CHAPTER 4

The Epigenetics of Non-coding RNA

Lesley J. Collins[1], Barbara Schönfeld[2,3], and Xiaowei Sylvia Chen[4]
[1]Massey Genome Service, Massey University, Palmerston North, New Zealand
[2]Allan Wilson Centre for Molecular Ecology and Evolution, Massey University, Palmerston North, New Zealand
[3]Institute of Molecular BioSciences, Massey University, Palmerston North, New Zealand
[4]Department of Biochemistry, University of Otago, Dunedin 9054, New Zealand

INTRODUCTION

Non-protein-coding RNAs (ncRNAs) are RNAs that are transcribed from DNA but are not translated into proteins. Many are functional and are involved in the processing and regulation of other RNAs such as mRNA, tRNA, and rRNA. Processing-type ncRNAs include small nuclear RNAs (snRNAs) involved in splicing, small nucleolar RNAs (snoRNAs) that modify nucleotides in rRNAs and other RNAs, and RNase P that cleaves pre-tRNAs. Other small ncRNAs such as microRNAs (miRNAs) and short interfering RNAs (siRNAs) are involved in the regulation of target mRNAs and chromatin. Although many of these latter ncRNA classes are grouped under the term RNA interference (RNAi), it has become clear that there are many different ways that ncRNAs can interact with genes to up-regulate or down-regulate expression, to silence translation, or guide methylation [1–3]. Adding to these classes are long ncRNAs (typically >200 nt) that have also been implicated in gene regulation [4]. All of these ncRNAs form a network of processes, the RNA-infrastructure [2] that spans the cell not only spatially as RNAs move across the cell, but also temporally as the RNAs regulate gene processes during the cell cycle. Thus, the regulation of RNA processes may not only be transcriptional or translational, but also from their biogenesis and processing pathways [2]. However, when talking about gene regulation, it is RNAi that immediately comes to mind (especially in multicellular organisms) and it appears that RNAi-based ncRNAs and some longer ncRNAs have roles in epigenetic processes [5]. Some of these roles have been known for some time (e.g. X-chromosome inactivation [6] and gene imprinting [7]) but other roles in non-developmental mechanisms and cancer are only just coming to light.

We can cover only some of these mechanisms here but further reviews are available [5,7–10]. Although work in this area has clearly concentrated on mammalian examples there are many interesting mechanisms coming to light from non-mammalian species which we will cover to a small extent here. Presently we can divide the epigenetic-related classes of ncRNAs into two main groups; the long ncRNAs, and short ncRNAs including miRNAs, siRNAs, and Piwi-interacting RNAs (piRNAs). This chapter reviews both the long and short classes of ncRNAs involved in epigenetic regulation: those that generally act as *cis*-acting silencers, but also as *trans*-acting regulators of site specific modification and imprinted gene-silencing (Table 4.1). As the examples in the following sections will show, we are still very much in the early days

Handbook of Epigenetics: The New Molecular and Medical Genetics. DOI: 10.1016/B978-0-12-375709-8.00004-6
© 2011 Elsevier Inc. All rights reserved.

TABLE 4.1 ncRNAs Discussed in This Chapter and Their Abbreviations

ncRNA		Length	Short Description	Suggested Reviews and Examples
miRNA	Micro RNA	21–23 nt	ssRNA folds into dsRNA structure; after processing and binding to RISC complex they target mRNAs to regulate translation.	Reviews [12,13,15,25] Figure 4.1A Rtl1 [16–19] Figure 4.2 miR-290 [20,21]
siRNA	Short interfering or silencing RNA	20–25 nt	Regulate a specific gene using complementary sequence. Post-Transcriptional Gene Silencing (PTGS) and Transcriptional Gene Silencing (TGS) pathways. Plants also use RNA-directed DNA methylation (RdDM).	Reviews [9,25–28] Figure 4.1B FLC gene [9,31,32]
piRNA	Piwi-interacting RNA	27–30 nt	Interact with PIWI proteins for chromatin regulation and transposon silencing. Scan RNAs (scnRNAs) are a type of piRNA.	Reviews [33–36] Figure 4.1C
XiRNAs	XCI inactivation linked small RNAs	24–42 nt	Produced from Xist and Tsix long ncRNAs, required for controlling methylation of the future inactive X chromosome and of the Xist promoter region on the future active X chromosome.	XiRNAs in XCI [45,51,52] Figure 4.3
Long ncRNAs		>200 nt	Many have specific targets and are critical for X chromosome inactivation in mammals (XCI), meiotic sex chromosome inactivation (MSCI), RoX (RNA on X) +system in insects, and Hox gene regulation.	XCI [40,41,45, 49,52] Figure 4.3 MSCI [54,55] RoX [56–58] HOX [63,65] Figure 4.4

of investigating how many characterized ncRNAs work to regulate processes such as RNA editing and methylation.

SHORT ncRNAs AND EPIGENETICS

RNAi is a mechanism by which short double-stranded RNAs (dsRNA) are used for sequence-specific regulation of gene expression, where some of the nucleotides on the ncRNA bind to either the coding or promoter region of an mRNA. This binding interferes with normal mRNA processing and consequently silences the expression of the mRNA. The three major classes are microRNA (miRNA), short interfering RNA (siRNA) and Piwi-interacting RNA (piRNA) which differ in their biogenesis and modes of target regulation [11] (Fig. 4.1). Although best known for roles in regulating mRNA transcripts, these short ncRNAs are also directly involved in other cellular processes including chromatin-mediated gene silencing and DNA rearrangements [2,12]. We will go through each class in turn highlighting how they are different, and review recent studies that indicate their use in epigenetics.

