, 1mM

EDTA, 0.1mM EGTA, 0.5mM spermidine, 0.15mM spermine, cOmplete EDTA-free protease inhibitor cocktail, 1mM DTT, 80U SUPERase-in), and centrifugation through 4mL glycerol buffer at 1,000 x g for 15 minutes at 4°C. The nuclei pellet was then washed twice with ice-cold PBS + 0.05% Tween, then further crosslinked with 3% formaldehyde (Fisher Scientific, cat #: BP531-500) in PBS + 0.05% Tween for 30 minutes at room temperature. Formaldehyde was then washed out twice with ice-cold PBS + 0.05% Tween, then resuspended in freshly prepared 250µL Sonication Buffer (50mM HEPES pH 7.5, 75mM NaCl, 0.5% *N*-lauroylsarcosine solution, 0.1% sodium-deoxycholate, 0.1mM EGTA, cOmplete EDTA-free protease inhibitor cocktail, 1mMDTT, 300U SUPERase-in). Chromatin was sonicated to 2-10kb fragments using a Bioruptor Plus sonication device (Diagenode; sonication parameters: 30 sec on, 30 sec off cycles on high setting), then spun at max speed for 30 minutes at 4°C to retrieve soluble chromatin.

For *Airn* CHART, we designed 22-nucleotide complementary oligos that tile across the first 75 kb of the *Airn* RNA sequence using the ChIRP Probe Designer (LGC Biosearch Technologies) under parameters of high masking for specificity and ≥500-nt spacings (Table S5). The resulting 51 oligo probes were then mixed to a 100µM pool for in-house oligo biotin labeling [60]. Briefly, 20µM oligo probe mix was labeled with 300µM biotin-16-dUTP (Roche, cat #: 11093070910) and 30U terminal deoxynucleotidyl transferase (Thermo Scientific, cat #: EP0161), then labeled for 15 minutes at 37°C and inactivated for 20 minutes at 75°C. Biotinylated oligo probes were then purified using the QIAquick Nucleotide Removal kit (Qiagen, cat #: 28304), eluting to ~20µM.

Per *Airn* CHART reaction, 12.5 million cells worth of chromatin was mixed with 0.5 volume PAB (8M Urea, 100mM HEPES pH 7.5, 200mM NaCl, 2% SDS) and 1.5 volumes of freshly prepared Hybridization Buffer (1.5M NaCl, 1.12M Urea, 10X Denhardt's solution

[Invitrogen, cat #: 750018], 10mM EDTA), then pre-cleared with 50µL Dynabeads M-280 Streptavidin beads (Invitrogen, cat #: 11205D) for 1 hour at room temperature with rotation. The pre-cleared chromatin was then isolated from the beads by spinning at 1,000 x g for 30 seconds. 1% of pre-cleared chromatin was saved as "Input" sample, and the remaining pre-cleared chromatin was incubated with 750pmol biotinylated oligo probes overnight at room temperature. The next day, the sample was spun at max speed for 20 minutes at 20°C, and the supernatant was incubated with 200µL worth of Dynabeads MyOne Streptavidin C1 beads (Invitrogen, cat #: 65001) resuspended in 125µL 2:1 diluted PAB overnight at room temperature with rotation. CHART beads were then placed on the magnet and washed 4 times with 400µL CHART Wash Buffer (250mM NaCl, 10mM HEPES pH 7.5, 2mM EDTA, 2mM EGTA, 0.2% SDS, 0.1% *N*-lauroylsarcosine) and once with RNase H Elution Buffer (75mM NaCl, 50mM HEPES pH 7.5, 0.02% sodium-deoxycholate, 0.1% *N*-lauroylsarcosine, 10mM DTT, 3mM $MgCl_2$, 200U SUPERase-in). Beads were resuspended in 100µL RNase H Elution Buffer and treated with 2µL RNase H (NEB, cat #: M0297) for 10 mins at room temperature. To stop the RNase H reaction, 1/4 volume of naXLR (166.7mM Tris-HCl pH 7.2, 1.67% SDS, 83.3mM EDTA, 600µg Proteinase K [Bioline, cat #: BIO-37084]) was added, and the CHART eluate was then subject to proteinase K digestion and reverse-crosslinking for 1 hour at 55°C, followed by 1 hour at 65°C. The sample was then split for DNA (90%) and RNA (10%) analysis.

The *Airn* CHART-enriched RNA sample was treated with 1mL TRIZol and 200µL chloroform, and the RNA was DNase-treated and purified with Zymo Research RNA Clean & Concentrator kit (Zymo Research, cat #: 50-125-1669). To determine the extent of *Airn* RNA enrichment, 50% of both Input and CHART RNA samples were reverse transcribed using High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, cat #: 4368814). qPCR was

performed using iTaq Universal SYBR Green Supermix (Bio-Rad, cat #: 1725125) and primers targeting 45 kb downstream of the *Airn* TSS and *Gapdh* (Table S5).

The *Airn* CHART-enriched DNA sample was extracted with 1 volume of phenol:chloroform:isoamyl and purified with ethanol precipitation and TE resuspension as above for ChIP DNA. qPCR was performed as above with the same primers to check *Airn* DNA enrichment.

CHART-Seq libraries were prepared as above for ChIP-Seq libraries. However, prior to library prep, CHART-enriched DNA samples were further sonicated to 100-500bp fragments using the Bioruptor Plus with 10 cycles of 30 sec on, 30 sec off on high setting, then purified with one round of 1:1 AMPure XP beads size selection purification. Single-end, 75-bp sequencing was performed using an Illumina NextSeq 500/550 High Output v2.5 kit on a NextSeq 500 System.

2.5.8 RNA isolation, RT-qPCR, RNA-Seq

TSCs were passaged once off of irMEFs as above onto a single well of a 6-well plate. ESCs were grown on a single well of a 6-well plate. Both were grown to ≥75% confluency prior to RNA harvest using 1mL TRIzol, followed by the addition of 200µL chloroform, which were vortexed and subsequently spun at max speed for 5 minutes at 4°C for phase separation. The aqueous layer was collected and combined with 1 volume of 100% isopropanol and 5µL linear acrylamide. Precipitation was achieved at -80°C for 1 hour, followed by a max speed spin for 30 minutes at 4°C and one wash of the RNA pellet with ice-cold 80% ethanol. The pellet was then resuspended in 100µL H_2O and quantified via Qubit (Invitrogen, cat #: Q32855).

For RT-qPCR assays in Figures 5A and S4E, 1μg of RNA was reverse transcribed using the High-Capacity cDNA Reverse Transcription Kit, and qPCR was performed using iTaq Universal SYBR Green Supermix and custom primers (Table S5).

RNA-Seq libraries were prepared from 1μg of total RNA using KAPA RNA HyperPrep Kit with Ribose Erase (Kapa Biosystems, cat #: KR1351) according to the manufacturer instructions. Single-end, 75-bp sequencing was performed using an Illumina NextSeq 500/550 High Output v2.5 kit on a NextSeq 500 System.

2.5.9 Sequence alignment and processing

All mouse reference NCBI build 37/mm9 genome annotations were obtained from the UCSC genome browser [61]. Variant sequence data was obtained from the Sanger Institute (http://www.sanger.ac.uk/resources/mouse/genomes/; [25]. The CAST/EiJ (CAST) pseudogenome creation was performed as in [40,55]. Hi-C reads were aligned using BWA as a part of the Juicer pipeline (Durand et al. 2016a). ChIP- and CHART-Seq reads were aligned using bowtie2 with default parameters [62]. RNA-Seq reads were aligned using STAR with default parameters [63].

For Hi-C analyses in this study, read pairs that had a mapping quality greater than or equal to 10 were used for allelic TSC analysis and read pairs that had greater than or equal to 30 were used for ESC analysis. For all ChIP-, CHART- and RNA-Seq analyses in this study, reads that had a mapping quality greater than or equal to 30 were extracted with samtools [64], and allele-specific read retention (i.e., reads that overlap at least one B6 or CAST SNP) was performed as in [40,55] using a custom perl script (intersect_reads_snps16.pl: see github).

2.5.10 Chromosome tiling density plots

For all chromosome-scale tiling density plots in Figures 2-6 and S2-6, reads were summed in 10kb bins across each chromosome. MAPQ≥30 aligned reads were then divided by the total number of reads in the dataset and divided by a million (i.e., RPM). For allelic TSC ChIP- and CHART-Seq data, binned counts were divided by the number of B6/CAST SNPs detected in the bin genomic coordinates (i.e., SNP-norm RPM). Finally, bins were averaged every 9 bins in 1bin increments. For allelic TSC Hi-C viewpoint data, we excluded bins whose aggregate SNP-overlapping read count across merged Hi-C datasets fell within the bottom quintile relative to bins in the rest of the genome. The allelic data in this group of bins were too sparse to be interpreted with confidence. For the same reason, for allelic CHART- and ChIP-Seq data, only bins with greater or equal to 25 SNPs were plotted. Bins of replicate datasets were averaged.

For total, non-allelic ESC *Airn* CHART-Seq and H3K27me3 ChIP-Seq data, reads were RPM converted, binned, and averaged as above. For all ESC data in Figure 4, the same genomic bins as the allelic TSC ChIP- and CHART-Seq data were plotted.

All plots were generated using ggplot2 [65] in RStudio.

2.5.11 Tiling density correlations

To derive Spearman's ρ and p values for tiling density correlations in Figures 2 and 4, reads were processed as described in "Chromosome tiling density plots" and every 10th bin along the genomic region of interest was correlated.

To determine significant changes in H3K27me3 and H2AK119ub density across genotypes in Figures 5 and S6, all binned SNP-norm RPM values over the genomic region of analysis were subjected to a Welch t-test in RStudio.

2.5.12 Genome browser density tracks

Wiggle density files were created using a custom perl script (bigbowtie_to_wig3_mm9.pl; see github) and loaded into a UCSC Genome Browser session to generate the graphics in Figures 3 and 7. All density tracks were auto-scaled to data view and set to a maximum windowing function with 3-pixel smoothing.

2.5.13 Hi-C analysis

2.5.13.1 Juicer processing, quality control, and allele-specific read retention

Hi-C analyses were carried out using a combination of Juicer (default parameters) and Hi-C Explorer [27,66,67]. For exact commands used, see github.

For quality control, Hi-C statistics of each dataset were generated by Juicer (Table S1) and were referenced to the standard guidelines in [26]. In addition, long-range DNA interactions (25kb-10Mb) were correlated between Hi-C replicate datasets at 10kb and 25kb resolutions using Hi-C Explorer's hicCorrelate [68] with the following parameters: --method=pearson --range 25000:10000000.

For allele-specific retention of Hi-C contacts, read pairs in which at least one read end overlaps a B6/CAST SNP were extracted from the Juicer output merged_nodups.txt file using a modified Juicer diploid script (juicer_diploid_v6.sh; see github).

2.5.13.2 2D contact heatmaps

Allelic 2D contact heatmaps in Figures 1, 6, 7, S1, S7 were generated with Juicebox [67]. Contact matrices of observed counts were viewed in KR (Knight-Ruiz) balance mode at 5kb and 50kb resolutions. Subtraction heatmaps, where maternal contacts were subtracted from paternal contacts, were viewed under the same conditions.

2.5.13.3 Loop calling

DNA loops were detected using the Significant Interaction Peak (SIP) caller [43] with the following parameters: -factor 4 -g 2.0 -t 2000 -fdr 0.05. The output finalLoops.txt file (i.e., a merged list of unique DNA loop anchors detected across 5k, 10kb, and 25kb resolutions) was used to determine loops over the 2D contact heatmaps in Juicebox for Figure 7.

2.5.13.4 "A" and "B" Compartmentalization

"A" and "B" chromosome compartments for each allele were delineated by eigenvector analysis using the Juicer *eigenvector* tool with the following parameters: -p KR 17 BP 50000. Extracted eigenvector values were then visualized and plotted with Juicebox in Figures 1 and 6.

2.5.13.5 Allelic viewpoints

Allelic viewpoints of locus-specific contact matrices were extracted at 10 or 25kb resolution for observed and observed-over-expected (O/E) counts using the Juicer *dump* tool with the following parameters: observed/oe NONE chr:start:end chr:start:end BP 10000/25000. If applicable, all viewpoint loci of interest were extended to 100kb lengths from their centers to

improve contact matrix coverage. Extracted counts were then processed and plotted as described in the Chromosome-scale tiling density plots section.

2.5.13.6 Correlation with FISH

Allelic observed contacts from *Airn* viewpoint data in C/B wildtype TSCs were summed over the genomic coordinates for each of the 9 FISH probes across the *Airn* target domain analyzed in [14]. For each allele, the summed Hi-C counts at the probe locations were then correlated by Spearman's test (using in RStudio) with the corresponding average distance to the *Airn* probe as measured by RNA/DNA co-FISH in C/B TSCs [14]. Scatter plots in Figure 2 were generated with ggplot2 in RStudio [65].

2.5.14 ChIP-Seq analysis

2.5.14.1 Calibrated ChIP-Seq spike-in normalization

Reads from calibrated ChIP-Seq samples were also aligned to human GRCh37/hg19 genome build. A normalization factor was calculated for each sample using the formula 1/h, where h is the number of hg19-aligned reads in millions, as described previously [69,70]. Raw B6 and CAST reads were then scaled by multiplying the corresponding normalization factor. For tiling density plots, binned spike-in normalized reads were divided by the total number of B6/CAST SNPs.

2.5.14.2 Peak calling

ChIP-Seq peaks were called from non-allelic reads against an H3 ChIP-Seq dataset (from [40]) using the MACS2 peak calling algorithm [71] with the following parameters: -g mm –broad –broad-cutoff 0.01.

2.5.14.3 Allelic enrichment at CpG islands and other features

A statistical permutation test was performed to determine how significantly enriched PRC components, CTCF, SMC1A/Cohesin, and epigenetic marks are at loci of interest relative to the rest of the genome (see Table S3). All datasets analyzed were generated from C/B TSCs. H3K27ac, H3K4me2, CTCF, and SMC1A data were generated in previous studies, as a part of [14,40]. If applicable, all genomic features of interest were standardized to 1.5kb lengths (i.e., the largest CGI of interest) relative to their center positions. Using bedtools' 'shuffle' [72], we created a list of 80,000 1.5kb regions randomly selected from within 'gene' coordinates from gencode.vM1.annotation.gtf [73] with 100kb extended start and end sites while excluding any regions that fell within 2.5kb of a region annotated by MACS as an H3K27me3 or PRC subunit peak. Shuffled coordinates were filtered to retain regions encompassing at least one B6/CAST SNP, leaving 67,262 shuffled regions. B6- and CAST-overlapping ChIP-Seq reads were then counted over the features of interest and shuffled coordinates using a custom script (ase_analyzer10_adj2.pl; see github), then divided by the number of B6/CAST SNPs detected in the genomic coordinates (SNP-norm counts). The features were then ranked by SNP-norm counts for each allele in each dataset (1 = highest allelic signal), and a percentile rank was used to determine an empirical p-value for allele-specific enrichment of the ChIP target at the loci of interest.

2.5.15 RNA-Seq analysis

For non-allelic expression analysis in Table S2, featureCounts [74] was used to count reads over 'gene' entries in gencode.vM1.annotation.gtf [73]. Counts were then divided by total reads in the dataset and divided by a million (RPM).

For allelic expression analysis in Table S2 and Figures 5, S6, a custom perl script (ase_analyzer10.pl; see github) was used to count B6- and CAST SNP-overlapping reads over 'gene' entries in gencode.vM1.annotation.gtf [73]. Read counts were divided by the total number of reads in the dataset, divided by a million, and divided by the number of SNPs detected within the 'gene' coordinates (SNP-norm RPM). To determine the relative paternal expression of *Airn* in Figure S5E, paternal SNP-norm RPM values over *Airn* for each genotype were divided by was divided by the averaged NTG value from all four NTG clone data. To determine the relative paternal bias of *Airn* target gene expression for each genotype in Figure 5, paternal SNP-norm RPM values for the 27 *Airn* target genes (Table S2; [14]) were divided by the sum of the paternal and maternal values. The paternal biases for each genotype were then divided by the averaged NTG value from all four NTG clone data, and a Welch t-test was used to determine the p-value of the relative change in expression relative to NTG in RStudio. Boxplots of these data were generated using GraphPad Prism v9.

2.5.16 Quantification and Statistical Analysis

In Figure 2B, we used Spearman's Rank Correlation to determine the relationship between KR-balanced Hi-C counts with the *Airn* viewpoint (normalized for SNP density) versus our previous measurements of spatial distance to the *Airn* locus made by DNA FISH in [14]. In Figure 2F, we used Spearman's Rank Correlation to determine the relationship between KR-

balanced Hi-C counts with the *Airn* viewpoint (normalized for SNP density) and the density of H3K27me3 and H2AK119ub on the paternal allele. In Figure 3, we used an empirical sampling of ChIP-Seq read density in intergenic regions to determine the likelihood that each analyzed histone modification and chromatin-associated factor was enriched within a given region above what we would have expected by chance ($p<0.05$). In Figures 4B-D, we used Spearman's Rank Correlation to determine the relationship the density of *Airn* CHART signal on the paternal allele and H3K27me3, RING1B, and EZH2 signal, respectively, also on the paternal allele. In Figure 4F, we used Spearman's Rank Correlation to determine the relationship the density of *Airn* CHART signal and H3K27me3 in ESCs (non-allelic). In Figure 5E, we used a Welch t-test to determine the likelihood ($p<0.05$) that the relative paternal bias of 27 Airn target genes from [14] changed in each individual genotype relative to NTG control. In Figure 7D, we used a Welch t-test to determine the likelihood that the allelic H3K27me3 and H2AK119ub density was altered in the Δ*Slc22a3* and Δ*Pde10a* genotypes relative to the NTG control. Throughout our study, the term "n" refers to a biological replicate. Statistical analyses were performed in R. Software packages used are listed in the Key Resources Table. Additional details can be found in the figure legends, the body of the manuscript, and in the METHOD DETAILS section above.

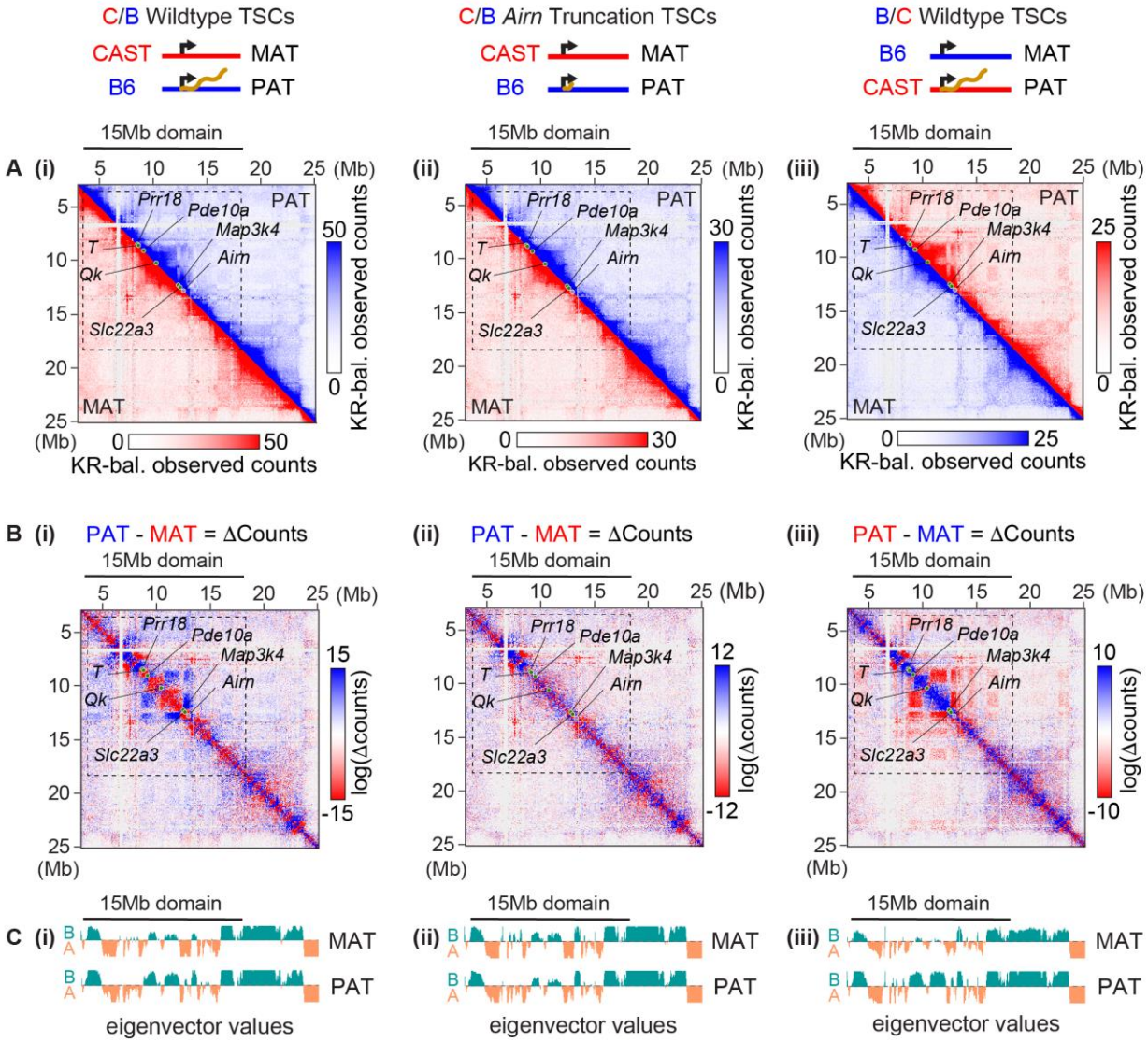


Figure 2.1 *Airn* expression induces large-scale changes to chromatin architecture. **(A)** Hi-C contact heatmaps of allelic observed counts in **(i)** C/B wildtype, **(ii)** C/B *Airn* truncation, and **(iii)** B/C wildtype TSCs, n = 2 or 3. Allelic heatmaps are partitioned and at 50kb resolution. KR bal., Knight-Ruiz balanced. **(B)** Subtraction contact heatmaps of log2 transformed [PAT minus MAT] observed counts, (i-iii) as (A)**. (C)** Eigenvectors at 50kb resolution for "A" and "B" chromosome compartmentalization. In heatmaps: dotted lines, 15Mb *Airn* target domain; purple circle, *Airn* gene; green circles, other loci of interest. See also Figure S1. See STAR Methods for detailed description of analyses. Datasets used are listed in Table S4.

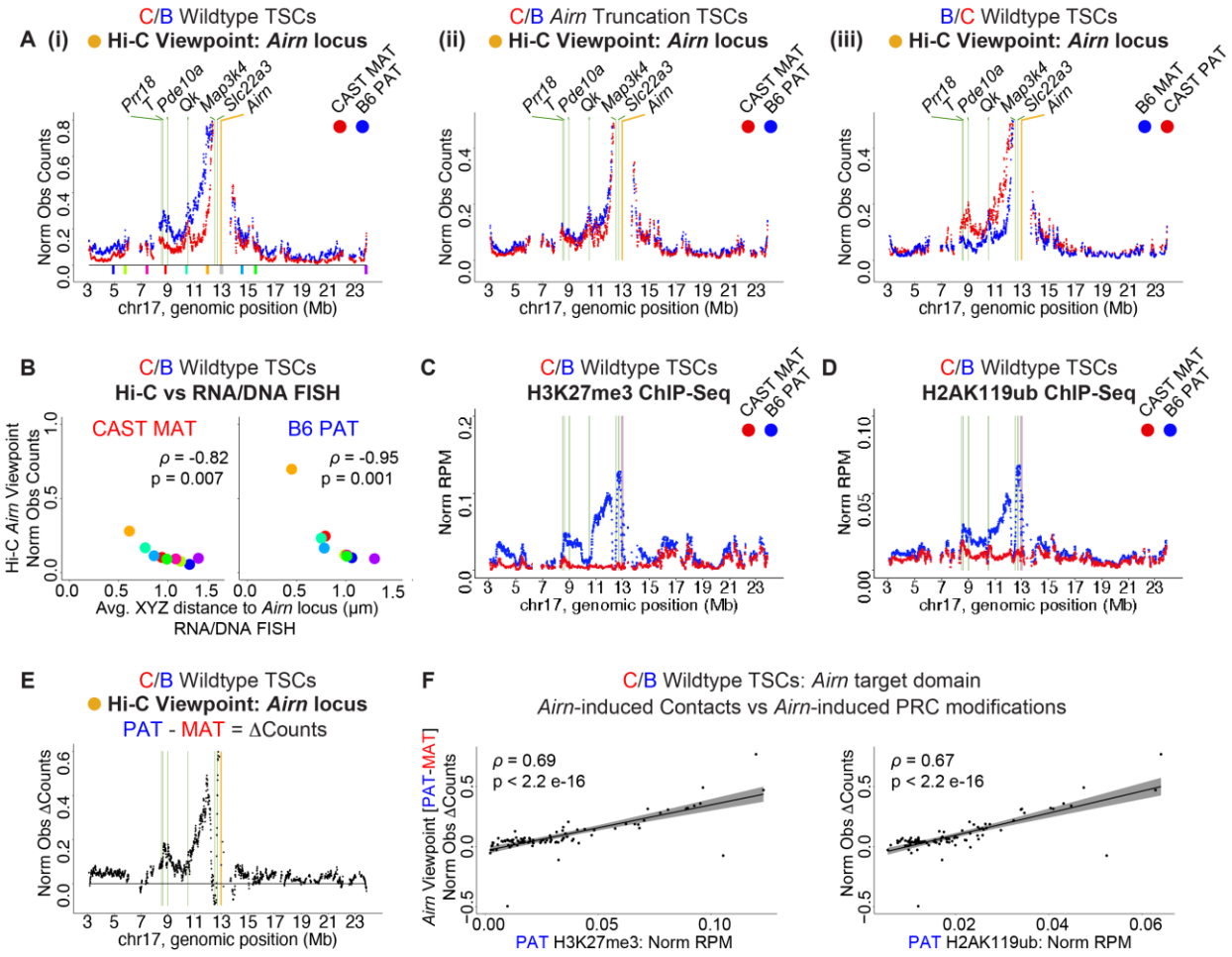


Figure 2.2 *Airn*-dependent changes in chromatin architecture coincide with the presence of PRC-deposited modifications.

