Published in final edited form as:

Wiley Interdiscip Rev RNA. 2012 ; 3(2): 159–177. doi:10.1002/wrna.1103.

Identification of RNA–protein interaction networks using PAR-CLIP

Manuel Ascano[†], Markus Hafner[†], Pavol Cekan, Stefanie Gerstberger, and Thomas Tuschl^{*}
Laboratory of RNA Molecular Biology, Howard Hughes Medical Institute, The Rockefeller University, New York, NY, USA

Abstract

All mRNA molecules are subject to some degree of post-transcriptional gene regulation (PTGR) involving sequence-dependent modulation of splicing, cleavage and polyadenylation, editing, transport, stability, and translation. The recent introduction of deep-sequencing technologies enabled the development of new methods for broadly mapping interaction sites between RNA-binding proteins (RBPs) and their RNA target sites. In this article, we review crosslinking and immunoprecipitation (CLIP) methods adapted for large-scale identification of target RNA-binding sites and the respective RNA recognition elements. CLIP methods have the potential to detect hundreds of thousands of binding sites in single experiments although the separation of signal from noise can be challenging. As a consequence, each CLIP method has developed different strategies to distinguish true targets from background. We focus on photoactivatable ribonucleoside-enhanced CLIP, which relies on the intracellular incorporation of photoactivatable ribonucleoside analogs into nascent transcripts, and yields characteristic sequence changes upon crosslinking that facilitate the separation of signal from noise. The precise knowledge of the position and distribution of binding sites across mature and primary mRNA transcripts allows critical insights into cellular localization and regulatory function of the examined RBP. When coupled with other systems-wide approaches measuring transcript and protein abundance, the generation of high-resolution RBP-binding site maps across the transcriptome will broaden our understanding of PTGR and thereby lead to new strategies for therapeutic treatment of genetic diseases perturbing these processes.

INTRODUCTION

RNA-binding proteins (RBPs) and ribonucleoprotein complexes (RNPs) recognize primary sequence and/or secondary structure elements within mature coding and non-coding RNAs and their precursors. Many RBPs and RNPs are required for constitutive processes, such as pre-mRNA splicing, cleavage, and polyadenylation. However, variations in the abundance of RNA targets and these RNA interacting factors can also influence the expression of specific genes. Furthermore, cell-type specific RBPs and non-coding RNAs regulate the flow of genetic information in more directed manners, e.g., by regulating mRNA stability or translation. The importance of post-transcriptional gene regulation (PTGR) is underscored by the wide range of diseases that result from genetic alterations within RBPs, constituents of RNPs, and/or their mRNA targets.¹

The human genome encodes at least 600 RBPs based on the presence of known RNA-binding domains (RBDs) as described in the Pfam² database (Figure 1 and Table 1). There are

an additional 1400 nucleic acid-binding proteins with zinc finger, helicase, or nuclease domains are known, many of which remain poorly characterized and lack information on whether they target DNA, RNA, or both. Despite the number of RBPs encoded in the human genome, the targets nor function for the vast majority are not well understood. RBDs represent evolutionary conserved peptide domains that recognize specific sequence or structural elements embedded in their target RNAs, which are referred to as RNA recognition elements (RREs). There are about 75 annotated RBDs and those that have been molecularly characterized mostly bind short 4–6 nt segments in a sequence and/or structural specific manner.³ The majority of RBPs (65%) contain only one RBD, whereas a smaller proportion (35%), often containing RNA recognition motif (RRM), hnRNP K homology (KH), dsrm, zf-CCHC, Piwi, Argonaut and Zwillie (PAZ), and Piwi domains, have additional repeats of the same RBD, other RBDs, and/or RNA processing domains. Given that most RBDs recognize short and often degenerate RNA sequences, the use of multiple types of binding domains should increase specificity of target recognition.

Non-coding RNAs assembled into RNPs may play multiple roles. They can act as scaffolds for recruitment of RBPs, which may further recruit auxiliary proteins to form a mature or functional RNP, or they may additionally act as catalysts in the processing of target RNAs. Finally, non-coding RNAs or segments of it may act as guide in recognizing complementary sequences within target RNAs. The ability to assemble non-coding RNAs with different guide sequences into RNPs of similar or identical protein composition provided an evolutionary advantage and is exploited in several processes related to mRNA maturation or regulation.

The RREs for many RBPs have not been determined and computational approaches alone are not sufficiently advanced to predict the primary sequence or structure of the RNA segment recognized by a given RBP. Comparative genomic approaches were effective in predicting short conserved sequence motifs within UTRs, about half of which corresponded to miRNA-binding sites, but the RBPs recognizing the remaining motifs have not been determined.²⁴ Furthermore, biochemical studies aimed at characterizing the sequence-specific contacts between RRE and RBD can be complicated by RBPs having a general electrostatic affinity to RNA wherein nucleotide residues flanking the core RRE could contribute to binding. A systems biology approach for dissection of RNA–RBP and RNA–RNP interactions therefore requires sensitive methods for identification of native RNA target sites, ultimately capturing, in a cell-type specific manner, the complete network of interactions. Considerable progress has been made toward this goal, driven by the development of powerful sequencing technologies adapted for the characterization of RNA segments bound by specific RBPs and RNPs. The methods applied toward this ambitious goal are reviewed, followed by a discussion on how these interactions are translated to functional outcomes on gene expression.

MOLECULAR AND BIOCHEMICAL APPROACHES FOR IDENTIFICATION OF RREs AND NATURAL TARGETS OF RBPs WITHOUT CROSSLINKING

Following the initial discovery of RBPs and their associated RNAs, computational and biochemical approaches were used to identify RBDs and RREs. RNA primary and secondary structure elements critical for RBP binding were predicted by sequence analysis of multiple targets and/or evolutionarily related targets. Subsequently, RNA segments comprising the motif and some flanking sequence as well as the mutants of the motif were assayed by biochemical approaches such as electrophoretic mobility shift assays or ultraviolet (UV) crosslinking of RBPs to target RNAs. Some investigators performed *in vitro* evolution experiments (also known as SELEX) to isolate high-affinity RNA ligands (aptamers) from pools of random RNA sequence.²⁵ RNA aptamers were identified for a

number of RBPs, among them the RRM-domain protein ELAVL2,²⁶ the KH-domain proteins Quaking (QKI),²⁷ and Nova.²⁸ It is presumed that the intrinsic affinity of RBPs to RNA facilitates the recovery of RNA ligands comprising their natural RRE. Similar to the discovery of RREs using natural target RNAs, the aptamer-defined RRE is derived by comparative analysis of independently selected sequences. Subsequently, natural targets are predicted based on the presence of the discovered RRE from the transcriptome of the species encoding the RBP. An array-based approach suitable for the recovery of low-complexity RREs has also been presented, where the RRE is derived by sequence enrichment from a library of about 200,000 distinct 29–38 nt RNAs from RBP–RNA coimmunoprecipitates.²⁹

To directly identify endogenous targets, various RBP-immunoprecipitation (IP) methods were established. They differ by the way they capture and analyze the RBP- or RNP-associated RNAs. IP and RNA isolation protocols are either designed to recover full-length RNA targets, such as mRNAs, or partially RNase-digested segments bound and protected by RBPs/RNPs. Subsequently, bound RNAs are examined by array hybridization assays, quantitative reverse transcription (RT) polymerase chain reaction (PCR) or by conversion to cDNA libraries for direct sequencing. The various approaches are listed in Table 2. It is critical that the experimental conditions of cell lysis and IP are adjusted to not destabilize the RNA–RBP/RNP interactions, e.g., avoiding high salt buffers or sonication.^{30,31} The separation of true target from background RNAs requires experimental controls, e.g., mock IP, to determine enrichment of RNA targets in the IP versus control. With the introduction of deep-sequencing technologies, comprehensive identification of RBP targets is now possible by sequencing cDNA libraries prepared from coimmunoprecipitated RNA.³² Large-scale genomic or transcriptomic alignments originating from short RNA segments are performed and have gained in importance over array-based cDNA analysis as it also facilitates *de novo* discovery of transcripts, editing, and alternative splicing events.

