CHAPTER 11

LONG NONCODING RNA AND EPIGENOMICS

Chandrasekhar Kanduri
Department of Genetics and Pathology, Uppsala University, Uppsala, Sweden
Email: kanduri.chandrasekhar@genpat.uu.se

Abstract: Accumulating evidence over the last decade has presented us with the intriguing observation that the majority of eukaryotic genomes are pervasively transcribed to encode a complex network of small and long noncoding RNAs. Long noncoding RNAs are of particular interest, as they were once thought to be restricted to housekeeping functions and are now linked to a wide variety of biological functions related to physiology, embryology and development. Emerging evidence indicates that a subset of long noncoding RNAs mediate their biological functions by using chromatin as a substrate, to index the genetic information encoded in the genome. This chapter will discuss how noncoding RNAs and the processes underlying their transcription mediate transcriptional regulation, by epigenetically regulating the structure of chromatin in various biological contexts.

INTRODUCTION

Until recently the noncoding portion of the genome in higher eukaryotes was considered "junk" and was thought to have evolved as a hot spot for mutations that permitted the coding portion of the genome to be highly conserved across the evolutionary spectrum. This view led to the belief that the protein-coding portion of the genome acted as the main architect of organismal development by controlling genetic programming. However, the advent of new high-throughput technologies such as high resolution tiling arrays and RNA sequencing (which can probe the transcriptional landscape of the entire genome with relative ease), has led to the realization that the protein-coding portion of the genome remains relatively constant across different eukaryotic species and that it is only the noncoding portion of the genome that has evolved, thus linking the noncoding

RNA Infrastructure and Networks, edited by Lesley J. Collins.

portion of the genome to eukaryotic complexity.[1] Recent evidence indicates that although more than 70% of the eukaryotic genome is transcribed, only approximately 1 to 2 % of the transcriptome contributes to protein-coding RNA, suggesting that transcription is not just limited to the protein-coding portion of the eukaryotic genome but is pervasive throughout it.[2,3] Several recent analyses have documented that widely transcribed noncoding RNAs (ncRNAs) play a critical role in various biological functions linked to development and differentiation.[4-9] Importantly, the majority of ncRNAs are expressed in a spatio-temporal manner and often exhibit precise sub-cellular localization. These observations lend support to the contention that transcription of the noncoding portion of the genome contributes to the evolution of complex organisms. However, the functional role of alternative splice forms of protein-coding mRNAs and posttranslational modification of proteins, which increase the diversity of their functions, have not been ruled out as also playing a role in the evolution of complex organisms.

Based on their size, ncRNAs can be classified as small, medium and long. Small ncRNAs range from 18-31 nucleotides in length, whereas long ncRNAs range in size from 200 bp to over several hundred kb. Medium length ncRNAs, measure between 31 and 200 mucleotides, contains of mainly snRNAs and snoRNAs. Unlike small and medium sized ncRNAs, long ncRNAs are not highly conserved at the primary sequence level. However, a subset of long intergenic ncRNAs (linc RNAs) identified using chromatin signatures, have been shown to be highly conserved at the sequence level.[10] Small ncRNAs such as small interfering RNAs (siRNAs), microRNAs (miRNAs) and Piwi-interacting RNAs (piRNAs) regulate gene activity at the transcriptional and/or posttranscriptional level through distinct mechanisms via the RNA interference (RNAi) pathway. miRNAs are 21-23-bp single-stranded RNA molecules that serve as posttranscriptional regulators of gene expression in plants and animals. They act by binding to complementary sites on target RNAs to induce cleavage or repression of productive translation. In contrast, siRNAs are double-stranded RNA molecules 20-25 nucleotides in length which have been shown to participate in numerous biological functions such as antiviral mechanisms and chromosome segregation by maintaining the integrity of centromeric heterochromatin.[11] Piwi-interacting RNAs or piRNAs are poorly conserved at the primary sequence level and are longer than both siRNA and miRNA, measuring in size between 26-31 nucleotides. piRNAs are implicated in transcriptional gene silencing of repeat elements such as of retrotransposons in germ line cells via methylating their promoters.[12,13]

Long ncRNAs were once thought to be restricted to housekeeping functions such as DNA replication, posttranscriptional processing and protein synthesis. However, increased focus on the functions of long ncRNA in eukaryotes resulted in the identification of a broad functional repertoire of ncRNAs that take part in important biological functions beyond the realm of housekeeping functions, such as X chromosome inactivation,[14,15] genomic imprinting,[16] sub-cellular structural organization,[17,18] telomere[19] and centromere organization[20,21] and nuclear trafficking.[22] Interestingly, accumulated evidence over the last few years suggests that the majority of functionally characterized ncRNAs act at the level of transcription as *cis* or *trans* elements by modulating chromatin structure or transcriptional programs.[6,23] In certain instances ncRNAs themselves have been observed to act as regulators, whereas in other instances their transcription and/or the RNA itself plays a critical role, indicating that no common theme applies to ncRNA-mediated functions.[16,24,25]

Relatively few ncRNAs have been studied so far, but since more than 70% (70% in nematode and 93% in humans) of the eukaryotic genome is transcribed, thousands of ncRNAs are likely to interact in a complex network. Although many of the functionally

implicated ncRNAs, as well as their target genes, are poorly conserved at the primary sequence level, there seem to be striking parallels in their mechanisms of action. For example, despite lacking homology at the sequence level, a subset of long ncRNAs mediates gene regulation of chromosomal domains[16,26] or chromosomes[14] at the transcriptional level by targeting chromatin remodeling complexes as *cis* or *trans* elements. This mechanism has been described in relatively well-characterized biological phenomena such as genomic imprinting and X-chromosome inactivation.

The act of ncRNA transcription (sense or antisense) is also sufficient to regulate the expression of neighboring genes. Contrary to widespread belief, ncRNA transcription does not always result in the silencing of flanking genes. In certain instances, it activates transcription.[27] How does ncRNA transcription control two distinctly regulated processes such as gene activation and silencing? The prevailing view has been that ncRNA transcription across the promoter region of a downstream protein-coding gene directly interferes with the transcription initiation complex, thus silencing the protein-coding gene. Interestingly, ncRNA transcription through the promoter regions of certain tumor suppressor genes in human disease conditions[28] and *GAL1-10* clusters in yeast[29] results in heterochromatinization and gene silencing. This gene silencing does not depend on the polarity of transcription, as it is induced by both sense[30] and antisense[31] ncRNA transcription, indicating that the act of transcription epigenetically regulates gene silencing by interfering with chromatin structure. In contrast, continuous ncRNA transcription through homeobox gene clusters in flies has been linked to homeobox gene activation by interfering with the binding of polycomb (PcG) proteins.[32] Similarly, ncRNA transcription upstream of the fbp1$^+$ gene in yeast induces chromatin remodeling and gene activation,[33] indicating that ncRNA transcription mediates transcriptional gene silencing or activation through chromatin-level regulation in a context-dependent manner. It remains unknown how the act of transcription specifically modifies the chromatin structure around the promoter region. Though the act of ncRNA transcription elicits transcriptional regulation through multiple pathways, transcriptional interference could be a common theme in these pathways.

Here, I focus on a network of RNA infrastructural ncRNAs that mediate gene regulation by epigenetically regulating chromatin structure in various biological contexts from fission yeast *S. pombe*, *Drosophila melanogaster*, mouse and human.

LONG ncRNAs AND TUMOR SUPPRESSOR GENES

DNA methylation of tumor suppressor gene promoters in the context of global hypomethylation is a common feature of several cancers. The mechanisms underlying the epigenetic silencing of tumor suppressor genes by CpG methylation are unknown, as loss of function mutations affecting the CpG methylation machinery have not been detected in the majority of cancers. Emerging evidence[28,34] from recent investigations implicates a functional role for long ncRNAs in epigenetic silencing of the tumor suppressor genes *p15* and *p21*. Both *p15* and *p21* genes are overlapped by antisense ncRNAs and ncRNA transcription through the *p15* and *p21* promoters is correlated with DNA methylation in *cis*. Although both the promoters become methylated in response to antisense ncRNA transcription, the available evidence to date suggests that the underlying mechanisms leading to their methylation differ.

In the case of the *p15* tumor suppressor gene, increased antisense ncRNA transcription was specifically detected in patients with acute myeloid leukemia and acute lymphatic

leukemia but not from normal individuals, indicating that it is aberrantly transcribed only in disease conditions.[28] Aberrant antisense transcription is associated with *p15* promoter methylation and gene silencing, indicating that the antisense transcription process and/or its RNA product epigenetically silences the *p15* promoter. By employing in vitro strategies, the antisense RNA was found to silence the *p15* promoter in *cis* or in *trans* and to be involved in the establishment but not in the maintenance of silencing, which persisted even in the absence of the antisense RNA. Interestingly, the dsRNA-mediated RNAi pathway was found not to play a functional role in the *p15* antisense RNA-mediated epigenetic silencing, indicating that the antisense RNA uses currently unknown mechanisms in its epigenetic silencing pathway. Since the in vitro system utilizes only a portion of the antisense RNA sequence, it is not clear whether the in vitro synthesized RNA is fully representative of the functions of the native antisense RNA.

In the case of *p21*, balanced bidirectional transcription from the *p21* sense and antisense promoters seems to play an important role in maintaining the activity of the *p21* promoter, as the silencing of either the *p21* sense or antisense RNA transcript affects the transcription level of the other.[34] For example, down regulation of the sense RNA leads to increased expression of its antisense RNA, followed by silencing of the *p21* promoter by epigenetic modification of chromatin, indicating that the antisense RNA, or its transcription, suppresses *p21* promoter activity by directing silent-state epigenetic marks. Interestingly, epigenetic modification of the *p21* promoter by the antisense RNA is dependent on Argonaute-1 protein, a key player in the small RNA induced gene silencing pathway, suggesting a functional role for the RNAi pathway in the antisense RNA-mediated epigenetic silencing of the *p21* promoter. Like in the majority of promoters, the *p21* promoter is also associated with promoter-associated RNAs (pRNAs) in both the sense and antisense directions. The appearance of the antisense pRNA is dependent on Argonaute-1, indicating a strong link between the *p21* antisense RNA and the Argonaute-1 protein in the generation of the antisense pRNA, which could play an important role in the epigenetic modification of the *p21* promoter. However, it remains to be seen whether the Argonaute-1-dependent silencing of the *p21* promoter has any resemblance to RNAi-mediated heterochromatinization in eukaryotes such as fission yeast *S. pombe*. The above studies clearly demonstrate that the *p21* and *p15* antisense RNAs exploit different strategies in executing silencing of their sense promoters.

