Published in final edited form as:

Trends Cell Biol. 2014 November ; 24(11): 651–663. doi:10.1016/j.tcb.2014.08.009.

Long non-coding RNAs: An emerging link between gene regulation and nuclear organization

Sofia Quinodoz¹ and Mitchell Guttman¹

Mitchell Guttman: mguttman@caltech.edu

¹Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125

Abstract

Mammalian genomes encode thousands of long non-coding RNAs (lncRNAs) that play important roles in diverse biological processes. As a class, lncRNAs are generally enriched in the nucleus and specifically within the chromatin-associated fraction. Consistent with their localization, many lncRNAs have been implicated in the regulation of gene expression and in shaping three-dimensional nuclear organization. Here, we discuss the evidence that many nuclear-retained lncRNAs can interact with various chromatin regulatory proteins and recruit them to specific sites on DNA to regulate gene expression. Furthermore, we discuss the role of specific lncRNAs in shaping nuclear organization and their emerging mechanisms. Based on these examples, we propose a model that explains how lncRNAs may shape aspects of nuclear organization to regulate gene expression.

Keywords

lncRNA; chromatin regulation; genome organization; nuclear domains

RNA and three-dimensional nuclear organization

Although the entire genome is present within the nucleus of every cell, distinct genes need to be accessed and expressed in different cellular conditions. Accordingly, the nucleus of each cell is a highly organized arrangement of DNA, RNA, and protein that is dynamically assembled and regulated in different cellular states^{1–3}. These dynamic nuclear structures are largely arranged around functionally-related roles and often occur across multiple chromosomes^{2–4}. These include large nuclear bodies (ie nucleolus^{5,6}, nuclear speckle⁷, and paraspeckle^{8,9}), gene-gene interactions (ie transcription factories^{10–12} and polycomb bodies^{13–16}), and promoter-enhancer interactions¹⁷. Yet, the molecular components involved in establishing this dynamic organization are still largely unknown^{1–3}.

© 2014 Elsevier Ltd. All rights reserved.

Correspondence to: Mitchell Guttman, mguttman@caltech.edu.

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

It has long been suspected that RNA might play a role in organizing the structure of the nucleus. Early studies of heterogeneous nuclear RNA (hnRNA) identified a large proportion of poly(A)-modified RNA that was retained in the nucleus and was of a distinct composition from messenger RNA (mRNA) and their precursors^{18–20}. Many of these RNAs were found to be localized to precise regions of the nucleus including nuclear speckles²¹ and other chromatin-associated regions^{21,22}. Subsequent studies showed that global disruption of RNA transcription, but not protein translation, led to visible rearrangements of nuclear organization²³. These studies led to the proposal that nuclear-retained RNAs might play an important structural role in the nucleus^{18,21,23}.

Over the last decade many thousands of functional lncRNAs have been identified^{24–27}. Recent work has highlighted that many of these lncRNAs can play important roles in diverse biological processes^{28–32, 33–37}. As a class, these lncRNAs are generally enriched in the nucleus and specifically within the chromatin-associated fraction^{27,38}. Accordingly, most work on lncRNAs have focused on their role in gene regulation and, specifically, in the recruitment of chromatin regulatory proteins to genomic DNA locations^{25,39,40}. In addition to this role, several recent studies have demonstrated another important role for lncRNAs in the nucleus – that is, several lncRNAs are essential for organizing distinct nuclear structures^{41–50}.

While lncRNAs are likely to fall into many different classes with different mechanisms^{25,39,40}, in this review we focus exclusively on nuclear-retained lncRNAs that are involved in the regulation of gene expression^{25,40} and in shaping three-dimensional nuclear organization^{4,35,42–45,51}. Here, we discuss the evidence demonstrating that several lncRNAs can interact with various chromatin regulatory proteins, recruit them to specific sites on DNA, modify chromatin, and regulate gene expression. Furthermore, we discuss the role of specific lncRNAs in shaping aspects of three-dimensional nuclear organization and the emerging mechanisms by which they perform this role. Based on these examples, we synthesize the observed data into a model that may explain how some lncRNAs can shape nuclear organization to regulate gene expression – highlighting how these two apparently distinct roles may indeed occur through a shared mechanism.

Mechanisms of lncRNA regulation of gene expression through chromatin regulation

It is becoming increasingly clear that many lncRNAs can act to affect various gene expression programs^{25,40}. Initial evidence for the role of lncRNAs in gene regulation came from studies of mammalian X-chromosome inactivation (XCI), a process that entails silencing of an entire X-chromosome in females during development⁵². This process is orchestrated by the Xist lncRNA, which is transcribed exclusively from the inactive X-chromosome (Xi)^{53–55} and coats the entire Xi⁵⁶. Importantly, genetic deletion of Xist prevents XCI⁵⁷, and induction of Xist is sufficient to initiate XCI on the same chromosome from which it is transcribed^{58,59}. This silencing capability is dependent on a discrete region of the lncRNA, the A-repeat domain, whose deletion prevents transcriptional silencing without affecting Xist localization across the X-chromosome⁶⁰.

There are numerous additional examples of lncRNAs that participate in the regulation of various genes. A classic example is the Air lncRNA, which is responsible for regulating the *Igfr2* gene to control genetic imprinting^{61–63}. In addition, HOTAIR affects the expression of genes in the *HoxD* cluster⁶⁴ among other genes throughout the genome^{33,65}. Recently, systematic studies exploring lncRNA function have shown that a large percentage of lncRNAs in the cell affect various gene expression programs^{29,30,66}, including those involved in embryonic development^{28–30}, cardiac function^{31,32}, immune responses^{67,68}, and cancer^{33–37}. Based on these gene expression studies, various regulatory strategies have been proposed for lncRNAs, including the activation^{47,69} and repression^{34,52,64} of genes in *cis*^{47,52} and in *trans*^{33,64}. Yet, whether lncRNAs directly or indirectly regulate these target genes remains unknown.

lncRNAs can recruit chromatin regulatory proteins to genomic DNA targets

Insights into how lncRNAs can regulate gene expression initially came from studies of Xist. Specifically, female embryos containing a deletion of a component of the Polycomb Repressive Complex 2 (PRC2), which places repressive histone modifications on chromatin⁷⁰, failed to maintain proper XCI⁷¹. It was subsequently shown that the PRC2 complex was recruited to the entire Xi³⁸ and that the timing of PRC2 recruitment tightly coincides with the induction of Xist during development^{52,72,73}. Importantly, deleting a discrete region of the Xist lncRNA, the B-F-repeat domain, causes a loss of PRC2 recruitment to the Xi without impacting transcriptional silencing or Xist localization across the Xi during the induction of XCI⁷⁴. Yet, there are still many open questions about Xist-mediated PRC2 recruitment. First, whether Xist physically interacts with the PRC2 complex^{75,76} or indirectly recruits PRC2^{74,77,78} is still debated⁷⁸ (see Box 1). Second, how Xist silences transcription and what role PRC2 may play during the induction of XCI is unclear⁷⁸ since PRC2 recruitment does not appear to be required for transcriptional silencing on the Xi⁷⁴. Specifically, Xist mutants that disrupt the ability to recruit PRC2 (B–F repeat mutants) can still silence transcription⁷⁴, mutants that fail to silence transcription (A-repeat mutants) can still recruit PRC2 across the Xi^{74,79}, and Xist can induce transcriptional silencing within cells containing a genetic deletion of PRC2^{80,81}. Despite these open questions, it is clear that Xist is required to recruit the PRC2 complex across the X-chromosome^{73,74,79}.

