CHAPTER 2

DNA, RNA, and Protein

SUMMARY

The central dogma of molecular biology states that genetic information flows from DNA to RNA and then to protein. The process of converting DNA to RNA is called transcription. A separate process, called translation, converts the information in RNA to the specific sequence of amino acids in proteins.

To understand transcription and translation, one needs to have a basic understanding of the typical gene structure. A typical gene contains three main regions: a promoter, a 5′ untranslated region (5′ UTR), and an open reading frame (ORF). The 5′ UTR contains sequences and elements needed during translation but is not actually translated into protein.

During transcription, a copy of DNA is made in the form of RNA. The major contributing enzyme is RNA polymerase. Transcription begins when RNA polymerase recognizes a specific region on the DNA called the promoter. Upon binding the promoter, transcription begins at the transcriptional start site where RNA polymerase inserts the first RNA nucleotide. Using one of the strands of the DNA double helix as a template, RNA polymerase transcribes a single-stranded RNA molecule that is complementary to the DNA template strand, with the exception that thymine is replaced with uracil in RNA. The copying of DNA into an RNA intermediary continues until RNA polymerase reaches a stop signal. The termination signal is often a hairpin loop structure that may or may not be flanked by a stretch of adenine nucleotides, which cause the DNA/RNA hybrid to become unstable. When polyadenylation is not present, a special protein called Rho unwinds the DNA/RNA hybrid to separate the two molecules.

Eukaryotic and prokaryotic genomes are arranged quite differently. Naturally, differences exist in both the mechanism and regulation of transcription between the two groups. Prokaryotic genes, particularly those for a single metabolic pathway, are often arranged next to each other and transcribed together as one unit from the same promoter. This arrangement is called an operon. In contrast, eukaryotes do not typically have operons. Eukaryotic transcription is far more complex than in prokaryotes. Prokaryotes typically have one type of RNA polymerase to transcribe all genes, but eukaryotes have three different types of RNA polymerase to transcribe different types of genes. The promoter regions of eukaryotic genes are also more complex and contain more elements, such as the initiator box, TATA box, and other regions that bind to transcription factors.

Several elements exist to ensure that RNA polymerase recognizes the correct gene to transcribe. Some genes are constitutive, meaning that they are constantly expressed and are used in the housekeeping of the cell. Other genes are expressed only at certain times in the cell cycle, under specific environmental conditions, or in the presence of some nutrients. A significant amount of control exists for transcription of many genes so that the cell is not wasting valuable energy and resources on transcribing genes that may not be necessary in a given moment. This is true for both prokaryotes and eukaryotes.

In prokaryotes, activator proteins or repressor proteins are involved in regulation of transcription. Activator proteins bind to the DNA and help RNA polymerase find the promoter. Repressor proteins are already bound to DNA and block RNA polymerase. The gene is expressed only when the repressor is removed. Additionally, a component of RNA polymerase, called a sigma factor, enables the enzyme to bind to different promoters. There are numerous sigma factors, and each one recognizes promoters for sets of genes involved in specific responses of the cell, such as housekeeping, high temperature, and the stationary phase. The classic example of regulation in prokaryotes is the lactose operon.

E. coli prefers to use glucose as a carbon source, even in the presence of other usable sources such as lactose. The genes in the *lac* operon are needed for the uptake and utilization of lactose. These genes include *lacI*, which encodes a repressor, and the structural genes *lacZ* (β-galactosidase), *lacY* (permease), and *lacA* (acetylase), which are transcribed as a polycistron.

Since it is wasteful for the cell to express genes that are not currently needed, the *lac* operon is repressed by LacI, which binds to the promoter of *lacZYA* and prevents transcription. When lactose becomes available to the cell and all the glucose has been used, a small metabolite called *allo*-lactose binds to LacI and induces a shape change, which causes the repressor to dissociate from the promoter. This allows RNA polymerase to transcribe *lacZYA*. Additionally, levels of cyclic adenosine monophosphate (cAMP) increase when glucose is not present. Cyclic AMP binds to the cAMP receptor protein (CRP) and then to DNA to activate transcription of *lacZYA*. The cAMP-CRP complex is a global regulator of transcription and can modulate gene expression of many genes within the organism.

Transcription factors and epigenetic changes both contribute to the complexity that surrounds regulation of transcription in eukaryotes and even in some prokaryotes. Some transcription factors are general, which means they can initiate transcription at all promoters. Others are specific for certain genes or environmental conditions. Regardless, all are assemblies of proteins that both bind DNA and initiate RNA polymerase activity. In addition, epigenetic changes can alter the expression of genes. These changes include histone modification, DNA methylation, nucleosome remodeling, and RNA-associated silencing. DNA is wrapped around histone proteins within the membrane-bound nucleus to form nucleosomes, which may be tightly or loosely packed. The degree of chromosomal condensation plays a direct role in the expression of genes from those regions. Cells regulate the density of the chromosomes by acetylation. Acetylation loosens the chromosome structure. Also, methylating portions of the eukaryotic DNA can prevent gene expression from occurring in these regions, which is called silencing. During nucleosome remodeling, a variety of events occur at the histones. These events include sliding, removing, remodeling, and alterations to spacing. Histones are moved or remodeled to allow access to DNA. Some histones are replaced to mark specific areas of active gene expression. Lastly, noncoding RNAs can regulate epigenetic modifications, such as X-inactivation in females. Review the case study for more information on the role of RNAs in gene expression.

In prokaryotes, transcription and translation are coupled. The reason is that prokaryotes, by definition, do not contain membrane-bound organelles, including a nucleus. Once a gene is transcribed, it can be immediately and simultaneously translated into protein. Also, every portion of the prokaryotic mRNA codes for protein. The two processes in eukaryotes occur in different parts of the cell and, consequently, cannot be coupled. Eukaryotic genes are first transcribed into a primary transcript within the nucleus. The primary transcript is processed into mRNA prior to leaving the nucleus to prepare it for translation. Three major modifications of the transcript occur: addition of a 5′ cap, addition of a 3′ poly (A) tail, and excision of the noncoding regions called introns. Only the exons code for protein.

Three types of RNA are involved in translation. Messenger RNA (mRNA) is the only type that codes for protein. Transfer RNA (tRNA) carries amino acids to the ribosome, which functions as the protein factory. Ribosomal RNA (rRNA) is a structural component, along with various proteins, of the ribosome itself. The genetic code is read in triplets of nucleotides on mRNA, called codons. Each set of three bases is recognized by the anticodon region of a tRNA and represents a single amino acid. The genetic code is redundant, which means that there are more codons than amino acids. The code is also universal, although there are a few exceptions.

A ribosome consists of both a small and large subunit and contains three binding sites for tRNA molecules. The A-site accepts incoming charged tRNAs (tRNAs with amino acids). The P-site holds the tRNA bound to the growing polypeptide chain. Finally, uncharged tRNA molecules (tRNAs without amino acids) exit the ribosome at the E-site. Initiation of translation in prokaryotes begins when the small subunit of the ribosome recognizes and binds to the Shine–Dalgarno sequence, also called the ribosomal binding site, on the mRNA. The first codon to be translated is almost always AUG, which codes for a special methionine called N-formyl-methionine (fMet). The initiator tRNA carries fMet to the small ribosomal subunit,

forming the initiation complex. Other protein complexes called initiation factors help with the assembly of the initiation complex. The anticodon loop of incoming charged tRNAs recognizes and binds to the next codon within the ribosome's A-site. The ribosome catalyzes the peptide bond formation between the first and second amino acids. This activity, called peptidyl transferase, is carried out by the rRNA structural component of the ribosome and not the protein portion. Elongation factors are separate proteins that supervise the entry of charged tRNAs, the translocation of the ribosome to the next codon, and the exit of uncharged amino acids from the ribosome. Translation continues until a stop codon (UGA, UAG, UAA) is encountered by the ribosome. There are no tRNAs that recognize a stop codon. Release factors bind to the A-site of the ribosome and cause it to release the polypeptide chain. The ribosome also releases the mRNA and the two ribosomal subunits dissociate.

Slight differences exist between translation in prokaryotes and eukaryotes. Generally, more proteins are involved in the process. Specifically, there are more initiation factors in eukaryotes than prokaryotes. Also, eukaryotic mRNA does not contain a Shine–Dalgarno. Instead, eukaryotic ribosomes recognize the 5′ cap for initiation and insert an unmodified methionine as the first amino acid of the polypeptide. Additionally, the mitochondria and chloroplasts that might be present in eukaryotes contain their own genomes and direct synthesis of their own proteins. The symbiotic theory argues that mitochondria and chloroplasts were once free-living prokaryotes (aerobic, heterotroph, and cyanobacteria, respectively) that formed a symbiotic relationship with an early eukaryotic host cell. Eventually, these symbionts lost the ability to live freely from their host but have retained some ancestral identities. Size of the organelles plus genome and ribosomal similarities suggest a common ancestry with bacteria.

Case Study Dogma Derailed: The Many Influences of RNA on the Genome

Leah R. Sabin et al. (2013). *Molecular Cell* 49, 783–794.

Small RNAs and long noncoding RNAs (lncRNAs) are two classes of functional RNAs that are known to regulate gene expression. Small RNAs may be derived from viral sources or transposable elements. Some small RNAs work in unison with the Argonaute family of protein effectors. The mechanism of action involves binding of the small RNA to the effector protein, which then base-pairs to the complementary target nucleic acid sequence. The result may include post-transcriptional gene silencing (PTGS), in which case the target is an RNA molecule. In PTGS, the target RNA is cleaved or translationally repressed. Further evidence indicates that the target might also be complementary DNAs and that the small RNA bound to the effector protein actually targets the DNA for transcriptional gene silencing (TGS) through DNA methylation or histone modification. In most cases, TGS is designed to protect the integrity of the genome. However, in some systems small RNAs moderate the excision of DNA segments during rearrangements.

Long noncoding RNAs are defined as RNA molecules over 200 bases in length that do not code for protein. The function of lncRNAs is not well defined. Analyses of lncRNA sequences indicate they are poorly conserved. Evidence from the few lncRNAs that have been characterized indicates a role for these RNA molecules in chromatin-level gene regulation through interactions with histone modifiers, DNA methylation proteins, or transcriptional regulators.

In this review, the authors discuss the mechanism of action for several small RNAs and lncRNAs, including histone modification, DNA methylation, and DNA cleavage.

In the *Schizosaccharomyces pombe* system, the authors discuss a "self-perpetuating feedforward loop" in regards to the small RNA–mediated heterochromatin formation in the yeast system. What is meant by this?

