13
Non-Coding RNAs as Regulators of Transcription and Genome Organization

Katalin Fejes Toth and Greg Hannon

13.1
Introduction

"Junk DNA" was used as the provisional label for the portions of a genome for which no discernible function had been identified [1]. In a 1980 review in *Nature* Leslie Orgel and Francis Crick described junk DNA as having "little specificity and conveying little or no selective advantage to the organism" [2]. For decades, scientist considered the non-protein-coding portions of the genome as dispensable. Protein-coding genes comprise an astonishingly small part of eukaryotic genomes (less than 2% in humans, corresponding to roughly 20 000 genes; Chapter 1). However, as less complex eukaryotes, such as *Caenorhabditis elegans*, have a very similar number of protein-coding genes, it is clear that the developmental and physiological complexity of humans cannot be explained solely by the number of protein-coding genes. Alternative splicing and post-translational modification of proteins increase the diversity and functionality of the proteome likely explaining at least part of its increased complexity. However, it seems evident that much of the regulatory complexity that may contribute to the development of more complex organisms can be established by non-coding RNAs (ncRNAs), as discussed in further detail in Chapter 14.

In accord with the notion that much of the genome represents an evolutionary junk heap, many non-coding transcripts have been proposed to result from leaky or pervasive non-specific transcription. Recent high throughput studies indicate that the transcription of our genomes extends far beyond the limited sequences comprising protein-coding genes, with almost the every base in the genome appearing in non-coding RNAs or ncRNAs. Only about 1.5% of the human and mouse genomes carry protein-coding information, while roughly 60–80% is transcribed into long, polyadenylated transcripts [3, 4], by far surpassing the coding fraction. The non-polyadenylated fraction is also highly complex and extends over a large fraction of total genomic space. It seems highly unlikely that these transcripts are without function.

Genome Organization And Function In The Cell Nucleus, First Edition. Edited by Karsten Rippe.

Published 2011 by WILEY-VCH Verlag GmbH & Co. KGaA

Conserved, non-coding sequences in mammals are estimated to contribute as much as 10% of the genome, much higher than the fraction of protein-coding sequences [5–7]. Similarly, upstream regions and promoters of ncRNAs are also conserved [4]. Generally, small RNAs such as miRNAs are very conserved, while longer ncRNAs such as Xist and Air show less or no conservation [8]. However, some long transcripts, such as MALAT1, show unexpectedly high conservation [9]. It seems that ncRNA sequences, secondary structure and splice site motifs have been subject to purifying selection. [4, 10]. ncRNAs are often developmentally regulated and frequently found next to genes known to be subject to tight transcriptional control. The expression of many ncRNAs is precisely regulated: (i) unannotated intronic genes as well as microRNAs and piRNAs shows high tissue specificity [11–13], (ii) some ncRNAs are activated via signaling pathways, including sonic hedgehog, notch, and BMP [14], and (iii) others respond to retinoic acid treatment [15, 16].

While the roles of some ncRNAs have been characterized, the vast majority still lacks functional annotation. There is an inherent difficulty in functional analysis of ncRNAs. RNA genes are immune to frame-shift or nonsense mutations and are often small and present in multiple copies in the genome. This makes them difficult targets for recessive mutational screens that have provided so much insight into the biology of protein-encoding genes. ncRNAs are also very hard to predict from genomic sequences. Many are highly unstable and have a rapid turnover making biochemical analysis challenging. Interestingly, many of these unstable ncRNAs, often called cryptic transcripts are associated with gene promoters and some have been shown to have functions, raising awareness that what was long considered as "junk" might actually regulate the protein output of our cells. Accordingly, in recent years the transcriptome was increasingly viewed as an RNA machine, wherein most information is expressed as ncRNAs in a developmentally regulated manner to orchestrate the precise patterns of gene expression during mammalian ontogeny [17].

In this chapter, we review several selected classes of non-coding transcripts and their diverse functions in transcriptional regulation, genome architecture, and maintaining genomic integrity. Post-transcriptional gene silencing by small RNAs in the cytoplasm is outside of the scope of this chapter and is described comprehensively in a number of excellent reviews, for example, [18, 19].

13.2 Classification of Non-Coding RNAs

Non-coding RNAs are defined as transcripts with a high density of stop codons that lack extensive open reading frames and are not translated into proteins. Some classes of non-coding RNAs, like transfer RNAs (tRNA) and ribosomal RNAs (rRNA), have been long studied and are well defined in both their structure and function. With recent advances in high-throughput transcriptome analysis techniques, the number of ncRNAs has increased dramatically, but a systematic classification is fraught with difficulties since for the vast majority their function

and mode of action is unknown. As a matter of fact, the current nomenclature is confusing and frequently a single ncRNA can be simultaneously assigned to several different classes. The main classification schemes are based on function, cellular localization, structure/chemical properties, associated proteins, and size. Difficulties are also reflected in the diversity of names both of classes and individual ncRNAs. For example, microRNAs were clearly named according to their size, small nucleolar (sno) RNAs indicate their (usually) nucleolar localization, Piwi-interacting (piRNAs) got their name from their protein partner Piwi, while others like Xist reveal the functional role of this RNA in X inactivation. Some ncRNAs have been functionally classified, for example, miRNAs, siRNAs, tRNAs, and so on, but this task becomes much more complicated in the case of long RNAs with very diverse functions and impossible in the case of the thousands of transcripts, which are still awaiting functional characterization.

An alternative way of classifying ncRNAs is according to their cellular localization. A significant proportion of unannotated ncRNAs is exclusively detected in either the nucleus or cytoplasm. Intronic ncRNAs, such as almost all of the snoRNAs, seem to be predominantly nuclear, but some, like intronically encoded miRNAs, are primarily cytoplasmic. Only very few intronic ncRNAs are found in both compartments. Interestingly, most of the long transcripts observed in the cytoplasm are polyadenylated, while the nuclear fraction seems to be packed with both polyA$^+$ and polyA$^-$ ncRNAs. Some nuclear ncRNAs have been specifically associated with chromatin, the nucleolus, or nuclear bodies, such as Cajal bodies.

Non-coding RNAs also show a great diversity in structure. For example rRNAs have very defined secondary structures and associate with their protein partners in a precisely regulated manner. They undergo extensive RNA editing and folding in the nucleolus to achieve their functional state. The structure of tRNAs is inherently defined by their sequence and the length and relative location of their arms are essential for proper amino acid recognition and translation. miRNA processing also requires a hairpin secondary structure that is recognized by Dicer and defines the cleavage site. Other ncRNAs, such as mature small regulatory RNAs or many long RNAs, have little or no evident structures and their functional role is mainly defined by their sequence (Figure 13.1).

Some ncRNAs can be classified according to their biogenesis, processing and chemical properties. Depending on which RNA polymerase or processing enzyme is responsible for their transcription and maturation, these ncRNAs possess different terminal structures. Primary RNA polymerase (RNAP) II transcripts bear a 5' cap structure and many are polyadenylated on the 3' end. Interestingly, it seems like capping does not have to occur cotranscriptionally and recently long and small RNAs have been described that are likely derived from long primary transcripts by cleavage and subsequent capping [20]. Transcripts from RNAP I and RNAP III carry triphosphate 5' ends. Small regulatory RNAs, such as miRNAs and siRNAs, are processed from longer transcripts through endonucleolytic cleavage by a type III endonuclease (RNAseIII-like) resulting in characteristic 5' monophosphate ends. piRNAs, another class of small RNAs, also have this 5' monophosphate end, but in addition they also carry a 2' O-methyl group on their 3' ribose [21, 22]. In

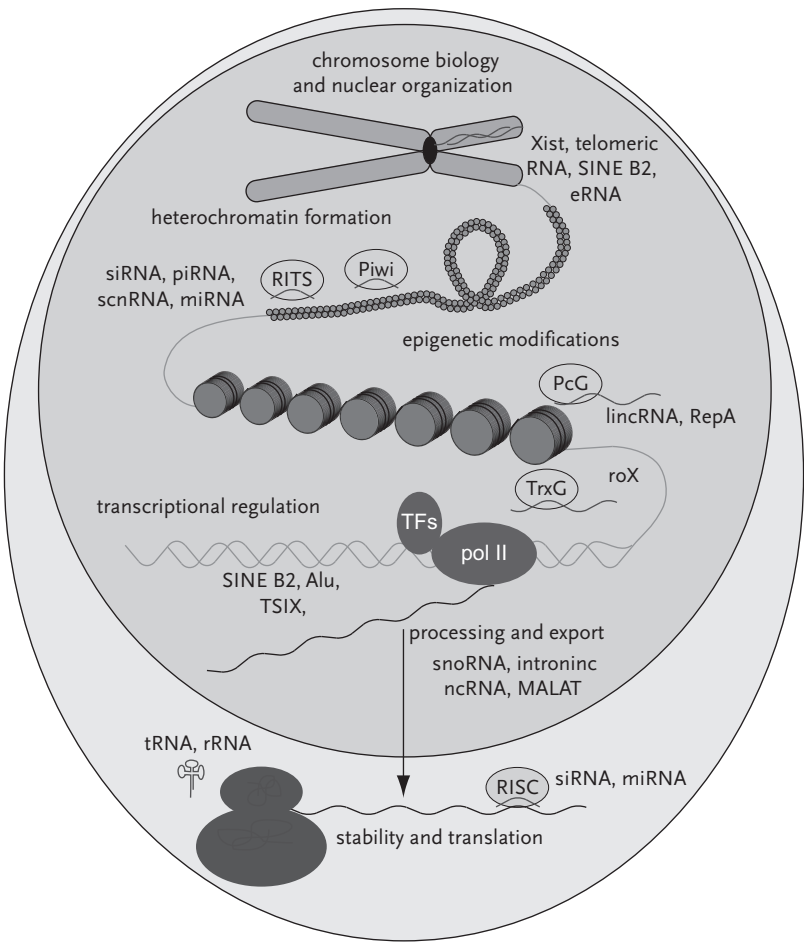


Figure 13.1 ncRNAs involved in eukaryotic gene expression.

most cases and especially for long ncRNAs, their biogenesis and structures still remain unresolved and in some instances ncRNAs of the same class might even be synthesized by different mechanisms. For example, precursors of siRNAs in many organisms such as *Schizosaccharomyces pombe*, *C. elegans* and plants are first synthesized by RNAP II, but then this primary transcript is used as a template by the RNA-dependent RNA polymerase (RDRP) to generate a second strand. The processing of the two complementary strands results in two small RNAs from the same class having two different biogenesis mechanisms.

The most robust overall classification of ncRNAs, though perhaps the least informative, is based on their size (Figure 13.2). Generally, RNAs have been divided into small and long RNAs using an arbitrary threshold around 200 nt. Within the small RNAs two further groups have been distinguished. A group of very short RNAs, ranging from 18 to 30 nt in length, are represented by the recently discovered microRNAs (miRNA), small interfering (si) RNAs, and piwi-

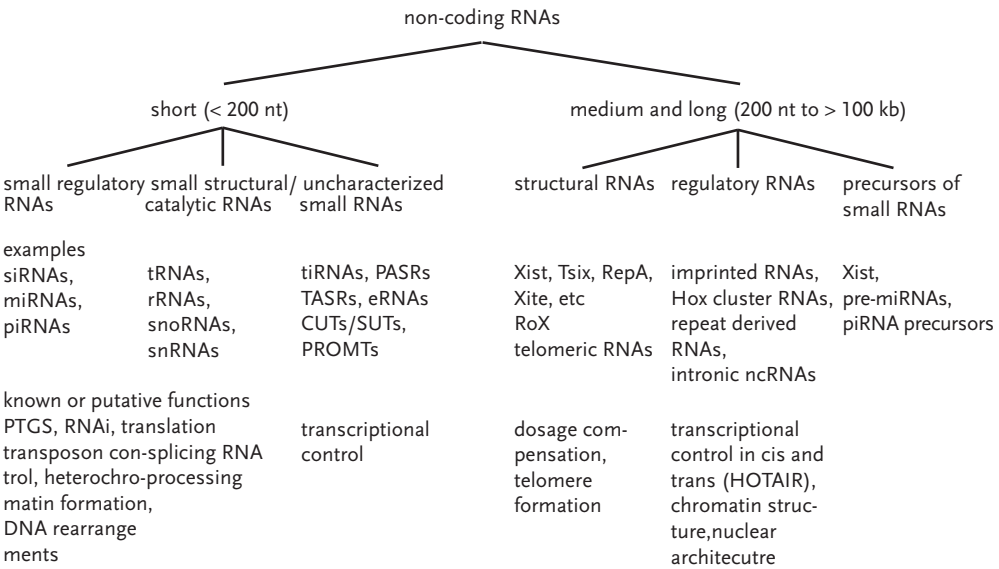


Figure 13.2 Classification of ncRNAs according to size.

interacting (pi) RNAs. These three main classes are well distinguished in terms of biogenesis mechanisms, protein partners, and functions in transcription, chromatin structure, mRNA stability, and translational control. The RNAs ranging in size between 30 and 200 nt include tRNAs, small nuclear (sn) RNAs, and small nucleolar (sno) RNAs, whose function and structure are more or less well known. The recently described genic ncRNAs such as promoter-associated small RNAs (PASRs), terminus associated small RNAs (TASRs), and a wide group of intra- and intergenic RNAs are less well characterized [20, 23]. The latter groups have been identified through genome-wide transcriptome analysis efforts, and they likely represent a pool of multiple distinct functional entities. The class of medium and large RNAs ranging from about 300 nt to over 100 kb in size is the most diverse in length, structure, and function. They participate in imprinting, dosage compensation, DNA methylation, regulation of transcription, chromatin structure and cytoskeleton, and many other processes (see Sections 13.4–13.6).

Multiple examples exist which demonstrate that one class of ncRNAs can regulate the expression or processing of another class. In addition, many long ncRNAs also serve as precursors for other classes of ncRNA. For example, MALAT1 and Xist, two long ncRNAs are processed into small RNAs, and it is not currently clear whether the long or the short transcript or both are functional [24, 25]. Another interesting example is the 2.4-kb unspliced, polyadenylated nuclear ncRNA mrh1 in mouse, which is processed by Drosha to yield an 80-nt small RNA. This small RNA is retained in the nucleus and is thus not processed further by Dicer. It associates with discrete chromatin foci, although its role is not fully understood [26].

In some cases long ncRNAs interact with small RNAs to modulate small RNA action. For example, in *Arabidopsis* the 550-nt IPS1 ncRNA is poorly conserved

except for a 23-nt site that is complementary to mir399 with a mismatched at its expected cleavage site [27]. This mismatch leads to a non-cleavable product that can sequester the microRNA thereby acting as a competitive inhibitor. Rncs-1 is an 800-nt transcript in *C. elegans*, which does not serve as substrate for, but instead inhibits Dicer activity *in trans*. Overexpression or deletion of rncs-1 leads to decrease or increase of certain siRNAs respectively [28]. Finally, antisense transcripts can lead to processing of functional protein-coding mRNAs into small RNAs. Transcripts from pseudogenes can cause the cleavage of the corresponding mRNA, thereby down-regulating gene activity [29, 30]. Similarly, natural antisense transcripts have been shown to generate dsRNA with overlapping genes leading to enod-siRNA production in flies [31–34]. There is also evidence for small RNAs that are complementary to protein-coding genes and transcribed by RNA polymerase (RNAP) III leading to a sense–antisense-based regulatory network in which RNAP III transcripts control their RNAP II counterparts. An example is the ncRNA 21A that regulates the expression of CENP-F *in trans* by a complementarity-based mechanism [35].

To make the annotation chaos worse, some RNAs have been previously defined as ncRNAs, but turned out to encode for proteins or peptides. The tarsal-less (tal) gene in *Drosophila* encodes a 1.5-kb transcript that only contains open reading frames (ORFs) shorter than 50 amino acids (aa). Although it was originally classified as a ncRNA, it was later found that several ORFs of 33 nt or shorter are translated into 11 amino acid peptides, which are key regulators of morphogenesis and pattern formation [36]. In other cases a single gene can encode both proteins and ncRNAs. The steroid receptor RNA activator (SRA), for example, has multiple isoforms, some of which can be translated. Both the protein and the ncRNA affect the transcription enhancing activity of the estrogen receptor in breast cancer cells [37]. A similarly intriguing case is the *Drosophila* Oscar gene, which apart from its protein coding mRNA also gives rise to a ncRNA from its 3'UTR. This transcript is required for *Drosophila* oogenesis independently of the protein-coding capacity of the gene [38]. In mice many 3'UTRs are expressed independently and discordantly from their mRNAs in a developmentally regulated fashion.