This chromatin protein recruitment model may be more general beyond Xist and has been proposed for several other lncRNAs. For example, HOTAIR is thought to physically interact with PRC2, and loss-of-function of HOTAIR leads to a reduction of the PRC2-dependent H3K27me3 repressive modifications across the *HoxD* gene cluster^{33,64}, suggesting that HOTAIR recruits PRC2 to these genes and may be involved in silencing their expression. Another example is HOTTIP, which is thought to physically interact with the WDR5 protein and whose loss-of-function leads to a reduction of its associated H3K4me3 active histone modifications on chromatin across the *HoxA* gene cluster⁴⁷, suggesting that HOTTIP recruits WDR5 to these genes and may be involved in activating their expression. More generally, a large percentage of lncRNAs are thought to physically interact with various chromatin regulatory proteins including PRC2^{30,38,82,83}, WDR5^{47,84}, and other readers^{30,35,83,85}, writers^{30,35,63,86}, and erasers^{30,87} of chromatin modifications (Figure 1B).

These examples highlight how lncRNAs may both activate and repress gene expression through a common chromatin-centric recruitment mechanism (Figure 1A).

Recently, it has been suggested that the PRC2 complex may interact with all RNAs in the cell – including lncRNAs and mRNAs^{88,89}. There is considerable debate about how many of the identified interactions between lncRNAs and chromatin proteins are specific^{76,88,89} and whether these physical interactions occur through direct RNA-protein or indirect protein-protein contacts^{78,90} (see Box 1). Yet, it is increasingly clear that at least some of these lncRNA-chromatin interactions are important for lncRNA- and chromatin-mediated gene regulation. For example, mutating the RNA binding domain of WDR5 eliminates its chromatin modification and gene regulatory activities at its target sites without impacting its catalytic activity⁸⁴. More generally, RNAi-mediated loss-of-function of several lncRNAs impact the same genes as those impacted by loss-of-function of their associated chromatin regulatory proteins^{30,38}.

Together, these results suggest that many lncRNAs may recruit chromatin regulatory complexes to specific targets on genomic DNA to control gene expression (Figure 1A).

lncRNAs may scaffold multiple chromatin proteins to coordinate discrete functions

Individual lncRNAs may interact with multiple chromatin proteins simultaneously to coordinate multiple functional roles that are required to properly regulate gene expression (Figure 1B). For example, HOTAIR is thought to interact with both the PRC2 histone methyltransferase and the LSD1 histone demethylase complex⁸⁷. This interaction may be important for coordinating the removal of activating marks (LSD1) and the addition of repressive marks (PRC2) on chromatin. More generally, over 30 of the lncRNAs expressed in embryonic stem cells (ESCs) are thought to interact simultaneously with multiple chromatin regulatory complexes that can read, write, and erase functionally related chromatin marks³⁰. Indeed, some of these ESC lncRNAs can interact with the Jarid1c histone demethylase complex, the PRC2 histone methyltransferase complex, and the ESET histone methyltransferase complex³⁰. This interaction may be important for coordinating the removal of activating marks (Jarid1c) and the addition of different repressive marks on chromatin (PRC2, ESET) (Figure 1C). Importantly, many of these chromatin proteins have been shown to co-localize at specific sets of genes in ESCs even though these proteins are not thought to directly interact with each other^{91–93}.

Furthermore, the Xist lncRNA is capable of coordinating at least three discrete functions to carry out its role in XCI. These functions are mediated by distinct genetic domains of the lncRNA that are required for silencing transcription (A-repeat)⁶⁰, recruitment of PRC2 (B-F-repeat)⁷⁴, and localization to chromatin (C-repeat)^{94,95} – all of which are required for proper XCI. Despite these clear genetic roles, the exact molecular mechanisms by which Xist coordinates these functions remains unclear because the proteins that directly interact with Xist are still largely unknown.

Together, these results suggest that some lncRNAs may create unique assemblies of chromatin regulatory complexes and other protein complexes that do not normally form protein-protein interactions (Figure 1B). By acting as a scaffold for regulatory proteins,

lncRNAs may coordinate the regulation of gene expression by recruiting a set of proteins that are required in combination for the shared regulation of a specific set of target genes (Figure 1C).

Mechanisms of lncRNA recruitment to genomic DNA

lncRNAs can recognize specific genomic DNA sites through diverse mechanisms

Three general mechanisms have been proposed for how lncRNAs that recruit protein complexes to genomic DNA can recognize specific target sites (Figure 1A). (i) RNA polymerase can tether a lncRNA to its site of transcription and, from this location, a lncRNA can act on its neighboring genes. This mechanism may explain the localization of the Neat1 lncRNA, which requires transcription to act even when large amounts of non-nascent mature RNA is present⁴². (ii) lncRNAs can interact with DNA through direct nucleic-acid hybridization. This can include traditional base-pairing interactions, which explains the specificity of the telomerase RNA component for telomeric DNA repeats⁹⁶.

Additionally, lncRNAs can interact with DNA through triplex-mediated interactions, which may explain the localization of specific ncRNAs to ribosomal DNA promoters⁹⁷. (iii) lncRNAs can physically interact with DNA binding proteins. Indeed the localization of *Drosophila* roX lncRNAs are dependent on their interaction with the CLAMP DNA binding protein to recognize specific DNA binding sites^{65,98–100}. This mechanism may also explain the localization of Xist and Firre; both lncRNAs are thought to interact with the hnRNPU/SAF-A DNA binding protein, which is required for their localization to DNA^{45,101}.

However, these mechanisms—polymerase tethering, hybridization, and DNA-binding protein mediated recruitment—alone may not be sufficient to explain how a lncRNA localizes to specific sites. For example, the roX lncRNAs localize through their interaction with CLAMP, but they do not interact at all sites throughout the genome where CLAMP is localized^{98,102}. Similarly, both Xist and Firre interact with hnRNPU/SAF-A^{45,101}, yet each localize to very different genomic DNA sites. This argues that specificity may not depend on a single factor, but may involve multiple independent factors, including those described above, that together provide localization specificity. Despite these examples, it remains largely unknown how most lncRNAs recognize and localize to genomic DNA.

lncRNAs can exploit the three-dimensional conformation of the nucleus to search for targets

Recent results are pointing to a potentially general mechanism by which lncRNAs search for regulatory targets by exploiting the three-dimensional conformation of the nucleus. For example, Xist utilizes three-dimensional nuclear organization to locate DNA target sites by first localizing to genomic sites that are in close spatial proximity to its own transcription locus^{44,103}. Moving Xist to a different genomic location leads to its relocalization to new genomic target sites that are defined by their close spatial proximity to the new Xist integration site⁴⁴. Other lncRNAs have also been shown to use spatial proximity to identify target sites^{37,98,104}. HOTTIP localizes across the HoxA cluster, which is in close spatial proximity to its own transcription locus⁴⁷.

This interplay between proximity-guided search and lncRNA localization is not restricted to interactions that occur on the same chromosome, but can also occur across chromosomes because regions that are present on different chromosomes can be in close spatial proximity in the nucleus¹⁰⁵ (Figure 2B). Indeed, the CISTR-ACT lncRNA localizes to sites present on the same chromosome as well as to sites on different chromosomes that are in close spatial proximity to its transcription locus⁴⁶. This proximity-guided model may also explain the localization of HOTAIR, the first example of a *trans* regulatory lncRNA. HOTAIR is transcribed from the HoxC locus and regulates the expression of genes in the HoxD locus, which is present on a different chromosome⁶⁴. Indeed, the Hox gene loci, despite being present on different chromosomes, often interact with each other in close spatial proximity within the nucleus^{13,106,107}. Such a proximity-guided search model may explain the apparent observations of both *cis* and *trans* mediated regulatory mechanisms of various lncRNAs and may suggest that these apparently divergent mechanisms share a common principle of proximity within the nucleus (Figure 2).

