From Discovery to Function: The Expanding Roles of Long NonCoding RNAs in Physiology and Disease

Miao Sun and W. Lee Kraus

Laboratory of Signaling and Gene Regulation, Cecil H. and Ida Green Center for Reproductive Biology Sciences and Division of Basic Reproductive Biology Research, Department of Obstetrics and Gynecology, University of Texas Southwestern Medical Center, Dallas, Texas 75390

Long noncoding RNAs (lncRNAs) are a relatively poorly understood class of RNAs with little or no coding capacity transcribed from a set of incompletely annotated genes. They have received considerable attention in the past few years and are emerging as potentially important players in biological regulation. Here we discuss the evolving understanding of this new class of molecular regulators that has emerged from ongoing research, which continues to expand our databases of annotated lncRNAs and provide new insights into their physical properties, molecular mechanisms of action, and biological functions. We outline the current strategies and approaches that have been employed to identify and characterize lncRNAs, which have been instrumental in revealing their multifaceted roles ranging from *cis*- to *trans*-regulation of gene expression and from epigenetic modulation in the nucleus to posttranscriptional control in the cytoplasm. In addition, we highlight the molecular and biological functions of some of the best characterized lncRNAs in physiology and disease, especially those relevant to endocrinology, reproduction, metabolism, immunology, neurobiology, muscle biology, and cancer. Finally, we discuss the tremendous diagnostic and therapeutic potential of lncRNAs in cancer and other diseases. **(*Endocrine Reviews* 36: 25–64, 2015)**

ISSN Print 0163-769X ISSN Online 1945-7189
Printed in U.S.A.

Received April 29, 2014. Accepted November 21, 2014.
First Published Online November 26, 2014

Abbreviations: AD, Alzheimer's disease; agRNA, antigene RNA; ALC-1, atrial myosin light chain 1; AR, androgen receptor; ceRNA, competing endogenous RNA; CHART, capture hybridization analysis of RNA targets; ChIP-seq, chromatin immunoprecipitation sequencing; ChIRP, chromatin isolation by RNA purification; ciRNA, circular intronic long ncRNA; circRNA, circular RNA; CoREST, RE1-silencing transcription factor corepressor 1; CPC, coding potential calculator; CSF, codon substitution frequency; CTCF, CCCTC-binding factor; Dlx, distal-less homeobox; DMD, Duchenne muscular dystrophy; DNMT1, DNA (cytosine-5)-methyltransferase 1; eRNA, enhancer RNA; ESC, embryonic stem cell; FSHD, facioscapulohumeral muscular dystrophy; GABA, γ-aminobutyric acid; GR, glucocorticoid receptor; GRO-seq, global nuclear run-on sequencing; HD, Huntington's disease; hnRNP, heterogeneous nuclear ribonucleoprotein; HuR, human antigen R; lincRNA, long intergenic ncRNA; lncRNA, long ncRNA; LSD1, lysine-specific demethylase 1; β-MHC, myosin heavy chain β; MLL, mixed-lineage leukemia protein; MYH, myosin heavy chain; NAT, natural antisense transcript; ncRNA, noncoding RNA; NF, nuclear factor; nt, nucleotide; ORF, open reading frame; Pol II, polymerase II; 3P-seq, polyadenosine position profiling by sequencing; PPARγ, peroxisome proliferator-activated receptor γ; PR, progesterone receptor; PRC2, Polycomb repressive complex 2; PWS, Prader-Willi syndrome; RAP, RNA antisense purification; RIP, RNA immunoprecipitation; rRNA, ribosomal RNA; SCA, spinocerebellar ataxia; snoRNA, small nucleolar RNA; SRA, steroid receptor RNA activator; STAU1, Staufen double-stranded RNA binding protein 1; TSS, transcription start site; WDR5, WD repeat-containing protein 5; XCI, X-chromosome inactivation.

I. Introduction

Genome-wide transcriptome analyses conducted over the past decade, including recent studies by the ENCODE (Encyclopedia of DNA Elements) Consortium, have revealed that mammalian genomes are pervasively, but not indiscriminately, transcribed, giving rise to a wide variety of coding and noncoding RNA (ncRNA) transcripts (1–3). The cellular repertoire of ncRNAs consists of small housekeeping RNAs such as ribosomal RNAs (rRNAs) and transfer RNAs, microRNAs, and long ncRNAs (lncRNAs) including antisense RNAs and enhancer RNAs (eRNAs). The functions of many of these ncRNAs are poorly understood, but interests in uncovering their biological functions and molecular mechanisms of action are intense. In this review, we focus on lncRNAs, presenting the most current information on their discovery, annotation, molecular actions, and biological functions, especially as they relate to hormonal signaling systems.

Figure 1.

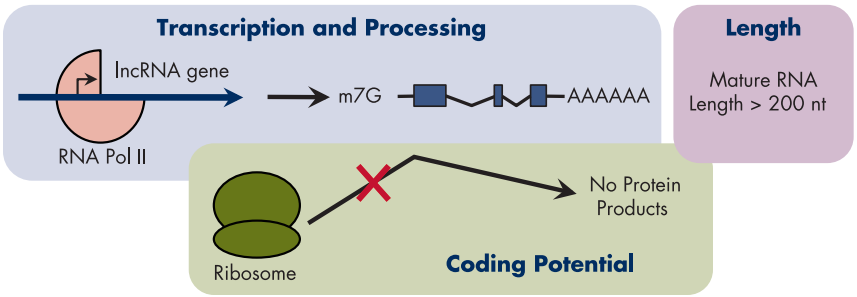

Figure 1. Molecular features of lncRNAs. LncRNAs are generally, but not exclusively, transcribed by RNA Pol II, spliced, 5′-capped (m7G), and 3′-polyadenylated (AAAAAA). By definition, they have a mature length of >200 nt, and low or no coding potential.

II. Defining LncRNAs

LncRNAs, defined as non-protein-coding RNA transcripts longer than 200 nucleotides (nt), are emerging as key regulators of diverse cellular processes (4–12). To date, a limited, but fast-growing number of lncRNAs have been functionally characterized through gene-specific studies. To further expand our understanding of lncRNAs, rapid advancements in genomic methods and analyses have spearheaded recent efforts in the large-scale identification of lncRNAs across multiple biological systems. Nevertheless, accurate identification demands a clear definition and sufficient knowledge of the features of lncRNAs.

A. An evolving definition of lncRNAs

The definition of lncRNAs continues to evolve. A universal classification scheme does not exist, and there have been various synonyms describing either very similar or slightly differing lncRNA-like molecules, adding to the confusion. The basic features are represented in the name lncRNA: they are obligate ncRNAs and are relatively long (>200 nt) (4, 7, 8, 10, 13–17) (Figure 1). Some definitions include an intergenic feature (ie, long intergenic ncRNA [lincRNAs]; by definition, they do not overlap in any way with annotated protein-coding transcription units) (9, 18–22) (Figure 2A).

1. Length

Although the current pool of known lncRNAs display a wide range of transcript length (13), the lower bound for long is somewhat arbitrarily set to be greater than 200 nt in an attempt to facilitate distinction from most other well-characterized groups of small ncRNA transcripts, such as rRNAs, transfer RNAs, small nuclear RNAs, small nucleolar RNAs (snoRNAs), and microRNAs. This length was chosen for practical considerations as well, because this threshold allows empirical separation of RNAs in common experimental procedures. The 200-nt cutoff, however, does not make clear biological distinctions, creating potential gray areas in our understanding.

2. Coding potential

The absolute requirement for being noncoding also invites controversy. Some studies have suggested that ncRNAs may engage ribosomes and produce small polypeptides (23); others have suggested that lncRNAs do not encode proteins (24). Of course, a lncRNA may code

Figure 2.

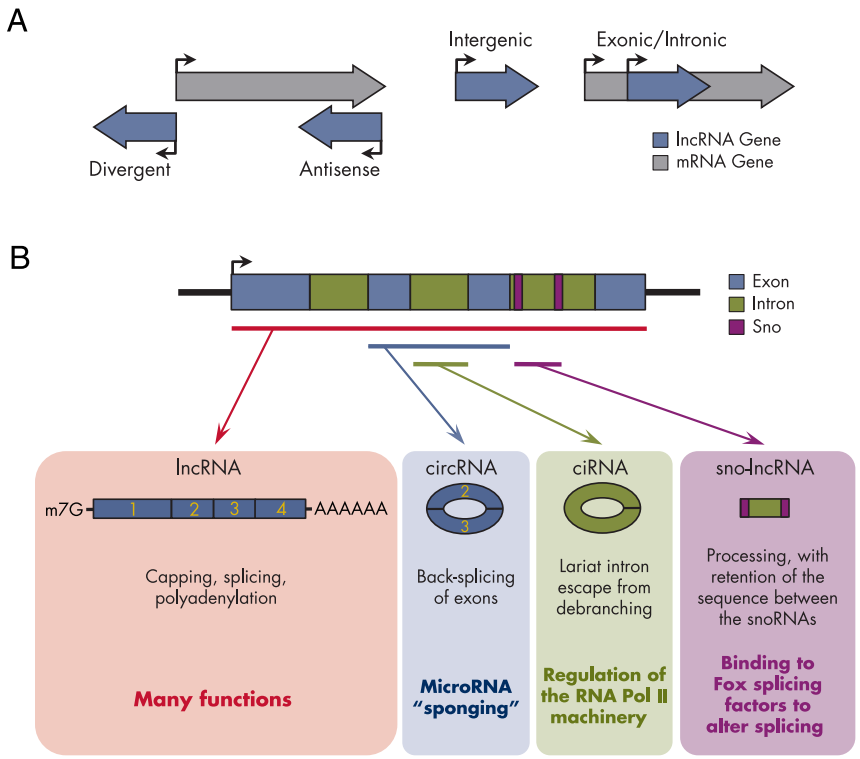

Figure 2. Biogenesis of lncRNAs. A, LncRNAs can be intergenic or genic (approximately one-third to one-half of lncRNAs overlap a protein-coding gene). Some intergenic lncRNAs are transcribed divergently to a protein-coding gene. Genic lncRNAs can be further divided into those that overlap a protein-coding gene in the sense vs antisense direction, and overlap exonic or intronic regions of a protein-coding gene. B, Many lncRNAs are transcribed and processed like mRNAs, whereas others originate from atypical processing of RNA transcripts. CircRNAs originate from back-spliced exons, whereas ciRNAs originate from lariat introns that escape from debranching. Sno-lncRNAs are processed on both ends by the snoRNA (sno) machinery, but retain the sequences between the snoRNAs, leading to the production of lncRNAs flanked by snoRNA sequences on either side, but lacking 5′-caps and 3′-polyadenosine tails.

for a polypeptide but also have coding-independent functions, as shown for the steroid receptor RNA activator (*SRA*), a well-characterized lncRNA involved in the nuclear receptor-mediated regulation of gene expression (described in *Section II.A.1*) (25). The *SRA* gene produces a functional ncRNA as well as a protein-coding variant (26). Therefore, instead of excluding any lncRNA-like transcripts due to a potential to code for a polypeptide product, a more reasonable approach may be to use a definition of noncoding that focuses on a coding-independent functional role of the untranslated RNA transcript. Thus, the key feature is that a lncRNA must function as an RNA transcript, whether or not it may also code for a polypeptide.

Interestingly, recent studies in yeast, flies, and fish have suggested that short polypeptides comprising a few to tens of amino acids encoded by short open reading frames (ORFs), like those found in mammalian lncRNAs, may have cellular functions (27–34). For example, the *ELA (*aka *Toddler*) gene encodes a conserved hormone, ELABELA, of 32 amino acids in zebrafish that acts through the G protein-coupled apelin receptor (29, 32). ELABELA, which is also expressed in human embryonic stem cells (ESCs), appears at the onset of zebrafish zygotic transcription and is required for early cardiovascular development (29). The extent to which short polypeptides encoded by short ORFs contribute to the function of lncRNAs, however, has yet to be determined.

3. Transcription and processing

In many respects, lncRNAs resemble protein-coding mRNAs; they are generally, but not exclusively, spliced, 5′-capped and 3′-polyadenylated, and transcribed by RNA polymerase II (Pol II) (13, 16, 18, 35, 36) (Figure 1). The results of a recent study, however, suggest that transcription is quantitatively different for lncRNAs and mRNAs, with transcription of the former being controlled by a canonical DGCR8 (DiGeorge syndrome critical region gene 8)-Dicer-microRNA pathway that supports robust transcriptional initiation and elongation (37). Pol II is likely responsible for the transcription of most lncRNA genes due to its higher processivity, whereas RNA Pols I and III are generally limited to the transcription of shorter housekeeping RNA transcripts. The polyadenylation of lncRNAs is consistent with transcription by Pol II, and it helps to stabilize the transcripts to preserve their functional roles. Nonetheless, nonpolyadenylated, Pol III-transcribed, ncRNA transcripts, such as *BC200* (38) and *asOct4-pg5* (39), have been identified. Both are functional RNAs, playing roles in the regulation of translation and chromatin structure respectively, and are commonly referred to as lncRNAs in the literature. Although *BC200* is 200 nt long, barely fulfilling the minimum length requirement of lncRNAs, the actual length of *asOct4-pg5* has not been evaluated and may be even shorter than 200 nt. Thus, the notion that Pol I and Pol III transcripts are too short to meet the length criterion of a lncRNA may still hold true; *BC200* may just be a rare

exception that marginally escapes the arbitrary length cutoff.

Some lncRNAs may originate from atypical processing of RNA transcripts. Recent studies have identified circular lncRNAs, including circular RNAs (circRNAs) originating from back-spliced exons (40–43) and circular intronic long ncRNAs (ciRNAs) originating from lariat introns that escape from debranching (44) (Figure 2B). CircRNAs are thought to antagonize the actions of microRNAs, whereas ciRNAs may act to regulate the RNA Pol II transcription machinery. For example, the ciRNA *ci-ankrd52* accumulates at its own sites of transcription and positively regulates Pol II transcription (44). The sno-lncRNAs represent another type of nuclear-enriched intron-derived lncRNA (45). They are processed on both ends by the snoRNA machinery, but retain the sequences between the snoRNAs, leading to the production of lncRNAs flanked by snoRNA sequences on either side, but lacking 5′-caps and 3′-polyadenosine tails (45) (Figure 2B). As these examples illustrate, lncRNA biogenesis occurs through multiple distinct mechanisms, which may direct specific functional outcomes.

4. Gene location and orientation

Historically, the focus has been on those lncRNAs encoded by genes that are well separated from genes encoding known protein-coding transcripts (Figure 2A), hence the name lincRNAs, as noted in *Section II.A* (9, 18–21). Nonetheless, as discovered in the large-scale discovery efforts noted in *Section III.A*, genic lncRNAs are emerging as a prevalent class, with approximately one-third to one-half of lncRNAs overlapping protein-coding genes (13, 36, 46) (Figure 2A). Genic lncRNAs can be further divided into those that overlap protein-coding loci in the sense vs antisense direction, and overlap exonic or intronic regions of the protein-coding gene (Figure 2A). More specifically, transcripts running on the opposite strand of protein loci form an abundant class of lncRNAs often known as natural antisense transcripts (NATs) (47, 48). Although a number of these NATs have been demonstrated to play a repressive role to regulate the expression of their sense mRNAs (49–53), the functional consequences of these distinctions associated with gene locations, at a global level, are unclear.

5. Conservation and evolution

Although some lncRNAs are conserved across related species, others lack strong evidence of conserved homologs (54). In fact, mammalian lncRNAs lack known orthologs in species outside of vertebrates (9). In such cases, positional and structural conservation may be more important than sequence conservation, as was recently shown in a study comparing lncRNAs in zebrafish and humans (55). Despite their rapid evolution, lncRNAs exhibit detectable signatures of natural selection, although these are weak (9). Tracking the evolutionary history of lncRNA genes from ancient to more recent species may provide information about the functions of the genes and how they may have changed over time. For example, one may be able to track across species the initial event of spurious transcription that gives rise to the birth of a lncRNA gene, which may gain, and perhaps subsequently lose, coding potential over evolutionary time (9, 28, 56).

B. A Working Definition of LncRNAs

As illustrated here, questions remain regarding a unifying definition for lncRNAs. The field, however, has reached the point of having a solid working definition for lncRNAs. For the purpose of convenience and simplicity in identifying lncRNAs and distinguishing them from other major classes of RNA transcripts, RNA molecules longer than 200 nt and having little coding potential are often classified as lncRNAs. They are very likely transcribed by Pol II and, in many cases, are capped, spliced, and polyadenylated.

III. Identifying and Cataloging LncRNAs

The earliest efforts to identify lncRNAs were mostly gene-specific, starting with the discovery of a novel transcript associated with a specific biological function and followed by the surprising realization that the function of the transcript is independent of the production of a protein product. More recently, significant advances in high-throughput sequencing technology and bioinformatics have revolutionized ncRNA discovery (Figure 3). Consistent with the definition of lncRNAs, the general strategy involves 2 major steps: 1) the identification of novel transcripts that pass the 200-nt length threshold and 2) evaluation of their coding potential. The newly acquired information has been consolidated into public databases, thus feeding back into the discovery process to facilitate identification of greater number of lncRNAs with higher confidence.

A. Identification of lncRNA transcripts: omics approaches

A number of different groups and consortia have used high-throughput sequencing technology and bioinformatics to facilitate ncRNA discovery.

1. cDNA cloning

RIKEN's FANTOM (Functional Annotation of the Mammalian Genome) consortium pioneered the genome-

Figure 3.

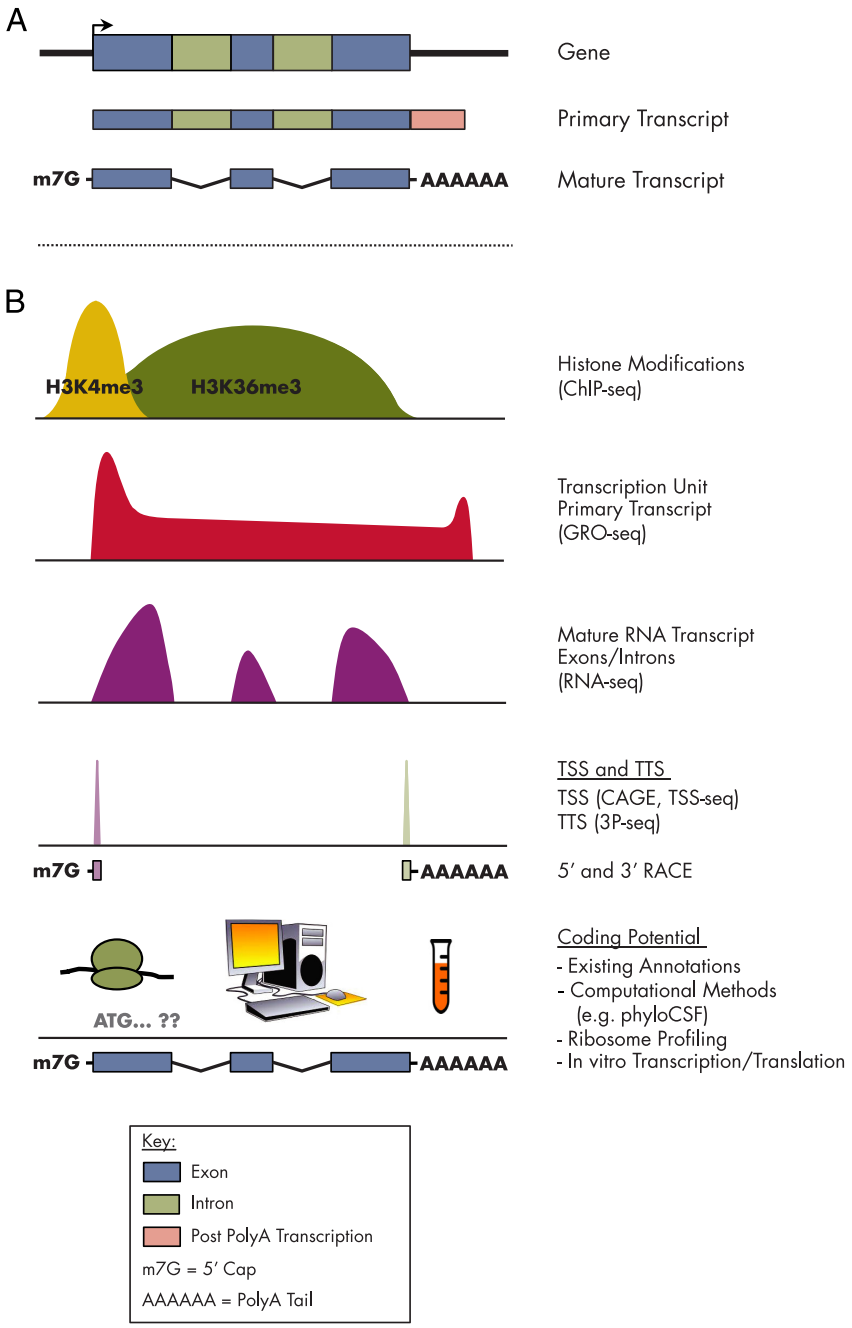

Figure 3. Omics approaches for identifying and annotating lncRNAs. A, Primary lncRNA transcripts are produced from lncRNA genes and are further processed to mature lncRNA transcripts. B, A variety of omics approaches have been used to identify and annotate lncRNA genes and transcripts, including ChIP-seq for histone modifications (H3K4me3, which marks active promoters, and H3K36me3, which marks transcribed gene bodies), GRO-seq, RNA-seq, and others, as illustrated. Abbreviations: TTS, transcription termination site; CAGE, cap analysis of gene expression; RACE, rapid amplification of cDNA ends.

wide discovery of lncRNAs, publishing a set of 34 030 polyadenylated lncRNAs from the mouse in 2005 (57). In addition, they isolated and cloned mouse full-length cDNA libraries for 5′- and 3′-sequencing and developed their own bioinformatics methods to map these transcripts to the mouse genome, resulting in 102 281 cDNAs as the starting point of lncRNA identification. To evaluate the coding potential of these cDNA transcripts, they searched for the presence of 1) protein-domain-like regions from Pfam (58) and SUPERFAMILY databases (59, 60) and 2) transmembrane regions predicted by the TMHMM program (61), coiled-coil regions predicted by the NCOIL program, and signal peptides predicted by the SignalP program (62). The absence of such protein-domain-like regions and the lack of an ORF longer than 100 amino acids were used to annotate one-third of the cDNA transcripts as lncRNAs.

2. Histone modification signatures

In 2009, Guttman and colleagues (18) proposed a different strategy that used global histone modification signatures to identify novel lncRNAs (Figure 3B). Using this approach, lncRNAs were defined as polyadenylated Pol II transcripts whose entire transcription units are longer than 5 kb and are well separated from known protein-coding and microRNA genes (the lincRNA definition provided in *Section II.A*). Using chromatin immunoprecipitation followed by massively parallel sequencing (ChIP-seq), the authors generated genome-wide histone modification maps, focusing on those signatures associated with Pol II transcription (ie, H3K4me3 at the promoter and H3K4me36 along the gene body), which served as markers of transcription units genome-wide. Polyadenylated exons were identified by microarray analyses with polyadenylated RNA across a random sample of 350 regions out of all 1675 transcription units identified. Due to a lack of splicing information, a conservative length cutoff of 5 kb was used to fulfill the length requirement of >200 nt. Codon substitution frequency (CSF)

(63), a measure of coding potential that examines evolutionary signatures characteristic to alignments of conserved coding regions, was evaluated for all intergenic transcripts in all 3 reading frames to confirm that most of the putative transcripts lack significant coding potential. Using this approach, the authors were able to identify approximately 1600 putative lncRNAs across 4 mouse cell types and about 3300 lncRNAs across 6 human cell types. An additional study has shown that histone modification signatures around transcription start sites may distinguish functionally distinct classes of lncRNAs (64).

3. Identification of transcription units

Another approach that can be used to identify novel ncRNAs is the identification of transcription units. Methods that detect the densities of elongating RNA polymerases along the genome, such as global nuclear run-on sequencing (GRO-seq) (65, 66) or native elongating transcript sequencing (67), can be used to define the transcription units and serve as the basis for transcript discovery (Figure 3B). A modified version of GRO-seq that incorporates initial steps of rapid amplification of 5′-ends (68) is useful in determining the exact transcription start sites (TSSs) of all transcripts. Additional genome-wide methods that facilitate the determination of TSSs include new cap-analysis gene expression (69) and TSS-seq (70, 71) (Figure 3B). Gene identification signature and gene signature cloning ditag technologies, as shown in FANTOM3 (57), can be used for the identification of sequences corresponding to both the TSS and the transcription termination sites. A method known as polyadenosine position profiling by sequencing (3P-Seq) can also be used to more precisely determine the directionality and end position of the polyadenylated transcription units (55, 72, 73) (Figure 3B). Nevertheless, although these methods delineate the transcriptional landscape of potential lncRNA genes, information from RNA-seq will still be essential in elucidating the structure of the mature RNA transcripts contained within the corresponding transcription units, which will in turn reveal the exact reading frame, allowing subsequent evaluation of coding potential.

