The Emergence of lncRNAs in Cancer Biology

John R. Prensner\textsuperscript{1,2} and Arul M. Chinnaiyan\textsuperscript{1,5}

INTRODUCTION

A central question in biology has been, Which regions of the human genome constitute its functional elements: those expressed as genes or those serving as regulatory elements? In the 1970s and 1980s, early cloning-based methods revealed more than 7,000 human genes (1), but in the 1990s large-scale analyses of expressed sequence tags suggested that the number of human genes was in the range of 35,000 to 100,000 (2). The completion of the Human Genome Project narrowed the focus considerably by highlighting the surprisingly small number of protein-coding genes, which is now conventionally cited as <25,000 (3).

While the number of protein-coding genes (20,000–25,000) has maintained broad consensus, recent studies of the human transcriptome have revealed an astounding number of noncoding RNAs (ncRNA). These transcribed elements, which lack the capacity to code for a protein, are bafflingly abundant in all organisms studied to date, from yeast to humans (4–6). Yet, over the past decade, numerous studies have shown that ncRNAs have distinct biologic functions and operate through defined mechanisms. Still, their sheer abundance—some reports estimate that up to 70% of the human genome is transcribed into RNA (4)—has sparked debates as to whether ncRNA transcription reflects true biology or by-products of a leaky transcriptional system. Encompassed within these studies are the broad questions of what constitutes a human gene, what distinguishes a gene from a region that is simply transcribed, and how we interpret the biologic meaning of transcription.

DEFINING lncRNAs AS DISTINCT TRANSCRIPTS

Currently, lncRNAs are emerging as a fundamental aspect of biology. However, recent estimates that up to 70% of the human genome may be transcribed have complicated the interpretation of the act of transcription. Although some have argued that many of the transcribed RNAs may reflect a “leaky” transcriptional system in mammalian cells, lncRNAs have largely avoided these controversies due to their strongly defined identity. We have indicated several common features of lncRNAs that confirm their biologic robustness:

• Epigenetic marks consistent with a transcribed gene (H3K4me3 at the gene promoter, H3K36me3 throughout the gene body)
• Transcription via RNA polymerase II
• Polyadenylation
• Frequent splicing of multiple exons via canonical genomic splice site motifs
• Regulation by well-established transcription factors
• Frequent expression in a tissue-specific manner

The discovery of numerous noncoding RNA (ncRNA) transcripts in species from yeast to mammals has dramatically altered our understanding of cell biology, especially the biology of diseases such as cancer. In humans, the identification of abundant long ncRNA (lncRNA) >200 bp has catalyzed their characterization as critical components of cancer biology. Recently, roles for lncRNAs as drivers of tumor suppressive and oncogenic functions have appeared in prevalent cancer types, such as breast and prostate cancer. In this review, we highlight the emerging impact of ncRNAs in cancer research, with a particular focus on the mechanisms and functions of lncRNAs.

Significance: lncRNAs represent the leading edge of cancer research. Their identity, function, and dysregulation in cancer are only beginning to be understood, and recent data suggest that they may serve as master drivers of carcinogenesis. Increased research on these RNAs will lead to a greater understanding of cancer cell function and may lead to novel clinical applications in oncology. Cancer Discovery. 1(5):391–407. ©2011 AACR.
ncRNA: A NEW KIND OF GENE

ncRNAs are RNA transcripts that do not encode for a protein. In the past decade, a great diversity of ncRNAs has been observed. Depending on the type of ncRNA, transcription can occur by any of the three RNA polymerases (RNA Pol I, RNA Pol II, or RNA Pol III). General conventions divide ncRNAs into two main categories: small ncRNAs (<200 bp) and lncRNAs (>200 bp) (10). Within these two categories, there are also many individual classes of ncRNAs (Table 1), although the degree of biologic and experimental support for each class varies substantially and, therefore, classes should be evaluated individually.

These developments have been matched by equally insightful discoveries analyzing the role of ncRNAs in human diseases, especially cancer, lending support to the importance of their cellular functions (7, 8). Initial evidence suggests that ncRNAs, particularly long ncRNAs (lncRNA), have essential roles in tumorigenesis (7), and that lncRNA-mediated biology occupies a central place in cancer progression (9). With the number of well-characterized cancer-associated lncRNAs growing, the study of lncRNAs in cancer is now generating new hypotheses about the biology of cancer cells. Here, we review the current understanding of ncRNAs in cancer, with particular focus on lncRNAs as novel drivers of tumorigenesis.

Table 1. Types of ncRNAs known in humans

<table>
<thead>
<tr>
<th>Category</th>
<th>Name</th>
<th>Quality of supporting data</th>
<th>Specific role in carcinogenesis</th>
<th>Aberration in cancer</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Housekeeping RNAs</td>
<td>Transfer RNAs</td>
<td>High</td>
<td>No</td>
<td>No</td>
<td>10, 11</td>
</tr>
<tr>
<td></td>
<td>Ribosomal RNAs</td>
<td>High</td>
<td>No</td>
<td>No</td>
<td>10, 11</td>
</tr>
<tr>
<td></td>
<td>Small nucleolar RNAs</td>
<td>High</td>
<td>No</td>
<td>No</td>
<td>10, 11</td>
</tr>
<tr>
<td></td>
<td>Small nuclear RNAs</td>
<td>High</td>
<td>No</td>
<td>No</td>
<td>10, 11</td>
</tr>
<tr>
<td>Small ncRNAs (&lt;200 bp in size)</td>
<td>MicroRNAs</td>
<td>High</td>
<td>Yes</td>
<td>Amplification, deletion, methylation, gene expression</td>
<td>12, 13</td>
</tr>
<tr>
<td></td>
<td>Tiny transcription initiation RNAs</td>
<td>High</td>
<td>Not known</td>
<td>Not known</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Repeat-associated small interfering RNAs</td>
<td>High</td>
<td>Not known</td>
<td>Not known</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Promoter-associated short RNAs</td>
<td>High</td>
<td>Not known</td>
<td>Not known</td>
<td>4, 6, 11</td>
</tr>
<tr>
<td></td>
<td>Terminator-associated short RNAs</td>
<td>High</td>
<td>Not known</td>
<td>Not known</td>
<td>4, 6, 11</td>
</tr>
<tr>
<td></td>
<td>Antisense terminator-associated short RNAs</td>
<td>High</td>
<td>Not known</td>
<td>Not known</td>
<td>6, 10</td>
</tr>
<tr>
<td></td>
<td>Transcription start site antisense RNAs</td>
<td>Moderate</td>
<td>Not known</td>
<td>Not known</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Retrotransposon-derived RNAs</td>
<td>High</td>
<td>Not known</td>
<td>Not known</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>3'UTR-derived RNAs</td>
<td>Moderate</td>
<td>Not known</td>
<td>Not known</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Splice-site RNAs</td>
<td>Poor</td>
<td>Not known</td>
<td>Not known</td>
<td>11</td>
</tr>
<tr>
<td>Long ncRNAs (&gt;200 bp in size)</td>
<td>Long or large intergenic ncRNAs</td>
<td>High</td>
<td>Yes</td>
<td>Gene expression, translocation</td>
<td>79, 25, 101</td>
</tr>
<tr>
<td></td>
<td>Transcribed ultraconserved regions</td>
<td>High</td>
<td>Yes</td>
<td>Gene expression</td>
<td>18, 19</td>
</tr>
<tr>
<td></td>
<td>Pseudogenes</td>
<td>High</td>
<td>Yes</td>
<td>Gene expression, deletion</td>
<td>15, 81</td>
</tr>
<tr>
<td></td>
<td>Enhancer RNAs</td>
<td>High</td>
<td>Yes</td>
<td>Not known</td>
<td>17, 29</td>
</tr>
<tr>
<td></td>
<td>Repeat-associated ncRNAs</td>
<td>High</td>
<td>Not known</td>
<td>Not known</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Long intronic ncRNAs</td>
<td>Moderate</td>
<td>Not known</td>
<td>Not known</td>
<td>10, 11</td>
</tr>
<tr>
<td></td>
<td>Antisense RNAs</td>
<td>High</td>
<td>Yes</td>
<td>Gene expression</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Promoter-associated long RNAs</td>
<td>Moderate</td>
<td>Not known</td>
<td>Not known</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Long stress-induced noncoding transcripts</td>
<td>Moderate</td>
<td>Yes</td>
<td>Gene expression</td>
<td>10, 11</td>
</tr>
</tbody>
</table>
Small ncRNAs