RNA–PROTEIN CROSSLINKING FOR IDENTIFICATION OF RRE AND NATURAL TARGETS OF RBPs

Crosslinking and immunoprecipitation (CLIP) protocols share the same basic outline (Figure 2). RNA is covalently crosslinked to interacting proteins, preferably *in vivo* and prior to cell lysis to eliminate any possibility of post-lysis reassociation of interaction partners.³¹ Cell lysates are treated with RNase to increase the yield of the IP by fragmenting mRNPs into small RNA–protein subcomplexes and to shorten RNA lengths to ideally simplify RRE identification. Immobilization of crosslinked RBPs during IP provides an opportunity for various biochemical manipulations including the radiolabeling of RBP-associated RNAs. The recovery of RBPs under non-denaturing conditions or incomplete RNase treatment can lead to the copurification of directly or indirectly interacting additional RBPs crosslinked to their respective targets. Therefore, recovered complexes are further fractionated using protein-denaturing sodium dodecyl sulfate (SDS) gels. Covalently attached RNA in the 25-nt size range is expected to alter the molecular mass of the RBP by about 8 kDa, which represents a small change in total mass when working with RBPs above 50 kDa. The SDS gel also separates non-covalently linked RNA segments because they migrate faster than the crosslinked RBP–RNA complexes. The RNAs crosslinked to the protein of the expected molecular mass are recovered after a Proteinase K treatment, leaving a covalently bound amino acid or oligopeptide at the crosslinked nucleotides. A cDNA library is prepared followed by deep sequencing. Finally, sequence reads are aligned to the genome and/or transcriptome. Clusters of overlapping sequence reads are selected for further determination of the RRE and for obtaining a list of precise binding sites within gene targets.

CLIP approaches differ by the use of natural versus non-natural nucleotides each having a different optimal UV wavelength for crosslinking. In addition, there are differences in the

selection and titration of specific RNases, the recovery methods of crosslinked RBP–RNA complexes, the construction of cDNA libraries, and the computational data analyses.

UV-254 CLIP PROTOCOLS

Exposure of nucleic acids in lysates, cultured cells, or tissues to short-wavelength UV light of 254 nm (UV-254) generates intra and intermolecular nucleic acid- and nucleic acid-protein crosslinks.⁶¹ One early version of the protocol was developed for characterizing RNA–protein interaction sites within abundant spliceosomal RNPs of known protein and RNA composition.⁶² In these studies, crosslinked RNA–protein complexes were recovered by IP, the RNA was isolated, and analyzed by RT/primer extension. Mapping of crosslinking sites *in vitro* is made possible because reverse transcriptase stalls at the crosslinking site.

To identify endogenous target sites of mRNA-specific RBPs, UV-254 crosslinked RNAs can be converted to cDNA libraries and sequenced (Table 2 and Figure 2(a)).³⁶ Efficient cDNA synthesis requires the joining of adapter sequences to the isolated RNA for RT-PCR and sequencing. Crosslinking between RNA and RBP enables additional purification steps after IP, such as denaturing SDS gel size fractionation and nitrocellulose membrane transfer to separate non-crosslinked background RNA. As an alternative to the use of antibodies, hexahistidine tagging the RBP for cell culture experiments allowed investigators to purify RBP–RNA complexes under denaturing conditions using metal-chelating matrices^{53,54,56} (Figure 2(a), right panel). Collectively, these purification steps enrich for crosslinked RNAs, but because the crosslinked RNA population is difficult to reverse transcribe, residual non-crosslinked RNAs (background RNA) confound the analysis since these can be reverse transcribed with high efficiency and therefore can be overrepresented. When reverse transcriptase stalls at crosslinking sites, it infrequently reads through the lesion generated by the photoadduct introducing characteristic mutations and/or deletions in the cDNA. Alignment of cDNA sequence reads can thus identify clusters of reads enriched for mutations. Two recent publications report the increased frequency of single-nucleotide deletions in sequence reads originating from UV-254 crosslink sites, and mapping of these errors provided important evidence for enrichment of crosslinked binding sites.^{63,64} The positional information of these deletions within clustered sequence reads can then be used to identify the underlying RRE(s) providing insights into RBP–RRE molecular interactions. Sequence analysis of UV-254 crosslinked cDNA libraries demonstrated that the sites of deletions correspond almost exclusively to uridines, indicating that either not all four nucleotides can participate in UV-254 crosslinking or that the reverse transcriptase cannot read past or does not induce mutations with guanosine, cytidine, and adenosine photoadducts. In summary, RT is a critical step for the generation of the mutation or deletion corresponding to the crosslink evidence featured in some CLIP methods. Several important questions remain to be answered. During cDNA preparation, what is the fraction of read-through events compared to premature terminations at different types of photoadduct lesions? What is the spectrum of cDNA changes observed for structurally defined photoadduct lesion?

Experimental controls designed to assess background RNA signal, such as absence of UV irradiation, mock IP experiments, or the use of cell lysates from RBP knockout cells, are often recommended but may be less informative than intended. In these cases, the amount of background RNA that can be carried through an experiment is often insufficient to reproducibly generate a cDNA library resulting in either the absence or distortion of the background. Therefore an alternative way to assess background may be to perform a parallel CLIP experiment on an unrelated RBP with similar abundance and subcellular localization but with different target RNA specificity. Such experiments would help remove signals

derived from UV-damaged RNA,⁶⁵ intramolecular RNA crosslinking, single-nucleotide polymorphisms interpreted as crosslinking events, and short fragmented RNA background generated by RNase digestion.

The iCLIP method (Figure 2(a), middle panel)⁵³ was developed to take advantage of the frequent reverse transcriptase termination events at crosslinked sites. It relies upon the ligation of a single adapter that provides the RT primer-binding site at the 3' end of crosslinked RNAs. First strand cDNAs are synthesized by RT using primers complementary to the adapter, which also contains a restriction site, barcode sequence, and a 5' primer-binding site for Solexa sequencing. The stalled cDNA fragments are then circularized, positioning the 5' primer-binding site upstream of the cDNA segment. Following second strand generation, the resulting dsDNA is linearized, PCR amplified, and deep sequenced. The nucleotide position immediately upstream of the 5' end of sequence read clusters are presumed to correspond to the crosslink sites that originally stalled the reverse transcriptase. However, without a means of measuring enrichment of reads derived from crosslinked sites over background, it remains difficult to assess what fraction of reads correspond to crosslinked versus non-crosslinked RNA sequences.

