

# Chapter 15

## Noncoding RNAs in Chromatin Organization and Transcription Regulation: An Epigenetic View

Karthigeyan Dhanasekaran, Sujata Kumari, and Chandrasekhar Kanduri

**Abstract** The Genome of a eukaryotic cell harbors genetic material in the form of DNA which carries the hereditary information encoded in their bases. Nucleotide bases of DNA are transcribed into complimentary RNA bases which are further translated into protein, performing defined set of functions. The central dogma of life ensures sequential flow of genetic information among these biopolymers. Noncoding RNAs (ncRNAs) serve as exceptions for this principle as they do not code for any protein. Nevertheless, a major portion of the human transcriptome comprises noncoding RNAs. These RNAs vary in size, as well as they vary in the spatio-temporal distribution. These ncRNAs are functional and are shown to be involved in diverse cellular activities. Precise location and expression of ncRNA is essential for the cellular homeostasis. Failures of these events ultimately results in numerous disease conditions including cancer. The present review lists out the various classes of ncRNAs with a special emphasis on their role in chromatin organization and transcription regulation.

---

K. Dhanasekaran • S. Kumari

Transcription and Disease Laboratory, Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Jakkur P.O., Bangalore 560064, India

C. Kanduri (✉)

Science for Life Laboratory, Department of Immunology, Genetics and Pathology, Rudbeck Laboratory, Uppsala University, Dag HammarskjöldsVäg 20, S75185 Uppsala, Sweden

Department of Medical and Clinical genetics, Institute of Biomedicine, The Sahlgrenska Academy, University of Gothenburg, Gothenberg, Sweden  
e-mail: kanduri.chandrasekhar@igp.uu.se

## 15.1 Overview of ncRNA

Noncoding RNAs (ncRNAs) are a large group of functional RNA transcribed by RNA polymerases, but never translated into protein. A decade before the discovery of these noncoding transcripts, most of the sequences in our genome were believed to be part of “Junk DNA”. Soon the genome sequencing projects revealed that only a mere 2% of the human genome codes for protein coding genes and almost 98% of the human transcriptome represents the ncRNA. They include well-characterized transfer RNAs and ribosomal RNAs involved in the process of translation, as well as a huge class of other regulatory ncRNAs which have been shown to play a crucial role in gene regulation. ncRNAs in general function as adaptors for the recognition of a particular nucleotide sequence in the target which is later positioned into the enzymatic molecule associated with the specific class of ncRNA. These functional ncRNA are involved in key cellular processes including transcriptional regulation, RNA processing and modification, protein trafficking, genome stability, mRNA stability, and even protein degradation (Hüttenhofer et al. 2005).

Terminologies like “Junk DNA” and “Transcriptional noise” have been challenged since the discovery of regulatory RNAs that are transcribed by both Pol II and Pol III. The list of known ncRNAs is growing larger in numbers since the completion of whole transcriptome analyses like the ENCODE Pilot Project (Birney et al. 2007), the mouse cDNA project FANTOM, and a series of other large scale transcriptome studies performed to fish these transcribed fragments (‘transfrags’) using various forms of high throughput tiling arrays, ESTs, SAGE tags and RACE techniques. Ultimately the present scenario of the transcriptome is that ncRNA are vast in number covering a huge proportion of the genome consisting of overlapping, bi-directional transcripts. The major obstacle in the identification of these transcripts is that majority of them are expressed in a short spatio-temporal frame, thus it is difficult to recognize such transcripts even after employing the most advanced deep sequencing techniques. Though the ncRNAs have coexisted with protein coding genes they were kept under curtain mainly because of the model organism taken for such studies. Also the mutations that were considered were often expected to have a major impact on the phenotypic outcome of such genetic screens. Mutation in a protein-coding gene can have severe effects on the structure and function of the protein which ultimately shows an altered phenotype, which is often visible in the regular genetic screening due to the high penetrance. On the other hand recessive mutation phenotypes and single-base mutations are harder to identify in comparison to insertions/deletions (Eddy 2001; Kavanaugh and Dietrich 2009). The reason for a shift in the focus of our research towards protein coding genes is that most of the genetic screening techniques and the methodologies followed during the early era of genomics had an inherent bias toward the scanning of known exons and the flanking sequences in that region. Most of the bioinformatic search tools were based on the signatures of protein coding genic regions which does not hold good for the ncRNA prediction especially when they are transcribed from the intergenic deserts. The other reason is that comparative studies on the sequences of ncRNA have failed to show any healthy

conservation among the known transcripts reported till date. However, studies on a handful of functional ncRNAs indicate that they carry out common functions via conserved secondary structures, indicating that despite having no sequence similarity they seem to harbor conserved secondary structures. Even though a large number of genomes have been sequenced, the number and diversity of ncRNA-encoding genes is largely unknown, especially due to the incompleteness of the list of various ncRNAs (van Bakel et al. 2010; Farh et al. 2005).

## 15.2 Classification and Evolution of ncRNA

The existing classes of ncRNAs are in general transcribed by all three possible modes of transcription. The pre-rRNA (28S, 18S, 5.8S) are transcribed by Pol I, and some of the snRNAs and LINEs are transcribed by Pol II where as SINES, snRNA, 7SL RNA, etc. are transcribed by Pol III transcription machinery. Broadly, ncRNAs can be classified into “housekeeping” and “regulatory” ncRNAs (Morey and Avner 2004). Housekeeping ncRNAs are constitutively expressed and involved in processes like translation, RNA processing, RNA modifications, protein trafficking and genomic stability required for normal cell viability, whereas the regulatory ncRNAs, often expressed in certain specified tissues during different stages of development or in response to an external stimuli and they are comprised of RNAs involved in the process of gene expression/regulation and chromatin organization. The other way to classify ncRNA is based on the size of the functional transcripts as long (9,999–10,000nt), medium (200–999nt), small (24–199nt) and micro (18–31nt) ncRNAs. Noncoding transcripts can also be classified based on the sequence origin as sense or antisense transcripts from the genic, intronic and intergenic region of the genome.

The human genome has approximately 27,161 genes (Flicek et al. 2008) in total, of which about 4,421 are ncRNA genes. Literature on the existing ncRNAs indicates an unequivocal correlation between the rise in the number of noncoding transcripts and the complexity of an organism (Amaral and Mattick 2008). Events responsible for such evolution are by gene duplication, mutation, horizontal transfer and integration of genetic material between different pathogen and host across various phases of evolution. Especially the non-genic deserts are more prone towards such events since the pressure to preserve the functionality of the protein coding genes does not apply to ncRNA coding intergenic regions. Also the drastic mutations can be well tolerated since the constraint for most of the ncRNA is to maintain its secondary structure. Moreover the base change in one strand is often compensated by a complementary mutation across the paired strand. Though ncRNAs in general are rapidly evolving they are earmarked by conservation in their secondary structure and are often found associated with regions spanning promoters, splice junctions, and other regions with specific chromatin signatures in relation to the spatiotemporal expression and subcellular localization pattern (Pang et al. 2006; Bradley et al. 2009).

### 15.3 Housekeeping ncRNA

**rRNAs:** Generally eukaryotes have many copies of the rRNA genes organized in tandem repeats; in humans approximately 300–400 rDNA repeats are present in five clusters. All mammalian cells possess two mitochondrial (12S and 16S) and four cytoplasmic rRNA (the 28S, 5.8S, 18S, and 5S subunits) transcribed by RNA polymerase I except 5S rRNA which is transcribed by RNA polymerase III. Most of these rRNAs constitute the active site of ribosomes and also aids in maintaining the fidelity of translation.

**tRNAs:** tRNAs are the adaptor molecules that aid in sequence specific incorporation of various amino acids according to the code present in an mRNA. According to the tRNADB, which is a curated database of tRNA, there are 22 known mitochondrial tRNA genes and 497 nuclear tRNA genes known in humans but the number varies a lot in different organisms (Abe et al. 2011). These genes are found on all chromosomes, except chromosome number 22 and Y chromosome of humans (Lander et al. 2001). tRNAs are transcribed by RNA polymerase III as pre-tRNAs in the nucleus (Dieci et al. 2007) which undergo extensive posttranscriptional modifications. The adaptor function of tRNA lies in its three-dimensional structure wherein one end of the tRNA carries the anticodon that serves as a genetic code to recognize the codon in mRNA during protein biosynthesis. Transfer RNA-like structures (tRNA-like structures) are a separate class of RNA sequences transcribed from the genome of many plant RNA viruses, which have a tRNA like tertiary structure (Crick 1968). These tRNA-like structures mimic some tRNA functions, such as aminoacylation, but only three aminoacylation specificities, valine, histidine and tyrosine have been reported till date (Dreher 2009). Such tRNA-like structures are also known to increase the stability of RNA viruses by encapsulating its RNA genome (Mans et al. 1991). In addition, they act as 3'-translational enhancers (Matsuda and Dreher 2004) and regulators of minus strand synthesis.

**tel-sRNAs:** Telomere specific small RNAs called as tel-sRNA are found exclusively in the telomeric region of the genome. These small  $\Pi$ -like RNAs are associated asymmetrically to the G-rich strand of telomers which are Dicer-independent, 2'-O-methylated at the 3' terminus, and conserved from protozoa to mammalian cells. tel-sRNAs were shown for the first time in mouse genome where they aid in the establishment and maintenance of heterochromatin in the telomeric loci (Cao et al. 2009).

**tmRNA:** The bacterial *tmRNA* has both tRNA-like and mRNA-like function e.g., 10Sa RNA or SsrA. *tmRNA* engages the problematic messenger RNAs and recycles the 70S ribosomes ultimately that incorporates a series of alanine residues which are earmarked for the degradation of those incomplete peptides (Gillet and Felden 2001). For more information about tmRNAs refer to tmRDB, an exclusive database for tmRNAs.

**SRP RNA:** The RNA component of the signal recognition particle (SRP) ribonucleoprotein complex also known as 7SL, 6S, ffs, or 4.5S RNA, is a universally conserved ncRNA (Rosenblad et al. 2009) that directs the newly synthesized

proteins within a cell to the endoplasmic reticulum either co-translationally or post-translationally thereby allowing them to be secreted.

**snRNAs:** small nuclear RNAs can be broadly classified into two. Firstly, the Sm-class of snRNA that possess a 5'-trimethylguanosine cap, 3'-stem-loop and heteroheptameric ring structure that binds to sm-proteins. These non-polyadenylated snRNAs are transcribed by Pol II and processed by integrator. The processed mature snRNAs finally aid in splicing out introns from the pre-mRNA. Secondly, Lsm-class RNAs that possess a monomethylphosphate cap and a 3'-stem-loop with uridine rich heteroheptameric ring that binds Lsm-proteins. Pol III transcribes such Lsm-snRNAs using external promoters and Uridine stretch as terminator (Segref et al. 2001). Almost all Lsm-snRNPs are assembled in the nucleus within the cajal body for a brief period after which they diffuse out in the nucleoplasm till they reach their specific nuclear domains like, perichromatin fibrils and interchromatin granule clusters. snRNPs containing such ncRNAs form the core of the spliceosome which are the catalytic centers for splicing introns from pre-mRNA (Matera et al. 2007). Among the snRNAs U7 snRNA needs a special mention which is involved in the processing of 3' end of histone genes of eukaryotes which possess a unique stem-loop structure instead of a poly-A tail. However, snRNAs are not just restricted to splicing events alone as they have been shown to regulate transcription, independent of their splicing function. 7SK is one such snRNA which mediates Pol II transcriptional inhibition via its interaction with P-TEFb. Apart from the above mentioned snRNAs there are numerous other snRNAs which carryout important biological functions like, RNA Pol III transcribed snaR-A RNA, Intergenic spacer RNA (IGS RNAs) etc., are discussed under regulatory ncRNAs section.

