Genome Regulation by Long Noncoding RNAs

John L. Rinn[1] and Howard Y. Chang[2]

[1]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts 02138; email: john_rinn@harvard.edu

[2]Howard Hughes Medical Institute and Program in Epithelial Biology, Stanford University School of Medicine, Stanford, California 94305; email: howchang@stanford.edu

Annu. Rev. Biochem. 2012. 81:145–66

The *Annual Review of Biochemistry* is online at biochem.annualreviews.org

This article's doi:
10.1146/annurev-biochem-051410-092902

0066-4154/12/0707-0145$20.00

Keywords

functional genomics, chromatin, histone modifications, epigenetics

Abstract

The central dogma of gene expression is that DNA is transcribed into messenger RNAs, which in turn serve as the template for protein synthesis. The discovery of extensive transcription of large RNA transcripts that do not code for proteins, termed long noncoding RNAs (lncRNAs), provides an important new perspective on the centrality of RNA in gene regulation. Here, we discuss genome-scale strategies to discover and characterize lncRNAs. An emerging theme from multiple model systems is that lncRNAs form extensive networks of ribonucleoprotein (RNP) complexes with numerous chromatin regulators and then target these enzymatic activities to appropriate locations in the genome. Consistent with this notion, lncRNAs can function as modular scaffolds to specify higher-order organization in RNP complexes and in chromatin states. The importance of these modes of regulation is underscored by the newly recognized roles of long RNAs for proper gene control across all kingdoms of life.

Contents

INTRODUCTION

Long noncoding RNA (lncRNA): an RNA that functions as a large RNA gene

The centrality of RNA in the flow of genetic information came to light in Jacob & Monod's 1961 paper "Genetic Regulatory Mechanisms in the Synthesis of Proteins," establishing the concept of messenger RNAs (1). In the 50 years since this landmark paper, numerous regulatory RNAs of all shapes and sizes have been discovered (2, 3). Long noncoding RNAs (lncRNAs) biochemically resemble mRNAs posited by Jacob & Monod, yet do not template protein synthesis. Rather, lncRNAs function as RNA genes to orchestrate genetic regulatory outputs. Today, lncRNA transcripts have emerged as a cryptic, but critical, layer in the genetic regulatory code (**Figures 1** and **2**).

Studies over the past several decades have pointed to the presence of large amounts of RNA that was transcribed but did not encode proteins (4–8). Some of this RNA was later explained by mRNA splicing and RNA genes comprising translation machinery and its regulation (i.e., ribosomal RNA, tRNA, RNase P, SRP-7S), yet the vast majority was still unexplained. Biochemical experiments were able to characterize many abundant structural and regulatory RNAs by cellular localization and sequence similarity (5, 8–12), and genetic studies identified a few lncRNA genes involved in imprinting and other cellular processes (*XIST*, *H19*, *AIR*) (13, 14). Additional genetic studies also pointed to an emerging class of small regulatory RNAs, such as microRNAs (miRNAs) (12, 15–18), that regulate the translation of mRNAs to fine-tune key genetic pathways. Collectively, these classical studies identified a diverse repertoire of RNAs but may have only scratched the surface of RNAs' functions in the cell.

The advent of full genome sequences enabled an unprecedented survey of the genomic landscape for new genes. Surprisingly, this prospecting for genes led to the discovery of numerous lncRNA genes, but not many more protein genes. At the same time, DNA microarray technology revealed that the genome encodes at least as many lncRNAs as the known protein-coding genes (19–22). In fact, further advancements in RNA sequencing and microarray technology allowed a consortium-wide effort to define all the transcribed bases in the genome. At present, lncRNAs are operationally defined as RNA genes larger than 200 bp that do not appear to have coding potential. Although this working definition is somewhat arbitrary, the size cutoff clearly distinguishes lncRNAs

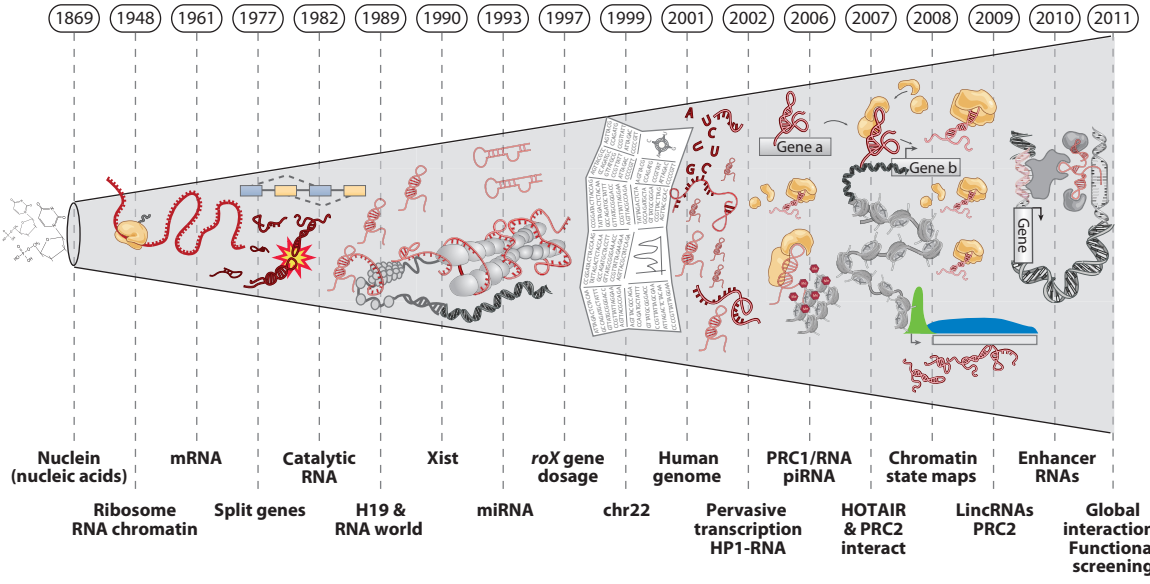

Figure 1

Timeline of discoveries of RNAs in biological regulation.

from small regulatory RNAs, such as miRNAs or piRNAs. Some classically defined small nuclear RNAs are in fact greater than 200 nucleotides, but the lncRNA designation is only prospectively applied to newly recognized transcripts. The conclusion was that a vast majority of the genome was transcribed (23).

In contrast to the substantial progress made in mapping lncRNAs, the functional roles for lncRNAs remained mostly elusive. In fact, the notion of such a widespread abundance of transcription was becoming controversial (23–25). More recently, dozens of functional examples have emerged implicating lncRNAs in numerous cellular processes ranging from embryonic stem cell (ESC) pluripotency, cell-cycle regulation, and diseases, such as cancer. Although lncRNAs exert a diverse spectrum of regulatory mechanisms across a variety of cellular pathways, a common theme is emerging: lncRNAs drive the formation of ribonucleic-protein complexes, which in turn influence the regulation of gene expression.

Here, we discuss multiple lines of evidence that point to lncRNAs as key regulatory layers in global gene regulation. We review the the technological approaches for genome-wide discovery and characterization of lncRNAs, as well as highlight emerging mechanistic themes based on well-studied examples from diverse model systems. We further evaluate several emerging studies indicating important roles for lncRNAs in the etiology of a wide spectrum of diseases.

GENOMIC DISCOVERY OF LONG NONCODING RNAS

At the turn of the twenty-first century, the scientific community was abuzz with great anticipation of the human genome project (26, 27). Perhaps at center stage was the burning question: How many genes are there in the human genome? Can the complexity of different organisms be explained by the sheer number of classic protein-coding genes, their splicing diversity, or perhaps new types of regulation? This simple yet profound question drove the progress of many technologies, such as microarray and DNA sequencing,

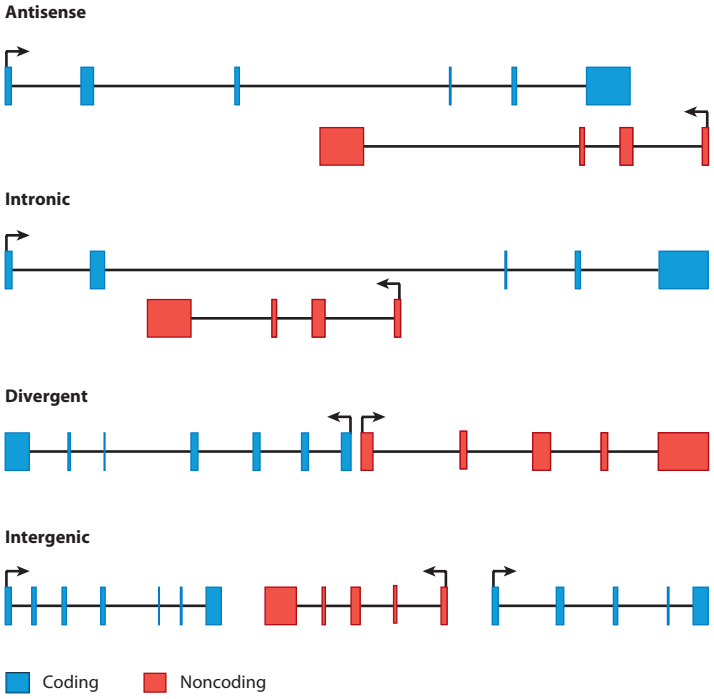

Figure 2

Anatomy of long noncoding RNA (lncRNA) loci. lncRNAs are often defined by their location relative to nearby protein-coding genes. Antisense lncRNAs are lncRNAs that initiate inside or 3′ of a protein-coding gene, are transcribed in the opposite direction of protein-coding genes, and overlap at least one coding exon. Intronic lncRNAs are lncRNAs that initiate inside of an intron of a protein-coding gene in either direction and terminate without overlapping exons. Bidirectional lncRNAs are transcripts that initiate in a divergent fashion from the promoter of a protein-coding gene; the precise distance cutoff that constitutes bidirectionality is not defined but is generally within a few hundred base pairs. Finally, intergenic lncRNAs (also termed large intervening noncoding RNAs or lincRNAs) are lncRNAs with separate transcriptional units from protein-coding genes. One definition required lincRNAs to be 5 kb away from protein-coding genes (44).

