Shaping the Genome with Non-Coding RNAs

Xue Q.D. Wang, Jennifer L. Crutchley and Josée Dostie*

Department of Biochemistry, and Goodman Cancer Research Center, McGill University, 3655 Promenade Sir-William-Osler, Room 815A, Montréal, Québec, H3G1Y6, Canada

Abstract: The human genome must be tightly packaged in order to fit inside the nucleus of a cell. Genome organization is functional rather than random, which allows for the proper execution of gene expression programs and other biological processes. Recently, three-dimensional chromatin organization has emerged as an important transcriptional control mechanism. For example, enhancers were shown to regulate target genes by physically interacting with them regardless of their linear distance and even if located on different chromosomes. These chromatin contacts can be measured with the "chromosome conformation capture" (3C) technology and other 3C-related techniques. Given the recent innovation of 3C-derived approaches, it is not surprising that we still know very little about the structure of our genome at high-resolution. Even less well understood is whether there exist distinct types of chromatin contacts and importantly, what regulates them. A new form of regulation involving the expression of long non-coding RNAs (lncRNAs) was recently identified. lncRNAs are a very abundant class of non-coding RNAs that are often expressed in a tissue-specific manner. Although their different subcellular localizations point to their involvement in numerous cellular processes, it is clear that lncRNAs play an important role in regulating gene expression. How they control transcription however is mostly unknown. In this review, we provide an overview of known lncRNA transcription regulation activities. We also discuss potential mechanisms by which ncRNAs might exert three-dimensional transcriptional control and what recent studies have revealed about their role in shaping our genome.

INTRODUCTION

Organisms with large genomes face the interesting problem of having to contain their genetic material within very small cellular volumes. For example, a human diploid genome measures approximately 2 meters in length and must fit within a micron-sized nucleus. Our genomic DNA must consequently be extensively packaged in a functional manner that allows activities like transcription, DNA replication and repair to occur properly. Although exactly how our genome is functionally organized *in vivo* is poorly understood, several key features appear common to most cell types (Fig. (**1**)). First, chromosomes are known to occupy distinct nuclear areas that are termed "chromosome territories" (CT). Gene-rich chromosomes tend to localize to the center of the nucleus and gene-poor near the periphery (reviewed in [1]). Genomic domains with similar activities and co-regulated genes often co-localize *in vivo* (reviewed in [2]). Moreover, genes located near anchor points where interphase chromosomes attach to the nuclear matrix tend to be poorly transcribed (reviewed in [3]). Together, these observations clearly reveal the existence of an intimate relationship between the organization of our genome in the nuclear space and gene activity.

At the molecular level, spatial genome organization can critically affect gene expression. *In vivo*, genomic DNA exists in the form of chromatin where DNA is associated with histone and non-histone proteins. While it is well known that chromatin composition can directly shape gene activity [4, 5], three-dimensional chromatin organization is also emerging as an important gene regulation mechanism. For example, it was shown that distal control DNA elements could regulate target genes located on the same or on different chromosomes by physically interacting with them [6, 7]. This type of long-range regulation has been reported for genes located throughout the genome that are involved in various cellular pathways [8-12]. Consequently, the genome is now viewed as a functional network of physical contacts within (*cis*) and between (*trans*) chromosomes. These physical DNA contacts can be mapped at high-resolution with technologies such as the "chromosome conformation capture" (3C) and other 3C-related methods including 3C-Carbon Copy (5C) and Hi-C [13-17]. Given the recent innovation and sophistication of these approaches, it is not surprising that we still know very little about genome organization at the molecular level. In fact, most questions remain unanswered: How many types of contacts exist? What are their relative contributions to overall genome organization and to various genome functions? What are their tissue-specificities? What are their roles in the establishment and/or maintenance of cellular identity? What is their involvement in human disease? Answers to these questions will require extensive mapping and classification

*Address correspondence to this author at the Department of Biochemistry, and Goodman Cancer Research Center, McGill University, 3655 Promenade Sir-William-Osler, Room 815A, Montréal, Québec, H3G1Y6, Canada; Tel: 1-(514)-398-4975; Fax: 1-(514)-398-7384; E-mail: josee.dostie@mcgill.ca
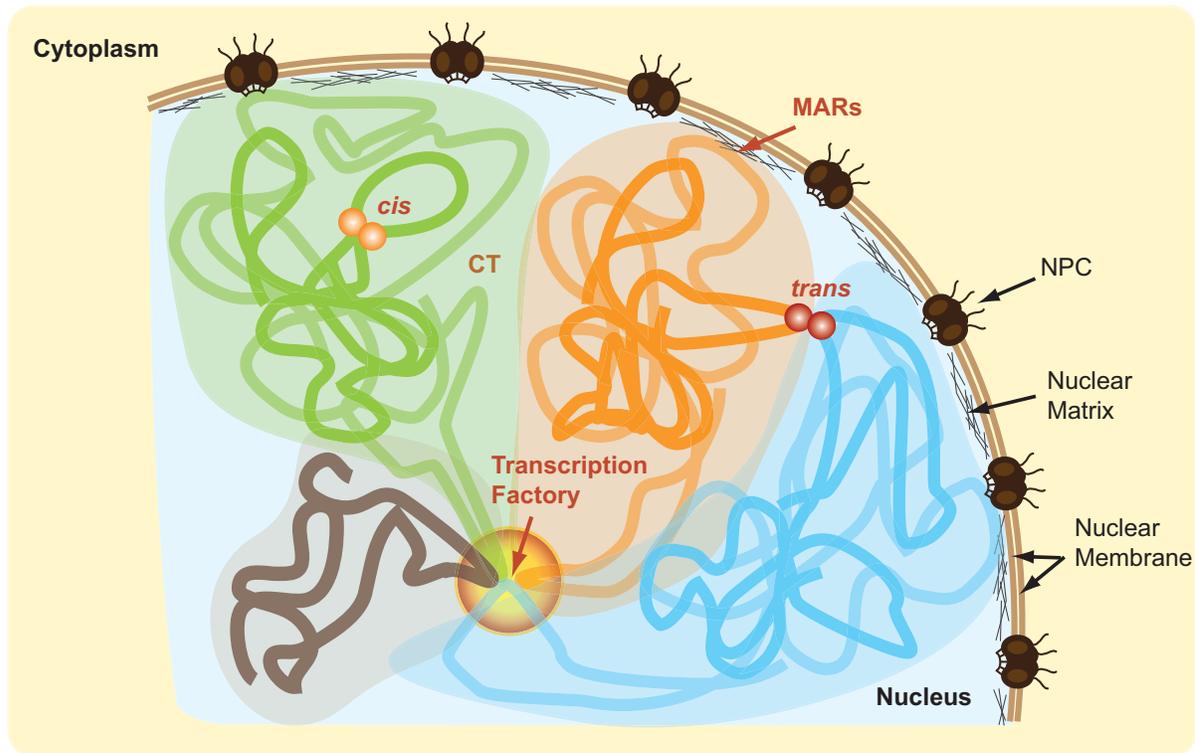
Fig. (1). *In vivo genome organization.* Individual chromosomes are illustrated by thick green, orange, blue and grey strokes. Respective chromosome territories (CT) are highlighted in corresponding colors. Chromatin interactions within (*cis*) and between (*trans*) chromosomes are represented by orange and red spheres. Transcription factory is shown as a graded circle. MARs: matrix attachment regions. NPC: nuclear pore complex. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper).

of chromatin interactions genome-wide under various cellular conditions. Only then will we be able to distinguish between regulated and structural contacts, and understand the outcome of altering either type on cellular function.

Even less well understood is what mediates chromatin interactions and how they are regulated. It was shown that chromatin contacts could be mediated by tissue-specific transcription factors. Such is the case for the beta-globin cluster where DNA looping between the locus control region (LCR) and activated beta-globin genes was shown to require the GATA-1 transcription factor and co-factor FOG-1 [18]. This type of long-range control mechanism where enhancers and distal promoters are brought in close physical proximity to each other by transcription factors has been found at other loci. These include many estrogen-regulated genes [19], the B lymphocyte-specific induction of *CIITA* by PU.1 [20], and *MYC* activation by TCF and the beta-catenin co-activator complex [21].

Other chromatin-binding proteins such as the CCCTC-binding factor (CTCF) and cohesin also appear to play critical roles in genome organization and gene expression [22, 23]. Mammalian CTCF is a DNA-binding protein associated with insulator sequences, boundary elements and imprinting control regions, all of which are thought to organize our genome into functional subdomains. CTCF is known to form networks of chromatin loops genome-wide and is viewed as a master regulator of spatial genome organization. It was also shown that CTCF co-localizes

extensively with cohesin during interphase by directly binding to its STAG1/2 component [24]. Cohesin is a four protein subunit ring complex important in DNA repair, chromosome segregation and transcription regulation. Within some genomic regions, CTCF-co-localized cohesin rather than solely CTCF itself appears to be responsible for chromatin contacts. For example, cohesin knockdown was shown to abolish long-range *cis* chromatin contacts associated with *IFNG* locus expression in T helper 1 cells, and to reduce gene expression without affecting CTCF binding [25]. Similar cohesin-dependent loops and gene expression effects have also been observed at other loci. These include the beta-globin locus [26] and the apolipoprotein gene cluster [27]. Although they are often found together on the chromatin, CTCF and cohesin do not however exclusively co-localize with each other. Since cohesin does not bind DNA directly, this observation points to the existence of CTCF-independent cohesin recruitment and looping mechanisms. One such mechanism could involve tissue-specific transcription factors. Indeed, CTCF-independent cohesin binding sites frequently co-localize with tissue-specific transcription factors [28]. Alternatively, transcriptional coactivators such as mediator could recruit cohesin [29].

There are many ways by which chromatin interactions could be regulated. First, by modifying the DNA itself with cytosine methylation and consequently altering protein association. An example for this type of regulation is found at the imprinted *Igf2/H19* locus. On the maternal allele,

expression of the *Igf2* gene is inhibited through the formation of a repressive loop encompassing the gene. This loop requires the recruitment of cohesin through CTCF bound at the DMR1 (DNA methylated region 1) and ICR (imprinted control region) sequences located upstream of *Igf2* and *H19* respectively [30]. Bound CTCF was also shown to recruit the polycomb repressive complex 2 (PRC2) at the *Igf2* promoter through its Suz12 subunit, thereby leading to allele-specific trimethylation of histone H3 lysine 27 (H3K27me3) and suppression of the maternal *Igf2* allele. In contrast, binding of CTCF to the ICR is negatively regulated by DNA methylation on the paternal allele. This lack of binding in turn prevents formation of the repressive loop, and enables access of the *Igf2* promoter to an enhancer located downstream of *H19* [31, 32].

Chromatin contacts could also be regulated by controlling access to DNA sequences with post-translational histone modifications (PTMs), the use of histone variants or by altering nucleosome positioning. Similarly, post-translational modification or changes in expression level of non-histone chromatin-binding proteins could represent important mechanisms to regulate chromatin contacts. Additionally, non-coding RNAs (ncRNAs) and their protein complexes could regulate the three-dimensional architecture of our genome. ncRNAs are a broad class of transcripts consisting of structural (rRNAs, tRNAs, snRNAs, snoRNAs, etc.), regulatory (miRNAs, piRNAs, etc.), and of sense/anti-sense transcripts, whose functions remain mostly uncharacterized. The latter RNA class includes transcriptional "features" (eRNAs, tiRNAs), and a very large number of long non-coding RNAs (lncRNAs) ranging in length from 200 nt to 100 kb. Although lncRNAs can be transcribed nearby known protein-coding genes or from their introns, they are often produced intergenically and hence, are sometimes referred to as "lincRNAs". While a number of transcripts falling in that category have been studied for some time (e.g. Xist, see below), the realization that lncRNAs represent a very abundant RNA subclass is relatively recent and originates from genome-wide transcriptome studies [33-40]. lncRNAs are estimated to be at least as abundant as mRNAs, and like mRNAs, their expression levels and patterns are often tissue-specific. The heterogeneity in sequence, structure and size, along with the widely different subcellular localizations of lncRNAs points to their involvement in numerous cellular processes [41]. In the nucleus, a few lncRNAs were shown to inhibit or activate transcription, which is in contrast to other ncRNAs thought to represent basic features of transcription.