This proximity-guided search model exploits a feature that is unique for RNA, relative to proteins, which is its ability to function immediately upon transcription. In this model, the local concentration of a lncRNA depends primarily on its spatial distance from its transcription locus, such that sites that are close will have high concentration and sites that are far will have low concentration. Yet, proximity alone is not sufficient to explain interaction, because mRNAs are also present at high concentration, but do not act, in spatial proximity to their transcription locus. Similarly, the Firre lncRNA interacts with specific DNA sites that are in spatial proximity to the Firre locus, but does not interact with all sites in spatial proximity⁴⁵. Instead, other mechanisms, such as tethering, hybridization, or DNA binding interactions, are likely to be required for proper localization of the lncRNA to specific sites. Indeed, the roX lncRNAs interact with specific DNA sites, defined by the presence of CLAMP DNA elements, only when these sites are present in close spatial proximity^{98,104}. These two components – proximity and sequence specificity – may explain the localization of many lncRNAs. Specifically, a lncRNA will have a high probability of interacting with a target site within a region of high concentration, but it will have a low probability of interacting with a target site within a region of low concentration – even if it has a high affinity for that site (Figure 2C). Importantly, such a strategy might explain how lncRNAs, which are generally of lower abundance, could reliably identify their target genes by searching in close spatial proximity to their transcription locus rather than searching across the entire nucleus.

lncRNAs are essential for the establishment and maintenance of nuclear domains

Recent studies have highlighted another link between lncRNAs and nuclear organization – that is, several lncRNAs have been shown to play a critical role in bringing together DNA, RNA, and proteins to actively shape some aspects of three-dimensional nuclear organization. We discuss examples of lncRNAs that establish nuclear domains across various levels of organization from nuclear bodies to enhancer-promoter interactions below (Figure 3).

Nuclear bodies: Neat1 establishes the paraspeckle

The paraspeckle consists of various RNAs and proteins that are spatially co-localized and is thought to be the site of nuclear retention of adenosine-to-inosine edited mRNAs^{8,9}. Recent studies have demonstrated an essential role for the Neat1 lncRNA in forming paraspeckles^{42,43,108}. Specifically, Neat1 has been found to localize within the paraspeckle^{43,108,109}, and its loss of function leads to a loss of the paraspeckle domain^{43,108}. Conversely, induction of Neat1 is sufficient to establish the paraspeckle domain⁴². Furthermore, recruitment of Neat1 to a transgenic site is sufficient to create paraspeckles at that location^{42,110}. Indeed, synthetically tethering Neat1 to a genomic DNA region is sufficient to form paraspeckles, but tethering the paraspeckle-associated proteins, such as PSP1 to DNA is not sufficient to assemble paraspeckles¹¹⁰. Neat1 transcription is required for establishing and maintaining paraspeckles by recruiting paraspeckle-associated proteins to the *Neat1* genomic locus⁴². Accordingly, disruption of Neat1 transcription, even without a reduction in overall Neat1 levels, leads to the loss of paraspeckles⁴².

Together, these studies demonstrate that Neat1 plays an architectural role in the establishment and maintenance of the paraspeckle nuclear domains by seeding at its transcription locus and recruiting associated proteins to create an RNA-protein nuclear compartment.

Intrachromosomal regulatory domains: Xist compacts the X-chromosome

During XCI, the inactive X-chromosome is compacted and relocated to the periphery of the nucleus to form an intrachromosomal domain, termed the Barr body⁵². This three-dimensional restructuring of the Xi is carried out by the Xist lncRNA^{52,111}. Indeed, integrating Xist into transgenic locations, including on autosomes, is sufficient to silence, compact, and reposition the chromosome on which Xist is integrated^{59,112}. Xist spreads from its transcription locus to initial sites that are in close spatial proximity⁴⁴. From these sites, Xist then spreads across the entire X-chromosome. This spreading process is known to involve significant changes to chromosome architecture across the X-chromosome^{52,111}. These structural changes depend on the A-repeat domain of Xist, the same domain required for silencing transcription, because deletion of the A-repeat leads to the exclusion of actively transcribed regions from the silenced X-chromosome territory^{44,51}.

Together, these studies demonstrate that Xist is necessary for restructuring genomic DNA regions to establish an RNA-mediated silenced nuclear compartment. Xist performs this function by spreading across the X-chromosome and repositioning genes into the silenced Xist compartment⁴⁴.

Interchromosomal regulatory domains: Firre forms a trans-chromosomal compartment

Multiple genes that are present on different chromosomes can often localize within shared regions of the nucleus. These interchromosomal nuclear domains are often defined by the presence of genes with shared functional roles or regulation by common factors^{105,113,114}. Recently, a lncRNA, termed Firre was identified based on its role in adipogenesis⁶⁶. This lncRNA was shown to localize within a single nuclear domain containing many genes previously implicated in energy metabolism^{45,66}. This single nuclear domain includes the

Firre transcription locus on the X-chromosome as well as at least 5 genes that are located on different chromosomes including chromosomes 2, 9, 15, and 17⁴⁵. Importantly, deletion of the *Firre* locus results in reduced co-localization of the *trans*-chromosomal contacts within this nuclear domain⁴⁵. Random integration of Firre into different chromosomal regions leads to the emergence of new nuclear foci, suggesting that Firre may be sufficient to create a nuclear compartment at its integration sites⁴⁵. Taken together, these results suggest that Firre is required to maintain and may even be required to establish the formation of a *trans*-chromosomal nuclear compartment containing target genes of shared function.

Enhancer-Promoter interactions: eRNAs can promote chromosomal looping

Gene regulation involves physical interactions between distal enhancer regions and the promoters that they regulate. Recent studies have shown that several active enhancer regions can be transcribed to produce eRNAs. Several eRNAs have been proposed to play a role in mediating chromosomal interactions between an enhancer region and its associated promoter^{35,115,116,37,48,115,117}. For example, estrogen-induced¹¹⁶ and androgen-induced eRNAs³⁵ have been shown to maintain DNA looping between enhancer and promoter regions and, through this interaction, promote gene activation of estrogen-responsive genes and androgen-receptor-activated genes, respectively. Together, these results suggest that some eRNAs are required to maintain the three-dimensional chromosomal looping between an enhancer and its associated promoter.

While much is still unknown about how these eRNAs work, initial insights are emerging from two specific eRNAs that are highly expressed in prostate cancer³⁵. The PRNCR1 eRNA binds to the enhancer regions of androgen-receptor regulated genes and is thought to recruit the DOT1L histone methyltransferase to the enhancer. This chromatin protein recruitment in turn recruits a second lncRNA, PCGEM1, to the same region. The PCGEM1 lncRNA is thought to interact with Pygo2, an H3K4me3 reader that can recognize methylation groups on active promoter regions³⁵. Through the recruitment of these proteins, these eRNAs appear to facilitate looping between the enhancer and promoter regions, leading to the subsequent activation of the target gene³⁵. These results suggest that eRNAs may recruit chromatin regulatory proteins to create high affinity interactions between different regions of DNA and, through this, act to reposition enhancer and promoter regions into close spatial proximity.

Collectively, these results and others⁴¹ demonstrate that several lncRNAs play an important role in establishing and maintaining higher-order nuclear structures across various levels of nuclear organization from nuclear bodies to enhancer-promoter interactions.