A functional role for RNA *per se* is evident in the case of *p15* silencing through an RNAi-independent pathway. However in the case of *p21*, the functional role played by the act of ncRNA transcription or the ncRNA itself, as well as the possible existence of a functional link between ncRNA and RNAi components, have not yet been determined.

LONG ncRNAs AND HOMEOBOX GENES

Homeobox genes (*Hox*) genes (which control embryonic body development in all bilaterally symmetrical animals), are organized into gene clusters and are coordinately regulated by common long-range *cis* acting regulatory elements. Epigenetic mechanisms, involving histone methylation, demethylation and ncRNAs, have recently been shown to be critical for the fidelity of *Hox* gene expression in a spatio-temporal manner along the developmental axis. Epigenetic regulation of *Hox* gene expression occurs primarily due to the interplay between two protein complexes: trithorax- and polycomb-group proteins. Trithorax-group proteins, which possess enzymatic activity to mediate the

methylation of the lysine 4 residue of histone H3 (H3K4), are required to maintain *Hox* gene activation, whereas polycomb-group proteins, comprising PRC2 complex members, maintain the repressive state of the *Hox* genes by mediating the methylation of histone H3 at lysine 27 (H3K27). Although *cis* acting DNA sequences (polycomb responsive elements, PRE) that recruit polycomb/trithorax proteins have been well-characterized in *Drosophila melanogaster*, in other organisms the recruitment of these proteins to the *Hox* gene cluster in a spatio-temporal manner is poorly understood.[35,36] The interplay between polycomb and trithorax proteins generates active or inactive chromatin domains in the *Hox* gene clusters along the developmental axis. Evidence from the last few years indicates that noncoding transcripts are prevalent among the *Hox* gene clusters, interweaving with genic and intergenic regions and that they have a key role in configuring this epigenetic landscape.[26,37]

Although ncRNAs were detected over two decades ago in *Drosophila Hox* gene clusters, their mode of action in regulating the spatio-temporal expression of *Hox* genes remained obscure. However, recent studies addressing the role of ncRNA in *Hox* gene regulation have revealed that ncRNA itself or its transcription, affect *Hox* gene activation and silencing through multiple pathways. The functional role of ncRNAs has been thoroughly characterized in the *Drosophila* bithorax complex (*BX-C*) region, which is implicated in two-thirds of the body plan. Two recent reports investigated the functional role of short processed transcripts from the *bxd* region, which flanks the Ultrabithorax (*Ubx*) *Hox* gene.[38,39] Both studies implicate *bxd* ncRNA as playing a functional role in the regulation of *Ubx* expression, but they contradict each other regarding how the regulation is brought about. One report suggests that *bxd* ncRNAs encoded from polycomb/trithorax elements upstream of the *Ubx* gene are retained at the site of transcription through sequence homology and this sequence-based interaction is required for the RNA mediated recruitment of trithorax protein ASH1 to the downstream *Ubx* gene promoter.[38] The RNA-mediated recruitment of ASH1 results in the formation of active chromatin around the *Ubx* promoter, thus making it more accessible to transcription initiation machinery. Interestingly, this study further shows that ectopic expression of *bxd* transcripts in larval imaginal discs increases the activation of the *Ubx* promoter, suggesting that the *bxd* transcript itself and not just its transcription, is crucial for *Ubx* promoter activation (Fig. 1A). In contrast, the other report suggests that ncRNA transcription from the *bxd* region extends into the *Ubx* promoter region, which negatively regulates *Ubx* promoter activity.[39] Importantly, ectopic expression of the regions that span the *bxd* transcripts do not activate *Ubx* transcription, suggesting that the act of transcription interferes with the initiation of transcription at the *Ubx* promoter (Fig. 1B). From these investigations, the mode of action of *bxd* transcripts in *Ubx* gene regulation is difficult to infer, but nevertheless, both studies clearly implicate a functional role for ncRNAs.

Similarly, several ncRNAs have been mapped to the genic and intergenic regions of the mouse and human *Hox* gene clusters.[26,37] In many cases their expression correlated with the active status of the neighboring *Hox* gene, indicating that *Hox* ncRNAs in mammals probably mediate the activation of flanking genes. However, a detailed examination of one of the human *Hox* ncRNAs, *HOTAIR* encoded from the *HOXC* locus, revealed that it regulates gene expression of the *HOXD* locus *in trans* by targeting one of the PRC2 complex members, Ezh2.[26] Although *trans* function of small ncRNAs such as microRNAs is mechanistically well defined, the *trans* function of long ncRNAs is largely unknown. Thus, this observation is of considerable importance given the fact that *HOTAIR* mediates gene silencing of the entire *HOXD* locus *in trans* despite a lack of sequence homology with

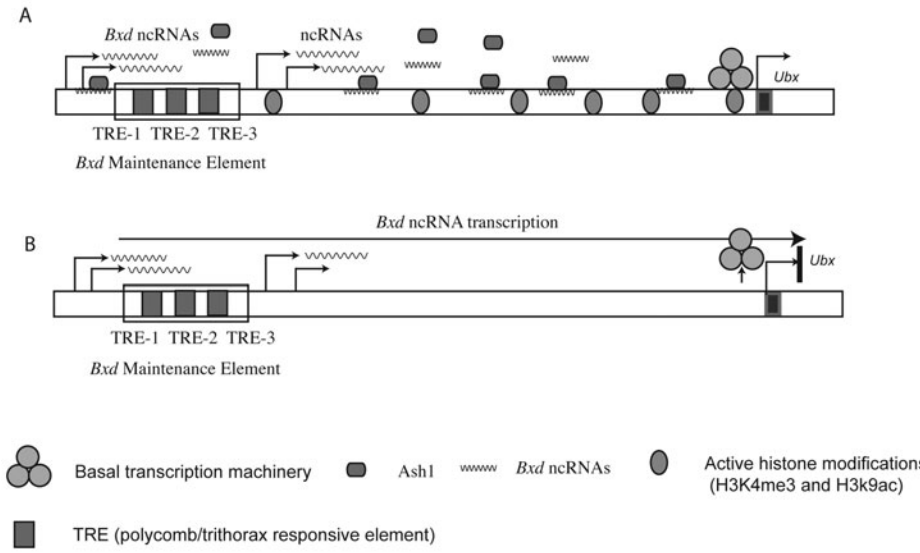

Figure 1. Models explaining the functional role of long ncRNAs from the *Bxd* region on the flanking *Ubx* gene expression. A) *Bxd* ncRNAs are retained at the *Ubx* locus via RNA-DNA interactions and recruit trithorax protein ASH1. The RNA dependent recruitment of ASH1 facilitates the formation of active chromatin structure and *Ubx* transcription. B) This model explores a functional role for noncoding transcription, rather than ncRNA per se, in the control of *Ubx* gene expression. According to this model, *Bxd* ncRNA transcription through the 5′ regulatory elements of the *Ubx* promoter results in the transcriptional repression of the *Ubx* gene via transcriptional interference or promoter occlusion.

the target sequences. The *trans* function of *HOTAIR* ncRNA mirrors the trans-activation of the *Ubx* promoter by *bxd* transcripts through the targeting of the trithorax protein ASH1.

Likewise, a recent study in mice has functionally linked two long antisense ncRNAs, *Evx1a*s and *Hoxb5/6as*, to *Hox* gene regulation.[40] The *Evx1as* and *Hoxb5/6as* ncRNAs showed concordant expression with their sense partners *Evx1* and *Hox5/6*, respectively, during the primitive streak phase of EB differentiation. These two ncRNAs were enriched in H3K4me3-precipitated chromatin. Moreover, both of these ncRNAs have been shown to interact with MLL1 histone methyltransferase, which is responsible for the establishment of H3K4me3, indicating that these antisense ncRNAs activate their associated sense protein-coding genes by establishing active chromatin structures.

Taken together, the above observations from various organisms imply that homeobox gene regulation by long ncRNAs is a conserved transcriptional regulatory mechanism across the evolutionary ladder, involving ncRNA and/or the act of its transcription.