Until recently, DNA sequence and associated proteins were thought to mainly dictate spatial genome organization. With the abundance and diversity of lncRNAs rivaling that of mRNAs, it seems likely that they also play an essential role in regulating chromatin organization. In fact, there is already evidence for the implication of ncRNA in general genome architecture [42], and in maintaining the integrity of nuclear bodies [43]. Conversely, chromatin conformation is thought to play a role in coordinating gene expression with lncRNAs [44]. Below, we report on the nuclear activity of a number of best-characterized lncRNAs, and discuss how lncRNAs might be instrumental in shaping the three-dimensional organization of our genome.

lncRNAs AS INHIBITORS OF TRANSCRIPTION

Although most lncRNAs have only recently been discovered, a few transcripts belonging to this category have already been scrutinized for some time. Amongst these, the X-inactive specific transcript (Xist) responsible for equalizing gene expression between mammalian males and females was first identified in the early 90's [45, 46]. As described below, this lncRNA silences genes in *cis* by coating the chromosome. More recently, a lncRNA termed HOTAIR was shown to function in *trans*. Like the Xist transcript, HOTAIR associates with the polycomb repressive complex 2 (PRC2) and is thought to repress transcription by increasing H3K27me3 levels at specific genes. How these transcripts bind and help target polycomb complexes specifically to specific regions however is unknown. lncRNAs are also important for gene imprinting. For example, the Air lncRNA was shown to silence 3 genes in *cis* on the paternal allele at the *Kcnq-1* imprinted region. Like Xist, the Air transcript coats adjacent chromatin, but instead recruits the G9a histone 3 lysine 9 methyltransferase (H3K9me3) to repress transcription. Interestingly, other lncRNAs such as Kcnq1ot1 silence by binding both PRC2 and G9a to increase the levels of H3K9me3 and H3K27me3. Although repression by PRC2 binding and elevated H3K27me3 is thus far the most common mode of inhibition (see also ANRIL below), we can expect that many types of mechanisms will mediate lncRNA activity in *cis* and *trans* (Fig. (**2**)). In fact, alternative mechanisms have already been identified. For example, the p53-inducible lincp21 ncRNA was shown to bind hnRNP-K and inhibit gene expression genome-wide following p53 induction. We report below on the function and mechanism of some of these inhibitory lncRNAs.

Tsix Repression of Xist

In mammals, flies and worms, females differ from males by having two X chromosomes rather than one. The abundance of X-linked gene products must therefore be equalized between sexes in the somatic cells of these organisms. In mammals, gene dosage compensation is achieved by inactivating one X chromosome in females through a process called X chromosome inactivation (XCI). XCI requires the Xist lncRNA expressed in *cis*, which coats one female X chromosome and inactivates it by recruiting polycomb proteins (reviewed in [47]). Xist is expressed from a region on the X chromosome called X inactivation center (*Xic*) where its transcription is controlled by multiple lncRNAs. Indeed, the *Xic* encodes at least 6 other lncRNAs, 2 of which are known to antagonistically regulate Xist expression. On the one hand, the Tsix lncRNA transcribed antisense to *Xist* is known to repress Xist expression on the active X (Xa). Conversely, the Jpx lncRNA activates *Xist* on the inactive X (Xi). Tsix represses *Xist* transcription *via* multiple mechanisms that include recruiting DNA methyltransferase activity and the RNA interference machinery locally. Tsix is also thought to change the epigenetic chromatin state of *Xist*, thereby preventing Xist and RepA (a short transcript corresponding to the 5'end of
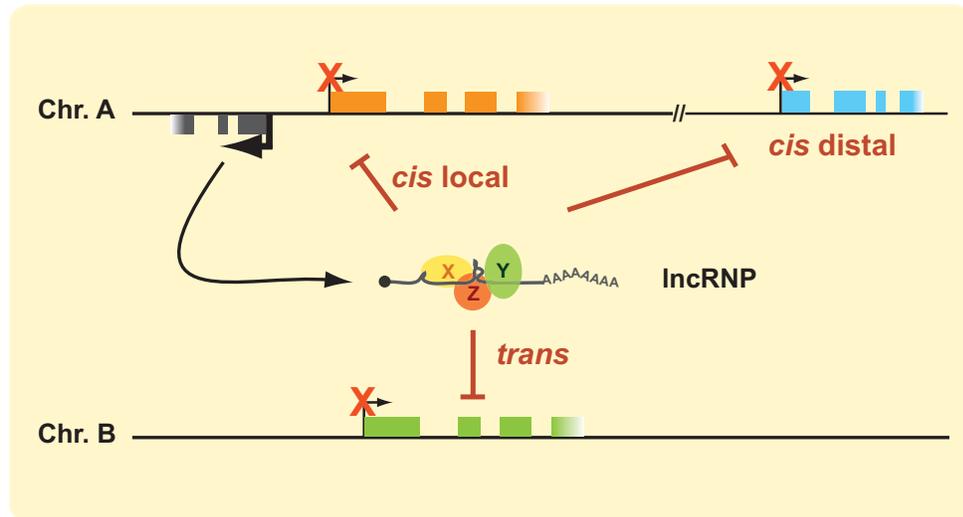
Fig. (2). *lncRNAs as inhibitors of transcription.* lncRNAs expressed from anywhere in the genome can be packaged with transcript-specific proteins into functional lncRNPs (long non-coding ribonucleoproteins). The X, Y and Z ovals represent putative transcript-specific proteins associated with lncRNAs. lncRNPs could act locally or distally in *cis*, or in *trans*. Protein-coding genes are represented by orange, blue and green boxes. ncRNA gene is illustrated in grey. Transcription start sites (TSSs) are indicated by bent arrows, with thick arrows showing greater transcription activity. TSS of inhibited genes are indicated by thin bent arrows marked with an "X". (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper).

Xist) from recruiting PRC2 (reviewed in [47]). Interestingly, the *Tsix* gene itself is positively regulated on Xa by another *Xic*-encoded lncRNAs called Xite. How Jpx regulates *Xist* will be described below in the next section. Together, XCI studies reveal that lncRNAs can regulate the transcription of other ncRNAs in addition to protein-coding genes. lncRNAs may therefore form very complex regulatory networks that control transcription genome-wide.

HOTAIR

In 2007, Chang and colleagues reported the first comprehensive analysis of the transcriptional landscapes of the four human *HOX* clusters in 11 different fibroblast cell types [48]. In this study, 231 Hox ncRNAs were identified at 5-bp resolution on custom tilling arrays. Like *HOX* genes, these ncRNAs displayed differential expression along developmental axes and marked chromosomal domains with different histone methylation patterns and RNA pol II accessibility. The *HOX* clusters encode highly conserved eukaryotic transcription factors with pivotal roles in body patterning during development. *HOX* genes were originally identified in *Drosophila melanogaster* and their mammalian counterparts are known to have similar developmental functions. For example, studies in mouse models demonstrated that dysregulated *HOX* gene expression in early development could lead to severe limb and skeletal malformations [49-51]. *HOX* gene expression must also be controlled in adult tissues as overexpression of some *HOX* genes was found to be a hallmark of certain cancers such as leukemia [52, 53]. Although *HOX* regulation is clearly essential during development and for human health, how these genes are controlled remains poorly understood to this day. *HOX* clusters are regulated epigenetically and in pluripotent cells rest in so-called bi-valent domains where activating H3K4me3 and repressive H3K27me3 histone marks co-exist [54, 55]. The study led by Chang and

colleagues uncovered not only a promising new level of *HOX* regulation, but also showed for the first time that lncRNAs can regulate genes in *trans*. Indeed, a lncRNA transcribed at a boundary between active and repressive chromatin states in the *HOXC* cluster was found to repress the expression of 5' end *HOXD* genes located on another chromosome. This RNA is transcribed in an anti-sense manner to the *HOXC* genes and was named HOTAIR for "*HOX* Antisense Intergenic RNA". HOTAIR is a 2,158 nt long spliced and polyadenylated RNA that binds PRC2 and silences genes by increasing H3K27me3 levels at specific sites [48, 56]. While the 5' end of HOTAIR was shown to bind PRC2 directly through its Suz12 subunit, its 3' end was also shown to bind the LSD1/CoREST/REST H3K4 demethylase complex. Thus, HOTAIR is thought to provide a scaffold for enzyme complexes that increase H3K27me3 (PRC2) and decrease H3K4me3 levels (LSD1/coREST/ REST) to silence gene expression [56].

In addition to the *HOX* genes themselves, which are involved in certain types of leukemia, Hox ncRNAs were also shown to be dysregulated in breast cancer [57]. Amongst these, high HOTAIR expression was found to correlate well with increased metastasis and lower patient life expectancy. Conversely, loss of HOTAIR expression reduced cancer invasiveness suggesting that lncRNAs, particularly HOTAIR, might significantly alter the epigenome of cancer cells. The association of lncRNAs with polycomb complexes sparks numerous questions: How can so many different lncRNA sequences bind directly to PRC2? How do these transcripts help target protein complexes to specific genes? Does the actual lncRNA sequence actively participate in targeting the silencing complexes to given genes? Might the local accumulation of lncRNAs around their transcription sites be sufficient to recruit histone-modifying enzymes and alter chromatin states?

Local INK4/ARF Locus Control by the ANRIL Non-Coding Transcript

The human *INK4/ARF* locus encodes three tumor suppressor genes named $p14^{ARF}$, $p15^{INK4B}$, and $p16^{INK4A}$ that are often altered in cancers [58]. In normal tissues, these genes are epigenetically repressed by PRC2-mediated deposition of H3K27me3, and are activated upon cellular stress to halt proliferation by inhibiting CDK4, CDK6 and MDM2. In 2007, a lncRNA transcribed anti-sense to the three tumor suppressor genes was identified along the *INK4/ARF* locus in human testes [59]. The mature 3,834-nt transcript was named ANRIL for "Antisense Non-coding RNA in the *INK4* Locus" and was derived from a 126 kb gene composed of 19 exons. In addition to full-length ANRIL, a shorter 2,659-nt form was also found in the same tissue, and 8 additional splice variants were later identified in several cell lines [60]. Although the roles of individual forms are unknown, full-length ANRIL was shown to repress $p15^{INK4B}$ expression [61]. Indeed, ANRIL knockdown in actively growing cells significantly increased expression of $p15^{INK4B}$. Similarly, induction of $p15^{INK4B}$ and $p16^{INK4A}$ by oncogenic Ras was found to repress ANRIL expression. It was known that the three genes were silenced by PRC2 under normal cellular condition, and ANRIL knockdown significantly reduced PRC2 occupancy and H3K27me3 at the *INK4/ARF* locus. ANRIL was also found to co-immunoprecipitate with Suz12 and CBX7, which are components of PRC2 and the polycomb repressive complex 1 (PRC1), respectively [61, 62]. Like PRC2, PRC1 is a multisubunit protein complex and can bind H3K27me3 through its CBX components [63]. PRC1 is also responsible for the ubiquitination of H2AK119 known to inhibit RNA pol II transcription elongation [64-66]. Thus, the lncRNA ANRIL silences the $p15^{INK4B}$ gene by recruiting both polycomb repressive complexes to halt RNA pol II transcription elongation.

Another anti-sense transcript named p15-antisense (p15AS) was identified at the *INK4/ARF* locus in 2008 [67]. Unlike ANRIL, this RNA measures 34.8 kb in length and overlaps only the $p15^{INK4B}$ gene. Consistent with the frequent epigenetic silencing of $p15^{INK4B}$ in leukemia patient samples, higher p15AS levels were often found in acute lymphoblastic (ALL) and myeloid leukemias (AML). Thus, it was suggested that p15AS could facilitate cancer progression by silencing the $p15^{INK4B}$ tumor suppressor gene. Surprisingly, p15AS overexpression increased the repressive H3K9me2 mark rather than H3K27me3, and decreased the levels of activating H3K4me2 near the TSS of the endogenous and exogenous $p15^{INK4B}$ gene. Similar histone modifications were also found at the $p15^{INK4B}$ promoter in cells where the gene is silent and no DNA methylation changes were observed. These results indicate that p15AS is capable of regulating its parental gene in *cis* and in *trans* through the formation of a type of heterochromatin. Although the silencing mechanism has not yet been identified, p15AS appears to trigger silencing but not maintain it since removal of the transcript did not reverse its effects.

Gene Imprinting by Air and Kcnq1ot1

RNA transcripts important for gene imprinting were amongst the first lncRNAs identified in human and mouse. Gene imprinting is an epigenetic process that selectively represses one or more genes from a parental allele. This process was originally thought to result mainly from differential DNA methylation at CpG islands [68], but there is now overwhelming evidence that histone-modifying complexes and lncRNAs also play a significant role in this process. The Air and Kcnq1ot1 transcripts are two well-studied lncRNAs involved in gene imprinting [69, 70]. Interestingly, these transcripts differ from Xist, HOTAIR and ANRIL by their much larger sizes and absence of splicing [70, 71]. The Air transcript (antisense *Igf2r* RNA) is a 107,796-nt long antisense RNA transcribed from the second intron of the *Igf2r* gene to the 3' end of the *Mas* gene [72]. Air is expressed only from the paternal allele where it silences the overlapping *Igf2r* gene and two other genes (*Slc22a2 and Slc22a3*) downstream of *Igf2r* [73]. The three genes are not consecutive and span approximately 300 kb indicating that Air acts specifically in *cis*. Although it is not spliced, Air is transcribed by RNA pol II and thus bears a 5' m(7)GpppN cap structure [74]. The full-length RNA transcript appears to be required for imprinting since deletion of its promoter or gene truncation prevented paternal silencing of the three genes [70]. Similarly to Xist, the Air lncRNA was found to coat its target genes but instead recruits the G9a histone methyltransferase rather than the subunits of PRC2 [75]. G9a is a SET domain-containing histone methyltransferase similar to polycomb Ezh2 but preferentially methylates H3 on lysine 9 (H3K9me3) [76-78]. Truncation substantially decreased Air lncRNA concentration around target genes, reduced G9a recruitment at the *Igf2r* locus, and failed to inhibit the *Igf2r*, *Slc22a2 and Slc22a3* genes [75].

