

Chapter 1

History, Discovery, and Classification of lncRNAs

Julien Jarroux, Antonin Morillon, and Marina Pinskaya

Abstract The RNA World Hypothesis suggests that prebiotic life revolved around RNA instead of DNA and proteins. Although modern cells have changed significantly in 4 billion years, RNA has maintained its central role in cell biology. Since the discovery of DNA at the end of the nineteenth century, RNA has been extensively studied. Many discoveries such as housekeeping RNAs (rRNA, tRNA, etc.) supported the messenger RNA model that is the pillar of the central dogma of molecular biology, which was first devised in the late 1950s. Thirty years later, the first regulatory non-coding RNAs (ncRNAs) were initially identified in bacteria and then in most eukaryotic organisms. A few long ncRNAs (lncRNAs) such as *H19* and *Xist* were characterized in the pre-genomic era but remained exceptions until the early 2000s. Indeed, when the sequence of the human genome was published in 2001, studies showed that only about 1.2% encodes proteins, the rest being deemed “non-coding.” It was later shown that the genome is pervasively transcribed into many ncRNAs, but their functionality remained controversial. Since then, regulatory lncRNAs have been characterized in many species and were shown to be involved in processes such as development and pathologies, revealing a new layer of regulation in eukaryotic cells. This newly found focus on lncRNAs, together with the advent of high-throughput sequencing, was accompanied by the rapid discovery of many novel transcripts which were further characterized and classified according to specific transcript traits.

In this review, we will discuss the many discoveries that led to the study of lncRNAs, from Friedrich Miescher’s “nuclein” in 1869 to the elucidation of the human genome and transcriptome in the early 2000s. We will then focus on the biological relevance during lncRNA evolution and describe their basic features as genes and transcripts. Finally, we will present a non-exhaustive catalogue of lncRNA classes, thus illustrating the vast complexity of eukaryotic transcriptomes.

Keywords Non-coding RNA • Classification • RNA World • Central dogma

J. Jarroux • A. Morillon (✉) • M. Pinskaya
ncRNA, epigenetic and genome fluidity, Institut Curie, Centre de Recherche, CNRS UMR 3244, PSL Research University and Université Pierre et Marie Curie, Paris, France
e-mail: julien.jarroux@curie.fr; antonin.morillon@curie.fr; marina.pinskaya@curie.fr

The deep complexity of eukaryotic transcriptomes and the rapid development of high-throughput sequencing technologies led to an explosion in the number of newly identified and uncharacterized lncRNAs. Many challenges in lncRNA biology remain, including accurate annotation, functional characterization, and clinical relevance. All these topics will be thoroughly discussed throughout the book. But to start with, we will detail the discovery of RNA as life's indispensable molecule. The long journey for the biological characterization of non-coding RNAs is summed up in Fig. 1.1, and this history will be described over the first half of this chapter, from the DNA to the first non-coding transcripts. Then, we will discuss how global genomic and transcriptomic studies changed our view on the role of RNA in regulatory circuits, biodiversity, and complexity. Finally, we will include a summary of the extensive classification of lncRNAs.

1.1 A Hundred-Years History of RNA Biology

Before the ever-expanding catalogues of lncRNAs that we have today, a long experimental and theoretical journey was required to prove the importance of RNA molecules in cell biology. It began in 1869 with the discovery of nucleic acids, and it took over a hundred years for researchers to finally identify non-coding transcripts and begin proposing regulatory roles for them.

1.1.1 *From “Nuclein” to Nucleic Acids and to the Double Helix*

At the end of the nineteenth century, a few pivotal discoveries foreshadowed the molecular biology era. In 1869, Friedrich Miescher isolated a material from nuclei that he called “nuclein” and which he described as highly acidic: in fact he had discovered DNA [1]. In contrast with proteins that were the main focus at the time, its content was low in sulfur and very high in phosphorus and could not be digested by protease treatment. Later, once the chemical composition of the “nuclein” isolated from different organisms had been discovered, it was realized that “thymus nucleic acid” consisted of DNA, while “yeast nucleic acid” was composed of RNA. In the early 1900s, several scientists proposed the chemical composition and the first structures for DNA and RNA, though the biological differences between these two molecules were still not apparent. Ironically, the discovery of “nuclein” by Miescher happened only a few years after Gregor Mendel published his work on the laws of heredity in 1866, but nevertheless many scientists thought proteins were the carriers of genetic information. Thus, the link between Mendel's model and

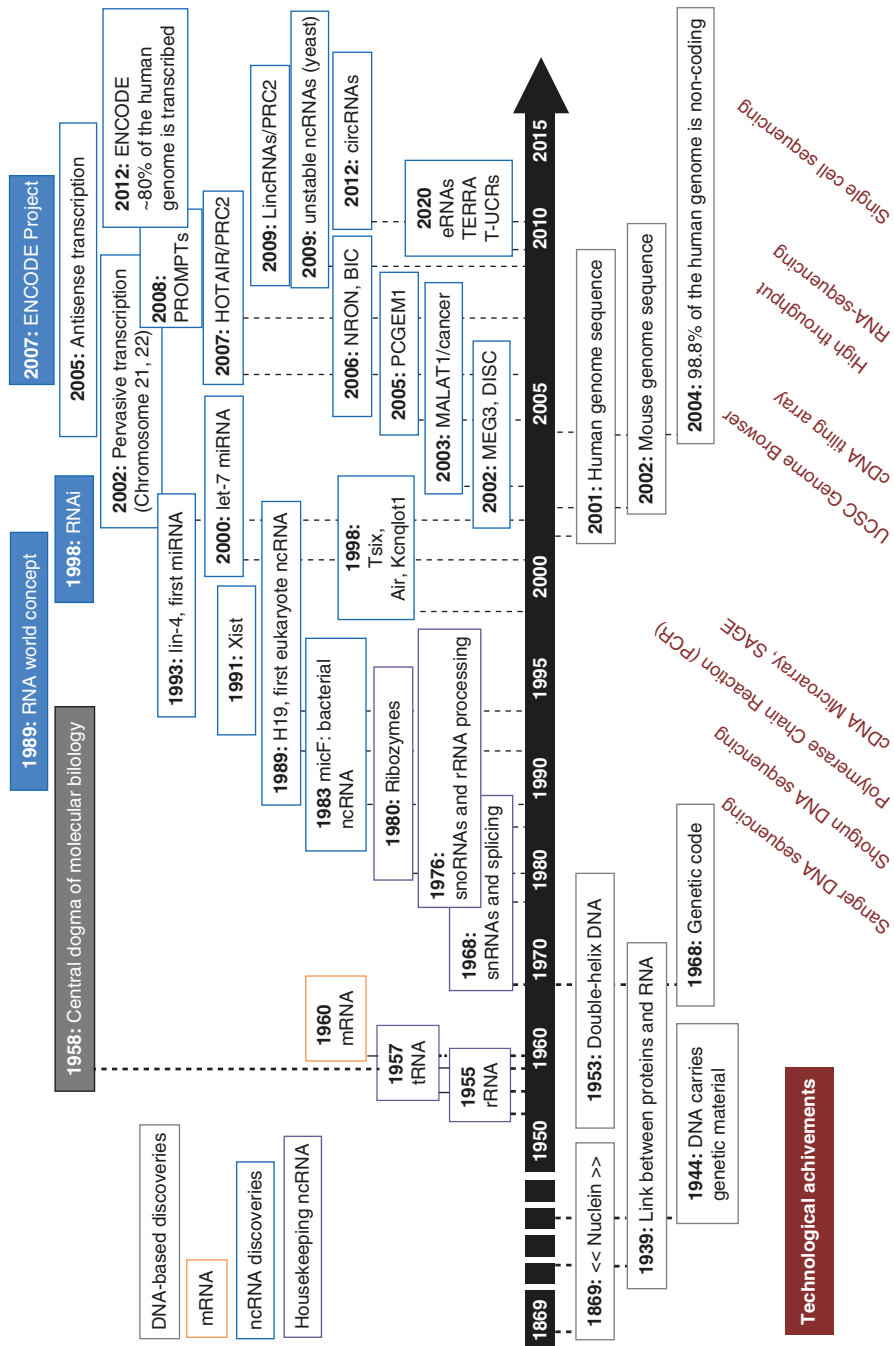


Fig. 1.1 The timeline of principle discoveries in nucleic acid biology and, in particular, eukaryotic ncRNAs

Miescher's "nuclein" remained missing until 1944 when Oswald Avery proposed DNA as a carrier of genetic information [2].

The link between DNA and RNA was established in the late 1950s as Elliot Volkin and Lawrence Astrachan thoroughly described RNA as a DNA-like molecule synthesized from DNA. This discovery was then further elaborated into a molecular concept of RNA and DNA synthesis [3, 4]. Indeed, following the X-ray crystallographic studies of Rosalind Franklin and the establishment of the double-helix structure of DNA by James Watson and Francis Crick in 1953, it was proposed in 1961 that RNA could be an intermediate molecule in the information flow from DNA to proteins [5]. First devised in 1958 by Francis Crick and then by François Jacob and Jacques Monod, the *Central Dogma of Molecular Biology* comprised transcription of a DNA gene into RNA in the nucleus followed by protein synthesis in the cytoplasm. It was also stated that the information flow can only proceed from DNA to RNA and then from RNA to protein, but never from protein to nucleic acids [5]. The mediating role of RNA became a new focus of research which has been pivotal for the development of modern molecular biology.

1.1.2 A Central Role for RNA in Cell Biology: The RNA World Concept

In 1939, Torbjörn Caspersson and Jean Brachet showed independently that the cytoplasm is very rich in RNA. They also showed that cells producing high amount of proteins seemed to have high amounts of RNA as well [5]. This was a first hint for the requirement of RNA during protein synthesis and its role as a link between DNA and proteins. In 1955, Georges Palade identified the very first ncRNA that makes part of the very abundant cytoplasmic ribonucleoprotein (RNP) complex: the ribosome. In his "Central Dogma" Crick also theorized that there was an "adapter" molecule for the translation of RNA to amino acids. This second class of ncRNAs was discovered in 1957 by Mahlon Hoagland and Paul Zamecnik: the transfer (t)RNA. In 1960, François Jacob and Jacques Monod first coined the term "messenger RNA" (mRNA) as part of their study of inducible enzymes in *Escherichia (E.) coli*. Indeed, they showed the existence of an intermediate molecule carrying the genetic information leading to protein synthesis. Shortly after, the work of Crick helped establish that the genetic code is a comma-less, non-overlapping triplet code in which three nucleotides code for one amino acid. It was later deciphered *in vitro* as well as *in vivo* and shown to be universal across all living organisms [6]. In the late 1960s, rather different from mRNAs, a new class of short-lived nuclear RNAs was found: heterogeneous nuclear (hn)RNAs. These long RNA molecules, which were in fact precursors for mature rRNAs and mRNAs, led to the study of rRNA processing and the discovery of splicing [7, 8]. During that period, small nuclear (sn)RNAs

which are part of the spliceosome, the RNP machinery responsible for intron splicing from pre-mRNAs, were discovered [9]; as well as small nucleolar (sno) RNAs, which are involved in the processing and maturation of ribosomal RNAs in the nucleolus [10].

Although Jacob, Monod, and Crick had already mentioned independently that RNA was not just a messenger, many scientists considered it as a mere unstable intermediating molecule, overlooking the active roles of other classes of ncRNAs. However, this view partially changed in 1980 when Thomas Cech and Sidney Altman discovered that RNA molecules could act as catalysts for a chemical reaction. Initially, Cech's group found an intron from an mRNA in *Tetrahymena thermophila* that is able to perform its own splicing through an RNA-catalyzed cleavage [11]. Subsequently, Altman's group showed that the RNA component of the ribonucleoprotein RNase P is responsible for its activity in degrading RNA [12]. These RNA enzymes were called ribozymes and have been shown since then to be key actors of the genetic information flow, making part of both the ribosome and the spliceosome [13, 14].

The discovery of catalytic RNA also led scientists to develop the RNA World theory, which states that prebiotic life revolved around RNA, since it appeared before DNA and protein. Indeed, the extensive studies of its roles in cell biology revealed that RNA is necessary for DNA replication and that its ribonucleotides are precursors for DNA's deoxyribonucleotides. Moreover, as it was previously mentioned, RNA plays an important role in every step of protein synthesis, both as scripts (mRNAs) and actors (ncRNAs: rRNAs, tRNAs, etc.) (Fig. 1.2) [15]. Remarkably, the latter ones are constitutively expressed in the cell and are necessary for vital cellular functions, constituting a class of housekeeping ncRNAs. Being

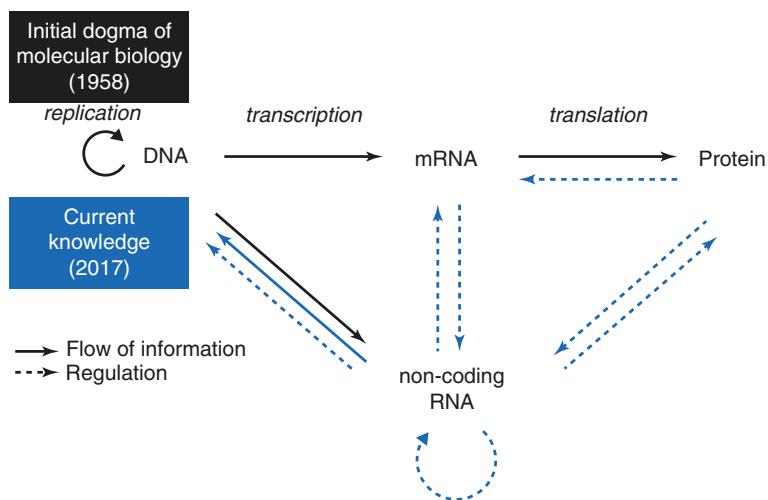


Fig. 1.2 Initial and current dogma of molecular biology

extensively studied, housekeeping ncRNAs are the subject of many specialized publications and will not be described here. Instead, other classes of regulatory ncRNAs that were discovered in the early 1990s will be discussed. These ncRNAs are characterized by very specific expression during certain developmental stages, in certain tissues or disease states, and play multiple roles in gene expression regulation.

1.1.3 Bacterial sRNAs: Pioneers of Regulatory ncRNAs

The very first regulatory ncRNA to be discovered and characterized was *micF* from the bacteria *E. coli*. It was described as the first RNA regulating gene expression through sense-antisense base pairing in 1984 by the team of Masayuki Inoue [16] and represents the major class of bacterial regulatory ncRNAs, small (s) RNAs. The *micF* ncRNA was shown to repress the translation of a target mRNA encoding a porin (outer membrane protein F, OmpF), involved in passive transport through the cell membrane. First discovered through multicopy plasmid experiments, the transcript was isolated 3 years later and shown to be an independent gene. When transcription of *micF* is activated, it inhibits the expression of the *ompF* gene at both mRNA and protein levels. Subsequently, following the characterization of the RNA duplex structure *in vitro*, *micF* was shown to bind to the ribosome-binding site (RBS) of the *ompF* mRNA, thus inhibiting ribosome binding and translation.

More recently, it was shown that the regulation of gene expression by *micF* through base pairing extends to other genes, among which is the *lrp* mRNA [17]. Lrp (leucine-responsive protein) is a transcription factor that vastly regulates gene expression in *E. coli* in response to osmotic changes and nutrient availability. Remarkably, Lrp regulates *micF* expression as well, thus creating a feedback and proving the important role of *micF* in global gene regulation and metabolism. The same mechanisms were also found in *Salmonella*, supporting the evolutionary conservation of this regulatory pathway [18]. Since then many other sRNAs ranging in length from 50 to 500 nucleotides (nt) have been discovered, including *trans*- or *cis*-encoded ncRNAs, RNA thermometers, and riboswitches. They all act by pairing, thus inhibiting translation of targeted mRNAs and inducing their degradation.

1.1.4 MicroRNAs and RNA Interference

In the early 1990s, several scientists observed independently and in different eukaryotic organisms, through experiments of transgene co-expression or viral infection, an intriguing phenomenon of RNA-mediated inhibition of protein synthesis. The regulatory effects of these RNA molecules reshaped the views of RNA as a mere messenger. The very first studies described the phenomenon as “co-suppression” in

plants, as “posttranscriptional gene silencing” in nematodes, or as “quelling” in fungi, but none of them suspected RNA to be the key actor until the identification of the first micro (mi)RNA in the nematode *Caenorhabditis (C.) elegans* in 1993 by Victor Ambros and coworkers. Ambros discovered that the *lin-4* gene produces small RNAs of 22 and 61 nt from a longer non-protein-coding precursor. The longer RNA forms a stem-loop structure, which is cut to generate the shorter RNA with antisense complementarity to the 3'-untranslated region (UTR) of the *lin-14* transcript [19]. The *lin-4* RNA pairing to *lin-14* mRNA was proposed as a molecular mechanism of “posttranscriptional gene silencing”, thus decreasing LIN-14 protein levels at first larval stages of nematode development [20]. Michael Wassenegger observed a similar phenomenon occurs in plants which he described as “homology-dependent gene silencing” or “transcriptional gene silencing”; this process is mediated by the incorporation of viroid RNA which induces the methylation of the viroid cDNA and gene silencing [21]. Ultimately the entire process of RNA-mediated gene silencing was elucidated in 1998 by Andrew Fire and Craig Mello in similar experiments with the *unc-22* gene of *C. elegans*.

In 2000, another essential miRNA was identified in *C. elegans*. This miRNA, *let-7*, was shown to have homologues in several other organisms, including humans [22, 23]. The biogenesis as well as the molecular mechanisms of miRNA-mediated gene silencing has been extensively characterized. In 2001, Thomas Tuschl showed that, in *C. elegans*, long double-stranded RNA is processed into shorter fragments of 21–25 nts. Since this discovery, it has been demonstrated that premature transcripts in the nucleus are processed into hairpin-structured RNA by the Drosha-containing microprocessor complex and then exported to the cytoplasm where they are cleaved into a double-stranded RNA by Dicer. One of the strands of this double-stranded RNA is loaded to the RISC complex and then targeted to an mRNA molecule by complementarity, thus inducing translational repression [23]. This simplified scheme constitutes the mechanistic basis of RNA interference (RNAi) and presently unites all gene silencing phenomena at transcriptional and posttranscriptional levels, mediated by small ncRNAs including miRNAs, small interfering (si)RNAs, and Piwi-interacting (pi)RNAs, all of which are processed from double-stranded RNA precursors [24, 25].

Although the focus on RNAi resulted in a breakthrough for modern biology and biotechnology, as well as provided a deeper understanding of gene regulation, development, and disease, the relevance of lncRNAs remained largely unexplored. Nevertheless, some lncRNAs were investigated in the late 1980s such as *H19* and *Xist*, the milestones of dosage compensation in mammals.

1.2 lncRNA Discovery in the Pre-genomic Era

In the 1980s, scientists were using differential hybridization screens of cDNA libraries to clone and study genes with tissue-specific and temporal patterns of expression. Initially, efforts were focused on genes producing known proteins;

subsequently, an *a posteriori* approach was adopted without regard to the coding potential of RNA. Through this approach, the first non-coding gene was discovered, *H19*, even though at that time it was first classified as an mRNA [26].