**SmY-RNA:** These ncRNAs belong to a Small nuclear class of ncRNAs in nematodes SmY-RNA were discovered in *Ascaris lumbricoides* during the year 1996 (Maroney et al. 1996). Based on the evidence obtained from the studies carried out in a related species i.e., *C. elegans* SmY-RNA is believed to be in complex with the spliced-leader RNA and involved in mRNA trans-splicing (MacMorris et al. 2007).

**snoRNAs:** Small nucleolar RNAs, as the name implies, are retained within the nucleolus and aid as guide strands for incorporating the specific modification like methylations and pseudouridylations, onto other RNA molecules like tRNA, rRNA, snRNA etc. snoRNAs can be further classified into C/D Box RNAs, H/ACA Box RNAs, composite C/D Box and H/ACA Box RNAs and Orphan snoRNAs (Bachellerie et al. 2002; Samarsky et al. 1998). In general C/D box members guide 2'-O-ribose-methylations and H/ACA members guide pseudouridylation. snoRNAs are defined by the characteristic secondary structure formed by the signature sequences which varies slightly in the composite snoRNAs. The composite snoRNA contains both C/D and H/ACA box and are retained in the cajal bodies and hence, named as "scaRNAs" (Jády and Kiss 2001). U85 a typical example of composite snoRNA, functions in both 2'-O-ribose methylation and pseudouridylation of snRNA. On the contrary, there are snoRNAs with unidentified substrates that are grouped under the Orphan snoRNAs. Apart from their function in guiding modifications for maintaining a stable pool of ncRNA, some members are even known to act like

miRNAs with exclusive regulatory functions and hence they are discussed under the regulatory RNAs section.

## 15.4 Regulatory RNAs

Regulatory RNAs comprise a subset of both long and small mRNAs having gene expression regulatory function. Regulatory ncRNAs especially the small ncRNAs in general, form base pairs with other RNA or DNA and constitute RNA:RNA or RNA:DNA duplexes. These duplexes are recognised by different complexes like RNA induced silencing complex (RISC), RNA induced transcriptional silencing (RITS) or RNA editing enzymes which act to decipher downstream consequences. These cis-acting regulatory sequences are generally found in non-coding regions of mRNAs and pre-mRNAs. Untranslated regions (UTRs) of mRNA generally act as binding sites for some trans acting regulatory RNAs, though they are also known to form secondary structures facilitating binding of regulatory proteins that in turn control stability, function or localization of mRNAs (Gebauer and Hentze 2004; Moore 2005). Splice junctions provide yet another cis regulatory sequences which along with the aid of spliceosomal snRNAs and other components of spliceosome, a ribonucleoprotein (RNP) complex that controls splicing of the primary transcript (Nilsen 2003; Valadkhan et al. 2007). Regulatory ncRNAs are generally categorized in two classes namely the small (<200 nucleotide) and large (>200 nucleotide) regulatory ncRNAs.

### 15.4.1 Small Regulatory RNA

There are numerous regulatory RNAs that are <200nt long and show unique spatio-temporal expression in comparison to housekeeping ncRNAs. Some of the well characterised small regulatory RNAs are discussed in this section.

Regulatory Small nucleolar RNAs (SnoRNAs): Some SnoRNAs show tissue specific expression, like tandemly arranged repeated intron-encoded C/D snoRNA genes in the region downstream from the GTL2 gene at 14q32 show brain specificity. These snoRNA genes associate with human imprinted 14q32 domain suggesting their regulatory role in epigenetic imprinting process (Cavaillé et al. 2002). Two other brain specific snoRNAs, HBII-52 and HBII-85 were reported to be absent from the cortex of a patient with Prader-Willi syndrome (PWS), which is a neuro-genetic disease resulting from a deficiency of paternal gene expression, indicating their role in the etiology of PWS (Cavaillé et al. 2000). Further, it was shown that the snoRNA HBII-52 regulates alternative splicing of the Serotonin Receptor 2C. Lack of HBII-52 in PWS patients generate different messenger RNA (mRNA) isoforms which leads to the loss of high-efficacy serotonin receptor, which could contribute to the disease (Kishore and Stamm 2006).

Apart from regulating mRNA transcription which is central for the regulation of gene expression, other biological reactions comprehending gene expression like mRNA turnover, gene silencing and translation are also controlled by ncRNAs like miRNAs and short interfering RNAs. miRNAs and siRNAs are 21–25 nt long RNAs derived from double stranded RNA precursors. Origin of miRNA is endogenous from short hairpin precursor RNAs whereas siRNA are mostly exogenous from double stranded RNAs or long hairpins. These small RNAs regulate gene expression through translational suppression (post transcriptional) and/or mRNA degradation (transcriptional) by perfect/non-perfect match formed between miRNA and target mRNA (Mattick and Makunin 2005; Yekta et al. 2004; Mansfield et al. 2004). siRNA is also known to regulate gene expression by modulating chromatin structure (discussed in next section). MicroRNA genes are generally transcribed by RNA polymerase II generating primary miRNA (pri-miRNA). Pri-miRNAs are several kilobases long and possess stemloop structure. Pri-miRNAs are cleaved by RNase III, enzyme Droscha, containing multiprotein complex, producing ~70-nt hairpin precursor miRNA (pre-miRNA). Pre-miRNA is exported to cytoplasm where it is processed into ~22 nt miRNA duplex by another RNase III enzyme, Dicer (Bushati and Cohen 2007 for review). Dicer along with protein argonaute form a complex triggering the assembly of ribonucleoprotein complex called as RNA-induced silencing complex (RISC). One strand of miRNA gets incorporated into RISC and guides the complex to target RNA for base pairing. In case of perfect match with target RNA it is cleaved and if base pairing is imperfect and the binding is strong enough to hold, then the translation is repressed. Major mode of action of animal miRNA involves translational repression rather than RNA degradation unlike the plant miRNAs (Millar and Waterhouse 2005a). Target recognition of miRNA mainly depends on the stringency of base pair match at the 5' end of miRNA called as the "seed region". Nevertheless, when the 5' sites are dominant, it can function with or without 3' pairing support. In case of insufficient 5' pairing in some miRNAs, the 3' compensatory sites play their part by strong pairing with the seed region sequence.

esiRNAs: Initially endogenous siRNAs (esiRNAs) have been detected only in organisms that possess RNA-dependent RNA polymerases (RDRPs) and absent in others which lack endogenous dsRNA (Millar and Waterhouse 2005a, b). However, other sources of dsRNAs including long hairpin structures generated from the palindromic sequences and dsRNAs generated by the annealing of complementary RNAs that are synthesized by two opposing transcription units in the same loci. Such dsRNAs have now proven to be the source of esiRNAs in both *D. melanogaster* and mice (Watanabe et al. 2008; Tam et al. 2008). In *Drosophila* esiRNAs have been shown to play important role in the formation of heterochromatin within the somatic tissues (Fagegaltier et al. 2009). esiRNAs have also been implicated in suppressing the expression of mobile genetic elements. Mice deficient for Dicer showed elevated expression of only certain transposable elements which are believed to be affected by the esiRNA pathway but the exact mechanism is yet to be discovered (Nilsen 2008).

Viral miRNAs: These are the viral transcripts that are generally employed in processes like immune recognition, cell survival, angiogenesis, proliferation and cell

differentiation upon infection of the host cells (Pfeffer et al. 2004; Gottwein et al. 2007; Grey et al. 2010). A recent review on viral miRNAs has listed the known viral miRNAs from different viral species (Plaisance-Bonstaff and Renne 2011). miRNAs in general show a higher degree of conservation but viral miRNAs on the other hand shows very poor sequence homology between viruses. Viral miRNAs targets only a small sub population of viral transcripts and obviously they target the majority of host mRNA transcripts thereby regulating their expression to substantial level to create a conducive, microenvironment for their survival and proliferation of viruses (Grey et al. 2010). Virus encoded miRNAs are known to act as suppressors of RNAi, modulating the host miRNAs and also incorporates epigenetic changes in the host which may aid in the viral oncogenesis (Scaria and Jadhav 2007).

**Y RNAs:** These are small noncoding RNAs that function as integral part of the Ro RNP. The Ro RNP was discovered by Lerner et al. in systemic lupus erythematosus patients. So far four Y RNA species have been discovered in humans namely hY1 (hY2 is a truncated form of hY1), hY3, hY4, and hY5 RNAs ranging in size from 83 to 112 nucleotides (Hendrick et al. 1981). Y RNAs are expressed in all vertebrate species studied (Perreault et al. 2007). Among the invertebrates Y RNA orthologues have been reported in *Caenorhabditis elegans* (Van Horn et al. 1995; Boria et al. 2010) and *Deinococcus radiodurans* (Chen et al. 2000), but no orthologues in yeasts, plants, or insects. In *Deinococcus*, Y RNAs are reported to be involved in 23S rRNA maturation (Chen et al. 2007). while the human Y RNAs (hY RNAs) aid in the process of chromosomal DNA replication which ultimately ensures a completely semi-conservative mode of replication throughout the genome. They have been implicated in either the initiation steps to establish an active replication forks or for elongation steps during DNA replication fork progression (Christov et al. 2006). Recently hY RNAs were shown to be even overexpressed in solid tumours, that aids in cell proliferation (Christov et al. 2008). Nevertheless the cause and consequences are not yet completely deciphered.

**TSSa-RNAs:** Transcription start site-associated RNAs as their name suggests are transcribed either as sense or antisense transcripts from region flanking the active promoters, with peaks of antisense and sense short RNAs peaking between nucleotides -100 and -300 nucleotides upstream and 0 to +50 nucleotides downstream of TSS, respectively. In yeast such TSSa-RNAs are called as cryptic unstable transcripts (CUTs) and stable unannotated transcripts (SUTs) (Neil et al. 2009; Wyers et al. 2005; Xu et al. 2009). TSSa-RNAs are 20–90 nt (Seila et al. 2008) in length and have been proposed to aid in maintaining poised chromatin state at the promoter regions for downstream transcriptional regulatory steps. The transcription initiation factors, RNAPII and the K4-trimethylated histone H3, occupy the same position over the chromatin where TSSa-RNA; whereas, K79-dimethylated histone H3, is located downstream of TSSs. Recently, a long promoter associated ncRNA (pncRNA) has been identified which repress the protein coding transcripts in cis via an RNA binding protein called TLS (Translocated in liposarcoma) that mediates transcription repression through HAT inhibition. Refer Sect. 15.4.3 for more details.



vRNAs: vault RNAs are integral part of the vault particles that were discovered as a vault ribonucleoprotein complex implicated in multidrug resistance and intracellular transport. Generally these are 100 bases long and transcribed by Pol III. vRNAs via a DICER mechanism generate small vault RNAs (svRNAs) that act like miRNAs in downregulating the expression of CYP3A4, an enzyme essential for drug metabolism (Persson et al. 2009).

### 15.4.2 Long Regulatory ncRNA

Long Regulatory ncRNA (lncRNAs) as mentioned earlier are greater than 200 nt long and both polyadenylated, and nonpolyadenylated transcripts have been reported. Apart from intronic and intergenic (linc RNAs) lncRNAs, they are also encoded from genomic regions enriched with repetitive elements, such as telomeric repeats (TelRNAs), long terminal repeat retrotransposon elements (LINE RNAs), and short interspersed nuclear elements (B2 RNA). lncRNAs often overlap with, or intersperse between the protein-coding and noncoding transcripts. Promoter-associated transcripts, such as promoter-associated long RNAs (PALRs) and promoter upstream transcripts (PROMPTs) have been recently added to the growing list of lncRNAs. Often PROMPTs overlap with PALRs in terms of the size and the distance from promoter. Also they resemble the cryptic unstable transcripts (CUTs) seen in yeasts (Neil et al. 2009). The major roles of most of the lncRNAs are implicated in transcription regulation by altering the enhancers, promoters and other regulatory regions of a gene. This is achieved either by modulating the chromatin structure around these loci or by directly binding to the transcription factors associated to these elements. In general most of these lncRNAs seem to act in a gene-specific manner and recent evidence that lncRNAs themselves may have enhancer activity was suggested by a handful of studies which still remains open for further investigation (Mondal et al. 2010; Ørom et al. 2010).

lincRNAs: The large intergenic non-coding RNAs (lincRNAs) are one among the largest members of lncRNAs which are evolutionarily highly conserved (Guttman et al. 2009). HOTAIR, was the the first lincRNA, identified by Rinn et al. (2007), showing that HOTAIR could influence gene expression in trans by binding PRC2 and targeting it to the HOXD cluster, thereby silencing target genes in HOXD cluster (Rinn et al. 2008). More than 8,000 lincRNAs are known to exist and are well conserved across mammals (Rinn et al. 2008). They are involved in diverse biological processes, like cell-cycle regulation, immune surveillance and in the maintenance of stem cell pluripotency. Often lincRNAs associate with repressive chromatin modifying complexes hence, act as repressors in transcriptional regulatory networks. The typical example being the p53 mediated global gene repression via the lincRNA-p21 triggering apoptosis by recruiting the hnRNP-K on to the defined set of p53 responsive genes (Huarte et al. 2010). Sabine Loewer et al. later described the role of lincRNA in reprogramming events during derivation of human iPSCs

which is presently being described as lincRNA-RoR for ‘regulator of reprogramming’ (Loewer et al. 2010). These observations indicate that ncRNA has wide reach in regulation of various biological functions.

