Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression

Jesse M. Engreitz[1,2], Noah Ollikainen[3]* and Mitchell Guttman[3]*

Abstract | Over the past decade, it has become clear that mammalian genomes encode thousands of long non-coding RNAs (lncRNAs), many of which are now implicated in diverse biological processes. Recent work studying the molecular mechanisms of several key examples — including Xist, which orchestrates X chromosome inactivation — has provided new insights into how lncRNAs can control cellular functions by acting in the nucleus. Here we discuss emerging mechanistic insights into how lncRNAs can regulate gene expression by coordinating regulatory proteins, localizing to target loci and shaping three-dimensional (3D) nuclear organization. We explore these principles to highlight biological challenges in gene regulation, in which lncRNAs are well-suited to perform roles that cannot be carried out by DNA elements or protein regulators alone, such as acting as spatial amplifiers of regulatory signals in the nucleus.

X chromosome inactivation (XCI). A process in early embryonic development whereby gene expression from one of the two X chromosomes in females is silenced to achieve balanced expression levels with X-linked genes in males.

[1]Broad Institute of Harvard and Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts 02142, USA.
[2]Harvard–MIT Division of Health Sciences and Technology, MIT, Cambridge, Massachusetts 02142, USA.
[3]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, California 91125, USA.
**These authors contributed equally to this work.*
Correspondence to M.G.
mguttman@caltech.edu

doi:10.1038/nrm.2016.126
Published online 26 Oct 2016

In the past decade, advances in genome sequencing and analysis[1–4] have led to the discovery of tens of thousands of RNA transcripts that have similar properties to mRNAs[5,6] but are not translated into proteins[7–9]. These transcripts are collectively referred to as long non-coding RNAs (lncRNAs), a term that is generally applied to any transcript that has a primary sequence that is longer than 200 nucleotides and lacks protein-coding potential[3,5,6]. Because of this broad definition, lncRNAs are heterogeneous in their biogenesis[10], stability[11], abundance[5,6] and evolution[12–15], and thus are also likely to be heterogeneous in their mechanisms of function. Indeed, whereas some lncRNAs act as functional RNA molecules, others seem to be non-functional byproducts of underlying *cis* regulatory elements such as enhancers[16–20]. Although it remains unclear precisely what proportion of lncRNAs are functional, dozens of lncRNAs are now known to act as regulators of gene expression programmes in diverse biological processes[21,22].

One of the best-studied examples of a lncRNA that regulates gene expression is Xist (X inactive specific transcript), which orchestrates X chromosome inactivation (XCI)[23,24]. Xist was initially identified because it is expressed from the inactive X chromosome (Xi) but not the active X (Xa)[25,26], and was subsequently shown to lack a conserved open reading frame[25,27], to localize in the nucleus[28] and to form a unique nuclear compartment that 'coats' the Xi[29]. Importantly, deletion of Xist leads to failure to initiate XCI[30,31], deletion of a single region of the Xist RNA ablates its silencing role[32], disruption of Xist localization on chromatin prevents silencing[33], and forced expression of Xist on X chromosomes or autosomes (in females or males) is sufficient to trigger gene silencing in *cis*[34,35]. These studies demonstrated that Xist acts as a functional RNA molecule that is required for transcriptional silencing of X chromosome genes.

Like Xist, other lncRNAs have been implicated in gene regulation (for recent reviews, see REFS 21,22). Many lncRNAs localize preferentially in the nucleus[5,6,36–38], and loss-of-function studies have suggested that lncRNAs can have broad effects on gene expression[39–42]. In this Review, we discuss emerging insights derived from Xist and other lncRNAs (see Supplementary information S1 (table)) for a summary of some key lncRNAs) that illuminate the molecular mechanisms by which lncRNAs can regulate gene expression in the nucleus. We highlight three key mechanisms: the ability to scaffold and recruit multiple regulatory proteins, the ability to localize to specific targets on genomic DNA, and the ability to utilize and shape three-dimensional (3D) nuclear structure. By integrating these properties, lncRNAs can carry out complex regulatory tasks that extend beyond what DNA elements or proteins are able to do, and we discuss several classes of biological processes that are ideally suited for lncRNA-mediated regulation of gene expression.

lncRNAs scaffold regulatory proteins

Many lncRNAs carry out their cellular functions by interacting with proteins to form macromolecular complexes. These interactions are mediated by specific

Polycomb group (PcG) proteins
A protein family involved in modifying histones and modulating chromatin structure to silence gene expression. PcG proteins comprise two main complexes — Polycomb repressive complexes 1 and 2.

elements — or 'domains' — in the RNA sequence, including short RNA sequence motifs or larger secondary or tertiary structures (for review, see REF. 43). An important feature of lncRNAs is that they often contain several discrete domains that interact with different proteins[21]. Here we discuss the evidence that lncRNAs can serve as scaffolds that coordinate the function of distinct transcription regulatory complexes.

Xist scaffolds multiple proteins to enable transcriptional silencing during XCI. XCI involves coordinated recruitment of many regulatory factors to the X chromosome. At the onset of XCI, induction of Xist expression triggers a cascade of events including loss of histone modifications associated with transcriptional activity[44–48], recruitment of the repressive Polycomb group (PcG) proteins and associated chromatin modifications[49,50], and chromosome compaction[51,52] — resulting in transcriptional silencing of most genes across the chromosome (FIG. 1). Early evidence suggested that Xist might coordinate these varied functions through interactions with multiple protein complexes because discrete regions of the RNA sequence are required for transcriptional silencing (A-repeat)[32], Polycomb repressive complex 2 (PRC2) recruitment (B-F repeat)[53], and localization to DNA (several, including the C-repeat)[32,33,54]. Furthermore, the functions of some of these regions appeared to be independent of one another: for example, deletion of the A-repeat precluded transcriptional silencing but did not affect Xist localization to genomic DNA[32]. Yet, which proteins were required for these various roles remained unknown for many years[24,55].

Recent advances in biochemical purification approaches enabled identification of the proteins that directly interact with Xist[56–58]. These studies, along with several independent functional studies[59,60], identified SMRT/HDAC1-associated repressor protein (SHARP; also known as SPEN and Msx2-interacting protein) as a direct Xist-interacting protein that is required for chromosome-wide transcriptional silencing on

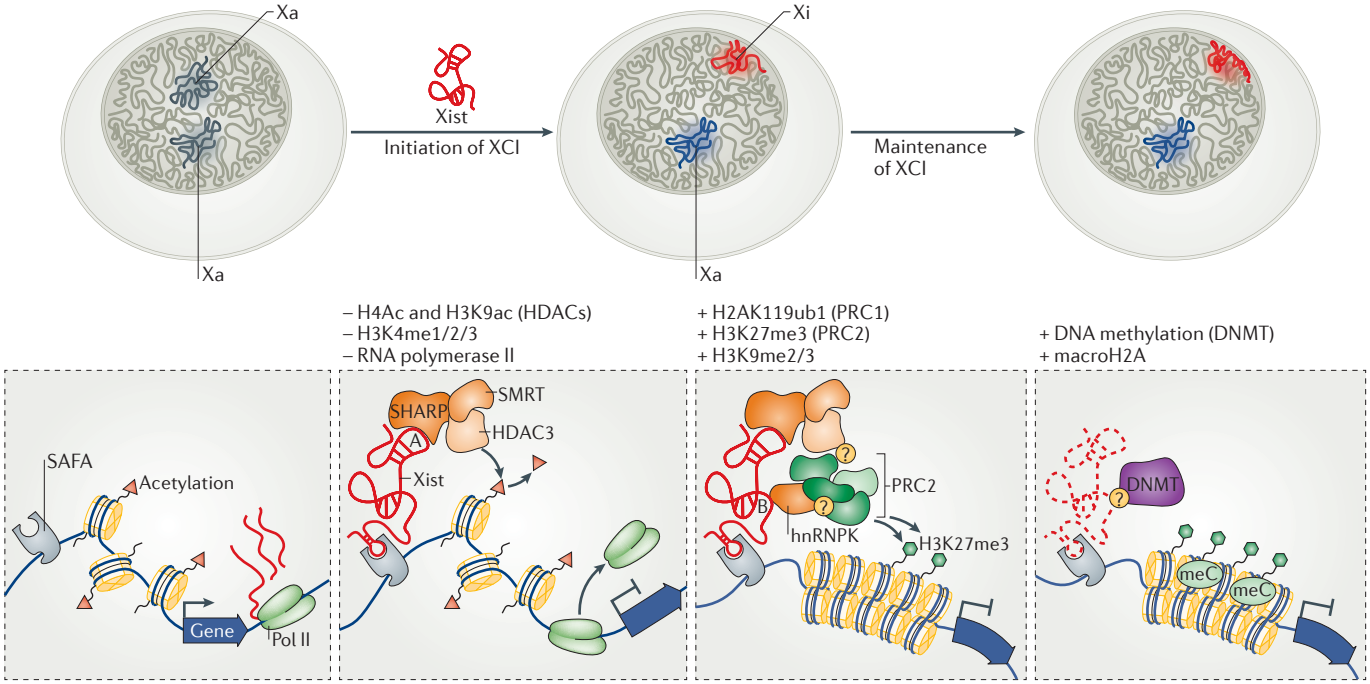


Figure 1 | **Xist recruits chromatin regulators to silence gene expression on the inactive X chromosome.** Xist (X inactive specific transcript) performs multiple roles during X chromosome inactivation (XCI) by recruiting various regulatory complexes at different stages to enact and maintain chromosome-wide transcriptional silencing. In embryonic stem cells, both X chromosomes are actively transcribed (Xa) and are marked by associated active histone modifications, including methylation of histone H3 at Lys 4 (H3K4me1; not shown) and acetylation of several residues on H3 and H4. During the initiation phase of XCI, early in embryonic development, Xist expression is activated on one of the two X chromosomes. Xist gradually spreads across the inactive X chromosome (Xi) by binding to chromatin through interactions with scaffold attachment factor A (SAFA), and recruits chromatin regulatory complexes to enact transcriptional silencing. Xist directly recruits SMRT/HDAC1-associated repressor protein (SHARP) through its A-repeat region (A) to promote histone deacetylation by histone deacetylase 3 (HDAC3), which is accompanied by demethylation of H3K4 and ejection of RNA polymerase II. Xist also recruits multiple complexes associated with gene repression, including polycomb repressive complex 1 (PRC1; not shown) and PRC2; the latter is recruited through an unknown mechanism (question marks) that requires binding of heterogeneous nuclear ribonucleoprotein K (hnRNPK) to the B-F repeat (B) of Xist as well as the presence of SHARP. The two PRC complexes deposit ubiquitylated H2AK119 (not shown) and H3K27me3, respectively, whereas the histone methyltransferase SETDB1 deposits the repressive H3K9me2 and H3K9me3 marks (not shown). In differentiated cells, the maintenance phase of XCI is characterized by stable gene repression through DNA cytosine methylation (meC) by DNA methyltransferase (DNMT), and the incorporation of the histone variant macroH2A; the mechanisms by which Xist recruits DNMT and macroH2A are unknown (dashed Xist and question mark). During this maintenance phase, the Xist RNA appears to be dispensable for transcriptional silencing, suggesting that the repressive marks present at this stage are sufficient to maintain silencing. SMRT, silencing mediator for retinoid and thyroid hormone receptors. The bottom left three panels are adapted from REF. 57, Nature Publishing Group.

the X chromosome. SHARP is an RNA-binding protein (RBP) that binds to the A-repeat of Xist[56,61]. SHARP was initially identified in humans as an interaction partner of the SMRT (silencing mediator for retinoid and thyroid hormone receptors) co-repressor complex[62], which recruits and activates histone deacetylase 3 (HDAC3)[63–65]. Consistent with these observations, both SMRT and HDAC3 are required for the exclusion of RNA polymerase II from the X-chromosome territory and subsequent transcriptional silencing[57]. Indeed, deacetylation across the future Xi is one of the earliest detectable events following the initiation of XCI[66]. These results indicate that Xist initiates transcriptional silencing by directly binding to SHARP, thereby recruiting SMRT and HDAC3 to trigger deacetylation on the X chromosome (FIG. 1).