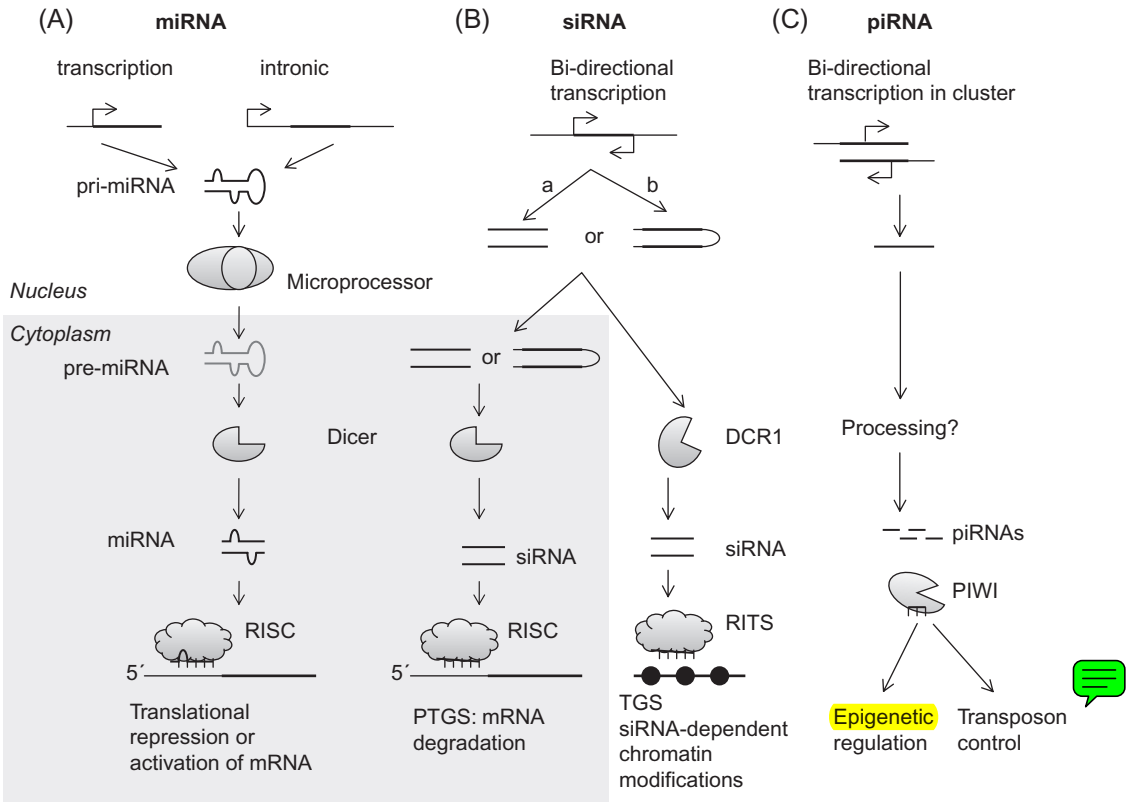

FIGURE 4.1

Processing pathways of small regulatory ncRNAs. (A) miRNAs are initially single-stranded RNAs (ssRNAs) produced via transcription or through splicing, which fold into stem-loop structures to form imperfect double-stranded RNA molecules (dsRNAs). These are then processed by the RNase III endoribonuclease (generally Dicer) before being denatured. One of the RNA strands (usually the less stable of the two) binds to the RNA-induced silencing complex (RISC), which then binds to a specific target mRNA that contains sequence complementary to the miRNA, to induce either cleavage or degradation, or block translation. (B) siRNAs are produced as dsRNAs, and can enter the Post-Transcriptional Gene Silencing (PTGS) pathway, which leads to mRNA degradation in the cytoplasm, or the Transcriptional Gene Silencing (TGS) pathway involved in chromatin modification. (C) piRNAs are ssRNAs produced in clusters and cleaved to individual units through an as yet undefined processing mechanism. They then bind to PIWI proteins to induce epigenetic regulation and transposon control.

miRNAs are perhaps the most well known of the regulatory ncRNA classes, and the general miRNA processing pathway is shown in Figure 4.1. Yet we cannot assume that all miRNAs within one species will regulate their genes in the same way in another species. Studies have shown that not only can a single miRNA down-regulate expression of hundreds of its target genes [13], but some miRNAs use alternative methods of down-regulation, such as accelerated deadenylation of the polyA tail [14]. Other studies revealed that animal miRNAs can induce translational up-regulation, and that some plant miRNAs can function as translational inhibitors contrary to their original functional descriptions [reviewed in Ref. 15]. However, miRNAs are not merely regulating mRNA targets, but are also involved in intricate mechanisms that involve feedback, self-regulation and in some cases methylation.

An example (Fig. 4.2) comes from the mouse Dlk1-Dio3 region in which three protein genes, i.e. delta-like 1 (Dlk1), retrotransposon gene (Rtl1), and Dio3, are expressed exclusively from the paternal chromosome [16]. On the maternal chromosome these protein-coding genes are normally repressed, and several other transcripts are produced including one antisense to the Rtl1 gene. Regional imprinting of Rtl1 is predetermined by the methylation status of the nearby intergenic differentially methylated region (IG-DMR), which is methylated in the paternal chromosome, but not in the maternal. The maternally inherited unmethylated

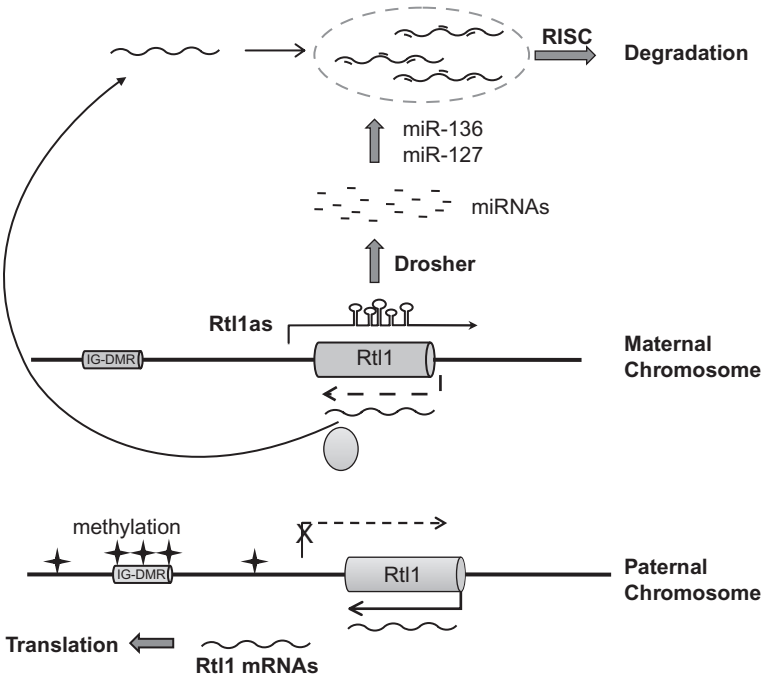

FIGURE 4.2
miRNAs and imprinting. Methylation of the IG-DMR region on the paternal chromosome represses the expression of the Rtl1a antisense transcript, allowing expression of the Rtl1 transcript. Without this methylation, as on the maternal strand, the Rtl1a is processed to produce miRNAs including miR-136 and miR-127, which complementarily bind to the Rtl1 transcript and induce degradation.

state of IG-DMR is essential for maintaining the repression of the protein-coding genes and for the expression of the antisense transcript [16,17]. The Rtl1as (antisense) transcript (also known as antiPeg11) forms hairpin structures from which after processing, miRNAs are released including miR-127 and miR-136 [18]. These miRNAs are located near 2 CpG islands in the Rtl1 transcript, and regulate the expression of Rtl1 *in trans* by guiding RISC-mediated cleavage of any maternal transcript. Aberrant epigenetic reprogramming of miR-127, miR-136, or Rtl1 result in late-fetal and/or neonatal lethality [19].