1.2.1 *H19: The Very First Eukaryote lncRNA Gene*

In the late 1980s, elegant genetic and molecular studies discovered a phenomenon of genomic imprinting or parent-of-origin-specific expression which constitutes part of the dosage compensation mechanisms. Independently, two imprinted genes were identified: the paternally expressed protein-coding *Igf2r* and the maternally expressed *H19*. Both genes were localized to mouse chromosome 7 in proximity to each other forming the *H19/IGF2* cluster [27, 28]. What made *H19* unusual was the absence of translation even though the gene contained small open reading frames. *H19* showed high sequence conservation across mammals, and the abundant transcript presented features of mRNAs: transcribed by RNA polymerase II, spliced, 3' polyadenylated, and localized to the cytoplasm [29]. The expression of *H19* in transgenic mice revealed to be lethal in prenatal stages, suggesting not only that the dosage of this lncRNA is tightly controlled but that it has an important role in embryonic development. However, the function of *H19* as an RNA molecule in its own right remained a mystery until the functional characterization of another lncRNA involved in dosage compensation in mammals, *Xist*. Since that time, *H19* has been thoroughly investigated and represents the prototype of a multitasking lncRNA.

1.2.2 *X Inactivation: Existence of Xist*

In living organisms, sex can be determined by many ways; it is defined in mammals by the X and Y chromosomes, while males only have one X and Y chromosome, females have two X chromosomes in their karyotype. However, the X chromosome carries many genes, most of which have functions that are not involved in sex determination. Hence, there is a need for dosage compensation between males and females. Although the mechanism of choice in *Drosophila* is to double the transcription of the single X chromosome in males, it is the opposite in mammals: one of the female X chromosomes is inactivated. This phenomenon, called X-chromosome inactivation (XCI), was first discovered in mouse by Mary Lyon in 1961 [30] and further generalized to other mammals. XCI is established early in development and is initiated by a unique locus, the X-inactivation center (*Xic*).

In the early 1990s, this locus was found to produce a long non-coding RNA, *Xist* (X-inactive-specific transcript). It is expressed at very low levels in both, male and female, mouse undifferentiated embryonic stem (ES) cells. Upon differentiation *Xist* expression is activated in a monoallelic way in female cells, from

the future inactive X (Xi) to initiate the onset of random XCI. Being retained in the nucleus, *Xist* triggers gene silencing in *cis* by physically localizing and spreading broadly on the future Xi [31–33]. In contrast to *H19* and other lncRNAs involved in dosage compensation, *Xist* is highly unusual since it triggers the silencing of the entire chromosome. The propagation of *Xist* along the Xi, called “coating,” implicates the RNA wrapping around the X and the recruitment of multiple factors, including the polycomb repressive complexes 1 and 2 (PRC1 and PRC2). This triggers a cascade of chromatin changes and a global spatial reorganization of the Xi and, ultimately, the stable repression of nearly all Xi-linked genes throughout development and adult life [34]. While *Xist* expression is critical for the initiation of the XCI, in somatic cells, *Xist* and the whole *Xic* were shown to be dispensable for the maintenance of silencing in mouse [35]. In 1999 human *XIST*, ectopically expressed from the artificially inserted transgene on mouse autosomes, was demonstrated to function as *Xic* and to initiate XCI even in undifferentiated mouse ES cells, unlike the mouse counterpart. This result suggested differences in the developmental regulation of *Xist* and in the initiation of the XCI process between mouse and humans. In addition, the inactivation by ectopic human *XIST* was observed only in a portion of mouse male ES cells, thus confirming that *Xist* is not a unique actor of stable inactivation [36]. Indeed, the *Xic* was initially defined in mouse as the minimal region of the X chromosome that contains all sequences both necessary and sufficient for the initiation of XCI. *Xic* extends over 1 Mb, and transcriptomic studies revealed that this region contains several protein-coding and non-coding genes, including *Linx*, *Ftx*, and others. Remarkably, some non-coding genes within *Xic* and beyond show poor primary sequence conservation between human and mouse, and this includes the sequence of *Xist* itself [37]. In particular, the *Tsix* lncRNA is an antisense transcript which overlaps the whole *Xist* gene and its promoter in mouse. In humans, key regulatory elements are truncated, and the transcript overlaps *XIST* only at 3' end. These differences abolish the *TSIX* function in transcriptional repression of *XIST* on the future active X in humans [38, 39]. Recently, another lncRNA, *XACT*, was discovered in human ES cells. This gene is located within the intergenic region, outside of *Xic*, and it is not conserved in mice. In female human ES cells, *XACT* is expressed from and coats both X chromosomes. This lncRNA seems to be specific to pluripotent cells and is proposed to ensure peculiar control of XCI in humans [40]. The biogenesis of *Xist*, its structure, and the molecular mechanism of XCI have been the focus of many studies in different mammals and extensively documented in other publications [34].

The pioneering studies of *H19* and *Xist* revolutionized our view of non-protein-coding gene functions and on the biological relevance of lncRNAs in general. These examples demonstrated the complexity and versatility of regulatory circuits orchestrated by a single lncRNA. They also stimulated the discovery and suggested potential mechanisms for other, yet uncharacterized, non-coding transcripts. A global effort toward lncRNA identification and characterization began in the 2000s, as a plethora of novel non-coding transcripts during the sequencing of the complete human genome.

1.3 From Non-coding Genome to Non-coding Transcriptome: The Genomic Era

Our modern view of eukaryotic transcriptomes was preceded by comprehensive investigations of genomic DNA and the discovery that, in addition to protein-coding (PC) sequences and regulatory elements essential for PC gene (PCG) transcription, the majority of the genome contains sequences that were considered to be useless evolutionary fossils. To differentiate these sequences from PC sequences, this DNA was named non-coding and referred to as selfish or junk DNA for almost 20 years [41].

1.3.1 *The Human Genome Project: Genomic DNA Is Mostly Non-coding*

In 1978, using the sequencing technique he had developed, Frederick Sanger generated the first ever full genomic sequence: the viral genome of the bacteriophage ϕ X174 [42]. Since then, Sanger sequencing has been routinely used worldwide, and its discovery and development earned the Nobel Prize in Chemistry for Sanger, along with Walter Gilbert. During the following years, several viral genomes were sequenced and by the end of 1990, a worldwide sequencing effort, the Human Genome Project (HGP), was established by the National Institute of Health (NIH, USA) to completely sequence the human genome. In parallel, the American biochemist and entrepreneur Craig Venter founded his own company and sought private funding to achieve the same goal. This put pressure on the public groups involved in the HGP, and the race to unravel the human genome began. The first bacterial genome was published in 1995 [43]. It was followed in 1999 by the sequence of the euchromatic portion of human chromosome 22 [44], which covered approximately 65% of what is now known to be the full chromosome 22. This sequence was thought to contain 545 protein-coding genes (whether known or predicted), with PC exons spanning a mere 3% of the full sequence.

Finally, using clone-by-clone methodology, the first draft of the complete human genome was published in *Nature* in 2001 covering 96% of the euchromatin [45], followed the next day by Craig Venter's publication in *Science* of the whole-genome sequence obtained by the shotgun-cloning method [46]. Regular updates completed the human genome sequence in 2003. In the meantime, the genomes of several other organisms had already been released, notably yeast [47], pufferfish [48], worm [49], fruit fly [50], and mouse [51], thus allowing comparative studies to be performed.

The first surprise from this comprehensive genomic sequencing effort was the rather low number of PCGs compared to what was initially expected. Indeed, early studies that looked at the repartition of CpG islands predicted 70,000–80,000 genes in the human genome [52], a figure close to the well-admitted 100,000 genes from the mid-1980s. However, the HGP predicted around 31,000 PCGs in 2001 reduced

to 22,287 PCGs in 2004 [45, 53]. In general, only 1.2% of the human genome represents PC exons, whereas 24 and 75% were attributed to intronic and intergenic non-coding DNA.

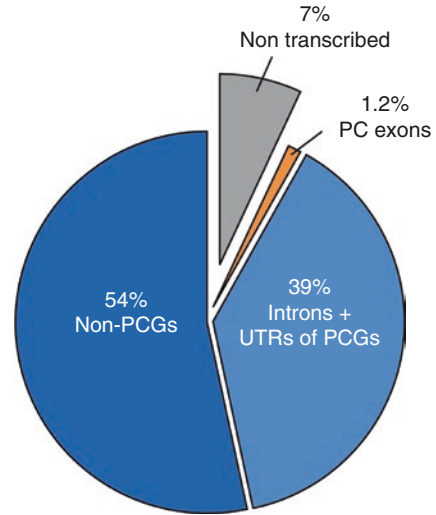
1.3.2 *Pervasive Transcription and the Dark Matter of the Genome*

The HGP also revealed that most of the genome is actually transcribed, whether it encodes proteins or not. Indeed, a tiling array with oligonucleotide probes spanning human chromosomes 21 and 22 revealed that 90% of detected cytosolic polyadenylated transcripts map to non-coding genomic regions and not to exons [54]. Similar results were found by the FANTOM and RIKEN consortia when analyzing the transcriptome in both human [55] and mouse [56]. They sequenced more than 60,000 full-length cDNAs from mouse in a standardized manner to generate accurate maps of the 5' and 3' boundaries of all transcripts, thus defining transcription start (TSS) and termination (TTS) sites. Remarkably, cap analysis gene expression (CAGE) sequencing, a technique that sequences 5' ends of capped transcripts, revealed over 23,000 ncRNAs originating from both sense and antisense transcription representing approximately two thirds of the mouse genome [57]. For the first time, antisense transcription was proposed to contribute to the regulation of gene expression at transcriptional level in mammals.

These results were later confirmed by even larger-scale studies conducted in humans by the ENCODE (Encyclopedia of DNA Elements) consortium. This project compiled over 200 experiments in its pilot phase [58] and up to 1640 datasets from 147 different cell lines in its later release [59]. Through various sequencing techniques, landscapes of DNase I hypersensitive sites, histone modifications, transcription factor binding sites, and the whole transcriptome were defined. Conclusions from these studies estimated that 93% of the human genome is actively transcribed and associated with at least one primary transcript (i.e., coding and non-coding exons and introns); among these transcripts, approximately 39% of the genome represented PCGs (from promoter to poly(A) signal) and 1% protein-coding exons, while the other 54% mapped outside of PCGs (Fig. 1.3). However, many lncRNAs overlap with PCG annotations in both sense, coding and antisense strands. More recently, the mouse counterpart of the ENCODE Consortium confirmed previous reports by publishing a similar analysis which showed that 46% of the mouse genome produces mRNAs while at least 87% of its genome is transcribed [60, 61].

Many studies aiming to characterize non-coding transcription were also performed in other eukaryotes, including *Saccharomyces cerevisiae*. Even in this primitive unicellular eukaryote, about 85% of the genome is transcribed [62]. This phenomenon is often referred to as “pervasive transcription” and is widespread among eukaryotes. An expanding body of literature details its function [63, 64]. The identification and characterization of non-coding transcripts as unique ncRNAs extended the former definition of a “gene” beyond its coding function. Furthermore,

Fig. 1.3 Proportion of transcribed protein-coding and non-coding sequences (introns, UTRs, and others) in the human genome according to ENCODE [59]



the discovery of the non-coding genome and transcriptome gave rise to heated debates in the scientific community concerning the biological significance and functional relevance of these non-coding DNA and RNA, still perceived as a junk [63, 65, 66]. These debates challenged the Central Dogma of Watson and Crick, promoting ncRNAs to the epicenter of the cellular processes as a driver of biological complexity through evolution.

1.4 Non-coding RNAs: Junk or Functional

Polemics around the biological and functional relevance of lncRNAs were oriented toward understanding the origin, conservation, and diversification of lncRNA species across evolution.

1.4.1 Origin of lncRNA Genes

Non-coding genes were proposed to arise through various mechanisms including DNA-based or RNA-based duplications of existing genomic sequences, the metamorphosis of PCGs by loss of protein-coding potential, transposable element exaptation, or non-coding DNA exaptation [67]. Homologous non-coding genes arise from duplications of already existing lncRNA genes. Pseudogenes are an example of PCG metamorphosis during which a duplicated ancestral open reading frame had accumulated disruptions destroying its potential to be translated. Once transcribed, pseudogenes often produce lncRNAs, as in the case of *PTENP1*.

Pseudogenization of a PCG, due to mutations deleterious to translation, can also produce lncRNA genes that do not have an apparent protein-coding “homologue”. An example is *Xist* which is derived from an ancestral *Lnx3* gene and which has acquired several frame-shifting mutations during early evolution of placental mammals [68]. Exaptation or co-option of RNA-derived transposable elements (TE) into non-coding genes is another frequent mechanism of lncRNA origination. In humans TEs constitute a large portion of the genome (40–45%) [45]. Most of them are genomic remnants that are currently defunct but are often embedded into non-coding transcripts. TEs are considered as major contributors to the origin and diversification of lncRNAs in vertebrates [69]. Together with local repeats, they provide lncRNA genes with TSS, splicing, polyadenylation, RNA editing, RNA binding sites, nuclear retention signals or particular secondary structures for protein binding [70–72].

Finally, pervasive transcription of the genome may generate cryptic RNAs that, if maintained through evolution, can give rise to lncRNA genes with novel functions. In particular, exaptation of non-coding sequences into lncRNAs can occur through the acquisition of regulatory elements within a silent region, thereby promoting transcription. However, the *de novo* origin of lncRNAs remains difficult to prove and is represented by few examples, such as the testis-specific lncRNA *Poldi* [73]. Interestingly in humans, the testis and cerebral cortex are the most enriched tissues for the expression of PCGs and non-coding genes of *de novo* origin. This particularity was suggested to contribute to phenotypic traits that are unique to humans, such as an improved cognitive ability [74, 75].

1.4.2 Evolutionary Conservation of lncRNAs

Genomic and transcriptomic studies across the eukaryotic kingdom allowed the analysis of the primary sequence conservation of protein-coding and non-coding loci. These studies revealed that the human genome is highly dynamic, and only 2.2% of its DNA sequence is subjected to conservation constraints [76]. Remarkably, non-coding genes are among the least conserved with more than 80% of lncRNA families being of primate origin [77]. This finding raised skepticism regarding the functionality and biological relevance of lncRNAs and initiated a search for other conservation constraints [78, 79]. If the criterion of primary sequence conservation is too restrictive in regard to lncRNA genes, other features such as structure, function, and expression from syntenic loci constitute multidimensional factors that are more applicable for evolutionary studies of lncRNAs [80]. Recently, a study looking at the non-coding transcriptome of 17 different species (16 vertebrates and the sea urchin) showed that although the body of non-coding genes tends not to be conserved, short patches of conserved sequences could be found at their 5' ends. This confirmed a higher conservation of TSS and synteny, as well as expression patterns in different tissues, especially in those involved in development [81]. Indeed, the most conserved are developmentally regulated lncRNAs of the lincRNAs subfamily.

These lncRNAs have a remarkably strong conservation of spatiotemporal and syntenic loci expression, suggesting that it is selectively maintained and crucial for developmental processes [77, 82, 83].

1.4.3 Role of lncRNAs in Biological Diversity

The identification of new lncRNAs in the last decade continues to increase and, as anticipated in the past, largely exceeds that of protein-coding transcripts. The diversity of the non-coding transcriptome is considered as an argument to explain the remarkable phenotypic differences observed among species given a relatively similar numbers of protein-coding genes among fruit fly (13,985; BDGP release 4), nematode worm (21,009; Wormbase release 150), and human (23,341; NCBI release 36) [84]. In 2001, John Mattick and Michael Gagen proposed, for the very first time, that non-coding transcripts named “efference” RNA, together with introns, constitute an endogenous network enabling dynamic gene-gene communications and the multitasking of eukaryotic genomes. In contrast to core proteomic circuits, this higher-order regulatory system is based on RNA and operates through RNA-DNA, RNA-RNA, and RNA-protein interactions to promote the evolution of developmentally sophisticated multicellular organisms and the rapid expansion of phenotypic complexity. A direct correlation between the portion of non-coding sequences in the genome and organism complexity was hypothesized [85, 86]. Interestingly comparative genomics allowed the identification of a few regions in the human genome that have high divergence when compared to other species [87, 88]. These human accelerated regions (HAR) contain many lncRNA genes and have been suggested to be involved in the acquisition of human-specific traits during evolution. In 2006, a first lncRNA from these regions was shown to be expressed during cortical brain development [89]. Since then, many mutations involved in diseases were identified in these non-coding regions and shown to be associated with regulatory elements in the brain [90]. A more recent study showed that mutations of HAR enhancer elements could be involved in the development of autism, thus supporting the hypothesis that some HAR could be involved in human-specific behavioral traits and cognitive or social disorders when mutated [91]. However, the functionality of non-coding transcripts was and still remains hotly debated. Nevertheless, the conception of developmental and evolutionary significance has stimulated an exhaustive molecular characterization of lncRNA genes and transcripts.

1.5 The General Portrait of lncRNA Genes and Transcripts

lncRNAs have been identified in all species which have been studied at the genomic level, including animals, plants, fungi, prokaryotes, and even viruses. Genome-wide studies continue to enlarge the catalogue of lncRNAs continuously

reshaping the specific features of lncRNAs as transcription units. Here, we will summarize the main features of lncRNAs that distinguish them from mRNAs (Table 1.1).

Table 1.1 Comparison of lncRNA and mRNA features

Feature	lncRNA	mRNA
Transcription		
	RNA polymerase II	RNA polymerase II
	RNA polymerase III (<i>B2-SINE</i> ; <i>NDM29</i> [101, 102])	
	RNA polymerases IV and V (plants, [103])	
Chromatin modifications		
H3K4me3	Low (eRNA, PROMPTs)	High
	High (others)	
H3K4me1	High (eRNA, PROMPTs)	Low
	Low (others)	
H3K27ac	High	Low
H3K36me	Moderate/high	High
H3K79me2	Enriched (bidirectional lncRNAs)	Low
H3K27me3	Present at bivalent and repressed promoters	Present at bivalent and repressed promoters
5'-Cap	Present (7-methylguanosine, m7G)	Present (m7G)
Poly(A) tail	Present or not	Present
	Bimorphic	
Length	200—>100 kb (10 kb mean)	5 kb mean
Exon-intron composition	Yes	Yes
	Exons are longer	
Splicing	Yes or less efficient	Yes
	No (macro lncRNA, vlincRNAs)	
RNA stability	Variable, globally lower than mRNA	Variable
	Highly unstable (eRNA, XUTs, CUTs, PROMPTs)	
Evolutionary conservation	High (lincRNAs)	High
	Low or not conserved (others)	
Protein-coding potential	Non or very low (sORFs)	Yes
Structure	Versatile, multi-modular	Kozak hairpin at the 5' end
Subcellular localization	Nucleus	Cytosol
	Cytosol	
	Mitochondria	
Expression specificity	high, including interindividual variability of expression	low to high
Transcript abundance	Very low or low	Moderate to high
	High (for few)	

1.5.1 Coding Potential of lncRNA Genes

As dictated by the acronym, lncRNA genes do not encode proteins. Cytosol-localized lncRNAs were found associated with mono- or polyribosomal complexes [92], but this association is not necessarily linked to translation but rather proposed to determine lncRNA decay [93, 94]. Some lncRNAs include short open reading frames (sORFs) and undergo translation, though only a minority of such translation events results in stable and functional peptides [95, 96]. This is the case of *DWORF*, a muscle-specific lncRNA that encodes a functional peptide of 34 amino acids [97–100]. Proteomic studies will undoubtedly introduce a new “coding” aspect to lncRNAs, expanding our conception of “coding” and leading to a possible concept of bifunctionality.

