Large non-coding RNAs: missing links in cancer?

Maite Huarte¹,² and John L. Rinn¹,²,*

¹The Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA and ²Department of Pathology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, 02215, USA

Received July 17, 2010; Revised August 6, 2010; Accepted August 16, 2010

Cellular homeostasis is achieved by the proper balance of regulatory networks that if disrupted can lead to cellular transformation. These cell circuits are fine-tuned and maintained by the coordinated function of proteins and non-coding RNAs (ncRNAs). In addition to the well-characterized protein coding and microRNAs constituents, large ncRNAs are also emerging as important regulatory molecules in tumor-suppressor and oncogenic pathways. Recent studies have revealed mechanistic insight of large ncRNAs regulating key cancer pathways at a transcriptional, post-transcriptional and epigenetic level. Here we synthesize these latest advances within the context of their mechanistic roles in regulating and maintaining cellular equilibrium. We posit that similar to protein-coding genes, large ncRNAs are a newly emerging class of oncogenic and tumor-suppressor genes. Our growing knowledge of the role of large ncRNAs in cellular transformation is pointing towards their potential use as biomarkers and targets for novel therapeutic approaches in the future.

INTRODUCTION

The ultimate cause of cancer is the alteration of the balanced harmony of cellular networks and gene expression programs that maintain cellular homeostasis. Even the slightest perturbation of these pathways can result in cellular transformation. For decades, genetic studies have revealed the mutational alteration of genes that control these critical pathways such as DNA damage response, growth arrest, cell survival or the apoptotic pathway (1).

Genes controlling such balance can be classified into two major groups: tumor-suppressor genes and oncogenes. Tumor-suppressor genes protect cells against deleterious mutations and cellular regulation that could prime transformation. Conversely, genes that initiate the cellular transformation process upon inappropriate activation comprise oncogenes. Recent research points to the need for an expanded definition beyond just protein-coding genes to also include ‘tumor-suppressor non-coding RNAs (ncRNAs)’ and ‘oncogenic ncRNAs’.

Indeed, numerous profiling and characterization studies of microRNAs have identified critical roles for ncRNAs in cancer (2–5). MicroRNA alterations have been involved in the initiation and progression of human cancer. Furthermore, microRNA-expression profiling of human tumors has identified signatures associated with diagnosis, staging, progression, prognosis and response to treatment (4,6–12). Given that microRNAs primarily function as post-transcriptional regulators (13), they can act as tumor suppressor or oncogenes depending on their target genes (14). But in addition to the relatively well-described microRNAs, the growing knowledge of the mammalian non-coding transcriptome is revealing that the genome is also replete with large ncRNAs, which could have a major role in the development and progression of cancer, although their mechanisms of function remain less well understood (15–19).

Besides genetic mutations of tumor suppressor or oncogenes, a great deal of evidence indicates that epigenetic alteration is also a major factor contributing to tumor transformation and cancer (20). Intriguingly, a number of studies suggest that large non-coding RNAs are key components of the epigenetic regulatory networks.

For example, it is now well established that some large ncRNAs such as XIST, HOTAIR, AIR and KCNQ1OT1 interact with chromatin-remodeling complexes targeting them to specific genes to exert their functions (16,21–25). It has been proposed that different ncRNAs may serve as molecular scaffolds for those complexes so they can function in an appropriate spatial and temporal manner (26–28). In support of this hypothesis, a recent study shows that as many as 20% of large intergenic ncRNAs (lincRNAs) expressed in a given cell associate with chromatin-repressive complexes.
such as polycomb repressive complex 2 (PRC2) and that many of these lincRNAs are bound by multiple chromatin factors. Moreover, depletion of these lincRNAs affects the ability of PRC2 to regulate a specific subset of genes (27,28). More recently, HOTAIR was also discovered to bridge several chromatin-modifying complexes (29). Collectively, these studies point to an emerging theme of large ncRNAs interfacing with chromatin regulation by serving as molecular scaffolds.

