Elements and machinery of non-coding RNAs: toward their taxonomy

Tetsuro Hirose¹, Yuichiro Mishima^{2,3} & Yukihide Tomari^{2,3,*}

Abstract

Although recent transcriptome analyses have uncovered numerous non-coding RNAs (ncRNAs), their functions remain largely unknown. ncRNAs assemble with proteins and operate as ribonucleoprotein (RNP) machineries, formation of which is thought to be determined by specific fundamental elements embedded in the primary RNA transcripts. Knowledge about the relationships between RNA elements, RNP machinery, and molecular and physiological functions is critical for understanding the diverse roles of ncRNAs and may eventually allow their systematic classification or “taxonomy.” In this review, we catalog and discuss representative small and long non-coding RNA classes, focusing on their currently known (and unknown) RNA elements and RNP machineries.

Keywords CRISPR; lncRNA; miRNA; piRNA; siRNA

DOI 10.1002/embr.201338390 | Received 19 December 2013 | Revised 4 March 2014 | Accepted 10 March 2014 | Published online 14 April 2014

EMBO Reports (2014) 15: 489–507

Introduction

Recent studies have uncovered numerous ncRNAs, which account for a large proportion of eukaryotic transcriptomes [1]. ncRNAs are literally defined as RNAs that do not code proteins, that is, all RNAs except mRNAs. These include traditional classes of ncRNAs such as rRNAs, tRNAs, snRNAs, as well as more recently discovered classes such as microRNAs (miRNAs) and long non-coding RNAs (lncRNAs). Prokaryotes also have their own ncRNAs including Hfq-binding small RNAs and CRISPR RNAs (crRNAs). In eukaryotes, many ncRNAs are first transcribed by RNA polymerase II (RNAPII), and as such, their primary transcripts are thought to be capped and poly(A)-tailed just like mRNAs. Yet, some transcripts are selectively processed into miRNAs to exert post-transcriptional gene silencing, while others are retained in the nucleus to form specific foci and/or mediate epigenetic regulation. What then determines their diverse fates and functions? This question seems analogous to asking what determines diverse fates and functions

of proteins, which are all synthesized by the ribosome. Like domains and motifs harbored in proteins, RNAs should also embed some fundamental features or “RNA elements”—such as RNA sequences, structures, and chemical modifications—determining their various roles. However, there is one big difference between proteins and RNAs; while domains and motifs of proteins can usually function by themselves, elements of ncRNAs almost always need their protein partners to form operating “RNP machineries.”

Roughly speaking, eukaryotic ncRNAs can be divided into small RNAs (~20–30 nucleotides (nt); e.g., miRNAs), intermediate-sized RNAs (~30–200 nt; e.g., snRNAs), and lncRNAs (>~200 nt, e.g., Xist RNA). RNA elements in primary transcripts of small RNAs are usually bound by RNA-processing proteins, which cleave the primary transcripts into small RNA pieces that are finally assembled into the operating RNP machinery (Fig 1). This RNP machinery is called RNA-induced silencing complex (RISC), and at the heart of RISC generally lie Argonaute (Ago) family proteins. The Ago family can be divided into the AGO subfamily that binds to miRNAs and siRNAs and the PIWI subfamily that binds to piRNAs. Both subfamilies have the common function to silence target genes complementary to their small guiding RNAs. In contrast, RNA elements in lncRNAs are often, if not always, bound by RNA-binding proteins without processing activity, which remain associated with lncRNAs to form the operating RNP machinery (Fig 1). lncRNAs participate in diverse cellular functions including chromatin modification, transcription, splicing, mRNA decay, translation, protein transport, and assembly, and their RNA elements and RNP machineries are also thought to be extremely diverse.

Despite such differences between small RNAs and lncRNAs, systematic studies on the relationship between fundamental RNA elements, operating RNP machineries, and molecular and physiological functions are important for understanding how ncRNAs work. In this review, we catalog representative small and long ncRNA classes in eukaryotes and the CRISPR system in prokaryotes, focusing on their currently known (and unknown) RNA elements and RNP machineries. For ncRNA classes not covered here, we refer the reader to other comprehensive reviews [2,3].

1 Institute for Genetic Medicine, Hokkaido University, Sapporo, Hokkaido, Japan

2 Institute of Molecular and Cellular Biosciences, The University of Tokyo, Bunkyo-ku, Tokyo, Japan

3 Department of Medical Genome Sciences, The University of Tokyo, Bunkyo-ku, Tokyo, Japan

*Corresponding author. Tel: +81 3 5841 7839; Fax: +81 3 5841 8485; E-mail: tomari@iam.u-tokyo.ac.jp

Glossary

Ago	Argonaute	NEAT1	nuclear paraspeckle assembly transcript 1
Air	antisense Igf2r RNA	NOT	negative on TATA
Aub	Aubergine	NR	nuclear receptor
C3PO	component 3 promoter of RISC	P-body	processing body
Cas	CRISPR-associated genes	PACT	protein activator of the interferon-induced protein kinase
Cascade	CRISPR-associated complex for antiviral defense	PAM	protospacer adjacent motif
CBP	CREB-binding protein	PAN2/3	poly(A)-specific ribonuclease subunit 2/3
CCR4	carbon catabolite repression 4	PAPI	partner of PIWIs
circRNA	circular RNA	piRNA	PIWI-interacting RNA
Cmr	Cas module receptor activity-modifying proteins	PIWI	P-element-induced wimpy testis
CoREST	REST corepressor 1	PRC2	polycomb repressive complex 2
CRISPR	clustered regularly interspaced short palindromic repeats	pre-miRNA	precursor miRNA
crRNA	CRISPR RNA	pri-miRNA	primary miRNA
Csm	CRISPR-Cas subtype Mtube	PSPs	paraspeckle protein components
DGCR8	DiGeorge syndrome critical region gene 8	R2D2	two dsRNA-binding domains, associated with DCR-2
EZH2	enhancer of zeste homolog 2	RdRP	RNA-dependent RNA polymerase
GAS5	growth arrest-specific 5	RIP	RNA immunoprecipitation
Gasz	germline gene containing ankyrin repeat, SAM, and basic leucine zipper domain	RISC	RNA-induced silencing complex
GR	glucocorticoid receptor	RNAi	RNA interference
Gtsf1	gametocyte-specific factor 1	RNAPII	RNA polymerase II
GW182	glycine-tryptophan protein of 182 KDa	RNP	ribonucleoprotein
H3K27me3	trimethylate histone H3 at lysine 27	rRNA	ribosomal RNA
HAP	hnRNP A1-interacting protein	RRS	ribosome release score
HAT	histone acetyltransferase	SatIII	satellite III
Hen1	Hua enhancer1	SF2/ASF	pre-mRNA-splicing factor 2/alternative splicing factor 1
hnRNP	heterogeneous nuclear ribonucleoprotein	siRNA	small interfering RNA
HOTAIR	HOX antisense intergenic RNA	SMD	Staufen-mediated mRNA decay
Hottip	HOXA transcript at the distal tip	snoRNA	small nucleolar RNA
HOXA	homeobox A	snRNA	small nuclear RNA
HP1	heterochromatin protein 1	SRA	steroid receptor RNA activator
Hsc70	heat shock cognate protein 70	STAU1	Staufen 1
Hsp90	heat shock protein 90	SUZ12	suppressor of zeste 12 homolog
Igf2r	insulin-like growth factor 2 receptor	tasiRNA	trans-acting siRNA
IGS lncRNA	ribosomal intergenic spacer long non-coding RNA	Tdrkh	Tudor and KH domain-containing protein
Kcnq1	potassium voltage-gated channel, KQT-like subfamily, member 1	TINCR	terminal differentiation-induced ncRNA
Kcnq1ot1	KCNQ1 opposite strand/antisense transcript 1	TLS	translocated in liposarcoma
lncRNA	long non-coding RNA	TNRC6	trinucleotide repeat containing 6
Loqs-PB	PB isoform of loquacious	tracrRNA	trans-acting CRISPR RNA
Loqs-PD	PD isoform of loquacious	TRBP	trans-activation response RNA-binding protein
LSD1	lysine-specific demethylase 1A	tRNA	transfer RNA
Malat-1	metastasis-associated lung adenocarcinoma transcript 1	UAP56	U2AF65-associated protein
miRNA	microRNA	UPF1	up-frameshift 1
MitoPLD	mitochondrial phospholipase D	UPF2	up-frameshift 2
ncRNA	non-coding RNA	XIST	X-inactive specific transcript
		Yb	female sterile (1) Yb
		1/2-sbsRNA	half-STAU1-binding site RNA

microRNAs

Functions

First discovered as a peculiar form of gene regulation in *Caenorhabditis elegans* [4,5], miRNAs are now recognized as a major class of small RNAs encoded in the genome of animals, plants, some unicellular eukaryotes, and their viruses [6]. As of March 2014, more than 24,000 miRNAs have been registered in the miRNA database, miRBase (<http://www.mirbase.org>), and the human genome alone is thought to encode more than 2,000 miRNAs. Mature miRNAs are ~22-nt single-stranded RNAs with 5' monophosphate, which bind to complementary sequences in mRNAs and direct silencing of protein synthesis from them. Most animal miRNAs are only partially

complementary to their targets; base pairing within a limited region of 6–7 nt in length at the 5' end of the miRNA, termed the “seed,” makes a disproportionately large contribution to determine target selection [7]. In addition, some non-seed target sites with base pairing in the central and/or 3' region(s) of the miRNA are also reported [8,9]. The imperfect base pairing with mRNAs allows a single miRNA to regulate hundreds of targets at both mRNA and protein levels [10–12]. Many miRNAs and their complementary target mRNAs are conserved across species and comprise a complex gene regulatory network [6]. Yet, genetic studies sometimes reveal that only a few among many miRNA targets are in fact responsible for obvious phenotypes (e.g., [13]), obscuring the definition of bona fide miRNA targets [14].

Figure 1. Formation of operating RNP machineries for small RNAs and lncRNAs.

For both small RNAs and lncRNAs, specific RNA elements that determine their fate and functions are thought to be embedded in their primary transcripts. In the case of small RNAs, those RNA elements are recognized and cleaved by RNA-processing proteins into small RNA fragments, which are finally assembled into the operating RNP machinery. In contrast, RNA-binding proteins often remain associated with RNA elements in lncRNAs to form the RNP machinery.

RNA elements

miRNAs are encoded either in their own genes or in introns (or sometimes exons) of regular protein-coding genes, and are transcribed by RNAPII as long pri-miRNAs [15–17]. Pri-miRNAs form one or more hairpin structures, each of which typically contains a stem region of ~33 base pairs (bp) and a terminal loop of ~10 nt on average [18] (Fig 2A i). The hairpin structure acts as the fundamental RNA element of miRNAs and is first recognized by the nuclear RNase III enzyme Drosha and its cofactor DGCR8 in mammals (Pasha in insects and worms) [19,20,20–23] (Fig 2A ii). Drosha and DGCR8, also known as Microprocessor, cleaves the stem of the hairpin at ~11 bp away from the junction between the stem and the flanking single-stranded regions, liberating the ~22-bp short hairpin structure with a 2-nt 3' overhang, called pre-miRNA [18]. Pre-miRNAs are then exported by Exportin 5 from the nucleus to the cytoplasm [24–27], where they are recognized and further cleaved by another RNase III enzyme Dicer and its partner protein TRBP or PACT in mammals [28–31] (Fig 2A ii). Flies have a specialized Dicer, Dicer-1, and its partner protein Loqs-PB for processing of pre-miRNAs [32–35]. The resultant ~22-nt double-stranded RNAs with 2-nt 3' overhangs at both ends are termed miRNA/miRNA* duplexes, which are ready for RISC assembly. However, these processing steps by the RNase III enzymes can be skipped in some cases, diversifying the miRNA biogenesis pathways. For example, pre-miRNA-like hairpins can be generated by splicing and debranching of introns independently of Drosha processing [36,37]. In contrast, the hairpin of pre-miR-451 is too short to be processed by Dicer but instead directly assembles into RISC [38,39]. Chemically synthesized miRNA/miRNA* duplexes can properly form RISC

independently of Drosha or Dicer, as long as they are appropriate in size, structure, and chemistry.