4. Mature RNA structure

Characterization of the exon structure of lncRNAs has been facilitated by the development of bioinformatics algorithms that perform ab initio transcriptome reconstruction. Using programs such as Cufflinks (74, 75) or Scripture (19), entire transcriptomes of mammalian cells, and of cells extracted from multiple organs and various species, can be reconstructed using only RNA-seq reads and the genome sequence. RNA-seq reads directly reflect the position and structure of mature RNA transcripts (Figure 3). Compared with histone modification signature-based transcript determination, RNA-seq analysis gives a more accurate measurement of the length of mature RNA transcripts, and information about exon-intron structure reveals the actual reading frame, allowing for more accurate calculation of coding potential. In the initial report using Scripture for transcript annotation, the authors identified over 1000 novel lncRNAs in 3 mouse cell types (19). These lncRNAs are polyadenylated and multiexonic, and have an average mature transcript length of 859 nt with very low coding potential. A recent report describing the use of Cufflinks as an initial step in lncRNA annotation with subsequent incorporation of evolutionary information identified over 13 500 polyadenylated lncRNAs across 11 tetrapod species (75). The authors examined lncRNAs in a number of nonmodel organisms and further expanded the repertoires of lncRNAs.

Both Cufflinks and Scripture have also been used to assemble transcripts from RNA-seq datasets of very high sequencing depth in an attempt to accurately identify comprehensive lists of lncRNAs. One study examined lncRNAs in 24 human tissues and cell types and cataloged the results in the Human Body Map lncRNA database (20). Another study looked across 8 time points during zebrafish embryogenesis (76). The combined use of 2 independent assembly programs, together with high sequencing depth on multiple cell types or across multiple developmental stages strengthens the confidence of the discovery process, especially because lncRNAs, as a group, have low expression levels, are highly cell-type-specific, and are tightly regulated during development. In both studies, low CSF scores and the absence of Pfam domains were absolutely required for designation as a lncRNA, introducing extra criteria to ensure the noncoding status of identified lncRNAs.

5. Integration of approaches

Researchers have developed and improvised a variety of strategies to identify and annotate lncRNAs genome-wide. Moreover, they have integrated elements from these pipelines to facilitate lncRNA discovery. For example, Sigova and colleagues (36) assembled transcripts from RNA-seq reads but added the requirement of H3K4me3 enrichment to indicate the presence of high-confidence TSSs. In an effort to identify lncRNAs genome-wide in zebrafish, Ulitsky and colleagues (55) also used H3K4me3 and H3K36me3 to mark promoters and gene bodies but supplemented the histone modification maps with 3P-seq to more precisely map the polyadenylated end positions. They also incorporated existing transcriptome datasets, such as RNA-seq, annotated expressed sequence tags, and full-length cDNAs, to partially compensate for the lack of

accurate mature RNA structures. A coding potential calculator (CPC) (77) was used to determine the coding potential of each transcript. Collectively, the authors bioinformatically integrated multiple genomic datasets and identified 550 distinct lncRNAs in zebrafish.

B. Evaluation of coding potential

By definition, lncRNAs are unable to code for proteins. Determining the coding potential of a lncRNA, however, can be difficult. Three determinants have commonly been used for distinguishing ncRNAs from all identified RNAs: the length of the longest ORF, the bioinformatically calculated coding potential, and the presence of coding potential for conserved protein domains. Among them, calculation of coding potential is the least straightforward. It involves the analysis of DNA alignments and codon usage across multiple species, favoring changes in amino acids that will preserve structural similarity vs changes that may lead to dramatic alterations in protein structure. In addition to the CSF and CPC scores mentioned in *Sections III.A.2* and *III.A.5*, other computational approaches examining coding potential include CSTminer (78), QRNA (79), and CRITICA (80). CONC (Coding Or NonCoding) is a program that was developed based on support vector machines and can be used to classify transcripts according to features including peptide length, amino acid composition, predicted secondary structure content, predicted percentage of exposed residues, compositional entropy, number of homologs from database searches, and alignment entropy (81). The identification and characterization of a growing set of lncRNAs has allowed experimental validation of these bioinformatic approaches (Figure 3B).

Although most studies of lncRNAs have used the aforementioned bioinformatic approaches to evaluate their coding potential, ribosomal profiling is a direct experimental approach that can be used to address this issue. It was first developed to investigate the process of translation with subcodon resolution and involves deep sequencing of ribosome-protected RNA fragments (23, 82). It was then adapted to distinguish polyribosome-associated RNAs that are likely being translated from other RNAs that are more likely to be noncoding (Figure 3B). Nam and Bartel (73) identified polyadenylated transcripts in *Caenorhabditis elegans* using both RNA-seq and 3P-seq. Over 300 lncRNAs were identified from these transcripts after filtering through the coding potential threshold calculated from the CPC program and removing those that can be detected in ribosome profiling experiments. However, ribosomal profiling requires further testing and validation, because association with the ribosome alone cannot be taken as the absolute evidence of protein coding potential. For example, both *H19* and *TUG1*, 2 well-characterized lncRNAs, can be detected in association with the ribosome (23, 83). Some researchers have argued that instead of simply eliminating all transcripts identified in association with ribosomes (ie, from ribosome profiling experiments), a more careful examination of the preferential usage of specific coding frames and features conferred by the release of the ribosomal complex at the site of the stop codon should be used to determine whether the transcript is productively translated (24).

C. Gene-specific validations

High-throughput sequencing and bioinformatics methods have led to tremendous progress in the large-scale identification of lncRNAs. Nevertheless, empirical validation of lncRNAs using a set of classical molecular biology techniques is still required. After learning the approximate location of a potential lncRNA transcript using global approaches, 5′- and 3′-rapid amplification of cDNA end (RACE) experiments can be carried out to determine the exact transcription initiation and termination sites and to examine the presence or the absence of the 5′-cap and 3′-polyadenosine tail (Figure 3B). PCR-based approaches can be used to isolate full-length cDNAs for those lncRNAs whose cDNAs are not available from public repositories, followed by traditional Sanger sequencing to obtain precise information on the exact exon-intron structure of the mature lncRNA transcript. Validation of the noncoding status of a putative lncRNA is less straightforward. In vitro transcription-translation assays have been used (Figure 3B), but may give inconclusive results. In the case of *SRA*, functional outcomes associated with the RNA transcript were monitored after the introduction of different missense and frameshift mutations, illustrating how one can prove that a lncRNA functions in a coding-independent manner (25). Nevertheless, this approaches demands prior knowledge of the functions of the identified lncRNAs.

D. Cataloging lncRNAs in public databases

The identification and characterization of a growing set of lncRNAs has provided additional insights into the properties of lncRNAs as a group, which facilitate subsequent efforts in lncRNA research. To make better use of the power of recursion, a number of lncRNA databases have been developed to consolidate and summarize the growing body of information (Table 1). These include 1) ncRNAdb, one of the first lncRNA databases, which focuses on functional ncRNA transcripts that perform regulatory roles in the cell (84); 2) fRNAdb (85) and NONCODE (86), more recent databases that compile and integrate existing information of ncRNAs, including

Table 1. Publicly Available LncRNA Databases

Database	Species	Number of LncRNAs	Website	Last Updated	References
ChIPBase[a]	Multiple	NA[b]	http://deepbase.sysu.edu.cn/chipbase/	2012	90
fRNAdb[a]	Multiple	137 363	http://www.ncrna.org/frnadb/catalog_taxonomy/index.html	2014	85
GENCODE[a]	Human, mouse	26 414	http://www.gencodegenes.org/	2014	13
Human Body Map lincRNA	Human	>8000	http://www.broadinstitute.org/genome_bio/human_lincrnas/	2011	20
LNCipedia	Human	32 183	http://www.lncipedia.org/	2013	89
lncRNAdb	Multiple	>150	http://www.lncrnadb.org/	2012	88
ncRNAdb	Multiple	>30 000	http://biobases.ibch.poznan.pl/ncRNA/	2006	84, 334
NONCODE	Multiple	210 831	http://www.noncode.org/	2014	86, 335
NRED	Human, mouse	NA[b]	http://nred.matticklab.com/cgi-bin/ncrnadb.pl	2008	87

Abbreviation: NA, not available.

[a] These databases also contain short ncRNAs.

[b] These databases link lncRNAs to expression microarray data (NRED) and ChIP-seq data (ChIPBase), rather than cataloging lncRNAs per se.

lncRNAs; 3) a number of public repositories for lncRNAs, such as Noncoding RNA Expression Database (NRED) (87), lncRNAdb (88), and LNCipedia (2013) (89); and 4) ChIPBase, a recently developed database that has extracted lncRNAs from lncRNAdb and incorporated transcription factor binding maps taken from 543 ChIP-seq experiments. The result is the identification of tens of thousands of regulatory relationships between transcription factors and lncRNAs in a wide variety of tissues and cell lines from 6 organisms (90).

Other catalogs collect lncRNAs using in-house annotation pipelines, such the Human Body Map lincRNA database (described above) (20) and GENCODE (13), which are the most current and comprehensive. GENCODE, which is part of the ENCODE project, seeks to annotate all evidence-based gene features (cDNA, expressed sequence tag sequences) in the entire human and mouse genomes at a high accuracy, and generates annotations of both protein-coding and noncoding genes, including a large number of lncRNAs. The latest version of GENCODE for human (June 2014 freeze, GRCh38) contains 26 414 lncRNA transcripts produced from 15 877 genes.

At this point, a substantial proportion of all polyadenylated lncRNAs expressed in humans have already been annotated, but these annotations need additional refinement and validation. In addition, given the tissue and species specificity of lncRNAs, there are most certainly more to be discovered. With the existing annotations and functional databases, molecular biologists interested in the functional characterization of lncRNAs are no longer tied to the requirement of bioinformatics expertise and the high cost of deep sequencing associated with de novo identification of lncRNAs.

IV. Functional Characterization of LncRNAs

Assigning molecular, cellular, and physiological functions to well-annotated lncRNAs is the next great challenge in the field. Classical biochemical and molecular biology techniques have been instrumental in gene-specific functional characterization of lncRNAs. Gain-of-function and loss-of-function experiments can be used to validate the role of lncRNAs in modulating specific cellular processes. But, it is often challenging to determine whether an uncharacterized lncRNA plays an important functional role, or which cellular process can be probed to yield an observable phenotype. Indeed, more efficient functional analyses, including high-throughput approaches linking lncRNAs to their probable functions, are required to keep pace with the tremendous progress made in lncRNA discovery.

A. Expression profiling across spatial and temporal gradients

The expression of lncRNAs is often cell type-, tissue-, and context-dependent. Therefore, the involvement of lncRNAs in specific cellular processes may be inferred by their differential expression patterns across tissues and across different developmental- or signal-regulated time points. For instance, Klattenhoff et al (91) identified lncRNAs that play critical roles in cardiovascular lineage commitment by reasoning that such candidates should demonstrate expression patterns restricted to specific cell types during ESC differentiation. They measured lncRNA expression in mouse ESCs and in differentiated tissues using RNA-seq and focused on 47 candidates whose expression levels were elevated in ESCs compared with other differentiated tissues. Among them, *Braveheart*, a lncRNA with higher expression in the heart relative to

other tissues, was selected and characterized as a mediator of epigenetic regulation of cardiac commitment.

Similarly, Kretz et al (92, 93) focused on lncRNAs in keratinocyte differentiation, performing RNA-seq in primary keratinocytes during a calcium-induced differentiation time course. *ANCR* and *TINCR* were 2 of the candidates chosen for further characterization. *ANCR*, one of the most strongly suppressed RNAs during differentiation, is required to enforce the undifferentiated cell state within the epidermis (93). In contrast, *TINCR*, is one of the most highly induced annotated lncRNAs during differentiation, is required for somatic tissue differentiation. It acts by binding to differentiation-specific mRNAs to stabilize their steady-state levels (92). Interestingly, lncRNA genes may be spatially correlated with genes encoding key transcription factors, as shown for lncRNAs in the lung and foregut endoderm (94). These lncRNA genes are located adjacent to, and show similar expression patterns as, adjacent genes encoding critical developmental transcription factors (eg, *Nkx2.1*, *Gata6*, *Foxa2*, and *Foxf1*) (94).

These are just a few examples where transcriptome profiling experiments across spatial or temporal gradients generate clues to the functions of annotated lncRNAs. Because most lncRNA discovery approaches incorporate transcriptome profiling, it can be easily envisaged that when carefully designed, such efforts will not only yield information on the annotation and expression of novel and existing lncRNAs genes but also shed light on the probable functions of a selected group of newly annotated lncRNAs.

B. Coding-noncoding coexpression relationships: guilt-by-association

Although the spatial and temporal gradients are helpful in choosing and characterizing a selected group of lncRNAs, additional approaches are needed for other situations. Guttman and colleagues (18) have proposed a genomic approach to allow global functional characterization of lncRNAs, also known as guilt-by-association, which relies on correlation and clustering analysis performed on mRNA expression profiling data and gene ontology or functional pathway analyses (Figure 4). In this approach, groups of lncRNAs of unknown function are associated with groups of protein-coding mRNAs known to be involved in a specific cellular process based on a common expression pattern across cell types and tissues. A positive correlation between the expression profile of a lncRNA and mRNAs suggests a common function in the same cellular process. In their original paper, *lincRNA-p21* was predicted to associate with p53-mediated DNA damage responses, with *lincRNA-p21* later validated as a p53 target that modulates apoptotic responses upon DNA damage (95). The guilt-by-association approach is a useful first pass in assigning putative biological functions to lncRNAs and provides a working hypothesis for targeted perturbation experiments.

Zhao and colleagues (96) have expanded the analysis of gene coexpression relationships into a coding-noncoding coexpression network (CNC), making computational prediction of lncRNA functions through the evaluation of network characteristics. In addition to the coexpression network, colocalization relationships were also taken into consideration in their analysis. They focused on mouse lncRNAs annotated by FANTOM3 and extracted gene expression information from reannotated Affymetrix Mouse Genome Array data. Ultimately, they predicted functions for 349 lncRNAs and further streamlined the application into a practical user interface called the Noncoding RNA Function Annotation Server (ncFANs) (97). ncFANs is a useful tool for global prediction of lncRNA function, forming the basis of functional annotation in the NONCODE database, but its application is limited to annotated lncRNAs associated with corresponding microarray-based gene expression data.

C. A role for lncRNAs in the *cis*-regulation of gene expression

One rationale behind the use of colocalization relationships in CNC-based functional characterization is that many lncRNAs have been shown to play a *cis*-regulatory role in the expression of nearby genes. For example, the gene for the lncRNA *ANRIL* overlaps and runs antisense to the gene encoding p15, mediating its gene silencing (98). Moreover, *linc-HOXA1* is located ~50 kb from the *HoxA* gene cluster in mouse ESCs and functions to repress *Hoxa1* by recruiting purine-rich element-binding protein b as a transcription cofactor (99). In contrast, a chromatin-associated lncRNA *CAR intergenic 10* is coexpressed with its flanking coding genes, *FANK1* and *Adam12*, and helps to maintain their expression by establishing active chromatin structures (100).

The *cis*-regulatory function of *HOTTIP* involves an additional element. It is a lncRNA transcribed from the *5′*-end of the *HoxA* gene cluster and functions to activate the expression of neighboring genes (101). Nevertheless, its influence extends to multiple distal *HoxA* genes due to chromosome looping, as suggested by chromosome conformation capture carbon copy (5C), a high throughput method to identify physical chromatin interaction. These results suggest a model of how a *cis*-acting lncRNA can affect distal genes.

In another study, Ørom et al (6), using lncRNAs from the GENCODE database, uncovered an enhancer-like function for several lncRNAs, which they termed ncRNA-activating (ncRNA-a). These ncRNA-as enhance the ex-

Figure 4.

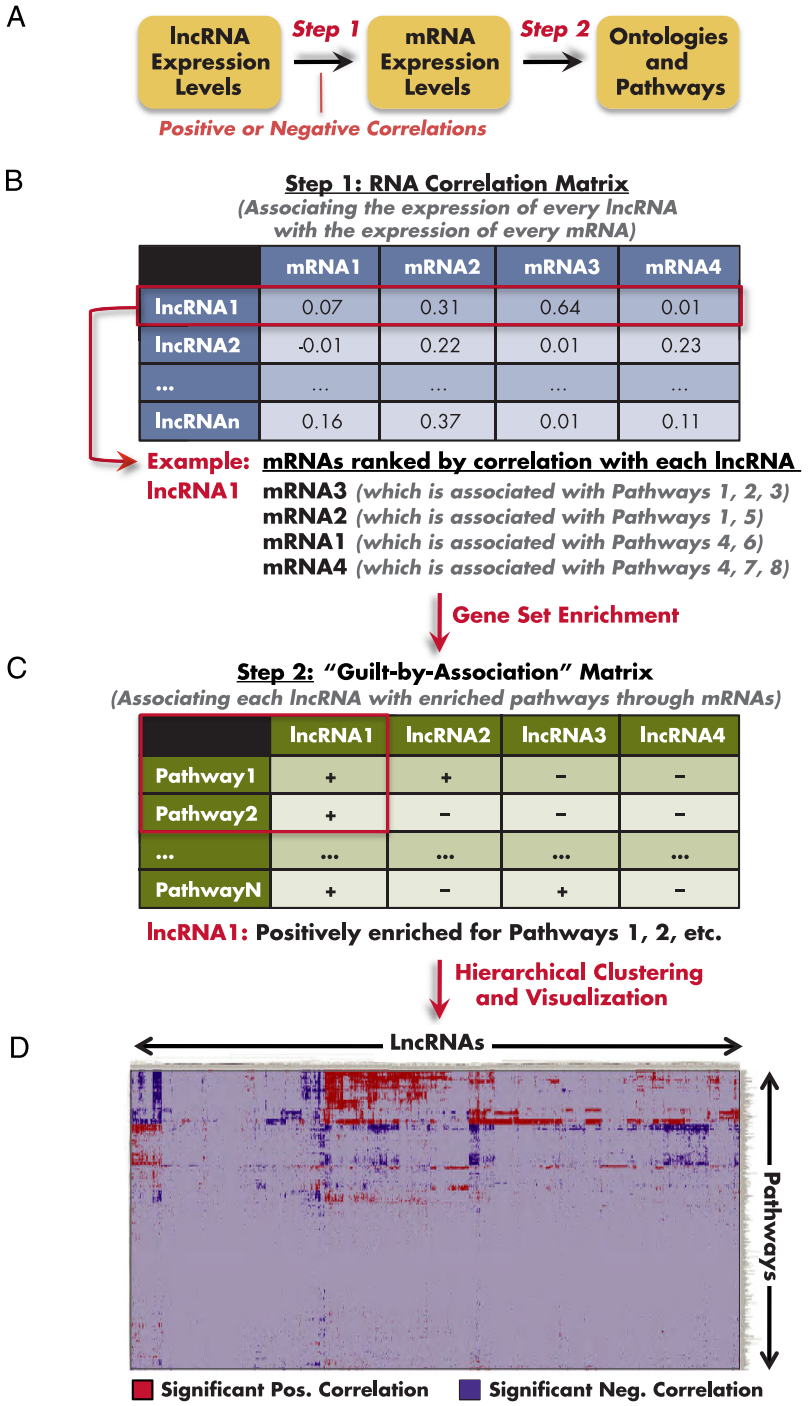

Figure 4. Guilt-by-association analyses link lncRNA expression patterns to gene ontologies and pathways through mRNA expression patterns. A, Overview of the guilt-by-association approach. B, The RNA correlation matrix links the expression of each lncRNA to the expression of all mRNAs (and their associated pathways and ontologies). C, The guilt-by-association matrix links pathways with each lncRNAs through gene set enrichment analyses. The assumption is lncRNAs that share expression patterns with mRNAs will also share pathways and ontologies. D, A heatmap provides a graphical representation of the results from the guilt-by-association matrix, showing significant positive and negative correlation between each lncRNA and each pathway. Hierarchical clustering groups lncRNAs that have similar expected functions. Abbreviations: pos., positive; neg., negative.

pression of neighboring coding genes in short interfering RNA-mediated knockdown experiments and heterologous reporter gene assays. Similar to *HOTTIP*, some ncRNA-a genes associate with their target genes through long-range chromatin loops. For example, Lai et al (102) have demonstrated that ncRNA-as recruit the Mediator

complex to their targets genes, where it plays an important role in forming DNA loops between the promoters of lncRNA genes and the promoters of target genes, as well as mediating ncRNA-a-dependent target gene activation (Figure 5). Looping events have also been observed between the promoters of lncRNA genes and nearby enhancers. For example, the promoter of the gene encoding the lncRNA *CARLo-5* (cancer-associated region long noncoding RNA-5) physically interacts with the *MYC* enhancer in the chromosome 8q24 region, possibly to regulate *CARLo-5* expression (103).

Colocalization relationships have been exploited even further. To study lncRNAs involved in cell cycle regulation, Hung and colleagues (104) looked in the proximity of known cell cycle genes and designed their approaches based on both guilt-by-association strategy and the *cis*-regulatory model. They used an ultra-high-density array that tiles the promoters of 56 cell cycle genes to interrogate 108 samples representing diverse conditions and perturbations, identifying 216 putative lncRNA transcripts originating proximal to these cell cycle gene promoters. Subsequently, they examined the coding-noncoding coexpression map across the conditions and clustered lncRNAs into different cell cycle-associated functions. The lncRNA *PANDA* (p21-associated ncRNA DNA damage activated) was selected for further analysis and was shown to regulate apoptosis, consistent with the prediction.

D. A role for lncRNAs in the *trans*-regulation of gene expression

When coexpression and colocalization relationships are used as the basis for functional prediction, direct perturbation experiments are required to validate the prediction. Therefore, Guttman and colleagues (105) suggested a more direct approach for the functional characterization of lncRNAs, performing RNA interference-based loss-of-function experiments and monitoring consequent changes in global gene expression. They focused on previously identified lncRNAs expressed in ESCs and were able to successfully knock down the expression of 147 lncRNAs using custom-designed short hairpin RNAs. For 137 lncRNAs, knockdown resulted in significant global changes in gene expression as shown in microarray analysis, and the majority had little effect on neighboring genes, suggesting that these lncRNAs most likely affect gene expression in *trans* (105).

These were not the first lncRNAs to be associated with *trans*-regulation. *HOTAIR*, a well-characterized lncRNA involved in developmental processes, is coexpressed with the *HoxC* genes, interacts with the chromatin-modifying Polycomb repressive complex 2 (PRC2) complex, and functions in trans to repress *HoxD* expression (106). Interactions between *HOTAIR* and PRC2 proteins have been verified in both RNA-pulldown (captures proteins associated with a RNA bait) and RNA immunoprecipitation (RIP) (captures RNAs that are associated with proteins of interest using specific antibodies). Indeed, there are many other lncRNAs that have been shown to interact with PRC2, including *Braveheart* (described earlier) (91) and *XIST*, which coats the X chromosome to initiate and propagate X-inactivation (107–109).

Expanding on these observations, Khalil et al (21) coupled RIP to a microarray analysis (RIP-chip) to query many lncRNAs simultaneously. Among the 3300 human

Figure 5.

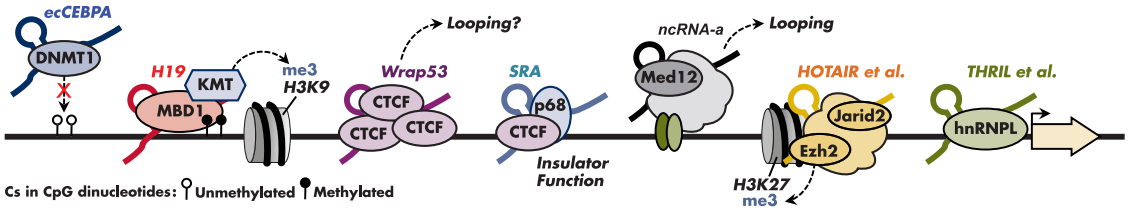

Figure 5. LncRNA-protein interactions drive molecular outcomes in gene regulation. Some lncRNAs function as molecular scaffolds that promote the assembly of complexes containing chromatin- and transcription-modulating factors. These interactions are driven by specific interactions between lncRNAs and proteins. The schematic diagram illustrates and generalizes specific lncRNA-protein interactions that have been observed in specific gene regulation contexts. From left to right: The lncRNA *ecCEBPA* interacts with the DNA methyltransferase DNMT1 to block DNA methylation and control gene expression outcomes (121). *H19* binds to the methyl-CpG-binding protein MBD1 to control gene expression by recruiting a histone lysine methyltransferase (KMT) to add repressive histone marks to the differentially methylated regions of imprinted genes (122). *Wrap53*, a natural antisense transcript of *TP53*, interacts with the insulator protein and transcriptional regulator CTCF to control gene expression (123). CTCF also interacts with *SRA* and its associated DEAD-box RNA helicase p68 to form a complex with CTCF that is essential for insulator function (124). ncRNA-a lncRNAs interact with the Med12 subunit of the Mediator complex to promote gene looping and target gene activation (102). *HOTAIR* interacts with the histone methyltransferase Ezh2, a key component of the PRC2 complex, to mediate chromatin-dependent gene regulation (21, 107, 112). *HOTAIR* also interacts with Jarid2, a PRC2-associated factor, to promote the targeting of PRC2 to chromatin (116, 117). *THRIL* binds to hnRNPL, a component of hnRNP complexes, and the THRIL-hnRNPL complex regulates transcription by binding to target gene promoters (125).

lncRNAs queried, PRC2 or RE1-silencing transcription factor corepressor 1 (CoREST) complexes were found to associate with 38% of them, suggesting that lncRNAs interacting with chromatin-associated complexes could be a common mechanism. In addition, although RIP-chip requires prior knowledge of lncRNA sequences, Zhao et al (107) improved the method by coupling it to high-throughput sequencing, which allows for unbiased identification of lncRNAs that interact with candidate proteins. In this case, they tested their method on PRC2 and identified a genome-wide pool of >9000 PRC2-interacting RNAs in mouse ESCs. Not surprisingly, *XIST* was highly enriched in the PRC2 RIP-seq experiments, serving as a good positive control. The types of lncRNA-protein interactions described here are likely to be a key component of *trans*-regulation pathways.