The kcnq1ot1 transcript is another well-studied lncRNA involved in imprinting [79, 80]. Kcnq1ot1 (also called lit1 for "long QT intronic transcript 1") is an antisense transcript that shares many common attributes with Air. First, it is a 91 kb RNA transcribed by RNA pol II, is capped, polyadenylated, but not spliced [71]. The transcript is expressed from an unmethylated promoter on the paternal allele of the *kcnq1* (potassium voltage-gated channel, KQT-like subfamily, member 1) gene and stays in the nucleus to silence protein coding genes in *cis* within the region [79, 81, 82]. In contrast to Air, which represses three genes at the *Igf2r* locus, Kcnq1ot1 can silence up to 10 genes in the *Kcnq1* cluster. Surprisingly, Kcnq1ot1 interacts with both G9a and the Ezh2 and Suz12 components of the PRC2 complex, and silences the surrounding genes by spreading H3K9me3 and H3K27me3 marks [71, 83]. A silencing domain was also identified at the 5' end of the Kcnq1ot1 and further analysis revealed that Kcnq1ot1 silences gene expression not only by recruiting HMTs to remodel chromatin structure, but also by re-localizing its target genes near heterochromatin at the perinucleolar region [84].

LincRNA-p21

The *p53* tumor suppressor gene encodes a key transcription factor, which is mutated in over 50% of human cancers. p53 is induced by many types of cellular stresses and exerts its tumor suppressor properties primarily through its ability to act as a sequence-specific transcription factor [85-87]. Many genes are upregulated or downregulated

during the p53 transcriptional response. While p53-mediated gene activation is relatively well understood, how this transcription factor could lead to gene repression was unknown until recently. As lncRNAs had been shown to silence genes by recruiting chromatin-modifying complexes, it seemed possible that p53 might repress genes by activating lncRNAs. In 2010, Rinn and colleagues used high-resolution tiling arrays featuring predicted lncRNAs to probe the ncRNA transcriptome of p53-induced cells [88]. The custom tiling arrays included 400 genomic regions outside known protein-coding genes that were selected based on the presence of active H3K4me3 and H3K36me3 histone marks [89, 90]. Over 30 lncRNAs were induced by p53 in two different cellular systems indicating that like protein-coding genes, numerous lncRNAs are also temporally regulated during the p53 transcriptional response. Two lncRNAs containing canonical p53 binding sites in their promoters were shown to be *bona fide* p53 transcription targets and one was retained for further analysis. The lncRNA named lincRNA-p21 was interesting because of its curious genomic position approximately 15 kb upstream of the p21 cell cycle regulator, which is a well-known p53 target gene. lincRNA-p21 is a 3.1 kb spliced and polyadenylated RNA transcribed from the opposite strand to *p21*, and is conserved in human and mouse. Individual knockdown of p53 and lincRNA-p21 revealed a strong overlap in derepressed genes indicating that lincRNA-p21 participates in p53-mediated gene repression. Common derepressed genes were enriched in those involved in cell cycle arrest and apoptosis although further analysis indicated that lincRNA-p21 increases apoptosis without any significant effects on the cell cycle. Intriguingly, lincRNA-p21 was found to interact with hnRNP-K through a 780 nt region at its 5'end. Although deletion of this region abolished the ability of lincRNA-p21 to induce apoptosis, the region alone was not sufficient to trigger it. Furthermore hnRNP-K knockdown also showed a strong overlap in derepressed genes with lincRNA-p21 and p53, indicating that hnRNP-K is involved in the repression of many p53 downregulated genes likely through lincRNA-p21. Accordingly, the authors found that hnRNP-K is at the promoter of many p53-repressed genes by chromatin immunoprecipitation. How hnRNP-K can repress transcription however is unknown and it will be interesting to see how it is specifically recruited at gene promoters.

lncRNAs AS ACTIVATORS OF TRANSCRIPTION

Transcription activation by lncRNAs is a much more recent discovery than inhibition. Antisense ncRNAs transcribed intergenically from the *HOXA* cluster were amongst the first transcripts found to induce protein-coding genes. Like HOTAIR, *HOXA*-encoded lncRNAs are tissue-specific, transcribed by RNA pol II, spliced and polyadenylated. However, these transcripts appear to mainly induce the expression of neighboring genes in *cis*. In fact, no lncRNA has yet been shown to induce protein-coding genes in *trans*. Activating lncRNAs have since been reported genome-wide in a number of cell types. As for *HOXA* lncRNAs, they are often tissue-specific and transcribed by RNA pol II. However, some of these transcripts were found to induce distal genes within at least 300 kb away without necessarily affecting the ones adjacent to them. This type of

specific long-range regulation suggests that three-dimensional chromatin architecture might be important for their activity. Activation by lncRNAs does not seem to be restricted to protein-coding genes since XCI was recently shown to require this type of mechanism. There will likely be many mechanisms by which lncRNAs can activate transcription (Fig. (**3**)) although most are currently unknown. Below are examples of lncRNAs thought to play a role in *cis* transcription activation.

Hox-Encoded ncRNAs, HOTAIRM1, and HOTTIP

An interesting aspect to *HOX* cluster regulation is reflected in the developmental expression pattern of their genes. During development, *HOX* genes are expressed in a spatio-temporal manner that is colinear with their positions along chromosomes [91, 92]. For example, genes located at the 3' end of the clusters are usually expressed more anteriorly and earlier in the embryo than those at the 5' end. Although genetic evidence suggests a role for three-dimensional chromatin organization in colinearity, how this type of *HOX* regulation occurs is unclear [93]. In early 2007, Orlando and colleagues used the well-described NT2/D1 cell differentiation system to investigate the role of intergenic *HOXA* ncRNAs in regulating *cis* gene expression [94]. NT2/D1 are human pluripotent embryonic carcinoma cells that can be terminally differentiated into neural-like cell lineages upon treatment with retinoic acid (RA). RA treatment also recapitulates the 3'-5' induction pattern of the *HOXA1* to *A5* 3' end genes in developing axial systems. The study showed that RA induces antisense *HOXA* ncRNAs in a manner that follows the collinear activation pattern of the protein-coding genes. Importantly, ncRNA induction was also accompanied by local changes in histone modifications and loss of repressive PcG complexes. It was proposed that antisense ncRNA transcription is fundamentally important to open and maintain the active state of *HOX* clusters during RA induction. In this model, the process of transcription itself rather than transcripts *per se* could represent an anti-silencing mechanism. The authors suggested that ncRNA genes might contain regulative epigenetic DNA elements and that transcription through them may counteract repression and maintain RA activation. Although we still do not know which elements could be at these sites, an attractive possibility is that transcription displaces chromatin-looping factors and disrupts repressive three-dimensional conformations. We previously reported the presence of long-range chromatin loops in the transcriptionally silent *HOXA* cluster in THP-1 and NT2/D1 cells [95, 96]. We found that RA treatment abolished these loops in NT2/D1 cells and suggested that the insulator-binding protein CTCF might be an important looping factor [96]. It will be interesting to determine the role of CTCF in regulating *HOXA* gene expression.

HOXA ncRNA expression is not restricted to the induction of pluripotent cells with RA. *HOX* cluster transcriptome analysis in human fibroblasts actually identified hundreds of ncRNAs throughout the four clusters [48]. By using cells corresponding to various anatomical sites along developmental axes, this study revealed that ncRNA transcription is often tissue-specific and might demarcate silent and active chromatin domains. A tissue-
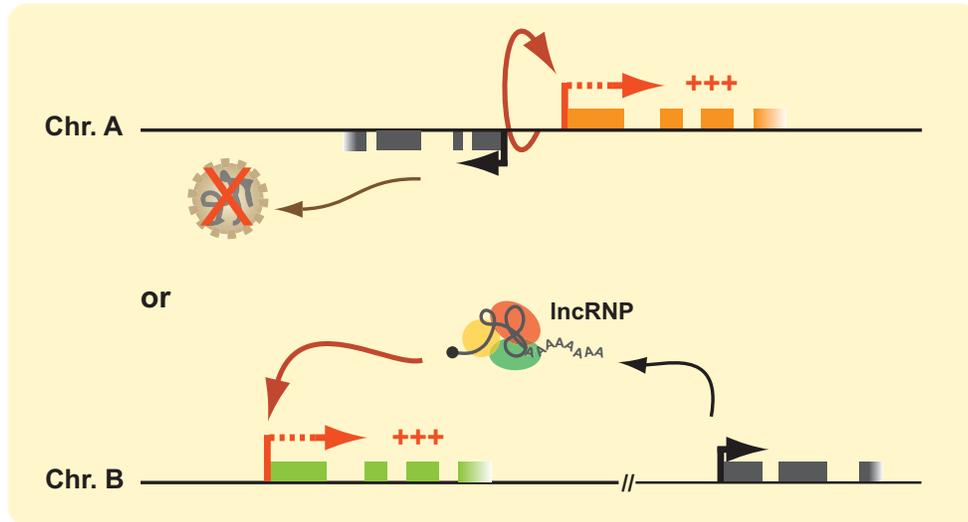
Fig. (3). *lncRNAs as activators of transcription.* (*top*) The process of lncRNA transcription and not the transcript itself could be responsible for *cis* activation of neighboring genes by maintaining or initiating the chromatin structure in an active state. (*bottom*) lncRNAs can be packaged with transcript-specific proteins into lncRNPs to selectively activate gene expression in *cis*. Protein-coding genes are represented by orange and green boxes. lncRNA genes are illustrated in grey. Transcription start sites (TSSs) are indicated by bent arrows. Thick red bent arrows indicate active protein coding genes and thick black arrows active lncRNA genes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper).

specific antisense lncRNA transcribed between *HOXA1* and *A2* was more recently identified in the myeloid cell lineage [97]. Because of its myeloid-specific expression, the transcript was termed HOTAIRM1 for "*HOX* Antisense Intergenic RNA Myeloid 1". HOTAIRM1 is 500 nt in length, transcribed by RNA pol II, spliced and polyadenylated. It is the most highly expressed *HOXA* lncRNA product following RA differentiation of NB4 cells into granulocytes and is highly expressed in leukocytes during normal human hematopoiesis. HOTAIRM1 knockdown in NB4 cells prevented RA induction of *HOXA1* and *A4*, and attenuated the induction of the *CD11b* and *CD18* granulocyte maturation genes. These results therefore suggest that transcription activation by HOTAIRM1 may not be restricted to its immediate *HOXA* neighboring genes, and it will be intriguing to determine its role in hematopoiesis and leukemogenesis.

An article recently published by the Chang group also reports a lncRNA at the distal 5' end of the *HOXA* cluster termed HOTTIP for "*HOXA* Transcript at the distal TIP" [44]. This 3,764-nt ncRNA is spliced, polyadenylated, and preferentially expressed at distal anatomical sites where 5' end *HOX* genes are usually expressed. For example, HOTTIP was detected in human foreskin fibroblasts but not in lung. It was also expressed in mouse limbs at E13.5 and in chick limb buds. HOTTIP was shown to positively regulate the expression of 5' *HOXA* genes in human distal fibroblasts and developing chick limbs. Indeed, knockdown of HOTTIP in foreskin fibroblasts resulted in lower 5' *HOXA* gene expression and reduced H3K4me3 at promoters, but did not affect the levels of H3K27me3 along the cluster. Consistent with the role of 5' end *HOXA* genes in limb development, HOTTIP shRNA injection in chicken embryos interfered with the formation of distal bony elements of the wing. These results show that HOTTIP might promote 5' end

HOXA transcription epigenetically by increasing the levels of activating histone marks. The mixed lineage leukemia proteins MLL1 and MLL2 are primarily responsible for trimethylation of H3K4 at the *HOX* clusters and HOTTIP was found to interact directly with the WDR5 component of the MLL modifying complex. HOTTIP knockdown reduced MLL-1 and WDR5 binding at the 5' end *HOX* gene promoters suggesting that the lncRNA is required to recruit the methyltransferase complex to the chromatin. However, ectopic expression of HOTTIP in non-expressing cells did not activate the transcription of 5' end *HOXA* genes. Similarly, HOTTIP overexpression in foreskin fibroblasts did not further enhance *HOXA* gene expression or rescue the effects of depleting endogenous nascent HOTTIP RNA. These results were explained with the suggestion that HOTTIP might only regulate genes when expressed immediately in *cis* of its targets and by hijacking pre-existing higher-order chromosomal structures that group regulated genes in close physical proximity to nascent HOTTIP RNA. Accordingly, 5C analysis of the *HOXA* cluster in fibroblasts indicated that HOTTIP expression correlates with enhanced chromatin interactions at the cluster 5' end between *HOXA13* and *HOXA7*. Frequent chromatin contacts corresponded to higher H3K4me3 levels and low H3K27me3 marks, and HOTTIP knockdown had no effect on three-dimensional chromatin organization. Thus, HOTTIP might represent an example whereby spatial chromatin organization is required to regulate gene expression by lncRNAs in *cis*. We have previously shown that the transcriptionally silent *HOX* clusters are extensively folded in different cell lines [95, 96]. Our findings along with the *HOXA* structural data summarized above and the exquisite richness of *HOX* clusters in lncRNAs suggest that the clusters might be regulated in a cell type-specific manner by different mechanisms involving both chromatin architecture and lncRNAs.