The diversity of small ncRNAs has perhaps grown the most; several dozen classes of small ncRNAs have been proposed (10, 11). These include well-characterized housekeeping ncRNAs (transfer RNA and some ribosomal RNA) essential for fundamental aspects of cell biology; splicing RNAs (small nuclear RNA); and a variety of recently observed RNAs associated with protein-coding gene transcription, such as tiny transcription-initiation RNAs, promoter-associated short RNAs, termini-associated short RNAs, 3′-untranslated region (UTR)–derived RNAs, and antisense termini-associated short RNAs (10).

To date, the most extensively studied small RNAs in cancer are microRNAs (miRNA). Elegant studies over the past 15 years have defined an intricate mechanistic basis for miRNA-mediated silencing of target gene expression through the RNA-induced silencing complex (RISC), which employs Argonaute family proteins (such as AGO2) to cleave target mRNA transcripts or inhibit the translation of that mRNA (Fig. 1A) (12). Aberrant expression patterns of miRNAs in cancer have been well documented in most tumor types (Fig. 1B), and detailed work from many laboratories has shown that numerous miRNAs, including miR-10b, let-7, miR-101, and the miR-15a-16-1 cluster, possess oncogenic or tumor suppressive functions (Fig. 1C) (12, 13).

Long ncRNAs

Recent observations of novel long ncRNA species have led to the establishment of a complex set of terms and terminologies used to describe a given long ncRNA. These include antisense RNAs, which are transcribed on the opposite strand from a protein-coding gene and frequently overlap that gene (14); transcribed ultraconserved regions (T-UCR), which originate in regions of the genome showing remarkable conservation across species; and ncRNAs derived from intronic transcription.

Although many RNA species are >200 bp long, such as repeat or pseudogene-derived transcripts (15), the abbreviated term IncRNA (also referred to as lncRNA, for long intergenic ncRNA) does not uniformly apply to all of these (Box 1). Although the nomenclature is still evolving, IncRNA typically refers to a polyadenylated long ncRNA that is transcribed by RNA polymerase II and associated with epigenetic signatures common to protein-coding genes, such as trimethylation of histone 3 lysine 4 (H3K4me3) at the transcriptional start site (TSS) and trimethylation of histone 3 lysine 36 (H3K36me3) throughout the gene body (16). This description also suits many T-UCRs and some antisense RNAs, and the overlap between these categories may be substantial. IncRNAs also commonly exhibit splicing of multiple exons into a mature transcript, as do many antisense RNAs, but not RNAs

![Figure 1](https://example.com/figure1.png)
transcribed from gene enhancers [enhancer RNAs (eRNA)] or T-UCRs (17,19). Transcription of lncRNAs occurs from an independent gene promoter and is not coupled to the transcription of a nearby or associated parental gene, as with some classes of ncRNAs (promoter/termini-associated RNAs, intronic ncRNAs) (10). In this review, we use the term IncRNA in this manner. When the data are supportive, we include specific T-UCRs and antisense RNAs under the IncRNA umbrella term, and we distinguish other long ncRNAs, such as eRNAs, where appropriate.

Identification of Long ncRNAs

Many initial IncRNAs, such as XIST and H19, were discovered in the 1980s and 1990s by searching cDNA libraries for clones of interest (20, 21). In these studies, the intention was generally to identify new genes important in a particular biologic process—X chromosome inactivation in the example of XIST—by studying their expression patterns. At the time, most of the genes discovered were protein coding, and this tended to be the assumption, with a few exceptions, such as XIST, which were subsequently determined to be noncoding as a secondary observation (20).

In the past decade, however, large-scale analyses have focused on comprehensively identifying ncRNA species. This paradigm shift has been mediated by dramatic advances in high-throughput technologies, including DNA tiling arrays and next-generation RNA sequencing (RNA-Seq) (9, 22–25). These platforms provide systems with which RNA transcription can be observed in an unbiased manner and have thereby highlighted the pervasive transcription of ncRNAs in cell biology (Box 2). Moreover, whereas conventional cDNA microarrays detected only the transcripts represented by probes on the array, the introduction and popularization of RNA-Seq as a standard tool in transcriptome studies has removed many obstacles affected the field of ncRNA research, in large part due to the ability to detect and monitor ncRNA expression in a global and unbiased manner. However, because the processing and interpretation of high-throughput data can be challenging, extensive validation by wet-lab assays is still an important part of confirming initial nominations. The following are the most commonly used methods to discover and validate ncRNAs.

<table>
<thead>
<tr>
<th>Discovery methods</th>
<th>Validation methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA tiling arrays</td>
<td>PCR</td>
</tr>
<tr>
<td>RNA-sequencing (RNA-Seq)</td>
<td>Immunohistochemistry</td>
</tr>
<tr>
<td>Custom microarrays</td>
<td>Northern blot</td>
</tr>
<tr>
<td></td>
<td>Rapid amplification of cDNA ends (RACE)</td>
</tr>
</tbody>
</table>

With the advent of high-throughput technologies, more and more ncRNA species are being discovered and characterized in mammalian systems. In this way, advancing technological achievements have dramatically affected the field of ncRNA research, in large part due to the ability to detect and monitor ncRNA expression in a global and unbiased manner. However, because the processing and interpretation of high-throughput data can be challenging, extensive validation by wet-lab assays is still an important part of confirming initial nominations. The following are the most commonly used methods to discover and validate ncRNAs.

**LONG ncRNAs IN CANCER**

Emerging evidence suggests that lncRNAs constitute an important component of tumor biology (Table 2). Dysregulated expression of lncRNAs in cancer marks the spectrum of disease progression (9) and may serve as an independent predictor for patient outcomes (28). Mechanistically, most well-characterized IncRNAs to date show a functional role in gene expression regulation, typically transcriptional rather than posttranscriptional regulation. This can occur by targeting either genomically local (cis-regulation) or genomically distant (trans-regulation) genes. Recently, a new type of long ncRNAs at gene enhancers, termed eRNAs, have also been implicated in transcriptional regulation (29).