miRNAs have also been shown to be important in stem-cell self-renewal and differentiation (reviewed in Ref. 12). There are two types of stem cell, tissue stem cells (which include somatic and germline cells which develop, maintain, and repair tissues in developing and adult organisms), and embryonic stem cells (ES) which develop from an embryo to give rise to the fetus. Self-renewal (or self-replication) in tissue cells results in asymmetrical division, whereby one daughter cell retains the stem-cell properties, and the other daughter cell is committed to a differentiated function. This behavior is controlled inter-cellularly (between cells by cell signalling), as well as intra-cellularly through epigenetic, transcriptional, translational, and post-translational mechanisms. Recently, miRNAs have been found to be important players in controlling stem-cell fate and behavior. One example is the mouse miR-290–295 miRNA cluster, a group of miRNAs that share a 5′ proximal AAGUGC motif [20]. The expression of this cluster increases during pre-implantation development and remains high in undifferentiated ES cells, but then decreases after ES cell differentiation [20]. The miR-290 miRNAs act as post-transcriptional regulators of retinoblastoma-like 2 (Rbl2), which in turn acts as a transcriptional repressor of DNA methyl transferases (DNMTs), Dnmt3a and Dnmt3b. DNMTs epigenetically silence OCT4, a key transcription factor of ES cell renewal and differentiation [20,21]. Repression of Dnmt3a and Dnmt3b results in hypomethylation of the genome and especially the telomeres, leading to the appearance of long telomeres and increased telomere recombination. Alternatively, if Dicer is knocked out, miRNAs are depleted and the methylation of the Oct4 promoter is severely impaired during differentiation [20]. Many other candidate targets of the AAGUGC seed-containing miRNAs have been identified as well as many indirectly regulated targets [20], but it remains to be seen how other aspects of self-renewal and differentiation are affected by the miR-290 cluster.

This is only one example of many that show how miRNAs are directly or indirectly regulating key self-renewal or differentiating genes by either directly or indirectly affecting methylation processes. It is also possible for a miRNA to regulate another miRNA. An example of this action is miR-184, which negatively regulates miR-205 in human epithelial cells. Interfering with miR-205 dampens the Akt signaling pathway and is associated with a marked increase in keratinocyte apoptosis and cell death [22]. Current research (e.g. 22–24) is finding that more and more of such miRNAs are subjected to feedback from their target genes, and serve as a warning that what may appear at first to be "simple" regulation of an mRNA by an miRNA, may in fact have hidden features only revealed upon a detailed investigation of a mechanism.

As with miRNAs there are many subclasses of siRNAs that can be processed either as sense–antisense pairs (e.g. bidirectional promoter produced; Fig. 4.1 – siRNA pathway A), or as double-stranded transcripts which are subsequently cleaved by Dicer (Fig. 4.1 – siRNA pathway B) [25]. siRNA-based mechanisms have been well-studied in plants and fission yeast [9,26]. Although at first siRNAs in animals were mostly considered to be from foreign DNA or RNA (i.e. viral-induced exo-siRNAs), recent studies have characterized many more endogenously encoded siRNAs (endo-siRNAs) that appear to have a role in transposon control [reviewed in Ref. 27]. However, concentrating on the more well-known mechanisms in plants, one group of endo-siRNAs are the RDR2-dependent siRNAs which are preferentially associated with transposons, retroelements, and repetitive DNA, but also appear to guide methylation of specific DNA regions [26]. In plants, fission yeast, and to a small extent in mammals [28], both the transcriptional (TGS) as well as post-transcriptional gene silencing (PTGS) pathways are activated by dsRNAs. With the PTGS pathway, siRNAs direct mRNA degradation in the cytoplasm with no epigenetic incidence. However, TGS acts in the nucleus and is associated with chromatin modifications that silence transcription, and are maintained throughout the phases of the cell cycle [9].

In fission yeast more complicated models have been characterized. During TGS, the RITS (RNAi-Induced Transcriptional Silencing) complex is similar to RISC in containing Argonaute, but (unlike RISC) RITS localizes exclusively to the nucleus and contains at least one chromatin-binding module called a chromodomain [9]. Bound to a siRNA it mediates sequence specific heterochromatin formation and histone methylation. Recent models propose that RITS and RDRC (RNA-directed RNA polymerase Complex) are recruited to the site of intended heterochromatin formation when their associated siRNAs bind to a nascent RNA being transcribed at that site [29]. Thus, the binding of RITS to chromatin initiates heterochromatin formation which in turn results in TGS. Assembly of heterochromatin at a given genomic site comes with a heritable silencing of transcription. In fission yeast this mechanism is widely used to regulate heterochromatin formation, and a positive-feedback loop involving RDRP couples siRNA production to chromatin modifications [9].

In plants, siRNAs are involved in RNA-directed DNA Methylation (RdDM), which was first observed in viroid infected tobacco plants where sequences similar in sequence to the viral genes became methylated [30]. The exact mechanism for RdDM has not yet been characterized but in a general model [9], the plant specific RNA Polymerase IV is (somehow) recruited to a target genomic site; once there it synthesizes an ssRNA which RDR2 uses as a template to construct dsRNA that is processed by DCL3 (plant dicer) into siRNAs that bind AGO4 proteins. An AGO4 protein bound to an siRNA is thought to form a complex with PollVb and DRM2 to guide DNA and chromatin methylation at the target genomic region [9,26]. One example is the FLC gene (Flowering locus gene C), a key MADS box transcription factor with key cell differentiation roles similar to that of HOX in animals. FLC gene expression is low during flowering in *Arabidopsis thaliana*, maintained by the Polycomb group of silencing proteins. A transposon in an intron of FLC is believed to nucleate formation of silent chromatin by attracting DNA and H3K9 methylation [9,31]. siRNAs complementary to

the 3′ end of the FLC gene have now been detected and their accumulation requires DCL2, RDR2, and PolIVa [32]. However, the siRNAs do not depend on the transposon but instead on antisense transcription of FLC 3′ UTR by a mechanism that is not yet clear [9].

The use of piRNAs (PIWI-interacting RNAs) in epigenetic processes is (like the siRNAs) only just coming under detailed investigation. Although also found in mammals and some ciliates piRNAs have been studied in greater detail in *Drosophila melanogaster*, in both germline and, more recently, soma cells [33,34], where they play critical roles in transposon "control" (i.e. preventing transposon activation and hence keeping the levels of transposons interrupting genes to a minimum) [35]. *Drosophila* piRNAs reside in clusters usually within heterochromatin or at heterochromatin–euchromatin boundaries. These piRNA clusters are repeat-rich regions composed of ancient fragmented transposon copies representing all major classes and element families [35]. Unlike miRNAs and siRNAs, piRNAs are not produced by "Dicing" (Fig. 4.1), but mainly by bi-directional promoters and what is known as the "ping pong" cycle of biogenesis and amplification. This cycle is initiated by primary piRNAs arising from piRNA clusters. Those piRNAs that are antisense to expressed transposons identify and cleave their targets, resulting in a set of new sense piRNAs in an AGO3 complex termed secondary piRNA. The AGO3-bound piRNA targets any transposon target that contains antisense transposon sequences. This cleavage then generates additional antisense piRNAs and the cycle can continue. This forms an effective small ncRNA-based transposon immune system.

piRNAs are now being proposed as possible vectors for carrying epigenetic inheritance [36]. An example comes from *Drosophila* strains that differ in the presence of a specific transposon, where crosses produce sterile progeny (hybrid dysgenesis), but only if the transposon is paternally inherited. Maternally inherited piRNAs are thought to play a role in this transposon silencing [36]. Both PIWI and Aubergine (Aub) proteins are deposited into developing oocytes and accumulate in the pole plasm suggesting a mechanism of transfer of maternal piRNAs into the germ lines of their progeny [36]. piRNA clusters alone have been shown to be insufficient to inactivate some transposons within a single generation. Instead maternally inherited siRNAs appear to prime the "resistance"-type control system at each generation to achieve full immunity. It is also thought that, since environment can influence the content of maternal small RNA populations, these RNAs could epigenetically alter the phenotype of progeny [35].