**Totally intronic ncRNAs (TIN):** E.M. Reis et al. identified the transcribed intronic ncRNAs (Reis et al. 2005) which are lincRNAs of approximately 0.6–2 kb in length. Later Helder I Nakaya et al. based on the *in silico* predictions available on data sets in different ncRNA databases and using the combined intron/exon oligoarrays they were able to point the intronic regions as key sources of potentially regulatory ncRNAs (Nakaya et al. 2007). They showed that TINs have tissue-specific expression signatures for human liver, prostate and kidney. The antisense TIN RNAs were transcribed from introns of protein-coding genes which are reported to be enriched in the ‘Regulation of transcription’ Gene Ontology category. Intronic RNAs are believed to regulate the abundance or the pattern of exon usage in protein-coding mRNAs. It has been proposed that TINs regulate the corresponding protein coding genes through transcriptional interference at promoters or through the epigenetic modulation of the chromatin architecture (Louro et al. 2009).

**T-UCRs:** David Haussler et al. (Bejerano et al. 2004) discovered a group of highly conserved transcripts called T-UCRs (Transcribed Ultra Conserved Regions) which do not code for any protein. There are about 481 such transcripts longer than 200 (bp) with 100% identity between the orthologous regions of the human, rat, and mouse genomes. Since these UCRs are often located at fragile sites in the chromatin and also associated to the genomic regions involved in cancers it is not surprising to link T-UCRs with tumorigenesis. It is also known that some of the UCRs’ expression is regulated by microRNAs abnormally expressed in human chronic lymphocytic leukemia, and the inhibition of UCR which is overexpressed in colon cancer could even induces apoptosis (Calin et al. 2007). T-UCR expression landscape in neuroblastoma suggests widespread T-UCR involvement in diverse cellular processes that are deregulated in the process of tumourigenesis.

**PROMPTs:** In mammals certain long, unstable promoter upstream transcripts (PROMPTs) initiate bidirectionally ~0.5–2.5 kb upstream of transcription start sites that are longer than the TSSa-RNAs (Preker et al. 2008). This class of RNA often overlaps with another class of bidirectional promoter-associated long RNAs known as PALRs which are longer than 200 nucleotides (Kapranov et al. 2007) and are distinct from PROMTs. Interestingly, siRNA targeted to promoter upstream regions often resulted in transcriptional gene silencing. Given that promoter upstream regions associated with bidirectional transcripts, siRNA could have mediated transcriptional silencing via promoter associated transcripts targeting to RNAi pathway (Han et al. 2007). However, the functional link between the expression of PROMTs and PALRs with cognate genes is not yet clear.

**GRC-RNAs:** A polypurine triplet repeat-rich lincRNAs, designated as GAA repeat-containing RNAs, are ~1.5 to ~4 kb long and localize to numerous intra nuclear punctate foci that associate with GAA.TTC-repeat containing genomic regions. These foci drop in number with more differentiation of the cell type. GRC-RNAs are components

of the nuclear matrix and interact with various nuclear matrix-associated proteins. In mitotic cells, GRC-RNAs localize to the midbody. The interesting part of GRC-RNA foci is that the number increases during cellular transformation (Zheng et al. 2010).

eRNAs: RNA polymerase II binding was noticed over 25% of the gene enhancers which later turned out that these occupancies were not mere landing pads rather they were more of transcription foci for the novel class of ncRNAs without polyadenylation called the Enhancer RNAs (eRNAs) (Kim et al. 2010; Ren 2010). eRNAs synthesis requires a functional promoter but the requirement of other general transcription factors or the mediator complex proteins is yet to be identified. The expression of eRNAs in the enhancer regions generally correlate with the gene activity of neighboring promoters, indicating that these transcripts may be necessary to activate the nearby promoters either by facilitating the formation of more open chromatin or via promoting enhancer promoter communications. Currently RNAi strategies are employed to decipher the precise mechanism of this class of regulatory ncRNAs.

mlncRNA: The mRNA-like ncRNAs are transcribed by Pol II and poladenylated at 3' and capped at 5' ends. Most of the members are known to be dysregulated in expression during the pathogenesis of multiple human diseases but their functional roles are yet to be assigned. Studies done so far strongly suggest that their expression is tightly regulated to specific subcellular compartments of specific tissues like brain but the exact role of these RNAs are not known (Inagaki et al. 2005; Jiang et al. 2011).

### ***15.4.3 Small and Long Noncoding RNAs in Transcription Regulation***

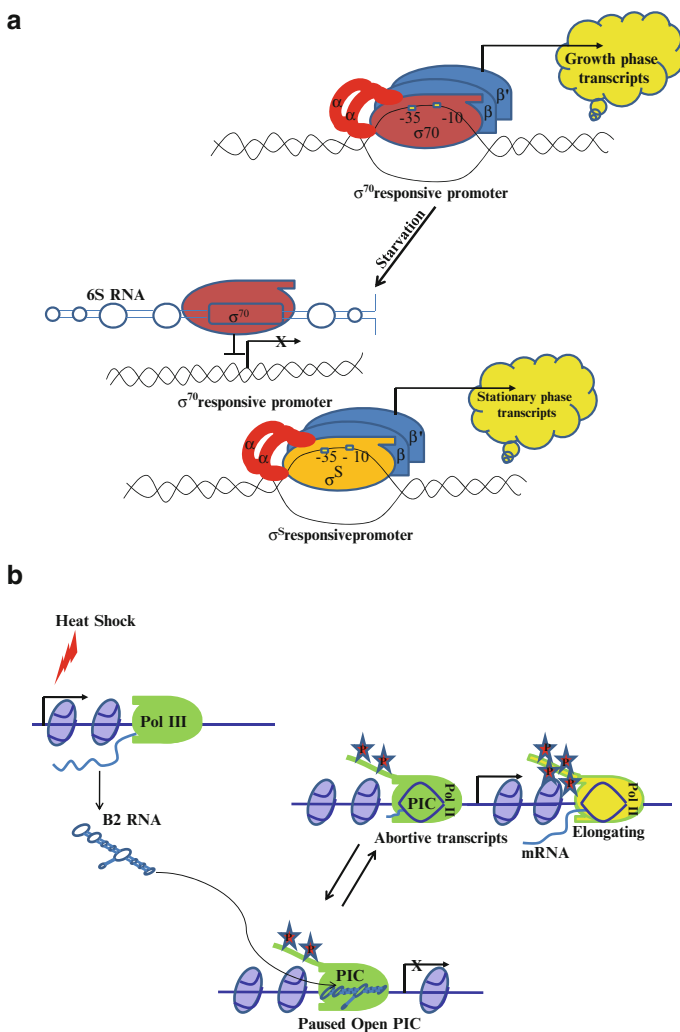
ncRNAs modulating transcription are abundant and were first to be discovered. Noncoding RNAs as transcriptional regulators target different components of transcription. Mostly such RNAs act in cis or trans and target general transcription factors, RNA polymerase, transcriptional activators or repressors. Here we are providing a few examples of ncRNA which regulate different steps of transcriptional process:

Bacterial 6S RNA: The *E. Coli* 6S RNA is one among the first ncRNAs to be discovered. About four decades ago 6S RNA was sequenced. It is 184 nucleotide long RNA having a conserved secondary structure containing largely double stranded and a central single stranded bulge. 6S RNA forms a stable complex with active polymerase tangled with promoter specificity factor  $\sigma^{70}$ . *E. Coli* 6S RNA was shown to interact with RNAP- $\sigma^{70}$  complex but not with free  $\sigma^{70}$ , thereby suppressing transcription (Trotochaud and Wassarman 2005). Interestingly, this repression of transcription was true for only a subset of promoters, as 6S RNA can activate transcription at promoters requiring Enzyme-  $\sigma^S$  complex (E-  $\sigma^S$  is required for survival

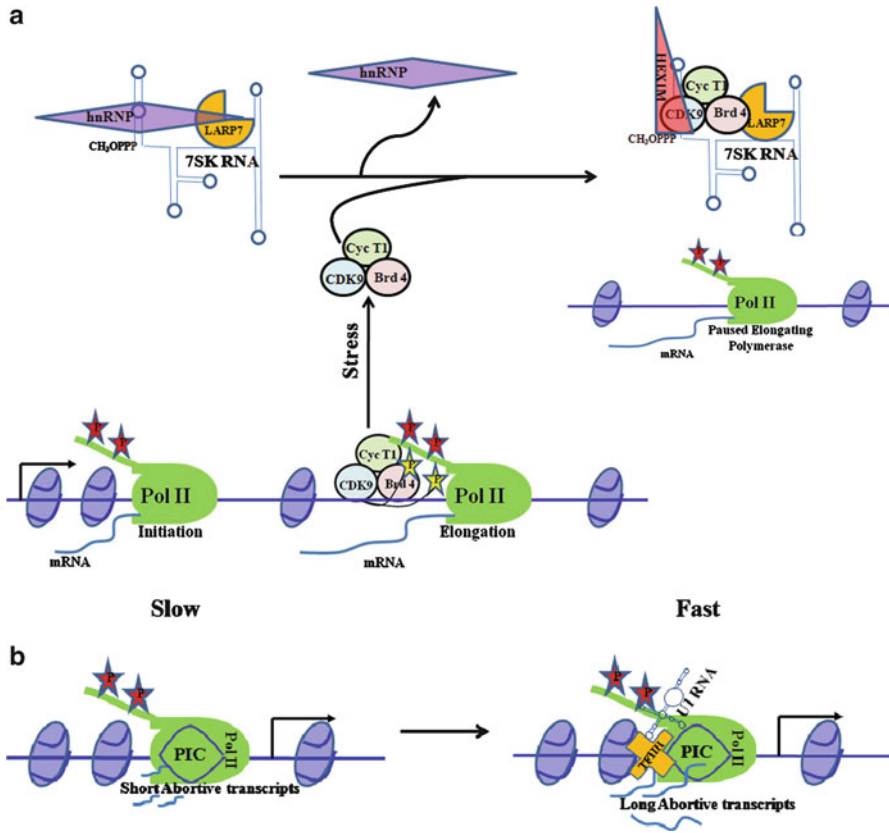
during stationary phase), indicating that 6S RNA regulates transcriptional process at multiple levels. Secondary structure of 6S RNA is essential for its activity and notably, single stranded bulge region was found to be critical for its RNAP binding and transcription modulation activity. Furthermore, 6S RNA structure mimic open promoter complex structure seen during transcriptional initiation (as shown in Fig. 15.1a) and thus proposed to inhibit transcription incorporating competition between promoter DNA and the E-  $\sigma^{70}$  (Barrick et al. 2005).