The centromeres of the *S. pombe* yeast contain a high number of repeating sequences and mobile genetic elements. Small RNA–mediated histone modifications repress the repeating sequences and mobile genetic elements present in the centromeric region. Specifically, short-interfering RNAs (siRNAs) silence this region. These siRNAs are also derived from centromere transcripts that have been processed by an enzyme called Dcr1. Once processed, the siRNAs are loaded onto Ago1, a protein effector of the Argonaute family and, together with several other proteins, produce the RNA-induced initiation of transcriptional silencing (RITS) complex. The siRNA within the complex binds to target sequences at the centromere, and the result is an accumulation of several proteins at the targeted region and TGS. Furthermore, the RITS complex recruits the RNA-directed RNA polymerase (RDRC) complex to generate more siRNA from the same region. In this manner, the biogenesis of siRNA is a self-perpetuating feedforward loop.

In addition to the repressive effects of the *S. pombe* siRNAs, are there other small RNA systems addressed by the authors that induce repressive histone modifications?

Yes. Piwi proteins are a group of Argonaute proteins that function to control mobile genetic elements in flies, humans, and other organisms. The authors discussed piRNA-directed transcriptional silencing in flies. Piwi-interacting RNAs (piRNAs) are specific to germline cells and function to silence the repetitive regions containing mobile DNAs by introducing and maintaining repressive histone modifications.

How does the biogenesis of centromere-silencing siRNAs from the *S. pombe* system differ from the biogenesis of piRNAs in the germline *Drosophila* cells?

The siRNAs in the yeast system discussed in the paper are produced from centromere transcripts by the action of Dcr1. In contrast, piRNAs are derived from the transcription and processing of piRNA clusters, which are mostly composed of inactive transposon fragments, in a Dicer-independent manner.

What role do lncRNAs play in regulation of gene expression in the human model?

Although the roles of lncRNAs are still quite vague, a few have been characterized. They include *Xist*, which is involved in X-chromosome inactivation and imprinting. Additionally, many lncRNAs copurify with chromatin remodeling complexes, including PRC2. This indicates a role for lncRNAs in scaffolding or targeting remodeling machinery to specific loci. Of the lncRNAs known to copurify with PRC2, several function in human cancers. *HOTAIR* is an lncRNA in humans that, when overexpressed, leads to increased invasiveness and poor outcome for several cancers. The results from knockdown experiments suggest that *HOTAIR* acts as a scaffold between PRC2 and LSD1, a histone tail demethylase. In addition to *HOTAIR*, *ANRIL* is overexpressed in human leukemias and prostate cancer due to epigenetic silencing of *p15*, a tumor suppressor gene.

Two other lncRNAs, *Air* and *Kcnq1ot1*, are involved in genomic imprinting and silencing of one allele in diploid human cells. *HOTAIRM1* is involved in myeloid differentiation. *HOTTIP* and *Mistral* interact with chromatin modifiers that remodel the chromatin structure and activate transcription of genes rather then repress genes.

In terms of epigenetics, RNA-mediated modifications were first discovered in plants but occur in a wide range of organisms, including mammals. How do small RNAs and lncRNAs guide methylation patterns in DNA?

A majority of methylation in *Arabidopsis*, a plant model, occurs in transposons and repetitive elements near the centromeres. A specialized RNA polymerase produces long transcripts from these regions. A second RNA polymerase generates a complementary copy of the RNA transcript to produce a dsRNA molecule, which is then processed by a ribonuclease to generate siRNAs. These siRNAs bind to an Argonaute family effector protein that recruits methylation enzymes to the specific region. The result is methylation

(*Continued*)

Case Study Dogma Derailed: The Many Influences of RNA on the Genome—cont'd

and silencing of those regions. Additionally, intergenic noncoding transcripts produced by a recently discovered RNA polymerase act as a scaffold between the Argonaute protein and the recently discovered RNA polymerase itself.

RNA-mediated methylation also occurs in mammalian male germ cells through the action of piRNA. The piRNAs are expressed at different times during germ cell development and spermatogenesis. Failure of the piRNAs to regulate DNA methylation in the developing cells results in sterility. The exact molecular mechanism for this is not yet understood.

DNA methylation is also mediated by lncRNAs, although the processes have not yet been investigated on a global scale. The authors discuss some lncRNAs that have been published. Specifically, the *Tsix* lncRNA regulates expression of *Xist* during X chromosome inactivation by interacting with chromatin-associated complex PRC2 and recruiting it to the *Xist* locus. Also, *Tsix* recruits the methylation enzyme to the *Xist* promoter, thus silencing the *Xist* gene. Other genes that are influenced by noncoding RNAs include those that silence ribosomal RNA genes through the possible action of an RNA:DNA:DNA triplex structure, although this has just been suggested and not confirmed *in vivo*.

Which organisms extensively use RNA-directed DNA cleavage as opposed to transcriptional repression to silence regions of the genome?

Prokaryotes (bacteria and archaea), along with a few eukaryotic protists (ciliated protozoans), use DNA cleavage mediated by RNA molecules to silence gene expression.

How is DNA cleavage mediated by RNA in ciliates?

Ciliates, which have two nuclei (micronucleus and macronucleus), remove internal eliminated segment (IES) sequences through the action of scnRNAs and also through the modification of histone proteins. The scnRNAs are generated from micronucleus sequences during conjugation and are used to scan transcripts from the macronucleus for complementary sequences. Once found, the sequences are degraded. Any scnRNA that is not bound to macronucleus transcripts is used to find homologous sequences and target those homologs for elimination. Elimination includes the use of Argonaute family effector proteins and involves a mechanism similar to transposon silencing in *S. pombe* as described previously.

The CRISPR system in prokaryotes is used to help protect against viral pathogens. How are the pathogens targeted through the action of RNAs?

The CRISPR system retains remnants of the viral pathogens and exogenous DNA as "memories" to help combat future infections. These "memories" are termed CRISPR loci. CRISPR RNAs (crRNAS) are the small RNAs that mediate immunity to the viral pathogens. Small pieces of phage and foreign DNA are retained by the prokaryote and integrated into the CRISPR locus. Expression of the CRISPR loci yields long precursor RNAs that are further processed into crRNAs, which target incoming foreign DNA. Effector proteins are recruited when crRNAs bind to their foreign DNA targets. The result is DNA cleavage. Several types of CRISPR systems have been characterized and differ in the proteins used to process the long precursor RNAs into crRNAs.

From the simplest prokaryotes to more advanced organisms, RNA molecules play diverse roles within cells. Most often RNAs are implicated in post-transcriptional control of gene expression. Recently, the roles of RNAs within the cells have expanded with the discovery that epigenetic modifications, including histone modification, DNA methylation, and DNA cleavage, are guided by noncoding RNAs. These RNAs control gene expression by inducing changes within chromatin structure and methylation patterns. RNA can also target DNA for rearrangement and elimination. The consequences of RNA-mediated control of gene expression even have health impacts for humans and represent a potential angle for development of therapeutics.

Molecular Cell
Review

Dogma Derailed: The Many Influences of RNA on the Genome

Leah R. Sabin,[1,2,*] M. Joaquina Delás,[1,2] and Gregory J. Hannon[1,2,*]
[1]Watson School of Biological Sciences
[2]Howard Hughes Medical Institute
Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA
*Correspondence: lsabin@cshl.edu (L.R.S.), hannon@cshl.edu (G.J.H.)
http://dx.doi.org/10.1016/j.molcel.2013.02.010

Epigenetic control of gene expression is a critical component of transcriptional regulation. Remarkably, the deposition of epigenetic modifications is often guided by noncoding RNAs. Although noncoding RNAs have been most often implicated in posttranscriptional gene silencing, these molecules are now emerging as critical regulators of gene expression and genomic stability at the transcriptional level. Here, we review recent efforts to understand the mechanisms by which RNA controls the expression or content of DNA. We discuss the role of both small RNAs and long noncoding RNAs in directing chromatin changes through histone modifications and DNA methylation. Furthermore, we highlight the function of RNA in mediating DNA cleavage during genome rearrangements and pathogen defense. In understanding the mechanisms of RNA control over DNA, the power of RNA may one day be harnessed to impact gene expression in a therapeutic setting.

Introduction

Since each cell within an organism contains an identical copy of the genome, regulation of the output of the genome is responsible for determining cellular identity and allowing complex organisms to develop and function. On a cellular level, organisms face two main challenges: to maintain genome integrity in the face of mutagens and mobile genetic elements, and to express a specific repertoire of genes at the proper level and with the appropriate timing. Disruptions of either of these two processes can have catastrophic consequences, such as infertility or malignant transformation. Therefore, organisms have evolved elegant mechanisms to monitor the stability of the genome and fine-tune gene expression. In recent years, it has become increasingly evident that many of these regulatory systems rely on RNA to mediate their effects. This review will discuss the various classes of noncoding RNAs that exert control over DNA, focusing on those that maintain genomic stability or regulate DNA structure and organization through chromatin modifications or DNA cleavage.

The catalog of functional noncoding RNAs is continuously expanding, due in part to the development of next-generation sequencing technologies. Two important classes of functional RNAs responsible for mediating effects on DNA are small RNAs and long noncoding RNAs (lncRNAs). In general, small RNAs are generated from longer precursors, which can derive from both endogenous and exogenous sources, including acute viral infections and transposable elements (TEs). Following biogenesis, small RNAs are loaded into an Argonaute family member within a large effector protein complex. Two classes of Argonaute proteins exist in most animals: the ubiquitously expressed Argonaute (Ago) clade proteins, which are defined by their relationship to *Arabidopsis* AGO1, and members of the Piwi clade, which bear similarity to *Drosophila* Piwi and whose expression is largely restricted to the germline (Hutvágner and Simard, 2008). In many organisms, small RNAs are amplified to promote a more robust response; this amplification can occur through a variety of mechanisms.

The canonical role of small RNAs is to mediate posttranscriptional gene silencing (PTGS) of target RNA transcripts. During PTGS, base pairing between the small RNA, bound to its effector complex, and the target results in target cleavage or translational repression. However, seminal studies in plants and yeast, as well as more recent work in other systems, have established that small RNAs are also capable of directing transcriptional gene silencing (TGS), which can be achieved through DNA methylation or the deposition of repressive histone modifications. In these cases, the function of TGS is often to protect genomic integrity by maintaining a repressive heterochromatic state in repetitive regions of the genome, most notably those regions which harbor mobile genetic elements. Arguably the most extreme mechanism by which the content and expression of DNA can be controlled by small RNAs is DNA elimination. In some ciliates, small RNAs guide the excision of DNA elements, such as transposons, during genome rearrangements. Moreover, small RNAs in bacteria and archaea orchestrate the clustered regularly interspaced short palindromic repeat (CRISPR) pathway, which directs sequence-specific DNA cleavage of plasmids or invading phage. In the following sections, we will describe the mechanistic details of these small RNA-guided pathways and the recent advances in our understanding of their functions.