In mammals, transposon control by TGS occurs using PIWI-type proteins Milli and Miwi2 (Line-1 non-LTR), and IAP (LTR) retrotransposons, along with DNA methylation during embryogenesis in male germ cells (prospermatognia) [35]. Like AGO3, Mili binds preferentially to piRNAs corresponding to transposon sense strands while Miwi2 contains mainly antisense piRNAs. piRNAs in prospermatogonia are derived from transposon rich piRNA clusters. There is evidence for a ping-pong amplification cycle as seen in *Drosophila*, but as yet its involvement in epigenetic inheritance is not characterized. Although the transmission of phenotype via piRNAs has only been demonstrated to date in *Drosophila*, the accumulation of small RNAs in the oocytes of other species is known and opens the way for this phenomenon to be more widespread [35].

LONG ncRNAs AND EPIGENETICS

During the last few years, evidence of complex, long ncRNA mediated epigenetic control systems has increased dramatically [3,37,38]. In a famous example, X chromosome inactivation (XCI) studied largely in mice, ensures only one of the two X chromosomes in XX females is expressed during development, and involves two long ncRNAs: Xist (17 kb) and its antisense transcript Tsix (40 kb) (reviewed in Refs 6,39,40). Xist RNA is expressed at a low level in both females and males before differentiation [41], but upon cell differentiation, Xist RNA coats the future inactive X chromosome (Xi) triggering extensive histone methylation

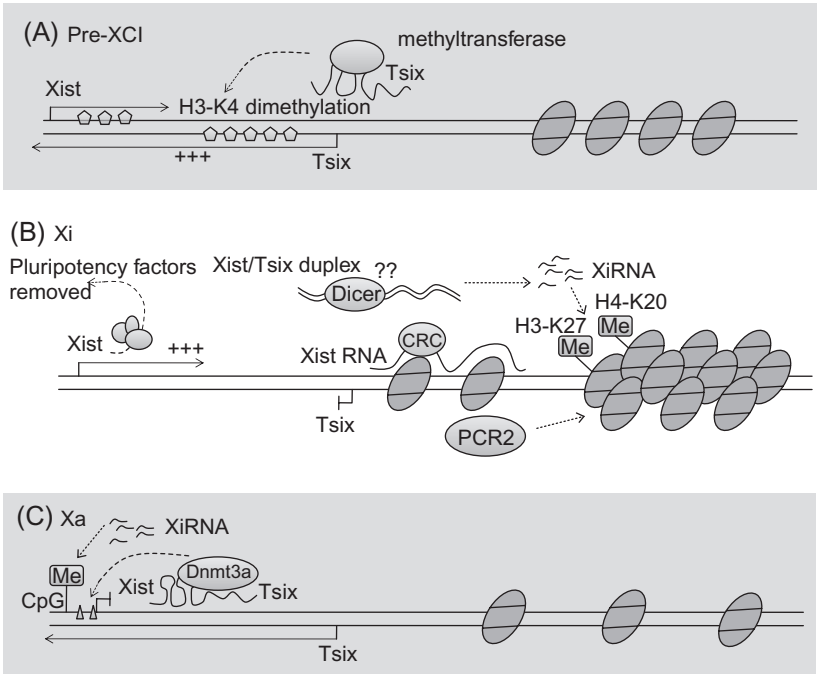

FIGURE 4.3

A general model of placental mammal X Chromosome Inactivation (XCI). (A) Prior to XCI, Tsix is expressed at a high level and triggers H3-K4 dimethylation in itself and the Xist gene, leading to active transcription of Xist and Tsix. This results in an equal chance for transcription and ensures random initiation of XCI. (B) During XCI, expression of Xist is elevated upon removal of the pluripotency factors from the first intron on Xist. Xist RNA then coats the future Xi *in cis* and recruits the chromatin repressive complex (CRC) to Xi. Xist RNA also forms RNA duplex with Tsix RNA and is processed into 24 to 42 nt XiRNAs through the possible action of Dicer. XiRNA then directs H3-K27 trimethylation and H4-K20 monomethylation on the future Xi. The Xi status is maintained by the Polycomb repressive complex PCR2. (C) On the future Xa, Tsix is associated with methyltransferase Dnm3a and directs methylation on the Xist promoter to ensure expression of X-linked genes through repression of Xist. XiRNA is also involved by directing methylation of Xist on CpG islands.

[42], whereas Tsix appears to restrict Xist activity on the future active X chromosome (Xa) [43]. Recent studies, especially in mice, have revealed a more complex regulatory network of XCI which involves the interaction of long and short ncRNAs (Fig. 4.3).

To explain in more detail, in mice pre-XCI embryonic stem cells (ES) (Fig. 4.3A), Tsix is transcribed at a much higher level than Xist and triggers cytosine methylation within both *Tsix* and *Xist* genes, resulting in epigenetically equal competency for transcription and random X-inactivation [44]. The transcriptional level of Xist is elevated when the major pluripotency factors Nanog, Oct3/4, and Sox2 dissociate from intron 1 within *Xist* initiating XCI [44] (Fig. 4.3B and C). The coating of Xist on the future inactive X chromosome (Xi) (Fig. 4.3B) forms a silent chromatin compartment where X-linked genes become "localized" through Xist binding [45]. Xist RNA is required for chromosome-wide methylation in undifferentiated ES cells during the onset of X inactivation; however, once established, the maintenance of the heterochromatic state is independent of Xist RNA [46]. In contrast, the Polycomb repressive complex PCR2 is recruited by the RepA (a 1.6 kb ncRNA within Xist), and responsible for the maintenance of Xi [47,48]. On the future active X chromosome Xa (Fig. 4.3C), the level of *Xist* expression is largely controlled by its antisense transcript Tsix. The expression of Tsix is restricted to Xa [49] and associates with the DNA methyltransferase *Dnmt3a* to direct methylation on *Xist* promoter [50]. However, this methylation event is transient and does not play a role during the initiation of XCI [50]. Besides directing histone modification, the Tsix RNA can also down-regulate *Xist* expression through antisense binding. It has been shown that Tsix transcription across the Xist promoter is crucial for Xist regulation [44].