RNP machinery

After dicing, miRNA/miRNA* duplexes are loaded into AGO proteins as double-strand RNAs [40,41] (Fig 2A iii). In mammals, miRNA/miRNA* duplexes are incorporated into all Ago1–4 proteins [42,43], while they are mainly but not exclusively sorted into Ago1 in flies [44–46]. These AGO proteins strictly recognize the 5' monophosphate on the miRNA strand and prefer uridine as the 5'-end nucleotide [47–50]. Loading of miRNA/miRNA* duplexes into AGO proteins does not occur autonomously but requires ATP and the Hsc70/Hsp90 chaperon complex [51–53], which likely support a productive conformation of AGO for duplex loading [54]. Subsequently, the miRNA* strand is expelled from AGO in a manner independent of ATP or the chaperone complex [40,41,51,53], producing mature RISC composed of AGO and the single-stranded miRNA. The seed region of the miRNA embedded in AGO is arranged in an ordered quasi-helical geometry suitable for target RNA binding, explaining why this region plays the central role in target recognition [55–58]. AGO proteins (Ago1–4 in mammals and Ago1 [but not Ago2] in flies) directly interact with the P-body protein TNRC6 in mammals (GW182 in flies), which recruits the CCR4-NOT and PAN2-PAN3 deadenylase complexes [59–63] (Fig 2A iv). This way, miRNAs promote shortening of the poly(A) tail and decay of their target mRNAs. In addition, miRNAs mediate the repression of translation initiation independently of deadenylation [64–72], but the underlying molecular mechanism remains unclear.

Small interfering RNAs

Functions

Small interfering RNAs (siRNAs) are another well-characterized class of small ncRNAs. Like miRNAs, siRNAs are eventually matured into ~21-nt-long single-stranded RNAs with 5' monophosphate [73]. Unlike miRNAs, however, siRNAs usually pair with target mRNAs with perfect complementarity and direct their

endonucleolytic cleavage and destruction, a phenomenon called RNAi [74]. In insects and plants, siRNAs are naturally made from genetic invaders such as viruses, transgenes, and transposable elements, and RNAi is critical for protection against those invaders [75]. The antiviral role of mammalian RNAi had long been elusive, until evidence for its existence was recently reported [76,77]. Endogenous siRNAs are also produced from double-stranded RNA regions of structured loci, bidirectional transcripts, and even mRNAs

Figure 2. Elements and machinery for representative small RNA classes.

(A) miRNAs are transcribed as long pri-miRNAs containing hairpin structures as an RNA element (i, red). The hairpin structure is cleaved by the nuclear RNase III enzyme Drosha at the stem (ii, dark blue) and by the cytoplasmic RNase III enzyme Dicer at the loop (ii, light blue), generating a ~22-nt miRNA/miRNA* duplex (ii, asterisk). The duplex is loaded into AGO proteins (iii, orange) and the passenger strand (miRNA*) is ejected from AGO, producing mature RISC containing the guide strand (miRNA) as an RNP machinery. RISC binds mRNAs containing the seed-matched sites and recruits GW182/TNRC6 (light green) and CCR4–NOT complex (dark green) to induce translation repression, deadenylation, and mRNA decay (iv). (B) In the siRNA pathway, long double-stranded RNAs act as an RNA element (i, red), which is recognized and cleaved by Dicer (ii, light blue), producing ~21-nt double-stranded siRNAs (ii, asterisk). Upon loading of the siRNAs duplex into catalytic AGO (iii, yellow), the passenger strand is expelled and mature RISC containing the guide strand is formed as an RNP machinery. RISC mediates RNAi by cleaving perfectly complementary target RNAs (iv). (C) Primary piRNAs are transcribed from piRNA clusters as long, single-stranded piRNA precursor transcripts that harbor unknown or no apparent RNA elements (i). piRNA precursor transcripts are processed by nucleases including Zucchini (ii, turquoise), generating piRNA precursors of intermediate size (ii, asterisk). The cleaved RNA containing uridine at the 5' end is preferentially loaded into a PIWI protein (iii, yellow-green) followed by 3' end trimming by a Trimmer (blue). The resultant piRNA-containing RISC acts as an RNP machinery and silences transposon RNAs by post-transcriptional cleavage and transcriptional silencing (iv). (D) During secondary piRNA biogenesis, RNA sequences that are complementary to primary piRNAs act as RNA elements (i, red) and are cleaved by primary piRNA-containing PIWI (e.g., Ago3; blue) and mature into secondary piRNAs by 3' end trimming (iii). Cleavage of the original piRNA precursor transcript by the secondary piRNA drives the ping-pong cycle of the piRNA amplification loop (iv). (E–G) crRNAs are transcribed as long CRISPR transcripts (pre-CRISPR) containing repetitive sequences as RNA elements (i, red). The repeat is cleaved by specific Cas6 nucleases (types I and III) or tracrRNA and RNase III (type II) (ii, turquoise and purple, respectively). Cleaved crRNAs containing a protospacer and a part of the repeat (ii, asterisk) may be trimmed by unknown nucleases and form a complex with effector Cas proteins (iii) that act as an RNP machinery with sequence-specific nuclease activity (iv).

pairing with their antisense transcripts (e.g., pseudogenes) and may play a role in the regulation of mRNA expression [78]. From a practical point of view, induction of RNAi by exogenous, synthetic siRNA duplexes can effectively silence virtually any gene in many organisms, and chemically synthesized siRNAs are used as convenient experimental tools and potential therapeutic agents.

RNA elements

siRNAs are naturally generated as ~21-nt double-stranded RNAs with 2-nt 3' overhangs from long double-stranded RNAs [79], which can thus be defined as the fundamental RNA elements for siRNA biogenesis (Fig 2B i). In flies, a specialized Dicer, Dicer-2, and its partner proteins R2D2 or Loqs-PD generate siRNA duplexes from long double-stranded RNAs [32,80–85], while mammals use the single Dicer and its partner proteins TRBP or PACT for both miRNA and siRNA production [31,86–88] (Fig 2B ii). Mammalian Dicer is more efficient in processing pre-miRNA hairpins than long double-stranded RNAs [89]. However, mouse oocytes express a specific Dicer isoform lacking the N-terminal helicase domain, which is highly active in siRNA production [90]. On the other hand, in adult *C. elegans* soma, Dicer is proteolytically truncated and the resultant small Dicer isoform enhances exogenous RNAi [91]. After dicing, siRNA duplexes are assembled into RISC. Chemically synthesized siRNA duplexes can also form RISC by bypassing the dicing step [92].

RNP machinery

Like miRNAs, siRNAs are also loaded into AGO subfamily proteins as double strands (Fig 2B iii) in a manner dependent on the guide strand 5' monophosphate and the action of the Hsc70/Hsp90 chaperone complex [51–53, 93,94]. While siRNA duplexes are incorporated into all Ago1–4 proteins in mammals [42,43], they are almost exclusively sorted into Ago2 in flies [44–46]. Fly Ago2 requires Dicer-2—the same protein required for siRNA production—and R2D2 for loading of siRNA duplexes [80,81,95], in addition to the chaperone complex. In contrast, Dicer and TRBP are not essential for duplex loading in mammals [96,97], although they may tune the loading process [98]. When the siRNA duplex is loaded into mammalian or fly Ago2, the passenger strand is cleaved at the phosphodiester bond between nucleotides 9 and 10 (across from nucleotides 10 and 11 of the siRNA guide strand) by the endonucleolytic

“slicer” activity of Ago2, as if it were the first target RNA for Ago2 [99–102]. The cleaved passenger strand is discarded from Ago2 independently of ATP and the chaperone machinery, producing the mature RISC containing the single-stranded guide strand [41,51]. The Translin–Trax endonuclease complex, also known as C3PO, is reported to accelerate this slicer-dependent RISC maturation process [97,103]. In the absence of passenger strand cleavage (e.g., when bound to human Ago1, 3 or 4, which lack slicer activity), the two strands of the siRNA duplex can still be separated, albeit slowly [41,99,104]. In flies, Ago2-loaded siRNA guide strands are 2'-O-methylated at their 3' ends by the methyltransferase Hen1 [105], which protects them from degradation upon target binding [106]. Mature RISC containing slicer-competent Ago2 mediates RNAi by cleaving perfectly complementary target RNAs [42–44,107] (Fig 2B iv). siRNAs with central mismatches to the target site or siRNAs bound to slicer-deficient AGO proteins do not cleave but can still silence their target mRNAs [108,109], most likely through miRNA-like mechanisms such as translational repression and deadenylation. Conversely, natural miRNAs loaded into Ago2 can cleave perfectly complementary target RNAs [43,110,111]. In this regard, operating RNP machineries for siRNAs and miRNAs share many common features, compared to the considerable differences in the upstream small RNA biogenesis pathways.

PIWI-interacting RNAs

Functions

Mobile genetic elements such as transposons are potential threats to genome integrity in living organisms. In animals, small RNAs called piRNAs play a pivotal role in protecting the germline genome from transposons [118–120]. piRNAs are abundantly expressed in gonads and are longer than AGO-interacting small RNAs, ranging from 20 to 35 nt in length. Unlike miRNAs, the sequences of piRNAs are highly heterogeneous and not well conserved across species. Intriguingly, however, many piRNAs are complementary to genomic regions in which transposon-related repetitive sequences are encoded [121–124]. piRNAs arise from a subset of these genomic loci, called piRNA clusters, where multiple copies of transposon remnants reside in sense and antisense orientation. Based on genetic, biochemical, and bioinformatic studies, it is proposed that

piRNAs derived from piRNA clusters (primary piRNAs) act as guide molecules to silence complementary transposon RNAs either by post-transcriptional cleavage or by transcriptional silencing [118–120]. Target cleavage also triggers the production of secondary piRNAs, thereby boosting the amount of piRNAs, via the so-called ping-pong amplification loop [125,126]. Loss or reduction of piRNAs in nematode, fly, zebrafish, and mouse results in up-regulation of transposon RNAs in the germline [125,127–130]. As a result, transposon insertions, DNA damage and apoptosis are increased and gametogenesis is compromised [127,129,131,132].

Box 1: Secondary siRNAs

siRNAs are not only made from exogenous or endogenous double-stranded RNAs; they can be amplified through the action of RNA-dependent RNA polymerases (RdRPs). In plants, worms, and fungi (but not in insects and mammals), target binding of siRNAs or target cleavage triggers the recruitment of RdRPs, often assisted by their co-factors. RdRPs use the target RNAs as templates to synthesize long double-stranded RNAs, which are then processed into a new pool of siRNAs (“secondary siRNAs”) by Dicer proteins. In worms, some RdRPs can directly synthesize ~22-nt secondary siRNAs with 5′ triphosphate in a Dicer-independent manner [112–114]. Production of secondary siRNAs can be triggered by small RNA species other than siRNAs. An extreme case is known as tasiRNAs in plants, the production of which is directed by miRNA-mediated cleavage of long ncRNAs [115]. In *C. elegans*, 21U RNAs whose characteristics are similar to insect and mammalian piRNAs trigger RdRP-dependent production of secondary siRNAs called 22G RNAs [116,117]. These examples highlight how distinct pathways of ncRNAs intersect and interact with each other, illustrating their remarkable complexity.