(A) Tiling density plots of allelic Hi-C *Airn* viewpoint observed contact counts in **(i)** C/B wildtype, **(ii)** C/B *Airn* truncation, and **(iii)** B/C wildtype TSCs. Colored blocks in (i), FISH probes analyzed in (B; [14]). Norm Obs Counts, counts normalized for SNP density. **(B)** Allelic Hi-C *Airn* viewpoint contacts from (A, panel (i)) vs average distance to *Airn* measured by RNA/DNA FISH in C/B wildtype TSCs from [14]. Spearman's ρ and p values are shown. **(C, D)** Tiling density plot of allelic **(C)** H3K27me3 ChIP-Seq and **(D)** H2AK119ub ChIP-Seq signal in C/B wildtype TSCs. Data from (Schertzer et al. 2019a), n = 4 and 2. Norm RPM, Reads per Million total reads normalized for SNP density. **(E)** Tiling density plot of allelic *Airn* viewpoint [PAT minus MAT] observed counts in C/B wildtype TSCs from (A, panel (i)). **(F)** Scatter plots of *Airn* viewpoint [PAT minus MAT] observed counts vs (left) H3K27me3 and (right) H2AK119ub. Spearman's ρ and p values are shown. In tiling plots: yellow/purple bar, *Airn* viewpoint/gene; green bars, other loci of interest. See also Figure S2. See STAR Methods for detailed description of analyses. Datasets used are listed in Table S4.

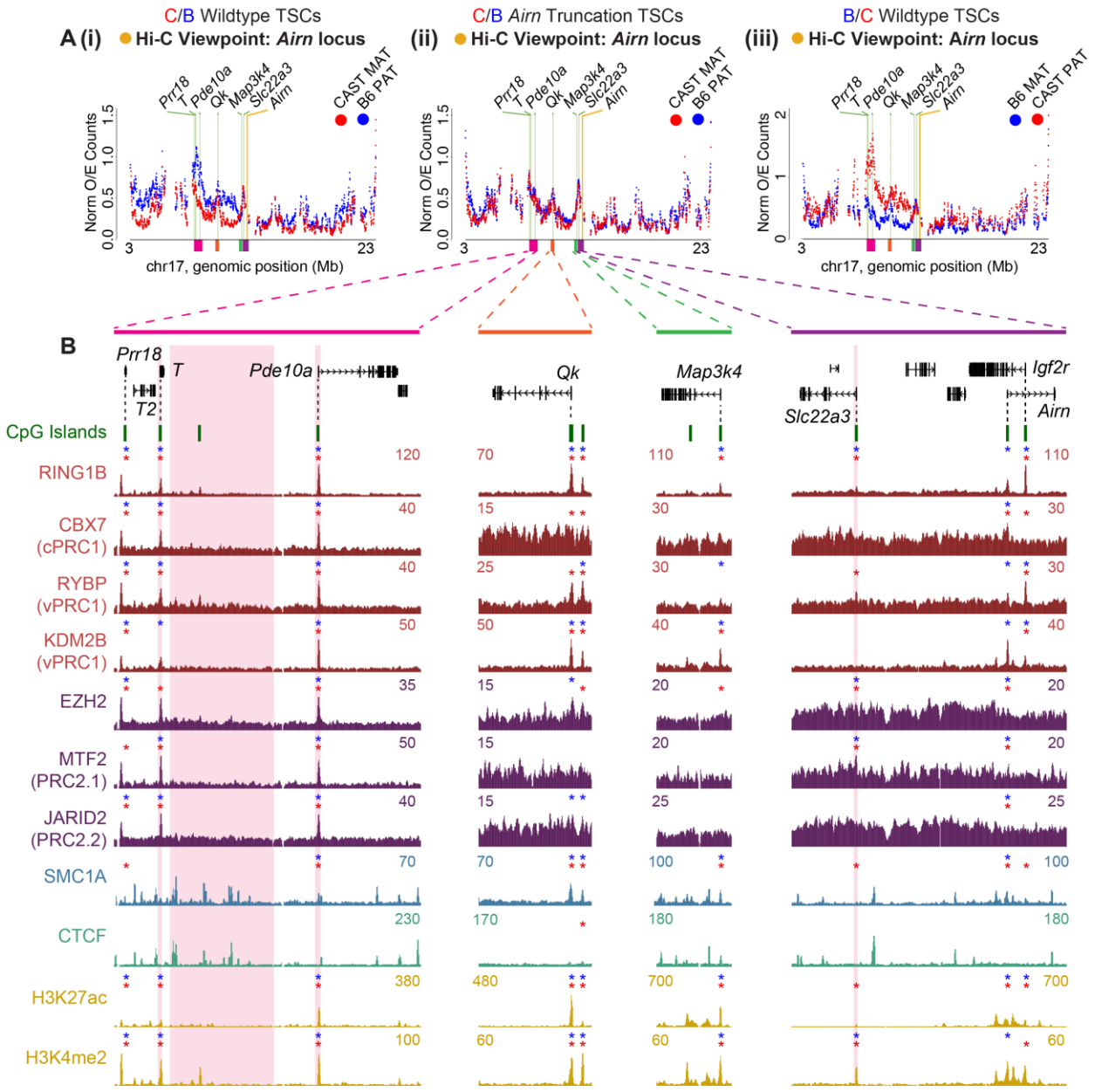


Figure 2.3 *Airn*-dependent repression centers around regions that form pre-existing contacts with the *Airn* locus and harbor CGIs bound by vPRC1.

(A) Tiling density plots of allelic Hi-C *Airn* viewpoint Observed-over-Expected (O/E) counts in **(i)** C/B wildtype, **(ii)** C/B *Airn* truncation, and **(iii)** B/C wildtype TSCs. Y-axes as in Figure 2. Yellow bar, viewpoint. Green bars, other loci of interest. **(B)** Genome browser graphics of regions harboring peaks of O/E contact. ChIP-Seq tracks, non-allelic read density from C/B wildtype TSCs, n = 2 or 3. Red or blue asterisks, significant enrichment on maternal or paternal alleles, respectively ($p<0.05$, permutation). See also Figure S4 and Table S3. Pink rectangles, DNA regions deleted in Figures 5-7. See STAR Methods for detailed description of analyses. Datasets used are listed in Table S4.

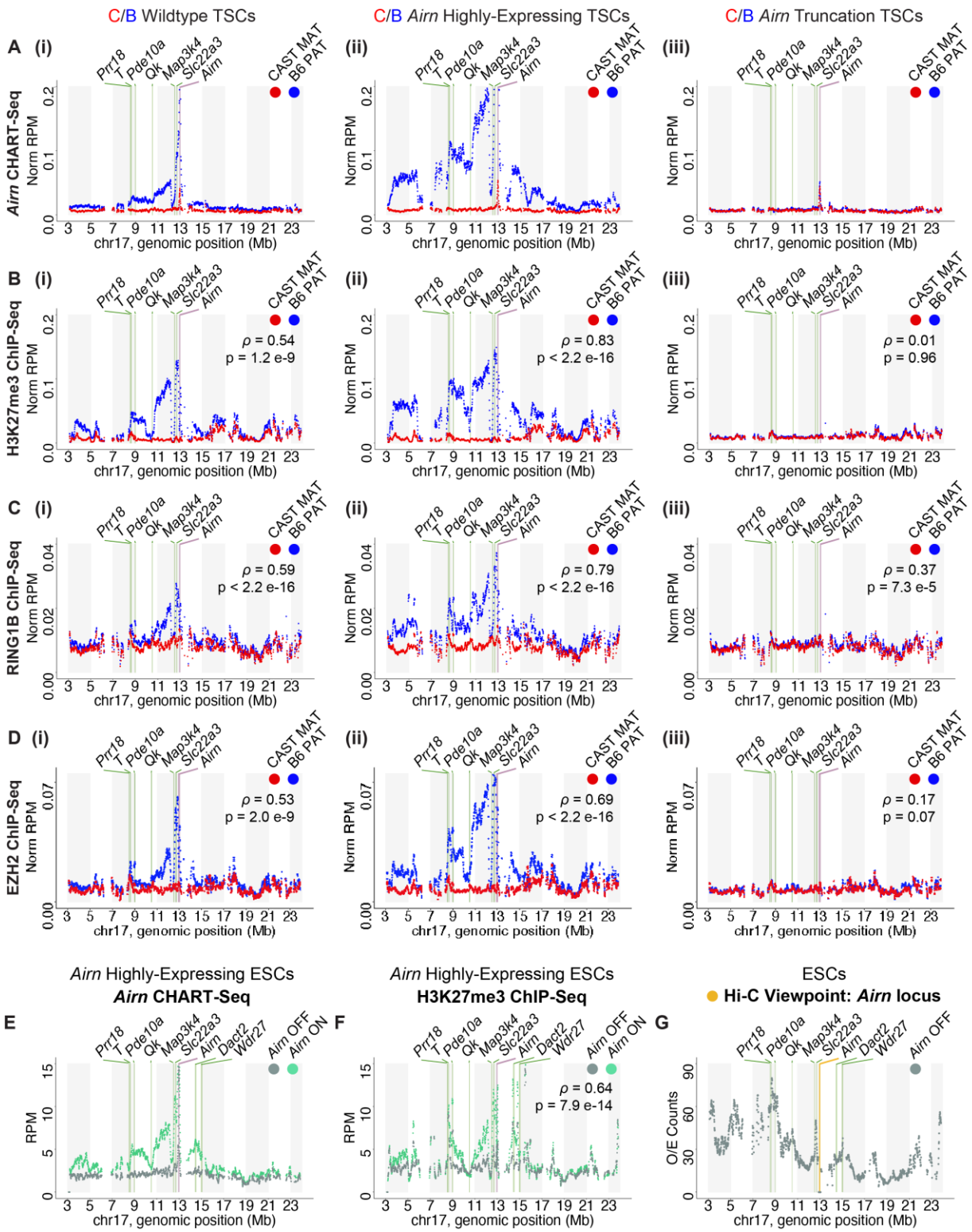

C/B Wildtype TSCs
C/B Airn Highly-Expressing TSCs
C/B Airn Truncation TSCs
A (i)
(ii)
(iii)
Airn CHART-Seq
Norm RPM
Prr18
T
Pde10a
Qk
Map3k4
Slc22a3
Airn
CAST MAT
B6 PAT
chr17, genomic position (Mb)
B (i)
(ii)
(iii)
H3K27me3 ChIP-Seq
ρ = 0.54
p = 1.2 e-9
ρ = 0.83
p < 2.2 e-16
ρ = 0.01
p = 0.96
C (i)
(ii)
(iii)
RING1B ChIP-Seq
ρ = 0.59
p < 2.2 e-16
ρ = 0.79
p < 2.2 e-16
ρ = 0.37
p = 7.3 e-5
D (i)
(ii)
(iii)
EZH2 ChIP-Seq
ρ = 0.53
p = 2.0 e-9
ρ = 0.69
p < 2.2 e-16
ρ = 0.17
p = 0.07
Airn Highly-Expressing ESCs
Airn CHART-Seq
Airn Highly-Expressing ESCs
H3K27me3 ChIP-Seq
ESCs
Hi-C Viewpoint: Airn locus
E
F
G
Dact2
Wdr27
Airn OFF
Airn ON
RPM
O/E Counts
ρ = 0.64
p = 7.9 e-14

Figure 2.4 Presence of *Airn* lncRNA on chromatin correlates with presence of PRC1 and PRC2 and centers around pre-existing contacts with the *Airn* locus.
(A-D) Tiling density plots of allelic **(A)** *Airn* CHART-Seq, **(B)** H3K27me3 ChIP-Seq, **(C)** RING1B ChIP-Seq, and **(D)** EZH2 ChIP-Seq signal in C/B **(i)** wildtype, **(ii)** *Airn* High-Expressing (H-E), and **(iii)** *Airn* truncation TSCs, n = 1 or 2. Y-axes as in Figure 2. H3K27me3 and RING1B ChIP-Seq data from [14]. **(E, F)** Tiling density plots of **(E)** *Airn* CHART-Seq and **(F)** H3K27me3 ChIP-Seq signal in untreated (*Airn* OFF) or dox-treated (*Airn* ON) H-E ESCs, n = 1. **(G)** Tiling density plot of Hi-C *Airn* viewpoint O/E counts in ESCs [42]. Yellow/purple bar, *Airn* viewpoint/gene; green bars, other loci of interest. See also Figure S2 and S5. See STAR Methods for detailed description of analyses. Datasets used are listed in Table S4.

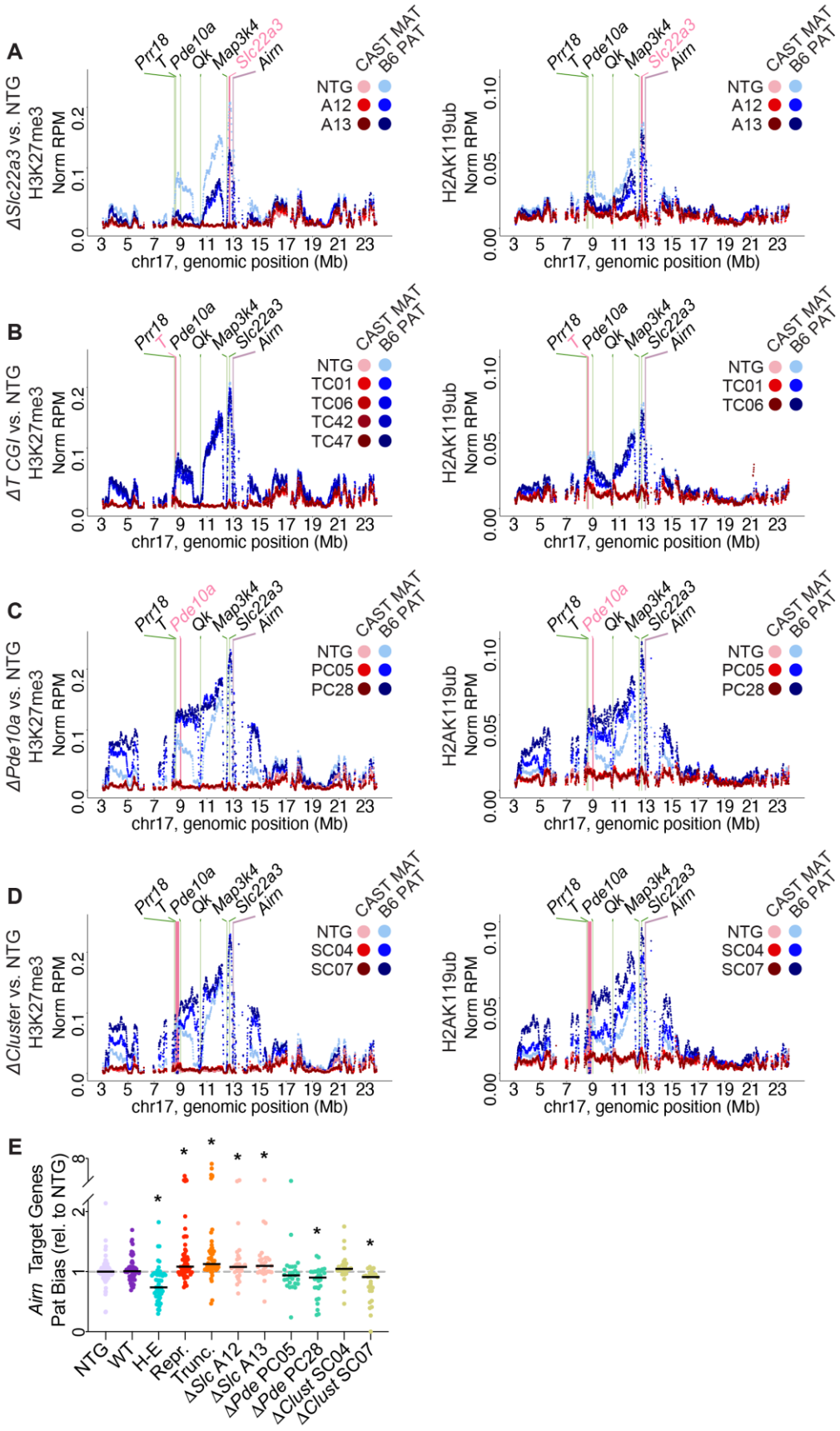

A
ΔSlc22a3 vs. NTG
H3K27me3
Norm RPM
H2AK119ub
Prr18
T
Pde10a
Qk
Map3k4
Slc22a3
Airn
CAST MAT
B6 PAT
NTG
A12
A13
chr17, genomic position (Mb)
B
ΔT CGI vs. NTG
TC01
TC06
TC42
TC47
C
ΔPde10a vs. NTG
PC05
PC28
D
ΔCluster vs. NTG
SC04
SC07
E
Airn Target Genes
Pat Bias (rel. to NTG)
NTG
WT
H-E
Repr.
Trunc.
ΔSlc A12
ΔSlc A13
ΔPde PC05
ΔPde PC28
ΔClust SC04
ΔClust SC07

Figure 2.5 DNA regulatory element deletions alter levels of PRC-directed modifications and gene repression throughout the *Airn* target domain.

(A-D) Tiling density plots of allelic (left) H3K27me3 and (right) H2AK119ub ChIP-Seq signal in C/B **(A)** Δ*Slc22a3*, **(B)** *ΔT*, **(C)** Δ*Pde10a* and **(D)** Δ*Cluster* TSCs, n = 1 per clonal line. Y-axes as in Figure 2. Data from NTG clones were averaged, n = 4. **(E)** Paternal expression of the 27 *Airn* target genes across genotypes relative to NTG, n = 1, 2, or 4. Asterisks, significant changes relative to NTG ($p<0.05$, Welch t-test). In tiling plots: purple bar, *Airn* gene; green bars, other loci of interest; pink bars, DNA regions deleted. See also Figure S6 and Table S2. See STAR Methods for detailed description of analyses. Datasets used are listed in Table S4.

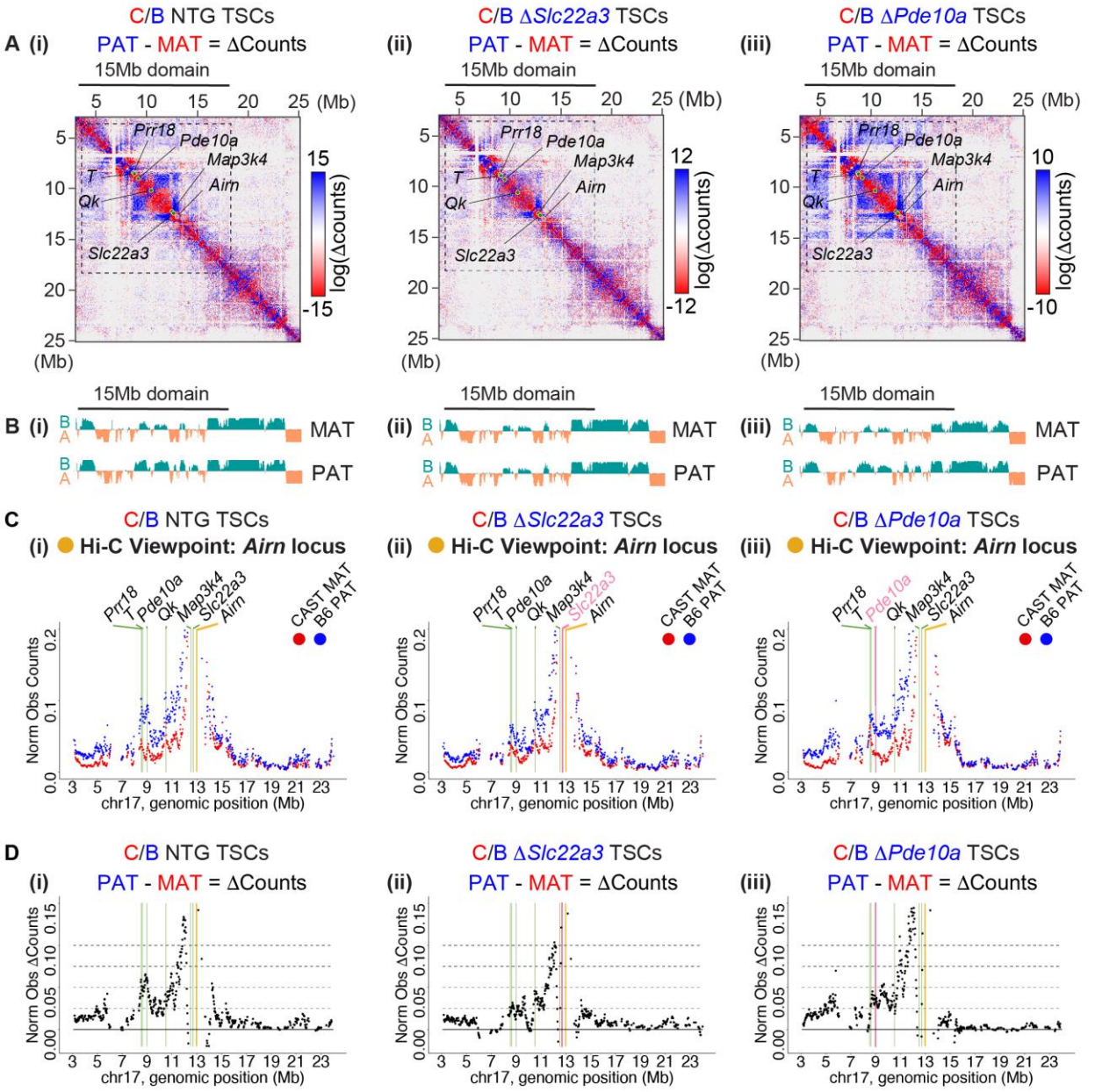


Figure 2.6 Changes in DNA contacts with *Airn* mirror changes in PRC activity caused by regulatory element deletion.

(A) Hi-C subtraction contact heatmaps of log2 transformed [PAT minus MAT] observed counts in C/B **(i)** NTG, **(ii)** Δ*Slc22a3*, and **(iii)** Δ*Pde10a* TSCs, n = 2. **(B)** Eigenvectors at 50kb resolution for "A" and "B" chromosome compartmentalization, (i-iii) as (A). **(C)** Tiling density plots of allelic *Airn* viewpoint observed contact counts, (i-iii) as (A). Y-axes as in Figure 2. **(D)** Tiling density plots of allelic *Airn* viewpoint [PAT minus MAT] observed counts, (i-iii) as (A). In heatmaps: dotted lines, 15Mb *Airn* target domain; purple circles, *Airn* gene; green circles, other loci of interest. In tiling plots: yellow bar, *Airn* locus viewpoint; green bars, other loci of interest; pink bar, DNA region deleted. See also Figure S7. See STAR Methods for detailed description of analyses. Datasets used are listed in Table S4.

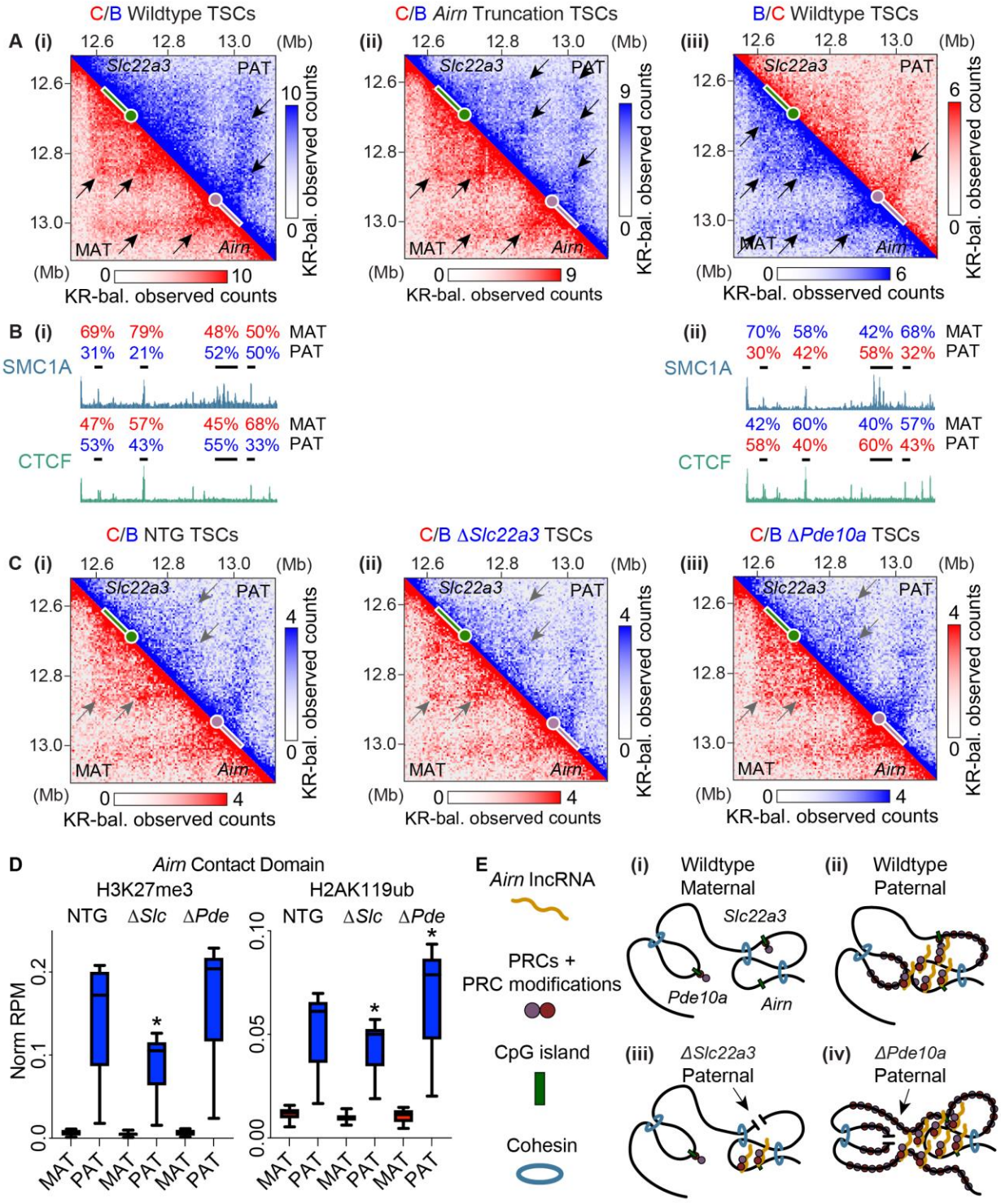


Figure 2.7 *Airn* expression is coincident with dissolution of DNA loops encasing *Slc22a3* and a local increase in PRC-directed modifications.