It is noteworthy that chromatin-regulatory complexes are linked with the aberrant proliferation of cancer cells. For instance, SUZ12, a subunit of PRC2 complex is overexpressed in colon and breast cancers and EZH2 is upregulated in many tumors, including Hodgkin lymphoma, prostate and breast cancer (30). Moreover, EZH2 expression is associated with poor prognosis and is an indication for the metastatic character of the disease (30). Besides PcG proteins, misregulation of other chromatin complexes is associated with cancer. For example, ~10% of leukemias bear chromosomal translocations of the trithorax group histone methyltransferase mixed lineage leukemia (31). Collectively, these findings point to an important interplay between ncRNAs and chromatin regulation, which might be relevant for the control of gene expression networks critical for the process of cell transformation. These studies point to the interface of RNA and chromatin representing a new dimension in our understanding of cellular transformation.

Here we synthesize recent advances in our understanding of large ncRNAs in cancer pathways. These include examples with a wide range of molecular mechanisms involved in gene regulation. Although there is a rich literature of the roles of miRNAs and other small RNA pathways in cancer, here we specifically focus on large ncRNAs with particular emphasis on their molecular mechanisms in tumorigenesis.

MECHANISMS OF CELL TRANSFORMATION AND TUMOR SUPPRESSION BY LARGE ncRNAs

One of the first steps to the identification of ncRNAs relevant to disease is the profiling of their expression across normal and tumor samples. To this end, different profiling strategies have been applied to identify ncRNAs in cancer. Some studies have analyzed the available gene expression data sets in search for tumor-specific ncRNAs (32–34). Other groups have designed tiling arrays covering non-coding sequences of the genome to profile tumor cells (34–36), whereas others have focussed on the identification of ncRNA genes that are differentially methylated in tumors and may thus have a role in cell transformation (37,38). These approaches have led to the identification of several long ncRNAs, whose expression and/or DNA methylation are significantly associated with cancerous tissues. However, this effort has been greatly limited by the incomplete representation of non-coding sequences on DNA microarrays. More recently, this limitation has been overcome by the advances in massively parallel RNA sequencing combined with new computational methods. This has allowed the reconstruction of transcripts that originate the sequence reads (39,40). Application of these methods can result in significant progress in the study of currently poorly annotated ncRNAs, such as lincRNAs, and their splicing isoform diversity.

The studies described above have identified numerous large ncRNAs that exhibit differential expression between normal and tumor states. Although such alterations could be due in some cases to secondary effects of the tumor progression, numerous experimental studies (summarized in Tables 1 and 2) have suggested that ncRNAs play important roles in controlling cellular pathways involved in cellular transformation, thus acting as potential onco- or tumor-suppressor RNAs. A challenging task is to determine how these RNA molecules are able to modulate those pathways. Here, we describe recent studies that have shed new light on the functional and mechanistic roles of large ncRNAs in cancer.

Oncogenic ncRNAs

Similar to protein-coding oncopgenes, large ncRNAs can also promote cellular pathways that lead to tumorigenesis. One example of such an oncogenic lincRNA is HOTAIR. HOTAIR underscores the importance of understanding the relationship between epigenetic regulation by ncRNAs and cancer. HOTAIR is expressed from the HOXC locus and was initially discovered as a gene repressor of HOXD genes. This repressive action is conferred by the interaction of HOTAIR with the PRC2 complex, imparting PRC2 localization and repression of the HOXD locus (22) (Fig. 1A). A new study has found that HOTAIR is significantly overexpressed in breast tumors (34). Furthermore, HOTAIR expression level in primary breast tumors is a powerful predictor of patient outcomes such as metastasis and death (34). This phenotype seems to be tightly associated with PRC2-dependent gene repression induced by HOTAIR. Enforced expression of HOTAIR results in an altered pattern of H3K27 methylation and increased invasiveness, whereas the depletion of HOTAIR causes the opposite cellular phenotype (34). Collectively, these studies demonstrate how oncogenic lincRNAs can hijack the epigenetic machinery to reshape the epigenetic landscape leading to cancer.