PAR-CLIP PROTOCOL

The Application of Photoactivatable Ribonucleosides for RNA–Protein Crosslinking

As an alternative to the crosslinking of natural nucleic acids with short-wavelength UV light, modification of RNA with photoactivatable nucleoside analogs has been shown to increase the efficiency of nucleic acid crosslinking. The best characterized photoactivatable ribonucleoside analogs (Figure 3(a)) are halogenated pyrimidines, such as 5-iodouridine and 5-iodocytosine,⁶⁶ and thione-containing ribonucleosides, such as 4-thiouridine (4SU) and 6-thioguanosine.⁶⁷ Photoactivatable ribonucleosides are excited with long-wavelength UV (>310 nm), where natural nucleotides no longer crosslink, and yield photoadducts of distinct structure (Figure 3(b)) depending on their reaction mechanisms involving either shortlived free radicals as intermediates or direct photoaddition without reaction intermediates. Amino acid side chains within proteins show different reactivity toward the different nucleoside analogs with the reactions involving predominantly aromatic amino acids such as phenylalanine, tyrosine, and tryptophan but also lysines and cysteine.⁶⁶

Initially, photoactivatable nucleosides were introduced into RNA by *in vitro* transcription from DNA templates or by chemical synthesis and used to probe RNA–RNA and RNA–protein interactions.^{68–71} RNPs are biochemically reconstituted from cell extracts or with recombinant proteins and modified RNA. After long-wavelength UV irradiation, the crosslink sites within RNAs were detected by monitoring accumulation of prematurely terminated cDNA products from stalling of reverse transcriptase in primer extension assays. The reconstitution of RNPs was also accomplished in living cells by transfection of modified RNAs followed by crosslinking and identification of the crosslinked protein partners.⁷²

Some photoactivatable nucleosides are readily bioavailable by providing it in tissue culture medium for uptake.^{73,74} 4SU is imported into metazoan cells by at least one member of equilibrative nucleoside transporters (ENT1 and SLC29A1).⁷⁵ Presumably, 6SG is taken up by the same or a related mechanism. 4SU-labeling of RNA can also be achieved in mammalian cells as well as tissue- and developmental stage-specific fashion using transgenic animals by conditionally expressing *Toxoplasma* uracil phosphoribosyltransferase (UPRT) and feeding of 4-thiouracil.^{76–78} UPRT catalyzes the conversion of uracil analogs and 5-phosphoribosyl-1-R-diphosphate to uridine analog monophosphates. Thioribonucleosides are randomly incorporated into nascent RNA,

substituting a fraction of the corresponding natural nucleosides. The incorporation rate of modified nucleosides depends on various factors including the identity of the nucleoside, nucleoside concentration, incubation time, and cell type. Consequently, modified nucleoside incorporation rates should be determined for a given experimental condition by recovering RNA and quantifying the nucleotide composition (Figure 3(c)). Ideally, the fraction of incorporation is adjusted to accomplish one crosslink for every interacting RNA segment, but bioavailability and toxicity are limiting. Incorporation rates of up to 4% 4SU relative to uridine were obtained in cultured human cells (i.e., 1 substitution for every 25 Us) without measurable change in gene expression or toxicity.^{59,78} The direct application of 4SU in whole animal models like mice has not been thoroughly tested. We suspect that achieving 4SU incorporation in the tissue(s) of choice must balance optimal delivery with toxicity. Thus the use of photoactivatable nucleosides remains a tractable approach for cell culture. One way to circumvent this limitation is to work with cultured primary cells or cell lines that closely approach or match the transcriptome of interest.

Photoactivatable ribonucleoside-enhanced CLIP (PAR-CLIP)⁵⁹ takes advantage of the random transcriptome-wide incorporation of photoactivatable ribonucleoside analogs into newly synthesized RNA in living cells and their subsequent crosslinking to interacting proteins. The current versions of PAR-CLIP use 4SU and 6SG (Figure 4(a)). The photocrosslinking efficiency is dependent on the ribonucleoside analog, the sequence composition of the RRE and its flanking sequence, and the amino acid composition of the protein surface. In general, crosslinking efficiency of modified RNA exceeds the efficiency of UV-254 crosslinking using the same energy dose.^{59,63} However, only a small number of proteins were analyzed in a comparative manner. PAR-CLIP of IGF2BP1, with a 4SU substitution ratio between 2 and 4%, yielded several hundred folds more crosslinked RNA than UV-254 irradiated cells grown without 4SU.⁵⁹ Under similar conditions, another study reported AGO2 crosslinked twofold better while ELAVL1 (HuR) crosslinked with similar efficiency as the UV-254 method.⁶³ The crosslinking efficiencies were estimated based on the intensity of the radioactive RNA signals obtained upon resolving the crosslinked and IPed RBPs on SDS-denaturing protein gels. Ultimately, however, the yield of reverse transcribed full-length cDNA from UV-254 or -365 crosslinked RNA needs to be determined to better judge which of the crosslinking methods is more efficient at binding site determination.

Use of 4SU and 6SG leads to cDNA Sequence Changes that Define the Position of Crosslinking

Sequence analysis following a 4SU PAR-CLIP experiment unexpectedly revealed frequent, non-random mutations of T to C in clusters of overlapping sequence reads that aligned to the reference genome. Typically, more than 50–70% of the sequence reads corresponding to a crosslinked RBP-binding site show T to C changes at U positions within or in close proximity of the RRE (Table 3). Sequence reads without the characteristic T to C change likely represent reads from non-crosslinked RNA background. An analysis using synthetic 100% 4SU-substituted oligoribonucleotides demonstrated that RT of non-crosslinked 4SU leads to about 20% misincorporation of dG across from 4SU, likely due to tautomerism between the thiol and thione form. The dG misincorporation ratio increased to over 90% incorporation upon crosslinking of 4SU to protein, presumably due to changed hydrogen donor–acceptor and base stacking properties (Figure 4(b)).⁵⁹ Non-crosslinked RNA recovered from 4SU-treated cells is therefore expected to yield less than 1% T to C changes assuming equal nucleotide composition and a 4% 4SU incorporation ratio.

When using 6SG, the crosslinking sites were discerned as a G to A change, albeit the frequency of these changes was much reduced (26% G to A frequency for a 6SG IGF2BP1 PAR-CLIP).⁵⁹ The crosslinked 6SG residue represents a structurally different adduct for RT

as compared to 4SU. Therefore, crosslinked 6SG either shows an intrinsically altered deoxynucleotide misincorporation frequency, or it is the result of a reduced probability to read through the crosslinking lesion and a dominance of non-crosslinked background signal.