Induction of Xist expression also leads to the recruitment of PRC2, which deposits the repressive histone 3 Lys 27 trimethylation (H3K27me3) modification across the Xi (FIG. 1). Although genetic deletion of PRC2 has no effect on the initiation of transcriptional silencing[67,68], it is required to maintain transcriptional silencing during the imprinted phase of XCI[69]. PRC2 was initially thought to interact with Xist directly[70], but recent studies have identified two proteins that directly bind Xist and seem to be required for PRC2 recruitment: SHARP and heterogeneous nuclear ribonucleoprotein K (hnRNPK)[56,57,59], although the precise details of how these proteins lead to recruitment of PRC2 are unclear. The Xist B-F repeat is required for PRC2 recruitment[71] and binds directly to hnRNPK[56] (M. Blanco and M. G., unpublished observations). Notably, the silencing functions of PRC2 and of SHARP-recruited HDACs may have distinct dynamics: histone deacetylation enables rapid transcriptional silencing, whereas PRC2-mediated changes provide more stable epigenetic memory[72]. Thus, Xist may coordinate the initiation and maintenance of XCI by recruiting complementary regulatory complexes through distinct domains in its sequence.

In addition to SHARP and hnRNPK, the Xist lncRNA binds to many proteins (10 to >100, depending on purification conditions), including a wide range of transcriptional silencers, DNA-binding proteins, lamina-associated proteins and RNA-modifying enzymes[56–58], some of which we discuss below. Although the roles of many of these proteins remain to be determined, this catalogue likely includes additional components that enable Xist to accomplish its varied tasks during XCI.

Other lncRNAs that interact with and scaffold regulatory proteins. Besides Xist, many nuclear lncRNAs have been reported to interact with a diverse range of proteins (FIG. 2a). Chromatin regulatory proteins reported to bind to RNA include those involved in histone modification (such as PcG proteins[73–75], G9a[76], NoRC[77], Lys-specific histone demethylase 1 (LSD1)[78] and WD-repeat-containing protein 5 (WDR5)[79,80]), DNA methylation (such as DNA methyltransferase 1 (DNMT1)[81]) and nucleosome remodelling (such as SWI/SNF[82]). Several lncRNAs have been reported to serve as co-activators for sequence-specific transcription factors: the lncRNA RMST (rhabdomyosarcoma 2-associated transcript) is involved in the proper localization of the transcription factor SOX2 during neurogenesis[83], and the lncRNA SRA1 (steroid receptor RNA activator 1) co-activates steroid receptor-dependent transcription[84]. Importantly, although many proteins have been reported to interact with lncRNAs, in most cases the exact functions of these interactions remain unclear; for example, PRC2 and some other chromatin regulatory complexes seem to interact promiscuously with many RNA molecules, including mRNAs[73,85]. Nevertheless, various studies have shown that in at least some cases these RNA–protein interactions can: recruit chromatin regulatory complexes to specific genomic sites to regulate gene expression[80,86–88]; competitively or allosterically modulate the functions of protein complexes[89–91]; and/or combine and coordinate the functions of multiple independent protein complexes[39,78] (FIG. 2b).

Indeed, many lncRNAs appear to serve as physical scaffolds, exploiting their large size to interact with multiple regulatory complexes simultaneously. Examples include Xist (FIG. 2c) and HOX transcript antisense RNA (HOTAIR), which regulates the HOXD gene cluster during limb development. HOTAIR associates with both PRC2, which deposits the repression-associated H3K27me3 modification, and the LSD1–CoREST–REST complex, which erases the activation-associated H3K4me2 modification, thereby coordinating two distinct yet functionally related chromatin-modifying activities[78]. The lncRNA Kcnq1ot1 (KCNQ1 opposite strand/antisense transcript 1) (Supplementary information S1 (table)) associates with both PRC2 and G9a, which write different repressive modifications (H3K27me3 and H3K9me3, respectively)[92]. A large-scale study found that many lncRNAs in mouse embryonic stem cells associate with multiple chromatin regulatory complexes[39].

In addition to scaffolding different complexes with distinct functions, some lncRNAs encode repetitive RNA domains that might enable high-avidity and/or multivalent interactions with one specific protein or complex (FIG. 2b). As an example, the lncRNA Firre (functional intergenic repeating RNA element) (Supplementary information S1 (table)) includes 12 repeated exons that evolved through segmental duplications in its genomic locus[93,94] (FIG. 2d). Each of these exons can interact with the nuclear scaffolding protein scaffold attachment factor A (SAFA; also known as hnRNPU), which is an abundant component of the nuclear matrix. SAFA contains both DNA- and RNA-binding protein domains[93,94] and is responsible for nuclear retention of many RNAs[95,96]. The repetitive structure of Firre might enable high-affinity interactions with chromatin and/or the formation of an interconnected, multimeric network of Firre RNA molecules at multiple sites on genomic DNA (FIG. 2d, see below).

LncRNA-mediated complexes are thought to be flexible, enabling tethering of multiple complexes in a modular manner such that their interactions are physically independent (FIG. 2a). Genetic dissection of yeast TLC1, known as telomerase RNA component (TERC) in humans (Supplementary information S1 (table)),

Imprinted phase of XCI
A process occurring on the paternally derived X chromosome of two- and four-cell stage embryos. Extra-embryonic tissues retain this imprinted pattern of XCI, whereas embryonic tissues reverse this imprinted pattern and then randomly inactivate one X chromosome.

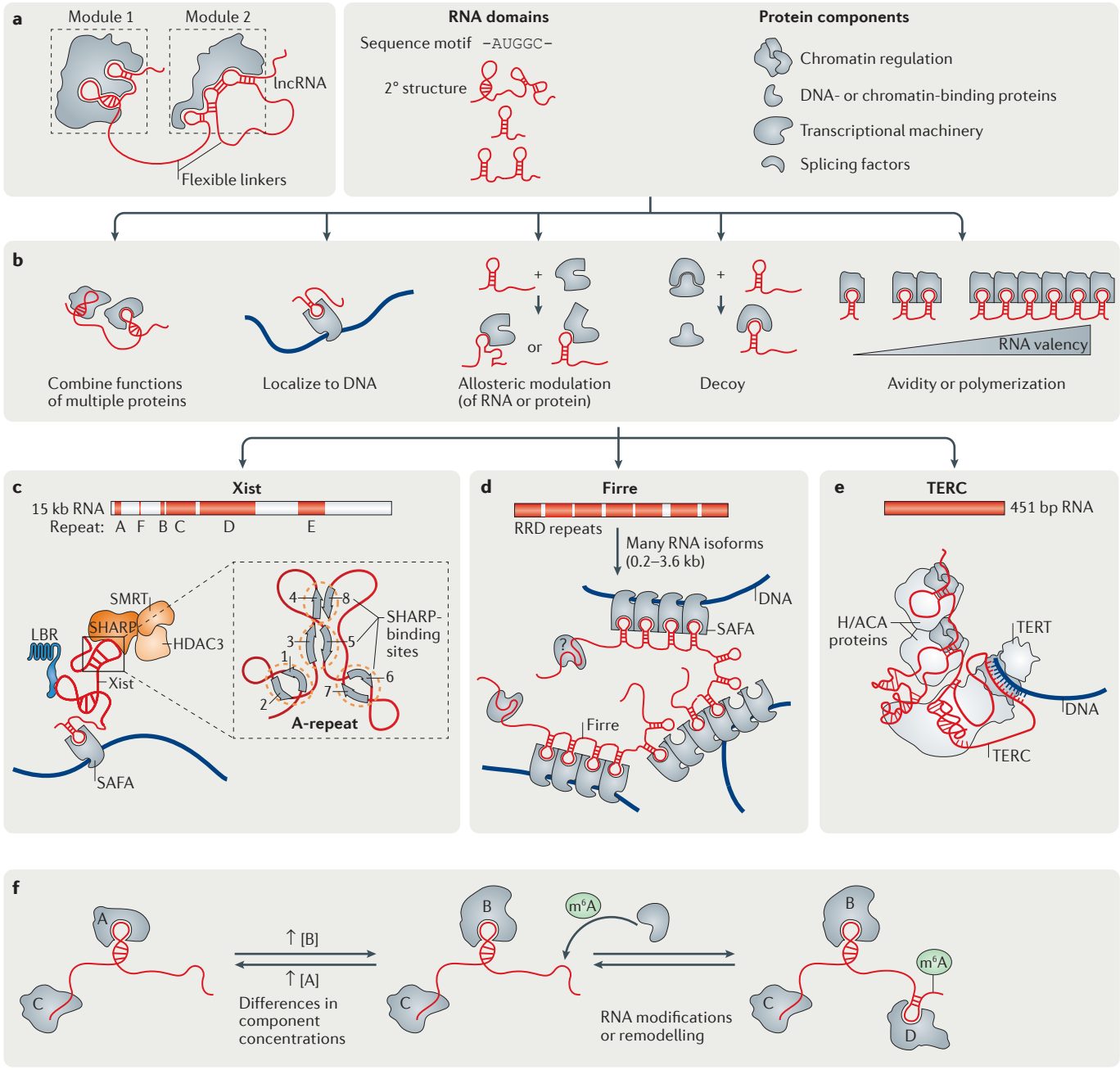
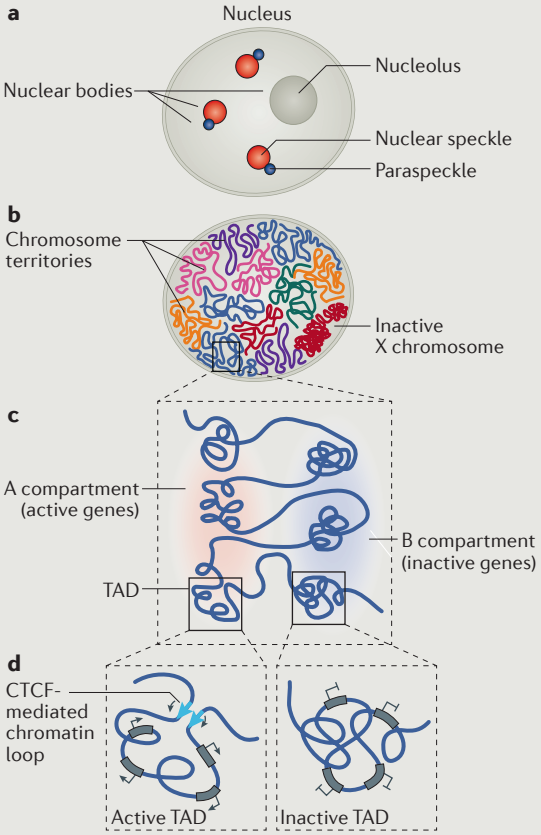


Figure 2 | **Long non-coding RNAs as scaffolds that integrate diverse regulatory proteins. a** | Various regions of the long non-coding RNA (lncRNA) molecule — each encoding specific sequence motifs or structural elements — can interact with distinct protein complexes in a modular manner. Functional regions can be connected by flexible linkers. **b** | LncRNA–protein interactions can have various functions, including: combining the functions of multiple proteins; localizing lncRNAs to genomic DNA; modifying the structure of lncRNAs or proteins; competitively inhibiting protein function (as decoys); and providing a multivalent platform, for example, to increase the avidity of protein interactions or to promote RNA–protein complex polymerization. **c** | Multiple modules can together create lncRNAs with diverse functions. For example, the mouse Xist contains repetitive sequences (A-F repeats) that interact with distinct protein complexes. An interaction with scaffold attachment factor A (SAFA) tethers Xist to DNA[56,57,95]. Interactions with SMRT/HDAC1-associated repressor protein (SHARP)[56–58] lead to transcriptional silencing. Xist interacts with SHARP through its A-repeat domain, which encodes eight tandem repeats that can hybridize to one another to provide multiple SHARP-binding sites[61]. Interactions with lamin B receptor (LBR) lead to repositioning of Xist-targeted loci to the lamina[136]. **d** | The Firre lncRNA is encoded at a locus that produces dozens of RNA isoforms with multiple copies of a repeating RNA domain (RRD). Multivalent interactions between Firre and SAFA anchor Firre to chromatin[93] and may form a scaffold that organizes chromosomal architecture. Firre may recruit additional proteins to this compartment. **e** | Telomerase RNA component (TERC) has distinct modules that interact with the protein components of the telomerase complex, such as telomerase reverse transcriptase (TERT) and the H/ACA proteins. These independent modules are connected by flexible linker regions[97]. TERC also encodes the sequence template for reverse transcription and extension of telomeric DNA. **f** | LncRNA–protein complexes can be dynamically regulated, for example, by altering the concentration of components that recognize the same RNA domain, or by altering the structure of the RNA through post-transcriptional RNA modifications or RNA remodelling (for example, by RNA helicases) so it can be recognized by new proteins. m^{6}A, N6-methyladenosine.