LONG ncRNAs AND DOSAGE COMPENSATION

Dosage compensation is an epigenetic process that ensures equal dosage of X-linked gene products between male and female organisms and it is achieved by either hyper-activation or inactivation of the X chromosome. In *Drosophila*, in which, like mammals, males have one X chromosome and females have two X chromosomes,

equal levels of X-linked gene products are achieved between males and females through hyper-activation of the X chromosome in males.[41,42] In mammals, the equal dosage of X-linked gene products between males and females is achieved through inactivation of one of the two X chromosomes in females during early embryonic development.[6,43,44] As *Drosophila* and mammals are distantly related organisms, it is interesting to note that the dosage compensation process is brought about by ncRNAs in both species, while the underlying mechanisms seem to differ significantly. In mammals, X-chromosome inactivation has been investigated in depth using the mouse as a model system. In XX female mammals, genes on one of the X chromosomes undergo transcriptional silencing in a chromosome-wide manner during early embryonic development, which occurs in nonrandom or random fashion and in a stage- and tissue-specific manner. Nonrandom or imprinted X-chromosome inactivation (the paternal chromosome is always selected for inactivation) occurs in very early preimplantation embryos and extra-embryonic tissues, whereas random X-chromosome inactivation (where both parental chromosomes have an equal probability of being inactivated) is restricted to the epiblast lineage i.e., the embryo.[45]

A 500-kb complex locus known as the X-inactivation center (*Xic*) is implicated in both random and nonrandom X-chromosome inactivation pathways (Fig. 2A). This region is replete with genes that transcribe ncRNAs and a few protein-coding RNAs. The ncRNA genes, which span a region of approximately 100-200 kb, control most of the X-chromosome inactivation pathways.[14] Five ncRNA genes have been identified in the *Xic*: *Xist*, *Tsix*, *Xite*, *RepA* and *DXPas34*. *Xist* is 15-17 kb in length and transcribed exclusively from the future inactive X chromosome.[46] *Tsix*, the antisense partner of *Xist*, is transcribed from the 3′ end of the *Xist* gene and covers the entire *Xist* coding region as well as its promoter region. It is about 40 kb in length and exclusively expressed by the future active X chromosome (Fig. 2C).[47] *Xite*, an upstream activator sequence, is located 10 to 15 kb upstream of the *Tsix* promoter and encodes several ncRNAs.[48] *RepA* is a 1.6-kb transcript encoded from the Repeat A region of *Xist. RepA* selectively transactivates the *Xist* promoter on the future inactive X chromosome (Fig. 2B).[49] A microsatellite repeat element, *DXPas34*, at the 5′ end of the *Tsix* transcript encodes short bidirectional transcripts which have been shown to regulate *Tsix* transcription.[50,51] In addition, bidirectional transcription over the *Xist* locus leads to the generation of several 25-45-bp small RNAs (xiRNAs).[52] Below, the functional interplay between these ncRNAs at various stages of the X-inactivation process, resulting in chromosomal asymmetry leading to active and inactive X chromosomes, will be discussed in greater detail (see Chapter 1, Figure 5, page 12 for a summary of XCI mechanism).

In the mouse, initiation of X-chromosome inactivation coincides with the onset of *Xist* expression. *Xist* expression is first detected on the paternal X chromosome at the two-cell stage of embryogenesis and in subsequent stages of embryogenesis *Xist* RNA spreads from its site of synthesis and blankets the entire X chromosome in cis (Fig. 2B).[53] The paternal-specific *Xist* expression coincides with transcriptional gene silencing and accumulation of heterochromatic marks along the paternal X chromosome, indicating a link between *Xist* expression and imprinted X inactivation. However, the functional role of *Xist* in the establishment of imprinted of X chromosome inactivation is not very clear, as a recent investigation documented that transcriptional silencing of the paternal chromosome occurs independent of *Xist*, because the genes on paternal chromosomes lacking the *Xist* gene were still repressed.[54] However, the maintenance of the imprinted form of X-chromosome inactivation in extra-embryonic tissues requires the presence of

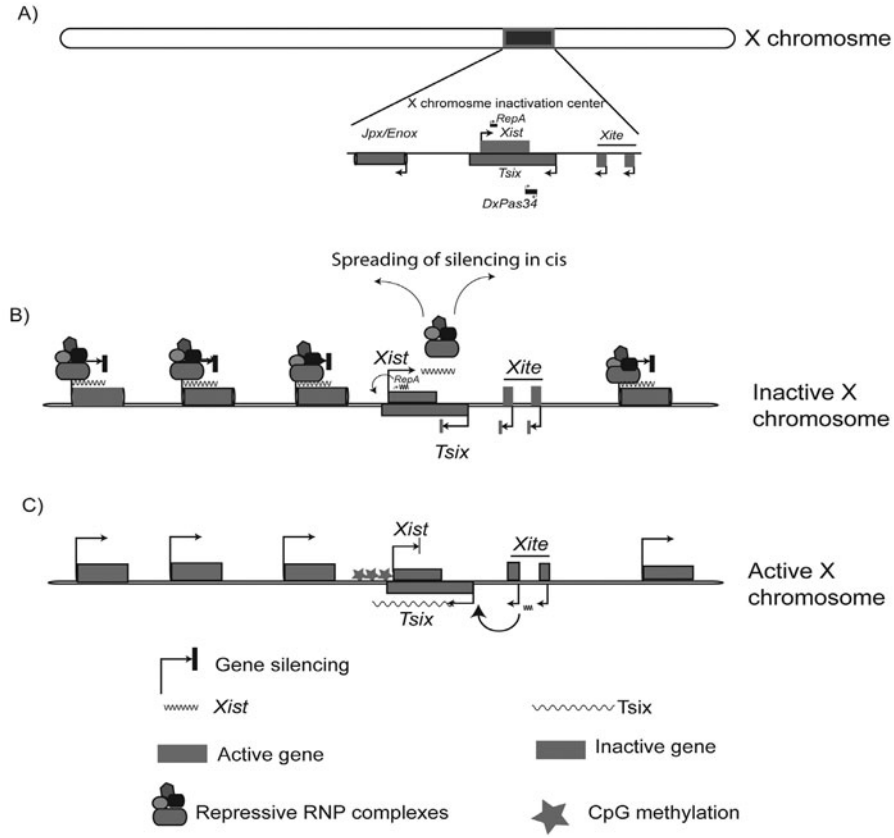

Figure 2. A) Doage compensation in mammals. X chromosome showing the X inactivation center (*Xic*). B) Inactive X chromosome showing *Xist* expression and its accumulation along the X chromsome. C) Active X chromosome showing *Tsix* activation, *Xist* inactivation and gene activation along the future active X chromosome. Maps are not drawn to scale.

Xist. In the embryonic lineage, the paternally inactivated X chromosome is reactivated and one of the parental chromosomes is randomly chosen for inactivation.

The random X inactivation process has been investigated using differentiating mouse ES cells as a model system, which faithfully recapitulates most of the X chromosome inactivation steps that occur during postimplantation development. In female ES cells prior to X-chromosome inactivation, *Tsix* is expressed at high levels and *Xist* is expressed at low levels on both chromosomes. Pluripotent transcription factors such as *Oct4*, *Nanog* and *Sox2* have been implicated in the repression of *Xist*, but it is not clear how *Tsix* transcript levels are maintained at higher levels in cells prior to X-chromosome inactivation.[55] Abundant *Tsix* expression across the *Xist* locus is correlated with establishment of active chromatin over the *Xist* gene body, which is thought to be involved in resetting the epigenetic marks prior to the onset of random X inactivation.[56] In addition, a recent investigation has demonstrated that *Tsix* transcription in cells prior to X-chromosome activation helps to limit the spread of heterochromatin

into the *Xist* promoter region from the so-called heterochromatic hot spot, which is located further upstream of the *Xist* promoter (Fig. 2), indicating that *Tsix* transcription performs multiple tasks which set the stage for the onset of random X chromosome inactivation.[57] *Xite* ncRNA positively regulates *Tsix* transcription. The truncation of *Xite* ncRNA using transcription termination signals results in downregulation of *Tsix* transcription,[48] indicating that *Xite* transcription probably creates transcriptionally competent chromatin at the *Tsix* promoter.

The onset of random inactivation involves two important steps: counting the number of X chromosomes and selection of the future active and inactive X chromosomes. These steps ensure inactivation of only one X chromosome while keeping the other X chromosome in an active form. The counting step is initiated only when the ratio of X chromosomes to autosomes is one or more (X:A ≥ 1) and hence X chromosome inactivation occurs only in females but not males, in which the X chromosome to autosome ratio is less than one (X:A = 0.5). A recent investigation implicated the X chromosome-encoded transcription factor E3 ubiquitin ligase RNF12 in the counting step of X chromosome inactivation. It has been shown that over expression of RNF12 in male and female ES cells resulted in inactivation of one and two X chromosomes, respectively, indicating that the RNF12 is an activator of X chromosome inactivation process.[58] Based on these observations, RNF12 has been speculated to directly activate *Xist* transcription in female ES cells during X-chromosome inactivation. However, no evidence to date has documented a direct interaction between RNF12 and the *Xist* promoter, let alone its mode of action.

Transient X chromosome pairing and eventual molecular cross-talk between the X chromosomes during the initiation of X chromosome inactivation have been suggested to play an important role in the selection of the active and inactive X chromosomes. This process has been shown to be mediated by autosome-encoded transcription factors such as OCT4, CTCF and YY1 by trans-activating *Tsix* transcription.[59] Importantly, *Tsix* RNA synthesis is critical for chromosome pairing,[60,61] indicating that RNA-protein interactions determine the transient pairing of two X chromosomes. However, the mechanisms by which this physical association leads to the selection of active and inactive X chromosomes in a mutually exclusive manner in the same nuclear milieu are unclear. In addition to *Xite*, *Xist* and *Tsix* ncRNAs, *RepA* also been shown to play an important role in X chromosome inactivation. Prior to X chromosome inactivation, *RepA* is expressed by both X chromosomes. At the onset of X chromosome inactivation, *Tsix* RNA is downregulated on the future inactive X chromosome by as yet unidentified mechanisms, while *RepA* remains active, interacts with the PRC2 complex and tethers it to the *Xist* promoter on the future inactive X chromosome to establish transient local heterochromatin structure with H3K27me3 marks. Paradoxically, this tethering aids in the *trans*-activation of the *Xist* promoter, leading to accumulation of *Xist* along the future inactive X chromosome and the subsequent recruitment of heterochromatin modifiers, resulting in an RNA polymerase II-deficient heterochromatin compartment.[49] However, on the future active X-chromosome, *Tsix* is exclusively expressed and represses *in cis* the *Xist* and *RepA* promoters by targeting CpG methylation through interactions with heterochromatin machinery containing Dnmt3a.[62]

The functional link between ncRNA and X-chromosome inactivation has been investigated in more depth than other biological pathways. However, several outstanding questions remain to be answered. For example, the factors involved in guiding *Xist* along the future inactive X-chromosome *in cis* have yet to be determined. Furthermore,

how does the expression of *Xist* and *Tsix* become exclusively restricted to inactive and active X chromosomes, respectively? Addressing this important issue will have implications in our understanding of the chromosomal counting and selection processes of X-chromosome inactivation.

