Curr Mol Bio Rep (2015) 1:10–18
DOI 10.1007/s40610-015-0002-6

GENE REGULATION (JL WORKMAN AND SM ABMAYR, SECTION EDITORS)

Non-coding RNAs in Transcriptional Regulation

Yung-Chia Ariel Chen · Alexei A. Aravin

Published online: 17 February 2015

Abstract Transcriptional gene silencing guided by small RNAs is a process conserved from protozoa to mammals. Small RNAs loaded into Argonaute family proteins direct repressive histone modifications or DNA cytosine methylation to homologous regions of the genome. Small RNA-mediated transcriptional silencing is required for many biological processes, including repression of transposable elements, maintaining the genome stability/integrity, and epigenetic inheritance of gene expression. Here, we will summarize the current knowledge about small RNA biogenesis and mechanisms of transcriptional regulation in plants, *Drosophila*, *Caenorhabditis elegans*, and mice. Furthermore, a rapidly growing number of long non-coding RNAs (lncRNAs) have been implicated as important players in transcription regulation. We will discuss current models for long non-coding RNA-mediated gene regulation.

Keywords Small RNA · lncRNA · Argonaute · PIWI · TGS · RdDM

Introduction

RNA interference (RNAi) is an important cellular process that uses small RNA molecules between 20 and 30 nt in length for sequence-specific recognition and regulation of complementary target sequences. RNAi was first shown to act post-transcriptionally by degrading targeted messenger RNAs (mRNAs) in the cytoplasm [1]. This process of post-transcriptional gene silencing (PTGS) has been shown well-conserved among all the model Metazoa organisms. In yeasts and plants, small RNAs were also shown to elicit transcriptional gene silencing (TGS) in the nucleus by guiding repressive epigenetic modifications such as DNA cytosine methylation and histone modifications to complementary genomic targets. Although TGS was initially thought to be absent in Metazoa, recent studies proved that similar mechanisms are active in germ cells of *Caenorhabditis elegans*, *Drosophila*, and mammals. These findings suggest that small RNA-mediated TGS is a conserved mechanism in many eukaryotes.

Small RNAs that can operate TGS includes two main types: small interfering RNAs (siRNAs) and PIWI-interacting RNAs (piRNAs). siRNAs are processed from double-stranded siRNA precursors by an RNase-III-like Dicer family protein. After being generated, siRNAs are loaded into an effector Argonaute (AGO) protein to operate TGS or PTGS functions. Epigenetic modification directed by siRNA has been most studied in plants and yeasts, but increasing evidence shows that this process might also occur in metazoan somatic cells [2]. In contrast to siRNAs, piRNA biogenesis is independent of Dicer, and piRNA precursors do not have an obvious secondary structure. The effector proteins for piRNA are PIWI proteins, a germline-specific clade of the AGO protein family. piRNAs and PIWI proteins are not present in plants or fungi, which use siRNA for TGS processes (Fig. 1).

An essential and conserved function for siRNA in plants and piRNAs in Metazoa is to protect genome integrity by repressing the expression of transposable elements (TEs). Indeed, failure of piRNA repression in flies and mice leads to transposon activation, formation of double-stranded DNA breaks, and sterility. Moreover, small RNAs expressed in the germline cells of *C. elegans* and *Drosophila* are inherited by

This article is part of the Topical Collection on *Gene Regulation*

Y.-C. A. Chen · A. A. Aravin (✉)
Division of Biology and Biological Engineering, California Institute of Technology, 147-75, 1200 E. California Blvd., Pasadena, CA 91125, USA
e-mail: aaa@caltech.edu