Activating lncRNAs (ncRNA-a)

In 2010, Orom and colleagues used GENCODE annotation data to search for uncharacterized human lncRNAs [98, 99]. By focusing on intergenic transcripts classified to have no coding potential by the HAVANA group (Human and Vertebrate Analysis and Annotation; Sanger Institute), they identified 3019 putative long non-coding RNAs. These transcripts displayed chromatin signatures similar to protein-coding genes with high H3K4me3 at their 5' ends and H3K36me3 along their bodies when transcriptionally active. In contrast to mRNAs however, the identified lncRNAs were shorter with an average transcript size of 800 nt. Their genes were also less complex than protein-coding genes with nearly 50% of them featuring a single intron. lncRNA expression analysis on custom arrays indicated that approximately 20% were ubiquitously expressed while over 33% displayed differential expression patterns across cell lines. Furthermore, it was found that the expression of many lncRNAs was regulated during cellular differentiation. Interestingly, functional knockdown analysis revealed an enhancer-like function for a set of lncRNAs. Indeed it was found that depletion of lncRNAs at multiple loci could result in a specific decrease in the expression of protein-coding genes in *cis*. Transcription activation was shown to be RNA-dependent, as evidenced by heterologous transcription assays. Importantly, some of these "ncRNA-a" were found to activate the transcription of critical regulators of development and cellular differentiation. For example, the Snai1 and Snai2 master regulators of hematopoiesis were downregulated following depletion of neighboring lncRNAs. Specifically, ncRNA-a7 RNAi knockdown was accompanied with less Snai1 expression and recapitulated the effects of Snai1 depletion itself. Interestingly, ncRNA-a7 depletion also led to a decrease in the expression of the *Aurora-kinase A* gene located 6 Mb away from the lncRNA gene. In fact, transcriptional effects were not restricted to immediate neighboring genes for many ncRNA-a, and could extend at least 300 kb away from the transcription units without necessarily affecting all genes in between. Thus, like *HOX*-encoded lncRNAs, these results suggest that spatial chromatin organization might play an important role in the specific activation of target genes.

Jpx Activation of Xist

In addition to activating the transcription of protein-coding genes, lncRNAs can also induce the expression of other lncRNAs. This type of regulation was found to be essential for XCI where Xist expression is repressed by Tsix on the Xa but induced by Jpx on Xi. Jpx is one of the 6 ncRNAs encoded at the Xic that was recently shown to be required for XCI [100]. Specifically, Jpx was found to derive mainly from the Xi and its expression to correlate with Xist transcription. Importantly, deleting Jpx in mouse blocked XCI and was female lethal. This phenotype was rescued by truncating Tsix suggesting that antagonistic lncRNA switches control XCI. Although it can work in *trans*, Jpx was found to have a mild preference for *cis* activation. Interestingly, previous 3C analysis of the Xic revealed higher-order chromatin contacts between the Jpx gene 5' end and the Xist gene [101]. This result suggests that similarly to

HOTTIP, three-dimensional chromatin organization might participate in preferentially targeting the Jpx lncRNA to its *cis*-located Xist allele. This study also reinforces an important principle in lncRNA function: that lncRNAs can regulate the expression of other lncRNAs. As such, it seems very likely that this type of regulatory network will play a role in the regulation of protein-coding genes.

ncRNAs AS BASIC TRANSCRIPTIONAL FEATURES

Ascribing a function to long ncRNAs that are capped, spliced, polyadenylated and tissue-specific is intuitively easier to accept than to do the same for low abundance, short, and apparently unprocessed transcripts. In the past, short ncRNAs were often viewed as likely degradation products or as transcriptional "noise", and were routinely discarded from datasets. Several groups have now taken a closer look at these RNAs and show that their non-random distribution is incompatible with aberrant transcription or with the truncated fragments of highly expressed genes. Short ncRNAs tend to be derived specifically from enhancer and promoter regions, and their expression levels, although very low, usually rises with increasing transcription (Fig. (**4**)). Regardless of whether these transcripts are simple manifestations of transcription activity or have actual regulatory functions, we can certainly learn much about transcription by taking a closer look at them.

Enhancer-Associated RNAs (eRNAs)

Enhancer sequences are scattered throughout mammalian genomes and are often activated in a tissue-specific manner. Enhancers can be identified by their association with the p300/CBP transcriptional co-activator, by the presence of histone 3 lysine 4 monomethylation marks (H3K4me1), and by their separate position to known transcription start sites (TSSs) [102]. Greenberg and colleagues recently used these enhancer traits to identify thousands of regulated enhancers during cortical neuron stimulation [103]. The study reports that most regulated p300/CBP binding sites are at least 1 kb away from known TSSs. Thus, transcription-dependent p300/CBP binding mainly occurs at enhancers rather than promoters in this system. Stimulation did not significantly change the histone modification landscape suggesting that chromatin is likely maintained permissive for induction in these cells. However, many transcription factors (TFs) were found to co-localize at enhancers, thereby raising the possibility that TFs might regulate enhancer function partly by recruiting p300/CBP. Strikingly, roughly 25% of enhancer-bound p300/CBP was found to bind RNA pol II in a transcription-dependent manner. Furthermore, short RNAs (< 2kb) transcribed bi-directionally from the center of enhancer domains could be detected from these sites. These transcripts were termed eRNAs for "enhancer RNAs" and appear to be non-polyadenylated. An interesting feature about eRNA transcription is that their levels correlate with the activity of nearby protein coding genes during neuronal stimulation. As such, enhancers may not strictly be recruitment sites of TFs, p300/CBP and RNA pol II but also sites of transcription. Changes in eRNA expression at actively transcribed enhancers were strongly correlated with regulated mRNA levels at nearby genes, and suggested that eRNA synthesis may only occur when enhancers interact
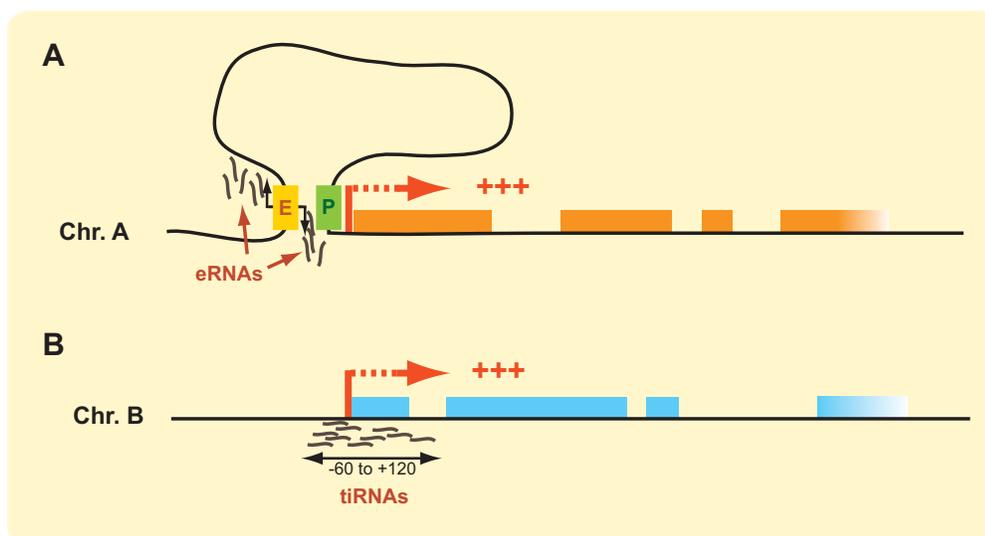
Fig. (4). *ncRNAs as basic transcriptional features.* (**A**) Bi-directional short ncRNAs produced from the center of enhancer sequences can be detected in cortical neurons. These "enhancer RNAs" (eRNAs) are only found near active genes and require RNA pol II. eRNA production would presumably require contact between the enhancer and transcribed gene. (**B**) Short (18 nt) transcription initiation RNAs (tiRNAs) can be detected at active genes. tiRNAs preferentially map around TSSs of highly transcribed genes and sites of RNA pol II binding. Protein-coding genes are represented by orange or blue boxes. Active transcription start sites (TSSs) are indicated thick red bent arrows. Enhancer sequence (E) is shown as a yellow box and promoter (P) as a green rectangle. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper).

with promoters. This hypothesis was confirmed by the absence of eRNA production from the active *Arc* gene enhancer in cells lacking the *Arc* promoter and most of the gene. Whether the transcriptional process itself or actual eRNA transcripts play a role in enhancer activation by modifying the chromatin state remains unknown. Nonetheless, this study not only clearly demonstrates that enhancers can be transcribed but also indicates that long-range looping rather than transcription-dependent RNA pol II tracking is a likely more widespread mechanism by which distal enhancers regulate target genes.

Transcription Initiation RNAs (tiRNAs)

Short ncRNAs mapping nearby gene promoters have been reported in several species over the past five years [104-106]. These studies together led to the hypothesis that genomes might encode multifunctional sequences that sometimes function as regulatory DNA elements and sometimes generate multiple independently regulated transcripts [107]. In 2009, the Mattick group examined small RNA deep-sequencing libraries from human, chicken and *Drosophila* to further characterize TSS-associated transcripts [108]. This study uncovered a new class of short ncRNAs mapping within 60 upstream to 120 bp downstream of annotated TSSs that were named tiRNAs for "transcription initiation RNAs". The 5' ends of tiRNAs peaked at 10 to 30 nts downstream from the TSS of their parent genes indicating that they are not simply RNA pol II runoff or truncated 5' capped ends. They were found in all 3 species, and although slightly different in lengths, most were shorter than 22 nt with over a quarter measuring 18 nucleotides. tiRNAs were usually transcribed from the same strand as their corresponding TSS and were preferentially derived from GC-rich promoters. Consistently, promoters with

tiRNAs were enriched in RNA pol II and binding of the Sp1 transcription factor. Curiously, tiRNAs were more frequently associated with highly expressed genes but their abundance did not directly correlate with gene expression levels. Furthermore, tiRNA synthesis was independent from the RNAi processing machinery indicating that they are likely derived directly from the RNA pol II activity at promoters. At least in *Drosophila*, tiRNA expression was only weakly associated with stalled transcripts produced at poised promoters. The authors therefore suggested that tiRNAs might represent products of RNA pol II backtracking, a process known to require the elongation and transcription cleavage factor TFIIS [109]. In a more recent paper, the same group reported a 40-fold nuclear enrichment of tiRNAs and demonstrated their association with transcription initiation marks [110]. Interestingly, tiRNAs were found to possess a 3'-guanine nucleotide bias in all 3 species, which further supports their non-random origin. tiRNA production is therefore a basic transcriptional feature in metazoans that may function in regulating chromatin modifications to fine-tune transcription.

lncRNAs AS REGULATORS OF SPATIAL GENOME ORGANIZATION

The histone-modifying activity of lncRNPs points to their pivotal role in regulating spatial genome organization. Although lncRNAs have thus far been predominantly associated with repressive complexes that methylate H3K27 or H3K9, it seems safe to predict that many more enzymatic activities will likely associate with these transcripts. Post-translational histone modifications may affect three-dimensional chromatin structure by altering the binding of nucleosomes to the DNA, the contacts between nucleosomes, or by changing the recruitment of non-histone
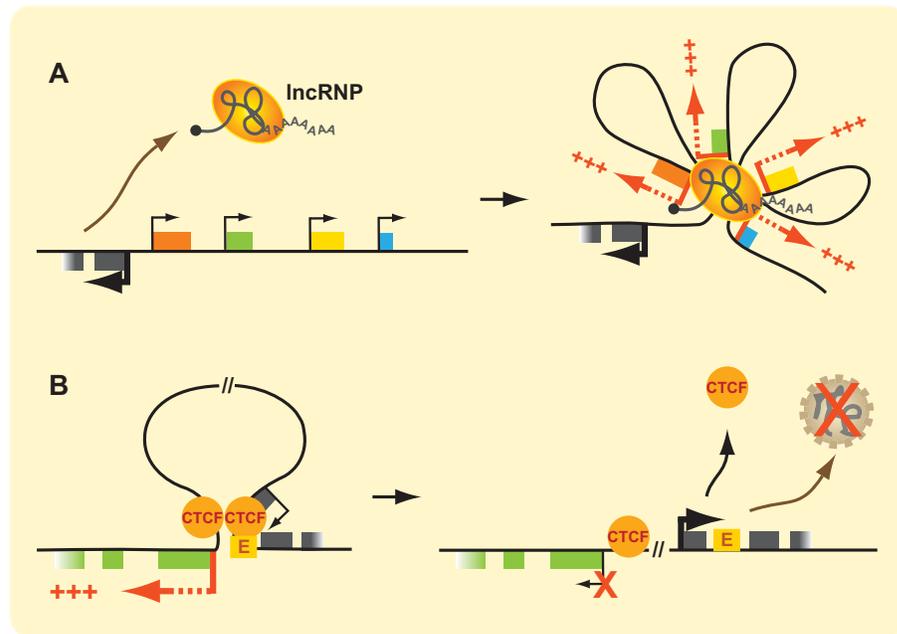
Fig. (5). *lncRNAs as regulators of spatial genome organization.* (**A**) A lncRNP produced in *cis* or in *trans* could drive the formation of activating contacts that bridge enhancers to promoters, or induce the formation of a transcription factory. This example illustrates how a *cis* lncRNA could lead to the coordinated activation of gene clusters. (**B**) Activating long-range looping contacts between enhancers and promoters could be disrupted by transcription of a lncRNA. In this example, lncRNA transcription itself rather than the transcript could be responsible for transcription inhibition of neighboring genes. This example shows one CTCF protein displaced by transcription and disruption of an activating loop. Protein-coding genes are represented by colored boxes. Transcription start sites (TSSs) are indicated by bent arrows, with thick red bent arrows indicating active protein coding genes and thick black arrows active lncRNA genes. TSS of inhibited gene is indicated by a thin bent arrow marked with an "X". (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper).

proteins. For example, changes in histone charge introduced by acetylation or phosphorylation might significantly affect *in vivo* nucleosome interactions and chromatin compaction levels. Many protein complexes recognizing specific PTMs have now been identified and usually include enzymatic activities that further modify chromatin. Enzymatic activities include HDACs, HATs, HDMTs, HMTs, ubiquitin ligases, and chromatin remodelers. How histone PTMs can tether histone-modifying complexes has been extensively described elsewhere ([5] and references therein).