E. LncRNA-protein interactions drive molecular outcomes in *cis* and *trans* gene regulation

LncRNAs are thought to function as molecular scaffolds that promote the assembly of complexes containing chromatin- and transcription-modulating factors (8, 10, 21, 110). These scaffolding effects are driven by specific interactions between lncRNAs and proteins (Figure 5). Determining specific and direct interactions between lncRNAs and their protein partners can be challenging. Interactions detected under native conditions may reflect nonspecific interactions or may be indirect (ie, mediated by another protein). UV or photoactivatable cross-linkers can help resolve both of these issues. In addition, reconstituting the interactions in biochemical assays, such as EMSAs, with subsequent validation of the interactions using mutants of the lncRNAs and proteins, can also be an effective tool. Ultimately, lncRNA-protein interactions should be explored using structural biology, which can reveal novel insights into the functions of lncRNAs (111).

The histone methyltransferase Ezh2 is a key lncRNA-binding component of the aforementioned PRC2 complex, promoting interactions with *XIST* and *HOTAIR,* and mediating chromatin-dependent gene regulation (21, 107, 112, 113) (Figure 5). Likewise, the PRC2-associated AT-rich interaction domain-containing chromatin regulator Jarid2 (113–115) also binds *XIST* and *HOTAIR* to promote the targeting of PRC2 to chromatin (116, 117), whereas the embryonic ectoderm development subunit regulates the affinity of Ezh2 for RNA, increasing the specificity to PRC2-RNA interactions (113). Interestingly, although RNA is important for targeting PRC2 to chromatin, it also inhibits Ezh2's catalytic activity; JARID2 attenuates the binding of PRC2 to RNA and relieves this inhibition (113). These results illustrate the complexity of the regulatory interactions between lncRNAs and PRC2.

Other lncRNAs have been shown to interact with additional chromatin-modifying complexes. For example, *HOTTIP* binds to and targets the WD repeat-containing protein 5 (WDR5)/mixed-lineage leukemia protein (MLL) complex across the *HoxA* cluster to maintain active chromatin and coordinate homeotic gene expression (101). Other lncRNAs, including some known to be important for gene expression in ESCs, also bind WDR5/MLL to specify cell fate outcomes (118). In addition, the tissue-specific lncRNA *Fendrr* has been shown to bind both the PRC2 and trithorax group/MLL complexes, modulating chromatin signatures and gene activities to ensure the proper development of heart and body wall in mouse (119). LncRNAs may also affect the expression of rRNA genes, which are transcribed by RNA Pol I. For example, *PAPAS* (promoter and pre-rRNA antisense) lncRNAs generated from the rDNA promoter mediate the recruitment of the H4K20 methyltransferase Suv4–20h2, increased H4K20me3, and chromatin compaction at the rDNA promoter in growth-arrested cells (120).

Recent studies have also demonstrated specific functional interactions of lncRNAs with other types of gene-regulating proteins (Figure 5). For example, the Med12 subunit of the Mediator complex interacts with lncRNAs (ncRNA-as, as noted in *Section IV.C*) to promote gene looping and target gene activation (102). LncRNAs that interact with the DNA (cytosine-5)-methyltransferase 1 (DNMT1) block DNA methylation to control gene expression outcomes for specific target genes (121) (Figure 5). In addition, the lncRNA *H19* binds to the methyl-CpG-binding domain protein 1 to control the expression of 5 genes in the Imprinted Gene Network (IGN), which is involved in growth control of the embryo, by adding repressive histone marks (eg, H3K9me3) to the differentially methylated regions of these imprinted genes (122) (Figure 5). Another study has shown that interactions between *Wrap53*, a natural antisense transcript of *TP53*, the gene encoding the tumor suppressor p53, interacts with the insulator protein and transcriptional regulator CCCTC-binding factor (CTCF) to control p53 expression (123) (Figure 5). CTCF also interacts with *SRA* and its associated DEAD-box RNA helicase p68 to form a complex with CTCF that is essential for insulator function (124). Furthermore, the lncRNA *THRIL* (TNFα and hnRNPL-related immunoregulatory LincRNA) binds to hnRNPL, a component of heterogeneous nuclear ribonucleoprotein (hnRNP) complexes that associate with nascent transcripts and regulate mRNA processing (125). The *THRIL*-hnRNPL complex regulates transcription of *TNFA*, the gene encoding the proinflammatory cytokine TNFα, by binding to its promoter (125) (Figure 5). Finally, human *Alu* RNA, a modular RNA that is transcribed from

Figure 6.

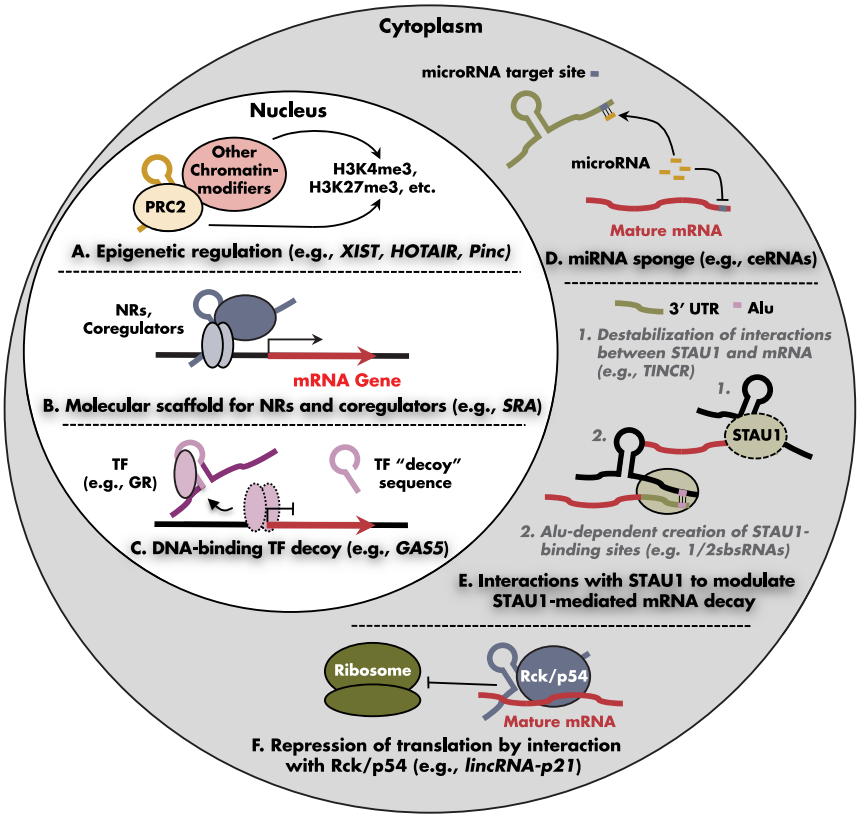

Figure 6. Gene regulation by lncRNAs occurs through nuclear and cytoplasmic mechanisms that affect transcriptional, posttranscriptional, and translational events. LncRNAs mediate their functional roles by regulating gene expression at many different levels, through a variety of molecular mechanisms both in the nucleus and in the cytoplasm. The nuclear functions of lncRNAs include interactions with chromatin-modifying complexes to alter epigenetic modifications (A); interactions with transcription factors (TFs) (B), such as nuclear receptors (NRs), and additional transcriptional coregulators, to alter their gene regulatory activities; and actions as molecular decoys to titrate away and inhibit the activity of DNA-binding TFs (C). The cytoplasmic functions of lncRNAs include sponging of microRNAs to reduce microRNA targeting of mRNAs (D), as in the case of ceRNAs; interactions with STAU1 and regulation of STAU1-dependent mRNA stability (E); and interactions with the cytoplasmic RNA-binding protein Rck/p54 to inhibit translation (F).

short interspersed elements, blocks transcription by binding RNA Pol II, thus providing a direct means for regulating transcription (126).

Taken together, the connections between lncRNAs and nuclear proteins that regulate chromatin, gene looping, transcription, and RNA processing suggest a number of appealing models of lncRNA-dependent gene regulation involving all steps in the gene regulatory processes (Figure 6, A–C). The further development of RIP-based tools to identify lncRNA-protein interactions, like those described above, as well as better computational tools for predicting and analyzing lncRNA-protein interactions (eg, lncPro, RBP-lncRNA Base) (127, 128) should enhance the ease and accuracy of identification, as well as functional studies. Furthermore, new genome-wide approaches to elucidate the structure and possibly the domain architecture of lncRNAs, such as those described recently by the Chang and Weissman labs (129, 130), will undoubtedly reveal the key sequence and structural elements of lncRNAs that allow them to interact specifically with their protein partners.

F. Methods for the detection of lncRNA interaction sites across the genome

RIP-based experiments have helped to establish direct interactions between lncRNAs and proteins, suggesting that lncRNAs can act as molecular scaffolds to guide chromatin-modifying complexes to their target genomic locations. Coupled with profiles of changes in chromatin signatures by ChIP-seq, the target sites of lncRNA action can be deduced. For example, changes in H3K4me3 and H3K27me3 were observed in *HOTAIR* knockdown foreskin fibroblasts, consistent with the modes of action of *HOTAIR* in targeting lysine-specific demethylase 1 (LSD1) and PRC2 to specific genomic locations to affect histone modifications (106). Nevertheless, direct methods that capture the interaction between lncRNAs and chromatin sites have been developed recently, including 1) chromatin isolation by RNA purification (ChIRP) (131), 2) capture hybridization analysis of RNA targets (CHART) (132), and 3) RNA antisense purification (RAP) (133). They are based on affinity capture of target lncRNA:chromatin complexes using tiling antisense oligonucleotides in ChIRP and RAP or preselected oligonucleotides targeting RNase-H-sensitive regions of the lncRNA in CHART to generate a genomic map of lncRNA binding sites. In comparison with CHIRP, RAP uses longer oligonucleotides as probes, which allows more stringent conditions for removal of nonspecific interactions. CHIRP and CHART have been applied to transacting lncRNAs, such as the *Drosophila* lncRNA *roX2* and human *HOTAIR*, to confirm their genomic binding sites, whereas CHART and RAP have been applied to the mouse lncRNA *XIST* to visualize its spreading along the inactive X

chromosome (133, 134). Moreover, CHART has also been used to functionally characterize *Paupar*, a newly identified lncRNA in mouse neuroblastoma cells that affects neural differentiation, to elucidate its transcriptional regulatory activity in *trans* (135).

G. Beyond the nucleus: a broader view of lncRNA functions

LncRNAs play important roles in both *cis*- and *trans*-regulation of transcription (Figures 5 and 6, A–C), but continued studies are needed to determine the relative contributions of *cis* and *trans* mechanisms of lncRNA function. There is a strong bias in the field for this potential aspect of lncRNA function, leading to the common belief that lncRNAs as a group are mostly involved in transcriptional regulation. Although lncRNAs as a group may show a slight enrichment for the nuclear compartment, many lncRNAs are predominantly or even exclusively localized to the cytoplasm. Inherent biases in some previous analytical approaches, however, have propagated the emphasis on nuclear functions for lncRNAs. For example, PRC2 RIP-based methods have suggested that a large number of lncRNAs are involved in PRC2-mediated transcriptional repression (107). Nevertheless, the RIP protocol limited the analysis to nucleus-retained RNAs, leaving open the possibility that a large proportion of lncRNAs interact with cytoplasmic proteins. Indeed, there have been an increasing number of examples of cytoplasmic lncRNAs. Among them, half-Staufen double-stranded RNA binding protein 1 (STAU1)-binding site RNAs have been shown to transactivate the binding of STAU1 protein to its target mRNAs to facilitate mRNA decay (136) (Figure 6E). On the other hand, *TINCR*, another lncRNA that has been shown to bind STAU1, functions to stabilize the expression of differentiation mRNAs in a STAU1-dependent manner (92) (Figure 6E).

Furthermore, Huarte et al (95) explored the mechanisms of action of *lincRNA-p21* by identifying its interaction partners using RNA pull downs with nuclear extracts. They found that the nuclear RNA binding protein hnRNPK associates with this lncRNA to facilitate gene repression. Yoon and colleagues (137) confirmed this interaction using hnRNP-K RIP, whereas Dimitrova et al (138) explored the functions of *lincRNA-p21* in vivo using a knockout mouse. The latter found that *lincRNA-p21* functions predominantly as a cis activator of its neighboring gene, *p21* (138). Interestingly, as Yoon and colleagues (137) searched for RNA partners of the cytoplasmic RNA binding protein human antigen R (HuR) in a RIP experiment using whole-cell lysates, they observed that *lincRNA-p21* was enriched as well. This interaction accelerates the degradation of *lincRNA-p21*, which in turn derepresses the expression of a subset of target mRNAs. In the absence of HuR, *lincRNA-p21* is stable, accumulates, and associates with the DEAD-box helicase Rck/p54. Rck/p54 promotes the association of *lincRNA-p21* with *CTNNB1* and *JUNB* mRNAs, repressing their translation through a mechanism that includes reduced polysome size, suggesting an additional role of cytoplasmic *lincRNA-p21* as a posttranscriptional inhibitor of translation (Figure 6F).

A number of lncRNAs have been shown to function as competing endogenous RNAs (ceRNAs) that function in multiple cellular models as sponges that can bind and reduce the targeted effects of microRNAs on mRNAs (139–143) (Figure 6D). This includes circRNAs originating from back-spliced exons (40–42). One example of a ceRNA is *linc-MD1*, which sponges *miR-133* and *miR-135* to control the expression of the transcription factors mastermind-like protein 1 and myocyte-specific enhancer factor 2C, which activate a muscle-specific gene expression program (139). Another example is the *H19* lncRNA, which sponges *let-7* microRNAs to control muscle differentiation (143). Examples exist from a variety of other biological systems as well (40–42, 140–142). In addition to sponging, a recent study has shown that lncRNAs may also control pri-microRNA processing, as exemplified by the lncRNA, *Uc.283+A*, which prevents pri-microRNA cleavage by Drosha (144).

Many of the examples provided here, like *lincRNA-p21*, suggest that methods limited to the characterization of nucleus-retained lncRNAs are thus not sufficient to provide us with a complete spectrum of functional roles played by lncRNAs. Delineating the cellular localization of lncRNAs in an unbiased manner should be one of the first steps used for gathering more clues on their possible functional roles. Nucleus-retained lncRNAs are more likely to be involved in transcriptional regulation, whereas cytoplasmic lncRNAs may have other functions. RNA fluorescence in situ hybridization is a common method that has been used to visualize the cellular localization of lncRNAs (145–151), but challenges remain for a high-throughput fluorescence in situ hybridization approach that examines many lncRNAs simultaneously. Alternatively, lncRNAs can be extracted from each of the physically defined cellular compartments and then sequenced, revealing the relative amount of each lncRNA in the various cellular fractions. With modifications as described in Yoon et al (137), RIP-based methods can also be used with key cytoplasmic proteins that act in important cellular pathways to identify and characterize cytoplasmic lncRNAs involved in those pathways. Furthermore, Kretz et al (92), who characterized *TINCR*, used a protein microarray analysis containing approximately 9400 recom-

binant human proteins (Human Protoarray) to identify the *TINCR*-STAU1 interaction in the cytoplasm.

V. Lessons Learned from the Best-Characterized LncRNAs

Using methods described above and additional strategies, a growing number of lncRNAs have been characterized molecularly and functionally (Table 2). A limited few are as well-characterized as some protein-coding RNAs. Below, we summarize the current status of the few best-characterized lncRNAs to date and highlight the lessons learned from these examples.

A. *XIST*

The X-inactive-specific transcript (*XIST*) was one of the first lncRNAs to be discovered in mammals (152–155). It is responsible for the initiation and spreading of X-chromosome inactivation (XCI) in female somatic cells (108, 109, 156, 157). *XIST* is transcribed from the XCI loci and acts in concert with the transcription factor YY1 and several other lncRNAs from the same locus (eg, *RepA*, *Tsix*, and *Jpx/Enox*) to facilitate the loading of PRC2 and initiate DNA methylation and the subsequent chromosome-wide silencing (158–165) (Figure 6A). It is one of the best examples of multiple lncRNAs using their base complementarity properties to collaborate with each other and with proteins to achieve a common cellular function. This could be a recurring theme with lncRNAs, which may base pair with DNA in the genome or RNA elements in the transcriptome, creating unique interfaces for RNA-protein interactions. LncRNAs encompass RNA motifs with variable lengths, offering advantages over small protein motifs and allowing more specificity in targeting to unique addresses.

Even after more than 2 decades of extensive research, the exact mechanism of *XIST*-mediated spreading of XCI is yet to be fully elucidated. This is due, in part, to the lack of high-throughput approaches of sufficient resolution to distinguish allelic differences of the X chromosomes. To address this, Pinter et al (166) developed allele-specific ChIP-seq, mapping the positions of the PRC2 component Ezh2 and XCI-associated histone marks on the inactive (Xi) and active (Xa) X chromosomes separately over a developmental time course. The authors presented a model in which XCI is governed by a hierarchy of defined PRC2 stations that spread H3K27 methylation in *cis*. In addition, Engreitz et al (133) used RAP to examine the mechanism of localization of the *XIST* lncRNA, showing that it exploits the 3-dimensional genome architecture to spread across Xi. Furthermore, Simon et al (134) used CHART-seq, a method similar to RAP, to provide high-resolution maps of *XIST* on the X chromosome across a developmental time course. The authors showed that *XIST* lncRNA spreads to gene-rich and gene-poor regions sequentially in a stage-specific manner.

B. *MALAT1*

Metastasis-associated lung adenocarcinoma transcript 1 (*MALAT1*) was one of the first cancer-associated lncRNAs discovered (167). It is extremely abundant and highly conserved over its full length across all mammalian species, both properties that highlight its likely importance and conserved functions (167–169). *MALAT1* localizes to nuclear bodies known as nuclear speckles (170), suggesting functions in the nucleus. In cell-based models, *MALAT1* has been shown to regulate alternative splicing and gene expression at the molecular level (169–172), contributing to its association with metastatic lung adenocarcinoma. Given these preliminary results, the observation that *Malat1*-knockout mice display little observable phenotype, especially with respect to splicing or gene expression, is surprising (173). The field must address why lncRNAs that show cell-based phenotypes are not functional in vivo, as observed with *MALAT1*. Parallels between lncRNAs and the better understood class of microRNAs may help to explain such conundrums. Phenotypic evaluation of microRNA knockout mice has revealed similar disappointing phenotypes (174–177). But, new studies suggest that the most dramatic phenotypes often arise in response to specific cellular signals, such as special diet or stress, or under a compromised genetic background (178, 179), suggesting that the appropriate cellular context is essential.

In this regard, given the association between *MALAT1* and lung adenocarcinoma, it will be interesting to cross the *Malat1*-knockout mice with genetic models of lung cancer, or to generate lung-specific *Malat1*-knockout mice and treat them with tumorigenic agents, to determine whether any transcriptional and phenotypic consequences arise. In this regard, Gutschner et al (180) diminished *MALAT1* expression in A549 human lung adenocarcinoma cells using a zinc finger nuclease-mediated knockout approach. They observed changes in gene expression and impaired metastatic potential of these *MALAT1*-deficient cells in mouse xenograph experiments, once again establishing a critical role of *MALAT1* as a regulator of gene expression governing hallmarks of lung cancer metastasis (180). Not unlike the situation with microRNAs, when probing the in vivo functions of lncRNAs, it is important to find the right context to uncover the observable phenotype.

Table 2. List of LncRNAs Discussed in This Review

Gene Symbol	Gene Name
A. Hormonal signaling	
CTBP1-as	C-terminal-binding protein 1-antisense
GAS5	Growth arrest-specific 5
PCGEM1	Prostate-specific transcript 1, non-protein coding
PRNCR1	Prostate cancer-associated non-coding RNA 1
SRA	Steroid receptor RNA activator
B. Reproduction and development	
Pinc	Pregnancy-induced ncRNA
Zfas1	Zinc finger antisense 1
C. Adipogenesis	
Blnc1	Brown fat lncRNA 1
Lnc-RAP1, 2	LncRNA regulating adipogenesis 1, 2
PU.1 AS	PU.1 antisense
SRA	Steroid receptor RNA activator
D. Metabolism	
116HG	PWS locus lncRNA
HI-LNC25	β-Cell-specific lncRNA whose depletion downregulates *GLIS3* mRNA
IPW	Imprinted in Prader-Willi syndrome, non-protein-coding
E. Immune system function	
GAS5	Growth arrest-specific 5
Lethe	Pseudogene lncRNA that is selectively induced by proinflammatory cytokines via NF-κB or glucocorticoid receptor
LincR-Ccr2–5′AS	LincRNA-chemokine receptor type 2, 5′-antisense
LincRNA-Cox2	LncRNA induced by TLRs that mediates both activation and repression of immune response genes
Lnc-DC	LncRNA in dendritic cell
NeST	Nettoie Salmonella pas Theiler's (cleanup Salmonella not Theiler's); aka *TMEVPG1* (Tmevpg1 Theiler's murine encephalomyelitis virus persistence candidate gene 1) or *lincR-ifng-3′AS* (lincRNA-interferon γ-antisense)
THRIL	TNFα and hnRNPL-related immunoregulatory lincRNA
F. Nervous system function	
ARXN8OS	Ataxin 8 opposite strand
ASFMR	Antisense fragile X mental retardation
BACE1-AS	β-Secretase 1-antisense
BC1	Brain cytoplasmic 1
BC200	Brain cytoplasmic 200 nt, aka BCYRN1, brain cytoplasmic RNA 1
BDNF-AS	Brain-derived neurotrophic factor-antisense
Cyrano	A zebrafish lncRNA required for normal development
Dlx1as	Distal-less homeobox 1 antisense
Evf-2	ncRNA transcribed from the Dlx-5/6 ultraconserved region; aka *Dlx6as* (Dlx6 antisense)
FMR4	Fragile X mental retardation 4
Miat	Myocardial infarction-associated transcript; aka *RNCR2* (retinal ncRNA 2), or *Gomafu*
Paupar	Paired box6 (PAX6) upstream antisense RNA
RMST	Rhabdomyosarcoma 2 associated transcript
SCAANT1	SCA type 7 antisense noncoding transcript 1
Six3OS	SIX homeobox 3 opposite strand
TUG1	Taurine up-regulated 1, non-protein coding
TUNA	Tcl1 upstream neuron-associated lincRNA; aka megamind
UBE3A-ATS	Ubiquitin ligase E3A-antisense
utNgn1	Upstream transcript of *Neurog1*
Vax2OS	Ventral anterior homeobox 2 opposite strand
G. Cardiac function	
ANRIL	Antisense ncRNA in the INK4 locus; aka CDKN2B-AS (cyclin-dependent kinase 4 inhibitor B also known as multiple tumor suppressor 2-antisense)
Braveheart	LncRNA required for cardiovascular lineage commitment, abbreviated as Bvht
Fendrr	FOXF1 adjacent noncoding developmental regulatory RNA
MALAT1	Metastasis-associated lung adenocarcinoma transcript 1
MIAT	Myocardial infarction-associated transcript; aka *RNCR2* (retinal ncRNA 2), or *Gomafu*
MYHCB-AS	β-Myosin heavy chain-antisense
MYL4-AS	Myosin light chain 4-antisense
Linc-MYH	LincRNA-Myosin heavy chain
SRA	Steroid receptor RNA activator
TNNT2-AS	Cardiac troponin T type 2-antisense

(*Continued*)

Table 2. Continued

Gene Symbol	Gene Name
H. Skeletal muscle function	
DBE-T	D4Z4 binding element transcript, non-protein coding
H19	Imprinted materially expressed transcript, non-protein coding
Linc-MD1	LincRNA-muscle differentiation 1
I. Cancer	
ANRIL	Antisense ncRNA in the *INK4* locus; aka *CDKN2B-AS* (*CDKN2B* antisense)
CARLo-5	Cancer-associated region lncRNA; aka *CCAT1* (colon cancer-associated transcript 1)
FAL1	Focally amplified lncRNA on chromosome 1
GAS5	Growth arrest-specific 5
HOTAIR	HOX transcript antisense RNA
HOTTIP	HOXA transcript at the distal tip
lncRNA-ATB	LncRNA-activated by TGF-β
lincRNA-p21	lncRNA upstream and on the opposite strand to *Cdkn1a* (p21 gene)
Loc285194	LncRNA, p53-regulated tumor suppressor
MEG3	Maternally expressed 3
MRUL	MDR-related and upregulated lncRNA
PANDA	p21-associated ncRNA DNA damage activated
PCAT-1	Prostate cancer associated transcript 1
PCA3	Prostate cancer-associated 3
PCGEM1	Prostate-specific transcript 1, non-protein coding
Pint	p53-induced transcript
PRNCR1	Prostate cancer-associated non-coding RNA 1
PTENP1	Phosphatase and tensin homolog (PTEN) pseudogene 1
P21NAT	p21 natural antisense transcript
SChLAP1	Second chromosome locus associated with prostate 1
SRA	Steroid receptor RNA activator
TARID	TCF21 antisense RNA inducing demethylation
J. Other LncRNAs	
ANCR	Anti-differentiation ncRNA
asOct4-pg5	Antisense to Oct4 pseudogene 5
CAR intergenic 10	Chromatin-associated RNA 10
ci-ankrd52	ciRNAs in *Ankrd52*
Linc-HOXA1	ncRNA that represses *Hoxa1* transcription in cis
ncRNA-a	Noncoding RNA-activating
PAPAS	Promoter and pre-rRNA antisense
roX2	ncRNA present in the male-specific lethal (MSL) complex required for sex dosage compensation in *Drosophila*
TINCR	Terminal differentiation-induced ncRNA
Xist	X-inactive-specific transcript

C. *HOTAIR*

HOTAIR is a 2.2-kb lncRNA transcribed from the *HoxC* locus that functions to repress transcription in *trans* across 40 kb of the *HoxD* locus (106). Similar to *MALAT1*, *HOTAIR* is a lncRNA associated with a variety of cancers, including breast, colorectal, nasopharyngeal, and hepatocellular cancers (181–184), although its prognostic value in clinical oncology is still undetermined. *HOTAIR* was the first lncRNA found to associate with PRC2 complexes (106), initiating the subsequent characterization of a large number of PRC2-interacting RNAs later known as the PRC2 transcriptome (21, 107). It is also the first mammalian lncRNA to be screened by ChIRP, demonstrating its direct association with GA-rich regions of chromatin that nucleate broad domains of Polycomb and H3K27me3 occupancy (131) (Figure 6A). Tsai et al (110) showed that not only does the 5′-domain of *HOTAIR* bind to PRC2, but the 3′-domain binds to LSD1, a chromatin-modifying complex that promotes H3K4me3 demethylation, suggesting a role for *HOTAIR* as a molecular scaffold possessing distinct RNA domains for protein interactions. Targeted deletion of *Hotair* in mice has begun to reveal its functions in vivo (185). Consistent with the molecular functions of human *HOTAIR* in cell models, mouse *Hotair* binds PRC2 and LSD1 complexes to modulate epigenetic modifications. Hence, *Hotair* knockout in mice derepresses hundreds of genes, including those in the *HoxD* locus, as well as several imprinted genes, leading to homeotic transformation and skeletal malformations (185).