(A) Allelic Hi-C contact heatmaps of the *Airn* contact domain at 5kb resolution in **(i)** C/B wildtype, **(ii)** C/B *Airn* truncation, and **(iii)** B/C wildtype TSCs. Black arrows, SIP-called DNA loops. **(B)** SMC1A/Cohesin and CTCF ChIP-Seq. Percent parental biases at loops, (i and ii) as in

(A). **(C)** Contact heatmaps of *Airn* contact domain at 5kb resolution in C/B **(i)** NTG, **(ii)** Δ*Slc22a3*, and **(iii)** Δ*Pde10a* TSCs. Grey arrows, DNA loops of interest from (A). In heatmaps: circles, 5' end of genes/CGIs; rectangles, gene bodies; purple, *Airn* gene; green, *Slc22a3* gene. **(D)** Boxplots of average allelic (left) H3K27me3 and (right) H2AK119 ChIP-Seq signal over *Airn* contact domain in NTG, Δ*Slc22a3*, and Δ*Pde10a* TSCs, n = 2. Asterisks, significant changes relative to NTG ($p<0.05$, Welch t-test). Error bars, datapoints outside the interquartile range. **(E)** Model: DNA regulatory elements modulate frequency of contact with and repression by *Airn*. **(i)** On maternal allele, pre-existing contacts with *Airn* locus render certain regions more susceptible to repression than others. **(ii)** On paternal allele, PRCs and PRC-deposited modifications to chromatin increase contacts with *Airn* locus, which in turn increase intensity of long-distance repression. **(iii)** Loss of *Slc22a3* CGI attenuates repression by reducing local PRC recruitment and contact with distal regions. **(iv)** Loss of *Pde10a* CGI increases frequency with which surrounding regions contact *Airn*. See STAR Methods for detailed description of analyses. Datasets used are listed in Table S4.

2.6 Supplemental Table Legends

Supplemental Table 2.1. Hi-C statistics and quality control. Related to Figures 1, 2, 6, 7, S1, S4, S7. Table shows statistical summary of Hi-C datasets by total read depth, Juicer quality control statistics of Hi-C contacts, and HiCExplorer Pearson correlation of long-range contacts. Genotypes are separated by Sheets. Total read depth and Hi-C quality control sections were obtained from the Juicer output inter.txt (i.e., all read pairs) and inter_30.txt (i.e., MAPQ≥30 read pairs) files and referenced to the standard guidelines in [26]. Pearson correlations of long-range Hi-C contacts (>25kb) at 10 and 25kb resolutions were determined by Hi-C Explorer's hicCorrelate tool [68].

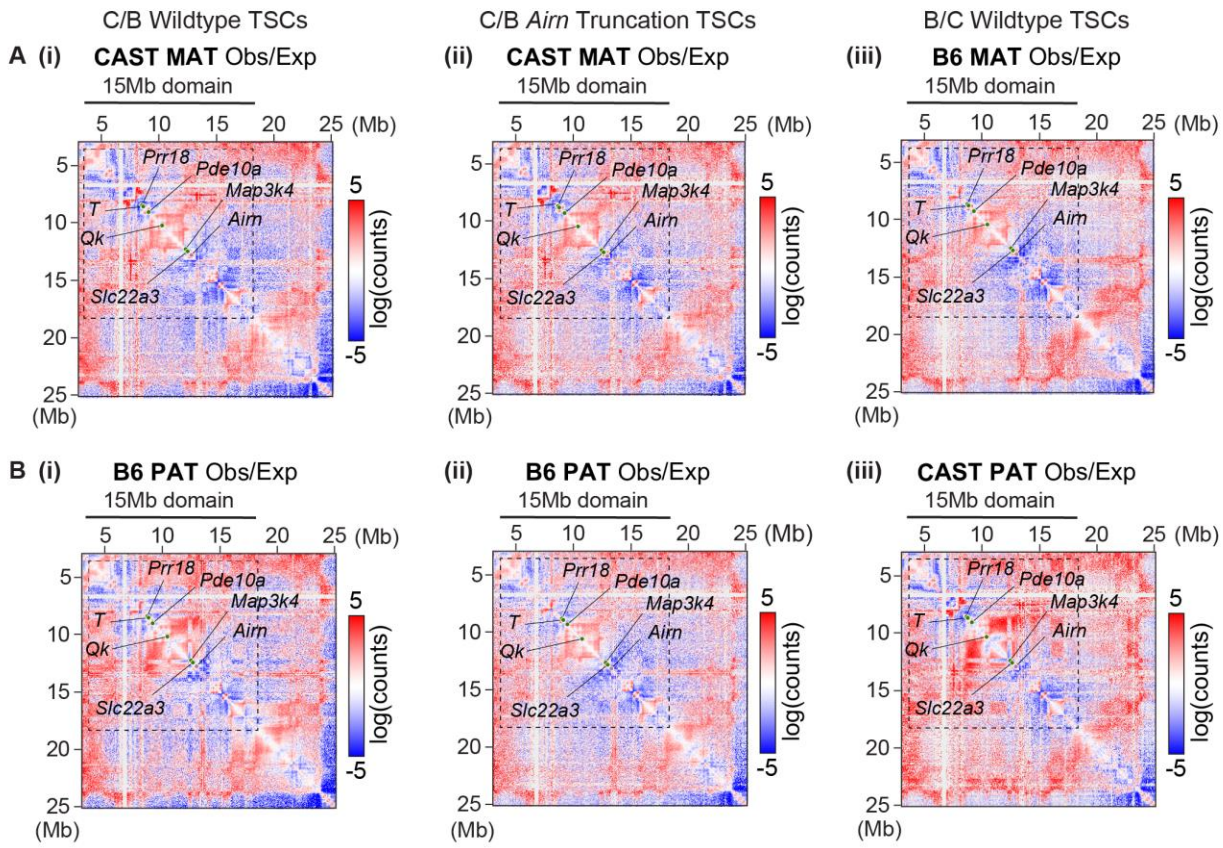
Supplemental Table 2.2. RNA-Seq gene expression changes of *Airn* target genes. Related to Figures 5E and S6E. Table shows gene expression changes across genotypes measured via RNA-Seq. Gene categories, separated by Sheets, include all chromosomes ('All Chr'), the 27 genes in the *Airn* target domain that significantly change between *Airn* highly-expressing and *Airn* truncation TSCs ('Airn Target Genes'; from [14]), and the CGI-promoter genes of interest ('CGI Genes'). All gene annotations are from gencode.vM1.annotation.gtf [73]. Reads aligning to introns are included. For total, non-allelic expression analysis, featureCounts [74] was used to count reads over exon coordinates (i.e., Counts, Total), and then divided by total reads in the dataset and multiplied by a million (i.e., RPM, Total). For allelic expression analysis, a custom perl script (ase_analyzer10.pl; see github) was used to count B6- and CAST SNP-overlapping reads over gene coordinates (i.e., Counts, CAST and B6).

Supplemental Table 2.3. Allelic detection of PRC components, architectural factors, and epigenetic marks. Related to Figure 3B. Table shows the statistical summary of a permutation test used to determine whether factors of interest were significantly enriched at loci of interest relative to dataset specific background. All ChIP-Seq data were acquired from C/B TSCs. H3K27ac, H3K4me2, CTCF, and SMC1A data were generated in previous studies [14,40]. If applicable, all genomic features of interest were standardized to 1.5kb lengths (i.e., the largest CGI of interest) relative to their center positions. A list of 80,000 1.5kb regions were randomly selected from within 'gene' coordinates from gencode.vM1.annotation.gtf [73] with 100kb extended start and end sites while excluding any regions that fell within 2.5kb of a region annotated by MACS as an H3K27me3 or PRC subunit peak. Shuffled coordinates were then filtered to retain regions encompassing at least one B6/CAST SNP, leaving 67,262 shuffled regions. B6- and CAST-overlapping ChIP-Seq reads were then counted over the features of interest and shuffled coordinates using a custom script (ase_analyzer10_adj2.pl; see github), then divided by the number of B6/CAST SNPs detected in the genomic coordinates (SNP-norm counts). The features were then ranked by SNP-norm counts for each allele in each dataset (1 = highest allelic signal), and a percentile rank was used to determine an empirical p-value for allele-specific enrichment of the ChIP target at the loci of interest. The raw allelic counts, rank, and empirical p-values for the genomic features of interest relative to the shuffled regions are separated by Sheets.

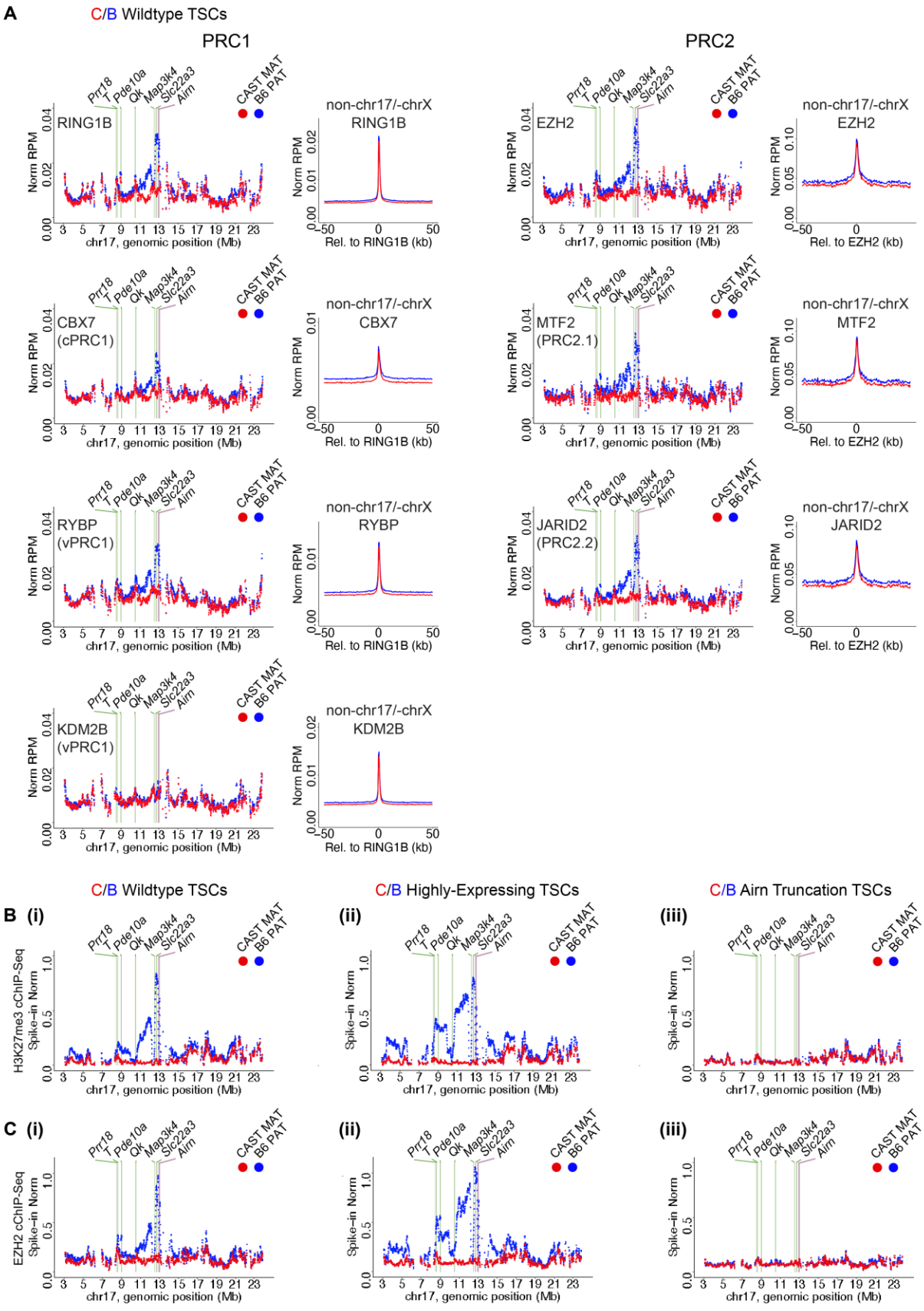
Supplemental Table 2.4. All high throughput sequencing datasets used. Related to Figures 1-7, S1-S7 and Tables S1, S2, S3. Table gives all high throughput sequencing genomic datasets used in this study and is divided into 2 sections: "Datasets generated in this study" and "Publicly available datasets". Under each section, if applicable: "File ID" gives the name of the dataset;

"Data Type" gives the experimental method (RNA-Seq, ChIP-Seq, CHART-Seq, Hi-C); "Cell Type" gives the cell type and strain information when relevant; "Spike-in" says whether ERCC or HEK293 chromatin spike-ins were included; "Read Length" describes read length and single versus paired end sequencing; "Figures and Tables" lists the figures and tables in the manuscript where each dataset was used; and "GEO" gives the GEO database reference for the data.

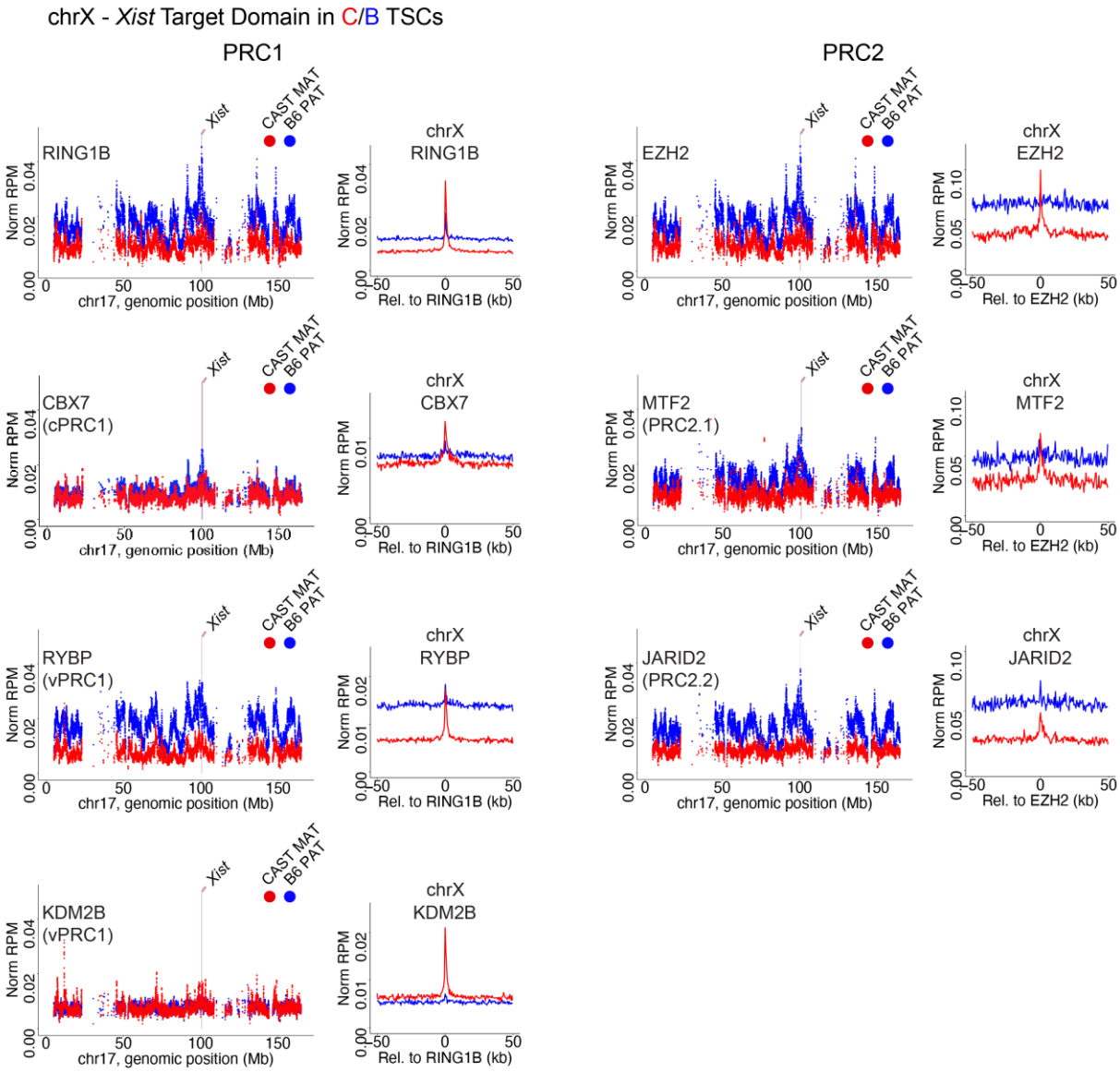
Supplemental Table 2.5. Oligonucleotides, Related to Figures 2.2, 2.4-2.7, Supplemental Figures 2.5-2.7. Table gives all oligonucleotide sequences used in this study. "ID" gives a descriptive name for the oligonucleotide; "Assay" describes the experimental method in which the oligonucleotide was used (PCR genotyping, qPCR, CRISPR, CHART); and "Sequence" gives the oligonucleotide sequence in 5' to 3' order.



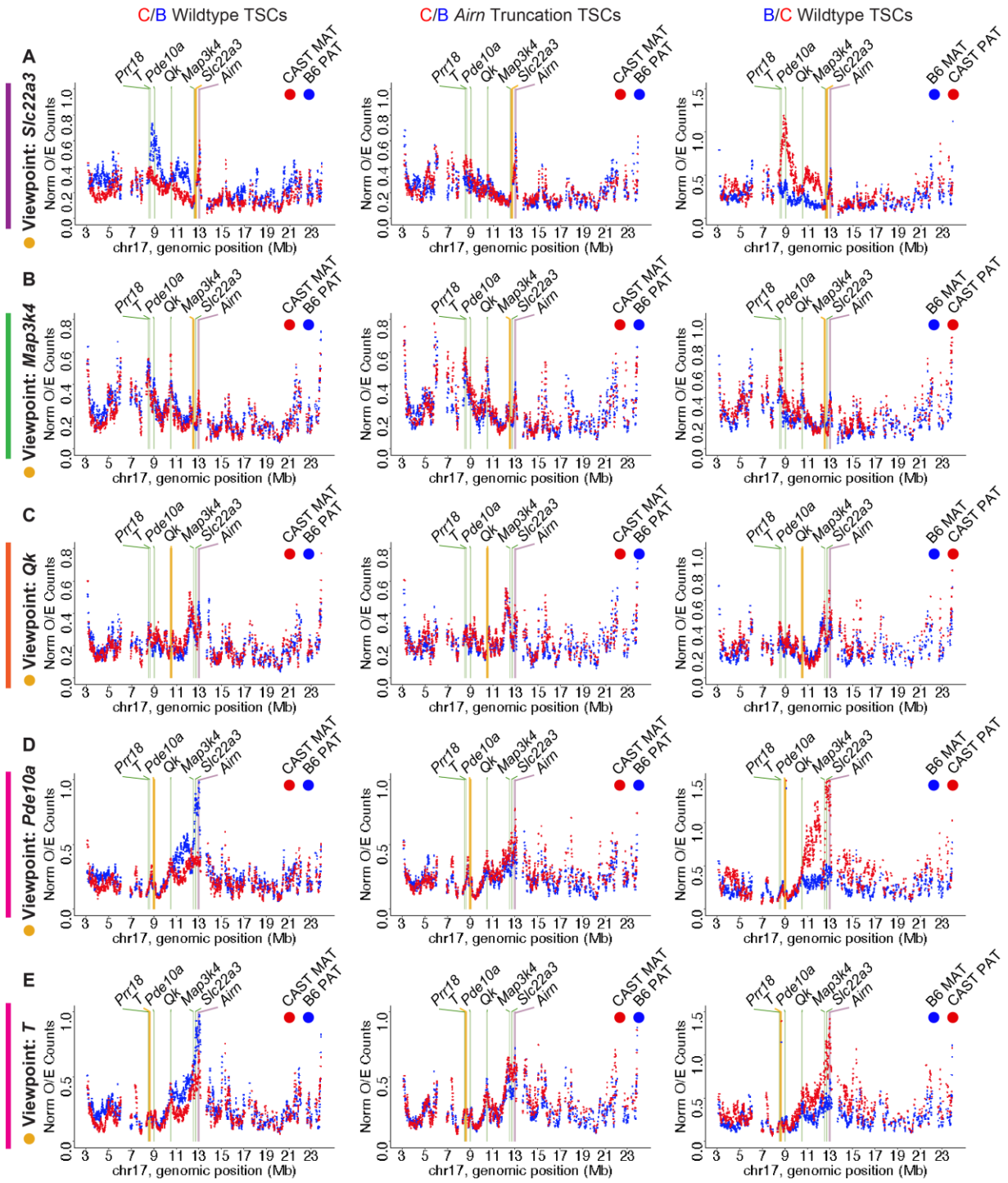
Supplemental Figure 2.1, Related to Figure 2.1.
(A, B) Hi-C contact heatmaps of **(A)** maternal and **(B)** paternal Observed-over-Expected (O/E) counts in **(i)** C/B wildtype, **(ii)** C/B *Airn* truncation, and **(iii)** B/C wildtype TSCs. Dotted lines, 15Mb *Airn* target domain. Purple circle, *Airn* gene. Green circles, other loci of interest. See STAR Methods for detailed description of analyses. Datasets used in this figure are listed in Table S4.


A
C/B Wildtype TSCs
PRC1
PRC2
RING1B
EZH2
CBX7
(cPRC1)
MTF2
(PRC2.1)
RYBP
(vPRC1)
JARID2
(PRC2.2)
KDM2B
(vPRC1)
Prr18
T
Pde10a
Qk
Map3k4
Slc22a3
Airn
CAST MAT
B6 PAT
Norm RPM
chr17, genomic position (Mb)
non-chr17/-chrX
Rel. to RING1B (kb)
Rel. to EZH2 (kb)
B (i)
C/B Wildtype TSCs
(ii)
C/B Highly-Expressing TSCs
(iii)
C/B Airn Truncation TSCs
H3K27me3 cChIP-Seq
Spike-in Norm
C (i)
EZH2 cChIP-Seq
Spike-in Norm

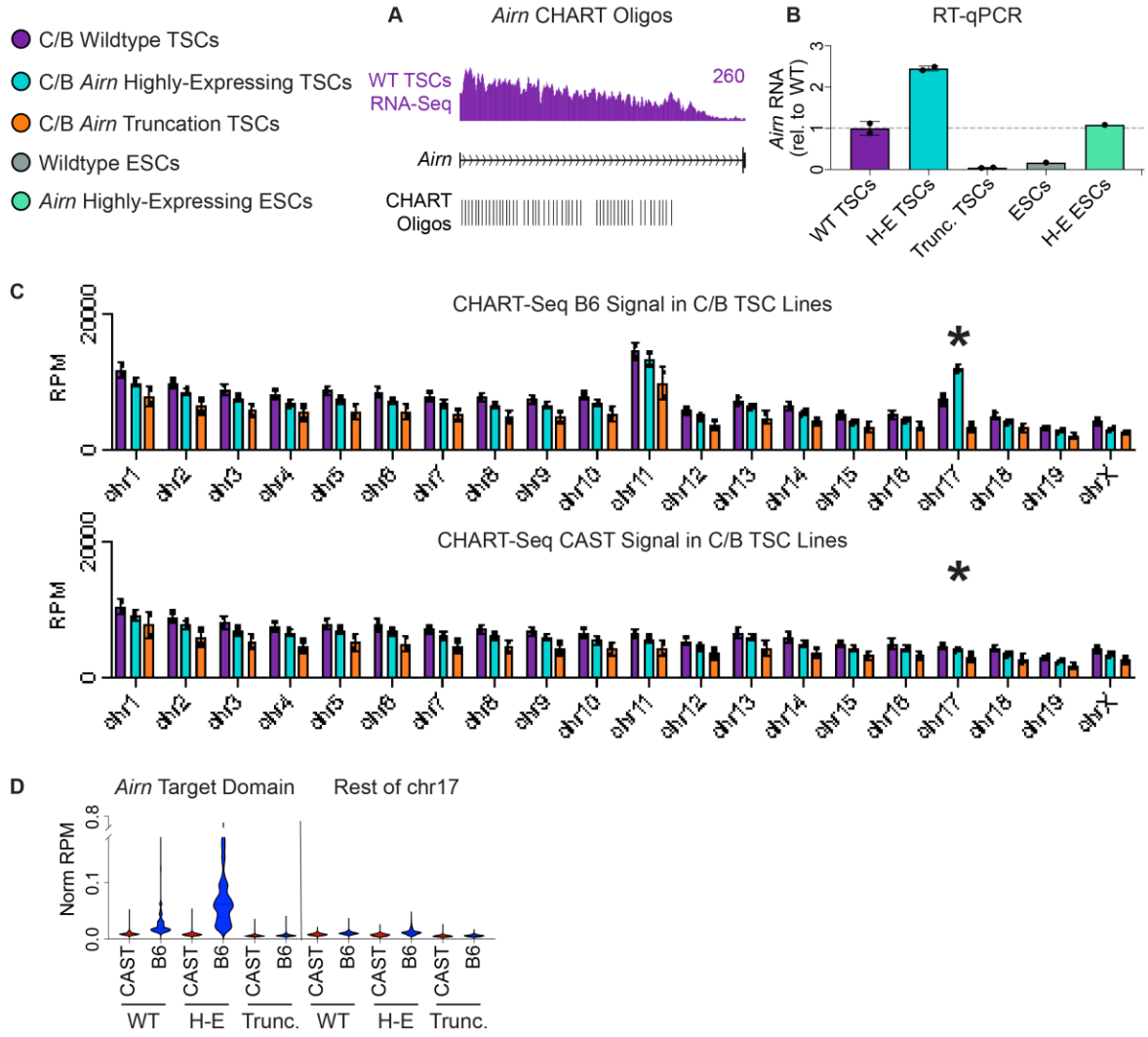
Supplemental Figure 2.2, Related to Figures 2.2 and 2.4.
(A) Tiling density plots of allelic ChIP-Seq signal for (left) PRC1 and (right) PRC2 components over the *Airn* target domain on chr17 in C/B wildtype TSCs, n = 2 or 3. Shown to right, metagenes of allelic signal for PRC subunits relative to all non-X, non-chr17 RING1B or EZH2 peaks. Y-axes as in Figure 2. **(B, C)** Tiling density plots of allelic **(B)** H3K27me3 and **(C)** EZH2 calibrated ChIP-Seq signal in **(i)** C/B wildtype, **(ii)** C/B *Airn* High-expressing (H-E), and **(iii)** C/B *Airn* truncation TSCs, n = 2. Spike-in Norm, allelic mouse reads scaled by human spike-in normalization factor and normalized for SNP density. In tiling plots: purple bar, *Airn* gene; green bars, other loci of interest. See STAR Methods for detailed description of analyses. Datasets used are listed in Table S4.