at an unprecedented rate. One of the first applications of automated Sanger sequencing in the mid-1990s was the mapping of expressed sequence tags (28, 29) that identify fragments of genomic regions that were being actively transcribed. This first glimpse into the transcriptome in 1996 revealed an intriguing new notion that many genes would be lurking in yet undefined regions of the human genome (29). Yet limited by short sequence reads, lower coverage, and an incomplete reference human genome to align expressed sequence tags, what these new genes may encode remained elusive.

Pervasive transcription: the phenomenon, recognized in the early twenty-first century, showing that a vast majority of the genome is transcriptionally active

Tiling Microarrays

In addition to sequencing advances, new technologies were emerging to understand the regulation of gene expression and de novo identification of new genes. In particular, the advent of DNA microarray technology provided the ability to survey on the order of 20,000 gene or genomic loci. In parallel, the first complete human chromosome 22 sequence was released in 1999 (30). The combined power of microarrays and draft genome sequences provided the first glimpse into pervasive transcription of noncoding RNAs. Specifically, two independent studies reported initial estimates that there may be as many lncRNA genes as protein-coding genes (21, 22). These studies used DNA microarrays with tiled or nested target sequences comprising the entirety of a chromosomal DNA sequence and allowed an unbiased survey of transcribed regions. Some limitations of tiling array studies included the potential for cross hybridization, the lack of strand-specific information if cDNAs were hybridized to the array, and the connectivity between transcribed regions not being known. Nonetheless, both studies were able to confirm expression from numerous noncoding loci by reverse transcription polymerase chain reaction, RNA-blot analysis, and evolutionary conservation studies, yet these findings were met with healthy skepticism that they may simply represent transcriptional noise. In a potentially interesting historical parallel, the discovery of "DNA-like RNA" upon phage infection in 1956 was a critical clue leading to the "mRNA hypothesis" (31). Yet the significance of this finding was not recognized for at least five years because the DNA-like RNA was less than 1% of total RNA (mostly rRNA) and thus assumed to be irrelevant noise.

Thus, one of the fruits of the Human Genome Project was the discovery of numerous new RNA genes, but not new protein-coding genes. For example, the number of human miRNAs quickly rose from a handful to nearly 1,000 (32–34). In fact, the further advancements in RNA sequencing, cDNA cloning, and

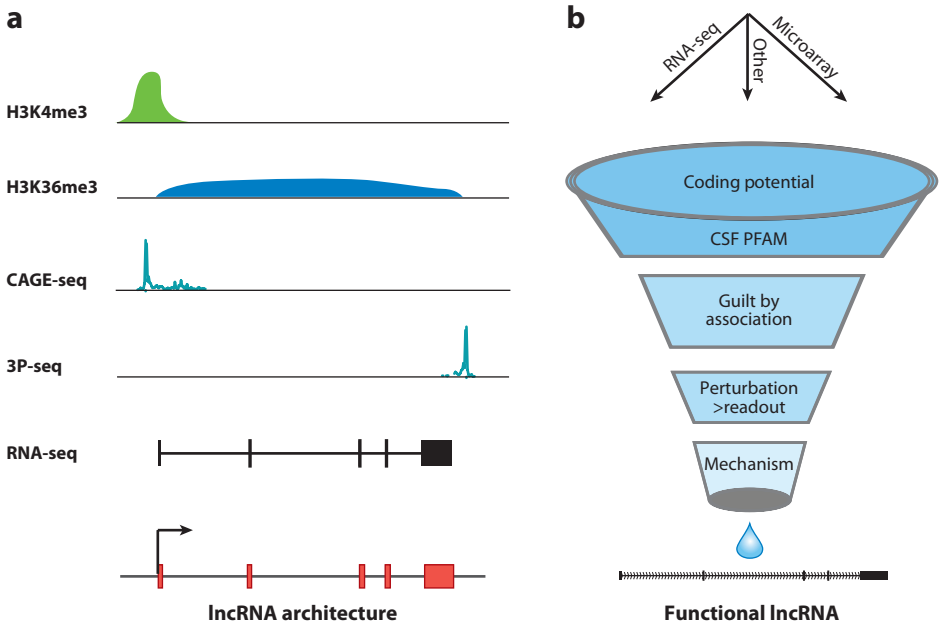

Figure 3

Functional discovery pipeline of long noncoding RNAs (lncRNAs). (*a*) Genome-wide discovery of lncRNAs. Chromatin marks of transcription initiation (histone H3 lysine 4 trimethylation, H3K4me3) and elongation (H3 lysine 36 trimethylation, H3K36me3) define transcribed regions of the genome, and sequencing of capped RNA fragments (CAGE-seq) or polyadenylation ends (3P-seq) define the precise beginnings and ends of transcripts. RNA sequencing (RNA-seq) can directly define the primary structure of lncRNAs. (*b*) Multiple bioinformatic tools and functional studies refine the set of lncRNAs associated with specific biological functions. Abbreviations: CSF, codon substitution frequency analysis; PFAM, Protein Family Database.

microarray technology of the next decade allowed a consortium-wide effort to define all the transcribed bases in the human genome. The conclusion was that a vast majority of the genome was transcribed (23). Despite the observed pervasive transcription throughout the genome, pinpointing functional RNA molecules was tantamount to finding needles in a haystack. The notion of such a widespread abundance of transcription was becoming increasingly controversial (24, 25, 35–38).

Chromatin Marks

A critical clue for hunting RNA genes came from chromatin, the DNA-protein complex where all eukaryotic genes reside (**Figure 3*a***). With the full genome sequence in hand, chromatin immunoprecipitation followed by massively parallel sequencing generated genomic maps of the chromatin architecture that have been termed the epigenome (39–41). Massively parallel sequencing of DNA sites occupied by histones and their modifications revealed numerous interesting domains of genomic architecture (39, 42). This included a clear signature of polymerase II–transcribed genes occupied by histone H3 lysine 4 trimethylation (H3K4me3) at the promoters of genes, followed by histone H3 lysine 36 trimethylation along the transcribed unit (K4-K36 domains) (42–44).

Surprisingly, surveying the entire mouse and human genomes by chromatin marks in several cell types revealed that approximately 5,000 K4-K36 domains represented lncRNAs (44). These lncRNAs had discrete gene loci that reside in previously unannotated intergenic

Large intergenic noncoding RNA (lincRNA): a lncRNA that does not overlap protein-coding genes

regions between protein-coding genes; hence, these RNAs were named large intergenic noncoding RNAs (lincRNAs). Further analysis of these loci revealed highly conserved promoter regions that recruit the binding and direct regulation of key transcription factors (24, 44, 45). lincRNAs show more sequence conservation throughout evolution over introns or untranscribed intergenic sequences, suggestive of their functionality (24, 44, 45), and syntenic lincRNAs in different species can have conserved functions (45a). Moreover, expression patterns of lncRNAs are associated with numerous key cellular processes, such as pluripotency, immune response, and regulation of the cell cycle (44, 46–48). More recently, approximately a third of lincRNAs were found to associate with chromatin-modifying complexes (49) and to modulate key cellular pathways (48, 50–52).

RNA Sequencing

The advent of deep sequencing technology led to the ability to sequence cDNA at an unprecedented scale and throughput, termed RNA-seq (53–57). These approaches have been coupled to computational methods, allowing the reconstruction of transcripts and their isoforms at single-nucleotide resolution (55–57). These studies have provided an unbiased identification of noncoding transcripts across many cell types and tissues (56, 58).

In addition to full-length reconstruction algorithms, several applications have emerged to perform RNA-seq. For example, a method termed 3-seq targets the polyadenylated tail of cDNA to quantitatively measure the abundance of transcripts using more affordable short sequence reads (59). Moreover, a variant of this method can be employed to precisely map 3′ ends of transcripts (60). Recently, metabolically labeled mRNAs have been utilized to measure nascent transcription, thereby providing insights into the pausing of polymerase and transcriptional dynamics (61). These and many other emerging technologies are providing ever deeper insights into the dynamic transcriptome (**Figure 3*a***).