At least in chicken and *Drosophila*, there is also evidence that ncRNAs themselves contribute to higher-order genome organization [42]. The Azorin group showed that polyA- RNA associates with chromatin where it represents approximately 2 to 5% of total nucleic acids. When they treated chromatin with RNase A, fragments sedimented less rapidly on sucrose gradients and were more sensitive to micrococcal nuclease. These structural changes were not correlated with any significant alteration in histone composition and were observed at euchromatin and heterochromatin. It was suggested that RNA might stabilize the binding of non-histone proteins to chromatin as was previously shown for HP1 at pericentric heterochromatin [111, 112]. A similar type of mechanism was also recently reported at repressed PcG target genes where short transcripts produced around TSSs can bind PRC2 complexes and might strengthen their association with the genes [113].

Specific ncRNAs may also nucleate the formation of functional nuclear subdomains or be required to maintain their integrity. The ncRNA NEAT1 (Nuclear Enriched Abundant Transcripts 1) is a very good example for this type of activity [114]. NEAT1 is a ubiquitously expressed nuclear lncRNA of 3.7 kb that is unspliced but polyadenylated [115]. It is highly expressed and localizes mainly to a type of membraneless nuclear body termed "paraspeckle". The function of paraspeckles is not well defined but thought to represent nuclear retention sites of adenosine-to-inosine (A-I) edited transcripts [116]. NEAT1 knockdown was shown to disrupt paraspeckles and its overexpression to result in a larger number of them [117]. Another lncRNA termed NEAT2 for "Nuclear Enriched Abundant Transcript 2" was found to preferentially localize to splicing speckles, which is another type of nuclear subcompartment enriched in splicing factors [115]. NEAT2 is a 8.7 kb unspliced, polyadenylated transcript with a similar expression pattern to NEAT1 that represent the bulk of polyA+ RNA in speckles. Although NEAT2 does not appear to play a structural role in these bodies as detected by epifluorescence [117], there is evidence suggesting that it may regulate mRNA alternative splicing by altering the ratio of phosphorylated and unphosphorylated SR splicing factor inside the speckles [118].

In addition to their involvement in the function of nuclear subdomains, ncRNPs could also control higher-order chromatin architecture and transcription activity by mediating long-range contacts. There are many ways by which lncRNAs could affect gene expression through long-range interactions. For example, lncRNAs might cluster to

form nuclear organization centers that coordinate gene expression. Good candidates for this type of activity are PcG-bound lncRNAs since polycomb complexes are known to form long-range *cis* and *trans* interactions genome-wide [119]. In this case, physical lncRNP networks would facilitate the coordinated down-regulation of multiple genes. Given the increasing complexity of RNA sequences found to interact with PRC2, it will be interesting to see which lncRNAs may be involved in this type of distal regulation. lncRNP-dependent physical networks could also be important for the formation of transcription factories (Fig. (**5A**)). Alternatively, lncRNA transcription might enhance or inhibit transcription of neighboring genes by displacing proteins from the chromatin and altering physical loops (Fig. (**5B**)). In this case, it is the transcription process itself rather than the transcripts that would mediate changes in gene expression. The lncRNAs genes themselves may also regulate chromatin architecture. For instance, lncRNA genes might cluster with other genes to promote the formation of transcription factories and gene expression. Additionally, lncRNA genes could compete with other genes (protein-coding or otherwise) for the same enhancers or other type of regulatory DNA element.

CONCLUSION

The discovery that lncRNAs represent a very large portion of our transcriptome has uncovered a new level of gene expression regulation. Although we know very little about them, their sequence diversity, potential splice variants, and cellular distribution suggests an important role in many aspects of RNA transcription, processing, and metabolism. The tissue-specificity and low conservation of lncRNAs amongst higher eukaryotes also points to their potential contribution in establishing or maintaining cell identity. Because of the overwhelming lncRNA complexity, recent studies have mainly focused on the function of individual transcripts in the nucleus. These studies clearly demonstrate that lncRNAs can regulate genes by various epigenetic mechanisms involving chromatin-modifying complexes. An interesting possibility is that lncRNAs also regulate transcription by altering chromatin architecture. Such spatial transcription regulation mechanisms could play an important role in human health and disease and it will be interesting to see how lncRNA function intersects with spatial chromatin architecture.

ABBREVIATIONS

3C	=	Chromosome conformation capture
rRNAs	=	Ribosomal RNAs
tRNAs	=	Transfer RNAs
snRNAs	=	Small nuclear RNAs
snoRNAs	=	Small nucleolar RNAs
miRNAs	=	MicroRNAs
piRNAs	=	Piwi-interacting RNAs
eRNAs	=	Enhancer RNAs
tiRNAs	=	Transcription initiation RNAs
spliRNAs	=	Splice-site RNAs
lincRNAs	=	Long intergenic non-coding RNAs
lncRNPs	=	Long non-coding ribonucleoprotein complexes
Igf2r	=	Insulin-like growth factor II receptor
HMTs	=	Histone methyl transferases
TSSs	=	Transcriptional start sites
TFs	=	Transcription factors
RNAi	=	RNA interference
PTMs	=	Post-translational modifications

REFERENCES

[1] Cremer, T.; Cremer, M. Chromosome territories. *Cold Spring Harb. Perspect. Biol.*, **2010**, *2*, a003889.
[2] Pombo, A.; Branco, M.R. Functional organisation of the genome during interphase. *Curr. Opin. Genet. Dev.*, **2007**, *17*, 451-455.
[3] Towbin, B.D.; Meister, P.; Gasser, S.M. The nuclear envelope--a scaffold for silencing? *Curr. Opin. Genet. Dev.*, **2009**, *19*, 180-186.
[4] Li, B.; Carey, M.; Workman, J.L. The role of chromatin during transcription. *Cell*, **2007**, *128*, 707-719.
[5] Kouzarides, T. Chromatin modifications and their function. *Cell*, **2007**, *128*, 693-705.
[6] Tolhuis, B.; Palstra, R.J.; Splinter, E.; Grosveld, F.; de Laat, W. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell*, **2002**, *10*, 1453-1465.
[7] Spilianakis, C.G.; Lalioti, M.D.; Town, T.; Lee, G.R.; Flavell, R.A. Interchromosomal associations between alternatively expressed loci. *Nature*, **2005**, *435*, 637-645.
[8] Gheldof, N.; Smith, E.M.; Tabuchi, T.M.; Koch, C.M.; Dunham, I.; Stamatoyannopoulos, J.A.; Dekker, J. Cell-type-specific long-range looping interactions identify distant regulatory elements of the CFTR gene. *Nucleic Acids Res.*, **2010**, *38*, 4325-4336.
[9] Chavanas, S.; Adoue, V.; Mechin, M.C.; Ying, S.; Dong, S.; Duplan, H.; Charveron, M.; Takahara, H.; Serre, G.; Simon, M. Long-range enhancer associated with chromatin looping allows AP-1 regulation of the peptidylarginine deiminase 3 gene in differentiated keratinocyte. *PLoS ONE*, **2008**, *3*, e3408.
[10] Bugge, A.; Siersbaek, M.; Madsen, M.S.; Gondor, A.; Rougier, C.; Mandrup, S. A novel intronic peroxisome proliferator-activated receptor gamma enhancer in the uncoupling protein (UCP) 3 gene as a regulator of both UCP2 and -3 expression in adipocytes. *J. Biol. Chem.*, **2010**, *285*, 17310-17317.
[11] Amano, T.; Sagai, T.; Tanabe, H.; Mizushina, Y.; Nakazawa, H.; Shiroishi, T. Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. *Dev. Cell*, **2009**, *16*, 47-57.
[12] Lomvardas, S.; Barnea, G.; Pisapia, D.J.; Mendelsohn, M.; Kirkland, J.; Axel, R. Interchromosomal interactions and olfactory receptor choice. *Cell*, **2006**, *126*, 403-413.
[13] Dekker, J.; Rippe, K.; Dekker, M.; Kleckner, N. Capturing Chromosome Conformation. *Science*, **2002**, *295*, 1306-1311.
[14] Dostie, J.; Richmond, T.A.; Arnaout, R.A.; Selzer, R.R.; Lee, W.L.; Honan, T.A.; Rubio, E.D.; Krumm, A.; Lamb, J.; Nusbaum,

C.; Green, R.D.; Dekker, J. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.*, **2006,** *16,* 1299-1309.

[15] Dostie, J.; Zhan, Y.; Dekker, J. Chromosome conformation capture carbon copy technology. *Curr. Protoc. Mol. Biol.*, **2007,** *Chapter 21,* Unit 21 14.

[16] Lieberman-Aiden, E.; van Berkum, N.L.; Williams, L.; Imakaev, M.; Ragoczy, T.; Telling, A.; Amit, I.; Lajoie, B.R.; Sabo, P.J.; Dorschner, M.O.; Sandstrom, R.; Bernstein, B.; Bender, M.A.; Groudine, M.; Gnirke, A.; Stamatoyannopoulos, J.; Mirny, L.A.; Lander, E.S.; Dekker, J. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science,* **2009,** *326,* 289-293.

[17] Crutchley, J.L.; Wang, X.Q.; Ferraiuolo, M.A.; Dostie, J. Chromatin conformation signatures: ideal human disease biomarkers? *Biomark. Med.*, **2010,** *4,* 611-629.

[18] Vakoc, C.R.; Letting, D.L.; Gheldof, N.; Sawado, T.; Bender, M.A.; Groudine, M.; Weiss, M.J.; Dekker, J.; Blobel, G.A. Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol. Cell,* **2005,** *17,* 453-462.

[19] Fullwood, M.J.; Liu, M.H.; Pan, Y.F.; Liu, J.; Xu, H.; Mohamed, Y.B.; Orlov, Y.L.; Velkov, S.; Ho, A.; Mei, P.H.; Chew, E.G.Y.; Huang, P.Y.H.; Welboren, W.-J.; Han, Y.; Ooi, H.S.; Ariyaratne, P.N.; Vega, V.B.; Luo, Y.; Tan, P.Y.; Choy, P.Y.; Wansa, K.D.S.A.; Zhao, B.; Lim, K.S.; Leow, S.C.; Yow, J.S.; Joseph, R.; Li, H.; Desai, K.V.; Thomsen, J.S.; Lee, Y.K.; Karuturi, R.K.M.; Herve, T.; Bourque, G.; Stunnenberg, H.G.; Ruan, X.; Cacheux-Rataboul, V.; Sung, W.-K.; Liu, E.T.; Wei, C.-L.; Cheung, E.; Ruan, Y. An oestrogen-receptor-[agr]-bound human chromatin interactome. *Nature,* **2009,** *462,* 58-64.

[20] Yoon, H.; Boss, J.M. PU.1 binds to a distal regulatory element that is necessary for B cell-specific expression of CIITA. *J. Immunol.*, **2010,** *184,* 5018-5028.

[21] Yochum, G.S.; Sherrick, C.M.; Macpartlin, M.; Goodman, R.H. A beta-catenin/TCF-coordinated chromatin loop at MYC integrates 5' and 3' Wnt responsive enhancers. *Proc. Natl. Acad. Sci. U. S. A.*, **2009,** *107,* 145-150.

[22] Phillips, J.E.; Corces, V.G. CTCF: master weaver of the genome. *Cell,* **2009,** *137,* 1194-1211.

[23] Merkenschlager, M. Cohesin: a global player in chromosome biology with local ties to gene regulation. *Curr. Opin. Genet. Dev.*, **2010,** *20,* 555-561.

[24] Rubio, E.D.; Reiss, D.J.; Welcsh, P.L.; Disteche, C.M.; Filippova, G.N.; Baliga, N.S.; Aebersold, R.; Ranish, J.A.; Krumm, A. CTCF physically links cohesin to chromatin. *Proc. Natl. Acad. Sci. U. S. A.*, **2008,** *105,* 8309-8314.

[25] Hadjur, S.; Williams, L.M.; Ryan, N.K.; Cobb, B.S.; Sexton, T.; Fraser, P.; Fisher, A.G.; Merkenschlager, M. Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature,* **2009,** *460,* 410-413.

[26] Hou, C.; Dale, R.; Dean, A. Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proc. Natl. Acad. Sci. U. S. A.*, **2010,** *107,* 3651-3656.