In addition small ncRNAs are also involved in XCI. Dicer-dependent XiRNAs are produced from both the Xist and Tsix ncRNA transcripts [44] and are required for methylation along the future inactive X chromosome, and methylation of the CpG island of the Xist promoter region in the future active X chromosome Xa [51]. Although XiRNAs are produced with Dicer, RNAi is not directly involved in X chromosome inactivation; instead it appears to maintain the steady-state level of the Xist RNA [52]. Adding to this increasingly complex network, RepA has been found to mediate the heterochromatic configuration of the Xist promoter through recruiting PRC2 [53].

Although most of the studies on XCI have been in placental mammals (and especially mice), recent work in marsupials has shown that a very similar mechanism exists although marsupials do not have the Xist RNA [54]. Here it is thought that male meiotic sex chromosome inactivation (MSCI) plays a greater role in dosage compensation. In mice, (reviewed in Ref. 55), MSCI silencing of the X chromosome genes is initiated during male meiosis, but unlike XCI, MSCI is transient, occurring during each round of spermatogenesis with some X-linked genes reactivating, then subsequently becoming silenced in the female. In marsupials, this can be demonstrated since XCI appears not to result from inheriting an X chromosome already inactivated by MSCI, but instead the inactivation takes place in the female (although the exact timing is not as yet known) [54]. It is also suggested [55] that some X-linked miRNAs escape MSCI and may contribute to the mechanisms regulating MSCI in an RNAi-like manner. Commonalities between the placental mammal and marsupial models (including enrichment of H3K27 trimethylation on the Xi and association of the Xi with the nucleolus [54]), indicate that aspects of the XCI system may be more conserved than originally thought [54].

In insects however, dosage compensation is achieved not by silencing but by a 2-fold increase of X-linked genes in males, relative to females [56–58]. In this mechanism the male-specific-lethal (MSL) complex (consisting of MSL1, MSL2, MSL3, MOF (*males absent on first*), and MLE (*maleless*)) binds to genes along the male X chromosome. Associated with this complex are two long ncRNAs, roX1 and roX2 (*RNA on X*), that direct activation, rather than silencing, of their target genes [56,59]. roX1 and roX2 transcripts spread along the X chromosome recruiting the histone deacetylation protein complex, which generates an open chromatin conformation to facilitate active transcription [60,61]. How roX RNA regulates changes in the localization and activity of the MSL complex, is still poorly understood [62], and likewise how the MSL complex achieves dosage compensation [58]. Studies are beginning to indicate that target genes are enriched at the 3′ end and not at promoter sites leading to a model that the MSL complex affects elongation, resulting perhaps in hyper-transcription of the targeted genes or chromatin looping [58]. What is clear is that the RoX RNAs are a key part of the insect dosage compensation mechanism, and more study is needed to uncover the finer details [58].

Recent studies have also revealed long non-coding RNAs regulating the *Hox* gene cluster in insects and vertebrates (reviewed in Ref. 63). First found in *Drosophila*, the Hox family of proteins are critical determinants of correct patterning of the axis during embryonic development [64]. A large number of non-coding transcripts have been identified within the *Hox* gene cluster [63], the majority of which are found as antisense transcripts from intergenic regions, and are coordinately induced with their 3′-end *Hox* genes [65]. In *Drosophila*, the Bithorax *Hox* gene cluster (BX-C) regulation is extremely complex, with the regulatory region containing enhancers, silencers, maintenance elements, boundary elements, and possibly other elements not yet characterized [63].

Included in this regulation are the long ncRNAs, bxd RNAs, and iab RNAs [64], involved in regulation of their downstream Hox genes, *Ultrabithorax* (*Ubx*), *Abdominal-A* (*abd-A*), and *Abdominal-B* (*abd-B*) [63,66]. Bxd RNAs are expressed in different cells and germ layers, consistent with each bxd ncRNA having a unique role [63]. This spatial regulation may account for the observed mosaic expression pattern of the Hox genes in early embryos

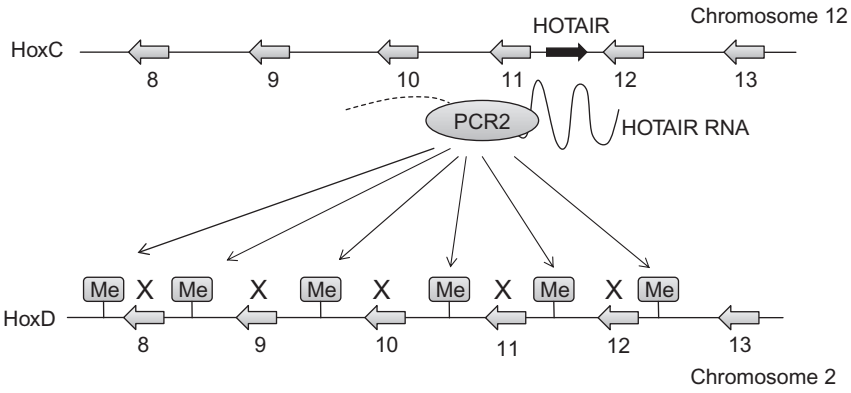

FIGURE 4.4

Long ncRNA regulation of Hox genes. Human HOTAIR RNA is expressed on the antisense strand within the HoxC gene cluster on chromosome 12. The HOTAIR RNA associates with Polycomb repressive complex PCR2 which triggers methylation along the HoxD gene cluster on chromosome 2, leading to silencing of the HoxD genes.

[63,66]. In particular there has been some attention focused on one of these ncRNAs, HOTAIR, identified as regulating chromatin silencing of the adjacent *Hox* locus [65]. Figure 4.4 illustrates the mechanism by which the HOTAIR RNA regulates expression of *HOX* gene clusters through epigenetic control. HOTAIR, a 2158 nt spliced and polyadenylated long ncRNA is transcribed as a single copy on the antisense strand of the *HOXC* gene [65]. siRNA knockdown of HOTAIR results in transcriptional activation of the *HOXD* gene locus spanning four genes on a different chromosome. The HOTAIR RNA is physically associated with the Polycomb Repressive Complex 2 (PCR2), and is required for H3K27me3 modification and transcriptional silencing at HoxD [48,65]. Interestingly, HOTAIR transcription is linked to Polycomb group protein deposition and HOXD silencing on a different chromosome, demonstrating the action of an ncRNA *in trans* [48,65]. This action, but *in cis*, is seen in other long ncRNAs such as RepA, Kcnq1ot1, and AIR [48].

