



Published in final edited form as:

ACS Chem Biol. 2016 August 19; 11(8): 2091–2100. doi:10.1021/acscchembio.6b00366.

## Chemical Tools for Dissecting the Role of lncRNAs in Epigenetic Regulation

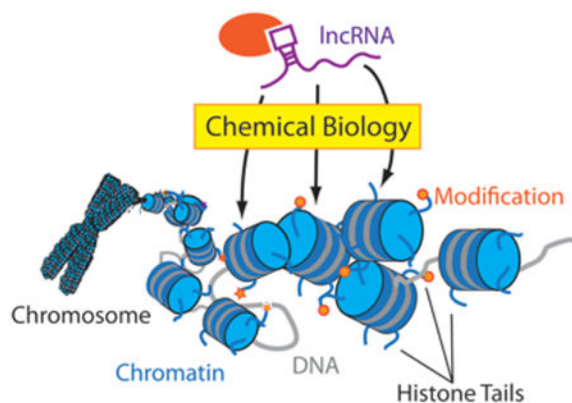
Sarah Nainar, Chao Feng, and Robert C. Spitale\*

Department of Pharmaceutical Sciences, University of California, Irvine. Irvine, California 92697, United States

### Abstract

Proper control and maintenance of gene expression is critical for cellular identity and maintenance. Transcription of RNA from the genome is intimately controlled by post-translational chemical modification of histone tails and DNA. Recent studies have demonstrated that chromatin-remodeling complexes seek out their target genomic loci through the help of noncoding RNA molecules. Within this Review, we will outline how the use of biochemical techniques has shed light on the mechanisms employed by RNA to guide these complexes and therefore control gene expression.

### Graphical abstract



Proper control of gene expression is essential for the cellular response to extracellular cues or stressors and ultimately organismal survival. While DNA content from cell to cell is largely the same, distinct gene sets are turned on or off in concert to orchestrate a systems-level reaction and ensure proper cellular function. To control such gene sets, DNA is organized into a proteinaceous unit termed chromatin. Chromatin structures, which are modulated through DNA looping, nuclear bodies, promoter-enhancer elements, and transcription factories, must be tightly controlled to ensure proper gene expression.<sup>1,2</sup>

\*Corresponding Author: rspitale@uci.edu.

Notes: The authors declare no competing financial interest.

Recent methods have been developed to understand the structure and functional mechanisms that regulate the chromatin state on a global level.<sup>3,4</sup>

Chemical modification of histone proteins is a key mechanism for controlling gene expression. Many histone modifications have been reported to date.<sup>5</sup> The two most abundant modifications—methylation and acetylation—have been widely studied and are well characterized. Their dynamic regulation has been demonstrated to be critical to chromatin maintenance across many species.<sup>6</sup> Acetylation and deacetylation occur on lysine residues near the N-terminal tail of histone proteins.<sup>7</sup> Histone methylation and demethylation have been implicated in both transcriptional activation (H3K4, K36, K79) and silencing (H3K9, H3K27, H4K20).<sup>8</sup> Chromatin-modification enzymes control the deposition and removal of chromatin modifications for modulating RNA expression. Histone acetyl transferases (HATs) and histone deacetylases (HDACs) are responsible for lysine acetylation (Figure 1A). Histone methyltransferases and demethylases control histone methylation status (Figure 1B).<sup>8</sup> Histone acetylation eliminates the positively charged lysines on histone tails, the result of which is decreased interaction with DNA. The DNA is subsequently more loosely wrapped around histones and is available for the binding of transcription factors and proteins that initiate transcription.<sup>1</sup> The opposite is true for histone deacetylation, wherein the restored positive charge on lysine causes greater interaction with DNA and thus condenses chromatin to prevent transcription.<sup>7</sup> Chromatin condensation can also be brought about by methylation via indirect mechanisms. Relaxed, transcriptionally active DNA is referred to as euchromatin, and transcriptionally silent DNA (condensed) is heterochromatin. The dynamic nature of chromatin chemical modification has been explored for many decades. However, because the chemical and structural composition of histones is very similar from gene to gene, the mechanism for loci selection was largely unknown.

Transcription factors are well-known for their ability to bind to specific sequences or structures in genomic DNA.<sup>9</sup> However, many chromatin-modifying complexes do not have proteins that bind to DNA directly.<sup>10</sup> In some cases transcription factors can guide chromatin-modifying complexes. For example, recent studies have recently shown that the Jumanji protein Jarid2 is responsible for recruiting the histone demethylase polycomb repressive complex 2 (PRC2) to its target sites in mouse embryonic stem cells.<sup>11</sup> However, Jarid2 expression is very low in differentiated cells.<sup>12</sup> Therefore, for some time, it was unknown how PRC2 is targeted to chromatin in other cell types. In addition, there are many chromatin-modifying complexes, which target specific loci, without discovered DNA-binding proteins.<sup>13</sup> This represents just one example but is a common gap in our understanding of how chromatin modification is controlled and targeted.

The sequencing of higher-order mammalian transcriptomes has revealed that a substantial fraction of DNA is transcribed to yield many short or long noncoding RNAs (lncRNAs) with limited protein-coding capacity. Long noncoding RNAs are a set of noncoding transcripts that are greater than 200 nt in length. Similar to coding mRNAs (mRNAs), lncRNAs are transcribed by RNA Polymerase II, capped, spliced and polyadenylated.<sup>14</sup> Profiling unique chromatin marks (K4–K36 bivalency) and reconstructing RNA-Seq maps have definitively demonstrated that lncRNAs are unique transcripts.<sup>15</sup> Additional sequencing technologies capturing both the 5′-7-methylguano-sine cap and polyA tail have further enhanced our

understanding of the complexity of lncRNA expression.<sup>16,17</sup> RNA sequencing coupled with protein-coding potential prediction programs have added additional levels of stringency for predicting lncRNAs and separating them from protein-coding mRNAs.<sup>18</sup> Each of these methods have developed robust and complementary maps of lncRNA expression. Comprehensive lncRNA expression analysis in several mouse and human cell types has revealed that lncRNA expression is very unique to cell-type and may even be more specific than the mRNA expression profile.<sup>19</sup> Molecular profiling and genetic screens have identified lncRNAs that play a role in dosage compensation,<sup>20</sup> imprinting,<sup>21</sup> developmental gene expression, and reprogramming of human induced pluripotent stem cells.<sup>22</sup>

Perhaps the first well-characterized lncRNA was X-inactive-specific transcript, or XIST.<sup>23</sup> XIST is a 17 kb participant in X-chromosome inactivation during early development in females.<sup>24</sup> XIST RNA spreads along the X-chromosome *in cis* and recruits different chromatin remodeling complexes, most notably polycomb repressive complex 2 (PRC2), to enforce gene silencing. PRC2 is a chromatin modifying complex consisting of H3K27 methyltransferase subunits EZH2 (Enhancer of Zeste 2 Polycomb Repressive Complex 2 Subunit), SUZ12 (Suppressor of Zeste 12), and EED (Embryonic Ectoderm Development).<sup>25</sup> The observation that XIST interacts *directly* with PRC2 has been challenged, as PRC2 has been demonstrated to have a high level of nonspecific promiscuity, and super-resolution microscopy has revealed a spatial separation of XIST and PRC2 on the X-chromosome.<sup>26–28</sup> Recent findings have also demonstrated that XIST may instead bind the protein SHARP to recruit HDAC3 (histone deacetylase 3) and initiate histone deacetylation and PRC2 recruitment to exclude Pol II across the X chromosome.<sup>29</sup> The molecular mechanism of XIST targeting is an evolving story; nevertheless, XIST has served as a paradigm for understanding the mechanisms by which lncRNAs can control chromatin state.