Recent studies have utilized RNA sequencing and transcript abundance estimations to identify specific properties of distinct classes of large RNA genes. For example, one study (58) identified 8,000 lincRNAs in the human genome by integrating numerous annotation sources in combination with RNA sequencing. This study revealed several global properties of lncRNAs, including a tendency for location next to developmental regulators, enrichment of tissue-specific expression patterns, identification of thousands of orthologous lincRNAs between human and mouse, and localization of hundreds of lincRNAs in gene deserts associated with genetic traits (58). Leveraging the ever increasing depth of sequencing and read lengths has allowed some of the first steps toward characterizing lncRNAs on a global scale.

GENOMIC CHARACTERIZATION OF LONG NONCODING RNAS

We next discuss experimental and computational approaches to identify, map, and derive hypotheses for lncRNA function (**Figure 3**). By combining the above technical approaches, it is now possible to identify all transcribed loci (K4-K36 chromatin domains) as well as to precisely map the primary structure of the RNA products (RNA-seq) (44, 45, 56). These combined layers of information are synergistic where chromatin modifications identify stably transcribed gene loci, and RNA sequencing allows detection of even very low-abundance transcripts that alone could be argued as transcriptional noise. The additional chromatin information also indicates the promoter region of a given locus (H3K4me3) and the transcribed unit (H3K36me3), thereby assisting in the mapping of RNA transcript 5′ and 3′ ends. Through the successive addition of additional layers of information (such as conservation, coding potential patterns, and anatomical properties) progress is being made toward identifying lncRNA gene families (43–45).

On the basis of the anatomical properties of their gene loci, lncRNAs have been further classified: for example, antisense lncRNAs that

overlap known protein-coding genes, intronic lncRNAs that are encoded within introns of protein-coding genes, lncRNAs that overlap protein-coding genes termed overlapping transcripts, and lincRNAs that are encoded completely within the intergenic genomic space between protein-coding loci (**Figure 2**). Although this anatomic characterization has been used initially, it is likely that research will show many of these lncRNAs share similar mechanistic and functional roles.

Excluding Protein-Coding Potential

Whether an RNA transcript functions by coding for protein is fundamental to the definition of lncRNA, but obtaining the answer is a challenging task. Many studies have assessed lncRNA-coding potential by translating each lncRNA in all 3′ frames and performing homology queries (i.e., BLASTX) across large protein family and domain databases (i.e., Swissprot and PFAM). These informatic analyses are good initial indications of protein-coding capacity but may miss newly evolved protein sequences or very small open reading frames (<50 amino acids). To address the former issue, codon substitution frequency analyses have been used to determine if codons for amino acids are preferentially conserved through evolution, indicating preservation of protein-coding potential (62, 63). Yet even these two methods combined could still miss small open reading frames buried in these long transcripts. Ribosomal profiling, an experimental menthod, which identifies putative RNAs that are bound and scanned by the ribosome, has provided additional insights into those RNAs that may encode small peptides (64). Moreover, this method identifies the region of ribosomal occupancy, thereby further homing in on potential translated regions that can be used as refined input into informatics predictions, such as codon substitution frequency and BLASTX. Although some portion of lncRNAs may encode small peptides (65), we note this does not rule out the potentially dual nature of lncRNAs acting through RNA and their protein products. This has been exemplified by numerous mRNAs that contain regulatory noncoding RNA elements (i.e., p53, Sgrs, Oskar, VegT, and others) (66–71).

Inference of Long Noncoding RNA Functions by Coexpression: Guilt by Association

With the mapping of thousands of lncRNA loci, the next challenge is to determine what lncRNAs do. A first step in hypothesis generation is to use the expression patterns of lncRNAs to identify specific cell types or biological processes associated with each candidate lncRNA. Some of the first expression studies of lncRNAs identified lncRNAs that are highly expressed in certain brain regions. In situ hybridization studies further confirmed these expression patterns, revealing exquisite patterns of expression in specific substructures of the mouse brain [see Mercer et al. (72)]. A similar study by this group identified numerous lncRNAs that were tightly correlated with pluripotency transcription factors, suggesting that many lncRNAs may function in stem cell pluripotency transcriptional networks (73).

More recently, an informatic method termed guilt by association allowed a global understanding of lncRNAs and protein-coding genes that are tightly coexpressed and thus presumably coregulated (44). This method, using gene-expression analyses, identifies protein-coding genes and pathways significantly correlated with a given lncRNA. Thus, on the basis of the known functions of the coexpressed protein-coding genes, hypotheses are generated for the functions and potential regulators of the candidate lncRNA. Moreover, this analysis revealed families of lncRNAs on the basis of the pathways with which they do and do not associate. This approach has predicted diverse roles for lncRNAs, ranging from stem cell pluripotency to cancer (44). For example, numerous lncRNAs that were tightly correlated with p53 were induced in a p53-dependent manner, many more than would be expected by chance (44, 47, 48). These lncRNAs also were enriched with the p53-binding motif

Codon substitution frequency: determines the evolutionary pressure to preserve synonymous amino acid content

Guilt by association: hypothesis generation of lncRNA function by the coexpression of lncRNAs with protein-coding mRNAs; can group lncRNAs into the pathways they may regulate

Loss-of-function (LOF) approaches: define lncRNA functional roles

in their promoters. Moreover, one of these lncRNAs, termed lincRNA-p21 and predicted to be associated with p53, was found to be directly regulated by p53 and subsequently formed a lncRNA-ribonucleoprotein (RNP) with a nuclear factor to act as a global transcriptional repressor, facilitating p53-mediated apoptosis (48). Similarly, several lncRNAs predicted to be associated with adipogenesis and pluripotency have recently been identified as requirements for maintaining these cellular states (46; M. Guttman and L. Sun, personal communications).

Other expression correlation analyses have revealed additional functional roles of lncRNAs. For example, a recent study profiling lncRNAs across more than 130 breast cancers contained varying grades of the tumor and clinical information (74). This study identified numerous lncRNAs that are specifically up- or downregulated in tumor subtypes. For example, it was identified that a lncRNA termed HOTAIR, encoded in the HOXC cluster, was a strong predictor of breast cancer metastasis. In fact, enforced expression of HOTAIR was sufficient to drive breast cancer metastasis. More global expression studies of lncRNAs overlapping promoter regions of protein-coding genes identified numerous lncRNAs associated with cell-cycle regulation (47). This led to the functional characterization of a lncRNA named PANDA, which plays a critical role in inhibiting p53-mediated apoptosis. The guilt-by-association methods are universally applicable to any biological system. For example, a family of telomere-encoded lncRNAs in the malaria parasite (*Plasmodium falciparum*) was identified by their stage-specific coexpression with PfsiP2, a key virulence transcription factor (75).

These and other correlative studies have started to identify specific roles of lncRNAs in global transcriptional regulation. Homing in on the pathways with which lncRNAs are associated has allowed for (or provided) hypothesis-driven experiments that identified functional lncRNAs. Yet the full scope of lncRNA transcriptional regulation and function is far from understood. To understand the more global regulatory roles of lncRNAs, comprehensive gain- or loss-of-function (LOF) experiments need to be performed.

High-Throughput Loss of Function by RNA Inference

Indeed, a very recent study performed a LOF study across most (237) of the lincRNAs expressed in mouse ESCs and characterized the resulting effects on global gene expression (52). The authors demonstrated that knockdowns of lincRNAs have major consequences on gene expression patterns, comparable to knockdown of well-known ESC regulators. Intriguingly, this global screen determined that lincRNAs primarily affect gene expression in *trans*. Perhaps more importantly, dozens of lincRNAs were found to be functionally required in the maintenance of the pluripotent state. Further investigation into the molecular circuitry of ESCs showed that lincRNA genes are regulated by key transcription factors and that lincRNA transcripts physically bind to multiple chromatin regulatory proteins to affect shared gene expression programs (52). This study provided the first glimpse of global lincRNA functional properties and mechanisms and highlights their key role in the circuitry controlling the ESC state.

LONG NONCODING RNAS IN GENE REGULATION

Long Noncoding RNAs Bind to and Target Chromatin Regulators

The intimate connection between RNA and chromatin—the DNA-protein complex where all eukaryotic genes reside—was recognized over 40 years ago (76). In 1975, Paul & Duerksen (5) made the surprising finding that biochemically purified chromatin contained twice as much RNA as DNA, raising the idea that RNA may influence chromatin structure and gene regulation. Through the years, it has been demonstrated that RNA is required for proper chromatin structure and recruitment of

the chromatin-modifying complexes to DNA (14). Yet the specific RNA species associated remained elusive. Genetic studies in the ensuing decades revealed a few lncRNAs that were associated with heterochromatin formation and imprinting [i.e., Xist (77), Air (78), H19 (79)]. Breakthroughs over the past few years have revealed numerous examples of lncRNAs in controlling the access or dismissal of regulatory proteins from chromatin (**Table 1**). Here, we first focus on the protein-binding partners of lncRNAs, next review the targeting mechanism of lncRNAs, and lastly discuss the emerging mechanistic themes of lncRNAs in gene regulation.