In summary, several approaches exists to classify ncRNAs all of them with shortcomings. Due to the many gaps in our knowledge with respect to the function, biogenesis, and structure of these molecules, the current nomenclature is, at best, a work in progress.

13.3
Small Regulatory RNAs and Their Diverse Nuclear Functions

13.3.1
Heterochromatin Formation and Maintenance by siRNAs in Yeast and Plants

Nuclear regulation of gene expression by small RNAs via heterochromatin formation is best understood in *S. pombe* and plants. Heterochromatin formation depends on the di/trimethylation of lysine 9 of histone H3 (H3K9me2/3) and the

binding of heterochromatin protein 1 (HP1) and its yeast homologs Swi6, Chp1, and Chp2 [39, 40] (Chapter 8). Deletion experiments in fission yeast have pointed to an important role of RNA interference in providing the sequence specificity of this process, at least at some critical genomic sites [41–44]. Two complexes have been found essential in yeast for small RNA formation and function: RNA-induced transcriptional silencing (RITS) and the RNA-dependent RNA polymerase complex (RDRC). RITS consists of Ago1, the yeast Argonaute protein carrying slicer activity, Chp1, the adapter protein Tas3, and the associated siRNA [44]. Chp1 is a structural component of heterochromatin, and its deletion leads to loss of H3K9me2/3 [39, 40, 45]. RDRC consists of Rdp1, the RNA-dependent RNA polymerase Hrr1, a putative RNA helicase, and Cid12, a polyA-polymerase [46].

In the originally proposed model for RNA-directed heterochromatin formation RITS binds to chromatin as guided by an siRNA, most likely through RNA : RNA base pairing with a nascent chromatin-tethered transcript. This leads to recruitment of the Clr4 histone methyltransferase and methylation of histone H3K9. Methylation in turn attracts Swi6 and leads to heterochromatin formation [39, 40]. Recent findings have increased the complexity of this model: RITS, RDRC, the RNAP II complex, and Clr4 all assemble on chromatin and their tight interaction is needed for heterochromatin initiation, assuring that only regions with both corresponding small RNAs and histone methylation are silenced. RITS can bind cooperatively to chromatin through: (i) the siRNA recognizing its complementary nascent RNA sequence, (ii) Chp1 recognizing methylated histones, and (iii) Ago1 interaction with the RNAP II C-terminal domain (CTD), the deletion of which disrupts centromeric heterochromatin [47]. RITS binding to chromatin recruits Clr4 and the simultaneous presence of both Clr4 and the siRNA leads to binding of the RDRC complex [46]. RDRC then synthesizes the second strand complementary to the locally transcribed nascent transcript. Dcr1, the yeast dicer homolog binds to RDRC at the site of transcription and processes the long double-stranded transcript into new siRNAs, creating a positive feedback loop [48]. RDRP is only recruited by chromatin-associated RITS. This leads to the chromatin dependence of siRNA biogenesis and assures that only transcripts targeted for heterochromatin formation produce siRNAs [48]. Three processes ensure that transcription at these regions does not result in mature transcripts:

1. The RNAi machinery itself cuts up transcripts to produce more siRNAs [49, 50].
2. A RNAi-independent surveillance machinery involving the TRAMP polyadenylation complex and the exosome competes with the RNAi machinery for nascent non-coding transcripts and degrades them [51–54].
3. Transcription in heterochromatin is cell cycle-dependent [55, 56].

During the G2/M phase serine at position 10 of histone H3 is phosphorylated and the condensin subunit cut3 is recruited to pericentromeric heterochromatin repeats. This leads to Swi6 dissociation at these sites allowing transcription. In S phase Cut3 levels drop and Ago1 and Rik1 (a Clr4 component) bind leading to reestablishment of H3K9me2/3 and Swi6 binding [57–60]. During G1 phase

transcription is enabled leading to the recruitment of RITS and the Clr4 complex, which then facilitates heterochromatin formation [55, 56, 61].

Contrary to pericentromeric regions, which fully depend on the RNAi machinery, deletion of RITS or RDRC components does not impair heterochromatin formation at the mating-type locus and the telomeres, indicating the existence of an alternative RNA-independent pathway, which needs to be inhibited simultaneously to obtain a complete loss of heterochromatin [41, 57, 62, 63] (Figure 13.3).

Plants also utilize siRNAs to establish repressive chromatin at repetitive regions. Contrary to yeast, heterochromatin is marked by DNA methylation. Plant DNA

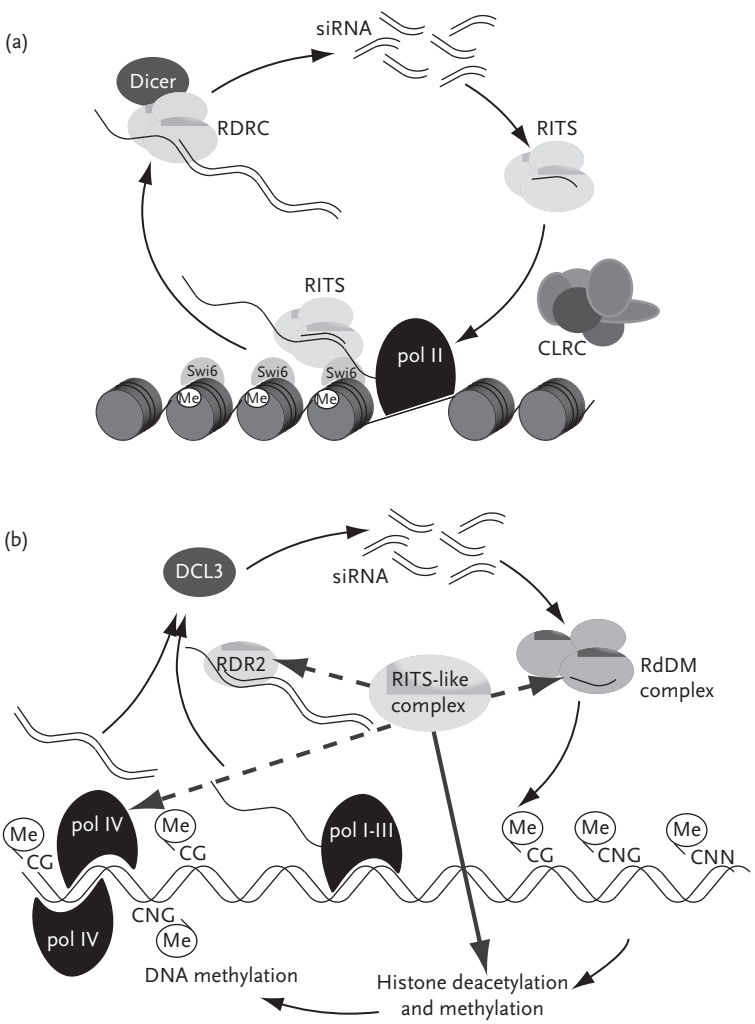


Figure 13.3 RNAi-mediated heterochromatin formation in *S. pombe* (a) and plants (b).

methylation occurs throughout the genome but mainly on repetitive sequences. It depends on the DNA methyltransferase DRM2, AGO4, DCL3 (a Dicer protein), RDR2 (RNA-dependent RNA polymerase), a histone deacetylase, a histone methyltransferase homologous to Clr4, the DRD1 chromatin-remodeling factor, and a plant-specific DNA-dependent RNA polymerase IV. The latter has two isoforms one involved in siRNA biogenesis, while the other binds to AGO4 and this interaction is essential for DNA methylation [64]. It is thought that RDR2 and/or RNAP IV recognize aberrant transcripts, convert them into dsRNAs, which then are processed into siRNAs by DCL3 and loaded onto AGO4. AGO4 forms a complex with the second isoform of RNAP IV and DRM2 and guides DNA methylation to sites from which the transcripts were derived [65]. Currently it is unclear whether the siRNAs in AGO4 base pair with nascent transcript such as in *S. pombe* or directly with DNA or possibly both.

In plants siRNA-mediated gene silencing is not limited to repetitive sequences. Constitutive expression of dsRNA mapping to promoter regions results in production of corresponding siRNAs and in *de novo* DNA methylation and gene silencing [66, 67]. This indicates a more general function of small RNA pathways in plant transcriptional control. In addition, organisms have developed strategies to utilize small RNA pathways by endowing repeats important functions. In fission yeast the pericentromeric repeats are important for proper chromosome segregation by initiating the RNAi-dependent heterochromatin formation. In plants repeats can influence development by nucleating RNAi-dependent methylation and silencing of surrounding protein-coding genes. Finally, plants siRNAs seem to regulate the expression of transposable elements (TEs) in pollen [68], similarly to piRNAs in animal germ cells (see below). Pollen consists of two nuclei, one that will form the sperm and be transmitted to the offspring, and an accompanying vegetative nucleus which will not contribute its DNA to the progeny. TEs are reactivated and transpose in the vegetative nucleus of the pollen and contribute to the formation of mature siRNAs, which can freely diffuse into the sperm nucleus. It is thought, that this protects the sperm from activation of TEs [68] and establishes the correct methylation pattern in the offspring. This function would possibly correspond to the maternally deposited piRNAs observed in flies, which are thought to direct transposon repression in the offspring.

13.3.2
Targeting of Chromatin Signatures by Endogenous siRNAs and piRNAs in Animals

Small RNAs are also emerging as important regulators of chromatin structure, nuclear organization, and transcription in animals. Deep sequencing of small RNA populations has revealed that a large fraction of endo-siRNAs in flies and mammals match transposons and other repeats [29–31, 33, 34]. Similar to yeast, inactivation of Dicer in mammals leads to aberrant heterochromatin formation, suggesting that siRNAs might participate in heterochromatin formation. In mice, knockout of Dicer leads to accumulation of satellite repeat RNA and L1 transposons, loss of heterochromatin, and chromosome defects. Similar observations have

been made in chicken cell lines containing a human chromosome 21 [69–71]. Furthermore, recent studies indicate that the spreading of silencing on the inactive X chromosome requires the transcription of LINE transposable elements and their processing into siRNAs [72]. This points to a role of small RNAs for regulating large-scale chromatin structure. However, the mechanism by which small RNAs silence transposons and establish heterochromatin is currently unknown. It seems that, like the model described in *S. pombe*, small RNA complexes might closely interact with chromatin-modifying machineries. In *C. elegans* transgene-induced gene silencing requires both chromatin modifiers and RNAi [73, 74]. The connection between the two is probably not limited to repeat-induced silencing [75] but might play a role in the regulation of gene expression.

In animals, a specialized small RNA pathway, called the piRNA pathway, functions to suppress transposable elements. It is principally active in the gonads and is required for maintaining germline integrity [76, 77]. The core of the pathway consists of members of the piwi clade of Argonaute proteins and their associated small RNAs, called piwi-interacting (pi)RNAs, that are 24–30 nt in length. These do not depend on Dicer and Drosha for their biogenesis, mostly map to transposable elements, and are far more diverse in sequence than siRNAs and miRNAs. Although the best studied function of the piwi pathway is post-transcriptional suppression in the cytoplasm, recent studies point to an additional role in the nucleus. In all animals at least one member of the family shows nuclear localization. In addition, Piwi, the founding member of the clade, has been found to interact with HP1 and binds to chromatin [78].

In mouse, like in plants, transposons are transcriptionally silenced by DNA methylation. As TEs constitute about 40% of the mouse genome, substantial changes in the methylation state of TEs might greatly alter overall chromatin structure in addition to the damage that expression of transposons might cause to the genome. TE repression and the corresponding DNA methylation is lacking in mice deficient in Mili or Miwi2, both of which are mouse piwi proteins [79–82]. The piRNA pathway is highly active during the prenatal developmental stage during which DNA methylation is established on transposons. The expression timing of one of these proteins, the nuclear Miwi2, exactly overlaps the developmental window during which *de novo* methylation of repeats occurs [82]. As piRNA populations are not affected by mutations in the DNA methylation machinery they likely act upstream to direct methylation to target sites.

Beyond the regulation of heterochromatin formation, small RNAs also influence nuclear organization. Nuclear positioning seems to be regulated by the RNAi machinery both in fission yeast and in flies [83–85], although the mechanisms underlying this effect are poorly understood. Dicer mutants show severe chromosome segregation defects during oocyte maturation indicating that Dicer products either directly regulate chromosome segregation or act indirectly by regulating genes that are important for segregation [70]. In some cases small RNAs are more likely to play an indirect role in regulating chromosome structure, as for example many endo-siRNA target genes in mouse oocytes are involved in controlling microtubule dynamics [29]. Similarly, mutations in the mouse Piwi

family members Mili and Miwi2 lead to arrest in spermatogenesis during meiosis with chromosome alignment defects [81, 86]. This might be due to a secondary effect of activating the DNA damage checkpoint; in *Drosophila* polarity and axis specification defects observed in Aubergine mutants are suppressed by the simultaneous mutation of ATR and Chk2 checkpoint kinases, which are responsible for meiotic arrest in the presence of non-repaired double-strand breaks [87].

13.3.3
DNA Rearrangement/Elimination in Ciliates Involves Scanning RNAs

The most extreme form of gene silencing is seen in some unicellular organisms that literally eliminate substantial unnecessary genomic information from their "somatic" nuclei. *Tetrahymena*, *Paramecium*, and other unicellular organisms have an unconventional genetic configuration, possessing a germline genome, located in the micronucleus (mic), which is transcriptionally silent during vegetative growth, and a somatic genome, found in the macronucleus (mac) from which genes are expressed [88, 89]. The genomes of the two nuclei greatly differ both in content and in structure. The macronucleus has active histone modifications and variants, while the micronucleus is devoid of most chromatin modifications [90–94]. In addition, the macronucleus contains multiple copies of fragmented chromosomes, which substantially lack repetitive sequences that are common in the micronucleus. During mating, the existing somatic mac of both parents is destroyed and the new mac and mic are formed from zygotic nuclei containing copies of the germline-derived genomes [95, 96]. Interestingly, *Tetrahymena* uses RNA guides to target the heterochromatin modifications to the thousands of loci to be excised (Figure 13.4).

The source of these guide RNAs appears to be bidirectional transcription of the germline genome during meiosis and subsequent cleavage of the dsRNAs by a Dicer-like enzyme. This generates 27–30 nt small RNAs, called scan RNAs or scnRNAs, that associate with the Piwi family protein, Twi1p [97–101]. scnRNAs are produced prior to formation of the zygotic genome but submit information to the newly forming somatic genome several hours later, guiding genome reorganization. Although the mechanism by which the small RNAs direct genome rearrangement and elimination is still elusive, some data suggests that the RNAi machinery interacts with the replication fork and acts as a histone-deposition apparatus. Deposition of H3K9 methylation marks seems crucial for accurate excision since mutation of the Dicer-like protein DCL1 lead to an aberrant H3K9 methylation pattern and inaccurate excision [98].