The T to C (or G to A) transition frequency of clusters of sequence reads is used to separate clusters derived from crosslinked RNA segments from those originating from non-crosslinked background. The background is typically composed of error-free reads matching abundant cellular RNAs, e.g., rRNA, miRNA, tRNA, or bacterial RNAs present in preparations of recombinant RNAligases used in the protocol. For typical mRNA-binding RBPs, such as IGF2BPs, PUM2, or QKI, the fraction of reads mapping to mRNAs (intronic and exonic) varied between 1.6 and 7.6% relative to all reads mapping to the human genome. However, whereas the reads corresponding to each category of background RNAs are basically error-free, clusters of reads corresponding to mRNAs show the 50–70% T to C transition frequency characteristic of crosslinked RNA and prototypical mRBPs.

The frequency of deletions induced by UV-254 in regular CLIP was reported to be as high as 20% in clusters of sequence reads.⁶⁴ However, this number was determined by counting the non-redundant sequences per cluster of reads and cannot be directly compared to those reported for PAR-CLIP—since PAR-CLIP considers read frequency in its accounting. Given its high rate of T to C changes, PAR-CLIP requires less sequence reads to capture crosslink evidence compared to UV-254 CLIP approaches. It readily identifies thousands to hundreds of thousands of binding sites⁵⁷ and is well suited for characterization of RBPs binding with lower sequence specificity and to many sites. Since UV-254 CLIP also generates a signal predominantly based on nucleotide deletions from crosslinked uridines, the sequence bias intrinsic to the use of 4SU in PAR-CLIP is the same. The use of 6SG by PAR-CLIP already represents one approach on how to overcome a uridine bias and it is conceivable that other modified nucleotides may reduce the current nucleotide biases in defining crosslinking sites.

Computational Analysis

Solexa sequencing of a cDNA library generated by the CLIP procedure currently yields over 20 Mio sequence reads and several bioinformatic pipelines for data analysis have been described.^{79,80} Sequence reads are categorized by alignment to genome and reference databases annotating mRNAs, rRNAs, tRNAs, and so forth. Mapping of sequence reads has to allow for at least one error (substitution, insertion, or deletion) to capture the reads arising from crosslinked RNA and to calculate the crosslink-induced mutation ratios (T to C and G to A). Allowing for more than one error can increase the percentage of mapped sequences, but can also increase the incidence of assigning sequence reads to multiple genomic locations, only one of which represents its origin. Sequence reads of >20 nt length are preferable to minimize mapping to multiple locations. Aligning the sequences to both transcriptome and genome reference sequences allows the annotation of intronic, exonic, exon–intron, and exon–exon sequences, the ratios of which can provide insight into the potential roles and subcellular localizations of RBPs. PAR-CLIP experiments using mRBPs typically yield thousands to hundreds of thousands of clusters of overlapping sequence reads representing RBP-binding sites (Figure 5).

Occasional single-nucleotide T to C (or G to A) polymorphisms present within the background of non-crosslinked RNAs may be misinterpreted as crosslinking sites, and clusters of reads with precisely 50 or 100% mutation frequency at a single site in diploid cell lines should be considered cautiously. RNAseq analysis of the transcriptome of the corresponding cell line or tissue can assist in identifying these polymorphic variations and in masking these clusters. Relating RNAseq transcript abundance to read numbers from crosslinked sites also provides a measure for the affinity of a RBP to its target sites and it is

also useful to determine the level of transcript abundance required for a saturating discovery of RNA-binding sites.⁵⁸

To facilitate the discovery of underlying RREs by motif-finding programs, it may be useful to rank clusters and only use a few hundred of the top clusters. PAR-CLIP clusters can be ranked according to the number of different positions that show the characteristic sequence change upon crosslinking, the fraction of reads with the crosslink-induced mutation frequency, and the total number of reads defining a cluster. When multiple binding sites occur in close proximity, the corresponding sequence reads will overlap and generate large clusters. These clusters can be resolved into individual binding sites by crosslink peak-finding approaches. One software tool that accomplishes this task is PARalyzer, developed by Öhler and colleagues (<http://www.genome.duke.edu/labs/ohler/research/PARalyzer/>). The program was first applied to isolate short oligoU-rich RREs found in AU-rich regions of 3' UTRs and intronic sequences bound by ELAVL1.⁵⁷ This approach is also useful for studying RBPs that contain more than one RBD, and/or repetitions of the same domain (Figure 1(b)).

There are several algorithms that can search for common sequence motifs, among them (Phylo)Gibbs, MEME, MDScan, Consensus, Weeder, cERMIT, or Gimsan.^{81–88} Many RBPs contain RBDs that typically recognizes short RREs, e.g., 4 nt in KH, 4–5 nt in RRM.³ For example, NOVA proteins, which contain 3 KH domains recognize the degenerate HYCAU (where H denotes A, C, or U and Y denotes C or U),²⁸ the IGF2BPs with 4 KH and 2 RRM domains recognize CAUH, and QKI with its single KH domain binds to AYUAAU-containing sequences⁵⁹ (Figure 5).

Motif-finding based on primary sequence may not always be successful because certain RBPs, such as those containing dsrm RBDs predominantly recognize structural elements without primary sequence conservation. Secondary structure prediction algorithms such as Mfold⁸⁹ may be used in conjunction with an analysis of the effects of sequence mutation on the RBP–RNA interaction. In binding sites that exhibit sufficient evolutionary conservation, it may be possible to determine if base-pair covariation⁹⁰ reliably predicts RNA secondary structure elements within crosslinked sequences.

It may ultimately be necessary to confirm predicted RREs and the contribution of flanking sequences to protein binding. For these validation experiments, recombinant protein is required such that gel shift or similar binding studies can be performed. Analyses using di- and tri-nucleotide repeat sequences corresponding to identified RREs, as well as the testing of natural binding sites and its sequence variants will ideally corroborate the results of the PAR-CLIP.

As the various CLIP methods gain broader use, they will likely identify a wider variety of sequences that are recognized by a single RBP, which has been underestimated since only *in vitro* characterized sequences with highest affinity have been considered. The fact that CLIP methods deeply capture RNA–protein interactions *in vivo* changes the relevance of motif discovery from target prediction to characterization of RRE usage and occupancy, affinity differences between RBDs, and consequences to PTGR.

TECHNICAL CHALLENGES AND LIMITATIONS IN PAR-CLIP AND UV-254 CLIP PROTOCOLS

Crosslinking Efficiency and the Influence of RNA and Protein Sequence and Structure

Irradiation of 4SU- and 6SG-modified RNAs by long-wavelength UV triggers photoaddition reactions, which require electronic orbital overlap between reactive amino acid side chains

and the excited nucleotide. Therefore not only nucleotide and protein composition are critical, but also the relative orientation of the reaction partners. This has been well documented from analysis of PAR-CLIP crosslinking patterns.⁵⁹ QKI protein with its RRE AYUAAAY preferably crosslinked to U at position 2 (U2), but rarely to the U at position 3 (U3). QKI-binding sites having cytidine at position 2 (C2) were identified from crosslinking to non-conserved Us immediately adjacent to the RRE, but not by crosslinking to U3. Therefore, U3 is situated in a structural environment not optimal for crosslinking. Similarly, a preference for crosslinking to U7 but not U1 or U3 of the PUM2 RRE UGUANAUA was observed with sequence variants at position 7 showing crosslinking predominantly outside of the RRE. These examples illustrate that 4SU residues flanking the RRE can participate in crosslinking and it is not necessary to have U-rich RREs for identification of binding sites. It is interesting to note that only 0.4% of the human transcriptome is composed of 32-nt windows devoid of any U residue. Although it was previously argued that 4SU PAR-CLIP is a U-biased method in comparison to UV-254 crosslinking, the predominant signal in the UV-254 protocol also involves crosslinked uridines that characteristically lead to deletions. The use of 6SG provides some remedy to this shortcoming, but as discussed before, the G to A transition events in cDNA are less frequent and may require deeper sequencing. It will therefore be interesting to explore other photoactivatable nucleotides and their respective signature on crosslinking and RT.

