HHS Public Access

Author manuscript

Biochim Biophys Acta. Author manuscript; available in PMC 2017 January 01.

Published in final edited form as:

Biochim Biophys Acta. 2016 January ; 1859(1): 121–127. doi:10.1016/j.bbagrm.2015.09.004.

Insight into lncRNA biology using hybridization capture analyses

Matthew D. Simon

Dept. of Molecular Biophysics & Biochemistry, Yale University, New Haven, CT 06516 Chemical Biology Institute, Yale West Campus, West Haven, CT, 06511

Abstract

Despite mounting evidence of the importance of large non-coding RNAs (lncRNAs) in biological regulation, we still know little about how these lncRNAs function. One approach to understand the function of lncRNAs is to biochemically purify endogenous lncRNAs from fixed cells using complementary oligonucleotides. These hybridization capture approaches can reveal the genomic localization of lncRNAs, as well as the proteins and RNAs with which they interact. To help researchers understand how these tools can uncover lncRNA function, this review discusses the considerations and influences of different parameters, (*e.g.*, crosslinking reagents, oliognucleotide chemistry and hybridization conditions) and controls to avoid artifacts. By examining the application of these tools, this review will highlight the progress and pitfalls of studying lncRNAs using hybridization capture approaches.

Keywords

lncRNA; CHART; ChIRP; RAP; formaldehyde; chromatin; genomic localization

Introduction

Functional non-coding RNAs have been known since the advent of molecular biology [1], but the scope of non-coding RNA function continues to expand at a surprising rate [2]. In mammals, the long non-coding RNAs (lncRNAs; also referred to as large non-coding or long intergenic non-coding RNAs) have drawn particular interest. One of the first mammalian lncRNA to be discovered was Xist, which spreads across chromatin of an entire chromosome (the inactive X in female cells), leading to remodeling of chromatin structure and composition [3, 4]. While Xist provides one good example of a lncRNA with an important biological role, it appears to be the tip of the iceberg; the numerous recent reports of lncRNA functions [5, 6, 7] support the notion that lncRNAs have fundamental roles in most biological processes.

*Corresponding author matthew.simon@yale.edu (Matthew D. Simon).

Despite many recent studies that highlight the importance of lncRNA functions [6], the foundations of the field have not yet solidified. While many specific classes of lncRNAs have been proposed, the field is far from consensus regarding how to best draw these delineations. Further, many of the choices for how to study and classify lncRNAs remain operational and pragmatic rather than being based on processing or biological pathways (for discussions of classification systems, see [7, 8, 9]). For example, there are only operational definitions of what defines a lncRNA; generally lncRNAs are referred to as lncRNAs if they are >200 nt long, but the term lncRNA has additional implications. While the majority of the genome can be transcribed, the terms lncRNA (and sometimes more specifically the intergenic or interleaving subclass of lncRNAs, lincRNAs) are reserved for those transcribed loci that resemble mRNAs in so far as they are RNA Pol II transcripts that are capped, spliced and polyadenylated. Unlike mRNAs, lncRNAs do not have extensive open reading frames. More progress will be needed to determine the importance of these distinctions and refine them to classify lncRNAs based on less arbitrary criteria.

Compared with our knowledge of proteins, little is known about the specific biochemical interactions that govern lncRNA function. To some extent, this uncertainty reflects the recent focus on lncRNAs, but it is also due to the relative paucity of tools to study these RNAs. It has been hard to reach consensus on how to interpret even classical biochemical analyses of lncRNA biochemistry [10]. Therefore, central to our knowledge of lncRNA biology is developing a critical understanding of the experiments that are most central to the field.

Of the numerous methods to study lncRNAs, there has been rapid progress in many areas that have been reviewed elsewhere [11, 12], including new functional approaches to direct and perturb lncRNA function, new approaches to study RNA structure, methods to study the RNA targets of proteins and methods to install genetic tags into lncRNAs. Here I will focus on advances in the use of hybridization capture approaches to study lncRNA targets.

There is a long history of studying cellular non-coding nucleic acids using labeled complementary nucleotides that hybridize specifically to their targets. As an early example, in situ hybridization (ISH) experiments were developed in which labeled nucleotides were used in cytology to localize complementary ribosomal sequences [13, 14]. Eventually, these experiments were extended to hybridization capture experiments using biotinylated oligonucleotides to biochemically enrich RNPs from native cell extracts leading to purification of splicing complexes [15, 16], 7SK RNP [17] and an active telomerase complex [18] among others.

Given the heightened interest that has developed around the roles of lncRNAs, it is no surprise that there has been a renaissance in hybridization capture methods to study these RNAs, and especially the adaptation of these techniques to use with crosslinked extracts [19], to deep sequencing [20, 21, 22], and to proteomic approaches [23, 24, 25, 26]. While the availability of these techniques represents a step forward for the field, maximizing the biological insights from the results of these techniques requires an expanded and critical understanding of the parameters and possible artifacts that these approaches produce.