The bidirectional transcription leading to small RNA formation is promiscuous and extends beyond germline-limited sequences [99], raising a question of how resulting RNAs can guide precise excision of only a subset of sequences. The answer is proposed to lie in the interaction of the small RNAs with the parental macronucleus. According to the current model, prior to formation of the zygotic genome the Twi1p-bound small RNAs visit the old mac and their sequences are compared to the genome (or transcripts derived from it). Small RNAs

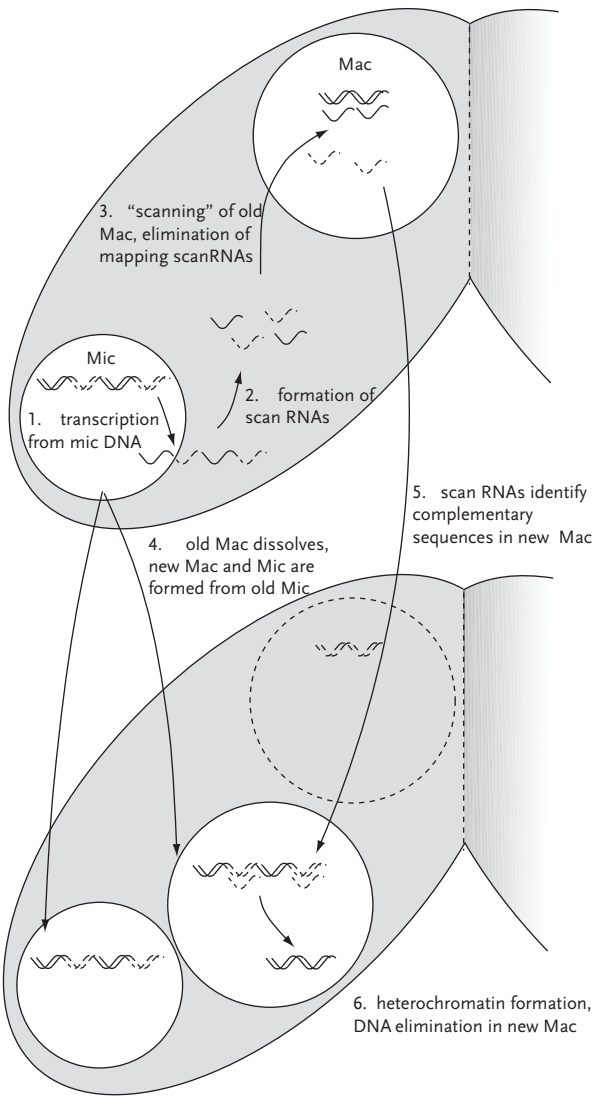


Figure 13.4 ScnRNAs regulate programmed DNA elimination in ciliates.

corresponding to sequences within the old mac are removed from the small RNA pool that will later direct heterochromatin formation and DNA elimination [100, 102], assuring that genomic sequences required for somatic functions are maintained in the mac of the next generation.

In *Oxytricha trifallax* 95% of the germline genome is destroyed during macronucleus formation and the remaining fragmented chromosomes are unscrambled in a precise manner through permutation or inversion [103]. There is strong

support for an RNA-template model for the rearrangements involving intracellular genome comparisons by means of RNA transcripts from the maternal macronucleus [104]. Indeed, injection of RNA (or DNA) sequences corresponding to modified rearrangement products during conjugation leads to altered rearrangement of macronuclear genome. This in turn leads to stable epigenetic inheritance of alternative DNA rearrangements [105]. These data indicate that both genome elimination and rearrangement in ciliates is directed by small non-coding RNAs and depends on the parental macronuclei to specify target loci.

13.3.4
Transcriptional Repression Mediated by siRNAs and miRNAs

Most studies have found that small RNA pathways impact transcription through alterations in local chromatin state via histone modifications (Chapter 4) or DNA methylation (Chapter 2). It is thought that low level of bidirectional transcription through loci results in the production of dsRNA, which gets chopped into siRNAs that associate with their Argonaute partners. Subsequently, these complexes target different chromatin modifying factors such as HDACs, histone methylases and possibly DNA methyltransferase to silence a given locus. Transcriptional silencing by exogenous siRNAs is accompanied by an accumulation of Argonaute 1 at target promoters followed by increase of H3K9 dimethylations and H3K27-trimethylation [106], while knockdown of Ago1 results in loss of H3K9me2/3 from targeted loci. PcG proteins such as EzH2 also associate with silenced promoters upon targeting by exogenous siRNAs [106]. In addition Ago1 was found to colocalize with EzH2 and H3K27-me3 at natural PcG target sites such as the MYT gene. Maintenance of heterochromatin region between the chicken beta-globin locus and the folate receptor gene also requires the expression of Ago2 and Dicer. Lack of either of these resulted in increased accessibility of DNA to restriction digest and increased H3K4 acetylation. This locus also shows enrichment of Ago2 in ChIP experiments compared to the neighboring coding regions [107].

In some cases the small RNAs were observed to also mediate DNA methylation. Small RNAs were shown to co-immunoprecipitate with DNMT3a at the promoter of some genes targeted by exogenous siRNAs [108]. Similarly DNMT3a recruitment to the EF1a promoter was dependent on siRNAs. At this site low copy-number EF1a transcripts initiating further upstream and spanning the promoter region were found, and blockade of this transcript impaired siRNA-directed silencing [109]. It was proposed, that siRNAs bind to nascent transcripts from the promoter and direct Dnmt3a to the site. The molecular mechanism underlying this phenomenon, however, is still not clear and further investigation is needed to determine whether the small RNA complex directly interacts with the DNA methyltransferase or induced histone modifications result in Dnmt3a recruitment.

Silencing can also occur in an Ago-2 dependent fashion without the involvement of epigenetic changes through obstruction of RNA polymerase binding if a small RNA directly targets the transcription initiation site [110–112]. Interestingly, in some studies, it was shown that only an antisense small RNA is capable of

inhibition [106, 108]. siRNAs and the RNAi machinery have also been shown to interfere with transcriptional elongation in *C. elegans* [113]. The nuclear NRDE-3 Argonaute protein targets the nuclear RNAi defective-2 (NRDE-2) protein to nascent transcripts corresponding to the associated siRNA, This association leads to accumulation of RNAP II at target sites and a decreased occupancy and transcriptional activity downstream of targets.

Do small RNAs target nascent transcripts or do they interact directly with chromatin? Although numerous arguments have been made supporting each model, recent findings such as the strand specificity of initiation repression and the direct inhibition of elongation argue for recognition of nascent transcript. Furthermore, inhibition of RNAP II with alpha-amanitin impairs siRNA-mediated H3K9 methylation indicating that transcription is probably required for siRNA-directed transcriptional gene silencing (TGS) [108]. Finally, RNAP II has been shown to co-immunoprecipitate with Ago-1 at RNA-targeted gene promoters in mammalian cell culture [106]. Nevertheless, as was proposed in plants, it is quite possible that both the interaction with chromatin and nascent RNA play a role and that different machineries utilize one or the other mechanism.

It is worth mentioning that the other well-studied class of small regulatory RNAs, microRNAs, also appear to regulate gene expression on the transcriptional level. Although most miRNAs act in the cytoplasm, some plant miRNAs may directly promote DNA methylation [114, 115]. Recent studies describe a role of miRNAs in transcriptional gene silencing and promoting heterochromatin formation in human cells [116, 117]. The mechanism of such transcriptional regulation remains elusive but it could occur through the tethering of effector complexes.

13.3.5 siRNA-Mediated Activation

In contrast to the aforementioned observations of small RNA-mediated repression, siRNAs have recently been shown to also activate transcription of a select set of genes. A few prominent examples are the E-cadherin, p21, and progesterone receptor genes [118–120], of which the best studied is the progesterone receptor (PR). Initial reports showed suppression of PR by a synthetic small RNA duplex but under certain conditions the same duplex was shown to activate. It seems that the effect of the small RNA depends on the expression level of the mRNA itself: in cells that highly express PR such as T47D, small RNAs seem to repress, while changing the growth conditions to obtain lower PR expression or using cells with low PR expression such as MCF7 results in small RNA-induced activation [110, 111, 118, 120].

Currently, two models have been proposed for explaining the dual small RNA function of activation and repression. According to the first model, long antisense transcripts are the key regulators of gene expression, and the exogenous small RNAs target these antisense transcripts. Accordingly, the effect of the small RNA depends on the original level of the antisense transcript. The exact mechanism by

which the long transcripts regulate the expression of the coding gene is not known. The second model suggests that an imbalance in bidirectional transcription levels may determine whether a promoter-targeted siRNA results in activation or silencing of transcription [121, 122]. In the case of p21, steady-state endogenous expression of p21 is associated with comparable levels of both sense and antisense transcripts across the promoter region. It was proposed that reducing the level of antisense transcripts by exogenous siRNAs would result in a reduction of the H3K27 methylation mark and an increase in transcription [119, 121]. Current data indicate that this activating effect of small RNAs is characteristic for promoters with low GC content and high complexity [106, 120] implying that subsets of genes might utilize very different mechanisms for regulating gene expression and long antisense transcripts might be restricted to only a few genes. This could possibly explain the varying effect on gene expression observed with exogenous small RNAs.

13.3.6
Promoter- and Gene-Derived Transcripts in Yeast and Animals

Recent genome-wide transcriptome analysis resulted in the recognition of genic and promoter-associated long and small RNAs in many species. These have been designated as cryptic unstable transcripts (CUTs), stable unannotated transcripts (SUTs), promoter upstream transcripts PROMPTs, and promoter-associated small RNAs (PASRs) or transcription initiation RNAs (tiRNAs). Although the genomic origins of these species may be similar, they appear to greatly differ in their putative functions.

In yeast, a class of ncRNAs has been identified, which is virtually invisible in wild-type cells but becomes abundant if the exosome or the associated polyadenylation complex, the TRAMP complex, are impaired [53]. Accordingly, these transcripts have been named cryptic unstable transcripts or CUTs. They are about 200–600 nt long, are capped, and seem to be promoter-associated [123]. In most cases they are transcribed in a direction opposite to the annotated mRNA [124, 125], although sense CUTs are also observed. Most originate about 200–300 nt upstream of the transcription start site (tss) but their expression level does not seem to correlate with the promoter strength of the nearby gene. A somewhat longer class of promoter-derived RNAs, with a median length of 760 nt, has also been observed, but these seem to be more stable and thus were termed stable unannotated transcripts or SUTs [124]. Both CUTs and SUTs have a well defined tss and seem to be transcribed from nucleosome-free regions enriched around promoters. It is believed that they are generated due to the promiscuous transcriptional activity of RNA polymerase II but are quickly recognized as non-coding and degraded. There is little evidence for any function of antisense CUTs, while sense CUTs have been shown to interfere with transcription of the downstream gene [125]. It was proposed that the sense CUTs might be generated from alternative transcription start sites and either a regulatory mechanism determines which promoter is used or they themselves are part of a regulatory pathway which controls the coding gene [126–129].

It is worth mentioning that, upon inhibiting TRAMP components by RNAi in human cells, a CUT-like population of small RNAs appeared, mapping hundreds of base pairs upstream of transcription start sites. These RNAs were termed PROMPTs, and regions where they map show enrichment of RNAP II indicative of transcription initiation but lacked histone marks associated with elongation. Contrary to CUTs, PROMPT expression levels correlated with the promoter strength of the downstream gene, and they were especially prominent at CpG-rich promoters.

A seemingly very different class of genic small RNAs was observed in multiple animals including humans, mice, chicken embryos, and *Drosophila* by microarray and deep-sequencing studies [20, 23, 130–132]. These have been termed promoter associated small RNAs (PASRs) or transcription initiation RNAs (tiRNAs). They are roughly 20–200 nt and 18 nt long, respectively, as determined by Northern blot and sequence analysis, and have a 5' cap [130]. Like CUTs they are probably short-lived but originate at or just downstream of the transcription initiation site mostly in a sense orientation. This suggests that some might be products of stalled RNAP II. However, depletion of core exosome factors by RNAi does not stabilize PASRs [133]. Thus, they are either degraded by an alternative pathway or they are stabilized in order to fulfill some regulatory function. It is also conceivable that purely the act of transcription of PASRs is relevant to maintain an open chromatin state or that a pool of RNAP II is concentrated at the site of PASR production and available for rapid local transcriptional activation. Contrary to CUTs, which show no correlation with promoter strength, PASRs are more abundant at the promoters of highly active genes with broad TSS regions and CpG islands than at promoters with single dominant TSS associated with a TATA box. A similar class of small RNAs was also found to be present at the 3'UTR of genes, which were accordingly termed terminus-associated small RNAs (TASRs). Currently, no function or correlation with gene expression has been identified for TASRs.

A last class of genic small RNAs was found to be mapping mostly in sense orientation to internal exons of genes. These small RNAs also bear 5' caps but seem to be generated from long mature mRNA transcripts by RNA cleavage and subsequent 5' capping by a recently described cytoplasmic capping complex [20, 134].

13.4
ncRNAs in Dosage Compensation

The discovery of the 17-kb Xist ncRNA in 1991 [135–137] resulted in the first example of a ncRNA with regulatory function: dosage compensation or the assurance of equal gene expression in both genders despite different number of sex chromosomes. Subsequently, the identification of the 40-kb Tsix antisense transcript of Xist [138, 139] lead to the paradigm for sense–antisense RNA relationships and long-range control of chromatin function by ncRNAs. The best studied dosage compensation mechanism is X chromosome inactivation (XCI) in mammals. Females inactivate one X chromosome resulting in the heterochromatic and largely genetically inactive Barr body (Figure 13.5).

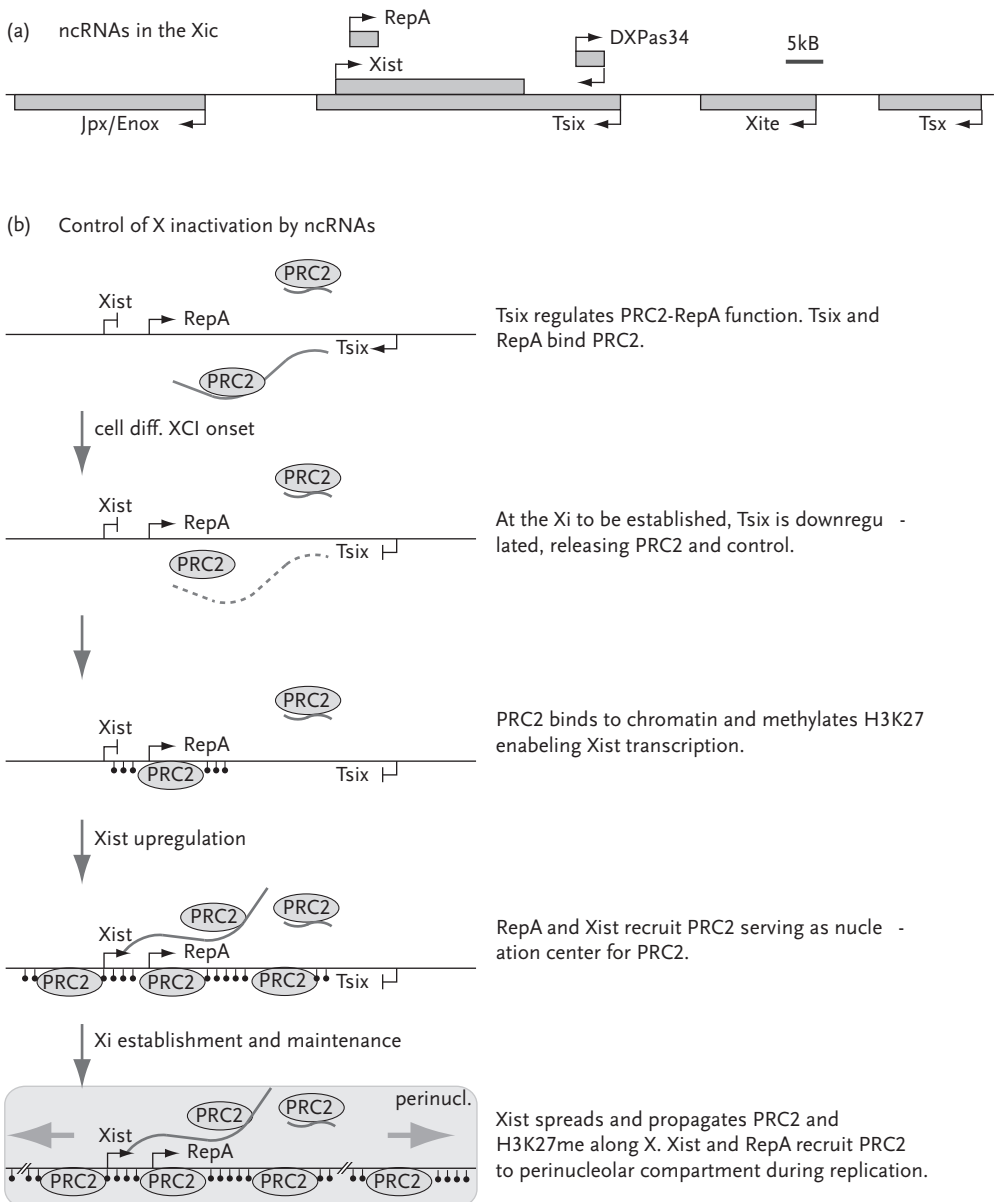


Figure 13.5 NcRNAs in mammalian dosage compensation. (a) Non-coding genes of the Xic. The large non-coding elements Xist, Tsix, and Xite are well established as regulators of XCI. More recently, shorter internal transcription units have been identified from the Xic locus. These include the "sense" locus RepA and the bidirectionally transcribed locus DXPas34. (b) The initiation of XCI controlled by interaction of Tsix, RepA, and Xist RNAs with PRC2. Adapted from [140, 141].