Box 1 | **Eukaryotic nuclei exhibit multiple, hierarchical levels of spatial organization**



The spatial localization of RNA, DNA and proteins in the nucleus is crucial for the proper functioning of nuclear processes and, consequently, it is highly organized and regulated. The nucleus contains several different nuclear bodies with high concentrations of specific RNAs and proteins, including: nucleoli, which are involved in both transcribing ribosomal RNA and assembling ribosomes[176]; nuclear speckles, which are involved in the regulation of pre-mRNA processing and splicing[155]; and the nuclear speckle-associated paraspeckles, the function of which is poorly understood (see the figure, part a). Nuclear bodies can be highly dynamic structures; for example, the formation of paraspeckles depends on the expression of the long non-coding RNA NEAT1 (nuclear enriched abundant transcript 1)[149,150]. In addition to the compartmentalization of RNA and proteins into nuclear bodies, DNA is organized in a hierarchical manner across many length scales, starting at the level of entire chromosomes down to individual genes. Chromosomes occupy discrete regions in the nucleus called chromosome territories (see the figure, part b) and generally avoid extensive intermingling[111]. Each chromosome is segregated into two spatial compartments, called A and B, that correspond to chromosomal domains and are transcriptionally active or inactive, respectively[177] (see the figure, part c). Both A and B compartments contain self-associating structures called topologically associating domains (TADs), which have a median size of 800 kb (REF. 137). In many cases, TADs are bordered by the protein CCCTC-binding factor (CTCF)[126,137], which binds to linearly distant DNA sequences and brings them into close three-dimensional proximity, resulting in the formation of a chromatin loop (see the figure, part d). TADs are highly conserved across cell types and species[137], and may provide a mechanism whereby neighbouring genes can be co-regulated by concentrating regulatory proteins or RNAs within a defined space.

revealed that this lncRNA comprises multiple domains connected by flexible linkers that are relatively unconstrained in their specific sequence and length[97] (FIG. 2e). Similarly, genetic deletions of specific domains in Xist, HOTAIR, Firre and others[32,78,93] have demonstrated some degree of functional independence of lncRNA domains. These results are consistent with the notion that lncRNAs can bring together unique combinations of functional complexes that may not interact or colocalize with one another except when engaged by the lncRNA. In most cases, the precise functions enabled by these unique combinations remain unclear and are under active investigation.

The composition of lncRNA–protein complexes can be dynamically controlled (FIG. 2f), including by changing the concentration of RBPs that recognize the same RNA element[98] or by altering the structure or post-transcriptional modifications of the lncRNA. For example, N^6-methyladenosine (m^6A) modifications of Xist are required for the recruitment of YTHDC1, which is an m^6A-binding protein, and for transcriptional silencing through an unknown mechanism[99]. Like other RNAs, which seem to have alternative structures *in vivo*[61], the 7SK non-coding RNA has two different RNA structures that interact with distinct sets of proteins[100]. The dynamic regulation of lncRNA–protein complexes remains relatively unexplored, and may best be revealed by single-molecule imaging or biochemical techniques that can define and characterize compositionally distinct lncRNA complexes that coexist in the same cell.

lncRNAs localize to specific DNA sites

Many lncRNAs, including Xist, regulate transcription by recruiting regulatory protein complexes to precise genomic locations. Proper localization of Xist and other lncRNAs on chromatin is crucial for their functions. Recent studies highlight two strategies to control how lncRNAs identify and interact with regulatory target sites on chromatin: affinity interactions to chromatin- or DNA-binding proteins, and proximity interactions, which are mediated by the 3D architecture of the genome (BOX 1; FIG. 3a).

Protein interactions provide lncRNA affinity for chromatin. To bind to chromatin on the X chromosome, Xist interacts with the nuclear matrix protein SAFA[56,57,95] (FIGS 2c,3a). High-resolution maps of Xist binding show that the RNA localizes broadly across the X chromosome, rather than at focal sites, likely mirroring the

relatively ubiquitous localization of SAFA[101]. Knockdown of SAFA or deletion of its RNA-binding domain leads to diffuse Xist localization throughout the nucleoplasm and loss of transcriptional silencing on the X chromosome[56,57,95]. Notably, SAFA localizes not only on the X chromosome but also on autosomes, and therefore on its own cannot account for the specific localization of Xist to the X chromosome (see below). Nonetheless, SAFA seems to provide the physical link that tethers the Xist lncRNA to genomic DNA.

Other lncRNAs use various mechanisms to achieve affinity for specific regions of the genome, including interactions with DNA- or chromatin-binding proteins[93,102]. In some cases, these mechanisms enable very punctate patterns of lncRNA localization to chromatin, in contrast to the broad localization of Xist. For example, the roX (RNA on X chromosome) non-coding RNAs, which mediate X chromosome dosage compensation in *Drosophila* spp. males[103] (Supplementary information S1 (table)), interact with chromatin through CLAMP (chromatin-linked adaptor for MSL proteins), a sequence-specific DNA-binding protein[102]. CLAMP recognizes ~150 copies of a specific motif spread across the X chromosome, leading to a punctate pattern of roX RNA localization at these sites[104–106]. Other examples include the highly abundant ~7 kb lncRNA MALAT1 (metastasis associated lung adenocarcinoma transcript 1), which co-purifies with DNA at active genes across the genome and shows a specific pattern of enrichment at the 3′ ends of genes[107,108] (FIG. 3a). MALAT1 seems to be recruited to active genes as part of the transcriptional machinery and shows

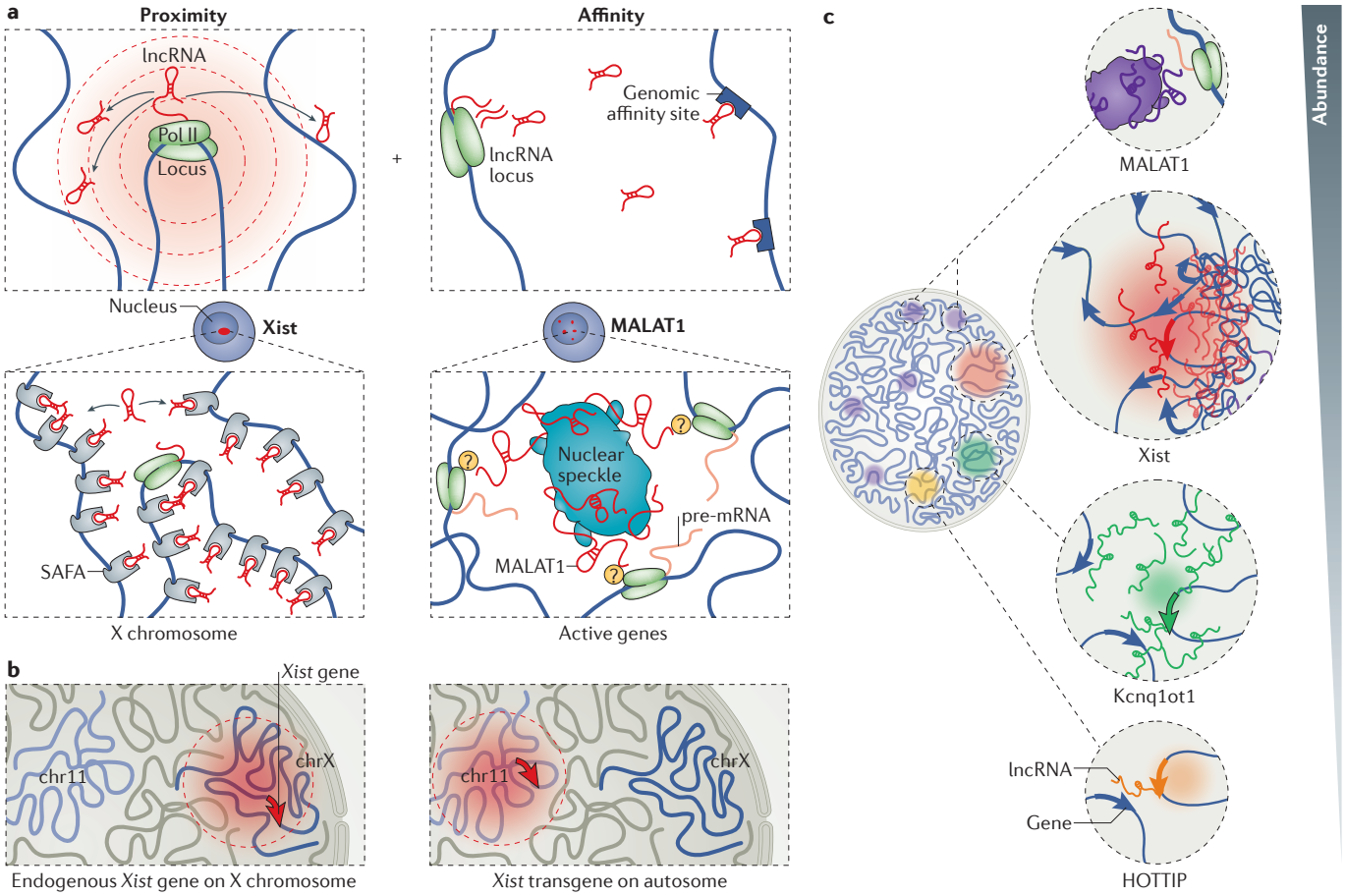


Figure 3 | **Patterns and mechanisms of long non-coding RNA localization to chromatin. a** | Two main strategies are used by long non-coding RNAs (lncRNAs) to identify and interact with specific sites on chromatin: three-dimensional (3D) proximity (relative to their sites of transcription) and affinity interactions with chromatin (mediated by interactions with chromatin- or DNA-bound proteins). LncRNAs use various combinations of these two strategies to identify regulatory targets. For example, Xist (X inactive specific transcript) displays a localization pattern dominated by 3D proximity. Its affinity for chromatin is dependent on interactions with scaffold attachment factor A (SAFA), which binds DNA relatively ubiquitously; thus, Xist spreads to sites in close 3D proximity to its locus and forms a single compartment in the nucleus. By contrast, MALAT1 (metastasis associated lung adenocarcinoma transcript 1) has a localization pattern that is dominated by affinity interactions: MALAT1 localizes to nuclear speckles and co-purifies with chromatin specifically at sites of active transcription. **b** | Genetic experiments involving changing the location of the Xist gene (from its endogenous locus to other locations on the X chromosome or on autosomes) demonstrated that 3D proximity defines Xist localization on chromatin. In ectopic loci on autosomes, Xist not only localizes to chromatin that is in proximity to its transgenic locus but also silences the expression of nearby genes (red circle), indicating that the specificity of its function is determined by its localization. **c** | Features intrinsic to lncRNAs, such as their abundance, might control the extent to which they spread to and regulate nearby genes. HOTTIP, HOXA transcript at the distal tip; Kcnq1ot1, KCNQ1 opposite strand/antisense transcript 1.

stronger interactions on chromatin with genes that are spliced and polyadenylated[107]. Thus, the specificity of localization and function of lncRNAs is in part determined by the proteins through which they attain affinity for chromatin.

Three-dimensional proximity can guide lncRNAs to their target sites. Protein affinity alone often does not entirely determine lncRNA localization. Xist, for example, localizes and functions exclusively in *cis*, on the same chromosome from which it is transcribed. Based solely on affinity, it is unclear how Xist could localize across the entire X chromosome yet not spread to other chromosomes — including the other X chromosome, which contains the same sequences. Indeed, neither DNA-binding specificity nor chromatin state can explain the restriction of Xist in *cis*, because at the initiation of XCI both X chromosomes are competent for silencing[66]. Instead, Xist localization and function appear to be controlled primarily by the location of its transcription on the genome (FIG. 3b).

To understand how genomic location leads to *cis*-restricted function of Xist, several groups characterized Xist localization during the initiation of XCI at high resolution[109,110] and found that Xist initially localizes to the DNA regions that most frequently contact the *Xist* genomic locus in 3D space[109]. Moving Xist to a new location of the X chromosome, which exhibits a distinct pattern of proximity contacts, led to a new localization pattern that reflected its new location[109] (FIG. 3b). Thus, Xist exploits 3D genome organization to find its regulatory targets. This proximity-guided localization strategy likely explains how Xist initially localizes to the Xi and not to other chromosomes, because DNA regions on a given chromosome are physically closest to other sites on the same chromosome, in what is termed a chromosome territory[111] (BOX 1). As it spreads from its site of transcription, Xist modifies chromatin state and chromosome architecture (see below); selective localization to the Xi may involve multiple iterations of Xist spreading in proximity to its transcription locus and subsequently changing nuclear architecture to draw in more distant regions of the Xi[109].