RNA elements

Primary piRNAs are transcribed from piRNA clusters by RNAPII as single-stranded long RNAs (piRNA precursor transcripts) [125,133]. piRNA precursor transcripts are then processed into intermediate-sized RNAs by a Dicer-independent mechanism [122,127]. Because piRNA precursor transcripts lack consensus structures and sequence elements except for the complementarity to transposons, how piRNA precursor transcripts are distinguished from other cellular mRNAs has not been revealed (Fig 2C i). Recently, the mitochondrial protein Zuc, also known as MitoPLD in mammals, has been identified as an RNA-processing protein required for piRNA biogenesis [134–138] (Fig 2C ii). Yet, Zuc cleaves single-stranded RNA with no apparent sequence preference *in vitro* [137,138], leaving the fundamental RNA elements in piRNA precursor transcripts unclear. *In vivo*, Zuc might recognize as-yet-uncharacterized elements embedded in piRNA precursor transcripts with the help of co-factors such as the DEAD box helicase Armitage, Yb, and Gasz [134,135,139,140]. Alternatively, piRNA precursor transcripts might be determined by their genomic origin rather than an RNA element embedded in the piRNA precursor transcript itself. In fly ovaries, a subset of bidirectional piRNA clusters is bound by Rhino, a variant of heterochromatin protein HP1, and requires Rhino for efficient piRNA precursor transcription [141]. A nuclear DEAD box helicase UAP56 co-localizes with Rhino and binds piRNA precursor transcripts that are derived from the genomic piRNA clusters [142]. Interestingly, UAP56 localizes in close proximity to the perinuclear

cytoplasmic compartment called nuage, where piRNA biogenesis and amplification are thought to take place [142,143]. A tempting model is that piRNA precursor transcripts are marked co-transcriptionally by RNA-binding proteins such as UAP56 and are selectively transferred to the piRNA biogenesis machinery. However, UAP56 is not a piRNA-specific factor but is also involved in pre-mRNA splicing and mRNA export [144]. Transcription factors that direct piRNA precursor expression also appear to be rather general and variable among species [133,145,146]. Indeed, some piRNAs are generated even from protein-coding mRNAs [147,148]. Therefore, with the current knowledge, it is difficult to define a universal RNA element that discriminates piRNA precursors from other transcripts.

RNP machinery

As their name suggests, piRNAs form the operating RNP machinery with PIWI proteins, an animal gonad-specific subfamily of Ago proteins [149]. Unlike siRNAs and miRNAs, piRNAs are thought to be loaded into PIWI as single-stranded RNAs (Fig 2C iii). Based on *in vitro* experimental evidence, putative precursor piRNAs generated by Zuc processing supposedly have a random nucleotide at their 5′ end [137,138]. In contrast, mature piRNAs naturally tend to have uridine at the 5′ end according to deep sequencing data [125,126]. This bias is most likely produced by preferential loading of precursor piRNAs containing 5′ U into cognate PIWI proteins [150] (Fig 2C iii). 5′ U may be a sequence constraint during the formation of the RNP machinery. After being loaded into PIWI proteins, precursor piRNAs are trimmed from the 3′ end by an exonuclease called Trimmer whose identity remains unknown (Fig 2C iii) [150]. This reaction, which is modulated by the mitochondrial protein PAPI in insects or Tdrkh in mice [151,152], is coupled with methylation by Hen1, generating 22- to 30-nt mature piRNAs with a 2′-O-methyl group at their 3′ end [105,150,153,154]. The resultant piRNA-containing RISC acts as a sequence-specific RNA-binding module to induce post-transcriptional and transcriptional silencing of transposon RNAs (Fig 2C iv).

The post-transcriptional silencing of transposon RNAs is mediated by PIWI's slicer activity, which is a conserved feature of Ago family proteins [126,149,155]. In addition, the slicer activity is important for amplifying piRNAs via the “ping-pong” cycle [125,126]. For example in *Drosophila melanogaster* oocytes, primary antisense piRNAs that are loaded into Aubergine (Aub), one of the fly PIWI proteins, cleave complementary sense transposon RNAs (Fig 2D i and ii). The cleaved fragment with a 5′ monophosphate is incorporated by another PIWI protein, Ago3, and matured as a secondary piRNA presumably by 3′ end trimming (Fig 2D-iii). The Ago3-loaded sense piRNA then cleaves the original antisense piRNA precursor transcript (Fig 2D iv) [125,126]. This cyclic process is thought to take place in nuage with help of the conserved germline-specific RNA helicase Vasa and Tudor domain-containing proteins [156,157]. As a consequence, the amount of piRNAs that are complementary to active transposon RNAs is amplified [125,126,130]. Therefore, piRNA-RISC can also act as an RNA-processing complex in the secondary piRNA biogenesis pathway.

Loss of piRNAs or PIWI proteins not only stabilizes transposon RNAs but also reduces repressive epigenetic marks and activates transcription at transposon loci [158–163]. In addition, some PIWI proteins translocate into the nucleus [162,164], indicating a

function of piRNA-RISC in the transcriptional silencing process, although the exact action of nuclear piRNA-RISC needs to be clarified. So far, it has been shown that *Drosophila* Piwi functionally and genetically interacts with several nuclear factors including heterochromatin protein HP1a [160,165,166], zinc finger protein Gtsf1/Asterix [167–169], and Maelstrom [161]. Recent genome-wide screenings for novel piRNA pathway components in *Drosophila* identified additional candidate genes with nuclear or cytoplasmic functions [139,140,169]. Future studies on these factors will fill in the gaps and provide a more complete picture of the piRNA machinery that maintains the integrity of the germline genome.

CRISPR-derived RNAs

Functions

Analogous to piRNAs in animals, bacteria and archaea have a unique genome defense mechanism, and small RNAs lie at its core. CRISPR, together with Cas proteins, forms a widespread DNA segment found in about half of all bacteria and approximately 90% of archaea [170]. CRISPR are characterized by the presence of 20–50-nt repetitive sequences separated by spacer sequences [171,172]. Each spacer is unique and diverse in sequence, but shares high homology with plasmids and viral genomes that the host cells encounter in the environment [173–175]. Upon infection of viruses and phages, a short fragment of the invading DNA called protospacer is excised either upstream or downstream of a DNA motif called PAM [176,177]. The protospacer is then integrated into the 5' end of the CRISPR following a repeat sequence, adding a new repeat–spacer pair as a memory of infection [178–180]. The CRISPR locus is transcribed from the 5' leader sequence and processed into small RNAs called CRISPR-derived RNAs (crRNAs), each of which contains a single spacer sequence that acts as a guide to counteract the foreign DNA [172,181]. Hence, the CRISPR-Cas system is an RNA-based adaptive immune system against foreign DNA elements in prokaryotes [179,182].

Due to its relatively simple structure and the ease of modification, the CRISPR-Cas system has recently been developed into an engineered nuclease that targets specific sequences in the genome [183–185]. Details of this novel genome editing technique are described elsewhere [186].

RNA elements

Similar to the biogenesis of piRNAs, multiple crRNAs are processed from long CRISPR transcripts (pre-CRISPR) containing the characteristic repetitive sequences as fundamental RNA elements (Fig 2E i, F i and G i). The biogenesis pathways of pre-CRISPR RNAs have some variations and can be divided into three types. In the type I system, each CRISPR repeat is palindromic and forms an approximately 20-nt stem-loop structure when transcribed as a single-stranded RNA (Fig 2E i) [170]. Endonucleases Cas6e/Cas6f, in complex with Cascade, recognize this stem-loop structure and cleave at the 3' end of the stem (Fig 2E ii) [187–189]. This reaction generates a mature crRNA containing a single protospacer flanked by a ~8-nt upstream sequence and a ~17- to 20-nt downstream stem-loop, both of which derived from CRISPR repeats. Similarly, in the type III system, Cas6 recognizes the single-stranded region within the repeat and cleaves

8 nt upstream from the repeat–spacer junction (Fig 2G ii) [190,191]. This is followed by trimming of the 3' end, generating 39- to 45-nt mature crRNA. In contrast, the type II system requires another small RNA termed tracrRNA for crRNA recognition and maturation (Fig 2F ii) [192]. tracrRNA is encoded upstream of the CRISPR locus in antisense orientation and contains a ~25-nt sequence complementary to the CRISPR repeat. RNase III cleaves the duplexed tracrRNA and pre-CRISPR at both strands in the presence of Cas9 [192], a core protein in this system. A part of the repeat sequence left at the 5' end of the protospacer is trimmed, generating mature crRNA containing ~40-nt protospacer sequence followed by ~20-nt sequence derived from the adjacent repeat.

RNP machinery

Similar to RISC formed by small RNAs and Ago proteins, mature crRNAs form an operating RNP machinery with Cas proteins acting as sequence-specific nucleases (Fig 2E–G iii and iv). The effector Cas proteins are diverse among the three systems described above. In the type I system, Cascade remains associated with the mature crRNA and is involved in the recognition of the homologous DNA through base pairing of the 5' end of the protospacer (Fig 2E iii) [193]. Cascade–crRNA complexes induce R-loop formation at the homologous DNA sequence [194], which recruits the single-stranded DNA nuclease Cas3 that cleaves the looped, non-complementary DNA strand (Fig 2E iv) [195,196]. crRNAs in the type II system remain incorporated in Cas9, which has HNH and RuvC nuclease domains (Fig 2F iii) [170,183]. These domains are responsible for the cleavage of the complementary and non-complementary target DNA strands, respectively, 3 nt upstream from the 3' end of the protospacer (Fig 2F iv) [183]. Cleavage by Cas9 also requires tracrRNA, indicating that Cas9–crRNA–tracrRNA complexes act as the minimal operating RNP machinery in the type II system [183,197]. crRNA and tracrRNA work as a form of chimeric guide RNA (gRNA), allowing the use of the two-component Cas9–gRNA nuclease for genome engineering [183–185]. Type III systems can be

Box 2: Judgment of coding or non-coding by ribosome profiling

Ribosome profiling is a recently developed method that utilizes digested RNA and sequencing of the portion that is bound by 80S ribosomes to give a profile of ribosome occupancy along transcripts [201]. This method was applied to experimentally analyze the protein-coding potential of lncRNAs and revealed that a large number of mammalian lncRNAs are bound by ribosomes [202], raising the possibility that these transcripts encode small proteins. Indeed, several functional small proteins are produced from transcripts that were originally categorized as lncRNAs [203–206]. Conclusions from the ribosome profiling data were revised, however, when a metric termed the ribosome release score (RRS) was considered [207]. Ribosomes that are engaged in translation are released from transcripts when they reach a stop codon. As a result, when ribosome profiling is applied to protein-coding transcripts, a sharp drop in ribosome occupancy is seen at the start of the 3' untranslated regions (UTRs). In contrast, translational termination should not occur for ncRNAs. The RRS was therefore developed to distinguish between coding and non-coding transcripts. Indeed, the RRS categorizes numerous lncRNAs together with well-established lncRNAs, which are distinguished from protein-coding mRNAs, indicating that lncRNAs generally are not translated into proteins.

divided into two subclasses, type III-A and type III-B, each of which loads crRNA to different Cas complexes (Fig 2G iii). In the type III-A system, crRNA is incorporated in the Csm complex that targets homologous DNA with help of Cas10 [182,198]. In contrast, in the type III-B system, crRNA is incorporated in the Cmr complex [199]. Interestingly, crRNA-Cmr targets RNA rather than DNA, thus representing the diversity and flexibility of the CRISPR-Cas system (Fig 2G iv) [199,200].

lncRNAs in epigenetic regulation

Functions

In 2013, the GENCODE ver. 7 catalog of human lncRNAs annotated 9,640 manually curated lncRNA loci [208]. Among them, multiple lncRNAs interact with various epigenetic regulatory factors (Fig 3A). Xist RNA, involved in X-chromosome dosage compensation in mammals, is expressed from one of the two X-chromosomes in female cells, which results in epigenetic silencing of this chromosome. Xist directly interacts with the Polycomb repressive complex 2 (PRC2) that leads to proper localization of the PRC2 to the inactive X-chromosome to trimethylate histone H3 at lysine 27 (H3K27me3) [209,210]. Analogously, human HOTAIR lncRNA physically associates with PRC2 and modulates PRC2 and H3K27me3 localization at hundreds of genomic sites [211,212]. Furthermore, RNA immunoprecipitation (RIP) with the antibody against EZH2, a subunit of PRC2, captured 9,788 RNA species including lncRNAs in mouse embryonic stem cells [213]. Air and Kcnq1ot1 lncRNAs interact with histone H3 lysine 9 methylase G9a to mediate epigenetic silencing of Igf2r and Kcnq1 loci, respectively

[214,215]. In contrast, other lncRNAs bind activating epigenetic regulatory proteins. HOTTIP and Mistral act as enhancer-like lncRNAs by interacting with the mixed lineage leukemia (MLL) complex to activate target genes in the HOXA locus [216,217]. Other enhancer-like lncRNAs, ncRNA-a7 and a5, interact with the Mediator complex for their enhancer-like functions [218]. Thus, lncRNAs possibly participate in various aspects of epigenetic regulation through interacting with various classes of epigenetic regulators including writers, erasers, and readers of histone as well as DNA marks (Fig 3A) [219,220].