A proposed model for lncRNA-mediated organization of nuclear structure

While there is some evidence that lncRNAs can recruit chromatin regulators, modify chromatin structure, regulate gene expression, search in spatial proximity, and reposition genes into a nuclear domain, how these mechanisms work together to create a dynamic nuclear compartment remains unclear. Important insights can be derived from studies of nuclear body formation, which depends on many molecules – including RNA, DNA, and protein – coming together into a single nuclear region^{4,118,119}. This process requires the

localization of an initial nucleating factor, which seeds organization and recruits other factors to this location^{4,119}. For example, tethering individual RNAs or proteins that are present in the Cajal bodies to a random location in the genome is sufficient to seed the formation of a new Cajal body at that site^{110,120}. In the context of well-studied nuclear bodies, such as the Cajal body and the nucleolus, the proteins involved have domains that allow them to self-interact, thereby creating preferential interactions between molecules of the same identity^{118,119,121,122}. This self-organization creates a high local concentration of a defined set of molecules within a spatially confined region around the initial nucleating factor¹¹⁹.

This process may also explain the assembly of other functional nuclear structures; DNA containing common chromatin modification patterns¹⁶ or DNA that is bound by shared proteins, such as PRC2^{13–15} or various transcription factors^{10–12}, can cluster together in three-dimensional proximity. While the exact mechanism that leads to the formation of these particular long-range interactions is largely unclear¹²³, it appears that a similar self-organization property may be involved because molecules of shared identity preferentially interact in three-dimensional proximity^{124,125}.

Based on the studies discussed above, we propose a model for how lncRNAs may organize nuclear architecture (Figure 4). This model is an extension of those originally proposed for Neat1⁴² and Xist⁴⁴. In this model, lncRNAs can seed organization by creating domains of high local lncRNA concentration near their site of transcription. This would allow the lncRNAs to scaffold various protein complexes and thereby nucleate lncRNA-protein complex assembly, increasing the effective concentration of proteins within this domain^{4,42}. lncRNAs can then interact with high affinity target sites to achieve specificity and recruit lncRNA-protein complexes to specific target sites. At these targets, lncRNA-protein complexes may modify the chromatin state and, through this, may act to reposition DNA sites into new nuclear domains of shared chromatin modification or protein occupancy⁴⁴. Importantly, whether chromatin modifications or other mechanisms, such as self-organization based on the recruitment of shared protein complexes, are what drive repositioning remains to be tested. This proposed model may not be restricted to the formation of DNA compartments, but may also explain the spatial assembly of RNA and protein domains in the nucleus through a similar lncRNA-centric mechanism^{42,110,122}.

This process of lncRNA spreading and repositioning may involve iterative steps by which the lncRNA, while actively transcribed, can continue to seed, nucleate, modify, and reposition genes into an expanding nuclear domain⁴⁴. For example, Xist spreads to new sites on the X-chromosome by interacting in spatial proximity with the genes that have not yet been silenced and then repositioning these genes into the growing silenced nuclear compartment⁴⁴. Once established, the lncRNA may maintain this domain through continued transcription from a location in close spatial proximity to the newly formed compartment, similar to how Neat1 is required to maintain paraspeckles through an ongoing process of transcription⁴².

Possible implications of lncRNA-mediated nuclear organization in gene regulation

It is increasingly clear that there are functional nuclear domains that contain shared chromatin modification patterns¹⁶ or protein occupancy^{2,10,12,14}. However, not all DNA that is modified or bound by a specific protein in the nucleus is spatially localized within a single nuclear domain^{13,16}. We hypothesize that different lncRNAs may establish these specific nuclear domains by scaffolding and recruiting distinct combinations of proteins. (Figure 5A). For example, the nucleus contains multiple discrete functional domains that are enriched for polycomb protein occupancy (polycomb bodies)^{14,15}; one such domain is the inactive X-chromosome⁷⁹, which is established by a specific lncRNA and is spatially distinct from other polycomb-enriched domains in the nucleus.

While nuclear organization is known to be highly dynamic between cell states, how this organization is dynamically established during various processes, such as cellular differentiation, remains unclear^{11,106,126,127}. We hypothesize that some lncRNAs might act as “organizational centers” to establish cell-type specific nuclear domains that organize genes of similar function in close three-dimensional proximity. Such a role is consistent with the observation that lncRNAs exhibit extraordinary cell-type specificity^{26,27,128}, in contrast to proteins, which are often reused in multiple cellular contexts¹²⁹. In this model, nuclear compartments can be dynamically organized simply through the activation or repression of a single lncRNA gene (Figure 5B).

This hypothesized role of lncRNAs as organizational centers might represent an ideal strategy for how nuclear-localized lncRNAs could act to regulate gene expression. Because lncRNAs are generally expressed at low abundance, the probability of coordinately finding multiple target genes that are distributed throughout the nucleus would be low, potentially leading to heterogeneous expression of these genes. There are two theoretical solutions: increase lncRNA abundance or cluster target genes in spatial proximity. While both approaches solve the challenge of finding distributed genes, increasing the levels of a lncRNA may not be an optimal solution because this may lead to sub-saturation of a lncRNA scaffold with its associated regulatory proteins (Figure 1B). Therefore, lncRNA-regulation of multiple distributed genes requires a tradeoff between the optimality of finding all genes (high lncRNA expression) with the optimality of interacting with all required regulatory proteins (low lncRNA expression). Spatial clustering would provide an ideal solution because it would enable a lncRNA to easily find all of its targets based on spatial proximity, where the lncRNA is in high local concentration, while ensuring saturation of the lncRNA regulatory complexes to coordinately regulate all of its target genes.

Concluding remarks

While the role of lncRNAs in establishing nuclear organization is attractive, many questions remain. Currently, there are only a few examples of lncRNAs that organize nuclear domains and even for these few lncRNAs, how they organize these nuclear domains is largely unknown. Future studies will be required to determine whether this role may be a more general role for nuclear-retained lncRNAs and whether there may be general mechanistic

principles by which lncRNAs act to shape nuclear domains. In particular, it will be important to identify additional lncRNA-mediated nuclear domains and characterize the dynamics of their formation across various cellular conditions. Such examples will allow us to dissect the precise mechanisms by which lncRNAs can organize nuclear domains and determine the various components required for domain assembly. To address these questions, it will be important to develop experimental systems, such as inducible lncRNA systems that enable precisely controlled formation of the associated nuclear domain, to dissect dynamic nuclear organization at the molecular level. Such experimental systems will enable the systematic perturbation of a lncRNA, including deletion of specific protein binding regions, and the measurement of their roles in the establishment and maintenance of nuclear domains. Finally, it will be essential to determine the role that lncRNA-mediated regulation of nuclear organization plays in the control of gene expression. While much work remains to be done, it is now clear that the roles of lncRNAs in regulating gene expression and establishing nuclear organization may be more tightly linked than previously appreciated.

Acknowledgments

We thank Mario Blanco, Jesse Engreitz, Manuel Garber, Shari Grossman, Colleen McHugh, and Klara Stefflova for helpful discussions, critical reading, and suggestions as well as Howard Chang for discussions about roX localization, John Rinn for discussions about proximity mechanisms, Kathrin Plath for discussions about Xist and nuclear organization, and Sigrid Knemeyer for illustrations and assistance with figures. SQ is a Howard Hughes Medical Institute Gilliam Fellow. Work in the Guttman lab is funded by an NIH Director's Early Independence Award, the Searle Scholars program, the Edward Mallinckrodt Jr. Foundation, and the Sidney Kimmel Foundation.