Dosage compensation in flies has the similar objective of equalizing X-linked gene products between male and female flies with unequal numbers of X chromosomes. However, in contrast to silencing one of the two X chromosomes in female mammals, dosage compensation in the flies is achieved through hyper-transcription of the single X chromosome in males (Fig. 3A-B). Nevertheless, like in mammals, the dosage compensation process in flies is carried out by long ncRNAs, here *rox1* and *rox2*, along with a defined set of male-specific lethal proteins: MSL1, MSL2, MSL3, MLE (maleless) and H4K16 acetyltransferase MOF. This ribonucleoprotein complex is referred to as the Dosage Compensation Complex or the MSL complex.[63] Like *Xist*, *rox* ncRNA blankets the male X chromosome *in cis* (Fig. 3A). Surprisingly, *rox* can coat the X chromosome *in trans* even when the *rox* gene has been translocated onto an autosome, indicating marked mechanistic differences in the mechanisms by which *Xist* and *rox* effect dosage compensation. *Rox* ncRNAs are encoded by RNA polymerase II and localized to the nuclear compartment. Although both *rox* ncRNAs are located on the X chromosome and perform similar functions, they differ structurally in size and sequence. *rox1*, the longer transcript and *rox2* are 3.7 kb and 0.6 kb in length, respectively.

The *rox* ncRNAs are exemplary of the ability of diverse ncRNAs to carry out a similar function despite a lack of sequence similarity. Single mutations involving one of the *rox* ncRNAs have no apparent phenotypic effect on either the RNA chromosomal coating or hyper-transcription of the male X chromosome, indicating that they are functionally redundant. However, flies with a deletion or mutation of *both rox* ncRNAs demonstrated mislocalization of the SL complex and reduction in the transcription of the X chromosome, indicating that the *rox* ncRNAs are integral parts of the MSL complex.[64] Both *rox1* and *rox2* harbor evolutionarily conserved short sequence stretches at the 3′end which can form stem-loop structures. Deletion or mutation of the putative stem-loop structures from *rox1* or *rox2* resulted in the reduction of male viability.[44] These stem-loop structures are thought to play an important role in the recruitment of the MSL complex, but evidence is still lacking as to the actual role of *rox* ncRNAs in the functional organization of the MSL complex.

The question remains as to how *rox* ncRNAs target the MSL complex along the male X chromosome. Using ChIP-on-chip technology with antibodies directed against components of the MSL complex, several high-affinity MSL binding sites were mapped along the X chromosome and found to be twice as prevalent there in comparison to autosomes.[65] If the MSL complex is recruited through high-affinity DNA sequences, then *rox* ncRNA simply acts as a RNA scaffold for integrating the MSL complex. Interestingly, most MSL high-affinity binding sites map to the 3′ end of protein-coding genes. In light of these observations, tethering of the MSL complex to the 3′ end of genes has been proposed to facilitate transcriptional elongation, resulting in hyper-transcription of the targeted genes.[66] The MSL complex could also be targeted by *rox* ncRNA through RNA-RNA or RNA-DNA interactions with the nascent transcripts or the promoter DNA sequences of the X-linked genes *in cis* to achieve hyper-transcription. However, experimental support for the latter hypothesis is lacking. Understanding the role of *rox* ncRNA in the functional integration of the MSL complex and its targeting along

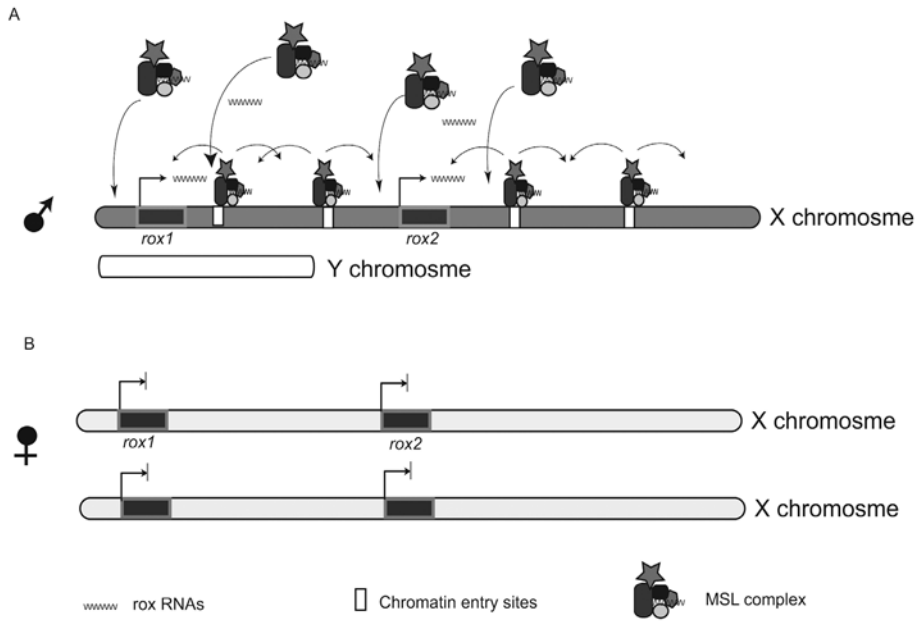

Figure 3. Dosage compensation in Drosophila. A) Single X chromosome in males showing transcriptionally active *rox* genes (*rox1* and *rox2*) and chromatin entry sites. The MSL complex specifically activates *rox* genes on the male X chromosome. *rox* RNAs are incorporated into the MSL complexes and the *rox*-MSL complex spread along the male X chromosomes via chromatin entry sites. The hyper-transcriptional activity on the male X chromosome is shown in dark color. B) Drosophila females have two X chromosomes. The *rox* genes are inactive on the female X chromosome and the low level of transcriptional activity on two female X chromosomes is shown in light color.

the male X chromosome will provide greater insight into how long ncRNAs achieve transcriptional activation.

LONG ncRNAs AND PLURIPOTENCY

The transcriptional network centered on a cocktail of transcription factors including Oct4, Nanog, c-Myc and Sox2, has been well-established as playing an essential role in the establishment and maintenance of the pluripotent state of embryonic stem cells. Interestingly, some of these transcription factors have been implicated in the regulation of small ncRNAs whose level is critical for maintaining pluripotency. For example, c-Myc regulates microRNAs such as miR-141, miR-200 and miR-429 in ES cells, but not in differentiated cells. Downregulation of these miRNAs accelerates the process of ES cell differentiation,[67] indicating that they are critical for maintaining ES cell pluripotency and that c-Myc regulates the pluripotent state of ES cells through ncRNA levels. Likewise, miR-145 represses pluripotency in human embryonic stem cells by regulating OCT4, SOX2 and KLF4 transcription factors.[68] Loss of miR-145 impairs differentiation and elevates OCT4, KLF4 and SOX2 levels. Intriguingly, miR-145 is, in turn, repressed in ES cells by OCT4 and downregulation of OCT4 in ES cells increases the level of miR-145 and

promotes differentiation. These findings indicate that small ncRNAs and the pluripotent transcriptional program maintain pluripotency through a double-negative feedback regulatory loop that switches embryonic stem cells between self-renewal and differentiation (Fig. 4).

Emerging evidence suggests a functional link between long ncRNAs and pluripotency. In a recent analysis, four conserved long ncRNAs were identified based on their proximity to Oct4 and Nanog binding sites; two of these were found to be direct targets of Oct4 and Nanog.[69] Similar to miRNAs, long ncRNAs and pluripotent transcription factors form a regulatory feedback loop. For example, Oct4 positively regulates one of the long ncRNAs, which in turn activates Oct4 transcription. These data establish another hidden regulatory loop in the maintenance of pluripotency. The handful of long ncRNAs mentioned above only represents the tip of the iceberg and a comprehensive analysis of long ncRNA transcription across the ES cell genome is needed in order to determine the extent of ncRNA involvement in the maintenance of pluripotency. Along these lines, a recent investigation using genome-wide analyses identified 945 long ncRNAs in differentiating ES cells of which 174 were differentially expressed. In-depth analysis of two novel differentially expressed ncRNAs revealed that they play an important

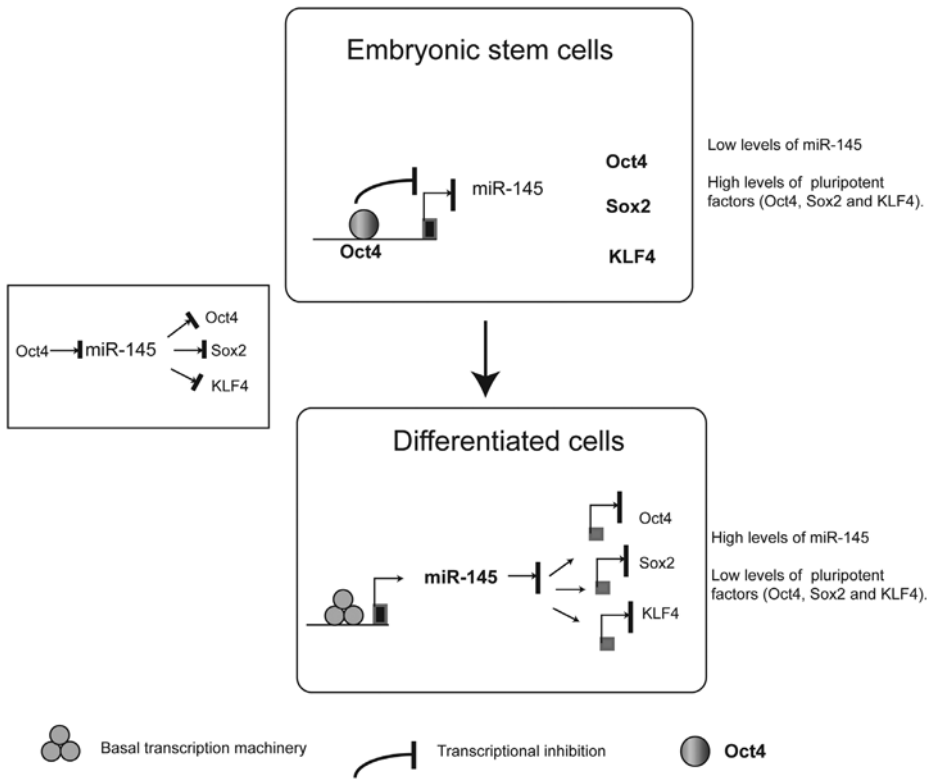

Figure 4. Model depicting the double-negative feedback loop by miR-145 and three pluripotent transcription factors Oct4, Sox2 and KLF4. miR-145 maintains the differentiated state of a cell via repressing the pluripotency transcription factors. Like-wise, Oct4 controls the pluripotency of embryonic stem cells by repressing miR-145. Bold font represents higher levels, whereas regular font represents low levels.

role in pluripotency and differentiation by regulating the chromatin structure through interactions with chromatin-modifying complexes.[40] This study is indeed encouraging and future research along these lines will likely uncover more pluripotent transcriptional programs involving long ncRNAs.