Another of the first functionally characterized lncRNAs, HOTAIR, was shown to associate with and target PRC2 to distantly located genes.<sup>30</sup> Unlike XIST, which targets gene repression *in cis*, HOTAIR works by silencing distant loci *in trans*. This work therefore provided evidence that lncRNAs may globally affect gene expression by bringing complexes to *repress* gene transcription. Another RNA, HOTTIP, was also shown to bind to the mixed-lineage leukemia (MLL) histone H3K4 methyltransferase complex, bringing MLL to specific sites on the genome to activate transcription.<sup>31</sup> These observations have been extended by others to show that many well-characterized intergenic RNAs are indeed binding to these complexes,<sup>32,33</sup> thereby suggesting that lncRNAs may serve as guides to bring chromatin remodeling complexes to specific loci (outlined in Figure 1C).

lncRNAs have been implicated to control many diverse biological functions, beyond chromatin modification. For example, the lncRNA *NRON* complexes with an importin protein to regulate the subcellular trafficking of NFAT.<sup>34</sup> A *UCL1* antisense lncRNA complexes with the *UCL1* mRNA to regulate protein synthesis of the *UCL1* protein.<sup>35</sup> Alu elements within cytoplasmic lncRNAs can form complementary duplexes with Alu elements in the 3'-UTRs of mRNAs to recruit the Staufen1 complex and enhance mRNA decay.<sup>36</sup> The many functions of lncRNAs, from chromatin remodeling to translation control,

have prompted the need to develop biochemical tools to study lncRNA function within living cells.

There are many questions that remain to be answered in order to fully characterize the content of chromatin-associated RNAs. Where and when do lncRNAs bind to the genome? What are the proteins that interact with lncRNAs, and what is the RNA–protein interface? How does the structure of lncRNAs contribute to their biological function? The recent developments in methods to address these questions have foundations in chemical biology, with chemical reagent design being a critical component to the progress of the field. Within this Review, we discuss the recent utilization of chemical biology and chemical techniques to increase our understanding of how lncRNAs work and target specific chromatin modifying complexes to regulate gene expression.

## Chemical Cross-Linking to Decipher Where and When RNAs Bind to the Genome

The observation that lncRNAs can bind to specific loci on the genome has prompted the development of biochemical methods that can assay their specific locations and the macromolecules they associate with. Many of these methods are rooted in the same principals that were first applied to assay chromatin marks and the localization of DNA-binding proteins and chromatin-modification enzymes. Chromatin immunoprecipitation followed by deep sequencing, or ChIP-Seq,<sup>37</sup> is a common method used to assay the genomic localization of chromatin chemical modifications and proteins. In a ChIP experiment, a protein is cross-linked to DNA using formaldehyde. An antibody is then used to enrich the protein of interest. The associated DNA is then isolated and transformed into a sequencing library. ChIP has since been used as an experimental paradigm for the design of protocols used to isolate regions of the genome associated with an RNA of interest.

A few different protocols have been developed to assay the chromosomal location of noncoding RNAs associated with the regulation of epigenetic processes. Each of these methods begins with a critical chemical cross-linking step that “freezes” the localization of an RNA on the genome (Figure 2, A). The first of these, chromatin isolation by RNA purification or ChIRP,<sup>38</sup> first relies on in-cell glutaraldehyde treatment (Figure 2B; compound **2**). Glutaraldehyde forms a variety of adducts upon reacting with biomolecules, particularly with primary amines, such as lysine.<sup>39</sup> Glutaraldehyde undergoes a Mannich-type reaction to form imines, which tautomerize to enamines rendering the products irreversible. The products of these reactions can be both direct and indirect targets of an RNA.

Capture hybridization analysis of RNA targets (CHART) was an analogous technique developed in parallel to ChIRP.<sup>40</sup> In CHART, formaldehyde is used as the *in vivo* cross-linking chemical reagent (Figure 2B; compound **1**). Extensive formaldehyde treatment (3% for 30 min) is expected to yield both direct and indirect cross-linking products containing RNA. Unlike glutaraldehyde cross-linking, formaldehyde forms an intermediary Schiff-base adduct, which is reversible.<sup>41</sup> Such reversibility may be advantageous, as the cross-linked nucleic acid and proteins can be easily retrieved in their near-native form by hydrolysis.

The most stringent chemical cross-linking protocol developed is associated with the methodology termed RNA antisense purification, or RAP.<sup>42</sup> RAP utilizes disuccinimidyl glutarate, or DSG (Figure 2B; compound 3), along with formaldehyde to cross-link RNA-containing complexes *in vivo*. DSG has two *N*-hydroxysuccinimide esters that are activated for nucleophilic attack with stronger nucleophiles, such as primary amines. The cross-linked products are formed *via* highly stable amides, and they are thought to be a mixture of RNA–protein, RNA–DNA, and DNA–protein. Each of these chemical cross-linking compounds has unique characteristics for their reactivity and stability, and thus they have all found utility in methods that are focused on describing the lncRNA–genome interface.

In all of the methods mentioned above, *in vivo* cross-linking promotes the formation of stable complexes that can preserve both protein and DNA interactions with RNAs. In each protocol (as summarized in Figure 2A), the cells are subsequently lysed. Following lysis, biotin-appended antisense oligos can be used to enrich for an RNA of interest. Once an RNA is enriched, the cross-linked complex can be digested to obtain cross-linked proteins or genomic coordinates. Below, we summarize how such powerful biochemical techniques have increased our understanding of the ways lncRNAs bind to the genome.

The utilization of ChIRP to study RNA genomic binding sites was employed on the long noncoding RNA HOTAIR.<sup>38</sup> HOTAIR associates with the chromatin modifying complex PRC2 to deposit methylation marks through its lysine methyltransferase activity, which leads to chromatin silencing. Incredibly, an unbiased comparison between the sites of HOTAIR binding demonstrated enrichment for genes involved in pattern specification processes, consistent with the model that HOTAIR enforces the epigenomic state of distal and posterior positional identity. Such a strong comparison is consistent with the suggestion that the HOTAIR–chromatin interaction is associated with PRC2 relocalization and gene silencing. Overlap of the HOTAIR ChIRP signal with the genomic occupancy of the PRC2 components (Figure 2C) clearly shows very focal organization of HOTAIR genomic binding. The overlap between PRC2 target genes and HOTAIR occupancy suggested two models: HOTAIR may actively recruit PRC2 to its target genes or HOTAIR is a scaffold that is transported with PRC2. The power of ChIRP was further demonstrated upon EZH2 (a core component of PRC2 that binds to histones) knockdown. The results of this test showed that HOTAIR is able to find its target loci, independent of PRC2. These ChIRP results were the first to show how focal lncRNA binding is, which is highly suggestive of lncRNAs' ability to seek out their target loci. Overall, such hybridization-based techniques demonstrated their power in revealing the intricacies of RNA targeting to chromatin. The additional utility of cross-linker selection has been extended to further increase our mechanistic understanding of how RNAs choose their respective chromatin targets.

The mechanisms for how RNAs choose to target specific chromatin loci is still largely unresolved. Choosing an appropriate cross-linker to trap unique complexes inside the cell can aid in attaining a deeper understanding. As one example, for some time it was hypothesized that intermolecular RNA–RNA interactions can be used by many ncRNAs to achieve their diverse functions (for example, snoRNAs<sup>43</sup>). However, this mechanism had not been proven in the context of genomic targeting. Using this approach was critical for unraveling the RNA–RNA interactions which provide evidence for RNA-mediated targeting

to the genome. Psoralen, and derivatives such as 4'aminomethyltrioxsalen or AMT, can crosslink two opposing uridine nucleobases in double-stranded RNA regions (Figure 2D). Upon light-exposure, AMT generates interstrand cross-links between uridine bases in RNA without reacting with proteins.<sup>44</sup> In a modified RAP protocol (see above), AMT was used to cross-link RNA–RNA interactions inside cells. To showcase the utility of RNA–RNA cross-linking by AMT, two ncRNAs implicated in RNA processing were investigated: U1 small nuclear RNA, a component of the spliceosome, and Malat1, a large ncRNA that localizes to nuclear speckles. U1 snRNA forms base-pair contacts with pre-mRNAs at 5' splice sites. RAP confirmed that U1 directly hybridizes to 5' splice sites and 5' splice site motifs throughout introns. It also revealed that Malat1 interacts with pre-mRNAs indirectly through protein intermediates. Interactions with nascent pre-mRNAs cause U1 and Malat1 to localize proximally to chromatin at active genes, demonstrating that ncRNAs can use RNA–RNA interactions, which come from other RNAs that are being actively transcribed during their activation, to target specific pre-mRNAs and genomic sites. Thus, RAP-RNA (using AMT cross-linking) can accurately and specifically identify RNA–RNA interactions mediated by direct hybridization. As illustrated by this section, there are now several effective methods and techniques that can be used to identify where, and to some extent how, ncRNAs can localize to specific regions of the genome. The next logical steps following such characterization are focused on elucidating the structure of ncRNAs and the RNA–protein interface that controls ncRNA function.