[27] Mishiro, T.; Ishihara, K.; Hino, S.; Tsutsumi, S.; Aburatani, H.; Shirahige, K.; Kinoshita, Y.; Nakao, M. Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. *EMBO J.*, **2009,** *28,* 1234-1245.

[28] Schmidt, D.; Schwalie, P.C.; Ross-Innes, C.S.; Hurtado, A.; Brown, G.D.; Carroll, J.S.; Flicek, P.; Odom, D.T. A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res.*, **2010,** *20,* 578-588.

[29] Kagey, M.H.; Newman, J.J.; Bilodeau, S.; Zhan, Y.; Orlando, D.A.; van Berkum, N.L.; Ebmeier, C.C.; Goossens, J.; Rahl, P.B.; Levine, S.S.; Taatjes, D.J.; Dekker, J.; Young, R.A. Mediator and cohesin connect gene expression and chromatin architecture. *Nature,* **2010,** *467,* 430-435.

[30] Nativio, R.; Wendt, K.S.; Ito, Y.; Huddleston, J.E.; Uribe-Lewis, S.; Woodfine, K.; Krueger, C.; Reik, W.; Peters, J.M.; Murrell, A. Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus. *PLoS Genet.*, **2009,** *5,* e1000739.

[31] Kurukuti, S.; Tiwari, V.K.; Tavoosidana, G.; Pugacheva, E.; Murrell, A.; Zhao, Z.; Lobanenkov, V.; Reik, W.; Ohlsson, R. CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proc. Natl. Acad. Sci. U. S. A.*, **2006,** *103,* 10684-10689.

[32] Li, T.; Hu, J.F.; Qiu, X.; Ling, J.; Chen, H.; Wang, S.; Hou, A.; Vu, T.H.; Hoffman, A.R. CTCF regulates allelic expression of Igf2 by orchestrating a promoter-polycomb repressive complex 2 intrachromosomal loop. *Mol. Cell. Biol.*, **2008,** *28,* 6473-6482.

[33] Kapranov, P.; Cawley, S.E.; Drenkow, J.; Bekiranov, S.; Strausberg, R.L.; Fodor, S.P.; Gingeras, T.R. Large-scale transcriptional activity in chromosomes 21 and 22. *Science,* **2002,** *296,* 916-919.

[34] Rinn, J.L.; Euskirchen, G.; Bertone, P.; Martone, R.; Luscombe, N.M.; Hartman, S.; Harrison, P.M.; Nelson, F.K.; Miller, P.; Gerstein, M.; Weissman, S.; Snyder, M. The transcriptional activity of human Chromosome 22. *Genes Dev.*, **2003,** *17,* 529-540.

[35] Bertone, P.; Stolc, V.; Royce, T.E.; Rozowsky, J.S.; Urban, A.E.; Zhu, X.; Rinn, J.L.; Tongprasit, W.; Samanta, M.; Weissman, S.; Gerstein, M.; Snyder, M. Global identification of human transcribed sequences with genome tiling arrays. *Science,* **2004,** *306,* 2242-2246.

[36] Cheng, J.; Kapranov, P.; Drenkow, J.; Dike, S.; Brubaker, S.; Patel, S.; Long, J.; Stern, D.; Tammana, H.; Helt, G.; Sementchenko, V.; Piccolboni, A.; Bekiranov, S.; Bailey, D.K.; Ganesh, M.; Ghosh, S.; Bell, I.; Gerhard, D.S.; Gingeras, T.R. Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science,* **2005,** *308,* 1149-1154.

[37] Carninci, P.; Kasukawa, T.; Katayama, S.; Gough, J.; Frith, M.C.; Maeda, N.; Oyama, R.; Ravasi, T.; Lenhard, B.; Wells, C.; Kodzius, R.; Shimokawa, K.; Bajic, V.B.; Brenner, S.E.; Batalov, S.; Forrest, A.R.; Zavolan, M.; Davis, M.J.; Wilming, L.G.; Aidinis, V.; Allen, J.E.; Ambesi-Impiombato, A.; Apweiler, R.; Aturaliya, R.N.; Bailey, T.L.; Bansal, M.; Baxter, L.; Beisel, K.W.; Bersano, T.; Bono, H.; Chalk, A.M.; Chiu, K.P.; Choudhary, V.; Christoffels, A.; Clutterbuck, D.R.; Crowe, M.L.; Dalla, E.; Dalrymple, B.P.; de Bono, B.; Della Gatta, G.; di Bernardo, D.; Down, T.; Engstrom, P.; Fagiolini, M.; Faulkner, G.; Fletcher, C.F.; Fukushima, T.; Furuno, M.; Futaki, S.; Gariboldi, M.; Georgii-Hemming, P.; Gingeras, T.R.; Gojobori, T.; Green, R.E.; Gustincich, S.; Harbers, M.; Hayashi, Y.; Hensch, T.K.; Hirokawa, N.; Hill, D.; Huminiecki, L.; Iacono, M.; Ikeo, K.; Iwama, A.; Ishikawa, T.; Jakt, M.; Kanapin, A.; Katoh, M.; Kawasawa, Y.; Kelso, J.; Kitamura, H.; Kitano, H.; Kollias, G.; Krishnan, S.P.; Kruger, A.; Kummerfeld, S.K.; Kurochkin, I.V.; Lareau, L.F.; Lazarevic, D.; Lipovich, L.; Liu, J.; Liuni, S.; McWilliam, S.; Madan Babu, M.; Madera, M.; Marchionni, L.; Matsuda, H.; Matsuzawa, S.; Miki, H.; Mignone, F.; Miyake, S.; Morris, K.; Mottagui-Tabar, S.; Mulder, N.; Nakano, N.; Nakauchi, H.; Ng, P.; Nilsson, R.; Nishiguchi, S.; Nishikawa, S.; Nori, F.; Ohara, O.; Okazaki, Y.; Orlando, V.; Pang, K.C.; Pavan, W.J.; Pavesi, G.; Pesole, G.; Petrovsky, N.; Piazza, S.; Reed, J.; Reid, J.F.; Ring, B.Z.; Ringwald, M.; Rost, B.; Ruan, Y.; Salzberg, S.L.; Sandelin, A.; Schneider, C.; Schonbach, C.; Sekiguchi, K.; Semple, C.A.; Seno, S.; Sessa, L.; Sheng, Y.; Shibata, Y.; Shimada, H.; Shimada, K.; Silva, D.; Sinclair, B.; Sperling, S.; Stupka, E.; Sugiura, K.; Sultana, R.; Takenaka, Y.; Taki, K.; Tammoja, K.; Tan, S.L.; Tang, S.; Taylor, M.S.; Tegner, J.; Teichmann, S.A.; Ueda, H.R.; van Nimwegen, E.; Verardo, R.; Wei, C.L.; Yagi, K.; Yamanishi, H.; Zabarovsky, E.; Zhu, S.; Zimmer, A.; Hide, W.; Bult, C.; Grimmond, S.M.; Teasdale, R.D.; Liu, E.T.; Brusic, V.; Quackenbush, J.; Wahlestedt, C.; Mattick, J.S.; Hume, D.A.; Kai, C.; Sasaki, D.; Tomaru, Y.; Fukuda, S.; Kanamori-Katayama, M.; Suzuki, M.; Aoki, J.; Arakawa, T.; Iida, J.; Imamura, K.; Itoh, M.; Kato, T.; Kawaji, H.; Kawagashira, N.; Kawashima, T.; Kojima, M.; Kondo, S.; Konno, H.; Nakano, K.; Ninomiya, N.; Nishio, T.; Okada, M.; Plessy, C.; Shibata, K.; Shiraki, T.; Suzuki, S.; Tagami, M.; Waki, K.; Watahiki, A.; Okamura-Oho, Y.; Suzuki, H.; Kawai, J.; Hayashizaki, Y. The transcriptional landscape of the mammalian genome. *Science,* **2005,** *309,* 1559-1563.

[38] Katayama, S.; Tomaru, Y.; Kasukawa, T.; Waki, K.; Nakanishi, M.; Nakamura, M.; Nishida, H.; Yap, C.C.; Suzuki, M.; Kawai, J.; Suzuki, H.; Carninci, P.; Hayashizaki, Y.; Wells, C.; Frith, M.; Ravasi, T.; Pang, K.C.; Hallinan, J.; Mattick, J.; Hume, D.A.; Lipovich, L.; Batalov, S.; Engstrom, P.G.; Mizuno, Y.; Faghihi, M.A.; Sandelin, A.; Chalk, A.M.; Mottagui-Tabar, S.; Liang, Z.; Lenhard, B.; Wahlestedt, C. Antisense transcription in the mammalian transcriptome. *Science,* **2005,** *309,* 1564-1566.

[39]   Birney, E.; Stamatoyannopoulos, J.A.; Dutta, A.; Guigo, R.; Gingeras, T.R.; Margulies, E.H.; Weng, Z.; Snyder, M.; Dermitzakis, E.T.; Thurman, R.E.; Kuehn, M.S.; Taylor, C.M.; Neph, S.; Koch, C.M.; Asthana, S.; Malhotra, A.; Adzhubei, I.; Greenbaum, J.A.; Andrews, R.M.; Flicek, P.; Boyle, P.J.; Cao, H.; Carter, N.P.; Clelland, G.K.; Davis, S.; Day, N.; Dhami, P.; Dillon, S.C.; Dorschner, M.O.; Fiegler, H.; Giresi, P.G.; Goldy, J.; Hawrylycz, M.; Haydock, A.; Humbert, R.; James, K.D.; Johnson, B.E.; Johnson, E.M.; Frum, T.T.; Rosenzweig, E.R.; Karnani, N.; Lee, K.; Lefebvre, G.C.; Navas, P.A.; Neri, F.; Parker, S.C.; Sabo, P.J.; Sandstrom, R.; Shafer, A.; Vetrie, D.; Weaver, M.; Wilcox, S.; Yu, M.; Collins, F.S.; Dekker, J.; Lieb, J.D.; Tullius, T.D.; Crawford, G.E.; Sunyaev, S.; Noble, W.S.; Dunham, I.; Denoeud, F.; Reymond, A.; Kapranov, P.; Rozowsky, J.; Zheng, D.; Castelo, R.; Frankish, A.; Harrow, J.; Ghosh, S.; Sandelin, A.; Hofacker, I.L.; Baertsch, R.; Keefe, D.; Dike, S.; Cheng, J.; Hirsch, H.A.; Sekinger, E.A.; Lagarde, J.; Abril, J.F.; Shahab, A.; Flamm, C.; Fried, C.; Hackermuller, J.; Hertel, J.; Lindemeyer, M.; Missal, K.; Tanzer, A.; Washietl, S.; Korbel, J.; Emanuelsson, O.; Pedersen, J.S.; Holroyd, N.; Taylor, R.; Swarbreck, D.; Matthews, N.; Dickson, M.C.; Thomas, D.J.; Weirauch, M.T.; Gilbert, J.; Drenkow, J.; Bell, I.; Zhao, X.; Srinivasan, K.G.; Sung, W.K.; Ooi, H.S.; Chiu, K.P.; Foissac, S.; Alioto, T.; Brent, M.; Pachter, L.; Tress, M.L.; Valencia, A.; Choo, S.W.; Choo, C.Y.; Ucla, C.; Manzano, C.; Wyss, C.; Cheung, E.; Clark, T.G.; Brown, J.B.; Ganesh, M.; Patel, S.; Tammana, H.; Chrast, J.; Henrichsen, C.N.; Kai, C.; Kawai, J.; Nagalakshmi, U.; Wu, J.; Lian, Z.; Lian, J.; Newburger, P.; Zhang, X.; Bickel, P.; Mattick, J.S.; Carninci, P.; Hayashizaki, Y.; Weissman, S.; Hubbard, T.; Myers, R.M.; Rogers, J.; Stadler, P.F.; Lowe, T.M.; Wei, C.L.; Ruan, Y.; Struhl, K.; Gerstein, M.; Antonarakis, S.E.; Fu, Y.; Green, E.D.; Karaoz, U.; Siepel, A.; Taylor, J.; Liefer, L.A.; Wetterstrand, K.A.; Good, P.J.; Feingold, E.A.; Guyer, M.S.; Cooper, G.M.; Asimenos, G.; Dewey, C.N.; Hou, M.; Nikolaev, S.; Montoya-Burgos, J.I.; Loytynoja, A.; Whelan, S.; Pardi, F.; Massingham, T.; Huang, H.; Zhang, N.R.; Holmes, I.; Mullikin, J.C.; Ureta-Vidal, A.; Paten, B.; Seringhaus, M.; Church, D.; Rosenbloom, K.; Kent, W.J.; Stone, E.A.; Batzoglou, S.; Goldman, N.; Hardison, R.C.; Haussler, D.; Miller, W.; Sidow, A.; Trinklein, N.D.; Zhang, Z.D.; Barrera, L.; Stuart, R.; King, D.C.; Ameur, A.; Enroth, S.; Bieda, M.C.; Kim, J.; Bhinge, A.A.; Jiang, N.; Liu, J.; Yao, F.; Vega, V.B.; Lee, C.W.; Ng, P.; Yang, A.; Moqtaderi, Z.; Zhu, Z.; Xu, X.; Squazzo, S.; Oberley, M.J.; Inman, D.; Singer, M.A.; Richmond, T.A.; Munn, K.J.; Rada-Iglesias, A.; Wallerman, O.; Komorowski, J.; Fowler, J.C.; Couttet, P.; Bruce, A.W.; Dovey, O.M.; Ellis, P.D.; Langford, C.F.; Nix, D.A.; Euskirchen, G.; Hartman, S.; Urban, A.E.; Kraus, P.; Van Calcar, S.; Heintzman, N.; Kim, T.H.; Wang, K.; Qu, C.; Hon, G.; Luna, R.; Glass, C.K.; Rosenfeld, M.G.; Aldred, S.F.; Cooper, S.J.; Halees, A.; Lin, J.M.; Shulha, H.P.; Xu, M.; Haidar, J.N.; Yu, Y.; Iyer, V.R.; Green, R.D.; Wadelius, C.; Farnham, P.J.; Ren, B.; Harte, R.A.; Hinrichs, A.S.; Trumbower, H.; Clawson, H.; Hillman-Jackson, J.; Zweig, A.S.; Smith, K.; Thakkapallayil, A.; Barber, G.; Kuhn, R.M.; Karolchik, D.; Armengol, L.; Bird, C.P.; de Bakker, P.I.; Kern, A.D.; Lopez-Bigas, N.; Martin, J.D.; Stranger, B.E.; Woodroffe, A.; Davydov, E.; Dimas, A.; Eyras, E.; Hallgrimsdottir, I.B.; Huppert, J.; Zody, M.C.; Abecasis, G.R.; Estivill, X.; Bouffard, G.G.; Guan, X.; Hansen, N.F.; Idol, J.R.; Maduro, V.V.; Maskeri, B.; McDowell, J.C.; Park, M.; Thomas, P.J.; Young, A.C.; Blakesley, R.W.; Muzny, D.M.; Sodergren, E.; Wheeler, D.A.; Worley, K.C.; Jiang, H.; Weinstock, G.M.; Gibbs, R.A.; Graves, T.; Fulton, R.; Mardis, E.R.; Wilson, R.K.; Clamp, M.; Cuff, J.; Gnerre, S.; Jaffe, D.B.; Chang, J.L.; Lindblad-Toh, K.; Lander, E.S.; Koriabine, M.; Nefedov, M.; Osoegawa, K.; Yoshinaga, Y.; Zhu, B.; de Jong, P.J. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, **2007**, *447*, 799-816.