Several studies have shown that lncRNAs can target several chromatin modification complexes involved in gene silencing (**Table 1**). One of the most dramatic examples of lncRNA-mediated chromatin regulation occurs during X chromosome dosage compensation in mammals. Briefly, dosage compensation refers to the process whereby the gene expression level of the two X chromosomes in female cells is made equal to the single X in male cells. The lncRNA Xist is expressed from one of the two X chromosomes in female cells and results in altering the chromatin structure of an entire chromosome—the inactive X—where most genes are transcriptionally silenced (reviewed in Reference 80). Importantly, Xist physically associates with the Polycomb repressive complex 2 (PRC2) through a structured domain termed Repeat A, resulting in the localization of the PRC2 and its cognate histone mark histone H3 lysine 27 trimethylation (H3K27me3) to the inactive X chromosome (81). In an analogous fashion, plants control the seasonal timing of flowering (a process termed vernalization) by a cold-inducible intronic lncRNA termed COLDAIR; COLDAIR recruits the PRC2 in *cis* to silence the flowering regulator gene *FLC* (82). PRC2 recruitment by lncRNA can also regulate distantly located genes throughout the genome. Human lncRNA HOTAIR was the first of such RNAs recognized. HOTAIR physically associates with the PRC2 and modulates the PRC2 and H3K27me3 localization of hundreds of sites throughout the genome (83, 84). Several additional studies have identified HOTAIR and Xist as interfacing with the PRC2 via the catalytic methyltransferase subunit EZH2 (81, 85), although other proteins are likely also involved (86). The precise molecular interactions between lncRNAs and the Polycomb complex have yet to be defined.

PRC1 and -2: Polycomb repressive complex 1 and 2

In addition, lncRNAs can target diverse chromatin regulators. In imprinting, the paternally and maternally inherited alleles are differentially expressed, and lncRNAs are often involved in distinguishing the two alleles. Both Air and Kcnq1ot1 are lncRNAs that are transcribed from the silenced paternal allele, and they specifically bind to and recruit the histone H3 lysine 9 methylase G9a in *cis* to mediate H3K9me3 and transcriptional silencing of *Kcnq1* or *Igf2r* loci, respectively (87, 88). DNA methylation can be regulated by lncRNA. In plants, the interplay of small interfering RNAs and nascent lncRNAs in targeting DNA methylation is well known (89), but a different mechanism, apparently independent of small regulatory RNAs, also operates in mammalian cells. Transcriptional repression of the repetitive ribosomal RNA gene loci (*rDNA*) depends in part on an ncRNA termed pRNA, which recruits DNMT3b to mediate cystosine methylation (90). Additionally, ANRIL, a lncRNA that is associated with cardiac disease, associates with CBX7 of the PRC1, facilitating the H3K27me3-based silencing of the *INK4a* locus (91). Two p53-regulated lncRNAs, lincRNA-p21 and PANDA, have been recently identified as interfacing with DNA-binding proteins such as hnRNP-K and NF-YA, and these interactions result in transcriptional repression at specific genomic loci (47, 48). Finally, beyond chromatin modifications, SRA, a lncRNA, can interact with and enhance the function of the insulator protein CTCF (92); CTCF can control higher-order chromosomal looping and "insulate" specific genes from the effects of long-range enhancers and regulatory elements. On the basis of these numerous examples, it is clear that many chromatin regulatory complexes moonlight as

Table 1 Protein partners of long noncoding RNAs[a]

Proteins[b]	Long noncoding RNAs	Long noncoding RNA functions	References
CTCF	SRA	Enhances insulator function of CTCF	92
DNMT3b	pRNA	Targets DNMT3b in *cis* to the rRNA locus via an RNA:DNA:DNA triplex for cytosine methylation and gene silencing	90
G9a	Kcnq1ot1, Air	Targets H3K9 methylase G9a in *cis* for imprinting	87, 88
Glucocorticoid receptor	Gas5	Binds to glucocorticoid receptor as a decoy and titrates GR away from target genes	102
hnRNP-K	lincRNA-p21	Targets hnRNP-K in *trans* to mediate p53-dependent gene repression	48
LSD1-CoREST	HOTAIR, many others	Targets the LSD1 complex to demethylate H3K4me2 to enforce gene silencing	49, 84
MLL-WDR5	HOTTIP, some eRNAs?	Binds to and localizes the MLL complex and H3K4me3 via chromosomal looping for gene activation	95, 96
NF-YA	PANDA	p53 inducible and titrates away NF-YA to favor survival over cell death during DNA damage	47
PRC1	ANRIL, Xist	Targets PRC1 in *cis* for gene silencing. ANRIL influences p16INK4a expression and cell senescence	9, 91
PRC2	Xist, HOTAIR, ANRIL, COLDAIR, Gtl2, Kcnq1ot1, many others	Targets PRC2 either in *cis* or *trans* to mediate H3K27 methylation and gene silencing for dosage compensation, imprinting, and developmental gene expression	49, 81, 83, 88, 109
Serine/arginine-rich splicing factors	MALAT1	Sequesters serine/arginine splicing factors to regulate alternative splicing	125
Staufen	1/2 SBS RNAs	Pairs with mRNAs via Alu repeats and targets them into a nonsense-mediated decay pathway	126
Set1 and Hda1/2/3 HDACs	CUTs, XUTs	Antisense RNAs repress sense transcription via control of H3K4me3 and histone deacetylation	127–130
hnRNP-A	TERRA	Controls telomerase access to telomeres in a cell-cycle phase-specific manner	131
TFIIB	DHFR minor	Titrates away TFIIB during cell quiescence to decrease *DHFR* transcription	99
TLS	CCND1 promoter ncRNA	Allosterically binds TLS to inhibit CREB binding protein and p300 activity, leads to repression of the *CCND1* gene	132
YY1	Xist	YY1 binding nucleates Xist on the inactive X chromosome	100

[a]Many of these long noncoding RNA-protein complexes function at chromatin.
[b]Abbreviations: 1/2 SBS, half Staufen binding site; CUTs, cryptic unstable transcripts; MLL, mixed lineage leukemia protein; TLS, translocated in liposarcoma; XUTs, Xrn1-sensitive unstable transcripts.

RNA-binding proteins; the ability to bind lncRNAs endows them with condition- or allele-specific recognition of target gene chromatin.

Enhancer RNAs

For historical reasons, many of the initial studies focused on RNAs associated with repressive chromatin-modifying complexes. Yet several other studies have also demonstrated that active chromatin states are associated with lncRNAs. Genome-scale mapping of histone modifications and enhancer-binding proteins have provided an additional layer of information to identify lncRNAs involved in gene activation. Chromatin immunoprecipitation followed by massively parallel sequencing analysis of H3K4me1, H3K27ac, and p300, several marks associated with gene-activating enhancers, showed these regions also produce lncRNA transcripts. Many such enhancer RNAs were bidirectional, lacked a polyA tail, and had a very low copy number (93, 94). Although many of these transcripts were initially thought to be by-products of polymerase II transcription or enhancer-promoter interaction, more evidence is pointing to functional roles of the lncRNAs. A recent study performed LOF experiments and found 7 of 12 lncRNAs affected expression of their cognate neighboring genes (95). The authors continued to demonstrate it was not the act of transcription rather the RNA itself that was important for gene enhancer activation. Although this trend of lncRNAs affecting transcription of neighboring genes is not a universal phenomenon (47, 52), these studies clearly demonstrate a functional role for the RNA molecule beyond that of a simple by-product of transcription in enhancer regions.

More recently, HOTTIP, an enhancer-like lncRNA, has been discovered to directly interact with the WDR5 protein, a key component of the mixed lineage leukemia/Trx complex, which catalyzes the activating H3K4me3 mark (96). HOTTIP is encoded on the distal 5′ end of the *HOXA* gene cluster. Chromosomal looping of the 5′ end of *HOXA* in an enhancer-like manner brings HOTTIP into spatial proximity with multiple *HOXA* genes, enforcing the maintenance of H3K4me3 and gene activation. Remarkably, in vivo HOTTIP LOF experiments silenced *HOXA* expression and altered limb morphology, consistent with its role in activating *HOXA* genes (96). Collectively, these studies demonstrate the critical importance of lncRNAs interfacing with chromatin-modifying machinery, resulting in enhancer-based gene activation, and these findings raise the possibility that many other enhancer-like RNAs may operate via similar mechanisms. Thus, a critical path forward for understanding the functions of lncRNAs is to understand the repertoire of lncRNA-binding proteins.

The RNA-Chromatin Interface

How does a lncRNA interface with selective regions of the genome? Several hypotheses have been forwarded, including (*a*) formation of an RNA:DNA:DNA triplex; (*b*) RNA binding to a sequence-specific DNA-binding protein; (*c*) an RNA:DNA hybrid that displaces a single strand of DNA (so-called R-loop); and (*d*) an RNA:RNA hybrid of lncRNA with a nascent transcript (97, 98). Mechanisms (*a*) and (*b*) have been experimentally demonstrated in several systems (**Figure 4**).