**cis-Regulatory IncRNAs**

By recruiting histone modification complexes to specific areas of the genome, cis-regulation by IncRNAs contributes to local control of gene expression (Fig. 2). This effect either can be highly specific to a particular gene, such as the regulation of IGF2 by IncRNAs (30), or can encompass a wide chromosomal region, such as X-chromosome inactivation in women through XIST. Historically, cis-regulation through IncRNAs was studied earlier than trans-regulation, because several cis-regulatory IncRNAs, including H19, AIR, KCNQ1OT1, and XIST, were earlier discoveries (20, 21, 31). Several cis-regulatory IncRNAs, including H19, AIR, and KCNQ1OT1, are also functionally related through their involvement in epigenetic imprinting regions.

**Imprinting IncRNAs**

The involvement of IncRNA in imprinted regions of the genome is critical for maintaining parent-of-origin-specific gene expression. In particular, an imprinted region of human chromosome 11 (orthologous to mouse chromosome 7) has been observed that many lncRNAs show tissue-specific expression, researchers speculate that the human genome may harbor nearly as many lncRNAs as protein-coding genes (perhaps ~15,000 IncRNA), although only a fraction are expressed in a given cell type.
The Emergence of IncRNAs in Cancer Biology

Together, these data support a general role for H19 in cancer, although its precise biologic contributions are still unclear. Other imprinting-associated lncRNAs are only tangentially associated with cancer. Although loss of imprinting is observed in many tumors, the role for lncRNAs in this process is not well defined. For example, Beckwith-Wiedemann syndrome (BWS), a disorder of abnormal development with an increased risk of cancer, displays aberrant imprinting patterns of \textit{KCNQ1OT1} (32, 37), but a direct association or causal role for \textit{KCNQ1OT1} in cancer is not described (37). Conversely, aberrant \textit{H19} methylation in BWS seems to predispose to cancer development more strongly (37).

\textbf{XIST} \textit{XIST}, perhaps the most well-studied lncRNA, is transcribed from the inactivated X chromosome in order to facilitate that chromosome's inactivation and manifests as multiple table.

### Table 2. Examples of IncRNAs in cancer

<table>
<thead>
<tr>
<th>IncRNA</th>
<th>Function</th>
<th>Cancer type</th>
<th>Cancer phenotype</th>
<th>Molecular interactors</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>HULC</td>
<td>Biomarker</td>
<td>Hepatocellular</td>
<td>Not known</td>
<td>Unknown</td>
<td>10</td>
</tr>
<tr>
<td>PCA3</td>
<td>Biomarker</td>
<td>Prostate</td>
<td>Not known</td>
<td>Unknown</td>
<td>82, 83</td>
</tr>
<tr>
<td>ANRIL/p15AS</td>
<td>Oncogenic</td>
<td>Prostate, leukemia</td>
<td>Suppression of senescence via LNK4A</td>
<td>Binds PRC1 and PRC2</td>
<td>46–48, 68</td>
</tr>
<tr>
<td>HOTAIR</td>
<td>Oncogenic</td>
<td>Breast, hepatocellular</td>
<td>Promotes metastasis</td>
<td>Binds PRC2 and LSD1</td>
<td>28, 55, 56</td>
</tr>
<tr>
<td>MALAT1/NEAT2</td>
<td>Oncogenic</td>
<td>Lung, prostate, breast, colon</td>
<td>Unclear</td>
<td>Contributory to nuclear paraspeckle function</td>
<td>76–79</td>
</tr>
<tr>
<td>PCAT-1</td>
<td>Oncogenic</td>
<td>Prostate</td>
<td>Promotes cell proliferation; inhibits BRCA2</td>
<td>Unknown</td>
<td>9</td>
</tr>
<tr>
<td>PCGEM1</td>
<td>Oncogenic</td>
<td>Prostate</td>
<td>Inhibits apoptosis; promotes cell proliferation</td>
<td>Unknown</td>
<td>7, 10</td>
</tr>
<tr>
<td>TUC338</td>
<td>Oncogenic</td>
<td>Hepatocellular</td>
<td>Promotes cell proliferation and colony formation</td>
<td>Unknown</td>
<td>19</td>
</tr>
<tr>
<td>uc.73a</td>
<td>Oncogenic</td>
<td>Leukemia</td>
<td>Inhibits apoptosis; promotes cell proliferation</td>
<td>Unknown</td>
<td>18</td>
</tr>
<tr>
<td>H19</td>
<td>Oncogenic; tumor suppressive</td>
<td>Breast, hepatocellular</td>
<td>Promotes cell growth and proliferation; activated by cMYC, downregulated by prolonged cell proliferation</td>
<td>Unknown</td>
<td>30, 34–36</td>
</tr>
<tr>
<td>GASS</td>
<td>Tumor suppressive</td>
<td>Breast</td>
<td>Induces apoptosis and growth arrest; prevents GR-induced gene expression</td>
<td>Binds GR</td>
<td>57, 58</td>
</tr>
<tr>
<td>linc-p21</td>
<td>Tumor suppressive</td>
<td>Mouse models of lung, sarcoma, lymphoma</td>
<td>Mediates p53 signaling; induces apoptosis</td>
<td>Binds hnRNP-k</td>
<td>73</td>
</tr>
<tr>
<td>MEG3</td>
<td>Tumor suppressive</td>
<td>Meningioma, hepatocellular, leukemia, pituitary</td>
<td>Mediates p53 signaling; inhibits cell proliferation</td>
<td>Unknown</td>
<td>69–72</td>
</tr>
<tr>
<td>PTENP1</td>
<td>Tumor suppressive</td>
<td>Prostate, colon</td>
<td>Binds PTEN-suppressing miRNAs</td>
<td>Unknown</td>
<td>81</td>
</tr>
</tbody>
</table>

**Abbreviations:** GR, glucocorticoid receptor; PRC1, polycomb repressive complex 1; PRC2, polycomb repressive complex 2.
**cis-Regulation**

A  Maternal allele

B  X chromosome

**trans-Regulation**

C  Chromosome 12

D  Chromosome 2

**E**  Cell proliferation

**Figure 2.** Gene expression regulation by lncRNAs. A and B, cis-regulation of gene expression results in local control of genes neighboring, or on the same chromosome as, lncRNA transcription. A, H19 and KCNQ1OT1 are imprinted lncRNAs on chromosome 11 associated with allele-specific expression of IGF2 and KCNQ1. B, XIST transcription facilitates inactivation of an individual X chromosome in women by recruiting PRC2. C and D, trans-regulation of gene expression results in control of genomically distant genes. C, HOXAIR is transcribed from the HOX cluster on chromosome 12 but represses the HOX loci via PRC2-mediated epigenetic modifications. D, PCAT-1 is transcribed from chromosome 8 but regulates target genes such as BRCA2, CENPE, and CENPF, thereby impacting cell proliferation. E, The GAS5 lncRNA binds GR and sequesters it, preventing upregulation of GR-target genes across the genome. The blue lines represent the GAS5 transcript.

isoforms (38, 39). On the active X allele, XIST is repressed by its antisense partner ncRNA, TSIX (39). XIST contains a double-hairpin RNA motif in the RepA domain, located in the first exon, which is crucial for its ability to bind polycomb repressive complex 2 (PRC2) and propagate epigenetic silencing of an individual X chromosome (Fig. 2B) (40).