Overview of Hybridization Capture Approaches

To study a lncRNA, it would be ideal to be able to purify the endogenous RNA from a biological extract and analyze the proteins and other targets with which it interacts (*e.g.*, where it binds in the genome). Unlike the study of proteins, in which case specific antisera must be raised, the predictability of Watson-Crick base-pairing means that complementary oligonucleotides can be used for hybridization capture experiments. Generally, hybridization capture experiments are performed as depicted in Fig. 1. To successfully accomplish such a purification, several parameters need to be determined: how to fix the cells, design the capture oligonucleotides, perform the hybridization, enrich the target molecules and interpret the results. Conditions for the first three reports of hybridization capture methods that provided successful enrichment in genome-wide analyses are provided in Table 1, including Capture Hybridization Analysis of RNA Targets (CHART) [20], Chromatin Isolation by RNA Purification (ChIRP)[21] and RNA Affinity Purification (RAP) [22]. Equally important to the hybridization capture conditions is the design of appropriate controls and an understanding of sources of background signals.

Cell systems

Hybridization capture experiments are generally performed on a similar scale to the analogous protein-centric experiments. For example, these experiments use on the order of 10^7 cells to determine genomic binding sites, similar to chromatin immunoprecipitation (ChIP) experiments. While ideally the analysis is directed at endogenous levels of the RNA, induced over-expression of the lncRNA can lead to higher yields, albeit at the potential cost of biological relevance. While eventually it will be helpful to establish principles for how much material is required for these experiments and how highly the RNA needs to be expressed, it is likely that the diversity of RNA structure, the nature of different RNA targets, and the purpose of the experiments will make generalization difficult. Furthermore, while analysis of nucleic acids allows for amplification of the enriched targets, analysis of proteins does not allow for such enrichment and therefore demands greater amounts of starting material.

Crosslinking conditions

In some cases, native conditions (without crosslinking) have been used to purify RNAs together with their associated complexes using hybridization capture [15, 16, 17, 18]. This approach offers several advantages, including the ability to purify catalytically active RNP complexes. If the aim of the study is to identify RNA targets, however, the ability to *covalently* attach an RNA to its targets can offer several advantages over native purifications. As long as the crosslinking is rapid, it is reasonable to assume that the covalent complexes represent a “snap-shot” of the factors bound to the RNA at the time of cross-linking. Under these conditions, sufficient rinsing is used to reduce artifacts caused by non-covalent interactions that occur after the cells are disrupted (the post-lysis interactions will be non-covalent whereas the real targets will be covalently connected to the RNA). Several cross-linking conditions have been developed for RNA hybridization capture experiments, each with their own considerations. These conditions include exposure to formaldehyde,

glutaraldehyde, disuccinimidyl gluterate, psoralen derivatives and UV light (Fig. 2). These crosslinking conditions, described in detail below, can be evaluated by their specificity, yield of crosslinking, stability and reversibility.

Formaldehyde

When studying factors that bind chromatin, the success of ChIP offers important precedent. The most common ChIP protocols employ formaldehyde crosslinking. Formaldehyde can be a very specific crosslinking reagent. Early studies using formaldehyde to examine protein-nucleic acid interactions demonstrated that it specifically captures bona fide interactions and demonstrates low levels of non-specific crosslinking (*e.g.*, DNA does not crosslink to irrelevant proteins such as bovine serum albumen, even at high concentrations in vitro) [27].

Formaldehyde alone is used in CHART [20], but this protocol uses more extensive formaldehyde treatment than ChIP (3% formaldehyde for 30 min in CHART rather than the ~1% formaldehyde for 10 min used in ChIP). In principle extensive cross-linking might trap non-specific interactions, but there is little evidence of such artifacts. In fact, the extensive formaldehyde treatment used in CHART was derived from a similar hybridization capture strategy designed to capture fragments of DNA (Proteomics of Intact Chromatin, PICh), which when applied to telomeres led to very specific retrieval of real telomere-localized proteins [28]. It is worth noting that formaldehyde is expected to lead to the retrieval of both molecules in direct contact with the RNA, as well as molecules that are indirectly cross-linked through other factors. Formaldehyde crosslinking is largely reversible, which offers downstream advantages. When analyzing proteins, this reversibility provides access to conventional SDS-PAGE and western blot analyses. When analyzing nucleic acids the formaldehyde crosslinking is reversed and does not substantially interfere with downstream library preparation or sequencing (as with ChIP-Seq).