1.5.2 lncRNA Transcription and Transcript Organization

The majority of eukaryotic lncRNAs are produced by RNA polymerase II, with some exceptions, for example, the murine heat-shock induced B2-SINE RNAs [101] or the human neuroblastoma associated *NDM29* [102], which are synthesized by RNA polymerase III. However, the last two examples are not strictly considered as lncRNAs because the transcript length is below the arbitrary threshold of 200 nts. In plants, two specialized RNA polymerases, Pol IV and Pol V, transcribe some lncRNA genes [103]. Many lncRNAs are capped at the 5' end, except those processed from longer precursors (intronic lncRNAs or circRNAs). However, some ambiguities exist concerning the presence of a cap, especially for highly unstable and low-abundant transcripts, since they can't be captured by the CAGE-seq technique. lncRNAs may or may not be 3'-end polyadenylated; in addition, they may also be present as both forms, such as bimorphic transcripts like *NEATI* and *MALATI* [104, 105]. lncRNAs with a polyadenylation signal have higher stability than those that are poorly or not polyadenylated, with the exception of lncRNAs bearing specific 3'-end structures as in case of *MALATI* [106]. Of note, poly(A)+ transcriptomic studies exclude the possibility of discovery of non-polyadenylated transcripts and introduce a quantitative bias in the identification of such lncRNAs. This point should be taken into account in comparative studies or in selection of RNA-seq strategies, favoring the use of total RNAs instead of the more customary used poly(A)+ RNA fraction.

Similar to PCGs, transcription of many lncRNA genes requires canonical factors assisting the RNA polymerase machinery such as the pre-initiation complex (PIC), Mediator, transcription elongation complex, and also specific transcription factors that in turn could define the specificity of lncRNA expression in different biological contexts. However, some particularities in lncRNA promoters have been demonstrated. In humans lncRNA promoters are more enriched in A/T mono-, di-, and trinucleotide stretches and are characterized by reduced CG and almost depleted AT

skews (CG and AT compositional strand biases); this is contrary to PCGs suggesting a distinct regulation of transcription for these two groups of genes [107]. Promoters of PROMPTs are devoid of transcription initiation factors such as TAFI, TAFII, p250, and E2F1 and are believed to initiate transcription without the use of conventional PIC [108]. eRNAs require the Integrator complex for the 3'-end cleavage of primary transcripts [109], and lncRNA precursors of small ncRNAs were shown to be processed by specific endonucleases [110, 111]. Some unstable lncRNAs such as yeast NUTs and CUTs are terminated by the Nrd1-dependent pathway, thus targeting them for rapid degradation by the exosome [112–114].

lncRNA genes can have a multi-exonic composition with similar splicing signals as PCGs and therefore could undergo splicing into several different isoforms with distinct functional outcomes and clinical relevance [115–117]. However, they usually comprise fewer and slightly longer exons than PCGs [118, 119].

1.5.3 Chromatin Signatures of lncRNAs Genes

As RNA polymerase II transcribes most of the lncRNA genes, their genomic regions present a chromatin organization resembling that of PCGs, with some differences. This could be due to the globally low expression of lncRNAs, which is a consequence of either low rate of transcription, lower stability, or both. Globally, lncRNA TSS reside within the DNase I hypersensitive sites suggesting nucleosome depletion from this region. lncRNA promoters have lower levels of histone H3K4 trimethylation (H3K4me3), which is in accordance with their low transcription rate. eRNAs and PROMPTs present high levels of histone H3K4 monomethylation (H3K4me1) and K27 acetylation (H3K27ac) at promoters, which is considered as a specific signature of enhancer- and promoter-associated unstable transcripts; these signatures exist in the following ratios: H3K4me3 over H3K4me1 as a mark of PROMPTs and H3K4me1 over H3K4me2 as a mark of eRNAs [120]. The body of most lncRNA genes with the exception of eRNAs and PROMPTs is marked by histone H3K36 trimethylation (H3K36me3). In yeast, sense-antisense transcription was reported to be associated with particular chromatin architecture: reduced histone H2B ubiquitination, H3K36me3, and histone H3K79 trimethylation, as well as increased levels of H3ac, chromatin remodeling enzymes, histone chaperones, and histone turnover [121]. In mouse, bidirectional transcription, which is often associated with developmental genes and genes involved in transcription regulation, was found to harbor high H3K79 dimethylation (H3K79me2) and elevated RNA polymerase II levels. This signature is characteristic of intensified rates of early transcriptional elongation within a region transcribed in both directions [122].

It is anticipated that single cell studies will resolve the problem of signal variability in a population of cells, allowing transcriptional events to be directly linked to specific chromatin modifications. Such efforts have already been initiated for transcriptome profiling [123–125] but remain challenging for epigenomic studies [126].

1.5.4 Expression Pattern of lncRNAs: Stability, Specificity, and Abundance

Several genome-wide studies addressed lncRNA stability and, depending on the employed experimental approach, revealed some discrepancy for different species of lncRNAs. In mouse, the measurements of the lncRNA half-life ($t_{1/2}$) and decay rates were performed through transcription inhibition by actinomycin B treatment. In this case, lncRNAs showed a half-life range from 30 min to 48 h, which is similar to mRNAs; however, a mean $t_{1/2}$ of 4.8 versus 7.7 h for mRNAs suggests that lncRNAs possess a lower stability. A high percentage of lncRNAs was classified as unstable ($t_{1/2} < 2$ h), e.g., *Neat1*, and a few as highly stable ($t_{1/2} > 12$ h) [127]. Comparison of the stability of different lncRNA species revealed that intronic or promoter-associated lncRNAs are less stable than either intergenic, antisense, or 3' UTR-associated lncRNAs. Single-exon transcripts, a class of nuclear-localized lncRNAs, are overrepresented among unstable transcripts. In human HeLa cells, the same approach of transcriptional inhibition was used and revealed that antisense lncRNAs are more stable than mRNAs (median $t_{1/2} = 3.9$ versus 3.2 h, respectively), whereas intronic lncRNAs included both stable ($t_{1/2} > 3$ h) and unstable ($t_{1/2} < 1$ h) transcripts with the $t_{1/2}$ median of 2.1 h [128]. Recently discovered circular RNAs are examples of highly stable lncRNAs with the median $t_{1/2}$ of 18.8–23.7 h and which is at least 2.5 times longer than their linear counterparts [129].

Nuclear and cytoplasmic exosomes, cytoplasmic Xrn1, and nonsense-mediated decay (NMD), as well as RNAi pathways, are known to control lncRNA abundance in the cell. Circular RNAs are intrinsically protected from any exonucleolytic- or polyadenylation-dependent decay pathways. Of note, actinomycin D treatment has a large impact on cells, and this can particularly influence lncRNA decay because of the very high sensitivity of lncRNAs to stress. Indeed, the measurements of $t_{1/2}$ for single lncRNAs could significantly vary from one experiment to another, pointing to the necessity of multiple approaches including *de novo* RNA labeling to achieve more accurate and confident conclusions.

Multiple transcriptome profiling globally highlighted a highly specific spatio-temporal, lineage, tissue- and cell-type expression patterns for lncRNAs compared to PCGs; only a minority are ubiquitously present across all tissues or cell types, such as *TUG1* or *MALAT1* [105, 130, 131]. Curiously, the brain and testis represent a very rich source of uniquely expressed lncRNAs supporting the hypothesis that such transcripts are important for the acquisition of specific phenotypic traits [82, 130]. The ubiquitously expressed lncRNAs are often highly abundant, whereas specific lncRNAs present in one tissue or cell type tend to be expressed at low levels [132]. Moreover, interindividual expression analysis in normal human primary granulocytes revealed increased variability in lncRNA abundance compared to mRNAs [133]. Some disease-associated single-nucleotide polymorphisms (SNPs) within lncRNA genes and their promoters were linked to altered lncRNA expression, thus supporting their functional relevance in pathologies [134].

The high specificity of lncRNA expression argues in favor of important regulatory roles that these molecules can play in different biological contexts, including normal and pathological development.

1.5.5 Subcellular Localization of lncRNAs

Globally, unlike mRNAs, many lncRNAs have nuclear residence with focal or dispersed localization pattern (*NEATI*) [135]. However, others were also found both in the nucleus and in the cytosol (*TUG1*, *HOTAIR*) or in the cytosol exclusively (*DANCR*) [105]. Multiple determinants, such as a specific RNA motif (*BORG*) [136] or RNA-protein assemblies, may dictate the subcellular localization of lncRNAs and define their function [137]. Remarkably, environmental changes or infection can induce lncRNA delocalization (or active trafficking) from one cellular compartment to another, as in the case of stress-induced lncRNAs [138]. HuR and GRSF1 modulate nuclear export and mitochondrial localization of the nuclear-encoded *RMRP* lncRNA [139].

1.5.6 Structure of lncRNAs

RNA is a highly flexible and dynamic molecule that adopts complex secondary structures. The folding of lncRNAs defines their cellular decay and functional versatility, enabling their nuclear localization, stability, and interaction with proteins [140]. A growing number of examples demonstrate that the RNA secondary structure constitutes the primary functional unit and evolutionary constraint bypassing poor interspecies lncRNA sequence conservation [141]. One such example is the lncRNA *HOTAIR* which exists only in mammals, sharing 58% of homology between human and mouse [142, 143]. Covariance analysis across 33 mammalian sequences of *HOTAIR* revealed a significant number of covariant base pairs and half-flips, which maintained a similar structure regardless of the changed sequence; this was especially true in regions surrounding proposed protein-binding segments of the lncRNA [144]. On the other hand, low sequence conservation that induces changes in structure can drive acquisition of new functions and specialization of the lncRNA-mediated regulatory circuit. This is the case of human accelerated region 1 (*HAR1*)-derived lncRNAs expressed in developing neocortex in primates where the capacity to form a stable cloverleaf-like structure has arisen only in humans [89, 145]. However, we are still far from understanding the function of this lncRNA in human brain development.

Numerous structure prediction tools, such as Rfold, have been developed to give guidance for further functional studies. Structural analysis of RNA has increased our understanding of mechanistic aspects of lncRNA action; however, X-ray crystallography, nuclear magnetic resonance (NMR), and cryo-electron microscopy

require purified and stable, nearly static, molecules and are not adapted to highly dynamic and flexible RNA. Very recently, new technologies based on high-throughput sequencing have evolved enabling both an *in vitro* and *in vivo* view of RNA conformation [140].

1.6 Classification of lncRNAs

Advances in deep sequencing technologies gave rise to a plethora of novel transcripts requiring a universal standardized system for lncRNA classification and functional annotation. The state of lncRNA annotations is still at its beginning, and different classifications based on their length, transcript properties, location in respect to known genomic annotations, regulatory elements, and function have been proposed. Here, we review a non-exhaustive cataloguing of eukaryotic lncRNAs summarized in Table 1.2.

Table 1.2 Classification of lncRNAs (adapted from [279])

lncRNA category	Abbreviation	Examples	Refs
Classification according to lncRNA length			
Long non-coding RNA	lncRNA		[105, 118]
Large non-coding RNA			
Very long intergenic ncRNA	vlinRNA		[280]
Macro lncRNA		<i>Xist, Airn, Kcnq1ot1</i>	[148]
Classification according to lncRNA location with respect to PCGs			
<i>Intergenic lncRNAs</i>			
Long intergenic ncRNA	lincRNA	<i>MALAT1, NEAT1, GAS5, CYRANO</i>	[83, 258, 281, 282]
Large intervening ncRNA		<i>Frigidair, lincRNA-COX2, XACT</i>	[40, 150]
Long intervening ncRNA			[283]
<i>Antisense lncRNAs</i>			
Natural antisense transcripts	NAT	<i>ZEB2NAT</i>	
	<i>cis</i> -NAT	<i>BACE1-AS</i>	
	<i>trans</i> -NAT		
Antisense lncRNA	asRNA		
	ancRNA		
SINE B2 containing RNAs	SINEUP	<i>AS-Uchl1</i>	[160]
Bidirectional lncRNAs	Long upstream antisense transcript	LUAT	[122]
	Upstream antisense RNA	uaRNA	[170]

<i>Intronic lncRNAs</i>			
Totally intronic RNA	TIN		[178]
Circular intronic RNA	ciRNA	<i>ci-ankrd52</i>	[175]
Circular RNA	circRNA	<i>CDR1as/ciRS-7, cANRIL</i>	[182, 184, 285, 286]
Exonic circRNA	ecircRNA		
Exon-intron circRNA	EIcircRNA		
Stable intronic sequence RNA	sisRNA	<i>sisR-1</i>	[287–289]
Switch RNAs			[177]
Overlapping sense transcripts	Sense ncRNA	<i>HLXB9-lncRNA, SOX2-OT</i>	[153, 181]
Classification according to lncRNA residence within specific DNA regulatory elements and loci			
Pseudogenes		<i>PTEPN1, Lethe</i>	[193, 265, 290]
Telomeres and subtelomeres		<i>TERRA</i>	[203]
		<i>subTERRA</i>	[206–209, 291]
Centromeres		<i>centromeric alpha-satellite RNA</i>	[210–212, 214]
Transcripts from ultraconserved regions	T-UCR	<i>Uc.283+A</i>	[195, 196, 200, 292]
		<i>Eyf2</i>	
rDNA loci		<i>PAPAS</i>	[215]
Promoter-associated ncRNAs	pancRNA	<i>CCND1-lncRNA</i>	[218, 223]
	PALRs		
Promoter upstream transcripts	PROMPT		[108]
Upstream antisense RNA	uaRNA ^a		[170]
Enhancer lncRNA	eRNA	<i>IL1β-eRNA, FOXC1e</i>	[216, 293]
3'-UTR-associated RNAs	uaRNA ^a		[224]
Classification according to lncRNA biogenesis pathway			
Stable unannotated transcript	SUT		[163]
Cryptic unstable transcripts	CUT	<i>PHO84 CUT, PROMPT, eRNAs</i>	[163, 225, 228], [219]
Cytoplasmically degraded CUTs	CD-CUT		[229]
Meiotic unannotated transcripts	MUT		[230, 231]
Nrd1-terminated transcripts	NUT		[113]
Xrn1-sensitive unannotated transcripts	XUT	<i>RTL, XUT1678 (ARG1-AS)</i>	[94, 226, 227]
Classification according to lncRNA subcellular localization or origin			
Nuclear lncRNAs		<i>NEATI</i>	[128]
GAA repeat-containing RNAs	GRC-RNA		[240]
Chromatin-enriched RNAs	cheRNA		[237]
Chromatin-associated RNAs	CAR		[238]
Mitochondrial ncRNAs	mtncRNA	<i>ASncmtRNA-2</i>	[243, 294]

(continued)

Table 1.2 (continued)

Classification according to lncRNA function			
Scaffold lncRNA		<i>HOTAIR, LINP1, NORAD</i>	[251, 252, 295]
<i>Architectural lncRNAs</i>	arcRNAs	<i>NEAT1</i>	[254]
Guide lncRNA		<i>MEG3, Khps1</i>	[256, 257]
Ribo-activator	ncRNA-a	<i>SRA, Lnc-DC, NeST</i>	[259, 260], [262, 263]
	eRNAs	<i>FOXC1e</i>	[217]
Ribo-repressor or decoy		<i>GAS5, CCND1-lncRNA, PANDA, Lethe</i>	[168, 223, 258, 290]
Competing endogenous RNA	ceRNAs	<i>PTENP1, HULC, CDR1as/ciRS-7</i>	[182, 265, 267]
lncRNA precursors	endo-si-lncRNA	<i>H19 (miR-675)</i>	[110, 268, 270, 271]
	pi-lncRNA	<i>MALAT1 (mascRNA)</i>	
	mi-lncRNA	<i>P5CDH-SRO5</i>	
Classification according to lncRNA association with specific biological processes			
Hypoxia induced	HINCUT		[197, 275]
Senescence-associated lncRNAs	SAL	<i>vlinRNAs, SALNR</i>	[147, 274]
Stress-induced lncRNAs	si-lncRNA		[138]
Non-annotated stem transcripts	NAST		[276]
Prostate cancer-associated transcripts	PCATs	<i>PCA3, PCAT1</i>	[277]

^aIn the literature, the term “uaRNA” has been attributed to two distinct groups of transcripts: upstream antisense RNAs and UTR-associated RNAs

1.6.1 Classification According to lncRNA Length

By convention, a length of 200 nt constitutes a bottom line for discrimination of long or large ncRNAs from small or short ncRNAs. However, lncRNAs vary significantly in size, and those that exceed the length of 10 kb belong to the groups of **very long intergenic (vlinc)RNAs** and **macro lncRNAs**. These transcripts possess some particular features that distinguish them from other lncRNAs: they are poorly or not spliced, weakly polyadenylated at 3' end, and are produced by particular genomic loci. The majority of vlincRNAs are localized in close proximity or within PCG promoters on the same or opposite strand and function in *cis* as positive regulators of nearby gene transcription. Interestingly, some vlincRNA promoters harbor LTR sequences that are highly regulated by three major pluripotency-associated transcription factors, suggesting a possible role in early embryonic development [146]. Others are specifically induced by senescence and are required for the maintenance of senescent features that in turn control the transcriptional response to environmental changes [147]. Macro lncRNAs are often antisense to PCGs and are produced from imprinted clusters in a parent-of-origin-specific manner. Macro lncRNAs silence nearby imprinted genes either through their lncRNA product triggering epigenetic chromatin modifications or by a transcriptional interference mechanism [148].

1.6.2 Classification According to lncRNA Location with Respect to PCGs

This attribute is commonly used by the GENCODE/Ensembl portal in transcript biotype annotations, but also employed on an individual scale by consortia and laboratories for newly assembled lncRNA transcripts. Initially transcripts are classified as either intergenic or intragenic (Fig. 1.4). **Long or large intergenic non-coding (linc)RNAs** do not intersect with any protein-coding and ncRNA gene annotations. This category also includes the adopted GENCODE and homonymous biotype of long or large intervening ncRNAs that were originally defined by specific histone H3K4-K36 chromatin signatures within evolutionary conserved genomic loci [149, 150]. LincRNAs are usually shorter than PCGs, transcribed by RNA polymerase II, 5' capped, 3' polyadenylated, and spliced. Although several highly conserved lincRNAs exist, the majority possess modest sequence conservation comprising short, 5'-biased patches of conserved sequence nested in exons [81]. Highly conserved lincRNAs are believed to contribute to biological processes that are common to many lineages, such as embryonic development [77], while others are proposed to assure phenotypic and functional variations at individual and interspecies levels. Many, if not most, lincRNAs are localized in the nucleus where they exercise their regulatory functions. One such example is *lincRNA-p21* which is induced by p53 upon DNA damage [151]. *lincRNA-p21* physically associates with and recruits the nuclear factor hnRNP-K to specific promoters mediating p53-dependent transcriptional responses.

Intragenic lncRNAs overlap with PCG annotations and can be further classified into antisense, bidirectional, intronic, and overlapping sense lncRNAs.