In the embryo proper X inactivation occurs in a random manner upon a cell differentiation trigger if more than one X inactivation center (Xic) is present. The Xic is a multi-megabase region of the X chromosome that contains several non-coding RNA genes. Currently seven have been identified, and although the function of some remains elusive, these RNAs all control different steps of X inactivation *in cis*.

The two Xics pair just prior to the onset of XCI, allowing for sensing and counting of X-chromosome copies and subsequent choice determination. This is proposed to lead to asymmetric localization of factors on the previously identical loci, which results in only one chromosome remaining fully active per diploid autosome set [142, 143]. This last step involves the initiation and spreading of silencing marks along the whole inactive X chromosome. All three steps – sensing, counting, and choice – are controlled by the Xic.

Xist is the best characterized of the ncRNAs originating from Xic. It is expressed from the two-cell stage of embryogenesis, being first repressed on the maternally derived active X chromosome and expressed from the paternal X (Xp) [144]. After reactivation it is expressed at low levels of about three copies per cell. Upon XCI Xist becomes silenced on the active X (Xa) and becomes upregulated about 100-fold on the inactive X in a process that requires the repressive H3K27-me3 mark and the down-regulation of pluripotency factors, such as Oct4, Nanog, and Sox2 [140, 145]. Xist coats the inactivated X chromosome, which is thought to result in exclusion of RNA polymerase II [146]. Gene silencing begins within one or two cell cycles of Xist up-regulation [147–150] through recruitment of PCR1 and PCR2 polycomb (Pc) complexes leading to histone H2A-K119 ubiquitinilation and H3K27methylation [151, 152]. It also results in H4K20 monomethylation, via Prset7 [153] and incorporation of the histone variant macroH2A [154]. Stable maintenance of gene silencing is established by methylation of promoters on the inactive X (Xi) [155]. Xi is transiently present at a perinucleolar compartment during mid-to-late S phase and this localization depends on Xist. This compartment is enriched for the Snf2h component of the ISWI remodeling complex and localization to this region during replication probably allows for maintaining the epigenetic factors associated with the inactive X [156]. While establishment and maintenance of all these epigenetic marks depend on Xist, currently it is not clear how Xist directs the necessary factors to target sites.

Xist is negatively regulated by its antisense gene partner Tsix that originates 12 kb downstream of Xist and in mice extends through the whole Xist locus. Tsix is detected beginning at the eight-cell stage and is oppositely imprinted to Xist, with its expression coming from the active X chromosome. Tsix expression is regulated by Xite, another locus in the Xic 10 kb upstream of Tsix, which bears a Tsix-specific enhancer and also expresses a non-coding RNA [157]. Tsix and Xite seem to be mediators of X chromosom pairing and loss of these RNAs results in either loss or severe delay of pairing with resulting consequences for counting and altered patterns of XCI in female cells [157–159].

In mice, transcription across the Xist promoter seems necessary for silencing, and Xist expression is regulated by Tsix via a change in chromatin structure,

although the mechanisms proposed are somewhat controversial [160–162]. Interestingly, in humans, antisense transcription does not cover the whole Xist region and clearly does not extend to its promoter region, indicating that in humans Xist transcription is regulated in a Tsix-independent manner [163]. Recently Xist and Tsix have been reported to form dsRNA, which during X-inactivation is processed into small RNAs (21–24 nt) via a Dicer-dependent mechanism [24]. Small RNAs map to complementary regions of the two transcripts as well as to the promoter regions and the 5'-end of Xist. Reduction of dicer activity in ES cells compromises small RNA production and leads to reduced DNA methylation at the Xist promoter and to its derepression, resulting in blockage of H3K27 methylation and inactivation. This suggests that Xist and Tsix can act both as long ncRNAs and serve as templates for small RNA production, although Xi defects could not be recapitulated in dicer-null mouse embryos (unpublished data).

Polycomb-mediated silencing of the Xi is induced by a highly conserved region of Xist termed repeat A (RepA) [164]. RepA is an independent transcription unit composed of 7.5 tandem repeats of two stem-loop structures. RepA RNA is produced only from Xi and recruits PRC2 to the Xic. A 28-nt stem-loop structure of RepA interacts directly with the catalytic subunit of Ezh2 histone methyltransferase in the PRC2 complex [140], and its ectopic expression recruits Ezh2 *in vivo*, indicating a role in recruiting Pc-complexes to Xi. However, Pc recruitment to Xi occurs even in mutants lacking the repA regions [164, 165], arguing that it is dispensable for Pc binding. It is likely that two independent pathways have evolved to recruit Ezh2 to the inactive X, and RepA is only involved in one of these. PRC2 directly binds to all three ncRNAs, Xist, Tsix, and RepA, and these binding events seem to interfere with each other by competing away PRC2. Thus it seems that ncRNAs control Pc proteins in several ways: RepA directs PRC2 to the Xic, Xist then spreads PRC2 along the future Xi, and Tsix blocks these activities on the future active X (Xa) by interfering with RepA-PRC2 function. After recruitment of PRC2 by RepA RNA to the Xic of the future Xi upon differentiation, PRC2 catalyzes H3K27 trimethylation. This is proposed to generate a permissive chromatin state for Xist expression leading to Xist upregulation, spreading of Xist, and consequent spreading of PRC2 [140].

Dosage compensation in *Drosophila* does the same job of equalizing gene output from an uneven number of sex chromosomes, but flies do this via a very different mechanism: a two-fold increase in the transcription of the single X chromosome in males. This is directed by a ribonucleoprotein complex, the dosage compensation complex (DCC) also called the male specific lethal (MSL) complex. This complex is responsible for acetylating H4K14 at target sites. It harbors two redundant large ncRNAs, roX1 (RNA on X) and roX2, which likely are responsible for directing the complex to its taget sites. roX1 is a 3.7-kb RNA, while roX2 is 0.6 kb long. Both are stably expressed from the X chromosome in males, but their stability and localization depend on the presence of the MSL complex [166–168]. Although functionally redundant, there is only a 30-nt region of strong similarity between the two RNAs, with an additional 110-nt stretch showing reduced but recognizable similarity. Of interest, deletion of these regions does not impact roX

function [168]. This indicates that roX RNAs function either through a degenerate primary, secondary, or even a tertiary structure or that they harbor further, as yet unidentified, redundant elements. In contrast to Xist, roX1 can function *in trans*. If the gene is translocated to an autosome, the roX RNA still localizes to the X chromosome [168]. Not only is roX required for proper binding and spreading of MSL to the X chromosome but it also enhances the acetyltransferase activity of MSL which seems to depend on a discrete 59-base stretch near the 3'-end that folds into a stem loop [169–171].

13.5
Developmental Regulation of Hox Clusters by Cis- and Trans-Acting ncRNAs

Hox genes are crucial regulators that specify the anterior–posterior axis and segment identity of metazoans during early embryonic development. Genes are arranged in clusters (in humans HOXA-D) in the same physical order along the chromosome as is their expression along the antero-posterior axis and their temporal expression during development [172, 173]. Regions within the cluster show differences in their epigenetic profile, which are maintained with sharp boundaries by Polycomb and trithorax complexes [174]. The different genes in the cluster are controlled by several enhancer elements present in the intergenic regions that are thought to act *in cis* as targets of regulatory proteins. In addition, hundreds of ncRNAs are transcribed from intergenic regions of Hox clusters mainly in the antisense orientation [175]. Out of the five classes of mutations that affect expression of the homeotic protein Ubx in *Drosophila* only one affects the protein-coding sequence. The others are located in introns or the upstream bxd region. The latter produces a 27-kb ncRNA transcript whose expression is highly regulated during embryogenesis in a pattern that is partially reflective of the Ubx transcript. There are currently different models for ncRNA function in hox clusters, and these are discussed below. These models are not mutually exclusive, and it is easily possible that ncRNA regulation of HOX genes operates by a combination of different mechanisms.

13.5.1
The Act of Transcription Leads to Altered DNA Accessibility

According to Sessa and colleagues, intergenic transcripts themselves have no effect on Hox gene expression and only the act of ncRNA transcription is required for regulation [176]. Genetic studies suggested that transcription of ncRNAs alters the accessibility of DNA sequences important for TrxG and PcG binding. Intergenic transcription is thought to enable trithorax-mediated activation of downstream hox genes and prevent Polycomb-mediated silencing [177, 178]. In human teratocarcinoma cell lines, induction of differentiation by retinoic acid leads to the timely regulated activation of HOX genes and also triggers the transcription of

non-coding intergenic RNAs. This is accompanied by histone modifications and loss of interaction with the PcG repressive complex PRC2 at the non-coding transcription units. Since ncRNA transcripts respond much faster to RA treatment than the sense transcript it is thought that the activation of the sense transcript is the result of RA mediated transcriptional activation of the corresponding ncRNAs [176]. This hypothesis is further strengthened by the observation that the expression patterns of ncRNAs in human tissues coincide with those of the adjacent HOXA mRNAs that are collinearly expressed along the antero-posterior axis [179]. The antisilencing by transcription described above is likely a widespread and fundamental process used by the cell as a memory system to initiate or maintain the active state of developmentally regulated genes and is probably not limited to hox gene regulation.

13.5.2
Transcription Through Hox Gene Promoters Interferes with Their Expression

Transcriptional interference occurs when the transcription of one RNA overlaps with the promoter or the coding sequence of another gene and inhibits its transcriptional initiation or elongation. Recent publications indicate that the *Drosophila* Ubx Hox gene and the upstream bxd non-coding RNAs are expressed in a non-overlapping pattern in embryos and the imaginal discs suggesting that transcription of these ncRNAs is associated with repression of the coding gene via transcriptional interference [180]. According to this model, transcription of an upstream ncRNAs into the promoter of downstream Hox genes prevents Hox gene expression leading to silencing *in cis*. Mutations of the bxd ncRNA lead to derepression of Ubx, while elimination of the ncRNA by RNAi had no effect on Ubx expression confirming that the act of transcription is important for repression [180].

13.5.3
Non-Coding RNA Transcripts Bind Regulatory Factors

Several ncRNAs transcribed upstream of the *Drosophila* Hox gene Ubx interact with the DNA at the locus and recruit the TrxG protein Ash1 to the Ubx promoter inducing an active chromatin state and Ubx transcription [181]. This observation implies that, at least in some instances, the RNA itself and not the act of transcription is relevant for regulation. The aforementioned study reported that ectopic expression of hox ncRNAs activates hox gene expression *in trans*, however, this observation was not confirmed by other studies [180].

13.5.4
Non-Coding RNA Regulation in trans

The hox cluster harbors yet another unexpected mode of ncRNA-mediated transcriptional regulation. The recently described ncRNA, HOTAIR (HOX antisense

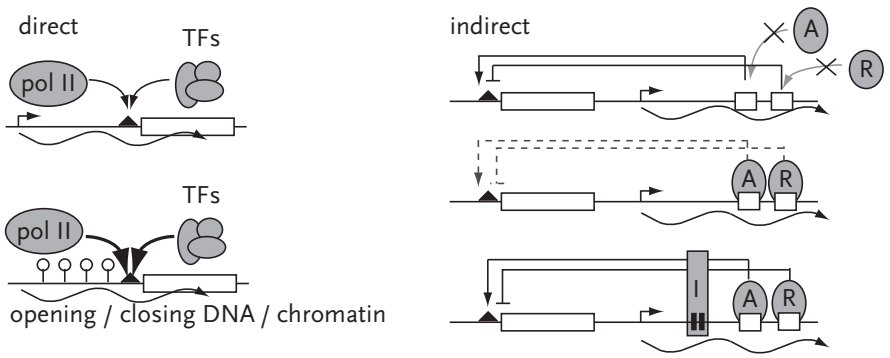
intergenic RNA), is a highly conserved 2158-nt spliced and polyadenylated ncRNA derived from a regulatory boundary in the HOXC locus. It functions to regulate the HOXD locus *in trans*. Knockdown of HOTAIR by RNAi did not effect the expression of the HOXC locus but resulted in loss of repression of a 40-kb region in the HOXD cluster, with loss of H3K27 methylation and the PcG complex PRC2 from this locus [175]. Pull-down of the Pcr2 showed specific interaction of the complex with HOTAIR. Although the mechanism by which HOTAIR targets silencing complexes to the HOXD locus is still poorly understood, it is the first long ncRNA shown to act *in trans* and thus represents a landmark in our understanding of long ncRNA biology.

13.6
Mechanisms of Transcriptional Regulation by Long ncRNAs

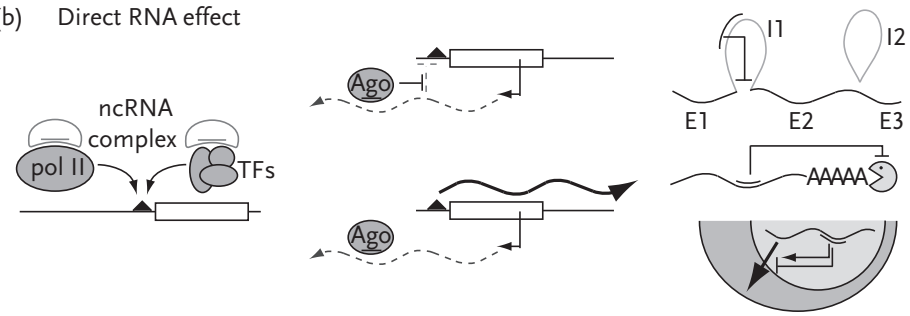
Based on the discussion of dosage compensation (Section 13.4) and HOX genes (Section 13.5), we can identify a number of different regulatory mechanisms by which these and other long ncRNA (see also review by [182]) are thought to operate (Figure 13.6, Table 13.1):

Figure 13.6 A selection of current models for transcriptional regulation by ncRNAs. (a) Transcription-based gene regulation models that do not require the ncRNA product itself, but only its transcription can either be direct or indirect. Direct effects are seen when transcription through a promoter region or gene influences transcription or initiation rate either by interfering with the transcriptional machinery or by changing the local chromatin state. Indirect effects are observed when the ncRNA is not transcribed through the gene but rather impacts a *cis*-regulatory element. ncRNA transcription can displace binding proteins from activators or repressors that lie in the ncRNA gene body. The ncRNA transcription can also inhibit the long-range activity of the bound activator or repressor. Finally, transcription can activate insulator elements. ncRNA transcription is proposed to activate an insulator or boundary element (located in the ncRNA gene body), which then enables binding of an insulator protein. This would prevent the interaction between the domain activator/repressor and the coding genes. (b) ncRNAs can act directly alter transcriptional activity by interfering with RNAP II activity or interacting with specific transcription factors leading to either activation or repression. For siRNAs mapping to promoters two models have been proposed, both of which depend on long ncRNAs transcribed in antisense direction. It is believed that either the antisense transcript regulates sense transcription or the equilibrium between sense and antisense transcription is important for proper expression. Small RNAs regulate the sense gene indirectly by "titrating out" the antisense transcript. NcRNAs can also influence gene expression by regulating splicing, processing and nuclear export of coding sequences. (c) Model of ncRNA regulation of chromatin domains via histone modification enzymes. Transcription of ncRNAs *in cis* may change the accessibility of activating or repressing proteins or directly recruit them, leading to changes in histone marks or DNA methylation and altered transcriptional activity of the downstream genes. In some instances recruitment of chromatin modifying factors is programmed by ncRNAs produced *in trans*. In most instances these are small ncRNAs although HOTAIR and roX are examples of long ncRNAs acting *in trans*. (d) Some ncRNAs have been implicated in influencing global nuclear architecture or higher-order chromatin structure leading to changes in transcription of larger genomic loci or whole chromosomes.