## Using Chemical Probing to Understand the Structure of Lncrnas

Each individual RNA can have unique primary sequences that can give rise to different structural components and motifs. Each of these motifs can perform different biological functions. For example, single-stranded regions often serve as landing pads for proteins.<sup>45</sup> Extended RNA structures have been implicated in RNA-based diseases,<sup>46</sup> and disease-associated mutations have been demonstrated to alter the structural landscape of RNA.<sup>47</sup> Therefore, discerning the function of noncoding RNAs and their control over epigenetic regulation critically relies on a more comprehensive understanding of their structures.

RNA structure mapping has been employed for decades to understand the physical properties of several RNAs. Several reagents have been developed for obtaining high confidence structure maps of RNAs. RNases that are specific for an RNA structural motif are used to obtain low-resolution measurements of RNA secondary structure.<sup>48,49</sup>

As an alternative, structure probing using small molecules has a rich history of providing RNA structure measurements at high resolution. Chemical probing of RNA structure first appeared in the 1980s.<sup>50,51</sup> In these early experiments, dimethylsulfate was used to interrogate the solvent accessibility of guanosine and interrogate the base-pairing characteristics of adenosine and cytosine (Figure 3A). N1-methylation of adenosine and N3-methylation of cytosine can be read out through reverse transcription, from which a corresponding cDNA sequence is created that is truncated at the modification site (explained below).<sup>52,53</sup> These early investigations established the paradigm of using chemical methods to probe RNA structure.



Additional chemical methods have more recently become available. Hydroxyl radical cleavage of solvent exposed residues, or metal catalyzed cleavage, are currently the most thorough methods for determining the secondary and tertiary structure of RNAs. Cleavage results in the formation of hydroxyl radicals that abstract the hydrogen from the C5' - position of the backbone, one product of which is a 3' -phosphate and 5' -aldehyde.<sup>49,54-56</sup>

Selective Hydroxyl Acylation analyzed by Primer Extension, or SHAPE, is the most widely used method of chemical modification to probe RNA structure (Figure 3A). SHAPE takes advantage of two key observations: first, that electrophilic reagents can acylate the 2' - hydroxyl group of RNA nucleotides and, second, that the reactivity of these 2' -hydroxyl groups differs depending on local RNA structure. Nucleotides at flexible positions sample conformations that transiently enhance their nucleophilicity and thus are more highly modified.<sup>57,58</sup> All together, these chemical methods provide an increasingly holistic approach to probing RNA secondary structure.

RNA chemical modifications that present as adduct formation or strand cleavage products are traditionally analyzed by denaturing gel electrophoresis.<sup>59</sup> For adduct formation, reverse transcription is performed to synthesize cDNA molecules. Each cDNA can be identified on a gel and mapped back to the primary sequence, where the intensity of cDNA stops is related to the number of modifications on the RNA at a given position. For RNA cleavage, the RNA is traditionally radiolabeled at the 5' -end. It is then directly loaded into a denaturing gel, and the cut sites can be identified and mapped back to the sequence (Figure 3B). More comprehensive reviews on methods to analyze RNA structure have recently been published.<sup>59</sup>

Interpretation of RNA structure data (from denaturing gels or sequencing) has its roots in the merger of RNA structure predictions and chemical probing experiments. RNA structure predictions originate in careful stability measurements on model RNAs; such measurements have given rise to the prevailing nearest-neighbor model developed by Turner and colleagues.<sup>60-62</sup> To expand this model, free energy calculations for base pairs, helix stacking, mismatches, and even modified nucleosides have been empirically determined through optical melting experiments. A more comprehensive review on this important subject was very recently published.<sup>63</sup> The merger of RNA structure measurements and structure prediction using nearest neighbor parameters can dramatically increase the true positive rates of base pairs and accuracy of secondary structure predictions.<sup>64,65</sup> RNA structure prediction algorithms, such as RNAstructure,<sup>66</sup> UNAFold,<sup>67</sup> or RNAfold,<sup>68</sup> can now be routinely used with chemical probing experiments to create reasonably accurate RNA structure models.

RNA structure probing by use of chemical probes has been utilized to understand the structure of entire viral RNA genomes,<sup>69</sup> to characterize RNA-protein interactions,<sup>70</sup> and to characterize RNA folding and structural transitions.<sup>57</sup> Further, the information gained from studying RNA folding transitions with chemical probing<sup>71</sup> and RNA-protein interactions inside cells for well-characterized RNAs<sup>72</sup> set a strong experimental outline for investigations into the structures of lncRNAs and their complexes.

The sites of RNA–protein interaction, which are critical for epigenetic regulation, have recently been elucidated through RNA structure probing techniques. Chemical probing was used to evaluate the complete secondary structure<sup>73</sup> of all 2148 nucleotides of HOTAIR, a long-noncoding RNA involved in the targeted recruitment of the polycomb repressive complex (PRC2) to silence gene expression.<sup>30</sup> Refolded HOTAIR RNA was subjected to RNA structure probing using SHAPE, DMS, and terbium-catalyzed strand cleavage.<sup>74</sup> These results revealed that HOTAIR has several isolated regions that form into modular structured domains. Structure probing comparisons of individual domains to the full-length RNA revealed similar profiles, further demonstrating the stability of modular domains in RNA structure. Overall, it was demonstrated that HOTAIR is highly structured, as there are 56 helical segments, 38 terminal loops, 34 internal loops, and 19 junction regions. Furthermore, greater than 50% of the nucleotides in HOTAIR are involved in base pairing.

The molecular interface between HOTAIR and PRC2 has yet to be fully determined. Previous deletion experiments on HOTAIR narrowed its protein interaction sites down to two modular regions: (1) a 300-nt-long region at the 5′ end of HOTAIR (nucleotides 1–300, HOTAIR300), which binds PRC2, and (2) a 646-nt-long region at the 3′ end of HOTAIR, which binds LSD1/CoREST/REST.<sup>75</sup> Analysis of primary sequence conservation and covariation analyses revealed that the regions associated with putative PRC2 binding are highly structured and also have highly conserved predicted secondary structures. A parallel analysis of the HOTAIR–PRC2 interface further demonstrated that the PRC2 interaction site is composed of structured motifs that are critical for PRC2 recognition.<sup>76</sup> These results demonstrate the overall utility of chemical probing to interrogate the structure and function of ncRNAs and proteins critical for epigenetic regulation.

Chemical probing of RNA structure has also been used to reveal the mechanism by which RNAs control the epigenetic state via recruitment of chromatin-modifying complexes. As one example, structure probing and RNA structure conservation analyses revealed the critical nature of RNA structure elements responsible for dosage compensation in *Drosophila*.<sup>77,78</sup> Dosage compensation is an epigenetic phenomenon wherein proteins and lncRNAs associate for the purpose of transcriptional upregulation of genes on the male X chromosome. In order for this to occur, two RNAs called roX1 and roX2 control X chromosome-wide histone acetylation at H4 lysine 16. Dosage compensation complex (DCC) contains the male-specific lethal proteins MSL1–MSL3, a helicase MLE, and the histone acetyltransferase.<sup>79,80</sup> A key missing piece in this epigenetic puzzle was how the structures of roX1 and roX2 RNAs were precisely formed and remodeled to control the nucleation of the dosage compensation complex.