Antisense lncRNAs, asRNAs or ancRNAs, were first discovered in single gene studies, but the recent development of stranded tiling and RNA-seq technologies has identified them as a common genome-wide feature of eukaryotic transcriptomes [152–154]. This group encompasses so-called natural antisense transcripts, NATs, which are in turn subdivided into *cis*-NATs, which affect the expression of the corresponding sense transcripts and into *trans*-NATs, which regulate expression of non-paired genes from other genomic locations [155–157]. A very recent

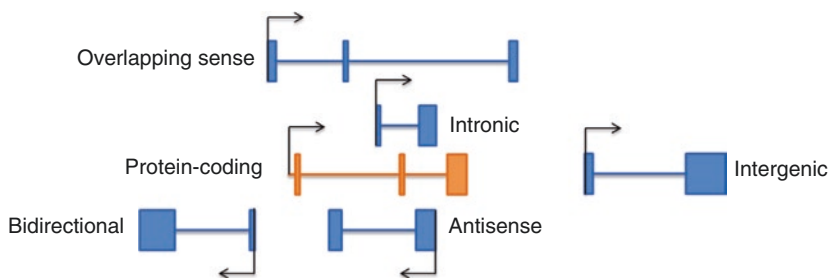


Fig. 1.4 Annotation of non-coding transcripts according to their genomic position relative to a protein-coding gene (orange box, protein-coding exon; blue box, non-coding exon)

study has pointed to a higher specificity of expression and an increased stability of asRNAs compared to lincRNAs and sense intragenic lincRNAs [128]. Due to sequence complementarity to sense-paired mRNAs or pre-mRNAs, asRNAs can act through RNA-RNA pairing, thereby ensuring specific targeting of the asRNA regulatory activity. This is the case of *BACE1-AS* that is highly expressed in Alzheimer's disease patients. It stabilizes the *BACE1* mRNA resulting in an increased expression of the *BACE1*-encoded beta-secretase and the accumulation of amyloid-beta peptides in the brain [158]. Antisense transcription across intron regions has been shown to regulate the local chromatin organization and environment, thus affecting co-transcriptional splicing of sense-paired pre-mRNAs [159]. Some NATs contain the inverted short interspersed nuclear element B2 (SINEB2), such as *AS-Uchl1* [160]. These NATs, called **SINEUPS**, are able to stimulate sense mRNA translation through lincRNA-mRNA pairing thanks to a complementary 5' overlapping sequence to the paired-sense protein-coding gene. Recently, SINEUPS were proposed as a synthetic reagent for biotechnological applications and in therapy of haploinsufficiencies [161, 162]. In spite of the poor evolutionary conservation of sense-antisense transcription, some subgroups of lincRNAs, such as senescence-associated vlincRNAs and macro lincRNAs in mammals or XUTs in yeast, are mostly constituted of antisense transcripts, which suggests potential antisense-mediated regulatory pathways in control of cellular homeostasis, stress response, and disease [154].

The discovery of bidirectional transcription as an intrinsic feature of the eukaryotic transcriptional machinery has given rise to the identification of **bidirectional lincRNAs** [153, 163–166]. Originating from the opposite strand of a PCG strand, these transcripts do not overlap or only partially overlap with the 5' region of paired PCGs, as is the case of promoter-associated (pa)ncRNAs, long upstream antisense transcripts (LUATs), and upstream antisense transcripts (uaRNA) [122, 167–170]. Presently, the number of bidirectional lincRNAs is largely underestimated not only because of the inaccurate annotation of transcriptional start sites (TSS) and promoters in the genome but also because of the highly unstable nature of these ncRNAs and the corresponding difficulty to detect them. Genomic studies have revealed that bidirectional promoters display distinct sequences and epigenetic features; moreover, they can be found near genes involved in specific biological processes such as developmental transcription factors or cell cycle regulation [122, 168, 169, 171, 172]. An imbalance in bidirectional transcription constitutes an endogenous fine-tuning mechanism that is particularly operative when facultative gene activation or repression is required [173, 174].

Intronic lincRNAs are restricted to PCG introns and could be either stand-alone unique transcripts or by-products of pre-mRNA processing. Examples of pre-mRNA-derived intronic transcripts are **circular intronic (ci)RNAs** produced from lariat introns which have escaped from debranching [175] and sno-lincRNAs produced from introns with two embedded snoRNA genes [176]. Such lincRNAs are proposed to positively regulate the transcription of the host PCG or its splicing by accumulating near the transcription locus. Another example of intronic lincRNAs of lariat origin, named **switch RNAs**, is produced by transcription through the

immunoglobulin switch regions. They are folded into G-quadruplex structures to bind and recruit the activation-induced cytidine deaminase AID to DNA in a sequence-specific manner, thereby ensuring proper class switch recombination in the germ line [177]. Stand-alone intronic transcripts, expressed independently of the PCG hosts, are believed to be the most prevalent class of intronic lncRNAs, including so-called totally intronic ncRNAs, **TINs** [178, 179]. Expression of a certain TIN is activated during inflammation, but the exact function of these lncRNAs is still poorly understood [180].

Overlapping sense transcripts encompass exons or whole PCGs within their introns without any sense exon overlap and are transcribed in the same sense direction. This annotation includes the GENCODE-adopted homonymous biotype and has been attributed to a number of transcripts, denoted as “*GENENAME-OT*.” One such example is *SOX2-OT* that harbors in its intron one of the major pluripotency regulators, the *SOX2* gene. *SOX2-OT* is dynamically expressed and is alternatively spliced not only during differentiation but also in cancer cells where it was proposed to regulate *SOX2* [181].

Intronic and overlapping sense lncRNAs could form **circular lncRNAs (circRNAs)** due to head-to-tail noncanonical splicing [182, 183]. Some sequence features such as the presence of repetitive elements within introns could be decisive for activation of noncanonical splicing and generation of a circular RNA molecule [184]. For example, Alu elements within introns are proposed to participate in RNA circularization via RNA-RNA pairing [185]. Remarkably, such events seem to be tissue or cell type specific, restricted to a certain developmental stage or pathological context [186, 187]. More generally, circRNAs function in the cytosol as miRNA sponges, as the case of CDR1as/ciRS-7 which is an RNA sponge of miR-7 [182, 183]. Some circRNAs, termed **exon-intron circRNAs (EIciRNAs)**, still contain unspliced introns and are retained in the nucleus, where they are able to interact with U1 snRNP and promote transcription of their parental genes [188]. The most remarkable property of circRNAs is their high stability which makes them eligible as potent diagnostic markers and therapeutic agents [189].

1.6.3 Classification According to lncRNA Residence Within Specific DNA Regulatory Elements and Loci

In addition to PCGs, mammalian genomes contain tens of thousands of pseudogenes, which are genomic remnants of ancient PCGs that have lost their coding potential throughout evolution. Importantly, many of them are transcribed in both sense and antisense directions into lncRNAs. Given high sequence similarity with parental genes, **pseudogene-derived lncRNAs** can regulate PCG expression via RNA-RNA pairing by acting as miRNA sponges, by producing endogenous siRNAs, or by interacting with mRNAs [190–192]. *PTENP1*, a lncRNA pseudogene derived from the tumor-suppressor gene *PTEN*, was among the first reported non-coding miRNA sponges with a function in cancer [193].

Ultra-conserved regions (UCRs) are genome segments that exhibit 100% DNA sequence conservation between human, mouse, and rat. The human genome contains 481 UCRs within intragenic (39%), intronic (43%), and exonic (15%) sequences [194]. These regions are extensively transcribed into ***T-UCR lncRNAs*** [195, 196]. Remarkably, expression of *T-UCRs* is induced by cancer-related stresses such as retinoid treatment or hypoxia. They are aberrantly expressed in different cancers and some are associated with poor prognosis [196–198]. Given high specificity of expression, *T-UCRs* were proposed as molecular markers for cancer diagnosis and prognosis [199]. The function of *T-UCRs* is still poorly understood. *Evf2* (or *Dlx6as*) is an example of *T-UCR* which acts as a decoy. It interacts with the transcription activator DLX1 increasing its association with key DNA enhancers but also with the SWI/SNF-like chromatin remodeler brahma-related gene 1 (BRG1) inhibiting its ATPase activity. As a result, *Evf2* induces chromatin remodeling and Dlx5/Dlx6 enhancers decommissioning with a final repression of transcription [200, 201].

Telomeres, which are protective nucleoprotein structures at the ends of chromosomes, are transcribed into non-coding **telomeric repeat-containing RNAs, *TERRA***, in all eukaryotes. This family of transcripts is generated from both Watson and Crick strands in a cell cycle-dependent manner [202, 203]. Formation of RNA-DNA hybrids by *TERRA* at chromosome ends promotes recombination and, hence, delays senescence. However, in cells lacking telomerase- and homology-directed repair, *TERRA* expression induces telomere shortening and accelerates senescence [204, 205]. Subtelomeric regions are also actively transcribed [206–208]. In budding yeast, this heterogeneous population of lncRNAs, named *subTERRA*, is transiently accumulating in late G2/M and G1 phases of the cell cycle in wild-type cells or in asynchronous cells deleted for the Xrn1 exoribonuclease [209]. The exact function of *subTERRA* is not yet clear though it has been proposed to have a regulatory role in telomere homeostasis.

Recent findings in different eukaryotes including human revealed that centromeric repeats are actively transcribed into lncRNAs during the progression from late mitosis to early G1 [210–214]. These **centromeric lncRNAs** physically interact with different centromere-specific nucleoprotein components, such as CENP-A/CENP-C and HJURP, and are required for correct kinetochore assembly and the maintenance of centromere integrity.

Ribosomal (r)DNA loci were shown to be transcribed by RNA polymerase II, antisense to the rRNA genes, into a heterogeneous population of lncRNAs, called ***PAPAS*** (promoter and pre-rRNA antisense). Their expression is induced in quiescent cells and triggers the recruitment of histone H4K20 methyltransferase Suv4-20h2 to ribosomal RNA genes for histone modification and transcriptional silencing [215]. *PAPAS* also allow heterochromatin formation and gene silencing in growth-arrested cells.

Promoters and enhancers constitute fundamental *cis*-regulatory elements for the control of PCG expression, serving as platforms for the recruitment of transcription factors and transcription machinery and the establishment of particular chromatin organization. Remarkably, many, if not all, functional enhancers and promoters are pervasively transcribed, respectively, into **eRNAs** and PALRs, in both sense and

antisense directions. Transcribed enhancer and promoter regions possess particular histone modification signatures that distinguish them from other transcription units. Such signatures include increased histone H3K27ac and H3K4me1 as compared with other lncRNA and PCGs. The termination of enhancer-derived lncRNAs, eRNAs, depends on the Integrator complex which ensures 3'-end transcript cleavage. The result is that eRNAs are poorly or not polyadenylated and highly unstable. Their expression is specific to cell type, tissue, or stages of development and can be activated by external or internal stimuli. Enhancer transcription was proposed to mark functional, active enhancer elements. However, eRNA function as stand-alone transcripts is still controversial, and the function of only few eRNAs, such as *FOXCIe* or *NRIP1e* [216], has been demonstrated. Specifically, it is proposed that these eRNAs control promoter chromatin environment, enhancer-promoter looping, RNA polymerase II loading and pausing, and “transcription factor trapping”; all these events contribute to a robust transcription activation of nearby and distant genes [217].

Promoter-associated lncRNAs or PALRs are transcribed in sense and antisense directions at promoter regions and can partially overlap the 5' end of a gene [218]. This class of transcripts includes highly unstable **PROMPTs** (promoter upstream transcripts) and **upstream antisense RNAs** (uaRNAs) that are more easily detectable in a context where the nuclear exosome has been depleted [108, 170, 219]. Polyadenylation-dependent degradation of PROMPTs was proposed to ensure directional RNA production from otherwise bidirectional promoters [220]. The presence of a splicing competent intron within uaRNAs was shown to facilitate gene looping placing termination factors at the vicinity of a bidirectional promoter for termination and thereby ensuring RNA polymerase II directionality toward a PCG [221]. Some PALRs were shown to negatively regulate transcription of the nearby genes. One such example is a PALR from the *CCND1* gene promoter which represses transcription by recruiting TLS and locally inhibiting CBP/p300 histone acetyltransferase activity on the downstream target gene, cyclin D1 [222, 223].

The 3'-untranslated regions (UTRs) of eukaryotic genes can be transcribed into independent transcription units or **UTR-associated (ua)RNAs** [224]. They are generated either by an independent transcriptional event from the upstream PCG or by posttranscriptional processing of a pre-mRNA. Expression of uaRNAs is regulated in a developmental stage- and tissue-specific fashion and is evolutionarily conserved; nevertheless, the functional relevance of such transcripts has not yet been explored.

1.6.4 Classification According to lncRNA Biogenesis Pathways

In budding yeast, since many lncRNAs are highly unstable or “cryptic,” the commonly employed classification of lncRNAs is based on their decay or biogenesis features. However, some so-called stable unannotated transcripts (**SUTs**) were identified in a wild-type genetic background [163]. Others are only detectable under specific stress conditions or in RNA-decay mutant strains. These latter transcripts

are roughly divided into three classes: **cryptic unstable transcripts (CUTs)**, which are sensitive to the nuclear RNA decay pathway [163, 225]; **Nrd1-terminated lncRNAs (NUTs)** [113]; and **Xrn1-sensitive unstable transcripts (XUTs)**, which are degraded by the cytoplasmic 5'–3' exoribonuclease, Xrn1 [226, 227]. The majority of XUTs are transcribed antisense to PCGs. CUTs are often bidirectional or overlapping sense transcripts, but can also be antisense, as is the case of the *PHO84* CUT [228]. Beyond each class definition, there is a considerable overlap between CUTs and NUTs but also XUTs and SUTs [94, 112]. Some CUTs have been reported to escape nuclear RNA decay and are exported to the cytoplasm where they are taken in charge by Xrn1 or by nonsense-mediated mRNA decay (NMD), as is the case of cytoplasmically degraded CUTs or CD-CUTs [229]. CD-CUTs bear a 5' extension originating upstream from the *bona fide* promoter and which partially or completely overlaps PCGs. CD-CUT transcription is proposed to control the expression of a subset of genes from subtelomeric regions and, in particular, metal homeostasis genes. Another subclass of CUTs includes meiotically induced lncRNAs, **meiotic unannotated transcripts (MUTs)**, that are degraded by the nuclear exosome Rrp6 and the exosome targeting complex TRAMP [230, 231]. The key difference between CUTs, XUTs, and SUTs is determined by their distinct subcellular fates. CUTs are transcribed and degraded in the nucleus, while SUTs and XUTs are exported to the cytoplasm where many XUTs are degraded by Xrn1 unless they escape degradation by pairing to complementary mRNAs [94]. In this case, they could be protected from NMD-mediated degradation and eventually translated into peptides, giving rise to new putatively functional molecules [232]. Notably, CUTs and XUTs are conserved among yeast species [233], (Wery et al., unpublished).

In other eukaryotes, some highly unstable lncRNAs have been reported, for example, above mentioned PROMPTs and eRNAs which could be considered to be human analogues of CUTs, since they are highly stabilized upon RNA exosome depletion [108, 234]. The RNA exosome is proposed to play a role in resolving deleterious RNA/DNA hybrids (R-loops) arising from active enhancers to prevent recombination. So far, the existence of mammalian XUTs has not been reported; however, in humans, XRN1 was shown to be sequestered by some RNA viruses [235, 236]. Their genomic RNA possesses a structured module in the 3'-UTR that traps and inhibits XRN1 catalytic activity. This action gives rise to the stabilization of the subgenomic flavivirus (sf)RNA which is important for the pathogenicity of the virus but could also result in a global stabilization of transcripts, including yet uncovered, highly unstable lncRNAs analogous to yeast XUTs.

1.6.5 Classification According to lncRNA Subcellular Localization or Origin

Knowing the subcellular localization of a particular lncRNA provides important insights into its biogenesis and function. lncRNAs could be exclusively cytosolic (*DANCR* and *OIP5-AS1*) or nuclear (*NEATI*) or have a dual localization (*HOTAIR*)

[128]. Several subgroups of lncRNAs with a precise subcellular localization have been defined, such as **chromatin-enriched (che)RNAs** [237] and **chromatin-associated lncRNAs, CARs** [238]. Many nuclear and chromatin functions have been proposed for such lncRNAs, including the assembly of subnuclear domains or RNP complexes, the guiding of chromatin modifications, and the activation or repression of protein activity [239]. **GAA repeat-containing RNAs, GRC-RNAs**, represent a subclass of nuclear lncRNAs that show focal localization in the mammalian interphase nucleus, where they are a part of the nuclear matrix. They have been suggested to play a role in the organization of the nucleus by assembling various nuclear matrix-associated proteins [240].

The mitochondrial genome is also transcribed into **mitochondrial ncRNAs, ncmtRNAs** [241–243]. Their biogenesis is dependent on nuclear-encoded mitochondrial processing proteins. After synthesis, some ncmtRNAs are exported from the mitochondria to the nucleus [244]. Importantly, expression of ncmtRNAs is altered in cancers promoting them as potential targets for cancer therapy [245, 246].

1.6.6 Classification According to lncRNA Function

To highlight a regulatory role, lncRNAs are often classified based on their function. Several archetypal activities of lncRNAs are used for classification: scaffolds, guides, decoys or ribo-repressors, ribo-activators, sponges, and precursors of small ncRNAs. Here we present examples of functional lncRNA classifications that regroup several lncRNAs into subclasses with a common operating mode.

lncRNA scaffolds function in the assembly of RNP complexes. The structural plasticity of lncRNAs allows them to adopt complex and dynamic three-dimensional structures with high affinity for proteins [247]. lncRNA scaffolds are often actors of epigenetic and transcriptional control of gene expression regulation. In this case, a lncRNA can act in *trans* or in *cis* in respect to its transcription site [248]. They are known to associate with a multitude of histone- or DNA-modifying and nucleosome remodeling complexes [249, 250]. lncRNA-mediated assembly of these complexes reshapes the epigenetic landscape and the organization of chromatin domains, thus allowing the modulation of all DNA-based processes including transcription, recombination, DNA repair, as well as RNA processing [159, 177, 251, 252]. *HOTAIR* is one example of a scaffold lncRNA which recognizes numerous targets. *HOTAIR* adopts a four-module secondary structure [144] which interacts in the nucleus with the PRC2 and Lsd1/REST/coREST complexes through its 5' and 3' modules, respectively [253]; it then targets them to specific genomic locations to affect histone modifications and gene silencing. In the cytoplasm, *HOTAIR* associates with the E3 ubiquitin ligases, Dzip3 and Mex3b, facilitating ubiquitination and proteolysis of their respective substrates, Ataxin-1 and Snurportin-1, in senescent cells [251].

Architectural lncRNAs (arcRNAs) represent a subclass of lncRNA scaffolds that are essential for the assembly of particular nuclear substructures [254].

Presently, five lncRNAs are classified as arcRNAs, and among them is *NEAT1*, which assembles more than 60 different RNA-binding proteins and transcription factors in paraspeckles [255]. ArcRNAs are highly enriched in repetitive sequences indicative of complex RNA folding that is essential for their scaffold function. They could be temporarily regulated by stress, during development, or in disease. ArcRNAs often sequester regulatory proteins, thereby changing gene expression. A detailed molecular role of scaffold and arcRNAs will be discussed in the forthcoming chapter.

Guide lncRNAs can recruit RNP complexes to specific chromatin loci. Remarkably, a guide function of one and the same lncRNA depends on the biological context (cell-/tissue-type, developmental stage, pathology) and often cannot be explained by a simple RNA/DNA sequence complementarity. For some lncRNA guides the formation of a triple helix structure between DNA and the lncRNA was experimentally proven, as in the case of *Khps1* which anchors the CBP/p300 complex to the proto-oncogene *SPHK1* [256]. Another example is *MEG3* which guides the EZH2 subunit of PRC2 to TGF β -regulated genes [257].

lncRNA decoys play the role of **ribo-repressors** for protein activities through the induction of allosteric modifications, the inhibition of catalytic activity, or by blocking the binding sites. One classical example of a ribo-repressor lncRNA is *GAS5* (growth arrest-specific 5), which acts as a decoy for a glucocorticoid receptor (GR) by mimicking its genomic DNA glucocorticoid response element (GRE). The interaction of *GAS5* with GR prevents it from binding to the GRE and ultimately represses GR-regulated genes, thus influencing many cellular functions including metabolism, cell survival, and response to apoptotic stimuli [258].

lncRNAs can also act as **ribo-activators** essential for or enhancing protein activities. One such example is the *lnc-DC* lncRNA which promotes the phosphorylation and activation of the STAT3 transcription factor [259]. Another subclass is the lncRNA transcriptional co-activators, also called **activating ncRNAs (ncRNA-a)**, which possess enhancer-like properties [260]. They were shown to interact with and regulate the kinase activity of Mediator, hence facilitating chromatin looping and transcription [261]. In addition to Mediator-interacting RNAs, other lncRNAs are able to upregulate transcription and could also be considered as ncRNA-a. Among them is the steroid receptor RNA activator *SRA* which interacts with and enhances the function of the insulator protein CTCF [262], and *NeST* which binds to and stimulates the activity of a subunit of the histone H3 Lysine 4 methyltransferase complex [263].