**Mouse B2 RNA:** B2 RNA is RNAP III encoded transcript, which is transcribed from short interspersed elements (SINE) of mouse genome and it represses RNAP II transcription in response to heat shock (Allen et al. 2004). B2 RNA is 178 nucleotide long and its expression increases many fold upon heat shock. B2 RNA interacts with a RNA docking site on RNAP II and assembles into the preinitiation complex at the promoter disrupting critical contacts between RNAP II and the promoter DNA, thereby inhibiting initiation of transcription (Espinoza et al. 2004). B2 RNA mediated RNAP II transcription repression shows promoter specificity. Recent investigations have explored the mechanisms underlying the B2 RNA mediated repression of RNAPII dependent transcription and found that B2 RNA targets early steps of transcription initiation like the Ser 5 phosphorylation by TFIIF (Espinoza et al. 2007). B2 RNA blocks CTD phosphorylation by TFIIF, only when RNAP II is in a transcriptionally repressed complex over the promoter DNA in an open state (Fig. 15.1b shown in green) prior to the formation of closed (Fig. 15.1b shown in yellow) complex (Yakovchuk et al. 2011).

**7SK RNA:** The human 7SK RNA is an abundant ( $2 \times 10^5$  copies/cell) evolutionarily conserved nuclear RNA of 331 nucleotides and is transcribed by RNAP III (Murphy et al. 1987 and Zieve et al. 1977). 7SK RNA controls RNAP II elongation by modulating the activity of transcription elongation factor P-TEFb (Nguyen et al. 2001). P-TEFb activates transcriptional elongation by phosphorylating C-Terminal Domain (CTD) of RNAPII. P-TEFb is a heterodimer comprising CDK9 and cyclin T1. In addition to general elongation factor, P-TEFb also functions as an HIV-1 Tat-specific transcription factor. P-TEFb interacts with Tat and the transactivating responsive (TAR) RNA structure located at the 5' end of the nascent viral transcript thus stimulating HIV-1 transcription. 7SK RNA binds to P-TEFb and represses transcription by abrogating its kinase activity. Association of P-TEFb and 7SK RNA is found to be reversible as ultraviolet irradiation and actinomycin D treatment disrupted P-TEFb/7SK RNA complex which can restore transcription (Yang et al. 2001). Further studies showed that inactivation of P-TEFb by 7SK RNA requires their association with other proteins namely MAQ1/HEXIM1 (hexamethylene bisacetamide-induced protein 1) which form the essential components of 7SK RNP. HEXIM1 was shown to inhibit P-TEFb in a 7SK-dependent manner while 7SK serves as a scaffold to mediate the HEXIM1:P-TEFb interaction (Fig. 15.2b) (Yik et al. 2003; Michels et al. 2003, 2004). A recent investigation has demonstrated that 7SK interacts with chromatin with high affinity (Mondal et al. 2010). The latter observation is consistent with the suggestion that 7SK by interacting with the chromatin serves as a scaffold for recruiting HEXIM1:P-TEFb proteins thereby inhibiting transcriptional elongation.



**Fig. 15.1** (a) 6S RNA mimics open promoter complex. 6S RNA targets the specificity factor  $\sigma^{70}$  in *E. coli* during stationary phase and sequesters from the active polymerase complex and but not free  $\sigma^{70}$  and hence, blocking transcription during stationary phase. On the other hand during stationary phase 6S RNA activates transcription at promoters requiring Enzyme-  $\sigma^S$  complex essential for the survival of bacteria. (b) B2RNA docks with RNAP II Preinitiation complex and blocks transcription initiation. In response to heat shock, B2 RNA is transcribed by RNAP III which binds RNA docking site of RNAP II within the paused open preinitiation complex over the promoter prior to the formation of closed complexes. This event blocks the critical contacts between RNAP II and the promoter DNA, and also represses the CTD phosphorylation (depicted in red stars: 2 stars and 4 stars:) by TFIIF thereby inhibiting the initiation of transcription by RNAP II



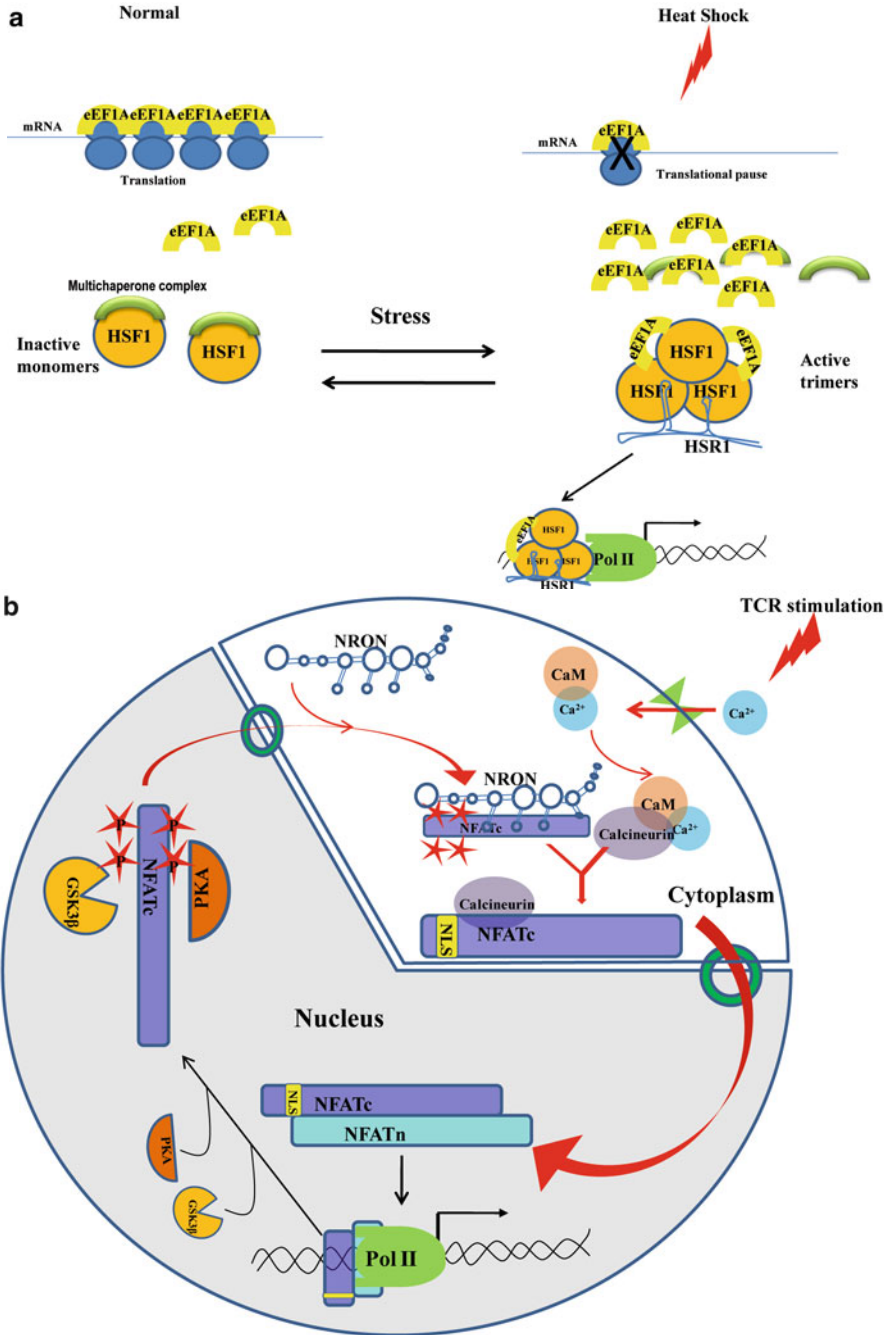
**Fig. 15.2** (a) 7SK RNA facilitates HEXIM1 mediated inhibition of P-TEFb. P-TEFb activates transcriptional elongation by phosphorylating (depicted in yellow stars) C-Terminal Domain (CTD) of RNAP II. P-TEFb consists of a kinase CDK9 and cyclin T1 heterodimer along with Brd4. Upon stress, the 7SK snRNP is released from hnRNP complex and binds to P-TEFb thereby abrogating its kinase activity and repression of transcription elongation. This inactivation of P-TEFb by 7SK RNA requires their association with other proteins namely HEXIM1 (hexamethylene bisacetamide-induced protein 1) and LARP7 (La ribonucleoprotein domain family, member 7) which form the essential components of 7SK snRNP upon stress. 7SK acts as a scaffold to mediate the HEXIM1:P-TEFb interaction that in turn blocks transcription elongation. (b) U1snRNA associates with TFIIF and enhances the transcription initiation rate. U1snRNA binds directly to the cyclin-H subunit of TFIIF and stimulates the kinase activity of TFIIF to phosphorylate C-terminal domain (CTD) of RNAP II, thereby stimulating the rate of initiation

**U1snRNA:** U1snRNA is approximately 160 nucleotide long ncRNA, transcribed by RNAP II. U1snRNA is one among the five small nuclear RNAs (snRNAs) U1-U6 that exist in snRNPs. These snRNPs facilitate splicing by forming the spliceosome together with many other proteins (Kramer 1996; Burge et al. 1999, for reviews). U1 snRNA has been shown to be associated with one of the general transcription factors TFIID, thereby influencing transcriptional initiation, a critical regulatory stage of gene expression. Specifically it binds directly to cyclin-H subunit of TFIID and stimulates kinase activity of TFIID that phosphorylates C-terminal domain (CTD) of RNAP II. Association of TFIID with U1snRNA stimulates the rate of initiation (rate of formation of first phosphodiester bond) by RNAP II (Fig. 15.2b). Addition of 5' splice site adjacent to promoter stimulates reinitiation of transcription in TFIID dependent manner indicating an important role for U1snRNA in transcriptional regulation by RNAP II apart from its well established role in RNA processing (Kwek et al. 2002).

**SRA RNA:** The steroid receptor RNA activator (SRA) is approximately 700 nucleotide long natural ncRNA. It exists in ribonucleoprotein complexes and functions as transcriptional coactivators of several steroid-hormone receptors (Lanz et al. 1999). Characterization of distinct RNA substructures within the SRA molecule reveals six RNA motifs critical for coactivation (Lanz et al. 2002). It is not clear whether RNA motifs execute transactivation at the RNA level or in cooperation with RNA binding proteins.

**HSR1:** Heat-shock RNA-1 (HSR1) is a ncRNA which modulates the activity of heat-shock transcription factor 1 (HSF1) upon heat shock response. In response to heat-shock, HSF1 induces the expression of heat shock proteins. In unstressed conditions, HSF1 exist in an inactive monomeric form and upon activation they acquire trimer formation ability and DNA binding properties. HSR1 and translation elongation factor eEF1A (present as ribonucleoprotein complex) are required for HSF1 activation (Shamovsky et al. 2006). eEF1A when free, is available for interaction with HSR1 and HSF1 which as a complex can initiate the heat-shock response. HSR1-eEF1A complexes when formed would capture HSF1 released from the HSP90 complex and assist its assembly into trimers and/or increase the stability of HSF1 trimers which is considered as the active form, which triggers the transcription of heat shock responsive genes (Fig. 15.3a).

**NRON RNA:** An RNAi based strategy employed to fish out ncRNAs modulating the activity of nuclear factor of activated T cells (NFAT) led to the identification of NRON RNA (Willingham et al. 2005). The nuclear factor of activated T cells (NFAT) refers to a family of transcription factors important in immune responses. These factors are sensitive to calcium signalling and upon activation calcineurin dephosphorylates NFAT resulting in its nuclear import essential for activating transcription. NRON size ranges from 0.8 to 4 kb based on alternative splicing. NRON represses NFAT activity by regulating its nuclear trafficking probably with aid of various transport factors (Fig. 15.3b). Thus, NRON ncRNA provides example of transcriptional regulation not via RNA-protein interactions or activity modulation of activator but through altering subcellular localisation of the latter.