Glutaraldehyde

Glutaraldehyde is also a fast-acting, membrane-permeable small molecule that can be used to covalently crosslink biomolecules. Gluteraldehyde is a popular crosslinker in the field of electron microscopy, and was used as the basis of the original ChIRP protocol [21]. Glutaraldehyde is a longer molecule than formaldehyde and contains two aldehydes available for crosslinking, increasing the distance between residues that it can crosslink. In aqueous solution, gluteraldehyde can form several diverse cyclized, oligomerized, and polymerized species, many of which can react with biomolecules, especially primary amines such as lysine. Around neutral pH, some glutaraldehyde adducts are irreversible [29]. This is not surprising as the imine adducts of glutaraldenhyde have reactive hydrogens alpha to the imine which have the potential to tautomerize to enamines and undergo Mannich-type reactions with irreversible products. In the context of a hybridization capture experiment, this irreversibility prevents conventional protein analyses such as SDS-PAGE, but protein enrichment can still be analyzed using slot blots [21]. Like formaldehyde, glutaraldehyde can crosslink both direct and indirect targets of an RNA.

Disuccinimidyl glutarate (DSG)

Another membrane-permeable crosslinker is DSG, which is a homobifunctional crosslinker that forms irreversible adducts with proteins. DSG was used in the original RAP protocol [22]. DSG contains two activated carboxylic acids in the form of N-hydroxysuccinimide (NHS) esters. NHS esters are stable enough in aqueous media to provide reasonable crosslinking times and concentrations, but are still highly reactive towards nucleophiles such as primary amines. In the case of reaction with primary amines, the product of DSG crosslinking is an amide bond which would be very challenging to reverse without destroying other target biomolecules. Therefore DSG has many of the same advantages and limitations as glutaraldehyde.

Psoralen and UV

Nucleic acids are photoreactive, and UV light has long been used to crosslink nucleic acids to their targets. UV crosslinking is the basis of some popular protein-centric analyses of protein-RNA interactions including CLIP [30] and PAR-CLIP [31]. UV crosslinking can be performed without additives, in which case the nucleobases of DNA or RNA absorb the UV light (near 260 nm), leading to a high-energy excited state which can lead to crosslinks to nearby biomolecules. This strategy has been used when performing hybridization capture reactions [25, 26], and has proven especially powerful for the isolation of protein targets of an RNA.

Both the strength and the weakness of UV crosslinking is its specificity. UV tends to crosslink only molecules in direct contact with the RNA of interest, which can provide an advantage as the adducts identified by hybridization capture are likely to be direct targets of the RNA. On the other hand, the efficiency of UV crosslink formation can be low and is thought to be very sensitive to local functional groups and structural orientation. Therefore it is generally assumed that UV crosslinking provides high specificity for real interactions, but also has a high false-negative rate, missing many interactions that exist in the cell but do not efficiently crosslink. It will be interesting to see how recent hybridization capture protocols, which expose cells to very high levels of UV crosslinking, impacts the rate of false-negative results.

As an alternative to direct excitation of the RNA nucleobases, the cell can be treated with photoreactive small molecules such as psoralen (or analogues thereof) which can intercalate into double-stranded nucleic acids and cause crosslinks when exposed to UV light, frequently using lower energy light (near 365 nm). Psoralen-mediated crosslinking for hybridization capture has been used to study networds of RNA-RNA interactions [32]. One major consideration with this approach is that the specificity of intercalation may dominate where crosslinks are formed. For example, psoralen has a strong preference to crosslink T/U-rich sequences. Nonetheless, the advantage of psoralen include the longer wavelength light that can be used in crosslinking (365 nm), leadning to less nucleic acid damage during the crosslinking reaction. Furthermore, unlike direct crosslinking of nucleic acids with short wavelength light, psoralen crosslinks can be reversed. The extent and wavelength of UV light used in crosslinking is a particularly important consideration for hybridization capture approaches (such as the genomic targets of an RNA). UV treatment at shorter wavelengths

will cause DNA and RNA to accumulate crosslinks, complicating efforts to use sequencing methodologies to characterize RNA targets.

Generation of extracts

After crosslinking, the cellular material is solubilized prior to hybridization. The ease of extract preparation depends on what crosslinking conditions were used. One advantage of UV crosslinking in comparison with formaldehyde or the other chemical crosslinkers described above is that the extract formed upon cell lysis is soluble under conditions developed for native extracts (*e.g.*, standard buffers and detergents and minimal mechanical perturbation). In contrast, for cells or nuclei crosslinked with chemical reagents including formaldehyde, making a soluble extract requires high levels of mechanical (*e.g.*, sonication) or enzymatic disruption (and sometimes both). When analyzing chromatin-bound lncRNA, this step also serves to fragment the genome. While in ChIP, the resulting chromatin fragments are frequently 200 bp long, the higher degree of crosslinking used in hybridization capture experiments can lead to the use of larger chromatin fragments. When mechanically fragmenting crosslinked nuclei, not only the DNA but also the target RNA is sheared. This problem can be partially mitigated by using of enzymes, including DNase, to help decrease the need for extensive mechanical shearing. Although shearing the target RNA can be problematic as it compromises the covalent integrity of the RNA, shearing also offers advantages when analyzing substructures within the target RNA [33].