Competing endogenous RNAs (ceRNAs), also known as lncRNA sponges, are represented by lncRNAs and circRNAs that share partial sequence similarity to PCG transcripts; they function by competing for miRNA binding and posttranscriptional control [264]. Pseudogene-derived lncRNAs represent an important source of ceRNAs as they are particularly enriched in miRNA response elements, as is the case of the already mentioned *PTENP1* [265]. The subcellular balance between ceRNA, one or multiple miRNAs, and mRNA targets constitutes a complex network allowing a fine-tuning of the regulation of gene expression during adaptation, stress response, and development [266, 267].

Many lncRNAs host small RNA genes and serve as **precursor lncRNAs** for shorter regulatory RNAs, in particular, those involved in the RNAi pathway (mi/si/piRNAs). Many lncRNAs were identified and functionally studied before their precursor function was known. Such is the case for *H19*, one of the first discovered lncRNA genes and which contains two conserved microRNAs, miR-675-3p and miR-675-5p. In undifferentiated cells, *H19* acts as a ribo-activator interacting with and promoting the activity of the ssRNA-binding protein KSRP (K homology-type splicing regulatory protein) to prevent myogenic differentiation [268]. During development, and, in particular, during skeletal muscle differentiation, *H19* is processed into miRNAs ensuring the posttranscriptional control of the anti-differentiation transcription factors Smad [269]. Some piRNA clusters were found to map to lncRNA genes, mostly in exonic but also in non-exonic regions enriched in mobile elements thereby constituting putative *pi-lncRNA* precursors [270]. Putative endo-siRNAs can be produced from inverted repeats within lncRNA genes or from any double-stranded lncRNA-RNA precursors originated from sense-antisense convergent transcription [271, 272]. Endo-siRNAs have been documented in many eukaryotes, including fly, nematode, and mouse. Overlapping and bidirectional transcription is an abundant and conserved phenomenon among eukaryotes [154, 218]. However, in mammals, processing of sense-antisense paired transcripts into siRNA and their functional relevance is still controversial and requires experimental evidence, specifically at the single cell level. lncRNA processing into small RNA molecules could depend on different cellular machineries such as RNase P- and RNase Z-mediated cleavage of the small cytoplasmic mascRNA from *MALAT1* [110] or Drosha-DGCR8-driven termination and 3' end formation for lnc-pri-miRNAs [111]. The possible coexistence of two operational modes combining a long, precursor lncRNA and a derived small RNA adds additional complexity in lncRNA-mediated regulatory circuits.

1.6.7 Classification According to lncRNA Association with Specific Biological Processes

Examination of the non-coding transcriptome in different biological contexts has resulted in the discovery of lncRNAs specifically associated with particular biological states or pathologies. lncRNAs differentially expressed during replicative senescence represent **senescence-associated lncRNAs, or SAL** [273]. One such example, *SALNR*, is able to delay oncogene-induced senescence by its interaction with and inhibition of the NF90 posttranscriptional repressor [274]. Hypoxia, one of the classic features of the tumor microenvironment, induces the expression of many lncRNAs, in particular those from UCRs, named **HINCUTs** [197, 275]. Oxidative stress induces the production of stress-induced lncRNAs, **si-lncRNAs**, that accumulate at polysomes in contrast to mRNAs, which are depleted [138]. Deep sequencing transcriptome analysis of mammalian stem cells identified **non-annotated stem transcripts, or NASTs**, that appear to be important for maintaining pluripotency

[276]. Finally, with the progression of clinical and diagnostic studies, a growing number of specific disease-associated lncRNAs have been detected. An example is the **prostate cancer-associated transcripts (PCATs)**, such as *PCAT1*, that were shown to have a role in cancer biology but also as potent prognostic markers [277].

1.6.8 *Future Challenges in lncRNA Annotation and Classification*

Presently, the discovery of a novel lncRNA is an everyday occurrence, and proper annotation and classification are a necessity. In addition to catchy nicknames, various classifications of lncRNAs that rely on certain properties of the transcript, its origin, or possible function are proposed in oral and written communications. However, and in the aim of universalization, a “gold standard” of annotation should be sought. Repositories such as RNAcentral and other consortia are working on the challenging task of integrating the unambiguous annotations of all transcripts and genes, including numerical identifiers in addition to unique transcript names such as “GENENAME”. Recently, John Mattick and John Rinn have proposed some rules for lncRNA annotation. In particular, it has been recommended to refer to intergenic lncRNAs as “LINC-X,” where X represents a number and to all intragenic lncRNAs as “GENENAME” corresponding to overlapping PCG annotations with a prefix “AS-” for antisense, “BI-” for bidirectional, “OT-” for overlapping sense, and “INT-” for intronic transcripts in order to provide them with a positional criterion [278]. Respecting this guideline, *OT-SOX2-(1)* would correspond to the first isoform of the *SOX2-OT1* lncRNA overlapping in sense orientation the *SOX2* gene, while *HOTAIR* should take the name of *AS-HOXC11-(1)* to designate the largest lncRNA antisense to the *HOXC11* gene. However, the descriptive nickname of experimentally assigned lncRNAs should be preserved on condition of its uniqueness as a gene name. To avoid confusion, the renaming of transcripts should be accurately marked in all lncRNA repositories.

Identification, annotation, and classification are the first steps toward unraveling lncRNA biology. This work is still in its early days and requires novel thinking and methodologies, in parallel with the development of new and more accurate technologies and improved tools for the discovery and assembly of transcripts. In particular, the twenty-first century has been marked by the emergence of new technologies in regard to genomics and integrative system biology. These new approaches will allow researchers to build a comprehensive framework of regulatory circuits embedding both coding and non-coding transcripts, thereby deciphering a bit more the puzzle of life biodiversity and complexity.

Acknowledgments We thank Edith Heard, Mike Schertzer, and members of the lab for attentively reading the manuscript and apologize to colleagues whose works are not discussed and cited due to space limitation.

References

1. Dahm R (2005) Friedrich Miescher and the discovery of DNA. *Dev Biol* 278:274–288. doi:[10.1016/j.ydbio.2004.11.028](https://doi.org/10.1016/j.ydbio.2004.11.028)
2. Avery OT, MacLeod CM, McCarty M (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcal types induction of transformation by a desoxyribonucleic acid fraction isolated from *Pneumococcus* type III. *J Exp Med* 79:137–158. doi:[10.1084/jem.79.2.137](https://doi.org/10.1084/jem.79.2.137)
3. Ochoa S (1980) A pursuit of a hobby. *Annu Rev Biochem* 49:1–31. doi:[10.1146/annurev.bi.49.070180.000245](https://doi.org/10.1146/annurev.bi.49.070180.000245)
4. Griffiths AJF, Miller JH, Suzuki DT, et al. *An Introduction to Genetic Analysis*. 7th edition. New York: W. H. Freeman; 2000. Transcription and RNA polymerase. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK22085/>
5. Cobb M (2015) Who discovered messenger RNA? *Curr Biol* 25:R526–R532. doi:[10.1016/j.cub.2015.05.032](https://doi.org/10.1016/j.cub.2015.05.032)
6. Crick FHC (1968) The origin of the genetic code. *J Mol Biol* 38:367–379. doi:[10.1016/0022-2836\(68\)90392-6](https://doi.org/10.1016/0022-2836(68)90392-6)
7. Lewis JB, Atkins JF, Anderson CW et al (1975) Mapping of late adenovirus genes by cell-free translation of RNA selected by hybridization to specific DNA fragments. *Proc Natl Acad Sci* 72:1344–1348
8. Berk AJ (2016) Discovery of RNA splicing and genes in pieces. *Proc Natl Acad Sci* 113:801–805. doi:[10.1073/pnas.1525084113](https://doi.org/10.1073/pnas.1525084113)
9. Weinberg RA, Penman S (1968) Small molecular weight monodisperse nuclear RNA. *J Mol Biol* 38:289–304. doi:[10.1016/0022-2836\(68\)90387-2](https://doi.org/10.1016/0022-2836(68)90387-2)
10. Zieve G, Penman S (1976) Small RNA species of the HeLa cell: metabolism and subcellular localization. *Cell* 8:19–31. doi:[10.1016/0092-8674\(76\)90181-1](https://doi.org/10.1016/0092-8674(76)90181-1)
11. Kruger K, Grabowski PJ, Zaug AJ et al (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell* 31:147–157. doi:[10.1016/0092-8674\(82\)90414-7](https://doi.org/10.1016/0092-8674(82)90414-7)
12. Guerrier-Takada C, Gardiner K, Marsh T et al (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* 35:849–857. doi:[10.1016/0092-8674\(83\)90117-4](https://doi.org/10.1016/0092-8674(83)90117-4)
13. Cech TR (2000) structural biology: enhanced: the ribosome is a ribozyme. *Science* 289:878–879. doi:[10.1126/science.289.5481.878](https://doi.org/10.1126/science.289.5481.878)
14. Butcher SE (2009) The spliceosome as ribozyme hypothesis takes a second step. *Proc Natl Acad Sci U S A* 106:12211–12212. doi:[10.1073/pnas.0906762106](https://doi.org/10.1073/pnas.0906762106)
15. Bernhardt HS (2012) The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others). *Biol Direct* 7:23. doi:[10.1186/1745-6150-7-23](https://doi.org/10.1186/1745-6150-7-23)
16. Inouye M, Delihast N (1988) Small RNAs in the prokaryotes: a growing list of diverse roles. *Cell* 53:5–7. doi:[10.1016/0092-8674\(88\)90480-1](https://doi.org/10.1016/0092-8674(88)90480-1)
17. Corcoran CP, Podkaminski D, Papenfort K et al (2012) Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA: new MicF targets. *Mol Microbiol* 84:428–445. doi:[10.1111/j.1365-2958.2012.08031.x](https://doi.org/10.1111/j.1365-2958.2012.08031.x)
18. Delihast N (2015) Discovery and characterization of the first non-coding RNA that regulates gene expression, *micF* RNA: a historical perspective. *World J Biol Chem* 6:272. doi:[10.4331/wjbc.v6.i4.272](https://doi.org/10.4331/wjbc.v6.i4.272)
19. Lee RC, Feinbaum RL, Ambros V (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75:843–854. doi:[10.1016/0092-8674\(93\)90529-Y](https://doi.org/10.1016/0092-8674(93)90529-Y)
20. Wightman B, Ha I, Ruvkun G (1993) Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* 75:855–862. doi:[10.1016/0092-8674\(93\)90530-4](https://doi.org/10.1016/0092-8674(93)90530-4)
21. Wassenegger M, Heimes S, Riedel L, Sanger HL (1994) RNA-directed de novo methylation of genomic sequences in plants. *Cell* 76:567–576. doi:[10.1016/0092-8674\(94\)90119-8](https://doi.org/10.1016/0092-8674(94)90119-8)

22. Ameres SL, Zamore PD (2013) Diversifying microRNA sequence and function. *Nat Rev Mol Cell Biol* 14:475–488. doi:[10.1038/nrm3611](https://doi.org/10.1038/nrm3611)
23. He L, Hannon GJ (2004) MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet* 5:522–531. doi:[10.1038/nrg1379](https://doi.org/10.1038/nrg1379)
24. Montgomery MK (2004) RNA interference. In: Gott JM (ed) *RNA Interf Ed Modif. Humana*, Totowa, pp 3–21
25. Castel SE, Martienssen RA (2013) RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nat Rev Genet* 14:100–112. doi:[10.1038/nrg3355](https://doi.org/10.1038/nrg3355)
26. Pachnis V, Belayew A, Tilghman SM (1984) Locus unlinked to alpha-fetoprotein under the control of the murine raf and Rif genes. *Proc Natl Acad Sci U S A* 81:5523–5527
27. Bartolomei MS, Zemel S, Tilghman SM (1991) Parental imprinting of the mouse H19 gene. *Nature* 351:153–155. doi:[10.1038/351153a0](https://doi.org/10.1038/351153a0)
28. Barlow DP, Stöger R, Herrmann BG et al (1991) The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. *Nature* 349:84–87. doi:[10.1038/349084a0](https://doi.org/10.1038/349084a0)
29. Brannan CI, Dees EC, Ingram RS, Tilghman SM (1990) The product of the H19 gene may function as an RNA. *Mol Cell Biol* 10:28–36. doi:[10.1128/MCB.10.1.28](https://doi.org/10.1128/MCB.10.1.28)
30. Lyon MF (1961) Gene action in the X-chromosome of the mouse (*Mus musculus* L.) *Nature* 190:372–373
31. Borsani G, Tonlorenzi R, Simmler MC et al (1991) Characterization of a murine gene expressed from the inactive X chromosome. *Nature* 351:325–329. doi:[10.1038/351325a0](https://doi.org/10.1038/351325a0)
32. Brown CJ, Ballabio A, Rupert JL et al (1991) A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* 349:38–44. doi:[10.1038/349038a0](https://doi.org/10.1038/349038a0)
33. Brockdorff N, Ashworth A, Kay GF et al (1991) Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. *Nature* 351:329–331. doi:[10.1038/351329a0](https://doi.org/10.1038/351329a0)
34. Gendrel A-V, Heard E (2014) Noncoding RNAs and epigenetic mechanisms during X-chromosome inactivation. *Annu Rev Cell Dev Biol* 30:561–580. doi:[10.1146/annurev-cellbio-101512-122415](https://doi.org/10.1146/annurev-cellbio-101512-122415)
35. Brown CJ, Willard HF (1994) The human X-inactivation centre is not required for maintenance of X-chromosome inactivation. *Nature* 368:154–156. doi:[10.1038/368154a0](https://doi.org/10.1038/368154a0)
36. Heard E, Mongelard F, Arnaud D et al (1999) Human XIST yeast artificial chromosome transgenes show partial X inactivation center function in mouse embryonic stem cells. *Proc Natl Acad Sci* 96:6841–6846. doi:[10.1073/pnas.96.12.6841](https://doi.org/10.1073/pnas.96.12.6841)
37. Chureau C, Prissette M, Bourdet A et al (2002) Comparative sequence analysis of the X-inactivation center region in mouse, human, and bovine. *Genome Res* 12:894–908. doi:[10.1101/gr.152902](https://doi.org/10.1101/gr.152902)
38. Lee JT, Lu N (1999) Targeted mutagenesis of Tsix leads to nonrandom X inactivation. *Cell* 99:47–57. doi:[10.1016/S0092-8674\(00\)80061-6](https://doi.org/10.1016/S0092-8674(00)80061-6)
39. Migeon BR, Lee CH, Chowdhury AK, Carpenter H (2002) Species differences in TSIX/Tsix reveal the roles of these genes in X-chromosome inactivation. *Am J Hum Genet* 71:286–293. doi:[10.1086/341605](https://doi.org/10.1086/341605)
40. Vallot C, Huret C, Lesecque Y et al (2013) XACT, a long noncoding transcript coating the active X chromosome in human pluripotent cells. *Nat Genet* 45:239–241. doi:[10.1038/ng.2530](https://doi.org/10.1038/ng.2530)
41. Orgel LE, Crick FHC (1980) Selfish DNA: the ultimate parasite. *Nature* 284:604–607. doi:[10.1038/284604a0](https://doi.org/10.1038/284604a0)
42. Sanger F, Coulson AR, Friedmann T et al (1978) The nucleotide sequence of bacteriophage phiX174. *J Mol Biol* 125:225–246
43. Fleischmann RD, Adams MD, White O et al (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269:496–512
44. Dunham I, Shimizu N, Roe BA et al (1999) The DNA sequence of human chromosome 22. *Nature* 402:489–495. doi:[10.1038/990031](https://doi.org/10.1038/990031)
45. Lander ES, Linton LM, Birren B et al (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860–921. doi:[10.1038/35057062](https://doi.org/10.1038/35057062)

46. Venter JC (2001) The sequence of the human genome. *Science* 291:1304–1351. doi:[10.1126/science.1058040](https://doi.org/10.1126/science.1058040)
47. Goffeau A, Barrell BG, Bussey H et al (1996) Life with 6000 genes. *Science* 274:546, 563–546, 567
48. Crollius HR (2000) Characterization and repeat analysis of the compact genome of the freshwater pufferfish *Tetraodon nigroviridis*. *Genome Res* 10:939–949. doi:[10.1101/gr.10.7.939](https://doi.org/10.1101/gr.10.7.939)
49. Waterston R, Sulston J (1995) The genome of *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A* 92:10836–10840
50. Adams MD, Celniker SE, Holt RA et al (2000) The genome sequence of *Drosophila melanogaster*. *Science* 287:2185–2195
51. Chinwalla AT, Cook LL, Delehaunty KD et al (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* 420:520–562. doi:[10.1038/nature01262](https://doi.org/10.1038/nature01262)
52. Antequera F, Bird A (1993) Number of CpG islands and genes in human and mouse. *Proc Natl Acad Sci U S A* 90:11995–11999
53. International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature* 431:931–945. doi:[10.1038/nature03001](https://doi.org/10.1038/nature03001)
54. Kapranov P (2002) Large-scale transcriptional activity in chromosomes 21 and 22. *Science* 296:916–919. doi:[10.1126/science.1068597](https://doi.org/10.1126/science.1068597)
55. The FANTOM Consortium (2005) The transcriptional landscape of the mammalian genome. *Science* 309:1559–1563. doi:[10.1126/science.1112014](https://doi.org/10.1126/science.1112014)
56. Okazaki Y, Furuno M, Kasukawa T et al (2002) Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* 420:563–573. doi:[10.1038/nature01266](https://doi.org/10.1038/nature01266)
57. Katayama S, Tomaru Y, Kasukawa T et al (2005) Antisense transcription in the mammalian transcriptome. *Science* 309:1564–1566. doi:[10.1126/science.1112009](https://doi.org/10.1126/science.1112009)
58. ENCODE Project Consortium, Birney E, Stamatoyannopoulos JA et al (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447:799–816. doi:[10.1038/nature05874](https://doi.org/10.1038/nature05874)
59. ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489:57–74. doi:[10.1038/nature11247](https://doi.org/10.1038/nature11247)
60. Mouse ENCODE Consortium, Stamatoyannopoulos JA, Snyder M et al (2012) An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome Biol* 13:418. doi:[10.1186/gb-2012-13-8-418](https://doi.org/10.1186/gb-2012-13-8-418)
61. Yue F, Cheng Y, Breschi A et al (2014) A comparative encyclopedia of DNA elements in the mouse genome. *Nature* 515:355–364. doi:[10.1038/nature13992](https://doi.org/10.1038/nature13992)
62. David L, Huber W, Granovskaia M et al (2006) A high-resolution map of transcription in the yeast genome. *Proc Natl Acad Sci* 103:5320–5325. doi:[10.1073/pnas.0601091103](https://doi.org/10.1073/pnas.0601091103)
63. Dinger ME, Amaral PP, Mercer TR, Mattick JS (2009) Pervasive transcription of the eukaryotic genome: functional indices and conceptual implications. *Brief Funct Genomic Proteomic* 8:407–423. doi:[10.1093/bfgp/elp038](https://doi.org/10.1093/bfgp/elp038)
64. Berretta J, Morillon A (2009) Pervasive transcription constitutes a new level of eukaryotic genome regulation. *EMBO Rep* 10:973–982. doi:[10.1038/embor.2009.181](https://doi.org/10.1038/embor.2009.181)
65. Mattick JS (2003) Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. *Bioessays* 25:930–939. doi:[10.1002/bies.10332](https://doi.org/10.1002/bies.10332)
66. Clark MB, Choudhary A, Smith MA et al (2013) The dark matter rises: the expanding world of regulatory RNAs. *Essays Biochem* 54:1–16. doi:[10.1042/bse0540001](https://doi.org/10.1042/bse0540001)
67. Marques AC, Ponting CP (2014) Intergenic lncRNAs and the evolution of gene expression. *Curr Opin Genet Dev* 27:48–53. doi:[10.1016/j.gde.2014.03.009](https://doi.org/10.1016/j.gde.2014.03.009)
68. Duret L (2006) The Xist RNA gene evolved in Eutherians by pseudogenization of a protein-coding gene. *Science* 312:1653–1655. doi:[10.1126/science.1126316](https://doi.org/10.1126/science.1126316)
69. Ganesh S, Svoboda P (2016) Retrotransposon-associated long non-coding RNAs in mice and men. *Pflüg Arch - Eur J Physiol* 468:1049–1060. doi:[10.1007/s00424-016-1818-5](https://doi.org/10.1007/s00424-016-1818-5)
70. Kapusta A, Kronenberg Z, Lynch VJ et al (2013) Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet* 9:e1003470. doi:[10.1371/journal.pgen.1003470](https://doi.org/10.1371/journal.pgen.1003470)