RNA elements and RNP machinery

As shown in Table 1, several functional elements or regions within lncRNAs required for epigenetic regulation have been identified. Xist lncRNA has been shown to include several distinct functional elements [221–224]. PRC2 directly interacts with unique AUCG tetraloop structures present at the A-repeat region at the 5' terminus of Xist [221,222]. Since this interaction is essential for Xist lncRNA function in X-chromosome inactivation, the A-repeat can be considered as the RNA element that forms the functional machinery responsible for histone modification. EZH2 and SUZ12 in PRC2 interact with the highly structured 5' domain of HOTAIR, suggesting that this RNA structure acts as the RNA element that constitutes the functional machinery for H3K27 methylation [212]. Intriguingly, HOTAIR was shown to contain a distinct domain required for interaction with the CoREST–LSD1 complex that is responsible for demethylation of histone H3K4 [212]. The ability to tether two distinct complexes enables RNA-mediated assembly of PRC2 and LSD1 and coordinates targeting of PRC2 and LSD1 to chromatin for coupled histone H3K27 methylation and K4 demethylation. This suggests

Figure 3. Categorization of lncRNA actions.

(A) lncRNAs recruit histone modification enzyme complexes such as PRC2 to specific chromosomal loci. The structural domain(s) within the lncRNA directly interacts with a subunit of the complex (e.g., EZH2). It remains to be elucidated how the lncRNAs recognize specific chromosomal loci. (B) lncRNAs act as transcriptional regulators. As an example, the lncRNA that acts as a transcriptional co-activator required for assembly of the transcription initiation complex is shown. An RNA modification shown by a red circle as well as distinct RNA structures are critical elements for the assembly. (C) lncRNAs act as structural cores of subcellular bodies (e.g., paraspeckles). The subcellular body is formed upon the assembly of multiple lncRNA–protein subcomplexes. (D) lncRNAs directly interact with specific mRNAs by base pairing. The base-paired duplex provides a binding platform for regulatory proteins (e.g., STAU1) that modulate mRNA stability and/or translation.

Table 1. Elements and machinery of lncRNAs

RNA species	RNA elements	Element functions	RNA-binding proteins	RNP Machinery functions	Reference
XIST	Repeat A (tetraloop)	Recruitment of histone methylase	EZH2 (PRC2)	X-chromosome inactivation	[222]
	Repeat A (tetraloop)	lncRNA processing	ASF/SF2		[223]
	Repeat F	Bridge between lncRNA and DNA	YY1		[224]
HOTAIR	5' domain (100- to 300-nt stem-loops)	Recruitment of histone methylase	PRC2	Epigenetic silencing	[212]
	3' domain (1500- to 2146-nt stem-loops)	Recruitment of histone demethylase	LSD1/CoREST/REST		[212]
HOTTIP	Exons 1–2	Recruitment of histone methylase	WDR5	Epigenetic activation	[217]
Mistral	ABM (stem-loop)	Recruitment of histone methylase	MLL1	Epigenetic activation	[216]
ecCEBPA	R2, R5, R6 (stem-loop (with CpG))	Binding of DNA methylase	DNMT1	Epigenetic activation	[219]
SRA	STR7, 10, 11 (stem-loop)	Transcriptional coactivation	?	Transcriptional activation	[228]
	U206 in STR5 (pseudouridine)	Transcriptional coactivation	Pus1 p, Pus3p (isomerize U206)		[229]
GAS5	GRE-1, GRE-2 (stem-loop)	Binding of nuclear receptor	Glucocorticoid receptor	Transcriptional suppression	[226]
ncRNA _{CCND1}	-454S, -341a (GGUG)	Inhibition of histone acetyltransferase	TLS/FUS	Transcriptional suppression	[227]
NEAT1 (NEAT1_2)	ENE-like (triple helix)	lncRNA stabilization	?	Nuclear body architecture	[232, 243, 244]
	tRNA-like structure	lncRNA 3' end processing	RNase P (RNA processing)		
PAN	ENE (triple helix)	lncRNA stabilization	?	Viral gene expression	[245, 246]
MALAT1	ENE-like (triple helix)	lncRNA stabilization	?	Epigenetic activation, splicing regulation	[243, 244]
	tRNA-like structure	lncRNA 3' end processing	RNase P (RNA processing)		[242]
1/2-sbsRNA	Alu	mRNA binding	STAU1 (with target mRNA)	Control of mRNA decay or translation	[250]
TINCR	TINCR box (complementary sequence)	mRNA binding	STAU1	Control of mRNA decay or translation	[252]

that lncRNAs may serve as “scaffolds” by providing binding surfaces to assemble selective histone-modifying enzymes, thereby specifying the pattern of histone modifications on target genes [212]. No common elements are known among lncRNAs that interact with epigenetic modifying complexes.

lncRNAs in transcriptional regulation

Functions

Steroid receptor RNA activator (SRA) is a lncRNA that acts as a co-activator of nuclear receptor (NR)-dependent transcription [225]. SRA acts as a “scaffold” for other regulatory proteins for NR-dependent transcription (Fig 3B). In contrast, GAS5 lncRNA is a negative regulator of NR-dependent transcription [226]. GAS5 lncRNA binds to the DNA-binding domain of the glucocorticoid receptor (GR) and acts as a “decoy” glucocorticoid response element (GRE). As a unique example, the promoter-associated ncRNA_{CCND1} serves as a molecular “ligand” for a specific RNA-binding protein, namely TLS, causing an allosteric effect to release it from an inactive conformation

[227]. This in turn permits gene-specific interaction between TLS and the histone acetyltransferase (HAT) CBP/p300 resulting in the inhibition of HAT function and the repression of transcription.

RNA elements and RNP machinery

SRA contains multiple stem-loop structures, some of which are required for its co-activator function [228]. Those structures function as RNA elements to form the RNP machinery and may provide the binding platform for the protein factors for NR-dependent transcriptional regulation (Fig 3B). SRA was reported to interact with two RNA pseudouridine synthases, Pus1p and Pus3p, which likely pseudouridylate the SRA transcript [229]. Significantly, the pseudouridylation at U206 within a stem-loop of SRA is required for its co-activator function (Fig 3B). The significance of specific RNA structures is illustrated by the decoy function of the GAS5 lncRNA that captures the GR, thus competing with DNA GR elements for binding to the GR [226]. TLS association and HAT inhibition by the ncRNA_{CCND1} are putatively mediated by the GGUG sequences which correspond to the TLS binding sequence identified by *in vitro* selection procedures [227].

Architectural lncRNAs

Functions

Recently, a number of relatively abundant lncRNAs have been found to localize to specific subnuclear structures. The nuclear paraspeckle assembly transcript 1 (NEAT1) lncRNA was found to localize specifically to paraspeckles where it forms an essential structural component (Fig 3C) [230–232]. The paraspeckle structure is formed with more than 40 paraspeckle protein components (PSPs), which mostly exhibited features of typical RNA-binding proteins and some of which directly interact with NEAT1. NEAT1 serves as “scaffold” to bring >40 PSPs into the paraspeckle as a huge RNP machinery (~0.36 μ m in diameter) (Fig 3C) [233,234]. The paraspeckles reportedly act to regulate gene expression by two different mechanisms through sequestration of mRNA and protein factors [235,236]. Therefore, it also acts as a functional “decoy.” Other RNA-dependent nuclear bodies have been reported. Nuclear stress bodies are formed in response to heat shock. Their formation is initiated through the synthesis of lncRNA derived from pericentric tandem repeats of satellite III (Sat III) sequences [237]. The Sat III lncRNA sequesters several splicing-related factors such as SF2/ASF and hnRNP HAP, suggesting that it affects global gene expression. Formation of a distinct subnuclear structure termed the detention center (DC) is involved in structural and functional adaptations of the nucleolus to heat shock or physiological acidosis. This process depends on the expression of the IGS lncRNA. The formation of DC redistributes nucleolar factors and arrests ribosomal biogenesis. Stress termination causes a decrease in IGS lncRNA levels and a return to the active nucleolar conformation, suggesting that IGS lncRNA regulates the structure and function of the nucleolus [238,239]. Thus, at least several lncRNAs function in mammalian cells as architectural RNAs in subnuclear structures where specific regulatory proteins are sequestered.

RNA elements and RNP machinery

The functional elements for architectural lncRNAs are not well understood. Proper synthesis and accumulation of the NEAT1_2 lncRNA isoform is essential for paraspeckle construction [231,233,240,241]. Despite it being transcribed by RNAPII, NEAT1_2 lacks a canonical poly(A) tail. The non-canonical 3' terminus is generated by RNase P-mediated cleavage of the tRNA-like structure adjacent to the 3' terminus of NEAT1_2 [232,242]. The non-polyadenylated 3' terminus forms a characteristic triple helix structure that is critical for NEAT1_2 stabilization [243,244]. Interestingly, a similar structure was first discovered as the expression and nuclear retention element (ENE) at 3' termini of the Kaposi's sarcoma-associated herpesvirus (KSHV)-encoded polyadenylated nuclear (PAN) lncRNA [245,246]. The related ENE-like elements were identified in the transcripts from six evolutionarily diverse viruses as well as in the cellular Malat-1 lncRNA [243,244,247]. Thus, the ENE-like element is considered the first family of functional lncRNA elements promoting stabilization. Extensive RNAi analysis revealed seven RNA-binding proteins that are essential for paraspeckle formation [233]: Among these, one has a role in alternative 3' end processing required for NEAT1_2 synthesis, three promote NEAT1_2 stabilization, and the remaining three function in paraspeckle assembly, suggesting that at

least three distinct elements required for paraspeckle formation may reside in the NEAT1_2 lncRNA. SatIII and IGS lncRNAs likely possess functional elements required for the formation of distinct nuclear bodies, although they remain uncharacterized. Interestingly, these lncRNAs contain repeats of short sequence stretches that may be the elements for efficient sequestration of multiple proteins on a single lncRNA molecule. This is reminiscent of the sequestration of specific RNA-binding proteins into nuclear foci by a nuclear-retained mRNA with aberrant triplet-nucleotide expansion, which causes neuromuscular diseases such as myotonic dystrophy 1 [248]. The sequestration of specific RNA-binding proteins into the nuclear foci is achieved by direct interactions with the expanded repeats.

lncRNAs that base pair with mRNAs

Functions

Besides their nuclear functions, lncRNAs can contribute to distinct regulatory events in the cytoplasm. Recent studies reported

Figure 4. Taxonomy of ncRNAs.

In a conventional view, all transcripts but protein-coding mRNAs are defined as ncRNAs in the lump (top). Understanding of fundamental RNA elements and operating RNP machineries may allow for the systematic classification or “taxonomy” of ncRNAs based on the relationships between RNA elements, RNP machinery, and molecular and physiological functions (bottom).

Box 3: Circular RNAs

Circular RNAs (circRNAs) are produced by unique head-to-tail splicing between a donor site at the 3' end of a downstream exon and an acceptor site at the 5' end of an upstream exon. Thousands of circRNAs are made from regular protein-coding genes as well as from non-coding loci, but their functions have long been unclear. Strikingly, two recent reports revealed that circRNAs can function as natural miRNA inhibitors or "sponges," with dozens of conserved miRNA-binding sites [253,254]. Because of the lack of their free ends, circRNAs are generally stable and abundantly expressed and therefore well suited to counteract miRNAs that are also abundant in cells. More generally, circRNAs with particular cis-acting elements, for example, specific motifs for RNA-binding proteins, could potentially have broader roles in cells.

cytoplasmic lncRNAs that control gene expression through base pairing with mRNAs (Fig 3D). STAU1 mediates mRNA decay (SMD) by binding to dsRNA structures in the 3' UTR of translationally active mRNAs [249]. It was found that STAU1-binding sites can be formed by base pairing between the target mRNA and a cytoplasmic polyadenylated lncRNA termed 1/2-sbsRNA. 1/2-sbsRNA thereby promotes the degradation of target mRNAs [250]. Alternatively, a recent report arguing for a general role for STAU1 in modulating translation elongation through structured mRNA coding regions [251] raised another possibility that the lncRNAs regulate translation rather than stability of the base-paired mRNAs. The lncRNA TINCR controls human epidermal differentiation through interacting with a range of differentiation-related mRNAs [252], and STAU1-deficient tissue recapitulated the impaired differentiation observed upon TINCR depletion [252]. Unlike 1/2-sbsRNA, however, TINCR directly binds to STAU1 without base pairing with other RNAs. UPF1 and UPF2, both of which are required for SMD, did not affect differentiation, suggesting that STAU1 and TINCR function is

independent of the SMD pathway. Indeed, the TINCR-STAU1 complex seems to mediate the stabilization of differentiation mRNAs, suggesting a new UPF1/2-independent role for STAU1 in differentiation [233].