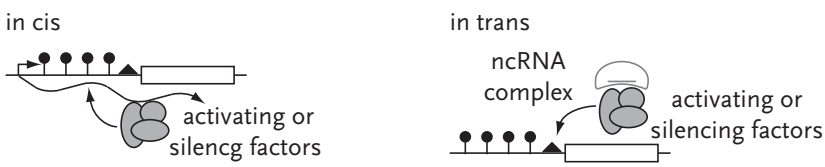
(a) Act of transcription



(b) Direct RNA effect



(c) Altering chromatin structure



(d) Higher order chromatin structure and nuclear organization

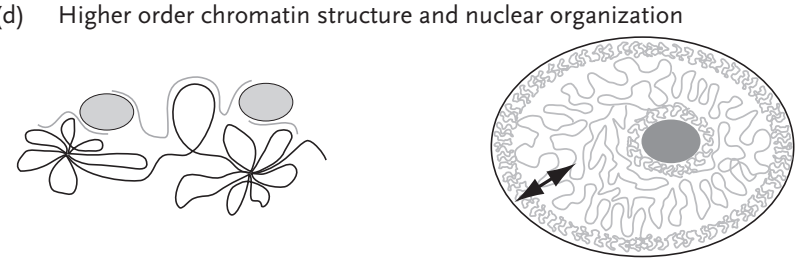


Table 13.1 Mechanisms of gene regulation by small and long non-coding RNAs.

ncRNA	Description of proposed mode of action	Reference
Transcription itself has regulatory function		
Yeast PHO5 and SER3 upstream transcripts	Transcription of nc RNA changes chromatin plasticitiy	[183, 184]
β-Globin region, Ig-heavy chain V region	DNA becomes accessible to binding proteins	[185, 186]
Tsix-mediated Xist regulation	Readthrough Xist gene/promoter leads to altered chromatin and Xist repression	[160–162]
Hox genes	Continuous transcription prevents heterochromatin formation	[177, 178, 180]
Imprinted genes	Transcription through regulatory region results in altered binding/activity of activating/repressing factors	[187]
Yeast fbp1+ locus upsteam transcript	Opening chromatin in fbp1+ locus allowing accessibility	[188]
Interaction with the transcription machinery or recruitment of regulatory factors		
Exogenous siRNAs to initiation sites	Inhibition of RNAP II binding to TATA box	[110–112]
siRNAs	Inhibition of RNAP II elongation in *C. elegans*	[113]
B2 and Alu repeat sequences	Inhibit closed complex formation, regulate RNAP II activity	[189–191]
7SK RNA	Inhibits kinase activity P-TEFb transcription factor and represses RNAP II transcription	[156, 192, 193]
DHFR	TFIIB inhibition, DNA:RNA triplex formation	[194]
Exo- and endogenous siRNAs to promoters	Titrating out endogenous regulatory antisense transcript	[121, 122]
RNA processing		
N-myc antisense	Stabilizes transcripts that retain intron 1	[195]
SAF	1500-nt partially intronic transcript that is important alteration in FAS alternative splicing	[106]
Targeting DNA methylation and histone modifications		
Endogenous siRNAs	Heterochromatin formation in yeast and plants	[41–44]
Plant endogenous siRNAs to promoters	DNA methylation and repression	[66, 67]
Exogenous siRNAs	Accumulation of Argonaute 1 at target promoters increases H3K9me2 and H3K27me3	[106]
Air	Recruits histone modifiers during Igf2r imprinting	[196]
RepA	H3K27-methyl recruitment during XCI	[197]
Hox genes	Recruitment of Ash1	[181]

ncRNA	Description of proposed mode of action	Reference
Xist	Recruitment of PCR1, PCR2, Pset resulting in repressive chromatin marks, incorporation of macroH2A, recruitment of HP1	[151–154]
Endogenous siRNA to MYT promoter	Ago1-mediated EzH2 recruitment and H3K27 trimethylation	[109]
RoX	H4K14 acetylation, activation	[167]
HOTAIR	Recruitment of Pc proteins *in trans*	[175]
Hemoglobin A2 locus	Promoter methylation	[198]
Changing higher order chromatin organization		
Xist	Perinuclear localization, coating and exclusion of RDRP II, formation of Barr body	[146, 156]
eRNAs	Enhancer–promoter looping	[199]
Chicken giant α-globin RNA	Part of nuclear matrix	[200]
SINE B2	Defines hetero/euchromatin boundary	[201]

1. The process of ncRNA transcription itself leads to interference with transcription of other genes (Figure 13.6a).
2. The ncRNA targets the transcriptional machinery or recruits a regulatory factor (Figure 13.6b).
3. They may also affect downstream processes such as splicing, nuclear export and mRNA stability.
4. The ncRNA regulates nucleosome structure (Figure 13.6c).
5. The higher order chromatin organization and nuclear localization (Figure 13.6d) of its target and thereby influences gene expression.

13.6.1 Transcriptional Read through Leading to Activation or Silencing

Transcriptional interference (TI) and activation (TA) are phenomena, wherein the polymerase transcribing one gene directly impacts the transcription initiation or elongation of an overlapping gene (Figure 13.6a). Direct TI requires read-through of the regulated promoter (or entire gene) and mainly leads to inhibition. Transcription of a ncRNA through a cis-regulatory element such as an activator, repressor, or insulator element may result in indirect changes of mRNA transcription. A well characterized insulator affected by transcriptional read-through lies in the Igf2 cluster. Activation was also proposed to act by making the DNA accessible to binding proteins, as in the case of intergenic transcription in multigene clusters of the beta-globin and immunoglobulin heavy chain V region [185, 186]. Transcriptional interference seems, however, to be the most common effect of overlapping transcriptional units. For example, the yeast PHO5 locus seems to be regulated by

transcription of a ncRNA via changes of local chromatin plasticity [183]. The ncRNA is rapidly degraded by the exosome and its expression *in trans* does not affect Pho transcription indicating the need of local transcription. Similarly, the yeast SER3 gene is regulated by an upstream transcription [184]. Finally, transcriptional elongation of ncRNAs across regulatory regions or genes can cause changes in histone modifications. For example, continuous ncRNA transcription has been suggested to prevent the silencing of certain Hox genes by PcG proteins [178].

13.6.2
Non-Coding RNAs Directly Regulating Transcription and RNA Processing

In many cases transcription is not sufficient and the non-coding transcript regulates transcription by for example modulating RNAP II activity, recruiting transcription factors, inducing changes in chromatin state or directly interacting with the DNA (Figure 13.6b). In addition, ncRNAs have also been shown to regulate alternative splicing and stability of coding mRNAs (Chapter 11). Contrary to transcriptional interference, ncRNA-mediated regulation can also act *in trans*.

A direct interaction of long ncRNAs with RNAP II has been described for repeat-derived sequences. Binding of the murine B2 ncRNA to RNAP II upon heat shock leads to transcriptional inhibition of other mRNAs *in trans* [189]. B2 is a mouse tRNA-derived small RNA (178 nt) transcribed by RNAP III upon heat shock from the SINE repeat element. It contains a 51-nt core sequence, which binds with high affinity to an RNA docking site in the core of RNAP II. Antisense sequences to B2 attenuate its inhibitory effect indicating that the RNA itself is responsible for the function. Similarly, in humans Alu repeat derived transcripts are also expressed upon stress and bind directly and tightly to RNAP II at promoters of repressed genes. Both B2 and Alu prevent RNAP II from establishing contacts with the promoter upstream and downstream of the TATA box during closed complex formation [190, 191].

NcRNAs often interact with transcription factors. For example, the 7SK RNA, a negative regulator of RNAP II transcription in higher mammals, binds to the phosphorylated P-TEFb transcription factor and inhibits its kinase activity, thereby repressing transcription [156, 192, 193]. Another example is the human dihydrofolate reductase (DHFR) gene, which is repressed by a ncRNA from an upstream alternative promoter. The ncRNA forms a RNA–DNA triplex structure with the DHFR promoter and directly interacts with TFIIB, resulting in disruption of the preinitiation complex both *in cis* and *in trans* [194].

Intronic ncRNAs can coordinate waves of gene expression important for particular cellular processes functionally related to the protein of the same locus. For example a study showed that expression of three different intronic sequences of the CFTR gene that did not contain any known miRNAs or predicted stem-loop structures caused extensive and specific transcriptional changes in genes mainly linked to CFTR function [202]. These affected genes were distributed at spatially diverse sites within the genome and each of the three intronic sequences induced unique, highly reproducible changes in HeLa, where CFTR itself is normally not expressed. How

these ncRNAs regulate distant genes is currently not known. However, since their expression is normally linked to that of the CFTR gene and the affected proteins have been linked to CFTR, it is likely that many regulatory functions attributed to CFTR are in fact carried out by the non-coding transcripts. This phenomena is possibly more widespread than anticipated and not unique to the CFTR gene.

13.6.3
RNA Processing

The role of ncRNAs in gene regulation reaches beyond that of transcription and influences downstream processes such as splicing, mRNA stability, and nuclear export. Non-coding RNAs can guide splice site selection leading to alternative splicing (Chapter 11). For example, an antisense RNA transcribed from the first intron of N-myc modulates RNA splicing by forming an RNA–RNA duplex and preserving a population of N-myc mRNA with retained intron1 [195]. Overexpression of SAF, a 1500-nt partially intronic ncRNA from the first intron of the FAS gene, causes a functionally important alteration in FAS alternative splicing [203]. The frequency of exon-skipping for exons overlapped by or located immediately 3' of an intronic ncRNAs is higher than the average, also emphasizing the importance of ncRNAs in identification of correct splice sites [204]. A recent study reported the role of the ncRNA MALAT in alternative splicing: it influences the distribution of splicing factors in nuclear speckle domains and modulates the level of active, phosphorylated, serine/argininge splicing factors [205]. NcRNAs can stabilize protein-coding RNAs. The B-cell lymphoma specific bcl-2/IgH antisense ncRNA contributes to the upregulation of bcl-2 expression probably by masking the AU-rich motifs in BCL2 3' UTR that would otherwise induce degradation [206].

13.6.4
Regulation of Gene Expression via ncRNA-Induced Epigenetic Modifications

Many ncRNAs regulate transcription indirectly by modifying local chromatin accessibility. Some examples, such as the process of dosage compensation and the regulation of HOX genes, are described above (Figure 13.6c). Both activation and repression have been reported, and these can involve DNA methylation as well as histone modifications, which can lead to long-lasting heritable effects. Several examples indicate a role for ncRNAs in regulating DNA methylation. For example, a deletion in the globin locus relocates the constitutively expressed LUC7L gene 300 bp downstream of the HBA2 gene, giving rise to an antisense RNA that overlaps the promoter of the HBA2 gene. This results in methylation and silencing of the HBA2 gene *in cis* [198]. Another example is the KHPS1 ncRNA that originates from the CpG island of sphingosine kinase 1 (SPHK1). This ncRNA overlaps a sense regulatory element, the tissue-dependent differentially methylated region (T-DMR). Overexpression of the KHPS1 fragments caused DNA methylation of T-DMR thereby regulating SPHK1 expression. Finally, as described previously in this chapter, synthetic small RNAs, such as siRNAs, have been described to induce

DNA methylation and regulate gene expression in mammalian cells when targeted to promoter regions [207–209]. According to one model, these are proposed to interfere with the function endogenous low-copy promoter-associated RNAs.

Long ncRNAs have been implicated in gene silencing through altering local histone methylation at the p15 locus with important implications for cell differentiation and tumorigenesis [210]. The P15AS long (35 kb) antisense RNA encompasses the entire p15 locus, however a small fragment overlapping the 1st exon of p15 is sufficient to control its chromatin structure. Nonetheless, the mechanism is Dicer-independent indicating that P15AS does not serve as a trigger for an siRNA-mediated pathway. In some cases it has been shown that the RNA alters the activity of chromatin modifying enzymes rather than recruiting them to target sites. For example, ncRNAs from the cyclin D1 promoter region are allosteric effectors of an RNA-binding protein, TLS. These ncRNAs are variable in length, are induced upon DNA damage, and remain bound to the chromatin at the cyclin D1 promoter region. Association of TLS with the ncRNA changes it from an inactive to an active conformation, resulting in inhibition of histone acetylases CBP and p300 and consequent silencing of the cyclin D1 locus [211].

A special case of ncRNA epigenetic regulation that warrants additional consideration is that of imprinted genes. Genomic imprinting is a phenomenon by which certain genes are expressed depending on their parent of origin. About 80 mouse genes, most of which are grouped into clusters of 3–15 genes, are subject to imprinting. Most imprinted genes encoding proteins; however, each cluster also contains at least one ncRNA gene. Imprinted expression of the cluster is controlled by a *cis*-acting imprint control element (ICE), which carries parental information in the form of DNA methylation acquired during gametogenesis and maintained only on one parental allele after fertilization [212, 213]. Six clusters, named after their founding imprinted gene, have been particularly well characterized. Two (Igf2, Dlk) are paternally imprinted and four (Igf2r, Kcnq1, Gnas, PWS-AS) are maternally imprinted. One parental chromosome carries the unmethylated ICE and expresses the ncRNA, indicating that, with one exception, the Rasgfr1 ncRNA, the ICE is a positive regulator of ncRNA expression. Expression of the ncRNA correlates with the repression *in cis* of some or all of the imprinted protein-coding genes. Almost all known antisense RNAs located in imprinted domains are expressed on the paternal chromosome, while all methylation-based imprints occur on the maternal allele, leading to the speculation that ncRNAs might have evolved as a paternal substitute for the methylation imprint [214].

Within the Igf2r cluster, the 108-nt Air transcript is the best characterized ncRNA involved in imprinting. Air is a nuclear retained, unspliced but capped, and polyadenylated unstable transcript from an antisense promoter located in the second intron of Igf2r. Air expression is imprinted by DNA methylation on the maternal chromosome. On the paternal chromosome from which Air is expressed, the promoter carries activating histone H3K4-methyl and H3K9 acetyl marks, while the downstream silent Igf2r promoter lacks these modifications [215]. In the embryo proper, Air expression only results in silencing of the paternal Igf2r gene *in cis*. In the placenta, two other distant genes are also silenced *in cis* by Air,

although how this difference is established is not understood [196]. The ncRNA does not overlap with all silenced genes. Thus, transcription interference is unlikely to be the origin of silencing, although it is possible that transcription through a regulatory domain indirectly acts to silence genes downstream. Truncating Air leads to desilencing of target genes indicating that the full-length transcript is critical. In addition the expression level of Air is relevant for silencing as exogenous promoters yielding high Air expression can silence Igf2r *in cis*, while low expression does not [216]. Air probably silences target loci by recruiting histone modifiers. Interestingly, there is indication that Air expression induces imprinted Igf2r expression, not simply by silencing the paternal Igf2r but by generating an expression bias between the two parental alleles resulting in increased maternal and decreased paternal expression [217].

13.6.5
ncRNA Regulation through Changes of Large-Scale Chromatin Structure

The aforementioned examples list instances of gene-specific regulation of local chromatin marks that lead to altered gene expression. However in many cases ncRNAs can change the global chromatin structure (Figure 13.6d). This is thought to alter the accessibility of surrounding genes and thereby impact transcription levels. Proposed mechanisms involve recruiting of factors that influence higher order structure or ncRNAs might act more directly to serve as structural components of chromatin. The best characterized example of ncRNA-induced higher order chromatin structure is the formation of a RNAP II-deficient compartment by Xist as the first step of X inactivation (see Section 13.4). In the case of the chicken giant alpha-globin RNAs appear to form part of the nuclear matrix and are essential for nuclear genome organization [200].

Another mechanism for ncRNA mediated chromatin looping between promoters and regulatory elements comes from recent high-throughput sequencing, which revealed substantial ncRNA transcription from enhancer regions. Enhancers are DNA regions that regulate expression of genes from a distance. They are maintained in an open chromatin conformation characterized by the absence of H3K27-me3 marks but are lacking H3K4 methylation typical for promoter regions. Many enhancers have been shown to bind RNAP II, and this association increases upon activation. The identified transcripts generated from enhancers, so called eRNAs, are smaller than 2 kb and do not have a polyA tail. They are often transcribed in both directions from the enhancer center and show a dynamic increase in level upon activating signals, which strongly correlate with changes in mRNA level at nearby genes. Interestingly, recruitment of RNAP II to the enhancer does not seem to be sufficient for eRNA synthesis and it is likely that an interaction between enhancer and promoter of the nearby target gene is required for both mRNA and eRNA synthesis [199]. It is tempting to speculate that the simultaneous transcription of both promoters and enhancers facilitates their interaction and that this proximity might further enhance transcription thereby forming a positive feedback loop in response to an activating signal.

Functional organization of chromatin can also be regulated by ncRNAs derived from repetitive elements. In mice, bidirectional transcription of the retrotransposon, SINE B2, regulates expression of the associated growth hormone by relocating the locus into specific nuclear compartments and locally defining the heterochromatin/euchromatin boundary during organogenesis [201]. Our genomes are comprised to a large extent of repetitive sequences much of which is transcribed. It is plausible to assume that this transcription is at least in part responsible for nuclear organization. In fact, small RNAs derived from centromeric regions have been shown to be responsible for centromeric heterochromatin organization and it is very likely that long transcripts fulfill similar functions.