Two studies have demonstrated that the association of lncRNAs and chromatin complexes can also include recruitment to DNA. The first of such studies demonstrated that a lncRNA encoded upstream of the *DHFR* gene forms a triplex structure with the promoter, which binds to and sequesters the general transcription factor IIB and prevents transcription of *DHFR* (99). More recently, another study identified a 150–250-nucleotide species of ncRNA, termed pRNA, which also forms a triplex at rDNA loci to recruit DNMT3b to this location through the DNA-RNA triplex (90). In each of these cases, purified ncRNA is able to bind to the cognate DNA sequence to form a triplex structure in vitro, but it is difficult to demonstrate that such a triplex forms in living cells. Nonetheless, on the basis of these precedents,

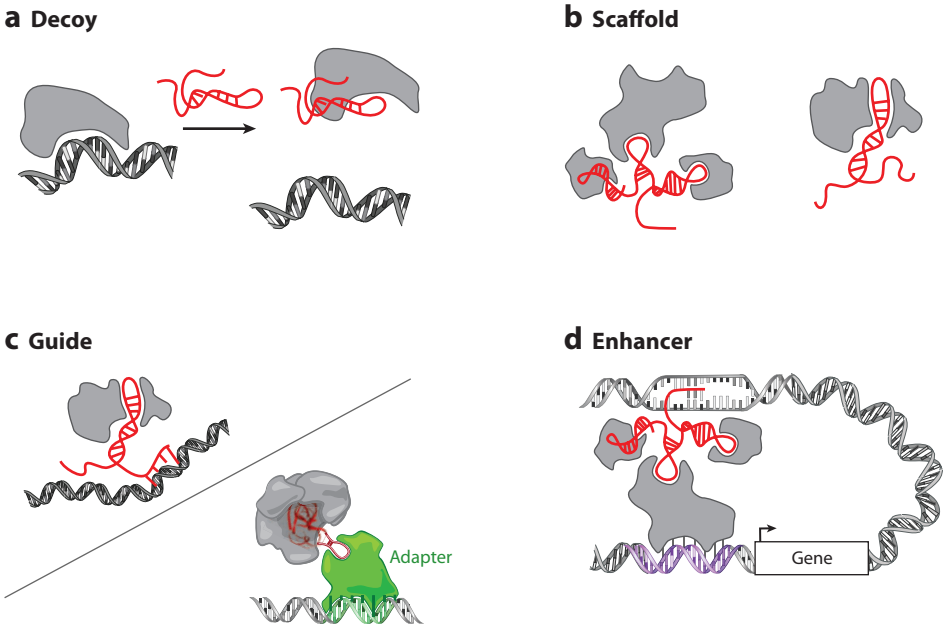

Figure 4

Models of long noncoding RNA (lncRNA) mechanisms of action. (*a*) The lncRNAs can act as decoys that titrate away DNA-binding proteins, such as transcription factors. (*b*) These lncRNAs may act as scaffolds to bring two or more proteins into a complex or spatial proximity and (*c*) may also act as guides to recruit proteins, such as chromatin modification enzymes, to DNA; this may occur through RNA-DNA interactions or through RNA interaction with a DNA-binding protein. (*d*) Such lncRNA guidance can also be exerted through chromosome looping in an enhancer-like model, where looping defines the *cis* nature and spread of the lncRNA effect.

it is likely that many more DNA-RNA interactions will be identified that serve as molecular beacons to recruit specific protein complexes.

By contrast, the ability of Xist to localize to the inactive X chromosome depends on the ability of Xist to bind to the sequence-specific transcription factor, YY1 (100). When nascent Xist is transcribed, the interaction of the Xist repeat C region with YY1 on the inactive X chromosome captures and nucleates Xist. Conversely, ectopic insertion of multiple copies of the YY1 motif can mobilize Xist from the inactive X chromosome to the ectopic sites. Why Xist is not captured by the numerous YY1 sites on autosomes remains a mystery.

In addition to the four hypothesized targeting strategies, the recent example of HOTTIP introduced a new concept for lncRNA targeting via chromosomal looping (**Figure 4**) (96). Mature HOTTIP RNA appears to have no ability to seek out the *HOXA* locus if HOTTIP is ectopically produced elsewhere in the genome. But endogenous, nascent HOTTIP RNA is brought to its target genes via chromosomal looping. In this way, the lncRNA can serve as a faithful conduit to transform spatial information in chromosome conformation into chemical information in histone modifications.

Emerging Mechanistic Themes: Decoys, Scaffolds, and Guides

The ability of lncRNAs to bind to protein partners endows them with several regulatory capacities. Despite our limited knowledge from just dozens of characterized examples, several mechanistic themes of lncRNAs' functions have

emerged (101). Three main themes encompass many of the examples discussed thus far (**Figure 4**).

1. Decoys: First, and at the simplest level, lncRNAs can serve as decoys that preclude the access of regulatory proteins to DNA. For example, the lncRNA Gas5 is induced upon growth factor starvation; Gas5 contains a hairpin sequence motif that resembles the DNA-binding site of the glucocorticoid receptor (102). Thus, upon starvation conditions, Gas5 is induced and serves as a decoy to release the receptor from DNA to prevent transcription of metabolic genes. A more recent lncRNA decoy example was identified, termed PANDA, which associates with the transcription factor NF-YA to prevent p53-mediated apoptosis (47). NF-YA transactivates several key genes for apoptosis, but PANDA binding to NF-YA titrates NF-YA away from target gene chromatin.
2. Scaffold: The lncRNAs can serve as adaptors to bring two or more proteins into discrete complexes (103). The telomerase RNA TERC is a classic example of an RNA scaffold that assembles the telomerase complex (104). Another prime example of lncRNA scaffolds is HOTAIR, which can simultaneously bind both the PRC2 and the LSD1-CoREST complex via specific domains of the RNA structure (84). This combination of interactions coordinates H3K27 methylation and H3K4me2 demethylation, ensuring gene silencing. Additional examples include ANRIL, which combines the PRC2 and the PRC1 (91, 105); and Kcnq1ot1, which interacts with both the PRC2 and G9a to promote H3K27me3 and H3K9me3, two different silencing histone marks (88). Both of these combinations are likely to reinforce the transcriptionally silent state. Importantly, the concept of RNA as molecular scaffold is likely to generalize more globally as hundreds of lncRNAs have been identified to form ribonucleic-protein interactions with multiple protein partners (49, 52).
3. Guides: As described above (**Table 1**), many lncRNAs are individually required for the proper localization of specific protein complexes. The lncRNAs involved in dosage compensation and imprinting (Xist, Kcnq1ot1, Air) serve as guides to target gene silencing activity in an allele-specific fashion. HOTAIR also serves as a guide to localize the PRC2 in developmental and cancer-related gene expression (74, 83). As another example, lincRNA-p21 is directly induced by p53 upon DNA damage and, in turn, physically associates with nuclear factor hnRNP-K to reroute this protein to specific promoters (48). Guide lncRNAs thereby combine two basic molecular functions—binding of a protein partner plus a mechanism to interface with selective regions of the genome.

Decoy: the notion that lncRNAs can associate with DNA-binding proteins to prevent their binding to DNA recognition elements

Scaffold: the formation of lncRNA-RNPs where the RNA joins several proteins together in a complex

Guide: the formation of a lncRNA-RNP that imparts specificity to genomic locations through either DNA-protein or RNA-DNA recognition rules

The concept of guide lncRNAs has previously been parsed by whether the guidance occurs in *cis* (on neighboring genes) or in *trans* (on distantly located genes). The *cis* actors have been assumed to occur in a cotranscriptional manner, leading to the analogy of lncRNAs as tethers (106). But recent experiments, where ectopically supplied lncRNAs are shown to seek out their cognate target sites, show that even *cis*-acting lncRNAs have the capacity to act in *trans* (90, 99, 100). The *cis* actions also do not simply correlate with distance from the site of lncRNA synthesis (87, 96), perhaps reflecting the important role of chromosomal looping in so-called *cis* effects. Interestingly, all of the classic long and small ncRNAs that have been well characterized also work in *trans* through interactions with proteins (i.e., rRNA, tRNA, snoRNA, RNase P, TERC). Two groups have recently developed methods to map the genomic binding sites of specific lncRNAs in a comprehensive manner (106a, 106b). Future studies that allow global mapping of lncRNA

RNA immunoprecipitation (RIP) and cross-linked immunoprecipitation (CLIP): procedures to identify lncRNAs associated with protein coding genes such as PRC2

binding sites may better define the *cis* versus *trans* nature of lncRNA actions.

These three modes of action encompass many of the recently discovered lncRNA mechanistic themes, yet there are likely many other mechanisms to be uncovered. It is clear from the above examples that, as additional protein partners and targeting mechanisms (not just to DNA but perhaps also to other cellular structures) are discovered, it is possible to build complex regulatory scripts out of lncRNAs (107). Essential to understanding the meaning of such scripts will be a systematic understanding of the individual parts of lncRNA and their relevant interactions—akin to deciphering the codons of messenger RNAs.