71. Johnson R, Guigo R (2014) The RIDL hypothesis: transposable elements as functional domains of long noncoding RNAs. *RNA* 20:959–976. doi:[10.1261/rna.044560.114](https://doi.org/10.1261/rna.044560.114)
72. Hacisuleyman E, Shukla CJ, Weiner CL, Rinn JL (2016) Function and evolution of local repeats in the Fire locus. *Nat Commun* 7:11021. doi:[10.1038/ncomms11021](https://doi.org/10.1038/ncomms11021)
73. Heinen TJAJ, Staubach F, Häming D, Tautz D (2009) Emergence of a new gene from an intergenic region. *Curr Biol* 19:1527–1531. doi:[10.1016/j.cub.2009.07.049](https://doi.org/10.1016/j.cub.2009.07.049)
74. D-D W, Irwin DM, Zhang Y-P (2011) De novo origin of human protein-coding genes. *PLoS Genet* 7:e1002379. doi:[10.1371/journal.pgen.1002379](https://doi.org/10.1371/journal.pgen.1002379)
75. Durruthy-Durruthy J, Sebastiano V, Wossidlo M et al (2015) The primate-specific noncoding RNA HPAT5 regulates pluripotency during human preimplantation development and nuclear reprogramming. *Nat Genet* 48:44–52. doi:[10.1038/ng.3449](https://doi.org/10.1038/ng.3449)
76. Rands CM, Meader S, Ponting CP, Lunter G (2014) 8.2% of the Human genome is constrained: variation in rates of turnover across functional element classes in the human lineage. *PLoS Genet* 10:e1004525. doi:[10.1371/journal.pgen.1004525](https://doi.org/10.1371/journal.pgen.1004525)
77. Necsulea A, Soumillon M, Warnefors M et al (2014) The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* 505:635–640. doi:[10.1038/nature12943](https://doi.org/10.1038/nature12943)
78. Young RS, Ponting CP (2013) Identification and function of long non-coding RNAs. *Essays Biochem* 54:113–126. doi:[10.1042/bse0540113](https://doi.org/10.1042/bse0540113)
79. Ponting CP, Oliver PL, Reik W (2009) Evolution and functions of long noncoding RNAs. *Cell* 136:629–641. doi:[10.1016/j.cell.2009.02.006](https://doi.org/10.1016/j.cell.2009.02.006)
80. Diederichs S (2014) The four dimensions of noncoding RNA conservation. *Trends Genet* 30:121–123. doi:[10.1016/j.tig.2014.01.004](https://doi.org/10.1016/j.tig.2014.01.004)
81. Hezroni H, Koppstein D, Schwartz MG et al (2015) Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep* 11:1110–1122. doi:[10.1016/j.celrep.2015.04.023](https://doi.org/10.1016/j.celrep.2015.04.023)
82. Washietl S, Kellis M, Garber M (2014) Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. *Genome Res* 24:616–628. doi:[10.1101/gr.165035.113](https://doi.org/10.1101/gr.165035.113)
83. Ulitsky I, Shkumatava A, Jan CH et al (2011) Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* 147:1537–1550. doi:[10.1016/j.cell.2011.11.055](https://doi.org/10.1016/j.cell.2011.11.055)
84. Willingham AT, Gingeras TR (2006) TUF Love for “Junk” DNA. *Cell* 125:1215–1220. doi:[10.1016/j.cell.2006.06.009](https://doi.org/10.1016/j.cell.2006.06.009)
85. Mattick JS, Gagen MJ (2001) The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. *Mol Biol Evol* 18:1611–1630
86. Mattick JS (2001) Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep* 2:986–991. doi:[10.1093/embo-reports/kve230](https://doi.org/10.1093/embo-reports/kve230)
87. Pollard KS, Salama SR, King B et al (2006) Forces shaping the fastest evolving regions in the human genome. *PLoS Genet* 2:e168. doi:[10.1371/journal.pgen.0020168](https://doi.org/10.1371/journal.pgen.0020168)
88. Bird CP, Stranger BE, Liu M et al (2007) Fast-evolving noncoding sequences in the human genome. *Genome Biol* 8:R118. doi:[10.1186/gb-2007-8-6-r118](https://doi.org/10.1186/gb-2007-8-6-r118)
89. Pollard KS, Salama SR, Lambert N et al (2006) An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* 443:167–172. doi:[10.1038/nature05113](https://doi.org/10.1038/nature05113)
90. Bae B-I, Tietjen I, Atabay KD et al (2014) Evolutionarily dynamic alternative splicing of GPR56 regulates regional cerebral cortical patterning. *Science* 343:764–768. doi:[10.1126/science.1244392](https://doi.org/10.1126/science.1244392)
91. Doan RN, Bae B-I, Cubelos B et al (2016) Mutations in human accelerated regions disrupt cognition and social behavior. *Cell* 167:341–354.e12. doi:[10.1016/j.cell.2016.08.071](https://doi.org/10.1016/j.cell.2016.08.071)
92. van Heesch S, van Iterson M, Jacobi J et al (2014) Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. *Genome Biol* 15:R6. doi:[10.1186/gb-2014-15-1-r6](https://doi.org/10.1186/gb-2014-15-1-r6)
93. Juna Carlevaro-Fita, Anisa Rahim, Roderic Guigo, Leah Vardy, Rory Johnson (2015) Widespread localisation of long noncoding RNAs to ribosomes: Distinguishing features and evidence for regulatory roles. *bioRxiv* 013508; doi: <https://doi.org/10.1101/013508>

94. Wery M, Describes M, Vogt N et al (2016) Nonsense-mediated decay restricts lncRNA levels in yeast unless blocked by double-stranded RNA structure. *Mol Cell* 61:379–392. doi:[10.1016/j.molcel.2015.12.020](https://doi.org/10.1016/j.molcel.2015.12.020)
95. Housman G, Ulitsky I (2016) Methods for distinguishing between protein-coding and long noncoding RNAs and the elusive biological purpose of translation of long noncoding RNAs. *Biochim Biophys Acta* 1859:31–40. doi:[10.1016/j.bbagr.2015.07.017](https://doi.org/10.1016/j.bbagr.2015.07.017)
96. Andrews SJ, Rothnagel JA (2014) Emerging evidence for functional peptides encoded by short open reading frames. *Nat Rev Genet* 15:193–204. doi:[10.1038/nrg3520](https://doi.org/10.1038/nrg3520)
97. Banfai B, Jia H, Khatun J et al (2012) Long noncoding RNAs are rarely translated in two human cell lines. *Genome Res* 22:1646–1657. doi:[10.1101/gr.134767.111](https://doi.org/10.1101/gr.134767.111)
98. Ruiz-Orera J, Messegue X, Subirana JA, Alba MM (2014) Long non-coding RNAs as a source of new peptides. *Elife*. doi:[10.7554/eLife.03523](https://doi.org/10.7554/eLife.03523)
99. Ji Z, Song R, Regev A, Struhl K (2015) Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins. *Elife*. doi:[10.7554/eLife.08890](https://doi.org/10.7554/eLife.08890)
100. Nelson BR, Makarewich CA, Anderson DM et al (2016) A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. *Science* 351:271–275. doi:[10.1126/science.aad4076](https://doi.org/10.1126/science.aad4076)
101. Espinoza CA, Goodrich JA, Kugel JF (2007) Characterization of the structure, function, and mechanism of B2 RNA, an ncRNA repressor of RNA polymerase II transcription. *RNA* 13:583–596. doi:[10.1261/rna.310307](https://doi.org/10.1261/rna.310307)
102. Massone S, Ciarlo E, Vella S et al (2012) NDM29, a RNA polymerase III-dependent non coding RNA, promotes amyloidogenic processing of APP and amyloid β secretion. *Biochim Biophys Acta Res* 1823:1170–1177. doi:[10.1016/j.bbamcr.2012.05.001](https://doi.org/10.1016/j.bbamcr.2012.05.001)
103. Ariel F, Romero-Barrios N, Jégu T et al (2015) Battles and hijacks: noncoding transcription in plants. *Trends Plant Sci* 20:362–371. doi:[10.1016/j.tplants.2015.03.003](https://doi.org/10.1016/j.tplants.2015.03.003)
104. Yang L, Duff MO, Graveley BR et al (2011) Genomewide characterization of non-polyadenylated RNAs. *Genome Biol* 12:R16. doi:[10.1186/gb-2011-12-2-r16](https://doi.org/10.1186/gb-2011-12-2-r16)
105. Djebali S, Davis CA, Merkel A et al (2012) Landscape of transcription in human cells. *Nature* 489:101–108. doi:[10.1038/nature11233](https://doi.org/10.1038/nature11233)
106. Wilusz JE, JnBaptiste CK, LY L et al (2012) A triple helix stabilizes the 3' ends of long non-coding RNAs that lack poly(A) tails. *Genes Dev* 26:2392–2407. doi:[10.1101/gad.204438.112](https://doi.org/10.1101/gad.204438.112)
107. Alam T, Medvedeva YA, Jia H et al (2014) Promoter analysis reveals globally differential regulation of human long non-coding RNA and protein-coding genes. *PLoS One* 9:e109443. doi:[10.1371/journal.pone.0109443](https://doi.org/10.1371/journal.pone.0109443)
108. Preker P, Almvig K, Christensen MS et al (2011) PROMoter uPstream transcripts share characteristics with mRNAs and are produced upstream of all three major types of mammalian promoters. *Nucleic Acids Res* 39:7179–7193. doi:[10.1093/nar/gkr370](https://doi.org/10.1093/nar/gkr370)
109. Lai F, Gardini A, Zhang A, Shiekhattar R (2015) Integrator mediates the biogenesis of enhancer RNAs. *Nature* 525:399–403. doi:[10.1038/nature14906](https://doi.org/10.1038/nature14906)
110. Wilusz JE, Freier SM, Spector DL (2008) 3' End processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell* 135:919–932. doi:[10.1016/j.cell.2008.10.012](https://doi.org/10.1016/j.cell.2008.10.012)
111. Dhir A, Dhir S, Proudfoot NJ, Jopling CL (2015) Microprocessor mediates transcriptional termination of long noncoding RNA transcripts hosting microRNAs. *Nat Struct Mol Biol* 22:319–327. doi:[10.1038/nsmb.2982](https://doi.org/10.1038/nsmb.2982)
112. Fox MJ, Gao H, Smith-Kinnaman WR et al (2015) The Exosome component Rps6 is required for RNA Polymerase II termination at specific targets of the Nrd1-Nab3 pathway. *PLoS Genet* 11:e1004999. doi:[10.1371/journal.pgen.1004999](https://doi.org/10.1371/journal.pgen.1004999)
113. Schulz D, Schwab B, Kiesel A et al (2013) Transcriptome surveillance by selective termination of noncoding RNA synthesis. *Cell* 155:1075–1087. doi:[10.1016/j.cell.2013.10.024](https://doi.org/10.1016/j.cell.2013.10.024)
114. Porrua O, Libri D (2015) Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat Rev Mol Cell Biol*. doi:[10.1038/nrm3943](https://doi.org/10.1038/nrm3943)
115. Spurlock CF, Tossberg JT, Guo Y et al (2015) Expression and functions of long noncoding RNAs during human T helper cell differentiation. *Nat Commun* 6:6932. doi:[10.1038/ncomms7932](https://doi.org/10.1038/ncomms7932)

116. Hoffmann M, Dehn J, Droop J et al (2015) Truncated isoforms of lncRNA ANRIL are over-expressed in bladder cancer, but do not contribute to repression of INK4 tumor suppressors. *Non-Coding RNA* 1:266–284. doi:[10.3390/ncrna1030266](https://doi.org/10.3390/ncrna1030266)
117. Meseure D, Vacher S, Lallemand F et al (2016) Prognostic value of a newly identified MALAT1 alternatively spliced transcript in breast cancer. *Br J Cancer* 114:1395–1404. doi:[10.1038/bjc.2016.123](https://doi.org/10.1038/bjc.2016.123)
118. Derrien T, Johnson R, Bussotti G et al (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* 22:1775–1789. doi:[10.1101/gr.132159.111](https://doi.org/10.1101/gr.132159.111)
119. Bogu GK, Vizán P, Stanton LW et al (2016) Chromatin and RNA maps reveal regulatory long noncoding RNAs in mouse. *Mol Cell Biol* 36:809–819. doi:[10.1128/MCB.00955-15](https://doi.org/10.1128/MCB.00955-15)
120. Marques AC, Hughes J, Graham B et al (2013) Chromatin signatures at transcriptional start sites separate two equally populated yet distinct classes of intergenic long noncoding RNAs. *Genome Biol* 14:R131. doi:[10.1186/gb-2013-14-11-r131](https://doi.org/10.1186/gb-2013-14-11-r131)
121. Murray SC, Haenni S, Howe FS et al (2015) Sense and antisense transcription are associated with distinct chromatin architectures across genes. *Nucleic Acids Res* 43:7823–7837. doi:[10.1093/nar/gkv666](https://doi.org/10.1093/nar/gkv666)
122. Lepoivre C, Belhocine M, Bergon A et al (2013) Divergent transcription is associated with promoters of transcriptional regulators. *BMC Genomics* 14:914. doi:[10.1186/1471-2164-14-914](https://doi.org/10.1186/1471-2164-14-914)
123. Kim DH, Marinov GK, Pepke S et al (2015) Single-cell transcriptome analysis reveals dynamic changes in lncRNA expression during reprogramming. *Cell Stem Cell* 16:88–101. doi:[10.1016/j.stem.2014.11.005](https://doi.org/10.1016/j.stem.2014.11.005)
124. Liu SJ, Nowakowski TJ, Pollen AA et al (2016) Single-cell analysis of long non-coding RNAs in the developing human neocortex. *Genome Biol*. doi:[10.1186/s13059-016-0932-1](https://doi.org/10.1186/s13059-016-0932-1)
125. Ma Q, Chang HY (2016) Single-cell profiling of lncRNAs in the developing human brain. *Genome Biol*. doi:[10.1186/s13059-016-0933-0](https://doi.org/10.1186/s13059-016-0933-0)
126. Rotem A, Ram O, Shores N et al (2015) Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat Biotechnol* 33:1165–1172. doi:[10.1038/nbt.3383](https://doi.org/10.1038/nbt.3383)
127. Clark MB, Johnston RL, Inostroza-Ponta M et al (2012) Genome-wide analysis of long non-coding RNA stability. *Genome Res* 22:885–898. doi:[10.1101/gr.131037.111](https://doi.org/10.1101/gr.131037.111)
128. Ayupe AC, Tahira AC, Camargo L et al (2015) Global analysis of biogenesis, stability and sub-cellular localization of lncRNAs mapping to intragenic regions of the human genome. *RNA Biol* 12:877–892. doi:[10.1080/15476286.2015.1062960](https://doi.org/10.1080/15476286.2015.1062960)
129. Enuka Y, Lauriola M, Feldman ME et al (2016) Circular RNAs are long-lived and display only minimal early alterations in response to a growth factor. *Nucleic Acids Res* 44:1370–1383. doi:[10.1093/nar/gkv1367](https://doi.org/10.1093/nar/gkv1367)
130. Ward M, McEwan C, Mills JD, Janitz M (2015) Conservation and tissue-specific transcription patterns of long noncoding RNAs. *J Hum Transcr* 1:2–9. doi:[10.3109/23324015.2015.1077591](https://doi.org/10.3109/23324015.2015.1077591)
131. Li F, Xiao Y, Huang F et al (2015) Spatiotemporal-specific lncRNAs in the brain, colon, liver and lung of macaque during development. *Mol Biosyst* 11:3253–3263. doi:[10.1039/C5MB00474H](https://doi.org/10.1039/C5MB00474H)
132. Jiang L, Zhao L (2016) Identifying and functionally characterizing tissue-specific and ubiquitously expressed human lncRNAs. *Oncotarget*. doi:[10.18632/oncotarget.6859](https://doi.org/10.18632/oncotarget.6859)
133. Kornienko AE, Dotter CP, Guenzl PM et al (2016) Long non-coding RNAs display higher natural expression variation than protein-coding genes in healthy humans. *Genome Biol*. doi:[10.1186/s13059-016-0873-8](https://doi.org/10.1186/s13059-016-0873-8)
134. Kumar V, Westra H-J, Karjalainen J et al (2013) Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. *PLoS Genet* 9:e1003201. doi:[10.1371/journal.pgen.1003201](https://doi.org/10.1371/journal.pgen.1003201)
135. Cabili MN, Dunagin MC, McClanahan PD et al (2015) Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol* 16:20. doi:[10.1186/s13059-015-0586-4](https://doi.org/10.1186/s13059-015-0586-4)