**pncRNA:** Cyclin D1 (CCND1) promoter is associated with lncRNAs (range in size between 200 and 400nt) which are induced in response to genotoxic factors like ionizing radiation (Wang et al. 2008). The CCND1 pncRNA interacts with an RNA binding protein TLS (Translocated in Liposarcoma) and allosterically modify its activity such that this RNA-Protein interactions exert transcriptional repression by blocking the histone acetyl transferase (HAT) activity of CPB/p300 at the repressed CCND1 promoter.

**NRSE dsRNA:** Neuron-restrictive silencer element double-stranded RNA (NRSE dsRNA) shares sequence complementarity to promoter element that is bound by NRSF/REST (neuron-restrictive silencing factor/RE-1-silencing transcription factor). NRSF/REST is a repressor protein known to silence neuronal genes in non neuronal cells and restricts neuronal gene expression to neurons. NRSE is a small 20 bp double stranded RNA found to activate neural gene expression thus directing neuronal lineage in stem cells (Kuwabara et al. 2004). Interestingly, activation function of NRSE dsRNA is not via base pairing to promoter element with which it shares sequence homology. Rather, it interacts with NRSF/REST and converts this repressor into transcriptional activator. It is proposed that this RNA:protein interaction might prevent association of NRSF/REST with other corepressor proteins thereby switching neuronal gene expression from repressed state in stem cells to activated state in differentiating cells.

**piRNAs:** Piwi-interacting RNAs (piRNAs) (24–30 nt) are yet another class of small regulatory RNAs whose functions are not fully understood. Piwi family proteins are a subtype of Argonaute proteins and forms RNA protein complex with piRNA. piRNA are found in both vertebrate and invertebrate class of animal kingdom. The best studied function of the piRNA pathway is shown in germline cells where it is involved in transcriptional silencing of retrotransposons (Aravin et al. 2007). Unlike miRNAs and siRNAs, piRNA biogenesis does not involve Dicer or RISC. Not much is known about piRNA biogenesis however, recently it has been shown that a conserved primary piRNA biogenesis pathway that acts selectively on the 3' UTRs of messenger RNAs having a functional role in gonadal and germline development (Robine et al. 2009).



**Fig. 15.3** (a) HSR/HSF1/eEF1A trio complex induces transcription of heat shock responsive genes. In normal unstressed condition, HSF-1 exist in an inactive monomeric form along with the multi-chaperone complex, while the translation elongation factor eEF1A (present in ribonucleoprotein complex) aid in translation process. Upon heat shock, the eEF1A is no more engaged in translation and so they are free to interact with the HSF1 pool and the HSR1-eEF1A complex could assist its assembly into trimers. The ncRNA HSR1 interact with eEF1A-HSF1 trimers to increase their stability and induce the expression of the downstream heat shock responsive genes. (b) NRON blocks NFAT shuttling and inhibits NFAT mediated transcription. In normal resting condition NFAT (nuclear factor of activated T cells) remains phosphorylated and associated with the ncRNA NRON as a complex. In response to TCR stimulation the calcium ion entry activates the phosphatase calcineurin. Calcineurin further dephosphorylates NFAT and exposes the NLS, resulting in its nuclear import essential for activating transcription. Further the cytoplasmic pool is restored upon phosphorylation by kinases like GSK3 $\beta$  and PKA

raasiRNA : Repeat associated small interfering RNA (raasiRNA) is considered to be a subclass of piRNA and associate with both the Ago and Piwi Argonaute protein subfamily unlike piRNA which associates only with the Piwi Argonaute subfamily (Girard et al. 2006; Faehnle and Joshua-Tor 2007). Like piRNAs, rasiRNAs are abundant in germline cells and function in silencing transposons and retrotransposons as well as maintaining heterochromatin structure by controlling repeat sequences transcription (Matzke et al. 2004; Lippman and Martienssen 2004; Aravin and Tuschl 2005).

NanoRNA: NanoRNAs are one among the most recently discovered class of functional small RNAs that are believed to affect gene expression through direct incorporation into a target RNA transcript rather than through a traditional antisense-based mechanism. These nanoRNAs were discovered in *Pseudomonas aeruginosa* as 2–4 nt long oligonucleotides that function as primers for initiating transcription from a set of promoters (Goldman et al. 2011). Still the exact molecular events of gene expression, regulatory role remains open for investigation.

#### 15.4.4 *lncRNAs in Genomic Imprinting*

Genomic imprinting is an epigenetic phenomenon which restricts expression of some genes to one of the two parental chromosomes. So far more than 100 imprinted genes have been identified and most of them are clustered in large chromosomal domains. The allelic expression of imprinted genes is controlled by imprint control element (ICE). ICE is epigenetically modified by DNA methylation and histone modification to regulate the expression of imprinted genes. Only unmethylated ICE is active in inducing repression of flanking genes. ICE attains methylation during gametogenesis and this germline DNA methylation is established by de novo DNA methyltransferases DNMT3A/DNMT3L (Bourc'his and Proudhon 2008). Subsequent maintenance of methylation at ICE requires maintenance DNA methyltransferase DNMT1 (Hirasawa et al. 2008). In addition, other protein factors (specific for each ICE) also contribute to the establishment and maintenance of ICE methylation (Li et al. 2008). Histone modifications for methylated and unmethylated ICEs are found different. In general, repressive marks like H3K9Me3, H4K20me3 are associated with DNA-methylated ICE and active marks like H3K4me and H3/H4 acetylation with those of unmethylated ICE.

The mechanism by which ICE is proposed to function is either by constituting an insulator region that prevents promoter enhancer interaction or by activating ncRNA transcription. As seen in the *Igf2* imprinted cluster, a methylation sensitive insulator in the ICE regulates its expression. The chromatin insulator protein CTCF (11-zinc finger protein or CCCTC-binding factor) binds to unmethylated the ICE and prevents the communication between the enhancers downstream of the H19 gene and *Igf2* promoters (Kanduri et al. 2000a, b; Bell and Felsenfeld 2000; Hark et al. 2000). DNA methylation of the ICE prevents CTCF binding and allows the enhancer-*Igf2* promoter communication to facilitate its transcription (Kanduri et al. 2001).

In most of the imprinted gene clusters there is at least one macro ncRNA gene. Some of the tested imprinted macro ncRNA have been shown to be indispensable for the imprinted expression of whole cluster (Pauler et al. 2007; Braidotti et al. 2004). Macro ncRNAs are transcribed from unmethylated ICE. These RNAs possess some unusual features such as low intron/exon ratio i.e. reduced splicing potential, nuclear retention and accumulation at the site of transcription (Pandey et al. 2008; Braidotti et al. 2004; Terranova et al. 2008; Nagano et al. 2008). ncRNA mode of regulation is seen more common in imprinted gene cluster expression in contrast to CTCF dependent chromatin insulation mechanism. *Igf2r* and *Kcnq1* imprinted clusters have been used extensively to investigate the role of macro ncRNAs in genomic imprinting. *Igf2r* cluster harbours four imprinted genes in about 500 kb region on chromosome 17: one macro ncRNA *Airn* is exclusively expressed on the paternal chromosome and three neighboring protein coding imprinted genes, *Igf2r*, *Slc22a2* and *Slc22a3* expressed only from the maternal chromosome (Brandeis et al. 1993; Stoger et al. 1993; Lucifero et al. 2002). The unmethylated ICE on the paternal chromosome serves as promoter for paternally expressed ncRNA, *Airn* (Antisense *Igf2r* RNA) that overlaps *Igf2r* in antisense orientation. *Airn* ncRNA is about 108 kb long, unspliced and polyadenylated transcript. Targeted deletion of ICE, comprising *Airn* promoter, resulted in loss of silencing of all three neighboring genes on the paternal chromosome, indicating that *Airn* ncRNA plays important role in gene silencing (Wutz et al. 1997).

*Kcnq1* domain is a one mega-base imprinted domain containing 8–10 imprinted protein coding genes, which are exclusively expressed from the maternal chromosome, and one lncRNA *Kcnq1ot1* expressed from the paternal chromosome. Expression of *Kcnq1ot1* on the paternal chromosome is linked to silencing of the imprinted protein coding genes (Fitzpatrick et al. 2002; Kanduri et al. 2006; DiNardo et al. 2006). However, on the maternal chromosome the imprinted protein coding genes are expressed due to silencing of *Kcnq1ot1* ncRNA promoter by CpG methylation. It has been shown that *Kcnq1ot1* itself mediates transcriptional gene silencing through interacting with chromatin remodeling machinery such as PRC2 complex members and G9a. Furthermore they are targeted specifically to imprinted gene promoters in a tissue-specific fashion thereby organizing higher order chromatin structure devoid of RNAP II (Pandey et al. 2008; Terranova et al. 2008).

Several recent studies have linked differential ncRNA expression to developmental and tissue specific expression of imprinted genes. One such study reveals that neurons do not show imprinted *Igf2r* expression due to lack of *Airn* ncRNA whereas, glial cells which express *Airn* ncRNA shows imprinting of *Igf2r* expression (Yamasaki et al. 2005). Placenta is another example of tissue specific imprinted expression. Several studies indicate the direct involvement of *Airn* and *Kcnq1ot1* macro ncRNAs in placental genes silencing. *Kcnq1ot1* physically localise to several silent genes lying away from promoter (Pandey et al. 2008). It also interacts with polycomb group proteins and establishes repressive marks. Similarly *Airn* ncRNA bind to H3K9 methyltransferase and lies in close proximity to silent *Slc22a3* promoter of *Igf2r* cluster. Deletion experiments involving *G9a* and polycomb group proteins *EZH2* and *RNF2* shows loss of placental tissue specific imprinted expression in these clusters (Nagano et al. 2008; Terranova et al. 2008; Wagschal et al. 2008).

### 15.4.5 *lncRNAs and X-chromosome Inactivation*

A best known phenomenon involving the lncRNA is X-chromosome inactivation (XCI). XCI occurs in mammalian females to ensure equal X-linked gene products between two sexes. Inactivated X chromosome expresses a ncRNA called the inactive X-specific transcript (*Xist*) that localizes and coats one of the X chromosome in cis and bring about gene silencing by establishing a higher order heterochromatic compartment. Recent studies have shown that *Xist* interacts with polycomb group proteins like *EZH2* which induces repressive marks like H3K27me and aid in gene silencing (Silva et al. 2003; Plath et al. 2003). The mechanism by which these repressive chromatin modifiers are recruited to inactive X-chromosome is unknown. On the active X-chromosome, *Xist* is repressed and its repression is carried out by a long ncRNA, *Tsix* which overlaps *Xist* in antisense orientation (Wutz and Gribnau 2007). *Tsix*, unlike *Xist*, silences only the *Xist* promoter on the active X chromosome. However, the mechanisms by which *Tsix* specifically regulates *Xist* repression is currently not clear. *Tsix* has also been shown to interact with epigenetic regulators such as polycomb proteins (Zhao et al. 2008) and DNA methyltransferases (Sun et al. 2006) and this interaction has been suggested to be crucial for the *Xist* repression on the active X chromosome.