A number of models have been studied to reveal the roles of long ncRNAs in imprinted gene clusters. Short ncRNAs clearly have a role in epigenetic imprinting. In mouse, for example, 80 genes are grouped into clusters [7,67] and in many cases, one or more ncRNAs expressed from within a gene cluster play a crucial role in regulating the expression the gene cluster [68]. This regulation directs chromatin modification forming an "epigenetic memory" within the same cell lineage [69]. Expression of genes in an imprinted cluster is generally controlled by a *cis*-regulatory region, the Imprint Control Element (ICE), which carries parental information in the form of DNA methylation [70]. Several gene clusters controlled by ICE are insulin-like growth factor 2 (*Igf2*), insulin-like growth factor 2 receptor (*Igf2r*), potassium voltage-gated channel (*Kcnq1*), and guanine nucleotide binding protein α stimulating factor (*Gnas*). Each of these clusters carries one ncRNA gene on the parental chromosome with unmethylated ICE [70]. The paternally imprinted *Igf2* cluster contains a 2.5 kb spliced long ncRNA H19, which correlates with the methylation silencing of *Igf2* genes [71], despite not having a direct role in maintaining silencing of the *Igf2* cluster [72]. Although not necessary for the silencing of the *Igf2* cluster, transgenic studies have revealed that H19 expression is sufficient for acquiring paternal-allele-specific methylation of the *Igf2* genes [73]. It is possible that some of these long imprinting ncRNAs are in fact miRNA precursors [7], but as with many of the ncRNAs discussed here, details will emerge as further investigations proceed.

CONCLUSIONS

The investigation of ncRNA-related epigenetic mechanisms is at this point relatively new, but the rise of new sequencing technologies has already revealed epigenetic regulation at the

genome level. Deep-sequencing technologies (also known as Next Generation Sequencing or NGS) have not only enabled analysis of histone modifications and methylation sites across entire genomes [74], but are enabling the detection of ncRNAs important in the regulation of these modifications. An example is Wang et al. (2009) [75] where maize organ specific distributions of canonical miRNAs and endogenous siRNAs have been linked to epigenetic modifications, H3K27me3, and DNA methylation [75]. This provides an important link between the epigenome and the transcriptome.

An interesting titbit is that the role of ncRNA in epigenetics has even been investigated in space. Spaceflight is a unique environment comprising of cosmic irradiation, microgravity, and space magnetic fields. A study of rice plants germinated from seed subjected to spaceflight showed altered methylation patterns and gene expression in six transposable elements and 11 cellular genes including siRNA related proteins Ago1 and Ago4 [76]. All of the detected alterations in the cellular genes were hypermethylation events occurring at CNG sites. This is consistent with the idea that plant CNG methylation is more prone to perturbation by environmental stresses [76,77].

While most studies have been conducted on major model organisms, there is now some information on ncRNA-based epigenetic mechanisms in protist lineages. piRNA-type RNAs (scan RNAs or scnRNAs) from ciliates are produced during the reorganization of the macronucleus during sexual development when some exons can become "scrambled" [35]. In *Tetrahymena thermophila* ~6000 IES internal eliminated sequences consisting of transposon-like and other repeats are targeted for removal by RNA-directed heterochromatin marking by scnRNAs [78]. Although the actual molecular mechanism is not as yet known, scnRNAs pair with either DNA or RNA from the parental macronucleus to be sorted, and then "selected" transcripts are moved to the newly developed macronucleus where they induce heterochromatin formation on the IES prior to elimination.

Recently, long ncRNAs have been found in the malaria parasite *Plasmodium falciparum*, where sterile sense and antisense RNAs are transcribed from the *var* virulence gene family and coat chromatin in a similar way to the animal Xist RNA and the *Drosophila* roX RNAs [79]. The regulation of the VSP genes involved in antigenic switching in the Diplomonad *Giardia lamblia* is also thought to be epigenetically regulated [80]. Subsequently there has been the identification of key RNAi proteins [81,82], some miRNAs [83,84], and a little on the regulatory mechanism of the VSP genes [85]; however, nothing is known as yet about chromatin modifications in *Giardia*. Further studies on protists are essential if we are to understand how ncRNAs in general regulate epigenetics and to understand how such mechanisms evolved.

Despite the individual variances in these pathways, miRNAs, siRNAs, and piRNAs all share several key protein components including Argonaute, PIWI, RDRP, and Dicer. Many RNA-directed epigenetic regulation events thus appear to be sharing protein and RNA components with the RNAi pathway if not dependent on the latter. We also note that evolution of ncRNAs by duplication could allow epigenetic states (e.g. methylation and imprinting) between the two copies to differ [1,86]. Since there are instances where a single *trans*-acting siRNA may have ~2300 predicted gene targets [7], this type of duplication could possibly result in a significant change in phenotype [1]. There is no doubt that the next few years will see more a greater understanding of ncRNA-related epigenetic mechanisms and perhaps then we can move on to constructive evolutionary analysis.

ACKNOWLEDGEMENTS

Many thanks go to David Penny, Peter Lockhart, and all, at the Allan Wilson Centre for Molecular Ecology and Evolution in Palmerston North, for support and encouragement. Funding for this project was from the Allan Wilson Centre, with funding also supplied to LJC by the Health Research Council of New Zealand.