136. Zhang B, Gunawardane L, Niazi F et al (2014) A novel RNA motif mediates the strict nuclear localization of a long noncoding RNA. *Mol Cell Biol* 34:2318–2329. doi:[10.1128/MCB.01673-13](https://doi.org/10.1128/MCB.01673-13)
137. Chen L-L (2016) Linking long noncoding RNA localization and function. *Trends Biochem Sci* 41:761–772. doi:[10.1016/j.tibs.2016.07.003](https://doi.org/10.1016/j.tibs.2016.07.003)
138. Giannakakis A, Zhang J, Jenjaroenpun P et al (2015) Contrasting expression patterns of coding and noncoding parts of the human genome upon oxidative stress. *Sci Rep* 5:9737. doi:[10.1038/srep09737](https://doi.org/10.1038/srep09737)
139. Noh JH, Kim KM, Abdelmohsen K et al (2016) HuR and GRSF1 modulate the nuclear export and mitochondrial localization of the lncRNA *RMRP*. *Genes Dev*. doi:[10.1101/gad.276022.115](https://doi.org/10.1101/gad.276022.115)
140. Lu Z, Chang HY (2016) Decoding the RNA structurome. *Curr Opin Struct Biol* 36:142–148. doi:[10.1016/j.sbi.2016.01.007](https://doi.org/10.1016/j.sbi.2016.01.007)
141. Johnsson P, Lipovich L, Grandér D, Morris KV (2014) Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochim Biophys Acta* 1840:1063–1071. doi:[10.1016/j.bbagen.2013.10.035](https://doi.org/10.1016/j.bbagen.2013.10.035)
142. He S, Liu S, Zhu H (2011) The sequence, structure and evolutionary features of HOTAIR in mammals. *BMC Evol Biol* 11:102. doi:[10.1186/1471-2148-11-102](https://doi.org/10.1186/1471-2148-11-102)
143. Bhan A, Mandal SS (2015) LncRNA HOTAIR: a master regulator of chromatin dynamics and cancer. *Biochim Biophys Acta* 1856:151–164. doi:[10.1016/j.bbcan.2015.07.001](https://doi.org/10.1016/j.bbcan.2015.07.001)
144. Somarowthu S, Legiewicz M, Chillón I et al (2015) HOTAIR forms an intricate and modular secondary structure. *Mol Cell* 58:353–361. doi:[10.1016/j.molcel.2015.03.006](https://doi.org/10.1016/j.molcel.2015.03.006)
145. Beniaminov A, Westhof E, Krol A (2008) Distinctive structures between chimpanzee and human in a brain noncoding RNA. *RNA* 14:1270–1275. doi:[10.1261/rna.1054608](https://doi.org/10.1261/rna.1054608)
146. St Laurent G, Shtokalo D, Dong B et al (2013) VlincRNAs controlled by retroviral elements are a hallmark of pluripotency and cancer. *Genome Biol* 14:R73. doi:[10.1186/gb-2013-14-7-r73](https://doi.org/10.1186/gb-2013-14-7-r73)
147. Lazorthes S, Vallot C, Briois S et al (2015) A vlincRNA participates in senescence maintenance by relieving H2AZ-mediated repression at the INK4 locus. *Nat Commun* 6:5971. doi:[10.1038/ncomms6971](https://doi.org/10.1038/ncomms6971)
148. Guenzl PM, Barlow DP (2012) Macro lncRNAs: a new layer of *cis*-regulatory information in the mammalian genome. *RNA Biol* 9:731–741. doi:[10.4161/rna.19985](https://doi.org/10.4161/rna.19985)
149. Khalil AM, Guttman M, Huarte M et al (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci* 106:11667–11672. doi:[10.1073/pnas.0904715106](https://doi.org/10.1073/pnas.0904715106)
150. Guttman M, Amit I, Garber M et al (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458:223–227. doi:[10.1038/nature07672](https://doi.org/10.1038/nature07672)
151. Huarte M, Guttman M, Feldser D et al (2010) A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* 142:409–419. doi:[10.1016/j.cell.2010.06.040](https://doi.org/10.1016/j.cell.2010.06.040)
152. Goodman AJ, Daugharthy ER, Kim J (2013) Pervasive antisense transcription is evolutionarily conserved in budding yeast. *Mol Biol Evol* 30:409–421. doi:[10.1093/molbev/mss240](https://doi.org/10.1093/molbev/mss240)
153. Kapranov P (2005) Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. *Genome Res* 15:987–997. doi:[10.1101/gr.3455305](https://doi.org/10.1101/gr.3455305)
154. Wood EJ, Chin-Inmanu K, Jia H, Lipovich L (2013) Sense-antisense gene pairs: sequence, transcription, and structure are not conserved between human and mouse. *Front Genet*. doi:[10.3389/fgene.2013.00183](https://doi.org/10.3389/fgene.2013.00183)
155. Magistri M, Faghihi MA, St Laurent G, Wahlestedt C (2012) Regulation of chromatin structure by long noncoding RNAs: focus on natural antisense transcripts. *Trends Genet* 28:389–396. doi:[10.1016/j.tig.2012.03.013](https://doi.org/10.1016/j.tig.2012.03.013)

156. W-Y S, Xiong H, Fang J-Y (2010) Natural antisense transcripts regulate gene expression in an epigenetic manner. *Biochem Biophys Res Commun* 396:177–181. doi:[10.1016/j.bbrc.2010.04.147](https://doi.org/10.1016/j.bbrc.2010.04.147)
157. Yuan C, Wang J, Harrison AP et al (2015) Genome-wide view of natural antisense transcripts in *Arabidopsis thaliana*. *DNA Res* 22:233–243. doi:[10.1093/dnares/dsv008](https://doi.org/10.1093/dnares/dsv008)
158. Faghihi MA, Modarresi F, Khalil AM et al (2008) Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of β -secretase. *Nat Med* 14:723–730. doi:[10.1038/nm1784](https://doi.org/10.1038/nm1784)
159. Gonzalez I, Munita R, Agirre E et al (2015) A lncRNA regulates alternative splicing via establishment of a splicing-specific chromatin signature. *Nat Struct Mol Biol*. doi:[10.1038/nsmb.3005](https://doi.org/10.1038/nsmb.3005)
160. Carrieri C, Cimatti L, Biagioli M et al (2012) Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature* 491:454–457. doi:[10.1038/nature11508](https://doi.org/10.1038/nature11508)
161. Zucchelli S, Fasolo F, Russo R et al (2015) SINEUPs are modular antisense long non-coding RNAs that increase synthesis of target proteins in cells. *Front Cell Neurosci*. doi:[10.3389/fncel.2015.00174](https://doi.org/10.3389/fncel.2015.00174)
162. Indriero A, Grimaldi C, Zucchelli S et al (2016) Synthetic long non-coding RNAs [SINEUPs] rescue defective gene expression in vivo. *Sci Rep* 6:27315. doi:[10.1038/srep27315](https://doi.org/10.1038/srep27315)
163. Xu Z, Wei W, Gagneur J et al (2009) Bidirectional promoters generate pervasive transcription in yeast. *Nature* 457:1033–1037. doi:[10.1038/nature07728](https://doi.org/10.1038/nature07728)
164. Scruggs BS, Gilchrist DA, Nechaev S et al (2015) Bidirectional transcription arises from two distinct hubs of transcription factor binding and active chromatin. *Mol Cell* 58:1101–1112. doi:[10.1016/j.molcel.2015.04.006](https://doi.org/10.1016/j.molcel.2015.04.006)
165. Wei W, Pelechano V, Järvelin AI, Steinmetz LM (2011) Functional consequences of bidirectional promoters. *Trends Genet* 27:267–276. doi:[10.1016/j.tig.2011.04.002](https://doi.org/10.1016/j.tig.2011.04.002)
166. Seila AC, Calabrese JM, Levine SS et al (2008) Divergent transcription from active promoters. *Science* 322:1849–1851. doi:[10.1126/science.1162253](https://doi.org/10.1126/science.1162253)
167. Hamazaki N, Uesaka M, Nakashima K et al (2015) Gene activation-associated long non-coding RNAs function in mouse preimplantation development. *Development* 142:910–920. doi:[10.1242/dev.116996](https://doi.org/10.1242/dev.116996)
168. Hung T, Wang Y, Lin MF et al (2011) Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat Genet* 43:621–629. doi:[10.1038/ng.848](https://doi.org/10.1038/ng.848)
169. Uesaka M, Nishimura O, Go Y et al (2014) Bidirectional promoters are the major source of gene activation-associated non-coding RNAs in mammals. *BMC Genomics* 15:35. doi:[10.1186/1471-2164-15-35](https://doi.org/10.1186/1471-2164-15-35)
170. Flynn RA, Almada AE, Zamudio JR, Sharp PA (2011) Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. *Proc Natl Acad Sci* 108:10460–10465. doi:[10.1073/pnas.1106630108](https://doi.org/10.1073/pnas.1106630108)
171. Hu H, He L, Khaitovich P (2014) Deep sequencing reveals a novel class of bidirectional promoters associated with neuronal genes. *BMC Genomics* 15:457. doi:[10.1186/1471-2164-15-457](https://doi.org/10.1186/1471-2164-15-457)
172. Sigova AA, Mullen AC, Molinie B et al (2013) Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proc Natl Acad Sci* 110:2876–2881. doi:[10.1073/pnas.1221904110](https://doi.org/10.1073/pnas.1221904110)
173. Morris KV, Santos S, Turner A-M et al (2008) Bidirectional transcription directs both transcriptional gene activation and suppression in human cells. *PLoS Genet* 4:e1000258. doi:[10.1371/journal.pgen.1000258](https://doi.org/10.1371/journal.pgen.1000258)
174. Kambara H, Gunawardane L, Zebrowski E et al (2015) Regulation of interferon-stimulated gene BST2 by a lncRNA transcribed from a shared bidirectional promoter. *Front Immunol*. doi:[10.3389/fimmu.2014.00676](https://doi.org/10.3389/fimmu.2014.00676)
175. Zhang Y, Zhang X-O, Chen T et al (2013) Circular intronic long noncoding RNAs. *Mol Cell* 51:792–806. doi:[10.1016/j.molcel.2013.08.017](https://doi.org/10.1016/j.molcel.2013.08.017)
176. Yin Q-F, Yang L, Zhang Y et al (2012) Long noncoding RNAs with snoRNA ends. *Mol Cell* 48:219–230. doi:[10.1016/j.molcel.2012.07.033](https://doi.org/10.1016/j.molcel.2012.07.033)

177. Zheng S, Vuong BQ, Vaidyanathan B et al (2015) Non-coding RNA generated following Lariat Debranching mediates targeting of AID to DNA. *Cell* 161:762–773. doi:[10.1016/j.cell.2015.03.020](https://doi.org/10.1016/j.cell.2015.03.020)
178. Nakaya HI, Amaral PP, Louro R et al (2007) Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription. *Genome Biol* 8:R43. doi:[10.1186/gb-2007-8-3-r43](https://doi.org/10.1186/gb-2007-8-3-r43)
179. Louro R, El-Jundi T, Nakaya HI et al (2008) Conserved tissue expression signatures of intronic noncoding RNAs transcribed from human and mouse loci. *Genomics* 92:18–25. doi:[10.1016/j.ygeno.2008.03.013](https://doi.org/10.1016/j.ygeno.2008.03.013)
180. St Laurent G, Shtokalo D, Tackett MR et al (2012) Intronic RNAs constitute the major fraction of the non-coding RNA in mammalian cells. *BMC Genomics* 13:504. doi:[10.1186/1471-2164-13-504](https://doi.org/10.1186/1471-2164-13-504)
181. Shahryari A, Jazi MS, Samaei NM, Mowla SJ (2015) Long non-coding RNA SOX2OT: expression signature, splicing patterns, and emerging roles in pluripotency and tumorigenesis. *Front Genet*. doi:[10.3389/fgene.2015.00196](https://doi.org/10.3389/fgene.2015.00196)
182. Memczak S, Jens M, Elefsinioti A et al (2013) Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 495:333–338. doi:[10.1038/nature11928](https://doi.org/10.1038/nature11928)
183. Hansen TB, Jensen TI, Clausen BH et al (2013) Natural RNA circles function as efficient microRNA sponges. *Nature* 495:384–388. doi:[10.1038/nature11993](https://doi.org/10.1038/nature11993)
184. Kramer MC, Liang D, Tatomer DC et al (2015) Combinatorial control of *Drosophila* circular RNA expression by intronic repeats, hnRNPs, and SR proteins. *Genes Dev* 29:2168–2182. doi:[10.1101/gad.270421.115](https://doi.org/10.1101/gad.270421.115)
185. Hadjiargyrou M, Delilhas N (2013) The intertwining of transposable elements and non-coding RNAs. *Int J Mol Sci* 14:13307–13328. doi:[10.3390/ijms140713307](https://doi.org/10.3390/ijms140713307)
186. Rybak-Wolf A, Stottmeister C, Glažar P et al (2015) Circular RNAs in the mammalian brain are highly abundant, conserved, and dynamically expressed. *Mol Cell* 58:870–885. doi:[10.1016/j.molcel.2015.03.027](https://doi.org/10.1016/j.molcel.2015.03.027)
187. Peng L, Yuan X, Li G (2015) The emerging landscape of circular RNA ciRS-7 in cancer (Review). *Oncol Rep*. doi:[10.3892/or.2015.3904](https://doi.org/10.3892/or.2015.3904)
188. Li Z, Huang C, Bao C et al (2015) Exon-intron circular RNAs regulate transcription in the nucleus. *Nat Struct Mol Biol* 22:256–264. doi:[10.1038/nsmb.2959](https://doi.org/10.1038/nsmb.2959)
189. Li J, Yang J, Zhou P et al (2015) Circular RNAs in cancer: novel insights into origins, properties, functions and implications. *Am J Cancer Res* 5:472–480
190. Milligan MJ, Lipovich L (2015) Pseudogene-derived lncRNAs: emerging regulators of gene expression. *Front Genet*. doi:[10.3389/fgene.2014.00476](https://doi.org/10.3389/fgene.2014.00476)
191. Zheng D, Frankish A, Baertsch R et al (2007) Pseudogenes in the ENCODE regions: consensus annotation, analysis of transcription, and evolution. *Genome Res* 17:839–851. doi:[10.1101/gr.5586307](https://doi.org/10.1101/gr.5586307)
192. Grandér D, Johnsson P (2015) Pseudogene-expressed RNAs: emerging roles in gene regulation and disease. In: Morris KV (ed) Long non-coding RNAs human disease. Springer, Cham, pp 111–126
193. Polisenio L, Salmena L, Zhang J et al (2010) A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* 465:1033–1038. doi:[10.1038/nature09144](https://doi.org/10.1038/nature09144)
194. Bejerano G (2004) Ultraconserved elements in the human genome. *Science* 304:1321–1325. doi:[10.1126/science.1098119](https://doi.org/10.1126/science.1098119)
195. Mestdagh P, Fredlund E, Pattyn F et al (2010) An integrative genomics screen uncovers ncRNA T-UCR functions in neuroblastoma tumours. *Oncogene* 29:3583–3592. doi:[10.1038/onc.2010.106](https://doi.org/10.1038/onc.2010.106)
196. Watters KM, Bryan K, Foley NH et al (2013) Expressional alterations in functional ultraconserved non-coding rnas in response to all-transretinoic acid – induced differentiation in neuroblastoma cells. *BMC Cancer*. doi:[10.1186/1471-2407-13-184](https://doi.org/10.1186/1471-2407-13-184)
197. Ferdin J, Nishida N, Wu X et al (2013) HINCUTs in cancer: hypoxia-induced noncoding ultraconserved transcripts. *Cell Death Differ* 20:1675–1687. doi:[10.1038/cdd.2013.119](https://doi.org/10.1038/cdd.2013.119)

198. Fassan M, Dall'Olmo L, Galasso M et al (2014) Transcribed ultraconserved noncoding RNAs (T-UCR) are involved in Barrett's esophagus carcinogenesis. *Oncotarget* 5:7162–7171. doi:[10.18632/oncotarget.2249](https://doi.org/10.18632/oncotarget.2249)
199. Scaruffi P, Stigliani S, Moretti S et al (2009) Transcribed-ultra conserved region expression is associated with outcome in high-risk neuroblastoma. *BMC Cancer*. doi:[10.1186/1471-2407-9-441](https://doi.org/10.1186/1471-2407-9-441)
200. Feng J (2006) The *Evf-2* noncoding RNA is transcribed from the *Dlx-5/6* ultraconserved region and functions as a *Dlx-2* transcriptional coactivator. *Genes Dev* 20:1470–1484. doi:[10.1101/gad.1416106](https://doi.org/10.1101/gad.1416106)
201. Cajigas I, Leib DE, Cochrane J et al (2015) *Evf2* lncRNA/BRG1/DLX1 interactions reveal RNA-dependent inhibition of chromatin remodeling. *Development* 142:2641–2652. doi:[10.1242/dev.126318](https://doi.org/10.1242/dev.126318)
202. Feuerhahn S, Iglesias N, Panza A et al (2010) TERRA biogenesis, turnover and implications for function. *FEBS Lett* 584:3812–3818. doi:[10.1016/j.febslet.2010.07.032](https://doi.org/10.1016/j.febslet.2010.07.032)
203. Porro A, Feuerhahn S, Reichenbach P, Lingner J (2010) Molecular dissection of telomeric repeat-containing RNA biogenesis unveils the presence of distinct and multiple regulatory pathways. *Mol Cell Biol* 30:4808–4817. doi:[10.1128/MCB.00460-10](https://doi.org/10.1128/MCB.00460-10)
204. Balk B, Maicher A, Dees M et al (2013) Telomeric RNA-DNA hybrids affect telomere-length dynamics and senescence. *Nat Struct Mol Biol* 20:1199–1205. doi:[10.1038/nsmb.2662](https://doi.org/10.1038/nsmb.2662)
205. Balk B, Dees M, Bender K, Luke B (2014) The differential processing of telomeres in response to increased telomeric transcription and RNA–DNA hybrid accumulation. *RNA Biol* 11:95–100. doi:[10.4161/rna.27798](https://doi.org/10.4161/rna.27798)
206. Greenwood J, Cooper JP (2012) Non-coding telomeric and subtelomeric transcripts are differentially regulated by telomeric and heterochromatin assembly factors in fission yeast. *Nucleic Acids Res* 40:2956–2963. doi:[10.1093/nar/gkr1155](https://doi.org/10.1093/nar/gkr1155)
207. Trofimova I, Chervyakova D, Krasikova A (2015) Transcription of subtelomere tandemly repetitive DNA in chicken embryogenesis. *Chromosome Res* 23:495–503. doi:[10.1007/s10577-015-9487-3](https://doi.org/10.1007/s10577-015-9487-3)
208. Broadbent KM, Broadbent JC, Ribacke U et al (2015) Strand-specific RNA sequencing in *Plasmodium falciparum* malaria identifies developmentally regulated long non-coding RNA and circular RNA. *BMC Genomics*. doi:[10.1186/s12864-015-1603-4](https://doi.org/10.1186/s12864-015-1603-4)
209. Kwapisz M, Ruault M, van Dijk E et al (2015) Expression of subtelomeric lncRNAs links telomeres dynamics to RNA decay in *S. cerevisiae*. *Non-Coding RNA* 1:94–126. doi:[10.3390/ncrna1020094](https://doi.org/10.3390/ncrna1020094)
210. Wong LH, Brettingham-Moore KH, Chan L et al (2007) Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. *Genome Res* 17:1146–1160. doi:[10.1101/gr.6022807](https://doi.org/10.1101/gr.6022807)
211. Quénet D, Dalal Y (2014) A long non-coding RNA is required for targeting centromeric protein A to the human centromere. *Elife*. doi:[10.7554/eLife.03254](https://doi.org/10.7554/eLife.03254)
212. Blower MD (2016) Centromeric transcription regulates Aurora-B localization and activation. *Cell Rep* 15:1624–1633. doi:[10.1016/j.celrep.2016.04.054](https://doi.org/10.1016/j.celrep.2016.04.054)
213. Chan FL, Marshall OJ, Saffery R et al (2012) Active transcription and essential role of RNA polymerase II at the centromere during mitosis. *Proc Natl Acad Sci* 109:1979–1984. doi:[10.1073/pnas.1108705109](https://doi.org/10.1073/pnas.1108705109)
214. Rošić S, Köhler F, Erhardt S (2014) Repetitive centromeric satellite RNA is essential for kinetochore formation and cell division. *J Cell Biol* 207:335–349. doi:[10.1083/jcb.201404097](https://doi.org/10.1083/jcb.201404097)
215. Bierhoff H, Dammert MA, Brocks D et al (2014) Quiescence-induced lncRNAs trigger H4K20 trimethylation and transcriptional silencing. *Mol Cell* 54:675–682. doi:[10.1016/j.molcel.2014.03.032](https://doi.org/10.1016/j.molcel.2014.03.032)
216. Li W, Notani D, Ma Q et al (2013) Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* 498:516–520. doi:[10.1038/nature12210](https://doi.org/10.1038/nature12210)
217. Li W, Notani D, Rosenfeld MG (2016) Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nat Rev Genet* 17:207–223. doi:[10.1038/nrg.2016.4](https://doi.org/10.1038/nrg.2016.4)