LONG ncRNAs AND GENOMIC IMPRINTING

Genomic imprinting is an epigenetically controlled gene regulatory mechanism by which expression of a sub-group of autosomal genes is restricted to either the maternal or paternal allele. So far, more than 100 imprinted genes have been identified in mouse and the majority of imprinted genes are organized in clusters containing both protein-coding and noncoding mRNA genes. The noncoding mRNA genes contain both small and long ncRNAs. The small RNAs include miRNAs, C/D small RNAs and piRNAs which range in size from 19 to 71 nucleotides and the long ncRNAs measure from 100 bp to several hundred kb.[70,71] Genes in imprinted clusters are coregulated by long-range *cis*-acting elements, which are short stretches of DNA sequence (1 to 3 kb) that often carry epigenetic marks such as CpG methylation and/or histone modifications on only one parental chromosome. These differentially marked regions are called DMRs or ICRs (imprinting control regions), as their deletion in mice was shown to affect imprinting in clusters.[72] The parent-specific epigenetic profile at the ICR or DMR is established during the germline (in the majority of cases it occurs in female germline but in some cases in male germline) or postimplantation stages of mouse development. Although imprinted clusters carry common epigenetic features, the mechanisms by which they attain these features differ across clusters. Among the 13-14 imprinted clusters identified so far, only the mechanisms of seven that contain differentially methylated imprinting control regions have been investigated. The mechanisms underlying the ICR-mediated long-range gene regulatory mechanism have been investigated intensely over the last several years. These studies have shown that ICRs regulate genomic imprinting in clusters either by acting as chromatin insulators[73,74] or by harboring the promoters for ncRNA.[24] Since ncRNA and chromatin architecture is the theme of this chapter, only the imprinted clusters regulated by ncRNAs will be discussed further (Fig. 5).

Kcnq1, *Igf2r* and *Gnas* imprinted clusters contain very long ncRNAs encoded from the promoters that map to ICRs (Fig. 5).[70] For example, *Kcnq1ot1* ncRNA is a 91-kb, poorly spliced, moderately stable transcript localized to the nucleus. Its promoter maps to an imprinting control region (*Kcnq1* ICR), which lies in intron 10 of the *Kcnq1* gene on mouse chromosome 7.[75] *Kcnq1ot1* transcribes in the antisense direction to the *Kcnq1* gene, starting in intron 10 and ending in intron 11 and thus it does not span across the promoter region of the *Kcnq1* gene (Fig. 5A). Targeted deletion of the *Kcnq1ot1* promoter or truncation of the *Kcnq1ot1* transcript from 91 kb to a size range of 0.7-2.6 kb resulted in the loss of silencing of 8 to 9 maternally expressed protein-coding mRNA genes spread over one mega-base region on either side of the *Kcnq1ot1* promoter in several different studies, indicating that *Kcnq1ot1* RNA is a bidirectional silencer.[76-80] However, these experiments did not distinguish between the functional role of the ncRNA itself and the act of ncRNA transcription. By interfering with the stability of episome-encoded *Kcnq1ot1* transcript in human cells by flanking *Kcnq1ot1* with the destabilizing domain of highly unstable *c-fos* transcript, *Kcnq1ot1* was also found to play an important role in bidirectional silencing.[75] Recent data from several independent investigations have

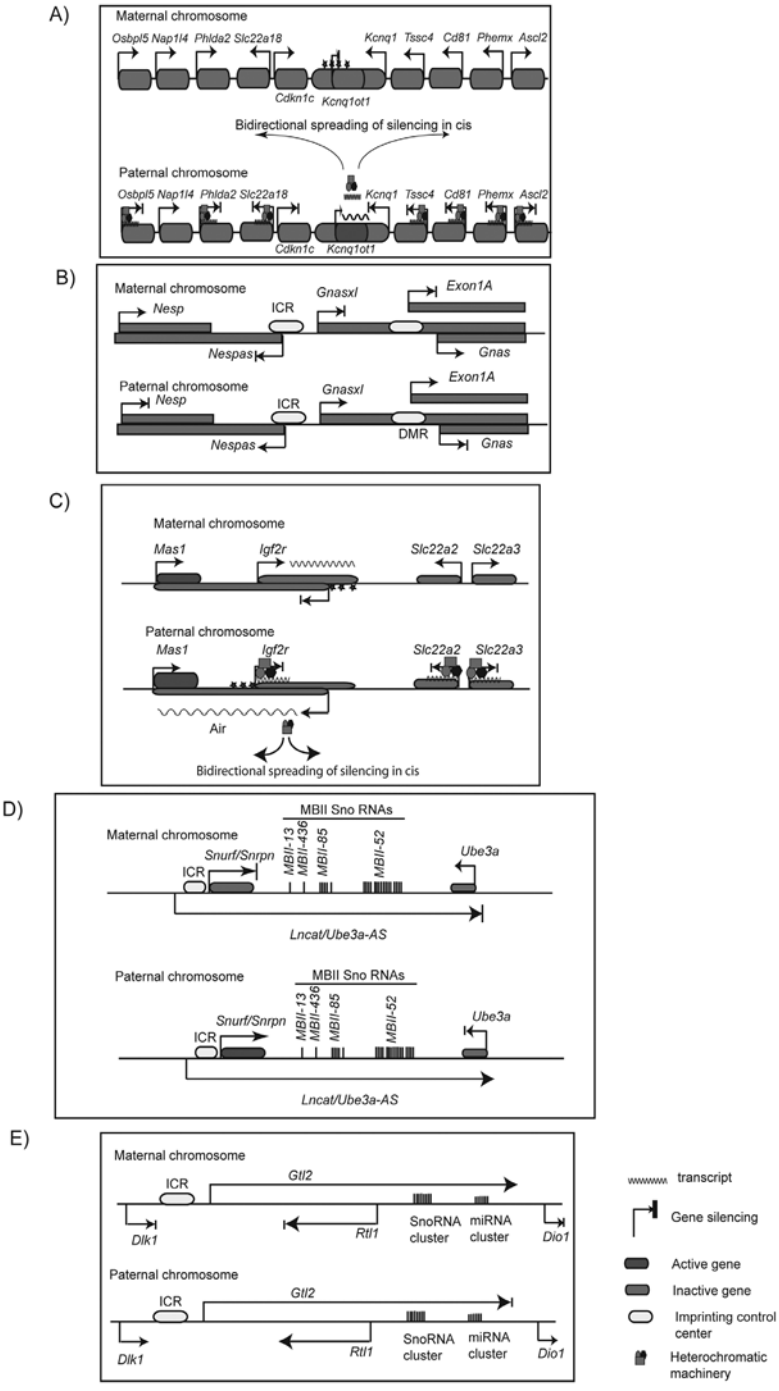

Figure 5. A-E) The mouse imprinted gene clusters *Kcnq1* (A), *Gnas* (B), *Igf2r* (C), *Snurf/Snrpn* (D) and *Dlk1/Gtl2* (E) containing small and/or long noncoding RNAs. Arrows show the direction of transcription. Maps are not drawn to scale.

shown that *Kcnq1ot1* interacts with chromatin and recruits heterochromatin complexes containing PRC2 and G9a in a gene-specific manner.[75] This mode of action of *Kcnq1ot1* has parallels with the mode of action of *Xist* ncRNA, which mediates transcriptional gene silencing by coating the future inactive X chromosome.

Airn, an antisense transcript to *Igf2r,* is 108 kb in length and localized to the nuclear compartment. *Airn* is spliced, but the splice variants constitute an insignificant portion of the total *Airn* transcription and moreover are localized to the cytoplasmic compartment.[81] The *Airn* promoter maps to a differentially methylated imprinting control region located in intron 2 of the *Igf2r* gene. Similar to *Kcnq1ot1*, selective deletion of the *Airn* promoter or truncation of its transcript from 108 kb to 3 kb resulted in loss of silencing of 3 protein-coding genes, *Slc22a3*, *Slc22a2* and *Igf2r*, on either side of the *Airn* promoter, indicating that *Airn* is also a bidirectional silencer (Fig. 5C).[82] It is not clear whether the RNA itself or its transcription is crucial for bidirectional silencing. However, a recent investigation has documented that *Airn* interacts with G9a histone methyltransferase and specifically targets the nonoverlapping *Slc22a3* promoter but not the overlapping *Igf2r* promoter, indicating that *Airn* probably uses different mechanisms to silence the overlapping and nonoverlapping promoters.[83] *Airn* is transcribed through the overlapping *Igf2r* promoter in all tissues and, unlike nonoverlapping genes silenced only in extra-embryonic tissues in a developmentally regulated manner, the overlapping *Igf2r* gene is constitutively silenced on the paternal chromosome. Based on these observations, the act of transcription has been proposed to silence the overlapping gene *in cis* through transcriptional interference, whereas *Airn* itself mediates nonoverlapping gene silencing by acting as a heterochromatin scaffold.[16]

The *Gnas* imprinted cluster is approximately 70 kb in length and accommodates three protein-coding (*Gnasxl*, *Nesp* and *Gnas*) and two noncoding mRNA (*Exon1A* and *Nespas*) genes (Fig. 5B).[70] Unlike the *Kcnq1* and *Igf2r* clusters, in which protein-coding mRNA is expressed only by the maternal allele, the *Gnas* cluster contains both paternally and maternally expressed protein-coding mRNAs. Its two ncRNAs are expressed exclusively by the paternal allele. The promoters of these two ncRNAs map to germline-derived differentially methylated regions. Targeted deletion of exon 1A DMR affected the imprinting of the flanking *Gnasxl* gene and deletion of the DMR containing the *Nespas* promoter resulted in loss of imprinting of all genes, indicating that *Nespa*s could play an important role in the imprinting of the *Gnas* cluster.[84] However, no experiments have been published to date that implicate a functional role of *Nespas* and/or its transcriptional process in imprint acquisition and maintenance.