RNase Treatment

The choice of RNase and the extent of RNase treatment affect the RNA background as well as the recovery and recognition of crosslinked sites. RNase digestion of cell lysates is required for fragmentation of large mRNPs to smaller subunits facilitating IP and reducing the risk of detection of RNA-dependent interactions with other RBPs. Furthermore, reduction of the RNA fragments close to the borders protected by the footprint of the RBP facilitates the identification of the binding sites. We find that RNA of 20–35 nt length is optimal for transcriptome-wide definition of binding sites for multiple reasons. (1) RNA of this size window avoids spreading of the RBP–RNA band on the denaturing SDS gels. (2) Shorter RNAs are more efficiently converted to small RNA cDNA libraries due to better adapter ligation. (3) Sequencing reads of up to 50 nt are obtained using the cost- and time-effective Solexa short-read sequencing approach. Full-length sequencing is necessary to capture the mutations characteristic for crosslinking. (4) Sequence reads of 20 or more nucleotides tend to map with little ambiguity to single genomic loci at the required error distance of 1. (5) Short RNA segments also reduce the chances of copurifying other RBPs independently bound in vicinity to RNA segments occupied by the RBP of interest.

RNases are highly processive enzymes and it is non-trivial to adjust reaction conditions to reproducibly generate a desired 20–35 nt crosslinked RNA peak. The RNases must be sufficiently active given that these enzymes will be used in cell lysates containing abundant non-coding RNAs, such as rRNA and tRNA that will compete with mRNAs and pre-mRNAs. However, to minimize the chances for overdigesting the input RNAs, RNases with sequence specificity are preferable, such as RNase A, which cleaves at the 3' end of pyrimidines or RNase T1 that cleaves at the 3' end of guanosines; both enzymes lead to the formation of products with 3' phosphates via a 2',3'-cyclic phosphate intermediate and 5' hydroxyls. Our laboratory prefers the use of RNase T1 because of its high activity and good specificity cleaving only after G. Even when RNase digestion is allowed to proceed to completion, crosslinked RNA segments, which are free of Gs around the borders of the RRE are recovered, and allow for the initial identification of the unknown RRE. Subsequently, the RNase digestion conditions may be altered to capture all possible target sites. The initial definition of the RRE under stringent RNase T1 digestion conditions also facilitates the

annotation of sites obtained under less stringent conditions and the realization of the presence of other sites due to contaminating crosslinked RBPs.

Other laboratories developed a strategy of using a variety of other endoribonucleases in place of, or in conjunction with, RNase T1, such as micrococcal nuclease,⁵⁰ RNase I,⁵⁴ and RNase A.⁴⁵ To date, only one study has compared the potential biases in the recovery of RNA-binding sites introduced by various RNases or digestion conditions using the RBP ELAVL1,⁶³ which recognizes AU-rich sequence elements. As anticipated, overdigestion by RNase T1 led to marked depletion of G-rich clusters as compared to conditions using minimal RNase T1 digestion alone or in combination with micrococcal nuclease.

PRIORITIZING TARGETS BOUND BY RBPs

Application of CLIP or PAR-CLIP frequently identified thousands to tens of thousands of clusters of sequence reads considered to represent binding sites for a given RBP (see references in Table 2). The ranking of targets to explain a specific phenotype resulting from depletion, overexpression, or mutation of the RBP represents a challenge and additional assays are required. For example, the role of RBPs in regulation of transcript stability or alternative splicing has been demonstrated by the correlation of crosslinking sites with regulated targets upon overexpression or knockdown of the RBP of interest using mRNA microarrays (whole genome or exon arrays) or RNAseq experiments.^{59,91,92} Crosslinking captures stable as well as transient RNA–RBP interactions, and ranking target sites according to their affinity or occupancy by the RBP may provide further clues to the identification of the most physiologically important targets. RIP-Chip analyses and its correlation with CLIP target sites can provide such ranking.⁵⁷ The increasing availability of quantitative high-throughput proteomic methods, including (pulsed) stable isotope labeling with amino acids in cell culture (SILAC and pSILAC), enables the analysis of translational regulation by RBPs.⁵⁸ Results from proteomic analysis of miRNA knockdown in HEK293 cells⁹³ were shown to correlate with targets comprising PAR-CLIP-defined miRNA-binding sites.⁵⁹ Ribosome profiling based on sequencing may also offer an alternative way to proteomic methods and an assessment of translational regulation.^{94,95} The functional annotations of RBPs participating in mRNA transport and localization will require the development or application of more specific assays.

CONCLUSION

As CLIP technologies continue to be applied and the number of RBP–RNA interactions increases, one of the next challenges for the field is the spatiotemporal resolution of these interactions. Figure 6 shows an example of an interaction map organized by nuclear/cytoplasmic distributions of RBPs. Furthermore, different cell types express different subsets of transcripts as well as RBPs, and different cellular growth conditions (stress, external stimuli, etc.) may result in post-translational modifications of RBPs, such as those known to occur for ELAVL1/HuR⁹⁶ and FMR1.^{97,98} Given that abundant RBPs are frequently expressed as multimember gene families, assessment of the specific biochemical as well as regulatory function will be labor-intensive and similar to studies addressing the role of multicopy and/or cisronically organized miRNAs.⁹⁹

REFERENCES

1. Cooper TA, Wan L, Dreyfuss G. RNA and disease. *Cell*. 2009; 136:777–793. [PubMed: 19239895]
2. Finn RD, Mistry J, Tate J, Coghill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, et al. The Pfam protein families database. *Nucleic Acids Res*. 2009; 38:D211–D222. [PubMed: 19920124]