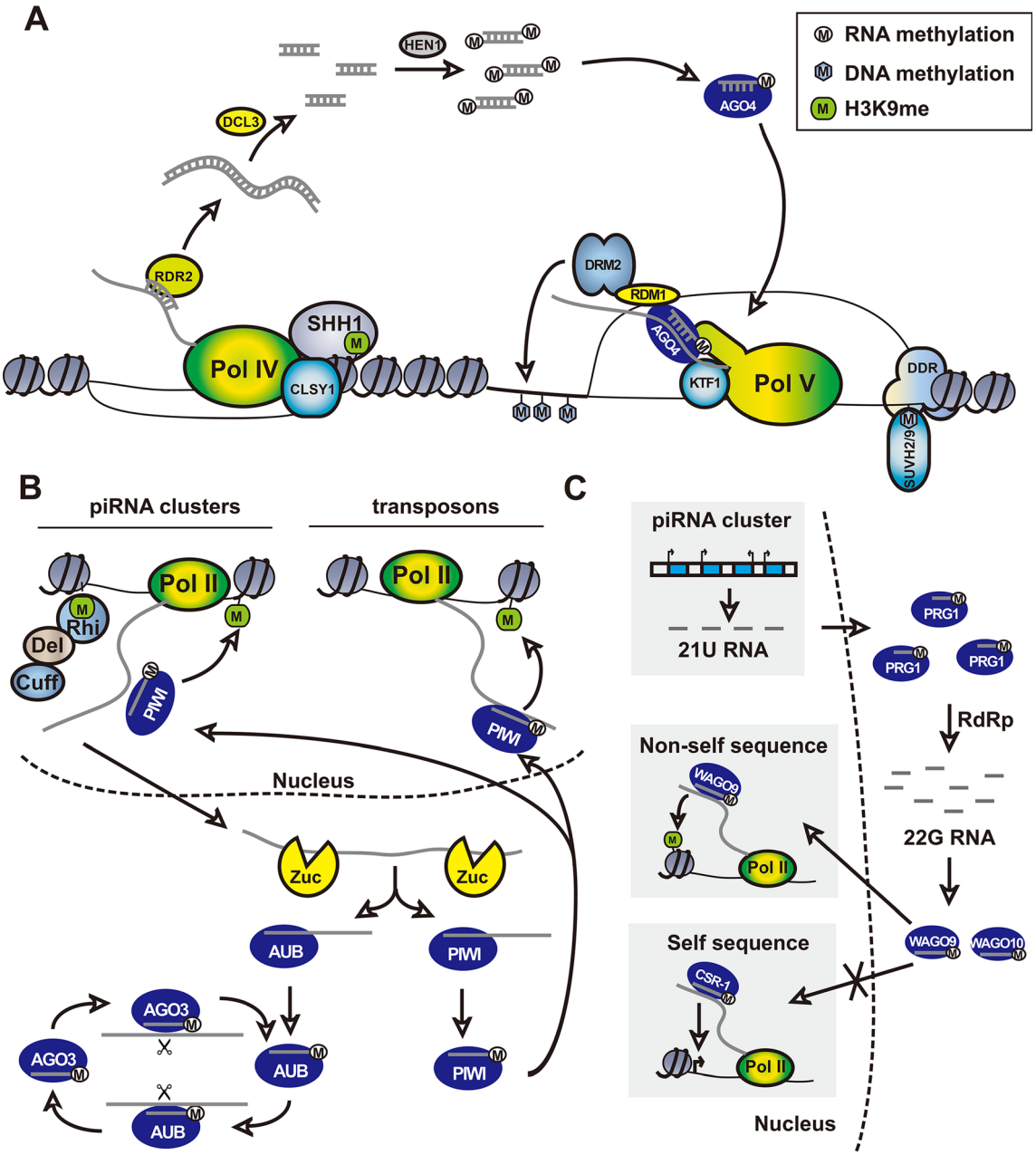

Fig. 1 Small RNA-directed transcriptional regulation in three model organisms. **a** RNA-directed DNA methylation in *Arabidopsis thaliana*. RNA Pol IV produces siRNA precursor transcripts which are further converted into double-stranded RNAs, processed into 24-nt siRNAs and loaded into AGO4. The AGO4-siRNA complexes then target to the Pol V-transcribed non-coding RNA through sequence complementarity and guide the de novo DNA methyltransferase, DRM2, to methylate DNA. **b** piRNA-mediated transcriptional silencing in *Drosophila melanogaster*. Long, polycistronic piRNA precursors are transcribed by Pol II, processed into 24–27-nt piRNAs and loaded into PIWI or AUB proteins. PIWI then translocate into nuclei to target the complementary nascent transcripts, leading to deposition of H3K9me2/3 repressive marks on surrounding chromatin regions. **c** Small RNA-mediated transcriptional regulation in *Caenorhabditis elegans*. piRNAs are transcribed, trimmed to 21 nt in length and loaded into PRG1 proteins. PRG1 in turn recruits RNA-dependent RNA polymerase that makes secondary 22-nt sRNAs (22G RNA) which are loaded into WAGO9 or WAGO10. WAGO9-22G RNA complex then translocates into the nucleus, scanning the whole transcriptome to find the complementary transcripts for H3K9me3 deposition on the chromatin. To protect host genes from silencing, CSR-1 binds to another group of 22G RNAs which are complementary to host genes. CSR-1 targeting of host genes increase their transcription and protect them from PRG-1-mediated silencing

the progeny and might provide an epigenetic signal for inheritance of certain traits in the next generation.

Besides small RNA, studies of long non-coding RNAs (lncRNAs) revealed that some lncRNAs play roles in

transcriptional regulation of gene expression. In this review, we outline the biogenesis and mechanisms of transcriptional regulation by different classes of non-coding RNAs.

RNA-Directed DNA Methylation in *Arabidopsis*

In plants, the process of RNA-directed DNA methylation (RdDM) requires two atypical DNA-dependent RNA polymerases-Pol IV and Pol V (review in [3, 4•]). Pol IV is responsible for making siRNA precursor transcripts, which are converted into double-stranded RNAs (dsRNAs) by RNA-dependent RNA polymerase 2 (RDR2). RDR2 physically interacts with Pol IV and is nonfunctional without Pol IV, suggesting that the conversion of nascent transcripts to dsRNAs is channeled [5, 6]. The dsRNAs are then cleaved into 24-nt siRNAs by Dicer-like 3 (DCL3), followed by methylation of their 3′ ends by HEN1 methyl transferase. Thereafter, one strand of 24-nt siRNAs is loaded into Argonaute 4 (AGO4) in the cytoplasm and the AGO4/siRNA complex is transported into the nucleus [7]. While Pol IV transcribes substrates for siRNA generation, another polymerase, Pol V, transcribes RNA that is targeted by the AGO4/siRNA complex. AGO4 interacts with Pol V and KTF1 protein that binds nascent transcripts generated by Pol V. The AGO4/Pol V/KTF1 complex was proposed to guide the de novo DNA methyltransferase DRM2 to chromatin at targeted loci promoting cytosine methylation at sites containing CG, CHG, and CHH motifs. Interactions between AGO4 and DRM2 were demonstrated in vivo, supporting this hypothesis (Fig. 1a) [8••].

The mechanism of DRM2-mediated DNA methylation is not yet fully understood. A recent study reported the crystal structure of DRM2 and showed that it forms a homodimer critical for catalytic activity. Furthermore, DRM2 preferentially methylates one DNA strand, likely the strand acting as the template for RNA polymerase V-mediated transcription. Taken together, tethering of the AGO4/siRNA complex to Pol V recruits DRM2 to methylate the Pol V-associated strand of DNA. DRM2 methylation seems to prefer double-stranded DNA, so how strand information for DRM2 target selection is delivered remains elusive [8••]. A multifunctional protein RDM1 may also play an important role in the recruitment of DRM2, since it is the only known protein that interacts with both AGO4 and DRM2, in addition to binding methylated single-stranded DNA [9].

How Pol IV and Pol V are specified to transcription of certain genomic regions is not fully understood. Analysis of Pol IV occupancy on DNA by ChIP-seq indicated that it is enriched at pericentromeric heterochromatin, which are regions enriched in transposons and other repetitive elements [10••]. Pol IV is recruited to a large subset of its genomic targets by a Pol IV-interacting protein Sawadee-Homeodomain Homolog1/DNA-Binding Transcription Factor 1 (SHH1/DTF1) [10••, 11••]. The SAWADEE domain of SHH1/DTF1 forms a unique tandem Tudor-like fold and functions as a dual lysine reader to probe both methylated H3K9 and unmethylated H3K4 residues of chromatin [10••, 11••]. The binding specificity of SAWADEE domains is consistent with SHH1/DTF1-dependent Pol IV ChIP-seq peaks, which are present in genomic regions enriched for H3K9me2 and depleted for H3K4me marks [10••]. The key residues in both lysine binding pockets of SHH1/DTF1 are required for Pol IV occupancy and activation of RdDM, suggesting that Pol IV is recruited to genomic regions through SHH1/DTF1 [10••]. SHH1/DTF1 may require chromatin-remodeling proteins like CLSY1 to change chromatin structure, facilitating the loading of Pol IV onto DNA [6, 10••, 11••].