In contrast to small RNAs, the study of lncRNAs as a defined class of molecules is still in its relative infancy; indeed, the fact that the human genome is pervasively transcribed, yet that protein coding genes comprise only ~10% of its content, is a relatively recent revelation. Unlike small RNAs, there appear to be no unifying structural, biochemical, or functional characteristics that define a given transcript as a lncRNA; rather, the simplest definition of a lncRNA is merely an RNA transcript greater than 200 nucleotides in length with no coding potential (Ponting et al., 2009). Over the last 10 years, RNA-Seq data

Table 1. Chromatin Regulatory Machinery

Organism	H3K9 Methyltransferases	H3K27 Methyltransferases	H3K4 Methyltransferases	Chromodomain Proteins	De Novo DNMTs	Maintenance DNMTs
S. pombe	Clr4	Activity is present, no HMT identified	Set1	Swi6 Chp1 Chp2	No cytosine methylation	No cytosine methylation
Arabidopsis	KYP SUVH5 SUVH6	CLF SWN MEA	ATX1 ATX2 ATXR7	LHP1	DRM2	MET1 (CG sites) CMT3 (CHG sites)
Drosophila	Su(var)3-9 egg	E(z)	trx	HP1	No cytosine methylation	No cytosine methylation
Mammals	Suv39h1 G9a	EZH2	MLL	CBX7	DNMT3A DNMT3B	DNMT1
Tetrahymena	Ezl1p	Ezl1p	Activity is present, no HMT identified	Pdd1p Pdd3p	?	?

This table lists known orthologs of histone and DNA methyltransferases, as well as chromodomain-containing proteins responsible for binding methylated histones and nucleating protein complexes on chromatin.

and chromatin maps from a staggering number of cell types and tissues have been used in large-scale efforts to catalog thousands of novel lncRNAs in organisms from plants to humans (Hu et al., 2012). lncRNAs are often poorly conserved at the sequence level, initially leading to uncertainty as to whether they represent active entities or transcriptional noise. However, as a growing number of lncRNAs are characterized, it has been established that orthologs can be identified in other species through synteny; although these "syntelogs" bear little to no sequence similarity, their position relative to neighboring protein coding genes has been maintained through evolution (Ulitsky et al., 2011). Nevertheless, the functions of most of the identified lncRNAs remain largely uncharacterized.

One theme that has emerged during large-scale characterization efforts is that lncRNAs are commonly involved in mediating chromatin-level gene regulation through interactions with histone modifiers, the DNA methylation machinery, or transcriptional regulators. Many other lncRNAs have been identified as key regulators of essential cellular processes but are not known to be involved in chromatin regulation and structure, and we will omit these from our discussion. However, we invite readers to refer to several excellent reviews on the subject (Ponting et al., 2009; Hu et al., 2012; Rinn and Chang, 2012). In the following sections, we highlight the functions of several diverse classes of small and large noncoding RNAs according to the mechanism by which the RNAs exert their control over DNA, specifically through histone modifications, DNA methylation, and DNA cleavage.

RNA-Directed Histone Modifications

Histones are responsible for the organization and regulation of DNA structure and are subject to a variety of modifications on their N-terminal tails, such as methylation, acetylation, and ubiquitination, that dynamically influence chromatin function. Histone modifications can function in activating or repressing gene expression; a common mark associated with active chromatin is the trimethylation of histone 3 (H3) at lysine 4 (K4), or H3K4me3, which is often found at promoters of actively transcribed genes (Black et al., 2012). Conversely, marks associated with silenced heterochromatin include di- and trimethylated H3K9, as well as trimethylation of H3K27 (Black et al., 2012). Histone methyltransferase (HMT) enzymes deposit methyl marks onto histone tails, while chromodomain-containing proteins, which directly bind to methylated lysines, often work in concert with HMTs and are involved in assembling protein complexes on DNA. Table 1 lists selected orthologs of HMTs and chromodomain-containing proteins operating within the organisms discussed in this review.

Small RNA-Mediated Heterochromatin Formation in Yeast

A defining characteristic of eukaryotic centromeres is their high density of repetitive DNA elements, such as satellite repeats and transposons. Repressive histone modifications, particularly methylated H3K9 (H3K9me), are essential for controlling these mobile genetic elements within centromeric regions. The cores of fission yeast centromeres are flanked by tRNA genes and tandem copies of repetitive elements termed *dg* and *dh* repeats (Castel and Martienssen, 2013). Seminal work by Volpe and colleagues first established a role for small interfering (si) RNAs in directing transcriptional silencing of centromeric repeats in *Schizosaccharomyces pombe*. Mutants in RNA silencing factors are defective in centromeric silencing and accumulate centromere-derived sense and antisense transcripts (Volpe et al., 2002). Furthermore, loss of RNAi components leads to a loss of both H3K9me and the chromodomain-containing protein Swi6 from centromeric repeats, coupled with an increase in H3K4me3 and transcriptional activation of repeat sequences (Volpe et al., 2002).

It is now well-established that centromeric transcripts are processed into siRNAs by Dcr1, which are then loaded into Ago1, the core component of the RNA-induced initiation of transcriptional silencing (RITS) complex (Verdel et al., 2004; Castel and Martienssen, 2013) (Figure 1). Other protein components of RITS include Chp1, a centromere-associated chromodomain

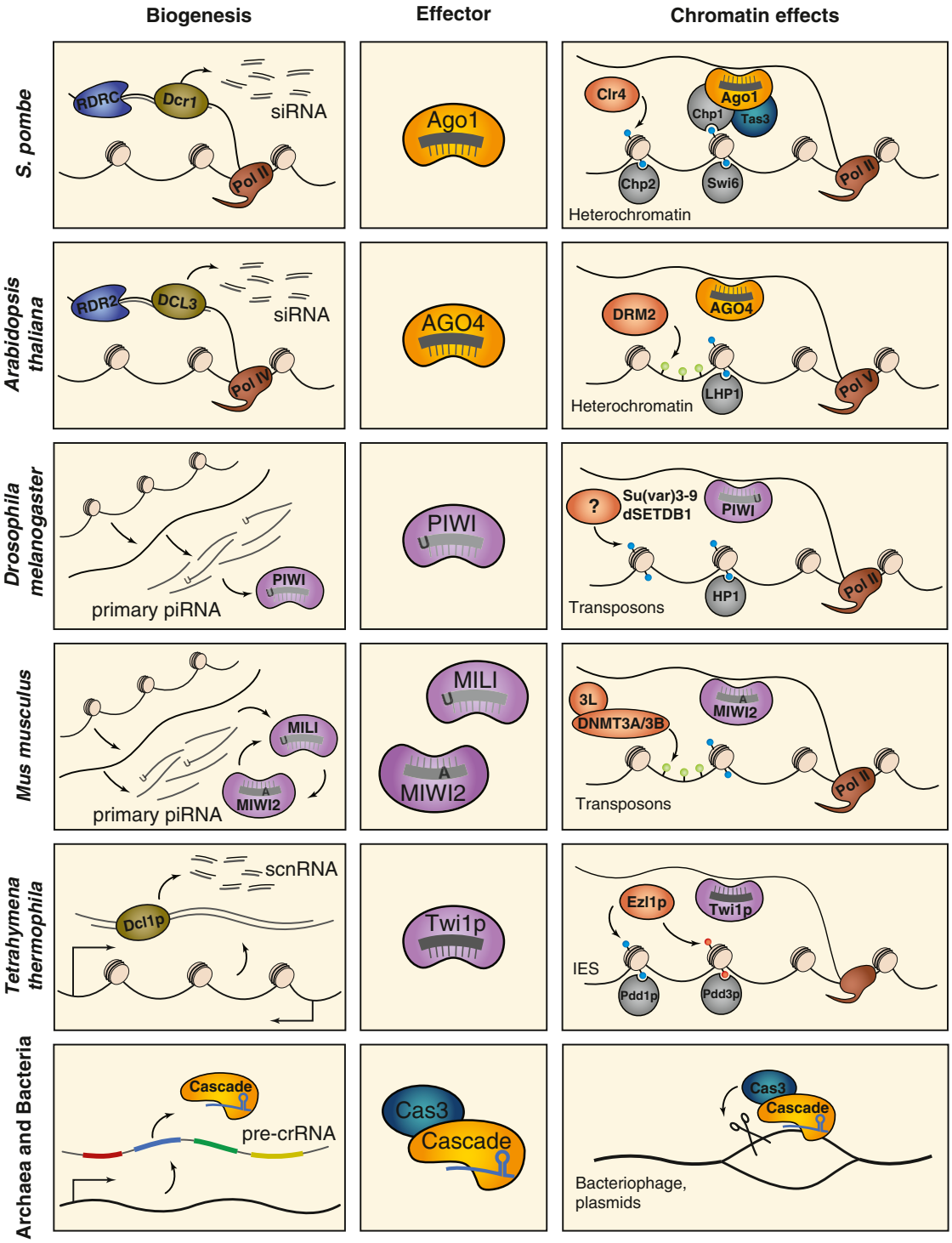

Figure 1. Parallels in Small RNA-Mediated Chromatin Effects

DNA silencing or cleavage is initiated by the biogenesis of small RNAs from host-encoded precursors, which are then bound by effector proteins and lead to the transcriptional silencing or elimination of complementary sequences. Orthologs are indicated as follows: Dicers, green; RNA-dependent RNA polymerase complexes, blue; polymerases, brown; Ago clade Argonautes, yellow; Piwi clade Argonautes, purple; histone and DNA methyltransferases, orange; and chromodomain-containing proteins, gray. Epigenetic modifications are represented as follows: H3K9me, blue; H3K27me, orange; and DNA methylation, green. The targets of small RNA-mediated TGS in each organism are indicated in the right panels.

protein that binds dimethylated H3K9, and Tas3, a GW domain-containing protein (Verdel et al., 2004). GW motifs are a common feature of Argonaute interactors; proteins containing this domain can also be found in Argonaute complexes in plants, flies, and mammals (Hutvágner and Simard, 2008). According to the current model for TGS in *S. pombe*, and in most other examples of small RNA-directed TGS in other species, target identification occurs through base pairing interactions between the bound siRNA and nascent transcripts at the target locus; in fission yeast, RITS associates with long, centromere-derived RNAs in a Dcr1-dependent manner (Motamedi et al., 2004). Moreover, RITS remains tethered to silent loci by the interaction of Chp1 with H3K9me, which is deposited by the HMT Clr4 (Noma et al., 2004). Clr4-dependent RITS tethering to heterochromatin is essential for maintaining transcriptional silencing and for generating new siRNAs (Noma et al., 2004).