The characterization of *HOTAIR* illustrates 3 aspects of lncRNA function. First, it provides a model for the function of a lncRNA that regulates transcription in *trans*, by tethering to chromatin regions and recruiting chromatin-modifying complexes. Second, it shows that lncRNAs

Figure 7.

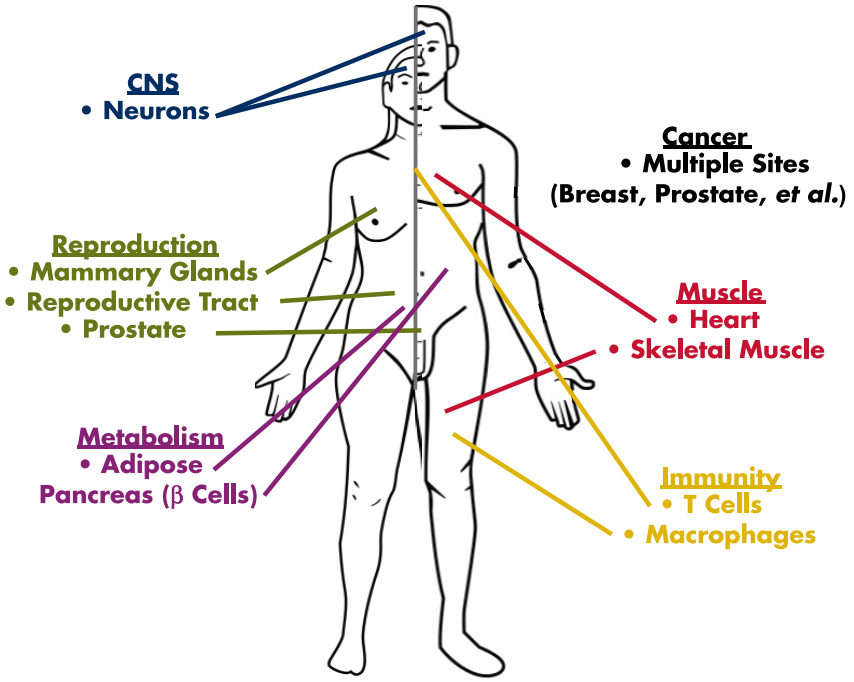

Figure 7. Physiological and pathophysiological functions of lncRNAs. Recent studies have identified important roles for lncRNAs in the physiology and pathophysiology of the endocrine, reproductive, metabolic, immune, nervous, and cardiovascular systems in both females and males. Moreover, lncRNAs are emerging as key regulators of cell proliferation and cell death, which are often associated with cancer.

can be modular, similar to proteins, with functions that can be separated into independent molecular domains that act in collaboration. Lastly, it provides an example of conservation of lncRNA functions with little sequence similarity between human and mice, but with conservation in synteny and RNA structures (185). These results suggest ways of studying lncRNAs in a manner similar to studying proteins. We can draw hints and insights from the prediction or biochemical mapping of RNA structures, as well as from information of evolutionary conversation, and perhaps can even work toward building a database of lncRNA domains or motifs, which will help to elucidate the functions of lncRNAs, much in the same way Pfam (58) and PROSITE (186) have done for proteins.

VI. The Biology of LncRNAs in Endocrine-Related Systems

As described in *Sections I-V*, recent studies of lncRNAs have yielded rapid advances in our understanding of this new class of RNAs. Recent studies have shown that lncRNAs are required for life (187) and that they are likely to be functionally involved in a wide variety of cellular processes in both the nucleus and cytoplasm (as described in detail in *Sections VI* and *VII*) (188). Both gene-specific and high-throughput studies have identified important roles for lncRNAs in the physiology and pathophysiology of the endocrine, reproductive, metabolic, immune, nervous, and cardiovascular systems (Figure 7). Moreover, lncRNAs are emerging as key regulators of cell proliferation and cell death, which are often associated with cancer. Some of the developmental effects of lncRNAs may be related to their effects on the pluripotency and lineage commitment of ESCs (105, 118, 189, 190). In the following sections, we review some of the recently uncovered functions of lncRNAs in these aspects of biology, beginning with endocrine-related systems.

A. LncRNAs and hormonal signaling: regulators, coregulators, and modulators of steroid receptors

Steroid receptors are members of a superfamily of DNA-binding transcription factors (nuclear receptors), many of which are regulated by the binding of small-molecule ligands, which play key roles in a wide variety of biological processes, including metabolism, reproduction, and development (191, 192). Their generally restricted expression patterns make them well suited to the control of tissue-specific biological responses. A number of reasonably well-characterized lncRNAs have been associated with steroid receptor functions (Figure 8). This may occur through 1) the regulation of nuclear receptor expression or activity by the lncRNAs or 2) the regulation of lncRNA expression by steroid receptors. Given the diverse array of biological functions controlled by nuclear receptors, the regulatory actions of lncRNAs could have a broad impact on physiology and disease.

1. SRA*: a nuclear receptor coregulator*

The first link between a lncRNA and hormone receptor-associated pathways was established in 1999 with the discovery of *SRA* (25). *SRA* was initially described as an RNA transcript specifically expressed in steroid hormone target tissues, which functions as a steroid receptor co-

Figure 8.

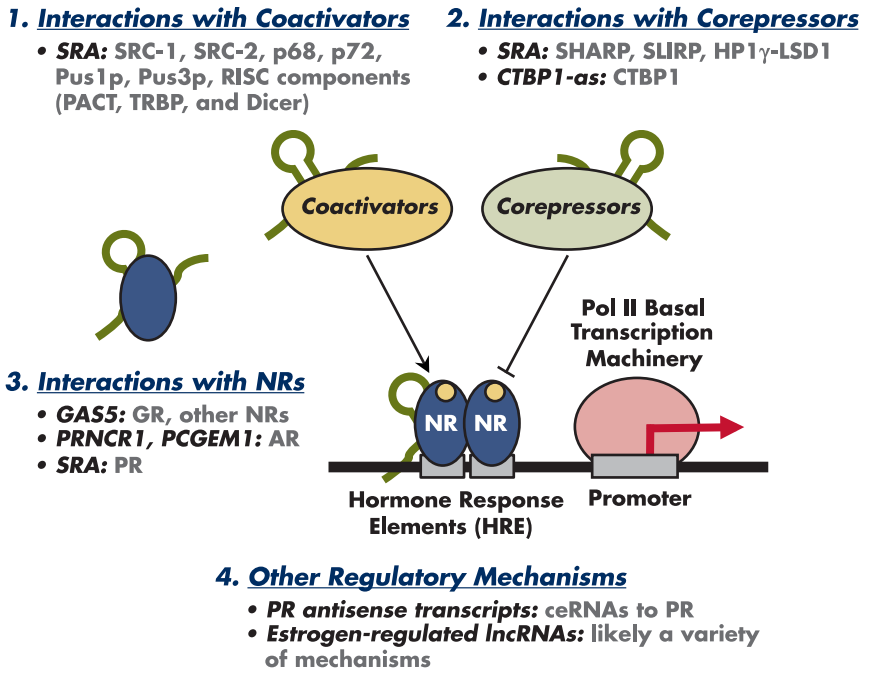

Figure 8. LncRNAs act as regulators, coregulators, and modulators of nuclear receptors. A number of lncRNAs have been implicated in the regulation of nuclear receptor (NR) functions, ultimately controlling receptor-mediated transcriptional programs. The regulation occurs through direct interactions with NR-associated coactivators and corepressors (as in the case of the lncRNAs *SRA* and *CTBP1-as*) (illustrations 1 and 2), direct interactions with NRs (as in the case of the lncRNAs *GAS5*, *PRNCR1*, and *PCGEM1*) (illustration 3), and additional transcriptional mechanisms acting either upstream or downstream of the NRs (eg, PR antisense transcripts, estrogen-regulated lncRNAs) (illustration 4).

activator. *SRA* interacts with steroid receptor coactivators 1 and 2 and facilitates ligand-dependent transactivation in reporter gene assays. In biochemical and cell-based assays, mutations that introduce early stop codons in *SRA* and inhibitors of protein synthesis convincingly demonstrate that the coactivator function of *SRA* is independent of translated protein products. Subsequent studies have substantiated the earlier findings and identified additional interaction partners involving both coactivators (eg, p68, p72, Pus1p and Pus3p, and components of the RNA-induced silencing complex, such as protein activator of the interferon-induced protein kinase, TAR RNA binding protein, and Dicer) (193–195) and corepressors (eg, SMRT- and HDAC1-associated repressor protein and stem-loop-interacting RNA-binding protein) (196, 197), thus expanding the role of *SRA* as a transcriptional coregulator (Figure 8, illustrations 1 and 2). More recently, *SRA* was identified in a chromatin-associated complex with unliganded progesterone receptor (PR) and an heterochromatin protein 1γ-LSD1 repressive complex at the promoter of hormone-regulated genes, where it functions to stabilize the complex and maintain the repressive state of the target genes before hormone induction (198) (Figure 8, illustration 3).

SRA has also been shown to be posttranscriptionally modified by pseudouridylation, a C-glycoside isomerization of the nucleoside uridine, which is mediated by pseudouridine synthase family members Pus1p and Pus3p (194, 199). Pus1p and Pus3p modify partially overlapping, but distinct, positions on *SRA*, altering *SRA*'s coregulator activity with nuclear receptors (194, 199). The extent to which other lncRNAs are posttranscriptionally modified has not been examined in great detail. This could be an important regulatory mechanisms across this class of regulatory RNAs, which should be studied in greater detail.

Interestingly, protein-coding isoforms of *SRA* containing an extended first exon have also been identified and characterized (26, 200), making *SRA* an interesting case of an RNA with roles as both a lncRNA and a protein-coding RNA. Nevertheless, the noncoding isoform displays differential expression patterns across different breast cancer cell lines and may play an oncogenic role in breast tumorigenesis (201–203), making the studies of such lncRNAs highly relevant to endocrine-related cancer research.

2. GAS5: *a nuclear receptor decoy*

Growth arrest-specific 5 (*GAS5*) is another lncRNA that has been shown to regulate the activity and function of multiple nuclear receptors, including the glucocorticoid, androgen, mineralocorticoid, and progesterone receptors (204). Unlike *SRA*, which participates in coactivator complexes as a scaffold, *GAS5* forms an RNA stem-loop structure to mimic a nuclear receptor DNA response element. *GAS5* interacts with the glucocorticoid receptor (GR) DNA binding domain and acts as a decoy GR response element, titrating GR away from its sites of transcriptional activity in a ligand-dependent manner (Figure 8, illustration 3). The *GAS5* RNA accumulates in fasting and growth arrested cells, thus functioning as a starvation- or growth arrest-linked riborepressor for GR and possibly other nuclear receptors that share the same DNA response element, facilitating steroid-modulated cell survival and

metabolism (204, 205). In adherent human cell lines, including 293T and MCF-10A, overexpression of *GAS5* suppresses cell growth and promotes apoptosis (206). *GAS5* is found at reduced levels in human breast carcinoma samples compared with their matched controls, suggesting a role for *GAS5* as a tumor suppressor (206).

3. PR *gene antisense transcripts: targets for antigene RNAs*

Some nuclear receptor-related lncRNAs may be more receptor-specific and may function at the level of the gene. Corey and colleagues have examined the transcriptional landscape of the *PR* gene and showed the existence of antisense RNA transcripts overlapping the *PR* gene promoter (207). They are likely to be lncRNAs, and at least 1 of them is spliced and polyadenylated. Although the coding potential of these transcripts has not been explicitly evaluated, they appear to act as RNAs, which can serve as targets for antigene RNAs (agRNAs). *PR* gene agRNAs are double-stranded RNAs complementary to the *PR* gene promoter, which act to increase expression of *PR* mRNA and protein levels after transfection into human breast cancer cells through a mechanism that involves the agRNAs and the recruitment of Argonaute proteins to the *PR* antisense transcripts (207, 208). As such, *PR* antisense transcripts are required for the agRNA-mediated activation of the *PR* gene, possibly through base pairing with the agRNAs (207, 209) (Figure 8, illustration 4).

The possibility that *PR* antisense lncRNAs modulate *PR* gene expression in response to endogenous agRNA-like molecules is an attractive one, and Corey and colleagues continue to search for RNA molecules that might mediate these effects (209). MicroRNAs are possible candidates; indeed, the inhibitory effects of *mir123b* on *PR* gene expression can be inhibited by *PR* antisense lncRNAs (209). These results suggest a role for *PR* antisense lncRNAs acting as ceRNAs to sequester microRNAs, adding to the growing list of lncRNAs that can serve as ceRNAs in multiple cellular models (139–142).

4. CTBP1-as, PRNCR1, *and* PCGEM1*: androgen-regulated lncRNAs*

The examples noted above illustrate how direct or indirect interactions between lncRNAs and nuclear receptors (or their genes) can affect receptor activity or expression. Other lncRNAs function as downstream targets of the gene-regulating activities of nuclear receptors but may also function in feedback loops that impact the activity of the receptor. *CTBP1-as*, an androgen-regulated lncRNA, is a NAT of *CTBP1*, which functions as a corepressor of androgen receptor (AR). In prostate cancer cells, androgen upregulates the expression of *CTBP1-as*, which in turn recruits the RNA-binding transcriptional repressor PTB-associated splicing factor and histone deacetylases to the promoter of the *CTBP1* gene, as well as additional target promoters, to mediate gene repression (49). Thus, *CTBP1-as* antagonizes the repressive functions of CTBP1 by limiting its expression, thereby facilitating the gene regulatory activities of AR. Moreover, *CTBP1-as* antagonizes the expression of additional target genes, including tumor suppressor genes, to promote tumor growth (49). Collectively, the available data suggest that androgen-regulated expression of *CTBP1-as* helps the AR bypass CTBP1-dependent repression to promote prostate cancer progression (Figure 8, illustration 2).

Similar to *CTBP1-as*, prostate cancer-associated noncoding RNA 1 (*PRNCR1*) and prostate-specific transcript 1 (nonprotein coding) (*PCGEM1*) add to the list of androgen-regulated lncRNAs that exhibit positive feedback in the androgen signaling pathway and, as such, affect prostate cancer progression (210). *PRNCR1* and *PCGEM1* interact directly and sequentially with AR in an androgen-dependent manner (210). These interactions enhance AR activity, facilitate AR gene activation programs, and drive prostate cancer cell proliferation (210). Furthermore, loss of these lncRNAs in castration-resistant prostate cancer cells leads to impaired tumor xenograph growth in vivo (210). These androgen-regulated lncRNAs suggest a model in which hormone signaling alters the expression or the activity of lncRNAs, which in turn modulate steroid receptor functions and ultimately control endocrine signaling outcomes. This model, however, has been challenged by Prensner et al (211), who have reported that 1) *PRNCR1* is not associated with prostate cancer, 2) neither *PRNCR1* nor *PCGEM1* is associated with poor patient outcomes, and 3) neither interact with AR, raising questions about their role in AR signaling.

5. Estrogen-regulated lncRNAs: components of a mitogenic program

Like the androgen-regulated lncRNAs *CTBP1-as*, *PRNCR1*, and *PCGEM1*, the expression of lncRNAs has been shown to be regulated by estrogens. In this regard, GRO-seq was recently used to explore the rapid effects of estrogen signaling on the entire transcriptome in MCF-7 human breast carcinoma cells, as well as identify thousands of novel estrogen-regulated ncRNAs, including ~1900 lncRNAs (66, 212, 213). Like a number of previously characterized estrogen-upregulated protein-coding genes, many estrogen-upregulated lncRNAs show estrogen-induced estrogen receptor-α binding in their proximal promoter regions, suggesting direct regulation by estrogen receptor-α (213). Those lncRNAs that show rapid and robust regulation in response to estrogen signaling are likely to play important roles in the estrogen signaling

pathway (Figure 8, illustration 4). Interestingly, knockdown of some of these estrogen-regulated lncRNAs inhibits the growth of breast cancer cells, suggesting potential roles for these lncRNAs in breast cancer cell proliferation (212, 213). Recent studies have also identified a whole host of estrogen-regulated eRNAs, some of which may meet the criteria for classification as lncRNAs (214, 215). Enhancer-derived lncRNAs have been shown to function as integral components of enhancers that drive the expression of distal target genes (6, 216).

B. LncRNAs and reproduction: regulators of mammary gland development

Mammary gland formation is a complex developmental program, much of which occurs after birth. During puberty, the formation of tubules within the mammary gland is coupled with branching, establishing the basic network of mammary ducts (217, 218). The gland continues to undergo dynamic changes during female reproductive cycles, with proliferation and regression of cells at specific times. Profound changes also occur during pregnancy, preparing the gland for lactogenesis and nursing at term and postpartum (217, 218). The mammary gland is a hormone-responsive organ, and many of the pregnancy-associated changes are mediated by the sex steroid hormones estrogen and progesterone (219). The lncRNAs described in the previous section (*Section VI.A*) that regulate the estrogen and progesterone signaling pathways are likely involved in mammary gland biology. For example, Lanz and colleagues (220) have shown that overexpression of human *SRA* lncRNA in mammary epithelial cells of MMTV-SRA transgenic mice leads to abnormal mammary gland development, implicating *SRA* in mammary gland biology. Results like this point to the potential importance of lncRNAs in the signaling programs that control key developmental processes.

Pregnancy-induced ncRNA (*Pinc*) and zinc finger antisense 1 (*Zfas1*) are two lncRNAs that have been identified and characterized as regulators of mammary gland proliferation and differentiation (221–223). Both lncRNAs show coordinated expression patterns along the mammary gland developmental axis in rats (*Pinc*) and mice (*Zfas1*) (223, 224). In HC11 mouse mammary epithelial cells, knockdown of *Pinc* isoforms increases differentiation in a morphological dome formation assay, whereas overexpression of *Pinc* results in reduced expression of differentiation marker genes (222). In contrast, knockdown of *Zfas1* in HC11 cells promotes cell proliferation, increases the expression of differentiation marker genes, and increases dome formation (223). The concerted actions of these lncRNAs in proliferation and differentiation ensure the proper developmental program in the mammary gland. Mechanistically, *Pinc* interacts with the chromatin-modifying PRC2 complex in RIP assays and may serve as an epigenetic regulator at the molecular level (222). The molecular basis for *Zfas1* function remains unclear.

C. LncRNAs and metabolism: adipogenesis and metabolic disorders

The adipose and pancreas exemplify organs in which endocrine and metabolic functions are well integrated (225–227). Impaired functions in these organs lead to metabolic dysregulation and can be an underlying cause of important health concerns, such as obesity and diabetes (228, 229). Adipose consists primarily of adipocytes and is a major site for lipid storage and metabolism. As an endocrine organ, it releases adipokines that signal to the other organs and tissues in the body to regulate lipid and glucose homeostasis (230). The pancreas contains multiple cell types, including β-cells, which produce and secrete insulin as part of the endocrine pancreas, and acinar cells, which produce and secrete digestive enzymes and part of the exocrine pancreas (231). Loss of β-cells is a direct cause of type 1 diabetes, and insufficient insulin production from β-cells is a major contributing factor for type 2 diabetes (232, 233). Thus, lncRNAs that alter adipose and pancreas function can have profound implications on metabolism. The aforementioned lncRNA *SRA*, which acts as a coregulator of multiple steroid receptors and has been implicated in endocrine and reproductive functions, also coactivates peroxisome proliferator-activated receptor γ (PPARγ), a nuclear receptor that serves as a master regulator of adipogenesis (234). Functional interactions between *SRA* and PPARγ facilitate the adipogenic transcriptional program associated with adipocyte differentiation. Knockout of the *SRA* gene in mice protects against diet-induced obesity and improves glucose tolerance (235), suggesting broad functions of SRA in metabolic processes.

1. Lnc-RAP*s,* PU.1 AS*, and* Blnc1*: adipogenic lncRNAs*

LncRNAs regulating adipogenesis have also been identified in large-scale genomic studies. By evaluating the differential expression of lncRNAs across primary brown and white adipocytes, preadipocytes, and cultured adipocytes, Sun and colleagues identified 175 lncRNAs that are specifically regulated during adipogenesis (12). They selected 20 lncRNAs that are likely regulated by PPARγ and CCAAT/enhancer-binding protein-α, master regulators of adipogenesis, to perform a loss-of-function screen, from which they showed that 10 of them, including *lncRAP-1* and *lncRAP-2*, function to modulate adipocyte differentiation (12). Depletion of these adipogenic lncRNAs causes dysregulation of adipogenic gene expression pro-

grams, suggesting an underlying transcriptional mechanism for these lncRNAs.

PU.1 antisense (*AS*) is another lncRNA that regulates adipogenesis. As the name implies, *PU.1 AS* is a NAT of the *PU.1* gene and, like other NATs, it functions to antagonize its sense transcript. Specifically, *PU.1 AS* forms an RNA duplex with, and inhibits the translation of, *PU.1* mRNA. PU.1 is a transcription factor that inhibits the differentiation of preadipocytes into adipocytes (236). By blocking the translation of *PU.1* mRNA, *PU.1 AS* allows adipogenesis to proceed (237). Interestingly, a nuclear, nonpolyadenylated, ncRNA originating upstream of, and encompassing the entire, *CEBPA* mRNA, called extra-coding CEBPA (*ecCEBPA*), interacts with the DNA methyltransferase DNMT1 to promote expression of the *CEBPA* mRNA (121). *CEBPA* encodes the adipogenic transcription factor CCAAT/enhancer-binding protein-α, a key driver of adipogenesis (238). Although not specifically demonstrated, this mode of regulation could play an important role in adipogenesis.

Blnc1 (brown fat lncRNA 1), which was identified through global profiling of lncRNA expression during thermogenic adipocyte formation, promotes the differentiation and function of brown and beige adipocytes (239). *Blnc1* forms a complex with transcription factor early β-cell factor 2 to stimulate a thermogenic gene expression program that promotes differentiation into brown and beige adipocytes. Interestingly, the *Blnc1* gene is a target of early β-cell factor 2, establishing a feed-forward regulatory loop to drive adipogenesis toward a thermogenic phenotype (239). LncRNAs regulating adipocyte differentiation, such as those described here, have important implications for endocrine-mediated metabolic functions.

2. HI-LNC25 *and islet cell lncRNAs: β-cell function and diabetes*

The endocrine pancreas is the key site for glucose metabolism and is essential for the maintenance of glucose homeostasis. To identify lncRNAs implicated in diabetes, Morán and colleagues (240) performed transcriptome analysis in human pancreatic islets and β-cells and reported the identification more than 1100 intergenic and antisense islet cell lncRNAs. Some of these lncRNAs are up- or downregulated in islet samples from individuals with type 2 diabetes, whereas others coincide with diabetes susceptibility loci. One intergenic lncRNA, *HI-LNC25*, is specifically expressed in β-cells (240). Knockdown of *HI-LNC25* in EndoC-β-H1 human β-cell cells, reduced the expression of *GLIS3*, a protein-coding transcription factor implicated in diabetes (240). Collectively, these results provide initial evidence that lncRNAs may be involved in β-cell function and diabetes pathophysiology.