THE GLOBAL RNA-CHROMATIN NETWORK

Sequencing of lncRNAs Bound to Proteins: RNA Immunoprecipitation and Cross-Linked Immunoprecipitation

Could the above described examples of Xist, HOTAIR, lincRNA-p21 and HOTTIP simply be a few quirky examples of lncRNAs interacting with chromatin-modifying complexes or are lncRNAs a more global phenomenon? Several recent studies point to the latter. These studies employ protein immunoprecipitation followed by microarray or deep sequencing to enumerate all RNAs associated with a protein complex of interest (RNA immunoprecipitation sequencing, RIP-seq) (49, 108, 109); often, cross-linking is performed to trap the relevant interactions in living cells (CLIP-seq) (110). For example, human PRC2 is associated with approximately 20% of lincRNAs expressed in a given cell type. Moreover, depletion of several PRC2-bound lincRNAs resulted in altered expression of PRC2-regulated gene loci in *trans* (49). Several of these PRC2-bound lincRNAs were independently identified as bound to the chromatin fraction (111), and a subset of these same lncRNAs are also associated with LSD1 complexes (49), raising the possibility of numerous lncRNA scaffolds. A similar study in mouse ESCs identified approximately 9,000 lncRNAs associated with PRC2 (109). Thus, numerous lncRNAs are associated with PRC2 and may serve as guides as do HOTAIR, Xist, Kcnq1ot1, and Air.

To address lncRNA-protein interactions at a more global level, a recent study systematically performed RIP in combination with LOF experiments by depleting the lincRNAs bound to a given chromatin-modifying complex, as well as depleting the chromatin-modifying complex itself (52). Interestingly, depletion of the lincRNAs associated with a given complex collectively phenocopied the depletion of the complex itself for PRC2 and several other complexes. These results strongly suggest that lincRNAs serve to modulate the targeting of the chromatin-modifying complex to specific genomic loci, which is an emerging mechanistic theme. Yet future studies will need to investigate the directness of these interactions and determine how lncRNAs may alter the activities of highly dynamic chromatin-modifying complexes.

Long Noncoding RNA Structure and Function

One of the most the intriguing features of RNA is the malleable adoption of secondary and tertiary structures that relate to function. Classic chemical probing and structural studies have resulted in a structural understanding of several lncRNAs, including the atomic structure of the largest known RNP complex–the ribosome (112). Several recent studies have developed genome-scale approaches to measure RNA secondary structures; these approaches also apply to lncRNAs (113). These studies use either chemical probes to acylate flexible RNA bases that are not participating in structural interactions or use specific enzymes that cleave structured and unstructured regions of RNAs. For example, a large-scale application of chemical probing followed by sequencing of

reverse-transcription products, termed SHAPE (selective 2′-hydroxyl acylation analyzed by primer extension), revealed the secondary structure of the entire RNA genome of the human immunodeficiency virus (114). More recently, deep sequencing of RNA fragments generated by enzymes that cleave single-stranded and double-stranded regions of RNA mapped the secondary structure of the entire budding yeast transcriptome, revealing several global structural properties (115). Notably, this study observed a triplet-based structural motif across gene bodies that correlates with translational efficiency. By contrast, 5′ and 3′ untranslated regions were observed to be much more lowly structured. In addition to the *Saccharomyces cerevisiae* transcriptome, this study also confirmed the known structural motifs and structural properties of the HOTAIR lncRNA. Two other studies successfully mapped the secondary structures of mouse small nuclear RNAs and compared wild-type and mutant RNase P (116, 117). With these new technologies in hand, it will be possible to gain a much needed understanding of the relationship between lncRNA structure and function, perhaps revealing common motifs of RNA structure that result in specific protein interactions or other functional properties.

Long Noncoding RNAs and Disease

Underscoring the importance of lncRNAs' regulatory roles is their emergence as key players in the etiology of several disease states (118). The strongest connection at present is with cancer (119). Dozens of lncRNAs have been documented to have altered expression in human cancers and are regulated by specific oncogenic and tumor-suppressor pathways, such as p53, MYC, and NF-κB (44, 47, 48). Hung et al. (47) recently described a class of lncRNAs that show periodic expression during the human cell cycle, and many of these are dysregulated in expression in human cancer samples. The lncRNA HOTAIR is highly induced in approximately one-quarter of human breast cancers, and HOTAIR expression is strongly predictive of eventual metastasis and death (74). HOTAIR overexpression in fact drives breast cancer metastasis in vivo, in part by relocalizing Polycomb occupancy patterns genome wide to alter the positional identity of cancer cells (74). Elevated HOTAIR level is also predictive of metastasis or progression in colon and liver cancers, suggesting a general oncogenic trait (120, 121). In effect, cancer cells reprogram themselves to act as if they belong in other anatomic sites (74). The concept of lncRNAs as disease markers is strongly bolstered by the notable discovery that lncRNAs, perhaps owing to their secondary structures, are stable in body fluids and enable noninvasive diagnoses (122). Chinnaiyan and colleagues (122) discovered a large set of lncRNAs in human prostate cancers by RNA-seq and also identified PCAT-1, a lncRNA involved in gene repression that can identify poor-prognosis patients on the basis of its level in urine. The human lncRNA ANRIL is located upstream of the *CDKN2A* tumor-suppressor locus encoding the p16 CDK inhibitor. Mutations in ANRIL are associated with cancer and cardiovascular disease and also lead to aberrant ANRIL transcripts and loss of p16 repression (123). Together, these examples illustrate diverse pathogenic mechanisms—from altering the epigenetic landscape (HOTAIR and ANRIL), modulation of the p53 pathway (lincRNA-p21 and PANDA), and alternative splicing that increases oncogenic protein production (Zeb2 antisense RNA) (124) to controlling the DNA damage response (*CCND1* promoter transcript) (132).

Although cancer has been the most studied, it is likely that lncRNAs are involved in the pathogenesis of many other diseases. Consistent with this notion, hundreds of genomic regions that do not contain protein-coding genes are strongly associated with a wide spectrum of human diseases. Future studies will need to pinpoint potential lncRNA transcripts in these regions and discern whether and how the noncoding genome contributes to human diseases.

SUMMARY POINTS

1. Several examples of functional lncRNAs have been identified that play key roles from pluripotency to cancer.
2. A common emerging theme of lncRNAs is that they form ribonucleic acid-protein interactions to carry out their functions by modulating chromatin-modifying complexes, by interaction with transcription factors, and likely by many additional mechanisms.
3. Enhancers transcribe RNAs. To date, two classes of enhancer RNAs have been identified: those that are by-products of transcription and lncRNAs that play a role in forming enhancer contacts to promote gene expression.

FUTURE ISSUES

1. Very little is understood about how specific lncRNAs seek out selective sites in the genome for interaction, the nature of lncRNA-chromatin interactions, and their possible functional roles in lncRNA biology.
2. Numerous annotation resources are available that will need to be compiled into parsimonious lncRNA transcript databases.
3. A deeper understanding of the sequence and structural elements that relate to lncRNA function will allow classification, or even prediction, of lncRNA families as has occurred for protein families with similar structural domains.
4. Research should identify potential ways of distinguishing *cis*-acting enhancer-like lncRNAs from lncRNAs that function in *trans*.
5. There is a need for large-scale LOF or gain-of-function studies to causally demonstrate lncRNA functions.
6. There is a need for detailed mapping and structural studies to understand the sequence and or structural basis of RNA-protein and RNA-DNA interactions.
7. The clear difference in conservation between protein-coding and lncRNA genes raises the question: How are these lncRNAs rapidly evolving?
8. Genomic regions genetically associated with disease contain only lncRNAs, pointing to their genetic importance to disease, yet the functional roles of lncRNA remain largely unresolved.
9. There is a clear need to develop genetic model systems to understand lncRNAs' function in vivo.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We apologize to colleagues whose work could not be discussed or cited owing to space limitations. J.L.R. is supported by National Institutes of Health (NIH) grants 1DP2OD00667-01

and 1P01GM099117-01. J.L.R. is a Damon Runyon-Rachleff, Searle, Smith Family, and Merkin Foundation scholar. H.Y.C. is supported by NIH grants R01-HG004361 and R01-CA118750 and by the California Institute for Regenerative Medicine. H.Y.C. is an Early Career Scientist of the Howard Hughes Medical Institute.