References

1. Collins LJ, Chen XS. Ancestral RNA: the RNA biology of the eukaryotic ancestor. *RNA Biology* 2009;**6**:1–8.
2. Collins LJ, Penny D. The RNA infrastructure: dark matter of the eukaryotic cell? *Trends Genet* 2009;**25**:120–8.
3. Amaral PP, Dinger ME, Mercer TR, Mattick JS. The eukaryotic genome as an RNA machine. *Science* 2008;**319**:1787–9.
4. Mattick JS. The genetic signatures of noncoding RNAs. *PLoS Genet* 2009;**5**:e1000459.
5. Kurokawa R, Rosenfeld MG, Glass CK. Transcriptional regulation through noncoding RNAs and epigenetic modifications. *RNA Biol* 2009;**6**.
6. Chow J, Heard E. X inactivation and the complexities of silencing a sex chromosome. *Curr Opin Cell Biol* 2009;**21**:359–66.
7. Royo H, Cavaille J. Non-coding RNAs in imprinted gene clusters. *Biol Cell* 2008;**100**:149–66.
8. Whitehead J, Pandey GK, Kanduri C. Regulation of the mammalian epigenome by long noncoding RNAs. *Biochim Biophys Acta* 2009;1790:936–47.
9. Verdel A, Vavasseur A, Le Gorrec M, Touat-Todeschini L. Common themes in siRNA-mediated epigenetic silencing pathways. *Int J Dev Biol* 2009;**53**:245–57.
10. Costa FF. Non-coding RNAs, epigenetics and complexity. *Gene* 2008;**410**:9–17.
11. Shrey K, Suchit A, Nishant M, Vibha R. RNA interference: emerging diagnostics and therapeutics tool. *Biochem Biophys Res Commun* 2009;**386**:273–7.
12. Gangaraju VK, Lin H. microRNAs: key regulators of stem cells. *Nat Rev Mol Cell Biol* 2009;**10**:116–25.
13. Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, et al. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* 2005;**433**:769–73.
14. Eulalio A, Huntzinger E, Nishihara T, Rehwinkel J, Fauser M, Izaurralde E. Deadenylation is a widespread effect of miRNA regulation. *RNA* 2009;**15**:21–32.
15. Zhang R, Su B. Small but influential: the role of microRNAs on gene regulatory network and 3′UTR evolution. *J Genet Genomics* 2009;**36**:1–6.
16. Lin SP, Coan P, da Rocha ST, Seitz H, Cavaille J, Teng PW, et al. Differential regulation of imprinting in the murine embryo and placenta by the Dlk1-Dio3 imprinting control region. *Development* 2007;**134**:417–26.
17. Kagami M, Sekita Y, Nishimura G, Irie M, Kato F, Okada M, et al. Deletions and epimutations affecting the human 14q32.2 imprinted region in individuals with paternal and maternal upd(14)-like phenotypes. *Nat Genet* 2008;**40**:237–42.
18. Seitz H, Youngson N, Lin SP, Dalbert S, Paulsen M, Bachellerie JP, et al. Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. *Nat Genet* 2003;**34**:261–2.
19. Cui XS, Zhang DX, Ko YG, Kim NH. Aberrant epigenetic reprogramming of imprinted microRNA-127 and Rtl1 in cloned mouse embryos. *Biochem Biophys Res Commun* 2009;**379**:390–4.
20. Sinkkonen L, Hugenschmidt T, Berninger P, Gaidatzis D, Mohn F, Artus-Revel CG, et al. microRNAs control de novo DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells. *Nat Struct Mol Biol* 2008;**15**:259–67.
21. Benetti R, Gonzalo S, Jaco I, Munoz P, Gonzalez S, Schoeftner S, et al. A mammalian microRNA cluster controls DNA methylation and telomere recombination via Rbl2-dependent regulation of DNA methyltransferases. *Nat Struct Mol Biol* 2008;**15**:268–79.
22. Yu J, Ryan DG, Getsios S, Oliveira-Fernandes M, Fatima A, Lavker RM. MicroRNA-184 antagonizes microRNA-205 to maintain SHIP2 levels in epithelia. *Proc Natl Acad Sci USA* 2008;**105**:19300–5.
23. Yang X, Feng M, Jiang X, Wu Z, Li Z, Aau M, et al. miR-449a and miR-449b are direct transcriptional targets of E2F1 and negatively regulate pRb-E2F1 activity through a feedback loop by targeting CDK6 and CDC25A. *Genes Dev* 2009;**23**:2388–93.
24. Zhao C, Sun G, Li S, Shi Y. A feedback regulatory loop involving microRNA-9 and nuclear receptor TLX in neural stem cell fate determination. *Nat Struct Mol Biol* 2009;**16**:365–71.
25. Carthew RW, Sontheimer EJ. Origins and mechanisms of miRNAs and siRNAs. *Cell* 2009;**136**:642–55.
26. Henderson IR, Jacobsen SE. Epigenetic inheritance in plants. *Nature* 2007;**447**:418–24.
27. Golden DE, Gerbasi VR, Sontheimer EJ. An inside job for siRNAs. *Mol Cell* 2008;**31**:309–12.
28. Kim DH, Saetrom P, Snove O Jr., Rossi JJ. microRNA-directed transcriptional gene silencing in mammalian cells. *Proc Natl Acad Sci USA* 2008;**105**:16230–5.
29. Motamedi MR, Verdel A, Colmenares SU, Gerber SA, Gygi SP, Moazed D. Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell* 2004;**119**:789–802.
30. Wassenegger M, Heimes S, Riedel L, Sanger HL. RNA-directed de novo methylation of genomic sequences in plants. *Cell* 1994;**76**:567–76.

31. Liu J, He Y, Amasino R, Chen X. siRNAs targeting an intronic transposon in the regulation of natural flowering behavior in *Arabidopsis*. *Genes Dev* 2004;**18**:2873–8.
32. Swiezewski S, Crevillen P, Liu F, Ecker JR, Jerzmanowski A, Dean C. Small RNA-mediated chromatin silencing directed to the 3′ region of the Arabidopsis gene encoding the developmental regulator, FLC. *Proc Natl Acad Sci USA* 2007;**104**:3633–8.
33. Li C, Vagin VV, Lee S, Xu J, Ma S, Xi H, et al. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell* 2009;**137**:509–21.
34. Malone CD, Brennecke J, Dus M, Stark A, McCombie WR, Sachidanandam R, et al. Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell* 2009;**137**:522–35.
35. Malone CD, Hannon GJ. Small RNAs as guardians of the genome. *Cell* 2009;**136**:656–68.
36. Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, Hannon GJ. An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* 2008;**322**:1387–92.
37. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 2009;**458**:223–7.
38. Kapranov P, Willingham AT, Gingeras TR. Genome-wide transcription and the implications for genomic organization. *Nat Rev Genet* 2007;**8**:413–23.
39. Lee JT. Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome. *Genes Dev* 2009;**23**:1831–42.
40. Erwin JA, Lee JT. New twists in X-chromosome inactivation. *Curr Opin Cell Biol* 2008;**20**:349–55.
41. Lee JT, Davidow LS, Warshawsky D. Tsix, a gene antisense to Xist at the X-inactivation centre. *Nat Genet* 1999;**21**:400–4.
42. Heard E, Rougeulle C, Arnaud D, Avner P, Allis CD, Spector DL. Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. *Cell* 2001;**107**:727–38.
43. Stavropoulos N, Lu N, Lee JT. A functional role for Tsix transcription in blocking Xist RNA accumulation but not in X-chromosome choice. *Proc Natl Acad Sci USA* 2001;**98**:10232–7.
44. Navarro P, Pichard S, Ciaudo C, Avner P, Rougeulle C. Tsix transcription across the Xist gene alters chromatin conformation without affecting Xist transcription: implications for X-chromosome inactivation. *Genes Dev* 2005;**19**:1474–84.
45. Chaumeil J, Le Baccon P, Wutz A, Heard E. A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev* 2006;**20**:2223–7.
46. Kohlmaier A, Savarese F, Lachner M, Martens J, Jenuwein T, Wutz A. A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. *PLoS Biol* 2004;**2**:E171.
47. Wang J, Mager J, Chen Y, Schneider E, Cross JC, Nagy A, et al. Imprinted X inactivation maintained by a mouse polycomb group gene. *Nat Genet* 2001;**28**:371–5.
48. Gieni RS, Hendzel MJ. Polycomb group protein gene silencing, non-coding RNA, stem cells, and cancer. *Biochem Cell Biol* 2009;**87**:711–46.
49. Lee JT, Lu N. Targeted mutagenesis of Tsix leads to nonrandom X inactivation. *Cell* 1999;**99**:47–57.
50. Sun BK, Deaton AM, Lee JT. A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization. *Mol Cell* 2006;**21**:617–28.
51. Ogawa Y, Sun BK, Lee JT. Intersection of the RNA interference and X-inactivation pathways. *Science* 2008;**320**:1336–41.
52. Kanellopoulou C, Muljo SA, Dimitrov SD, Chen X, Colin C, Plath K, et al. X chromosome inactivation in the absence of dicer. *Proc Natl Acad Sci USA* 2009;**106**:1122–27.
53. Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* 2008;**322**:750–6.
54. Mahadevaiah SK, Royo H, Vandeberg JL, McCarrey JR, Mackay S, Turner JM. Key features of the X inactivation process are conserved between marsupials and eutherians. *Curr Biol* 2009;**19**:1478–84.
55. Yan W, McCarrey JR. Sex chromosome inactivation in the male. *Epigenetics* 2009;**4**:452–6.
56. Bai X, Larschan E, Kwon SY, Badenhorst P, Kuroda MI. Regional control of chromatin organization by noncoding roX RNAs and the NURF remodeling complex in *Drosophila melanogaster*. *Genetics* 2007;**176**:1491–9.
57. 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**:117–22.
58. Ilik I, Akhtar A. roX RNAs: non-coding regulators of the male X chromosome in flies. *RNA Biol* 2009;**6**:113–21.
59. Deng X, Meller VH. Molecularly severe roX1 mutations contribute to dosage compensation in *Drosophila*. *Genesis* 2009;**47**:49–54.