Chromatography and immunopurification analysis demonstrated that MLE and the MSL proteins form direct interactions with roX RNAs. As such, MLE and MSL RNA binding sites were then characterized using individual-nucleotide resolution cross-linking and immunoprecipitation (iCLIP, discussed above). This analysis revealed that the MLE RNA helicase and MSL2 ubiquitin ligase bind to evolutionarily conserved domains containing tandem stem-loops in roX1 and roX2 RNAs *in vivo*. These sites are consistent with conserved elements known as roX boxes, which could play an important role in dosage compensation.<sup>81</sup> Concurrent structural analysis by SHAPE probing demonstrated that MLE



binding sites are embedded into long stretches of sequence that are less structured and might be functionally irrelevant. Overall structure models revealed a similar architecture between both roX RNAs: stable helices formed by paired roX box motifs that are strung together by flexible single-stranded linkers. Additionally, analysis between structure probing and structure conservation revealed that structural flexibility is inversely correlated with conservation. This highly suggests that the double stranded regions of the RNA are structurally important (see example in Figure 3C). The structural conservation of roX box motifs was further underscored by demonstrating that different roX RNA domains have overlapping function; only combinatorial mutations in the tandem stem-loops result in a severe loss of dosage compensation and consequently male-specific lethality. This analysis serves as the basis for a model where repetitive structural motifs in lncRNAs could provide plasticity during multiprotein complex assemblies to ensure efficient targeting in *cis* or in *trans* along chromosomes. This analysis revealed that a combination of chemical probing and biochemical methods to interrogate the RNA–protein interface could reveal the mechanism of epigenetic regulation by noncoding RNAs.

The methods for measuring RNA structure with small molecules have also been extended to incorporate transcriptome-spanning technologies, such as RNA sequencing. DMS and SHAPE have been used in this capacity to provide transcriptome-wide measurements of RNA structure.<sup>82–84</sup> These efforts have revealed the architecture of RNAs within cells and how RNA structure elements can influence post-transcriptional interactions, such as protein binding and methylation. Furthermore, more focused sequencing-based methods have been already used to reconstruct the entire secondary structure of the XIST noncoding RNA inside cells.<sup>85</sup> As such, the marrying of chemical and transcriptomic approaches will undoubtedly reveal the even more fascinating roles that RNA structure plays to regulate the chromatin state.

## Using Chemical Methods to Characterize the RNA–Protein Interface

Within cells, RNA molecules are rarely without a protein counterpart. RNA–protein interactions are critical for all aspects of RNA biology, from transcription to export, localization, and decay.<sup>86</sup> The prevailing model in the field is that chromatin-associated RNAs are intimately tied to proteinacious chromatin remodeling complexes. Therefore, gaining molecular insights of the RNA–protein interface is critical for a holistic understanding of RNA-based mechanisms for controlling gene expression.

Perhaps the most widely used method for analyzing RNA–protein interactions is RNA immunopurification, or RIP.<sup>87</sup> In RIP, formaldehyde cross-linking is used to freeze RNA–protein complexes. The complex is then purified by utilizing an antibody either against the native protein sequence or a protein containing tag. qRT-PCR or RNA sequencing then identifies the associated RNAs. Formaldehyde cross-linking can capture RNA–protein or protein–protein interactions that are not at the RNA–protein interface. As such, the associated RNAs may not be in direct contact with a protein of interest enriched through purification. This limitation can result in false-positives or an unnatural RNA-protein interaction that forms post lysis.<sup>88,89</sup> To overcome these methods, more precise assays have been developed to study the RNA–protein interface with high resolution.

Cross-linking and immunoprecipitation or CLIP of *in vivo* RNA targets of RNA-binding proteins is a recently developed method that takes advantage of chemical cross-linking to identify the sites of protein binding on RNA.<sup>90</sup> CLIP uses 254 nm ultraviolet light to induce excitation of the RNA nucleobases. This causes the formation of covalent bonds between proteins and nucleic acids that are in close proximity. The RNA–protein interface is then amenable to high stringency washes and denaturing gel electrophoresis. Isolated cross-link sites can then be cloned and identified by deep sequencing, where the actual binding site can be identified (outlined in Figure 4A). Although powerful, a major setback in CLIP experiments is the reliance upon the low-efficiency cross-linking between RNA and proteins, which is estimated to be <1%.

To overcome the limitations of CLIP, an alternative method termed Photoactivatable-Ribonucleoside-Enhanced Cross-linking and Immunoprecipitation or PAR-CLIP was developed. PAR-CLIP leverages the ability of thiol-containing nucleosides to selectively excite with 365 nm light.<sup>91</sup> Nascent RNA transcripts can be labeled with modified nucleosides by treating cells with photoreactive ribonucleoside analogs, such as 4-thiouridine (4-SU; Figure 4B) and 6-thioguanosine (6-SG), which will nascently incorporate. Irradiation of the cells by 365 nm light induces efficient cross-linking of RNAs containing photoreactive nucleoside RNAs with electron-rich amino acids that are closely associated (Figure 4B). For some RNA-binding proteins compared head-to-head, UV cross-linking efficiency has been demonstrated to increase from <1% (with uridine) to in some cases near quantitative.<sup>91</sup> One potential drawback to PAR-CLIP is the introduction of 4-SU into cellular RNA; however, 4-SU has been demonstrated to be robustly incorporated into many different cell types *in vitro*.<sup>91–93</sup>

PAR-CLIP was recently used to characterize the lncRNA-PRC2 interface.<sup>94</sup> PRC2 has histone methyltransferase activity and primarily trimethylates lysine 27 of histone H3 (i.e., H3K27me3), a mark of transcriptionally silent chromatin.<sup>95</sup> The biochemical function of PRC2 had been known for some time; however, how PRC2 is coordinated to the genome to locate methylation target sites was still not fully resolved. Using PAR-CLIP, it was revealed that PRC2 binds with low occupancy to a majority of promoters. PRC2 was demonstrated to bind to the 5′-end of nascent RNA transcribed at or near the promoter sites of PRC2 enrichment. As such, PAR-CLIP was instrumental for establishing the model in which PRC2 senses nascent RNA transcription and anchors itself at promoters, prompting methylation at those genetic loci. The numerous studies utilizing the methods of CLIP and PAR-CLIP have paved the way to an increased understanding of how RNAs interact with protein complexes and guide them to particular loci on the genome.

Additional tools continue to be developed with the hopes of expanding the methodologies used to probe and identify RNA-binding proteins. This is especially important in the context of RNA–protein interactions that are highly transient. Recent studies have demonstrated that many of the chromatin-associated RNAs are in complexes that act in transit and thus have very short residence time on the genome while complexed. As such, for furthered understanding of chromatin-associated RNAs and the proteins they associate with *in vivo*, there is a need to develop novel cross-linking methods that have higher efficiencies than CLIP or PAR-CLIP.

One such method was recently developed, and it relies on ligand-induced chemical cross-linking. *In vivo* proximity labeling (IPL) technology employs an affinity tag, combined with a photoactivatable probe to label RNAs nearby in space (length limited by the linker distance from BSA) to a protein of interest.<sup>96</sup> One side of the ligand probe is appended with a biotin, the other with an aryl azide (Figure 4C). Aryl azides are especially powerful for cross-linking with RNA due to their high efficiency. When an aryl azide is exposed to UV light (250 to 350 nm), it forms a nitrene group that can initiate addition reactions with double bonds, insertion into C–H and N–H sites, or subsequent ring expansion to react with a nucleophile (e.g., primary amines). IPL cross-linking relies on tagging of nearby RNAs to the BSA-tagged protein (Figure 4D). The power of IPL was showcased when it was used to identify RNAs associated with the SNRNP70 protein. RNA sequencing of tagged RNAs was used to identify the spliceosomal U1 small nuclear RNA (snRNA) but was unable to detect other highly abundant 5S rRNA and U2 small nuclear RNA (snRNA), both of which are known not to be associated with the SNRNP70 protein. As such, IPL is a novel alternative for utilizing chemical probes to identify protein–RNA interactions inside living cells.