RNA elements and RNP machinery

Imperfect base pairing between an Alu element in the 3' UTR of SMD-target mRNAs and the Alu element in 1/2-sbsRNA provides a binding platform for STAU1 [250]. TINCR-mRNA interaction occurs through a 25-nt "TINCR box" motif that is strongly enriched in the target mRNAs and required for TINCR binding. However, STAU1 binding to TINCR does not require target mRNAs [252], indicating that the RNP machineries of two cytoplasmic lncRNAs are built through distinct mechanisms. In contrast to the common base pairing interactions between small ncRNAs (e.g., miRNA and snoRNA) and their target RNAs, few lncRNAs including antisense transcripts have been demonstrated to directly interact with other RNAs through base pairing. The above-described lncRNAs acting as "guides" for STAU1-mediated control of target mRNAs are among those few examples.

Concluding remarks: toward taxonomy of ncRNAs

As described above, small RNA pathways are now relatively well understood, even though their actual RNA elements and RNP machineries are quite diverse. This is largely due to the remarkable commonality of their biogenesis pathways—multi-step processing of precursor transcripts into mature small RNAs—and their mode of action—guiding the effector RNP machinery to their target mRNAs via base pairing. On the other hand, such fundamental commonality is still vague among lncRNAs at this point. Thus, future characterization and categorization will be essential for a systematic understanding of the vast ncRNA world. This is an extremely challenging task akin to (or perhaps more complicated than) categorization of protein families and subfamilies, but further accumulation of knowledge will hopefully allow us to establish a "taxonomy" of ncRNAs by defining their RNA elements, RNP machinery, and molecular functions (Fig 4). To this end, recent technologies for genome-wide analyses of RNA structures [255–257], RNA modifications [258–263], and RNA–protein interactions [264–266] will be without doubt powerful tools for identifying RNA elements and RNP machineries. By integrating all those high-throughput data and experimental validations, we may eventually be able to input the information of an unknown ncRNA as a query and predict its function within minutes, just like we now routinely do for proteins.

Acknowledgments

We thank members of our laboratories for helpful discussion and critical reading of the manuscript. We are supported in part by a Grant-in-Aid for Scientific Research on Innovative Areas ("Functional machinery for non-coding RNAs") from the Japan Ministry of Education, Culture, Sports, Science and Technology and the Funding Program for Next Generation World-Leading Researchers from Japan Society for the Promotion of Science.

Sidebar A: In need of answers

- (i) Exactly how do miRNAs silence their targets? How are the contributions and kinetics of different modes of miRNA-mediated gene silencing (i.e., deadenylation and translational repression) determined?
- (ii) How can we determine best bona fide miRNA targets and the physiological outcome of their regulation?
- (iii) How can we best design and deliver siRNAs for efficient knockdown, especially *in vivo*?
- (iv) How are piRNA precursors discriminated from other transcripts? How are they processed into mature piRNAs? Exactly how do piRNAs silence their targets? What are the roles of subcellular platforms such as nuage and mitochondria for the piRNA pathway?
- (v) How are crRNAs matured?
- (vi) Are there other small RNA classes?
- (vii) What are the functions of circRNAs besides being miRNA sponges?
- (viii) How do different ncRNA pathways interact, cross-talk, and compete with each other?
- (ix) How do lncRNAs specifically interact with RNA-binding proteins?
- (x) How can we systematically determine RNA elements, RNP machineries, and functions of diverse lncRNAs? How can we categorize them?

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489: 57–74
- Vogel J, Luisi BF (2011) Hfq and its constellation of RNA. *Nat Rev Microbiol* 9: 578–589
- Matera AG, Terns RM, Terns MP (2007) Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat Rev Mol Cell Biol* 8: 209–220
- Wightman B, Ha I, Ruvkun G (1993) Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* 75: 855–862
- Lee RC, Feinbaum RL, Ambros V (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75: 843–854
- Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116: 281–297
- Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB (2003) Prediction of mammalian microRNA targets. *Cell* 115: 787–798
- Shin C, Nam JW, Farh KK, Chiang HR, Shkumatava A, Bartel DP (2010) Expanding the microRNA targeting code: functional sites with centered pairing. *Mol Cell* 38: 789–802
- Helwak A, Kudla G, Dudnakova T, Tollervey D (2013) Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* 153: 654–665
- Lewis BP, Burge CB, Bartel DP (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120: 15–20
- Selbach M, Schwanhauser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N (2008) Widespread changes in protein synthesis induced by microRNAs. *Nature* 455: 58–63
- Baek D, Villen J, Shin C, Camargo FD, Gygi SP, Bartel DP (2008) The impact of microRNAs on protein output. *Nature* 455: 64–71
- Ambros V (1989) A hierarchy of regulatory genes controls a larva-to-adult developmental switch in *C. elegans*. *Cell* 57: 49–57
- Seitz H (2009) Redefining microRNA targets. *Curr Biol* 19: 870–873
- Lee Y, Jeon K, Lee JT, Kim S, Kim VN (2002) MicroRNA maturation: step-wise processing and subcellular localization. *EMBO J* 21: 4663–4670
- Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, Kim VN (2004) MicroRNA genes are transcribed by RNA polymerase II. *EMBO J* 23: 4051–4060
- Cai X, Hagedorn CH, Cullen BR (2004) Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA* 10: 1957–1966
- Han J, Lee Y, Yeom KH, Nam JW, Heo I, Rhee JK, Sohn SY, Cho Y, Zhang BT, Kim VN (2006) Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* 125: 887–901
- Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S et al (2003) The nuclear RNase III Drosha initiates microRNA processing. *Nature* 425: 415–419
- Han J, Lee Y, Yeom KH, Kim YK, Jin H, Kim VN (2004) The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev* 18: 3016–3027
- Denli AM, Tops BB, Plasterk RH, Ketting RF, Hannon GJ (2004) Processing of primary microRNAs by the Microprocessor complex. *Nature* 432: 231–235
- Gregory RI, Yan KP, Amuthan G, Chendrimada T, Doratotaj B, Cooch N, Shiekhattar R (2004) The Microprocessor complex mediates the genesis of microRNAs. *Nature* 432: 235–240
- Landthaler M, Yalcin A, Tuschl T (2004) The human DiGeorge syndrome critical region gene 8 and its *D. melanogaster* homolog are required for miRNA biogenesis. *Curr Biol* 14: 2162–2167
- Yi R, Qin Y, Macara IG, Cullen BR (2003) Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev* 17: 3011–3016
- Bohnsack MT, Czaplinski K, Gorlich D (2004) Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA* 10: 185–191
- Lund E, Guttinger S, Calado A, Dahlberg JE, Kutay U (2004) Nuclear export of microRNA precursors. *Science* 303: 95–98
- Zeng Y, Cullen BR (2004) Structural requirements for pre-microRNA binding and nuclear export by Exportin 5. *Nucleic Acids Res* 32: 4776–4785
- Maniataki E, Mourelatos Z (2005) A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. *Genes Dev* 19: 2979–2990
- Gregory RI, Chendrimada TP, Cooch N, Shiekhattar R (2005) Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell* 123: 631–640
- Haase AD, Jaskiewicz L, Zhang H, Laine S, Sack R, Gatignol A, Filipowicz W (2005) TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. *EMBO Rep* 6: 961–967
- Lee Y, Hur I, Park SY, Kim YK, Suh MR, Kim VN (2006) The role of PACT in the RNA silencing pathway. *EMBO J* 25: 522–532
- Lee YS, Nakahara K, Pham JW, Kim K, He Z, Sontheimer EJ, Carthew RW (2004) Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* 117: 69–81
- Forstemann K, Tomari Y, Du T, Vagin VV, Denli AM, Bratu DP, Klattenhoff C, Theurkauf WE, Zamore PD (2005) Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS Biol* 3: e236
- Saito K, Ishizuka A, Siomi H, Siomi MC (2005) Processing of pre-microRNAs by the Dicer-1–Loquacious complex in *Drosophila* cells. *PLoS Biol* 3: e235
- Jiang F, Ye X, Liu X, Fincher L, McKearin D, Liu Q (2005) Dicer-1 and R3D1-L catalyze microRNA maturation in *Drosophila*. *Genes Dev* 19: 1674–1679
- Ruby JG, Jan CH, Bartel DP (2007) Intronic microRNA precursors that bypass Drosha processing. *Nature* 448: 83–86
- Okamura K, Hagen JW, Duan H, Tyler DM, Lai EC (2007) The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*. *Cell* 130: 89–100
- Cheloufi S, Dos Santos CO, Chong MM, Hannon GJ (2010) A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature* 465: 584–589
- Cifuentes D, Xue H, Taylor DW, Patnode H, Mishima Y, Cheloufi S, Ma E, Mane S, Hannon GJ, Lawson ND et al (2010) A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. *Science* 328: 1694–1698