In contrast to Pol IV, Pol V has been shown to preferentially associate with euchromatic regions, particularly at recently integrated small transposons and at promoters of genes which contain transposons or other repeats within their promoters, introns, or coding regions [12–16]. The recruitment of Pol V to target loci has recently been shown to be dependent on SUVH2 and SUVH9, both members of the Su(var)3-9 histone methyltransferase family [17••, 18••]. SUVH2 and SUVH9 showed no histone methyltransferase activity [19, 20]. Instead, they bind methylated DNA through their SRA domains, which are required for DNA methylation and transcriptional silencing functions of both SUVH2 and SUVH9 [19]. SUVH2 and SUVH9 physically interact with DMS3 and DRD1, which both belong to a chromatin-remodeling complex termed DDR (DRD1, DMS3, and RDM1) [17••, 18••]. The DDR complex was demonstrated to be essential for Pol V occupancy on chromatin and Pol V-dependent DNA methylation and siRNA accumulation [13, 14]. Therefore, SUVH2 and SUVH9 may facilitate Pol V binding to chromatin containing pre-existing DNA methylation, likely by recruiting the DDR complex and forming a positive feedback loop that further enhances DNA methylation.

An intriguing question of how new, invasive transposons are detected and silenced remained unknown because the RdDM pathway was previously investigated in steady-state systems containing long-established conditions of transcriptional silencing. However, a recent study described the sequence of events leading to de novo silencing of the Evadé (EVD) retrotransposon [21•]. At an early stage after activation of EVD (generations F_8–F_{11}), the numbers of new EVD insertions and expression level of its RNA were increased, in addition to an increase in 21–22-nt EVD siRNAs, which are processed by DCL2 and DCL4. siRNAs target EVD transcripts through PTGS; however, EVD RNA was only partly degraded, attributable to protection by retrotransposon-derived GAG protein. In generations F_{14}, genomic insertions of EVD reached a plateau of ~40 copies, coinciding with the appearance of EVD-derived 24-nt siRNA, LTR methylation and decrease of 21–22 nt siRNA. This suggested that EVD had undergone TGS at this stage and was successfully attenuated. The PTGS-to-TGS shift seems to be initiated by the appearance of 24-nt

siRNA, which may be generated by DCL3 when DCL2 and DCL4 that generate 21–22-nt siRNA are saturated. The 24-nt siRNA was loaded into AGO4 and induced de novo methylation of loci-expressing EVD transcripts, which in turn initiates antisense transcription and the spread of methylation toward the 5′-LTR [21•]. Pol IV and Pol V are believed to be responsible for antisense transcription, but a detailed mechanism is lacking.

Transcriptional Silencing by piRNA in *Drosophila melanogaster*

PIWI-interacting RNAs or piRNAs of length of 24 to 30 nt associate with PIWI clade Argonaute proteins and are expressed in germ cells and associated follicular cells in *Drosophila* ovary and testis. The population of piRNA in germ cells are highly complex as millions of distinct piRNA molecules were annotated. About 80 % of the piRNA population is mapped to annotated transposons or transposon remnants [22]. Mapping piRNAs to genomic regions revealed that they are derived from discrete genomic loci called piRNA clusters [22]. These clusters are mostly located in percentromeric and subtelomeric heterochromatic regions, which are enriched in transposable element sequences [22]. piRNA clusters are believed to be transcribed by RNA polymerase II (Pol II) as long, continuous transcripts, which have no obvious secondary structures and can span up to 200 kb in length. It is thought that piRNA biogenesis begins with the endonucleolytic cleavage in the long precursor transcript, generating shorter piRNA precursors. The cleavage is likely executed by the endonuclease (Zuc) [23–25]. After cleavage, the 5′ end of piRNA precursors is loaded into PIWI family proteins, PIWI and Aubergene (AUB). Once loaded into PIWI proteins, the piRNA precursor is trimmed at its 3′ end by an unidentified 3′-to-5′ exonuclease to mature piRNA length, followed by 2′-*O*-mehylation by Hen1 methyl transferase [26]. Loading of piRNA into AUB initiates biogenesis of secondary piRNA called ping-pong amplification, which simultaneously destroy transposon transcripts in cytoplasm and increases the production of piRNAs targeting active transposon sequences [22, 27]. Conversely, once loaded with piRNAs, PIWI translocates into nuclei to regulate chromatin structure, resulting in TGS (Fig. 1b).

The genome of *Drosophila* lacks DNA methylation, and the properties of chromatin are mainly defined by histone modifications and histone variants. It is believed that PIWI/piRNA mediates TGS by directing the deposition of repressive histone marks on targets recognized by piRNAs. Indeed, abolishment of the ability for PIWI to localize to the nucleus leads to transposon de-repression, coupled with an increase of active histone marks (H3K79me2 and H3K4me2) and a decrease of repressive marks (H3K9me2/3) over several transposable elements [28, 29]. These observations were further expanded by three recent genome-wide studies. Knockdown of PIWI leads to transcriptional de-repression of a significant fraction of transposons accompanied by loss of H3K9me2/3 marks and increase in RNA pol II occupancy on their sequences in both cell culture and in fly ovaries [30••, 31••, 32••].

PIWI is guided to genomic targets by piRNA. This is supported by the fact that the piRNA pathway does not target most host genes and that PIWI is required to be associated with piRNAs to repress TEs [31••]. Moreover, PIWI loaded with artificial piRNAs targeting a LacZ transgene leads to efficient lacZ silencing accompanied by accumulation of repressive chromatin marks and a decrease in both active chromatin marks and RNA pol II occupancy [31••]. Although the mechanism of target recognition by the PIWI/piRNA complex is unclear, it is believed that piRNAs recognize nascent transcripts through sequence complementarity. Indeed, PIWI knockdown causes a decrease in H3K9me3 signal only on actively transcribed TEs but not on untranscribed transposon remnants, which are transcriptionally inactive [30••].