## 15.5 ncRNA in Disease

### 15.5.1 *An Overview*

A wide variety of diseases have been discovered with altered expression or function of ncRNAs. Dyskeratosis congenita, Spinal muscular dystrophy, Autism, Alzheimer's, miR96 associated Hearing loss and Prader-Willi syndrome are some of the diseases where the small RNPs like snRNAs, miRNAs and snoRNAs are altered. The Sm-class snRNPs are not properly assembled in spinal muscular dystrophy (Selenko et al. 2001), and in dyskeratosis congenita mutations occur in telomerase RNA (Vulliamy et al. 2001). Duplication of snRNA SNORD115 is associated with Autism. In Alzheimer's disease, an antisense lncRNA (BACE1-AS) is implicated in increasing the steady state levels of its sense counterpart beta-secretase (BACE1) gene by enhancing its stability via masking certain crucial regulatory elements through sense and antisense interactions (Faghihi et al. 2008). This results in increased cleavage of amyloid precursor protein into amyloid beta1-42 which is a critical component in Alzheimer's disease. In case of Prader-Willi syndrome the paternal copies of the imprinted *SNRPN* and *Necdin* genes along with a cluster of 48 *SNORD116* coding region are deleted (Cavaillé et al. 2000; Skryabin et al. 2007; Ding et al. 2008). One other disorder where ncRNAs are implicated in the disease etiology is a rare forms of hearing disorder where the miRNA, miR-96 is aberrantly expressed (Lewis et al. 2009).

ncRNAs also mediate changes at an epigenetic level that ultimately contribute to certain disease etiology. In a rare form of  $\beta$ -thalassemia, a translocation juxtaposes distantly located *LUC57L* in close proximity to the  $\alpha$ -globin gene *HBA2*. This results in transcriptional read through from the truncated *LUC57L* transcription unit and specific methylation of *HBA2* gene thus causing transcriptional silencing of *HBA2* gene (Tufarelli et al. 2003). BC1/BC200 an mRNA like ncRNA is known to be altered in the fragile X syndrome, where the loss of function of FMRP (fragile X mental retardation protein) occurs due to the absence of BC200 binding where the subsequent loss of translational repression of mRNAs in the post synaptic area of such patients (Zalfa et al. 2005). Another related ncRNA which has ancestral similarity towards BC200 called as Psoriasis-related ncRNA (PRINS) (Sonkoly et al. 2005) that like BC200 possess two Alu repetitive sequences and was implicated in Psoriasis via the down-regulation of G1P3 (Szegedi et al. 2010) but the exact mechanism is still unknown. Recent reports have shown some SNPs within the non-coding regions associated with certain disease conditions but the complex patterns of ncRNA expression makes it particularly difficult to screen such SNPs (Mattick 2009a, b).

### 15.5.2 ncRNA and Cancer

In the recent past there is an increasing appreciation in exploring the functional link between ncRNA expression profiles and cancer. Genome wide association studies (GWAS) have now shifted their focus towards miRNAs and lncRNAs' expression patterns in various cancers. The evidences of altered ncRNAs are often correlated well with cancers to a great extent due to the statistically valid observations made from different geographic locations and gene pools. For some of the cancers, these ncRNAs presently, serve as markers for the diagnosis and scoring the treatment regime. snoRNAs, UCRs and miRNAs are some of the commonly reported class of ncRNAs used for such purposes in cancers (Galasso et al. 2010a, b, c). Numerous lncRNAs have been shown to be altered in multiple cancers. For further reference on the list of lncRNAs and the associated cancer type refer to table 3 in ref (recently reviewed by Gibb et al. 2011) and table 1 in a review by Mattick (2009a, b). T-UCRs (Transcribe ultraconserved Regions) are a class of ncRNA that have been reported to be altered in cancers like adult CLL, colorectal carcinoma, hepatocellular carcinoma and few neuroblastomas where these RNAs are currently being used to predict the patient prognosis with greater confidence (Braconi et al. 2011).

### 15.5.3 ncRNA and Therapeutics

As mentioned above, numerous ncRNAs have been implicated in the molecular pathogenesis of various human diseases, especially in cancer a special set of miRNAs

possess oncogenic properties which are named as “OncomiRs”. Targeting these ncRNAs has always been a valid approach to contain such disorders. Unfortunately the existing information on the functional mechanism involving these ncRNAs is incomplete. The major obstacle for this lack of information is the technical difficulties faced by researchers while performing knock-down of the very few ncRNAs that have been distinctly correlated to a disease state using rigorous screening procedures. siRNA based knock-down does not hold good for ncRNAs but the LNA and PNA based AntagomiRs, and the recently developed synthetic ribozyme based enzymes that cleave specific ncRNA population are showing encouraging results. Unfortunately the efficiency of such molecules is poor. Also the delivery of these antagomiRs poses another level of complexity. Currently people are trying to solve the delivery issues using various vehicles like liposome conjugation, cholesterol conjugation, viral vector based infection and other transgenic and nanomaterial approaches (Galasso et al. 2010a, b, c).

## 15.6 Outlook

The last decade has been a fruitful year for the investigations on noncoding portion of genome, which previously thought to represents a junk portion of the genome. With the development of several high throughput applications such as microarrays and massive parallel sequencing, it is realized that the majority of the noncoding portion of the genome is pervasively transcribed to encode several thousands of small and long transcripts. Though there is a discrepancy as to the extent of transcription across noncoding portion of the genome, the evidence from several independent investigations provides support to the fact that noncoding transcripts are present in several thousands. Early estimates suggest that existence of about 28,000 lncRNAs and their number could grow well beyond the suggested number. Especially when we consider intronic, antisense and promoter associated transcripts. One of the major challenges associated with this huge number is that detailed physical, structural and functional characterization of each transcript. This will enable us to know the extent of transcriptional noise versus functional noncoding transcripts. Unlike protein coding RNA, lncRNA are expressed at very low level, thus posing a problem in functional annotation of lncRNAs. Hence there is a need for technologies to annotate lncRNAs expressed at low levels. Unlike small RNA mediated silencing pathways, lncRNA mediated silencing and activation pathways are ill defined. Base pair interactions primarily define the specificity of small ncRNAs. Given the absence of sequence similarity between lncRNAs and their targets, it is not clear how lncRNAs specifically activate or silence target genes. This is one of the outstanding questions that remain to be investigated. In the recent past, expression profiles of lncRNA in various cancers have been explored to identify potential prognostic and/or diagnostic markers. Like, small RNAs, lncRNAs show distinct expression profiles in various cancers. However, there is not much progress in the treatment of

cancers using ncRNAs as targets. Moreover, the molecular pathways by which lncRNAs induce pathogenesis are not well investigated. Hence the molecular pathways that are affected in response to aberrant expression of lncRNAs need to be well investigated in order to devise better intervention strategies using ncRNAs as targets. Detailed functional annotation of ncRNA transcription across the genome is required in order to realize the potential of ncRNAs in mammalian development and disease.

**Acknowledgments** The authors would like to thank Prof. Tapas K Kundu, for his scientific and technical inputs during the preparation of this chapter. We thank Department of Science and Technology, Govt. of India, and Jawaharlal Nehru Centre for Advanced Scientific Research and Programme support, DBT, Govt. of India, for financial support. DK and SK are Senior research fellows of Council of Scientific and Industrial Research, India.

## References

- Abe T, Ikemura T, Sugahara J, Kanai A, Ohara Y, Uehara H, Kinouchi M, Kanaya S, Yamada Y, Muto A, Inokuchi H (2011) tRNADB-CE 2011: tRNA gene database curated manually by experts. *Nucleic Acids Res* 39:D210–D213
- Allen TA, Von Kaenel S, Goodrich JA, Kugel JF (2004) The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. *Nat Struct Mol Biol* 11:816–821
- Amaral PP, Mattick JS (2008) Noncoding RNA in development. *Mamm Genome* 19:454–492
- Aravin A, Tuschl T (2005) Identification and characterization of small RNAs involved in RNA silencing. *FEBS Lett* 579:5830–5840
- Aravin AA, Hannon GJ, Brennecke J (2007) The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* 318:761–764
- Bachellerie JP, Cavaillé J, Hüttenhofer A (2002) The expanding snoRNA world. *Biochimie* 84:775–790
- Barrick JE, Sudarsan N, Weinberg Z, Ruzzo WL, Breaker RR (2005) 6S RNA is a widespread regulator of eubacterial RNA polymerase that resembles an open promoter. *RNA* 11:774–784
- Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, Haussler D (2004) Ultraconserved elements in the human genome. *Science* 304:1321–1325
- Bell AC, Felsenfeld G (2000) Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature* 405:482–485
- Boria I, Gruber AR, Tanzer A, Bernhart SH, Lorenz R, Mueller MM, Hofacker IL, Stadler PF (2010) Nematode sbRNAs: homologs of vertebrate Y RNAs. *J Mol Evol* 70:346–358
- Bourc'his D, Proudhon C (2008) Sexual dimorphism in parental imprint ontogeny and contribution to embryonic development. *Mol Cell Endocrinol* 282:87–94
- Braconi C, Kogure T, Valeri N, Huang N, Nuovo G, Costinean S, Negrini M, Miotto E, Croce CM, Patel T (2011) microRNA-29 can regulate expression of the long non-coding RNA gene MEG3 in hepatocellular cancer. *Oncogene*. doi:10.1038/onc.2011.193
- Bradley RK, Uzilov AV, Skinner ME, Bendaña YR, Barquist L, Holmes I (2009) Evolutionary modeling and prediction of non-coding RNAs in *Drosophila*. *PLoS One* 4:e6478
- Braidotti G, Baubec T, Pauler F, Seidl C, Smrzka O, Stricker S, Yotova I, Barlow DP (2004) The air noncoding RNA: an imprinted cis-silencing transcript. *Cold Spring Harb Symp Quant Biol* 69:55–66
- Brandeis M, Kafri T, Ariel M, Chaillet JR, McCarrey J, Razin A, Cedar H (1993) The ontogeny of allele-specific methylation associated with imprinted genes in the mouse. *EMBO J* 12:3669–3677

- Burge CB, Tuschl T, Sharp PA (1999) Splicing of precursors to mRNAs by the spliceosome. In: Gesteland RF, Cech TR, Atkins JF (eds) *The RNA world*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp 525–560
- Bushati N, Cohen SM (2007) MicroRNA functions. *Annu Rev Cell Dev Biol* 23:175–205
- Calin GA, Liu CG, Ferracin M, Hyslop T, Spizzo R, Sevignani C, Fabbri M, Cimmino A, Lee EJ, Wojcik SE, Shimizu M, Tili E, Rossi S, Taccioli C, Pichiorri F, Liu X, Zupo S, Herlea V, Gramantieri L, Lanza G, Alder H, Rassenti L, Volinia S, Schmittgen TD, Kipps TJ, Negrini M, Croce CM (2007) Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell* 12:215–229
- Cao F, Li X, Hiew S, Brady H, Liu Y, Dou Y (2009) Dicer independent small RNAs associate with telomeric heterochromatin. *RNA* 15:1274–1281
- Cavaillé J, Buiting K, Kieffmann M, Lalande M, Brannan CI, Horsthemke B, Bachelier JP, Brosius J, Hüttenhofer A (2000) Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc Natl Acad Sci U S A* 97:14311–14316
- Cavaillé J, Seitz H, Paulsen M, Ferguson-Smith AC, Bachelier JP (2002) Identification of tandemly-repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader-Willi/Angelman syndrome region. *Hum Mol Genet* 11:1527–1538
- Chen X, Quinn AM, Wolin SL (2000) Ro ribonucleoproteins contribute to the resistance of *Deinococcus radiodurans* to ultraviolet irradiation. *Genes Dev* 14:777–782
- Chen X, Wurtmann EJ, Van Batavia J, Zybailov B, Washburn MP, Wolin SL (2007) An ortholog of the Ro autoantigen functions in 23S rRNA maturation in *D. radiodurans*. *Genes Dev* 21:1328–1339
- Christov CP, Gardiner TJ, Szüts D, Krude T (2006) Functional requirement of noncoding Y RNAs for human chromosomal DNA replication. *Mol Cell Biol* 26:6993–7004
- Christov CP, Trivier E, Krude T (2008) Noncoding human Y RNAs are overexpressed in tumours and required for cell proliferation. *Br J Cancer* 98:981–988
- Crick FH (1968) The origin of the genetic code. *J Mol Biol* 38:367–379
- Dieci G, Fiorino G, Castelnuovo M, Teichmann M, Pagano A (2007) The expanding RNA polymerase III transcriptome. *Trends Genet* 23:614–622
- Ding F, Li HH, Zhang S, Solomon NM, Camper SA, Cohen P, Francke U (2008) SnoRNA Snord116 (Pwcr1/MBII-85) deletion causes growth deficiency and hyperphagia in mice. *PLoS One* 3:e1709
- Dreher TW (2009) Role of tRNA-like structures in controlling plant virus replication. *Virus Res* 139:217–229
- Eddy SR (2001) Non-coding RNA genes and the modern RNA world. *Nat Rev Genet* 2:919–929
- ENCODE Project Consortium, Birney E, Stamatoyannopoulos JA, Dutta A et al (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447:799–816
- Espinoza CA, Allen TA, Hieb AR, Kugel JF, Goodrich JA (2004) B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. *Nat Struct Mol Biol* 11:822–829
- 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
- Faehle CR, Joshua-Tor L (2007) Argonautes confront new small RNAs. *Curr Opin Chem Biol* 11:569–577
- Fagegaltier D, Bougé AL, Berry B, Poisot E, Sismeiro O, Coppée JY, Théodore L, Voinnet O, Antoniewski C (2009) The endogenous siRNA pathway is involved in heterochromatin formation in *Drosophila*. *Proc Natl Acad Sci U S A* 106:21258–21263
- Faghihi MA, Modarresi F, Khalil AM, Wood DE, Sahagan BG, Morgan TE, Finch CE, St Laurent G 3rd, Kenny PJ, Wahlestedt C (2008) Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. *Nat Med* 14:723–730
- Farh KK, Grimson A, Jan C, Lewis BP, Johnston WK, Lim LP, Burge CB, Bartel DP (2005) The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. *Science* 310:1817–1821