LITERATURE CITED

1. Jacob F, Monod J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* 3:318–56
2. Amaral PP, Dinger ME, Mercer TR, Mattick JS. 2008. The eukaryotic genome as an RNA machine. *Science* 319:1787–89
3. Ponting CP, Oliver PL, Reik W. 2009. Evolution and functions of long noncoding RNAs. *Cell* 136:629–41
4. Weinberg RA, Penman S. 1968. Small molecular weight monodisperse nuclear RNA. *J. Mol. Biol.* 38:289–304
5. Paul J, Duerksen JD. 1975. Chromatin-associated RNA content of heterochromatin and euchromatin. *Mol. Cell. Biochem.* 9:9–16
6. Salditt-Georgieff M, Harpold MM, Wilson MC, Darnell JE Jr. 1981. Large heterogeneous nuclear ribonucleic acid has three times as many 5′ caps as polyadenylic acid segments, and most caps do not enter polyribosomes. *Mol. Cell. Biol.* 1:179–87
7. Salditt-Georgieff M, Darnell JE Jr. 1982. Further evidence that the majority of primary nuclear RNA transcripts in mammalian cells do not contribute to mRNA. *Mol. Cell. Biol.* 2:701–7
8. Nickerson JA, Krochmalnic G, Wan KM, Penman S. 1989. Chromatin architecture and nuclear RNA. *Proc. Natl. Acad. Sci. USA* 86:177–81
9. Bernstein E, Duncan EM, Masui O, Gil J, Heard E, Allis CD. 2006. Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Mol. Cell. Biol.* 26:2560–69
10. Maison C, Bailly D, Peters AH, Quivy JP, Roche D, et al. 2002. Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. *Nat. Genet.* 30:329–34
11. Warner JR, Soeiro R, Birnboim HC, Girard M, Darnell JE. 1966. Rapidly labeled HeLa cell nuclear RNA. I. Identification by zone sedimentation of a heterogeneous fraction separate from ribosomal precursor RNA. *J. Mol. Biol.* 19:349–61
12. Bartel DP. 2009. MicroRNAs: target recognition and regulatory functions. *Cell* 136:215–33
13. Nagano T, Fraser P. 2011. No-nonsense functions for long noncoding RNAs. *Cell* 145:178–81
14. Bernstein E, Allis CD. 2005. RNA meets chromatin. *Genes Dev.* 19:1635–55
15. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. 2001. Identification of novel genes coding for small expressed RNAs. *Science* 294:853–58
16. Lau NC, Lim LP, Weinstein EG, Bartel DP. 2001. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* 294:858–62
17. Lee RC, Ambros V. 2001. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* 294:862–64
18. Lee RC, Feinbaum RL, Ambros V. 1993. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75:843–54
19. Bertone P, Stolc V, Royce TE, Rozowsky JS, Urban AE, et al. 2004. Global identification of human transcribed sequences with genome tiling arrays. *Science* 306:2242–46
20. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, et al. 2005. The transcriptional landscape of the mammalian genome. *Science* 309:1559–63
21. Kapranov P, Cawley SE, Drenkow J, Bekiranov S, Strausberg RL, et al. 2002. Large-scale transcriptional activity in chromosomes 21 and 22. *Science* 296:916–19
22. Rinn JL, Euskirchen G, Bertone P, Martone R, Luscombe NM, et al. 2003. The transcriptional activity of human Chromosome 22. *Genes Dev.* 17:529–40

23. Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, et al. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447:799–816
24. Ponjavic J, Ponting CP, Lunter G. 2007. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res.* 17:556–65
25. Shoemaker DD, Schadt EE, Armour CD, He YD, Garrett-Engele P, et al. 2001. Experimental annotation of the human genome using microarray technology. *Nature* 409:922–27
26. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921
27. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. 2001. The sequence of the human genome. *Science* 291:1304–51
28. Kawai J, Shinagawa A, Shibata K, Yoshino M, Itoh M, et al. 2001. Functional annotation of a full-length mouse cDNA collection. *Nature* 409:685–90
29. Schuler GD, Boguski MS, Stewart EA, Stein LD, Gyapay G, et al. 1996. A gene map of the human genome. *Science* 274:540–46
30. Dunham I, Shimizu N, Roe BA, Chissoe S, Hunt AR, et al. 1999. The DNA sequence of human chromosome 22. *Nature* 402:489–95
31. Volkin E, Astrachan L. 1956. Intracellular distribution of labeled ribonucleic acid after phage infection of *Escherichia coli*. *Virology* 2:433–37
32. Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, et al. 2003. A uniform system for microRNA annotation. *RNA* 9:277–79
33. Bartel DP. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116:281–97
34. Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, et al. 2005. Identification of hundreds of conserved and nonconserved human microRNAs. *Nat. Genet.* 37:766–70
35. Clark MB, Amaral PP, Schlesinger FJ, Dinger ME, Taft RJ, et al. 2011. The reality of pervasive transcription. *PLoS Biol.* 9:e1000625
36. Struhl K. 2007. Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat. Struct. Mol. Biol.* 14:103–5
37. van Bakel H, Nislow C, Blencowe BJ, Hughes TR. 2010. Most "dark matter" transcripts are associated with known genes. *PLoS Biol.* 8:e1000371
38. van Bakel H, Nislow C, Blencowe BJ, Hughes TR. 2011. Response to "the reality of pervasive transcription." *PLoS Biol.* 9:e1001102
39. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, et al. 2007. High-resolution profiling of histone methylations in the human genome. *Cell* 129:823–37
40. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, et al. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125:315–26
41. Rando OJ, Chang HY. 2009. Genome-wide views of chromatin structure. *Annu. Rev. Biochem.* 78:245–71
42. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, et al. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448:553–60
43. Marson A, Levine SS, Cole MF, Frampton GM, Brambrink T, et al. 2008. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* 134:521–33
44. Guttman M, Amit I, Garber M, French C, Lin MF, et al. 2009. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458:223–27
45. Khalil AM, Guttman M, Huarte M, Garber M, Raj A, et al. 2009. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. USA* 106:11667–72
45a. Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP. 2011. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* 147:1537–50
46. Loewer S, Cabili MN, Guttman M, Loh YH, Thomas K, et al. 2010. Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat. Genet.* 42:1113–17

47. Hung T, Wang Y, Lin MF, Koegel AK, Kotake Y, et al. 2011. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat. Genet.* 43:621–29
48. Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, et al. 2010. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* 142:409–19
49. Khalil AM, Guttman M, Huarte M, Garber M, Raj A, et al. 2009. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. USA* 106:11667–72
50. Koziol MJ, Rinn JL. 2010. RNA traffic control of chromatin complexes. *Curr. Opin. Genet. Dev.* 20:142–48
51. Rinn JL, Huarte M. 2011. To repress or not to repress: This is the guardian's question. *Trends Cell Biol.* 21:344–53
52. Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, et al. 2011. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* 477:295–300
53. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5:621–28
54. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. 2008. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* 18:1509–17
55. Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25:1105–11
56. Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, et al. 2010. Ab initio reconstruction of cell type–specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat. Biotechnol.* 28:503–10
57. Garber M, Grabherr MG, Guttman M, Trapnell C. 2011. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat. Methods* 8:469–77
58. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, et al. 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 25:1915–27
59. Beck AH, Weng Z, Witten DM, Zhu S, Foley JW, et al. 2010. 3′-end sequencing for expression quantification (3SEQ) from archival tumor samples. *PLoS ONE* 5:e8768
60. Jan CH, Friedman RC, Ruby JG, Bartel DP. 2011. Formation, regulation and evolution of *Caenorhabditis elegans* 3′UTRs. *Nature* 469:97–101
61. Core LJ, Waterfall JJ, Lis JT. 2008. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 322:1845–48
62. Lin MF, Carlson JW, Crosby MA, Matthews BB, Yu C, et al. 2007. Revisiting the protein-coding gene catalog of *Drosophila melanogaster* using 12 fly genomes. *Genome Res.* 17:1823–36
63. Lin MF, Deoras AN, Rasmussen MD, Kellis M. 2008. Performance and scalability of discriminative metrics for comparative gene identification in 12 *Drosophila* genomes. *PLoS Comput. Biol.* 4:e1000067
64. Pueyo JI, Couso JP. 2011. Tarsal-less peptides control Notch signalling through the Shavenbaby transcription factor. *Dev. Biol.* 355:183–93
65. 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
66. Wadler CS, Vanderpool CK. 2007. A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide. *Proc. Natl. Acad. Sci. USA* 104:20454–59
67. Dinger ME, Pang KC, Mercer TR, Mattick JS. 2008. Differentiating protein-coding and noncoding RNA: challenges and ambiguities. *PLoS Comput. Biol.* 4:e1000176
68. Leygue E. 2007. Steroid receptor RNA activator (SRA1): unusual bifaceted gene products with suspected relevance to breast cancer. *Nucl. Recept. Signal.* 5:e006
69. Jenny A, Hachet O, Zavorszky P, Cyrklaff A, Weston MD, et al. 2006. A translation-independent role of *oskar* RNA in early *Drosophila* oogenesis. *Development* 133:2827–33
70. Kloc M, Wilk K, Vargas D, Shirato Y, Bilinski S, Etkin LD. 2005. Potential structural role of non-coding and coding RNAs in the organization of the cytoskeleton at the vegetal cortex of *Xenopus* oocytes. *Development* 132:3445–57