Supplemental Figure 2.3, Related to Figure 2.2.
Tiling density plots of average allelic ChIP-Seq signal for (left) PRC1 and (right) PRC2 components over chrX in C/B wildtype TSCs, n = 2 or 3. Shown to right, metagenes of allelic signal for PRC subunits relative to RING1B or EZH2 peaks on chrX. Y-axes as in Figure 2. In tiling plots: purple bar, *Xist* gene; green bars, other loci of interest. See STAR Methods for detailed description of analyses. Datasets used are listed in Table S4.

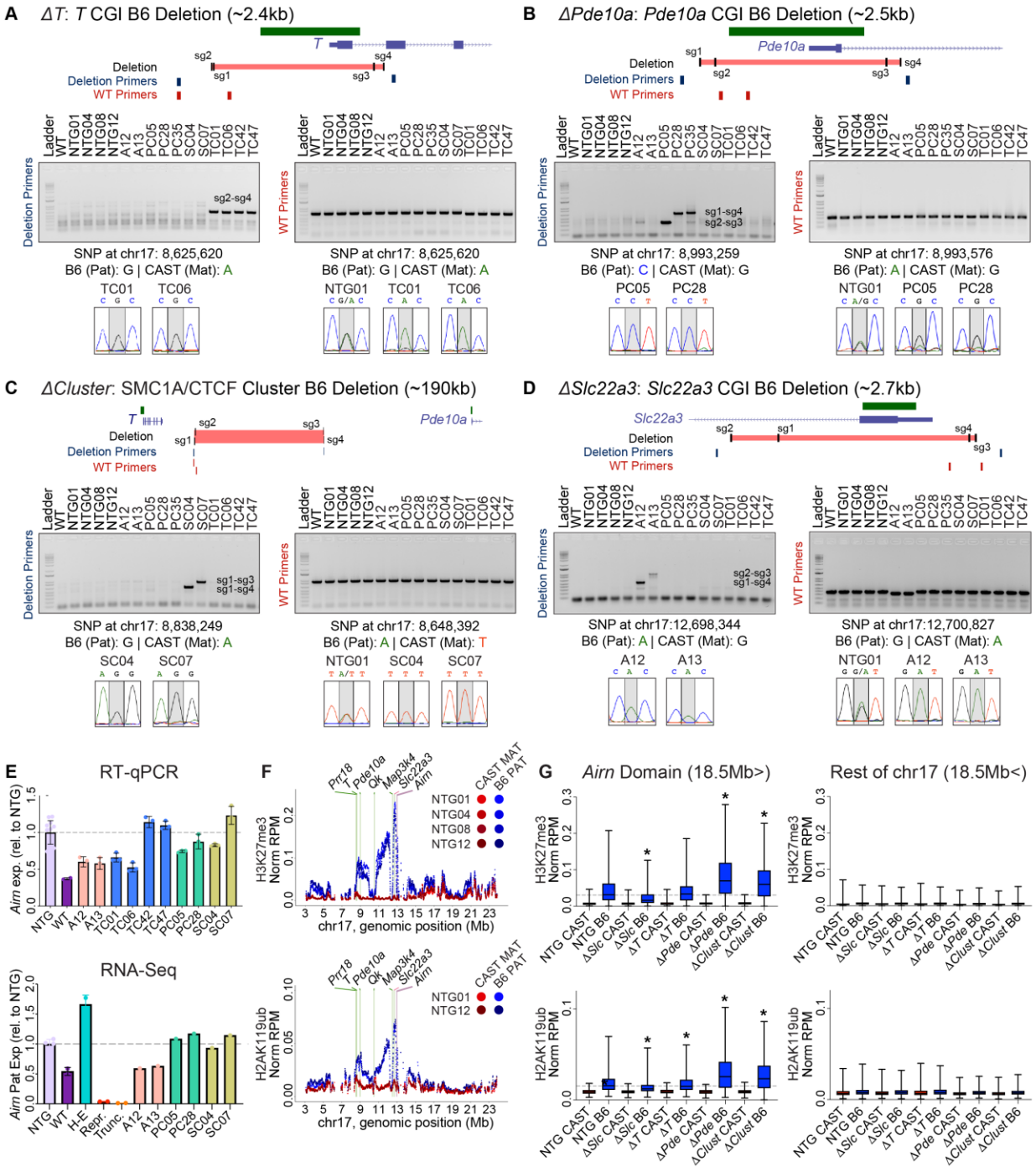


Supplemental Figure 2.4, Related to Figure 2.3.
(A-E) Tiling density plots of allelic Hi-C Observed-over-Expected (O/E) counts for viewpoints of **(A)** *Slc22a3*, **(B)** *Map3k4*, **(C)** *Qk*, **(D)** *Pde10a*, and **(E)** *T* in **(i)** C/B wildtype, **(ii)** C/B *Airn* truncation, **(iii)** B/C wildtype TSCs. Y-axes as in Figure 2. Yellow bar, viewpoint. Purple bar, *Airn* gene. Green bars, other loci of interest. Colored bars, regions analyzed in Figure 3. See STAR Methods for detailed description of analyses. Datasets used are listed in Table S4.



Supplemental Figure 2.5, Related to Figure 2.4.

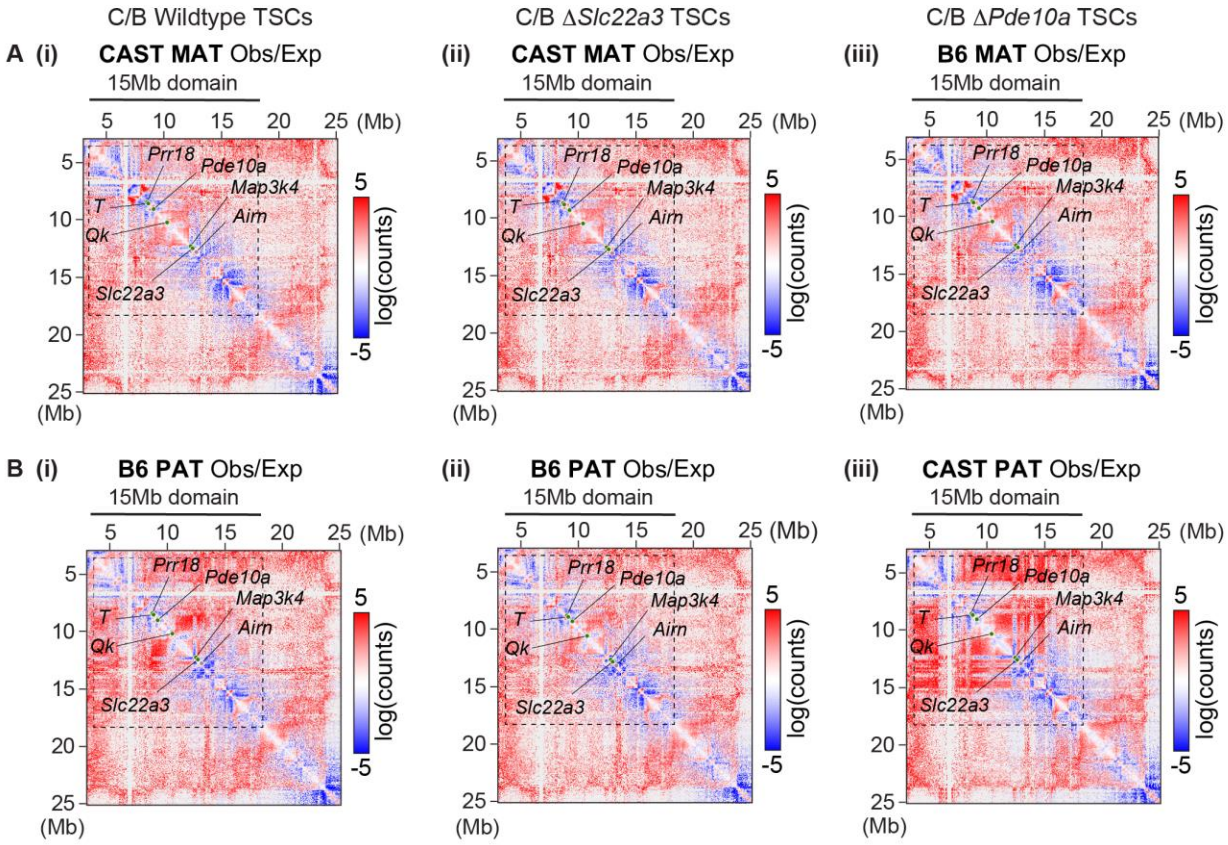
(A) Location of probes used for *Airn* CHART relative to an RNA-Seq density track of *Airn* in C/B wildtype TSCs. **(B)** RT-qPCR of *Airn* expression levels normalized by *Gapdh* and relative to wildtype TSCs. **(C)** Allelic B6 (top) and CAST (bottom) CHART-Seq signal per chromosome in C/B TSC lines used for *Airn* CHART-Seq, n = 2. RPM, Reads per Million total reads. The *Airn* gene is located on chromosome 17. **(D)** Allelic *Airn* CHART-Seq signal over the *Airn* target domain (≤18.5Mb) versus the rest of chromosome 17 (>18.5Mb) in C/B TSC lines used for *Airn* CHART, n = 2. Norm RPM, Reads per Million total reads normalized for SNP density. See STAR Methods for detailed description of analyses. Datasets used are listed in Table S4. Oligos used for *Airn* CHART probes are listed in Table S5.



Supplemental Figure 2.6, Related to Figure 2.5.

(A-D) Genotyping PCR gels and sequencing tracks for **(A)** Δ*T*, **(B)** Δ*Pde10a*, **(C)** Δ*Cluster*, and **(D)** Δ*Slc22a3* clonal TSC lines. Δ*Slc22a3* TSCs were generated in [14]. Gels of PCR product from either wildtype primers, which amplify the flanking end or internal region of the deletion site, or deletion primers, which amplify a sizeable region if the deletion occurred (Table S5). Sanger sequencing tracks of the PCR products show regions with B6/CAST SNPs for allelic identification. **(E)** RT-qPCR and RNA-Seq of *Airn* expression in clones relative to the averaged NTG value, n = 2 or 4. **(F)** Tiling density plots of average allelic (top) H3K27me3 and (bottom)

H2AK119ub ChIP-Seq signal in C/B NTG TSCs, n = 2 or 4. In both panels: purple bar, *Airn* gene; green bars, other loci of interest. Y-axes as in Figure 2. **(G)** Boxplots of average allelic (top) H3K27me3 and (bottom) H2AK119 ChIP-Seq signal over the (left) 15Mb *Airn* target domain versus the (right) rest of chr17 in C/B NTG and deletion TSCs, n = 2. Asterisks, significant changes relative to NTG ($p<0.05$, Welch t-test). Error bars, datapoints outside the interquartile range. See STAR Methods for detailed description of analyses. Datasets used are listed in Table S4. Oligos for CRISPR-Cas9 sgRNAs, qPCR, and PCR genotyping are listed in Table S5.



Supplemental Figure 2.7, Related to Figure 2.6.
Hi-C contact heatmaps of **(A)** maternal and **(B)** paternal Observed-over-Expected (O/E) counts in C/B **(i)** NTG, **(ii)** Δ*Slc22a3*, and **(iii)** Δ*Pde10a* TSCs. Dotted lines, 15Mb *Airn* target domain. Purple circle, *Airn* gene. Green circles, other loci of interest. See STAR Methods for detailed description of analyses. Datasets used are listed in Table S4.

REFERENCES

1. Monk, D., Mackay, D.J.G., Eggermann, T., Maher, E.R., and Riccio, A. (2019). Genomic imprinting disorders: lessons on how genome, epigenome and environment interact. Nat Rev Genet *20*, 235-248. 10.1038/s41576-018-0092-0.

2. Tucci, V., Isles, A.R., Kelsey, G., Ferguson-Smith, A.C., and Erice Imprinting, G. (2019). Genomic Imprinting and Physiological Processes in Mammals. Cell *176*, 952-965. 10.1016/j.cell.2019.01.043.

3. MacDonald, W.A., and Mann, M.R.W. (2020). Long noncoding RNA functionality in imprinted domain regulation. PLoS Genet *16*, e1008930. 10.1371/journal.pgen.1008930.

4. Lleres, D., Imaizumi, Y., and Feil, R. (2021). Exploring chromatin structural roles of non-coding RNAs at imprinted domains. Biochemical Society transactions *49*, 1867-1879. 10.1042/BST20210758.

5. Huang, R., Jaritz, M., Guenzl, P., Vlatkovic, I., Sommer, A., Tamir, I.M., Marks, H., Klampfl, T., Kralovics, R., Stunnenberg, H.G., et al. (2011). An RNA-Seq strategy to detect the complete coding and non-coding transcriptome including full-length imprinted macro ncRNAs. PLoS One *6*, e27288. 10.1371/journal.pone.0027288.

6. Lyle, R., Watanabe, D., te Vruchte, D., Lerchner, W., Smrzka, O.W., Wutz, A., Schageman, J., Hahner, L., Davies, C., and Barlow, D.P. (2000). The imprinted antisense RNA at the Igf2r locus overlaps but does not imprint Mas1. Nat Genet *25*, 19-21. 10.1038/75546.

7. Terranova, R., Yokobayashi, S., Stadler, M.B., Otte, A.P., van Lohuizen, M., Orkin, S.H., and Peters, A.H. (2008). Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. Dev Cell *15*, 668-679. 10.1016/j.devcel.2008.08.015.

8. Sleutels, F., Zwart, R., and Barlow, D.P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature *415*, 810-813. 10.1038/415810a

9. Killian, J.K., Nolan, C.M., Wylie, A.A., Li, T., Vu, T.H., Hoffman, A.R., and Jirtle, R.L. (2001). Divergent evolution in M6P/IGF2R imprinting from the Jurassic to the Quaternary. Hum Mol Genet *10*, 1721-1728. 10.1093/hmg/10.17.1721.

10. Yotova, I.Y., Vlatkovic, I.M., Pauler, F.M., Warczok, K.E., Ambros, P.F., Oshimura, M., Theussl, H.C., Gessler, M., Wagner, E.F., and Barlow, D.P. (2008). Identification of the human homolog of the imprinted mouse Air non-coding RNA. Genomics *92*, 464-473. 10.1016/j.ygeno.2008.08.004.

11. Suzuki, S., Shaw, G., and Renfree, M.B. (2018). Identification of a novel antisense noncoding RNA, ALID, transcribed from the putative imprinting control region of marsupial IGF2R. Epigenetics Chromatin *11*, 55. 10.1186/s13072-018-0227-8.

12. Seidl, C.I., Stricker, S.H., and Barlow, D.P. (2006). The imprinted Air ncRNA is an atypical RNAPII transcript that evades splicing and escapes nuclear export. EMBO J *25*, 3565-3575. 10.1038/sj.emboj.7601245.

13. Andergassen, D., Dotter, C.P., Wenzel, D., Sigl, V., Bammer, P.C., Muckenhuber, M., Mayer, D., Kulinski, T.M., Theussl, H.C., Penninger, J.M., et al. (2017). Mapping the mouse Allelome reveals tissue-specific regulation of allelic expression. eLife *6*. 10.7554/eLife.25125.

14. Schertzer, M.D., Braceros, K.C.A., Starmer, J., Cherney, R.E., Lee, D.M., Salazar, G., Justice, M., Bischoff, S.R., Cowley, D.O., Ariel, P., et al. (2019). lncRNA-Induced Spread of Polycomb Controlled by Genome Architecture, RNA Abundance, and CpG Island DNA. Mol Cell *75*, 523-537 e510. 10.1016/j.molcel.2019.05.028.

15. Latos, P.A., Pauler, F.M., Koerner, M.V., Senergin, H.B., Hudson, Q.J., Stocsits, R.R., Allhoff, W., Stricker, S.H., Klement, R.M., Warczok, K.E., et al. (2012). Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. Science *338*, 1469-1472. 338/6113/1469 [pii]10.1126/science.1228110.

16. Pauler, F.M., Barlow, D.P., and Hudson, Q.J. (2012). Mechanisms of long range silencing by imprinted macro non-coding RNAs. Current Opinion in Genetics & Development *22*, 283-289. 10.1016/j.gde.2012.02.005.

17. Andergassen, D., Muckenhuber, M., Bammer, P.C., Kulinski, T.M., Theussl, H.C., Shimizu, T., Penninger, J.M., Pauler, F.M., and Hudson, Q.J. (2019). The Airn lncRNA does not require any DNA elements within its locus to silence distant imprinted genes. PLoS Genet *15*, e1008268. 10.1371/journal.pgen.1008268.

18. Nagano, T., Mitchell, J.A., Sanz, L.A., Pauler, F.M., Ferguson-Smith, A.C., Feil, R., and Fraser, P. (2008). The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science *322*, 1717-1720. 1163802 [pii]10.1126/science.1163802.

19. Grosswendt, S., Kretzmer, H., Smith, Z.D., Kumar, A.S., Hetzel, S., Wittler, L., Klages, S., Timmermann, B., Mukherji, S., and Meissner, A. (2020). Epigenetic regulator function through mouse gastrulation. Nature *584*, 102-108. 10.1038/s41586-020-2552-x.

20. Andergassen, D., Smith, Z.D., Kretzmer, H., Rinn, J.L., and Meissner, A. (2021). Diverse epigenetic mechanisms maintain parental imprints within the embryonic and extraembryonic lineages. Dev Cell *56*, 2995-3005 e2994. 10.1016/j.devcel.2021.10.010.

21. Trotman, J.B., Braceros, K.C.A., Cherney, R.E., Murvin, M.M., and Calabrese, J.M. (2021). The control of polycomb repressive complexes by long noncoding RNAs. Wiley Interdiscip Rev RNA *12*, e1657. 10.1002/wrna.1657.

22. Blackledge, N.P., and Klose, R.J. (2021). The molecular principles of gene regulation by Polycomb repressive complexes. Nat Rev Mol Cell Biol *22*, 815-833. 10.1038/s41580-021-00398-y.

23. 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. 10.1126/science.aal2512.

24. Bousard, A., Raposo, A.C., Zylicz, J.J., Picard, C., Pires, V.B., Qi, Y., Gil, C., Syx, L., Chang, H.Y., Heard, E., and da Rocha, S.T. (2019). The role of Xist-mediated Polycomb recruitment in the initiation of X-chromosome inactivation. EMBO Rep *20*, e48019. 10.15252/embr.201948019.

25. Keane, T.M., Goodstadt, L., Danecek, P., White, M.A., Wong, K., Yalcin, B., Heger, A., Agam, A., Slater, G., Goodson, M., et al. (2011). Mouse genomic variation and its effect on phenotypes and gene regulation. Nature *477*, 289-294. nature10413 [pii]10.1038/nature10413.

26. Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665-1680. 10.1016/j.cell.2014.11.021.

27. Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., and Aiden, E.L. (2016). Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. Cell Syst *3*, 99-101. 10.1016/j.cels.2015.07.012.

28. Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst *3*, 95-98. 10.1016/j.cels.2016.07.002.

29. Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289-293. 326/5950/289 [pii]10.1126/science.1181369.

30. Blackledge, N.P., Fursova, N.A., Kelley, J.R., Huseyin, M.K., Feldmann, A., and Klose, R.J. (2020). PRC1 Catalytic Activity Is Central to Polycomb System Function. Mol Cell *77*, 857-874 e859. 10.1016/j.molcel.2019.12.001.

31. Boyle, S., Flyamer, I.M., Williamson, I., Sengupta, D., Bickmore, W.A., and Illingworth, R.S. (2020). A central role for canonical PRC1 in shaping the 3D nuclear landscape. Genes Dev *34*, 931-949. 10.1101/gad.336487.120.

32. Oksuz, O., Narendra, V., Lee, C.H., Descostes, N., LeRoy, G., Raviram, R., Blumenberg, L., Karch, K., Rocha, P.P., Garcia, B.A., et al. (2018). Capturing the Onset of PRC2-Mediated Repressive Domain Formation. Mol Cell *70*, 1149-1162 e1145. 10.1016/j.molcel.2018.05.023.

33. Isono, K., Endo, T.A., Ku, M., Yamada, D., Suzuki, R., Sharif, J., Ishikura, T., Toyoda, T., Bernstein, B.E., and Koseki, H. (2013). SAM domain polymerization links subnuclear

clustering of PRC1 to gene silencing. Dev Cell *26*, 565-577. 10.1016/j.devcel.2013.08.016.

34. Eeftens, J.M., Kapoor, M., Michieletto, D., and Brangwynne, C.P. (2021). Polycomb condensates can promote epigenetic marks but are not required for sustained chromatin compaction. Nature communications *12*, 5888. 10.1038/s41467-021-26147-5.

35. Kraft, K., Yost, K.E., Murphy, S.E., Magg, A., Long, Y., Corces, M.R., Granja, J.M., Wittler, L., Mundlos, S., Cech, T.R., et al. (2022). Polycomb-mediated genome architecture enables long-range spreading of H3K27 methylation. Proc Natl Acad Sci U S A *119*, e2201883119. 10.1073/pnas.2201883119.

36. Cai, Y., Zhang, Y., Loh, Y.P., Tng, J.Q., Lim, M.C., Cao, Z., Raju, A., Lieberman Aiden, E., Li, S., Manikandan, L., et al. (2021). H3K27me3-rich genomic regions can function as silencers to repress gene expression via chromatin interactions. Nature communications *12*, 719. 10.1038/s41467-021-20940-y.

37. Cuadrado, A., Gimenez-Llorente, D., Kojic, A., Rodriguez-Corsino, M., Cuartero, Y., Martin-Serrano, G., Gomez-Lopez, G., Marti-Renom, M.A., and Losada, A. (2019). Specific Contributions of Cohesin-SA1 and Cohesin-SA2 to TADs and Polycomb Domains in Embryonic Stem Cells. Cell Rep *27*, 3500-3510 e3504. 10.1016/j.celrep.2019.05.078.

38. Kriz, A.J., Colognori, D., Sunwoo, H., Nabet, B., and Lee, J.T. (2021). Balancing cohesin eviction and retention prevents aberrant chromosomal interactions, Polycomb-mediated repression, and X-inactivation. Mol Cell *81*, 1970-1987 e1979. 10.1016/j.molcel.2021.02.031.

39. Rhodes, J.D.P., Feldmann, A., Hernandez-Rodriguez, B., Diaz, N., Brown, J.M., Fursova, N.A., Blackledge, N.P., Prathapan, P., Dobrinic, P., Huseyin, M.K., et al. (2020). Cohesin Disrupts Polycomb-Dependent Chromosome Interactions in Embryonic Stem Cells. Cell Rep *30*, 820-835 e810. 10.1016/j.celrep.2019.12.057.

40. Calabrese, J.M., Sun, W., Song, L., Mugford, J.W., Williams, L., Yee, D., Starmer, J., Mieczkowski, P., Crawford, G.E., and Magnuson, T. (2012). Site-specific silencing of regulatory elements as a mechanism of X inactivation. Cell *151*, 951-963. S0092-8674(12)01300-1 [pii]10.1016/j.cell.2012.10.037.

41. Simon, M.D., Pinter, S.F., Fang, R., Sarma, K., Rutenberg-Schoenberg, M., Bowman, S.K., Kesner, B.A., Maier, V.K., Kingston, R.E., and Lee, J.T. (2013). High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. Nature *504*, 465-469. 10.1038/nature12719.

42. Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G.L., Lubling, Y., Xu, X., Lv, X., Hugnot, J.P., Tanay, A., and Cavalli, G. (2017). Multiscale 3D Genome Rewiring during Mouse Neural Development. Cell *171*, 557-572 e524. 10.1016/j.cell.2017.09.043.

43. Rowley, M.J., Poulet, A., Nichols, M.H., Bixler, B.J., Sanborn, A.L., Brouhard, E.A., Hermetz, K., Linsenbaum, H., Csankovszki, G., Lieberman Aiden, E., and Corces, V.G. (2020). Analysis of Hi-C data using SIP effectively identifies loops in organisms from C. elegans to mammals. Genome Res *30*, 447-458. 10.1101/gr.257832.119.

44. Mirny, L., and Dekker, J. (2022). Mechanisms of Chromosome Folding and Nuclear Organization: Their Interplay and Open Questions. Cold Spring Harb Perspect Biol *14*. 10.1101/cshperspect.a040147.

45. Hojfeldt, J.W., Laugesen, A., Willumsen, B.M., Damhofer, H., Hedehus, L., Tvardovskiy, A., Mohammad, F., Jensen, O.N., and Helin, K. (2018). Accurate H3K27 methylation can be established de novo by SUZ12-directed PRC2. Nat Struct Mol Biol *25*, 225-232. 10.1038/s41594-018-0036-6.

46. Arnold, P.R., Wells, A.D., and Li, X.C. (2019). Diversity and Emerging Roles of Enhancer RNA in Regulation of Gene Expression and Cell Fate. Front Cell Dev Biol *7*, 377. 10.3389/fcell.2019.00377.

47. Statello, L., Guo, C.J., Chen, L.L., and Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol *22*, 96-118. 10.1038/s41580-020-00315-9.

48. Galupa, R., Picard, C., Servant, N., Nora, E.P., Zhan, Y., van Bemmel, J.G., El Marjou, F., Johanneau, C., Borensztein, M., Ancelin, K., et al. (2022). Inversion of a topological domain leads to restricted changes in its gene expression and affects interdomain communication. Development *149*. 10.1242/dev.200568.

49. Greenwald, W.W., Li, H., Benaglio, P., Jakubosky, D., Matsui, H., Schmitt, A., Selvaraj, S., D'Antonio, M., D'Antonio-Chronowska, A., Smith, E.N., and Frazer, K.A. (2019). Subtle changes in chromatin loop contact propensity are associated with differential gene regulation and expression. Nature communications *10*, 1054. 10.1038/s41467-019-08940-5.

50. Zuin, J., Roth, G., Zhan, Y., Cramard, J., Redolfi, J., Piskadlo, E., Mach, P., Kryzhanovska, M., Tihanyi, G., Kohler, H., et al. (2022). Nonlinear control of transcription through enhancer-promoter interactions. Nature *604*, 571-577. 10.1038/s41586-022-04570-y.

51. Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-Andre, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934-947. 10.1016/j.cell.2013.09.053.

52. Gazal, S., Weissbrod, O., Hormozdiari, F., Dey, K.K., Nasser, J., Jagadeesh, K.A., Weiner, D.J., Shi, H., Fulco, C.P., O'Connor, L.J., et al. (2022). Combining SNP-to-gene linking strategies to identify disease genes and assess disease omnigenicity. Nat Genet *54*, 827-836. 10.1038/s41588-022-01087-y.