Capture oligonucleotide design

Hybridization capture generally uses biotinylated nucleic acids that can be immobilized on streptavidin-coated magnetic beads. The biotin-streptavidin interaction is robust to even highly denaturing conditions, but is generally difficult to elute. As an alternative, desthiobiotin (DSB) can also be used, and material bound to DSB-conjugated capture oligonucleotides can be eluted from a solid support using a biotin-containing buffer [28, 20].

There are several factors to balance when choosing capture oligonucleotide sequences for hybridization capture experiments. First is to determine regions that are sufficiently unique to the target RNA for hybridization. This has been accomplished in several ways, including using pipelines that were originally developed for in situ hybridization protocols [21]. Next, it is important that the oligonucleotides can hybridize to the target RNA in the context of a crosslinked extract. There are generally two solutions: (1) tile the entire RNA with capture oligonucleotides with the expectation that at least some sites will be available for hybridization [21, 22], or (2) use RNase H sensitivity in the presence of candidate capture oligonucleotides sequences to determine accessibility [20]. Both these approaches have proven successful under multiple hybridization capture conditions. For example, while CHART generally works using RNase H mapped capture oligonucleotides [20], using tiled ChIRP-like capture oligonucleotides provided nearly identical results [34]. Similarly, ChIRP can be adapted to fewer oligonucleotides [33]. In the extreme case, CHART has been successful using individual capture oligonucleotides [23]. While most recent experiments have used DNA chemistry for capture oligonucleotides, longer RNAs and O2-methyl RNA capture oligonucleotides have also been used in genome-wide mapping experiments [35].

The high cost of synthetically biotinylated oligonucleotides can become prohibitive, but it is possible to enzymatically biotinylate capture oligos individually or in pools [34], allowing the use of unmodified DNAs.

Hybridization and rinse conditions

Critical to the success of hybridization capture reactions is finding hybridization conditions that provide high yields while maintaining sufficient specificity. It is also important to find conditions that maintain the solubility of the extract, which can be a challenge when using chromatin extracts. The three approaches highlighted in Table 1 converged on similar solutions to this problem: using high levels of denaturants combined with high concentration of salts. The urea and high salt buffer used in CHART is specifically tuned to minimize double stranded DNA (dsDNA) denaturation (with high salt, which increases duplex melting temperature) while maintaining stringency (with high urea, which depresses duplex melting temperature) with respect to short capture oligonucleotide hybridization. The original ChIRP conditions accomplish a similar end with formamide rather than urea, but these conditions cause higher degrees of denaturation of dsDNA. In the case of RAP, the high salt and denaturation is accomplished by the same additive, guanidinium thiocyanate. The more forcefully denaturing conditions used in RAP are required for the longer RNA capture oligonucleotides that have a higher melting temperature.

Elution Conditions

Elution conditions for each of hybridization capture experiments are highly variable depending on the class of RNA target that will be analyzed. For example, the nucleic acids can be isolated by eluting from the resin using proteinase K, but this will destroy the target proteins. In contrast, nucleases can be used to elute the target proteins, but at the cost of the nucleic acids. In some cases specific elution has been successful at decreasing background. One example of this is the use of only RNase H to specifically cleave the target RNA hybridized to DNA capture oligonucleotides. In the case of protein analysis, the proteins can be eluted from the resin directly for the purposes of mass spectrometry.

Quality control and data analysis

The most challenging part of a hybridization capture experiment is determining which enriched molecules represent real targets of the RNA as opposed to background or artifacts. In some cases, previous experiments and biology create clear expectations that help in optimization and analysis. For lncRNAs where little is known, however, it can be difficult to determine if an experiment is working, and to determine what represents real signal rather than background in the experiment. To some extent these challenges can be mitigated using well-designed controls as discussed below. Ultimately further experimentation will be important to provide scrutiny and functional context for the results.

Distinguishing artifacts from signal using controls

Many of the considerations when using hybridization capture approaches overlap with other affinity purification techniques. There are also a set of artifacts that are unique when using

hybridization capture approaches. These two types of artifacts can be broken down into (1) artifacts from non-specific binding and (2) hybridization induced artifacts (Fig. 3). These artifacts can be problematic regardless of whether the goal of the experiment is to study protein, DNA, or RNA targets of a lncRNA.