In addition to intergenic large ncRNAs such as HOTAIR, global transcriptome analysis shows that up to 70% of protein-coding transcripts have antisense partners, and the perturbation of the antisense RNA can alter the expression of the sense gene (41). Some of these genes encode tumor-suppressor proteins that can become epigenetically silenced by the expression of the antisense ncRNA. Thus, misregulation of these antisense ncRNAs could lead to cellular transformation.

Indeed, the antisense ncRNA ANRIL controls expression in the INK4A/ARF locus comprising the tumor-suppressor genes INK4n/ARF/INK4a, p16/CDKN2A and p15/CDKN2B, which regulate cell cycle progression and senescence. ANRIL is transcribed antisense to the INK4n/ARF/INK4a promoter and overlaps with two exons of p15/CDKN2B. Independent studies have shown that overexpression of ANRIL in prostate cancer results in the silencing of INK4n/ARF/INK4a and p15/CDKN2B by heterochromatin formation (42,43). ANRIL interacts with CBX7, a component of the polycomb repressive complex 1 (PRC1), resulting in the targeting of this complex to the chromatin and the establishment of repressive epigenetic marks (43) (Fig. 1B). Another example of a tumor-suppressor gene that is epigenetically silenced by an antisense RNA is the cell cycle regulator p21/CDKN1A. In this case, the silencing...
<table>
<thead>
<tr>
<th>RNA</th>
<th>Organism</th>
<th>Size</th>
<th>Genomic location</th>
<th>Expression</th>
<th>Functional characteristics</th>
<th>Mechanism</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>p21 NAT/</td>
<td>Homo sapiens</td>
<td>≥423 bp</td>
<td>Antisense cdkn1a/p21</td>
<td>EST sequenced from neuroblastoma</td>
<td>Antisense of Cdkn1a/p21. Negative regulation of Cdkn1a/p21 by epigenetic silencing.</td>
<td>Requirement of Ago1 but not Ago2 for epigenetic silencing of Cdkn1a/p21</td>
<td>(44)</td>
</tr>
<tr>
<td>Bx332409</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOTAIR</td>
<td>H. sapiens</td>
<td>2.2 kb</td>
<td>Intergenic HoxC locus</td>
<td>Distal fibroblasts/metastatic breast tumors</td>
<td>Gene silencing in trans. Metastasis</td>
<td>Interaction with PRC2 and LSD1 complexes and targeting to repressed genes</td>
<td>(22,34)</td>
</tr>
<tr>
<td>MALAT-1</td>
<td>H. sapiens</td>
<td>8.7 kb</td>
<td>Intergenic Chr11</td>
<td>Overexpression in lung adenocarcinoma, breast, pancreas, colon, prostate and hepatocellular carcinomas</td>
<td>Gene silencing of INK4a/ARF/INK4a and p15/CDKN2B</td>
<td>Induction of GAGE6 proto-oncogene transcription by inhibition of PSF repressor</td>
<td>(35,49–51)</td>
</tr>
<tr>
<td>VL30-1</td>
<td>Mus musculus</td>
<td>4.9 kb</td>
<td>Retroelement ncRNA</td>
<td>Upregulated in prostate cancer</td>
<td>Gene silencing of INK4a/ARF/INK4a and p15/CDKN2B</td>
<td>Interaction with CBX7 component of PRC1 complex</td>
<td>(42,43,63–65)</td>
</tr>
<tr>
<td>ANRIL</td>
<td>H. sapiens</td>
<td>2.2 kb</td>
<td>Antisense of INK4a/ARF/INK4a and p15/CDKN2B</td>
<td>Overexpressed in a doxorubicin-resistant human squamous carcinoma subline</td>
<td>Inhibition of E-cadherin through induction of Zeb2 protein levels</td>
<td>Induction of Ras-mediated transformation of mouse fibroblasts</td>
<td>(47,60–62)</td>
</tr>
<tr>
<td>H19</td>
<td>H. sapiens</td>
<td>2.3 kb</td>
<td>Imprinted H19-Igf2 locus in Chr11</td>
<td>Bladder, human lung and breast carcinoma, choriocarcinoma</td>
<td>Control of imprinting. Oncogenic or tumor suppressor. Containing microRNA miR-675</td>
<td>Interaction with CBX7 component of PRC1 complex</td>
<td>(66–70)</td>
</tr>
<tr>
<td>CUDR</td>
<td>H. sapiens</td>
<td>2.2 kb</td>
<td>Intergenic Chr19</td>
<td>Overexpressed in a doxorubicin-resistant human squamous carcinoma subline</td>
<td>Inhibition of splicing of Zeb2 first exon containing IRES sequence</td>
<td>Unknown</td>
<td>(71)</td>
</tr>
<tr>
<td>Zeb2/Sip1</td>
<td>H. sapiens</td>
<td>680 bp</td>
<td>Antisense of Zeb2/Sip1</td>
<td>Overexpressed in human tumors with low E-cadherin expression</td>
<td>Inhibition of E-cadherin through induction of Zeb2 protein levels</td>
<td>Inhibition of splicing of Zeb2 first exon containing IRES sequence</td>
<td>(46)</td>
</tr>
<tr>
<td>NAT</td>
<td>H. sapiens</td>
<td>875 bp</td>
<td>Alternative splicing of SRA gene, loss of coding frame</td>
<td>In breast cancer cells, different balance of coding/non-coding splicing isoforms</td>
<td>Co-activator of steroid receptors and other transcription factors as MyoD. Increased levels of non-coding isoform associated with metastasis</td>
<td>Interaction in ribonucleoprotein complexes with several positive regulators, including SRC-1, p68 and p72, Pus1p and Pus3p, as well as negative regulators, such Sharp and SLIRP to be recruited to promoters of regulated genes</td>
<td>(72–79)</td>
</tr>
<tr>
<td>PCGEM1</td>
<td>H. sapiens</td>
<td>1.6 kb</td>
<td>Intergenic, Chr2</td>
<td>Upregulated in prostate cancer in African-American patients</td>
<td>Inhibition of apoptosis, promotion of cell growth increases proliferation, migration, invasion and drug resistance of human bladder cell line</td>
<td>Unknown</td>
<td>(80–84)</td>
</tr>
<tr>
<td>UCA1</td>
<td>H. sapiens</td>
<td>1.4 kb</td>
<td>Intergenic, Chr19</td>
<td>Embryonic development and bladder cancer associated</td>
<td>Unkown</td>
<td>Unknown</td>
<td>(85)</td>
</tr>
</tbody>
</table>
mechanism requires the component of the RNAi pathway Ago-1 (44) (Fig. 1C). Thus, perhaps similar to ANRIL, the p21 antisense ncRNA may also be upregulated in cancer rendering p21 inert, leading to cellular transformation.