53. Markaki, Y., Gan Chong, J., Wang, Y., Jacobson, E.C., Luong, C., Tan, S.Y.X., Jachowicz, J.W., Strehle, M., Maestrini, D., Banerjee, A.K., et al. (2021). Xist nucleates local protein gradients to propagate silencing across the X chromosome. Cell *184*, 6174-6192 e6132. 10.1016/j.cell.2021.10.022.

54. Jachowicz, J.W., Strehle, M., Banerjee, A.K., Blanco, M.R., Thai, J., and Guttman, M. (2022). Xist spatially amplifies SHARP/SPEN recruitment to balance chromosome-wide silencing and specificity to the X chromosome. Nat Struct Mol Biol *29*, 239-249. 10.1038/s41594-022-00739-1.

55. Calabrese, J.M., Starmer, J., Schertzer, M.D., Yee, D., and Magnuson, T. (2015). A survey of imprinted gene expression in mouse trophoblast stem cells. G3 (Bethesda) *5*, 751-759. 10.1534/g3.114.016238.

56. Quinn, J., Kunath, T., and Rossant, J. (2006). Mouse trophoblast stem cells. Methods Mol Med *121*, 125-148.

57. Concordet, J.P., and Haeussler, M. (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. Nucleic Acids Res *46*, W242-W245. 10.1093/nar/gky354.

58. Schertzer, M.D., Thulson, E., Braceros, K.C.A., Lee, D.M., Hinkle, E.R., Murphy, R.M., Kim, S.O., Vitucci, E.C.M., and Calabrese, J.M. (2019). A piggyBac-based toolkit for inducible genome editing in mammalian cells. RNA *25*, 1047-1058. 10.1261/rna.068932.118.

59. Davis, C.P., and West, J.A. (2015). Purification of specific chromatin regions using oligonucleotides: capture hybridization analysis of RNA targets (CHART). Methods Mol Biol *1262*, 167-182. 10.1007/978-1-4939-2253-6_10.

60. Flickinger, J.L., Gebeyehu, G., Buchman, G., Haces, A., and Rashtchian, A. (1992). Differential incorporation of biotinylated nucleotides by terminal deoxynucleotidyl transferase. Nucleic Acids Res *20*, 2382. 10.1093/nar/20.9.2382.

61. Lee, B.T., Barber, G.P., Benet-Pages, A., Casper, J., Clawson, H., Diekhans, M., Fischer, C., Gonzalez, J.N., Hinrichs, A.S., Lee, C.M., et al. (2022). The UCSC Genome Browser database: 2022 update. Nucleic Acids Res *50*, D1115-D1122. 10.1093/nar/gkab959.

62. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359. 10.1038/nmeth.1923.

63. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21. 10.1093/bioinformatics/bts635.

64. 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. 10.1093/bioinformatics/btp352.

65. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag).

66. Ramirez, F., Bhardwaj, V., Arrigoni, L., Lam, K.C., Gruning, B.A., Villaveces, J., Habermann, B., Akhtar, A., and Manke, T. (2018). High-resolution TADs reveal DNA sequences underlying genome organization in flies. Nature communications *9*, 189. 10.1038/s41467-017-02525-w.

67. Robinson, J.T., Turner, D., Durand, N.C., Thorvaldsdottir, H., Mesirov, J.P., and Aiden, E.L. (2018). Juicebox.js Provides a Cloud-Based Visualization System for Hi-C Data. Cell Syst *6*, 256-258 e251. 10.1016/j.cels.2018.01.001.

68. Wolff, J., Bhardwaj, V., Nothjunge, S., Richard, G., Renschler, G., Gilsbach, R., Manke, T., Backofen, R., Ramirez, F., and Gruning, B.A. (2018). Galaxy HiCExplorer: a web server for reproducible Hi-C data analysis, quality control and visualization. Nucleic Acids Res *46*, W11-W16. 10.1093/nar/gky504.

69. Orlando, D.A., Chen, M.W., Brown, V.E., Solanki, S., Choi, Y.J., Olson, E.R., Fritz, C.C., Bradner, J.E., and Guenther, M.G. (2014). Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. Cell Rep *9*, 1163-1170. 10.1016/j.celrep.2014.10.018.

70. Justice, M., Carico, Z.M., Stefan, H.C., and Dowen, J.M. (2020). A WIZ/Cohesin/CTCF Complex Anchors DNA Loops to Define Gene Expression and Cell Identity. Cell Rep *31*, 107503. 10.1016/j.celrep.2020.03.067.

71. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137. gb-2008-9-9-r137 [pii]10.1186/gb-2008-9-9-r137.

72. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842. 10.1093/bioinformatics/btq033.

73. Frankish, A., Diekhans, M., Jungreis, I., Lagarde, J., Loveland, J.E., Mudge, J.M., Sisu, C., Wright, J.C., Armstrong, J., Barnes, I., et al. (2021). Gencode 2021. Nucleic Acids Res *49*, D916-D923. 10.1093/nar/gkaa1087.

74. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923-930. 10.1093/bioinformatics/btt656.

CHAPTER 3: Ectopically expressed *Airn* lncRNA deposits Polycomb with a potency that rivals *Xist*[2]

3.1 Introduction

Long noncoding RNAs (lncRNAs) play critical roles in development by repressing transcription *in cis* (on the same chromosome from which they were transcribed [1]). In the most extreme example, expression of the lncRNA *Xist* represses transcription across one entire X chromosome during the essential process of X inactivation. Critical for stable repression by *Xist* are enzymes called the Polycomb Repressive Complexes, PRC1 and PRC2, which monoubiquitylate lysine 119 of Histone H2A (H2AK119ub) and trimethylate lysine 27 of Histone H3 (H3K27me3), respectively [2]. The mechanisms that *Xist* uses to silence gene expression and recruit PRCs to chromatin remain under investigation. However, there is little question that the *Xist* lncRNA product is functional; its ectopic expression from essentially any chromosomal location, not only the X, results in chromosome-wide transcriptional repression and PRC recruitment [1].

Beyond *Xist*, a handful of other *cis*-acting repressive lncRNAs have been identified. The most potent of these is a lncRNA called *Airn*, whose expression has been shown to repress genes and recruit PRCs within a genomic interval that spans roughly 15 megabases (Mb) in the extraembryonic tissues of the mouse [3,4]. *Airn* is an unusual RNA [5]. Unlike *Xist*, which is robustly spliced, polyadenylated, and stable (half-life of ~6 hours), *Airn* evades splicing, is incompletely

[2] This chapter previously appeared as a preprint article on bioRxiv. The original citation is as follows: Trotman, Braceros et al. "Ectopically expressed *Airn* lncRNA deposits Polycomb with a potency that rivals *Xist*," *bioRxiv* (May 2023).

polyadenylated, and is unstable (half-life of ~1 hour; [4,6]). Also, while *Xist* contains well-conserved sequence elements and is processed into distinct transcripts, *Airn* is more poorly conserved an produces transcripts of variable length that terminate gradually as distance from *Airn*'s transcription start site increases, upwards of 100kb away [7]. These unusual properties raise the question of whether it is the *Airn* lncRNA product or act of *Airn* transcription that is responsible for its regulatory effect. This same question has been posed for other putative regulatory lncRNAs. In many cases, the act of transcription, and not the lncRNA product, has been found to be the dominant mediator of regulation [8-11].

3.2 Results and Discussion

To determine whether the *Airn* lncRNA product is functional, we sought to induce its expression from an ectopic locus; this would allow us to investigate effects arising only from expression of the lncRNA itself, decoupled from any effects of endogenous *cis*-regulatory features or chromosomal structures that could have evolved in concert with the *Airn* gene. To this end, we used recombineering to insert the first ~89kb of the *Airn* gene downstream of a doxycycline (dox)-inducible promoter in a DNA vector that enabled its Cre-mediated insertion into the C57BL/6J (B6) allele of the *Rosa26* locus on chromosome 6 (chr6), in F1-hybrid mouse embryonic stem cells (ESCs) derived from a cross between B6 and CAST/EiJ (CAST) mice [12] (Figure 1A). We obtained two separate ESC clones containing the *Airn* gene inserted into B6 *Rosa26* (Figure S1A). RNA-Seq after treatment with 1000ng/mL dox demonstrated that in both clones, *Airn* expression was induced to levels that slightly surpassed its endogenous levels in trophoblast stem cells (TSCs; Figures 1B, S1B). Like endogenous *Airn* in TSCs [4], ectopically expressed *Airn* localized as a single focus at its presumptive site of transcription, underwent

transcriptional attrition across its gene body, rarely underwent splicing, and was unstable, with a half-life of ~45 minutes (Figures 1C-E, 1G). We conclude that ectopically expressed *Airn* retains properties that are similar to the endogenous transcript.

We next compared the relative expression levels of ectopically expressed *Airn* to those of *Xist*. We had previously used the same Cre-mediated approach to insert the 22kb *Xist* gene into *Rosa26* under control of the same dox-inducible promoter used to express *Airn* (Figure S1A) [12]. In these ESCs, after treatment with 1000ng/mL dox, we found that *Xist* was expressed at levels ~10-fold higher than *Airn*, consistent with its ~10-fold greater half-life (Figures 1B, 1E, 1F). After treatment with only 10ng/mL dox, *Xist* was expressed at levels on par with *Airn* expressed with 1000ng/mL dox (Figure 1B). RNA FISH demonstrated that at the 1000ng/mL condition, 65% and 86% of cells expressed detectable levels of *Airn* and *Xist*, respectively (Figure 1D). At the 10ng/mL condition, 40% of cells expressed detectable levels of *Xist* (Figure 1D). Expression levels of the reverse tetracycline-controlled transactivator (rtTA) were equivalent across genotypes, supporting the notion that each *Airn* and *Xist* ESC line was capable of inducing lncRNA expression to equivalent extents (Figure S1C). We conclude that when inserted into the same genomic locus, under control of the same promoter, and subject to the same level of promoter induction, ectopically expressed *Airn* is less stable and accumulates to levels lower than *Xist*.

We next performed ChIP-Seq to determine the extent to which ectopic expression of *Airn* altered surrounding levels of PRC2-deposited H3K27me3. In addition to examining H3K27me3 levels in *Airn*-expressing cells, we examined H3K27me3 levels in "Empty" control ESCs harboring a dox-inducible promoter at *Rosa26* but no inserted lncRNA, as well as in *Xist*-expressing cells treated without dox or with 10ng/mL or 1000ng/mL dox (Figure 2A). Relative to

Empty control ESCs, in *Airn*-expressing ESCs, we observed a striking increase of H3K27me3 over the majority of chr6, specifically on the B6 and not CAST allele (Figures 2A, 2B, S2A). Moreover, the increases in H3K27me3 induced by *Airn* were significantly larger than those induced by *Xist* when the two lncRNAs were expressed at near-equivalent levels (1000ng/mL dox for *Airn*, 10ng/mL dox for *Xist*; Figures 1B, 2A, 2B). The increases in H3K27me3 when *Xist* was expressed at 1000ng/mL dox surpassed those of *Airn* at the 1000ng/mL condition (Figures 2A, 2B). We conclude that the *Airn* lncRNA product is functional, its expression being sufficient to induce the deposition of H3K27me3 over the majority of chr6 when expressed from *Rosa26*. Moreover, in this context and when normalized for copy number, *Airn* deposits H3K27me3 with a potency that rivals or even surpasses that of *Xist*.

We next examined our RNA-Seq data to determine the extent to which *Airn* and *Xist* repressed gene expression. Despite the large-scale changes in H3K27me3, *Airn* did not induce repression of any chr6 genes (Figure 2C, Table S1). Likewise, prolonged, 7-day *Airn* expression did not induce repression (Table S1). In contrast, at 10ng/mL dox, *Xist* repressed expression of 21 chr6 genes, and at 1000ng/mL dox, *Xist* repressed expression of 242 chr6 genes (Figure 2C, Table S1). Thus, *Airn*'s ability to repress transcription is markedly weaker than that of *Xist*. Also, *Xist*'s ability to repress transcription correlates positively with its abundance.

Our study unequivocally demonstrates that the *Airn* lncRNA product is functional, being capable of inducing the accumulation of Polycomb over a genomic range and intensity that rivals that of *Xist*. This is despite the fact that *Airn* and *Xist* harbor diverse evolutionary origins and sequence contents [13]. We also demonstrate that under the conditions tested above, *Airn* does not induce transcriptional repression despite its ability to induce deposition of H3K27me3. Transcriptional repression and H3K27me3 deposition have been shown to be separable functions

within *Xist* [14-17]. Our observation that the same is true for regulation by *Airn* suggests that these principles are generalizable to other lncRNAs, and supports the possibility that in its endogenous locus, *Airn*, along with the Polycomb-directed marks that it helps to deposit, functions largely to maintain a memory of previous transcriptionally repressive events [18,19]. We speculate that the establishment of transcriptional memory may be a common function of *cis*-acting lncRNAs, perhaps more common than the type of robust, *de novo* gene silencing induced by *Xist*.

Lastly, our data underscore the importance of RNA abundance in moderating the extent of lncRNA-mediated repression [4]. In the context of *Rosa26,* when subject to identical extents of promoter activation, *Xist* accumulated to levels that were 10-fold higher than those of *Airn*. These elevated levels were necessary for *Xist*'s full repressive effect. We speculate that the evolution of full-scale X inactivation [20] may have been driven, in part, by the acquisition of features within *Xist* that led to punctuated increases in its nuclear abundance. Conversely, without such features, lncRNAs such as *Airn* effectively limit their influence on chromatin.

3.3 Methods

3.3.1 pCARGO-RMCE-Airn Plasmid construction

BAC clone RP23-309H20 harboring the Airn lncRNA gene was obtained from the BACPAC Resources Center and was modified by recombineering to generate the tetracycline-inducible pCARGO-RMCE-*Airn* donor vector. In brief: a first recombineering step was performed to remove the lox511 site in the pBACe3.6 vector backbone along with the ~56 kb of mouse genomic sequence located upstream of *Airn*/ NR_027772 exon 1 (mm10 chr17:12741311). The removed sequences were replaced by an ampicillin resistance cassette, a lox71 site, and a TreTight/CMV minimal promoter, all from the pCARGO-RMCE cargo vector described in [12], immediately upstream of *Airn* exon 1. A second recombineering step was performed to remove ~119 kb of mouse sequence downstream of the transcription end site for *Airn*/ NR_027772 (mm10 chr17:12830123) as well as the loxP site in the BAC vector backbone. The removed sequences were replaced by an SV40 polyadenylation sequence, an FRT-flanked PGK-EM7-Hygromycin resistance cassette, and a lox2272 site, again from the pCARGO-RMCE cargo vector described in [12]. The final BAC clone was fully sequence-validated. Large-scale preparation of pCARGO-RMCE-*Airn* plasmid was purified from cells grown overnight at 30°C with ampicillin selection (100µg/mL ampicillin sodium salt, Fisher Scientific BP1760), using the NucleoBond BAC100 Maxiprep Kit (Macherey-Nagel 740579). Pelleted plasmids were resuspended in Invitrogen TE Buffer before measuring DNA concentration with a Qubit 2.0 Fluorometer (dsDNA high sensitivity kit, Thermo Fisher Q32851). Integrity and sequence composition of plasmid preparations were confirmed by diagnostic digestion and by whole-plasmid Oxford Nanopore sequencing (Plasmidsaurus, File S1). Full details of the BAC recombineering approach are provided in File S2.

3.3.2 Cell culture

B6/CAST F1-hybrid *Rosa26*-RMCE ESCs from [12] were grown on gelatin-coated plastic dishes in a humidified Thermo Fisher Forma Series II water-jacketed incubator at 37°C and under 5% CO_2. Cells were grown in DMEM high glucose plus sodium pyruvate (Gibco 11995-065), 15% ESC-qualified fetal bovine serum (Gibco 26140-079), 0.1mM non-essential amino acids (Gibco 11140-050), 100U/mL penicillin-streptomycin (Gibco 15140-122), 2mM L-glutamine (Gibco 25030-081), 0.1mM β-mercaptoethanol (Sigma-Aldrich 63689), and 1:500 LIF conditioned media produced from Lif-1Cα (COS) cells.

TSCs were cultured as in (Quinn et al. 2006). Briefly, TSCs were cultured on gelatin-coated, pre-plated irradiated mouse embryonic fibroblast (irMEF) feeder cells in TSC media (RPMI [Gibco 11875093], 20% qualified FBS [Gibco 26140079], 0.1mM penicillin-streptomycin [Gibco 15140122], 1mM sodium pyruvate [Gibco 11360070], 2mM L-glutamine [Gibco 25030081], 100μM β-mercaptoethanol [Sigma-Aldrich 63689]) supplemented with 25ng/mL FGF4 (Gibco PHG0154) and 1μg/mL Heparin (Sigma-Aldrich H3149) just before use, at 37°C in a humidified incubator at 5% CO_2. At passage, TSCs were trypsinized with 0.125% Trypsin-EDTA in PBS solution (Gibco 25200-072) for ~4 minutes at room temperature and gently dislodged from the plate with a sterile, cotton-plugged Pasteur pipette. To deplete irMEFs from TSCs prior to all harvests, TSCs were pre-plated for 45 minutes at 37°C, transferred to a fresh culture plate, and then cultured for three days in 70% irMEF-conditioned TSC media supplemented with growth factors as above.

3.3.3 Generation of RMCE cell lines

Prior to electroporation, plasmid mixtures were prepared as follows: for RMCE-empty cells: 6.25µg pCARGO-RMCE-empty [12], 4.1µg pOG-Cre, 6µg PB-TRE-Cas9 (for transient hygromycin B resistance [22]; for RMCE-*Xist* cells: 19µg pCARGO-RMCE-*Xist* [12], 2.2µg pOG-Cre, 6µg PB-TRE-Cas9; for RMCE-*Airn* cells, two conditions that each ultimately yielded one colony: (19µg pCARGO-RMCE-*Airn*, 2.2µg pOG-Cre, 6µg pCL116 [22]) or (54 µg pCARGO-RMCE-*Airn*, 10.8µg pOG-Cre, 6µg pCL116). Each plasmid mixture was ethanol-precipitated, air-dried, and thoroughly resuspended in 10µL of TE Buffer (Invitrogen) after soaking overnight at 4°C. The 10-µL plasmid mixtures were mixed with 1 million B6/CAST F1-hybrid *Rosa26*-RMCE ESCs (in 100 µL Neon Buffer R) and electroporated using a Neon Transfection System (Invitrogen) with one 40-ms pulse of 1000 V before seeding sparsely onto a confluent 10-cm dish of gamma-irradiated drug-resistant (DR4) MEF feeder cells (ATCC SCRC-1045) in growth medium lacking penicillin-streptomycin. Growth medium was not changed the day after electroporation but was replaced daily thereafter. Two days after electroporation, growth medium was replaced with normal growth medium (containing penicillin-streptomycin). Starting three days after electroporation and lasting for one day (RMCE-*Airn*) or at least ten days (RMCE-Empty and RMCE-*Xist*), cells were selected in growth medium containing 150µg/mL hygromycin B (Roche 10843555001). Ganciclovir (InvivoGen sud-gcv) was added at a working concentration of 3µM starting five days (RMCE-*Airn* cells electroporated with 54µg pCARGO-RMCE-*Airn* plasmid), six days (RMCE-*Airn* cells electroporated with 19µg pCARGO-RMCE-*Airn* plasmid), or seven days (RMCE-Empty and RMCE-*Xist*) after electroporation and lasting for least six days. When clearly visible by eye (9-12 days after electroporation), individual colonies were picked with aid of a light microscope fitted with a 4X objective, dissociated with

0.125% trypsin, and each transferred to a 24-well plate well pre-seeded confluently with gamma-irradiated DR4 MEF feeder cells. Clonal lines of cells were grown until sufficiently confluent (>50%), at which point they were passaged for subsequent genotyping, preparing cryogenic freeze-down stocks, and introducing the reverse tetracycline-controlled transactivator (rtTA) via plasmid transfection.

For introduction of rtTA, four independent RMCE-Empty cell lines, four independent RMCE-*Xist* wild-type cell lines, and two independent RMCE-*Airn* cell lines were each seeded at a density of 750,000 cells per well of a 6-well plate. The following day, each cell line was given fresh growth medium and transfected with a plasmid mixture containing 1µg PB-rtTA (Addgene #126034; [22]) and 1µg pUC19-piggyBac transposase [23] using Lipofectamine 3000 (Thermo Fisher L3000001). Starting around 24h after transfection and lasting at least 10 days, cells were selected in growth medium containing 200µg/mL G418 sulfate (Thermo Fisher 10131035).

3.3.4 Genomic DNA preparation and genotyping PCR

To prepare genomic DNA, cells in a confluent well of 24-well plate were lysed by adding 488µL of a mixture containing 400µL lysis buffer (100 mM Tris-HCl pH 8.0, 5mM EDTA, 200mM NaCl, 0.2% (w/v) SDS), 80µL 20mg/mL proteinase K (Meridian Bioscience BIO-37084), and 8µL 5mg/mL linear acrylamide (Thermo Fisher AM9520) and incubating at 55 °C overnight. Samples were then heated at 100 °C for 1h, incubated at 4 °C for at least 30min, and centrifuged at 16,100 x g for 5min at 4°C. Supernatants were removed, and pellets were washed with 1mL ice-cold 80% ethanol (v/v). After centrifuging again as before, supernatants were completely removed, and pellets were air-dried before adding 100µL TE Buffer (Thermo Fisher

8019005) and incubating at 4°C overnight. Genomic DNA samples were mixed thoroughly by vortexing and pipetting before using for genotyping PCR.

To perform genotyping PCR, 25-µL reactions were prepared on ice in PCR strip-tubes, each containing the following: 1µL 10mM dNTPs, 0.25µL forward primer, 0.25µL reverse primer (see Table S3 for oligonucleotide sequences), 2.5µL 10X Ammonium Buffer, 0.75µL 50 mM MgCl2, 18.25 nuclease-free water, 0.5µL 1U/µL Apex Taq Polymerase (Genesee Scientific 42-401), and 1.5µL genomic DNA. Reactions were run on a thermocycler with the following program: 95 °C for 3min, 30 cycles of (95°C for 30s, 60°C for 30s, 72°C for 1min), 72°C for 5min. To each PCR product, 5µL 6X gel loading dye (NEB B7024S) was added, and a 10-µL sample of each was run on a 1.5% agarose-TAE-ethidium gel alongside 5µL Invitrogen 1 Kb Plus DNA Ladder (Thermo Fisher 10787018) before imaging with a Bio-Rad ChemiDoc MP Imaging System (Ethidium Bromide mode, 0.1s exposure).

3.3.5 RNA preparation

Prior to harvesting, RMCE cell lines (with rtTA) were treated with 0-1000ng/mL doxycycline (1000ng/mL if not specified otherwise) for 72 hours (if not specified otherwise). At harvest, RMCE cells or MEF feeder-free TSCs (see above) were washed twice with ice-cold PBS and harvested with 1 mL TRIzol (Thermo Fisher 15596026). Total RNA was prepared via standard TRIzol-chlorofom extraction, per manufacturer's instructions, except for the addition of 4µL 5µg/µL linear acrylamide (Thermo Fisher AM9520) to promote precipitation in isopropanol. RNA pellets were fully resuspended in RNase-free water, RNA concentrations were measured via Qubit 2.0 fluorometer (RNA high sensitivity kit, Thermo Fisher Q32852), and integrity of

RNA was confirmed by visualizing rRNA bands on an agarose gel. Working stocks of 100ng/μL total RNA were prepared, and all RNA samples were stored at -80°C.

3.3.6 Reverse transcription and quantitative PCR (RT-qPCR)

Using the Applied Biosystems High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher 4368814), 20-μL randomer-primed reverse transcription reactions were prepared in 0.2-mL PCR strip-tubes on ice, each containing 10μL 100ng/μL total RNA (1000 ng) and 0.5μL 40U/μL RNaseOUT RNase Inhibitor (Thermo Fisher 10777019). Reverse transcription was performed with the following thermal cycler program: 25°C for 10min, 37°C for 120min, 85°C for 5min, hold at 4°C. For robust quantification via qPCR, a standard curve was constructed by pooling a portion of representative reverse transcription product samples and serially diluting this mixture 4-fold with nuclease-free water to generate 6 standards. For quantification of each sample, reverse transcription products were each diluted 10-fold with nuclease-free water. For each qPCR primer pair, a qPCR master mix was generated, containing for each 10-μL qPCR reaction: 5μL iTaq Universal SYBR Green Supermix (Bio-Rad 1725124), 3μL nuclease-free water, 0.5μL 10μM forward primer, and 0.5μL 10μM reverse primer (see Table S3 for oligonucleotide sequences) [4,12]. In Hard-Shell 96-Well PCR Plates (Bio-Rad HSP9601), qPCR reactions were set up in technical triplicate by combining 9μL qPCR master mix and 1μL of standard-curve or 10-fold-diluted reverse transcription product. A standard curve was run for each primer pair on every plate. Plates were firmly sealed with Microseal 'B' PCR Plate Sealing Film (Bio-Rad MSB1001), mixed several times by inversion and low-intensity vortexing, and briefly pulsed down in a centrifuge (up to 2000 x g) to collect the volumes and remove bubbles. Reactions were analyzed on a Bio-Rad C1000 Touch Thermal Cycler equipped with a CFX96

Real-Time System with the following program: 95°C for 10min, 40 cycles of (95°C for 15s, 60°C for 30s, 72°C for 30s, plateread), followed by melt curve analysis: 95°C for 10s, 65°C for 31s, 61 cycles of (65°C for 5s +0.5°C/cycle Ramp 0.5°C/s, plateread). Data were analyzed using Bio-Rad CFX manager, first setting "Cq Determination Mode" to "Regression" and then using each primer pair's standard curve (Cq vs. starting quantity) to convert Cq values to starting quantities for each technical replicate. Average *Airn* starting quantities were normalized by dividing by the average *Gapdh* starting quantity for each biological sample, and *Gapdh*-normalized *Airn* expression data were plotted using GraphPad Prism.