In summation, the number of methods that can be utilized to study RNA–protein interactions continues to grow. This underscores the expanse of their utility in characterizing the RNA–protein interface inside living cells.

## Concluding Remarks

Studies focused on RNA biology are undergoing a renaissance. Characterizing their structure and the proteins they interact with was key to further understanding their function. We suspect that it is due in large part to the plethora of chemical tools used to functionally characterize lncRNAs, and their interaction with chromatin will move at a much faster pace. In addition, many other types of RNA are associated with chromatin. For example, antisense RNAs have been demonstrated to interact with chromatin complexes and control gene expression and even RNA decay.<sup>97,98</sup> Furthermore, small RNAs have been shown to form duplex structures with nascent RNA to guide protein complexes for chromatin modification.<sup>99</sup> The tools and techniques described herein have demonstrated great utility in furthering our understanding of the many types of RNAs controlling chromatin and the mechanisms used. Moreover, further development and optimization of biological protocols based on the principals of chemistry will immensely aid in this effort.

## Acknowledgments

We thank members of the Spitale lab for critical reading of the manuscript. S.N. is supported as a Vertex Fellow. RNA research in the Spitale lab is supported by startup funds through the University of California, Irvine and the National Institutes of Health Director's New Innovator Award (1DP2GM119164).

## References

1. Chen T, Dent SY. Chromatin modifiers and remodellers: regulators of cellular differentiation. *Nat Rev Genet.* 2014; 15:93–106. [PubMed: 24366184]
2. Venkatesh S, Workman JL. Histone exchange, chromatin structure and the regulation of transcription. *Nat Rev Mol Cell Biol.* 2015; 16:178–189. [PubMed: 25650798]

3. Ho JW, Bishop E, Karchenko PV, Negre N, White KP, Park PJ. ChIP-chip versus ChIP-seq: lessons for experimental design and data analysis. *BMC Genomics*. 2011; 12:134. [PubMed: 21356108]
4. Belton JM, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods*. 2012; 58:268–276. [PubMed: 22652625]
5. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell Res*. 2011; 21:381–395. [PubMed: 21321607]
6. Mersfelder EL, Parthun MR. The tale beyond the tail: histone core domain modifications and the regulation of chromatin structure. *Nucleic Acids Res*. 2006; 34:2653–2662. [PubMed: 16714444]
7. Verdin E, Ott M. 50 years of protein acetylation: from gene regulation to epigenetics, metabolism and beyond. *Nat Rev Mol Cell Biol*. 2015; 16:258–264. [PubMed: 25549891]
8. Greer EL, Shi Y. Histone methylation: a dynamic mark in health, disease and inheritance. *Nat Rev Genet*. 2012; 13:343–357. [PubMed: 22473383]
9. Spitz F, Furlong EE. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet*. 2012; 13:613–626. [PubMed: 22868264]
10. Khalil AM, Rinn JL. RNA-protein interactions in human health and disease. *Semin Cell Dev Biol*. 2011; 22:359–365. [PubMed: 21333748]
11. Peng JC, Valouev A, Swigut T, Zhang J, Zhao Y, Sidow A, Wysocka J. Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell*. 2009; 139:1290–1302. [PubMed: 20064375]
12. Pethe P, Nagvenkar P, Bhartiya D. Polycomb group protein expression during differentiation of human embryonic stem cells into pancreatic lineage in vitro. *BMC Cell Biol*. 2014; 15:18. [PubMed: 24885493]
13. Hassan AH, Neely KE, Vignali M, Reese JC, Workman JL. Promoter targeting of chromatin-modifying complexes. *Front Biosci, Landmark Ed*. 2001; 6:D1054–1064.
14. Wilusz JE, Sunwoo H, Spector DL. Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev*. 2009; 23:1494–1504. [PubMed: 19571179]
15. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*. 2009; 458:223–227. [PubMed: 19182780]
16. Kawaji H, Lizio M, Itoh M, Kanamori-Katayama M, Kaiho A, Nishiyori-Sueki H, Shin JW, Kojima-Ishiyama M, Kawano M, Murata M, Ninomiya-Fukuda N, Ishikawa-Kato S, Nagao-Sato S, Noma S, Hayashizaki Y, Forrest AR, Carninci P, Consortium F. Comparison of CAGE and RNA-seq transcriptome profiling using clonally amplified and single-molecule next-generation sequencing. *Genome Res*. 2014; 24:708–717. [PubMed: 24676093]
17. Beck AH, Weng Z, Witten DM, Zhu S, Foley JW, Lacroute P, Smith CL, Tibshirani R, van de Rijn M, Sidow A, West RB. 3'-end sequencing for expression quantification (3SEQ) from archival tumor samples. *PLoS One*. 2010; 5:e8768. [PubMed: 20098735]
18. Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, Barrette TR, Prensner JR, Evans JR, Zhao S, Poliakov A, Cao X, Dhanasekaran SM, Wu YM, Robinson DR, Beer DG, Feng FY, Iyer HK, Chinnaiyan AM. The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet*. 2015; 47:199–208. [PubMed: 25599403]
19. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev*. 2011; 25:1915–1927. [PubMed: 21890647]
20. Quinn JJ, Zhang QC, Georgiev P, Ilik IA, Akhtar A, Chang HY. Rapid evolutionary turnover underlies conserved lncRNA-genome interactions. *Genes Dev*. 2016; 30:191–207. [PubMed: 26773003]
21. Kanduri C. Long noncoding RNAs: Lessons from genomic imprinting. *Biochim Biophys Acta, Gene Regul Mech*. 2016; 1859:102–111.
22. Flynn RA, Chang HY. Long noncoding RNAs in cell-fate programming and reprogramming. *Cell stem cell*. 2014; 14:752–761. [PubMed: 24905165]

23. Brockdorff N, Ashworth A, Kay GF, McCabe VM, Norris DP, Cooper PJ, Swift S, Rastan S. The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell*. 1992; 71:515–526. [PubMed: 1423610]
24. Clemson CM, McNeil JA, Willard HF, Lawrence JB. XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. *J Cell Biol*. 1996; 132:259–275. [PubMed: 8636206]
25. Margueron R, Reinberg D. The Polycomb complex PRC2 and its mark in life. *Nature*. 2011; 469:343–349. [PubMed: 21248841]
26. Davidovich C, Cech TR. The recruitment of chromatin modifiers by long noncoding RNAs: lessons from PRC2. *RNA*. 2015; 21:2007–2022. [PubMed: 26574518]
27. Davidovich C, Wang X, Cifuentes-Rojas C, Goodrich KJ, Gooding AR, Lee JT, Cech TR. Toward a consensus on the binding specificity and promiscuity of PRC2 for RNA. *Mol Cell*. 2015; 57:552–558. [PubMed: 25601759]
28. Cerase A, Smeets D, Tang YA, Gdula M, Kraus F, Spivakov M, Moindrot B, Leleu M, Tattermusch A, Demmerle J, Nesterova TB, Green C, Otte AP, Schermelleh L, Brockdorff N. Spatial separation of Xist RNA and polycomb proteins revealed by superresolution microscopy. *Proc Natl Acad Sci U S A*. 2014; 111:2235–2240. [PubMed: 24469834]
29. McHugh CA, Chen CK, Chow A, Surka CF, Tran C, McDonel P, Pandya-Jones A, Blanco M, Burghard C, Moradian A, Sweredoski MJ, Shishkin AA, Su J, Lander ES, Hess S, Plath K, Guttman M. The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature*. 2015; 521:232–236. [PubMed: 25915022]
30. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Bruggmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*. 2007; 129:1311–1323. [PubMed: 17604720]
31. Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA, Wysocka J, Lei M, Dekker J, Helms JA, Chang HY. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature*. 2011; 472:120–124. [PubMed: 21423168]
32. Guttman M, Rinn JL. Modular regulatory principles of large non-coding RNAs. *Nature*. 2012; 482:339–346. [PubMed: 22337053]
33. Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A, Regev A, Lander ES, Rinn JL. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci U S A*. 2009; 106:11667–11672. [PubMed: 19571010]
34. Willingham AT, Orth AP, Batalov S, Peters EC, Wen BG, Aza-Blanc P, Hogenesch JB, Schultz PG. A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. *Science*. 2005; 309:1570–1573. [PubMed: 16141075]
35. Carrieri C, Cimatti L, Biagioli M, Beugnet A, Zucchelli S, Fedele S, Pesce E, Ferrer I, Collavin L, Santoro C, Forrest AR, Carninci P, Biffo S, Stupka E, Gustincich S. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature*. 2012; 491:454–457. [PubMed: 23064229]
36. Gong C, Maquat LE. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3'UTRs via Alu elements. *Nature*. 2011; 470:284–288. [PubMed: 21307942]
37. Park PJ. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet*. 2009; 10:669–680. [PubMed: 19736561]
38. Chu C, Qu K, Zhong FL, Artandi SE, Chang HY. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol Cell*. 2011; 44:667–678. [PubMed: 21963238]
39. Migneault I, Dartiguenave C, Bertrand MJ, Waldron KC. Glutaraldehyde: behavior in aqueous solution, reaction with proteins, and application to enzyme crosslinking. *Biotechniques*. 2004; 37(790-796):798–802.
40. Simon MD, Wang CI, Kharchenko PV, West JA, Chapman BA, Alekseyenko AA, Borowsky ML, Kuroda MI, Kingston RE. The genomic binding sites of a noncoding RNA. *Proc Natl Acad Sci U S A*. 2011; 108:20497–20502. [PubMed: 22143764]