RITS also physically interacts with another complex named RDRC, or the RNA-directed RNA polymerase complex, which contains Rdp1, an RNA-dependent RNA polymerase (Motamedi et al., 2004). The physical interaction between the RDRC and RITS complexes requires Dcr1 and Clr4 (Motamedi et al., 2004), suggesting that siRNA-driven RITS targeting to centromeric repeats, and stable H3K9 binding by Chp1, allows the recruitment of RDRC to chromatin. Moreover, the RDRC physically interacts with Dcr1, suggesting that siRNA biogenesis is coupled to dsRNA synthesis (Castel and Martienssen, 2013). Thus, the tethering of biogenesis, effector, and amplification complexes at the silenced loci likely organizes a self-perpetuating feedforward loop (Figure 1). The fundamental principles of siRNA-directed centromeric silencing that have been uncovered in *S. pombe* provide an excellent framework for understanding RNA-guided heterochromatin formation in other systems.

piRNA-Directed Transcriptional Silencing in Flies

Piwi-interacting RNAs (piRNAs) are a germline-specific class of small RNAs that bind to Piwi clade Argonaute proteins and constitute a transposon defense system in a variety of organisms, including humans. Defects in the piRNA pathway lead to uncontrolled transposition of mobile genetic elements, which can cause genomic instability and sterility (Malone and Hannon, 2009). piRNA-mediated transposon silencing has been studied most extensively in the *Drosophila* ovary. Although the mechanisms by which piRNAs repress transposon expression are still being elucidated, recent work in flies suggests that piRNAs induce and maintain heterochromatic silencing of these repetitive elements through the deposition of repressive histone modifications.

Drosophila piRNAs are between 23 and 29 nt in length and often bear a 5′ uridine (1U), compared to endogenous siRNAs and miRNAs, which tend to be 20–23 nt and show little sequence bias (Malone and Hannon, 2009). *Drosophila* encodes three nonredundant Piwi clade proteins that bind piRNAs: Piwi, the founding member of the clade, Aubergine (Aub), and Argonaute 3 (Ago3) (Hutvágner and Simard, 2008). Importantly, primary piRNAs are not generated by a Dicer protein. Although the precise mechanisms are only now beginning to be understood, it is clear that piRNA biogenesis is initiated by the transcription and processing of piRNA clusters, which are unique loci comprised of inactive fragments of several classes of transposons (Brennecke et al., 2007) (Figure 1). Following primary biogenesis, piRNAs enter into the ping-pong amplification cycle, which generates secondary piRNAs and is predominantly mediated by Aub and Ago3 in *Drosophila* ovarian germ cells. During ping-pong, a Piwi- or Aub-bound antisense piRNA recognizes the sense strand of an active transposon, leading to transcript cleavage. This cleavage generates the 5′ end of a new sense piRNA, which is bound by Ago3 and can recognize antisense transposon transcripts, generating new antisense piRNAs and thus continuing the amplification cycle (Brennecke et al., 2007; Gunawardane et al., 2007).

This elegant ping-pong amplification mechanism silences active transposons at the posttranscriptional level. How, then, does transposon repression occur in the somatic follicle cells of the ovary, where only Piwi and the primary piRNA pathway are known to operate? Several lines of evidence now point to a role for Piwi in mediating TGS in *Drosophila*. Piwi localizes to the nucleus, and cytoplasmic retention of Piwi results in massive transposon upregulation and sterility, which hints at a role in TGS rather than PTGS (Klenov et al., 2011). In addition, piRNA pathway mutants, including *piwi*, suppress position effect variegation (Malone and Hannon, 2009). Germline Piwi knockdown or cytoplasmic retention results in the depletion of H3K9me2 and HP1, the *Drosophila* homolog of Swi6, from a specific set of repetitive elements (Klenov et al., 2011; Wang and Elgin, 2011). In addition, nuclear run-on experiments in Piwi-depleted ovaries measured increased transcription of particular transposons as compared to wild-type tissue (Shpiz et al., 2011). Unexpectedly, nuclear run-on analysis of *armitage* mutants, in which Piwi does not bind piRNAs or enter the nucleus, did not detect differences in the transcription of repetitive elements between mutant and control ovaries, leading to doubts surrounding the role of Piwi in TGS (Malone and Hannon, 2009). Therefore, a direct comparison of chromatin states, nascent transcription, and steady-state transposon levels in the presence or absence of Piwi was necessary to unambiguously determine its role in TGS.

Recently, Sienski and colleagues reported such a study in an ovarian somatic follicle cell line, OSC, which does not express Aub or Ago3 and therefore allows the dissection of the Piwi-driven piRNA pathway in isolation. Following Piwi depletion, the authors found that transposon transcripts are upregulated, and Piwi-regulated transposons display enhanced Pol II occupancy and increased rates of nascent transcription (Sienski et al., 2012). By examining de novo insertions of individual TEs in the OSC genome, Sienski et al. found that active elements display Piwi-dependent H3K9me3. Moreover, two additional studies report similar effects on chromatin states and transposon suppression when Piwi is depleted in vivo (Le Thomas et al., 2013; Rozhkov et al., 2013).These data provide strong evidence that Piwi restricts transposon expression via TGS by directing the deposition of repressive histone modifications on TEs. Similar to heterochromatin formation in yeast, the recruitment of Piwi to TEs is likely mediated by base-pairing interactions between Piwi-bound piRNAs and nascent transposon transcripts, leading to the recruitment of H3K9 methyltransferases and associated chromatin binding factors to enforce a silent, heterochromatic state (Figure 1). Whether Piwi, like the yeast RITS complex, remains tethered to heterochromatic loci following H3K9 methylation has not yet been addressed.

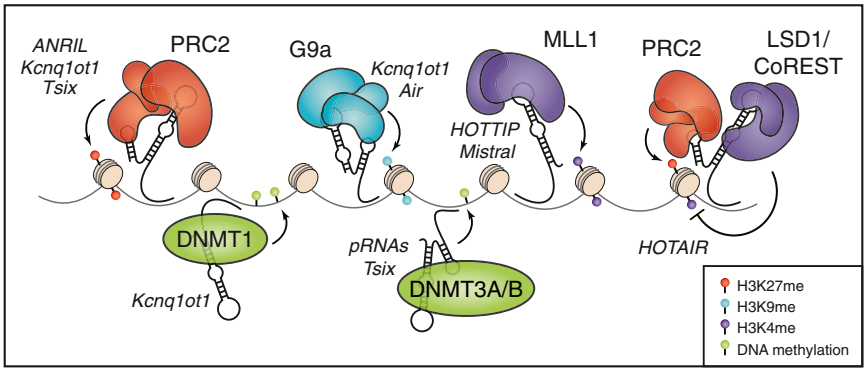

Figure 2. Effects of lncRNAs at the Chromatin Level

lncRNAs interact with several activating and repressive chromatin-modifying complexes. Representative lncRNAs are grouped according to the protein complex with which they interact. Depending on the lncRNA, genomic targeting of these chromatin modifiers can occur in *cis* or in *trans*.

Interactions of Long Noncoding RNAs with Chromatin Remodeling Machinery

Although the diverse functions of lncRNAs are only beginning to be uncovered, their potential ability to interact with and modulate the activity of chromatin regulatory complexes may allow lncRNAs to affect gene expression on a genome-wide scale. Several of the well-characterized lncRNAs, such as *Xist*, are involved in X chromosome inactivation (XCI) and imprinting, which require changes in chromatin structure on a massive scale (Lee, 2011). Moreover, many recently identified lncRNAs copurify with chromatin remodeling complexes, suggesting that they may function in targeting these complexes to genomic loci, or may serve as molecular scaffolds for complex assembly (Figure 2). Whether genomic targeting of protein complexes by lncRNAs occurs in *cis,* through interactions with nascent lncRNA transcripts, or in *trans*, perhaps through RNA:DNA duplexes or triplexes, is an important facet of lncRNA biology that future work must address. In a recent study, Guttman and colleagues characterized lncRNAs bound to 12 chromatin-associated complexes in mouse embryonic stem cells (Guttman et al., 2011). Interestingly, several lncRNAs purified with more than one complex, supporting the notion that lncRNAs serve as molecular scaffolds that bridge multiple regulatory units. One of these complexes, PRC2, has emerged as a common binding partner for lncRNAs in multiple organisms (Figure 2). EZH2 is the catalytic subunit of PRC2 in mammals and methylates H3K27 to enforce repressive heterochromatin (Table 1). Studies in multiple human and mouse cell types have found that hundreds of lncRNAs interact with PRC2, although the specificity of each of these interactions has not been rigorously explored (Khalil et al., 2009; Zhao et al., 2010).

One of the most well-known PRC2-interacting lncRNAs is *HOTAIR*, a 2.1 kb transcript derived from the human *HOXC* locus that represses the expression of genes within the *HOXD* locus, as well as other targets throughout the genome (Rinn et al., 2007). *HOTAIR* binds EZH2 and is required for PRC2-mediated H3K27 trimethylation and silencing of the *HOXD* locus in humans (Rinn et al., 2007). However, *HOTAIR* may play a different role in mice, since a deletion within the *HOXC* locus that encompasses *mHOTAIR* has no effect on expression of the *HOXD* locus (Schorderet and Duboule, 2011). *HOTAIR* overexpression has been linked to increased invasiveness and poorer outcomes in several human cancers (Rinn and Chang, 2012). Interaction of *HOTAIR* with PRC2 is mediated through a region in its 5′ terminus, while the 3′ terminus binds LSD1, a H3K4 demethylase that functions within the CoREST/REST complexes (Tsai et al., 2010). Overexpression of *HOTAIR* results in global changes in PRC2 occupancy and H3K27me3 marks (Rinn and Chang, 2012). Conversely, knockdown of *HOTAIR* alters the chromatin occupancy of PRC2 and LSD1 genome-wide, leading to reduced H3K27me3 and increased H3K4me2 at target loci (Tsai et al., 2010). Thus, *HOTAIR* appears to serve as a scaffolding molecule, bridging PRC2 with LSD1 (Figure 2).

Another lncRNA associated with human cancers is *ANRIL*, a long, antisense transcript found in the *INK4a/Arf* locus. *ANRIL* is overexpressed in human leukemias and prostate cancers, and its expression leads to epigenetic silencing of the nearby tumor suppressor *p15* (Yu et al., 2008; Yap et al., 2010). H3K27 trimethylation of the *INK4a/Arf* locus requires *ANRIL*, which was found to interact with SUZ12, a component of PRC2, and CBX7, a PRC1-associated chromodomain-containing protein (Rinn and Chang, 2012).