References

1. Gibcus JH, Dekker J. The hierarchy of the 3D genome. *Mol Cell*. 2013; 49:773–782. [PubMed: 23473598]
2. Cavalli G, Misteli T. Functional implications of genome topology. *Nat Struct Mol Biol*. 2013; 20:290–299. [PubMed: 23463314]
3. Misteli T. Beyond the sequence: cellular organization of genome function. *Cell*. 2007; 128:787–800. [PubMed: 17320514]
4. Mao YS, Zhang B, Spector DL. Biogenesis and function of nuclear bodies. *Trends Genet*. 2011; 27:295–306. [PubMed: 21680045]
5. McStay B, Grummt I. The epigenetics of rRNA genes: from molecular to chromosome biology. *Annu Rev Cell Dev Biol*. 2008; 24:131–157. [PubMed: 18616426]
6. Melese T, Xue Z. The nucleolus: an organelle formed by the act of building a ribosome. *Curr Opin Cell Biol*. 1995; 7:319–324. [PubMed: 7662360]
7. Lamond AI, Spector DL. Nuclear speckles: a model for nuclear organelles. *Nat Rev Mol Cell Biol*. 2003; 4:605–612. [PubMed: 12923522]
8. Fox AH, Lamond AI. Paraspeckles. *Cold Spring Harb Perspect Biol*. 2010; 2:a000687. [PubMed: 20573717]
9. Fox AH, et al. Paraspeckles: a novel nuclear domain. *Curr Biol*. 2002; 12:13–25. [PubMed: 11790299]
10. Schoenfelder S, et al. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat Genet*. 2010; 42:53–61. [PubMed: 20010836]
11. Wei Z, et al. Klf4 organizes long-range chromosomal interactions with the oct4 locus in reprogramming and pluripotency. *Cell Stem Cell*. 2013; 13:36–47. [PubMed: 23747203]
12. Osborne CS, et al. Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. *PLoS Biol*. 2007; 5:e192. [PubMed: 17622196]

13. Bantignies F, et al. Polycomb-dependent regulatory contacts between distant Hox loci in *Drosophila*. *Cell*. 2011; 144:214–226. [PubMed: 21241892]
14. Cheutin T, Cavalli G. Polycomb silencing: from linear chromatin domains to 3D chromosome folding. *Curr Opin Genet Dev*. 2014; 25C:30–37. [PubMed: 24434548]
15. Delest A, Sexton T, Cavalli G. Polycomb: a paradigm for genome organization from one to three dimensions. *Curr Opin Cell Biol*. 2012; 24:405–414. [PubMed: 22336329]
16. Sexton T, et al. Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell*. 2012; 148:458–472. [PubMed: 22265598]
17. Smallwood A, Ren B. Genome organization and long-range regulation of gene expression by enhancers. *Curr Opin Cell Biol*. 2013; 25:387–394. [PubMed: 23465541]
18. Herman RC, Williams JG, Penman S. Message and non-message sequences adjacent to poly(A) in steady state heterogeneous nuclear RNA of HeLa cells. *Cell*. 1976; 7:429–437. [PubMed: 59633]
19. Perry RP, Kelley DE, LaTorre J. Synthesis and turnover of nuclear and cytoplasmic polyadenylic acid in mouse L cells. *J Mol Biol*. 1974; 82:315–331. [PubMed: 4856346]
20. Brawerman G, Diez J. Metabolism of the polyadenylate sequence of nuclear RNA and messenger RNA in mammalian cells. *Cell*. 1975; 5:271–280. [PubMed: 1170955]
21. Huang S, Deerinck TJ, Ellisman MH, Spector DL. In vivo analysis of the stability and transport of nuclear poly(A)⁺ RNA. *J Cell Biol*. 1994; 126:877–899. [PubMed: 7519622]
22. He DC, Nickerson JA, Penman S. Core filaments of the nuclear matrix. *J Cell Biol*. 1990; 110:569–580. [PubMed: 2307700]
23. Nickerson JA, Krochmalnic G, Wan KM, Penman S. Chromatin architecture and nuclear RNA. *Proc Natl Acad Sci U S A*. 1989; 86:177–181. [PubMed: 2911567]
24. Guttman M, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*. 2009; 458:223–227. [PubMed: 19182780]
25. Guttman M, Rinn JL. Modular regulatory principles of large non-coding RNAs. *Nature*. 2012; 482:339–346. [PubMed: 22337053]
26. Cabili MN, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev*. 2011; 25:1915–1927. [PubMed: 21890647]
27. Derrien T, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res*. 2012; 22:1775–1789. [PubMed: 22955988]
28. Sauvageau M, et al. Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *Elife*. 2013; 2:e01749. [PubMed: 24381249]
29. Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell*. 2011; 147:1537–1550. [PubMed: 22196729]
30. Guttman M, et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature*. 2011; 477:295–300. [PubMed: 21874018]
31. Klattenhoff CA, et al. Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell*. 2013; 152:570–583. [PubMed: 23352431]
32. Han P, et al. A long noncoding RNA protects the heart from pathological hypertrophy. *Nature*. 2014
33. Gupta RA, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*. 2010; 464:1071–1076. [PubMed: 20393566]
34. Huarte M, et al. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell*. 2010; 142:409–419. [PubMed: 20673990]
35. Yang L, et al. lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. *Nature*. 2013; 500:598–602. [PubMed: 23945587]
36. Gutschner T, et al. The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res*. 2013; 73:1180–1189. [PubMed: 23243023]
37. Trimarchi T, et al. Genome-wide Mapping and Characterization of Notch-Regulated Long Noncoding RNAs in Acute Leukemia. *Cell*. 2014; 158:593–606. [PubMed: 25083870]

38. Khalil AM, et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci U S A*. 2009; 106:11667–11672. [PubMed: 19571010]
39. Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. *Mol Cell*. 2011; 43:904–914. [PubMed: 21925379]
40. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. *Annu Rev Biochem*. 2012; 81:145–166. [PubMed: 22663078]
41. Bergmann JH, Spector DL. Long non-coding RNAs: modulators of nuclear structure and function. *Curr Opin Cell Biol*. 2014; 26:10–18. [PubMed: 24529241]
42. Mao YS, Sunwoo H, Zhang B, Spector DL. Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. *Nat Cell Biol*. 2011; 13:95–101. [PubMed: 21170033]
43. Clemson CM, et al. An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol Cell*. 2009; 33:717–726. [PubMed: 19217333]
44. Engreitz JM, et al. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science*. 2013; 341:1237973. [PubMed: 23828888]
45. Hacisuleyman E, et al. Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. *Nat Struct Mol Biol*. 2014; 21:198–206. [PubMed: 24463464]
46. Maass PG, et al. A misplaced lncRNA causes brachydactyly in humans. *J Clin Invest*. 2012; 122:3990–4002. [PubMed: 23093776]
47. Wang KC, et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature*. 2011; 472:120–124. [PubMed: 21423168]
48. Lam MT, Li W, Rosenfeld MG, Glass CK. Enhancer RNAs and regulated transcriptional programs. *Trends Biochem Sci*. 2014; 39:170–182. [PubMed: 24674738]
49. Shichino Y, Yamashita A, Yamamoto M. Meiotic long non-coding meiRNA accumulates as a dot at its genetic locus facilitated by Mmi1 and plays as a decoy to lure Mmi1. *Open Biol*. 2014; 4:140022. [PubMed: 24920274]
50. Shimada T, Yamashita A, Yamamoto M. The fission yeast meiotic regulator Mei2p forms a dot structure in the horse-tail nucleus in association with the sme2 locus on chromosome II. *Mol Biol Cell*. 2003; 14:2461–2469. [PubMed: 12808043]
51. Chaumeil J, Le Baccon P, Wutz A, Heard E. A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes & development*. 2006; 20:2223–2237. [PubMed: 16912274]
52. Plath K, Mlynarczyk-Evans S, Nusinow DA, Panning B. Xist RNA and the mechanism of X chromosome inactivation. *Annual review of genetics*. 2002; 36:233–278.
53. Brown CJ, et al. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature*. 1991; 349:38–44. [PubMed: 1985261]
54. Brockdorff N, et al. Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. *Nature*. 1991; 351:329–331. [PubMed: 2034279]
55. Brockdorff N, et al. The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell*. 1992; 71:515–526. [PubMed: 1423610]
56. Clemson CM, McNeil JA, Willard HF, Lawrence JB. XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. *The Journal of cell biology*. 1996; 132:259–275. [PubMed: 8636206]
57. Penny GD, Kay GF, Sheardown SA, Rastan S, Brockdorff N. Requirement for Xist in X chromosome inactivation. *Nature*. 1996; 379:131–137. [PubMed: 8538762]
58. Rasmussen TP, Wutz AP, Pehrson JR, Jaenisch RR. Expression of Xist RNA is sufficient to initiate macrochromatin body formation. *Chromosoma*. 2001; 110:411–420. [PubMed: 11734999]
59. Wutz A, Jaenisch R. A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Molecular cell*. 2000; 5:695–705. [PubMed: 10882105]
60. Wutz A, Rasmussen TP, Jaenisch R. Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nature genetics*. 2002; 30:167–174. [PubMed: 11780141]