The *PWS/AS* and *Gtl2/Dlk1* imprinted clusters encode long ncRNAs containing large repeated arrays of C/D small nucleolar RNA (snoRNA) and miRNA (Fig. 5D and E). In mice, the PWS/AS locus encodes a huge paternally expressed 1-Mb poly-cistronic long ncRNA, Lncat.[85] The locus has approximately 148 exons that include ICR/DMR, paternally expressed protein-coding Snrpn/Snurf mRNAs, two repeated large arrays of MBII-52 and MBII-85 C/D SnoRNAs and an antisense RNA (Ube3a-as) to maternally expressed Ube3a transcript (Fig. 5D). The existence of a 1-Mb long ncRNA in the PWS/AS locus is controversial. Using high-resolution fluorescence *in situ* hybridization with oligonucleotide probes, a recent investigation has demonstrated that MBII-52 and MBII-85 C/D snoRNAs are not part of the large Lncat RNA but are instead derived from two independent noncoding transcripts.[86] This evidence was ascertained based on the fact that spliced MBII-52 and MBII-85 accumulate as "RNA clouds" in a non-overlapping fashion near the site of transcription. However, this observation cannot rule out the possibility

that Lncat is pervasively transcribed through protein-coding and noncoding transcription units at a very low level. The formation of RNA clouds by MBII-85 and MBII-52 near their transcription units is indicative of a functional role in the chromatin-based transcriptional regulation of neighboring genes. The functional significance of paternally expressed Ube3a-as in the paternal repression of its sense counterpart Ube3a has not yet been investigated.

Numerous imprinted microRNA genes are organized in two clusters at the *Dlk1-Gtl2* domain (Fig. 5E). A small group of 5-6 miRNAs (miR-431, miR-433, miR-127, miR-434 and miR-136) are processed from the maternally expressed antisense transcript that overlaps the paternally expressed *Rtl1* (retrotransposon-like 1) gene.[71,87] The miRNAs show perfect complementary to paternally expressed *Rtl1* mRNA and, importantly, several cleavage sites have been experimentally mapped within the *Rtl1* mRNA using 5′ RACE technology, suggesting that these miRNAs may mimic the mode of action of siRNA.[88] It remains to be seen whether these miRNAs have a role in the maternal-specific repression of *Rtl1 in cis*. In addition, the mechanism by which paternally expressed *Rtl1* transcript escapes the actions of miRNA *in trans* must still be established. Another larger cluster of nearly 40 miRNAs maps to the region downstream of the C/D snoRNA cluster. The functional role of this microRNA cluster in genomic imprinting of the *Dlk1/Gtl2* locus has not yet been investigated. The C/D snoRNA cluster is located between the two miRNA clusters upstream of the *Rtl1* gene. Its functional significance in the imprinting of *Dlk1/Gtl2* is also currently unknown.

LONG ncRNAs AND CENTROMERIC HETEROCHROMATIN

Spatial organization of heterochromatin and euchromatin in the mammalian genome is crucial for normal development and differentiation. While the euchromatic environment is conducive to active transcription, heterochromatin plays an important role in genome stability, chromosomal segregation and gene regulation. Defects in the molecular pathways that functionally demarcate the genome could lead to abnormal chromosome segregation and genome instability.[89] For example, the centromeric heterochromatin formation is crucial for faithful chromosome segregation during cell division and defective centromeres lead to abnormal chromosome segregation and aneuploidy.[90] The centromere is an epigenetic regulatory framework that plays a supportive role in mitotic spindle formation and sister chromatid cohesion. The centromere is made up of an array of repeat elements thought to organize heterochromatic structures by recruiting repressive chromatin remodeling complexes. Recent investigations have demonstrated that small and long ncRNAs play a functional role in the establishment as well as maintenance of centromeric heterochromatin.[20,91,92] In *S. pombe*, RNAi-based silencing mechanisms play a pivotal role in the establishment and maintenance of centromeric heterochromatin.[93-95] The centromere in *S. pombe* is compartmentalized into a central core (cnt) flanked by inner (*imr*) and outer (*otr*) repeats (Fig. 6). CENP-A (the centromere-specific histone H3 variant) nucleosomes in the central core region act as an anchor point for the kinetochore, which connects the mitotic spindle to the centromere. The outer repeat region contains dg-dh repeats which upon bidirectional transcription produce long double-stranded RNAs that are tethered to RNAi substrates to form heterochromatin at the *Otr* region (Fig. 6).[94] The heterochromatin at the *Otr*

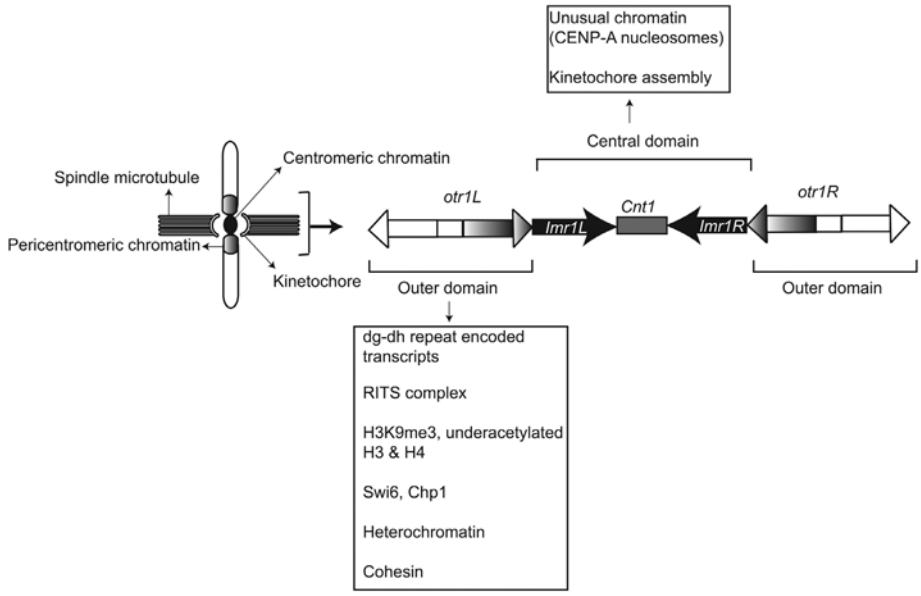

Figure 6. Centromere structure and organization in *S.pombe*. A-B) Schematic depiction of *S.pombe* centromere and pericentromere regions. Centromere1 (*cnt1*) contains a central core of nonrepetitive sequences flanked by innermost repeats (*imr1L and imr1R*) and outer repeats (*otr1L and otr1R*). *Cnt1* acts as an anchor point for kinetochore assembly. The *otr* region composed of dg-dh repeats and are transcribed in forward and reverse directions, thus giving rise to double stranded RNA (dsRNA). dsRNAs, derived from the dg-dh repeats, have been shown to take part in heterochromatin organization of the *otr* region by triggering RNAi via recruiting RITS-RDRP complexes. The heterochromatic modifications such as H3K9me2 and H3K9me3 in the *otr* region recruit Swi6 (HP1) and Chp1 heterochromatic proteins and they, in turn, are responsible for the recruitment of Cohesin. Cohesin is required for the proper orientation of centromeres during mitosis.

region is not only critical for sister chromatid cohesion but also helps in CENP-A chromatin formation in the *Cnt* region through provision of RNAi substrates.[96]

In eukaryotes, however, the molecular mechanisms by which the centromeric heterochromatin is established and maintained are not clear. In mammals, the centromere contains two domains: a central domain required for the kinetochore assembly and flanking pericentric heterochromatin required for sister chromatid cohesion. Thus, both structural and functional features of centromeres are highly conserved across the evolutionary spectrum. Previously, RNA was shown to define the higher order chromatin structure at pericentric heterochromatin by organizing heterochromatic components such as H3K9me3 and HP1 proteins. Recent analyses have further linked both RNAi components and long ncRNAs in the maintenance of pericentric heterochromatin. Chicken cell lines lacking Dicer, a dsRNA RNA-specific endonuclease and a crucial component of the RNA-induced silencing complex (RISC), showed reduced CpG methylation and H3K9me3 in the pericentric heterochromatin.[97] Whether Dicer plays a direct or indirect role in the maintenance of pericentric heterochromatin has not yet been determined. In Dicer mutants, a decrease in Dnmt1, a maintenance methyltransferase, levels occurs due to overexpression of Rbl2, which is caused by downregulation of its negative regulator miR-290.[98] Since DNA methylation is associated with H3K9me3

levels and Dicer depletion does not have any effect on CENPA and CENPC1 localization to the centromeres,[20,97] the functional role of the RNAi components in centromeric heterochromatin formation requires a thorough investigation in eukaryotes. However, recent analyses have identified a functional role for long ncRNAs in centromeric heterochromatin formation. In mammals, the centromere contains two classes of tandemly organized repeat arrays, termed minor and major satellite sequences. Major satellite sequences map to the pericentric portion of the centromere; minor satellite sequences are located at the primary constriction of condensed mitotic chromosomes. Minor (alpha) satellite sequences are transcribed, with transcripts ranging between 100-500 nt in size and localized specifically to the centromere.[92] Overexpression of 120-nt RNA molecules, blocking the 120-bp unit minor satellite sequence in mammalian cells, or treatment of mitotic cells with single-stranded RNA-specific nucleases leads to mislocalization of centromere-associated heterochromatin proteins, indicating that the minor-satellite repeat RNA acts as a scaffold in recruiting heterochromatin proteins.[92] In addition, minor satellite repeat RNA has been shown to directly interact with centromere-associated proteins such as CENPC1 and the centromere associated chromosomal passenger complex, which contains Aurora B kinase, Survivin and inner centromere protein (INCENP), indicating that this ribonucleoprotein complex is critical for proper centromere formation.[20,21] Taken together, the above data implicate ncRNA in mammalian centromere formation and thereby, genome stability.