Despite the body of research on XIST, a precise role for XIST in cancer has remained elusive (41). Some evidence initially suggested a role for XIST in hereditary BRCA1-deficient breast cancers (42, 43); data indicated that BRCA1 was not required for XIST function in these cells (44). It has also been reasoned that XIST may be implicated in the X chromosome abnormalities observed in some breast cancers. In addition, there have been surprising accounts of aberrant XIST regulation in other cancers, including lymphoma and male testicular germ-cell tumors, in which XIST hypomethylation is, unexpectedly, a biomarker (45). However, it remains unclear whether these observations reflect a passenger or driver status for XIST, because a well-defined function for XIST in cancer has yet to attain a broad consensus.

**ANRIL**

Located on Ch9p21 in the INK4A/ARF tumor suppressor locus, ANRIL was initially described by examining the

![Image of gene expression regulation by lncRNAs.](image-url)
deletion of this region in hereditary neural system tumors, which predispose for hereditary cutaneous malignant melanoma (46). ANRIL was subsequently defined as a polyadenylated lncRNA antisense to the CDKN2A and CDKN2B genes. In vitro data have suggested that ANRIL functions to repress the INK4A/INK4B isoforms (47), but not ARF. This repression is mediated through direct binding to CBX7 (47), a member of PRC1, and SUZ12 (48), a member of PRC2, which apply repressive histone modifications to the locus. However, these data are from studies done in different cell types, and it is not known whether ANRIL binds both complexes simultaneously.

ANRIL also displays a highly complicated splicing pattern, with numerous variants, including circular RNA isoforms (49). Currently, it is unclear whether these isoforms have tissue-specific expression patterns or unique functions, which may suggest a biologic basis for this variation. Through genome-wide association studies (GWAS), ANRIL has also been identified by single-nucleotide polymorphisms (SNP) correlated with a higher risk of atherosclerosis and coronary artery disease (50), and ANRIL expression has been noted in many tissues. The function and isoform-level expression of ANRIL in these tissue types is not yet elucidated but may shed light on its role in diverse disease processes.

**HOTTIP and HOTAIRM1**

An intriguing theme emerging in developmental biology is the regulation of HOX gene expression by lncRNAs. Highly conserved among metazoan species, HOX genes are responsible for determining tissue patterning and early development, and in humans HOX genes reside in four genomic clusters. Within these clusters, HOX genes display intriguing anterior-posterior and proximal-distal expression patterns that mirror their genomic position 5′ to 3′ in the gene cluster.

Two recently discovered lncRNAs, termed HOTTIP and HOTAIRM1, may help to explain this colinear patterning of HOX gene expression. HOTTIP and HOTAIRM1 are located at opposite ends of the HoxA cluster, and each helps to enhance gene expression of the neighboring HoxA genes (51, 52). HOTAIRM1, located at the 3′ end, coordinates HOX1 expression and has tissue-specific expression patterns identical to those of HOX1 (51). HOTTIP, by contrast, is at the 5′ end of the cluster and similarly enhances expression of the 5′ HoxA genes, most prominently HOX13 (52). Mechanistic studies of HOTTIP suggest that it binds WDR5 and recruits the MLL H3K4 histone methyltransferase complex to the HoxA cluster to support active chromatin confirmation (52). These observations distinguish HOTTIP and HOTAIRM1, because most lncRNAs to date facilitate gene repression.

Although HOTAIRM1 and HOTTIP have not been extensively studied in cancer, expression of these lncRNAs may have important roles in the development of cancer cells. For example, differentiation of myeloid cell lines, such as K562 and NB4, by treatment with small-molecule drugs led to an increase in HOTAIRM1 expression, implicating it in myeloid differentiation (51). Moreover, HoxA genes are broadly known to be important for many cancers, particularly HOXA9, which is essential for oncogenesis in leukemias harboring MLL rearrangements. Thus, HOTAIRM1 and HOTTIP also suggest a potential role for lncRNAs in MLL-rearranged leukemias.

**trans-Regulatory lncRNAs**

Like most cis-acting lncRNAs, trans-acting lncRNAs typically facilitate epigenetic regulation of gene expression. However, because trans-acting lncRNAs may operate at geographically distant locations of the genome, it is generally thought that the mature lncRNA transcript is the primary actor in these cases, as opposed to cis-regulating lncRNAs such as H19, AIR, and KCNQ1OT1, which may function through the act of transcription itself (34, 53, 54).

**HOTAIR**

The characterization of HOTAIR led to the widespread attention to trans-regulatory lncRNAs. First described in fibroblasts, HOTAIR is located in the HoxC cluster, but unlike HOTTIP and HOTAIRM1, HOTAIR was found to regulate HoxD cluster genes in a trans-regulatory mechanism (Fig. 2C) (55). These observations raise the question, Are all Hox clusters regulated by lncRNAs, either by a cis-regulatory or by a trans-regulatory mechanism?

In cancer, HOTAIR is upregulated in breast and hepatocellular carcinomas (10), and in breast cancer overexpression of HOTAIR is an independent predictor of overall survival and progression-free survival (28). Work by Howard Chang and colleagues has further defined a compelling mechanistic basis for HOTAIR in cancer. HOTAIR has two main functional domains, a PRC2-binding domain located at the 5′ end of the RNA, and an LSD1/CoREST1-binding domain located at the 3′ end of the RNA (55, 56). In this way, HOTAIR is thought to operate as a tether that links two repressive protein complexes in order to coordinate their functions. In breast cancer, HOTAIR overexpression facilitates aberrant PRC2 function by increasing PRC2 recruitment to the genomic positions of target genes. By doing so, HOTAIR mediates the epigenetic repression of PRC2 target genes, and profiling of repressive (H3K27me3) and active (H3K4me3) chromatin marks shows widespread changes in chromatin structure following HOTAIR knockdown (28).

Furthermore, HOTAIR dysregulation results in a phenotype in both in vitro and in vivo models. Ectopic overexpression of HOTAIR in breast cancer cell lines increases their invasiveness both in vitro and in vivo. Supporting these findings, in benign immortalized breast cells overexpressing EZH2, a core component of PRC2, knockdown of HOTAIR mitigated EZH2-induced invasion in vitro (28). Taken together, these data provide the most thorough picture of an lncRNA in cancer.

**PCAT-1**

Using RNA-Seq (i.e., transcriptome sequencing) on a large panel of tissue samples, our laboratory recently described approximately 1,800 lncRNAs expressed in prostate tissue, including 121 lncRNAs that are transcriptionally dysregulated in prostate cancer (9). These 121 prostate cancer–associated transcripts (PCAT) may represent an unbiased list of potentially functional lncRNAs associated with prostate cancer. Among these, we focused on PCAT-1, a 1.9-kb polyadenylated lncRNA comprising two exons and located in the Chr8q24 gene desert (9).
PCAT-1 shows tissue-specific expression and is selectively upregulated only in prostate cancer. Interestingly, PCAT-1, unlike HOTAIR, is repressed by PRC2, and PCAT-1 overexpression may define a molecular subtype of prostate that is not coordinated by PRC2 (9). In vitro and in vivo experiments showed that PCAT-1 supports cancer cell proliferation (J.R. Prensner and A.M. Chinnaiyan; unpublished data). Like HOTAIR, PCAT-1 functions predominantly as a transcriptional repressor by facilitating trans-regulation of genes preferentially involved in mitosis and cell division, including known tumor suppressor genes such as BRCA2 (Fig. 2D). Intriguingly, because loss of BRCA2 function is known to increase cell sensitivity to small-molecule inhibitors of PARP1, these data may suggest that PCAT-1 may impact cellular response to these drugs as well.