40. Kawamata T, Seitz H, Tomari Y (2009) Structural determinants of miRNAs for RISC loading and slicer-independent unwinding. *Nat Struct Mol Biol* 16: 953–960
41. Yoda M, Kawamata T, Paroo Z, Ye X, Iwasaki S, Liu Q, Tomari Y (2010) ATP-dependent human RISC assembly pathways. *Nat Struct Mol Biol* 17: 17–23
42. Liu J, Carmell MA, Rivas FV, Marsden CG, Thomson JM, Song JJ, Hammond SM, Joshua-Tor L, Hannon GJ (2004) Argonaute2 is the catalytic engine of mammalian RNAi. *Science* 305: 1437–1441
43. Meister G, Landthaler M, Patkaniowska A, Dorsett Y, Teng G, Tuschl T (2004) Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol Cell* 15: 185–197
44. Okamura K, Ishizuka A, Siomi H, Siomi MC (2004) Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev* 18: 1655–1666
45. Forstemann K, Horwich MD, Wee L, Tomari Y, Zamore PD (2007) *Drosophila* microRNAs are sorted into functionally distinct Argonaute complexes after production by Dicer-1. *Cell* 130: 287–297
46. Tomari Y, Du T, Zamore PD (2007) Sorting of *Drosophila* small silencing RNAs. *Cell* 130: 299–308
47. Czech B, Zhou R, Erlich Y, Brennecke J, Binari R, Villalta C, Gordon A, Perrimon N, Hannon GJ (2009) Hierarchical rules for Argonaute loading in *Drosophila*. *Mol Cell* 36: 445–456
48. Okamura K, Liu N, Lai EC (2009) Distinct mechanisms for microRNA strand selection by *Drosophila* Argonautes. *Mol Cell* 36: 431–444
49. Ghildiyal M, Xu J, Seitz H, Weng Z, Zamore PD (2010) Sorting of *Drosophila* small silencing RNAs partitions microRNA* strands into the RNA interference pathway. *RNA* 16: 43–56
50. Frank F, Sonenberg N, Nagar B (2010) Structural basis for 5′ nucleotide base-specific recognition of guide RNA by human AGO2. *Nature* 465: 818–822
51. Iwasaki S, Kobayashi M, Yoda M, Sakaguchi Y, Katsuma S, Suzuki T, Tomari Y (2010) Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. *Mol Cell* 39: 292–299
52. Johnston M, Geoffroy MC, Sobala A, Hay R, Hutvagner G (2010) HSP90 protein stabilizes unloaded argonaute complexes and microscopic P-bodies in human cells. *Mol Biol Cell* 21: 1462–1469
53. Iki T, Yoshikawa M, Nishikiori M, Jaudal MC, Matsumoto-Yokoyama E, Mitsuhashi I, Meshi T, Ishikawa M (2010) *In vitro* assembly of plant RNA-induced silencing complexes facilitated by molecular chaperone HSP90. *Mol Cell* 39: 282–291
54. Kawamata T, Tomari Y (2010) Making RISC. *Trends Biochem Sci* 35: 368–376
55. Wang Y, Sheng G, Juranek S, Tuschl T, Patel DJ (2008) Structure of the guide-strand-containing Argonaute silencing complex. *Nature* 456: 209–213
56. Nakanishi K, Weinberg DE, Bartel DP, Patel DJ (2012) Structure of yeast Argonaute with guide RNA. *Nature* 486: 368–374
57. Schirle NT, MacRae IJ (2012) The crystal structure of human Argonaute2. *Science* 336: 1037–1040
58. Elkayam E, Kuhn CD, Tocilj A, Haase AD, Greene EM, Hannon GJ, Joshua-Tor L (2012) The structure of human Argonaute-2 in complex with miR-20a. *Cell* 150: 100–110
59. Behm-Ansmant I, Rehwinkel J, Doerks T, Stark A, Bork P, Izaurralde E (2006) mRNA degradation by miRNAs and GW182 requires both CCR4: NOT deadenylase and DCP1:DCP2 decapping complexes. *Genes Dev* 20: 1885–1898
60. Eulalio A, Huntzinger E, Izaurralde E (2008) GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay. *Nat Struct Mol Biol* 15: 346–353
61. Braun JE, Huntzinger E, Fauser M, Izaurralde E (2011) GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. *Mol Cell* 44: 120–133
62. Chekulaeva M, Mathys H, Zipprich JT, Attig J, Colic M, Parker R, Filipowicz W (2011) miRNA repression involves GW182-mediated recruitment of CCR4-NOT through conserved W-containing motifs. *Nat Struct Mol Biol* 18: 1218–1226
63. Fabian MR, Cieplak MK, Frank F, Morita M, Green J, Srikumar T, Nagar B, Yamamoto T, Raught B, Duchaine TF et al (2011) miRNA-mediated deadenylation is orchestrated by GW182 through two conserved motifs that interact with CCR4-NOT. *Nat Struct Mol Biol* 18: 1211–1217
64. Humphreys DT, Westman BJ, Martin DI, Preiss T (2005) MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. *Proc Natl Acad Sci USA* 102: 16961–16966
65. Pillai RS, Bhattacharyya SN, Artus CG, Zoller T, Cougot N, Basyuk E, Bertrand E, Filipowicz W (2005) Inhibition of translational initiation by let-7 MicroRNA in human cells. *Science* 309: 1573–1576
66. Wu L, Fan J, Belasco JG (2006) MicroRNAs direct rapid deadenylation of mRNA. *Proc Natl Acad Sci USA* 103: 4034–4039
67. Mishima Y, Fukao A, Kishimoto T, Sakamoto H, Fujiwara T, Inoue K (2012) Translational inhibition by deadenylation-independent mechanisms is central to microRNA-mediated silencing in zebrafish. *Proc Natl Acad Sci USA* 109: 1104–1109
68. Fukaya T, Tomari Y (2011) PABP is not essential for microRNA-mediated translational repression and deadenylation *in vitro*. *EMBO J* 30: 4998–5009
69. Fukaya T, Tomari Y (2012) MicroRNAs Mediate Gene Silencing via Multiple Different Pathways in *Drosophila*. *Mol Cell* 48: 825–836
70. Djuranovic S, Nahvi A, Green R (2012) miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. *Science* 336: 237–240
71. Bazzini AA, Lee MT, Giraldez AJ (2012) Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science* 336: 233–237
72. Bethune J, Artus-Revel CG, Filipowicz W (2012) Kinetic analysis reveals successive steps leading to miRNA-mediated silencing in mammalian cells. *EMBO Rep* 13: 716–723
73. Ghildiyal M, Zamore PD (2009) Small silencing RNAs: an expanding universe. *Nat Rev Genet* 10: 94–108
74. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391: 806–811
75. Ding SW, Voinnet O (2007) Antiviral immunity directed by small RNAs. *Cell* 130: 413–426
76. Maillard PV, Ciaudo C, Marchais A, Li Y, Jay F, Ding SW, Voinnet O (2013) Antiviral RNA interference in mammalian cells. *Science* 342: 235–238
77. Li Y, Lu J, Han Y, Fan X, Ding SW (2013) RNA interference functions as an antiviral immunity mechanism in mammals. *Science* 342: 231–234
78. Okamura K, Lai EC (2008) Endogenous small interfering RNAs in animals. *Nat Rev Mol Cell Biol* 9: 673–678
79. Bernstein E, Caudy AA, Hammond SM, Hannon GJ (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 409: 363–366

80. Pham JW, Pellino JL, Lee YS, Carthew RW, Sontheimer EJ (2004) A Dicer-2-dependent 80S complex cleaves targeted mRNAs during RNAi in *Drosophila*. *Cell* 117: 83–94
81. Liu Q, Rand TA, Kalidas S, Du F, Kim HE, Smith DP, Wang X (2003) R2D2, a bridge between the initiation and effector steps of the *Drosophila* RNAi pathway. *Science* 301: 1921–1925
82. Hartig JV, Esslinger S, Bottcher R, Saito K, Forstemann K (2009) Endo-siRNAs depend on a new isoform of loquacious and target artificially introduced, high-copy sequences. *EMBO J* 28: 2932–2944
83. Zhou R, Czech B, Brennecke J, Sachidanandam R, Wohlschlegel JA, Perrimon N, Hannon GJ (2009) Processing of *Drosophila* endo-siRNAs depends on a specific Loquacious isoform. *RNA* 15: 1886–1895
84. Miyoshi K, Miyoshi T, Hartig JV, Siomi H, Siomi MC (2010) Molecular mechanisms that funnel RNA precursors into endogenous small-interfering RNA and microRNA biogenesis pathways in *Drosophila*. *RNA* 16: 506–515
85. Hartig JV, Forstemann K (2011) Loqs-PD and R2D2 define independent pathways for RISC generation in *Drosophila*. *Nucleic Acids Res* 39: 3836–3851
86. Zhang H, Kolb FA, Brondani V, Billy E, Filipowicz W (2002) Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *EMBO J* 21: 5875–5885
87. Provost P, Dishart D, Doucet J, Frenthewey D, Samuelsson B, Radmark O (2002) Ribonuclease activity and RNA binding of recombinant human Dicer. *EMBO J* 21: 5864–5874
88. Kok KH, Ng MH, Ching YP, Jin DY (2007) Human TRBP and PACT directly interact with each other and associate with dicer to facilitate the production of small interfering RNA. *J Biol Chem* 282: 17649–17657
89. Ma E, MacRae IJ, Kirsch JF, Doudna JA (2008) Autoinhibition of human dicer by its internal helicase domain. *J Mol Biol* 380: 237–243
90. Flemr M, Malik R, Franke V, Nejepinska J, Sedlacek R, Vlahovick K, Svoboda P (2013) A retrotransposon-driven dicer isoform directs endogenous small interfering RNA production in mouse oocytes. *Cell* 155: 807–816
91. Sawh AN, Duchaine TF (2013) A truncated form of dicer tilts the balance of RNA interference pathways. *Cell Rep* 4: 454–463
92. Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* 411: 494–498
93. Nykanen A, Haley B, Zamore PD (2001) ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell* 107: 309–321
94. Miyoshi T, Takeuchi A, Siomi H, Siomi MC (2010) A direct role for Hsp90 in pre-RISC formation in *Drosophila*. *Nat Struct Mol Biol* 17: 1024–1026
95. Tomari Y, Du T, Haley B, Schwarz DS, Bennett R, Cook HA, Koppetsch BS, Theurkauf WE, Zamore PD (2004) RISC assembly defects in the *Drosophila* RNAi mutant armitage. *Cell* 116: 831–841
96. Betancur JG, Yoda M, Tomari Y (2012) miRNA-like duplexes as RNAi triggers with improved specificity. *Front Genet* 3: 127
97. Ye X, Huang N, Liu Y, Paroo Z, Huerta C, Li P, Chen S, Liu Q, Zhang H (2011) Structure of C3PO and mechanism of human RISC activation. *Nat Struct Mol Biol* 18: 650–657
98. Noland CL, Doudna JA (2013) Multiple sensors ensure guide strand selection in human RNAi pathways. *RNA* 19: 639–648
99. Matranga C, Tomari Y, Shin C, Bartel DP, Zamore PD (2005) Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* 123: 607–620
100. Miyoshi K, Tsukumo H, Nagami T, Siomi H, Siomi MC (2005) Slicer function of *Drosophila* Argonautes and its involvement in RISC formation. *Genes Dev* 19: 2837–2848
101. Leuschner PJ, Ameres SL, Kueng S, Martinez J (2006) Cleavage of the siRNA passenger strand during RISC assembly in human cells. *EMBO Rep* 7: 314–320
102. Rand TA, Petersen S, Du F, Wang X (2005) Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell* 123: 621–629
103. Liu Y, Ye X, Jiang F, Liang C, Chen D, Peng J, Kinch LN, Grishin NV, Liu Q (2009) C3PO, an endoribonuclease that promotes RNAi by facilitating RISC activation. *Science* 325: 750–753
104. Gu S, Jin L, Zhang F, Huang Y, Grimm D, Rossi JJ, Kay MA (2011) Thermodynamic stability of small hairpin RNAs highly influences the loading process of different mammalian Argonautes. *Proc Natl Acad Sci USA* 108: 9208–9213
105. Horwich MD, Li C, Matranga C, Vagin V, Farley G, Wang P, Zamore PD (2007) The *Drosophila* RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr Biol* 17: 1265–1272
106. Ameres SL, Horwich MD, Hung JH, Xu J, Gildiyal M, Weng Z, Zamore PD (2010) Target RNA-directed trimming and tailing of small silencing RNAs. *Science* 328: 1534–1539
107. Rand TA, Ginalski K, Grishin NV, Wang X (2004) Biochemical identification of Argonaute 2 as the sole protein required for RNA-induced silencing complex activity. *Proc Natl Acad Sci USA* 101: 14385–14389
108. Doench JG, Petersen CP, Sharp PA (2003) siRNAs can function as miRNAs. *Genes Dev* 17: 438–442
109. Su H, Trombly MI, Chen J, Wang X (2009) Essential and overlapping functions for mammalian Argonautes in microRNA silencing. *Genes Dev* 23: 304–317
110. Hutvagner G, Zamore PD (2002) A microRNA in a multiple-turnover RNAi enzyme complex. *Science* 297: 2056–2060
111. Yekta S, Shih IH, Bartel DP (2004) MicroRNA-directed cleavage of HOXB8 mRNA. *Science* 304: 594–596
112. Pak J, Fire A (2007) Distinct populations of primary and secondary effectors during RNAi in *C. elegans*. *Science* 315: 241–244
113. Aoki K, Moriguchi H, Yoshioka T, Okawa K, Tabara H (2007) *In vitro* analyses of the production and activity of secondary small interfering RNAs in *C. elegans*. *EMBO J* 26: 5007–5019
114. Sijen T, Steiner FA, Thijssen KL, Plasterk RH (2007) Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. *Science* 315: 244–247
115. Allen E, Xie Z, Gustafson AM, Carrington JC (2005) microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* 121: 207–221
116. Lee HC, Gu W, Shirayama M, Youngman E, Conte DJ, Mello CC (2012) *C. elegans* piRNAs mediate the genome-wide surveillance of germline transcripts. *Cell* 150: 78–87
117. Bagijn MP, Goldstein LD, Sapetschnig A, Weick EM, Bouasker S, Lehrbach NJ, Simard MJ, Miska EA (2012) Function, targets, and evolution of *Caenorhabditis elegans* piRNAs. *Science* 337: 574–578
118. Aravin AA, Hannon GJ, Brennecke J (2007) The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* 318: 761–764
119. Malone CD, Hannon GJ (2009) Small RNAs as guardians of the genome. *Cell* 136: 656–668
120. Luteijn MJ, Ketting RF (2013) PIWI-interacting RNAs: from generation to transgenerational epigenetics. *Nat Rev Genet* 14: 523–534