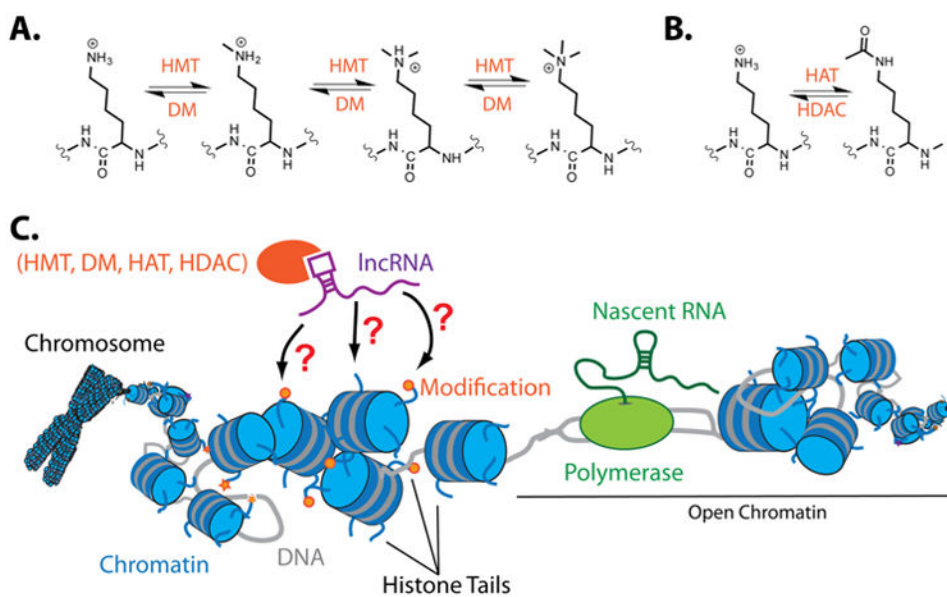


41. Thavarajah R, Mudimbaimannar VK, Elizabeth J, Rao UK, Ranganathan K. Chemical and physical basics of routine formaldehyde fixation. *J Oral Maxillofac Pathol*. 2012; 16:400–405. [PubMed: 23248474]
42. Engreitz JM, Pandya-Jones A, McDonel P, Shishkin A, Sirokman K, Surka C, Kadri S, Xing J, Goren A, Lander ES, Plath K, Guttman M. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science*. 2013; 341:1237973. [PubMed: 23828888]
43. Dieci G, Preti M, Montanini B. Eukaryotic snoRNAs: a paradigm for gene expression flexibility. *Genomics*. 2009; 94:83–88. [PubMed: 19446021]
44. Lustig Y, Wachtel C, Safro M, Liu L, Michaeli S. ‘RNA walk’ a novel approach to study RNA-RNA interactions between a small RNA and its target. *Nucleic Acids Res*. 2010; 38:e5. [PubMed: 19854950]
45. Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, Gueroussov S, Albu M, Zheng H, Yang A, Na H, Irimia M, Matzat LH, Dale RK, Smith SA, Yarosh CA, Kelly SM, Nabet B, Mecnas D, Li W, Laishram RS, Qiao M, Lipshitz HD, Piano F, Corbett AH, Carstens RP, Frey BJ, Anderson RA, Lynch KW, Penalva LO, Lei EP, Fraser AG, Blencowe BJ, Morris QD, Hughes TR. A compendium of RNA-binding motifs for decoding gene regulation. *Nature*. 2013; 499:172–177. [PubMed: 23846655]
46. Bernat V, Disney MD. RNA Structures as Mediators of Neurological Diseases and as Drug Targets. *Neuron*. 2015; 87:28–46. [PubMed: 26139368]
47. Halvorsen M, Martin JS, Broadaway S, Laederach A. Disease-associated mutations that alter the RNA structural ensemble. *PLoS Genet*. 2010; 6:e1001074. [PubMed: 20808897]
48. Trang P, Hsu AW, Liu F. Nuclease footprint analyses of the interactions between RNase P ribozyme and a model mRNA substrate. *Nucleic Acids Res*. 1999; 27:4590–4597. [PubMed: 10556315]
49. Tijerina P, Mohr S, Russell R. DMS footprinting of structured RNAs and RNA-protein complexes. *Nat Protoc*. 2007; 2:2608–2623. [PubMed: 17948004]
50. Peattie DA, Gilbert W. Chemical probes for higher-order structure in RNA. *Proc Natl Acad Sci U S A*. 1980; 77:4679–4682. [PubMed: 6159633]
51. Ehresmann C, Baudin F, Mougél M, Romby P, Ebel JP, Ehresmann B. Probing the structure of RNAs in solution. *Nucleic Acids Res*. 1987; 15:9109–9128. [PubMed: 2446263]
52. Stern S, Moazed D, Noller HF. Structural analysis of RNA using chemical and enzymatic probing monitored by primer extension. *Methods Enzymol*. 1988; 164:481–489. [PubMed: 2468070]
53. Been MD, Barford ET, Burke JM, Price JV, Tanner NK, Zaug AJ, Cech TR. Structures involved in Tetrahymena rRNA self-splicing and RNA enzyme activity. *Cold Spring Harbor Symp Quant Biol*. 1987; 52:147–157. [PubMed: 3454258]
54. Latham JA, Cech TR. Defining the inside and outside of a catalytic RNA molecule. *Science*. 1989; 245:276–282. [PubMed: 2501870]
55. Pan T. Probing RNA structure by lead cleavage. *Current Protocols in Nucleic Acid Chemistry*. 2000:6.3.1.
56. Ingle S, Azad RN, Jain SS, Tullius TD. Chemical probing of RNA with the hydroxyl radical at single-atom resolution. *Nucleic Acids Res*. 2014; 42:12758–12767. [PubMed: 25313156]
57. Wilkinson KA, Merino EJ, Weeks KM. RNA SHAPE chemistry reveals nonhierarchical interactions dominate equilibrium structural transitions in tRNA(Asp) transcripts. *J Am Chem Soc*. 2005; 127:4659–4667. [PubMed: 15796531]
58. Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM. RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc*. 2005; 127:4223–4231. [PubMed: 15783204]
59. Kubota M, Tran C, Spitale RC. Progress and challenges for chemical probing of RNA structure inside living cells. *Nat Chem Biol*. 2015; 11:933–941. [PubMed: 26575240]
60. Turner DH, Sugimoto N, Jaeger JA, Longfellow CE, Freier SM, Kierzek R. Improved parameters for prediction of RNA structure. *Cold Spring Harbor Symp Quant Biol*. 1987; 52:123–133. [PubMed: 2456874]