The discovery of PCAT-1 highlights the power of unbiased transcriptome studies to explore a rich set of IncRNAs associated with cancer. Although PCAT-1 is the first cancer IncRNA to be discovered by this method, we anticipate that many additional studies will use this approach.

**GAS5**

GAS5, first identified in murine NIH-3T3 cells, is a mature, spliced IncRNA manifesting as multiple isoforms up to 12 exons in size (57). Using HeLa cells engineered to express GAS5, Kino and colleagues (57) recently described an intriguing mechanism by which GAS5 modulates cell survival and metabolism by antagonizing the glucocorticoid receptor (GR). The 3’ end of GAS5 both interacts with the GR DNA-binding domain and is sufficient to repress GR-induced genes, such as clAP2, when cells are stimulated with dexamethasone. By binding to the GR, GAS5 serves as a decoy that prevents GR binding to target DNA sequences (Fig. 2E) (57).

In cancer, GAS5 induces apoptosis and suppresses cell proliferation when overexpressed in breast cancer cell lines, and in human breast tumors GAS5 expression is downregulated (58). Although it is unclear whether this phenotype is due to an interaction with GR, it is intriguing that GAS5 may also be able to suppress signaling by other hormone receptors, such as androgen receptor (AR), although this effect has not been seen with estrogen receptor (ER) (57).

**Other Long ncRNAs**

**eRNAs**

eRNAs are transcribed by RNA polymerase II at active gene enhancers (17). However, unlike IncRNAs, they are not polyadenylated and are marked by an H3K4me1 histone signature denoting enhancer regions (17) rather than the H3K4me3/H3K36me3 signature classically associated with IncRNAs. Although research on eRNAs is still in the earliest phases, an emerging role for them in hormone signaling is already being explored. Nuclear hormone receptors, such as AR and ER, are critical regulators of numerous cell growth pathways and are important in large subsets of prostate (AR), breast (ER), and thyroid (PPAR) cancers. To date, eRNAs have been most directly implicated in prostate cancer, in which they assist in AR-driven signaling and are maintained by FOXA1, a transcription factor that mediates cell lineage gene expression in several cell types (29).

**T-UCRs**

Ultraconserved regions in the genome were initially described as stretches of sequence >200 bp long with 100% conservation between humans and rodents but harboring no known gene (59). Because high levels of sequence conservation are hallmarks of exonic sequences in protein-coding genes, ultraconserved regions strongly suggest the presence of either a gene or a regulatory region, such as an enhancer. Subsequently, numerous ultraconserved sequences were found to be transcriptionally active, defining a class of T-UCRs as ncRNAs (18). Many transcripts from T-UCRs are polyadenylated and associated with H3K4me3 at their TSSs, indicating that many are likely IncRNAs according to our definition (60).

Aberrant expression of T-UCRs has been noted in several cancer types, including neuroblastoma (60), leukemia (18), and hepatocellular carcinoma (19). Most notably, one T-UCR gene, termed TUC338, has been shown to promote both cell proliferation and anchorage-independent growth in hepatocellular carcinoma cell lines (19), and TUC338 transcript is localized to the nucleus, suggesting a role in regulation of expression (19). Calin and colleagues (18) further showed that T-UCRs are targets for miRNAs. While T-UCRs remain poorly characterized as a whole, further exploration of the role and mechanism of these IncRNAs will likely elucidate novel aspects of tumor biology.

**FUNCTIONS AND MECHANISMS OF LONG ncRNAs**

Like protein-coding genes, there is considerable variability in the function of long ncRNAs, yet clear themes in the data suggest that many long ncRNAs contribute to associated biologic processes. These processes typically relate to transcriptional regulation or mRNA processing, which is reminiscent of miRNAs and may indicate a similar sequence-based mechanism akin to miRNA binding to seed sequences on target mRNAs. However, unlike miRNAs, long ncRNAs show a wide spectrum of biologic contexts that show greater complexity to their functions.

**Epigenetic Transcriptional Regulation**

The most dominant function explored in IncRNA studies relates to epigenetic regulation of target genes. This typically results in transcriptional repression, and many IncRNAs were first characterized by their repressive functions, including ANRIL, HOTAIR, H19, KCNQ1OT1, and XIST (10, 47, 55). These IncRNAs achieve their repressive function by coupling with histone-modifying or chromatin-remodeling protein complexes.

The most common protein partners of IncRNAs are the PRC1 and PRC2 polycomb repressive complexes. These complexes transfer repressive posttranslational modifications to specific amino acid positions on histone tail proteins, thereby facilitating chromatin compaction and heterochromatin formation in order to enact repression of gene transcription. PRC1 may comprise numerous proteins, including BMI1, RING1, RING2, and Chromobox (CBX) proteins, which act as a multiprotein complex to ubiquitinate histone H2A at
lysine 119 (61). PRC2 is classically composed of EED, SUZ12, and EZH2, the latter of which is a histone methyltransferase enzymatic subunit that trimethylates histone 3 lysine 27 (61). Both EZH2 and BMI1 are upregulated in numerous common solid tumors, leading to tumor progression and aggressiveness (13, 61).

Indeed, ANRIL, HOTAIR, H19, KCNQ1OT1, and XIST have all been linked to the PRC2 complex, and in all except H19, direct binding has been observed between PRC2 proteins and the ncRNA itself (40, 48, 55, 62, 63). Binding of IncRNAs to PRC2 proteins, however, is common and observed for ncRNAs, such as PCAT-1, which do not seem to function through a PRC2-mediated mechanism. It is estimated that nearly 20% of all IncRNAs may bind PRC2 (64), although the biologic meaning of this observation remains unclear. It is possible that PRC2 promiscuously binds IncRNAs in a nonspecific manner. However, if IncRNAs are functioning in a predominantly cis-regulatory mechanism—such as ANRIL, KCNQ1OT1, and XIST—then numerous IncRNAs may bind PRC2 to facilitate local gene expression control throughout the genome. Relatedly, studies of PRC2-DNA-binding properties have shown a putative PRC2-binding motif that includes a GC-rich double hairpin, indicating a structural basis for PRC2-DNA binding in many cases (40).

Similarly, PRC1 proteins, particularly CBX proteins, have been implicated in ncRNA-based biology. For example, ANRIL binds CBX7 in addition to PRC2 proteins, and this interaction with CBX7 recruits PRC1 to the INK4A/ARF locus to mediate transcriptional silencing (47). More broadly, work with mouse polycistronic proteins showed that treatment with RNase abolished CBX7 binding to heterochromatin on a global level, supporting the notion that IncRNAs are critical for PRC1 genomic recruitment (65).