[40]   He, Y.; Vogelstein, B.; Velculescu, V.E.; Papadopoulos, N.; Kinzler, K.W. The antisense transcriptomes of human cells. *Science*, **2008**, *322*, 1855-1857.

[41]   Clark, M.B.; Mattick, J.S. Long noncoding RNAs in cell biology. *Semin. Cell Dev. Biol.*, **2011**.

[42]   Rodriguez-Campos, A.; Azorin, F. RNA is an integral component of chromatin that contributes to its structural organization. *PLoS ONE*, **2007**, *2*, e1182.

[43]   Chen, L.L.; Carmichael, G.G. Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells:

functional role of a nuclear noncoding RNA. *Mol. Cell*, **2009**, *35*, 467-478.

[44]   Wang, K.C.; Yang, Y.W.; Liu, B.; Sanyal, A.; Corces-Zimmerman, R.; Chen, Y.; Lajoie, B.R.; Protacio, A.; Flynn, R.A.; Gupta, R.A.; Wysocka, J.; Lei, M.; Dekker, J.; Helms, J.A.; Chang, H.Y. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature*, **2011**, *472*, 120-124.

[45]   Brown, C.J.; Ballabio, A.; Rupert, J.L.; Lafreniere, R.G.; Grompe, M.; Tonlorenzi, R.; Willard, H.F. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature*, **1991**, *349*, 38-44.

[46]   Brown, C.J.; Lafreniere, R.G.; Powers, V.E.; Sebastio, G.; Ballabio, A.; Pettigrew, A.L.; Ledbetter, D.H.; Levy, E.; Craig, I.W.; Willard, H.F. Localization of the X inactivation centre on the human X chromosome in Xq13. *Nature*, **1991**, *349*, 82-84.

[47]   Lee, J.T. The X as model for RNA's niche in epigenomic regulation. *Cold Spring Harb. Perspect. Biol.*, **2010**, *2*, a003749.

[48]   Rinn, J.L.; Kertesz, M.; Wang, J.K.; Squazzo, S.L.; Xu, X.; Brugmann, S.A.; Goodnough, L.H.; Helms, J.A.; Farnham, P.J.; Segal, E.; Chang, H.Y. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*, **2007**, *129*, 1311-1323.

[49]   McGinnis, W.; Garber, R.L.; Wirz, J.; Kuroiwa, A.; Gehring, W.J. A homologous protein-coding sequence in drosophila homeotic genes and its conservation in other metazoans. *Cell*, **1984**, *37*, 403-408.

[50]   Dolle, P.; Izpisua-Belmonte, J.-C.; Falkenstein, H.; Renucci, A.; Duboule, D. Coordinate expression of the murine Hox-5 complex homoeobox-containing genes during limb pattern formation. *Nature*, **1989**, *342*, 767-772.

[51]   Wellik, D.M.; Capecchi, M.R. Hox10 and Hox11 Genes Are Required to Globally Pattern the Mammalian Skeleton. *Science*, **2003**, *301*, 363-367.

[52]   Starkova, J.; Zamostna, B.; Mejstrikova, E.; Krejci, R.; Drabkin, H.A.; Trka, J. HOX gene expression in phenotypic and genotypic subgroups and low HOXA gene expression as an adverse prognostic factor in pediatric ALL. *Pediatr. Blood Cancer*, **2010**, *55*, 1072-1082.

[53]   Wheadon, H.; Ramsey, J.M.; Dobbin, E.; Dickson, G.J.; Corrigan, P.M.; Freeburn, R.W.; Thompson, A. Differential Hox Expression in Murine Embryonic Stem Cell Models of Normal and Malignant Hematopoiesis. *Stem Cells Dev.*, **2011**, *0*, null.

[54]   Bernstein, B.E.; Mikkelsen, T.S.; Xie, X.; Kamal, M.; Huebert, D.J.; Cuff, J.; Fry, B.; Meissner, A.; Wernig, M.; Plath, K.; Jaenisch, R.; Wagschal, A.; Feil, R.; Schreiber, S.L.; Lander, E.S. A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell*, **2006**, *125*, 315-326.

[55]   Barber, B.A.; Rastegar, M. Epigenetic control of Hox genes during neurogenesis, development, and disease. *Annals of Anatomy - Anatomischer Anzeiger*, **2010**, *192*, 261-274.

[56]   Tsai, M.-C.; Manor, O.; Wan, Y.; Mosammaparast, N.; Wang, J.K.; Lan, F.; Shi, Y.; Segal, E.; Chang, H.Y. Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes. *Science*, **2010**, *329*, 689-693.

[57]   Gupta, R.A.; Shah, N.; Wang, K.C.; Kim, J.; Horlings, H.M.; Wong, D.J.; Tsai, M.-C.; Hung, T.; Argani, P.; Rinn, J.L.; Wang, Y.; Brzoska, P.; Kong, B.; Li, R.; West, R.B.; van de Vijver, M.J.; Sukumar, S.; Chang, H.Y. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*, **2010**, *464*, 1071-1076.

[58]   Gil, J.; Peters, G. Regulation of the INK4b-ARF-INK4a tumour suppressor locus: all for one or one for all. *Nat. Rev. Mol. Cell Biol.*, **2006**, *7*, 667-677.

[59]   Pasmant, E.; Laurendeau, I.; Héron, D.; Vidaud, M.; Vidaud, D.; Bièche, I. Characterization of a Germ-Line Deletion, Including the Entire INK4/ARF Locus, in a Melanoma-Neural System Tumor Family: Identification of ANRIL, an Antisense Noncoding RNA Whose Expression Coclusters with ARF. *Cancer Res.*, **2007**, *67*, 3963-3969.

[60]   Folkersen, L.; Kyriakou, T.; Goel, A.; Peden, J.; Mälarstig, A.; Paulsson-Berne, G.; Hamsten, A.; Franco-Cereceda, A.; Gabrielsen, A.; Eriksson, P.; Hugh Watkins on behalf of the, P.c. Relationship between CAD Risk Genotype in the Chromosome 9p21 Locus and Gene Expression. Identification of Eight New ANRIL Splice Variants. *PLoS ONE*, **2009**, *4*, e7677.

[61]   Kotake, Y.; Nakagawa, T.; Kitagawa, K.; Suzuki, S.; Liu, N.; Kitagawa, M.; Xiong, Y. Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15INK4B tumor suppressor gene. *Oncogene*, **2010**.

[62]   Yap, K.L.; Li, S.; Muñoz-Cabello, A.M.; Raguz, S.; Zeng, L.; Mujtaba, S.; Gil, J.; Walsh, M.J.; Zhou, M.-M. Molecular Interplay of the Noncoding RNA ANRIL and Methylated Histone H3 Lysine 27 by Polycomb CBX7 in Transcriptional Silencing of INK4a. *Mol. Cell*, **2010**, *38*, 662-674.

[63]   Cao, R.; Wang, L.; Wang, H.; Xia, L.; Erdjument-Bromage, H.; Tempst, P.; Jones, R.S.; Zhang, Y. Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. *Science*, **2002**, *298*, 1039-1043.

[64]   Cao, R.; Tsukada, Y.-i.; Zhang, Y. Role of Bmi-1 and Ring1A in H2A Ubiquitylation and Hox Gene Silencing. *Mol. Cell*, **2005**, *20*, 845-854.

[65]   de Napoles, M.; Mermoud, J.E.; Wakao, R.; Tang, Y.A.; Endoh, M.; Appanah, R.; Nesterova, T.B.; Silva, J.; Otte, A.P.; Vidal, M.; Koseki, H.; Brockdorff, N. Polycomb Group Proteins Ring1A/B Link Ubiquitylation of Histone H2A to Heritable Gene Silencing and X Inactivation. *Dev. Cell*, **2004**, *7*, 663-676.

[66]   Stock, J.K.; Giadrossi, S.; Casanova, M.; Brookes, E.; Vidal, M.; Koseki, H.; Brockdorff, N.; Fisher, A.G.; Pombo, A. Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat. Cell Biol.*, **2007**, *9*, 1428-1435.

[67]   Yu, W.; Gius, D.; Onyango, P.; Muldoon-Jacobs, K.; Karp, J.; Feinberg, A.P.; Cui, H. Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature*, **2008**, *451*, 202-206.

[68]   Monk, M. Genomic imprinting. *Genes Dev.*, **1988**, *2*, 921-925.

[69]   Wutz, A.; Smrzka, O.W.; Schweifer, N.; Schellander, K.; Wagner, E.F.; Barlow, D.P. Imprinted expression of the Igf2r gene depends on an intronic CpG island. *Nature*, **1997**, *389*, 745-749.

[70]   Sleutels, F.; Zwart, R.; Barlow, D.P. The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature*, **2002**, *415*, 810-813.

[71]   Pandey, R.R.; Mondal, T.; Mohammad, F.; Enroth, S.; Redrup, L.; Komorowski, J.; Nagano, T.; Mancini-DiNardo, D.; Kanduri, C. Kcnq1ot1 Antisense Noncoding RNA Mediates Lineage-Specific Transcriptional Silencing through Chromatin-Level Regulation. *Mol. Cell*, **2008**, *32*, 232-246.

[72]   Lyle, R.; Watanabe, D.; Vruchte, D.t.; Lerchner, W.; Smrzka, O.W.; Wutz, A.; Schageman, J.; Hahner, L.; Davies, C.; Barlow, D.P. The imprinted antisense RNA at the Igf2r locus overlaps but does not imprint Mas1. *Nat. Genet.*, **2000**, *25*, 19-21.

[73]   Zwart, R.; Sleutels, F.; Wutz, A.; Schinkel, A.H.; Barlow, D.P. Bidirectional action of the Igf2r imprint control element on upstream and downstream imprinted genes. *Genes Dev.*, **2001**, *15*, 2361-2366.

[74]   Seidl, C.I.M.; Stricker, S.H.; Barlow, D.P. The imprinted Air ncRNA is an atypical RNAPII transcript that evades splicing and escapes nuclear export. *EMBO J.*, **2006**, *25*, 3565-3575.

[75]   Nagano, T.; Mitchell, J.A.; Sanz, L.A.; Pauler, F.M.; Ferguson-Smith, A.C.; Feil, R.; Fraser, P. The Air Noncoding RNA Epigenetically Silences Transcription by Targeting G9a to Chromatin. *Science*, **2008**, *322*, 1717-1720.

[76]   Tachibana, M.; Sugimoto, K.; Fukushima, T.; Shinkai, Y. SET Domain-containing Protein, G9a, Is a Novel Lysine-preferring Mammalian Histone Methyltransferase with Hyperactivity and Specific Selectivity to Lysines 9 and 27 of Histone H3. *J. Biol. Chem.*, **2001**, *276*, 25309-25317.

[77]   Tachibana, M.; Ueda, J.; Fukuda, M.; Takeda, N.; Ohta, T.; Iwanari, H.; Sakihama, T.; Kodama, T.; Hamakubo, T.; Shinkai, Y. Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. *Genes Dev.*, **2005**, *19*, 815-826.