3.3.7 Single-molecule RNA FISH

Stellaris RNA FISH probes were designed against the first 40kb of *Airn* utilizing the Stellaris RNA FISH Probe Designer (LGC, Biosearch Technologies, Petaluma, CA) available online at www.biosearchtech.com/stellarisdesigner (version 4.2; parameters: masking level 5; max number of 48 probes; oligo length 20nt; min. spacing length 2nt; Quasar 570). Probes to *Xist* were designed by Stellaris (product # SMF-3011-1, Quasar 570). ESCs were hybridized with the Stellaris RNA FISH probes following the manufacturer's instructions for adherent cells, available online at www.biosearchtech.com/stellarisprotocols. Briefly, cells were grown on coverslips, fixed with 3.7% formaldehyde and permeabilized overnight in 70% ethanol. Cells were then incubated in Wash Buffer A with formamide for five minutes at room temperature, before hybridization with 125nM of *Airn* or *Xist* probes in hybridization buffer at 37°C for 4 hours, then washed twice for 30 minutes at 37°C with Wash Buffer A. Cells were then washed for 5 minutes at room temperature once with Wash Buffer B, then fixed with ProLong Gold containing DAPI (Thermo Fisher P36941) overnight at room temperature in the dark.

Two biological replicates per doxycycline condition were imaged. Images were taken with a Leica DMi8 inverted confocal microscope, using Leica Application Suite X software version 3.7.5.24914. Images used for counting *Xist* or *Airn* expression in nuclei were taken with 40x or 63x objectives. Representative images were taken with 100x objective. Z-slice sizes were 0.2 μm. Images were deconvolved with Huygens Essential version 20.04.0p3 64b (Scientific Volume Imaging, The Netherlands, http://svi.nl) using the standard deconvolution profile under batch express. Images were analyzed using FIJI [24]. To quantify nuclei with or without *Xist* or *Airn* expression, the 570 channel was turned off, DAPI channel was adjusted so that distinct nuclei were visible and then numbered. Then DAPI was turned off, and the 570 channel was adjusted so that signal inside nuclei, but not background signal was visible. Numbered nuclei were then counted as having *Xist*/*Airn* signal or not.

3.3.8 RNA half-life measurement

RNA half-lives were measured using Actinomycin D. *Airn*-RMCE and *Xist*-RMCE cells were induced for 2 days with 1000ng/mL doxycycline prior to start of the Actinomycin D time course. ESCs were then treated with 5μg/mL Actinomycin D (Sigma-Aldrich A9415-5MG) for 15 minutes, 30 minutes, 1 hour, 2 hours, 4 hours, 8 hours, and 12 hours (or without Actinomycin D, "0hr") and lysed with TRIzol. RNA was extracted and subjected to RT-qPCR (see above) to measure levels of *Airn*, *Xist*, and *Gapdh* at each time point (see Table S3 for oligonucleotide sequences). Time course experiments were done in biological replicate for both cell types.

RNA levels were normalized to *Gapdh* at each time point and calculated as the percentage of RNA relative to the 0hr time point. To estimate half-life, *Gapdh*-normalized data for each replicate were averaged and fit to a non-linear one-phase decay model in GraphPad

Prism using the equation Y=(Y0)*exp(-K*X). Error bars represent the standard deviation between biological replicates.

3.3.9 RNA-Seq

Libraries for RNA sequencing were prepared using the KAPA RNA HyperPrep Kit with RiboErase (HMR) (Roche 08098140702) and KAPA Unique Dual-Indexed Adapters (Roche 08861919702) or TruSeq adapters, with each library's input consisting of 9µL 100ng/µL total RNA (900 ng) mixed with 1µL of a 1:250 dilution of ERCC RNA spike-in controls (Thermo Fisher 4456740). RNA fragmentation was performed at 94°C for 6min, and libraries were amplified using 11 cycles of PCR. Libraries were quantified using a Qubit 2.0 fluorometer (dsDNA broad sensitivity kit, Thermo Fisher Q32850), pooled, and sequenced on an Illumina NextSeq500 platform with a 75-cycle NextSeq 500/550 High Output Kit v2.5 (Illumina 20024906).

Sequencing reads (fastq files) were mapped both to the GENCODE vM25 mouse (mm10) genome ("basic CHR" from https://www.gencodegenes.org/mouse/release_M25.html) [for B6] and to a version of the same genome modified to incorporate CAST single-nucleotide polymorphisms (SNPs) [for CAST]. Mapping was performed with STAR (v2.7.9a [25]), using two-pass mapping and the option "--outFilterMultimapNmax 1" to consider only uniquely mapping reads. Using a custom Perl script (intersect_reads_snps16.pl) and a file containing a list of CAST SNPs (downloaded from the Sanger Mouse Genomes Project on 7/30/2020 [26]), aligned reads in the B6 and CAST SAM files were parsed to identify reads clearly originating from either the B6 or CAST allele (i.e. reads that overlap at least 1 B6/CAST SNP). Reads marked as either B6 or CAST were then assigned to genes using a custom Perl script (ase_analyzer10.pl) and the GTF

file gencode.vM25.basic.annotation.gtf [27]. The number of B6-specific reads for each gene from each sample were then compiled. Prior to running DESeq2, a pre-filtering step was used in which genes with fewer than an average of five B6 reads per sample were excluded. Using R, DESeq2 [28] was used with default settings, except for changing the significance threshold to an adjusted p value of 0.01, to determine differential gene expression and to generate MA plots. For RPKM expression analysis, total, non-allelic reads were counted over gene annotations from gencode.vM25.annotation.gtf using featureCounts with default parameters [29]. Read counts for each gene were then divided by total number of reads in the dataset, divided by a million, and then divided by the gene length.

3.3.10 Quantification of spliced reads

Aligned reads with quality score over thirty were merged from all replicates. Samtools [30] was used to extract reads over *Airn* and *Xist*, and count the reads over each transcript. Reads with gap junctions greater than or equal to 125 bp were then extracted and labeled as spliced reads. The number of spliced reads over *Airn* and *Xist*, were then divided by total merged reads over each transcript, respectively.

3.3.11 H3K27me3 ChIP-Seq

Prior to crosslinking, RMCE cell lines (with rtTA) were treated with 0-1000ng/mL doxycycline (1000ng/mL if not otherwise specified) for 72 hours (or longer, 7 days). Adhered cells were crosslinked with 0.6% formaldehyde (Fisher Scientific BP531-500) in DMEM media with 5% FBS for 10min at room temperature, then quenched with 125mM glycine for 5min at room temperature. Crosslinked cells were then washed twice with ice-cold PBS and scraped with

ice-cold PBS with 0.05% Tween (Fisher Scientific EW-88065-31) and PIC (Sigma Aldrich P8340). The cells were then spun at 1,200 x g at 4°C to remove PBS, followed by resuspension in ice-cold PBS with PIC and divided into 10-million cell aliquots. Each ChIP was performed using 10 million cells, 10µL H3K27me3 antibody (Cell Signaling 9733), and 30µL Protein A/G agarose beads (Santa Cruz sc-2003). Antibody-conjugated beads were prepared by incubating antibody with beads in 300µL Blocking Buffer (PBS, 0.5% BSA [Invitrogen AM2616]) overnight at 4°C with rotation.

10 million crosslinked cells were resuspended in 1mL Lysis Buffer 1 (50mM HEPES pH 7.5, 140mM NaCl, 1mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, PIC) and incubated with rotation for 10min at 4°C. Cells were then resuspended in 1mL Lysis Buffer 2 (10mM Tris-HCl pH 8.0, 200mM NaCl, 1mM EDTA, 0.5 mM EGTA, PIC) for 10 minutes at room temperature. All buffer removal steps were performed with 5-min 1,200 x g spins at 4°C. The extracted nuclei pellet was then resuspended in 500µL Lysis Buffer 3 (10mM Tris-HCl pH 8.0, 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.1% sodium-deoxycholate, 0.5% N-lauroylsarcosine, PIC), and chromatin was sonicated to 100-500bp fragments using a Vibracell VX130 (Sonics) with the following parameters: 10 cycles of 30% intensity for 30 seconds with 1 minute of rest on ice between cycles.

Soluble chromatin was obtained with a 30-min max speed spin at 4°C, mixed with 1% Triton X-100, and then incubated with pre-conjugated antibody beads overnight at 4°C with rotation. The ChIP beads were washed five times in 1mL RIPA Buffer (50mM HEPES pH 7.5, 500mM LiCl, 1mM EDTA, 1% NP-40, 0.7% sodium-deoxycholate, PIC) and once with 1mL TE, each for 5 minutes at 4°C with rotation and spun at 2,000 x g for 2 minutes for buffer removal. The washed beads were then resuspended in Elution buffer (50mM Tris-HCl pH 8.0, 10mM

EDTA, 1% SDS) and placed on a 65°C heat block for 17 minutes with frequent vortexing. ChIP DNA was then reverse crosslinked in 0.5% SDS and 100mM NaCl overnight at 65°C, followed by a 1-h RNaseA (3µL; Thermo Scientific EN0531) treatment at 37°C and a 2.5-h Proteinase K (10µL; Invitrogen 25530015) treatment at 56°C. DNA was then extracted with 1 volume of phenol:chloroform:isoamyl alcohol (Sigma-Aldrich P3803) and precipitated with 2 volumes 100% ethanol, 1/10 volume 3M sodium-acetate pH 5.4, and 1/1000 volume linear acrylamide (Invitrogen AM9520) overnight at -20°C. Precipitated DNA was then extracted with a 30-minute max speed spin at 4°C, washed once with ice-cold 80% ethanol, and resuspended in TE.

ChIP-Seq libraries were prepared with NEBNext End Repair Module (NEB E6050S), A-tailing by Klenow Fragment (3'→5' exo-; NEB M0212S), and TruSeq 6-bp index adaptor ligation by Quick ligase (NEB M2200S), and NEBNext High-Fidelity 2X PCR Master Mix (NEB M0541S). All DNA clean-up steps were performed with AMPure XP beads (Beckman Coulter A63880). Single-end, 75-bp sequencing was performed using an Illumina NextSeq 500/550 High Output v2.5 kit (Illumina 20024906) on a NextSeq 500 System.

ChIP-Seq reads were aligned using bowtie2 with default parameters [31]. Aligned reads that had a mapping quality greater than or equal to 30 were extracted with samtools [30], and allele-specific read retention (i.e., reads that overlap at least one B6 or CAST SNP) was performed as in [32,33] using a custom Perl script (intersect_reads_snps16.pl). For chromosome-scale tiling density plots, B6- or CAST-specific reads were summed in 10kb bins across chr6. Binned counts were then divided by the total number of reads in the dataset and divided by a million (i.e., RPM), then divided by the number of B6/CAST SNPs detected in the bin genomic coordinates (i.e., SNP-norm RPM). Finally, bins were averaged every 9 bins in 1bin increments. For plotting, only bins with greater or equal to 25 B6/CAST SNPs were retained for higher confidence

interpretation of the allelic data. All plots were generated using ggplot2 [34] in RStudio. For statistical comparisons of H3K27me3 levels across conditions, SNP-norm RPM signal of chr6 bins were divided by quartiles, then subset for every 10th bin. An ANOVA-TukeyHSD test was applied to all B6 and CAST data for adjusted p values; two conditions were considered significantly different with an adjusted p value threshold of < 0.0001.

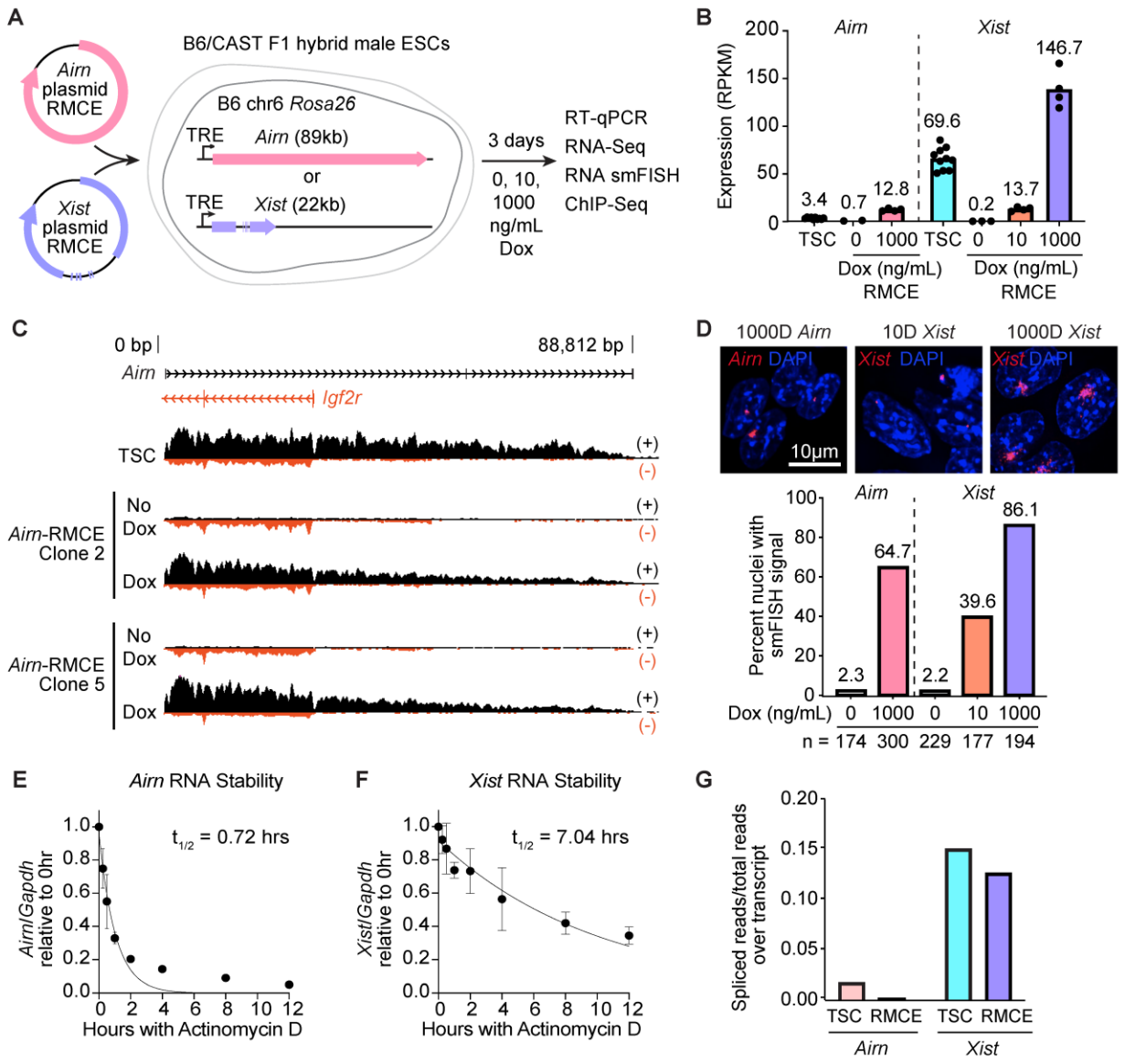


Figure 3.1. Ectopically expressed *Airn* resembles the endogenous *Airn* transcript.
(A) Schematic of *Airn* and *Xist* gene insertion into the *Rosa26* locus on B6 chr6 in F1-hybrid B6/CAST ESCs via RMCE. **(B)** RNA-Seq RPKM expression levels of *Airn* and *Xist* in TSCs and in RMCE ESCs following treatment with various concentrations of dox. RPKM, Reads Per Kilobase of transcript, per Million mapped reads. **(C)** Genome browser scaled view of strand-specific RNA-Seq signal from *Airn* in TSCs versus untreated and dox-treated *Airn*-RMCE ESC clones. **(D)** Representative *Xist* and *Airn* RNA smFISH images (top) and counts of nuclei with detectable smFISH signal (bottom) for RMCE ESCs treated with the indicated concentration of dox, labeled with Quasar 570-conjugated probes (red) and DAPI (blue). Scale, 10μm. **(E-F)** Half-life measurements for *Airn* and *Xist* RNA in RMCE ESCs determined by RT-qPCR following treatment with actinomycin D for the indicated time. **(G)** Extent of splicing of *Xist* and *Airn* transcripts in TSCs and RMCE ESCs.

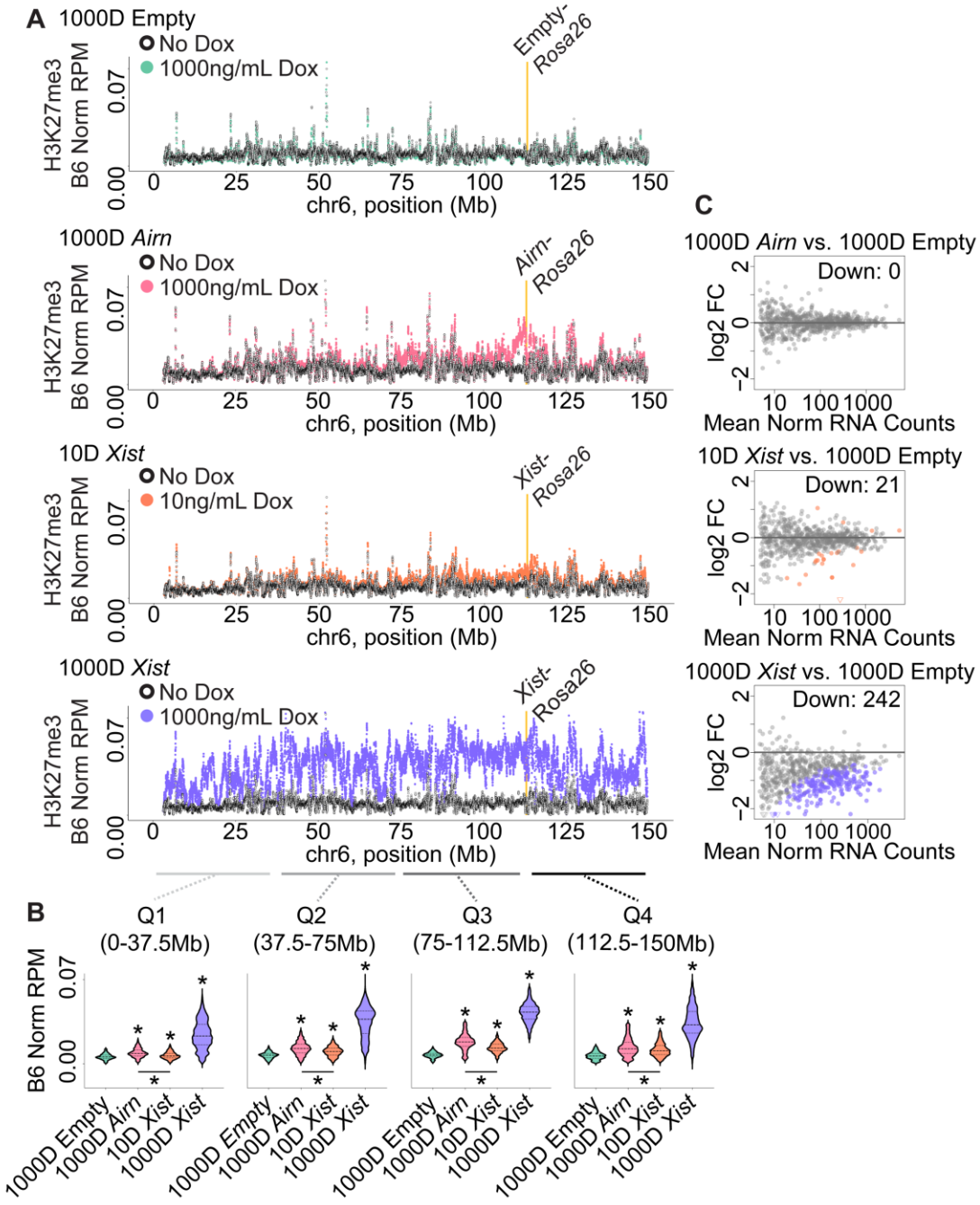


Figure 3.2 Changes in chr6 H3K27me3 density and gene expression induced by *Airn* and *Xist*.

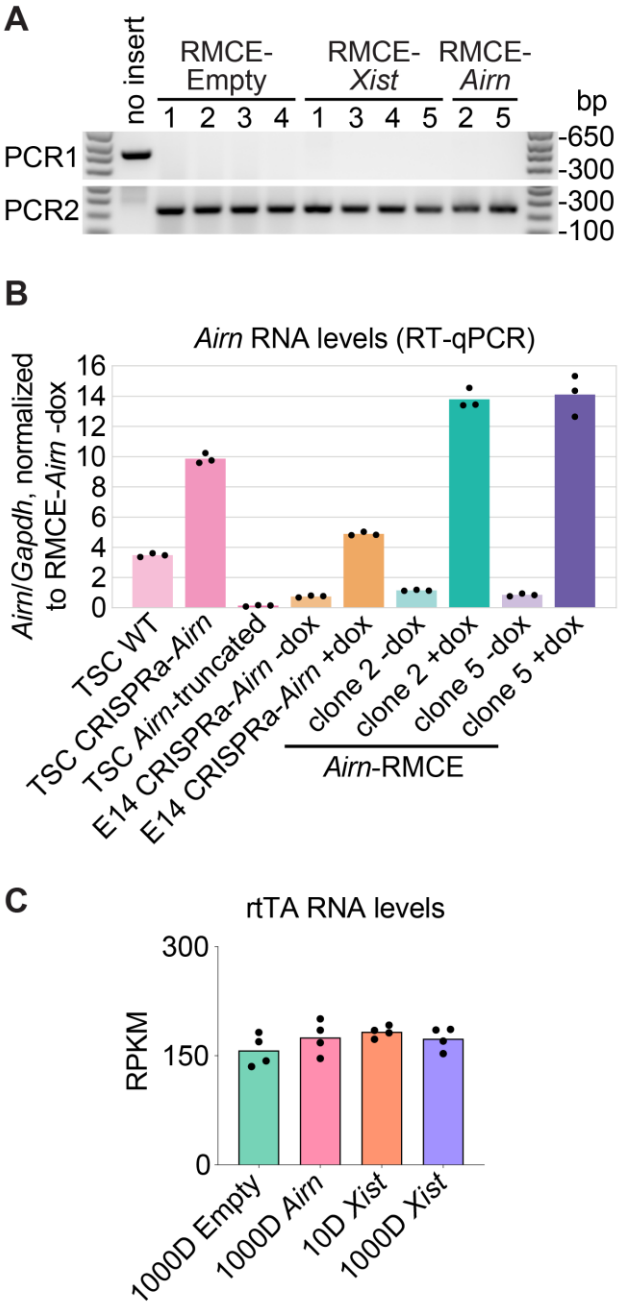
(A) Tiling density plots of H3K27me3 ChIP-Seq signal over B6 chr6 in untreated and dox-treated RMCE ESCs. Yellow bar, *Rosa26* locus: RMCE cargo insertion site. **(B)** H3K27me3 levels on B6 chr6 in dox-treated RMCE ESCs in binned quartiles. Asterisks, TukeyHSD adjusted p value < 0.0001; top asterisks represent comparisons to 1000D Empty. **(C)** MA plots showing differential expression of B6 chr6 genes relative to 1000ng/mL dox-treated empty-cargo RMCE ESCs. Log fold change was plotted against the mean of normalized B6 counts. Each dot is an expressed gene, and colored dots show those that are significantly changed as determined by DESeq2 analysis (adjusted p value < 0.01).

3.4 Supplementary Table Legends

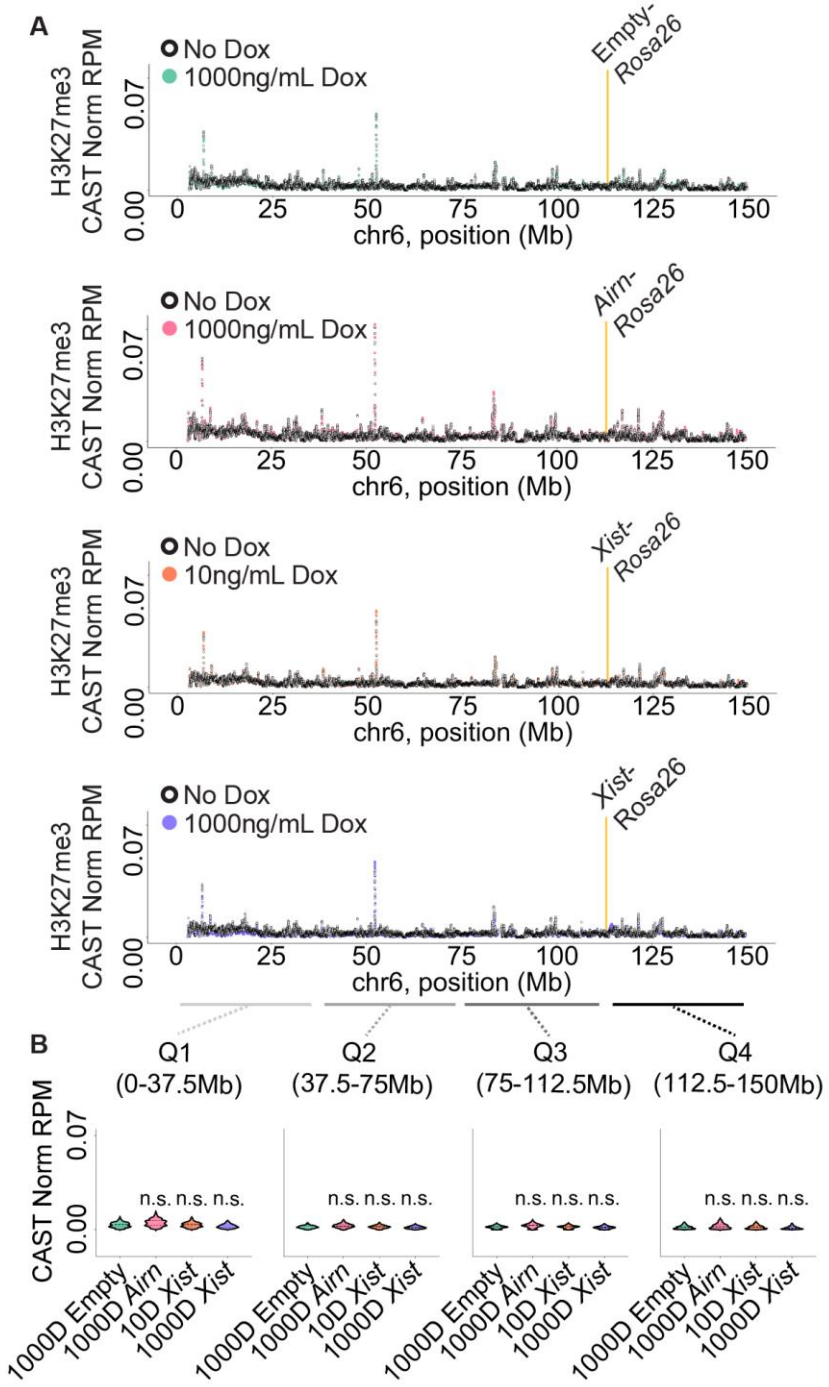
Supplementary Table 3.1. Differential expression analysis of RNA-seq data. Related to Figure 3.2. DESeq2 was used to analyze levels of B6-allele-specific RNA reads of all genes following 3- or 7-d expression of *Airn* (1000 ng/mL dox) or 3-d expression of *Xist* (10 or 1000 ng/mL dox) from the *Rosa26* locus of B6 chr6, relative to empty-cargo control (3-d treatment, 1000 ng/mL dox). For each comparison (different tabs of the spreadsheet), DESeq2 output was filtered to show only genes on B6 chr6 and then sorted by increasing adjusted p-value. chr6 genes with padj < 0.01 were considered significantly different (shown in color). The number of independent clones used for RNA-seq were as follows: 2 for "Airn3d", 2 for "Airn7d", 4 for "Xist10ngmL", 4 for "Xist1000ngmL", and 4 for "empty".