The artifacts caused by direct, non-specific binding are similar to conventional affinity purification reactions so will be discussed only briefly here. For example, basic proteins in the extract can bind non-specifically to the acidic oligonucleotides on beads. In this case, incorporating high salt rinses can help reduce this source of background. As these interactions are non-covalent, stringent rinses can help remove these interactions while leaving the covalently bound molecules that are the real targets of the RNA. The stringency of these rinses is primarily limited by the stability of the capture oligo - target RNA duplex. Most of these artifacts can be identified using control capture oligonucleotides that target a different RNA (or no RNA at all). These artifacts may persist in the absence of crosslinking, although it is frequently difficult to compare crosslinked and non-crosslinked extracts to each other.

When examining enrichment of proteins, this non-specific binding provides a challenge due to the dynamic range of protein concentrations in the proteome. Similar to other affinity enrichment approaches, Stable isotope labeling by amino acids in cell culture (SILAC) [36] in combination with the other controls discussed here can provide a powerful tool to measure protein enrichment in complex mixtures [25].

Direct hybridization to off-target DNA or RNA

One source of artifacts is caused by hybridization events in which the capture-oligonucleotide directly hybridizes to an off-target RNA or DNA. In the case of CHART the hybridization to off-target DNA is reduced by using hybridization conditions that leave genomic DNA in its double stranded, and any residual signal from capture oligonucleotides binding to DNA can be reduced using an RNase H elution, which elutes RNA bound to the capture oligonucleotides but not DNA. An alternative method that controls for both off-target RNA and DNA binding is to use two independent sets of capture oligonucleotides that bind the target RNA. Real signals are expected to be found in both experiments whereas signals that are found in only one experiment are interpreted as hybridization-induced artifacts. Using cells that do not express the target RNA is also a good way to control for this class of artifact.

Indirect hybridization to off-target DNA or RNA

A challenge when interpreting signals from hybridization capture experiments arise from molecules that artifactually hybridize to the target RNA. For example, if the hybridization capture conditions are not well optimized, the enriched DNA may include motifs that are directly complementary to the target lncRNA. This could be due to real biology (*i.e.*, the lncRNA forms R-loops with its target DNA), but it could also be caused by hybridization induced artifacts. This class of artifact is not addressed using the controls listed above. Using two sets of capture oligonucleotides targeting the same RNA will both result in beads coated with the target RNA; any artifacts that bind the target RNA will be interpreted as

"real" signal because they are present in both oligonucleotide mixtures. Similarly, using cells that lack the target RNA will remove these artifactual signals, making them appear as true positives.

While in general these indirect hybridization induced artifacts are difficult to distinguish from real signals, it is occasionally possible to distinguish real signal from these artifacts. The Xist lncRNA spreads across one of the two female X chromosomes. By using a cell line that allows allele-specific analysis [37, 38], it was possible to distinguish Xist spreading on the two different copies of the X chromosome. As the artifactual signals are not allele-specific but the biologically valid signals from hybridization capture analysis of Xist are allele-specific, it was possible to validate that Xist CHART signals were not caused by hybridization induced artifacts [34].

In summary, controls are helpful to avoid over-interpretation of capture hybridization data, yet even with well designed controls there is reason for caution. In some cases positive controls, or at least agreement with past literature, can assist in interpretation. For example, if a lncRNA colocalizes with a chromatin modifying complex, it is expected that there will be extensive overlap between ChIP signals of the target protein and signals from the hybridization capture coupled to DNA sequencing experiments. Examples of this were found with the roX2 lncRNA, which binds to the same discrete sites where the proteins in the fly dosage compensation complex bind [21, 20]. Similarly, overlap was observed between the ChIP signals from the polycomb repressive complex 2 with Xist CHART signals on the inactive X during the onset of X-chromosome inactivation [34]. Functional connections can also be used to help judge the validity of hybridization capture data. For example, both roX2 in fly and Xist in mouse were found to be highly enriched on the X-chromosome [21, 20, 22, 34], as expected based on their functions [39]. Similarly, SAF-A/hnRNP-U was previously reported to work together with Xist [40, 41, 42], so it is reassuring that three recent hybridization capture experiments enriched this protein [25, 24, 26]. Nonetheless, because so little is known about many lncRNAs, leaving few good positive controls, it is frequently challenging to optimize enrichment conditions or determine appropriate signal thresholds in many hybridization capture experiments.

Conclusions

Hybridization capture experiments have already provided insight into a wide range of lncRNA functions and mechanisms [12]. For example, several different hybridization capture approaches have collectively demonstrated the spreading dynamics of Xist [34, 22], and revealed proteins that bind to it [25, 24, 26]. While these some of these results agree well with each other (for example, Xist spreading profiles in female mouse embryonic stem cells) demonstrating convergence and progress in the field, other results have only partial overlap (the list of Xist-binding proteins in different analyses) underscoring the continued need for protocol development and vigilance. It is encouraging that the successes and lessons from the different hybridization capture approaches have influenced each other, have been expanded to new approaches, and continue to be optimized. As the field of lncRNA biology continues to expand, these hybridization capture approaches have the potential to be

a primary vehicle driving the integration of lncRNAs into our understanding of biological regulation.