3. 116HG*: energy imbalances in Prader-Willi syndrome*

Prader-Willi syndrome (PWS) is a genetic disorder with a spectrum of phenotypes, including childhood obesity, linked to the PWS critical region on chromosome 15 (241). Interestingly, the metabolic disorders in PWS have been linked to ncRNAs. Specifically, loss of the paternally imprinted *snord116* gene cluster, which encodes multiple *SNORD116* snoRNAs, sno-lncRNAs, and a lncRNA called *116HG*, underlies the symptoms of PWS in humans and mouse (45, 242–244). Not only does the *116HG* locus serve as the host for *SNORD116* snoRNAs, which are potential drivers of PWS phenotypes, the *116HG* lncRNA itself interacts with the transcriptional coactivator retinoblastoma binding protein 5 to regulate transcriptional programs underlying circadian energy homeostasis in postnatal neurons in mice (245). Results from ChIRP-seq and RNA-seq experiments showed that *116HG* likely prevents the binding of RBBP5 to the promoters of target genes, hence preventing the upregulation of genes, which encode proteins involved in chromatin modification and metabolic signaling (245). Consequently, Snord116del mice lacking *116HG* exhibit dysregulation of metabolic genes and diurnal energy expenditure in the brain, as shown in genomic and metabolic analyses, suggesting that the *116HG* contributes to the energy imbalance associated with PWS (245).

LncRNAs originating from the PWS critical region on chromosome 15 may affect the expression of imprinted genes at loci on other chromosomes. For example, the lncRNA *IPW*, whose gene is located in the PWS critical region, regulates the *DLK1-DIO3* region chromosome 14 (246). Overexpression of *IPW* in PWS promotes downregulation of maternally expressed genes in the *DLK1-DIO3* region through alterations of histone modifications, rather than DNA methylation (246). Whether this altered regulation contributes to the PWS metabolic phenotypes has not been determined, but the possibility exists. Collectively, the studies highlighted here illustrate the emerging roles of lncRNAs in adipogenesis and metabolic disorders.

D. LncRNAs in the immune system: innate and adaptive immune responses

The immune system comprises a collection of biological molecules, structures, cells, and processes that lead the defense against disease. The immune system has 2 overlapping and interactive components, referred to as the innate immune system (eg, phagocytosis, antimicrobial peptides, and the complement system) and the adaptive immune system (eg, antigen-specific responses and immunological memory) (247). Both are essential for maintaining homeostasis in the face of infections. Malfunctions in

Figure 9.

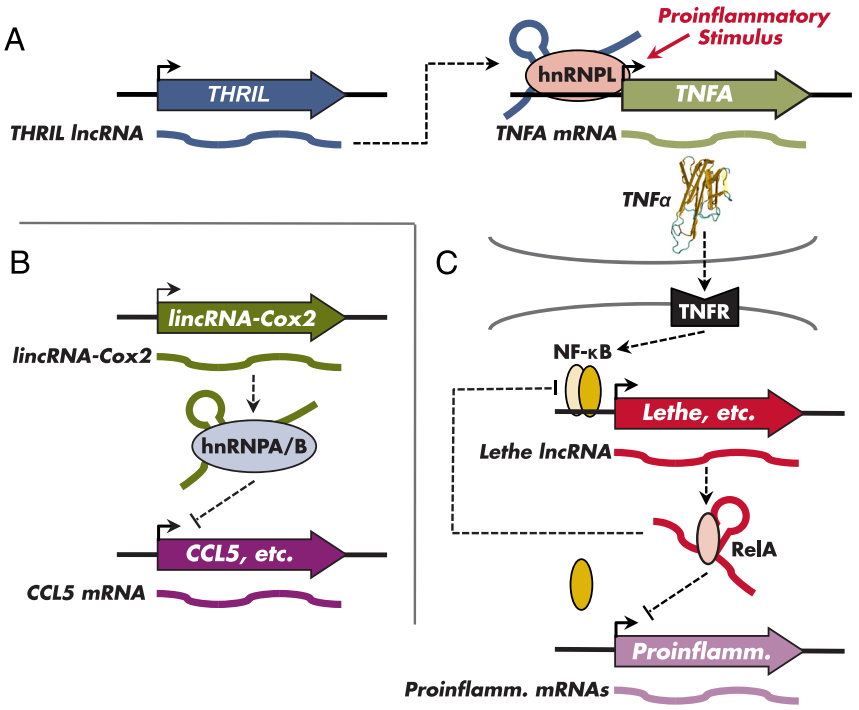

Figure 9. A possible integrated lncRNA network controlling inflammatory responses. Schematic representation of known proinflammatory (A and C) and anti-inflammatory (B) gene regulatory responses. Although the specific interrelationships illustrated here are not specifically known, they are inferred from the literature. A, The lncRNA *THRIL* regulates the expression of *TNFA*, the gene encoding TNFα (125), a key proinflammatory cytokine. B, *LincRNA-Cox2*, one of the most highly upregulated lincRNAs in Tlr4-stimulated mouse dendritic cells bone marrow-derived macrophages (18), mediates both activation and repression of important inflammatory genes, such as those encoding cytokines and chemokines (253). This repression involves the formation of a lncRNA-protein complex containing *lincRNA-Cox2* and hnRNP proteins (253). C, The expression of *Lethe* is regulated by the proinflammatory cytokine TNFα through the transcription factor NF-κB, a master regulator of inflammatory responses. *Lethe* acts as a negative regulator of NF-κB-dependent inflammatory signaling through physical interactions with the RelA subunit of NF-κB, forming a negative feedback loop to modulate inflammatory responses (254).

the innate and adaptive immune systems can lead to pathological inflammatory responses, autoimmunity, and immunodeficiency (248–251). Recent studies have implicated lncRNAs in the biology of the immune system, orchestrating both innate and adaptive immune responses to support host defense mechanisms against invading pathogens (Figure 9).

1. LincRNA-Cox2, Lethe, *and* THRIL*: regulators of inflammation in innate immunity*

The innate immune system is the first line of defense against infection. Its major functions include promoting inflammatory responses and organizing the recruitment and activation of various immune cell types, mostly through the well-controlled production and release of cellular mediators, such as cytokines and chemokines. Transcriptome profiling in the innate immune system have shown global differential expression of lncRNAs in response to influenza virus infection in mouse lung tissue samples (252). In this regard, *lincRNA-Cox2* has been shown to function as a regulator of inflammatory signaling in the innate immune system. *LincRNA-Cox2* is located proximal to the *Ptgs2* (*Cox2*) gene and is one of the most highly upregulated lincRNAs in toll-like receptor 4 stimulated mouse dendritic cells bone marrow-derived macrophages (18). *LincRNA-Cox2* mediates both activation and repression of important inflammatory genes, such as those encoding cytokines and chemokines (253). This repression involves the formation of a lncRNA-protein complex containing *lincRNA-Cox2* and hnRNP proteins (253) (Figure 9B).

Lethe is another lncRNA that has been shown to play a critical role in innate immune responses by regulating the inflammatory transcriptional program. Rapicavoli and colleagues (254) identified hundreds of lncRNAs that are regulated by the proinflammatory cytokine TNFα through the transcription factor nuclear factor (NF)-κB, a master regulator of inflammatory responses, in mouse embryonic fibroblasts. Among them, they focused on *Lethe*, a nuclear lncRNA produced from a pseudogene of *Rps15*. *Lethe* acts as a negative regulator of NF-κB-dependent inflammatory signaling through physical interactions with NF-κB protein, forming a negative feedback loop to modulate inflammatory responses (254) (Figure 9C). As noted above, the lncRNA *THRIL* regulates the expression of *TNFA*, the gene encoding TNFα (125), a regulatory event that may act upstream of *Lethe* in the aforementioned pathway (Figure 9A).

The expression of other cytokines is also regulated by lncRNAs. For example, the lncRNA *NEAT1*, which is required for the formation of nuclear body paraspeckles, facilitates the expression of antiviral cytokines, such as IL-8. *NEAT1* binds to the proline/glutamine-rich splicing factor splicing factor proline/glutamine-rich, a repressor of *IL8* gene expression, promoting the relocalization of splicing factor proline/glutamine-rich from the *IL8* promoter to paraspeckles, leading to transcriptional activa-

tion of *IL8* (255). LncRNA-mediated movement of genes between architectural structures in the nucleus has been noted in other systems as well (eg, movement between *TUG1*-containing Polycomb bodies and *MALAT1*-containing interchromatin granules in response to growth signals) (170), suggesting that this may be a general mode of regulation.

2. NeST, GAS5, lincR-Ccr2–5′AS, *and* lnc-DC*: regulators of T cells in adaptive immunity*

The adaptive immune system involves highly specialized immune cell types that act to contain and eliminate invading pathogens. Recent studies have shown that, in addition to innate immunity, lncRNAs are involved in the regulation of adaptive immune responses as well, particularly in the context of various subsets of T lymphocytes (256, 257), as illustrated by the lncRNAs *NeST* and *GAS5* (206, 258). *NeST*, also known as *TMEVPG1* or *lincR-ifng-3′AS*, is transcribed from a gene that is positioned next to the *IFNG* locus (257–259). *NeST* acts as a transcriptional enhancer to increase the expression of *IFNG* in Th1 cells and cytotoxic $CD8^+$ T cells (257, 258). Consequently, *NeST* modulates adaptive immune responses to pathogens, including Theiler's virus and *Salmonella enterica* in transgenic mouse models (258). Likewise, *GAS5*, another lncRNA implicated in T-cell biology, has been shown to play an essential role in the control of growth arrest, apoptosis, and cell cycle in both a human leukemic T-cell line and human peripheral blood T cells (206).

Large-scale transcriptome experiments have led to the identification of thousands of lncRNAs in specific T-cell populations. For example, Pang and colleagues (256) examined the expression profiles of lncRNAs in $CD8^+$ T cells by using existing microarray datasets from human and mouse $CD8^+$ T cells, as well as by performing custom expression microarrays on naive and activated $CD8^+$ T cells isolated from the mouse spleen. In addition, Hu and colleagues (257) performed RNA-seq on 42 T-cell populations, including $CD4^+$, $CD8^+$, and double-negative thymic T cells, thymus-derived regulatory T cells, and various types of in vitro differentiated $CD4^+$ T cells. Both studies showed cell-specific and dynamically regulated expression patterns of subsets of lncRNAs across the different the T-cell populations. Interestingly, the genes for many of the lncRNAs with T-cell-specific patterns of expression are located in proximity to protein-coding genes with immunological functions. For example, the expression of *lincR-Ccr2–5′AS* in T helper type 2 cells is highly correlated with the expression of nearby protein-coding genes involved in chemokine signaling pathways (257). Specifically, *LincR-Ccr2–5′AS* is required for expression of the neighboring chemokine genes *Ccr2* and *Ccr3*, which facilitate the migration of T helper type 2 cells into the lung tissues (257).

Finally, some lncRNAs support the important functional interplay between the innate and adaptive immune systems. For example, *lnc-DC* is a lncRNA expressed exclusively in dendritic cells, which are antigen-presenting cells that process antigens and present them on the cell surface to T cells. *Lnc-DC* supports the differentiation of dendritic cells from monocytes, as well as the ability of dendritic cells to stimulate T cell activation, by activating the transcription factor STAT3 (260). Collectively, the studies described in this section highlight the cell-type-specific expression of lncRNAs in T cells and identify 3 specific lncRNAs (ie, *NeST, GAS5*, and *lincR-Ccr2–5′AS*) that play key roles in T-cell function.

VII. LncRNAs in Other Biological Systems

A number of recent studies have begun to elucidate the roles of lncRNAs in a wide variety of other biological systems. Two of the best characterized with respect to lncRNA function are the nervous and cardiovascular systems (Figure 7). As observed with the other biological systems described above, the lncRNAs that impact the biology of the nervous system and the cardiovascular system illustrate important concepts about the structure, function, and biological roles of lncRNAs in health and disease.

A. LncRNAs in the nervous system: neural development and disorders

The nervous system is the most complex organ system in the human body, comprising numerous cell types and an exquisite regulatory network of cellular activity. Given their temporally and spatially controlled patterns of expression, lncRNAs are well suited for providing an additional regulatory layer to fine tune the cellular outcomes needed for proper neuronal development and function. Indeed, lncRNAs are abundantly expressed in the cells of the central nervous system, dynamically regulated across the developmental axis and in response to neuronal activity, and expressed in localized patterns in specific brain structures (261–263). In this regard, lncRNAs have been implicated in neuronal development and the differentiation of neurons as well as the pathogenesis of neurological disorders (264–267). Recent results from Rinn's lab (187) using a collection of 18 lncRNA knockout mouse lines have reinforced the important role that lncRNAs play in the brain.

1. Identification of lncRNAs important in neural development

A number of studies using high-throughput approaches have established a connection between lncRNAs and the nervous system. As mentioned earlier, Guttman and colleagues (105) screened for lncRNAs that alter global gene expression in ESCs through a loss-of-function RNA interference screen. Among the hits was a set of lncRNAs that function to negatively regulate neuroectodermal differentiation (105). Moreover, Khalil and colleagues found that many lncRNAs physically associate with the REST/CoREST complex in RIP-chip experiments (21). Given the prominent roles of this chromatin-modifying complex in neural development, it seems likely that some of these lncRNAs contribute to REST/CoREST-related neuronal functions.

Ramos and colleagues (268) examined lncRNAs expressed in the adult mouse subventricular zone neural stem cell lineage through an integrative genomic approach, including RNA-seq, RNA CaptureSeq, and ChIP-seq. They characterized 2 lncRNAs, *Six3OS* and *Dlx1as*, with roles in the glial-neuronal lineage specification of multipotent adult stem cells (268). The evolutionarily conserved lncRNA *TUNA* (Tcl1 upstream neuron-associated lincRNA, or *megamind*) controls pluripotency and neural lineage commitment in ESCs (190). Other studies have focused on the role of lncRNAs during corticogenesis, using a fluorescence reporter mouse line that allows transcriptome profiling in isolated subpopulations of proliferating neural stem cells, differentiating progenitors, and newborn neurons during brain development (269). A subset of lncRNAs show differential expression across the different subpopulations, and several of them are associated with neuronal phenotypes. For example, the lncRNA *Miat* functions to ensure the proper splicing of *Wnt7b* and regulate neurogenic differentiation and neuronal survival during brain development (269). In addition, Ulitsky and colleagues (55) used an integrative genomic approach to identify 550 lncRNAs in zebrafish. Morpholino antisense oligo-mediated knockdown of 2 of them, *cyrano* and *megamind*, causes neural defects in zebrafish in vivo (55). Interestingly, these defects can be rescued by injection of mammalian orthologs of the lncRNAs from either mouse or human, highlighting the conserved function of lncRNAs, even when sequence similarity may be limited (55).

Many other lncRNAs have been implicated in brain and retinal development. These have been reviewed elsewhere and are mentioned briefly here. LncRNAs with important roles in brain development include 1) *RMST*, a highly expressed lncRNA required for the differentiation of midbrain neurons (270) and 2) *utNgn1*, a lncRNA transcribed from the *Neurog1* enhancer region, which regulates *Neurog1* expression to drive development of the cortical region of the brain (271). LncRNAs with important roles in retinal development and differentiation include *Tug1*, *Vax2OS*, *Six3OS*, *RNCR2* (aka *Gomafu* or *Miat*) (272–275). In addition to the regulation of retinal cell fate, *Six3OS* and *RNCR2* have been shown to play roles in neural cell fate determination in the central nervous system, including a role in adult neurogenesis in subventricular zone neurons for *Six3OS* (268, 276).

2. LncRNAs in neuronal function and activity

A growing number of lncRNAs have been implicated in neuronal functions. For example, the highly abundant lncRNA *MALAT1* is induced in response to neuronal activity and has been shown to affect synapse function in cultured neurons (169). Although there is a lack of observable neurological phenotypes in *Malat1*-knockout mice (173), knockout of other lncRNAs is associated with neurological phenotypes in vivo. *Dlx1as* is a lncRNA that is transcribed antisense to the *Dlx1* gene from the locus harboring the distal-less homeobox (Dlx) genes, which have been implicated in the differentiation of multiple neuronal subtypes (277). Knockdown of *Dlx1as* downregulates the expression of *Dlx1* and *Dlx2* (50). Moreover, functional ablation of *Dlx1as* in mice without affecting *Dlx1* through a carefully designed gene-targeting approach results in more γ-aminobutyric acid (GABA)-ergic interneurons (277). In addition, the mice exhibit neurological phenotypes that are similar to those caused by an excess of Dlx1, suggesting that *Dlx1as* acts as a negative regulator of *Dlx1* to alter its mRNA transcript levels (50). *Evf-2* is another lncRNA transcribed from the *Dlx* locus, originating from an ultraconserved enhancer region between the *Dlx5* and *Dlx6* genes (278). *Evf-2* acts as a transcriptional coactivator of Dlx2 to modulate the activity of the enhancer and to maintain the expression of *Dlx5*, *Dlx6*, and *Gad1*, protein-coding genes required for the proper formation of the GABA-dependent neuronal circuitry (278, 279). In this regard, *Evf-2* mouse mutants show reduced numbers of GABAergic interneurons, suggesting a role of *Evf-2* in adult brain development (279).

3. Antisense lncRNAs and neurological disorders

Among the lncRNAs implicated in neurological disorders, there is a recurring theme of lncRNAs transcribed antisense to a protein-coding gene playing key roles in the pathophysiology of the disease. These NAT lncRNAs are thought to contribute to the respective diseases by modulating the levels of the sense protein-coding transcript. For example, *BACE1-AS* is transcribed antisense to the *BACE1* gene, which encodes β–amyloid, a key protein

component of the amyloid plaques that are central to the pathophysiology of Alzheimer's disease (AD) (280). *BACE1-AS* exhibits elevated expression levels in AD individuals, leading to increased levels of BACE1 protein by stabilizing the *BACE1* mRNA (280). In addition to *BACE1-AS*, the brain cytoplasmic RNAs, *BC200* in human and *BC1* in mouse, are lncRNAs with dysregulated expression in individuals with AD and mouse models of the disease, respectively (38). *BC200* and *BC1* regulate the translation of mRNAs encoding specific proteins that contribute to the pathogenesis of AD (38).

BDNF-AS is another antisense lncRNA that has been implicated in neurological disorders, including Huntington's disease (HD) and schizophrenia. *BDNF-AS* likely contributes to these diseases by antagonizing the expression of the sense *BDNF* transcript, which encodes brain-derived neurotrophic factor, an important neuronal growth factor shown to plays a crucial role in the pathogenesis of HD (281, 282). *UBE3A-ATS* is a lncRNA transcribed antisense to *UBE3A*, a gene encoding ubiquitin-protein ligase E3A, which has been implicated in Angelman syndrome, an imprinting-related neurodevelopmental disorder (283–285). *UBE3A-ATS* functions to regulate the epigenetic silencing of the imprinted sense allele, hence contributing to the disease pathology (286, 287). In all of these cases, the potential utility of targeting the antisense lncRNAs in vivo has been suggested as a therapeutic approach in treating the cognate neurological disorder.

4. LncRNAs and trinucleotide repeat expansion disorders

HD is one of a number of trinucleotide repeat disorders, genetic diseases caused by the abnormal expansion of trinucleotide repeat sequences in the associated genes. Several lncRNAs are associated with other trinucleotide repeat disorders, such as fragile X syndrome, fragile X tremor ataxia syndrome, and spinocerebellar ataxias (SCAs). The lncRNAs *FMR4* and *ASFMR1* are transcribed from the disease-causing *fragile X mental retardation 1* (*FMR1*) gene locus and are associated with fragile X syndrome, fragile X tremor ataxia syndrome, and possibly autism (288, 289). The lncRNA SCA type 7 antisense noncoding transcript 1 (*SCAANT1*) is transcribed antisense to the gene encoding ataxin-7 (*ATXN7*), a component of the SPT3-TAF(II)31-GCN5L acetylase coactivator complex, to regulate *ATXN7* expression and contribute to the disease phenotypes in SCA type 7 (290). In contrast, the lncRNA ataxin 8 opposite strand (*ATXN8OS*) is transcribed from within the trinucleotide repeat expanded region of the SCA type 8 disease locus and has been show to play a role in related disease phenotypes in *Drosophila* in vivo (291–293). Collectively, the studies described here illustrate the wide range of developmental and disease processes in neurons that are mediated or altered by tissue-specific expression of lncRNAs.

B. LncRNAs in cardiac and skeletal muscle: muscle development and pathologies

Muscle, a soft tissue representing one of the 4 main tissue types in the body (in addition to connective, nervous, and epithelial tissues), comprises 3 distinct types in vertebrates: cardiac, skeletal, and smooth muscle. Cardiac muscle (aka myocardium or heart muscle) and skeletal muscle are striated, containing sarcomeres that are arranged into highly regular bundles or myofibrils. In contrast to smooth muscle, which exhibits prolonged contractions, cardiac and skeletal muscles contract and relax in short, intense bursts. Cardiac muscle drives the contractions of the heart that circulate blood through the vascular system, whereas skeletal muscle allows skeletal movement through tendon-mediated anchors to the bone. Although distinct in their functions, different striated muscles share some structural, molecular, and cellular features that underlie the common aspects of their biology.

1. LncRNAs in the heart: development, physiology, and disease

The heart is an essential organ in the circulatory system. Its proper functioning is essential for life, but it is a common site of lethal disorders. In fact, heart disease is the leading cause of death in the United States. Increased understanding of the complex biology of the heart, including the cardiomyocytes and vascular cells within it, is needed to reduce the incidence and mortality of heart disease. Not surprisingly, lncRNAs have been shown to serve varied and important functions in cardiac development, physiology, and pathophysiology. Transcriptome analyses in an immortalized adult ventricular cardiomyocyte cell line have revealed a wide array of intergenic transcripts, many of which are lncRNAs (294). Interestingly, the expression of many of these lncRNAs, including *MALAT1*, is altered in response to TNFα, a stress-related cytokine that promotes proinflammatory responses and may contribute to cardiac pathologies. Other lncRNAs, such as *Braveheart* and *Fendrr*, have been implicated in normal heart development and physiology. *Braveheart* is expressed in mouse ESCs and adult heart cells and is required for the transition from nascent to cardiac mesoderm by controlling the expression of core cardiac transcription factors, including mesoderm posterior protein 1, to regulate the cardiovascular gene network (91). On the other hand, *Fendrr* is specifically expressed in the lateral plate mesoderm, where it modulates the epigenetic environment of mesoderm-

specific genes and plays an essential role in the development of heart and body wall in mice (119).

Interestingly, several lncRNAs with molecular functions initially characterized in other tissues have been implicated in cardiovascular diseases. *ANRIL*, originally associated with cancer, is a lncRNA located in the strongest genetic susceptibility locus for coronary artery disease, and its expression in patients correlates with the severity of atherosclerosis (295). *SRA*, a gene encoding a lncRNA initially described as a steroid receptor coactivator in steroid hormone target tissues, is located in a 600 kb region of linkage disequilibrium associated with cardiomyopathy. Furthermore, depletion of the *SRA* homolog in zebrafish results in severe myocardial dysfunction (296). *RCNR2* (aka *Gomafu* or myocardial infarction-associated transcript *MIAT*), identified in the nervous system as a regulator of retinal development, contains SNPs that are associated with myocardial infarction (297). How *MIAT* is linked mechanistically to the disease, however, remains unknown.

Other lncRNAs that have been identified specifically in the heart have functions directly related to cardiac biology. For example, *Mhrt* (myosin heavy-chain-associated RNA transcript) protects the heart from pathological hypertrophy by antagonizing the functions of BRM/SWI2-related gene 1, a chromatin-remodeling enzyme that promotes aberrant gene expression and cardiac myopathy in response to stress (298). *Linc-MYH*, a lncRNA whose gene shares a common enhancer with the genes encoding myosin heavy chain (MYH), plays a key role in adult fast-type myofiber specialization by preventing slow-type and enhancing fast-type *MYH* gene expression (299). This is achieved through the coordinated expression of fast MYHs and *linc-MYH* by Six1 homeoproteins acting at the common enhancer.

Similar to NATs in the brain, a number of antisense lncRNAs have been implicated in cardiac biology through interplay with their cognate sense genes. For example, lncRNAs transcribed antisense to genes encoding the essential cardiac proteins cardiac troponin T type 2, myosin heavy chain β (β-MHC), and atrial myosin light chain 1 (ALC-1) are required for their proper expression (300–302). Specifically, *TNNT2-AS,* which is transcribed antisense to the gene encoding cardiac troponin T type 2, forms a duplex with *TNNT2* mRNA to regulate translation of the mRNA (300). In addition, *MYHCB-AS*, which is transcribed antisense to the gene encoding β-MHC, regulates isoform switching between α-MHC and β-MHC (301). Finally, induced expression of *MYL4-AS,* which is transcribed antisense to the gene encoding ALC-1, is associated with reduced ALC-1 protein levels in hypertrophic ventricles (302). As these 3 examples illustrate, lncRNAs add an important layer of regulation to the cardiac gene network, contributing to cardiovascular function and pathophysiology. Nevertheless, the study of lncRNAs in the cardiovascular system is still in its infancy, and additional studies are needed to elucidate fully the underlying mechanisms that ultimately impact cardiac biology.

2. LncRNAs in skeletal muscle: muscle differentiation and muscular dystrophy

LncRNAs have also been implicated in muscle differentiation and myopathies in skeletal muscle. For example, *SRA*, a lncRNA that was described in other biological contexts in *Sections VI.A.1*, *VI.B*, *VI.C*, and *VII.B.1*, functions in a complex with p68/p72 to coregulate myogenic differentiation 1 and promote skeletal muscle differentiation (303). This process may be inhibited by the protein product of the *SRA* gene, SRA protein, which is thought to bind SRA and inhibits its coregulatory function (200), although direct interactions between SRA and SRA protein have been questioned by McKay et al (304). Other lncRNAs affect skeletal muscle differentiation by producing or inhibiting microRNAs. The lncRNA *H19* produces microRNAs *miR-675–3p* and *miR-675–5p*, which target mRNAs encoding Smad transcription factors to promote skeletal muscle differentiation and regeneration (305). *H19* also sponges *let-7* microRNAs to control muscle differentiation (143).