While PRC1 and PRC2 are perhaps the most notable partners of IncRNAs, numerous other epigenetic complexes are implicated in ncRNA-mediated gene regulation. For example, the 3’ domain of HOTAIR contains a binding site for the LSD1/CoREST, a histone deacetylase complex that facilitates gene repression by chromatin remodeling (Fig. 3A) (56). AIR is similarly reported to interact with G9a, an H3K9 histone methyltransferase (66). KCNQ1OT1 has been shown to interact with PRC2 (63), G9a (63), and DNMT1, which methylates CpG dinucleotides in the genome. More rarely, IncRNAs have been observed in the activation of epigenetic complexes. In a recent example, HOTTIP interacted with WDR5 to mediate recruitment of the MLL histone methyltransferase to the distal HoxA locus (52). MLL transfers methyl groups to H3K4me3, thereby generating open chromatin structures that promote gene transcription.

In some cases, the mere act of IncRNA transcription is critical for the recruitment of protein complexes. Studies on H19, KCNQ1OT1, and AIR suggest that transcriptional elongation of these genes is an important component of their function (34, 53, 54). By contrast, other IncRNAs, including HOTTIP as well as many trans-regulatory IncRNAs, do not show this relationship (52). For these IncRNAs, biologic function may be centrally linked to their role as flexible scaffolds. In this model, IncRNAs serve as tethers that rope together multiple protein complexes through a loose arrangement. Supporting this model are the multiple IncRNAs found to bind multiple protein complexes, such as ANRIL (binding PRC1 and PRC2) and HOTAIR (binding PRC2 and LSD1/CoREST) (Fig. 3A).

Enhancer-Associated Long ncRNAs

In addition to facilitating epigenetic changes that impact gene transcription, emerging evidence suggests that some ncRNAs contribute to gene regulation by influencing the activity of gene enhancers. For example, HOTTIP is implicated in chromosomal looping of active enhancers to the distal HoxA locus (52), but knockdown and overexpression of HOTTIP is not sufficient to alter chromosomal confirmations (52). There is also a report of local enhancer-like ncRNAs that typically lack the H3K4me1 enhancer histone signature but possess H3K4me3 and function to potentiate neighbor gene transcription in a manner independent of sequence orientation (67).

A major recent development has been the discovery of eRNAs, which are critical for the proper coordination of enhancer genomic loci with gene expression regulation. Although the mechanism of their action is still unclear, in prostate cancer cells, induction of AR signaling increased eRNA synthesis at AR-regulated gene enhancers, suggesting that eRNAs facilitate active transcription on induction of a signaling pathway (29). Using chromatin conformation assays, Wang and colleagues (29) showed that eRNAs are also important for the establishment of enhancer-promoter genomic proximity by chromosomal looping. Moreover, eRNAs work in conjunction with cell lineage specific transcription factors, such as FOXL1 in prostate cells, thereby creating a highly specialized enhancer network to regulate transcription of genes in individual cell types (Fig. 3B) (29). Future work in this area will likely provide insight into signaling mechanisms important in cancer.

Modulating Tumor Suppressor Activity

The role of many IncRNAs as transcriptional repressors lends itself to inquiry as a mechanism for suppression of tumor suppressor genes. Here, one particular hot spot is the chromosome 9p21 locus, harboring the tumor suppressor genes CDKN2A and CDKN2B, which give rise to multiple unique isoforms, such as p14, p15, and p16, and function as inhibitors of oncogenic cyclin-dependent kinases. Expression of this region is affected by several repressive ncRNAs, such as ANRIL (Fig. 3C, top) and the p15-Antisense RNA, the latter of which also mediates heterochromatin formation through repressive histone modifications and has been observed in leukemias (47, 68).

Several IncRNAs are implicated in the regulation of p53 tumor suppressor signaling. MEG3, a maternally expressed imprinted IncRNA on Chr14q32, has been shown to activate p53 and facilitate p53 signaling, including the enhancement of p53 binding to target gene promoters (69). MEG3 has also been linked to p53 signaling in meningioma (70), and MEG3 overexpression suppresses cell proliferation in meningioma and hepatocellular carcinoma cell lines (70, 71). In human tumors, MEG3 downregulation is widely noted, with frequent hypermethylation of its promoter observed in pituitary tumors (10) and leukemias (72). Taken together, these data implicate MEG3 as a putative tumor suppressor.
A recently described murine lncRNA located near the p21 gene, termed linc-p21, has also emerged as a promising p53-pathway gene. In murine lung, sarcoma, and lymphoma tumors, linc-p21 expression is induced on activation of p53 signaling and represses p53 target genes through a physical interaction with hnRNP-K, a protein that binds the promoters of genes involved in p53 signaling (Fig. 3C, bottom) (73). linc-p21 is further required for proper apoptotic induction (73). These data highlight linc-p21 as a candidate tumor suppressor gene. However, due to sequence differences among species, it is currently unclear whether the human homolog of linc-p21 plays a similarly important role in human tumor development.

Regulation of mRNA Processing and Translation

While many lncRNAs operate by regulating gene transcription, posttranscriptional processing of mRNAs is also critical to gene expression. A primary actor in these processes is the nuclear paraspeckle, a subcellular compartment found in the interchromatin space within a nucleus and characterized by PSP1 protein granules (74). Although nuclear paraspeckle functions are not fully elucidated, this structure is known to be involved in a variety of posttranscriptional activities, including splicing and RNA editing (74). Paraspeckles are postulated to serve as storage sites for mRNA prior to its export to the cytoplasm for translation.

**Figure 3.** Mechanisms of IncRNA function. A, IncRNAs, such as HOTAIR, may serve as a scaffolding base for the coordination of epigenetic or histone-modifying complexes, including Polycomb repressive complexes and LSD1/CoREST. B, IncRNAs transcribed from gene enhancers may facilitate hormone signaling by cooperating with lineage-specific complexes such as FOXA1 and AR. C, IncRNAs may directly affect tumor suppressor signaling either by transcriptional regulation of tumor suppressor genes through epigenetic silencing (e.g., ANRIL, top) or by mediating activation of tumor suppressor target genes (e.g., linc-p21, bottom). D, the MALAT1 IncRNA may be an integral component of the nuclear paraspeckle and may contribute to posttranscriptional processing of mRNAs. E, gene expression regulation may occur through direct IncRNA-mRNA interactions that arise from hybridization of homologous sequences and can serve as a signal for STAU1-mediated degradation of the mRNA. F, RNA molecules, including mRNAs, pseudogenes, and ncRNAs, can serve as molecular sponges for miRNAs. This generates an environment of competitive binding of miRNAs to achieve gene expression control based on the degree of miRNA binding to each transcript. The colored triangles represent different miRNA binding sites in a transcript. CDS, coding sequence.
and one study discovered a paraspeckle-retained, polyadenylated nuclear ncRNA, termed CTN-RNA, that is a counterpart to the protein-coding murine CAT2 (mCAT2) gene (75). CTN-RNA is longer than mCAT2, and under stress conditions, cleavage of CTN-RNA to the mCAT2 coding transcript resulted in increased mCAT2 protein (75).