Such a 3D proximity-guided localization mechanism highlights a unique role for lncRNAs in gene regulation. Unlike proteins, which must be imported into the nucleus after translation and thus inherently lack information about their loci of origin, lncRNAs can function immediately upon their transcription and processing in the nucleus and regulate the expression of genes that are in close 3D proximity on a specific allele (FIG. 3a). In addition to enabling allele-specific gene regulation, a proximity-mediated search strategy might also explain how a low-abundance lncRNA can reliably identify target genes in the nucleus, because it is present at a high concentration at sites that are proximal to its genomic locus (FIGS 3a,c).

Numerous lncRNAs have been reported to regulate the expression of one or several spatially proximal genes[80,92,112]. For example, Kcnq1ot1 is expressed in a monoallelic manner from the imprinted *Kcnq1* locus and silences the expression of several neighbouring genes in the same chromosomal domain[92], likely utilizing spatial proximity to enable allele-specific regulation in a manner similar to Xist. The lncRNA HOTTIP (HOXA transcript at the distal tip) (Supplementary information S1 (table)) is encoded in the *HOXA* gene cluster and recruits various transcriptional activators to spatially proximal *HOXA* genes[80]. Importantly, several lncRNAs have been demonstrated to regulate proximal gene expression when they are expressed adjacently to or recruited to a reporter gene on a plasmid[80,113,114], suggesting that, like Xist, the specificity of their localization and function is primarily encoded by physical location rather than by preferential affinity for their regulatory targets. Thus, spatial proximity may have a dominant role in guiding the specificity of many RNA regulators to their target genes.

Although the physical architecture of chromosomes generally favours interactions between loci on the same chromosome, regions on different chromosomes can also be located in close physical proximity in the nucleus[115] (BOX 1). There are now several examples where the principle of proximity-guided search can enable lncRNAs to interact with regions on different chromosomes. Notable examples include Firre, which is encoded on the X chromosome and forms a punctate cloud in the nucleus that includes multiple sites on other chromosomes[93] (see below), and the *CISTR–ACT* locus, which is encoded on chromosome 12 in humans but is frequently found to colocalize with the *SOX9* locus on chromosome 17 (REF. 116). Although the precise mechanisms by which these inter-chromosomal contacts affect gene expression are unknown, these findings show that lncRNAs can identify and interact with genomic targets on different chromosomes through 3D proximity, exploiting the spatial contacts of their genomic loci to identify and localize to sites throughout the genome.

Genomic localization by combination of affinity and three-dimensional proximity. Together, these data suggest that many lncRNAs balance the protein-affinity and 3D-proximity strategies to locate and interact with their target sites in the genome. For some lncRNAs, 3D proximity dominates their localization to chromatin (FIG. 3a). The localization of Xist during early XCI seems to be guided primarily by proximity[109], likely because it interacts with chromatin through a relatively ubiquitous factor, SAFA. For other lncRNAs, the affinity component dominates their localization: MALAT1 localizes to thousands of actively transcribed genes throughout the genome, most of which are not located in proximity to its genomic locus[107,108]. This pattern of interactions presumably arises because the proteins that interact with MALAT1 have a high affinity for actively transcribed loci. Xist and MALAT1 illustrate the two extremes, and other lncRNAs may localize to chromatin using combinations of affinity and proximity[93,116]. For example, roX2 binds to specific DNA locations that are defined in part by the binding pattern of CLAMP, but preferentially binds to CLAMP-binding sites that are in close 3D proximity to the *roX2* locus[104,117].

The precise balance between the two strategies of affinity and proximity may depend on the abundance and/or stability of the lncRNA (FIG. 3c). For example, low-abundance lncRNAs like HOTTIP (<1 copy per cell) only interact with and regulate genes in very close proximity, perhaps because the lncRNA cannot attain a high enough concentration at sites in the nucleus beyond its vicinity. By contrast, Xist is present at moderate levels (50–100 copies per cell[118]) and has low specificity; these properties may enable Xist to spread across the entire X chromosome while limiting its ability to spread to other chromosomes. Finally, MALAT1 is highly abundant (~3,000 copies per cell[119]) and stable[120], which may allow it to diffuse throughout the nucleus to search for sites with which it has high affinity. Thus, properties intrinsic to each lncRNA may control the extent of lncRNA diffusion and localization in the nucleus.

lncRNAs shape nuclear organization

The role of RNA in shaping nuclear architecture has long been speculated[121–124], and recent examples have identified nuclear-retained lncRNAs that are required for the formation of specific nuclear structures. Indeed, some lncRNAs seem to use their abilities to scaffold proteins and bind chromatin to manipulate the architecture of the nucleus. Below, we discuss the evidence that lncRNAs act to establish nuclear compartments, defined here as spatial locations that contain high concentrations of specific RNAs, proteins and/or genomic sites.

Xist shapes the three-dimensional structure of the inactive X chromosome. Expression of Xist during XCI leads not only to changes in chromatin structure and histone modifications but also to large-scale remodelling of the 3D architecture of the Xi[58,125–129] (FIG. 4). One of the most striking architectural changes associated with the induction of Xist is the repositioning of the entire Xi adjacently to the nuclear lamina[130], which is a nuclear compartment generally associated with transcriptionally inactive genes[131,132] (FIG. 4a). Recent proteomic studies revealed that Xist directly interacts with the lamin B receptor (LBR)[57,58], which is a transmembrane protein at the inner nuclear membrane that interacts with lamin B and is required to anchor chromatin to the nuclear lamina[133–135]. Knockdown of LBR or disruption of the Xist–LBR interaction leads to defective recruitment of the X chromosome to the nuclear lamina, demonstrating that this large-scale architectural change is directly mediated by Xist. Deletion of LBR or LBR-binding sites on Xist also prevents this lncRNA from spreading to actively transcribed genes across the X chromosome and abolishes Xist-mediated gene silencing[136]. These results indicate that Xist reshapes nuclear structure by tethering the X chromosome to the nuclear lamina, and that this tethering is required for Xist to access genes across the X chromosome and silence their transcription.

In addition to relocating to the nuclear lamina, the Xi adopts a unique 3D structure following XCI[58,125–127]. Normally, chromosomes form a series of self-interacting architectural domains, termed topologically associated domains (TADs), each of which includes hundreds of kilobases of DNA[137] (BOX 1). During XCI, the Xi seems to lose this typical TAD structure and instead forms two 'mega-domains' of interactions across the chromosome[126,128,129] (FIG. 4b). Whereas the patterns of contacts on most chromosomes show punctate interactions between specific loci, the mega-domains on the Xi exhibit a more random pattern of interactions[125–127]. The two mega-domains on the Xi are demarcated at the *DXZ4* locus[126], which is essential to maintain this boundary[128,129].

A final notable feature of the architecture of the Xi is the existence of 27 long-range, higher-order DNA looping interactions (termed superloops) between genomic loci that are separated by 7–74 Mb, including interactions between loci in separate mega-domains[126,128]. These superloops are only observed on the Xi but not the Xa[126] (FIG. 4a). Intriguingly, several of the loci that participate in these superloops encode lncRNAs, including the *DXZ4* locus (which encodes the lncRNAs DANT1 and DANT2 (REF. 138)), FIRRE, ICCE (also known as LOC550643) and XIST itself, each of which is expressed from the Xi[126,128]. It is unknown whether these lncRNAs mechanistically contribute to the formation of these unique 3D genomic contacts, or whether these contacts are mediated by the genomic DNA near these lncRNA loci and do not depend on the lncRNAs themselves.

The precise mechanisms by which Xist establishes or maintains this unique chromosomal architecture remain unclear. Deletion of Xist after establishment of XCI leads to a partial reversion of the Xi structure back to that of the Xa[125]. This occurs without reactivating the expression of silenced genes on the Xi, indicating that these structural effects are actively maintained by Xist and are not merely caused by transcriptional silencing. One possible explanation for these observations is that these structural effects are maintained by the Xist–LBR interaction. Xist may also affect chromosomal architecture by recruiting PcG proteins, which can influence 3D chromosome architecture by aggregating into PcG-associated domains[139–142], or by recruiting other structural proteins such as SMCHD1 (structural maintenance of chromosomes flexible hinge domain-containing protein 1), which is known to contribute to the higher-order structure of the Xi[143].

Other lncRNAs nucleate and organize nuclear compartments. In addition to Xist and its role in XCI, other lncRNAs are now known to organize higher-order chromosome architecture[87,93]. For example, Firre forms a punctate compartment in the nucleus that includes not only its locus on the X chromosome but also several specific loci on mouse chromosomes 2, 9, 15 and 17 (REF. 93) (FIG. 5a). Genetic deletion of Firre leads to loss of the observed colocalization of these interacting DNA loci, indicating that the lncRNA is required for these inter-chromosomal interactions, and also leads to broad transcriptional changes including many genes involved in RNA processing and electron transport chain metabolism[93,144]. Notably, Firre is required for adipogenesis[40], and several of the loci with the strongest Firre occupancy encode genes that are involved in energy metabolism and/or adipogenesis[93]. These observations suggest

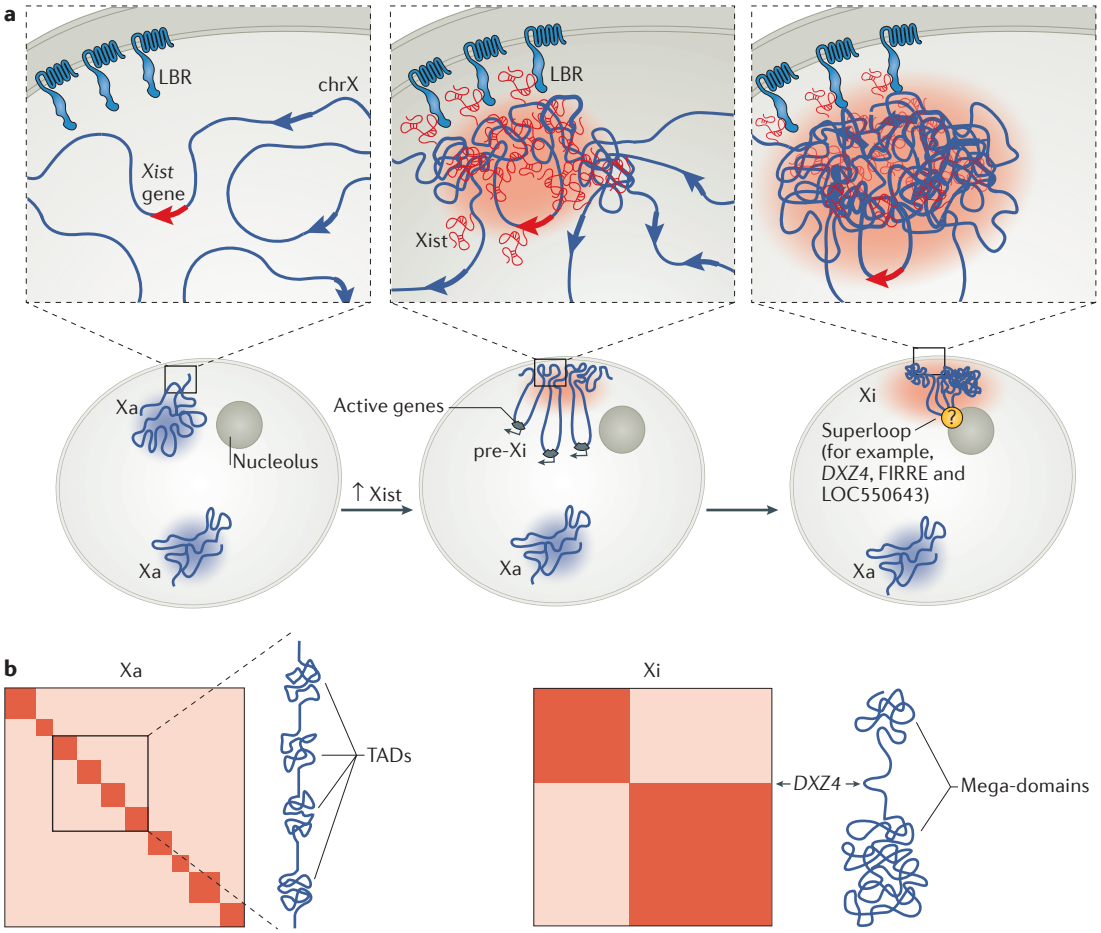


Figure 4 | **Architectural changes to the X chromosome during X chromosome inactivation.** The X chromosome undergoes major structural changes during X chromosome inactivation (XCI). Before the expression of Xist (X inactive specific transcript) in female cells, both X chromosomes are transcriptionally active, not strongly associated with the nuclear lamina (part **a**, left) and adopt a similar organization to that of the autosomes, including characteristic topologically associating domains (TADs; see BOX 1) (part **b**, left). Once Xist has been expressed by one allele, it begins to spread across the future inactive X chromosome (Xi), and its interaction with lamin B receptor (LBR) enables it to re-position the X chromosome to the nuclear lamina (**a**, middle). Finally, active genes are sequestered into the Xist compartment (red circle) and silenced (part **a**, right). During this process most TADs on the X chromosome are lost, and two large 'mega-domains' are formed, separated by a boundary at the *DXZ4* locus, which has been observed to associate with the nucleolus through an unknown mechanism (part **a and part b**, right). The *DXZ4* locus also forms long-range interactions (known as superloops) with specific loci on the inactive X chromosome, including loci encoding the lncRNAs FIRRE and LOC550643 (also known as ICCE)[128]. In part **b**, the heatmap diagram depicts contact frequency between genomic sites on the X chromosome, simulating the visualization of high-throughput chromosome conformation capture data. Red boxes denote TADs or mega-domains, which are chromosomal regions with elevated frequency of internal chromatin interactions. Top middle panel in part **a** is adapted with permission from REF. 136, AAAS.

that Firre may promote these architectural changes to spatially coordinate the regulation of genes involved in the same biological process.