Recently, two factors were proposed to be involved in PIWI-mediated TGS. Knockdown of Maelstrom (Mael) resulted in increased Pol II occupancy over transposons similar to changes seen in PIWI knockdown. Interestingly, in contrast to PIWI knockdown that resulted in strong H3K9me3 depletion on piRNA target loci, Mael knockdown does not significantly change H3K9me3 mark [30••]. Another factor Asterix (CG3893) is required for transposon silencing and PIWI-mediated establishment of H3K9me3 mark over at least of a subset of TEs [33]. Overall, while the details of piRNA-mediated transcriptional silencing remain to be elucidated, it is clear that PIWI plays an essential role in transcriptional silencing of TEs and establishment of a repressive genomic environment.

Chromatin structure also plays a central role in piRNA biogenesis, in addition to defining piRNA-generating genomic loci, such as piRNA clusters. Increasing evidence suggests that piRNA clusters have a unique chromatin environment. piRNA clusters show features of heterochromatin, which is enriched with repressive H3K9me3 marks and heterochromatin protein 1 (HP1) that interacts with H3K9me3 marks. Moreover, the heterochromatic environment seems to be required for expression and proper processing of piRNA precursors. Mutation or knockdown of the histone methyltransferase SetDB1/eggless that methylates H3K9 disrupts transcription of piRNA clusters [34]. Nevertheless, deposition of H3K9me3 cannot be sufficient for distinguishing piRNA clusters, because these features are present in many genomic loci that do not produce piRNAs. Interestingly, a germline-specific HP1 homolog protein, Rhino, interacts with H3K9me3 mark through its chromodomain, associates with a majority of piRNA clusters expressed in germ cells and seems to be absent from other genomic regions [35, 36•, 37•, 38•]. Furthermore, mutation of *rhino* results in piRNA loss and transposon de-repression, indicating its crucial role in

piRNA pathway. Several recent studies further suggest that Rhino binding to H3K9me3 may serve as a platform to initiate assembly of a complex required for piRNA precursor transcription and/or processing. Indeed, Rhino forms a complex with two previously identified piRNA pathway proteins, Deadlock and Cutoff, and is required for efficient transcription of piRNA clusters [36•, 37•]. In another study, Rhino was shown to work together with Cutoff and a nuclear DEAD-box RNA helicase protein, UAP56, to inhibit splicing of piRNA precursors. It was proposed that a stalled splicing complex differentiates piRNA precursors from mRNAs and directs their processing to small RNAs [38•, 39]. In both models, Cutoff seems to play a crucial role although the molecular function of Cutoff remains unknown. Further characterization of Cutoff function should help understand more about piRNA biogenesis.

In *Drosophila*, transgenerational inheritance of piRNA seems to provide a critical trigger that defines genomic regions for piRNA production in progeny. A study by de Vanssay et al. showed that a transgenic locus that generates piRNAs can induce piRNA production from a homologous locus that was originally incompetent for piRNA generation [40]. Significantly, the recipient locus is activated in the progeny of the females that express active locus even if the active locus itself is not inherited, indicating that activation is caused by epigenetic signal transmitted to the progeny independently of the genomic locus. It was proposed that piRNAs in the maternal cytoplasm are the epigenetic signal that activates piRNA biogenesis in the next generation. This hypothesis is consistent with the fact that PIWI/piRNA complexes that are expressed during oogenesis are deposited into the developing eggs and are present in the early embryo [22, 41, 42]. Transgenerationally inherited piRNAs may act as an epigenetic signal for identification of substrates for piRNA biogenesis using two different mechanisms. First, piRNAs associated with nuclear PIWI protein install the H3K9me3 mark on homologous genomic sequences, leading to recruitment of Rhino and other biogenesis components. Second, inherited piRNAs enhance processing of homologous transcripts into mature piRNAs by ping-pong amplification in the cytoplasm [36•, 43•].

piRNA-Mediated DNA and H3K9 Methylation in Mouse

In mouse, piRNA and PIWI proteins are expressed in male germ cells and their expression is tightly regulated during spermatogenesis. Abundant populations of piRNA are expressed during the pachytene stage of meiosis and are called pachytene piRNA. The function and targets of pachytene piRNAs remain unknown; however, it is unlikely that they are involved in transcriptional repression since pachytene piRNAs associate with two PIWI proteins, MIWI and MILI, which are exclusively localized in the cytoplasm. An interesting feature of the genomic regions that encode pachytene piRNAs is that they are transcribed divergently from a central promoter in both directions [44, 45]. A recent study has identified a transcription factor, A-myb, which regulates pachytene piRNA cluster transcription through the binding of a conserved sequence motif in the promoter. A-myb also controls its own transcription and that of several key proteins in the piRNA pathway including MIWI. This suggests that A-myb orchestrates the transcription of piRNA clusters along with expression of the proteins necessary to produce piRNA from those cluster transcripts [46•]. Precursors of pachytene piRNAs are transcribed by Pol II and have features resembling canonical mRNA transcription, such as 5′-cap and 3′-poly(A) tail.

In contrast to pachytene piRNAs, piRNAs expressed in developing germ cells during embryogenesis are involved in transcriptional repression of transposable elements in the nucleus. These embryonic piRNAs associate with two PIWI proteins, MIWI2 and MILI and the loss of either of these proteins leads to transposon activation and sterility [47, 48]. In mouse male germ cells, transposon silencing is established in a narrow window during embryonic germ cell development after global genome de-methylation. This is a critical stage of development, since the failure to establish CpG DNA methylation on TEs leads to their activation and subsequent meiotic failure and sterility [49]. As MIWI2 is located in the nucleus and expressed only in a narrow developmental window exactly overlapping the time of de novo DNA methylation, it is considered as the effector of piRNA-directed DNA methylation in mice. Indeed, CpG methylation patterns were not reestablished on retrotransposon sequences in MIWI2-deficient mice [50]. As MILI is exclusively cytoplasmic, it is unlikely to be the direct effector of DNA methylation. However, MILI is necessary for piRNA loading and nuclear localization of MIWI2, explaining its role in DNA methylation.