The *Air* and *Kcnq1ot1* lncRNAs are both involved in imprinting in mammals and use similar mechanisms to induce the deposition of repressive marks at silenced alleles. *Air* is transcribed from the *Igf2r* locus and mediates H3K9 trimethylation of the nearby *Slc22a3* promoter by recruitment of the repressive G9a chromatin-modifying complex in *cis* (Ponting et al., 2009). Similarly, *Kcnq1ot1*, which is transcribed antisense to the silenced paternal allele of *Kcnq1*, binds and recruits G9a and PRC2 to direct H3K9 and H3K27 trimethylation of the locus in *cis* (Pandey et al., 2008). Defects in silencing by either of these lncRNAs result in biallelic expression of their normally imprinted targets, underscoring their importance in maintaining monoallelic heterochromatic silencing.

In addition to the *HOXC* cluster-derived lncRNA *HOTAIR*, three *HOXA* locus-associated lncRNAs have been described. *HOTTIP* is expressed from the 5′ end of the locus, *Mistral* is encoded between *Hoxa6* and *a7*, and *HOTAIRM1* is located at the distal 3′ end of the cluster (Hu et al., 2012). Little is known about *HOTAIRM1*, apart from the fact that it is induced during myelopoiesis and is required for expression of *HOXA* genes during myeloid differentiation. *HOTTIP* and *Mistral*, on the other hand, mediate the transcriptional activation of *HOXA* genes through physical interactions with the MLL1 complex, a H3K4 methyltransferase complex which is known to bind to *HOX* gene promoters (Bertani et al., 2011; Wang et al., 2011). Depletion of either *HOTTIP* or *Mistral* results in a loss of MLL1 occupancy on *HOXA* target genes. Therefore, in the case of *HOTTIP* and *Mistral*, interaction of the lncRNA with chromatin modifiers

results in the deposition of activating chromatin marks rather than repressive marks (Figure 2).

In general, the ability of lncRNAs to recruit the activities of protein complexes to genomic loci may allow the cell to impart specificity to broadly acting chromatin-modifying machineries. It is clear that much work remains to be done in order to fully understand the complex interactions between lncRNAs and chromatin-modifying complexes, particularly the acetylating or ubiquitinating complexes. However, by characterizing the functions of these molecules through dissecting their protein partners and identifying the genomic loci with which they interact, the regulatory power of lncRNAs may one day be harnessed for use in chromatin-targeted therapeutic applications.

RNA-Directed DNA Methylation

DNA methylation was the first RNA-guided epigenetic modification to be discovered, and is utilized by organisms ranging from plants to mammals. In most systems, methylation occurs on cytosine residues in the context of a CG dinucleotide, or CpG motif, by the action of DNA methyltransferase (DNMT) proteins. DNA methylation is a dynamic epigenetic modification that often functions as a repressive mark to silence transcription (Law and Jacobsen, 2010). Two types of DNMT proteins perform DNA methylation, and differ in their preferred substrates: de novo DNMTs methylate completely unmethylated CpG dinucleotides, while maintenance methyltransferases act on hemimethylated DNA during DNA replication to methylate the newly replicated strand (Law and Jacobsen, 2010). Table 1 lists the DNMTs present in species relevant to this review. Notably, many species have lost the ability to methylate their DNA; these include *S. pombe* and *D. melanogaster*.

Small RNA-Directed DNA Methylation in Plants

RNA-directed DNA methylation (RdDM) was first described in plants, following the observation that viroid cDNAs become specifically methylated upon integration into the tobacco genome (Law and Jacobsen, 2010). Since its initial discovery, the pathway has been characterized extensively in *Arabidopsis* and now serves as the defining model for RdDM in eukaryotes. Genome-wide mapping of methylation in *Arabidopsis* has revealed that most methylation occurs on transposons and repetitive elements, which are concentrated near centromeres (Castel and Martienssen, 2013). RNA Pol IV, a specialized polymerase essential for RdDM, produces long RNA transcripts from these heterochromatic regions (Henderson and Jacobsen, 2007). Pol IV transcripts are substrates for RDR2, an RNA-dependent RNA polymerase, which produces the complementary strand to generate dsRNA (Law and Jacobsen, 2010). The RNase III enzyme DCL3 processes the dsRNA from repetitive regions into 24 nt heterochromatic siRNAs, which are bound by the Argonaute family member AGO4 (Henderson and Jacobsen, 2007). AGO4 then targets the de novo methyltransferase DRM2 to the corresponding genomic locus for methylation and silencing (Henderson and Jacobsen, 2007) (Figure 1).

Further molecular details of this process were uncovered following the identification of a second specialized polymerase complex in *Arabidopsis*: RNA polymerase V. Wierzbicki and colleagues found that Pol V transcribes intergenic noncoding transcripts within heterochromatin, which serve as scaffolds for AGO4 recruitment (Wierzbicki et al., 2008). AGO4 not only binds to nascent Pol V transcripts through siRNA-mediated base pairing but also interacts directly with the largest Pol V subunit, NRPE1 (Law and Jacobsen, 2010). In a sense, the mechanism guiding AGO4 to repetitive loci parallels the recruitment of *S. pombe* RITS or *Drosophila* Piwi to their targets: each of these processes is driven by an interaction with nascent transcripts at the locus.

The mechanism by which the methyltransferase DRM2 is recruited to AGO4-targeted regions has not been fully elucidated, although several recent studies hint that RDM1, a novel regulator of RdDM, may play a role in the process (Castel and Martienssen, 2013). RDM1 specifically binds to single-stranded methylated DNA and is required for the accumulation of Pol V transcripts. Astonishingly, RDM1 interacts not only with Pol V subunits but also with AGO4 and DRM2 (Castel and Martienssen, 2013). Therefore, RDM1 may recruit Pol V to heterochromatic loci, while mediating subsequent interactions between components of the RNA silencing pathway and methylation machinery at the locus. However, the sequence of these events and the molecular mechanisms that drive them have not yet been established.

piRNA-Mediated Methylation of Mobile Genetic Elements in Mammals

Although many of the fundamental principles of piRNA pathway function have been characterized in *Drosophila*, including the recent discovery of piRNA-guided histone modifications, the role of piRNAs in TGS was first described in mammals, where piRNAs direct the methylation of transposons in order to enforce their transcriptional repression. Mammalian piRNA pathways operate in male germ cells, and the expression of piRNAs occurs at two distinct stages during development. Pre-pachytene piRNAs, and their interacting Piwi proteins, Mili and Miwi2, are expressed in primordial germ cells and map to transposons and repetitive elements (Malone and Hannon, 2009). Conversely, pachytene piRNAs are expressed during adult spermatogenesis, and although some pachytene piRNAs align to repetitive genomic regions, the majority map uniquely throughout the mouse genome, and their function remains unknown (Aravin et al., 2007a). The mouse Piwi proteins expressed at this stage are Miwi and, to a lesser extent, Mili, which do not appear to direct transposon repression via TGS or to engage in ping-pong amplification (Malone and Hannon, 2009).

During mammalian embryonic germ cell development, the genome undergoes a process of global demethylation, followed by the reestablishment of methylation at repetitive elements. Prior to the identification of piRNAs, it was known that transposon silencing in primordial germ cells was dependent upon methylation of repetitive elements by de novo methyltransferases DNMT3A and 3B (Okano et al., 1999). DNMT3L, which is related to DNMT3A and DNMT3B in sequence but lacks their catalytic motifs, stimulates the activity of DNMT3A and DNMT3B and specifically regulates transposon silencing in the germline; DNMT3L is required for the re-establishment of DNA methylation on repetitive elements, and its mutation results in transposon upregulation and sterility (Malone and Hannon, 2009). Due to the overlapping phenotypes between *mili*, *miwi2*, and *dnmt3l* mutants, investigators asked whether piRNAs were

involved in the process of de novo methylation of transposons in developing germ cells. Indeed, transposon methylation is lost in *mili* and *miwi2* mutants, inducing a massive upregulation of TEs, which eventually leads to sterility (Carmell et al., 2007; Aravin et al., 2007b; Kuramochi-Miyagawa et al., 2008). DNMT3L acts downstream of Mili and Miwi2, since *dnmt3l* mutants display only minor defects in piRNA populations (Malone and Hannon, 2009). Therefore, transposon suppression in primordial germ cells depends on TGS, which is achieved through methylation of repetitive elements via Mili- and Miwi2-bound piRNAs and the action of de novo DNMTs (Figure 1).

Although a functional link between the piRNA pathway and transposon methylation is clear, the molecular mechanisms by which the mammalian DNA methylation machinery is recruited to target loci remain unknown. Given the well-characterized pathways governing small RNA-mediated TGS in other organisms, it is likely that Mili and Miwi2 complexes are recruited to repetitive elements through piRNA-guided base-pairing interactions with nascent TE transcripts. Once tethered to the locus, the complexes may recruit the methylation machinery directly. Alternatively, it is also possible that histone methylation serves as a signal for DNA methylation. Indeed, H3K9 methylation is a prerequisite for DNA methylation in *Neurospora crassa* and guides maintenance DNA methylation in *Arabidopsis* (Malone and Hannon, 2009; Law and Jacobsen, 2010). Mammalian DNMT proteins have been reported to interact with EZH2, as well as H3K9 methyltransferases SUV39H1, G9a, and SETDB1 (Jin et al., 2011) (Table 1). Thus, methylation of histones and DNA may synergistically enforce and perpetuate a repressive, heterochromatic state at target loci. Investigation of potential interactions between the piRNA machinery and histone-modifying complexes will be an important next step in understanding piRNA-directed transcriptional repression in mammalian systems.

DNA Methylation Mediated by Long Noncoding RNAs

Given the complex interplay between DNA methylation and histone modifications, combined with rapidly mounting evidence supporting lncRNA-mediated effects on chromatin structure, it is not surprising that several examples of lncRNA-guided DNA methylation have been reported in the literature (Figure 2). However, the interaction of lncRNAs with the DNA methylation machinery has not been investigated on a global scale.

Tsix is a lncRNA that regulates the expression of *Xist* during XCI. In addition to binding and recruiting PRC2 to the *Xist* locus (Lee, 2011), *Tsix* physically interacts with the de novo methyltransferase DNMT3A and recruits the protein to the *Xist* promoter in order to induce DNA methylation and silencing (Sun et al., 2006). Similarly, recent studies have uncoupled the role of *Kcnq1ot1* in tissue-specific imprinting from its role in regulating ubiquitously imprinted genes. A 890 nt domain within *Kcnq1ot1* regulates differential methylation of somatically imprinted genes through a physical interaction with the maintenance methyltransferase DNMT1 (Mohammad et al., 2010). Loss of this domain does not abolish interaction of *Kcnq1ot1* with G9a or EZH2, suggesting that the lncRNA utilizes a multifaceted approach to regulate its targets through interactions with both chromatin modifiers and DNA methylation machinery.