Whereas Xist and Firre mediate the formation of compartments that shape the organization of DNA, other lncRNAs nucleate and maintain nuclear bodies that concentrate specific proteins and/or RNAs[145–147] (BOX 1). One of the best examples is the 23 kb NEAT1 (nuclear enriched abundant transcript 1) lncRNA (Supplementary information S1 (table)), which nucleates the formation of compartments called paraspeckles, containing various mRNAs and RBPs[148–150] (FIG. 5b). The functions of paraspeckles remain poorly defined, but they are known to be dynamically regulated in various processes (including response to viral infection[151]), to lead to nuclear retention of certain mRNAs that had been subjected to high levels of adenosine-to-inosine editing (potentially to limit their translation in the cytoplasm[152]) and to concentrate and potentially sequester certain RBPs to limit their functions in the nucleus[151]. NEAT1 is required for the structural integrity of this nuclear body: knockdown or knockout of NEAT1 leads to the dispersion of paraspeckle proteins[149,150], and synthetic tethering of NEAT1 to a genomic site is sufficient for local paraspeckle formation[153]. Paraspeckle integrity requires not only the NEAT1 RNA itself but also its ongoing transcription,

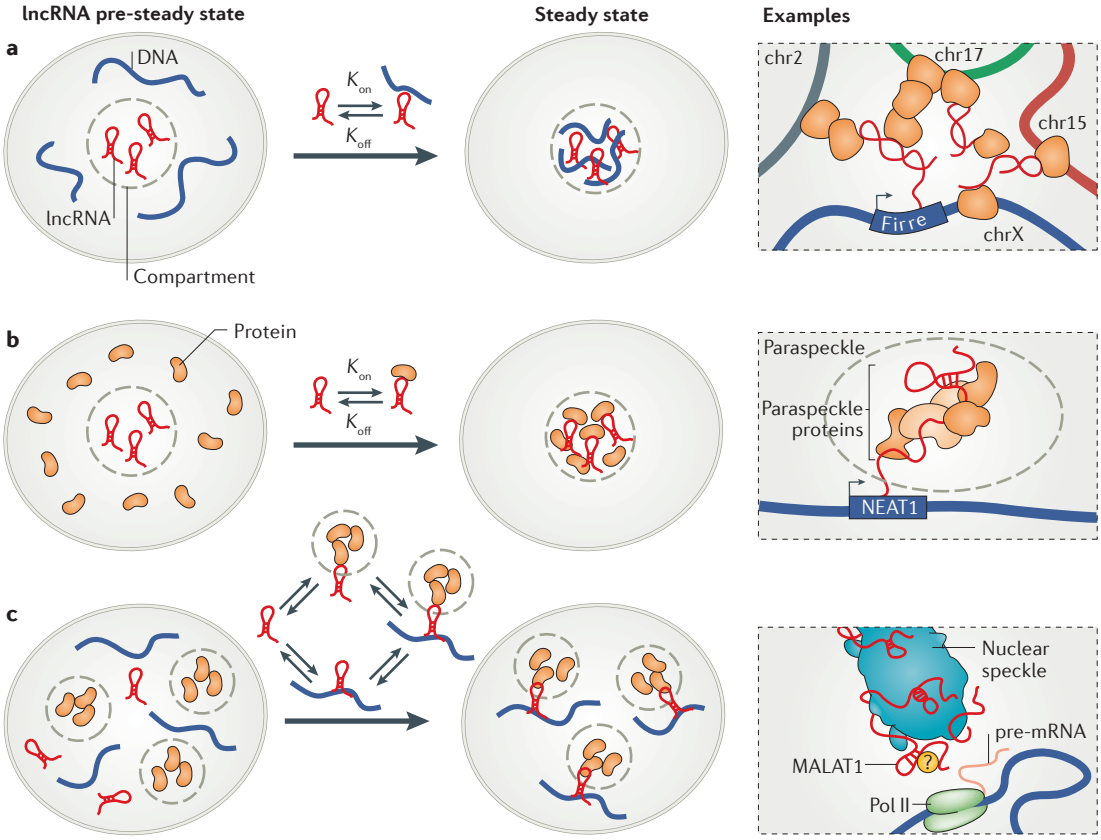


Figure 5 | **Long non-coding RNAs can form spatial compartments in the nucleus. a** | Interactions with proteins and DNA enable long non-coding RNAs (lncRNAs) to form dynamic spatial compartments in the nucleus. Assembly may be controlled by passive diffusion and interactions that are characterized by specific association and dissociation constants (K_{on} and K_{off}). The lncRNA Firre forms high-affinity interactions with specific DNA loci on different mouse chromosomes, resulting in the colocalization of these loci in three-dimensional (3D) space to facilitate the co-regulation of genes involved in energy metabolism and/or adipogenesis. **b** | The upregulation of the lncRNA NEAT1 (nuclear enriched abundant transcript 1), which initially localizes in close 3D proximity to its transcribed locus, nucleates the formation of paraspeckles. These are characterized by high local concentration of NEAT1 and of paraspeckle proteins. **c** | High affinity to both protein and DNA enables lncRNAs to recruit genomic loci to functional nuclear bodies to influence gene expression. For example, interactions between MALAT1 and protein components of pre-organized nuclear speckles could assist in promoting transcription or regulating mRNA splicing. Pol II, RNA polymerase II.

suggesting that paraspeckle formation is coupled to NEAT1 biogenesis[154]. This dependence on transcription enables both rapid assembly as well as rapid disassembly of paraspeckles[154], potentially providing a mechanism by which nuclear compartment formation and function might be dynamically controlled.

Another remarkable example is MALAT1 (Supplementary information S1 (table)), which localizes to compartments called nuclear speckles (FIG. 5c). Nuclear speckles contain various splicing, RNA-processing and transcription factors, including many SR proteins[155], and are thought to function as a storage–assembly–modification area for RNA-processing proteins when they are not actively engaged[156]. Actively transcribed genes associate with the periphery of nuclear speckles[157,158]. Genome-wide mapping of MALAT1 localization to DNA showed that it associates with all actively transcribed genes in a dynamic and transcription-dependent manner[107,108]. MALAT1 also interacts with dozens of RBPs, including multiple SR proteins[159–162] (Supplementary information S1 (table)). Although MALAT1 is not required for the overall integrity of nuclear speckles, MALAT1 loss-of-function mutation leads to improper localization of a subset of nuclear speckle proteins[162,163]. The functional consequences of this improper localization are unclear. Because MALAT1 associates with actively transcribed genes, its role might be to help reposition active genes to the periphery of nuclear speckles. According to this model, MALAT1 might organize these nuclear structures by scaffolding interactions between active genes, nuclear speckles and certain RNA-processing proteins (FIG. 5c).

A model for how lncRNAs can shape three-dimensional nuclear structures. The above examples indicate that, by virtue of their ability to serve as molecular scaffolds, lncRNAs can assemble nuclear compartments that contain multiple regulatory factors and their targets. This model of lncRNAs as scaffolds suggests that the dynamic assembly of nuclear compartments may involve passive,

SR proteins
Proteins that contain domains enriched in Arg–Ser dipeptides. Many SR proteins are nuclear and involved in RNA processing.

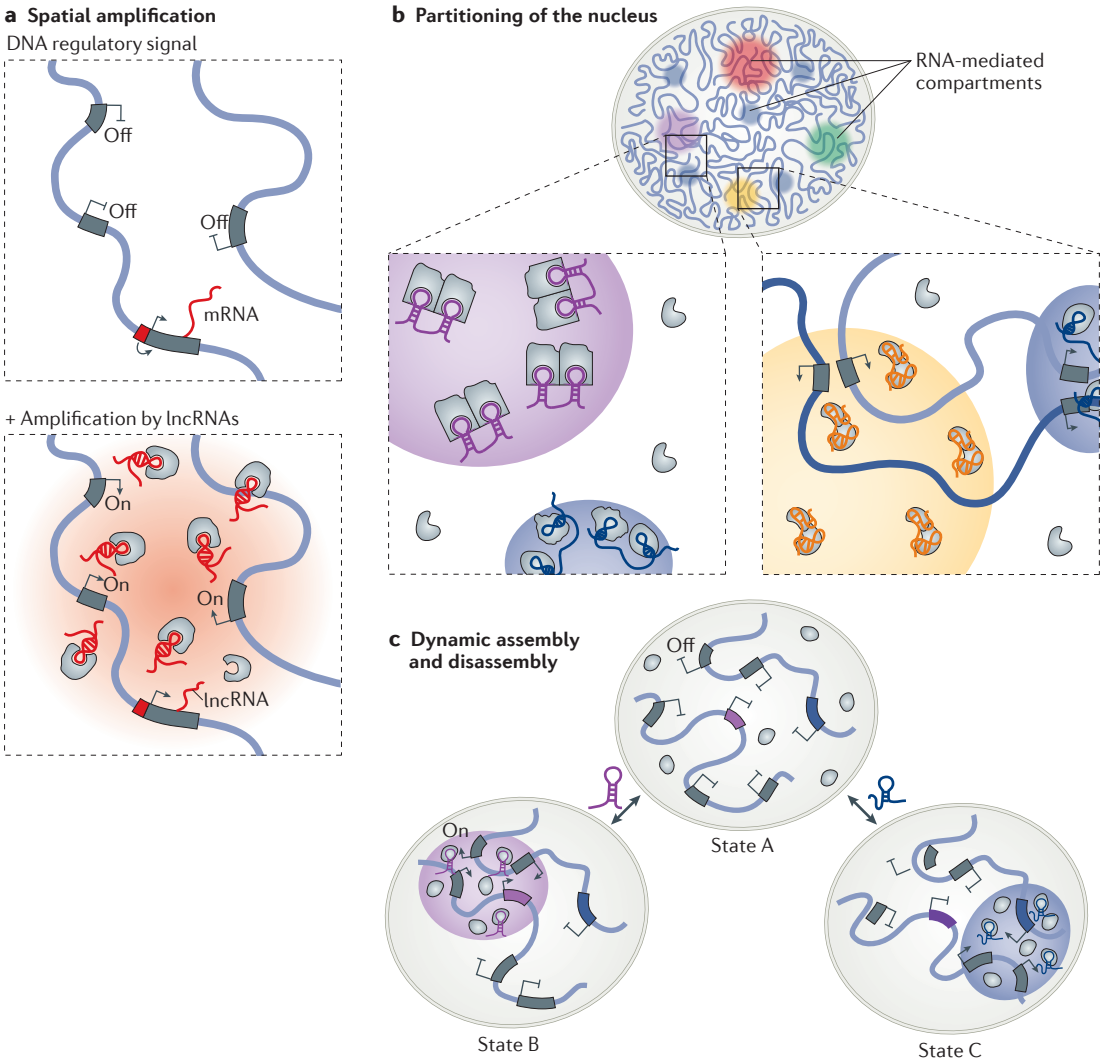


Figure 6 | **The unique abilities of long non-coding RNAs as regulators of gene expression and nuclear architecture.** **a** | Long non-coding RNAs (lncRNAs) can spatially amplify regulatory signals in the nucleus. Regulatory information encoded by the promoter of a protein-coding gene has limited ability to affect the expression of nearby genes, owing to topological constraints on chromatin structure (top). Regulatory information in the promoter of a lncRNA gene can be spatially amplified by diffusion of the lncRNA to genes in close spatial proximity (bottom). **b** | By virtue of their ability to encode multivalent structures and scaffold multiple types of regulatory protein, lncRNAs seem to be particularly suitable for nucleating the formation of cellular substructures. The formation of nuclear compartments by lncRNAs may help to partition the nucleus into distinct territories that enhance the efficiency of regulatory processes. These compartments may comprise regions of high concentration of specific regulatory factors (left pullout), or include specific genomic target sites that might reside on different chromosomes (right pullout). **c** | Expression of lncRNAs (purple and blue) may enable dynamic assembly and disassembly of functional nuclear compartments. These compartments may contain the same regulatory proteins (grey circles) but assemble around different genomic sites by virtue of the position of the lncRNA locus.

diffusion-based processes rather than active, ATP-dependent mechanisms. For example, how the entire X chromosome becomes associated with the nuclear lamina during XCI might be explained by the abilities of Xist to passively diffuse through the nucleus and to interact both with DNA (through SAFA) and with the nuclear lamina (through LBR). Once the Xist RNA tethers one region of the X chromosome to the lamina, the movement of the rest of the chromosome may be constrained[164,165], promoting further interactions mediated by other Xist molecules (FIG. 4a). Similarly, NEAT1 may establish a paraspeckle compartment co-transcriptionally, while being tethered to chromatin by RNA polymerase II, by attracting paraspeckle proteins that diffuse in the proximity of, and stably interact with, the NEAT1 RNA (FIG. 5b).