3. Mackay JP, Font J, Segal DJ. The prospects for designer single-stranded RNA-binding proteins. *Nat Struct Mol Biol.* 2011; 18:256–261. [PubMed: 21358629]
4. Sonnhammer EL, Eddy SR, Durbin R. Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins.* 1997; 28:405–420. [PubMed: 9223186]
5. Timchenko LT, Timchenko NA, Caskey CT, Roberts R. Novel proteins with binding specificity for DNA CTG repeats and RNA Cug repeats: implications for myotonic dystrophy. *Hum Mol Genet.* 1996; 5:115–121. [PubMed: 8789448]
6. Yang M. JAZ requires the double-stranded RNA-binding zinc finger motifs for nuclear localization. *J Biol Chem.* 1999; 274:27399–27406. [PubMed: 10488071]
7. Lai WS, Stumpo DJ, Blackshear PJ. Rapid insulin-stimulated accumulation of an mRNA encoding a proline-rich protein. *J Biol Chem.* 1990; 265:16556–16563. [PubMed: 2204625]
8. DuBois RN, McLane MW, Ryder K, Lau LF, Nathans D. A growth factor-inducible nuclear protein with a novel cysteine/histidine repetitive sequence. *J Biol Chem.* 1990; 265:19185–19191. [PubMed: 1699942]
9. Siomi H, Matunis MJ, Michael WM, Dreyfuss G. The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. *Nucleic Acids Res.* 1993; 21:1193–1198. [PubMed: 8464704]
10. Aravind L, Koonin EV. G-patch: a new conserved domain in eukaryotic RNA-processing proteins and type D retroviral polyproteins. *Trends Biochem Sci.* 1999; 24:342–344. [PubMed: 10470032]
11. Mayes AE, Verdone L, Legrain P, Beggs JD. Characterization of Sm-like proteins in yeast and their association with U6 snRNA. *EMBO J.* 1999; 18:4321–4331. [PubMed: 10428970]
12. Kiledjian M, Dreyfuss G. Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. *EMBO J.* 1992; 11:2655–2664. [PubMed: 1628625]
13. Bertolotti A, Lutz Y, Heard DJ, Chambon P, Tora L. hTAF(II)68, a novel RNA/ssDNA-binding protein with homology to the pro-oncoproteins TLS/FUS and EWS is associated with both TFIID and RNA polymerase II. *EMBO J.* 1996; 15:5022–5031. [PubMed: 8890175]
14. Jones MR, Quinton LJ, Blahna MT, Neilson JR, Fu S, Ivanov AR, Wolf DA, Mizgerd JP. Zcchc11-dependent uridylation of microRNA directs cytokine expression. *Nat Cell Biol.* 2009; 11:1157–1163. [PubMed: 19701194]
15. Heaton JH, Dlakic WM, Dlakic M, Gelehrter TD. Identification and cDNA cloning of a novel RNA-binding protein that interacts with the cyclic nucleotide-responsive sequence in the type-1 plasminogen activator inhibitor mRNA. *J Biol Chem.* 2001; 276:3341–3347. [PubMed: 11001948]
16. St Johnston D, Beuchle D, Nüsslein-Volhard C. Stauf, a gene required to localize maternal RNAs in the Drosophila egg. *Cell.* 1991; 66:51–63. [PubMed: 1712672]
17. Liu Q, Fischer U, Wang F, Dreyfuss G. The spinal muscular atrophy disease gene product, SMN, and its associated protein SIP1 are in a complex with spliceosomal snRNP proteins. *Cell.* 1997; 90:1013–1021. [PubMed: 9323129]
18. Ponting CP. Novel eIF4G domain homologues linking mRNA translation with nonsense-mediated mRNA decay. *Trends Biochem Sci.* 2000; 25:423–426. [PubMed: 10973054]
19. Kataoka Y, Takeichi M, Uemura T. Developmental roles and molecular characterization of a Drosophila homologue of Arabidopsis Argonaute1, the founder of a novel gene superfamily. *Genes Cells.* 2001; 6:313–325. [PubMed: 11318874]
20. Grishok A, Pasquinelli AE, Conte D, Li N, Parrish S, Ha I, Baillie DL, Fire A, Ruvkun G, Mello CC. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C elegans developmental timing. *Cell.* 2001; 106:23–34. [PubMed: 11461699]
21. Souillard M, Valle Della V, Siomi MC, Piñol-Roma S, Codogno P, Bauvy C, Bellini M, Lacroix JC, Monod G, Dreyfuss G. hnRNP G: sequence and characterization of a glycosylated RNA-binding protein. *Nucleic Acids Res.* 1993; 21:4210–4217. [PubMed: 7692398]
22. Ono Y, Ohno M, Shimura Y. Identification of a putative RNA helicase (HRH1), a human homolog of yeast Prp22. *Mol Cell Biol.* 1994; 14:7611–7620. [PubMed: 7935475]
23. Moss EG, Lee RC, Ambros V. The cold shock domain protein LIN-28 controls developmental timing in C. elegans and is regulated by the lin-4 RNA. *Cell.* 1997; 88:637–646. [PubMed: 9054503]

24. Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, Lindblad-Toh K, Lander ES, Kellis M. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature*. 2005; 434:338–345. [PubMed: 15735639]
25. Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*. 1990; 249:505–510. [PubMed: 2200121]
26. Levine TD, Gao F, King PH, Andrews LG, Keene JD. Hel-N1: an autoimmune RNA-binding protein with specificity for 3' uridylate-rich untranslated regions of growth factor mRNAs. *Mol Cell Biol*. 1993; 13:3494–3504. [PubMed: 8497264]
27. Galarneau A, Richard S. Target RNA motif and target mRNAs of the Quaking STAR protein. *Nat Struct Mol Biol*. 2005; 12:691–698. [PubMed: 16041388]
28. Buckanovich RJ, Darnell RB. The neuronal RNA binding protein Nova-1 recognizes specific RNA targets in vitro and in vivo. *Mol Cell Biol*. 1997; 17:3194–320. [PubMed: 9154818]
29. Ray D, Kazan H, Chan ET, Peña Castillo L, Chaudhry S, Talukder S, Blencowe BJ, Morris Q, Hughes TR. Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. *Nat Biotechnol*. 2009; 27:667–670. [PubMed: 19561594]
30. Keene JD, Komisarow JM, Friedersdorf MB. RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nat Protocols*. 2006; 1:302–307.
31. 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]
32. Zhao J, Ohsumi TK, Kung JT, Ogawa Y, Grau DJ, Sarma K, Song J-J, Kingston RE, Borowsky M, Lee JT. Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol Cell*. 2010; 40:939–953. [PubMed: 21172659]
33. Morris AR, Mukherjee N, Keene JD. Systematic analysis of posttranscriptional gene expression. *Wiley Interdiscip Rev Syst Biol Med*. 2010; 2:162–180. [PubMed: 20836020]
34. Sephton CF, Cenik C, Kucukural A, Dammer EB, Cenik B, Han Y, Dewey CM, Roth FP, Herz J, Peng J, et al. Identification of neuronal RNA targets of TDP-43-containing ribonucleoprotein complexes. *J Biol Chem*. 2011; 286:1204–1215. [PubMed: 21051541]
35. Peng S, Chen L-L, Lei X-X, Yang L, Lin H, Carmichael GG, Huang Y. Genome-wide studies reveal that lin28 enhances the translation of genes important for growth and survival of human embryonic stem cells. *Stem Cells*. 2011; 29:496–504. [PubMed: 21425412]
36. Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB. CLIP identifies Nova-regulated RNA networks in the brain. *Science*. 2003; 302:1212–1215. [PubMed: 14615540]
37. Sanford JR, Coutinho P, Hackett JA, Wang X, Ranahan W, Caceres JF. Identification of nuclear and cytoplasmic mRNA targets for the shuttling protein SF2/ASF. *PLoS ONE*. 2008; 3:e3369. [PubMed: 18841201]
38. Daughters RS, Tuttle DL, Gao W, Ikeda Y, Moseley ML, Ebner TJ, Swanson MS, Ranum LPW. RNA gain-of-function in spinocerebellar ataxia type 8. *PLoS Genet*. 2009; 5:e1000600. [PubMed: 19680539]
39. Wurtmann EJ, Wolin SL. A role for a bacterial ortholog of the Ro autoantigen in starvation-induced rRNA degradation. *Proc Natl Acad Sci U S A*. 2010; 107:4022–4027. [PubMed: 20160119]
40. Xu M, Medvedev S, Yang J, Hecht NB. MIWI-independent small RNAs (MSY-RNAs) bind to the RNA-binding protein, MSY2, in male germ cells. *Proc Nat Acad Sci*. 2009; 106:12371–12376. [PubMed: 19597149]
41. Tremblay GA, Richard S. mRNAs associated with the sam68 RNA binding protein. *RNA Biol*. 2006; 3:90–93. [PubMed: 17179751]
42. Guil S, Caceres JF. The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nat Struct Mol Biol*. 2007; 14:591–596. [PubMed: 17558416]
43. Polymenidou M, Lagier-Tourenne C, Hutt KR, Huelga SC, Moran J, Liang TY, Ling S-C, Sun E, Wancewicz E, Mazur C, et al. Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. *Nat Neurosci*. 2011; 14:459–468. [PubMed: 21358643]