71. Candeias MM, Malbert-Colas L, Powell DJ, Daskalogianni C, Maslon MM, et al. 2008. *p53* mRNA controls p53 activity by managing Mdm2 functions. *Nat. Cell Biol.* 10:1098–105
72. Mercer TR, Dinger ME, Sunkin SM, Mehler MF, Mattick JS. 2008. Specific expression of long noncoding RNAs in the mouse brain. *Proc. Natl. Acad. Sci. USA* 105:716–21
73. Dinger ME, Amaral PP, Mercer TR, Pang KC, Bruce SJ, et al. 2008. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res.* 18:1433–45
74. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, et al. 2010. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464:1071–76
75. Broadbent KM, Park D, Wolf AR, Van Tyne D, Sims JS, et al. 2011. A global transcriptional analysis of *Plasmodium falciparum* malaria reveals a novel family of telomere-associated lncRNAs. *Genome Biol.* 12:R56
76. Britten RJ, Davidson EH. 1969. Gene regulation for higher cells: a theory. *Science* 165:349–57
77. Brown CJ, Ballabio A, Rupert JL, Lafreniere RG, Grompe M, et al. 1991. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* 349:38–44
78. Barlow DP, Stoger R, Herrmann BG, Saito K, Schweifer N. 1991. The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the *Tme* locus. *Nature* 349:84–87
79. Bartolomei MS, Zemel S, Tilghman SM. 1991. Parental imprinting of the mouse *H19* gene. *Nature* 351:153–55
80. Wutz A. 2011. Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. *Nat. Rev. Genet.* 12:542–53
81. 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* 322:750–56
82. Heo JB, Sung S. 2010. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* 331:76–79
83. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, et al. 2007. Functional demarcation of active and silent chromatin domains in human *HOX* loci by noncoding RNAs. *Cell* 129:1311–23
84. Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, et al. 2010. Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 329:689–93
85. Kaneko S, Li G, Son J, Xu CF, Margueron R, et al. 2010. Phosphorylation of the PRC2 component Ezh2 is cell cycle–regulated and up-regulates its binding to ncRNA. *Genes Dev.* 24:2615–20
86. Maenner S, Blaud M, Fouillen L, Savoye A, Marchand V, et al. 2010. 2-D structure of the A region of Xist RNA and its implication for PRC2 association. *PLoS Biol.* 8:e1000276
87. Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, et al. 2008. The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* 322:1717–20
88. Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, et al. 2008. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol. Cell* 32:232–46
89. Law JA, Jacobsen SE. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* 11:204–20
90. Schmitz KM, Mayer C, Postepska A, Grummt I. 2010. Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes Dev.* 24:2264–69
91. Yap KL, Li S, Muñoz-Cabello AM, Raguz S, Zeng L, et al. 2010. Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol. Cell* 38:662–74
92. Yao H, Brick K, Evrard Y, Xiao T, Camerini-Otero RD, Felsenfeld G. 2010. Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. *Genes Dev.* 24:2543–55
93. Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, et al. 2010. Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465:182–87
94. De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, et al. 2010. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol.* 8:e1000384

95. Ørom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, et al. 2010. Long noncoding RNAs with enhancer-like function in human cells. *Cell* 143:46–58
96. Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, et al. 2011. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 472:120–24
97. Hung T, Chang HY. 2010. Long noncoding RNA in genome regulation: prospects and mechanisms. *RNA Biol.* 7:582–85
98. Bonasio R, Tu S, Reinberg D. 2010. Molecular signals of epigenetic states. *Science* 330:612–16
99. Martianov I, Ramadass A, Serra Barros A, Chow N, Akoulitchev A. 2007. Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature* 445:666–70
100. Jeon Y, Lee JT. 2011. YY1 tethers Xist RNA to the inactive X nucleation center. *Cell* 146:119–33
101. Wang KC, Chang HY. 2011. Molecular mechanisms of long noncoding RNAs. *Mol. Cell* 43:904–14
102. 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
103. Spitale RC, Tsai MC, Chang HY. 2011. RNA templating the epigenome: long noncoding RNAs as molecular scaffolds. *Epigenetics* 6:539–43
104. Zappulla DC, Cech TR. 2006. RNA as a flexible scaffold for proteins: yeast telomerase and beyond. *Cold Spring Harb. Symp. Quant. Biol.* 71:217–24
105. Kotake Y, Nakagawa T, Kitagawa K, Suzuki S, Liu N, et al. 2010. Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of $p15^{INK4B}$ tumor suppressor gene. *Oncogene* 30:1956–62
106. Lee JT. 2009. Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome. *Genes Dev.* 23:1831–42
106a. Chu C, Qu K, Zhong FL, Artandi SE, Chang HY. 2011. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol. Cell* 44:667–78
106b. Simon MD, Wang CI, Kharchenko PV, West JA, Chapman BA, et al. 2011. The genomic binding sites of a noncoding RNA. *Proc. Natl. Acad. Sci. USA* 08:20497–502
107. Delebecque CJ, Lindner AB, Silver PA, Aldaye FA. 2011. Organization of intracellular reactions with rationally designed RNA assemblies. *Science* 333:470–74
108. Gerber AP, Herschlag D, Brown PO. 2004. Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast. *PLoS Biol.* 2:E79
109. Zhao J, Ohsumi TK, Kung JT, Ogawa Y, Grau DJ, et al. 2010. Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol. Cell* 40:939–53
110. Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, et al. 2008. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* 456:464–69
111. Mondal T, Rasmussen M, Pandey GK, Isaksson A, Kanduri C. 2010. Characterization of the RNA content of chromatin. *Genome Res.* 20:899–907
112. Steitz TA. 2008. A structural understanding of the dynamic ribosome machine. *Nat. Rev. Mol. Cell Biol.* 9:242–53
113. Wan Y, Kertesz M, Spitale RC, Segal E, Chang HY. 2011. Understanding the transcriptome through RNA structure. *Nat. Rev. Genet.* 12:641–55
114. Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW Jr, et al. 2009. Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* 460:711–16
115. Kertesz M, Wan Y, Mazor E, Rinn JL, Nutter RC, et al. 2010. Genome-wide measurement of RNA secondary structure in yeast. *Nature* 467:103–7
116. Underwood JG, Uzilov AV, Katzman S, Onodera CS, Mainzer JE, et al. 2010. FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat. Methods* 7:995–1001
117. Lucks JB, Mortimer SA, Trapnell C, Luo S, Aviran S, et al. 2011. Multiplexed RNA structure characterization with selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc. Natl. Acad. Sci. USA* 108:11063–68
118. Wapinski O, Chang HY. 2011. Long noncoding RNAs and human disease. *Trends Cell Biol.* 21:354–61
119. Tsai MC, Spitale RC, Chang HY. 2011. Long intergenic noncoding RNAs: new links in cancer progression. *Cancer Res.* 71:3–7

120. Kogo R, Shimamura T, Mimori K, Kawahara K, Imoto S, et al. 2011. Long non-coding RNA HOTAIR regulates Polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. *Cancer Res.* 71:6320–26
121. Yang Z, Zhou L, Wu LM, Lai MC, Xie HY, et al. 2011. Overexpression of long non-coding RNA HOTAIR predicts tumor recurrence in hepatocellular carcinoma patients following liver transplantation. *Ann. Surg. Oncol.* 18:1243–50
122. Prensner JR, Iyer MK, Balbin OA, Dhanasekaran SM, Cao Q, et al. 2011. Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nat. Biotechnol.* 29:742–49
123. Burd CE, Jeck WR, Liu Y, Sanoff HK, Wang Z, Sharpless NE. 2010. Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk. *PLoS Genet.* 6:e1001233
124. Beltran M, Puig I, Peña C, Garcia JM, Alvarez AB, et al. 2008. A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. *Genes Dev.* 22:756–69
125. Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, et al. 2010. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell* 39:925–38
126. Gong C, Maquat LE. 2011. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3′ UTRs via Alu elements. *Nature* 470:284–88
127. van Dijk EL, Chen CL, d'Aubenton-Carafa Y, Gourvennec S, Kwapisz M, et al. 2011. XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature* 475:114–17
128. Berretta J, Pinskaya M, Morillon A. 2008. A cryptic unstable transcript mediates transcriptional *trans*-silencing of the Ty1 retrotransposon in *S. cerevisiae*. *Genes Dev.* 22:615–26
129. Camblong J, Beyrouthy N, Guffanti E, Schlaepfer G, Steinmetz LM, Stutz F. 2009. *Trans*-acting antisense RNAs mediate transcriptional gene cosuppression in *S. cerevisiae*. *Genes Dev.* 23:1534–45
130. Camblong J, Iglesias N, Fickentscher C, Dieppois G, Stutz F. 2007. Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in *S. cerevisiae*. *Cell* 131:706–17
131. Flynn RL, Centore RC, O'Sullivan RJ, Rai R, Tse A, et al. 2011. TERRA and hnRNPA1 orchestrate an RPA-to-POT1 switch on telomeric single-stranded DNA. *Nature* 471:532–36
132. Wang X, Arai S, Song X, Reichart D, Du K, et al. 2008. Induced ncRNAs allosterically modify RNA-binding proteins in *cis* to inhibit transcription. *Nature* 454:126–30

Annual Review of Biochemistry
Volume 81, 2012

Contents

Recent Advances in Biochemistry

Indexes

Errata

An online log of corrections to *Annual Review of Biochemistry* articles may be found at http://biochem.annualreviews.org/errata.shtml