In some cases, the ability of protein components within these compartments to self-interact may also contribute to compartment assembly and maintenance. For example, nuclear compartments can be formed through liquid–liquid phase separation, in which sets of cellular components passively de-mix to form separate compartments[166–168]. Such de-mixing processes are driven by interactions between multivalent components.

For instance, two subcompartments form in the nucleolus through phase separations driven by the FIB1 (rRNA 2′-*O*-methyltransferase fibrillarin) and nucleophosmin 1 proteins, both of which contain RNA-binding domains as well as low-complexity domains that can polymerize[169]. Many other RBPs contain low-complexity domains that facilitate self-aggregation[170,166], and in some cases RNA has been shown to facilitate this process[167,171]. LncRNAs might initiate such liquid–liquid phase transitions either by increasing local concentrations of such RBPs (for example, NEAT1-mediated recruitment of paraspeckle proteins; FIG. 5b) or by directly participating in heterotypic multivalent interactions through RNA sequence repeats (for example, the repetitive RNA domains of Firre and its interactions with SAFA; FIG. 2d).

The importance of self-reinforcing interactions within compartments is not necessarily limited to liquid–liquid phase separation. The Xist-mediated recruitment of DNA to the nuclear lamina may be aided by changes in chromatin state that enable DNA tethering to the lamina independently of continuous interactions with Xist, allowing a single, unengaged Xist molecule to bind to new parts of the X chromosome and recruit them to the nuclear lamina. Testing these lncRNA-mediated nuclear compartment dynamics will require the use of live-cell imaging of multiple compartments to determine how lncRNAs cause directed movement and organization of nuclear structures.

Principles of lncRNA regulation

The properties described above — including the abilities to scaffold proteins, localize to DNA across 3D nuclear territories and organize nuclear structures — may help to explain how many lncRNAs mechanistically control gene expression. Xist, for example, uses these abilities to induce XCI: it recruits multiple protein complexes to sites in close 3D proximity and alters nuclear architecture to initiate and maintain transcriptional silencing. Other lncRNAs may use various combinations of these capabilities to regulate gene expression. In this section, we explore how these properties enable unique types of regulatory function, which might explain why mammalian cells use lncRNAs to control certain gene expression programmes.

Spatial amplification of regulatory information. As discussed above, lncRNAs can localize and spread across chromatin in proximity to their genomic loci. Because of these properties, some lncRNAs, including Xist, Firre and Kcnq1ot1 (Supplementary information S1 (table)), can simultaneously regulate multiple genes that are spatially clustered. This form of regulation precludes the need for independent regulatory elements at each gene. Studies in bacteria, in which genes of similar function are organized in linear operons, suggest that this arrangement can increase efficiency by ensuring the coordinated expression of functionally dependent components[172]. LncRNAs may provide a similar means to achieve efficient regulation in eukaryotes by controlling clusters of genes that are assembled together based on shared function (for example, Firre and genes involved in adipogenesis) or regulation (for example, Xist and X chromosome genes during XCI). In such cases, the lncRNA effectively amplifies the regulatory information that is encoded in its promoter to control a broader gene expression programme.

LncRNAs are unique in their ability to spatially amplify regulatory information encoded by DNA. Unlike proteins, lncRNAs can act in close proximity to their site of transcription; and unlike DNA regulatory elements, lncRNAs can amplify DNA-encoded regulatory signals to different extents according to their expression levels (FIGS 3c,6a). Furthermore, lncRNAs are not necessarily restricted by topological constraints of the chromatin fibre, allowing them to diffuse to or mediate contacts at spatially proximal sites that might even reside on different chromosomes. This mode of control can also enable coordinated allele-specific expression of multiple proximal genes. The unique ability of RNA regulators to accomplish this task is highlighted by the convergent evolution of lncRNAs involved in chromosome-wide dosage compensation in other organisms, including roX in *Drosophila* spp.[103] and Rsx (RNA on the silenced X) in metatherians[173], as well as in genomic regions subject to genomic imprinting[174].

Spatial partitioning of the nucleus for efficient regulation. By virtue of their functions as scaffolds and signal amplifiers, lncRNAs can also provide unique platforms to guide regulatory complexes to specific locations in the nucleus (FIG. 6b). These functions might involve combining different regulatory factors, recruiting chromatin regulatory complexes to specific sites on DNA, or increasing the local concentration of a particular factor through multimerization. The resulting nuclear compartments may improve the kinetic efficiency of nuclear processes. For example, the concentration of regulatory factors and their targets in specific territories might enhance the efficiency of target search, while preventing the localization of these regulatory factors to regions of the nucleus where they are not needed. Examples such as Xist and NEAT1 suggest that lncRNAs can serve as the physical scaffolds to spatially partition the nucleus into functional compartments (Supplementary information S1 (table)).

Dynamic assembly and disassembly of functional compartments. Because lncRNAs are functional immediately upon transcription and can diffuse in the nucleus, regulating their transcription and/or degradation may enable rapid assembly or disassembly of nuclear compartments (FIG. 6c). For example, continued transcription of lncRNAs such as NEAT1 is required to maintain their associated nuclear compartments (paraspeckles in the case of NEAT1)[145–147,149,150]. This property may be particularly useful for combinatorial assembly of dynamic regulatory programmes in specific cell types or conditions, and indeed many aspects of nuclear structure are dynamically regulated. One example is the localization of DNA to the nuclear lamina: although ~40% of the genome is associated with the lamina in any given cell type, the specific regions of the genome vary from

Genomic imprinting
An epigenetic phenomenon in which expression of a gene is restricted to a single allele based on parental origin.

cell to cell[175]. LncRNAs may have an important role in this process, particularly given the newly appreciated ability of LBR to bind RNA[136]. Through this mechanism and others, some lncRNAs may act to dynamically assemble compartments that contain co-regulated genes.

Concluding remarks

LncRNAs can have specialized functions in diverse biological processes and can shape nuclear structure and regulate gene expression. They occupy a unique position in the regulatory landscape of the nucleus, with an intrinsic and efficient capacity for dynamically and locally amplifying DNA-encoded regulatory information. The examples presented in this Review provide a first glimpse into these varied functions and mechanisms of action of nuclear lncRNAs. Many questions remain, including which lncRNAs in fact act to regulate gene expression as well as how these principles are integrated to perform specific cellular functions. As functional and mechanistic dissection of lncRNAs progresses, we will be able to refine our understanding of these principles as well as uncover additional mechanisms by which lncRNAs address biological challenges in gene regulation.

1. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628 (2008).
2. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
3. Guttman, M. *et al. Ab initio* reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat. Biotechnol.* **28**, 503–510 (2010).
4. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
5. Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25**, 1915–1927 (2011).
6. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775–1789 (2012).
7. Guttman, M., Russell, P., Ingolia, N. T., Weissman, J. S. & Lander, E. S. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* **154**, 240–251 (2013).
8. Chew, G.-L. *et al.* Ribosome profiling reveals resemblance between long non-coding RNAs and 5′ leaders of coding RNAs. *Development* **140**, 2828–2834 (2013).
9. Bánfai, B. *et al.* Long noncoding RNAs are rarely translated in two human cell lines. *Genome Res.* **22**, 1646–1657 (2012).
10. Ayupe, A. C. *et al.* Global analysis of biogenesis, stability and sub-cellular localization of lncRNAs mapping to intragenic regions of the human genome. *RNA Biol.* **12**, 877–892 (2015).
11. Clark, M. B. *et al.* Genome-wide analysis of long noncoding RNA stability. *Genome Res.* **22**, 885–898 (2012).
12. Chen, J. *et al.* Evolutionary analysis across mammals reveals distinct classes of long non-coding RNAs. *Genome Biol.* **17**, 19 (2016).
13. Washietl, S., Kellis, M. & Garber, M. Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. *Genome Res.* **24**, 616–628 (2014).
14. Hezroni, H. *et al.* Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep.* **11**, 1110–1122 (2015).
15. Necsulea, A. *et al.* The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* **505**, 635–640 (2014).
16. Latos, P. A. *et al.* Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science* **338**, 1469–1472 (2012).
17. Ebisuya, M., Yamamoto, T., Nakajima, M. & Nishida, E. Ripples from neighbouring transcription. *Nat. Cell Biol.* **10**, 1106–1113 (2008).
18. Engreitz, J. M. *et al.* Local regulation of gene expression by lncRNA promoters, transcription, and splicing. *Nature* http://dx.doi.org/10.1038/nature20149 (2016).
19. Martens, J. A., Laprade, L. & Winston, F. Intergenic transcription is required to repress the *Saccharomyces cerevisiae* SER3 gene. *Nature* **429**, 571–574 (2004).
20. Melé, M. & Rinn, J. L. 'Cat's cradling' the 3D genome by the act of LncRNA transcription. *Mol. Cell* **62**, 657–664 (2016).
21. Guttman, M. & Rinn, J. L. Modular regulatory principles of large non-coding RNAs. *Nature* **482**, 339–346 (2012).
22. Rinn, J. L. & Chang, H. Y. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81**, 145–166 (2012).
23. Galupa, R. & Heard, E. X-Chromosome inactivation: new insights into *cis* and *trans* regulation. *Curr. Opin. Genet. Dev.* **31**, 57–66 (2015).
24. Wutz, A. Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. *Nat. Rev. Genet.* **12**, 542–553 (2011).
25. Brown, C. J. *et al.* A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* **349**, 38–44 (1991).
26. Brockdorff, N. *et al.* Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. *Nature* **351**, 329–331 (1991).
References 25 and 26 report the identification of the Xist RNA as a gene that is expressed exclusively from the Xi in both humans and mice.
27. 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* **71**, 515–526 (1992).
28. Brown, C. J. *et al.* The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* **71**, 527–542 (1992).
29. Clemson, C. M., McNeil, J. A., Willard, H. F. & Lawrence, J. B. XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. *J. Cell Biol.* **132**, 259–275 (1996).
This reports that the Xist RNA coats the entire territory of the Xi in the nucleus.
30. Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. & Brockdorff, N. Requirement for Xist in X chromosome inactivation. *Nature* **379**, 131–137 (1996).
This paper demonstrates a key role for the Xist RNA in XIC.
31. Marahrens, Y., Panning, B., Dausman, J., Strauss, W. & Jaenisch, R. Xist-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes Dev.* **11**, 156–166 (1997).
32. Wutz, A., Rasmussen, T. P. & Jaenisch, R. Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat. Genet.* **30**, 167–174 (2002).
This study identifies the A-repeat of Xist as essential for transcriptional silencing on the X chromosome and shows that different domains of the Xist RNA have separate functions.
33. Beletskii, A., Hong, Y. K., Pehrson, J., Egholm, M. & Strauss, W. M. PNA interference mapping demonstrates functional domains in the noncoding RNA Xist. *Proc. Natl Acad. Sci. USA* **98**, 9215–9220 (2001).
34. Lee, J. T. & Jaenisch, R. Long-range *cis* effects of ectopic X-inactivation centres on a mouse autosome. *Nature* **386**, 275–279 (1997).
35. Lee, J. T., Strauss, W. M., Dausman, J. A. & Jaenisch, R. A. 450 kb transgene displays properties of the mammalian X-inactivation center. *Cell* **86**, 83–94 (1996).
36. Cheng, J. *et al.* Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* **308**, 1149–1154 (2005).
37. Kapranov, P. *et al.* RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316**, 1484–1488 (2007).
38. Mondal, T., Rasmussen, M., Pandey, G. K., Isaksson, A. & Kanduri, C. Characterization of the RNA content of chromatin. *Genome Res.* **20**, 899–907 (2010).
39. Guttman, M. *et al.* lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477**, 295–300 (2011).
40. Sun, L. *et al.* Long noncoding RNAs regulate adipogenesis. *Proc. Natl Acad. Sci. USA* **110**, 3387–3392 (2013).
41. Alvarez-Dominguez, J. R. *et al.* Global discovery of erythroid long noncoding RNAs reveals novel regulators of red cell maturation. *Blood* **123**, 570–581 (2014).
42. Loewer, S. *et al.* Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat. Genet.* **42**, 1113–1117 (2010).
43. Leontis, N. B., Lescoute, A. & Westhof, E. The building blocks and motifs of RNA architecture. *Curr. Opin. Struct. Biol.* **16**, 279–287 (2006).
44. Keohane, A. M., Lavender, J. S., O'Neill, L. P. & Turner, B. M. Histone acetylation and X inactivation. *Dev. Genet.* **22**, 65–73 (1998).
45. Wakefield, M. J., Keohane, A. M., Turner, B. M. & Graves, J. A. Histone underacetylation is an ancient component of mammalian X chromosome inactivation. *Proc. Natl Acad. Sci. USA* **94**, 9665–9668 (1997).
46. Keohane, A. M. *et al.* X-Inactivation and histone H4 acetylation in embryonic stem cells. *Dev. Biol.* **180**, 618–630 (1996).
47. Csankovszki, G., Nagy, A. & Jaenisch, R. Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. *J. Cell Biol.* **153**, 773–784 (2001).
48. Boggs, B. A. *et al.* Differentially methylated forms of histone H3 show unique association patterns with inactive human X chromosomes. *Nat. Genet.* **30**, 73–76 (2002).
49. Plath, K. *et al.* Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300**, 131–135 (2003).
50. Silva, J. *et al.* Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. *Dev. Cell* **4**, 481–495 (2003).
51. Smeets, D. *et al.* Three-dimensional super-resolution microscopy of the inactive X chromosome territory reveals a collapse of its active nuclear compartment harboring distinct Xist RNA foci. *Epigenetics Chromatin* **7**, 8 (2014).
52. Naughton, C., Sproul, D., Hamilton, C. & Gilbert, N. Analysis of active and inactive X chromosome architecture reveals the independent organization of 30 nm and large-scale chromatin structures. *Mol. Cell* **40**, 397–409 (2010).
53. da Rocha, S. T. *et al.* Jarid2 is implicated in the initial Xist-induced targeting of PRC2 to the inactive X chromosome. *Mol. Cell* **53**, 301–316 (2014).
54. Jeon, Y. & Lee, J. T. YY1 tethers Xist RNA to the inactive X nucleation center. *Cell* **146**, 119–133 (2011).