Finally, beyond chromatin structure, certain nuclear structures and processes are sensitive to RNase treatment. For example, RNase treatment disrupts paraspeckles [218, 219]. Similarly, rRNAs have been shown to associate with the mitotic spindle and RNaseA treatment disrupts spindle assembly, while treatment with translational inhibitors does not, indicating that the rRNA (possibly along with other transcripts) contributes to spindle integrity [220].

13.7 Conclusions

From the accumulating data, it seems evident that non-coding RNAs govern the expression of our genomes on multiple levels from direct transcriptional interference to regulating higher order chromatin structure and nuclear organization. Though we are only starting to dissect the actual mechanisms of their action, we have already found evidence for a wide range of mechanisms utilized by ncRNAs, involving interactions with DNA, RNA, and proteins alike. It is apparent that the vast majority of protein coding genes is affected in one way or another by ncRNAs and understanding their mechanism of action will – and already does – enable us to harvest them as tools both in biology as in targeted therapy.

References

1 Ohno, S. (1972) So much "junk" DNA in our genome. *Brookhaven Symp Biol*, **23**, 366–370.

2 Orgel, L.E., and Crick, F.H. (1980) Selfish DNA: the ultimate parasite. *Nature*, **284**, 604–607.

3 Wheeler, D.L., Barrett, T., Benson, D. A., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., Dicuccio, M., Edgar, R., Federhen, S., Feolo, M., *et al.* (2008) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*, **36**, D13–D21.

4 Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., Kodzius, R., *et al.* (2005) The transcriptional landscape of the mammalian genome. *Science*, **309**, 1559–1563.

5 Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.

W., Richards, S., Weinstock, G.M., *et al.* (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res*, **15**, 1034–1050.

6 Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., Kamal, M., Clamp, M., Chang, J. L., Kulbokas, E.J. 3rd, Zody, M.C., Mauceli, E., *et al.* (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, **438**, 803–819.

7 Smith, N.G., Brandstrom, M., and Ellegren, H. (2004) Evidence for turnover of functional noncoding DNA in mammalian genome evolution. *Genomics*, **84**, 806–813.

8 Kawaji, H., Nakamura, M., Takahashi, Y., Sandelin, A., Katayama, S., Fukuda, S., Daub, C.O., Kai, C., Kawai, J., Yasuda, J., Carninci, P., *et al.* (2008) Hidden layers of human small RNAs. *BMC Genomics*, **9**, 157.

9 Lin, R., Maeda, S., Liu, C., Karin, M., and Edgington, T.S. (2007) A large noncoding RNA is a marker for murine hepatocellular carcinomas and a spectrum of human carcinomas. *Oncogene*, **26**, 851–858.

10 Torarinsson, E., Yao, Z., Wiklund, E. D., Bramsen, J.B., Hansen, C., Kjems, J., Tommerup, N., Ruzzo, W.L., and Gorodkin, J. (2008) Comparative genomics beyond sequence-based alignments: RNA structures in the ENCODE regions. *Genome Res*, **18**, 242–251.

11 Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E., Kuehn, M.S., *et al.* (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, **447**, 799–816.

12 Kampa, D., Cheng, J., Kapranov, P., Yamanaka, M., Brubaker, S., Cawley, S., Drenkow, J., Piccolboni, A., Bekiranov, S., Helt, G., Tammana, H., *et al.* (2004) Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res*, **14**, 331–342.

13 Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M. J., Kuramochi-Miyagawa, S., Nakano, T., Chien, M., *et al.* (2006) A novel class of small RNAs bind to MILI protein in mouse testes. *Nature*, **442**, 203–207.

14 Kohtz, J.D. and Fishell, G. (2004) Developmental regulation of EVF-1, a novel non-coding RNA transcribed upstream of the mouse Dlx6 gene. *Gene Expr Patterns*, **4**, 407–412.

15 Louro, R., Nakaya, H.I., Amaral, P.P., Festa, F., Sogayar, M.C., da Silva, A.M., Verjovski-Almeida, S., and Reis, E.M. (2007) Androgen responsive intronic non-coding RNAs. *BMC Biol*, **5**, 4.

16 Cawley, S., Bekiranov, S., Ng, H.H., Kapranov, P., Sekinger, E.A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A.J., Wheeler, R., *et al.* (2004) Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell*, **116**, 499–509.

17 Pheasant, M. and Mattick, J.S. (2007) Raising the estimate of functional human sequences. *Genome Res*, **17**, 1245–1253.

18 Bartel, D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.

19 Filipowicz, W., Jaskiewicz, L., Kolb, F. A., and Pillai, R.S. (2005) Post-transcriptional gene silencing by siRNAs and miRNAs. *Curr Opin Struct Biol*, **15**, 331–341.

20 Fejes-Toth, K., Sotirova, V., Sachidanandam, R., Assaf, G., Hannon, G., Kapranov, P., Foissac, S., Willingham, A., Duttagupta, R., Dumais, E., and Gingeras, T. (2009) Post-transcriptional processing generates a diversity of 5′-modified long and short RNAs. *Nature*, **457**, 1028–1032.

21 Kirino, Y. and Mourelatos, Z. (2007) Mouse Piwi-interacting RNAs are

2′-O-methylated at their 3′ termini. *Nat Struct Mol Biol*, **14**, 347–348.
22 Ohara, T., Sakaguchi, Y., Suzuki, T., Ueda, H., and Miyauchi, K. (2007) The 3′ termini of mouse Piwi-interacting RNAs are 2′-O-methylated. *Nat Struct Mol Biol*, **14**, 349–350.
23 Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermuller, J., Hofacker, I.L., Bell, I., *et al.* (2007) RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science*, **316**, 1484–1488.
24 Ogawa, Y., Sun, B.K., and Lee, J.T. (2008) Intersection of the RNA interference and X-inactivation pathways. *Science*, **320**, 1336–1341.
25 Wilusz, J.E., Freier, S.M., and Spector, D.L. (2008) 3′ end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell*, **135**, 919–932.
26 Ganesan, G. and Rao, S.M. (2008) A novel noncoding RNA processed by Drosha is restricted to nucleus in mouse. *RNA*, **14**, 1399–1410.
27 Franco-Zorrilla, J.M., Valli, A., Todesco, M., Mateos, I., Puga, M.I., Rubio-Somoza, I., Leyva, A., Weigel, D., Garcia, J.A., and Paz-Ares, J. (2007) Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat Genet*, **39**, 1033–1037.
28 Hellwig, S. and Bass, B.L. (2008) A starvation-induced noncoding RNA modulates expression of Dicer-regulated genes. *Proc Natl Acad Sci USA*, **105**, 12897–12902.
29 Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R.M., and Hannon, G.J. (2008) Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature*, **453**, 534–538.
30 Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T., Surani, M.A., *et al.* (2008) Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature*, **453**, 539–543.
31 Czech, B., Malone, C.D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J.A., Sachidanandam, R., Hannon, G.J., *et al.* (2008) An endogenous small interfering RNA pathway in *Drosophila*. *Nature*, **453**, 798–802.
32 Kawamura, Y., Saito, K., Kin, T., Ono, Y., Asai, K., Sunohara, T., Okada, T.N., Siomi, M.C., and Siomi, H. (2008) *Drosophila* endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature*, **453**, 793–797.
33 Ghildiyal, M., Seitz, H., Horwich, M. D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E.L., Zapp, M.L., Weng, Z., and Zamore, P.D. (2008) Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science*, **320**, 1077–1081.
34 Okamura, K., Balla, S., Martin, R., Liu, N., and Lai, E.C. (2008) Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in *Drosophila melanogaster*. *Nat Struct Mol Biol*, **15**, 581–590.
35 Pagano, A., Castelnuovo, M., Tortelli, F., Ferrari, R., Dieci, G., and Cancedda, R. (2007) New small nuclear RNA gene-like transcriptional units as sources of regulatory transcripts. *PLoS Genet*, **3**, e1.
36 Galindo, M.I., Pueyo, J.I., Fouix, S., Bishop, S.A., and Couso, J.P. (2007) Peptides encoded by short ORFs control development and define a new eukaryotic gene family. *PLoS Biol*, **5**, e106.
37 Leygue, E. (2007) Steroid receptor RNA activator (SRA1): unusual bifaceted gene products with suspected relevance to breast cancer. *Nucl Recept Signal*, **5**, e006.
38 Jenny, A., Hachet, O., Zavorszky, P., Cyrklaff, A., Weston, M.D., Johnston, D.S., Erdelyi, M., and Ephrussi, A. (2006) A translation-independent role of oskar RNA in early *Drosophila* oogenesis. *Development*, **133**, 2827–2833.

39 Grewal, S.I. and Jia, S. (2007) Heterochromatin revisited. *Nat Rev Genet*, **8**, 35–46.

40 Maison, C. and Almouzni, G. (2004) HP1 and the dynamics of heterochromatin maintenance. *Nat Rev Mol Cell Biol*, **5**, 296–304.

41 Hall, I.M., Shankaranarayana, G.D., Noma, K., Ayoub, N., Cohen, A., and Grewal, S.I. (2002) Establishment and maintenance of a heterochromatin domain. *Science*, **297**, 2232–2237.

42 Volpe, T., Schramke, V., Hamilton, G. L., White, S.A., Teng, G., Martienssen, R.A., and Allshire, R.C. (2003) RNA interference is required for normal centromere function in fission yeast. *Chromosome Res*, **11**, 137–146.

43 Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I., and Martienssen, R.A. (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science*, **297**, 1833–1837.

44 Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S.I., and Moazed, D. (2004) RNAi-mediated targeting of heterochromatin by the RITS complex. *Science*, **303**, 672–676.

45 Partridge, J.F., Borgstrom, B., and Allshire, R.C. (2000) Distinct protein interaction domains and protein spreading in a complex centromere. *Genes Dev*, **14**, 783–791.

46 Motamedi, M.R., Verdel, A., Colmenares, S.U., Gerber, S.A., Gygi, S.P., and Moazed, D. (2004) Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell*, **119**, 789–802.

47 Schramke, V., Sheedy, D.M., Denli, A. M., Bonila, C., Ekwall, K., Hannon, G. J., and Allshire, R.C. (2005) RNA-interference-directed chromatin modification coupled to RNA polymerase II transcription. *Nature*, **435**, 1275–1279.

48 Colmenares, S.U., Buker, S.M., Buhler, M., Dlakic, M., and Moazed, D. (2007) Coupling of double-stranded RNA synthesis and siRNA generation in fission yeast RNAi. *Mol Cell*, **27**, 449–461.

49 Buhler, M., Verdel, A., and Moazed, D. (2006) Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. *Cell*, **125**, 873–886.

50 Noma, K., Sugiyama, T., Cam, H., Verdel, A., Zofall, M., Jia, S., Moazed, D., and Grewal, S.I. (2004) RITS acts *in cis* to promote RNA interference-mediated transcriptional and post-transcriptional silencing. *Nat Genet*, **36**, 1174–1180.

51 LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A., and Tollervey, D. (2005) RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell*, **121**, 713–724.

52 Vanacova, S., Wolf, J., Martin, G., Blank, D., Dettwiler, S., Friedlein, A., Langen, H., Keith, G., and Keller, W. (2005) A new yeast poly(A) polymerase complex involved in RNA quality control. *PLoS Biol*, **3**, e189.

53 Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.C., Dufour, M.E., Boulay, J., Regnault, B., Devaux, F., Namane, A., Seraphin, B., Libri, D., *et al.* (2005) Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell*, **121**, 725–737.

54 Buhler, M., Haas, W., Gygi, S.P., and Moazed, D. (2007) RNAi-dependent and -independent RNA turnover mechanisms contribute to heterochromatic gene silencing. *Cell*, **129**, 707–721.

55 Chen, E.S., Zhang, K., Nicolas, E., Cam, H.P., Zofall, M., and Grewal, S.I. (2008) Cell cycle control of centromeric repeat transcription and heterochromatin assembly. *Nature*, **451**, 734–737.

56 Kloc, A., Zaratiegui, M., Nora, E., and Martienssen, R. (2008) RNA interference guides histone modification during the S phase of chromosomal replication. *Curr Biol*, **18**, 490–495.

57 Jia, S., Noma, K., and Grewal, S.I. (2004) RNAi-independent heterochromatin nucleation by the

stress-activated ATF/CREB family proteins. *Science*, **304**, 1971–1976.
58 Hong, E.J., Villen, J., Gerace, E.L., Gygi, S.P., and Moazed, D. (2005) A cullin E3 ubiquitin ligase complex associates with Rik1 and the Clr4 histone H3-K9 methyltransferase and is required for RNAi-mediated heterochromatin formation. *RNA Biol*, **2**, 106–111.
59 Horn, P.J., Bastie, J.N., and Peterson, C.L. (2005) A Rik1-associated, cullin-dependent E3 ubiquitin ligase is essential for heterochromatin formation. *Genes Dev*, **19**, 1705–1714.
60 Li, F., Goto, D.B., Zaratiegui, M., Tang, X., Martienssen, R., and Cande, W.Z. (2005) Two novel proteins, dos1 and dos2, interact with rik1 to regulate heterochromatic RNA interference and histone modification. *Curr Biol*, **15**, 1448–1457.
61 Zhang, K., Mosch, K., Fischle, W., and Grewal, S.I. (2008) Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin. *Nat Struct Mol Biol*, **15**, 381–388.
62 Petrie, V.J., Wuitschick, J.D., Givens, C.D., Kosinski, A.M., and Partridge, J. F. (2005) RNA interference (RNAi)-dependent and RNAi-independent association of the Chp1 chromodomain protein with distinct heterochromatic loci in fission yeast. *Mol Cell Biol*, **25**, 2331–2346.
63 Kim, H.S., Choi, E.S., Shin, J.A., Jang, Y.K., and Park, S.D. (2004) Regulation of Swi6/HP1-dependent heterochromatin assembly by cooperation of components of the mitogen-activated protein kinase pathway and a histone deacetylase Clr6. *J Biol Chem*, **279**, 42850–42859.
64 El-Shami, M., Pontier, D., Lahmy, S., Braun, L., Picart, C., Vega, D., Hakimi, M.A., Jacobsen, S.E., Cooke, R., and Lagrange, T. (2007) Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. *Genes Dev*, **21**, 2539–2544.
65 Henderson, I.R. and Jacobsen, S.E. (2007) Epigenetic inheritance in plants. *Nature*, **447**, 418–424.
66 Mette, M.F., Aufsatz, W., van der Winden, J., Matzke, M.A., and Matzke, A.J. (2000) Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *EMBO J*, **19**, 5194–5201.
67 Matzke, M., Aufsatz, W., Kanno, T., Daxinger, L., Papp, I., Mette, M.F., and Matzke, A.J. (2004) Genetic analysis of RNA-mediated transcriptional gene silencing. *Biochim Biophys Acta*, **1677**, 129–141.
68 Slotkin, R.K., Vaughn, M., Borges, F., Tanurdzic, M., Becker, J.D., Feijo, J.A., and Martienssen, R.A. (2009) Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell*, **136**, 461–472.
69 Kanellopoulou, C., Muljo, S.A., Kung, A.L., Ganesan, S., Drapkin, R., Jenuwein, T., Livingston, D.M., and Rajewsky, K. (2005) Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev*, **19**, 489–501.
70 Murchison, E.P., Partridge, J.F., Tam, O.H., Cheloufi, S., and Hannon, G.J. (2005) Characterization of Dicer-deficient murine embryonic stem cells. *Proc Natl Acad Sci USA*, **102**, 12135–12140.
71 Svoboda, P., Stein, P., Anger, M., Bernstein, E., Hannon, G.J., and Schultz, R.M. (2004) RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. *Dev Biol*, **269**, 276–285.
72 Chow, J.C., Ciaudo, C., Fazzari, M.J., Mise, N., Servant, N., Glass, J.L., Attreed, M., Avner, P., Wutz, A., Barillot, E., Greally, J.M., *et al.* (2010) LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. *Cell*, **141**, 956–969.
73 Robert, V.J., Sijen, T., van Wolfswinkel, J., and Plasterk, R.H. (2005) Chromatin and RNAi factors protect the *C. elegans*

germline against repetitive sequences. *Genes Dev*, **19**, 782–787.