61. Kulinski TM, Barlow DP, Hudson QJ. Imprinted silencing is extended over broad chromosomal domains in mouse extra-embryonic lineages. *Curr Opin Cell Biol.* 2013; 25:297–304. [PubMed: 23478214]
62. Yamasaki Y, et al. Neuron-specific relaxation of Igf2r imprinting is associated with neuron-specific histone modifications and lack of its antisense transcript Air. *Hum Mol Genet.* 2005; 14:2511–2520. [PubMed: 16037066]
63. Nagano T, et al. The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science.* 2008; 322:1717–1720. [PubMed: 18988810]
64. Rinn JL, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell.* 2007; 129:1311–1323. [PubMed: 17604720]
65. Chu C, Qu K, Zhong FL, Artandi SE, Chang HY. Genomic Maps of Long Noncoding RNA Occupancy Reveal Principles of RNA-Chromatin Interactions. *Mol Cell.* 2011
66. Sun L, et al. Long noncoding RNAs regulate adipogenesis. *Proc Natl Acad Sci U S A.* 2013; 110:3387–3392. [PubMed: 23401553]
67. Carpenter S, et al. A long noncoding RNA mediates both activation and repression of immune response genes. *Science.* 2013; 341:789–792. [PubMed: 23907535]
68. Wang P, et al. The STAT3-binding long noncoding RNA Inc-DC controls human dendritic cell differentiation. *Science.* 2014; 344:310–313. [PubMed: 24744378]
69. Orom UA, et al. Long noncoding RNAs with enhancer-like function in human cells. *Cell.* 2010; 143:46–58. [PubMed: 20887892]
70. Margueron R, Reinberg D. The Polycomb complex PRC2 and its mark in life. *Nature.* 2011; 469:343–349. [PubMed: 21248841]
71. Wang J, et al. Imprinted X inactivation maintained by a mouse Polycomb group gene. *Nat Genet.* 2001; 28:371–375. [PubMed: 11479595]
72. Mak W, et al. Mitotically stable association of polycomb group proteins eed and enx1 with the inactive x chromosome in trophoblast stem cells. *Current biology : CB.* 2002; 12:1016–1020. [PubMed: 12123576]
73. Silva J, et al. Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. *Developmental cell.* 2003; 4:481–495. [PubMed: 12689588]
74. da Rocha ST, et al. Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome. *Mol Cell.* 2014; 53:301–316. [PubMed: 24462204]
75. Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science.* 2008; 322:750–756. [PubMed: 18974356]
76. Cifuentes-Rojas C, Hernandez AJ, Sarma K, Lee JT. Regulatory Interactions between RNA and Polycomb Repressive Complex 2. *Mol Cell.* 2014; 55:171–185. [PubMed: 24882207]
77. Cerase A, et al. Spatial separation of Xist RNA and polycomb proteins revealed by superresolution microscopy. *Proc Natl Acad Sci U S A.* 2014; 111:2235–2240. [PubMed: 24469834]
78. Brockdorff N. Noncoding RNA and Polycomb recruitment. *RNA.* 2013; 19:429–442. [PubMed: 23431328]
79. Plath K, et al. Role of histone H3 lysine 27 methylation in X inactivation. *Science.* 2003; 300:131–135. [PubMed: 12649488]
80. Schoeftner S, et al. Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *The EMBO journal.* 2006; 25:3110–3122. [PubMed: 16763550]
81. Kalantry S, Magnuson T. The Polycomb group protein EED is dispensable for the initiation of random X-chromosome inactivation. *PLoS Genet.* 2006; 2:e66. [PubMed: 16680199]
82. Zhao J, et al. Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol Cell.* 2010; 40:939–953. [PubMed: 21172659]
83. Kaneko S, et al. Interactions between JARID2 and noncoding RNAs regulate PRC2 recruitment to chromatin. *Mol Cell.* 2014; 53:290–300. [PubMed: 24374312]
84. Yang YW, et al. Essential role of lncRNA binding for WDR5 maintenance of active chromatin and embryonic stem cell pluripotency. *Elife.* 2014; 3:e02046. [PubMed: 24521543]