55. Brockdorff, N. X-Chromosome inactivation: closing in on proteins that bind Xist RNA. *Trends Genet.* **18**, 352–358 (2002).
56. Chu, C. *et al.* Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404–416 (2015).
57. McHugh, C. A. *et al.* The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232–236 (2015).
58. Minajigi, A. *et al.* Chromosomes. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* **349**, 6245 (2015).
References 56–58 report the development of new biochemical methods to define the proteins that interact with the Xist lncRNA and the identification of key proteins involved in Xist-mediated silencing.
59. Monfort, A., Minin, G. Di, Postlmayr, A., Arieti, F. & Wutz, A. Identification of Spen as a crucial factor for Xist function through forward genetic screening in haploid embryonic stem cells. *Cell Rep.* **12**, 554–561 (2015).
60. Moindrot, B. *et al.* A pooled shRNA screen identifies Rbm15, Spen, and Wtap as factors required for Xist RNA-mediated silencing. *Cell Rep.* **12**, 562–572 (2015).
61. Lu, Z. *et al.* RNA duplex map in living cells reveals higher-order transcriptome structure. *Cell* **165**, 1267–1279 (2016).
62. Shi, Y. *et al.* Sharp, an inducible cofactor that integrates nuclear receptor repression and activation. *Genes Dev.* **15**, 1140–1151 (2001).
63. Guenther, M. G., Barak, O. & Lazar, M. A. The SMRT and N-CoR corepressors are activating cofactors for histone deacetylase 3. *Mol. Cell. Biol.* **21**, 6091–6101 (2001).
64. Guenther, M. G. *et al.* A core SMRT corepressor complex containing HDAC3 and TBL1, a WD40-repeat protein linked to deafness. *Genes Dev.* **14**, 1048–1057 (2000).
65. You, S. H. *et al.* Nuclear receptor co-repressors are required for the histone-deacetylase activity of HDAC3 *in vivo*. *Nat. Struct. Mol. Biol.* **20**, 182–187 (2013).
66. Plath, K., Mlynarczyk-Evans, S., Nusinow, D. A. & Panning, B. Xist RNA and the mechanism of X chromosome inactivation. *Annu. Rev. Genet.* **36**, 233–278 (2002).
67. Schoeftner, S. *et al.* Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *EMBO J.* **25**, 3110–3122 (2006).
68. Kalantry, S. & Magnuson, T. The Polycomb group protein EED is dispensable for the initiation of random X-chromosome inactivation. *PLoS Genet.* **2**, e66 (2006).
69. Kalantry, S. *et al.* The Polycomb group protein Eed protects the inactive X-chromosome from differentiation-induced reactivation. *Nat. Cell Biol.* **8**, 195–202 (2006).
70. Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J. & Lee, J. T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750–756 (2008).
71. da Rocha, S. T. *et al.* Jarid2 is implicated in the initial Xist-induced targeting of PRC2 to the inactive X chromosome. *Mol. Cell* **53**, 301–316 (2014).
72. Bintu, L. *et al.* Dynamics of epigenetic regulation at the single-cell level. *Science* **351**, 720–724 (2016).
73. Kaneko, S., Son, J., Shen, S. S., Reinberg, D. & Bonasio, R. PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1258–1264 (2013).
74. Kanhere, A. *et al.* Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. *Mol. Cell* **38**, 675–688 (2010).
75. Yap, K. L. *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* **38**, 662–674 (2010).
76. Nagano, T. *et al.* The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* **322**, 1717–1720 (2008).
77. Mayer, C., Neubert, M. & Grummt, I. The structure of NoRC-associated RNA is crucial for targeting the chromatin remodelling complex NoRC to the nucleolus. *EMBO Rep.* **9**, 774–780 (2008).
78. Tsai, M. C. *et al.* Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**, 689–693 (2010).
This study demonstrates that the HOTAIR lncRNA acts as a scaffold that brings together distinct regulatory complexes.
79. Yang, Y. W. *et al.* Essential role of lncRNA binding for WDR5 maintenance of active chromatin and embryonic stem cell pluripotency. *eLife* **3**, e02046 (2014).
80. Wang, K. C. *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120–124 (2011).
This paper identifies the lncRNA HOTTIP and shows that it activates genes in close proximity to its locus.
81. Di Ruscio, A. *et al.* DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* **503**, 371–376 (2013).
82. Prensner, J. R. *et al.* The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *Nat. Genet.* **45**, 1392–1398 (2013).
83. Ng, S. Y., Bogu, G. K., Soh, B. S. & Stanton, L. W. The long noncoding RNA RMST interacts with SOX2 to regulate neurogenesis. *Mol. Cell* **51**, 349–359 (2013).
84. Lanz, R. B. *et al.* A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. *Cell* **97**, 17–27 (1999).
85. Davidovich, C. *et al.* Toward a consensus on the binding specificity and promiscuity of PRC2 for RNA. *Mol. Cell* **57**, 552–558 (2015).
This reference and reference 73 show that PRC2 associates promiscuously with RNA in cells.
86. Rinn, J. L. *et al.* Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311–1323 (2007).
87. Lai, F. *et al.* Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* **494**, 497–501 (2013).
88. Maison, C. *et al.* Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. *Nat. Genet.* **30**, 329–334 (2002).
89. Wang, X. *et al.* Induced ncRNAs allosterically modify RNA-binding proteins in *cis* to inhibit transcription. *Nature* **454**, 126–130 (2008).
90. Kino, T., Hurt, D. E., Ichijo, T., Nader, N. & Chrousos, G. P. Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci. Signal.* **3**, ra8 (2010).
91. Yin, Q.-F. *et al.* Long noncoding RNAs with snoRNA ends. *Mol. Cell* **48**, 219–230 (2012).
92. Pandey, R. R. *et al.* Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol. Cell* **32**, 232–246 (2008).
93. Hacisuleyman, E. *et al.* Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. *Nat. Struct. Mol. Biol.* **21**, 198–206 (2014).
This study describes the Firre lncRNA, which localizes to sites on multiple chromosomes in a single nuclear compartment.
94. Hacisuleyman, E., Shukla, C. J., Weiner, C. L. & Rinn, J. L. Function and evolution of local repeats in the Firre locus. *Nat. Commun.* **7**, 11021 (2016).
95. Hasegawa, Y. *et al.* The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. *Dev. Cell* **19**, 469–476 (2010).
This paper defines SAFA as a crucial RBP that is required for tethering Xist to DNA.
96. Hall, L. L. *et al.* Stable C0T-1 repeat RNA is abundant and is associated with euchromatic interphase chromosomes. *Cell* **156**, 907–919 (2014).
97. Zappulla, D. C. & Cech, T. R. Yeast telomerase RNA: a flexible scaffold for protein subunits. *Proc. Natl Acad. Sci. USA* **101**, 10024–10029 (2004).
This paper demonstrates that the yeast TERC acts as a flexible scaffold of protein complexes.
98. Banani, S. F. *et al.* Compositional control of phase-separated cellular bodies. *Cell* **166**, 651–663 (2016).
99. Patil, D. P. *et al.* m6A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* **537**, 369–373 (2016).
100. Flynn, R. A. *et al.* 7SK-BAF axis controls pervasive transcription at enhancers. *Nat. Struct. Mol. Biol.* **23**, 231–238 (2016).
101. Helbig, R. & Fackelmayer, F. O. Scaffold attachment factor A (SAF-A) is concentrated in inactive X chromosome territories through its RGG domain. *Chromosoma* **112**, 173–182 (2003).
102. Soruco, M. M. L. *et al.* The CLAMP protein links the MSL complex to the X chromosome during *Drosophila* dosage compensation. *Genes Dev.* **27**, 1551–1556 (2013).
103. Meller, V. H. *et al.* Ordered assembly of roX RNAs into MSL complexes on the dosage-compensated X chromosome in *Drosophila*. *Curr. Biol.* **10**, 136–143 (2000).
104. Quinn, J. J. *et al.* Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification. *Nat. Biotechnol.* **32**, 933–940 (2014).
105. Simon, M. D. *et al.* The genomic binding sites of a noncoding RNA. *Proc. Natl Acad. Sci. USA* **108**, 20497–20502 (2011).
106. Quinn, J. J. & Chang, H. Y. *In situ* dissection of RNA functional subunits by domain-specific chromatin isolation by RNA purification (dChIRP). *Methods Mol. Biol.* **1262**, 199–213 (2015).
107. Engreitz, J. M. *et al.* RNA–RNA interactions enable specific targeting of noncoding RNAs to nascent pre-mRNAs and chromatin sites. *Cell* **159**, 188–199 (2014).
108. West, J. A. *et al.* The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Mol. Cell* **55**, 791–802 (2014).
109. Engreitz, J. M. *et al.* The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973 (2013).
This study maps the genomic localization of the Xist lncRNA at high resolution upon initiation of XIC and shows that it spreads by using the 3D structure of the X chromosome.
110. Simon, M. D. *et al.* High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. *Nature* **504**, 465–469 (2013).
111. Cremer, T. & Cremer, C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* **2**, 292–301 (2001).
112. Maamar, H., Cabili, M. N., Rinn, J. & Raj, A. linc-HOXA1 is a noncoding RNA that represses Hoxa1 transcription in cis. *Genes Dev.* **27**, 1260–1271 (2013).
113. Mohammad, F. *et al.* Kcnq1ot1/Lit1 noncoding RNA mediates transcriptional silencing by targeting to the perinucleolar region. *Mol. Cell. Biol.* **28**, 3713–3728 (2008).
114. Shechner, D. M., Hacisuleyman, E., Younger, S. T. & Rinn, J. L. Multiplexable, locus-specific targeting of long RNAs with CRISPR-display. *Nat. Methods* **12**, 664–670 (2015).
115. Dekker, J. & Misteli, T. Long-range chromatin interactions. *Cold Spring Harb. Perspect. Biol.* **7**, a019356 (2015).
116. Maass, P. G. *et al.* A misplaced lncRNA causes brachydactyly in humans. *J. Clin. Invest.* **122**, 3990–4002 (2012).
117. Ramírez, F. *et al.* High-affinity sites form an interaction network to facilitate spreading of the MSL complex across the X chromosome in *Drosophila*. *Mol. Cell* **60**, 146–162 (2015).
118. Sunwoo, H., Wu, J. Y. & Lee, J. T. The Xist RNA-PRC2 complex at 20-nm resolution reveals a low Xist stoichiometry and suggests a hit-and-run mechanism in mouse cells. *Proc. Natl Acad. Sci. USA* **112**, E4216–E4225 (2015).
119. Dodd, D. W., Gagnon, K. T. & Corey, D. R. Digital quantitation of potential therapeutic target RNAs. *Nucleic Acid. Ther.* **23**, 188–194 (2013).
120. Tani, H., Nakamura, Y., Ijiri, K. & Akimitsu, N. Stability of MALAT-1, a nuclear long non-coding RNA in mammalian cells, varies in various cancer cells. *Drug Discov. Ther.* **4**, 235–239 (2010).
121. Brawerman, G. & Diez, J. Metabolism of the polyadenylate sequence of nuclear RNA and messenger RNA in mammalian cells. *Cell* **5**, 271–280 (1975).
122. Huang, S., Deerinck, T. J., Ellisman, M. H. & Spector, D. L. *In vivo* analysis of the stability and transport of nuclear poly(A)$^+$ RNA. *J. Cell Biol.* **126**, 877–899 (1994).
123. He, D. C., Nickerson, J. A. & Penman, S. Core filaments of the nuclear matrix. *J. Cell Biol.* **110**, 569–580 (1990).
124. Nickerson, J. A., Krochmalnic, G., Wan, K. M. & Penman, S. Chromatin architecture and nuclear RNA. *Proc. Natl Acad. Sci. USA* **86**, 177–181 (1989).
125. Splinter, E. *et al.* The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. *Genes Dev.* **25**, 1371–1383 (2011).
This paper identifies large-scale structural changes on the Xi that depend on the Xist RNA but are independent of its silencing function.