Supplemental Table 3.2. All high throughput sequencing datasets used. Related to Figures 3.1-3.2, Supplementary Figures 3.1-3.2, and Supplementary Table 3.1. Table gives all high throughput sequencing genomic datasets used in this study and is divided into 2 sections: "Datasets generated in this study" and "Publicly available datasets". Under each section, if applicable: "File ID" gives the name of the dataset; "Data Type" gives the experimental method (RNA-Seq versus ChIP-Seq); "Cell Type" gives the cell type and strain information when relevant; "Spike-in" says whether ERCC or HEK293 chromatin spike-ins were included; "Read Length" describes read length and single versus paired end sequencing; "Figures and Tables" lists the figures and tables in the manuscript where each dataset was used; and "GEO" gives the GEO database reference for the data.

Supplemental Table 3.3. Oligonucleotides, Related to Figures 3.1, Supplemental Figures 3.1. Table gives all oligonucleotide sequences used in this study. "ID" gives a descriptive name for the oligonucleotide; "Assay" describes the experimental method in which the oligonucleotide was used (RMCE genotyping versus RT-qPCR); and "Sequence" gives the oligonucleotide sequence in 5' to 3' order.



Supplemental Figure 3.1, Related to Figure 3.1.
(A) Genotyping PCR products confirming successful insertion of cargo sequences into B6 *Rosa26* via RMCE. Lanes are labeled with numbers corresponding to the ID of each clonal cell line. **(B)** RT-qPCR comparison of *Airn* levels in RMCE ESCs to those in TSC and ESC cell lines from previous studies [4,21]. **(C)** RNA-Seq RPKM expression levels of rtTA across dox-treated RMCE ESCs.



Supplemental Figure 3.2, Related to Figure 3.2.
(A) Tiling density plots of H3K27me3 ChIP-Seq signal over CAST chr6 in untreated and dox-treated RMCE ESCs. Yellow bar, *Rosa26* locus: RMCE cargo insertion site. **(B)** H3K27me3 levels in CAST chr6 across dox-treated RMCE ESCs in binned quartiles. Asterisks, TukeyHSD adjusted p value < 0.0001, as in Figure 2B; top asterisks represent comparisons to 1000D Empty.

REFERENCES

1. Trotman, J.B., Braceros, K.C.A., Cherney, R.E., Murvin, M.M. & Calabrese, J.M. The control of polycomb repressive complexes by long noncoding RNAs. *Wiley Interdiscip Rev RNA* **12**, e1657 (2021).

2. Andergassen, D., Smith, Z.D., Kretzmer, H., Rinn, J.L. & Meissner, A. Diverse epigenetic mechanisms maintain parental imprints within the embryonic and extraembryonic lineages. *Dev Cell* **56**, 2995-3005 e4 (2021).

3. Andergassen, D. *et al.* Mapping the mouse Allelome reveals tissue-specific regulation of allelic expression. *Elife* **6**(2017).

4. Schertzer, M.D. *et al.* lncRNA-Induced Spread of Polycomb Controlled by Genome Architecture, RNA Abundance, and CpG Island DNA. *Mol Cell* **75**, 523-537 e10 (2019).

5. Lyle, R. *et al.* The imprinted antisense RNA at the Igf2r locus overlaps but does not imprint Mas1. *Nat Genet* **25**, 19-21 (2000).

6. Seidl, C.I., Stricker, S.H. & Barlow, D.P. The imprinted Air ncRNA is an atypical RNAPII transcript that evades splicing and escapes nuclear export. *EMBO J* **25**, 3565-75 (2006).

7. Huang, R. *et al.* An RNA-Seq strategy to detect the complete coding and non-coding transcriptome including full-length imprinted macro ncRNAs. *PLoS One* **6**, e27288 (2011).

8. Anderson, K.M. *et al.* Transcription of the non-coding RNA upperhand controls Hand2 expression and heart development. *Nature* **539**, 433-436 (2016).

9. Engreitz, J.M. *et al.* Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* **539**, 452-455 (2016).

10. Groff, A.F. *et al.* In Vivo Characterization of Linc-p21 Reveals Functional cis-Regulatory DNA Elements. *Cell Rep* **16**, 2178-2186 (2016).

11. Winkler, L. *et al.* Functional elements of the cis-regulatory lincRNA-p21. *Cell Rep* **39**, 110687 (2022).

12. Trotman, J.B. *et al.* Elements at the 5' end of Xist harbor SPEN-independent transcriptional antiterminator activity. *Nucleic Acids Res* **48**, 10500-10517 (2020).

13. Sprague, D. *et al.* Nonlinear sequence similarity between the Xist and Rsx long noncoding RNAs suggests shared functions of tandem repeat domains. *RNA* **25**, 1004-1019 (2019).

14. Zylicz, J.J. *et al.* The Implication of Early Chromatin Changes in X Chromosome Inactivation. *Cell* **176**, 182-197 e23 (2019).

15. Dixon-McDougall, T. & Brown, C.J. Multiple distinct domains of human XIST are required to coordinate gene silencing and subsequent heterochromatin formation. *Epigenetics Chromatin* **15**, 6 (2022).

16. Kohlmaier, A. *et al.* A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. *PLoS Biol* **2**, E171 (2004).

17. Plath, K. *et al.* Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300**, 131-5 (2003).

18. Latos, P.A. *et al.* An in vitro ES cell imprinting model shows that imprinted expression of the Igf2r gene arises from an allele-specific expression bias. *Development* **136**, 437-48 (2009).

19. Holoch, D. *et al.* A cis-acting mechanism mediates transcriptional memory at Polycomb target genes in mammals. *Nat Genet* **53**, 1686-1697 (2021).

20. Lahn, B.T. & Page, D.C. Four evolutionary strata on the human X chromosome. *Science* **286**, 964-7 (1999).

21. Braceros, A.K. *et al.* Proximity-dependent recruitment of Polycomb Repressive Complexes by the lncRNA Airn. *bioRxiv* (2022).

22. Schertzer, M.D. *et al.* A piggyBac-based toolkit for inducible genome editing in mammalian cells. *RNA* **25**, 1047-1058 (2019).

23. Kirk, J.M. *et al.* Functional classification of long non-coding RNAs by k-mer content. *Nat Genet* **50**, 1474-1482 (2018).

24. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-82 (2012).

25. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).

26. Keane, T.M. *et al.* Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**, 289-94 (2011).

27. Frankish, A. *et al.* Gencode 2021. *Nucleic Acids Res* **49**, D916-D923 (2021).

28. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

29. Liao, Y., Smyth, G.K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-30 (2014).

30. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-9 (2009).

31. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-9 (2012).

32. Calabrese, J.M., Starmer, J., Schertzer, M.D., Yee, D. & Magnuson, T. A survey of imprinted gene expression in mouse trophoblast stem cells. *G3 (Bethesda)* **5**, 751-9 (2015).

33. Calabrese, J.M. *et al.* Site-specific silencing of regulatory elements as a mechanism of X inactivation. *Cell* **151**, 951-63 (2012).

34. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*, (Springer-Verlag, New York, 2016).

CHAPTER 4: Discussion and Future Directions

4.1 Main Conclusions

Repressive lncRNAs are critical regulators of gene expression with major roles in genomic imprinting. Until recently, our mechanistic understanding of lncRNA-mediated repression was largely based on *Xist* in XCI, but we and others have leveraged novel insights on *Airn* and its imprinted domain, which may better reflect other regulatory lncRNAs elsewhere in the genome. Our work takes advantage of F1-hybrid mouse cell lines, in which we profiled epigenetic events associated with *Xist* and *Airn* expression and their *cis*-acting nature in both endogenous and exogenous genomic contexts. In doing so, we report a series of intriguing relationships between 3D genome architecture, DNA regulatory elements, and PRC recruitment in *Airn*-mediated repression. Furthermore, our work sheds light on a fascinating distinction in repressive function between *Airn* and *Xist*, and that *Airn*'s relationship with PRCs is much more potent than originally thought. Altogether, the paradigms we established significantly advance our understanding of lncRNA-mediated repression and contributes to a generalizable model that likely apply to other repressive lncRNAs.

In Chapter 2 [32], we gained novel insights on the 15-Mb domain repressed by *Airn* in TSCs, including the CGI regulatory elements that we previously found to be bound by RING1B and EZH2 with high affinity on both alleles, to be in close proximity to the *Airn* locus in 3D space, and to be the most susceptible to *Airn*-induced PRC recruitment and PRC-catalyzed chromatin modifications [12,37]. Using allelic *in situ* Hi-C, we profiled, for the first time in high resolution, the 3D chromatin organization of the *Airn*-repressed domain and identified specific

long-distance interactions between *Airn* and target DNA that correlated in-step with extent of underlying *Airn*-recruited PRCs and PRC-deposited repressive marks. Consistent with our previous FISH microscopic observations, we found differences between *Airn*-expressing and non-expressing alleles, where the former exhibited both new and enhanced DNA interactions, and that the major regions of contact coincide with the PRC-bound CGIs [37]. Still, the DNA contacts on the non-expressing allele that persist and, in some cases occur with the same frequency as the *Airn*-expressing one, further supports the notion that lncRNAs exploit pre-existing genome structure to spread repression [37,65,68,71]. We also demonstrated concrete evidence of a role for *Airn*'s lncRNA product in long-distance repression by using CHART-Seq to show that *Airn* lncRNA associates with target chromatin to the same extent of which it forms 3D DNA contacts and recruits PRCs. Finally, upon further investigation of a functional role of PRC-bound CGIs that interact with the *Airn* locus, we unexpectedly found that deletion of specific CGIs alters long-distance transcriptional repression and PRC activity in a variable manner that correlated with changes in chromatin architecture. Collectively, our findings imply that the extent to which *Airn* expression recruits PRCs to chromatin is controlled by DNA regulatory elements that modulate proximity of the *Airn* lncRNA product to its target DNA. Importantly, our findings shed light on support of a model whereby *Airn* is a functional lncRNA and that DNA elements shape long-range contacts within the *Airn*-targeted domain in ways that extend beyond single loop-based models of regulation.

In Chapter 3 [86], we utilized our sophisticated mouse ESC system (referred as "RMCEs") to directly compare functional outcomes associated with ectopic *Xist* and *Airn* expression from the same chromosomal locus. We found similar transcriptional properties of ectopic *Xist* and *Airn* to their endogenous counterparts in TSCs [37]. Interestingly, the inducible component of the

ectopic expression for these lncRNAs revealed that *Xist* is subject to higher RNA production than *Airn*, which we allude to *Xist*'s inherit anti-termination activity as previously observed [12,83]. Together with this high RNA abundance, we see a chromosome-wide deposition of Polycomb by *Xist* through H3K27me3 analysis. We interpret these results as *Xist* anti-termination activity promoting enough RNA molecules to recruit and spread PRCs over the entire chromosome. Interestingly, upon expression to similar RNA abundance, we identified a much stronger capacity in *Airn* to induce Polycomb activity over the near entire length of the chromosome, doing so in a manner that matches or even surpasses that of *Xist*. While this is case, *Airn* does not induce transcriptional silencing of the underlying genes, thereby decoupling PRC activity and transcriptional repression as previously observed for *Xist* and other PRC-repressed regions [15,41,95–97]. Meanwhile, the degree of transcriptional silencing observed with *Xist* correlates with its RNA abundance, further supporting the notion that *Xist* accumulation helps it to achieve chromosome-wide silencing and underscoring the importance of RNA abundance in determining extent of repressive potency [37]. These data also raise the question of RNA-binding proteins (RBPs) that render initiation of transcriptional silencing by *Xist*, which motivates further investigation into the RBP-interactome of *Xist* versus *Airn*. Furthermore, because we do not observe transcriptional silencing with ectopic *Airn*, we interpret *Airn*'s function to PRCs as a mode of transcriptional memory of previously repressive events [98].

Altogether, our findings significantly advance the knowledge on the repressive action of *Airn* and highlight the importance of both genetic and epigenetic features in this process. While our work contribute great strides in this understanding, several areas within our collective findings raise new mechanistic questions worth investigating going forward. In what ways does PRC activity contribute to establishing and maintaining the *Airn*-repressed domain? What

genomic features are facilitating the unidirectionality observed with *Airn*-induced PRC recruitment and activity? How do genes escape lncRNA-mediated repression? We discuss these below.

4.2 PRC function in lncRNA-mediated repression

From our comparisons between lncRNA-repressed and non-repressed alleles across different genomics analyses, our findings suggest that 3D genome architecture, PRCs, and *Airn* work cooperatively to establish a compact and transcriptionally inactive domain. However, it is important to note that our observations are merely correlative and have not yet proven causation. In particular, it is unclear to what extent PRC function plays a role in both chromatin compaction and repression over the *Airn*-targeted domain.

In the presence of *Airn*, we see changes in 3D DNA interactions with its gene locus that correlate in-step with underlying *Airn* lncRNA association, PRC recruitment, and PRC-catalyzed chromatin marks across the targeted domain. These *Airn*-induced patterns illustrate a compact and repressed domain, which we attribute to the function of PRCs and importance for transcriptional memory. Indeed, PRC activity has previously been shown to induce chromatin compaction, both locally and distally, that propagate PRC spreading and PRC-catalyzed modifications in a positive feedforward fashion [16,63,99–101]. Thus, we hypothesize that *Airn*-recruited PRCs and the modifications they deposit on chromatin are responsible for the major changes in chromatin architecture dependent on *Airn* expression, thereby stabilizing the domain's repressive state.

Still, it is unclear to what extent these *Airn*-induced changes in chromatin architecture through PRCs is attributed to their catalyzed histone modifications or long-range interactions. To

address the former, we propose treating TSCs with catalytic inhibitors or creating mutations within the PRC catalytic subunits (RING1B and EZH2) that would disrupt its activity of histone mark deposition and assess how loss of these repressive marks would affect the chromatin changes by *Airn* [16,21,101–104]. However, given the steady-state imprinted nature of the *Airn*-repressed domain in TSCs, it is possible that no alterations would result from the loss of new PRC marks since these modifications are not required for the transcriptional silencing by *Airn*, nor the maintenance of PRC-mediated chromatin structures [16,105,106]. As an alternative approach, we propose performing these same experiments in *Airn*-inducible ESCs, either through upregulation of the endogenous promoter as in Chapter 2 or through ectopic expression as in Chapter 3, for temporal analyses. These experiments would demonstrate the role of PRC marks in establishing the chromatin architecture induced by *Airn* expression. To address the role of PRC-mediated chromatin organization in similar contexts, we suggest targeting the subunits associated with this function, namely CBX and PHCs in mammalian cPRC1, and monitor the same functional consequences in their absence [15,16,21,101–104]. These are just some of the possible approaches to dissect the function of PRCs in promoting *Airn*-mediated chromatin compaction and repression. On the other hand, our findings also suggest a role for PRC-mediated chromatin organization of the *Airn*-targeted domain prior to expression of *Airn*, which we attribute to the pre-bound PRCs at CGIs [12,32,52].

Given that we see pre-existing interactions between CGIs strongly bound by PRCs and the *Airn* locus, including the nearby PRC-bound CGI at *Igf2r*, we speculate these DNA contacts to be due, at least in part, by PRC-mediated chromatin organization activity. Such a 3D network of PRC-bound CGIs resemble what has been observed for PRCs at poised enhancer (PE) and silencer regulatory elements [16,27,105,107,108]. Indeed, both types of elements bound by PRCs exhibit

similar histone modification profiles and undergo long-range interactions with PRC target genes that control their repressive states in an interdependent manner. For example, in mouse ESCs, deletion of specific PRC2-bound silencers caused the interacting genes across multiple Mbs to undergo de-repression [16,27,107,109]. This is consistent with our own observation when we deleted the CGI at the *Slc22a3* gene promoter in TSCs, which also exhibited a multi-Mb reduction in transcriptional silencing and PRC activity across the interacting PRC-bound CGIs [32,52]. Thus, we believe that PRCs play an essential role in mediating the long-range interactions between PRC-bound CGIs and *Airn* that create a 3D proximity sufficient for *Airn* to recruit and spread more PRCs across its targeted domain. Whether PRC-independent factors are involved in this chromatin architecture is also of question.

4.3 Genomic factors that might influence the directionality of *Airn*-mediated repression

Our work led to a consistent observation of unidirectionality in *Airn*-induced repression towards the centromeric side of the lncRNA gene in all measurements [32,52,86]. Therefore, questions arise whether inherit features of the chromosome, such as the centromere itself or the presence of pre-existing long-range DNA interactions with the lncRNA locus, influence the direction of *Airn*-mediated gene silencing.

Given the structure of mouse chromosomes, which contain a centromere on one end and a telomere on the other, we speculate that the unidirectionality we observe for *Airn* in our mouse cell lines might be influenced by features of the centromere. Centromeres themselves are heterochromatic, or contain transcriptionally inactive and compact chromatin, similar to what is observed for the *Airn*-repressed domain, and like PRCs, heterochromatin can spread in a positive feedforward fashion [110–112]. Like PRC-directed chromatin modifications, mono-, di- and tri-

methylated lysine 9 of histone H3 (H3K9me1/2/3) are epigenetic marks of heterochromatin, which are deposited by specific SET domain-containing histone methyltransferases and recognized by other repressors that promote heterochromatin formation [111–114]. Indeed, recruitment of these H3K9me1/2/3 writer enzymes and their decorated histone modifications are observed over the *Airn*-repressed domain [114,115]. Thus, the inherit heterochromatin nature of the centromere could explain the directionality of *Airn*-induced repression and heterochromatin spreading towards the centromeric end. In addition, this could explain the exclusive unidirectionality effect observed by *Airn* and not by *Xist*, since repression and heterochromatin formation on the Xi does not require H3K9me1/2 writer, G9a [114]. To test if this is the case for *Airn*, we propose degrading G9a or inhibiting its methyltransferase activity, then measuring for functional consequences in directionality of *Airn*-induced PRC activity. Such findings would illuminate a new role for the centromere in heterochromatin formation and spreading.

Another explanation for the unidirectionality could be the orientation of the *Airn* gene itself. In both endogenous and exogenous chromosome contexts of our cell lines, the gene is located on the positive/sense strand, meaning euchromatin formation and transcription by Pol II over its 90-Kb length goes the opposite way towards the telomeric end. To determine if *Airn*'s silencing directionality is dependent on its gene transcription directionality, we propose inverting the gene and then measuring its effect on PRC activity. Indeed, if this is the case, it could contribute a new level of *Airn*'s repressive function through its act of transcription and would require further investigation into the potential role of Pol II.

Finally, it is clear that 3D genome architecture plays a key role in heterochromatin integrity, whereby higher-ordered structures, such as contact domains and DNA loops, help to control the promiscuous spread of heterochromatin [58,63,64,116,117]. This is demonstrated from our

Hi-C analysis over the *Airn*-targeted domain on chr17 in TSCs, where specific DNA contacts coincide with either inflection points or nucleation sites of PRC-occupied and PRC-modified chromatin when *Airn* is expressed. In general, the majority of pre-existing and *Airn*-enhanced interactions with the *Airn* locus are on the centromeric side of the gene, suggesting that these long-range structures could predict directionality of *Airn* silencing. From our preliminary analysis using ultra-deep Hi-C data from wildtype ESCs [105], we not only identify a higher density of specific pre-existing DNA interactions with the *Airn* gene on chr17 but also the *Rosa26* locus on chr6 (where *Airn* was inserted in our RMCEs) on the centromeric side (see Figures 2.4G and 4.1). Previous studies have demonstrated *Xist* to preferentially localize to target chromatin that undergo pre-existing DNA contacts with its gene locus, and these sites harbor a high density of active genes, which in turn predicts *Xist*-recruited PRCs and their directed modifications on the Xi [71,76]. Thus, we hypothesize that the regions interacting with *Rosa26* in RMCEs are highly dense with active genes that predict the location and directionality of *Airn*-mediated PRC deposition on chr6 in RMCEs. Cross-analyses using our current genomics data to address this hypothesis are in progress. The insights gained here could further strengthen the relationship between the 3D genome and transcription, particularly highlighting new ways of identifying regions elsewhere in the genome that are repressed by lncRNAs.

4.4 Epigenetic features that facilitate gene escape from lncRNA-mediated repression

Both *Xist* and *Airn*, among other lncRNAs, recruit PRCs and induce repression in a non-uniform fashion across their respective targeted domains [12,32,52,86,93]. Despite the focus on understanding how these lncRNAs establish repressed chromatin, the mechanisms that allow specific regions and the genes within them to escape silencing are poorly understood. Our

findings in Chapter 2 reveal that 3D genome architecture not only facilitates *Airn*-mediated repression of genes but also their escape from it. Two genes, *Qk* and *Map3k4*, located within the *Airn*-targeted domain on chr17 show strong, yet unchanged frequencies of interactions with the *Airn* locus. While they partially escape silencing by *Airn* at the transcriptional level, their gene bodies are completely devoid of *Airn* lncRNA, PRCs, and PRC-catalyzed histone modifications, resembling boundary elements that attenuate local spread of repression by *Airn* [32,52]. In addition, their CGIs exhibit robust biallelic epigenetic profiles of active chromatin marks, KDM2B-containing vPRC1, and SMC1A/Cohesin relative to other silenced CGIs, which implies that these regions likely undergo Cohesin-mediated interactions with the *Airn* locus and are perhaps signaled by their unique epigenetic features. Altogether, our findings suggest that variation in the genetic or epigenetic content of a DNA regulatory element has the potential to control gene expression by altering spatial equilibria between lncRNAs and their target genes.

The epigenetic features we identified in *Qk* and *Map3k4* beg the question of any shared mechanisms for escaping genes observed on the Xi that is repressed by *Xist*. On the Xi, ~5-15% of chrX genes invariably evade XCI in both humans and mice, which serve functional purpose in influencing sex-dependent phenotypic variability [118–123]. While we can identify these genes by allelic expression analysis of the Xi versus Xa, the mechanisms that facilitate their escape from *Xist* silencing remain unclear. In general, these genes exhibit genomic features that are generalizable to actively transcribed genes, including their Xa copies. Such features include Pol II binding, chromatin accessibility (i.e., DNase sensitivity), DNA hypomethylation, enrichment of active histone modifications, and depletion of inactive marks [67,93,118]. At least in the case of mice, escaping genes on the Xi are also associated with CTCF binding and have a higher propensity of being found exterior to the *Xist* RNA cloud in nuclear space [93]. Again, these

features are not mutually exclusive to escaping genes; they can also be observed for the same genes on the Xa, and they are not applicable to all escaping genes. Nonetheless, with the improvement and diversity of high-resolution genomics technologies over the years, we have the capacity to perform more comprehensive analyses of the genetic and epigenetic signatures associated with escaping genes, particularly in TSCs in which most of our experiments are performed in. We can also apply the insights we gained from the escaping genes of *Airn*-mediated repression and expand our assessments to escaping genes on chr6 under ectopic *Xist*-expressing conditions. For example, we can use the same Hi-C data from TSCs and ESCs to observe any pre-existing DNA interactions on chrX and chr6 that may be occurring with the lncRNA loci, as well as assess whether these interacting regions are concomitantly enriched for SMC1A/Cohesin and KDM2B-containing vPRC1 and depleted for PRCs, PRC-catalyzed histone marks, and *Xist* RNA. Speaking to the latter point, we have successfully performed *Xist* CHART-Seq in TSCs as we have done for *Airn*, which we can use to directly pinpoint genes that escape *Xist*-mediated silencing with an expected absence in CHART signal (Figure 4.2). These comprehensive analyses are currently ongoing in the lab. Overall, the perspectives gained from studying escape can open new doors to identifying features that both promote gene expression and escape within large repressed genomic regions, especially in diseases where genes are inappropriately repressed or de-repressed.

4.5 Final Thoughts

We report a series of pivotal evidence that support the model whereby *Airn* produces a functional lncRNA product that is responsible for the long-distance repression of imprinted genes. Our findings also highlight the importance of 3D genome architecture and DNA

regulatory elements in controlling *Airn*'s repressive potency through its relationship with PRCs. While *Airn* does not appear to induce transcriptional silencing through PRC deposition under ectopic conditions, it raises the questions of other factors, either genomic or repressive enzymes, that play a role in promoting the silencing over its endogenous targeted domain. If our hypothesis is correct in that *Airn*'s function through PRCs is important for maintaining transcriptional memory of genes, it suggests that mechanisms upstream or at the onset of *Airn* expression are involved in the initiation of *Airn* silencing. Nonetheless, given that PRC function is recognized for maintaining a poised yet repressed state of genes, it is plausible that *Airn*'s role in gene regulation is to recruit PRCs to keep genes in a silent yet flexible state that allow their reactivation when appropriate. This is especially critical in cases where cells need to differentiate, such as during embryonic development, stress response, and tissue wound repair [15,16]. Future studies are needed to definitively prove this notion and demonstrate the mechanism by which the *Airn* lncRNA might mediate its long-range repressive effects on chr17. The paradigms we established from the work discussed in this dissertation strongly support the notion that lncRNAs have critical functions in gene regulation and are essential for embryonic development and adult life. With the evidence we demonstrated on *Airn* and its repressive domain, we reveal both similarities and differences with *Xist* and the Xi that shed light on repressed domains elsewhere in the genome under the regulation by lncRNAs or locus control elements.

4.6 Methods

4.6.1 Chromosome tiling density plot of Hi-C O/E Rosa26 Viewpoint

Mouse ESC Hi-C data was obtained from [105] and processed as exactly described in Chapter 2.5.13.1, with the exception of using the mm10 genome build. Tiling density plot for chr6 at 10-Kb resolution showing O/E *Rosa26* viewpoint data was generated for total, non-allelic signal as detailed in Chapter 2.5.10 and 2.5.13.5.