121. Aravin AA, Lagos-Quintana M, Yalcin A, Zavolan M, Marks D, Snyder B, Gaasterland T, Meyer J, Tuschl T (2003) The small RNA profile during *Drosophila melanogaster* development. *Dev Cell* 5: 337–350
122. Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD (2006) A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* 313: 320–324
123. Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, Morris P, Brownstein MJ, Kuramochi-Miyagawa S, Nakano T et al (2006) A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* 442: 203–207
124. Girard A, Sachidanandam R, Hannon GJ, Carmell MA (2006) A germ-line-specific class of small RNAs binds mammalian Piwi proteins. *Nature* 442: 199–202
125. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ (2007) Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 128: 1089–1103
126. Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, Siomi H, Siomi MC (2007) A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* 315: 1587–1590
127. Houwing S, Kamminga LM, Berezikov E, Cronembold D, Girard A, van den Elst H, Filippov DV, Blaser H, Raz E, Moens CB et al (2007) A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell* 129: 69–82
128. Kasper DM, Gardner KE, Reinke V (2013) Homeland security in the germ line: insights into the biogenesis and function of piRNAs. *Epigenetics* 9: 62–74
129. Carmell MA, Girard A, van de Kant HJ, Bourc'is D, Bestor TH, de Rooij DG, Hannon GJ (2007) MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev Cell* 12: 503–514
130. Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ (2007) Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* 316: 744–747
131. Kalmykova AI, Klenov MS, Gvozdev VA (2005) Argonaute protein PIWI controls mobilization of retrotransposons in the *Drosophila* male germline. *Nucleic Acids Res* 33: 2052–2059
132. Chen Y, Pane A, Schupbach T (2007) Cutoff and aubergine mutations result in retrotransposon upregulation and checkpoint activation in *Drosophila*. *Curr Biol* 17: 637–642
133. Li XZ, Roy CK, Dong X, Bolcun-Filas E, Wang J, Han BW, Xu J, Moore MJ, Schimenti JC, Weng Z et al (2013) An ancient transcription factor initiates the burst of piRNA production during early meiosis in mouse testes. *Mol Cell* 50: 67–81
134. Saito K, Ishizu H, Komai M, Kotani H, Kawamura Y, Nishida KM, Siomi H, Siomi MC (2010) Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in *Drosophila*. *Genes Dev* 24: 2493–2498
135. Olivieri D, Sykora MM, Sachidanandam R, Mechtler K, Brennecke J (2010) An *in vivo* RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in *Drosophila*. *EMBO J* 29: 3301–3317
136. Watanabe T, Chuma S, Yamamoto Y, Kuramochi-Miyagawa S, Totoki Y, Toyoda A, Hoki Y, Fujiyama A, Shibata T, Sado T et al (2011) MITOPLD is a mitochondrial protein essential for nuage formation and piRNA biogenesis in the mouse germline. *Dev Cell* 20: 364–375
137. Nishimasu H, Ishizu H, Saito K, Fukuhara S, Kamatani MK, Bonnefond L, Matsumoto N, Nishizawa T, Nakanaga K, Aoki J et al (2012) Structure and function of Zucchini endoribonuclease in piRNA biogenesis. *Nature* 491: 284–287
138. Ipsaro JJ, Haase AD, Knott SR, Joshua-Tor L, Hannon GJ (2012) The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis. *Nature* 491: 279–283
139. Handler D, Meixner K, Pizka M, Lauss K, Schmied C, Gruber FS, Brennecke J (2013) The genetic makeup of the *Drosophila* piRNA pathway. *Mol Cell* 50: 762–777
140. Czech B, Preall JB, McGinn J, Hannon GJ (2013) A transcriptome-wide RNAi screen in the *Drosophila* ovary reveals factors of the germline piRNA pathway. *Mol Cell* 50: 749–761
141. Klattenhoff C, Xi H, Li C, Lee S, Xu J, Khurana JS, Zhang F, Schultz N, Koppetsch BS, Nowosielska A et al (2009) The *Drosophila* HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. *Cell* 138: 1137–1149
142. Zhang F, Wang J, Xu J, Zhang Z, Koppetsch BS, Schultz N, Vreven T, Meignin C, Davis I, Zamore PD et al (2012) UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. *Cell* 151: 871–884
143. Lim AK, Kai T (2007) Unique germ-line organelle, nuage, functions to repress selfish genetic elements in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 104: 6714–6719
144. Linder P, Stutz F (2001) mRNA export: travelling with DEAD box proteins. *Curr Biol* 11: R961–R963
145. Cecere G, Zheng GX, Mansisidor AR, Klymko KE, Grishok A (2012) Promoters recognized by forkhead proteins exist for individual 21U-RNAs. *Mol Cell* 47: 734–745
146. Goriaux C, Desset S, Renaud Y, Vaury C, Brasset E (2014) Transcriptional properties and splicing of the flamenco piRNA cluster. *EMBO Rep* doi: 10.1002/embr.201337898
147. Saito K, Inagaki S, Mituyama T, Kawamura Y, Ono Y, Sakota E, Kotani H, Asai K, Siomi H, Siomi MC (2009) A regulatory circuit for piwi by the large Maf gene traffic jam in *Drosophila*. *Nature* 461: 1296–1299
148. Robine N, Lau NC, Balla S, Jin Z, Okamura K, Kuramochi-Miyagawa S, Blower MD, Lai EC (2009) A broadly conserved pathway generates 3'UTR-directed primary piRNAs. *Curr Biol* 19: 2066–2076
149. Meister G (2013) Argonaute proteins: functional insights and emerging roles. *Nat Rev Genet* 14: 447–459
150. Kawaoka S, Izumi N, Katsuma S, Tomari Y (2011) 3' end formation of PIWI-interacting RNAs *in vitro*. *Mol Cell* 43: 1015–1022
151. Honda S, Kirino Y, Maragkakis M, Alexiou P, Ohtaki A, Murali R, Mourelatos Z, Kirino Y (2013) Mitochondrial protein BmPAPI modulates the length of mature piRNAs. *RNA* 19: 1405–1418
152. Saxe JP, Chen M, Zhao H, Lin H (2013) Tdrkh is essential for spermatogenesis and participates in primary piRNA biogenesis in the germline. *EMBO J* 32: 1869–1885
153. Kirino Y, Mourelatos Z (2007) The mouse homolog of HEN1 is a potential methylase for Piwi-interacting RNAs. *RNA* 13: 1397–1401
154. Saito K, Sakaguchi Y, Suzuki T, Suzuki T, Siomi H, Siomi MC (2007) Pimet, the *Drosophila* homolog of HEN1, mediates 2'O-methylation of Piwi-interacting RNAs at their 3' ends. *Genes Dev* 21: 1603–1608
155. Saito K, Nishida KM, Mori T, Kawamura Y, Miyoshi K, Nagami T, Siomi H, Siomi MC (2006) Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev* 20: 2214–2222
156. Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Takamatsu K, Chuma S, Kojima-Kita K, Shiromoto Y, Asada N, Toyoda A, Fujiyama A et al (2010)

- MVH in piRNA processing and gene silencing of retrotransposons. *Genes Dev* 24: 887–892
157. Siomi MC, Mannen T, Siomi H (2010) How does the royal family of Tudor rule the PIWI-interacting RNA pathway? *Genes Dev* 24: 636–646
 158. Klenov MS, Lavrov SA, Stolyarenko AD, Ryazansky SS, Aravin AA, Tuschl T, Gvozdev VA (2007) Repeat-associated siRNAs cause chromatin silencing of retrotransposons in the *Drosophila melanogaster* germline. *Nucleic Acids Res* 35: 5430–5438
 159. Shpiz S, Olovnikov I, Sergeeva A, Lavrov S, Abramov Y, Savitsky M, Kalmykova A (2011) Mechanism of the piRNA-mediated silencing of *Drosophila* telomeric retrotransposons. *Nucleic Acids Res* 39: 8703–8711
 160. Wang SH, Elgin SC (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
 161. Sienski G, Donertas D, Brennecke J (2012) Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* 151: 964–980
 162. Aravin AA, Sachidanandam R, Bourc'is D, Schaefer C, Pezic D, Toth KF, Bestor T, Hannon GJ (2008) A piRNA pathway primed by individual transposons is linked to *de novo* DNA methylation in mice. *Mol Cell* 31: 785–799
 163. Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totoki Y, Toyoda A, Ikawa M, Asada N, Kojima K, Yamaguchi Y, Ijiri TW 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
 164. Cox DN, Chao A, Lin H (2000) piwi encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. *Development* 127: 503–514
 165. Pal-Bhadra M, Leibovitch BA, Gandhi SG, Chikka MR, Bhadra U, Birchler JA, Elgin SC (2004) Heterochromatic silencing and HP1 localization in *Drosophila* are dependent on the RNAi machinery. *Science* 303: 669–672
 166. Brower-Toland B, Findley SD, Jiang L, Liu L, Yin H, Dus M, Zhou P, Elgin SC, Lin H (2007) *Drosophila* PIWI associates with chromatin and interacts directly with HP1a. *Genes Dev* 21: 2300–2311
 167. Ohtani H, Iwasaki YW, Shibuya A, Siomi H, Siomi MC, Saito K (2013) DmGTSF1 is necessary for Piwi-piRISC-mediated transcriptional transposon silencing in the *Drosophila* ovary. *Genes Dev* 27: 1656–1661
 168. Donertas D, Sienski G, Brennecke J (2013) *Drosophila* Gtsf1 is an essential component of the Piwi-mediated transcriptional silencing complex. *Genes Dev* 27: 1693–1705
 169. Muerdter F, Guzzardo PM, Gillis J, Luo Y, Yu Y, Chen C, Fekete R, Hannon GJ (2013) A genome-wide RNAi screen draws a genetic framework for transposon control and primary piRNA biogenesis in *Drosophila*. *Mol Cell* 50: 736–748
 170. Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF et al (2011) Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* 9: 467–477
 171. Horvath P, Barrangou R (2010) CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327: 167–170
 172. Jansen R, Embden JD, Gastra W, Schouls LM (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* 43: 1565–1575
 173. Bolotin A, Quinquis B, Sorokin A, Ehrlich SD (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extra-chromosomal origin. *Microbiology* 151: 2551–2561
 174. Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Soria E (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* 60: 174–182
 175. Pourcel C, Salvignol G, Vergnaud G (2005) CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* 151: 653–663
 176. Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Almendros C (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 155: 733–740
 177. Horvath P, Romero DA, Coute-Monvoisin AC, Richards M, Deveau H, Moineau S, Boyaval P, Fremaux C, Barrangou R (2008) Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J Bacteriol* 190: 1401–1412
 178. Garneau JE, Dupuis ME, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadan AH, Moineau S (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468: 67–71
 179. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315: 1709–1712
 180. Datsenko KA, Pougach K, Tikhonov A, Wanner BL, Severinov K, Semenova E (2012) Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat Commun* 3: 945
 181. Pougach K, Semenova E, Bogdanova E, Datsenko KA, Djordjevic M, Wanner BL, Severinov K (2010) Transcription, processing and function of CRISPR cassettes in *Escherichia coli*. *Mol Microbiol* 77: 1367–1379
 182. Marraffini LA, Sontheimer EJ (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 322: 1843–1845
 183. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337: 816–821
 184. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM (2013) RNA-guided human genome engineering via Cas9. *Science* 339: 823–826
 185. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA et al (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339: 819–823
 186. Mali P, Esvelt KM, Church GM (2013) Cas9 as a versatile tool for engineering biology. *Nat Methods* 10: 957–963
 187. Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuys RJ, Snijders AP, Dickman MJ, Makarova KS, Koonin EV, van der Oost J (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321: 960–964
 188. Haurwitz RE, Jinek M, Wiedenheft B, Zhou K, Doudna JA (2010) Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* 329: 1355–1358
 189. Gesner EM, Schellenberg MJ, Garside EL, George MM, Macmillan AM (2011) Recognition and maturation of effector RNAs in a CRISPR interference pathway. *Nat Struct Mol Biol* 18: 688–692
 190. Carte J, Wang R, Li H, Terns RM, Terns MP (2008) Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev* 22: 3489–3496
 191. Carte J, Pfister NT, Compton MM, Terns RM, Terns MP (2010) Binding and cleavage of CRISPR RNA by Cas6. *RNA* 16: 2181–2188

192. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 471: 602–607
193. Semenova E, Jore MM, Datsenko KA, Semenova A, Westra ER, Wanner B, van der Oost J, Brouns SJ, Severinov K (2011) Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci USA* 108: 10098–10103
194. Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, Waghmare SP, Wiedenheft B, Pul U, Wurm R, Wagner R et al (2011) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol* 18: 529–536
195. Beloglazova N, Petit P, Flick R, Brown G, Savchenko A, Yakunin AF (2011) Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference. *EMBO J* 30: 4616–4627
196. Sinkunas T, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V (2011) Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J* 30: 1335–1342
197. Gasiunas G, Barrangou R, Horvath P, Siksnys V (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci USA* 109: E2579–E2586
198. Hatoum-Aslan A, Maniv I, Samai P, Marraffini LA (2013) Genetic characterization of anti-plasmid immunity by a Type III-A CRISPR-Cas system. *J Bacteriol* doi: 10.1128/JB.01130-13
199. Hale CR, Zhao P, Olson S, Duff MO, Graveley BR, Wells L, Terns RM, Terns MP (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* 139: 945–956
200. Hale CR, Majumdar S, Elmore J, Pfister N, Compton M, Olson S, Resch AM, Glover CV, Graveley BR, Terns RM et al (2012) Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. *Mol Cell* 45: 292–302
201. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS (2009) Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* 324: 218–223
202. Ingolia NT, Lareau LF, Weissman JS (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147: 789–802
203. Galindo MI, Pueyo JI, Fouix S, Bishop SA, Couso JP (2007) Peptides encoded by short ORFs control development and define a new eukaryotic gene family. *PLoS Biol* 5: e106
204. Savard J, Marques-Souza H, Aranda M, Tautz D (2006) A segmentation gene in *tribolium* produces a polycistronic mRNA that codes for multiple conserved peptides. *Cell* 126: 559–569
205. Kondo T, Hashimoto Y, Kato K, Inagaki S, Hayashi S, Kageyama Y (2007) Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA. *Nat Cell Biol* 9: 660–665
206. Pauli A, Norris ML, Valen E, Chew GL, Gagnon JA, Zimmerman S, Mitchell A, Ma J, Dubrulle J, Reyón D et al (2014) Toddler: An Embryonic Signal That Promotes Cell Movement via Apelin Receptors. *Science* 343: 1248636
207. Guttman M, Russell P, Ingolia NT, Weissman JS, Lander ES (2013) Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* 154: 240–251
208. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG et al (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* 22: 1775–1789
209. Brockdorff N (2013) Noncoding RNA and Polycomb recruitment. *RNA* 19: 429–442
210. Froberg JE, Yang L, Lee JT (2013) Guided by RNAs: X-inactivation as a model for lncRNA function. *J Mol Biol* 425: 3698–3706
211. Rinn JL, Chang HY (2012) Genome regulation by long noncoding RNAs. *Annu Rev Biochem* 81: 145–166
212. Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, Shi Y, Segal E, Chang HY (2010) Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 329: 689–693
213. Zhao J, Ohsumi TK, Kung JT, Ogawa Y, Grau DJ, Sarma K, Song JJ, Kingston RE, Borowsky M, Lee JT (2010) Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol Cell* 40: 939–953
214. Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, Feil R, Fraser P (2008) The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* 322: 1717–1720
215. Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, Nagano T, Mancini-Dinardo D, Kanduri C (2008) Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell* 32: 232–246
216. Bertani S, Sauer S, Bolotin E, 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
217. Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA et al (2011) A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 472: 120–124
218. Lai F, Orom UA, Cesaroni M, Beringer M, Taatjes DJ, Blobel GA, Shiekhattar R (2013) Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* 494: 497–501
219. Di Ruscio A, Ebralidze AK, Benoukrat T, Amabile G, Goff LA, Terragni J, Figueroa ME, De Figueiredo Pontes LL, Alberich-Jorda M, Zhang P et al (2013) DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* 503: 371–376
220. Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L et al (2011) lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* 477: 295–300
221. Duszczek MM, Wutz A, Rybin V, Sattler M (2011) The Xist RNA A-repeat comprises a novel AUCC tetraloop fold and a platform for multimerization. *RNA* 17: 1973–1982
222. Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT (2008) Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* 322: 750–756
223. Royce-Tolland ME, Andersen AA, Koyfman HR, Talbot DJ, Wutz A, Tonks ID, Kay GF, Panning B (2010) The A-repeat links ASF/SF2-dependent Xist RNA processing with random choice during X inactivation. *Nat Struct Mol Biol* 17: 948–954
224. Jeon Y, Lee JT (2011) YY1 tethers Xist RNA to the inactive X nucleation center. *Cell* 146: 119–133
225. Lanz RB, McKenna NJ, Onate SA, Albrecht U, Wong J, Tsai SY, Tsai MJ, O'Malley BW (1999) A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. *Cell* 97: 17–27
226. Kino T, Hurt DE, Ichijo T, Nader N, Chrousos GP (2010) Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci Signal* 3: ra8
227. Wang X, Arai S, Song X, Reichart D, Du K, Pascual G, Tempst P, Rosenfeld MG, Glass CK, Kurokawa R (2008) Induced ncRNAs allosterically modify

- RNA-binding proteins in cis to inhibit transcription. *Nature* 454: 126–130
228. Lanz RB, Razani B, Goldberg AD, O'Malley BW (2002) Distinct RNA motifs are important for coactivation of steroid hormone receptors by steroid receptor RNA activator (SRA). *Proc Natl Acad Sci USA* 99: 16081–16086
 229. Zhao X, Patton JR, Ghosh SK, Fischel-Ghodsian N, Shen L, Spanjaard RA (2007) Pus3p- and Pus1p-dependent pseudouridylation of steroid receptor RNA activator controls a functional switch that regulates nuclear receptor signaling. *Mol Endocrinol* 21: 686–699
 230. Clemson CM, Hutchinson JN, Sara SA, Ensminger AW, Fox AH, Chess A, Lawrence JB (2009) An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol Cell* 33: 717–726
 231. Sasaki YT, Ideue T, Sano M, Mituyama T, Hirose T (2009) MENepsilon/ beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles. *Proc Natl Acad Sci USA* 106: 2525–2530
 232. Sunwoo H, Dinger ME, Wilusz JE, Amaral PP, Mattick JS, Spector DL (2009) MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res* 19: 347–359
 233. Naganuma T, Nakagawa S, Tanigawa A, Sasaki YF, Goshima N, Hirose T (2012) Alternative 3' end processing of long noncoding RNA initiates construction of nuclear paraspeckles. *EMBO J* 31: 4020–4034
 234. Souquere S, Beauclair G, Harper F, Fox A, Pierron G (2010) Highly ordered spatial organization of the structural long noncoding NEAT1 RNAs within paraspeckle nuclear bodies. *Mol Biol Cell* 21: 4020–4027
 235. Hirose T, Virnicchi G, Tanigawa A, Naganuma T, Li R, Kimura H, Yokoi T, Nakagawa S, Benard M, Fox AH et al (2014) NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies. *Mol Biol Cell* 25: 169–183
 236. Prasanth KV, Prasanth SG, Xuan Z, Hearn S, Freier SM, Bennett CF, Zhang MQ, Spector DL (2005) Regulating gene expression through RNA nuclear retention. *Cell* 123: 249–263
 237. Valgardsdottir R, Chiodi I, Giordano M, Rossi A, Bazzini S, Ghigna C, Riva S, Biamonti G (2008) Transcription of Satellite III non-coding RNAs is a general stress response in human cells. *Nucleic Acids Res* 36: 423–434
 238. Audas TE, Jacob MD, Lee S (2012) Immobilization of proteins in the nucleolus by ribosomal intergenic spacer noncoding RNA. *Mol Cell* 45: 147–157
 239. Jacob MD, Audas TE, Uniacke J, Trinkle-Mulcahy L, Lee S (2013) Environmental cues induce a long noncoding RNA-dependent remodeling of the nucleolus. *Mol Biol Cell* 24: 2943–2953
 240. Chen LL, Carmichael GG (2009) Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. *Mol Cell* 35: 467–478
 241. Nakagawa S, Naganuma T, Shioi G, Hirose T (2011) Paraspeckles are subpopulation-specific nuclear bodies that are not essential in mice. *J Cell Biol* 193: 31–39
 242. Wilusz JE, Freier SM, Spector DL (2008) 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell* 135: 919–932
 243. Brown JA, Valenstein ML, Yario TA, Tycowski KT, Steitz JA (2012) Formation of triple-helical structures by the 3' end sequences of MALAT1 and MENbeta noncoding RNAs. *Proc Natl Acad Sci USA* 109: 19202–19207
 244. Wilusz JE, Jnbaptiste CK, Lu LY, Kuhn CD, Joshua-Tor L, Sharp PA (2012) A triple helix stabilizes the 3' ends of long noncoding RNAs that lack poly(A) tails. *Genes Dev* 26: 2392–2407
 245. Conrad NK, Mili S, Marshall EL, Shu MD, Steitz JA (2006) Identification of a rapid mammalian deadenylation-dependent decay pathway and its inhibition by a viral RNA element. *Mol Cell* 24: 943–953
 246. Mitton-Fry RM, DeGregorio SJ, Wang J, Steitz TA, Steitz JA (2010) Poly (A) tail recognition by a viral RNA element through assembly of a triple helix. *Science* 330: 1244–1247
 247. Tycowski KT, Shu MD, Borah S, Shi M, Steitz JA (2012) Conservation of a triple-helix-forming RNA stability element in noncoding and genomic RNAs of diverse viruses. *Cell Rep* 2: 26–32
 248. Wojciechowska M, Krzyzosiak WJ (2011) Cellular toxicity of expanded RNA repeats: focus on RNA foci. *Hum Mol Genet* 20: 3811–3821
 249. Park E, Maquat LE (2013) Staufen-mediated mRNA decay. *Wiley Interdiscip Rev RNA* 4: 423–435
 250. Gong C, Maquat LE (2011) lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3'UTRs via Alu elements. *Nature* 470: 284–288
 251. Ricci EP, Kucukural A, Cenik C, Mercier BC, Singh G, Heyer EE, Ashar-Patel A, Peng L, Moore MJ (2014) Staufen1 senses overall transcript secondary structure to regulate translation. *Nat Struct Mol Biol* 21: 26–35
 252. Kretz M, Siprashvili Z, Chu C, Webster DE, Zehnder A, Qu K, Lee CS, Flockhart RJ, Groff AF, Chow J et al (2013) Control of somatic tissue differentiation by the long non-coding RNA TINCR. *Nature* 493: 231–235
 253. Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J (2013) Natural RNA circles function as efficient microRNA sponges. *Nature* 495: 384–388
 254. Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, Maier L, Mackowiak SD, Gregersen LH, Munschauer M et al (2013) Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 495: 333–338
 255. Kertesz M, Wan Y, Mazor E, Rinn JL, Nutter RC, Chang HY, Segal E (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature* 467: 103–107
 256. Rouskin S, Zubradt M, Washietl S, Kellis M, Weissman JS (2013) Genome-wide probing of RNA structure reveals active unfolding of mRNA structures *in vivo*. *Nature* 505: 701–705
 257. Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC, Assmann SM (2013) *In vivo* genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* 505: 696–700
 258. Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, Osenberg S, Cesarkas K, Jacob-Hirsch J, Amariglio N, Kupiec M et al (2012) Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* 485: 201–206
 259. Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR (2012) Comprehensive analysis of mRNA methylation reveals enrichment in 3'UTRs and near stop codons. *Cell* 149: 1635–1646
 260. Squires JE, Patel HR, Nousch M, Sibbritt T, Humphreys DT, Parker BJ, Suter CM, Preiss T (2012) Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Res* 40: 5023–5033
 261. Khoddami V, Cairns BR (2013) Identification of direct targets and modified bases of RNA cytosine methyltransferases. *Nat Biotechnol* 31: 458–464
 262. Hussain S, Sajini AA, Blanco S, Dietmann S, Lombard P, Sugimoto Y, Paramor M, Gleeson JG, Odom DT, Ule J et al (2013) NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs. *Cell Rep* 4: 255–261

263. Sakurai M, Ueda H, Yano T, Okada S, Terajima H, Mitsuyama T, Toyoda A, Fujiyama A, Kawabata H, Suzuki T (2014) A biochemical landscape of A-to-I RNA editing in the human brain transcriptome. *Genome Res* 24: 522–534
264. Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, Clark TA, Schweitzer AC, Blume JE, Wang X et al (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* 456: 464–469
265. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano MJ, Jungkamp AC, Munschauer M et al (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 141: 129–141
266. Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM et al (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* 149: 1393–1406