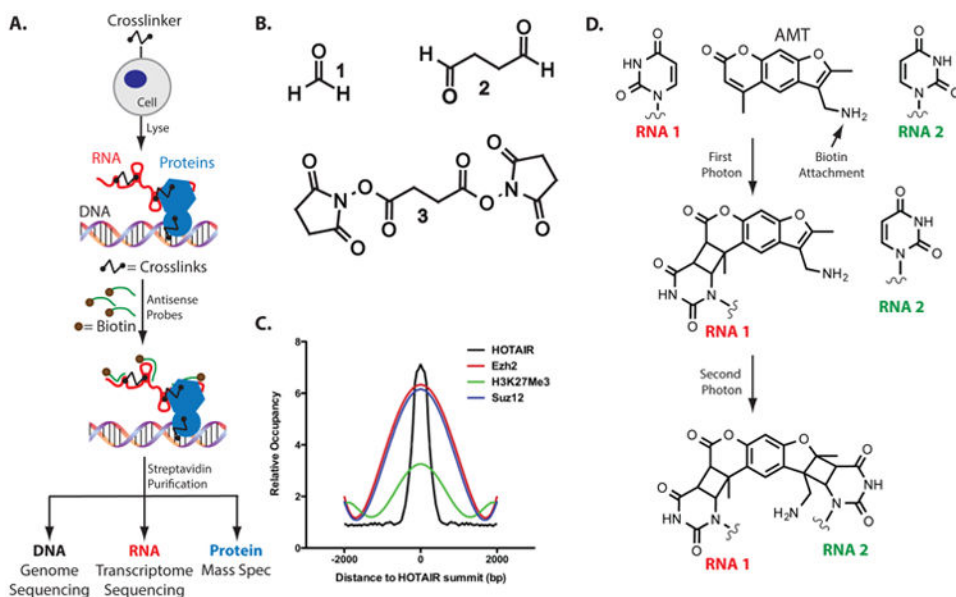


61. Jaeger JA, Turner DH, Zuker M. Predicting optimal and suboptimal secondary structure for RNA. *Methods Enzymol.* 1990; 183:281–306. [PubMed: 1690335]
62. Serra MJ, Turner DH. Predicting thermodynamic properties of RNA. *Methods Enzymol.* 1995; 259:242–261. [PubMed: 8538457]
63. Andronescu M, Condon A, Turner DH, Mathews DH. The determination of RNA folding nearest neighbor parameters. *Methods Mol Biol.* 2014; 1097:45–70. [PubMed: 24639154]
64. Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci U S A.* 2004; 101:7287–7292. [PubMed: 15123812]
65. Deigan KE, Li TW, Mathews DH, Weeks KM. Accurate SHAPE-directed RNA structure determination. *Proc Natl Acad Sci U S A.* 2009; 106:97–102. [PubMed: 19109441]
66. Mathews DH. RNA Secondary Structure Analysis Using RNAstructure. *Current Protocols in Bioinformatics.* 2014; 46:12 16 11–12 16 25.
67. Markham NR, Zuker M. UNAFold: software for nucleic acid folding and hybridization. *Methods in molecular biology.* 2008; 453:3–31. [PubMed: 18712296]
68. Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL. The Vienna RNA websuite. *Nucleic Acids Res.* 2008; 36:W70–74. [PubMed: 18424795]
69. Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW Jr, Swanstrom R, Burch CL, Weeks KM. Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature.* 2009; 460:711–716. [PubMed: 19661910]
70. McGinnis JL, Duncan CD, Weeks KM. High-throughput SHAPE and hydroxyl radical analysis of RNA structure and ribonucleoprotein assembly. *Methods Enzymol.* 2009; 468:67–89. [PubMed: 20946765]
71. Banerjee AR, Jaeger JA, Turner DH. Thermal unfolding of a group I ribozyme: the low-temperature transition is primarily disruption of tertiary structure. *Biochemistry.* 1993; 32:153–163. [PubMed: 8418835]
72. Clatterbuck Soper SF, Dator RP, Limbach PA, Woodson SA. In vivo X-ray footprinting of pre-30S ribosomes reveals chaperone-dependent remodeling of late assembly intermediates. *Mol Cell.* 2013; 52:506–516. [PubMed: 24207057]
73. Somarowthu S, Legiewicz M, Chillon I, Marcia M, Liu F, Pyle AM. HOTAIR forms an intricate and modular secondary structure. *Mol Cell.* 2015; 58:353–361. [PubMed: 25866246]
74. Somarowthu S, Legiewicz M, Chillon I, Marcia M, Liu F, Pyle AM. HOTAIR forms an intricate and modular secondary structure. *Mol Cell.* 2015; 58:353–361. [PubMed: 25866246]
75. Tsai MC, Manor O, Wan Y, Mosammapparast N, Wang JK, Lan F, Shi Y, Segal E, Chang HY. Long noncoding RNA as modular scaffold of histone modification complexes. *Science.* 2010; 329:689–693. [PubMed: 20616235]
76. Wu L, Murat P, Matak-Vinkovic D, Murrell A, Balasubramanian S. Binding interactions between long noncoding RNA HOTAIR and PRC2 proteins. *Biochemistry.* 2013; 52:9519–9527. [PubMed: 24320048]
77. Ilik IA, Quinn JJ, Georgiev P, Tavares-Cadete F, Maticzka D, Toscano S, Wan Y, Spitale RC, Luscombe N, Backofen R, Chang HY, Akhtar A. Tandem stem-loops in roX RNAs act together to mediate X chromosome dosage compensation in *Drosophila*. *Mol Cell.* 2013; 51:156–173. [PubMed: 23870142]
78. Maenner S, Muller M, Frohlich J, Langer D, Becker PB. ATP-dependent roX RNA remodeling by the helicase maleless enables specific association of MSL proteins. *Mol Cell.* 2013; 51:174–184. [PubMed: 23870143]
79. Amrein H, Axel R. Genes expressed in neurons of adult male *Drosophila*. *Cell.* 1997; 88:459–469. [PubMed: 9038337]
80. Meller VH, Wu KH, Roman G, Kuroda MI, Davis RL. roX1 RNA paints the X chromosome of male *Drosophila* and is regulated by the dosage compensation system. *Cell.* 1997; 88:445–457. [PubMed: 9038336]
81. Hallacli E, Lipp M, Georgiev P, Spielman C, Cusack S, Akhtar A, Kadlec J. Msl1-mediated dimerization of the dosage compensation complex is essential for male X-chromosome regulation in *Drosophila*. *Mol Cell.* 2012; 48:587–600. [PubMed: 23084835]