In cancer, two ncRNAs involved in mRNA splicing and nuclear paraspeckle function, MALAT1 and NEAT1, are over-expressed. MALAT1 and NEAT1 are genomic neighbors on Chr11q13 and both are thought to contribute to gene expression by regulating mRNA splicing, editing, and export (Fig. 3D) (76, 77). MALAT1 may further serve as a precursor to a small 61-bp ncRNA that is generated by RNase P cleavage of the primary MALAT1 transcript and exported into the cytoplasm (78). Although a unique role for MALAT1 in cancer is not yet known, its overexpression in lung cancer predicts for aggressive, metastatic disease (79).

**Regulatory RNA-RNA Interactions**

Recent work on mechanisms of RNA regulation has highlighted a novel role for RNA-RNA interactions between ncRNAs and mRNA sequences. These interactions are conceptually akin to miRNA regulation of mRNAs, because sequence homology between the ncRNA and the mRNA is important to the regulatory process.

This sequence homology may be derived from ancestral repeat elements that contribute sequence to either the untranslated sequences of a protein-coding gene or, less frequently, the coding region itself. For example, STAUI-mediated mRNA decay involves the binding of STAUI, an mRNA degradation protein, to protein-coding mRNAs that interact with IncRNAs containing ancestral Alu repeats. In this model, sequence repeats, typically Alus, in IncRNAs and mRNAs partially hybridize, forming double-stranded RNA complexes that then recruit STAUI to implement RNA degradation (Fig. 3E) (80). A related concept is found with XIST, which contains a conserved repeat sequence, termed RepA, in its first exon. RepA is essential for XIST function, and the RepA sequence is necessary to recruit PRC2 proteins for X-chromosome inactivation (40).

Poliseno and colleagues (81) recently posited another model for mRNA regulation in which they suggested that transcribed pseudogenes serve as a decoy for miRNAs that target the protein-coding mRNA transcripts of their cognate genes. Sequestration of miRNAs by the pseudogene then regulates the gene expression level of the protein-coding mRNA indirectly (Fig. 3F). In addition to pseudogenes, this model more broadly suggests that all long ncRNAs, as well as other protein-coding mRNAs, may function as molecular “sponges” that bind and sequester miRNAs in order to control gene expression indirectly. These researchers showed that pseudogenes of two cancer genes, PTEN and KRAS, may be biologically active, and that PTENP1, a pseudogene of PTEN that competes for mRNA binding sites with PTEN, itself functions as a tumor suppressor in in vitro assays and may be genomically lost in cancer (81). This intriguing hypothesis may shed new light on the functions of ncRNAs, pseudogenes, and even the untranscribed regions of a protein-coding gene.

**Implications of ncRNAs for Cancer Management**

**IncRNA Diagnostic Biomarkers**

For clinical medicine, IncRNAs offer several possible benefits. IncRNAs, such as PCAT1, commonly show restricted tissue-specific and cancer-specific expression patterns (9). This tissue-specific expression distinguishes IncRNAs from miRNAs and protein-coding mRNAs, which are frequently expressed from multiple tissue types. Although the underlying mechanism for IncRNA tissue specificity is unclear, recent studies of chromatin confirmation show tissue-specific patterns, which may affect ncRNA transcription (29, 52). Given this specificity, ncRNAs may be superior biomarkers to many current protein-coding biomarkers, both for tissue-of-origin tests and for cancer diagnostics.

A prominent example is PCA3, an IncRNA that is a prostate-specific gene and markedly overexpressed in prostate cancer. Although the biologic function of PCA3 is unclear, its utility as a biomarker has led to the development of a clinical PCA3 diagnostic assay for prostate cancer, and this test is already being clinically used (82, 83). In this test, PCA3 transcript is detected in urine samples from patients with prostate cancer, which contain prostate cancer cells shed into and passed through the urethra. Thus, monitoring PCA3 does not require invasive procedures (Fig. 4A) (82). The PCA3 test represents the most effective clinical translation of a cancer-associated ncRNA gene, and the rapid timeline of this development—only 10 years between its initial description and a clinical test—suggests that the use of ncRNAs in clinical medicine is only beginning. Noninvasive detection of other aberrantly expressed IncRNAs, such as upregulation of HULC, which occurs in hepatocellular carcinomas, has also been observed in patient blood sera (10); however, other IncRNA-based diagnostics have not been developed for widespread use.

**IncRNA-Based Therapies**

The transition from ncRNA-based diagnostics to ncRNA-based therapies is also showing initial signs of development. Although the implementation of therapies targeting ncRNAs is still remote for clinical oncology, experimental therapeutics employing RNA interference (RNAi) to target miRNAs have been tested in mice, cynomolgus monkeys, and humans (84), as part of a phase I clinical trial for patients with advanced cancer (Fig. 4B). Davis and colleagues (84) found that systemic administration of RNAi-based therapy was able to localize effectively to human tumors and reduce expression of its target gene mRNA and protein. Currently, ongoing clinical trials are further evaluating the safety and efficacy of RNAi-based therapeutics in patients with a variety of diseases, including cancer (85), and these approaches could be adapted to target IncRNA transcripts.

Other studies investigate an intriguing approach that employs modular assembly of small molecules to adapt to aberrant RNA secondary structure motifs in disease (86). This approach could potentially target aberrant ncRNAs, mutant mRNAs, as well as nucleotide triplet-repeat expansions seen in several neurologic diseases (such as Huntington disease). However, most RNA-based research remains in the early stages of development, and the potential for RNAi therapies targeting IncRNAs in cancer is still far from use in oncology clinics.
In the past decade, GWASs have become a mainstream way to identify germline SNPs that may predispose to myriad human diseases. In prostate cancer, more than 20 GWASs have reported 31 SNPs with reproducible allele-frequency changes in patients with prostate cancer compared with those without prostate cancer (87), and these 31 SNPs cluster into 14 genomic loci (87). In principle, profiling of these SNPs could represent an epidemiologic tool to assess patient populations with a high risk of prostate cancer.

Of the 14 genomic loci, the most prominent by far is the “gene desert” region upstream of the cMYC oncogene on chromosome 8q24, which harbors 10 of the 31 reproducible SNPs associated with prostate cancer (Fig. 4C). Several SNPs in the 8q24 region have been studied for their effect on enhancers (88), particularly for enhancers of cMYC (89), and
chromosome looping studies have shown that many regions within 8q24 may physically interact with the genomic position of the cMYC gene (90).

Recently, our identification of PCAT-1 as a novel chr8q24 gene implicated in the pathogenesis of prostate cancer further highlights the importance and complexity of this region (Fig. 4C) (9). Although the relationship between PCAT-1 and the 8q24 SNPs is not clear at this time, this discovery suggests that previously termed “gene deserts” may in fact harbor critical lncRNA genes, and that SNPs found in these regions may affect uncovered aspects of biology. Relatedly, GWAS analyses of atherosclerosis, coronary artery disease, and type 2 diabetes have all highlighted ANRIL on chr9p21 as an ncRNA gene harboring disease-associated SNPs (50).

Clinically, the use of GWAS data may identify patient populations at risk of cancer and may stratify patient disease phenotypes, such as aggressive versus indolent cancer, and patient outcomes (91). SNP profiles may also be used to predict a patient’s response to a given therapy (92). As such, the clinical translation of GWAS data remains an area of interest for cancer epidemiology.