- Fitzpatrick GV, Soloway PD, Higgins MJ (2002) Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. *Nat Genet* 32:426–431
- Flieck P, Aken BL, Beal K et al (2008) Ensembl 2008. *Nucleic Acids Res* 36:D707–D714
- Galasso D, Carnuccio A, Larghi A (2010a) Pancreatic cancer: diagnosis and endoscopic staging. *Eur Rev Med Pharmacol Sci* 14:375–385
- Galasso F, Giannella R, Bruni P, Giulivo R, Barbini VR, Disanto V, Leonardi R, Pansadoro V, Sepe G (2010b) PCA3: a new tool to diagnose prostate cancer (PCa) and a guidance in biopsy decisions. Preliminary report of the UrOP study. *Arch Ital Urol Androl* 82:5–9
- Galasso M, Elena Sana M, Volinia S (2010c) Non-coding RNAs: a key to future personalized molecular therapy? *Genome Med* 2:12–22
- Gebauer F, Hentze MW (2004) Molecular mechanisms of translational control. *Nat Rev Mol Cell Biol* 5:827–835
- Gibb EA, Brown CJ, Lam WL (2011) The functional role of long non-coding RNA in human carcinomas. *Mol Cancer* 10:38
- Gillet R, Felden B (2001) Emerging views on tmRNA-mediated protein tagging and ribosome rescue. *Mol Microbiol* 42:879–885
- Girard A, Sachidanandam R, Hannon GJ, Carmell MA (2006) A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* 442:199–202
- Goldman SR, Sharp JS, Vvedenskaya IO, Livny J, Dove SL, Nickels BE (2011) NanoRNAs prime transcription initiation in vivo. *Mol Cell* 42:817–825
- Gottwein E, Mukherjee N, Sachse C, Frenzel C, Majoros WH, Chi JT, Braich R, Manoharan M, Soutschek J, Ohler U, Cullen BR (2007) A viral microRNA functions as an orthologue of cellular miR-155. *Nature* 450:1096–1099
- Grey F, Tirabassi R, Meyers H, Wu G, McWeeney S, Hook L, Nelson JA (2010) A viral microRNA down-regulates multiple cell cycle genes through mRNA 5' UTRs. *PLoS Pathog* 6:e1000967
- Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458:223–227
- Han J, Kim D, Morris KV (2007) Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cells. *Proc Natl Acad Sci U S A* 104:12422–12427
- Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilghman SM (2000) CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* 405:486–489
- Hendrick JP, Wolin SL, Rinke J, Lerner MR, Steitz JA (1981) Ro small cytoplasmic ribonucleoproteins are a subclass of La ribonucleoproteins: further characterization of the Ro and La small ribonucleoproteins from uninfected mammalian cells. *Mol Cell Biol* 1:1138–1149
- Hirasawa R, Chiba H, Kaneda M, Tajima S, Li E, Jaenisch R, Sasaki H (2008) Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. *Genes Dev* 22:1607–1616
- Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, Khalil AM, Zuk O, Amit I, Rabani M, Attardi LD, Regev A, Lander ES, Jacks T, Rinn JL (2010) A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* 142:409–419
- Hüttenhofer A, Schattner P, Polacek N (2005) Non-coding RNAs: hope or hype? *Trends Genet* 21:289–297
- Inagaki S, Numata K, Kondo T, Tomita M, Yasuda K, Kanai A, Kageyama Y (2005) Identification and expression analysis of putative mRNA-like non-coding RNA in *Drosophila*. *Genes Cells* 10:1163–1173
- Jády BE, Kiss T (2001) A small nucleolar guide RNA functions both in 2'-O-ribose methylation and pseudouridylation of the U5 spliceosomal RNA. *EMBO J* 20:541–551
- Jiang ZF, Croshaw DA, Wang Y, Hey J, Machado CA (2011) Enrichment of mRNA-like noncoding RNAs in the divergence of *Drosophila* males. *Mol Biol Evol* 28:1339–1348
- Kanduri C (2001) Restriction enzyme BstZ17I is sensitive to cytosine methylation. *FEMS Microbiol Lett* 200:191–193

- Kanduri C, Holmgren C, Pilartz M, Franklin G, Kanduri M, Liu L, Ginjala V, Ullerås E, Mattsson R, Ohlsson R (2000a) The 5' flank of mouse H19 in an unusual chromatin conformation unidirectionally blocks enhancer-promoter communication. *Curr Biol* 10:449–457
- Kanduri C, Pant V, Loukinov D, Pugacheva E, Qi CF, Wolffe A, Ohlsson R, Lobanenkov VV (2000b) Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. *Curr Biol* 10:853–856
- Kanduri C, Thakur N, Pandey RR (2006) The length of the transcript encoded from the Kcnq1ot1 antisense promoter determines the degree of silencing. *EMBO J* 25:2096–2106
- Kapranov P, Willingham AT, Gingeras TR (2007) Genome-wide transcription and the implications for genomic organization. *Nat Rev Genet* 8:413–423
- Kavanaugh LA, Dietrich FS (2009) Non-coding RNA prediction and verification in *Saccharomyces cerevisiae*. *PLoS Genet* 5:e1000321
- Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, Markenscoff-Papadimitriou E, Kuhl D, Bito H, Worley PF, Kreiman G, Greenberg ME (2010) Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465:182–187
- Kishore S, Stamm S (2006) The snoRNA HBI1-52 regulates alternative splicing of the serotonin receptor 2C. *Science* 311:230–232
- Krämer A (1996) The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu Rev Biochem* 65:367–409
- Kuwabara T, Hsieh J, Nakashima K, Taira K, Gage FH (2004) A small modulatory dsRNA specifies the fate of adult neural stem cells. *Cell* 116:779–793
- Kwek KY, Murphy S, Furger A, Thomas B, O’Gorman W, Kimura H, Proudfoot NJ, Akoulitchev A (2002) U1 snRNA associates with TFIIF and regulates transcriptional initiation. *Nat Struct Biol* 9:800–805
- Lander ES, Linton LM, Birren B et al (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860–921
- Lanz RB, McKenna NJ, Onate SA, Albrecht U, Wong J, Tsai SY, Tsai MJ, O’Malley BW (1999) A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. *Cell* 97:17–27
- Lanz RB, Razani B, Goldberg AD, O’Malley BW (2002) Distinct RNA motifs are important for coactivation of steroid hormone receptors by steroid receptor RNA activator (SRA). *Proc Natl Acad Sci U S A* 99:16081–16086
- Lewis MA, Quint E, Glazier AM, Fuchs H, De Angelis MH, Langford C, van Dongen S, Abreu-Goodger C, Piipari M, Redshaw N, Dalmay T, Moreno-Pelayo MA, Enright AJ, Steel KP (2009) An ENU-induced mutation of miR-96 associated with progressive hearing loss in mice. *Nat Genet* 41:614–618
- Li X, Ito M, Zhou F, Youngson N, Zuo X, Leder P, Ferguson-Smith AC (2008) A maternal-zygotic effect gene, *Zfp57*, maintains both maternal and paternal imprints. *Dev Cell* 15:547–557
- Lippman Z, Martienssen R (2004) The role of RNA interference in heterochromatic silencing. *Nature* 431:364–370
- Loewer S, Cabili MN, Guttman M, Loh YH, Thomas K, Park IH, Garber M, Curran M, Onder T, Agarwal S, Manos PD, Datta S, Lander ES, Schlaeger TM, Daley GQ, Rinn JL (2010) Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat Genet* 42:1113–1117
- Louro R, Smirnova AS, Verjovski-Almeida S (2009) Long intronic noncoding RNA transcription: expression noise or expression choice? *Genomics* 93:291–298
- Lucifero D, Mertineit C, Clarke HJ, Bestor TH, Trasler JM (2002) Methylation dynamics of imprinted genes in mouse germ cells. *Genomics* 79:530–538
- MacMorris M, Kumar M, Lasda E, Larsen A, Kraemer B, Blumenthal T (2007) A novel family of *C. elegans* snRNPs contains proteins associated with trans-splicing. *RNA* 13:511–520
- Mancini-Dinardo D, Steele SJ, Levorse JM, Ingram RS, Tilghman SM (2006) Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. *Genes Dev* 20:1268–1282
- Mans RM, Pleij CW, Bosch L (1991) tRNA-like structures. Structure, function and evolutionary significance. *Eur J Biochem* 201:303–324

- Mansfield JH, Harfe BD, Nissen R, Obenauer J, Srineel J, Chaudhuri A, Farzan-Kashani R, Zuker M, Pasquinelli AE, Ruvkun G, Sharp PA, Tabin CJ, McManus MT (2004) MicroRNA-responsive 'sensor' transgenes uncover Hox-like and other developmentally regulated patterns of vertebrate microRNA expression. *Nat Genet* 36:1079–1083
- Maroney PA, Yu YT, Jankowska M, Nilsen TW (1996) Direct analysis of nematode cis- and trans-spliceosomes: a functional role for U5 snRNA in spliced leader addition trans-splicing and the identification of novel Sm snRNPs. *RNA* 2:735–745
- Matera AG, Terns RM, Terns MP (2007) Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat Rev Mol Cell Biol* 8:209–220
- Matsuda D, Dreher TW (2004) The tRNA-like structure of Turnip yellow mosaic virus RNA is a 3'-translational enhancer. *Virology* 321:36–46
- Mattick JS (2009a) Has evolution learnt how to learn? *EMBO Rep* 10:665
- Mattick JS (2009b) The genetic signatures of noncoding RNAs. *PLoS Genet* 5:e1000459
- Mattick JS, Makunin IV (2005) Small regulatory RNAs in mammals. *Hum Mol Genet* 14(Spec No 1):R121–32
- Matzke M, Aufsatz W, Kanno T, Daxinger L, Papp I, Mette MF, Matzke AJ (2004) Genetic analysis of RNA-mediated transcriptional gene silencing. *Biochim Biophys Acta* 1677:129–141
- Michels AA, Nguyen VT, Fraldi A, Labas V, Edwards M, Bonnet F, Lania L, Bensaude O (2003) MAQ1 and 7SK RNA interact with CDK9/cyclin T complexes in a transcription-dependent manner. *Mol Cell Biol* 23:4859–4869
- Michels AA, Fraldi A, Li Q, Adamson TE, Bonnet F, Nguyen VT, Sedore SC, Price JP, Price DH, Lania L, Bensaude O (2004) Binding of the 7SK snRNA turns the HEXIM1 protein into a P-TEFb (CDK9/cyclin T) inhibitor. *EMBO J* 23:2608–2619
- Millar AA, Waterhouse PM (2005a) Plant and animal microRNAs: similarities and differences. *Funct Integr Genomics* 5:129–135
- Millar AA, Waterhouse PM (2005b) Small RNAs: endo-siRNAs truly endogenous. *Nat Rev Mol Cell Biol* 9:426–427
- Mondal T, Rasmussen M, Pandey GK, Isaksson A, Kanduri C (2010) Characterization of the RNA content of chromatin. *Genome Res* 20:899–907
- Moore MJ (2005) From birth to death: the complex lives of eukaryotic mRNAs. *Science* 309:1514–1518
- Morey C, Avner P (2004) Employment opportunities for non-coding RNAs. *FEBS Lett* 567:27–34
- Murphy S, Di Liegro C, Melli M (1987) The in vitro transcription of the 7SK RNA gene by RNA polymerase III is dependent only on the presence of an upstream promoter. *Cell* 51:81–87
- Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, Feil R, Fraser P (2008) The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* 322:1717–1720
- Nakaya HI, Amaral PP, Louro R, Lopes A, Fachel AA, Moreira YB, El-Jundi TA, da Silva AM, Reis EM, Verjovski-Almeida S (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
- Neil H, Malabat C, d'Aubenton-Carafa Y, Xu Z, Steinmetz LM, Jacquier A (2009) Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* 457:1038–1042
- Nguyen VT, Kiss T, Michels AA, Bensaude O (2001) 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. *Nature* 414:322–325
- Nilsen TW (2003) The spliceosome: the most complex macromolecular machine in the cell? *Bioessays* 25:1147–1149
- Nilsen TW (2008) Endo-siRNAs: yet another layer of complexity in RNA silencing. *Nat Struct Mol Biol* 15:546–548
- Ørom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, Lai F, Zytnicki M, Notredame C, Huang Q, Guigo R, Shiekhattar R (2010) Long noncoding RNAs with enhancer-like function in human cells. *Cell* 143:46–58
- Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, Nagano T, Mancini-Dinardo D, Kanduri C (2008) Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell* 32:232–246