Two recent studies expand our knowledge of piRNA-mediated transposon silencing in mice [51•, 52•]. According to these studies, there are two waves of de novo DNA methylation during epigenetic reprogramming in male germ cells. The first wave of de novo methylation acts on most genomic DNA by default; however, a fraction of retrotransposon insertions evade this first wave and remain active. The active retrotransposon transcripts then enter the ping-pong amplification pathway to produce antisense piRNAs that are loaded into MIWI2, adaptively triggering the second wave of de novo DNA methylation in a sequence-specific manner [52•]. Besides DNA methylation, MIWI2 is also required to establish repressive H3K9me3 histone marks on transposon loci indicating that the nuclear functions of piRNAs are well conserved between flies and mammals. It is noteworthy that MIWI2 only targets full-length but not truncated copies of retrotransposons for deposition of repressive chromatin marks, supporting the model that piRNA require nascent transcription to change the chromatin state of transposons [51•].

Small RNA-Mediated Transcriptional Regulation in *Caenorhabditis elegans*

Small RNA pathways in *C. elegans* are very complex due to the fact that the worm genome contains more than 20 different Argonaute proteins compared with five in flies and eight in mouse [53]. Each of these Argonaute proteins associate with a distinct set of small RNAs, but there is substantial cross-talk between distinct Argonaute/small RNA pathways. One example of such cross-talk is a complex interaction between four Argonaute proteins, PRG1, WAGO9/10 and CSR1, and distinct small RNA populations associated with each of these proteins that together lead to stable transcriptional silencing of foreign genomic elements alongside protection of host genes.

PRG1 is the homolog of fly and mammalian PIWI proteins in worms. PRG1-bound piRNAs were also termed 21U-RNA for its precise 21-nt length and enrichment for uridine residue at the 5′ terminus [54–56]. In contrast to piRNA in fly and mouse that are processed from long precursor RNAs, each 21U-RNA represents a tiny, autonomous unit. The majority of 21U-RNAs are derived from two 21U-generating clusters on chromosome IV [57]. The *C. elegans*, the 21U-RNA cassette requires only the 21U-RNA sequence itself and two upstream motifs. One of the motifs is recognized by Forkhead transcription factors to enhance transcription of 21U-RNA precursors by Pol II [58•, 59•]. The transcription of 21U-RNA precursors begins at the second motif, making a ~26-nt capped-21U-RNA precursor followed by trimming extra nucleotides at 5′ and 3′ ends to produce mature 21U-RNAs. Recent studies have identified several 21U-RNA biogenesis factors in *C. elegans*, some of which are involved in precursor synthesis, while others are involved in precursor processing [58•, 60–62].

In contrast to fly and mammalian PIWI, PRG1 does not relocate into nucleus and is not directly involved in transcriptional silencing. Instead, it recruits an RNA-dependent RNA polymerase that makes a massive number of secondary 22-nt small RNAs (22G RNAs) that are bound by the worm-specific Argonaute proteins WAGO-9 (HRDE-1) and WAGO-10. WAGO-9 enters the nucleus, where it targets RNA transcripts with homology to 22G RNA, mediating TGS along with deposition of the H3K9me3 marks (Fig. 1c) [63•, 64•, 65•]. Once TGS is established, silencing of targets can be maintained for many generations independently of 21U-RNA or PRG-1 expression [63•, 65•, 66•].

Considering that there are more than 30,000 distinct 21U-RNA sequences, in addition to the fact that only partial complementarity is needed to identify targets, almost any foreign sequence can be targeted by PRG1/21U-RNA [67]. On the other hand, this also means that host genes can be potentially repressed. *C. elegans* has evolved a unique mechanism to protect its own DNA from WAGO-9 targeting: CSR-1 associates with RdRP-derived small RNAs (22G RNA) that are antisense to most (if not all) germline-expressed mRNAs [68, 69]. Strikingly, targeting of CSR-1 to host genes leads to their enhanced transcription and protection from PRG-1-mediated silencing [64•, 67, 70•, 71•, 72•]. Therefore, while fly and mouse use small RNAs exclusively for repression of foreign elements, the worm small RNAs scans throughout the genome, tags self-genes as "good" and represses the unlabeled "bad" invader genes (Fig. 1c). Since CSR-1-associated 22G-RNAs are transmitted into the progeny [69], the small RNA system in worm appears to be a molecular memory for both foreign and host genes.

Long Non-coding RNA in Transcriptional Regulation

To date, more than 10,000 intergenic long non-coding RNAs (lncRNAs) have been annotated in mammals, and a rapidly growing number of lncRNAs have been implicated in a variety of biological processes [73]. Many lncRNAs localize exclusively or partially within the nucleus. There are various genomic origins of lncRNAs, which can be roughly classified into five categories: (1) stand-alone lncRNAs that are intergenic, (2) antisense transcripts that occurs at the opposite DNA strand orientation to a protein coding gene, (3) pseudogenes which have lost their coding ability due to mutations, (4) long intronic ncRNAs, and (5) divergent transcripts transcribed from promoters or enhancers [74].

A group of enhancer RNAs (eRNAs) were demonstrated to modulate expression of neighboring protein coding genes [75, 76•, 77–79]. Two studies have reported that eRNAs associated with their site of synthesis and induced looping of the local chromosomes to bring the enhancers near to the promoters of nearby target genes [76•, 80•]. Interestingly, eRNAs are not limited to the adjacent genes but also a long-range DNA looping interaction up to 27 Mb in length has also been documented [76•]. Bridging of enhancer and promoter loci by eRNAs were thought to act *in-cis* on the same chromosome, however lncRNA can also modulate gene expression *in trans* [81–83]. In one of these cases, lncRNA was shown to inhibit DNA methylation of an enhancer to control expression of adjacent genes on the sister chromosome [82]. Moreover, determination of the genomic binding profiles of several lncRNAs has revealed that a single type of lncRNA can associate with up to thousands of genomic loci distributed on different chromosomes to control large-scale gene expression [84–89]. The mechanisms of genome targeting by lncRNAs are currently not well understood, but the spatial organization of the chromosomes in the nucleus may play a role. A process called "proximity transfer" is easily speculated in the cases of local gene expression regulation by nearby eRNAs. Recent reports further suggest that the three-dimensional conformation of genomic DNA may guide lncRNAs to distal but spatially localized target sites [88, 90, 91]. However, spatial

organization should not be the only factor to identify correct targets, since several nuclear lncRNAs can regulate transcription when expressed from ectopic loci [81, 83, 87, 92]. Thus, some lncRNAs may form complexes with proteins to bind specific regions for gene regulation. For example, NeST lncRNA interacts with WD repeat domain 5 (WDR5) and activate gene transcription of interferon gamma by altering the active histone mark H3K4me3 [92].