44. Ayala YM, De Conti L, Avendaño-Vázquez SE, Dhir A, Romano M, D'Ambrogio A, Tollervey J, Ule J, Baralle M, Buratti E, et al. TDP-43 regulates its mRNA levels through a negative feedback loop. *EMBO J*. 2011; 30:277–288. [PubMed: 21131904]
45. Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, Clark TA, Schweitzer AC, Blume JE, Wang X, et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature*. 2008; 456:464–469. [PubMed: 18978773]
46. Chi SW, Zang JB, Mele A, Darnell RB. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature*. 2009; 460:479–486. [PubMed: 19536157]
47. Leung AKL, Young AG, Bhutkar A, Zheng GX, Bosson AD, Nielsen CB, Sharp PA. Genome-wide identification of Ago2 binding sites from mouse embryonic stem cells with and without mature microRNAs. *Nat Struct Mol Biol*. 2011; 18:237–244. [PubMed: 21258322]
48. Zisoulis DG, Lovci MT, Wilbert ML, Hutt KR, Liang TY, Pasquinelli AE, Yeo GW. Comprehensive discovery of endogenous Argonaute binding sites in *Caenorhabditis elegans*. *Nat Struct Mol Biol*. 2010; 17:173–179. [PubMed: 20062054]
49. Sanford JR, Wang X, Mort M, Vanduy N, Cooper DN, Mooney SD, Edenberg HJ, Liu Y. Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Res*. 2009; 19:381–394. [PubMed: 19116412]
50. Yeo GW, Coufal NG, Liang TY, Peng GE, Fu X-D, Gage FH. An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat Struct Mol Biol*. 2009; 16:130–137. [PubMed: 19136955]
51. Xue Y, Zhou Y, Wu T, Zhu T, Ji X, Kwon Y-S, Zhang C, Yeo G, Black DL, Sun H, et al. Genome-wide analysis of PTB-RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping. *Mol Cell*. 2009; 36:996–1006. [PubMed: 20064465]
52. Wolf JJ, Dowell RD, Mahony S, Rabani M, Gifford DK, Fink GR. Feed-forward regulation of a cell fate determinant by an RNA-binding protein generates asymmetry in yeast. *Genetics*. 2010; 185:513–522. [PubMed: 20382833]
53. Konig J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, Turner DJ, Luscombe NM, Ule J. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol*. 2010; 17:909–915. [PubMed: 20601959]
54. Wang Z, Kayikci M, Briese M, Zarnack K, Luscombe NM, Rot G, Zupan B, Curk T, Ule J. iCLIP predicts the dual splicing effects of TIA-RNA interactions. *PLoS Biol*. 2010; 8:e1000530. [PubMed: 21048981]
55. Bohnsack MT, Martin R, Granneman S, Ruprecht M, Schleiff E, Tollervey D. Prp43 bound at different sites on the pre-rRNA performs distinct functions in ribosome synthesis. *Mol Cell*. 2009; 36:583–592. [PubMed: 19941819]
56. Granneman S, Kudla G, Petfalski E, Tollervey D. Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proc Natl Acad Sci U S A*. 2009; 106:9613–9618. [PubMed: 19482942]
57. Mukherjee N, Corcoran DL, Nusbaum JD, Reid DW, Georgiev S, Hafner M, Ascano M, Tuschl T, Ohler U, Keene JD. Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability. *Mol Cell*. 2011; 43:327–339. [PubMed: 21723170]
58. Lebedeva S, Jens M, Theil K, Schwanhauser B, Selbach M, Landthaler M, Rajewsky N. Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR. *Mol Cell*. 2011; 43:340–352. [PubMed: 21723171]
59. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M, Jungkamp A-C, Munschauer M, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*. 2010; 141:129–141. [PubMed: 20371350]
60. Hoell JI, Larsson E, Runge S, Nusbaum JD, Duggimpudi S, Farazi TA, Hafner M, Borkhardt A, Sander C, Tuschl T. RNA targets of wild-type and mutant FET family proteins. *Nature Struct Mol Biol*. 2011; 18:1428–1431. [PubMed: 22081015]