Noncoding RNAs are also involved in the silencing of clustered ribosomal RNA genes (rDNA) within mammalian genomes. rDNA is transcriptionally silenced through repressive histone modifications and DNA methylation of the promoter region, which is mediated by NoRC, a chromatin remodeling complex (Schmitz et al., 2010). RNA Pol I-derived transcripts from rDNA promoters, termed pRNAs, play an essential role in this process by binding to NoRC. Interestingly, pRNAs were recently found to mediate NoRC-independent de novo methylation and heterochromatic silencing of rDNA promoters. Through a series of elegant structural and functional studies, Schmitz and colleagues demonstrate that pRNAs bind to an element within the rDNA promoter termed T_0. In fact, pRNAs form an RNA:DNA:DNA triplex structure with T_0 in vitro (Schmitz et al., 2010). Moreover, DNMT3B, and to a lesser extent DNMT3A, preferentially binds RNA:DNA:DNA triplexes rather than DNA duplexes in vitro (Schmitz et al., 2010). Although the presence of this triplex has not been validated in vivo, the current model suggests that the triplex formed between pRNAs and T_0 directly recruits DNMT3B to the rDNA promoter to induce DNA methylation and silencing.

This study has important implications for the study of lncRNA-induced chromatin changes; not only does it suggest that lncRNAs may be capable of forming triplex structures with DNA in vivo, but it also demonstrates that cellular proteins, such as DNMTs, are able to specifically recognize and bind to RNA:DNA:DNA triplex structures. The development of predictive tools such as the Triplexator (Buske et al., 2012), to predict triplex structures genome-wide, will hopefully aid in determining whether lncRNAs commonly utilize triplexes when locating target loci. The recent generation of genome-wide DNA methylation maps for a diverse number of organisms and cell types supports the notion that differentially methylated regions may be important in orchestrating gene expression across tissues. Considering how little we currently understand about the role of lncRNAs in directing DNA methylation genome-wide, it will be of interest to investigate the extent to which lncRNAs control differential methylation in specific cell types, and whether depletion or ectopic expression of lncRNAs is capable of modulating these patterns.

RNA-Directed DNA Cleavage

Yeast, plants, and animals have evolved elegant small RNA-directed strategies to suppress the potentially catastrophic activity of mobile genetic elements harbored within their own genomes, as discussed above. Perhaps not surprisingly, small RNA-driven pathways have also evolved in bacteria, archaea, and protists, and share many features with their analogous counterparts in other systems. However, as we will discuss in the following sections, RNA-directed targeting by these ancient silencing pathways leads to DNA cleavage rather than transcriptional repression. Accordingly, small RNA-driven silencing must be carefully regulated, since DNA cleavage and elimination, unlike chromatin silencing, is an irreversible process.

DNA Elimination in Ciliates

Ciliated protozoans have historically served as fascinating models for the study of noncoding RNAs and programmed DNA rearrangements. Ciliates compartmentalize their DNA into

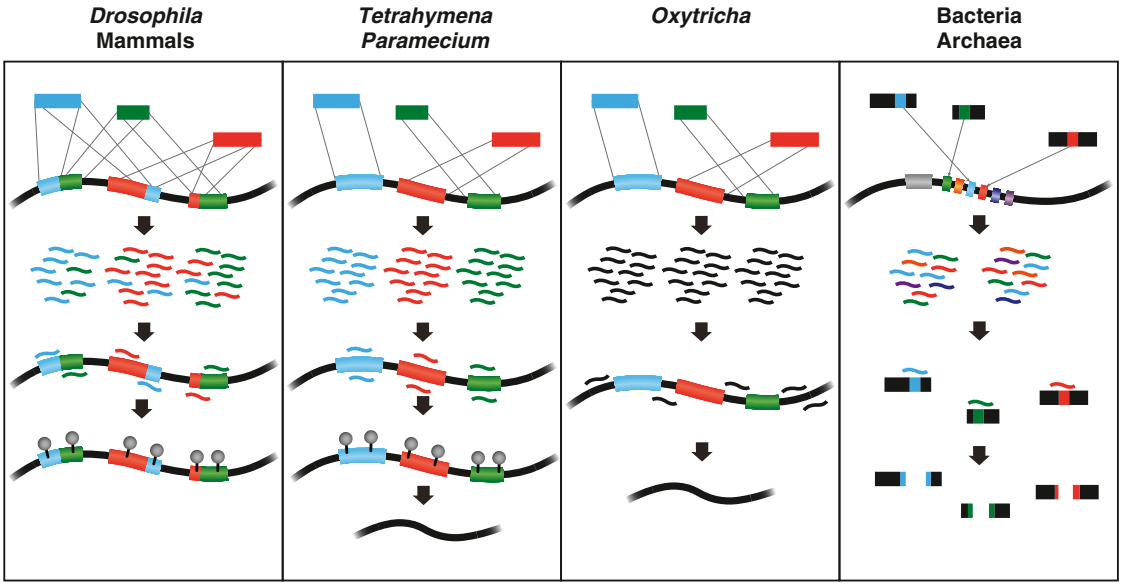

Figure 3. Small RNA-Mediated Genome Defense

Foreign transposable and repetitive sequences, represented by blue, green, and red segments, are present in the genomes of mammals, insects, and ciliates. Mammals and *Drosophila* utilize piRNAs to silence active transposons, either by DNA methylation or by repressive histone modifications (gray circles). In *Tetrahymena* and *Paramecium*, scnRNAs recognize IES regions, which include mobile genetic elements, and guide their genomic excision by marking them with repressive histone modifications. *Oxytricha* eliminates IES sequences using an orthogonal mechanism; piRNAs correspond to the retained sequences (black). Chromatin modifications that guide this process have not yet been identified. Bacteria and archaea incorporate sequences from foreign pathogens or plasmids (colored segments) into CRISPR loci. Expression of crRNAs from these loci results in cleavage of infecting bacteriophages and plasmid DNA.

two distinct nuclei: the germline micronucleus (MIC) and the somatic macronucleus (MAC). The diploid MIC stores all of the genetic information necessary for reproduction and development and is transcriptionally silent during vegetative growth. The MAC, on the other hand, contains a polyploid, stripped-down copy of the MIC genome that functions as a transcriptional machine to produce the cellular factors critical for vegetative growth. Within the MAC, much of the intergenic or repetitive content that is present in the MIC has been eliminated, and split genes have been reassembled. During the sexual process of conjugation, a new MIC and MAC are generated from the zygotic nucleus. As the details of conjugation are complex and lie beyond the scope of this review, we refer the reader to two excellent reviews (Matzke and Birchler, 2005; Nowacki et al., 2011). As the new MAC differentiates, it undergoes an extensive series of gene rearrangements that result in elimination of a substantial portion of the parental genome. The eliminated sequences are termed internal eliminated segment (IES) sequences, many of which are believed to be remnants of inactivated transposons (Nowacki et al., 2011). Different ciliate species undergo varying degrees of DNA elimination; *Tetrahymena thermophila* discards ~6,000 unique sequences, or ~15% of its genome, while *Paramecium tetraurelia* expels an estimated 60,000 IES regions. Finally, *Oxytricha trifallax* discards an astounding 95%–98% of its genome during MAC differentiation while simultaneously unscrambling the retained segments in order to assemble full-length, properly ordered genes from discontinuous fragments (Nowacki et al., 2011).

How can such a process occur in a controlled, programmable manner? The first molecular insight into this question was the observation that IES retention or elimination was a homology-dependent phenomenon that was maternally transmitted in both *Tetrahymena* and *Paramecium* (Duharcourt et al., 1995; Chalker and Yao, 1996). That is, sequences present in the parental MAC were retained in the offspring, while elements that were absent in the parent were eliminated from the new MAC. In short, retention of a given sequence in the new MAC is dictated by its presence in the parental MAC.

We now know that the process of DNA elimination in *Tetrahymena* and *Paramecium* is orchestrated by small RNAs, termed scanRNAs (scnRNAs), which are expressed during conjugation and direct chromatin modifications to mark IES sequences for elimination. scnRNAs are generated from long, overlapping MIC transcripts and scan the parental MAC for homologous sequences. If a scnRNA identifies a complementary sequence in the MAC, it is degraded, while unpaired scnRNAs are transported into the developing MAC to mark homologous sequences for elimination (Mochizuki et al., 2002) (Figure 3). Interestingly, scnRNA selection occurs through the scanning of long RNA transcripts derived from the parental MAC rather than through scanning the DNA itself (Mochizuki, 2010).

Genetic and biochemical characterization of the protein factors that drive scnRNA-mediated DNA elimination in *Tetrahymena* and *Paramecium* have revealed remarkable parallels with cognate RNA silencing pathways in other systems (Figure 1). For simplicity, we will now focus on the mechanisms known to operate within *Tetrahymena*, although an analogous pathway also exists in *Paramecium.* Similar to siRNA biogenesis in yeast, plants, and higher organisms, the production of scnRNAs requires the activity of a Dicer protein, Dcl1p, which generates

~27–30 nt products (Malone et al., 2005; Mochizuki and Gorovsky, 2005). Following biogenesis, mature scnRNAs bind Twi1p, a Piwi clade Argonaute protein (Mochizuki et al., 2002), and therefore resemble mammalian and *Drosophila* piRNAs.

Similar to transposon silencing in *S. pombe* and *Drosophila*, repressive histone modifications are a critical component of DNA rearrangements in ciliates. Both H3K9me and H3K27me are required for proper genome rearrangements and mark IES sequences for elimination (Mochizuki, 2010). The HMT Ezl1p, which is related to *Drosophila* E(z) and EZH2 in mammals, is responsible for methylating both H3K9 and H3K27 in *Tetrahymena*, and is required for DNA elimination (Liu et al., 2004; 2007). Importantly, Dcl1p and Twi1p are also required for the methylation of H3K9 and H3K27 (Mochizuki, 2010). These chromatin marks are bound by the chromodomain-containing proteins Pdd1p and Pdd3p, which are required for proper genome rearrangement (Mochizuki, 2010). Tethering Pdd1 to an inactive IES is sufficient to induce its excision (Taverna et al., 2002), providing strong evidence that repressive histone modifications serve as a signal to promote DNA elimination. Taken together, these discoveries suggest a model for DNA elimination whereby selected scnRNAs, which are generated by Dcl1p and bound by Twi1p, enter the nucleus of the developing MAC, and locate homologous sequences within IES regions through interactions with nascent transcripts. Tethered Twi1p then recruits Ezl1p to deposit repressive methyl marks on the proximal histones, which are bound by Pdd1p and Pdd3p and trigger elimination by the DNA excision machinery (Figure 1).