[78]   Rice, J.C.; Briggs, S.D.; Ueberheide, B.; Barber, C.M.; Shabanowitz, J.; Hunt, D.F.; Shinkai, Y.; Allis, C.D. Histone Methyltransferases Direct Different Degrees of Methylation to Define Distinct Chromatin Domains. *Mol. Cell*, **2003**, *12*, 1591-1598.

[79]   Mitsuya, K.; Meguro, M.; Lee, M.P.; Katoh, M.; Schulz, T.C.; Kugoh, H.; Yoshida, M.A.; Niikawa, N.; Feinberg, A.P.; Oshimura, M. LIT1, an Imprinted Antisense RNA in the Human KvLQT1 Locus Identified by Screening for Differentially Expressed Transcripts Using Monochromosomal Hybrids. *Hum. Mol. Genet.*, **1999**, *8*, 1209-1217.

[80]   Bliek, J.; Maas, S.M.; Ruijter, J.M.; Hennekam, R.C.M.; Alders, M.; Westerveld, A.; Mannens, M.M.A.M. Increased tumour risk for BWS patients correlates with aberrant H19 and not KCNQ1OT1 methylation: occurrence of KCNQ1OT1 hypomethylation in familial cases of BWS. *Hum. Mol. Genet.*, **2001**, *10*, 467-476.

[81]   Ohya, S.; Sergeant, G.P.; Greenwood, I.A.; Horowitz, B. Molecular Variants of KCNQ Channels Expressed in Murine Portal Vein Myocytes: A Role in Delayed Rectifier Current. *Circ. Res.*, **2003**, *92*, 1016-1023.

[82]   Thakur, N.; Tiwari, V.K.; Thomassin, H.; Pandey, R.R.; Kanduri, M.; Gondor, A.; Grange, T.; Ohlsson, R.; Kanduri, C. An Antisense RNA Regulates the Bidirectional Silencing Property of the Kcnq1 Imprinting Control Region. *Mol. Cell. Biol.*, **2004**, *24*, 7855-7862.

[83]   Kanduri, C.; Thakur, N.; Pandey, R.R. The length of the transcript encoded from the Kcnq1ot1 antisense promoter determines the degree of silencing. *EMBO J.*, **2006**, *25*, 2096-2106.

[84]   Mohammad, F.; Pandey, R.R.; Nagano, T.; Chakalova, L.; Mondal, T.; Fraser, P.; Kanduri, C. Kcnq1ot1/Lit1 Noncoding RNA Mediates Transcriptional Silencing by Targeting to the Perinucleolar Region. *Mol. Cell. Biol.*, **2008**, *28*, 3713-3728.

[85]   Lane, D.P. p53, guardian of the genome. *Nature*, **1992**, *358*, 15-16.

[86]   Vousden, K.H.; Lu, X. Live or let die: the cell's response to p53. *Nat. Rev. Cancer*, **2002**, *2*, 594-604.

[87]   Goh, A.M.; Coffill, C.R.; Lane, D.P. The role of mutant p53 in human cancer. *The Journal of Pathology*, **2011**, *223*, 116-126.

[88]   Huarte, M.; Guttman, M.; Feldser, D.; Garber, M.; Koziol, M.J.; Kenzelmann-Broz, D.; Khalil, A.M.; Zuk, O.; Amit, I.; Rabani, M.; Attardi, L.D.; Regev, A.; Lander, E.S.; Jacks, T.; Rinn, J.L. A Large Intergenic Noncoding RNA Induced by p53 Mediates Global Gene Repression in the p53 Response. *Cell*, **2010**, *142*, 409-419.

[89]   Mikkelsen, T.S.; Ku, M.; Jaffe, D.B.; Issac, B.; Lieberman, E.; Giannoukos, G.; Alvarez, P.; Brockman, W.; Kim, T.-K.; Koche, R.P.; Lee, W.; Mendenhall, E.; O'Donovan, A.; Presser, A.; Russ, C.; Xie, X.; Meissner, A.; Wernig, M.; Jaenisch, R.; Nusbaum, C.; Lander, E.S.; Bernstein, B.E. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, **2007**, *448*, 553-560.

[90]   Guttman, M.; Amit, I.; Garber, M.; French, C.; Lin, M.F.; Feldser, D.; Huarte, M.; Zuk, O.; Carey, B.W.; Cassady, J.P.; Cabili, M.N.; Jaenisch, R.; Mikkelsen, T.S.; Jacks, T.; Hacohen, N.; Bernstein, B.E.; Kellis, M.; Regev, A.; Rinn, J.L.; Lander, E.S. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, **2009**, *458*, 223-227.

[91]   Krumlauf, R. Hox genes in vertebrate development. *Cell*, **1994**, *78*, 191-201.

[92]   Duboule, D.; Morata, G. Colinearity and functional hierarchy among genes of the homeotic complexes. *Trends Genet.*, **1994**, *10*, 358-364.

[93]   Duboule, D.; Deschamps, J. Colinearity loops out. *Dev. Cell*, **2004**, *6*, 738-740.

[94]   Sessa, L.; Breiling, A.; Lavorgna, G.; Silvestri, L.; Casari, G.; Orlando, V. Noncoding RNA synthesis and loss of Polycomb group repression accompanies the colinear activation of the human HOXA cluster. *RNA*, **2007**, *13*, 223-239.

[95]   Fraser, J.; Rousseau, M.; Shenker, S.; Ferraiuolo, M.A.; Hayashizaki, Y.; Blanchette, M.; Dostie, J. Chromatin conformation signatures of cellular differentiation. *Genome Biol.*, **2009**, *10*, R37.

[96]   Ferraiuolo, M.A.; Rousseau, M.; Miyamoto, C.; Shenker, S.; Wang, X.Q.; Nadler, M.; Blanchette, M.; Dostie, J. The three-dimensional architecture of Hox cluster silencing. *Nucleic Acids Res.*, **2010**, *38*, 7472-7484.

[97]   Zhang, X.; Lian, Z.; Padden, C.; Gerstein, M.B.; Rozowsky, J.; Snyder, M.; Gingeras, T.R.; Kapranov, P.; Weissman, S.M.; Newburger, P.E. A myelopoiesis-associated regulatory intergenic noncoding RNA transcript within the human HOXA cluster. *Blood*, **2009**, *113*, 2526-2534.

[98]   Harrow, J.; Denoeud, F.; Frankish, A.; Reymond, A.; Chen, C.K.; Chrast, J.; Lagarde, J.; Gilbert, J.G.; Storey, R.; Swarbreck, D.; Rossier, C.; Ucla, C.; Hubbard, T.; Antonarakis, S.E.; Guigo, R. GENCODE: producing a reference annotation for ENCODE. *Genome Biol.*, **2006**, *7 Suppl 1*, S4 1-9.

[99]   Orom, U.A.; Derrien, T.; Beringer, M.; Gumireddy, K.; Gardini, A.; Bussotti, G.; Lai, F.; Zytnicki, M.; Notredame, C.; Huang, Q.;

Guigo, R.; Shiekhattar, R. Long noncoding RNAs with enhancer-like function in human cells. *Cell*, **2010**, *143*, 46-58.

[100] Tian, D.; Sun, S.; Lee, J.T. The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. *Cell*, **2010**, *143*, 390-403.

[101] Tsai, C.L.; Rowntree, R.K.; Cohen, D.E.; Lee, J.T. Higher order chromatin structure at the X-inactivation center *via* looping DNA. *Dev. Biol.*, **2008**, *319*, 416-425.

[102] Heintzman, N.D.; Hon, G.C.; Hawkins, R.D.; Kheradpour, P.; Stark, A.; Harp, L.F.; Ye, Z.; Lee, L.K.; Stuart, R.K.; Ching, C.W.; Ching, K.A.; Antosiewicz-Bourget, J.E.; Liu, H.; Zhang, X.; Green, R.D.; Lobanenkov, V.V.; Stewart, R.; Thomson, J.A.; Crawford, G.E.; Kellis, M.; Ren, B. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, **2009**, *459*, 108-112.

[103] Kim, T.K.; Hemberg, M.; Gray, J.M.; Costa, A.M.; Bear, D.M.; Wu, J.; Harmin, D.A.; Laptewicz, M.; Barbara-Haley, K.; Kuersten, S.; Markenscoff-Papadimitriou, E.; Kuhl, D.; Bito, H.; Worley, P.F.; Kreiman, G.; Greenberg, M.E. Widespread transcription at neuronal activity-regulated enhancers. *Nature*, **2010**, *465*, 182-187.

[104] Davis, C.A.; Ares, M., Jr. Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.*, **2006**, *103*, 3262-3267.

[105] Chekanova, J.A.; Gregory, B.D.; Reverdatto, S.V.; Chen, H.; Kumar, R.; Hooker, T.; Yazaki, J.; Li, P.; Skiba, N.; Peng, Q.; Alonso, J.; Brukhin, V.; Grossniklaus, U.; Ecker, J.R.; Belostotsky, D.A. Genome-wide high-resolution mapping of exosome substrates reveals hidden features in the Arabidopsis transcriptome. *Cell*, **2007**, *131*, 1340-1353.

[106] Guenther, M.G.; Levine, S.S.; Boyer, L.A.; Jaenisch, R.; Young, R.A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell*, **2007**, *130*, 77-88.

[107] Kapranov, P.; Willingham, A.T.; Gingeras, T.R. Genome-wide transcription and the implications for genomic organization. *Nat. Rev. Genet.*, **2007**, *8*, 413-423.

[108] Taft, R.J.; Glazov, E.A.; Cloonan, N.; Simons, C.; Stephen, S.; Faulkner, G.J.; Lassmann, T.; Forrest, A.R.; Grimmond, S.M.; Schroder, K.; Irvine, K.; Arakawa, T.; Nakamura, M.; Kubosaki, A.; Hayashida, K.; Kawazu, C.; Murata, M.; Nishiyori, H.; Fukuda, S.; Kawai, J.; Daub, C.O.; Hume, D.A.; Suzuki, H.; Orlando, V.; Carninci, P.; Hayashizaki, Y.; Mattick, J.S. Tiny RNAs associated with transcription start sites in animals. *Nat. Genet.*, **2009**, *41*, 572-578.

[109] Pal, M.; McKean, D.; Luse, D.S. Promoter clearance by RNA polymerase II is an extended, multistep process strongly affected by sequence. *Mol. Cell. Biol.*, **2001**, *21*, 5815-5825.

[110] Taft, R.J.; Simons, C.; Nahkuri, S.; Oey, H.; Korbie, D.J.; Mercer, T.R.; Holst, J.; Ritchie, W.; Wong, J.J.; Rasko, J.E.; Rokhsar, D.S.; Degnan, B.M.; Mattick, J.S. Nuclear-localized tiny RNAs are associated with transcription initiation and splice sites in metazoans. *Nat. Struct. Mol. Biol.*, **2010**, *17*, 1030-1034.

[111] Maison, C.; Bailly, D.; Peters, A.H.; Quivy, J.P.; Roche, D.; Taddei, A.; Lachner, M.; Jenuwein, T.; Almouzni, G. Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. *Nat. Genet.*, **2002**, *30*, 329-334.

[112] Muchardt, C.; Guilleme, M.; Seeler, J.S.; Trouche, D.; Dejean, A.; Yaniv, M. Coordinated methyl and RNA binding is required for heterochromatin localization of mammalian HP1alpha. *EMBO Rep.*, **2002**, *3*, 975-981.

[113] Kanhere, A.; Viiri, K.; Araujo, C.C.; Rasaiyaah, J.; Bouwman, R.D.; Whyte, W.A.; Pereira, C.F.; Brookes, E.; Walker, K.; Bell, G.W.; Pombo, A.; Fisher, A.G.; Young, R.A.; Jenner, R.G. Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. *Mol. Cell*, **2010**, *38*, 675-688.

[114] Chen, L.L.; Carmichael, G.G. Decoding the function of nuclear long non-coding RNAs. *Curr. Opin. Cell Biol.*, **2010**, *22*, 357-364.

[115] Hutchinson, J.N.; Ensminger, A.W.; Clemson, C.M.; Lynch, C.R.; Lawrence, J.B.; Chess, A. A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics*, **2007**, *8*, 39.

[116] Fox, A.H.; Lam, Y.W.; Leung, A.K.; Lyon, C.E.; Andersen, J.; Mann, M.; Lamond, A.I. Paraspeckles: a novel nuclear domain. *Curr. Biol.*, **2002**, *12*, 13-25.

[117] Clemson, C.M.; Hutchinson, J.N.; Sara, S.A.; Ensminger, A.W.; Fox, A.H.; Chess, A.; Lawrence, J.B. An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol. Cell*, **2009**, *33*, 717-726.

[118] Tripathi, V.; Ellis, J.D.; Shen, Z.; Song, D.Y.; Pan, Q.; Watt, A.T.; Freier, S.M.; Bennett, C.F.; Sharma, A.; Bubulya, P.A.; Blencowe, B.J.; Prasanth, S.G.; Prasanth, K.V. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell*, **2010**, *39*, 925-938.

[119] Tiwari, V.K.; Cope, L.; McGarvey, K.M.; Ohm, J.E.; Baylin, S.B. A novel 6C assay uncovers Polycomb-mediated higher order chromatin conformations. *Genome Res.*, **2008**, *18*, 1171-1179.