85. Yap KL, et al. Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol Cell*. 2010; 38:662–674. [PubMed: 20541999]
86. Yang L, et al. ncRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs. *Cell*. 2011; 147:773–788. [PubMed: 22078878]
87. Tsai MC, et al. Long noncoding RNA as modular scaffold of histone modification complexes. *Science*. 2010; 329:689–693. [PubMed: 20616235]
88. Kaneko S, Son J, Shen SS, Reinberg D, Bonasio R. PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. *Nat Struct Mol Biol*. 2013; 20:1258–1264. [PubMed: 24141703]
89. Davidovich C, Zheng L, Goodrich KJ, Cech TR. Promiscuous RNA binding by Polycomb repressive complex 2. *Nat Struct Mol Biol*. 2013; 20:1250–1257. [PubMed: 24077223]
90. McHugh CA, Russell P, Guttman M. Methods for comprehensive experimental identification of RNA-protein interactions. *Genome Biol*. 2014; 15:203. [PubMed: 24467948]
91. Bilodeau S, Kagey MH, Frampton GM, Rahl PB, Young RA. SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes Dev*. 2009; 23:2484–2489. [PubMed: 19884255]
92. Tahiliani M, et al. The histone H3K4 demethylase SMCX links REST target genes to X-linked mental retardation. *Nature*. 2007; 447:601–605. [PubMed: 17468742]
93. Cloos PA, Christensen J, Agger K, Helin K. Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. *Genes Dev*. 2008; 22:1115–1140. [PubMed: 18451103]
94. Beletskii A, Hong YK, Pehrson J, Egholm M, Strauss WM. PNA interference mapping demonstrates functional domains in the noncoding RNA Xist. *Proceedings of the National Academy of Sciences of the United States of America*. 2001; 98:9215–9220. [PubMed: 11481485]
95. Sarma K, Levasseur P, Aristarkhov A, Lee JT. Locked nucleic acids (LNAs) reveal sequence requirements and kinetics of Xist RNA localization to the X chromosome. *Proc Natl Acad Sci U S A*. 2010; 107:22196–22201. [PubMed: 21135235]
96. Harrington LA, Greider CW. Telomerase primer specificity and chromosome healing. *Nature*. 1991; 353:451–454. [PubMed: 1896088]
97. Schmitz KM, Mayer C, Postepska A, Grummt I. Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes Dev*. 2010; 24:2264–2269. [PubMed: 20952535]
98. Soruco MM, et al. The CLAMP protein links the MSL complex to the X chromosome during *Drosophila* dosage compensation. *Genes Dev*. 2013; 27:1551–1556. [PubMed: 23873939]
99. Simon MD, et al. The genomic binding sites of a noncoding RNA. *Proc Natl Acad Sci U S A*. 2011; 108:20497–20502. [PubMed: 22143764]
100. Wang CI, et al. Chromatin proteins captured by ChIP-mass spectrometry are linked to dosage compensation in *Drosophila*. *Nat Struct Mol Biol*. 2013; 20:202–209. [PubMed: 23295261]
101. Hasegawa Y, Brockdorff N, Kawano S, Tsutui K, Nakagawa S. The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. *Dev Cell*. 2010; 19:469–476. [PubMed: 20833368]
102. McElroy KA, Kang H, Kuroda MI. Are we there yet? Initial targeting of the Male-Specific Lethal and Polycomb group chromatin complexes in *Drosophila*. *Open Biol*. 2014; 4:140006. [PubMed: 24671948]
103. Simon MD, et al. High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. *Nature*. 2013; 504:465–469. [PubMed: 24162848]
104. Quinn JJ, et al. Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification. *Nat Biotechnol*. 2014
105. Williams A, Spilianakis CG, Flavell RA. Interchromosomal association and gene regulation in trans. *Trends Genet*. 2010; 26:188–197. [PubMed: 20236724]
106. Denholtz M, et al. Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. *Cell Stem Cell*. 2013; 13:602–616. [PubMed: 24035354]

107. Noordermeer D, et al. Temporal dynamics and developmental memory of 3D chromatin architecture at Hox gene loci. *Elife*. 2014; 3:e02557. [PubMed: 24843030]
108. Sunwoo H, et al. MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res*. 2009; 19:347–359. [PubMed: 19106332]
109. Chen LL, Carmichael GG. Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. *Mol Cell*. 2009; 35:467–478. [PubMed: 19716791]
110. Shevtsov SP, Dundr M. Nucleation of nuclear bodies by RNA. *Nat Cell Biol*. 2011; 13:167–173. [PubMed: 21240286]
111. Splinter E, et al. The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. *Genes & development*. 2011; 25:1371–1383. [PubMed: 21690198]
112. Lee JT, Jaenisch R. Long-range cis effects of ectopic X-inactivation centres on a mouse autosome. *Nature*. 1997; 386:275–279. [PubMed: 9069285]
113. Lancot C, Cheutin T, Cremer M, Cavalli G, Cremer T. Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat Rev Genet*. 2007; 8:104–115. [PubMed: 17230197]
114. Spilianakis CG, Lalioti MD, Town T, Lee GR, Flavell RA. Interchromosomal associations between alternatively expressed loci. *Nature*. 2005; 435:637–645. [PubMed: 15880101]
115. Lai F, et al. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature*. 2013; 494:497–501. [PubMed: 23417068]
116. Li W, et al. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature*. 2013; 498:516–520. [PubMed: 23728302]
117. Orom UA, Shiekhattar R. Long noncoding RNAs usher in a new era in the biology of enhancers. *Cell*. 2013; 154:1190–1193. [PubMed: 24034243]
118. Dundr M, Misteli T. Biogenesis of nuclear bodies. *Cold Spring Harb Perspect Biol*. 2010; 2:a000711. [PubMed: 21068152]
119. Misteli T. The concept of self-organization in cellular architecture. *J Cell Biol*. 2001; 155:181–185. [PubMed: 11604416]
120. Kaiser TE, Intine RV, Dundr M. De novo formation of a subnuclear body. *Science*. 2008; 322:1713–1717. [PubMed: 18948503]
121. Misteli T. Concepts in nuclear architecture. *Bioessays*. 2005; 27:477–487. [PubMed: 15832379]
122. Lewis JD, Tollervy D. Like attracts like: getting RNA processing together in the nucleus. *Science*. 2000; 288:1385–1389. [PubMed: 10827942]
123. Gonzalez I, Mateos-Langerak J, Thomas A, Cheutin T, Cavalli G. Identification of Regulators of the Three-Dimensional Polycomb Organization by a Microscopy-Based Genome-wide RNAi Screen. *Mol Cell*. 2014; 54:485–499. [PubMed: 24703951]
124. Schaaf CA, et al. Cohesin and polycomb proteins functionally interact to control transcription at silenced and active genes. *PLoS Genet*. 2013; 9:e1003560. [PubMed: 23818863]
125. Atchison ML. Function of YY1 in Long-Distance DNA Interactions. *Front Immunol*. 2014; 5:45. [PubMed: 24575094]
126. Zhang H, et al. Intrachromosomal looping is required for activation of endogenous pluripotency genes during reprogramming. *Cell Stem Cell*. 2013; 13:30–35. [PubMed: 23747202]
127. Phillips-Cremins JE, et al. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell*. 2013; 153:1281–1295. [PubMed: 23706625]
128. Guttman M, et al. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat Biotechnol*. 2010; 28:503–510. [PubMed: 20436462]
129. Ponten F, et al. A global view of protein expression in human cells, tissues, and organs. *Mol Syst Biol*. 2009; 5:337. [PubMed: 20029370]

130. Mili S, Steitz JA. Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *RNA*. 2004; 10:1692–1694. [PubMed: 15388877]
131. Darnell RB. HITS-CLIP: panoramic views of protein-RNA regulation in living cells. *Wiley Interdiscip Rev RNA*. 2010; 1:266–286. [PubMed: 21935890]
132. Ule J, Jensen K, Mele A, Darnell RB. CLIP: a method for identifying protein-RNA interaction sites in living cells. *Methods*. 2005; 37:376–386. [PubMed: 16314267]
133. Wang Z, Tollervey J, Briese M, Turner D, Ule J. CLIP: construction of cDNA libraries for high-throughput sequencing from RNAs cross-linked to proteins in vivo. *Methods*. 2009; 48:287–293. [PubMed: 19272451]
134. Brimacombe R, Stiege W, Kyriatsoulis A, Maly P. Intra-RNA and RNA-protein cross-linking techniques in *Escherichia coli* ribosomes. *Methods Enzymol*. 1988; 164:287–309. [PubMed: 3071669]
135. Singh G, Ricci EP, Moore MJ. RIPiT-Seq: a high-throughput approach for footprinting RNA:protein complexes. *Methods*. 2014; 65:320–332. [PubMed: 24096052]

Highlights

Many nuclear-retained lncRNAs can recruit chromatin regulatory proteins to genomic DNA to regulate gene expression

Several of these lncRNAs can search for regulatory targets by exploiting the three-dimensional conformation of the nucleus

Several lncRNAs are essential for the establishment and maintenance of three-dimensional nuclear domains

The role of lncRNAs in regulating gene expression and shaping nuclear organization may occur through some shared molecular mechanisms

Box 1: Experimental methods to define lncRNA-protein interactions

There are several common methods for purifying lncRNA-protein complexes including protein-based and RNA-based purification methods. For a more complete discussion of these methods and their strengths and limitations, see⁹⁰.