126. Rao, S. S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
This study maps the 3D architecture of the genome at 1 kb resolution and identifies unique architectural features of the Xi, including superloops.
127. Deng, X. *et al.* Bipartite structure of the inactive mouse X chromosome. *Genome Biol.* **16**, 152 (2015).
128. Darrow, E. M. *et al.* Deletion of *DXZ4* on the human inactive X chromosome alters higher-order genome architecture. *Proc. Natl Acad. Sci. USA* **113**, E4504–E4512 (2016).
129. Giorgetti, L. *et al.* Structural organization of the inactive X chromosome in the mouse. *Nature* **535**, 575–579 (2016).
130. Rego, A., Sinclair, P. B., Tao, W., Kireev, I. & Belmont, A. S. The facultative heterochromatin of the inactive X chromosome has a distinctive condensed ultrastructure. *J. Cell Sci.* **121**, 1119–1127 (2008).
131. Shevelyov, Y. Y. & Nurminsky, D. I. The nuclear lamina as a gene-silencing hub. *Curr. Issues Mol. Biol.* **14**, 27–38 (2012).
132. Amendola, M. & van Steensel, B. Mechanisms and dynamics of nuclear lamina–genome interactions. *Curr. Opin. Cell Biol.* **28**, 61–68 (2014).
133. Worman, H. J., Yuan, J., Blobel, G. & Georgatos, S. D. A lamin B receptor in the nuclear envelope. *Proc. Natl Acad. Sci. USA* **85**, 8531–8534 (1988).
134. Gruenbaum, Y., Margalit, A., Goldman, R. D., Shumaker, D. K. & Wilson, K. L. The nuclear lamina comes of age. *Nat. Rev. Mol. Cell. Biol.* **6**, 21–31 (2005).
135. Burke, B. & Stewart, C. L. The nuclear lamins: flexibility in function. *Nat. Rev. Mol. Cell. Biol.* **14**, 13–24 (2013).
136. Chen, C. K. *et al.* Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. *Science* http://dx.doi.org/10.1126/science.aae0047 (2016).
137. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
138. Figueroa, D. M., Darrow, E. M. & Chadwick, B. P. Two novel DXZ4-associated long noncoding RNAs show developmental changes in expression coincident with heterochromatin formation at the human (*Homo sapiens*) macrosatellite repeat. *Chromosome Res.* **23**, 733–752 (2015).
139. 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* **54**, 485–499 (2014).
140. Cheutin, T. & Cavalli, G. Polycomb silencing: from linear chromatin domains to 3D chromosome folding. *Curr. Opin. Genet. Dev.* **25C**, 30–37 (2014).
141. 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* **13**, 602–616 (2013).
142. Schoenfelder, S. *et al.* Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. *Nat. Genet.* **47**, 1179–1186 (2015).
143. Nozawa, R. S. *et al.* Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HBiX1 pathway. *Nat. Struct. Mol. Biol.* **20**, 566–573 (2013).
144. Bergmann, J. H. *et al.* Regulation of the ESC transcriptome by nuclear long noncoding RNAs. *Genome Res.* **25**, 1336–1346 (2015).
145. Biamonti, G. & Vourc'h, C. Nuclear stress bodies. *Cold Spring Harb. Perspect. Biol.* **2**, a000695 (2010).
146. Caudron-Herger, M. *et al.* Alu element-containing RNAs maintain nucleolar structure and function. *EMBO J.* **34**, 2758–2774 (2015).
147. Jacob, M. D., Audas, T. E., Uniacke, J., Trinkle-Mulcahy, L. & Lee, S. Environmental cues induce a long noncoding RNA-dependent remodeling of the nucleolus. *Mol. Biol. Cell* **24**, 2943–2953 (2013).
148. Prasanth, K. V. *et al.* Regulating gene expression through RNA nuclear retention. *Cell* **123**, 249–263 (2005).
149. Clemson, C. M. *et al.* An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol. Cell* **33**, 717–726 (2009).
150. Sunwoo, H. *et al.* MEN ε/β nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res.* **19**, 347–359 (2009).
151. Imamura, K. *et al.* Long noncoding RNA NEAT1-dependent SFPQ relocation from promoter region to paraspeckle mediates IL8 expression upon immune stimuli. *Mol. Cell* **53**, 393–406 (2014).
152. Chen, L. L. & Carmichael, G. G. Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. *Mol. Cell* **35**, 467–478 (2009).
153. Shevtsov, S. P. & Dundr, M. Nucleation of nuclear bodies by RNA. *Nat. Cell Biol.* **13**, 167–173 (2011).
154. Mao, Y. S., Sunwoo, H., Zhang, B. & Spector, D. L. Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. *Nat. Cell Biol.* **13**, 95–101 (2011).
References 149, 150, 153 and 154 show that NEAT1 is essential for establishing and maintaining the paraspeckle nuclear body and is sufficient to nucleate the formation of a paraspeckle.
155. Lamond, A. I. & Spector, D. L. Nuclear speckles: a model for nuclear organelles. *Nat. Rev. Mol. Cell Biol.* **4**, 605–612 (2003).
156. Spector, D. L. & Lamond, A. I. Nuclear speckles. *Cold Spring Harb. Perspect. Biol.* **3**, a000646 (2011).
157. Hall, L. L., Smith, K. P., Byron, M. & Lawrence, J. B. Molecular anatomy of a speckle. *Anat. Rec. A. Discov. Mol. Cell. Evol. Biol.* **288**, 664–675 (2006).
158. Khanna, N., Hu, Y. & Belmont, A. S. HSP70 transgene directed motion to nuclear speckles facilitates heat shock activation. *Curr. Biol.* **24**, 1138–1144 (2014).
159. Änkö, M.-L. *et al.* The RNA-binding landscapes of two SR proteins reveal unique functions and binding to diverse RNA classes. *Genome Biol.* **13**, R17 (2012).
160. Miyagawa, R. *et al.* Identification of *cis-* and *trans-*acting factors involved in the localization of MALAT-1 noncoding RNA to nuclear speckles. *RNA* **18**, 738–751 (2012).
161. Sanford, J. R. *et al.* Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Res.* **19**, 381–394 (2009).
162. Tripathi, V. *et al.* The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell* **39**, 925–938 (2010).
163. Bernard, D. *et al.* A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. *EMBO J.* **29**, 3082–3093 (2010).
164. Marshall, W. F., Marko, J. F., Agard, D. A. & Sedat, J. W. Chromosome elasticity and mitotic polar ejection force measured in living *Drosophila* embryos by four-dimensional microscopy-based motion analysis. *Curr. Biol.* **11**, 569–578 (2001).
165. Chubb, J. R., Boyle, S., Perry, P. & Bickmore, W. A. Chromatin motion is constrained by association with nuclear compartments in human cells. *Curr. Biol.* **12**, 439–445 (2002).
166. Kwon, I. *et al.* Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell* **155**, 1049–1060 (2013).
167. Fackelmayer, F. O., Dahm, K., Renz, A., Ramsperger, U. & Richter, A. Nucleic-acid-binding properties of hnRNP-U/SAF-A, a nuclear-matrix protein which binds DNA and RNA *in vivo* and *in vitro*. *Eur. J. Biochem.* **221**, 749–757 (1994).
168. Hyman, A. A., Weber, C. A. & Jülicher, F. Liquid–liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol.* **30**, 39–58 (2014).
169. Feric, M. *et al.* Coexisting liquid phases underlie nucleolar subcompartments. *Cell* **165**, 1686–1697 (2016).
170. Kato, M. *et al.* Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753–767 (2012).
This paper demonstrates that many RBPs contain low-complexity domains that undergo concentration-dependent phase transition, which may underlie the formation of some nuclear compartments.
171. Xiang, S. *et al.* The LC domain of hnRNPA2 adopts similar conformations in hydrogel polymers, liquid-like droplets, and nuclei. *Cell* **163**, 829–839 (2015).
172. Rocha, E. P. C. The organization of the bacterial genome. *Annu. Rev. Genet.* **42**, 211–233 (2008).
173. Grant, J. *et al.* Rsx is a metatherian RNA with Xist-like properties in X-chromosome inactivation. *Nature* **487**, 254–258 (2012).
174. Bartolomei, M. S. & Ferguson-Smith, A. C. Mammalian genomic imprinting. *Cold Spring Harb. Perspect. Biol.* **3**, 1–17 (2011).
175. Kind, J. *et al.* Genome-wide maps of nuclear lamina interactions in single human cells. *Cell* **163**, 134–147 (2015).
176. Melese, T. & Xue, Z. The nucleolus: an organelle formed by the act of building a ribosome. *Curr. Opin. Cell Biol.* **7**, 319–324 (1995).
177. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).

Acknowledgements
We thank members of the Guttman laboratory, especially Sofia Quinodoz, for helpful discussions, and Sigrid Knemeyer for help with the figures. J.M.E. is supported by the Fannie and John Hertz Foundation and the Broad Institute. N.O. is supported by a postdoctoral fellowship from the California Institute of Technology. M.G. is a New York Stem Cell Foundation Robertson Investigator, an investigator at the Heritage Medical Research Institute, a Searle Scholar, a Pew-Steward scholar and an Alfred P. Sloan fellow. Research in the Guttman laboratory is funded by the National Institutes of Health (NIH) 4DN programme, an NIH Director's Early Independence Award, the New York Stem Cell Foundation, the Edward Mallinckrodt Foundation, Sontag Foundation, Searle Scholars Program, Pew-Steward Scholars programme and funds from the California Institute of Technology.

Competing interests statement
The authors declare no competing interests.

SUPPLEMENTARY INFORMATION
See online article: S1 (table)

ALL LINKS ARE ACTIVE IN THE ONLINE PDFF