- Pang KC, Frith MC, Mattick JS (2006) Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends Genet* 22:1–5
- Pauler FM, Koerner MV, Barlow DP (2007) Silencing by imprinted noncoding RNAs: is transcription the answer? *Trends Genet* 23:284–292
- Perreault J, Perreault JP, Boire G (2007) Ro-associated Y RNAs in metazoans: evolution and diversification. *Mol Biol Evol* 24:1678–1689
- Persson H, Kvist A, Vallon-Christersson J, Medstrand P, Borg A, Rovira C (2009) The non-coding RNA of the multidrug resistance-linked vault particle encodes multiple regulatory small RNAs. *Nat Cell Biol* 11:1268–1271
- Pfeffer S, Zavolan M, Grässer FA, Chien M, Russo JJ, Ju J, John B, Enright AJ, Marks D, Sander C, Tuschl T (2004) Identification of virus-encoded microRNAs. *Science* 304:734–736
- Plaisance-Bonstaff K, Renne R (2011) Viral miRNAs. *Methods Mol Biol* 721:43–66
- Plath K, Fang J, Mlynarczyk-Evans SK, Cao R, Worringer KA, Wang H, de la Cruz CC, Otte AP, Panning B, Zhang Y (2003) Role of histone H3 lysine 27 methylation in X inactivation. *Science* 300:131–135
- Preker P, Nielsen J, Kammler S, Lykke-Andersen S, Christensen MS, Mapendano CK, Schierup MH, Jensen TH (2008) RNA exosome depletion reveals transcription upstream of active human promoters. *Science* 322:1851–1854
- Reis EM, Louro R, Nakaya HI, Verjovski-Almeida S (2005) As antisense RNA gets intronic. *OMICS* 9:2–12
- Ren B (2010) Transcription: enhancers make non-coding RNA. *Nature* 465:173–174
- Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Bruggmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129:1311–1323
- Rinn JL, Wang JK, Allen N, Bruggmann SA, Mikels AJ, Liu H, Ridky TW, Stadler HS, Nusse R, Helms JA, Chang HY (2008) A dermal HOX transcriptional program regulates site-specific epidermal fate. *Genes Dev* 22:303–307
- Robine N, Lau NC, Balla S, Jin Z, Okamura K, Kuramochi-Miyagawa S, Blower MD, Lai EC (2009) A broadly conserved pathway generates 3'UTR-directed primary piRNAs. *Curr Biol* 19:2066–2076
- Rosenblad MA, Larsen N, Samuelsson T, Zwieb C (2009) Kinship in the SRP RNA family. *RNA Biol* 6:508–516
- Samarsky DA, Fournier MJ, Singer RH, Bertrand E (1998) The snoRNA box C/D motif directs nucleolar targeting and also couples snoRNA synthesis and localization. *EMBO J* 17:3747–3757
- Scaria V, Jadhav V (2007) microRNAs in viral oncogenesis. *Retrovirology* 4:82
- Segref A, Mattaj JW, Ohno M (2001) The evolutionarily conserved region of the U snRNA export mediator PHAX is a novel RNA-binding domain that is essential for U snRNA export. *RNA* 7:351–360
- Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, Young RA, Sharp PA (2008) Divergent transcription from active promoters. *Science* 322:1849–1851
- Selenko P, Sprangers R, Stier G, Bühler D, Fischer U, Sattler M (2001) SMN tudor domain structure and its interaction with the Sm proteins. *Nat Struct Biol* 8:27–31
- Shamovsky I, Ivannikov M, Kandel ES, Gershon D, Nudler E (2006) RNA-mediated response to heat shock in mammalian cells. *Nature* 440:556–560
- Silva J, Mak W, Zvetkova I, Appanah R, Nesterova TB, Webster Z, Peters AH, Jenuwein T, Otte AP, Brockdorff N (2003) Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. *Dev Cell* 4:481–495
- Skryabin BV, Gubar LV, Seeger B, Pfeiffer J, Handel S, Robeck T, Karpova E, Rozhdestvensky TS, Brosius J (2007) Deletion of the MBII-85 snoRNA gene cluster in mice results in postnatal growth retardation. *PLoS Genet* 3:e235
- Sonkoly E, Bata-Csorgo Z, Pivarcsi A, Polyanka H, Kenderessy-Szabo A, Molnar G, Szentpali K, Bari L, Megyeri K, Mandi Y, Dobozy A, Kemeny L, Szell M (2005) Identification and characterization of a novel, psoriasis susceptibility-related noncoding RNA gene, PRINS. *J Biol Chem* 280:24159–24167

- Stöger R, Kubicka P, Liu CG, Kafri T, Razin A, Cedar H, Barlow DP (1993) Maternal-specific methylation of the imprinted mouse *Igf2r* locus identifies the expressed locus as carrying the imprinting signal. *Cell* 73:61–71
- Sun BK, Deaton AM, Lee JT (2006) A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization. *Mol Cell* 21:617–628
- Szegedi K, Sonkoly E, Nagy N, Németh IB, Bata-Csörgo Z, Kemény L, Dobozy A, Széll M (2010) The anti-apoptotic protein GIP3 is overexpressed in psoriasis and regulated by the non-coding RNA, PRINS. *Exp Dermatol* 19:269–278
- Tam OH, Aravin AA, Stein P, Girard A, Murchison EP, Cheloufi S, Hodges E, Anger M, Sachidanandam R, Schultz RM, Hannon GJ (2008) Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* 453:534–538
- Terranova R, Yokobayashi S, Stadler MB, Otte AP, van Lohuizen M, Orkin SH, Peters AH (2008) Polycomb group proteins *Ezh2* and *Rnf2* direct genomic contraction and imprinted repression in early mouse embryos. *Dev Cell* 15:668–679
- Trotochaud AE, Wassarman KM (2005) A highly conserved 6S RNA structure is required for regulation of transcription. *Nat Struct Mol Biol* 12:313–319
- Tufarelli C, Stanley JA, Garrick D, Sharpe JA, Ayyub H, Wood WG, Higgs DR (2003) Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nat Genet* 34:157–165
- Valadkhan S, Mohammadi A, Wachtel C, Manley JL (2007) Protein-free spliceosomal snRNAs catalyze a reaction that resembles the first step of splicing. *RNA* 13:2300–2311
- van Bakel H, Nislow C, Blencowe BJ, Hughes TR (2010) Most “dark matter” transcripts are associated with known genes. *PLoS Biol* 8:e1000371
- Van Horn DJ, Eisenberg D, O’Brien CA, Wolin SL (1995) *Caenorhabditis elegans* embryos contain only one major species of Ro RNP. *RNA* 1:293–303
- Vulliamy T, Marrone A, Goldman F, Dearlove A, Bessler M, Mason PJ, Dokal I (2001) The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita. *Nature* 413:432–435
- Wagschal A, Sutherland HG, Woodfine K, Henckel A, Chebli K, Schulz R, Oakey RJ, Bickmore WA, Feil R (2008) G9a histone methyltransferase contributes to imprinting in the mouse placenta. *Mol Cell Biol* 28:1104–1113
- Wang X, Arai S, Song X, Reichart D, Du K, Pascual G, Tempst P, Rosenfeld MG, Glass CK, Kurokawa R (2008) Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature* 454:126–130
- Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, Chiba H, Kohara Y, Kono T, Nakano T, Surani MA, Sakaki Y, Sasaki H (2008) Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* 453:539–543
- Willingham AT, Orth AP, Batalov S, Peters EC, Wen BG, Aza-Blanc P, Hogenesch JB, Schultz PG (2005) A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. *Science* 309:1570–1573
- Wutz A, Gribnau J (2007) X inactivation Xplained. *Curr Opin Genet Dev* 17:387–393
- Wutz A, Smrzka OW, Schweifer N, Schellander K, Wagner EF, Barlow DP (1997) Imprinted expression of the *Igf2r* gene depends on an intronic CpG island. *Nature* 389:745–749
- Wyers F, Rougemaille M, Badis G, Rousselle JC, Dufour ME, Boulay J, Régnault B, Devaux F, Namane A, Séraphin B, Libri D, Jacquier A (2005) Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* 121:725–737
- Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Münster S, Camblong J, Guffanti E, Stutz F, Huber W, Steinmetz LM (2009) Bidirectional promoters generate pervasive transcription in yeast. *Nature* 457:1033–1037
- Yakovchuk P, Goodrich JA, Kugel JF (2011) B2 RNA represses TFIIF phosphorylation of RNA polymerase II. *Transcription* 2:45–49
- Yamasaki Y, Kayashima T, Soejima H, Kinoshita A, Yoshiura K, Matsumoto N, Ohta T, Urano T, Masuzaki H, Ishimaru T, Mukai T, Niikawa N, Kishino T (2005) Neuron-specific relaxation of

- Igf2r imprinting is associated with neuron-specific histone modifications and lack of its antisense transcript air. *Hum Mol Genet* 14:2511–2520
- Yang Z, Zhu Q, Luo K, Zhou Q (2001) The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. *Nature* 414:317–322
- Yekta S, Shih IH, Bartel DP (2004) MicroRNA-directed cleavage of HOXB8 mRNA. *Science* 304:594–596
- Yik JH, Chen R, Nishimura R, Jennings JL, Link AJ, Zhou Q (2003) Inhibition of P-TEFb (CDK9/Cyclin T) kinase and RNA polymerase II transcription by the coordinated actions of HEXIM1 and 7SK snRNA. *Mol Cell* 12:971–982
- Zalfa F, Adinolfi S, Napoli I, Kühn-Hölsken E, Urlaub H, Achsel T, Pastore A, Bagni C (2005) Fragile X mental retardation protein (FMRP) binds specifically to the brain cytoplasmic RNAs BC1/BC200 via a novel RNA-binding motif. *J Biol Chem* 280:33403–33410
- Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT (2008) Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* 322:750–756
- Zheng R, Shen Z, Tripathi V, Xuan Z, Freier SM, Bennett CF, Prasanth SG, Prasanth KV (2010) Polypurine-repeat-containing RNAs: a novel class of long non-coding RNA in mammalian cells. *J Cell Sci* 123:3734–3744
- Zieve G, Benecke BJ, Penman S (1977) Synthesis of two classes of small RNA species in vivo and in vitro. *Biochemistry* 16:4520–4525