Briefly, most lncRNA-chromatin interactions^{34,38,82}, including the Xist-PRC2 interaction^{75,76}, have been identified using ‘native purification’ methods, which purify RNA-protein complexes under physiological conditions. The advantage of these methods is that they preserve the native complexes present in the cell. Yet, these methods also have several limitations including the potential for the identification of RNA-protein interactions that form in solution, which do not reflect interactions occurring in the cell^{130,131}. Because of these issues, there has been some debate about the biological significance of interactions detected by these methods^{76,78,89}, including the Xist-PRC2 interaction, with some arguing that they are non-specific⁷⁸.

One way to distinguish *in vivo* interactions from interactions that form subsequently in solution is by crosslinking RNA-protein complexes in the cell and purifying the complex under denaturing conditions¹³¹. Methods such as CLIP utilize UV crosslinking, which crosslinks directly interacting RNA and protein molecules, to purify complexes using high-stringency wash conditions followed by separation on a denaturing SDS-PAGE gel^{132,133}. A limitation of this method is that UV will not capture interactions that occur through a complex of multiple proteins¹³⁴. This has restricted its adoption for mapping many chromatin regulatory proteins because the precise protein within most chromatin-regulatory complexes that might directly interact with a lncRNA is unknown.

Other crosslinking methods such as formaldehyde, which crosslinks nucleic acid-protein as well as protein-protein interactions, can eliminate the need to know the exact interacting protein while enabling purification in high stringency conditions^{30,135}. Indeed, several studies have used this approach to map numerous chromatin regulatory proteins, including PRC2 and WDR5, and have identified a more specific set of interactions than previously identified by native purifications^{30,84}. Yet, adapting this formaldehyde approach to a denaturing strategy is challenging since a denaturing SDS-PAGE gel will no longer resolve the purified complex. Furthermore, because formaldehyde crosslinks across a larger physical distances than UV, many of the interactions identified by this method might not reflect physical interactions between a lncRNA and chromatin complex⁷⁸. For example, this approach will also identify chromatin proteins and lncRNAs that are in close proximity within a DNA locus; such proximity will likely occur for nascent transcripts and the many activating chromatin complexes bound near their transcription locus.

In the absence of the ability to define a lncRNA-protein interaction using direct crosslinking and denaturing conditions, it is unclear how to confidently define *in vivo* physical interactions using biochemical methods. In such cases, complimentary genetic methods are essential to demonstrate the functional importance of an identified lncRNA-protein interaction.

Figure 1. lncRNA-mediated regulation of gene expression through the recruitment of chromatin regulatory proteins

(a) Different cell types express distinct lncRNAs that can differentially recruit these same chromatin regulatory proteins, including the repressive PRC2 complex and the activating WDR5 chromatin modifying protein, to specific genes. Inset: lncRNAs can recruit these complexes by binding to target sites through three mechanisms: tethering to its nascent transcription locus (top panel), directly hybridizing to genomic targets (middle panel), or interacting with a DNA-binding protein (bottom panel). (b) Different lncRNAs can scaffold

unique assemblies of chromatin regulatory complexes. lncRNAs are generally expressed at lower levels relative their associated chromatin proteins (background). (c) lncRNAs may act to coordinate the regulation of gene expression at specific target locations. In this illustration, a lncRNA that can scaffold PRC2, Jarid1C, and ESET may act to remove H3K4me3 and place H3K27me3 and H3K9me2, thereby coordinating the repression of transcription.

Figure 2. IncRNAs can utilize a proximity-guided search to localize to target genes
(a) IncRNAs can regulate genes (green box) on its own chromosome (left panel). In the nucleus, this regulation can occur if the lncRNA locus is in close physical proximity to its target sites (middle panel). For instance, Xist localizes to genes across the X-chromosome (right panel). **(b)** IncRNAs can also regulate expression of genes on different chromosomes (blue box, left panel). In the nucleus, this can also occur when the lncRNA locus and its targets are in close proximity (middle panel). An example is Firre, which localizes to targets that are present across several chromosomes (right panel). **(c)** The concentration of a

lncRNA will be highest (dark red – inner circle) near its site of transcription and will decrease (light red – outer circles) the further the distance is from its site of transcription, creating a concentration gradient of lncRNA abundance (red cloud, intensity indicates average lncRNA concentration). This spatial gradient establishes a nuclear domain with a high lncRNA concentration, where they can interact with site-specific targets (dark blue arrows). Conversely, lncRNAs outside of the nuclear domain will have a lower probability of interacting with site-specific targets (light blue arrows) due to decreased lncRNA concentration.

Figure 3. lncRNAs can shape three-dimensional nuclear architecture across various levels of organization

(a) Actively transcribed Neat1 (red line) is required to establish the formation of the paraspeckle nuclear body (red cloud), which is an RNA-protein (gray) nuclear domain that is the site of nuclear retention of RNAs such as the CTN RNA (black). (b) Xist (red line) establishes an intrachromosomal nuclear domain (red cloud) by nucleating near its transcription site (white box) and spreading to DNA sites in spatial proximity to its locus. (c) Firre establishes an interchromosomal nuclear domain and brings together targets on

chromosomes 2, 15, and 17 into close physical proximity to its transcriptional locus on the X-chromosome. **(d)** Enhancer RNAs (eRNAs) maintain the interaction between enhancer and promoter regions and may do this by interacting with proteins that can modify chromatin.

Figure 4. A model for how lncRNAs can dynamically shape nuclear organization

The proposed steps involved in lncRNA mediated assembly of nuclear organization roughly based on the proposed models for the *Neat1*⁴² and *Xist*⁴⁴ lncRNAs. (i) Transcription of a lncRNA can seed the formation of a lncRNA nuclear domain. (ii) lncRNAs can bind to proteins in the nucleus (gray circles) to scaffold protein complexes. Formation of these complexes will nucleate the formation of a spatial compartment (red cloud, dashed lines) near the transcriptional locus of the lncRNA. (iii) lncRNAs can bind to specific DNA sites (white squares) to recruit lncRNA-protein complexes to target sites. (iv) By recruiting these

complexes to DNA, lncRNAs can guide chromatin modifications (blue histones), such as repressive histone modification (red marks). (v) Modified chromatin may be compressed and repositioned into a new nuclear region. (vi) As the lncRNA continues to be transcribed from its transcriptional locus, it may iteratively bind to DNA sites (green regions), modify target sites, and reposition DNA into the lncRNA nuclear domain. This continuous process may act to maintain the nuclear domain established by a lncRNA.

Figure 5. A hypothesis for how lncRNAs may act to assemble dynamic and specific nuclear domains

(a) Nuclear domains that share the same proteins can interact in different regions of the nucleus. Zoom-in panels: We hypothesize that different lncRNAs may act to distinguish between these domains by scaffolding and assembling distinct domains. (b) Through linear co-regulation, operons can simultaneously regulate sets of genes (A, B, C and D, E, F) with shared regulatory functions. Activators (pink triangles) and repressor (green boxes) control operon expression under a particular cell state. We hypothesize that through spatial co-

regulation, lncRNAs may nucleate the formation of nuclear domains to co-localize target genes upon induction of lncRNA expression. For instance, upon induction of lncRNA1, genes A, B, and C are co-regulated in a nuclear domain (red cloud, dashed lines). Under a different cell state, lncRNA1 expression is repressed, leading to the breakdown of the lncRNA1 nuclear domain and expression of lncRNA2 leads to formation of another nuclear domain (blue cloud, dashed lines) containing genes D, E, and F.