In contrast, the lncRNA *linc-MD1* functions as a ceRNA in mouse and human myoblasts to control the timing of muscle differentiation. It sequesters *miR-133* and *miR-135*, 2 microRNAs targeting key transcription factors involved in muscle-specific gene expression programs, ultimately affecting muscle differentiation (139). *Linc-MD1* is also the host transcript of *miR-133b*, a microRNA whose biogenesis is mutually exclusive with *linc-MD1*. HuR protein, whose levels are reduced by *miR-133*, promotes *linc-MD1* accumulation over *miR-133b* accumulation by binding to *linc-MD1* and repressing cleavage by Drosha. The sponging activity of *linc-MD1* stabilizes HuR levels in a feed-forward positive loop, reinforcing *linc-MD1* sponge activity (306). *Linc-MD1* has also been associated with Duchenne muscular dystrophy (DMD). It shows significantly reduced expression in muscle from DMD individuals compared with controls (139). The differentiation delay observed in DMD myoblasts can be rescued by the ectopic expression of *linc-MD1*, suggesting potential therapeutic utility for the lncRNA in DMD and possibly other muscle diseases.

The lncRNA *DBE-T* has been implicated in facioscapulohumeral muscular dystrophy (FSHD), one of the most common myopathies, through a mechanism that differs

from that observed for *linc-MD1* in DMD. *DBE-T* is transcribed specifically in FSHD patients from a gene proximal to the *D4Z4* locus, which contains many copies of D4Z4 repeat units (307). In unafflicted individuals, a sufficient number of the D4Z4 repeat units can recruits the Polycomb complex to ensure chromatin compaction and repression of nearby genes. In FSHD individuals, a reduced number of D4Z4 repeat units results in chromatin derepression, allowing transcription of *DBE-T* (307). The expressed *DBE-T* then initiates a positive feedback loop to recruit the Trithorax complex, which antagonizes the Polycomb complex to further derepress the expression of nearby genes and *DBE-T*, leading to FSHD (307). Thus, *DBE-T* plays a critical role in the etiology of FSHD, and the therapeutic potential of targeting *DBE-T* for the treatment of the disease is under investigation. Collectively, the studies described here have revealed key roles for a broad array of lncRNAs in muscle development, function, and disease.

VIII. LncRNAs in Cancer: Oncogenes and Tumor Suppressors

Some of the first biological functions ascribed to lncRNAs were related to their roles in cancers, such as the endocrine-related cancers noted above. Accumulating evidence supports a role for lncRNAs in a much broader array of cancer types. Du and colleagues (308) performed a global analysis of lncRNAs in cancer, comprehensively interrogating the expression of more than 10 000 lncRNA genes in over 1000 tumor samples. An integrative analysis of the expression profiles, somatic copy-number alterations, and clinical information allowed the authors to identify lncRNAs associated with specific cancer subtypes and clinical outcomes as well as lncRNAs that potentially function as drivers of oncogenic cell growth (308). Other studies have also worked toward the identification and functional analyses of lncRNAs that are expressed and function in a particular type of cancer or are expressed and function more broadly in cancers (309–311). Studies such as these highlight how the role of lncRNAs in cancers has emerged as a major focus in the field. This topic has already been reviewed extensively elsewhere (17, 312–315), so we will only provide a few highlights here.

A. LncRNAs and oncogenesis

A number of lncRNAs have been shown to promote oncogenic cell proliferation, including *SRA*, *PRNCR1*, and *PCGEM1*, which were described in *Sections VI.A.1* and *VI.A.4*. All 3 of these lncRNAs facilitate mitogenic transcriptional programs associated with nuclear receptors. *PCGEM1* has also been shown to play an antiapoptotic role in doxorubicin-treated prostate cancer cells, further establishing its role as an oncogenic lncRNA (316). Similarly, the DNA damage-induced lncRNA *PANDA* also functions in cancer cells to evade apoptotic cell death responses, but it does so by attenuating the expression of proapoptotic genes (104). Alternatively, other oncogenic lncRNAs enhance the proliferative potential of cancer cells by negatively regulating the expression of growth-inhibiting tumor suppressors. For example, *ANRIL* and *p21NAT* are antisense lncRNAs that function to downregulate the expression of their cognate sense tumor suppressor genes, *CDKN2B* (encoding p15) and *CDKN1A* (encoding p21), respectively (98, 317). Likewise, *FAL1* (focally amplified lncRNA on chromosome 1), a lncRNA whose gene is subject to somatic copy number alterations in cancers, associates with the epigenetic repressor β cell-specific Moloney murine leukemia virus integration site 1 to modulate the transcription of *CDKN1A* and other genes (310). Moreover, the prostate cancer-associated transcript 1 (*PCAT-1*) acts as a transcriptional repressor to lower the expression of tumor suppressing target genes, including *BRCA2* (318). Through these and other related mechanisms, lncRNAs can enhance cellular oncogenic potential and promote cancer formation.

B. LncRNAs and tumor suppression

Other lncRNAs have been associated with tumor suppressor functions. *MEG3* (maternally expressed 3), the first lncRNA suggested to function as a tumor suppressor, has been implicated in meningiomas (319). It functions to stimulate both p53-dependent and p53-independent pathways, which collectively inhibit cancer cell growth. Consistent with its tumor-suppressive role, *MEG3* is selectively downregulated in various brain cancers as well as human cancer cell lines. Reduced expression of *MEG3* is associated with hypermethylation of the *MEG3* gene regulatory regions in clinically nonfunctioning pituitary tumors (ie, benign neoplasms that do not secrete active hormones) (319). Although *MEG3* functions upstream of p53, *linc-p21* and *Pint* (p53-induced noncoding transcript) are direct transcriptional targets of p53. *Linc-p21* is required for p53-depdendent induction of apoptosis in cancer cell lines through transcriptional and translational regulation of p53 target genes (95, 137). *Pint*, whose expression is regulated by p53, promotes cell proliferation and survival by regulating the expression of genes in the TGFβ, MAPK, and p53 pathways (320). Similarly, *GAS5* sensitizes the cells to apoptosis to prevent tumorigenesis, doing so by negatively regulating the activity of GR, hence repressing the expression of apoptosis-inhibiting genes

(206). Thus, a number of lncRNAs act to repress tumorigenesis by promoting apoptosis.

A number of lncRNAs are involved in pathways that control the levels of tumor suppressor mRNA and protein. For example, *PTENP1* is a lncRNA produced from a pseudogene of *PTEN*, a classical tumor suppressor gene, and it sequesters *PTEN* mRNA-targeting microRNAs through a sponging mechanism to maintain *PTEN* levels (321). *PTENP1* also acts as a tumor suppressor in its own right. Interestingly, a pair of lncRNA isoforms transcribed antisense to *PTENP1* regulates the expression and stability of *PTENP1* RNA, thus controlling its microRNA sponge activity (322). Although *PTENP1* positively regulates a tumor suppressor through a microRNA intermediate, a recent study has shown that the lncRNA *loc285194* does so by directly antagonizing a growth-promoting microRNA, namely *miR-211* (323). *Loc285194* is upregulated by p53 and is a downstream target of *miR-211*. Consequently, the *loc285194* forms a reciprocal repression loop with *miR-211* upon p53 induction, acting to suppress tumor cell growth in vitro and in vivo (323). In contrast to *PTENP1* and *loc285194*, the lncRNA *TARID* (TCF21 antisense RNA inducing demethylation) activates the expression of the tumor suppressor gene *TCF21* by inducing promoter demethylation through interactions with the *TCF21* gene promoter and GADD45A, a regulator of DNA demethylation (324).

C. LncRNAs and metastasis

Although dysregulated cell growth and cell death responses often promote the initiation of oncogenesis, the tissue invasion and metastasis promote cancer progression and ultimately cancer-related death. Not surprisingly, several lncRNAs have been implicated in metastasis, such as the previously described *MALAT1* and *HOTAIR* (180, 181). Interestingly, *HOTAIR* is targeted and regulated by a microRNA, *miR-141*, in cancer cells, which antagonizes the positive effects of *HOTAIR* in proliferation and invasion (325). Recently, the lncRNA *SChLAP1* (second chromosome locus associated with prostate-1) has been shown to play critical roles in prostate cancer cell invasiveness in vitro and metastasis in vivo. *SChLAP1* exhibits elevated expression in a subset of prostate cancers that are associated with a poor prognosis (326). By attenuating the genomic localization of the tumor-suppressing SWItch/sucrose nonfermentable chromatin remodeling complex, *SChLAP1* promotes metastasis and leads to more aggressive forms of prostate cancer (326).

A recent study by Yuan et al (327) has connected *lncRNA-ATB* (lncRNA-activated by TGFβ), a lncRNA upregulated in hepatocellular carcinoma metastases and associated with poor prognosis, to the TGFβ signaling pathway. TGFβ is a key mediator of the epithelial-mesenchymal transition in metastasis. *LncRNA-ATB* induces epithelial-mesenchymal transition and invasion by upregulating the zinc finger E-box-binding homeobox 1 and zinc finger E-box binding homeobox 2 transcriptional regulators through a mechanism involving competitive binding of miR-200 family microRNAs. As a mediator of the TGFβ signaling pathway signaling, *lncRNA-ATB* expression may predispose hepatocellular carcinoma patients to metastases (327).

IX. The Therapeutic Potential of LncRNAs

The broad array of biological functions of lncRNAs as a class, in conjunction with their restricted cell-type-specific expression, make lncRNAs attractive as therapeutic targets. Recent studies have begun to explore the therapeutic potential of lncRNAs and other ncRNAs, many of which have focused on the diagnosis and treatment of cancers (328, 329).

As described in *Section VIII*, the study of lncRNAs has broadened our perspectives on fundamental aspects of cancer biology. More importantly, it also offers possibilities in medicine. Given their tissue-specific expression patterns (13, 20, 76), lncRNAs can serve as excellent biomarkers for certain types of cancer (35, 181, 330, 331). In the case of prostate cancer, a test based on the expression of the lncRNA *PCA3* has been developed and is being used clinically, capitalizing on the observation that the *PCA3* is overexpressed specifically in prostate cancer cells (332).

In addition to lncRNA-based diagnostics, lncRNAs also offer the possibility of lncRNA-based therapies. A growing number of lncRNAs have been shown to function as important oncogenes (eg, *SRA*, *PANDA*, *ANRIL*, and *p21NAT*) and tumor suppressors (eg, *MEG3*, *Pint*, *GAS5*, and *PTENP1*), hence providing us with new opportunities to approach cancer therapeutics. Molecular strategies that antagonize the levels and activities of oncogenic lncRNAs, including the administration of short interfering RNAs and antisense RNAs, or those that increase the levels and activities of the tumor-suppressive lncRNAs, have the potential to function as anticancer drugs. Similarly, methods that target lncRNAs involved in metastasis may also prove useful. For example, in preclinical studies, antisense oligos that attenuate the expression of *MALAT1* in EBC-1 lung cancer cells inhibits metastasis to the lung (180).

MRUL (multidrug-resistant [MDR]-related and upregulated lncRNA), a lncRNA that is upregulated in MDR gastric cancer cell sublines, is a good example of the therapeutic potential of lncRNAs. *MRUL* expression in gastric cancers is associated with a poor prognosis for gastric

cancer patients due to likely effects on multidrug resistance, the most common cause of chemotherapy failure in gastric cancer (333). *MRUL* acts to maintain the expression of *ABCB1*, a gene encoding a membrane-associated ATP-binding cassette transporter that transports small molecules across cellular membranes and is involved in multidrug resistance (333). Presumably, therapeutic approaches that target *MRUL* might help to reduce multidrug resistance.

Of course, given the plethora of functions served by lncRNAs in physiology, as highlighted in this review, it is likely that the therapeutic potential of lncRNAs extends well beyond cancer. For example, manipulating the levels of *HI-LNC25* and *116HG* may be useful in treating metabolic disorders. Molecular interventions that modulate the levels of *lincRNA-Cox2*, *Lethe*, and *THRIL* may be useful in modulating effects of inflammation. In addition, approaches that target the various antisense lncRNAs associated with neurological disorders and cardiac pathophysiology may be useful in treating the diseases of these systems. Collectively, the diagnostic and therapeutic potential of lncRNAs in various diseases, although still in its infancy, could be tremendous and warrants further investigation.

X. Summary, Conclusions, and Future Directions

In this review, we have discussed the evolving understanding of lncRNAs, a new class of noncoding regulatory RNAs, focusing on their discovery, annotation, physical properties, and molecular mechanisms of action. In addition, we highlight the biological functions of some of the best characterized lncRNA in physiology and disease, especially those relevant to endocrinology, reproduction, metabolism, immunology, neurobiology, muscle biology, and cancer.

A. Summary and Conclusions

The introduction of whole transcriptome sequencing methods and the advent of large-scale transcript mapping projects have transformed our perspectives on the variety and dynamic nature of lncRNAs. Studies characterizing the functions of an expanding set of lncRNAs have shown that they play central roles in various aspects of physiology. This relatively poorly characterized class of RNAs with little or no coding capacity has been implicated in endocrinology, reproduction, metabolism, immunology, neurobiology, muscle biology, and cancer. They function through molecular and biochemical mechanisms that range from *cis*- to *trans*-regulation of gene expression, and from epigenetic modulation in the nucleus to posttranscriptional control in the cytoplasm. Given their widespread functions throughout the body, it is not surprising that many lncRNAs have been associated with diseases of the tissues in which they are expressed. Efforts to tap the tremendous diagnostic and therapeutic potential of lncRNAs have begun to meet with some success, although this is an area that needs to be explored in considerably more detail.

B. Future directions

With the high level of interest in the field, our understanding of lncRNAs is constantly shifting and rapidly expanding (336). Yet, as we move forward from efforts to annotate lncRNAs to a greater focus on molecular function and biology, many questions and challenges await. We still do not fully understand the biological significance of lncRNAs as a group. Is the majority functional, or is the act of transcription at these genomic loci the relevant endpoint? To answer these questions, we need better tools to 1) determine the structure and elucidate the key structure-function relationships of lncRNAs, especially how they interact with their protein partners; 2) track lncRNA localization throughout cellular compartments and across the genome; 3) monitor the interactions of lncRNAs with proteins and nucleic acids; 4) detect and analyze the functions of posttranscriptional modifications of lncRNAs; and 5) perturb the cellular levels of lncRNAs in a fast and efficient manner. In addition, we need more effective high-throughput approaches for screening the physiological functions of lncRNAs in cells and animal models.

Moreover, we need better biological models to understand the key elements of evolutionary conservation (eg, sequence vs structure) as well as the different functions of lncRNAs in the nucleus, cytoplasm, and elsewhere in the cell. Finally, we need more effective ways to discern the diagnostic and therapeutic potential of lncRNAs. These are just a few of the many challenges faced by the field. Given the rapid and extensive progress that has been made over the past decade, there is every reason to be optimistic about the field's ability to address these questions and challenges and ultimately produce a greater understanding of the biology of lncRNAs.

Acknowledgments

We thank members of the Kraus lab for their critical comments and suggestions on this manuscript. Although we strove to make this review as comprehensive as possible, the vast and rapidly growing literature in this area is certain to result in the inevitable omission. We apologize to those whose work could not be included due to space and focus, and for any unintentional omissions.

Address requests for reprints to: W. Lee Kraus, PhD, Cecil H. and Ida Green Center for Reproductive Biology Sciences, The University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, TX 75390–8511. E-mail: LEE.KRAUS@utsouthwestern.edu.

The lncRNA research in the Kraus lab is supported by grants from the Cancer Prevention and Research Institute of Texas and the National Institutes of Health National Institute of Diabetes and Digestive and Kidney Diseases.

Current address for M.S.: Computational and Systems Biology, Genome Institute of Singapore, 60 Biopolis Street, Singapore 138672.

Disclosure Summary: The authors have nothing to disclose.

References

1. **Okazaki Y, Furuno M, Kasukawa T, et al.** Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature*. 2002;420(6915):563–573.
2. **ENCODE Project Consortium; Birney E, Stamatoyannopoulos JA, Dutta A, et al.** Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*. 2007;447(7146):799–816.
3. **Djebali S, Davis CA, Merkel A, et al.** Landscape of transcription in human cells. *Nature*. 2012;489(7414):101–108.
4. **Mercer TR, Dinger ME, Mattick JS.** Long non-coding RNAs: insights into functions. *Nat Rev Genet*. 2009;10(3):155–159.
5. **Moran VA, Perera RJ, Khalil AM.** Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs. *Nucleic Acids Res*. 2012;40(14):6391–6400.
6. **Ørom UA, Derrien T, Beringer M, et al.** Long noncoding RNAs with enhancer-like function in human cells. *Cell*. 2010;143(1):46–58.
7. **Wilusz JE, Sunwoo H, Spector DL.** Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev*. 2009;23(13):1494–1504.
8. **Rinn JL, Chang HY.** Genome regulation by long noncoding RNAs. *Annu Rev Biochem*. 2012;81:145–166.
9. **Ulitsky I, Bartel DP.** lincRNAs: genomics, evolution, and mechanisms. *Cell*. 2013;154(1):26–46.
10. **Batista PJ, Chang HY.** Long noncoding RNAs: cellular address codes in development and disease. *Cell*. 2013;152(6):1298–1307.
11. **Abdelmohsen K, Panda A, Kang MJ, et al.** SAL-RNAs: senescence-associated long non-coding RNAs. *Aging Cell*. 2013.
12. **Sun L, Goff LA, Trapnell C, et al.** Long noncoding RNAs regulate adipogenesis. *Proc Natl Acad Sci U S A*. 2013;110(9):3387–3392.
13. **Derrien T, Johnson R, Bussotti G, et al.** The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res*. 2012;22(9):1775–1789.
14. **Kapranov P, Cheng J, Dike S, et al.** RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science*. 2007;316(5830):1484–1488.
15. **Ma L, Bajic VB, Zhang Z.** On the classification of long non-coding RNAs. *RNA Biol*. 2013;10(6);925–933.
16. **Novikova IV, Hennelly SP, Sanbonmatsu KY.** Sizing up long non-coding RNAs: Do lncRNAs have secondary and tertiary structure? *Bioarchitecture*. 2012;2(6):189–199.
17. **Prensner JR, Chinnaiyan AM.** The emergence of lncRNAs in cancer biology. *Cancer Discov*. 2011;1(5):391–407.
18. **Guttman M, Amit I, Garber M, et al.** Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*. 2009;458(7235):223–227.
19. **Guttman M, Garber M, Levin JZ, et al.** Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat Biotechnol*. 2010;28(5):503–510.
20. **Cabili MN, Trapnell C, Goff L, et al.** Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev*. 2011;25(18):1915–1927.
21. **Khalil AM, Guttman M, Huarte M, et al.** Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci U S A*. 2009;106(28):11667–11672.
22. **Hangauer MJ, Vaughn IW, McManus MT.** Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. *PLoS Genet*. 2013;9(6):e1003569.
23. **Ingolia NT, Lareau LF, Weissman JS.** Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*. 2011;147(4):789–802.
24. **Guttman M, Russell P, Ingolia NT, Weissman JS, Lander ES.** Ribosome Profiling Provides Evidence that Large Noncoding RNAs Do Not Encode Proteins. *Cell*. 2013;154(1):240–251.
25. **Lanz RB, McKenna NJ, Onate SA, et al.** A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. *Cell*. 1999;97(1):17–27.
26. **Chooniedass-Kothari S, Emberley E, Hamedani MK, et al.** The steroid receptor RNA activator is the first functional RNA encoding a protein. *FEBS Lett*. 2004;566(1–3):43–47.
27. **Brar GA, Yassour M, Friedman N, Regev A, Ingolia NT, Weissman JS.** High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science*. 2012;335(6068):552–557.
28. **Carvunis AR, Rolland T, Wapinski I, et al.** Proto-genes and de novo gene birth. *Nature*. 2012;487(7407):370–374.
29. **Chng SC, Ho L, Tian J, Reversade B.** ELABELA: a hormone essential for heart development signals via the apelin receptor. *Dev Cell*. 2013;27(6):672–680.
30. **Kondo T, Plaza S, Zanet J, et al.** Small peptides switch the transcriptional activity of Shavenbaby during Drosophila embryogenesis. *Science*. 2010;329(5989):336–339.
31. **Reinhardt JA, Wanjiru BM, Brant AT, Saelao P, Begun DJ, Jones CD.** De novo ORFs in Drosophila are important to organismal fitness and evolved rapidly from previously non-coding sequences. *PLoS Genet*. 2013;9(10):e1003860.
32. **Pauli A, Norris ML, Valen E, et al.** Toddler: an embryonic signal that promotes cell movement via Apelin receptors. *Science*. 2014;343(6172):1248636.
33. **Bazzini AA, Johnstone TG, Christiano R, et al.** Identification of small ORFs in vertebrates using ribosome footprint-

ing and evolutionary conservation. *EMBO J*. 2014;33(9):981–993.
34. **Magny EG, Pueyo JI, Pearl FM, et al.** Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames. *Science*. 2013;341(6150):1116–1120.
35. **Gibb EA, Brown CJ, Lam WL.** The functional role of long non-coding RNA in human carcinomas. *Mol Cancer*. 2011;10:38.
36. **Sigova AA, Mullen AC, Molinie B, et al.** Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proc Natl Acad Sci U S A*. 2013;110(8):2876–2881.
37. **Zheng GX, Do BT, Webster DE, Khavari PA, Chang HY.** Dicer-microRNA-Myc circuit promotes transcription of hundreds of long noncoding RNAs. *Nat Struct Mol Biol*. 2014;21(7):585–590.
38. **Mus E, Hof PR, Tiedge H.** Dendritic BC200 RNA in aging and in Alzheimer's disease. *Proc Natl Acad Sci U S A*. 2007;104(25):10679–10684.
39. **Hawkins PG, Morris KV.** Transcriptional regulation of Oct4 by a long non-coding RNA antisense to Oct4-pseudogene 5. *Transcription*. 2010;1(3):165–175.
40. **Hansen TB, Jensen TI, Clausen BH, et al.** Natural RNA circles function as efficient microRNA sponges. *Nature*. 2013;495(7441):384–388.
41. **Jeck WR, Sorrentino JA, Wang K, et al.** Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA*. 2013;19(2):141–157.
42. **Memczak S, Jens M, Elefsinioti A, et al.** Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*. 2013;495(7441):333–338.
43. **Liang D, Wilusz JE.** Short intronic repeat sequences facilitate circular RNA production. *Genes Dev*. 2014;28(20):2233–2247.
44. **Zhang Y, Zhang XO, Chen T, et al.** Circular intronic long noncoding RNAs. *Mol Cell*. 2013;51(6):792–806.
45. **Yin QF, Yang L, Zhang Y, et al.** Long noncoding RNAs with snoRNA ends. *Mol Cell*. 2012;48(2):219–230.
46. **Jia H, Osak M, Bogu GK, Stanton LW, Johnson R, Lipovich L.** Genome-wide computational identification and manual annotation of human long noncoding RNA genes. *RNA*. 2010;16(8):1478–1487.
47. **Zhang Y, Liu XS, Liu QR, Wei L.** Genome-wide in silico identification and analysis of cis natural antisense transcripts (cis-NATs) in ten species. *Nucleic Acids Res*. 2006;34(12):3465–3475.
48. **Li JT, Zhang Y, Kong L, Liu QR, Wei L.** Trans-natural antisense transcripts including noncoding RNAs in 10 species: implications for expression regulation. *Nucleic Acids Res*. 2008;36(15):4833–4844.
49. **Takayama K, Horie-Inoue K, Katayama S, et al.** Androgen-responsive long noncoding RNA CTBP1-AS promotes prostate cancer. *EMBO J*. 2013;32(12):1665–1680.
50. **Kraus P, Sivakamasundari V, Lim SL, Xing X, Lipovich L, Lufkin T.** Making sense of Dlx1 antisense RNA. *Dev Biol*. 2013;376(2):224–235.
51. **Munroe SH, Lazar MA.** Inhibition of c-erbA mRNA splicing by a naturally occurring antisense RNA. *J Biol Chem*. 1991;266(33):22083–22086.
52. **Beltran M, Puig I, Peña C, et al.** A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. *Genes Dev*. 2008;22(6):756–769.
53. **Carrieri C, Cimatti L, Biagioli M, et al.** Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature*. 2012;491(7424):454–457.
54. **Washietl S, Kellis M, Garber M.** Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. *Genome Res*. 2014;24(4):616–628.
55. **Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP.** Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell*. 2011;147(7):1537–1550.
56. **Wu X, Sharp PA.** Divergent transcription: a driving force for new gene origination? *Cell*. 2013;155(5):990–996.
57. **Carninci P, Kasukawa T, Katayama S, et al.** The transcriptional landscape of the mammalian genome. *Science*. 2005;309(5740):1559–1563.
58. **Finn RD, Mistry J, Schuster-Bockler B, et al.** Pfam: clans, web tools and services. *Nucleic Acids Res*. 2006;34(Database issue):D247–D251.
59. **Wilson D, Pethica R, Zhou Y, et al.** SUPERFAMILY–sophisticated comparative genomics, data mining, visualization and phylogeny. *Nucleic Acids Res*. 2009;37(Database issue):D380–D386.
60. **Wilson D, Madera M, Vogel C, Chothia C, Gough J.** The SUPERFAMILY database in 2007: families and functions. *Nucleic Acids Res*. 2007;35(Database issue):D308–D313.
61. **Krogh A, Larsson B, von Heijne G, Sonnhammer EL.** Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol*. 2001;305(3):567–580.
62. **Nielsen H, Engelbrecht J, Brunak S, von Heijne G.** A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Int J Neural Syst*. 1997;8(5–6):581–599.
63. **Clamp M, Fry B, Kamal M, et al.** Distinguishing protein-coding and noncoding genes in the human genome. *Proc Natl Acad Sci U S A*. 2007;104(49):19428–19433.
64. **Marques AC, Hughes J, Graham B, Kowalczyk MS, Higgs DR, Ponting CP.** Chromatin signatures at transcriptional start sites separate two equally populated yet distinct classes of intergenic long noncoding RNAs. *Genome Biol*. 2013;14(11):R131.
65. **Core LJ, Waterfall JJ, Lis JT.** Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*. 2008;322(5909):1845–1848.
66. **Hah N, Danko CG, Core L, et al.** A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. *Cell*. 2011;145(4):622–634.
67. **Churchman LS, Weissman JS.** Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature*. 2011;469(7330):368–373.
68. **Frohman MA, Dush MK, Martin GR.** Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci U S A*. 1988;85(23):8998–9002.
69. **Shiraki T, Kondo S, Katayama S, et al.** Cap analysis gene expression for high-throughput analysis of transcriptional