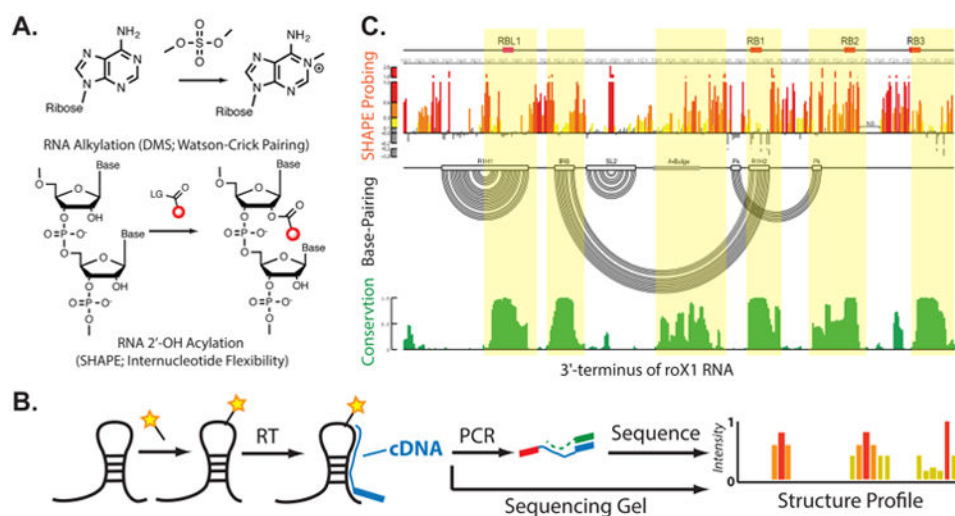
82. Spitale RC, Flynn RA, Zhang QC, Crisalli P, Lee B, Jung JW, Kuchelmeister HY, Batista PJ, Torre EA, Kool ET, Chang HY. Structural imprints in vivo decode RNA regulatory mechanisms. *Nature*. 2015; 519:486–490. [PubMed: 25799993]
83. Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC, Assmann SM. In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature*. 2014; 505:696–700. [PubMed: 24270811]
84. Rouskin S, Zubradt M, Washietl S, Kellis M, Weissman JS. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature*. 2014; 505:701–705. [PubMed: 24336214]
85. Fang R, Moss WN, Rutenberg-Schoenberg M, Simon MD. Probing Xist RNA Structure in Cells Using Targeted Structure-Seq. *PLoS Genet*. 2015; 11:e1005668. [PubMed: 26646615]
86. Lunde BM, Moore C, Varani G. RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol*. 2007; 8:479–490. [PubMed: 17473849]
87. Galgano A, Gerber AP. RNA-binding protein immunopurification-microarray (RIP-Chip) analysis to profile localized RNAs. *Methods Mol Biol*. 2011; 714:369–385. [PubMed: 21431753]
88. Riley KJ, Yario TA, Steitz JA. Association of Argonaute proteins and microRNAs can occur after cell lysis. *RNA*. 2012; 18:1581–1585. [PubMed: 22836356]
89. Mili S, Steitz JA. Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *RNA*. 2004; 10:1692–1694. [PubMed: 15388877]
90. Jensen KB, Darnell RB. CLIP: crosslinking and immunoprecipitation of in vivo RNA targets of RNA-binding proteins. *Methods Mol Biol*. 2008; 488:85–98. [PubMed: 18982285]
91. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M Jr, Jungkamp AC, Munschauer M, Ulrich A, Wardle GS, Dewell S, Zavolan M, Tuschl T. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*. 2010; 141:129–141. [PubMed: 20371350]
92. Rabani M, Levin JZ, Fan L, Adiconis X, Raychowdhury R, Garber M, Gnirke A, Nusbaum C, Hacohen N, Friedman N, Amit I, Regev A. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat Biotechnol*. 2011; 29:436–442. [PubMed: 21516085]
93. Miller C, Schwalb B, Maier K, Schulz D, Dumcke S, Zacher B, Mayer A, Sydow J, Marcinowski L, Dolken L, Martin DE, Tresch A, Cramer P. Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Mol Syst Biol*. 2011; 7:458. [PubMed: 21206491]
94. Kaneko S, Son J, Shen SS, Reinberg D, Bonasio R. PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. *Nat Struct Mol Biol*. 2013; 20:1258–1264. [PubMed: 24141703]
95. Jiao L, Liu X. Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2. *Science*. 2015; 350:aac4383. [PubMed: 26472914]
96. Beck DB, Narendra V, Drury WJ 3rd, Casey R, Jansen PW, Yuan ZF, Garcia BA, Vermeulen M, Bonasio R. In vivo proximity labeling for the detection of protein-protein and protein-RNA interactions. *J Proteome Res*. 2014; 13:6135–6143. [PubMed: 25311790]
97. Beltran M, Garcia de Herreros A. Antisense non-coding RNAs and regulation of gene transcription. *Transcription*. 2016; 7:39–43. [PubMed: 26985653]
98. Pelechano V, Steinmetz LM. Gene regulation by antisense transcription. *Nat Rev Genet*. 2013; 14:880–893. [PubMed: 24217315]
99. Moazed D. Small RNAs in transcriptional gene silencing and genome defence. *Nature*. 2009; 457:413–420. [PubMed: 19158787]



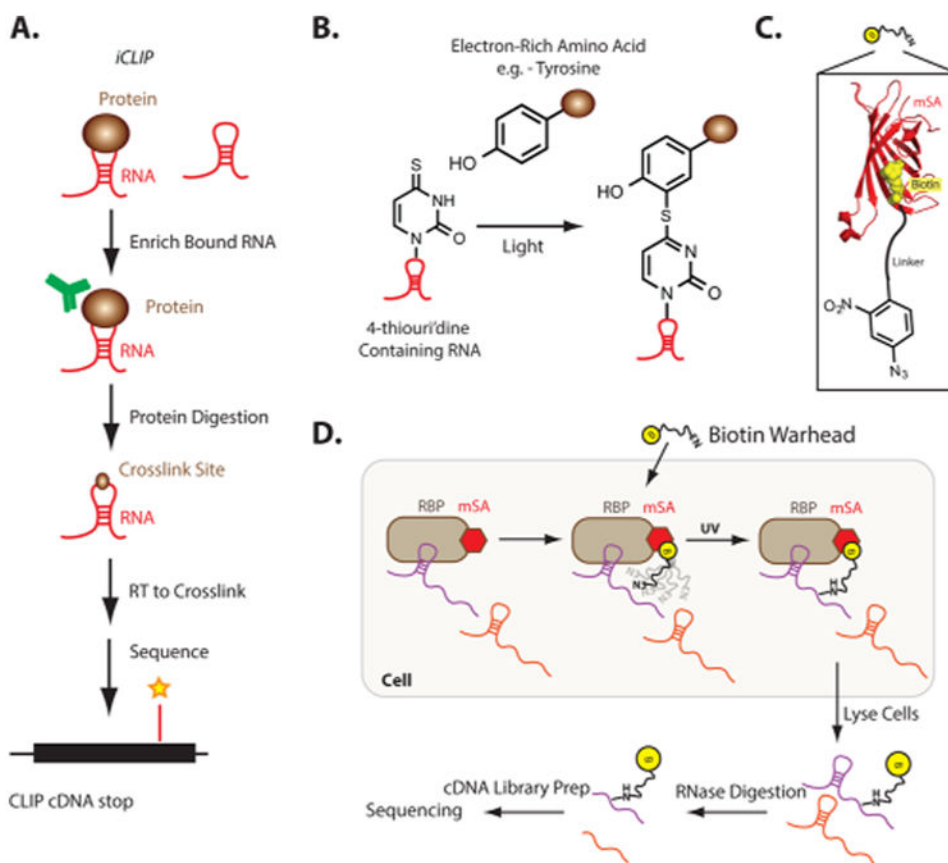
**Figure 1.** Mechanisms of chromatin modification and RNA-guided modification. (A) Methylation and demethylation of histone tails. (B) Acetylation and deacetylation of histone tails. (C) Depiction of guidance of histone modification enzymes by RNA.



**Figure 2.** Methods for assaying RNA localization on chromatin. (A) Schematic of methods to isolate RNAs, based on hybridization techniques. (B) Chemical structure of cross-linkers used to capture *in vivo* RNA interactions. (C) Localization profile demonstrating HOTAIR genomic occupancy in comparison to proteins in the PRC2 complex. (Figure adapted from ref 38. Copyright Elsevier 2011). (D) Chemical scheme depicting RNA–RNA cross-linking using AMT cross-linker.



**Figure 3.** Chemical and genomic methods for analyzing RNA structure. (A) Chemical modification by RNA structure readers DMS and SHAPE. (B) Schematic for analysis of adduct formation through RNA structure measurements. The experimenter can either analyze adducts through denaturing gel electrophoresis, or more recently, RNA sequencing. (C) SHAPE analysis of RoX RNA. A comparison between RNA structure probing and conservation suggests the lower SHAPE reactive sites (double stranded) are highly conserved and thus are important for function (Figure adapted from ref 77. Copyright Elsevier 2013).



**Figure 4.** Methods to interrogate the RNA–protein interface. (A) Schematic of CLIP. (B) Chemical cross-linking of 4-thiouridine and tyrosine in PAR-CLIP. (C) Schematic of aryl zide warhead in IPL. (D) Schematic of IPL.