CONCLUSION

Functional demarcation of the eukaryotic genome into active and inactive chromatin compartments enables spatio-temporal gene expression, which is crucial for the normal development of an organism. Accumulating evidence in recent years suggests that long ncRNAs and/or the process of their transcription act as scaffolds in the spatial organization of active and inactive chromatin compartments. Although the latter two processes have been implicated in the functional organization of the genome, a number of gaps remain in our understanding of the mechanisms they govern. In most cases the effects of ncRNA transcription are restricted to overlapping genes, whereas the effects of ncRNAs themselves can range from a single gene to an entire chromosome. Functional dissection of these mechanisms is imperative to gain further understanding of the contribution of ncRNAs to organismal complexity. Although, in the majority of instances, ncRNA sequences and their target genes differ across the evolutionary spectrum, their physical location, promoter sequences and, in part, their modes of action are conserved. For example, well-investigated ncRNAs such as *Xist* and *Kcnq1ot1* significantly differ in their primary sequence across different species, but their modes of action remain unchanged between mice and humans. This consistency indicates that the secondary structures of ncRNAs likely play a role in their functions.

Misexpression of several long ncRNAs deregulates clinically significant genes, as evidenced by the formation of focal heterochromatin structures in several diseases, For example, aberrant expression of antisense ncRNAs through the *p15* and *p21* promoters in certain diseases results in the inactivation of these promoters via CpG methylation. What triggers aberrant antisense ncRNA expression in certain diseases is not clear,

nor, importantly, is how their transcription across an overlapping or neighboring gene causes specific methylation of their promoters. Similarly, recent investigations have also exposed a link between genome stability and some ncRNAs, which play an important role in the maintenance of higher order chromatin structure at the centromere. Defects in the pathways that lead to functional centromeres result in loss of chromosome segregation and aneuploidy. The loss of genome stability triggers chromosome deletions and translocations, which, in turn, result in aberrant ncRNA transcription by cryptic promoters, through several genes with clinical significance.

Though the objectives of certain ncRNA-mediated epigenetic pathways are similar, the mechanisms used to achieve these objectives differ considerably depending on the organism. For example, dosage compensation pathways in flies and mammals employ the common theme of ncRNA but use different mechanisms, i.e., hyperactivation of one X chromosome in male flies and transcriptional silencing of one X chromosome in female mammals, to ensure an equal dosage of X-linked gene products in males and females. Interestingly, both actions involve coating the entire X chromosome with ncRNA.

The majority of long ncRNAs described here constitute a network that regulates various biological functions via modifying the chromatin structure in a spatio-temporal manner. Hence this network of long ncRNAs is a part of the RNA infrastructure dedicated towards maintaining chromatin structure in time and space.

Though the evidence collected over the last few years has enabled us to understand ncRNA as a hidden gene regulatory mechanism participating in diverse biological functions, much more remains to be learned about their actions before their real impact on organismal complexity can be fully appreciated.

ACKNOWLEDGEMENTS

We thank Radha Raman Pandey and Natalia Guseva for comments on the chapter. This work was supported by the grants from the Swedish Cancer Research foundation (Cancerfonden), Swedish Medical Research Council (VR-M) and Swedish Childhood Cancer Society (Barncancerfonden) to CK. CK is a Senior Research Fellow supported by VR-M.

REFERENCES

1. Mattick JS. RNA regulation: a new genetics? Nat Rev Genet 2004; 5(4):316-323.
2. Kapranov P, Cheng J, Dike S et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science 2007; 316(5830):1484-1488.
3. Cheng J, Kapranov P, Drenkow J et al. Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. Science 2005; 308(5725):1149-1154.
4. Prasanth KV, Spector DL. Eukaryotic regulatory RNAs: an answer to the 'genome complexity' conundrum. Genes Dev 2007; 21(1):11-42.
5. Mercer TR, Dinger ME, Mattick JS. Long noncoding RNAs: insights into functions. Nat Rev Genet 2009; 10(3):155-159.
6. Whitehead J, Pandey GK, Kanduri C. Regulation of the mammalian epigenome by long noncoding RNAs. Biochim Biophys Acta 2008.
7. Faghihi MA, Wahlestedt C. Regulatory roles of natural antisense transcripts. Nat Rev Mol Cell Biol 2009; 10(9):637-643.

8. Collins LJ, Penny D. The RNA infrastructure: dark matter of the eukaryotic cell? Trends Genet 2009; 25(3):120-128.
9. Collins LJ, Chen XS. Ancestral RNA: the RNA biology of the eukaryotic ancestor. RNA Biol 2009; 6(5):495-502.
10. Guttman M, Amit I, Garber M et al. Chromatin signature reveals over a thousand highly conserved large noncoding RNAs in mammals. Nature 2009; 458(7235):223-227.
11. Brosnan CA, Voinnet O. The long and the short of noncoding RNAs. Curr Opin Cell Biol 2009; 21(3):416-425.
12. Seto AG, Kingston RE, Lau NC. The coming of age for Piwi proteins. Mol Cell 2007; 26(5):603-609.
13. Aravin AA, Hannon GJ, Brennecke J. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science 2007; 318(5851):761-764.
14. Kanduri C, Whitehead J, Mohammad F. The long and the short of it: RNA-directed chromatin asymmetry in mammalian X-chromosome inactivation. FEBS Lett 2009; 583(5):857-864.
15. Lee JT. Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome. Genes Dev 2009; 23(16):1831-1842.
16. Mohammad F, Mondal T, Kanduri C. Epigenetics of imprinted long noncoding RNAs. Epigenetics 2009; 4(5):277-286.
17. Clemson CM, Hutchinson JN, Sara SA et al. An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. Mol Cell 2009; 33(6):717-726.
18. Sasaki YT, Ideue T, Sano M et al. MENepsilon/beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles. Proc Natl Acad Sci USA 2009; 106(8):2525-2530.
19. Schoeftner S, Blasco MA. A 'higher order' of telomere regulation: telomere heterochromatin and telomeric RNAs. EMBO J 2009; 28(16):2323-2336.
20. Wong LH, Brettingham-Moore KH, Chan L et al. Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. Genome Res 2007; 17(8):1146-1160.
21. Ferri F, Bouzinba-Segard H, Velasco G et al. Non-coding murine centromeric transcripts associate with and potentiate Aurora B kinase. Nucleic Acids Res 2009; 37(15):5071-5080.
22. Willingham AT, Orth AP, Batalov S et al. A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. Science 2005; 309(5740):1570-1573.
23. Mondal T, Rasmussen M, Pandey GK et al. Characterization of the RNA content of chromatin. Genome Res 2010; 20:899-907.
24. Kanduri C. Functional insights into long antisense noncoding RNA Kcnq1ot1 mediated bidirectional silencing. RNA Biol 2008; 5(4):208-211.
25. Wilusz JE, Sunwoo H, Spector DL. Long noncoding RNAs: functional surprises from the RNA world. Genes Dev 2009; 23(13):1494-1504.
26. Rinn JL, Kertesz M, Wang JK et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 2007; 129(7):1311-1323.
27. Katayama S, Tomaru Y, Kasukawa T et al. Antisense transcription in the mammalian transcriptome. Science 2005; 309(5740):1564-1566.
28. Yu W, Gius D, Onyango P et al. Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. Nature 2008; 451(7175):202-206.
29. Houseley J, Rubbi L, Grunstein M et al. A ncRNA modulates histone modification and mRNA induction in the yeast GAL gene cluster. Mol Cell 2008; 32(5):685-695.
30. Ligtenberg MJ, Kuiper RP, Chan TL et al. Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3′ exons of TACSTD1. Nat Genet 2009; 41(1):112-117.
31. Tufarelli C, Stanley JA, Garrick D et al. Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. Nat Genet 2003; 34(2):157-165.
32. Hekimoglu B, Ringrose L. Non-coding RNAs in polycomb/trithorax regulation. RNA Biol 2009; 6(2):129-137.
33. Hirota K, Miyoshi T, Kugou K et al. Stepwise chromatin remodelling by a cascade of transcription initiation of noncoding RNAs. Nature 2008; 456(7218):130-134.
34. Morris KV, Santoso S, Turner AM et al. Bidirectional transcription directs both transcriptional gene activation and suppression in human cells. PLoS Genet 2008; 4(11):e1000258.
35. Ringrose L, Paro R. Polycomb/Trithorax response elements and epigenetic memory of cell identity. Development 2007; 134(2):223-232.
36. Schuettengruber B, Cavalli G. Recruitment of polycomb group complexes and their role in the dynamic regulation of cell fate choice. Development 2009; 136(21):3531-3542.
37. Sessa L, Breiling A, Lavorgna G et al. Noncoding RNA synthesis and loss of Polycomb group repression accompanies the colinear activation of the human HOXA cluster. RNA 2007; 13(2):223-239.
38. Sanchez-Elsner T, Gou D, Kremmer E et al. Noncoding RNAs of trithorax response elements recruit Drosophila Ash1 to Ultrabithorax. Science 2006; 311(5764):1118-1123.