starting point and identification of promoter usage. *Proc Natl Acad Sci U S A*. 2003;100(26):15776–15781.
70. **Tsuchihara K, Suzuki Y, Wakaguri H, et al.** Massive transcriptional start site analysis of human genes in hypoxia cells. *Nucleic Acids Res*. 2009;37(7):2249–2263.
71. **Wakaguri H, Yamashita R, Suzuki Y, Sugano S, Nakai K.** DBTSS: database of transcription start sites, progress report 2008. *Nucleic Acids Res*. 2008;36(Database issue): D97–D101.
72. **Jan CH, Friedman RC, Ruby JG, Bartel DP.** Formation, regulation and evolution of Caenorhabditis elegans 3′UTRs. *Nature*. 2011;469(7328):97–101.
73. **Nam JW, Bartel DP.** Long noncoding RNAs in C. elegans. *Genome Res*. 2012;22(12):2529–2540.
74. **Trapnell C, Williams BA, Pertea G, et al.** Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010;28(5):511–515.
75. **Necsulea A, Soumillon M, Warnefors M, et al.** The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature*. 2014;505(7485):635–640.
76. **Pauli A, Valen E, Lin MF, et al.** Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res*. 2012;22(3):577–591.
77. **Kong L, Zhang Y, Ye ZQ, et al.** CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res*. 2007;35(Web Server issue):W345–W349.
78. **Castrignano T, Canali A, Grillo G, Liuni S, Mignone F, Pesole G.** CSTminer: a web tool for the identification of coding and noncoding conserved sequence tags through cross-species genome comparison. *Nucleic Acids Res*. 2004;32(Web Server issue):W624–W627.
79. **Rivas E, Eddy SR.** Noncoding RNA gene detection using comparative sequence analysis. *BMC Bioinformatics*. 2001;2:8.
80. **Badger JH, Olsen GJ.** CRITICA: coding region identification tool invoking comparative analysis. *Mol Biol Evol*. 1999;16(4):512–524.
81. **Liu J, Gough J, Rost B.** Distinguishing protein-coding from non-coding RNAs through support vector machines. *PLoS Genet*. 2006;2(4):e29.
82. **Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS.** Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*. 2009; 324(5924):218–223.
83. **Li YM, Franklin G, Cui HM, et al.** The H19 transcript is associated with polysomes and may regulate IGF2 expression in trans. *J Biol Chem*. 1998;273(43):28247–28252.
84. **Szymanski M, Erdmann VA, Barciszewski J.** Noncoding regulatory RNAs database. *Nucleic Acids Res*. 2003;31(1): 429–431.
85. **Mituyama T, Yamada K, Hattori E, et al.** The Functional RNA Database 3.0: databases to support mining and annotation of functional RNAs. *Nucleic Acids Res*. 2009; 37(Database issue):D89–D92.
86. **Bu D, Yu K, Sun S, et al.** NONCODE v3.0: integrative annotation of long noncoding RNAs. *Nucleic Acids Res*. 2012;40(Database issue):D210–D215.
87. **Dinger ME, Pang KC, Mercer TR, Crowe ML, Grimmond SM, Mattick JS.** NRED: a database of long noncoding RNA expression. *Nucleic Acids Res*. 2009;37(Database issue):D122–D126.
88. **Amaral PP, Clark MB, Gascoigne DK, Dinger ME, Mattick JS.** lncRNAdb: a reference database for long noncoding RNAs. *Nucleic Acids Res*. 2011;39(Database issue): D146–D151.
89. **Volders PJ, Helsens K, Wang X, et al.** LNCipedia: a database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Res*. 2013;41(Database issue):D246–D251.
90. **Yang JH, Li JH, Jiang S, Zhou H, Qu LH.** ChIPBase: a database for decoding the transcriptional regulation of long non-coding RNA and microRNA genes from ChIP-Seq data. *Nucleic Acids Res*. 2013;41(Database issue): D177–D187.
91. **Klattenhoff CA, Scheuermann JC, Surface LE, et al.** Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell*. 2013;152(3):570–583.
92. **Kretz M, Siprashvili Z, Chu C, et al.** Control of somatic tissue differentiation by the long non-coding RNA TINCR. *Nature*. 2013;493(7431):231–235.
93. **Kretz M, Webster DE, Flockhart RJ, et al.** Suppression of progenitor differentiation requires the long noncoding RNA ANCR. *Genes Dev*. 2012;26(4):338–343.
94. **Herriges MJ, Swarr DT, Morley MP, et al.** Long noncoding RNAs are spatially correlated with transcription factors and regulate lung development. *Genes Dev*. 2014;28(12): 1363–1379.
95. **Huarte M, Guttman M, Feldser D, et al.** A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell*. 2010;142(3):409–419.
96. **Liao Q, Liu C, Yuan X, et al.** Large-scale prediction of long non-coding RNA functions in a coding-non-coding gene co-expression network. *Nucleic Acids Res*. 2011;39(9): 3864–3878.
97. **Liao Q, Xiao H, Bu D, et al.** ncFANs: a web server for functional annotation of long non-coding RNAs. *Nucleic Acids Res*. 2011;39(Web Server issue):W118–W124.
98. **Kotake Y, Nakagawa T, Kitagawa K, et al.** Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. *Oncogene*. 2011;30(16):1956–1962.
99. **Maamar H, Cabili MN, Rinn J, Raj A.** linc-HOXA1 is a noncoding RNA that represses Hoxa1 transcription in cis. *Genes Dev*. 2013;27(11):1260–1271.
100. **Mondal T, Rasmussen M, Pandey GK, Isaksson A, Kanduri C.** Characterization of the RNA content of chromatin. *Genome Res*. 2010;20(7):899–907.
101. **Wang KC, Yang YW, Liu B, et al.** A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature*. 2011;472(7341):120–124.
102. **Lai F, Orom UA, Cesaroni M, et al.** Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature*. 2013;494(7438):497–501.
103. **Kim T, Cui R, Jeon YJ, et al.** Long-range interaction and correlation between MYC enhancer and oncogenic long noncoding RNA CARLo-5. *Proc Natl Acad Sci U S A*. 2014;111(11):4173–4178.
104. **Hung T, Wang Y, Lin MF, et al.** Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat Genet*. 2011;43(7):621–629.

105. **Guttman M, Donaghey J, Carey BW, et al.** lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature*. 2011;477(7364):295–300.
106. **Rinn JL, Kertesz M, Wang JK, et al.** Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*. 2007;129(7):1311–1323.
107. **Zhao J, Ohsumi TK, Kung JT, et al.** Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol Cell*. 2010;40(6):939–953.
108. **Penny GD, Kay GF, Sheardown SA, Rastan S, Brockdorff N.** Requirement for Xist in X chromosome inactivation. *Nature*. 1996;379(6561):131–137.
109. **Marahrens Y, Panning B, Dausman J, Strauss W, Jaenisch R.** Xist-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes Dev*. 1997;11(2):156–166.
110. **Tsai MC, Manor O, Wan Y, et al.** Long noncoding RNA as modular scaffold of histone modification complexes. *Science*. 2010;329(5992):689–693.
111. **Brown JA, Bulkley D, Wang J, et al.** Structural insights into the stabilization of MALAT1 noncoding RNA by a bipartite triple helix. *Nat Struct Mol Biol*. 2014;21(7):633–640.
112. **Kaneko S, Li G, Son J, et al.** Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA. *Genes Dev*. 2010;24(23):2615–2620.
113. **Cifuentes-Rojas C, Hernandez AJ, Sarma K, Lee JT.** Regulatory interactions between RNA and polycomb repressive complex 2. *Mol Cell*. 2014;55(2):171–185.
114. **Li G, Margueron R, Ku M, Chambon P, Bernstein BE, Reinberg D.** Jarid2 and PRC2, partners in regulating gene expression. *Genes Dev*. 2010;24(4):368–380.
115. **Son J, Shen SS, Margueron R, Reinberg D.** Nucleosome-binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin. *Genes Dev*. 2013;27(24):2663–2677.
116. **da Rocha ST, Boeva V, Escamilla-Del-Arenal M, et al.** Jarid2 is implicated in the initial Xist-induced targeting of PRC2 to the inactive X chromosome. *Mol Cell*. 2014;53(2):301–316.
117. **Kaneko S, Bonasio R, Saldaña-Meyer R, et al.** Interactions between JARID2 and noncoding RNAs regulate PRC2 recruitment to chromatin. *Mol Cell*. 2014;53(2):290–300.
118. **Yang YW, Flynn RA, Chen Y, et al.** Essential role of lncRNA binding for WDR5 maintenance of active chromatin and embryonic stem cell pluripotency. *Elife*. 2014;3:e02046.
119. **Grote P, Wittler L, Hendrix D, et al.** The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Dev Cell*. 2013;24(2):206–214.
120. **Bierhoff H, Dammert MA, Brocks D, Dambacher S, Schotta G, Grummt I.** Quiescence-induced LncRNAs trigger H4K20 trimethylation and transcriptional silencing. *Mol Cell*. 2014;54(4):675–682.
121. **Di Ruscio A, Ebralidze AK, Benoukraf T, et al.** DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature*. 2013;503(7476):371–376.
122. **Monnier P, Martinet C, Pontis J, Stancheva I, Ait-Si-Ali S, Dandolo L.** H19 lncRNA controls gene expression of the Imprinted Gene Network by recruiting MBD1. *Proc Natl Acad Sci U S A*. 2013;110(51):20693–20698.
123. **Saldana-Meyer R, Gonzalez-Buendía E, Guerrero G, et al.** CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. *Genes Dev*. 2014;28(7):723–734.
124. **Yao H, Brick K, Evrard Y, Xiao T, Camerini-Otero RD, Felsenfeld G.** Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. *Genes Dev*. 2010;24(22):2543–2555.
125. **Li Z, Chao TC, Chang KY, et al.** The long noncoding RNA THRIL regulates TNFα expression through its interaction with hnRNPL. *Proc Natl Acad Sci U S A*. 2014;111(3):1002–1007.
126. **Mariner PD, Walters RD, Espinoza CA, et al.** Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. *Mol Cell*. 2008;29(4):499–509.
127. **Lu Q, Ren S, Lu M, et al.** Computational prediction of associations between long non-coding RNAs and proteins. *BMC Genomics*. 2013;14:651.
128. **Li JH, Liu S, Zhou H, Qu LH, Yang JH.** starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res*. 2014;42(Database issue):D92–D97.
129. **Rouskin S, Zubradt M, Washietl S, Kellis M, Weissman JS.** Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature*. 2014;505(7485):701–705.
130. **Wan Y, Qu K, Zhang QC, et al.** Landscape and variation of RNA secondary structure across the human transcriptome. *Nature*. 2014;505(7485):706–709.
131. **Chu C, Qu K, Zhong FL, Artandi SE, Chang HY.** Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol Cell*. 2011;44(4):667–678.
132. **Simon MD, Wang CI, Kharchenko PV, et al.** The genomic binding sites of a noncoding RNA. *Proc Natl Acad Sci U S A*. 2011;108(51):20497–20502.
133. **Engreitz JM, Pandya-Jones A, McDonel P, et al.** The Xist lncRNA Exploits Three-Dimensional Genome Architecture to Spread Across the X Chromosome. *Science*. 2013;341(6147):1237973.
134. **Simon MD, Pinter SF, Fang R, et al.** High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. *Nature*. 2013;504(7480):465–469.
135. **Vance KW, Sansom SN, Lee S, et al.** The long non-coding RNA Paupar regulates the expression of both local and distal genes. *The EMBO journal*. 2014;33(4):296–311.
136. **Gong C, Maquat LE.** lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3′ UTRs via Alu elements. *Nature*. 2011;470(7333):284–288.
137. **Yoon JH, Abdelmohsen K, Srikantan S, et al.** LincRNA-p21 suppresses target mRNA translation. *Mol Cell*. 2012;47(4):648–655.
138. **Dimitrova N, Zamudio JR, Jong RM, et al.** LincRNA-p21 activates p21 in cis to promote Polycomb target gene expression and to enforce the G1/S checkpoint. *Mol Cell*. 2014;54(5):777–790.

139. **Cesana M, Cacchiarelli D, Legnini I, et al.** A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell*. 2011;147(2):358–369.
140. **Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP.** A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell*. 2011;146(3):353–358.
141. **Tay Y, Kats L, Salmena L, et al.** Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. *Cell*. 2011;147(2):344–357.
142. **Karreth FA, Tay Y, Perna D, et al.** In vivo identification of tumor- suppressive PTEN ceRNAs in an oncogenic BRAF-induced mouse model of melanoma. *Cell*. 2011;147(2):382–395.
143. **Kallen AN, Zhou XB, Xu J, et al.** The imprinted H19 lncRNA antagonizes let-7 microRNAs. *Mol Cell*. 2013;52(1):101–112.
144. **Liz J, Portela A, Soler M, et al.** Regulation of pri-miRNA processing by a long noncoding RNA transcribed from an ultraconserved region. *Mol Cell*. 2014;55(1):138–147.
145. **Clemson CM, Chow JC, Brown CJ, Lawrence JB.** Stabilization and localization of Xist RNA are controlled by separate mechanisms and are not sufficient for X inactivation. *J Cell Biol*. 1998;142(1):13–23.
146. **Duthie SM, Nesterova TB, Formstone EJ, et al.** Xist RNA exhibits a banded localization on the inactive X chromosome and is excluded from autosomal material in cis. *Hum Mol Genet*. 1999;8(2):195–204.
147. **Murakami K, Ohhira T, Oshiro E, Qi D, Oshimura M, Kugoh H.** Identification of the chromatin regions coated by non-coding Xist RNA. *Cytogenet Genome Res*. 2009;125(1):19–25.
148. **Pandey RR, Mondal T, Mohammad F, et al.** Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell*. 2008;32(2):232–246.
149. **Nagano T, Mitchell JA, Sanz LA, et al.** The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science*. 2008;322(5908):1717–1720.
150. **Murakami K, Oshimura M, Kugoh H.** Suggestive evidence for chromosomal localization of non-coding RNA from imprinted LIT1. *J Hum Genet*. 2007;52(11):926–933.
151. **Redrup L, Branco MR, Perdeaux ER, et al.** The long noncoding RNA Kcnq1ot1 organises a lineage-specific nuclear domain for epigenetic gene silencing. *Development*. 2009;136(4):525–530.
152. **Borsani G, Tonlorenzi R, Simmler MC, et al.** Characterization of a murine gene expressed from the inactive X chromosome. *Nature*. 1991;351(6324):325–329.
153. **Brockdorff N, Ashworth A, Kay GF, et al.** Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. *Nature*. 1991;351(6324):329–331.
154. **Brockdorff N, Ashworth A, Kay GF, et al.** The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell*. 1992;71(3):515–526.
155. **Brown CJ, Hendrich BD, Rupert JL, et al.** The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell*. 1992;71(3):527–542.
156. **Wutz A, Jaenisch R.** A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Mol Cell*. 2000;5(4):695–705.
157. **Wutz A, Rasmussen TP, Jaenisch R.** Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat Genet*. 2002;30(2):167–174.
158. **Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT.** Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science*. 2008;322(5902):750–756.
159. **Jeon Y, Lee JT.** YY1 tethers Xist RNA to the inactive X nucleation center. *Cell*. 2011;146(1):119–133.
160. **Lee JT, Lu N.** Targeted mutagenesis of Tsix leads to nonrandom X inactivation. *Cell*. 1999;99(1):47–57.
161. **Lee JT, Davidow LS, Warshawsky D.** Tsix, a gene antisense to Xist at the X-inactivation centre. *Nat Genet*. 1999;21(4):400–404.
162. **Sado T, Hoki Y, Sasaki H.** Tsix silences Xist through modification of chromatin structure. *Dev Cell*. 2005;9(1):159–165.
163. **Ogawa Y, Sun BK, Lee JT.** Intersection of the RNA interference and X-inactivation pathways. *Science*. 2008;320(5881):1336–1341.
164. **Tian D, Sun S, Lee JT.** The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. *Cell*. 2010;143(3):390–403.
165. **Sun S, Del Rosario BC, Szanto A, Ogawa Y, Jeon Y, Lee JT.** Jpx RNA activates Xist by evicting CTCF. *Cell*. 2013;153(7):1537–1551.
166. **Pinter SF, Sadreyev RI, Yildirim E, et al.** Spreading of X chromosome inactivation via a hierarchy of defined Polycomb stations. *Genome Res*. 2012;22(10):1864–1876.
167. **Ji P, Diederichs S, Wang W, et al.** MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene*. 2003;22(39):8031–8041.
168. **Hutchinson JN, Ensminger AW, Clemson CM, Lynch CR, Lawrence JB, Chess A.** A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics*. 2007;8:39.
169. **Bernard D, Prasanth KV, Tripathi V, et al.** A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. *EMBO J*. 2010;29(18):3082–3093.
170. **Yang L, Lin C, Liu W, et al.** ncRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs. *Cell*. 2011;147(4):773–788.
171. **Tripathi V, Ellis JD, Shen Z, et al.** The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell*. 2010;39(6):925–938.
172. **Tripathi V, Shen Z, Chakraborty A, et al.** Long noncoding RNA MALAT1 controls cell cycle progression by regulating the expression of oncogenic transcription factor B-MYB. *PLoS Genet*. 2013;9(3):e1003368.
173. **Zhang B, Arun G, Mao YS, et al.** The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. *Cell Rep*. 2012;2(1):111–123.
174. **Park CY, Jeker LT, Carver-Moore K, et al.** A resource for

the conditional ablation of microRNAs in the mouse. *Cell Rep*. 2012;1(4):385–391.
175. **Liu N, Bezprozvannaya S, Williams AH, et al.** microRNA-133a regulates cardiomyocyte proliferation and suppresses smooth muscle gene expression in the heart. *Genes Dev*. 2008;22(23):3242–3254.
176. **Williams AH, Valdez G, Moresi V, et al.** MicroRNA-206 delays ALS progression and promotes regeneration of neuromuscular synapses in mice. *Science*. 2009;326(5959):1549–1554.
177. **Jin ZB, Hirokawa G, Gui L, et al.** Targeted deletion of miR-182, an abundant retinal microRNA. *Mol Vis*. 2009;15:523–533.
178. **van Rooij E, Sutherland LB, Qi X, Richardson JA, Hill J, Olson EN.** Control of stress-dependent cardiac growth and gene expression by a microRNA. *Science*. 2007;316(5824):575–579.
179. **Callis TE, Pandya K, Seok HY, et al.** MicroRNA-208a is a regulator of cardiac hypertrophy and conduction in mice. *J Clin Invest*. 2009;119(9):2772–2786.
180. **Gutschner T, Hämmerle M, Eissmann M, et al.** The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res*. 2013;73(3):1180–1189.
181. **Gupta RA, Shah N, Wang KC, et al.** Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*. 2010;464(7291):1071–1076.
182. **Kogo R, Shimamura T, Mimori K, et al.** Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. *Cancer Res*. 2011;71(20):6320–6326.
183. **Nie Y, Liu X, Qu S, Song E, Zou H, Gong C.** Long noncoding RNA HOTAIR is an independent prognostic marker for nasopharyngeal carcinoma progression and survival. *Cancer Sci*. 2013;104(4):458–464.
184. **Geng YJ, Xie SL, Li Q, Ma J, Wang GY.** Large intervening non-coding RNA HOTAIR is associated with hepatocellular carcinoma progression. *J Int Med Res*. 2011;39(6):2119–2128.
185. **Li L, Liu B, Wapinski OL, et al.** Targeted disruption of Hotair leads to homeotic transformation and gene derepression. *Cell Rep*. 2013;5(1):3–12.
186. **Hulo N, Bairoch A, Bulliard V, et al.** The PROSITE database. *Nucleic Acids Res*. 2006;34(Database issue):D227–D230.
187. **Sauvageau M, Goff LA, Lodato S, et al.** Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *Elife*. 2013;2:e01749.
188. **Fatica A, Bozzoni I.** Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet*. 2014;15(1):7–21.
189. **Loewer S, Cabili MN, Guttman M, et al.** Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat Genet*. 2010;42(12):1113–1117.
190. **Lin N, Chang KY, Li Z, et al.** An evolutionarily conserved long noncoding RNA TUNA controls pluripotency and neural lineage commitment. *Mol Cell*. 2014;53(6):1005–1019.
191. **Evans RM.** The steroid and thyroid hormone receptor superfamily. *Science*. 1988;240(4854):889–895.
192. **Mangelsdorf DJ, Thummel C, Beato M, et al.** The nuclear receptor superfamily: the second decade. *Cell*. 1995;83(6):835–839.
193. **Watanabe M, Yanagisawa J, Kitagawa H, et al.** A subfamily of RNA-binding DEAD-box proteins acts as an estrogen receptor alpha coactivator through the N-terminal activation domain (AF-1) with an RNA coactivator, SRA. *EMBO J*. 2001;20(6):1341–1352.
194. **Zhao X, Patton JR, Ghosh SK, Fischel-Ghodsian N, Shen L, Spanjaard RA.** Pus3p- and Pus1p-dependent pseudouridylation of steroid receptor RNA activator controls a functional switch that regulates nuclear receptor signaling. *Mol Endocrinol*. 2007;21(3):686–699.
195. **Redfern AD, Colley SM, Beveridge DJ, et al.** RNA-induced silencing complex (RISC) Proteins PACT, TRBP, and Dicer are SRA binding nuclear receptor coregulators. *Proc Natl Acad Sci U S A*. 2013;110(16):6536–6541.
196. **Shi Y, Downes M, Xie W, et al.** Sharp, an inducible cofactor that integrates nuclear receptor repression and activation. *Genes Dev*. 2001;15(9):1140–1151.
197. **Hatchell EC, Colley SM, Beveridge DJ, et al.** SLIRP, a small SRA binding protein, is a nuclear receptor corepressor. *Mol Cell*. 2006;22(5):657–668.
198. **Vicent GP, Nacht AS, Zaurin R, et al.** Unliganded progesterone receptor-mediated targeting of an RNA-containing repressive complex silences a subset of hormone-inducible genes. *Genes Dev*. 2013;27(10):1179–1197.
199. **Zhao X, Patton JR, Davis SL, Florence B, Ames SJ, Spanjaard RA.** Regulation of nuclear receptor activity by a pseudouridine synthase through posttranscriptional modification of steroid receptor RNA activator. *Mol Cell*. 2004;15(4):549–558.
200. **Hubé F, Velasco G, Rollin J, Furling D, Francastel C.** Steroid receptor RNA activator protein binds to and counteracts SRA RNA-mediated activation of MyoD and muscle differentiation. *Nucleic Acids Res*. 2011;39(2):513–525.
201. **Leygue E.** Steroid receptor RNA activator (SRA1): unusual bifaceted gene products with suspected relevance to breast cancer. *Nucl Recept Signal*. 2007;5:e006.
202. **Chooniedass-Kothari S, Hamedani MK, Troup S, Hubé F, Leygue E.** The steroid receptor RNA activator protein is expressed in breast tumor tissues. *Int J Cancer*. 2006;118(4):1054–1059.
203. **Cooper C, Guo J, Yan Y, et al.** Increasing the relative expression of endogenous non-coding Steroid Receptor RNA Activator (SRA) in human breast cancer cells using modified oligonucleotides. *Nucleic Acids Res*. 2009;37(13):4518–4531.
204. **Kino T, Hurt DE, Ichijo T, Nader N, Chrousos GP.** Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci Signal*. 2010;3(107):ra8.
205. **Williams GT, Mourtada-Maarabouni M, Farzaneh F.** A critical role for non-coding RNA GAS5 in growth arrest and rapamycin inhibition in human T-lymphocytes. *Biochem Soc Trans*. 2011;39(2):482–486.
206. **Mourtada-Maarabouni M, Pickard MR, Hedge VL, Farzaneh F, Williams GT.** GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer. *Oncogene*. 2009;28(2):195–208.