74 Grishok, A., Sinskey, J.L., and Sharp, P.A. (2005) Transcriptional silencing of a transgene by RNAi in the soma of *C. elegans*. *Genes Dev*, **19**, 683–696.

75 Kim, J.K., Gabel, H.W., Kamath, R.S., Tewari, M., Pasquinelli, A., Rual, J.F., Kennedy, S., Dybbs, M., Bertin, N., Kaplan, J.M., Vidal, M., *et al.* (2005) Functional genomic analysis of RNA interference in *C. elegans*. *Science*, **308**, 1164–1167.

76 Cox, D.N., Chao, A., and Lin, H. (2000) piwi encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. *Development*, **127**, 503–514.

77 Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007) Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell*, **128**, 1089–1103.

78 Brower-Toland, B., Findley, S.D., Jiang, L., Liu, L., Yin, H., Dus, M., Zhou, P., Elgin, S.C., and Lin, H. (2007) *Drosophila* PIWI associates with chromatin and interacts directly with HP1a. *Genes Dev*, **21**, 2300–2311.

79 Aravin, A.A., Sachidanandam, R., Girard, A., Fejes-Toth, K., and Hannon, G.J. (2007) Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science*, **316**, 744–747.

80 Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., Asada, N., Kojima, K., Yamaguchi, Y., Ijiri, T.W., Hata, K., *et al.* (2008) DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev*, **22**, 908–917.

81 Carmell, M.A., Girard, A., van de Kant, H.J., Bourc'his, D., Bestor, T.H., de Rooij, D.G., and Hannon, G.J. (2007) MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev Cell*, **12**, 503–514.

82 Aravin, A.A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K.F., Bestor, T., and Hannon, G. J. (2008) A piRNA pathway primed by individual transposons is linked to *de novo* DNA methylation in mice. *Mol Cell*, **31**, 785–799.

83 Sugiyama, T., Cam, H., Verdel, A., Moazed, D., and Grewal, S.I. (2005) RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production. *Proc Natl Acad Sci USA*, **102**, 152–157.

84 Hall, I.M., Noma, K., and Grewal, S.I. (2003) RNA interference machinery regulates chromosome dynamics during mitosis and meiosis in fission yeast. *Proc Natl Acad Sci USA*, **100**, 193–198.

85 Grimaud, C., Bantignies, F., Pal-Bhadra, M., Ghana, P., Bhadra, U., and Cavalli, G. (2006) RNAi components are required for nuclear clustering of Polycomb group response elements. *Cell*, **124**, 957–971.

86 Kuramochi-Miyagawa, S., Kimura, T., Ijiri, T.W., Isobe, T., Asada, N., Fujita, Y., Ikawa, M., Iwai, N., Okabe, M., Deng, W., Lin, H., *et al.* (2004) Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. *Development*, **131**, 839–849.

87 Klattenhoff, C., Bratu, D.P., McGinnis-Schultz, N., Koppetsch, B.S., Cook, H. A., and Theurkauf, W.E. (2007) *Drosophila* rasiRNA pathway mutations disrupt embryonic axis specification through activation of an ATR/Chk2 DNA damage response. *Dev Cell*, **12**, 45–55.

88 Duharcourt, S., Butler, A., and Meyer, E. (1995) Epigenetic self-regulation of developmental excision of an internal eliminated sequence on *Paramecium tetraurelia*. *Genes Dev*, **9**, 2065–2077.

89 Duharcourt, S., Keller, A.M., and Meyer, E. (1998) Homology-dependent maternal inhibition of developmental excision of internal eliminated sequences in *Paramecium tetraurelia*. *Mol Cell Biol*, **18**, 7075–7085.

90 Hayashi, T., Hayashi, H., Fusauchi, Y., and Iwai, K. (1984) Tetrahymena histone H3. Purification and two variant sequences. *J Biochem*, **95**, 1741–1749.

91 Jackson, J.D. and Gorovsky, M.A. (2000) Histone H2A.Z has a conserved function that is distinct from that of the major H2A sequence variants. *Nucleic Acids Res*, **28**, 3811–3816.

92 Allis, C.D., Glover, C.V., Bowen, J.K., and Gorovsky, M.A. (1980) Histone variants specific to the transcriptionally active, amitotically dividing macronucleus of the unicellular eucaryote, *Tetrahymena thermophila*. *Cell*, **20**, 609–617.

93 Yu, L. and Gorovsky, M.A. (1997) Constitutive expression, not a particular primary sequence, is the important feature of the H3 replacement variant hv2 in *Tetrahymena thermophila*. *Mol Cell Biol*, **17**, 6303–6310.

94 Allis, C.D., Richman, R., Gorovsky, M. A., Ziegler, Y.S., Touchstone, B., Bradley, W.A., and Cook, R.G. (1986) hv1 is an evolutionarily conserved H2A variant that is preferentially associated with active genes. *J Biol Chem*, **261**, 1941–1948.

95 Yao, M.C., Zheng, K., and Yao, C.H. (1987) A conserved nucleotide sequence at the sites of developmentally regulated chromosomal breakage in *Tetrahymena*. *Cell*, **48**, 779–788.

96 Hamilton, E.P., Williamson, S., Dunn, S., Merriam, V., Lin, C., Vong, L., Russell-Colantonio, J., and Orias, E. (2006) The highly conserved family of *Tetrahymena thermophila* chromosome breakage elements contains an invariant 10-base-pair core. *Eukaryot Cell*, **5**, 771–780.

97 Yao, M.C. and Chao, J.L. (2005) RNA-guided DNA deletion in *Tetrahymena*: an RNAi-based mechanism for programmed genome rearrangements. *Annu Rev Genet*, **39**, 537–559.

98 Malone, C.D., Anderson, A.M., Motl, J. A., Rexer, C.H., and Chalker, D.L. (2005) Germ line transcripts are processed by a Dicer-like protein that is essential for developmentally programmed genome rearrangements of *Tetrahymena thermophila*. *Mol Cell Biol*, **25**, 9151–9164.

99 Chalker, D.L. and Yao, M.C. (2001) Nongenic, bidirectional transcription precedes and may promote developmental DNA deletion in *Tetrahymena thermophila*. *Genes Dev*, **15**, 1287–1298.

100 Mochizuki, K., Fine, N.A., Fujisawa, T., and Gorovsky, M.A. (2002) Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in *Tetrahymena*. *Cell*, **110**, 689–699.

101 Mochizuki, K. and Gorovsky, M.A. (2005) A Dicer-like protein in *Tetrahymena* has distinct functions in genome rearrangement, chromosome segregation, and meiotic prophase. *Genes Dev*, **19**, 77–89.

102 Mochizuki, K. and Gorovsky, M.A. (2004) Conjugation-specific small RNAs in *Tetrahymena* have predicted properties of scan (scn) RNAs involved in genome rearrangement. *Genes Dev*, **18**, 2068–2073.

103 Prescott, D.M. (1994) The DNA of ciliated protozoa. *Microbiol Rev*, **58**, 233–267.

104 Mollenbeck, M., Zhou, Y., Cavalcanti, A.R., Jonsson, F., Higgins, B.P., Chang, W.J., Juranek, S., Doak, T.G., Rozenberg, G., Lipps, H.J., and Landweber, L.F. (2008) The pathway to detangle a scrambled gene. *PLoS One*, **3**, e2330.

105 Nowacki, M., Vijayan, V., Zhou, Y., Schotanus, K., Doak, T.G., and Landweber, L.F. (2008) RNA-mediated epigenetic programming of a genome-rearrangement pathway. *Nature*, **451**, 153–158.

106 Kim, D.H., Villeneuve, L.M., Morris, K. V., and Rossi, J.J. (2006) Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. *Nat Struct Mol Biol*, **13**, 793–797.

107 Giles, K.E., Ghirlando, R., and Felsenfeld, G. (2010) Maintenance of a constitutive heterochromatin domain in vertebrates by a Dicer-dependent

mechanism. *Nat Cell Biol*, **12**, 94–99; sup pp 91–96.

108 Weinberg, M.S., Villeneuve, L.M., Ehsani, A., Amarzguioui, M., Aagaard, L., Chen, Z.X., Riggs, A.D., Rossi, J.J., and Morris, K.V. (2006) The antisense strand of small interfering RNAs directs histone methylation and transcriptional gene silencing in human cells. *RNA*, **12**, 256–262.

109 Han, J., Kim, D., and Morris, K.V. (2007) Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cells. *Proc Natl Acad Sci USA*, **104**, 12422–12427.

110 Janowski, B.A., Huffman, K.E., Schwartz, J.C., Ram, R., Hardy, D., Shames, D.S., Minna, J.D., and Corey, D.R. (2005) Inhibiting gene expression at transcription start sites in chromosomal DNA with antigene RNAs. *Nat Chem Biol*, **1**, 216–222.

111 Janowski, B.A., Huffman, K.E., Schwartz, J.C., Ram, R., Nordsell, R., Shames, D.S., Minna, J.D., and Corey, D.R. (2006) Involvement of AGO1 and AGO2 in mammalian transcriptional silencing. *Nat Struct Mol Biol*, **13**, 787–792.

112 Napoli, S., Pastori, C., Magistri, M., Carbone, G.M., and Catapano, C.V. (2009) Promoter-specific transcriptional interference and c-myc gene silencing by siRNAs in human cells. *EMBO J*, **28**, 1708–1719.

113 Guang, S., Bochner, A.F., Burkhart, K. B., Burton, N., Pavelec, D.M., and Kennedy, S. (2010) Small regulatory RNAs inhibit RNA polymerase II during the elongation phase of transcription. *Nature*, **465**, 1097–1101.

114 Bao, N., Lye, K.W., and Barton, M.K. (2004) MicroRNA binding sites in Arabidopsis class III HD-ZIP mRNAs are required for methylation of the template chromosome. *Dev Cell*, **7**, 653–662.

115 Bayne, E.H. and Allshire, R.C. (2005) RNA-directed transcriptional gene silencing in mammals. *Trends Genet*, **21**, 370–373.

116 Gonzalez, S., Pisano, D.G., and Serrano, M. (2008) Mechanistic principles of chromatin remodeling guided by siRNAs and miRNAs. *Cell Cycle*, **7**, 2601–2608.

117 Kim, D.H., Saetrom, P., Snove, O. Jr, and Rossi, J.J. (2008) MicroRNA-directed transcriptional gene silencing in mammalian cells. *Proc Natl Acad Sci USA*, **105**, 16230–16235.

118 Schwartz, J.C., Younger, S.T., Nguyen, N.B., Hardy, D.B., Monia, B.P., Corey, D.R., and Janowski, B.A. (2008) Antisense transcripts are targets for activating small RNAs. *Nat Struct Mol Biol*, **15**, 842–848.

119 Li, L.C., Okino, S.T., Zhao, H., Pookot, D., Place, R.F., Urakami, S., Enokida, H., and Dahiya, R. (2006) Small dsRNAs induce transcriptional activation in human cells. *Proc Natl Acad Sci USA*, **103**, 17337–17342.

120 Janowski, B.A., Younger, S.T., Hardy, D.B., Ram, R., Huffman, K.E., and Corey, D.R. (2007) Activating gene expression in mammalian cells with promoter-targeted duplex RNAs. *Nat Chem Biol*, **3**, 166–173.

121 Morris, K.V., Santoso, S., Turner, A.M., Pastori, C., and Hawkins, P.G. (2008) Bidirectional transcription directs both transcriptional gene activation and suppression in human cells. *PLoS Genet*, **4**, e1000258.

122 Hawkins, P.G. and Morris, K.V. (2008) RNA and transcriptional modulation of gene expression. *Cell Cycle*, **7**, 602–607.

123 Davis, C.A. and Ares, M. Jr (2006) Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA*, **103**, 3262–3267.

124 Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Munster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., and Steinmetz, L.M. (2009) Bidirectional promoters generate pervasive transcription in yeast. *Nature*, **457**, 1033–1037.

125 Neil, H., Malabat, C., d'Aubenton-Carafa, Y., Xu, Z., Steinmetz, L.M., and Jacquier, A. (2009) Widespread

bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature*, **457**, 1038–1042.

126 Steinmetz, E.J., Warren, C.L., Kuehner, J.N., Panbehi, B., Ansari, A.Z., and Brow, D.A. (2006) Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. *Mol Cell*, **24**, 735–746.

127 Kopcewicz, K.A., O'Rourke, T.W., and Reines, D. (2007) Metabolic regulation of IMD2 transcription and an unusual DNA element that generates short transcripts. *Mol Cell Biol*, **27**, 2821–2829.

128 Kuehner, J.N. and Brow, D.A. (2008) Regulation of a eukaryotic gene by GTP-dependent start site selection and transcription attenuation. *Mol Cell*, **31**, 201–211.

129 Jenks, M.H., O'Rourke, T.W., and Reines, D. (2008) Properties of an intergenic terminator and start site switch that regulate IMD2 transcription in yeast. *Mol Cell Biol*, **28**, 3883–3893.

130 Taft, R.J., Glazov, E.A., Cloonan, N., Simons, C., Stephen, S., Faulkner, G.J., Lassmann, T., Forrest, A.R., Grimmond, S.M., Schroder, K., Irvine, K., *et al.* (2009) Tiny RNAs associated with transcription start sites in animals. *Nat Genet*, **41**, 572–578.

131 Core, L.J. and Lis, J.T. (2008) Transcription regulation through promoter-proximal pausing of RNA polymerase II. *Science*, **319**, 1791–1792.

132 Seila, A.C., Calabrese, J.M., Levine, S. S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008) Divergent transcription from active promoters. *Science*, **322**, 1849–1851.

133 Jacquier, A. (2009) The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. *Nat Rev Genet*, **10**, 833–844.

134 Otsuka, Y., Kedersha, N.L., and Schoenberg, D.R. (2009) Identification of a cytoplasmic complex that adds a cap onto 5′-monophosphate RNA. *Mol Cell Biol*, **29**, 2155–2167.

135 Borsani, G., Tonlorenzi, R., Simmler, M.C., Dandolo, L., Arnaud, D., Capra, V., Grompe, M., Pizzuti, A., Muzny, D., Lawrence, C., *et al.* (1991) Characterization of a murine gene expressed from the inactive X chromosome. *Nature*, **351**, 325–329.

136 Brown, C.J., Ballabio, A., Rupert, J.L., Lafreniere, R.G., Grompe, M., Tonlorenzi, R., and Willard, H.F. (1991) A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature*, **349**, 38–44.

137 Brockdorff, N., Ashworth, A., Kay, G. F., McCabe, V.M., Norris, D.P., Cooper, P.J., Swift, S., and Rastan, S. (1992) The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell*, **71**, 515–526.

138 Lee, J.T., Davidow, L.S., and Warshawsky, D. (1999) Tsix, a gene antisense to Xist at the X-inactivation centre. *Nat Genet*, **21**, 400–404.

139 Lee, J.T. and Lu, N. (1999) Targeted mutagenesis of Tsix leads to nonrandom X inactivation. *Cell*, **99**, 47–57.

140 Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J., and Lee, J.T. (2008) Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science*, **322**, 750–756.

141 Lee, J.T. (2009) Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome. *Genes Dev*, **23**, 1831–1842.

142 Bacher, C.P., Guggiari, M., Brors, B., Augui, S., Clerc, P., Avner, P., Eils, R., and Heard, E. (2006) Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. *Nat Cell Biol*, **8**, 293–299.

143 Nicodemi, M. and Prisco, A. (2007) Symmetry-breaking model for X-chromosome inactivation. *Phys Rev Lett*, **98**, 108104.

144 Jeong, K.S., Park, J.H., and Lee, S. (2005) The analysis of X-chromosome inactivation-related gene expression from single mouse embryo with sex-determination. *Biochem Biophys Res Commun*, **333**, 803–807.

145 Navarro, P., Chambers, I., Karwacki-Neisius, V., Chureau, C., Morey, C., Rougeulle, C., and Avner, P. (2008) Molecular coupling of Xist regulation and pluripotency. *Science*, **321**, 1693–1695.

146 Chaumeil, J., Le Baccon, P., Wutz, A. and Heard, E. (2006) A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev*, **20**, 2223–2237.

147 Kay, G.F., Barton, S.C., Surani, M.A., and Rastan, S. (1994) Imprinting and X chromosome counting mechanisms determine Xist expression in early mouse development. *Cell*, **77**, 639–650.