218. Kapranov P, Willingham AT, Gingeras TR (2007) Genome-wide transcription and the implications for genomic organization. *Nat Rev Genet* 8:413–423. doi:[10.1038/nrg2083](https://doi.org/10.1038/nrg2083)
219. Preker P, Nielsen J, Kammler S et al (2008) RNA exosome depletion reveals transcription upstream of active human promoters. *Science* 322:1851–1854. doi:[10.1126/science.1164096](https://doi.org/10.1126/science.1164096)
220. Ntini E, Järvelin AI, Bornholdt J et al (2013) Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. *Nat Struct Mol Biol* 20:923–928. doi:[10.1038/nsmb.2640](https://doi.org/10.1038/nsmb.2640)
221. Agarwal N, Ansari A (2016) Enhancement of transcription by a splicing-competent intron is dependent on promoter directionality. *PLoS Genet* 12:e1006047. doi:[10.1371/journal.pgen.1006047](https://doi.org/10.1371/journal.pgen.1006047)
222. Wang X, Arai S, Song X et al (2008) Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature* 454:126–130. doi:[10.1038/nature06992](https://doi.org/10.1038/nature06992)
223. Song X, Wang X, Arai S, Kurokawa R (2012) Promoter-associated noncoding RNA from the CCND1 promoter. In: Vancura A (ed) *Transcription regulation*. Springer, New York, pp 609–622
224. Mercer TR, Wilhelm D, Dinger ME et al (2011) Expression of distinct RNAs from 3' untranslated regions. *Nucleic Acids Res* 39:2393–2403. doi:[10.1093/nar/gkq1158](https://doi.org/10.1093/nar/gkq1158)
225. Neil H, Malabat C, d'Aubenton-Carafa Y et al (2009) Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* 457:1038–1042. doi:[10.1038/nature07747](https://doi.org/10.1038/nature07747)
226. van Dijk EL, Chen CL, d'Aubenton-Carafa Y et al (2011) XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature* 475:114–117. doi:[10.1038/nature10118](https://doi.org/10.1038/nature10118)
227. Berretta J, Pinskaya M, Morillon A (2008) A cryptic unstable transcript mediates transcriptional trans-silencing of the Ty1 retrotransposon in *S. cerevisiae*. *Genes Dev* 22:615–626. doi:[10.1101/gad.458008](https://doi.org/10.1101/gad.458008)
228. Camblong J, Beyrouthy N, Guffanti E et al (2009) Trans-acting antisense RNAs mediate transcriptional gene cosuppression in *S. cerevisiae*. *Genes Dev* 23:1534–1545. doi:[10.1101/gad.522509](https://doi.org/10.1101/gad.522509)
229. Toesca I, Nery CR, Fernandez CF et al (2011) Cryptic transcription mediates repression of subtelomeric metal homeostasis genes. *PLoS Genet* 7:e1002163. doi:[10.1371/journal.pgen.1002163](https://doi.org/10.1371/journal.pgen.1002163)
230. Lardenois A, Liu Y, Walther T et al (2011) Execution of the meiotic noncoding RNA expression program and the onset of gametogenesis in yeast require the conserved exosome subunit Rrp6. *Proc Natl Acad Sci* 108:1058–1063. doi:[10.1073/pnas.1016459108](https://doi.org/10.1073/pnas.1016459108)
231. Frenk S, Oxley D, Houseley J (2014) The nuclear exosome is active and important during budding yeast meiosis. *PLoS One* 9:e107648. doi:[10.1371/journal.pone.0107648](https://doi.org/10.1371/journal.pone.0107648)
232. de Andres-Pablo A, Morillon A, Wery M (2016) lncRNAs, lost in translation or licence to regulate? *Curr Genet*. doi:[10.1007/s00294-016-0615-1](https://doi.org/10.1007/s00294-016-0615-1)
233. Vera JM, Dowell RD (2016) Survey of cryptic unstable transcripts in yeast. *BMC Genomics*. doi:[10.1186/s12864-016-2622-5](https://doi.org/10.1186/s12864-016-2622-5)
234. Pefanis E, Wang J, Rothschild G et al (2015) RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. *Cell* 161:774–789. doi:[10.1016/j.cell.2015.04.034](https://doi.org/10.1016/j.cell.2015.04.034)
235. Moon SL, Blackinton JG, Anderson JR et al (2015) XRN1 stalling in the 5' UTR of hepatitis C virus and bovine viral diarrhoea virus is associated with dysregulated host mRNA stability. *PLoS Pathog* 11:e1004708. doi:[10.1371/journal.ppat.1004708](https://doi.org/10.1371/journal.ppat.1004708)
236. Chapman EG, Moon SL, Wilusz J, Kieft JS (2014) RNA structures that resist degradation by Xrn1 produce a pathogenic Dengue virus RNA. *Elife*. doi:[10.7554/eLife.01892](https://doi.org/10.7554/eLife.01892)
237. Werner MS, Ruthenburg AJ (2015) Nuclear fractionation reveals thousands of chromatin-tethered noncoding RNAs adjacent to active genes. *Cell Rep* 12:1089–1098. doi:[10.1016/j.celrep.2015.07.033](https://doi.org/10.1016/j.celrep.2015.07.033)
238. Mondal T, Rasmussen M, Pandey GK et al (2010) Characterization of the RNA content of chromatin. *Genome Res* 20:899–907. doi:[10.1101/gr.103473.109](https://doi.org/10.1101/gr.103473.109)

239. Singh DK, Prasanth KV (2013) Functional insights into the role of nuclear-retained long non-coding RNAs in gene expression control in mammalian cells. *Chromosome Res* 21:695–711. doi:[10.1007/s10577-013-9391-7](https://doi.org/10.1007/s10577-013-9391-7)
240. Zheng R, Shen Z, Tripathi V et al (2010) Polypurine-repeat-containing RNAs: a novel class of long non-coding RNA in mammalian cells. *J Cell Sci* 123:3734–3744. doi:[10.1242/jcs.070466](https://doi.org/10.1242/jcs.070466)
241. Rackham O, Shearwood A-MJ, Mercer TR et al (2011) Long noncoding RNAs are generated from the mitochondrial genome and regulated by nuclear-encoded proteins. *RNA* 17:2085–2093. doi:[10.1261/rna.029405.111](https://doi.org/10.1261/rna.029405.111)
242. Burzio VA, Villota C, Villegas J et al (2009) Expression of a family of noncoding mitochondrial RNAs distinguishes normal from cancer cells. *Proc Natl Acad Sci* 106:9430–9434. doi:[10.1073/pnas.0903086106](https://doi.org/10.1073/pnas.0903086106)
243. Anandakumar S, Vijayakumar S, Centre for Advanced Study in Crystallography and Biophysics, University of Madras et al (2015) Mammalian mitochondrial ncRNA database. *Bioinformatics* 11:512–514. doi: [10.6026/97320630011512](https://doi.org/10.6026/97320630011512)
244. Landerer E, Villegas J, Burzio VA et al (2011) Nuclear localization of the mitochondrial ncRNAs in normal and cancer cells. *Cell Oncol* 34:297–305. doi:[10.1007/s13402-011-0018-8](https://doi.org/10.1007/s13402-011-0018-8)
245. Vidaurre S, Fitzpatrick C, Burzio VA et al (2014) Down-regulation of the antisense mitochondrial non-coding RNAs (ncRNAs) is a unique vulnerability of cancer cells and a potential target for cancer therapy. *J Biol Chem* 289:27182–27198. doi:[10.1074/jbc.M114.558841](https://doi.org/10.1074/jbc.M114.558841)
246. Lobos-González L, Silva V, Araya M et al (2016) Targeting antisense mitochondrial ncRNAs inhibits murine melanoma tumor growth and metastasis through reduction in survival and invasion factors. *Oncotarget*. doi:[10.18632/oncotarget.11110](https://doi.org/10.18632/oncotarget.11110)
247. Guo X, Gao L, Wang Y et al (2015) Advances in long noncoding RNAs: identification, structure prediction and function annotation. *Brief Funct Genomics*. doi:[10.1093/bfgp/elt022](https://doi.org/10.1093/bfgp/elt022)
248. Quinn JJ, Chang HY (2015) Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet* 17:47–62. doi:[10.1038/nrg.2015.10](https://doi.org/10.1038/nrg.2015.10)
249. Han P, Chang C-P (2015) Long non-coding RNA and chromatin remodeling. *RNA Biol* 12:1094–1098. doi:[10.1080/15476286.2015.1063770](https://doi.org/10.1080/15476286.2015.1063770)
250. Davidovich C, Cech TR (2015) The recruitment of chromatin modifiers by long noncoding RNAs: lessons from PRC2. *RNA* 21:2007–2022. doi:[10.1261/rna.053918.115](https://doi.org/10.1261/rna.053918.115)
251. Yoon J-H, Abdelmohsen K, Kim J et al (2013) Scaffold function of long non-coding RNA HOTAIR in protein ubiquitination. *Nat Commun*. doi:[10.1038/ncomms3939](https://doi.org/10.1038/ncomms3939)
252. Lee S, Kopp F, Chang T-C et al (2016) Noncoding RNA NORAD regulates genomic stability by sequestering PUMILIO proteins. *Cell* 164:69–80. doi:[10.1016/j.cell.2015.12.017](https://doi.org/10.1016/j.cell.2015.12.017)
253. Tsai M-C, Manor O, Wan Y et al (2010) Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 329:689–693. doi:[10.1126/science.1192002](https://doi.org/10.1126/science.1192002)
254. Chujo T, Yamazaki T, Hirose T (2016) Architectural RNAs (arcRNAs): a class of long noncoding RNAs that function as the scaffold of nuclear bodies. *Biochim Biophys Acta* 1859:139–146. doi:[10.1016/j.bbagr.2015.05.007](https://doi.org/10.1016/j.bbagr.2015.05.007)
255. Yamazaki T, Hirose T (2015) The building process of the functional paraspeckle with long non-coding RNAs. *Front Biosci* 7:1–47. doi:[10.2741/715](https://doi.org/10.2741/715)
256. Postepska-Igielska A, Giwojna A, Gasri-Plotnitsky L et al (2015) LncRNA Khps1 regulates expression of the proto-oncogene SPHK1 via triplex-mediated changes in chromatin structure. *Mol Cell* 60:626–636. doi:[10.1016/j.molcel.2015.10.001](https://doi.org/10.1016/j.molcel.2015.10.001)
257. Mondal T, Subhash S, Vaid R et al (2015) MEG3 long noncoding RNA regulates the TGF- β pathway genes through formation of RNA–DNA triplex structures. *Nat Commun* 6:7743. doi:[10.1038/ncomms8743](https://doi.org/10.1038/ncomms8743)
258. Kino T, Hurt DE, Ichijo T et al (2010) Noncoding RNA Gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci Signal* 3:ra8. doi:[10.1126/scisignal.2000568](https://doi.org/10.1126/scisignal.2000568)
259. Wang P, Xue Y, Han Y et al (2014) The STAT3-binding long noncoding RNA lnc-DC controls human dendritic cell differentiation. *Science* 344:310–313. doi:[10.1126/science.1251456](https://doi.org/10.1126/science.1251456)

260. Ørom UA, Derrien T, Beringer M et al (2010) Long noncoding RNAs with enhancer-like function in human cells. *Cell* 143:46–58. doi:[10.1016/j.cell.2010.09.001](https://doi.org/10.1016/j.cell.2010.09.001)
261. Lai F, Orom UA, Cesaroni M et al (2013) Activating RNAs associate with mediator to enhance chromatin architecture and transcription. *Nature* 494:497–501. doi:[10.1038/nature11884](https://doi.org/10.1038/nature11884)
262. Yao H, Brick K, Evrard Y et al (2010) Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. *Genes Dev* 24:2543–2555. doi:[10.1101/gad.1967810](https://doi.org/10.1101/gad.1967810)
263. Gomez JA, Wapinski OL, Yang YW et al (2013) The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon- γ locus. *Cell* 152:743–754. doi:[10.1016/j.cell.2013.01.015](https://doi.org/10.1016/j.cell.2013.01.015)
264. Szcześniak MW, Makałowska I (2016) lncRNA-RNA interactions across the human transcriptome. *PLoS One* 11:e0150353. doi:[10.1371/journal.pone.0150353](https://doi.org/10.1371/journal.pone.0150353)
265. An Y, Furber KL, Ji S (2016) Pseudogenes regulate parental gene expression *via* ceRNA network. *J Cell Mol Med*. doi:[10.1111/jcmm.12952](https://doi.org/10.1111/jcmm.12952)
266. Thomson DW, Dinger ME (2016) Endogenous microRNA sponges: evidence and controversy. *Nat Rev Genet* 17:272–283. doi:[10.1038/nrg.2016.20](https://doi.org/10.1038/nrg.2016.20)
267. Tay Y, Rinn J, Pandolfi PP (2014) The multilayered complexity of ceRNA crosstalk and competition. *Nature* 505:344–352. doi:[10.1038/nature12986](https://doi.org/10.1038/nature12986)
268. Giovarelli M, Bucci G, Ramos A et al (2014) H19 long noncoding RNA controls the mRNA decay promoting function of KSRP. *Proc Natl Acad Sci* 111:E5023–E5028. doi:[10.1073/pnas.1415098111](https://doi.org/10.1073/pnas.1415098111)
269. Dey BK, Pfeifer K, Dutta A (2014) 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* 28:491–501. doi:[10.1101/gad.234419.113](https://doi.org/10.1101/gad.234419.113)
270. Ha H, Song J, Wang S et al (2014) A comprehensive analysis of piRNAs from adult human testis and their relationship with genes and mobile elements. *BMC Genomics* 15:545. doi:[10.1186/1471-2164-15-545](https://doi.org/10.1186/1471-2164-15-545)
271. Carlile M, Swan D, Jackson K et al (2009) Strand selective generation of endo-siRNAs from the Na/phosphate transporter gene *Slc34a1* in murine tissues. *Nucleic Acids Res* 37:2274–2282. doi:[10.1093/nar/gkp088](https://doi.org/10.1093/nar/gkp088)
272. Werner A (2013) Biological functions of natural antisense transcripts. *BMC Biol* 11:31. doi:[10.1186/1741-7007-11-31](https://doi.org/10.1186/1741-7007-11-31)
273. Abdelmohsen K, Panda A, Kang M-J et al (2013) Senescence-associated lncRNAs: senescence-associated long noncoding RNAs. *Aging Cell* 12:890–900. doi:[10.1111/accel.12115](https://doi.org/10.1111/accel.12115)
274. C-L W, Wang Y, Jin B et al (2015) Senescence-associated long non-coding RNA (*SALNR*) delays oncogene-induced senescence through NF90 regulation. *J Biol Chem* 290:30175–30192. doi:[10.1074/jbc.M115.661785](https://doi.org/10.1074/jbc.M115.661785)
275. Choudhry H, Harris AL, McIntyre A (2016) The tumour hypoxia induced non-coding transcriptome. *Mol Aspects Med* 47–48:35–53. doi:[10.1016/j.mam.2016.01.003](https://doi.org/10.1016/j.mam.2016.01.003)
276. Fort A, Hashimoto K, Yamada D et al (2014) Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. *Nat Genet* 46:558–566. doi:[10.1038/ng.2965](https://doi.org/10.1038/ng.2965)
277. Prensner JR, Iyer MK, Balbin OA et al (2011) Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nat Biotechnol* 29:742–749. doi:[10.1038/nbt.1914](https://doi.org/10.1038/nbt.1914)
278. Mattick JS, Rinn JL (2015) Discovery and annotation of long noncoding RNAs. *Nat Struct Mol Biol* 22:5–7. doi:[10.1038/nsmb.2942](https://doi.org/10.1038/nsmb.2942)
279. St. Laurent G, Wahlestedt C, Kapranov P (2015) The landscape of long noncoding RNA classification. *Trends Genet* 31:239–251. doi:[10.1016/j.tig.2015.03.007](https://doi.org/10.1016/j.tig.2015.03.007)
280. Laurent GS, Vyatkin Y, Antonets D et al (2016) Functional annotation of the vlinc class of non-coding RNAs using systems biology approach. *Nucleic Acids Res* 44:3233–3252. doi:[10.1093/nar/gkw162](https://doi.org/10.1093/nar/gkw162)

281. Lin R, Maeda S, Liu C et al (2007) A large noncoding RNA is a marker for murine hepatocellular carcinomas and a spectrum of human carcinomas. *Oncogene* 26:851–858. doi:[10.1038/sj.onc.1209846](https://doi.org/10.1038/sj.onc.1209846)
282. Clemson CM, Hutchinson JN, Sara SA et al (2009) An architectural role for a nuclear non-coding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol Cell* 33:717–726. doi:[10.1016/j.molcel.2009.01.026](https://doi.org/10.1016/j.molcel.2009.01.026)
283. Nam J-W, Bartel DP (2012) Long noncoding RNAs in *C. elegans*. *Genome Res* 22:2529–2540. doi:[10.1101/gr.140475.112](https://doi.org/10.1101/gr.140475.112)
284. Beltran M, Puig I, Pena C et al (2008) A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. *Genes Dev* 22:756–769. doi:[10.1101/gad.455708](https://doi.org/10.1101/gad.455708)
285. Salzman J, Gawad C, Wang PL et al (2012) Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS One* 7:e30733. doi:[10.1371/journal.pone.0030733](https://doi.org/10.1371/journal.pone.0030733)
286. Jeck WR, Sorrentino JA, Wang K et al (2013) Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* 19:141–157. doi:[10.1261/rna.035667.112](https://doi.org/10.1261/rna.035667.112)
287. Gardner EJ, Nizami ZF, Talbot CC, Gall JG (2012) Stable intronic sequence RNA (sisRNA), a new class of noncoding RNA from the oocyte nucleus of *Xenopus tropicalis*. *Genes Dev* 26:2550–2559. doi:[10.1101/gad.202184.112](https://doi.org/10.1101/gad.202184.112)
288. Talhouarne GJS, Gall JG (2014) Lariat intronic RNAs in the cytoplasm of *Xenopus tropicalis* oocytes. *RNA* 20:1476–1487. doi:[10.1261/rna.045781.114](https://doi.org/10.1261/rna.045781.114)
289. Pek JW, Osman I, Tay ML-I, Zheng RT (2015) Stable intronic sequence RNAs have possible regulatory roles in *Drosophila melanogaster*. *J Cell Biol* 211:243–251. doi:[10.1083/jcb.201507065](https://doi.org/10.1083/jcb.201507065)
290. Rapicavoli NA, Qu K, Zhang J et al (2013) A mammalian pseudogene lncRNA at the interface of inflammation and anti-inflammatory therapeutics. *Elife*. doi:[10.7554/eLife.00762](https://doi.org/10.7554/eLife.00762)
291. Vembar SS, Scherf A, Siegel TN (2014) Noncoding RNAs as emerging regulators of *Plasmodium falciparum* virulence gene expression. *Curr Opin Microbiol* 20:153–161. doi:[10.1016/j.mib.2014.06.013](https://doi.org/10.1016/j.mib.2014.06.013)
292. Liz J, Portela A, Soler M et al (2014) Regulation of pri-miRNA processing by a long noncoding RNA transcribed from an ultraconserved region. *Mol Cell* 55:138–147. doi:[10.1016/j.molcel.2014.05.005](https://doi.org/10.1016/j.molcel.2014.05.005)
293. Ilott NE, Heward JA, Roux B et al (2014) Long non-coding RNAs and enhancer RNAs regulate the lipopolysaccharide-induced inflammatory response in human monocytes. *Nat Commun*. doi:[10.1038/ncomms4979](https://doi.org/10.1038/ncomms4979)
294. Bianchessi V, Badi I, Bertolotti M et al (2015) The mitochondrial lncRNA ASncmtRNA-2 is induced in aging and replicative senescence in Endothelial Cells. *J Mol Cell Cardiol* 81:62–70. doi:[10.1016/j.yjmcc.2015.01.012](https://doi.org/10.1016/j.yjmcc.2015.01.012)
295. Zhang Y, He Q, Hu Z et al (2016) Long noncoding RNA LINP1 regulates repair of DNA double-strand breaks in triple-negative breast cancer. *Nat Struct Mol Biol* 23:522–530. doi:[10.1038/nsmb.3211](https://doi.org/10.1038/nsmb.3211)