The molecular mechanism that allows targeting of lncRNAs to specific genomic loci are largely unknown, but at least three possible mechanisms were proposed: (1) forming RNA-DNA-DNA triplex, (2) RNA-RNA base-pairing, and (3) indirect recruitment through forming an RNA-protein-DNA complex [93•]. The results of lncRNA-targeting to genomic loci are only now beginning to be investigated. One of the possible outcomes of lncRNA function is the recruitment of protein factors that regulate the chromatin states to targeted loci. In addition, lncRNAs can directly interact with transcriptional machinery. It may act as a decoy or co-regulator for transcription factors or as an inhibitor of Pol II [74].

Conclusion

In the 1960s, the central dogma of molecular biology stated that RNA is a simple messenger between DNA and protein. Studies over the past several decades have revealed the presence of large numbers of non-coding RNAs that regulate diverse biological processes. The two major classes of regulatory RNAs described here, small RNAs and lncRNAs, show both similar and distinct features. They both need to collaborate with effector proteins to regulate gene expression. Similar to small RNAs, it was proposed that some of lncRNAs target genomic regions through sequence homology and regulate gene expression by altering histone marks. The functions of the vast majority of lncRNAs are currently unknown, and future studies should reveal the molecular mechanisms of their action.

Acknowledgments We thank members of the Aravin lab and particularly Alexandre Webster for discussion and comments. The work in Aravin laboratory is supported by grants from the National Institutes of Health (R01 GM097363 and DP2 OD007371A) and by the Damon Runyon-Rachleff, the Searle Scholar, and the Packard Fellowship Awards to A.A.A.

Compliance with Ethics Guidelines

Conflict of Interest Yung-Chia Ariel Chen and Alexei A. Aravin declare that they have no conflict of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:
• Of importance
•• Of major importance

1. Fire A, et al. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature. 1998;391(6669):806–11.
2. Castel SE, Martienssen RA. RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. Nat Rev Genet. 2013;14(2):100–12.
3. He XJ, Ma ZY, Liu ZW. Non-coding RNA transcription and RNA-directed DNA methylation in *Arabidopsis*. Mol Plant. 2014;7(9):1406–14.
4.• Matzke MA, Mosher RA. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat Rev Genet. 2014;15(6):394–408. *A nice review which summarize current models of mechnisms of lncRNA mediated transcriptional regulation*.
5. Haag JR, et al. In vitro transcription activities of Pol IV, Pol V, and RDR2 reveal coupling of Pol IV and RDR2 for dsRNA synthesis in plant RNA silencing. Mol Cell. 2012;48(5):811–8.
6. Law JA, et al. SHH1, a homeodomain protein required for DNA methylation, as well as RDR2, RDM4, and chromatin remodeling factors, associate with RNA polymerase IV. PLoS Genet. 2011;7(7):e1002195.
7. Ye R, et al. Cytoplasmic assembly and selective nuclear import of *Arabidopsis* Argonaute4/siRNA complexes. Mol Cell. 2012;46(6):859–70.
8.•• Zhong X, et al. Molecular mechanism of action of plant DRM de novo DNA methyltransferases. Cell. 2014;157(5):1050–60. *This study reported a crystal structure of DRM and provides a model of DRM2 recruitment to nascent RNA transcripts by Ago4-siRNA*.
9. Gao Z, et al. An RNA polymerase II- and AGO4-associated protein acts in RNA-directed DNA methylation. Nature. 2010;465(7294):106–9.
10.•• Law JA, et al. Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. Nature. 2013;498(7454):385–9. *Together with (11), these papers show SHH1/DTF1 binds to H3K9me and suggest such binding is required for Pol IV recruitment*.
11.•• Zhang H, et al. DTF1 is a core component of RNA-directed DNA methylation and may assist in the recruitment of Pol IV. Proc Natl Acad Sci U S A. 2013;110(20):8290–5. *Together with (10), these papers show SHH1/DTF1 binds to H3K9me and suggest such binding is required for Pol IV recruitment*.
12. Lee TF, et al. RNA polymerase V-dependent small RNAs in *Arabidopsis* originate from small, intergenic loci including most SINE repeats. Epigenetics. 2012;7(7):781–95.
13. Zhong X, et al. DDR complex facilitates global association of RNA polymerase V to promoters and evolutionarily young transposons. Nat Struct Mol Biol. 2012;19(9):870–5.
14. Wierzbicki AT, et al. Spatial and functional relationships among Pol V-associated loci, Pol IV-dependent siRNAs, and cytosine methylation in the *Arabidopsis* epigenome. Genes Dev. 2012;26(16):1825–36.
15. Zemach A, et al. The *Arabidopsis* nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell. 2013;153(1):193–205.
16. Zheng Q, et al. RNA polymerase V targets transcriptional silencing components to promoters of protein-coding genes. Plant J, 2012.

17.•• Liu ZW, et al. The SET domain proteins SUVH2 and SUVH9 are required for Pol V occupancy at RNA-directed DNA methylation loci. PLoS Genet. 2014;10(1):e1003948. *This study together with (18) show that SUVH2 and SUVH9 bind to methylated DNA to recruit Pol V.*

18.•• Johnson LM, et al. SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. Nature. 2014;507(7490):124–8. *This study together with (17) show that SUVH2 and SUVH9 bind to methylated DNA to recruit Pol V.*

19. Johnson LM, et al. SRA-domain proteins required for DRM2-mediated de novo DNA methylation. PLoS Genet. 2008;4(11): e1000280.

20. Kuhlmann M, Mette MF. Developmentally non-redundant SET domain proteins SUVH2 and SUVH9 are required for transcriptional gene silencing in *Arabidopsis thaliana*. Plant Mol Biol. 2012;79(6): 623–33.