207. **Schwartz JC, Younger ST, Nguyen NB, et al.** Antisense transcripts are targets for activating small RNAs. *Nat Struct Mol Biol*. 2008;15(8):842–848.
208. **Janowski BA, Younger ST, Hardy DB, Ram R, Huffman KE, Corey DR.** Activating gene expression in mammalian cells with promoter-targeted duplex RNAs. *Nat Chem Biol*. 2007;3(3):166–173.
209. **Janowski BA, Corey DR.** Minireview: Switching on progesterone receptor expression with duplex RNA. *Mol Endocrinol*. 2010;24(12):2243–2252.
210. **Yang L, Lin C, Jin C, et al.** lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. *Nature*. 2013;500(7464):598–602.
211. **Prensner JR, Sahu A, Iyer MK, et al.** The IncRNAs PCGEM1 and PRNCR1 are not implicated in castration resistant prostate cancer. *Oncotarget*. 2014;5(6):1434–1438.
212. **Gadad SS, Sun M, Kraus WL.** Identification and functional characterization of estrogen-regulated long noncoding RNAs in breast cancer cells: From genomics to molecular and cellular biology. Program of the Endocrine Society's 94th Annual Meeting and Expo; June 23–26, 2012; Houston, TX. Abstract SUN-565.
213. **Sun M, Danko GG, Gadad SS, Hah N, Kraus WL.** Characterization of estrogen-regulated non-coding RNAs. Program of the Endocrine Society's 93rd Annual Meeting and Expo; June 4–7, 2011; Boston, MA. Abstract OR14–12.
214. **Hah N, Murakami S, Nagari A, Danko CG, Kraus WL.** Enhancer transcripts mark active estrogen receptor binding sites. *Genome Res*. 2013;23(8):1210–1223.
215. **Li W, Notani D, Ma Q, et al.** Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature*. 2013;498(7455):516–520.
216. **Lai F, Shiekhattar R.** Enhancer RNAs: the new molecules of transcription. *Curr Opin Genet Dev*. 2014;25:38–42.
217. **Hens JR, Wysolmerski JJ.** Key stages of mammary gland development: molecular mechanisms involved in the formation of the embryonic mammary gland. *Breast Cancer Res*. 2005;7(5):220–224.
218. **Daniel CW, Smith GH.** The mammary gland: a model for development. *J Mammary Gland Biol Neoplasia*. 1999;4(1):3–8.
219. **Brisken C, O'Malley B.** Hormone action in the mammary gland. *Cold Spring Harb Perspect Biol*. 2010;2(12):a003178.
220. **Lanz RB, Chua SS, Barron N, Söder BM, DeMayo F, O'Malley BW.** Steroid receptor RNA activator stimulates proliferation as well as apoptosis in vivo. *Mol Cell Biol*. 2003;23(20):7163–7176.
221. **Ginger MR, Shore AN, Contreras A, et al.** A noncoding RNA is a potential marker of cell fate during mammary gland development. *Proc Natl Acad Sci U S A*. 2006;103(15):5781–5786.
222. **Shore AN, Kabotyanski EB, Roarty K, et al.** Pregnancy-induced noncoding RNA (PINC) associates with polycomb repressive complex 2 and regulates mammary epithelial differentiation. *PLoS Genet*. 2012;8(7):e1002840.
223. **Askarian-Amiri ME, Crawford J, French JD, et al.** SNORD-host RNA Zfas1 is a regulator of mammary development and a potential marker for breast cancer. *RNA*. 2011;17(5):878–891.
224. **Ginger MR, Gonzalez-Rimbau MF, Gay JP, Rosen JM.** Persistent changes in gene expression induced by estrogen and progesterone in the rat mammary gland. *Mol Endocrinol*. 2001;15(11):1993–2009.
225. **Frühbeck G, Gómez-Ambrosi J, Muruzábal FJ, Burrell MA.** The adipocyte: a model for integration of endocrine and metabolic signaling in energy metabolism regulation. *Am J Physiol Endocrinol Metab*. 2001;280(6):E827–847.
226. **Collombat P, Hecksher-Sørensen J, Serup P, Mansouri A.** Specifying pancreatic endocrine cell fates. *Mech Dev*. 2006;123(7):501–512.
227. **Suckale J, Solimena M.** Pancreas islets in metabolic signaling–focus on the beta-cell. *Front Biosci*. 2008;13:7156–7171.
228. **Hajer GR, van Haeften TW, Visseren FL.** Adipose tissue dysfunction in obesity, diabetes, and vascular diseases. *Eur Heart J*. 2008;29(24):2959–2971.
229. **Soleimanpour SA, Stoffers DA.** The pancreatic β cell and type 1 diabetes: innocent bystander or active participant? *Trends Endocrinol Metab*. 2013;24(7):324–331.
230. **Romacho T, Elsen M, Rohrborn D, Eckel J.** Adipose tissue and its role in organ crosstalk. *Acta Physiol (Oxf)*. 2014;210(4):733–753.
231. **Gittes GK.** Dev Biol of the pancreas: a comprehensive review. *Dev Biol*. 2009;326(1):4–35.
232. **Cnop M, Welsh N, Jonas JC, Jorns A, Lenzen S, Eizirik DL.** Mechanisms of pancreatic beta-cell death in type 1 and type 2 diabetes: many differences, few similarities. *Diabetes*. 2005;54(Suppl 2):S97–S107.
233. **Taylor R.** Insulin resistance and type 2 diabetes. *Diabetes*. 2012;61(4):778–779.
234. **Xu B, Gerin I, Miao H, et al.** Multiple roles for the noncoding RNA SRA in regulation of adipogenesis and insulin sensitivity. *PLoS One*. 2010;5(12):e14199.
235. **Liu S, Sheng L, Miao H, et al.** SRA gene knockout protects against diet-induced obesity and improves glucose tolerance. *J Biol Chem*. 2014;289(19):13000–13009.
236. **Wang F, Tong Q.** Transcription factor PU. 1 is expressed in white adipose and inhibits adipocyte differentiation. *Am J Physiol Cell Physiol*. 2008;295(1):C213–C220.
237. **Pang WJ, Lin LG, Xiong Y, et al.** Knockdown of PU.1 as lncRNA inhibits adipogenesis through enhancing PU.1 mRNA translation. *J Cell Biochem*. 2013;114(11):2500–2512.
238. **Freytag SO, Paielli DL, Gilbert JD.** Ectopic expression of the CCAAT/enhancer-binding protein alpha promotes the adipogenic program in a variety of mouse fibroblastic cells. *Genes Dev*. 1994;8(14):1654–1663.
239. **Zhao XY, Li S, Wang GX, Yu Q, Lin JD.** A long noncoding RNA transcriptional regulatory circuit drives thermogenic adipocyte differentiation. *Mol Cell*. 2014;55(3):372–382.
240. **Morán I, Akerman I, van de Bunt M, et al.** Human β cell transcriptome analysis uncovers lncRNAs that are tissue-specific, dynamically regulated, and abnormally expressed in type 2 diabetes. *Cell Metab*. 2012;16(4):435–448.
241. **Farooqi IS, O'Rahilly S.** Monogenic obesity in humans. *Annu Rev Med*. 2005;56:443–458.
242. **Duker AL, Ballif BC, Bawle EV, et al.** Paternally inherited microdeletion at 15q11.2 confirms a significant role for the SNORD116 C/D box snoRNA cluster in Prader-Willi syndrome. *Eur J Hum Genet*. 2010;18(11):1196–1201.

243. **Sahoo T, del Gaudio D, German JR, et al.** Prader-Willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. *Nat Genet*. 2008; 40(6):719–721.
244. **Vitali P, Royo H, Marty V, Bortolin-Cavaille ML, Cavaille J.** Long nuclear-retained non-coding RNAs and allele-specific higher-order chromatin organization at imprinted snoRNA gene arrays. *J Cell Sci*. 2010;123(Pt 1):70–83.
245. **Powell WT, Coulson RL, Crary FK, et al.** A Prader-Willi locus lncRNA cloud modulates diurnal genes and energy expenditure. *Hum Mol Genet*. 2013;22(21):4318–4328.
246. **Stelzer Y, Sagi I, Yanuka O, Eiges R, Benvenisty N.** The noncoding RNA IPW regulates the imprinted DLK1-DIO3 locus in an induced pluripotent stem cell model of Prader-Willi syndrome. *Nat Genet*. 2014;46(6):551–557.
247. **Chaplin DD.** Overview of the immune response. *J Allergy Clin Immunol*. 2010;125(2 Suppl 2):S3–S23.
248. **McCusker C, Warrington R.** Primary immunodeficiency. *Allergy Asthma Clin Immunol*. 2011;7 Suppl 1:S11.
249. **Galli SJ, Tsai M, Piliponsky AM.** The development of allergic inflammation. *Nature*. 2008;454(7203):445–454.
250. **Mantovani A, Allavena P, Sica A, Balkwill F.** Cancer-related inflammation. *Nature*. 2008;454(7203):436–444.
251. **Gregersen PK, Olsson LM.** Recent advances in the genetics of autoimmune disease. *Annu Rev Immunol*. 2009;27: 363–391.
252. **Peng X, Gralinski L, Armour CD, et al.** Unique signatures of long noncoding RNA expression in response to virus infection and altered innate immune signaling. *MBio*. 2010;1(5):e00206–10.
253. **Carpenter S, Aiello D, Atianand MK, et al.** A long noncoding RNA mediates both activation and repression of immune response genes. *Science*. 2013;341(6147):789–792.
254. **Rapicavoli NA, Qu K, Zhang J, Mikhail M, Laberge RM, Chang HY.** A mammalian pseudogene lncRNA at the interface of inflammation and anti-inflammatory therapeutics. *Elife*. 2013;2:e00762.
255. **Imamura K, Imamachi N, Akizuki G, et al.** Long Noncoding RNA NEAT1-Dependent SFPQ Relocation from Promoter Region to Paraspeckle Mediates IL8 Expression upon Immune Stimuli. *Mol Cell*. 2014;53(3):393–406.
256. **Pang KC, Dinger ME, Mercer TR, et al.** Genome-wide identification of long noncoding RNAs in $CD8^+$ T cells. *J Immunol*. 2009;182(12):7738–7748.
257. **Hu G, Tang Q, Sharma S, et al.** Expression and regulation of intergenic long noncoding RNAs during T cell development and differentiation. *Nat Immunol*. 2013;14(11): 1190–1198.
258. **Gomez JA, Wapinski OL, Yang YW, et al.** The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon-γ locus. *Cell*. 2013;152(4):743–754.
259. **Vigneau S, Rohrlich PS, Brahic M, Bureau JF.** Tmevpg1, a candidate gene for the control of Theiler's virus persistence, could be implicated in the regulation of gamma interferon. *J Virol*. 2003;77(10):5632–5638.
260. **Wang P, Xue Y, Han Y, et al.** The STAT3-binding long noncoding RNA lnc-DC controls human dendritic cell differentiation. *Science*. 2014;344(6181):310–313.
261. **Mercer TR, Dinger ME, Sunkin SM, Mehler MF, Mattick JS.** Specific expression of long noncoding RNAs in the mouse brain. *Proc Natl Acad Sci U S A*. 2008;105(2):716–721.
262. **Belgard TG, Marques AC, Oliver PL, et al.** A transcriptomic atlas of mouse neocortical layers. *Neuron*. 2011; 71(4):605–616.
263. **Mehler MF, Mattick JS.** Noncoding RNAs and RNA editing in brain development, functional diversification, and neurological disease. *Physiol Rev*. 2007;87(3):799–823.
264. **Qureshi IA, Mattick JS, Mehler MF.** Long non-coding RNAs in nervous system function and disease. *Brain Res*. 2010;1338:20–35.
265. **Pastori C, Wahlestedt C.** Involvement of long noncoding RNAs in diseases affecting the central nervous system. *RNA Biol*. 2012;9(6):860–870.
266. **Ng SY, Lin L, Soh BS, Stanton LW.** Long noncoding RNAs in development and disease of the central nervous system. *Trends Genet*. 2013;29(8):461–468.
267. **Niland CN, Merry CR, Khalil AM.** Emerging roles for long non-coding RNAs in cancer and neurological disorders. *Front Genet*. 2012;3:25.
268. **Ramos AD, Diaz A, Nellore A, et al.** Integration of genome-wide approaches identifies lncRNAs of adult neural stem cells and their progeny in vivo. *Cell Stem Cell*. 2013;12(5): 616–628.
269. **Aprea J, Prenninger S, Dori M, et al.** Transcriptome sequencing during mouse brain development identifies long non-coding RNAs functionally involved in neurogenic commitment. *EMBO J*. 2013;32(24):3145–3160.
270. **Ng SY, Bogu GK, Soh BS, Stanton LW.** The long noncoding RNA RMST interacts with SOX2 to regulate neurogenesis. *Mol Cell*. 2013;51(3):349–359.
271. **Onoguchi M, Hirabayashi Y, Koseki H, Gotoh Y.** A noncoding RNA regulates the neurogenin1 gene locus during mouse neocortical development. *Proc Natl Acad Sci U S A*. 2012;109(42):16939–16944.
272. **Young TL, Matsuda T, Cepko CL.** The noncoding RNA taurine upregulated gene 1 is required for differentiation of the murine retina. *Curr Biol*. 2005;15(6):501–512.
273. **Rapicavoli NA, Poth EM, Zhu H, Blackshaw S.** The long noncoding RNA Six3OS acts in trans to regulate retinal development by modulating Six3 activity. *Neural Dev*. 2011;6:32.
274. **Rapicavoli NA, Poth EM, Blackshaw S.** The long noncoding RNA RNCR2 directs mouse retinal cell specification. *BMC Dev Biol*. 2010;10:49.
275. **Meola N, Pizzo M, Alfano G, Surace EM, Banfi S.** The long noncoding RNA Vax2os1 controls the cell cycle progression of photoreceptor progenitors in the mouse retina. *RNA*. 2012;18(1):111–123.
276. **Mercer TR, Qureshi IA, Gokhan S, et al.** Long noncoding RNAs in neuronal-glial fate specification and oligodendrocyte lineage maturation. *BMC Neurosci*. 2010;11:14.
277. **Panganiban G, Rubenstein JL.** Developmental functions of the Distal-less/Dlx homeobox genes. *Development*. 2002; 129(19):4371–4386.
278. **Feng J, Bi C, Clark BS, Mady R, Shah P, Kohtz JD.** The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. *Genes Dev*. 2006;20(11):1470–1484.
279. **Bond AM, Vangompel MJ, Sametsky EA, et al.** Balanced

gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry. *Nat Neurosci*. 2009;12(8):1020–1027.
280. **Faghihi MA, Modarresi F, Khalil AM, et al.** Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. *Nat Med*. 2008;14(7):723–730.
281. **Xie Y, Hayden MR, Xu B.** BDNF overexpression in the forebrain rescues Huntington's disease phenotypes in YAC128 mice. *J Neurosci*. 2010;30(44):14708–14718.
282. **Modarresi F, Faghihi MA, Lopez-Toledano MA, et al.** Inhibition of natural antisense transcripts in vivo results in gene-specific transcriptional upregulation. *Nat Biotechnol*. 2012;30(5):453–459.
283. **Yamasaki K, Joh K, Ohta T, et al.** Neurons but not glial cells show reciprocal imprinting of sense and antisense transcripts of Ube3a. *Hum Mol Genet*. 2003;12(8):837–847.
284. **Matsuura T, Sutcliffe JS, Fang P, et al.** De novo truncating mutations in E6-AP ubiquitin-protein ligase gene (UBE3A) in Angelman syndrome. *Nat Genet*. 1997;15(1):74–77.
285. **Chamberlain SJ, Lalande M.** Angelman syndrome, a genomic imprinting disorder of the brain. *J Neurosci*. 2010;30(30):9958–9963.
286. **Landers M, Calciano MA, Colosi D, Glatt-Deeley H, Wagstaff J, Lalande M.** Maternal disruption of Ube3a leads to increased expression of Ube3a-ATS in trans. *Nucleic Acids Res*. 2005;33(13):3976–3984.
287. **Meng L, Person RE, Beaudet AL.** Ube3a-ATS is an atypical RNA polymerase II transcript that represses the paternal expression of Ube3a. *Hum Mol Genet*. 2012;21(13):3001–3012.
288. **Khalil AM, Faghihi MA, Modarresi F, Brothers SP, Wahlestedt C.** A novel RNA transcript with antiapoptotic function is silenced in fragile X syndrome. *PLoS One*. 2008;3(1):e1486.
289. **Ladd PD, Smith LE, Rabaia NA, et al.** An antisense transcript spanning the CGG repeat region of FMR1 is upregulated in premutation carriers but silenced in full mutation individuals. *Hum Mol Genet*. 2007;16(24):3174–3187.
290. **Sopher BL, Ladd PD, Pineda VV, et al.** CTCF regulates ataxin-7 expression through promotion of a convergently transcribed, antisense noncoding RNA. *Neuron*. 2011;70(6):1071–1084.
291. **Moseley ML, Zu T, Ikeda Y, et al.** Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8. *Nat Genet*. 2006;38(7):758–769.
292. **Daughters RS, Tuttle DL, Gao W, et al.** RNA gain-of-function in spinocerebellar ataxia type 8. *PLoS Genet*. 2009;5(8):e1000600.
293. **Koob MD, Moseley ML, Schut LJ, et al.** An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). *Nat Genet*. 1999;21(4):379–384.
294. **Luo X, Chae M, Krishnakumar R, Danko CG, Kraus WL.** Dynamic reorganization of the AC16 cardiomyocyte transcriptome in response to TNFα signaling revealed by integrated genomic analyses. *BMC Genomics*. 2014;15:155.
295. **Burd CE, Jeck WR, Liu Y, Sanoff HK, Wang Z, Sharpless NE.** Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk. *PLoS Genet*. 2010;6(12):e1001233.
296. **Friedrichs F, Zugck C, Rauch GJ, et al.** HBEGF, SRA1, and IK: Three cosegregating genes as determinants of cardiomyopathy. *Genome Res*. 2009;19(3):395–403.
297. **Ishii N, Ozaki K, Sato H, et al.** Identification of a novel non-coding RNA, MIAT, that confers risk of myocardial infarction. *J Hum Genet*. 2006;51(12):1087–1099.
298. **Han P, Li W, Lin CH, et al.** A long noncoding RNA protects the heart from pathological hypertrophy. *Nature*. 2014;514(7520):102–106.
299. **Sakakibara I, Santolini M, Ferry A, Hakim V, Maire P.** Six homeoproteins and a linc-RNA at the fast MYH locus lock fast myofiber terminal phenotype. *PLoS Genet*. 2014;10(5):e1004386.
300. **Podlowski S, Bramlage P, Baumann G, Morano I, Luther HP.** Cardiac troponin I sense-antisense RNA duplexes in the myocardium. *J Cell Biochem*. 2002;85(1):198–207.
301. **Luther HP.** Role of endogenous antisense RNA in cardiac gene regulation. *J Mol Med (Berl)*. 2005;83(1):26–32.
302. **Ritter O, Haase H, Schulte HD, Lange PE, Morano I.** Remodeling of the hypertrophied human myocardium by cardiac bHLH transcription factors. *J Cell Biochem*. 1999;74(4):551–561.
303. **Caretti G, Schiltz RL, Dilworth FJ, et al.** The RNA helicases p68/p72 and the noncoding RNA SRA are coregulators of MyoD and skeletal muscle differentiation. *Dev Cell*. 2006;11(4):547–560.
304. **McKay DB, Xi L, Barthel KK, Cech TR.** Structure and function of steroid receptor RNA activator protein, the proposed partner of SRA noncoding RNA. *J Mol Biol*. 2014;426(8):1766–1785.
305. **Dey BK, Pfeifer K, Dutta A.** The H19 long noncoding RNA gives rise to microRNAs miR-675–3p and miR-675–5p to promote skeletal muscle differentiation and regeneration. *Genes Dev*. 2014;28(5):491–501.
306. **Legnini I, Morlando M, Mangiavacchi A, Fatica A, Bozzoni I.** A feedforward regulatory loop between HuR and the long noncoding RNA linc-MD1 controls early phases of myogenesis. *Mol Cell*. 2014;53(3):506–514.
307. **Cabianca DS, Casa V, Bodega B, et al.** A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. *Cell*. 2012;149(4):819–831.
308. **Du Z, Fei T, Verhaak RG, et al.** Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer. *Nat Struct Mol Biol*. 2013;20(7):908–913.
309. **Trimarchi T, Bilal E, Ntziachristos P, et al.** Genome-wide mapping and characterization of Notch-regulated long noncoding RNAs in acute leukemia. *Cell*. 2014;158(3):593–606.
310. **Hu X, Feng Y, Zhang D, et al.** A functional genomic approach identifies FAL1 as an oncogenic long noncoding RNA that associates with BMI1 and represses p21 expression in cancer. *Cancer Cell*. 2014;26(3):344–357.
311. **Prensner JR, Iyer MK, Balbin OA, et al.** Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nat Biotechnol*. 2011;29(8):742–749.
312. **Gutschner T, Diederichs S.** The hallmarks of cancer: a long

non-coding RNA point of view. *RNA Biol*. 2012;9(6):703–719.
313. **Mitra SA, Mitra AP, Triche TJ.** A central role for long non-coding RNA in cancer. *Front Genet*. 2012;3:17.
314. **Nie L, Wu HJ, Hsu JM, et al.** Long non-coding RNAs: versatile master regulators of gene expression and crucial players in cancer. *Am J Transl Res*. 2012;4(2):127–150.
315. **Huarte M, Rinn JL.** Large non-coding RNAs: missing links in cancer? *Hum Mol Genet*. 2010;19(R2):R152–R161.
316. **Fu X, Ravindranath L, Tran N, Petrovics G, Srivastava S.** Regulation of apoptosis by a prostate-specific and prostate cancer-associated noncoding gene, PCGEM1. *DNA Cell Biol*. 2006;25(3):135–141.
317. **Morris KV, Santoso S, Turner AM, Pastori C, Hawkins PG.** Bidirectional transcription directs both transcriptional gene activation and suppression in human cells. *PLoS Genet*. 2008;4(11):e1000258.
318. **Prensner JR, Chen W, Iyer MK, et al.** PCAT-1, a long noncoding RNA, regulates BRCA2 and controls homologous recombination in cancer. *Cancer Res*. 2014;74(6):1651–1660.
319. **Balik V, Srovnal J, Sulla I, et al.** MEG3: a novel long noncoding potentially tumour-suppressing RNA in meningiomas. *J Neurooncol*. 2013;112(1):1–8.
320. **Marin-Bejar O, Marchese FP, Athie A, et al.** Pint lincRNA connects the p53 pathway with epigenetic silencing by the Polycomb repressive complex 2. *Genome Biol*. 2013;14(9):R104.
321. **Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP.** A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature*. 2010;465(7301):1033–1038.
322. **Johnsson P, Ackley A, Vidarsdottir L, et al.** A pseudogene long-noncoding-RNA network regulates PTEN transcription and translation in human cells. *Nat Struct Mol Biol*. 2013;20(4):440–446.
323. **Liu Q, Huang J, Zhou N, et al.** LncRNA loc285194 is a p53-regulated tumor suppressor. *Nucleic Acids Res*. 2013;41(9):4976–4987.
324. **Arab K, Park YJ, Lindroth AM, et al.** Long noncoding RNA TARID directs demethylation and activation of the tumor suppressor TCF21 via GADD45A. *Mol Cell*. 2014;55(4):604–614.
325. **Chiyomaru T, Fukuhara S, Saini S, et al.** Long non-coding RNA HOTAIR is targeted and regulated by miR-141 in human cancer cells. *J Biol Chem*. 2014;289(18):12550–12565.
326. **Prensner JR, Iyer MK, Sahu A, et al.** The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *Nat Genet*. 2013;45(11):1392–1398.
327. **Yuan JH, Yang F, Wang F, et al.** A long noncoding RNA activated by TGF-β promotes the invasion-metastasis cascade in hepatocellular carcinoma. *Cancer Cell*. 2014;25(5):666–681.
328. **Li CH, Chen Y.** Targeting long non-coding RNAs in cancers: progress and prospects. *Int J Biochem Cell Biol*. 2013;45(8):1895–1910.
329. **Ling H, Fabbri M, Calin GA.** MicroRNAs and other noncoding RNAs as targets for anticancer drug development. *Nat Rev Drug Discov*. 2013;12(11):847–865.
330. **He W, Cai Q, Sun F, et al.** linc-UBC1 physically associates with polycomb repressive complex 2 (PRC2) and acts as a negative prognostic factor for lymph node metastasis and survival in bladder cancer. *Biochim Biophys Acta*. 2013;1832(10):1528–1537.
331. **Kawakami T, Okamoto K, Ogawa O, Okada Y.** XIST unmethylated DNA fragments in male-derived plasma as a tumour marker for testicular cancer. *Lancet*. 2004;363(9402):40–42.
332. **Lee GL, Dobi A, Srivastava S.** Prostate cancer: diagnostic performance of the PCA3 urine test. *Nat Rev Urol*. 2011;8(3):123–124.
333. **Wang Y, Zhang D, Wu K, Zhao Q, Nie Y, Fan D.** Long noncoding RNA MRUL promotes ABCB1 expression in multidrug-resistant gastric cancer cell sublines. *Mol Cell Biol*. 2014;34(17):3182–3193.
334. **Szymanski M, Erdmann VA, Barciszewski J.** Noncoding RNAs database (ncRNAdb). *Nucleic Acids Res*. 2007;35(Database issue):D162–D164.
335. **Xie C, Yuan J, Li H, et al.** NONCODEv4: exploring the world of long non-coding RNA genes. *Nucleic Acids Res*. 2014;42(Database issue):D98–D103.
336. **Sun M, Kraus WL.** Long noncoding RNAs: new "links" between gene expression and cellular outcomes in endocrinology. *Mol Endocrinol*. 2013;27(9):1390–1402.