FUTURE DIRECTIONS
Defining the lncRNA Component of the Human Genome

Going forward, it is clear that the systematic identification and annotation of lncRNAs, as well as their expression patterns in human tissues and disease, is important to clarifying the molecular biology underlying cancer. These efforts will be facilitated by large-scale RNA-Seq studies followed by ab initio or de novo sequence data assembly to discover lncRNAs in an unbiased manner (9, 26).

However, it is increasingly appreciated that a number of annotated but uncharacterized transcripts are important lncRNAs; HOTTIP is one such example (52). Similarly, the STAU1-interacting lncRNAs described by Gong and colleagues (80) were also found by screening for annotated transcripts that contained prominent Alu repeats. Although these examples were annotated as noncoding genes, it is also possible that other annotated genes, enumerated in early studies as protein-coding but not studied experimentally, are mislabeled ncRNA genes. These may include the generic “open-reading frame” (ORF) genes (such as LOCxxx or CxxORFx genes) that have not been studied in detail.

Supporting this idea, Dinger et al. (93) recently argued that bioinformatically distinguishing between protein-coding and noncoding genes can be difficult and that traditional computational methods for doing this may have been inadequate in many cases. For example, XIST was initially identified as a protein-coding gene because it has a potential, unused ORF of nearly 300 amino acids (94). Additional complications further include an increasing appreciation of miRNA transcripts that function both by encoding a protein and at the RNA level, which would support miRNA sequestration hypotheses posited by Poliseno and colleagues (81), and of very small ORFs (encoding peptides <10 kDa) (95).

Elucidating the Role of lncRNA Sequence Conservation

In general, most protein-coding exons are highly conserved and most lncRNAs are poorly conserved. This is not always true; T-UCRs are prime examples of conserved ncRNAs. However, the large majority of lncRNAs exhibit substantial sequence divergence among species, and lncRNAs that do show strong conservation frequently exhibit this conservation in only a limited region of the transcript, and not in the remainder of the gene.

This conundrum has sparked many hypotheses, many of which have merit. Small regions of conservation could indicate functional domains of a given ncRNA, such as a binding site for proteins, miRNAs, mRNAs, or genomic DNA. Development of abundant ncRNA species could also suggest evolutionary advancement as species develop. In support of this latter proposition, many researchers have commented that complex mammalian genomes (such as the human genome) have a vastly increased noncoding DNA component of their genome compared with single-celled organisms and nematodes, whereas the complement of protein-coding genes varies less throughout evolutionary time (96).

For lncRNAs, the issue of sequence conservation is paramount. However, it is now well established that poorly conserved lncRNAs can be biologically important, but it is unclear whether these lncRNAs represent species-specific evolutionary traits or whether functional homologs have simply not been found. For example, AIR was initially described in mice in the 1980s, but a human homolog was not identified until 2008 (97).

Moreover, even lncRNAs with relatively high conservation, such as HOTAIR, may have species-specific function. Indeed, a study of murine HOTAIR (mHOTAIR) showed that mHOTAIR did not regulate the HoxD locus and did not recapitulate the functions observed in human cells (98). Other ncRNAs observed in mice, such as linc-p21, also show only limited sequence homology to their human forms and may have divergent functions as well. This may support hypotheses of rapid evolution of lncRNAs during the course of mammalian development. Additionally, this may suggest either that lncRNAs may have functions independent of conserved protein complexes (which have comparatively static functions throughout evolution) or that lncRNAs may adapt to cooperate with different protein complexes in different species.

Determining Somatic Alterations of lncRNAs in Cancer

To date, somatic mutation of lncRNAs in cancer is not well explored. Although numerous lncRNAs display altered expression levels in cancer, it is unclear to what extent cancers specifically target lncRNAs for genomic amplification/deletion, somatic point mutations, or other targeted aberrations.

In several examples, data suggest that lncRNAs may be a target for somatic aberrations in cancer. For example, approximately half of prostate cancers harbor gene fusions of the ETS family transcription factors (ERG, ETV1, ETV4, ETV5), which generally result in the translocation of an androgen-regulated promoter to drive upregulation of the ETS gene (99). One patient was initially found to have an ETV1 translocation to an intergenic androgen-regulated region (100),
which was subsequently found to encode a prostate-specific lncRNA (PCAT-14) (9), thereby creating a gene fusion between the lncRNA and ET1. Similarly, a GASS-BCL6 gene fusion, resulting from a chromosomal translocation and retaining the full coding sequence of BCL6, has been reported in a patient with B-cell lymphoma (101). Finally, Poliseno and colleagues (81) showed that the PTEN pseudogene, PTENP1, is genomically deleted in prostate and colon cancers, leading to aberrant expression levels of these genes.

These initial data suggest that somatic aberrations of lncRNAs do contribute to their dysregulated function in cancer, although most studies to date identify gene expression changes as the primary alteration in lncRNA function. However, the study of mutated lncRNAs in cancer will be an area of high importance in future investigations, because several prominent oncogenes, such as KRAS, show no substantial change in protein expression level in mutated compared with nonmutated cases.

Characterizing RNA Structural Motifs

Just as protein-coding genes harbor specific domains of amino acids that mediate distinct functions (e.g., a kinase domain), RNA molecules have intricate and specific structures. Among the most well-known RNA structures is the stem-loop-stem design of a hairpin, which is integral for miRNA generation (12). RNA structures are also known to be essential for binding to proteins, particularly PRC2 proteins (40). However, global profiles of lncRNA structures are poorly understood. Although it is clear that lncRNA structure is important to lncRNA function, few RNA domains are well characterized. Moreover, it is likely that RNA domains occur at the level of secondary structure, because lncRNA sequences are highly diverse yet may form similar secondary structures following RNA folding (102).

To this end, both computational and experimental advancements are beginning to address these topics. Although numerous computational algorithms have been proposed to predict RNA structures (102), perhaps the most dramatic advance in this area has been the development of RNA-Seq methods to interrogate aspects of RNA structure globally. Recently, Frag-Seq and PARS-Seq have shown the unbiased evaluation of RNA structures by treating RNA samples with positions (103, 104). These RNA fragments are then processed and sequenced to determine the nucleotide sites where RNA transcripts were cleaved, indirectly implying a secondary structure. This area of research promises to yield tremendous insight into the overall mechanics of lncRNA function.

CONCLUSIONS

In the past decade, the rapid discovery of ncRNA species by high-throughput technologies has accelerated current conceptions of transcriptome complexity. Although a biologic understanding of these ncRNAs has proceeded more slowly, increasing recognition of lncRNAs has defined these genes as critical actors in numerous cellular processes. In cancer, dysregulated lncRNA expression characterizes the entire spectrum of disease and aberrant lncRNA function drives cancer through disruption of normal cell processes, typically by facilitating epigenetic repression of downstream target genes. Thus, lncRNAs represent a novel, poorly characterized layer of cancer biology. In the near term, clinical translation of lncRNAs may assist biomarker development in cancer types without robust and specific biomarkers, and in the future, RNA-based therapies may be a viable option for clinical oncology.
The Emergence of lncRNAs in Cancer Biology


The Emergence of lncRNAs in Cancer Biology

REVIEW


100. Tomlins SA, Laxman B, Dhanasekaran SM, Helgeson BE, Cao X, Morris DS, et al. Distinct classes of chromosomal rearrangements...