60. Park Y, Kelley RL, Oh H, Kuroda MI, Meller VH. Extent of chromatin spreading determined by roX RNA recruitment of MSL proteins. *Science* 2002;**298**:1620–3.
61. Oh H, Park Y, Kuroda MI. Local spreading of MSL complexes from roX genes on the *Drosophila* X chromosome. *Genes Dev* 2003;**17**:1334–9.
62. Menon DU, Meller VH. Imprinting of the Y chromosome influences dosage compensation in roX1 roX2 *Drosophila melanogaster*. *Genetics* 2009;**183**:811–20.
63. Brock HW, Hodgson JW, Petruk S, Mazo A. Regulatory noncoding RNAs at Hox loci. *Biochem Cell Biol* 2009;**87**:27–34.
64. Lipshitz HD, Peattie DA, Hogness DS. Novel transcripts from the Ultrabithorax domain of the bithorax complex. *Genes Dev* 1987;**1**:307–22.
65. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 2007;**129**:1311–23.
66. Petruk S, Sedkov Y, Brock HW, Mazo A. A model for initiation of mosaic HOX gene expression patterns by non-coding RNAs in early embryos. *RNA Biol* 2007;**4**:1–6.
67. Verona RI, Mann MR, Bartolomei MS. Genomic imprinting: intricacies of epigenetic regulation in clusters. *Annu Rev Cell Dev Biol* 2003;**19**:237–59.
68. O'Neill MJ. The influence of non-coding RNAs on allele-specific gene expression in mammals. *Hum Mol Genet* 2005;**14**(Spec No 1):R113–20.
69. Martin C, Zhang Y. Mechanisms of epigenetic inheritance. *Curr Opin Cell Biol* 2007;**19**:266–72.
70. Lewis A, Reik W. How imprinting centres work. *Cytogenet Genome Res* 2006;**113**:81–9.
71. Feil R, Walter J, Allen ND, Reik W. Developmental control of allelic methylation in the imprinted mouse Igf2 and H19 genes. *Development* 1994;**120**:2933–43.
72. Thorvaldsen JL, Fedoriw AM, Nguyen S, Bartolomei MS. Developmental profile of H19 differentially methylated domain (DMD) deletion alleles reveals multiple roles of the DMD in regulating allelic expression and DNA methylation at the imprinted H19/Igf2 locus. *Mol Cell Biol* 2006;**26**:1245–58.
73. Cranston MJ, Spinka TL, Elson DA, Bartolomei MS. Elucidation of the minimal sequence required to imprint H19 transgenes. *Genomics* 2001;**73**:98–107.
74. Park PJ. Epigenetics meets next-generation sequencing. *Epigenetics* 2008;**3**:318–21.
75. Wang X, Elling AA, Li X, Li N, Peng Z, He G, et al. Genome-wide and organ-specific landscapes of epigenetic modifications and their relationships to mRNA and small RNA transcriptomes in maize. *Plant Cell* 2009;**21**:1053–69.
76. Ou X, Long L, Zhang Y, Xue Y, Liu J, Lin X, et al. Spaceflight induces both transient and heritable alterations in DNA methylation and gene expression in rice (*Oryza sativa* L.). *Mutat Res* 2009;**662**:44–53.
77. Boyko A, Kathiria P, Zemp FJ, Yao Y, Pogribny I, Kovalchuk I. Transgenerational changes in the genome stability and methylation in pathogen-infected plants: (virus-induced plant genome instability). *Nucleic Acids Res* 2007;**35**:1714–25.
78. Kurth HM, Mochizuki K. Non-coding RNA: a bridge between small RNA and DNA. *RNA Biol* 2009;**6**:138–40.
79. Epp C, Li F, Howitt CA, Chookajorn T, Deitsch KW. Chromatin associated sense and antisense noncoding RNAs are transcribed from the *var* gene family of virulence genes of the malaria parasite *Plasmodium falciparum*. *RNA* 2009;**15**:116–27.
80. Kulakova L, Singer SM, Conrad J, Nash TE. Epigenetic mechanisms are involved in the control of *Giardia lamblia* antigenic variation. *Mol Microbiol* 2006;**61**:1533–42.
81. Iyer LM, Anantharaman V, Wolf MY, Aravind L. Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. *Int J Parasitol* 2008;**38**:1–31.
82. Macrae IJ, Li F, Zhou K, Cande WZ, Doudna JA. Structure of Dicer and mechanistic implications for RNAi. *Cold Spring Harb Symp Quant Biol* 2006;**71**:73–80.
83. Chen XS, Collins LJ, Biggs PJ, Penny D. High throughput genome-wide survey of small RNAs from the parasitic protists *Giardia intestinalis* and *Trichomonas vaginalis*. *Genome Biol Evol* 2009;**1**:165–75.
84. Chen XS, White WT, Collins LJ, Penny D. Computational identification of four spliceosomal snRNAs from the deep-branching eukaryote *Giardia intestinalis*. *PLoS ONE* 2008;**3**:e3106.
85. Prucca CG, Slavin I, Quiroga R, Elias EV, Rivero FD, Saura A, et al. Antigenic variation in *Giardia lamblia* is regulated by RNA interference. *Nature* 2008;**456**:750–4.
86. Gu X, Su Z, Huang Y. Simultaneous expansions of microRNAs and protein-coding genes by gene/genome duplications in early vertebrates. *J Exp Zool B Mol Dev Evol* 2009;**312**:164–70.