4.6.2 Xist CHART-Seq in TSCs

C/B wildtype TSCs were cultured and prepared for CHART exactly as detailed in Chapter 2.5.1. *Xist* CHART-Seq was performed in duplicate as exactly described in the detailed methods in Chapter 2.5.7 with minor modifications. We took the 40 oligos tiled across the length of *Xist* from [70] and mixed them to a 100μM pool for in-house oligo biotin labeling as described earlier. Per *Xist* CHART reaction, 12.5 million cells worth of chromatin was incubated with 250pmol biotinylated oligo probe mix and purified with 200μL worth of Dynabeads MyOne Streptavidin C1 beads (Invitrogen, cat #: 65001).

Allelic *Xist* CHART-Seq data processing and chrX tiling density plot was performed as previously described using the mm10 and CAST pseudogenome genomes described in Chapter 3.3.9.

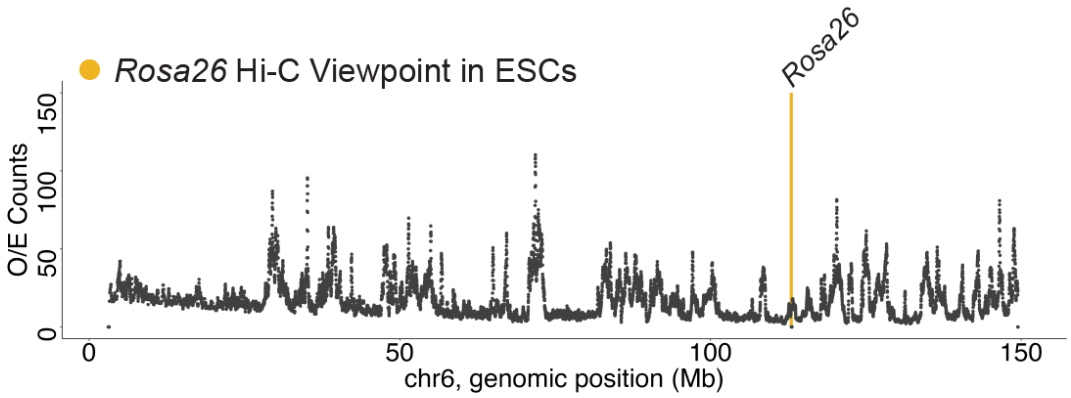


Figure 4.1 Pre-existing DNA interactions with the *Rosa26* locus in ESCs.
Tiling density plot of Hi-C *Rosa26* viewpoint O/E counts in ESCs [105]. O/E counts, observed-over-expected contact counts. Yellow bar, *Rosa26* locus.

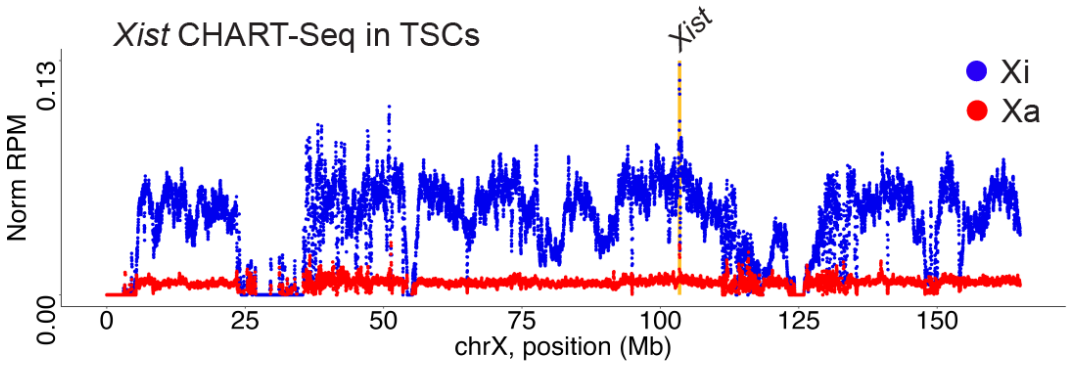


Figure 4.2 Intensity of chromatin association by *Xist* lncRNA in TSCs.
Tiling density plots of allelic *Xist* CHART-Seq, in C/B wildtype TSCs, n = 2. Norm RPM, Reads per Million total reads normalized for SNP density. Xi, inactive X-derived B6 signal. Xa, active X-derived CAST signal. Yellow bar, *Xist* locus.

REFERENCES

1. Theunissen, T. W. & Jaenisch, R. Mechanisms of gene regulation in human embryos and pluripotent stem cells. *Development* **144**, 4496–4509 (2017).

2. Latham, K. E. Preimplantation embryo gene expression: 56 years of discovery, and counting. *Mol. Reprod. Dev.* **90**, 169–200 (2023).

3. Monk, D., Mackay, D. J. G., Eggermann, T., Maher, E. R. & Riccio, A. Genomic imprinting disorders: lessons on how genome, epigenome and environment interact. *Nat. Rev. Genet.* **20**, 235–248 (2019).

4. Barlow, Denise P., Bartolomei, M. S. Genomic Imprinting in Mammals. *Cold Spring Harb. Perspect. Biol.* **6**, a018382 (2014).

5. Tucci, V. *et al.* Genomic Imprinting and Physiological Processes in Mammals. *Cell* **176**, 952–965 (2019).

6. Engreitz, J. M., Ollikainen, N. & Guttman, M. Long non-coding RNAs: Spatial amplifiers that control nuclear structure and gene expression. *Nature Reviews Molecular Cell Biology* **17**, 756–770 (2016).

7. Statello, L., Guo, C. J., Chen, L. L. & Huarte, M. Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* **22**, 96–118 (2021).

8. Autuoro, J. M., Pirnie, S. P. & Carmichael, G. G. Long noncoding RNAs in imprinting and X chromosome inactivation. *Biomolecules* **4**, 76–100 (2014).

9. MacDonald, W. A. & Mann, M. R. W. Long noncoding RNA functionality in imprinted domain regulation. *PLoS Genet.* **16**, 1–22 (2020).

10. Llères, D., Imaizumi, Y. & Feil, R. Exploring chromatin structural roles of non-coding RNAs at imprinted domains. *Biochem. Soc. Trans.* **49**, 1867–1879 (2021).

11. Yao, R. W., Wang, Y. & Chen, L. L. Cellular functions of long noncoding RNAs. *Nat. Cell Biol.* **21**, 542–551 (2019).

12. Trotman, J. B., Braceros, K. C. A., Cherney, R. E., Murvin, M. M. & Calabrese, J. M. The control of polycomb repressive complexes by long noncoding RNAs. *Wiley Interdiscip. Rev. RNA* **12**, 1–41 (2021).

13. Piunti, A. & Shilatifard, A. The roles of Polycomb repressive complexes in mammalian development and cancer. *Nat. Rev. Mol. Cell Biol.* **22**, 326–345 (2021).

14. Almeida, M., Bowness, J. S. & Brockdorff, N. The many faces of Polycomb regulation by RNA. *Current Opinion in Genetics and Development* **61**, 53–61 (2020).

15. Illingworth, R. S. Chromatin folding and nuclear architecture: PRC1 function in 3D.

Current Opinion in Genetics and Development **55**, 82–90 (2019).

16. Guo, Y. & Wang, G. G. Modulation of the high-order chromatin structure by Polycomb complexes. *Front. Cell Dev. Biol.* **10**, 1–10 (2022).

17. Vijayanathan, M., Trejo-Arellano, M. G. & Mozgová, I. Polycomb Repressive Complex 2 in Eukaryotes—An Evolutionary Perspective. *Epigenomes* **6**, 1–29 (2022).

18. Blackledge, N. P. & Klose, R. J. The molecular principles of gene regulation by Polycomb repressive complexes. *Nat. Rev. Mol. Cell Biol.* **22**, 815–833 (2021).

19. Simon, J. A. & Kingston, R. E. Occupying Chromatin: Polycomb Mechanisms for Getting to Genomic Targets, Stopping Transcriptional Traffic, and Staying Put. *Mol. Cell* **49**, 808–824 (2013).

20. Blackledge, N. P. *et al.* PRC1 Catalytic Activity Is Central to Polycomb System Function. *Mol. Cell* (2019). doi:10.1016/j.molcel.2019.12.001

21. Boyle, S. *et al.* A central role for canonical PRC1 in shaping the 3D nuclear landscape. *Genes Dev.* (2020). doi:10.1101/gad.336487.120

22. Blackledge, N. P. *et al.* Variant PRC1 Complex-Dependent H2A Ubiquitylation Drives PRC2 Recruitment and Polycomb Domain Formation. *Cell* **157**, 1445–1459 (2014).

23. Fursova, N. A. *et al.* Synergy between Variant PRC1 Complexes Defines Polycomb-Mediated Gene Repression. *Mol. Cell* (2019). doi:10.1016/j.molcel.2019.03.024

24. Healy, E. *et al.* PRC2.1 and PRC2.2 Synergize to Coordinate H3K27 Trimethylation. *Mol. Cell* (2019). doi:10.1016/J.MOLCEL.2019.08.012

25. Glancy, E. *et al.* PRC2.1- and PRC2.2-specific accessory proteins drive recruitment of different forms of canonical PRC1. *Mol. Cell* **83**, 1393-1411.e7 (2023).

26. Højfeldt, J. W. *et al.* Non-core Subunits of the PRC2 Complex Are Collectively Required for Its Target-Site Specificity. *Mol. Cell* (2019). doi:10.1016/J.MOLCEL.2019.07.031

27. Ngan, C. Y. *et al.* Chromatin interaction analyses elucidate the roles of PRC2-bound silencers in mouse development. *Nat. Genet.* **52**, 264–272 (2020).

28. Van Kruijsbergen, I., Hontelez, S. & Veenstra, G. J. C. Recruiting polycomb to chromatin. *Int. J. Biochem. Cell Biol.* **67**, 177–187 (2015).

29. Lee, J. T. & Bartolomei, M. S. X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell* **152**, 1308–1323 (2013).

30. Brockdorff, N. Noncoding RNA and Polycomb recruitment. *RNA* **19**, 429–442 (2013).

31. Bousard, A. *et al.* The role of Xist -mediated Polycomb recruitment in the initiation of X-chromosome inactivation . *EMBO Rep.* **20**, 119 (2019).

32. Braceros, A. K. *et al.* Proximity-dependent recruitment of Polycomb repressive complexes by the lncRNA Airn. *Cell Rep.* **42**, 112803 (2023).

33. Chu, C. *et al.* Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404–416 (2015).

34. Raposo, A. C., Casanova, M., Gendrel, A. V. & da Rocha, S. T. The tandem repeat modules of Xist lncRNA: A swiss army knife for the control of X-chromosome inactivation. *Biochem. Soc. Trans.* **49**, 2549–2560 (2021).

35. Nakamoto, M. Y., Lammer, N. C., Batey, R. T. & Wuttke, D. S. HnRNPK recognition of the B motif of Xist and other biological RNAs. *Nucleic Acids Res.* **48**, 9320–9335 (2020).

36. Pintacuda, G. *et al.* hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing. *Mol. Cell* **68**, 955-969.e10 (2017).

37. Schertzer, M. D. *et al.* lncRNA-Induced Spread of Polycomb Controlled by Genome Architecture, RNA Abundance, and CpG Island DNA. *Mol. Cell* **75**, 523-537.e10 (2019).

38. Colognori, D., Sunwoo, H., Kriz, A. J., Wang, C. Y. & Lee, J. T. Xist Deletional Analysis Reveals an Interdependency between Xist RNA and Polycomb Complexes for Spreading along the Inactive X. *Mol. Cell* **74**, 101-117.e10 (2019).

39. Wei, G., Almeida, M., Bowness, J. S., Nesterova, T. B. & Brockdorff, N. Xist Repeats B and C, but not Repeat A, mediate de novo recruitment of the Polycomb system in X chromosome inactivation. *Dev. Cell* **56**, 1234–1235 (2021).

40. Bousard, A. *et al.* The role of *Xist* -mediated Polycomb recruitment in the initiation of X-chromosome inactivation. *EMBO Rep.* (2019). doi:10.15252/embr.201948019

41. Żylicz, J. J. *et al.* The Implication of Early Chromatin Changes in X Chromosome Inactivation. *Cell* **176**, 182-197.e23 (2019).

42. Colognori, D., Sunwoo, H., Wang, D., Wang, C. Y. & Lee, J. T. Xist Repeat A contributes to early recruitment of Polycomb complexes during X-chromosome inactivation. *Dev. Cell* **56**, 1236–1237 (2021).

43. Andergassen, D. *et al.* The Airn lncRNA does not require any DNA elements within its locus to silence distant imprinted genes. (2019). doi:10.1371/journal.pgen.1008268

44. Andergassen, D. *et al.* Mapping the mouse Allelome reveals tissue-specific regulation of allelic expression. *Elife* **6**, 1–29 (2017).

45. Wutz, A., Smrzka, O. W. & Schweifer, N. Imprinted expression of the Igf2r gene depends on an intronic CpG island. *Nature* **389**, 745–749 (1997).

46. Sleutels, F., Zwart, R. & Barlow, D. P. The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* **415**, 810–813 (2002).

47. Yotova, I. Y. *et al.* Identification of the human homolog of the imprinted mouse Air non-coding RNA. *Genomics* **92**, 464–473 (2008).

48. Suzuki, S., Shaw, G. & Renfree, M. B. Identification of a novel antisense noncoding RNA, ALID, transcribed from the putative imprinting control region of marsupial IGF2R 06 Biological Sciences 0604 Genetics. *Epigenetics and Chromatin* **11**, 1–11 (2018).

49. Liu, X. *et al.* Genomic imprinting of the IGF2R/AIR locus is conserved between bovines and mice. *Theriogenology* **180**, 121–129 (2022).

50. Braidotti, G, Baubec, T, Seidl, C, Smrzka, O, Stricker, S, Yotova, I, Barlow, D. *The Air Noncoding RNA: An Imprinted cis-silencing Transcript*. (2004).

51. Seidl, C. I. M., Stricker, S. H. & Barlow, D. P. The imprinted Air ncRNA is an atypical RNAPII transcript that evades splicing and escapes nuclear export. *EMBO J.* **25**, 3565–3575 (2006).

52. Schertzer, M. D. *et al.* lncRNA-Induced Spread of Polycomb Controlled by Genome Architecture, RNA Abundance, and CpG Island DNA. *Mol. Cell* **75**, 523-537.e10 (2019).

53. Calabrese, J. M., Starmer, J., Schertzer, M. D., Yee, D. & Magnuson, T. A survey of imprinted gene expression in mouse trophoblast stem cells. *G3 Genes, Genomes, Genet.* **5**, 751–759 (2015).

54. Latos, P. A. *et al.* Airn Transcriptional Overlap, But Not Its lncRNA Products, Induces Imprinted Igf2r Silencing. *Science (80-.).* **338**, 1469–1472 (2012).

55. Pauler, F. M., Barlow, D. P. & Hudson, Q. J. Mechanisms of long range silencing by imprinted macro non-coding RNAs. *Curr. Opin. Genet. Dev.* **22**, 283–289 (2012).

56. Sleutels, F., Tjon, G., Ludwig, T. & Barlow, D. P. Imprinted silencing of Slc22a2 and Slc22a3 does not need transcriptional overlap between Igf2r and Air. *EMBO J.* **22**, 3696–3704 (2003).

57. Stricker, S. H. *et al.* Silencing and transcriptional properties of the imprinted Airn ncRNA are independent of the endogenous promoter. *EMBO J.* **27**, 3116–3128 (2008).

58. Li, Y., Hu, M. & Shen, Y. Gene regulation in the 3D genome. *Hum. Mol. Genet.* **27**, R228–R233 (2018).

59. van Steensel, B. & Furlong, E. E. M. The role of transcription in shaping the spatial organization of the genome. *Nat. Rev. Mol. Cell Biol.* **20**, 327–337 (2019).

60. Venkatesh, S. & Workman, J. L. Histone exchange, chromatin structure and the regulation of transcription. *Nat. Rev. Mol. Cell Biol.* **16**, 178–189 (2015).

61. Dekker, J. & Mirny, L. The 3D Genome as Moderator of Chromosomal Communication. *Cell* **164**, 1110–1121 (2016).

62. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).

63. Rao, S. S. P. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. (2014). doi:10.1016/j.cell.2014.11.021

64. Gibcus, J. H. & Dekker, J. The Hierarchy of the 3D Genome. *Mol. Cell* **49**, 773–782 (2013).

65. Yan, F., Wang, X. & Zeng, Y. 3D genomic regulation of lncRNA and Xist in X chromosome. *Semin. Cell Dev. Biol.* (2019). doi:10.1016/j.semcdb.2018.07.013

66. Gil, N. & Ulitsky, I. Regulation of gene expression by cis-acting long non-coding RNAs. *Nat. Rev. Genet.* **21**, 102–117 (2020).

67. Fang, H., Disteche, C. M. & Berletch, J. B. X Inactivation and Escape: Epigenetic and Structural Features. *Front. Cell Dev. Biol.* **7**, 1–12 (2019).

68. Pandya-Jones, A. & Plath, K. The "lnc" between 3D chromatin structure and X chromosome inactivation. *Semin. Cell Dev. Biol.* **56**, 35–47 (2016).

69. Giorgetti, L. *et al.* Structural organization of the inactive X chromosome in the mouse. *Nature* **535**, 575–579 (2016).

70. Simon, M. D. *et al.* High-resolution Xist binding maps reveal 2-step spreading during X-inactivaiton. *Nature* **504**, 465–469 (2014).

71. Engreitz, J. M. *et al.* The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science (80-.).* **341**, 1–9 (2013).

72. Deng, X. *et al.* Bipartite structure of the inactive mouse X chromosome. *Genome Biol.* **16**, 1–21 (2015).

73. Froberg, J. E., Pinter, S. F., Kriz, A. J., Jégu, T. & Lee, J. T. Megadomains and superloops form dynamically but are dispensable for X-chromosome inactivation and gene escape. *Nat. Commun.* **9**, 1–19 (2018).

74. Wang, C. Y., Jégu, T., Chu, H. P., Oh, H. J. & Lee, J. T. SMCHD1 Merges Chromosome Compartments and Assists Formation of Super-Structures on the Inactive X. *Cell* 406–421 (2018). doi:10.1016/j.cell.2018.05.007

75. Jansz, N. *et al.* Smchd1 Targeting to the Inactive X Is Dependent on the Xist-HnrnpK-PRC1 Pathway. *Cell Rep.* **25**, (2018).

76. Simon, M. D. *et al.* High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. *Nature* **504**, 465–469 (2013).

77. Cerase, A., Pintacuda, G., Tattermusch, A. & Avner, P. Xist localization and function:

New insights from multiple levels. *Genome Biol.* **16**, 1–12 (2015).

78. Brockdorff, N. Localized accumulation of Xist RNA in X chromosome inactivation. *Open Biol.* **9**, (2019).

79. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **25**, 1010–22 (2011).

80. Blackledge, N. P. & Klose, R. J. CpG island chromatin: A platform for gene regulation. *Epigenetics* **6**, 147–152 (2011).

81. Simon, J. A. & Kingston, R. E. Occupying Chromatin: Polycomb Mechanisms for Getting to Genomic Targets, Stopping Transcriptional Traffic, and Staying Put. *Mol. Cell* **49**, 808–824 (2013).

82. Ogiyama, Y., Schuettengruber, B., Papadopoulos, G. L., Chang, J. M. & Cavalli, G. Polycomb-Dependent Chromatin Looping Contributes to Gene Silencing during Drosophila Development. *Mol. Cell* **71**, 73-88.e5 (2018).

83. Trotman, J. B. *et al.* Elements at the 5′ end of Xist harbor SPEN-independent transcriptional antiterminator activity. *Nucleic Acids Res.* **48**, 10500–10517 (2020).

84. Holding, C. Autosome inactivation by ectopic XISTin differentiated cells. *Genome Biol.* **3**, (2002).

85. Hall, L. L. *et al.* An ectopic human XIST gene can induce chromosome inactivation in postdifferentiation human HT-1080 cells. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 8677–8682 (2002).

86. Trotman, J. B. *et al.* Ectopically expressed Airn lncRNA deposits Polycomb with a potency that rivals Xist. *bioRxiv* (2023). doi:10.1101/2023.05.09.539960

87. Engreitz, J. M. *et al.* Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* **539**, 452–455 (2016).

88. McAninch, D., Roberts, C. T. & Bianco-Miotto, T. Mechanistic insight into long noncoding RNAs and the placenta. *International Journal of Molecular Sciences* **18**, (2017).

89. Maclary, E. *et al.* PRC2 represses transcribed genes on the imprinted inactive X chromosome in mice. *Genome Biol.* **18**, 1–17 (2017).

90. Quinn, J., Kunath, T. & Rossant, J. Mouse trophoblast stem cells. *Methods Mol. Med.* **121**, 125–148 (2006).

91. Oda, M., Shiota, K. & Tanaka, S. Trophoblast Stem Cells. *Methods Enzymol.* **419**, 387–400 (2006).

92. Roberts, R. M. & Fisher, S. J. Trophoblast Stem Cells. *Biol. Reprod.* **84**, 412–21 (2011).

93. Calabrese, J. M. *et al.* Site-specific silencing of regulatory elements as a mechanism of x inactivation. *Cell* **151**, 951–963 (2012).

94. Rossant, J. & Cross, J. C. Placental development: Lessons from mouse mutants. *Nature Reviews Genetics* **2**, 538–548 (2001).

95. Plath, K. *et al.* Role of histone H3 lysine 27 methylation in X inactivation. *Science (80-.).* **300**, 131–135 (2003).

96. Kohlmaier, A. *et al.* A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. *PLoS Biol.* **2**, (2004).

97. Dixon-McDougall, T. & Brown, C. J. Multiple distinct domains of human XIST are required to coordinate gene silencing and subsequent heterochromatin formation. *Epigenetics and Chromatin* **15**, 1–18 (2022).

98. Holoch, D. *et al. A cis-acting mechanism mediates transcriptional memory at Polycomb target genes in mammals*. *Nature Genetics* **53**, (Springer US, 2021).

99. Terranova, R. *et al.* Polycomb Group Proteins Ezh2 and Rnf2 Direct Genomic Contraction and Imprinted Repression in Early Mouse Embryos. *Dev. Cell* **15**, 668–679 (2008).

100. Hernández-Romero, I. A. & Valdes, V. J. De Novo Polycomb Recruitment and Repressive Domain Formation. *Epigenomes* **6**, (2022).

101. Plys, A. J. *et al.* Phase separation of polycomb-repressive complex 1 is governed by a charged disordered region of CBX2. *Genes Dev.* **33**, 799–813 (2019).

102. Isono, K. *et al.* SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. *Dev. Cell* **26**, 565–577 (2013).

103. Shukla, S. *et al.* Small-molecule inhibitors targeting Polycomb repressive complex 1 RING domain. *Nat. Chem. Biol.* **17**, 784–793 (2021).

104. Danishuddin, Subbarao, N., Faheem, M. & Khan, S. N. Polycomb repressive complex 2 inhibitors: emerging epigenetic modulators. *Drug Discov. Today* **24**, 179–188 (2019).

105. Bonev, B. *et al.* Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171**, 557-572.e24 (2017).

106. Du, Z. *et al.* Polycomb Group Proteins Regulate Chromatin Architecture in Mouse Oocytes and Early Embryos. *Mol. Cell* **77**, 825-839.e7 (2020).

107. Oksuz, O. *et al.* Capturing the Onset of PRC2-Mediated Repressive Domain Formation. *Mol. Cell* **70**, 1149-1162.e5 (2018).

108. Crispatzu, G. *et al.* The chromatin, topological and regulatory properties of pluripotency-

associated poised enhancers are conserved in vivo. *Nat. Commun.* **12**, 1–17 (2021).

109. Kraft, K. *et al.* Polycomb-mediated genome architecture enables long-range spreading of H3K27 methylation. *Proc. Natl. Acad. Sci. U. S. A.* **119**, 1–10 (2022).

110. Bloom, K. S. Centromeric Heterochromatin: The Primordial Segregation Machine. *Annu. Rev. Genet.* **48**, 457–484 (2014).

111. Allshire, R. C. & Madhani, H. D. Ten principles of heterochromatin formation and function. *Nat. Rev. Mol. Cell Biol.* **19**, 229–244 (2018).

112. Cutter DiPiazza, A. R. *et al.* Spreading and epigenetic inheritance of heterochromatin require a critical density of histone H3 lysine 9 tri-methylation. *Proc. Natl. Acad. Sci. U. S. A.* **118**, 1–10 (2021).

113. Nicetto, D. & Zaret, K. Role of H3K9me3 Heterochromatin in Cell Identity Establishment and Maintenance. *Curr. Opin. Genet. Dev.* **55**, 1–10 (2019).

114. Shankar, S. R. *et al.* G9a, a multipotent regulator of gene expression. *Epigenetics* **8**, 16–22 (2013).

115. Nagano, T. *et al.* The Air Noncoding RNA Epigenetically Silences Transcription by Targeting G9a to Chromatin. *Science (80-.).* **322**, 1717–1720 (2008).

116. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, (2012).

117. Haws, S. A., Simandi, Z., Barnett, R. J. & Phillips-Cremins, J. E. 3D genome, on repeat: Higher-order folding principles of the heterochromatinized repetitive genome. *Cell* **185**, 2690–2707 (2022).

118. Navarro-Cobos, M. J., Balaton, B. P. & Brown, C. J. Genes that escape from X-chromosome inactivation: Potential contributors to Klinefelter syndrome. *Am. J. Med. Genet. Part C Semin. Med. Genet.* **184**, 226–238 (2020).

119. Zhang, Y. *et al.* Genes that escape X-inactivation in humans have high intraspecific variability in expression, are associated with mental impairment but are not slow evolving. *Mol. Biol. Evol.* **30**, 2588–2601 (2013).

120. Balaton, B. P., Dixon-Mcdougall, T., Peeters, S. B. & Brown, C. J. The eXceptional nature of the X chromosome. *Hum. Mol. Genet.* **27**, R242–R249 (2018).

121. Berletch, J. B., F, Y. & Disteche, C. M. Escape from X inactivation in mice and humans. *Genome Biol.* **11**, (2010).

122. Wainer Katsir, K. & Linial, M. Human genes escaping X-inactivation revealed by single cell expression data. *BMC Genomics* **20**, 1–17 (2019).

123. Carrel, L. & Brown, C. J. When the lyon(Ized chromosome) roars: Ongoing expression from an inactive X chromosome. *Philos. Trans. R. Soc. B Biol. Sci.* **372**, (2017).