The process of DNA elimination in *Oxytricha* resembles that of *Paramecium* and *Tetrahymena* in a number of ways, but one stark contrast between them has been the lack of evidence supporting a role for small RNAs in *Oxytricha* conjugation. However, recent work by Fang and colleagues has uncovered a surprising twist: *Oxytricha* expresses small RNAs during conjugation, but their function is orthogonal to scnRNAs; *Oxytricha* small RNAs mark complementary regions for *retention*, rather than elimination (Fang et al., 2012) (Figure 3). These conjugation-specific small RNAs, or piRNAs, are 27 nt in length and display a 1U bias, similar to piRNAs in other organisms (Fang et al., 2012; Zahler et al., 2012). The Piwi protein Otiwi1 binds *Oxytricha* piRNAs and is critical for the development of the new MAC. Accordingly, injection of a synthetic piRNA complementary to an eliminated IES programs its retention in the offspring (Fang et al., 2012). To unscramble split genes, *Oxytricha* utilizes a mechanism similar to scnRNA selection, in which transcription of the parental MAC produces long RNA transcripts that are used as templates to guide proper segment orientation for the reassembly of full-length genes in the developing MAC (Nowacki et al., 2011).

These insights into DNA elimination in *Oxytricha* resolve many of the uncertainties that have plagued the community for years, but they also raise new questions. For instance, are piRNAs generated from single- or double-stranded precursors, and which enzyme(s) is responsible for their biogenesis? Moreover, what are the signals deposited on macronuclear-destined sequences that mark them for retention? Intriguingly, conjugation triggers the methylation of *Oxytricha* DNA (Bracht et al., 2012). Cytosine methylation is enriched within the repetitive sequences that are eliminated during MAC development, suggesting that DNA methylation may be involved in marking specific elements for degradation. Further investigation of DNA methylation and other epigenetic marks present on retained and eliminated sequences during *Oxytricha* conjugation will be a critical next step.

CRISPR-Mediated Defense against Foreign DNA in Prokaryotes

Prokaryotic organisms face a constant barrage of viral pathogens; bacteriophages are the most abundant viruses on Earth and have the ability to rapidly evolve and adapt to their environments, creating the need for an antiviral defense strategy that is equally flexible and adaptable. Bacteria and archaea have responded by employing a small RNA-driven pathway that not only protects against invading phage but also retains a molecular memory of the pathogen through the incorporation of small portions of the foreign DNA into host loci termed CRISPRs. CRISPR loci can be found in ~40% of bacterial species, and in most archaea (Marraffini and Sontheimer, 2010). A CRISPR locus is a tandem array of short direct-repeat sequences, which are separated by unique spacer regions. On average, there are 20 repeat-spacer units in a locus. Repeat lengths range between 21 and 47 nt, while spacer lengths can be 20–72 nt (Karginov and Hannon, 2010). CRISPR-dependent immunity is mediated by small RNAs termed CRISPR RNAs (crRNAs) and is achieved in three phases: adaptation, expression, and interference (Makarova et al., 2011). CRISPR loci are flanked by a diverse array of *cas* genes, which are responsible for mediating these three phases of immunity.

During the process of adaptation, short pieces of phage and plasmid sequences, termed protospacers, are identified and incorporated into the CRISPR loci (Barrangou et al., 2007). Although the molecular mechanism by which novel spacer sequences are integrated into CRISPR loci remains unclear, in most systems, their selection relies on the presence of a protospacer-adjacent motif (PAM). PAMs are short sequences encoded within the phage or plasmid genome that lie adjacent to the region destined for integration into a CRISPR locus (Makarova et al., 2011). These sequences likely serve as recognition motifs for the as-yet-unidentified protospacer integration machinery, and aid in target recognition and cleavage during the interference phase of immunity. The highly conserved cas proteins Cas1 and Cas2 are suspected to be involved in protospacer acquisition (Barrangou et al., 2007; Brouns et al., 2008), but no direct evidence linking them to this process has been reported.

The expression phase of CRISPR-mediated antiviral defense involves the transcription of CRISPR loci to yield long precursor RNAs, which are then processed by specialized Cas proteins into mature crRNAs (Marraffini and Sontheimer, 2010). Although the mechanism of crRNA biogenesis varies between different types of CRISPR systems, the expression phase ultimately results in the generation of small RNAs bearing phage- or plasmid-derived spacer sequences, which are then funneled into the interference phase of the pathway (Figure 1 and Figure 3).

CRISPR-mediated interference is responsible for targeting and cleaving the DNA of invading phage or plasmids through crRNA-mediated base-pairing interactions, reminiscent of miRNA-mediated seed pairing in other systems. In order to mediate pathogen interference, crRNAs recruit effector complexes to complementary sequences within the foreign DNA,

which are then cleaved within the spacer sequence (Wiedenheft et al., 2012) (Figure 1 and Figure 3).

Three types of CRISPR loci have been characterized to date, and more than one type of locus can be found within a single organism (Makarova et al., 2011). Although the general phases of the pathway remain the same, the cas proteins responsible for carrying out these phases differ between CRISPR types. The hallmark of type I CRISPR-Cas systems is the presence of a *cas3* gene, along with a large, multisubunit effector complex known as the CRISPR-associated complex for antiviral defense, or Cascade (Brouns et al., 2008; Makarova et al., 2011). Cascade processes long, pre-crRNA transcripts into mature crRNAs through endonucleolytic cleavage (Brouns et al., 2008) (Figure 1). Next, the complex recognizes and binds a PAM motif within the target dsDNA, and recruits Cas3, which catalyzes target cleavage (Marraffini and Sontheimer, 2010).

Type II CRISPR-Cas systems, such as those found in *S. pyogenes*, rely heavily on a large cas protein, Cas9, and a *trans*-activating CRISPR RNA (tracrRNA) (Wiedenheft et al., 2012). The tracrRNA bears 24 nt of perfect sequence complementarity to the repeat regions within the CRISPR locus, and base pairs with crRNA precursor RNAs. This duplex is bound by Cas9 and cleaved by a host-encoded RNase III protein to generate mature crRNAs (Deltcheva et al., 2011). Apart from its role in crRNA biogenesis, Cas9 was recently implicated in crRNA-guided DNA cleavage. Cas9, in complex with a crRNA and tracrRNA, identifies the dsDNA target and catalyzes two ssDNA cleavage events within the target. Most importantly, Jinek and colleagues demonstrate that Cas9 can be programmed for dsDNA cleavage using a single hybrid crRNA-tracrRNA molecule, which combines the essential features from both RNAs and can, in theory, be engineered to cleave any DNA of interest (Jinek et al., 2012). Indeed, several recent studies have demonstrated that programmed Cas9 can mediate specific DNA cleavage in human cells, providing tangible evidence that this technology may one day be used clinically for genome editing or gene therapy (Cong et al., 2013; Mali et al., 2013).

The final type of CRISPR-Cas system is the type III pathway. Present in organisms such as *P. furiosis*, crRNA biogenesis is carried out by Cas6, which then transfers the small RNA to an effector complex to mediate target recognition and cleavage (Makarova et al., 2011). While some type III systems have been shown to target DNA, others appear to cleave complementary RNAs (Hale et al., 2009). By targeting RNA, these CRISPR-Cas systems bear an even closer resemblance to eukaryotic RNA-silencing pathways and may potentially be functioning in the degradation of viral transcripts.

It is clear that bacteria, archaea, and viral pathogens are engaged in a dynamic interplay that has shaped their evolutionary history. In order to gain an upper hand, bacteria and archaea have developed an intricate repertoire of small RNA-driven effector mechanisms based upon long-term immunological memory of previous viral challenge. Moreover, it seems that viruses have responded to the restrictive pressures placed upon them by CRISPR-Cas systems by evolving protein-based suppressors that inactivate CRISPR-mediated defenses during infection (Bondy-Denomy et al., 2013). Bacteriophage also engage in rapid sequence evolution by shuffling their genomes in 25 nt blocks (Andersson and Banfield, 2008). Given the size of the shuffled units, which approximate the size of many protospacers, it is tempting to speculate that the process is in some way linked to spacer acquisition in some CRISPR-Cas systems.

Molecular Connections between RNA and Chromatin: Future Perspectives

As connections between noncoding RNAs and chromatin-level regulatory mechanisms are uncovered, a number of striking similarities between diverse species have emerged. Although the RNAs themselves differ between organisms and cellular pathways, they converge on a common challenge: to regulate the content and expression of DNA. Both small and large noncoding RNAs are able to direct chromatin-modifying machinery to specific targets, often through base pairing with nascent transcripts at the locus. In many cases, interaction with genomic targets leads to the deposition of covalent modifications, or in the most extreme cases, DNA cleavage. Interactions between noncoding RNAs and chromatin-modifying machinery are key components of these regulatory systems; lncRNAs in particular can interact directly with both HMTs and DNMTs, which implicates this novel class of molecules in physically guiding regulatory protein complexes to genomic targets (Figure 2). Small RNAs, on the other hand, often associate with effector protein complexes, which then recruit chromatin modifiers to targets based on small RNA-mediated base pairing with nascent transcripts (Figure 1). In many small RNA-driven silencing pathways, the regulatory RNAs identify and mark potentially dangerous "nonself" elements for transcriptional silencing or elimination. However, in other systems, homology between the regulatory RNA and the target locus marks the region as "self" and protects it from elimination (Figure 3). Nevertheless, these complementary mechanisms both rely on signals directed by host-encoded small RNAs to maintain genomic stability.

Although many of the RNA-mediated pathways that we describe are responsible for preventing genetic lesions caused by activated transposons, another essential facet of genome stability is the ability to repair breaks in DNA once they arise. Perhaps not surprisingly, noncoding RNAs were recently implicated in double-strand break repair in plants and animals (Francia et al., 2012; Wei et al., 2012). Although the mechanistic details of this process have not been fully characterized, particularly in mammals, these findings highlight the fact that we likely have only scratched the surface of the myriad ways in which RNA can influence DNA. While noncoding RNAs clearly mediate the methylation of histones and DNA, the role of RNA in directing other covalent marks, such as acetylation or ubiquitination, remains largely unexplored. As we expand our understanding of chromatin structure and modifications, we will undoubtedly continue to uncover exciting new mechanisms of epigenetic regulation by noncoding RNAs.

ACKNOWLEDGMENTS

Given the broad scope of this review and the limitations in space, we apologize to our colleagues whose work we were unable to describe or reference. L.R.S is a Damon Runyon Fellow supported by the Damon Runyon Cancer Research Foundation (DRG-2016-12). M.J.D. is supported by a graduate studies fellowship from "la Caixa" Foundation and by the Watson School of Biological Sciences. G.J.H. is supported by grants from the NIH and by a kind gift from Kathryn W. Davis. G.J.H. is an investigator of the HHMI.