Acknowledgments

This work was supported by the NIH grant DP2 1DP2HD083992 and the Searle Scholars Program.

References

[1]. Darnell, JE. RNA: Life's Indispensable Molecule. Cold Spring Harbor Laboratory Press; 2011.

[2]. Cech TR, Steitz JA. The noncoding rna revolution-trashing old rules to forge new ones. Cell. 2014; 157(1):77–94. doi:10.1016/j.cell.2014.03.008. URL http://www.ncbi.nlm.nih.gov/pubmed/24679528. [PubMed: 24679528]

[3]. Jeon Y, Sarma K, Lee JT. New and Xisting regulatory mechanisms of X chromosome inactivation. Curr Opin Genet Dev. 2012; 22(2):62–71. [PubMed: 22424802]

[4]. Gendrel A-V, Heard E. Noncoding RNAs and Epigenetic Mechanisms During X-Chromosome Inactivation. Annu. Rev. Cell Dev. Biol. 2014; 30(1):561–580. [PubMed: 25000994]

[5]. Bergmann JH, Spector DL. Long non-coding RNAs: modulators of nuclear structure and function. Current Opinion in Cell Biology. 2014; 26:10–18. [PubMed: 24529241]

[6]. Vance KW, Ponting CP. Transcriptional regulatory functions of nuclear long noncoding RNAs. Trends Genet. 2014; 30(8):348–355. [PubMed: 24974018]

[7]. Mattick JS, Rinn JL. Discovery and annotation of long noncoding RNAs. Nat Struct Mol Biol. 2015; 22(1):5–7. [PubMed: 25565026]

[8]. Bonasio R, Shiekhattar R. Regulation of Transcription by Long Noncoding RNAs. Annu Rev Genet.

[9]. Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. CELL. 2013; 154(1):26–46. [PubMed: 23827673]

[10]. 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.

[11]. McHugh CA, Russell P, Guttman M. Methods for comprehensive experimental identification of RNA-protein interactions. Genome Biol. 2014; 15(1):203. [PubMed: 24467948]

[12]. Chu C, Spitale RC, Chang HY. Technologies to probe functions and mechanisms of long noncoding RNAs. Nat Struct Mol Biol. 2015; 22(1):29–35. [PubMed: 25565030]

[13]. Gall JG, Pardue ML. Formation and detection of RNA-DNA hybrid molecules in cytological preparations. Proc Natl Acad Sci USA. 1969; 63(2):378–383. [PubMed: 4895535]

[14]. John HA, Birnstiel ML, Jones KW. RNA-DNA hybrids at the cytological level. Nature. 1969; 223(5206):582–587. [PubMed: 5799530]

[15]. Blencowe BJ, Sproat BS, Ryder U, Barabino S, Lamond AI. Anti-sense probing of the human U4/U6 snRNP with biotinylated 2'-OMe RNA oligonucleotides. CELL. 1989; 59(3):531–539. [PubMed: 2478298]

[16]. Barabino SM, Sproat BS, Ryder U, Blencowe BJ, Lamond AI. Mapping U2 snRNP–pre-mRNA interactions using biotinylated oligonucleotides made of 2'-OMe RNA. EMBO J. 1989; 8(13):4171–4178. [PubMed: 2531659]

[17]. Wassarman DA, Steitz JA. Structural analyses of the 7SK ribonucleo-protein (RNP), the most abundant human small RNP of unknown function. Mol Cell Biol. 1991; 11(7):3432–3445. [PubMed: 1646389]

[18]. Lingner J, Cech TR. Purification of telomerase from Euplotes aediculatus: requirement of a primer 3' overhang. Proc Natl Acad Sci USA. 1996; 93(20):10712–10717. [PubMed: 8855245]

[19]. Mariner PD, Walters RD, Espinoza CA, Drullinger LF, Wagner SD, Kugel JF, Goodrich JA. Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. Mol Cell. 2008; 29(4):499–509. [PubMed: 18313387]

[20]. 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 USA. 2011; 108(51):20497–20502. [PubMed: 22143764]

[21]. 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(4):667–678. [PubMed: 21963238]

[22]. 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(6147):1237973. [PubMed: 23828888]

[23]. West JA, Davis CP, Sunwoo H, Simon MD, Sadreyev RI, Wang PI, Tolstorukov MY, Kingston RE. The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. Mol Cell. 2014; 55(5):791–802. [PubMed: 25155612]

[24]. Chu C, Zhang QC, da Rocha ST, Flynn RA, Bharadwaj M, Calabrese JM, Magnuson T, Heard E, Chang HY. Systematic discovery of xist RNA binding proteins. CELL. 2015; 161(2):404–416. [PubMed: 25843628]

[25]. McHugh CA, Chen C-K, 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.