21.• Mari-Ordonez A, et al. Reconstructing de novo silencing of an active plant retrotransposon. Nat Genet. 2013;45(9):1029–39. *This study provides a model of the sequence of events leading to de novo silencing of transposon in plant.*

22. Brennecke J, et al. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. Cell. 2007;128(6): 1089–103.

23. Nishimasu H, et al. Structure and function of Zucchini endoribonuclease in piRNA biogenesis. Nature. 2012;491(7423):284–7.

24. Voigt F, et al. Crystal structure of the primary piRNA biogenesis factor Zucchini reveals similarity to the bacterial PLD endonuclease Nuc. RNA. 2012;18(12):2128–34.

25. Ipsaro JJ, et al. The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis. Nature. 2012;491(7423): 279–83.

26. Kawaoka S, et al. 3′ end formation of PIWI-interacting RNAs in vitro. Mol Cell. 2011;43(6):1015–22.

27. Gunawardane LS, et al. A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in *Drosophila*. Science. 2007;315(5818):1587–90.

28. Saito K, et al. A regulatory circuit for piwi by the large Maf gene traffic jam in *Drosophila*. Nature. 2009;461(7268):1296–9.

29. Klenov MS, et al. Repeat-associated siRNAs cause chromatin silencing of retrotransposons in the *Drosophila melanogaster* germline. Nucleic Acids Res. 2007;35(16):5430–8.

30.•• Sienski G, Dönertas D, Brennecke J. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. Cell. 2012;151(5):964–80. *Together with (31) and (32), this study provides genome-wide evidence of Piwi-mediated transcriptional silencing.*

31.•• Le Thomas A, et al. Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. Genes Dev. 2013;27(4):390–9. *Together with (30) and (32), this study provides genome-wide evidence of Piwi-mediated transcriptional silencing.*

32.•• Rozhkov NV, Hammell M, Hannon GJ. Multiple roles for Piwi in silencing *Drosophila* transposons. Genes Dev. 2013;27(4):400–12. *Together with (30) and (31), this study provides genome-wide evidence of Piwi-mediated transcriptional silencing.*

33. Muerdter F, et al. A Genome-wide RNAi Screen Draws a Genetic Framework for Transposon Control and Primary piRNA Biogenesis in *Drosophila*. Mol Cell 2013;50(5):736–48.

34. Rangan P, et al. PiRNA production requires heterochromatin formation in *Drosophila*. Curr Biol. 2011;21(16):1373–9.

35. Klattenhoff C, et al. The *Drosophila* HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. Cell. 2009;138(6):1137–49.

36.• Le Thomas A. et al. Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. Genes Dev. 2014;28(15):1667–80. *This paper together with (43) show that the maternal deposition of piRNAs defines the piRNA clusters in progeny by deposition of H3K9me3 mark which recruits Rhino protein.*

37.• Mohn F, et al. The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in *Drosophila*. Cell. 2014;157(6):1364–79. *This paper shows that Rhino forms complex with Deadlock and Cutoff to define bi-directional piRNA cluster and suggests the function of the complex in noncanonical transcription.*

38.• Zhang Z, et al. The HP1 homolog rhino anchors a nuclear complex that suppresses piRNA precursor splicing. Cell. 2014;157(6):1353–63. *This paper proposes a model that Rhino and Cutoff inhibit splicing of piRNA precursors and thus direct them to piRNA processing pathway.*

39. Zhang F, et al. UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. Cell. 2012;151(4):871–84.

40. de Vanssay A, et al. Paramutation in *Drosophila* linked to emergence of a piRNA-producing locus. Nature. 2012;490(7418):112–5.

41. Harris AN, Macdonald PM. Aubergine encodes a *Drosophila* polar granule component required for pole cell formation and related to eIF2C. Development. 2001;128(14):2823–32.

42. Megosh HB, et al. The role of PIWI and the miRNA machinery in *Drosophila* germline determination. Curr Biol. 2006;16(19): 1884–94.

43.• Le Thomas A, Marinov GK, Aravin AA. A transgenerational process defines piRNA biogenesis in *Drosophila virilis*. Cell Rep. 2014;8(6):1617–23. *This paper together with (43) show that the maternal deposition of piRNAs defines the piRNA clusters in progeny by deposition of H3K9me3 mark.*

44. Aravin A, et al. A novel class of small RNAs bind to MILI protein in mouse testes. Nature. 2006;442(7099):203–7.

45. Girard A, et al. A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature. 2006;442(7099):199–202.

46.• Li XZ, et al. An ancient transcription factor initiates the burst of piRNA production during early meiosis in mouse testes. Mol Cell. 2013;50(1):67–81. *This study identified a transcription factor, A-myb, which regulates mouse pachytene piRNA cluster transcription.*

47. Aravin AA, et al. Developmentally regulated piRNA clusters implicate MILI in transposon control. Science. 2007;316(5825):744–7.

48. Carmell MA, et al. MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. Dev Cell. 2007;12(4):503–14.

49. Bourc'his D, Bestor TH. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature. 2004;431(7004):96–9.

50. Kuramochi-Miyagawa S, et al. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. Genes Dev. 2008;22(7):908–17.

51.• Pezic D, et al. piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells. Genes Dev. 2014;28(13):1410–28. *This study finds piRNA-mediated H3K9me3 deposition on transposons in mice.*

52.• Molaro A, et al. Two waves of de novo methylation during mouse germ cell development. Genes Dev. 2014;28(14):1544–9. *This study describes two waves of de novo methylation during epigenetic reprogramming in germ cells: first wave by default and second one mediated by MIWI2-piRNA.*

53. Carmell MA, et al. The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. Genes Dev. 2002;16(21):2733–42.

54. Batista PJ et al. PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in *C. elegans*. Mol Cell. 2008;31(1): 67–78.