In addition to the regulation of tumor-suppressor pathways by epigenetic silencing shown by previous examples, some antisense transcripts can also fine tune gene expression at the post-transcriptional level. E-cadherin is a gene correlated with gastric, breast, colorectal, thyroid and ovarian cancers. Its loss of function is thought to contribute to progression in cancer by increasing proliferation, invasion and/or metastasis (45). A strong association has been demonstrated between the expression of a particular natural antisense transcript (NAT) and human tumors with low E-cadherin expression (46). NAT overlaps with an intronic 5’ splice site of the Zeb2 gene and prevents its splicing. The retained intron contains an internal ribosome entry site (IRES) necessary for the increased translation of Zeb2 protein, which can subsequently function as a transcriptional repressor of E-cadherin (46) (Fig. 1D). Collectively, these studies provide strong impetus for further investigation of antisense ncRNAs in cancer pathways.

Other ncRNAs can induce the expression of the proto-oncogene Rab23 resulting in transformation and metastasis of skin fibroblasts (47). These ncRNAs act as inhibitory molecules, by complexing with the DNA and RNA-binding PSF (polypyrimidine tract-binding protein-associated splicing factor) protein to block its function as a transcriptional repressor, resulting in aberrant Rab23 expression. Interestingly, the interaction of ncRNAs with PSF is conserved between mouse and human, although species-specific ncRNAs are involved. In the mouse VL30-1 RNA, a member of the VL30 family of mouse retroelement ncRNAs mediates this mechanism (48), whereas in human five different RNAs can interact with PSF to induce the expression of the proto-oncogene GAGE6. These include the retroelements L1PA16 and MER11C, the mitochondrial gene HN encoding the humanin peptide and the ncRNA MALAT-1 (49) (Fig. 1E). Interestingly, many studies have identified the large ncRNA MALAT-1 as a tumor marker that is overexpressed in many different tumor types (35,50,51). However, it remains to be determined whether MALAT-1 acts exclusively through inhibition of the tumor-suppressor PSF.