39. Petruk S, Sedkov Y, Riley KM et al. Transcription of bxd noncoding RNAs promoted by trithorax represses Ubx in cis by transcriptional interference. Cell 2006; 127(6):1209-1221.
40. Dinger ME, Amaral PP, Mercer TR et al. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. Genome Res 2008; 18(9):1433-1445.
41. Kelley RL, Kuroda MI. Noncoding RNA genes in dosage compensation and imprinting. Cell 2000; 103(1):9-12.
42. Hallacli E, Akhtar A. X chromosomal regulation in flies: when less is more. Chromosome Res 2009; 17(5):603-619.
43. Heard E, Clerc P, Avner P. X-chromosome inactivation in mammals. Annu Rev Genet 1997; 31:571-610.
44. Ilik I, Akhtar A. roX RNAs: noncoding regulators of the male X chromosome in flies. RNA Biol 2009; 6(2):113-121.
45. Wutz A, Gribnau J. X inactivation Xplained. Curr Opin Genet Dev 2007; 17(5):387-393.
46. Brockdorff N, Ashworth A, Kay GF 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.
47. Lee JT, Davidow LS, Warshawsky D. Tsix, a gene antisense to Xist at the X-inactivation centre. Nat Genet 1999; 21(4):400-404.
48. Lee JT. Regulation of X-chromosome counting by Tsix and Xite sequences. Science 2005; 309(5735):768-771.
49. Zhao J, Sun BK, Erwin JA et al. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science 2008; 322(5902):750-756.
50. Cohen DE, Davidow LS, Erwin JA et al. The DXPas34 repeat regulates random and imprinted X inactivation. Dev Cell 2007; 12(1):57-71.
51. Vigneau S, Augui S, Navarro P et al. An essential role for the DXPas34 tandem repeat and Tsix transcription in the counting process of X chromosome inactivation. Proc Natl Acad Sci USA 2006; 103(19):7390-7395.
52. Ogawa Y, Sun BK, Lee JT. Intersection of the RNA interference and X-inactivation pathways. Science 2008; 320(5881):1336-1341.
53. Rougeulle C, Avner P. Controlling X-inactivation in mammals: what does the centre hold? Semin Cell Dev Biol 2003; 14(6):331-340.
54. Kalantry S, Purushothaman S, Bowen RB et al. Evidence of Xist RNA-independent initiation of mouse imprinted X-chromosome inactivation. Nature 2009; 460(7255):647-651.
55. Navarro P, Chambers I, Karwacki-Neisius V et al. Molecular coupling of Xist regulation and pluripotency. Science 2008; 321(5896):1693-1695.
56. Navarro P, Pichard S, Ciaudo C et al. Tsix transcription across the Xist gene alters chromatin conformation without affecting Xist transcription: implications for X-chromosome inactivation. Genes Dev 2005; 19(12):1474-1484.
57. Navarro P, Chantalat S, Foglio M et al. A role for noncoding Tsix transcription in partitioning chromatin domains within the mouse X-inactivation centre. Epigenetics Chromatin 2009; 2(1):8.
58. Monkhorst K, Jonkers I, Rentmeester E et al. X inactivation counting and choice is a stochastic process: evidence for involvement of an X-linked activator. Cell 2008; 132(3):410-421.
59. Donohoe ME, Silva SS, Pinter SF et al. The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. Nature 2009; 460(7251):128-132.
60. Xu N, Donohoe ME, Silva SS et al. Evidence that homologous X-chromosome pairing requires transcription and Ctcf protein. Nat Genet 2007; 39(11):1390-1396.
61. Augui S, Filion GJ, Huart S et al. Sensing X chromosome pairs before X inactivation via a novel X-pairing region of the Xic. Science 2007; 318(5856):1632-1636.
62. Sun BK, Deaton AM, Lee JT. A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization. Mol Cell 2006; 21(5):617-628.
63. Gilfillan GD, Dahlsveen IK, Becker PB. Lifting a chromosome: dosage compensation in Drosophila melanogaster. FEBS Lett 2004; 567(1):8-14.
64. Franke A, Baker BS. The rox1 and rox2 RNAs are essential components of the compensasome, which mediates dosage compensation in Drosophila. Mol Cell 1999; 4(1):117-122.
65. Alekseyenko AA, Peng S, Larschan E et al. A sequence motif within chromatin entry sites directs MSL establishment on the Drosophila X chromosome. Cell 2008; 134(4):599-609.
66. Kind J, Vaquerizas JM, Gebhardt P et al. Genome-wide analysis reveals MOF as a key regulator of dosage compensation and gene expression in Drosophila. Cell 2008; 133(5):813-828.
67. Lin CH, Jackson AL, Guo J et al. Myc-regulated microRNAs attenuate embryonic stem cell differentiation. EMBO J 2009; 28(20):3157-3170.
68. Xu N, Papagiannakopoulos T, Pan G et al. MicroRNA-145 regulates OCT4, SOX2 and KLF4 and represses pluripotency in human embryonic stem cells. Cell 2009; 137(4):647-658.

69. Sheik Mohamed J, Gaughwin PM, Lim B et al. Conserved long noncoding RNAs transcriptionally regulated by Oct4 and Nanog modulate pluripotency in mouse embryonic stem cells. RNA 2009.
70. Peters J, Robson JE. Imprinted noncoding RNAs. Mamm Genome 2008; 19(7-8):493-502.
71. Royo H, Cavaille J. Non-coding RNAs in imprinted gene clusters. Biol Cell 2008; 100(3):149-166.
72. Lewis A, Reik W. How imprinting centres work. Cytogenet Genome Res 2006; 113(1-4):81-89.
73. Kanduri C, Pant V, Loukinov D et al. Functional interaction of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. Curr Biol 2000; 10:853-856.
74. Kanduri C, Holmgren C, Franklin G et al. The 5′-flank of the murine H19 gene in an unusual chromatin conformation unidirectionally blocks enhancer-promoter communication. Curr Biol 2000; 10:449-457.
75. Pandey RR, Mondal T, Mohammad F et al. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell 2008; 32(2):232-246.
76. Pandey RR, Ceribelli M, Singh PB et al. NF-Y regulates the antisense promoter, bidirectional silencing and differential epigenetic marks of the Kcnq1 imprinting control region. J Biol Chem 2004; 279(50):52685-52693.
77. Thakur N, Tiwari VK, Thomassin H et al. An antisense RNA regulates the bidirectional silencing property of the kcnq1 imprinting control region. Mol Cell Biol 2004; 24(18):7855-7862.
78. Kanduri C, Thakur N, Pandey RR. The length of the transcript encoded from the Kcnq1ot1 antisense promoter determines the degree of silencing. EMBO J 2006; 25(10):2096-2106.
79. Mancini-Dinardo D, Steele SJ, Levorse JM et al. Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. Genes Dev 2006; 20(10):1268-1282.
80. Shin JY, Fitzpatrick GV, Higgins MJ. Two distinct mechanisms of silencing by the KvDMR1 imprinting control region. EMBO J 2008; 27(1):168-178.
81. Koerner MV, Pauler FM, Huang R et al. The function of noncoding RNAs in genomic imprinting. Development 2009; 136(11):1771-1783.
82. Sleutels F, Zwart R, Barlow DP. The noncoding Air RNA is required for silencing autosomal imprinted genes. Nature 2002; 415(6873):810-813.
83. Nagano T, Mitchell JA, Sanz LA et al. The Air Noncoding RNA Epigenetically Silences Transcription by Targeting G9a to Chromatin. Science. 2008.
84. Williamson CM, Ball ST, Nottingham WT et al. A cis-acting control region is required exclusively for the tissue-specific imprinting of Gnas. Nat Genet 2004; 36(8):894-899.
85. Landers M, Bancescu DL, Le Meur E et al. Regulation of the large (approximately 1000 kb) imprinted murine Ube3a antisense transcript by alternative exons upstream of Snurf/Snrpn. Nucleic Acids Res 2004; 32(11):3480-3492. Print 2004.
86. Vitali P, Royo H, Marty V et al. Long nuclear-retained noncoding RNAs and allele-specific higher-order chromatin organization at imprinted snoRNA gene arrays. J Cell Sci 123(Pt 1):70-83.
87. Lin SP, Youngson N, Takada S et al. Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the Dlk1-Gtl2 imprinted cluster on mouse chromosome 12. Nat Genet 2003; 35(1):97-102.
88. Davis E, Caiment F, Tordoir X et al. RNAi-mediated allelic trans-interaction at the imprinted Rtl1/Peg11 locus. Curr Biol 2005; 15(8):743-749.
89. Bernstein E, Allis CD. RNA meets chromatin. Genes Dev 2005; 19(14):1635-1655.
90. Pezer Z, Ugarkovic D. Role of noncoding RNA and heterochromatin in aneuploidy and cancer. Semin Cancer Biol 2008; 18(2):123-130.
91. Chen ES, Zhang K, Nicolas E et al. Cell cycle control of centromeric repeat transcription and heterochromatin assembly. Nature 2008; 451(7179):734-737.
92. Bouzinba-Segard H, Guais A, Francastel C. Accumulation of small murine minor satellite transcripts leads to impaired centromeric architecture and function. Proc Natl Acad Sci USA 2006; 103(23):8709-8714.
93. Djupedal I, Ekwall K. Epigenetics: heterochromatin meets RNAi. Cell Res 2009; 19(3):282-295.
94. Motamedi MR, Verdel A, Colmenares SU et al. Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. Cell 2004; 119(6):789-802.
95. Verdel A, Jia S, Gerber S et al. RNAi-mediated targeting of heterochromatin by the RITS complex. Science 2004; 303(5658):672-676.
96. Kagansky A, Folco HD, Almeida R et al. Synthetic heterochromatin bypasses RNAi and centromeric repeats to establish functional centromeres. Science 2009; 324(5935):1716-1719.
97. Fukagawa T, Nogami M, Yoshikawa M et al. Dicer is essential for formation of the heterochromatin structure in vertebrate cells. Nat Cell Biol 2004; 6(8):784-791.
98. Sinkkonen L, Hugenschmidt T, Berninger P et al. MicroRNAs control de novo DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells. Nat Struct Mol Biol 2008; 15(3):259-267.