55. Das PP, et al. Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the *Caenorhabditis elegans* germline. Mol Cell. 2008;31(1):79–90.
56. Wang G, Reinke V. A *C. elegans* Piwi, PRG-1, regulates 21U-RNAs during spermatogenesis. Curr Biol. 2008;18(12):861–7.
57. Ruby JG, et al. Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. Cell. 2006;127(6):1193–207.
58.• Cecere G, et al. Promoters recognized by forkhead proteins exist for individual 21U-RNAs. Mol Cell. 2012;47(5):734–45. *This study identifies the feature and a transcription factor of C. elegans piRNA genes.*
59.• Gu W, et al. CapSeq and CIP-TAP identify Pol II start sites and reveal capped small RNAs as *C. elegans* piRNA precursors. Cell. 2012;151(7):1488–500. *By developing Cap-seq technique, this study find the C. elegans piRNA precursors are 26 nt capped small RNAs.*
60. Weick EM, et al. PRDE-1 is a nuclear factor essential for the biogenesis of Ruby motif-dependent piRNAs in *C. elegans*. Genes Dev. 2014;28(7):783–96.
61. de Albuquerque BF, et al. PID-1 is a novel factor that operates during 21U-RNA biogenesis in *Caenorhabditis elegans*. Genes Dev. 2014;28(7):683–8.
62. Goh WS, et al. A genome-wide RNAi screen identifies factors required for distinct stages of *C. elegans* piRNA biogenesis. Genes Dev. 2014;28(7):797–807.
63.• Ashe A, et al. piRNAs can trigger a multigenerational epigenetic memory in the germline of *C. elegans*. Cell. 2012;150(1):88–99. *This study and others (63-66) demonstrate that in C. elegans, the piRNA can trigger TGS which becomes an epigenetic memory lasting for many generations.*
64.• Shirayama M, et al. piRNAs initiate an epigenetic memory of nonself RNA in the *C. elegans* germline. Cell. 2012;150(1):65–77. *This study and others (63-66) demonstrate that in C. elegans, the piRNA can trigger TGS which becomes an epigenetic memory lasting for many generations.*
65.• Buckley BA, et al. A nuclear Argonaute promotes multigenerational epigenetic inheritance and germline immortality. Nature. 2012;489(7416):447–51. *This study and others (63-66) demonstrate that in C. elegans, the piRNA can trigger TGS which becomes an epigenetic memory lasting for many generations.*
66.• Luteijn MJ, et al. Extremely stable Piwi-induced gene silencing in *Caenorhabditis elegans*. EMBO J. 2012;31(16):3422–30. *This study and others (63-66) demonstrate that in C. elegans, the piRNA can trigger TGS which becomes an epigenetic memory lasting for many generations.*
67. Lee HC, et al. *C. elegans* piRNAs mediate the genome-wide surveillance of germline transcripts. Cell. 2012;150(1):78–87.
68. Gu W, et al. Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the *C. elegans* germline. Mol Cell. 2009;36(2):231–44.
69. Claycomb JM, et al. The Argonaute CSR-1 and its 22G-RNA cofactors are required for holocentric chromosome segregation. Cell. 2009;139(1):123–34.
70.• Conine CC, et al. Argonautes promote male fertility and provide a paternal memory of germline gene expression in *C. elegans*. Cell. 2013;155(7):1532–44. *This study along with (71-72) show that targeting of CSR-1 to host genes enhance their transcription and protect them from PRG-1 mediated silencing.*
71.• Seth M, et al. The *C. elegans* CSR-1 argonaute pathway counteracts epigenetic silencing to promote germline gene expression. Dev Cell. 2013;27(6):656–63. *This study and (70,72) show that targeting of CSR-1 to host genes enhance their transcription and protect them from PRG-1 mediated silencing.*
72.• Wedeles CJ, Wu MZ, Claycomb JM. Protection of germline gene expression by the *C. elegans* Argonaute CSR-1. Dev Cell. 2013;27(6):664–71. *This study and (70-71) show that targeting of CSR-1 to host genes enhance their transcription and protect them from PRG-1 mediated silencing.*
73. Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. Cell. 2013;154(1):26–46.
74. Kung JT, Colognori D, Lee JT. Long noncoding RNAs: past, present, and future. Genetics. 2013;193(3):651–69.
75. Lam MT, et al. Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. Nature. 2013;498(7455):511–5.
76.• Li W, et al. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature. 2013;498(7455):516–20. *This study shows that ER-alpha-induced eRNAs are likely to increase the enhancer-promoter looping to regulate the neighboring genes.*
77. Melo CA, et al. eRNAs are required for p53-dependent enhancer activity and gene transcription. Mol Cell. 2013;49(3):524–35.
78. Mousavi K, et al. eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. Mol Cell. 2013;51(5):606–17.
79. Hah N, et al. Enhancer transcripts mark active estrogen receptor binding sites. Genome Res. 2013;23(8):1210–23.
80.• Lai F, et al. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature. 2013;494(7438):497–501. *This paper shows the presence of DNA looping between the ncRNA loci and its neighboring targets.*
81. Tian D, Sun S, Lee JT. The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. Cell. 2010;143(3):390–403.
82. Berghoff EG, et al. Evf2 (Dlx6as) lncRNA regulates ultraconserved enhancer methylation and the differential transcriptional control of adjacent genes. Development. 2013;140(21):4407–16.
83. Feng J, et al. The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. Genes Dev. 2006;20(11):1470–84.
84. Chu C, et al. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol Cell. 2011;44(4):667–78.
85. Rinn JL, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell. 2007;129(7):1311–23.
86. Yang L, et al. lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. Nature. 2013;500(7464):598–602.
87. Vance KW, et al. The long non-coding RNA Paupar regulates the expression of both local and distal genes. EMBO J. 2014;33(4):296–311.
88. Hacisuleyman E, et al. Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. Nat Struct Mol Biol. 2014;21(2):198–206.
89. Takayama K, et al. Androgen-responsive long noncoding RNA CTBP1-AS promotes prostate cancer. EMBO J. 2013;32(12):1665–80.
90. Engreitz JM, et al. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. Science. 2013;341(6147):1237973.
91. Wang KC, et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature. 2011;472(7341):120–4.
92. Gomez JA, et al. The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon-gamma locus. Cell. 2013;152(4):743–54.
93.• Vance KW, Ponting CP. Transcriptional regulatory functions of nuclear long noncoding RNAs. Trends Genet. 2014;30(8):348–55. *A nice review which summarize current models of mechanisms of lncRNA mediated transcriptional regulation.*