Collectively, these studies point to the possibility of ‘oncogenic large ncRNAs’ that upon misregulation could either silence tumor-suppressor genes or induce the expression of oncogenes priming the cell for transformation.

Tumor-suppressor ncRNAs

Tumor-suppressor ncRNAs could phenotypically affect cells by promoting tumor-suppressor pathways, and when their function is compromised, cells are prone to develop cancer. In support of this notion, a few new studies have elucidated several examples (Table 2) of ‘tumor-suppressor large ncRNAs’.

For example, recent studies identified numerous lincRNAs that are induced by the p53 tumor-suppressor pathway (17,36). When cells are subjected to stress, the transcription factor p53 initiates a tumor-suppressor program that involves
the expression and repression of many genes. Surprisingly, among the genes specifically induced by p53, there are many lincRNAs. In particular, one of these lincRNAs, named lincRNA-p21, is directly induced by p53 to play a critical role in the p53 transcriptional response. LincRNA-p21 is required for the global repression of genes that interfere with p53 function regulating cellular apoptosis. Interestingly, lincRNA-p21 can mediate gene repression by physically interacting with the protein hnRNP-K, allowing its localization to promoters of genes to be repressed in a p53-dependent manner (36) (Fig. 2A). This study underlines the importance of well-tuned regulation of lincRNAs to orchestrate transcriptional programs that maintain cellular homeostasis.

GAS5 (growth arrest-specific 5) represents another example of a large ncRNA that regulates the expression of a critical subset of genes with tumor suppressive consequences. GAS5 is induced under starvation conditions being highly expressed in cells whose growth is arrested (52,53). GAS5 functions by outcompeting the DNA-binding sites of the glucocorticoid receptor (GR), thus reducing cell metabolism (54). Specifically, the GAS5 RNA conformation mimics that of the glucocorticoid responsive element (GRE) DNA, blocking the ability of GR to bind gene promoters to induce their transcription (54) (Fig. 2B). Interestingly, GAS5 has also been observed to be downregulated in breast cancer, perhaps to keep cancer cells active even under low nutrient conditions (52,53).

Another tumor-suppressor ncRNA is involved in the regulation of cyclin D1/CCND1 gene expression. Cyclin D1 is a cell cycle regulator frequently mutated, amplified and overexpressed in a variety of tumors (55). When cells are subjected to DNA damage, ncRNAs are expressed from the 5′ regulatory regions of cyclin D1 gene, thereby mediating its transcriptional repression. Indeed, these ncRNAs interact with the TLS protein, inducing its allosteric modification. This conformational change allows the association of TLS to the cyclin D1 promoter, which inhibits transcriptional induction by histone acetyltransferases such as CBP and p300 (56) (Fig. 2C).

These studies show that tumor-suppressor ncRNAs can be rapidly induced by cellular stress to regulate gene expression. Possibly, RNA molecules, due to their quick turnover rate, are ideal effectors when a rapid response is required to protect cells from external insults.

**Figure 1.** Mechanisms of gene regulation by oncogenic large ncRNAs. (A) lincRNA HOTAIR recruits PRC2 to specific gene promoters for methylation of lysine 27 of histone 3 (H3K27me), inducing gene repression that leads to breast tumor metastasis. (B) Large ncRNA ANRIL is transcribed antisense of the p14/ARF and p15/CDKN2B genes. ANRIL mediates gene silencing of the locus by interaction and recruitment of CBX7, a component of PRC1 and GAGE6 proto-oncogene (mouse) and GAGE6 proto-oncogene (human) are repressed by PSF protein. This repression is relieved when VL30-1 ncRNA (mouse) or MALAT-1 and others (human) interact with PSF, displacing it from the promoter.