[26]. Minajigi A, Froberg JE, Wei C, Sunwoo H, Kesner B, Colognori D, Lessing D, Payer B, Boukhali M, Haas W, Lee JT. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. Science.

[27]. Solomon MJ, Varshavsky A. Formaldehyde-mediated DNA-protein crosslinking: a probe for in vivo chromatin structures. Proc Natl Acad Sci USA. 1985; 82(19):6470–6474. [PubMed: 2995966]

[28]. Déjardin J, Kingston RE. Purification of proteins associated with specific genomic Loci. CELL. 2009; 136(1):175–186. [PubMed: 19135898]

[29]. OKUDA K, URABE I, YAMADA Y, OKADA H. Reaction of Glu-taraldehyde with Amino and Thiol Compounds. Journal of Fermentation and Bioengineering. 1991; 71(2):100–105.

[30]. Darnell RB. HITSCLIP: panoramic views of protein–RNA regulation in living cells. WIREs RNA. 2010; 1(2):266–286. [PubMed: 21935890]

[31]. Ascano MM, Hafner MM, Cekan PP, Gerstberger SS, Tuschl TT. Identification of RNA-protein interaction networks using PARCLIP. WIREs RNA. 2012; 3(2):159–177. [PubMed: 22213601]

[32]. Engreitz JM, Sirokman K, McDonel P, Shishkin AA, Surka C, Russell P, Grossman SR, Chow AY, Guttman M, Lander ES. RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre- mRNAs and chromatin sites. CELL. 2014; 159(1):188–199. [PubMed: 25259926]

[33]. Quinn JJ, Ilik IA, Qu K, Georgiev P, Chu C, Akhtar A, Chang HY. Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification. Nat Biotech. 2014; 32(9):933–940.

[34]. Simon MD, Pinter SF, Fang R, Sarma K, Rutenberg-Schoenberg M, Bowman SK, Kesner BA, Maier VK, Kingston RE, Lee JT. High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. Nature. 2013; 504(7480):465–469. [PubMed: 24162848]

[35]. Lee N, Moss WN, Yario TA, Steitz JA. EBV noncoding RNA binds nascent RNA to drive host PAX5 to viral DNA. CELL. 2015; 160(4):607–618. [PubMed: 25662012]

[36]. Ong S-E, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M. Stable isotope labeling by amino acids in cell culture. SILAC, as a simple and accurate approach to expression proteomics., Mol Cell Proteomics. 2002; 1(5):376–386. [PubMed: 12118079]

[37]. Yildirim E, Sadreyev RI, Pinter SF, Lee JT. X-chromosome hyperactivation in mammals via nonlinear relationships between chromatin states and transcription. Nat Struct Mol Biol. 2012; 19(1):56–61. [PubMed: 22139016]

[38]. Pinter SF, Sadreyev RI, Yildirim E, Jeon Y, Ohsumi TK, Borowsky M, Lee JT. Spreading of X chromosome inactivation via a hierarchy of defined Polycomb stations. Genome Res. 2012; 22(10):1864–1876. [PubMed: 22948768]

[39]. Lucchesi JC, Kelly WG, Panning B. Chromatin remodeling in dosage compensation. Annu Rev Genet. 2005; 39:615–651. [PubMed: 16285873]

[40]. Helbig R, Fackelmayer FO. Scaffold attachment factor A (SAF-A) is concentrated in inactive X chromosome territories through its RGG domain. Chromosoma. 2003; 112(4):173–182. [PubMed: 14608463]

[41]. Pullirsch D, Härtel R, Kishimoto H, Leeb M, Steiner G, Wutz A. The Trithorax group protein Ash2l and Saf-A are recruited to the inactive X chromosome at the onset of stable X inactivation. Development. 2010; 137(6):935–943. [PubMed: 20150277]

[42]. Hasegawa Y, Brockdorff N, Kawano S, Tsutui K, Tsutui K, Nakagawa S. The Matrix Protein hnRNP U Is Required for Chromosomal Localization of Xist RNA. Dev Cell. 2010; 19(3):469–476. [PubMed: 20833368]

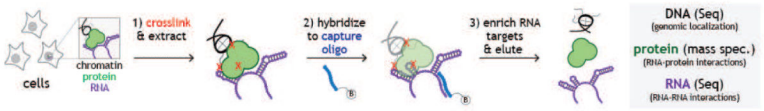

Figure 1.
Overview of hybridization capture methods to study endogenous RNAs.