148 Panning, B., Dausman, J. and Jaenisch, R. (1997) X chromosome inactivation is mediated by Xist RNA stabilization. *Cell*, **90**, 907–916.

149 Wutz, A. and Jaenisch, R. (2000) A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Mol Cell*, **5**, 695–705.

150 Sheardown, S.A., Duthie, S.M., Johnston, C.M., Newall, A.E., Formstone, E.J., Arkell, R.M., Nesterova, T.B., Alghisi, G.C., Rastan, S., and Brockdorff, N. (1997) Stabilization of Xist RNA mediates initiation of X chromosome inactivation. *Cell*, **91**, 99–107.

151 Silva, J., Mak, W., Zvetkova, I., Appanah, R., Nesterova, T.B., Webster, Z., Peters, A.H., Jenuwein, T., Otte, A. P., and Brockdorff, N. (2003) Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. *Dev Cell*, **4**, 481–495.

152 Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., and Zhang, Y. (2003) Role of histone H3 lysine 27 methylation in X inactivation. *Science*, **300**, 131–135.

153 Oda, H., Okamoto, I., Murphy, N., Chu, J., Price, S.M., Shen, M.M., Torres-Padilla, M.E., Heard, E., and Reinberg, D. (2009) Monomethylation of histone H4-lysine 20 is involved in chromosome structure and stability and is essential for mouse development. *Mol Cell Biol*, **29**, 2278–2295.

154 Perche, P.Y., Vourc'h, C., Konecny, L., Souchier, C., Robert-Nicoud, M., Dimitrov, S., and Khochbin, S. (2000) Higher concentrations of histone macroH2A in the Barr body are correlated with higher nucleosome density. *Curr Biol*, **10**, 1531–1534.

155 Sado, T., Fenner, M.H., Tan, S.S., Tam, P., Shioda, T., and Li, E. (2000) X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation. *Dev Biol*, **225**, 294–303.

156 Zhang, L.F., Huynh, K.D. and Lee, J.T. (2007) Perinucleolar targeting of the inactive X during S phase: evidence for a role in the maintenance of silencing. *Cell*, **129**, 693–706.

157 Xu, N., Tsai, C.L., and Lee, J.T. (2006) Transient homologous chromosome pairing marks the onset of X inactivation. *Science*, **311**, 1149–1152.

158 Xu, N., Donohoe, M.E., Silva, S.S., and Lee, J.T. (2007) Evidence that homologous X-chromosome pairing requires transcription and Ctcf protein. *Nat Genet*, **39**, 1390–1396.

159 Donohoe, M.E., Silva, S.S., Pinter, S.F., Xu, N., and Lee, J.T. (2009) The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. *Nature*, **460**, 128–132.

160 Navarro, P., Pichard, S., Ciaudo, C., Avner, P., and Rougeulle, C. (2005) Tsix transcription across the Xist gene alters chromatin conformation without affecting Xist transcription: implications for X-chromosome inactivation. *Genes Dev*, **19**, 1474–1484.

161 Sado, T., Hoki, Y., and Sasaki, H. (2005) Tsix silences Xist through modification of chromatin structure. *Dev Cell*, **9**, 159–165.

162 Sun, B.K., Deaton, A.M., and Lee, J.T. (2006) A transient heterochromatic state in Xist preempts X inactivation

choice without RNA stabilization. *Mol Cell*, **21**, 617–628.

163 Migeon, B.R., Chowdhury, A.K., Dunston, J.A., and McIntosh, I. (2001) Identification of TSIX, encoding an RNA antisense to human XIST, reveals differences from its murine counterpart: implications for X inactivation. *Am J Hum Genet*, **69**, 951–960.

164 Wutz, A., Rasmussen, T.P., and Jaenisch, R. (2002) Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat Genet*, **30**, 167–174.

165 Kohlmaier, A., Savarese, F., Lachner, M., Martens, J., Jenuwein, T., and Wutz, A. (2004) A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. *PLoS Biol*, **2**, E171.

166 Amrein, H. and Axel, R. (1997) Genes expressed in neurons of adult male Drosophila. *Cell*, **88**, 459–469.

167 Payer, B. and Lee, J.T. (2008) X chromosome dosage compensation: how mammals keep the balance. *Annu Rev Genet*, **42**, 733–772.

168 Franke, A. and Baker, B.S. (1999) The rox1 and rox2 RNAs are essential components of the compensasome, which mediates dosage compensation in *Drosophila*. *Mol Cell*, **4**, 117–122.

169 Stuckenholz, C., Meller, V.H., and Kuroda, M.I. (2003) Functional redundancy within roX1, a noncoding RNA involved in dosage compensation in *Drosophila melanogaster*. *Genetics*, **164**, 1003–1014.

170 Kotlikova, I.V., Demakova, O.V., Semeshin, V.F., Shloma, V.V., Boldyreva, L.V., Kuroda, M.I., and Zhimulev, I.F. (2006) The Drosophila dosage compensation complex binds to polytene chromosomes independently of developmental changes in transcription. *Genetics*, **172**, 963–974.

171 Legube, G., McWeeney, S.K., Lercher, M.J., and Akhtar, A. (2006) X-chromosome-wide profiling of MSL-1 distribution and dosage compensation in *Drosophila*. *Genes Dev*, **20**, 871–883.

172 Kmita, M. and Duboule, D. (2003) Organizing axes in time and space; 25 years of colinear tinkering. *Science*, **301**, 331–333.

173 Pearson, J.C., Lemons, D., and McGinnis, W. (2005) Modulating Hox gene functions during animal body patterning. *Nat Rev Genet*, **6**, 893–904.

174 Klymenko, T. and Muller, J. (2004) The histone methyltransferases Trithorax and Ash1 prevent transcriptional silencing by Polycomb group proteins. *EMBO Rep*, **5**, 373–377.

175 Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., and Chang, H.Y. (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*, **129**, 1311–1323.

176 Sessa, L., Breiling, A., Lavorgna, G., Silvestri, L., Casari, G., and Orlando, V. (2007) Noncoding RNA synthesis and loss of Polycomb group repression accompanies the colinear activation of the human HOXA cluster. *RNA*, **13**, 223–239.

177 Ringrose, L. (2007) Polycomb comes of age: genome-wide profiling of target sites. *Curr Opin Cell Biol*, **19**, 290–297.

178 Schmitt, S., Prestel, M., and Paro, R. (2005) Intergenic transcription through a polycomb group response element counteracts silencing. *Genes Dev*, **19**, 697–708.

179 Sasaki, Y.T., Sano, M., Kin, T., Asai, K., and Hirose, T. (2007) Coordinated expression of ncRNAs and HOX mRNAs in the human HOXA locus. *Biochem Biophys Res Commun*, **357**, 724–730.

180 Petruk, S., Sedkov, Y., Riley, K.M., Hodgson, J., Schweisguth, F., Hirose, S., Jaynes, J.B., Brock, H.W., and Mazo, A. (2006) Transcription of bxd noncoding RNAs promoted by trithorax represses Ubx in cis by transcriptional interference. *Cell*, **127**, 1209–1221.

181 Sanchez-Elsner, T., Gou, D., Kremmer, E., and Sauer, F. (2006) Noncoding RNAs of trithorax response elements

recruit *Drosophila* Ash1 to Ultrabithorax. *Science*, **311**, 1118–1123.
182 Ilik, I. and Akhtar, A. (2009) roX RNAs: non-coding regulators of the male X chromosome in flies. *RNA Biol*, **6**, 113–121.
183 Uhler, J.P., Hertel, C., and Svejstrup, J. Q. (2007) A role for noncoding transcription in activation of the yeast PHO5 gene. *Proc Natl Acad Sci USA*, **104**, 8011–8016.
184 Martens, J.A., Laprade, L., and Winston, F. (2004) Intergenic transcription is required to repress the *Saccharomyces cerevisiae* SER3 gene. *Nature*, **429**, 571–574.
185 Ashe, H.L., Monks, J., Wijgerde, M., Fraser, P., and Proudfoot, N.J. (1997) Intergenic transcription and transinduction of the human beta-globin locus. *Genes Dev*, **11**, 2494–2509.
186 Bolland, D.J., Wood, A.L., Johnston, C. M., Bunting, S.F., Morgan, G., Chakalova, L., Fraser, P.J., and Corcoran, A.E. (2004) Antisense intergenic transcription in V(D)J recombination. *Nat Immunol*, **5**, 630–637.
187 Pauler, F.M., Koerner, M.V., and Barlow, D.P. (2007) Silencing by imprinted noncoding RNAs: is transcription the answer? *Trends Genet*, **23**, 284–292.
188 Hirota, K., Miyoshi, T., Kugou, K., Hoffman, C.S., Shibata, T., and Ohta, K. (2008) Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. *Nature*, **456**, 130–134.
189 Espinoza, C.A., Goodrich, J.A., and Kugel, J.F. (2007) Characterization of the structure, function, and mechanism of B2 RNA, an ncRNA repressor of RNA polymerase II transcription. *RNA*, **13**, 583–596.
190 Espinoza, C.A., Allen, T.A., Hieb, A.R., Kugel, J.F., and Goodrich, J.A. (2004) B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. *Nat Struct Mol Biol*, **11**, 822–829.
191 Yakovchuk, P., Goodrich, J.A., and Kugel, J.F. (2009) B2 RNA and Alu RNA repress transcription by disrupting contacts between RNA polymerase II and promoter DNA within assembled complexes. *Proc Natl Acad Sci U S A*, **106**, 5569–5574.
192 Nguyen, V.T., Kiss, T., Michels, A.A., and Bensaude, O. (2001) 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. *Nature*, **414**, 322–325.
193 Yik, J.H., Chen, R., Nishimura, R., Jennings, J.L., Link, A.J., and Zhou, Q. (2003) Inhibition of P-TEFb (CDK9/Cyclin T) kinase and RNA polymerase II transcription by the coordinated actions of HEXIM1 and 7SK snRNA. *Mol Cell*, **12**, 971–982.
194 Martianov, I., Ramadass, A., Serra Barros, A., Chow, N., and Akoulitchev, A. (2007) Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature*, **445**, 666–670.
195 Krystal, G.W., Armstrong, B.C., and Battey, J.F. (1990) N-myc mRNA forms an RNA-RNA duplex with endogenous antisense transcripts. *Mol Cell Biol*, **10**, 4180–4191.
196 Nagano, T., Mitchell, J.A., Sanz, L.A., Pauler, F.M., Ferguson-Smith, A.C., Feil, R., and Fraser, P. (2008) The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science*, **322**, 1717–1720.
197 Zhao, R. and Daley, G.Q. (2008) From fibroblasts to iPS cells: induced pluripotency by defined factors. *J Cell Biochem*, **105**, 949–955.
198 Tufarelli, C., Stanley, J.A., Garrick, D., Sharpe, J.A., Ayyub, H., Wood, W.G., and Higgs, D.R. (2003) Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nat Genet*, **34**, 157–165.
199 Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., Markenscoff-Papadimitriou, E., *et al.* (2010) Widespread transcription at neuronal activity-regulated enhancers. *Nature*, **465**, 182–187.

200 Razin, S.V., Rynditch, A., Borunova, V., Ioudinkova, E., Smalko, V., and Scherrer, K. (2004) The 33 kb transcript of the chicken alpha-globin gene domain is part of the nuclear matrix. *J Cell Biochem*, **92**, 445–457.

201 Lunyak, V.V., Prefontaine, G.G., Nunez, E., Cramer, T., Ju, B.G., Ohgi, K.A., Hutt, K., Roy, R., Garcia-Diaz, A., Zhu, X., Yung, Y., *et al.* (2007) Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. *Science*, **317**, 248–251.

202 Hill, A.E., Hong, J.S., Wen, H., Teng, L., McPherson, D.T., McPherson, S.A., Levasseur, D.N., and Sorscher, E.J. (2006) Micro-RNA-like effects of complete intronic sequences. *Front Biosci*, **11**, 1998–2006.

203 Yan, M.D., Hong, C.C., Lai, G.M., Cheng, A.L., Lin, Y.W., and Chuang, S. E. (2005) Identification and characterization of a novel gene Saf transcribed from the opposite strand of Fas. *Hum Mol Genet*, **14**, 1465–1474.

204 Nakaya, H.I., Amaral, P.P., Louro, R., Lopes, A., Fachel, A.A., Moreira, Y.B., El-Jundi, T.A., da Silva, A.M., Reis, E. M., and Verjovski-Almeida, S. (2007) Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription. *Genome Biol*, **8**, R43.

205 Tripathi, V., Ellis, J.D., Shen, Z., Song, D.Y., Pan, Q., Watt, A.T., Freier, S.M., Bennett, C.F., Sharma, A., Bubulya, P. A., Blencowe, B.J., *et al.* (2010) The Nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell*, **39**, 925–938.

206 Capaccioli, S., Quattrone, A., Schiavone, N., Calastretti, A., Copreni, E., Bevilacqua, A., Canti, G., Gong, L., Morelli, S., and Nicolin, A. (1996) A bcl-2/IgH antisense transcript deregulates bcl-2 gene expression in human follicular lymphoma t (14;18) cell lines. *Oncogene*, **13**, 105–115.

207 Schwartz, Y.B. and Pirrotta, V. (2007) Polycomb silencing mechanisms and the management of genomic programmes. *Nat Rev Genet*, **8**, 9–22.

208 Han, L., Witmer, P.D., Casey, E., Valle, D., and Sukumar, S. (2007) DNA methylation regulates MicroRNA expression. *Cancer Biol Ther*, **6**, 1284–1288.

209 Morris, K.V., Chan, S.W., Jacobsen, S. E., and Looney, D.J. (2004) Small interfering RNA-induced transcriptional gene silencing in human cells. *Science*, **305**, 1289–1292.

210 Yu, W., Gius, D., Onyango, P., Muldoon-Jacobs, K., Karp, J., Feinberg, A.P., and Cui, H. (2008) Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature*, **451**, 202–206.

211 Wang, X., Arai, S., Song, X., Reichart, D., Du, K., Pascual, G., Tempst, P., Rosenfeld, M.G., Glass, C.K., and Kurokawa, R. (2008) Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature*, **454**, 126–130.

212 Spahn, L. and Barlow, D.P. (2003) An ICE pattern crystallizes. *Nat Genet*, **35**, 11–12.

213 Lewis, A. and Reik, W. (2006) How imprinting centres work. *Cytogenet Genome Res*, **113**, 81–89.

214 Reik, W. and Walter, J. (2001) Evolution of imprinting mechanisms: the battle of the sexes begins in the zygote. *Nat Genet*, **27**, 255–256.

215 Yang, Y., Li, T., Vu, T.H., Ulaner, G.A., Hu, J.F., and Hoffman, A.R. (2003) The histone code regulating expression of the imprinted mouse Igf2r gene. *Endocrinology*, **144**, 5658–5670.

216 Stricker, S.H., Steenpass, L., Pauler, F. M., Santoro, F., Latos, P.A., Huang, R., Koerner, M.V., Sloane, M.A., Warczok, K.E., and Barlow, D.P. (2008) Silencing and transcriptional properties of the imprinted Airn ncRNA are independent of the endogenous promoter. *EMBO J*, **27**, 3116–3128.

217 Latos, P.A., Stricker, S.H., Steenpass, L., Pauler, F.M., Huang, R., Senergin, B.H., Regha, K., Koerner, M.V.,

Warczok, K.E., Unger, C., and Barlow, D.P. (2009) An *in vitro* ES cell imprinting model shows that imprinted expression of the Igf2r gene arises from an allele-specific expression bias. *Development*, **136**, 437–448.

218 Fox, A.H., Bond, C.S., and Lamond, A. I. (2005) P54nrb forms a heterodimer with PSP1 that localizes to paraspeckles in an RNA-dependent manner. *Mol Biol Cell*, **16**, 5304–5315.

219 Prasanth, K.V., Prasanth, S.G., Xuan, Z., Hearn, S., Freier, S.M., Bennett, C. F., Zhang, M.Q., and Spector, D.L. (2005) Regulating gene expression through RNA nuclear retention. *Cell*, **123**, 249–263.

220 Blower, M.D., Nachury, M., Heald, R., and Weis, K. (2005) A Rae1-containing ribonucleoprotein complex is required for mitotic spindle assembly. *Cell*, **121**, 223–234.

Keywords: genome, organization, chromatin, histone, DNA methylation, imprinting