**FUTURE PERSPECTIVES**

The studies reviewed herein contribute to the growing evidence of the important roles of large ncRNAs in cancer, both by regulating tumor-suppressor and oncogenic pathways. Thus, some ncRNAs play a critical role in maintaining cellular homeostasis and, when we have a deeper understanding of their roles in cancer, they can be used as diagnostic tools in conjunction with protein-coding genes.

An intriguing common theme is emerging of large ncRNAs forming ribonucleic–protein complexes that impart key regulatory functions in cellular circuits. We have discussed HOTAIR, lincRNA-p21 ANRIL or MALAT-1 among others that share a common functionality of forming RNA–protein complexes with chromatin regulatory factors. However, a higher-resolution understanding of cancer will require a comprehensive identification of large ncRNAs misregulated across a spectrum of cancer types and their associated protein complexes. Both biochemical approaches combined with in vivo studies will be required to fully understand the mechanistic and phenotypic roles of large ncRNAs in cancer.

A key goal for future progress is to identify large ncRNAs that could potentially serve as biomarkers for specific disease states. A clear advantage in the diagnostic use of ncRNA detection versus that of protein-coding RNAs is that in the former the RNA itself is the effector molecule, thus its expression levels may be a better indicator of the intrinsic characteristics of the tumor. Indeed, microRNA expression profiling has been successfully used for cancer classification, reflecting the developmental lineage and differentiation state of the tumors (10–12). In the near future, the great technological advance and decrease in cost of parallel massive sequencing will allow the profiling of the entire transcriptome of every type of tumor, including small and large ncRNA molecules, allowing the most powerful and informative diagnosis. In fact, the application of the new genomic technologies to the profiling of multiple cancers is already a reality. The tremendous amount of data generated by these projects present great possibilities for prognosis and therapeutic application. This has called for the creation of the International Cancer Genome Consortium (ICGC) that will coordinate the international effort to systematically study more than 25,000 cancer genomes at the genomic, epigenomic and transcriptomic levels (57). We can easily predict that in the next few years a complete catalogue of the large ncRNA expression as well as the genetic mutations, amplifications and deletions in non-coding regions associated with different types of tumors will be available.

Besides the imminent use of our knowledge of cancer-associated large ncRNAs for diagnosis, therapeutic applications may be possible in a more distant future. The progress in the use of RNAi-mediated gene silencing for the treatment of different diseases is encouraging and could be applied to selectively silence oncogenic ncRNAs. Gene therapy could also be applied for the delivery to specific cells of tumor-suppressor large ncRNAs for the treatment of cancer. However, many technical challenges have to be overcome for a wider use of therapeutic RNAi and gene therapy, including the development of reliable delivery systems, dosage regimes and techniques to ameliorate RNAi off target effects.
When the technical limitations are overcome, ncRNAs may be ideal targets for therapy due to their high turnover rate as well as their direct and specific regulatory functions. Predictably, therapeutic targeting of ncRNAs will carry fewer negative effects than those of protein-coding genes, given that they function regulating specific facets of their protein interacting partners.

In summary, overwhelming evidence reveals that large ncRNAs are molecules that keep in perfect tune the balance of gene expression networks, and discordance in their function results in homeostatic imbalance, ultimately causing cellular transformation. Large ncRNAs are shedding new light on our understanding of these cancer pathways and may represent a ‘missing link’ in cancer.

ACKNOWLEDGEMENTS

We would like to thank Moran Cabili-Kalmar, Magdalena J. Koziol, Sabine Loewer and Bárbara Tazón-Vega for critical comments on the manuscript and Sigrid Hart (Broad Institute) for illustration support.

Conflict of Interest statement. None declared.
FUNDING
J.L.R. is a Damon Runyon-Rachleff, Searle and Smith Family Foundation Scholar and Richard Merkin awardee. M.H. is supported by the NIH Directors New Innovator Award (Grant number 1DP2OD00667-01), Smith Family Foundation, the Damon Runyon Cancer Foundation and Searle Scholar Program.

REFERENCES