REFERENCES

Andersson, A.F., and Banfield, J.F. (2008). Virus population dynamics and acquired virus resistance in natural microbial communities. Science *320*, 1047–1050.

Aravin, A.A., Hannon, G.J., and Brennecke, J. (2007a). The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science *318*, 761–764.

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

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709–1712.

Bertani, S., Sauer, S., Bolotin, E., and Sauer, F. (2011). The noncoding RNA Mistral activates Hoxa6 and Hoxa7 expression and stem cell differentiation by recruiting MLL1 to chromatin. Mol. Cell *43*, 1040–1046.

Black, J.C., Van Rechem, C., and Whetstine, J.R. (2012). Histone lysine methylation dynamics: establishment, regulation, and biological impact. Mol. Cell *48*, 491–507.

Bondy-Denomy, J., Pawluk, A., Maxwell, K.L., and Davidson, A.R. (2013). Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. Nature *493*, 429–432.

Bracht, J.R., Perlman, D.H., and Landweber, L.F. (2012). Cytosine methylation and hydroxymethylation mark DNA for elimination in Oxytricha trifallax. Genome Biol. *13*, R99. http://dx.doi.org/10.1186/gb-2012-13-10-r99.

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.

Brouns, S.J.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J.H., Snijders, A.P.L., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science *321*, 960–964.

Buske, F.A., Bauer, D.C., Mattick, J.S., and Bailey, T.L. (2012). Triplexator: detecting nucleic acid triple helices in genomic and transcriptomic data. Genome Res. *22*, 1372–1381.

Carmell, M.A., Girard, A., van de Kant, H.J.G., 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.

Castel, S.E., and Martienssen, R.A. (2013). RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. Nat. Rev. Genet. *14*, 100–112.

Chalker, D.L., and Yao, M.C. (1996). Non-Mendelian, heritable blocks to DNA rearrangement are induced by loading the somatic nucleus of Tetrahymena thermophila with germ line-limited DNA. Mol. Cell. Biol. *16*, 3658–3667.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature *471*, 602–607.

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.

Fang, W., Wang, X., Bracht, J.R., Nowacki, M., and Landweber, L.F. (2012). Piwi-interacting RNAs protect DNA against loss during Oxytricha genome rearrangement. Cell *151*, 1243–1255.

Francia, S., Michelini, F., Saxena, A., Tang, D., de Hoon, M., Anelli, V., Mione, M., Carninci, P., and d'Adda di Fagagna, F. (2012). Site-specific DICER and DROSHA RNA products control the DNA-damage response. Nature *488*, 231–235.

Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M.C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in Drosophila. Science *315*, 1587–1590.

Guttman, M., Donaghey, J., Carey, B.W., Garber, M., Grenier, J.K., Munson, G., Young, G., Lucas, A.B., Ach, R., Bruhn, L., et al. (2011). lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature *477*, 295–300.

Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M., and Terns, M.P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell *139*, 945–956.

Henderson, I.R., and Jacobsen, S.E. (2007). Epigenetic inheritance in plants. Nature *447*, 418–424.

Hu, W., Alvarez-Dominguez, J.R., and Lodish, H.F. (2012). Regulation of mammalian cell differentiation by long non-coding RNAs. EMBO Rep. *13*, 971–983.

Hutvágner, G., and Simard, M.J. (2008). Argonaute proteins: key players in RNA silencing. Nat. Rev. Mol. Cell Biol. *9*, 22–32.

Jin, B., Li, Y., and Robertson, K.D. (2011). DNA methylation: superior or subordinate in the epigenetic hierarchy? Genes Cancer *2*, 607–617.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816–821.

Karginov, F.V., and Hannon, G.J. (2010). The CRISPR system: small RNA-guided defense in bacteria and archaea. Mol. Cell *37*, 7–19.

Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B.E., van Oudenaarden, A., et al. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc. Natl. Acad. Sci. USA *106*, 11667–11672.

Klenov, M.S., Sokolova, O.A., Yakushev, E.Y., Stolyarenko, A.D., Mikhaleva, E.A., Lavrov, S.A., and Gvozdev, V.A. (2011). Separation of stem cell maintenance and transposon silencing functions of Piwi protein. Proc. Natl. Acad. Sci. USA *108*, 18760–18765.

Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., Asada, N., Kojima, K., Yamaguchi, Y., Ijiri, T.W., 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.

Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat. Rev. Genet. *11*, 204–220.

Lee, J.T. (2011). Gracefully ageing at 50, X-chromosome inactivation becomes a paradigm for RNA and chromatin control. Nat. Rev. Mol. Cell Biol. *12*, 815–826.

Le Thomas, A., Rogers, A.K., Webster, A., Marinov, G.K., Liao, S.E., Perkins, E.M., Hur, J.K., Aravin, A.A., and Toth, K.F. (2013). Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. Genes Dev. *27*, 390–399.

Liu, Y., Mochizuki, K., and Gorovsky, M.A. (2004). Histone H3 lysine 9 methylation is required for DNA elimination in developing macronuclei in Tetrahymena. Proc. Natl. Acad. Sci. USA *101*, 1679–1684.

Liu, Y., Taverna, S.D., Muratore, T.L., Shabanowitz, J., Hunt, D.F., and Allis, C.D. (2007). RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in Tetrahymena. Genes Dev. *21*, 1530–1545.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J.M., Wolf, Y.I., Yakunin, A.F., et al. (2011). Evolution and classification of the CRISPR-Cas systems. Nat. Rev. Microbiol. *9*, 467–477.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., Dicarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823–826.

Malone, C.D., and Hannon, G.J. (2009). Small RNAs as guardians of the genome. Cell *136*, 656–668.

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.

Marraffini, L.A., and Sontheimer, E.J. (2010). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat. Rev. Genet. *11*, 181–190.

Matzke, M.A., and Birchler, J.A. (2005). RNAi-mediated pathways in the nucleus. Nat. Rev. Genet. *6*, 24–35.

Mochizuki, K. (2010). DNA rearrangements directed by non-coding RNAs in ciliates. Wiley Interdiscip. Rev. RNA *1*, 376–387.

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.

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.

Mohammad, F., Mondal, T., Guseva, N., Pandey, G.K., and Kanduri, C. (2010). Kcnq1ot1 noncoding RNA mediates transcriptional gene silencing by interacting with Dnmt1. Development *137*, 2493–2499.

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.

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

Nowacki, M., Shetty, K., and Landweber, L.F. (2011). RNA-mediated epigenetic programming of genome rearrangements. Annu. Rev. Genomics Hum. Genet. *12*, 367–389.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247–257.

Pandey, R.R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D., and Kanduri, C. (2008). Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol. Cell *32*, 232–246.

Ponting, C.P., Oliver, P.L., and Reik, W. (2009). Evolution and functions of long noncoding RNAs. Cell *136*, 629–641.

Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. Annu. Rev. Biochem. *81*, 145–166.

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.

Rozhkov, N.V., Hammell, M., and Hannon, G.J. (2013). Multiple roles for Piwi in silencing Drosophila transposons. Genes Dev. *27*, 400–412.

Schmitz, K.M., Mayer, C., Postepska, A., and Grummt, I. (2010). Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. Genes Dev. *24*, 2264–2269.

Schorderet, P., and Duboule, D. (2011). Structural and functional differences in the long non-coding RNA hotair in mouse and human. PLoS Genet. *7*, e1002071. http://dx.doi.org/10.1371/journal.pgen.1002071.

Shpiz, S., Olovnikov, I., Sergeeva, A., Lavrov, S., Abramov, Y., Savitsky, M., and Kalmykova, A. (2011). Mechanism of the piRNA-mediated silencing of Drosophila telomeric retrotransposons. Nucleic Acids Res. *39*, 8703–8711.

Sienski, G., Dönertas, D., and Brennecke, J. (2012). Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. Cell *151*, 964–980.

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.

Taverna, S.D., Coyne, R.S., and Allis, C.D. (2002). Methylation of histone h3 at lysine 9 targets programmed DNA elimination in tetrahymena. Cell *110*, 701–711.

Tsai, M.-C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. Science *329*, 689–693.

Ulitsky, I., Shkumatava, A., Jan, C.H., Sive, H., and Bartel, D.P. (2011). Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. Cell *147*, 1537–1550.

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

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

Wang, S.H., and Elgin, S.C.R. (2011). Drosophila Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. Proc. Natl. Acad. Sci. USA *108*, 21164–21169.

Wang, K.C., Yang, Y.W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B.R., Protacio, A., Flynn, R.A., Gupta, R.A., et al. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature *472*, 120–124.

Wei, W., Ba, Z., Gao, M., Wu, Y., Ma, Y., Amiard, S., White, C.I., Rendtlew Danielsen, J.M., Yang, Y.-G., and Qi, Y. (2012). A role for small RNAs in DNA double-strand break repair. Cell *149*, 101–112.

Wiedenheft, B., Sternberg, S.H., and Doudna, J.A. (2012). RNA-guided genetic silencing systems in bacteria and archaea. Nature *482*, 331–338.

Wierzbicki, A.T., Haag, J.R., and Pikaard, C.S. (2008). Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. Cell *135*, 635–648.

Yap, K.L., Li, S., Muñoz-Cabello, A.M., Raguz, S., Zeng, L., Mujtaba, S., Gil, J., Walsh, M.J., and Zhou, M.-M. (2010). Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. Mol. Cell *38*, 662–674.

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.

Zahler, A.M., Neeb, Z.T., Lin, A., and Katzman, S. (2012). Mating of the stichotrichous ciliate Oxytricha trifallax induces production of a class of 27 nt small RNAs derived from the parental macronucleus. PLoS ONE *7*, e42371. http://dx.doi.org/10.1371/journal.pone.0042371.

Zhao, J., Ohsumi, T.K., Kung, J.T., Ogawa, Y., Grau, D.J., Sarma, K., Song, J.J., Kingston, R.E., Borowsky, M., and Lee, J.T. (2010). Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol. Cell *40*, 939–953.

Note Added in Proof

Since the submission of this manuscript, several additional papers have demonstrated the use of Cas9 as a programmable nuclease for genome editing in multiple organisms. These include the following:

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. Elife 2, e00471. http://dx.doi.org/10.7554/eLife.00471.

Cho, S.W., Kim, S., Kim, J.M., and Kim, J.-S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat. Biotechnol. Published online January 29, 2013. http://dx.doi.org/10.1038/nbt.2507.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.-R.J., and Joung, J.K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat. Biotechnol. Published online January 29, 2013. http://dx.doi.org/10.1038/nbt.2501.

Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L.A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat. Biotechnol. Published online January 29, 2013. http://dx.doi.org/10.1038/nbt.2508.