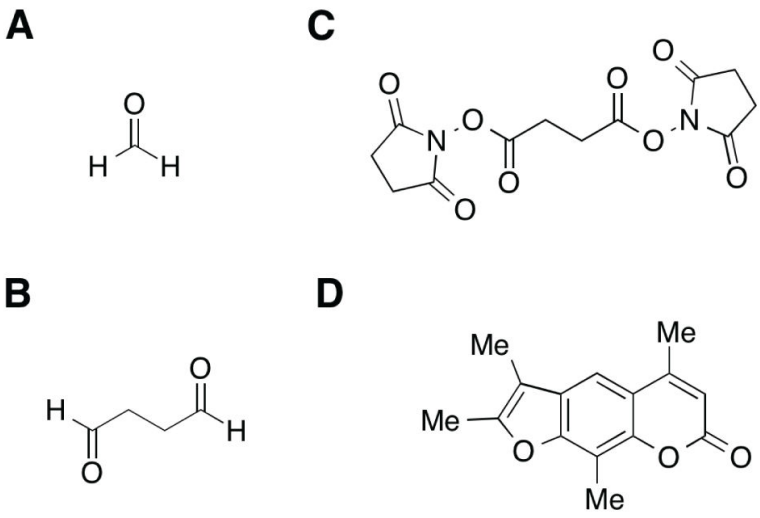

Figure 2.
Chemical structures of crosslinkers used in hybridization capture approaches. A. Formaldehyde. B. Glutaraldehyde. C. Disuccinimidyl glutarate. D. Trimethylpsoralen.

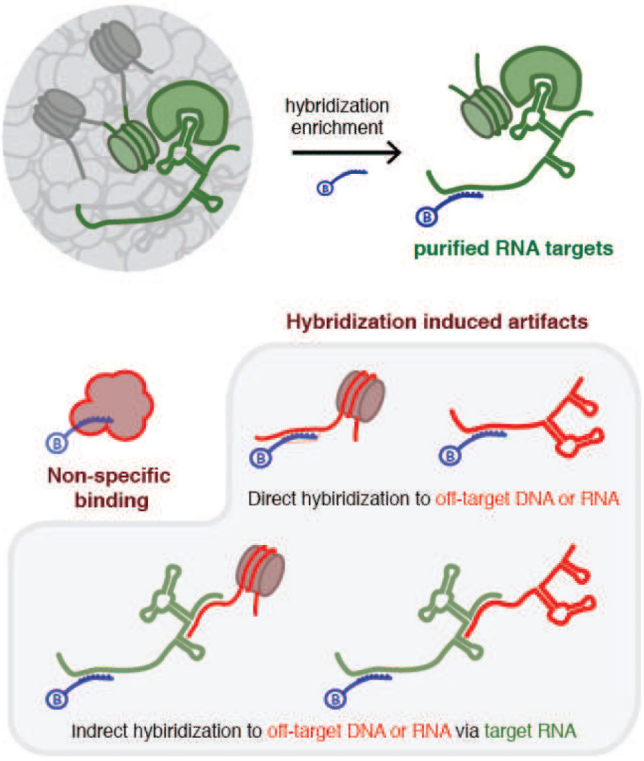

Figure 3.
Overview of enrichment of desired signals (green) as opposed to different types of artifacts (red) using hybridization capture experiments.

Table 1

Overview of three hybridization capture methods.

	Method		
	CHART[1]	**ChIRP**[2]	**RAP**[3]
crosslinking reagent(s)	3% formaldehyde	1% gluteraldehyde	2 mM DSG[a] and 3% formaldehyde
oligonucleotide chemistry	biotinylated or DSB[b] ~25 nt DNAs	biotinylated 20 nt DNA	biotinylated 120 nt RNAs
oligonucleotide design	RNase H mapping	full tiling	full tiling
hybridization conditions	2 M urea ~850 mM salt detergents[c]	10% formamide 500 mM salt 1% SDS	3 M guanidine thiocyanate detergents[d]
elution conditions	RNase H[e]	RNases A + H	Proteinase K[f]

[1] Conditions from original citation, Simon, 2011 [20].

[2] Conditions from original citation, Chu, 2011 [21].

[3] Conditions from original citation, Engreitz, 2013 [22].

[a] DSG: disuccinimidyl glutarate.

[b] DSB: desthiobiotinylated.

[c] Detergents used in CHART conditions: ~0.2% *N*-Lauroylsarcosine, 0.03% deoxycholate, 0.3% SDS.

[d] Detergents used in RAP: 1% NP-40, ~0.2% *N*-Lauroylsarcosine, 0.1% deoxycholate.

[e] Alternate elution with biotin buffer reported when using DSB.

[f] Alternate elution with heat (94° 5 min) reported for RNA analysis.