61. Brimacombe R, Stiege W, Kyriatsoulis A, Maly P. Intra-RNA and RNA-protein cross-linking techniques in *Escherichia coli* ribosomes. *Methods Enzymol.* 1988; 164:287–309. [PubMed: 3071669]
62. Urlaub H, Hartmuth K, Luhrmann R. A two-tracked approach to analyze RNA-protein crosslinking sites in native, nonlabeled small nuclear ribonucleoprotein particles. *Methods.* 2002; 26:170–181. [PubMed: 12054894]
63. Kishore S, Jaskiewicz L, Burger L, Hausser J, Khorshid M, Zavolan M. A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nat Methods.* 2011; 8:559–564. [PubMed: 21572407]
64. Zhang C, Darnell RB. Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat Biotechnol.* 2011; 29:607–614. [PubMed: 21633356]
65. Wurtmann EJ, Wolin SL. RNA under attack: cellular handling of RNA damage. *Crit Rev Biochem Mol Biol.* 2009; 44:34–49. [PubMed: 19089684]
66. Meisenheimer KM, Meisenheimer PL, Koch TH. Nucleoprotein photo-cross-linking using halopyrimidine-substituted RNAs A2. *Methods Enzymol.* 2000; 318:88–104. [PubMed: 10889982]
67. Favre A, Saintom C, Fourrey J-L, Clivio P, Lauga P. Thionucleobases as intrinsic photoaffinity probes of nucleic acid structure and nucleic acid-protein interactions. *J Photochem Photobiol B Biol.* 1998; 42:109–124.
68. Willis MC, Hicke BJ, Uhlenbeck OC, Cech TR, Koch TH. Photocrosslinking of 5-iodouracil-substituted RNA and DNA to proteins. *Science.* 1993; 262:1255–1257. [PubMed: 7694369]
69. Meisenheimer KM, Meisenheimer PL, Willis MC, Koch TH. High yield photocrosslinking of a 5-iodocytidine (IC) substituted RNA to its associated protein. *Nucleic Acids Res.* 1996; 24:981–982. [PubMed: 8600470]
70. Mishima Y, Steitz JA. Site-specific crosslinking of 4-thiouridine-modified human tRNA(3Lys) to reverse transcriptase from human immunodeficiency virus type I. *EMBO J.* 1995; 14:2679–2687. [PubMed: 7540137]
71. Sontheimer EJ, Steitz JA. The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science.* 1993; 262:1989–1996. [PubMed: 8266094]
72. Kirino Y, Mourelatos Z. Site-specific crosslinking of human microRNPs to RNA targets. *RNA.* 2008; 14:2254–2259. [PubMed: 18719246]
73. Melvin WT, Milne HB, Slater AA, Allen HJ, Keir HM. Incorporation of 6-thioguanosine and 4-thiouridine into RNA. *Eur J Biochem.* 1978; 92:373–379. [PubMed: 570106]
74. Favre A, Moreno G, Blondel MO, Kliber J, Vinzens F, Salet C. 4-Thiouridine photosensitized RNA-protein crosslinking in mammalian cells. *Biochem Biophys Res Commun.* 1986; 141:847–854. [PubMed: 2432896]
75. Yao SY, Ng AM, Vickers MF, Sundaram M, Cass CE, Baldwin SA, Young JD. Functional molecular characterization of nucleobase transport by recombinant human and rat equilibrative nucleoside transporters 1 and 2Chimeric constructs reveal a role for the ENT2 helix 5–6 region in nucleobase translocation. *J Biol Chem.* 2002; 277:24938–24948. [PubMed: 12006583]
76. Miller MR, Robinson KJ, Cleary MD, Doe CQ. TU-tagging: cell type-specific RNA isolation from intact complex tissues. *Nat Methods.* 2009; 6:439–441. [PubMed: 19430475]
77. Cleary MD, Meiering CD, Jan E, Guymon R, Boothroyd JC. Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay. *Nat Biotechnol.* 2005; 23:232–237. [PubMed: 15685165]
78. Dlken L, Ruzsics Z, Radle B, Friedel CC, Zimmer R, Mages J, Hoffmann R, Dickinson P, Forster T, Ghazal P, et al. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA.* 2008; 14:1959–1972. [PubMed: 18658122]
79. Yang J-H, Li J-H, Shao P, Zhou H, Chen Y-Q, Qu L-H. starBase: a database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. *Nucleic Acids Res.* 2011; 39:D202–D209. [PubMed: 21037263]
80. Khorshid M, Rodak C, Zavolan M. CLIPZ: a database and analysis environment for experimentally determined binding sites of RNA-binding proteins. *Nucleic Acids Res.* 2011; 39:D245–D252. [PubMed: 21087992]

81. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol.* 1994; 2:28–36. [PubMed: 7584402]
82. Siddharthan R, Siggia ED, van Nimwegen E. PhyloGibbs: a Gibbs sampling motif finder that incorporates phylogeny. *PLoS Comput Biol.* 2005; 1:e67. [PubMed: 16477324]
83. Lawrence CE, Altschul SF, Boguski MS, Liu JS, Neuwald AF, Wootton JC. Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. *Science.* 1993; 262:208–214. [PubMed: 8211139]
84. Liu XS, Brutlag DL, Liu JS. An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. *Nat Biotechnol.* 2002; 20:835–839. [PubMed: 12101404]
85. Stormo GD, Hartzell GW. Identifying protein-binding sites from unaligned DNA fragments. *Proc Nat Acad Sci.* 1989; 86:1183–1187. [PubMed: 2919167]
86. Pavesi G, Mauri G, Pesole G. An algorithm for finding signals of unknown length in DNA sequences. *Bioinformatics.* 2001; 17(suppl 1):S207–S214. [PubMed: 11473011]
87. Georgiev S, Boyle A, Jayasurya K, Ding X, Mukherjee S, Ohler U. Evidence-ranked motif identification. *Genome Biol.* 2010; 11:R19. [PubMed: 20156354]
88. Ng P, Keich U. GIMSAN: a Gibbs motif finder with significance analysis. *Bioinformatics.* 2008; 24:2256–2257. [PubMed: 18703586]
89. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 2003; 31:3406–3415. [PubMed: 12824337]
90. Eddy SR, Durbin R. RNA sequence analysis using covariance models. *Nucleic Acids Res.* 1994; 22:2079–2088. [PubMed: 8029015]
91. Ule J, Ule A, Spencer J, Williams A, Hu JS, Cline M, Wang H, Clark T, Fraser C, Ruggiu M, et al. Nova regulates brain-specific splicing to shape the synapse. *Nat Genet.* 2005; 37:844–852. [PubMed: 16041372]
92. Landthaler M, Gaidatzis D, Rothballer A, Chen PY, Soll SJ, Dinic L, Ojo T, Hafner M, Zavolan M, Tuschl T. Molecular characterization of human Argonaute-containing ribonucleoprotein complexes and their bound target mRNAs. *RNA.* 2008; 14:2580–2596. [PubMed: 18978028]
93. Fang Z, Rajewsky N. The impact of miRNA target sites in coding sequences and in 3'UTRs. *PLoS ONE.* 2011; 6:e18067. [PubMed: 21445367]
94. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science.* 2009; 324:218–223. [PubMed: 19213877]
95. Guo H, Ingolia NT, Weissman JS, Bartel DP. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature.* 2010; 466:835–840. [PubMed: 20703300]
96. Abdelmohsen K, Pullmann R Jr, Lal A, Kim HH, Galban S, Yang X, Blethrow JD, Walker M, Shubert J, Gillespie DA, et al. Phosphorylation of HuR by Chk2 regulates SIRT1 expression. *Mol Cell.* 2007; 25:543–557. [PubMed: 17317627]
97. Siomi MC, Higashijima K, Ishizuka A, Siomi H. Casein kinase II phosphorylates the fragile X mental retardation protein and modulates its biological properties. *Mol Cell Biol.* 2002; 22:8438–8447. [PubMed: 12446764]
98. Muddashetty RS, Nalavadi VC, Gross C, Yao X, Xing L, Laur O, Warren ST, Bassell GJ. Reversible inhibition of PSD-95 mRNA translation by miR-125a, FMRP phosphorylation, and mGluR signaling. *Mol Cell.* 2011; 42:673–688. [PubMed: 21658607]
99. Ventura A, Young AG, Winslow MM, Lintault L, Meissner A, Erkland SJ, Newman J, Bronson RT, Crowley D, Stone JR, et al. Targeted deletion reveals essential and overlapping functions of the miR-17~92 family of miRNA clusters. *Cell.* 2008; 132:875–886. [PubMed: 18329372]

FIGURE 1. RNA-binding proteins (RBPs) and their RNA-binding domains (RBDs). (a) The most frequently occurring RBDs present in approximately 400 human proteins are shown. Domain names are according to Pfam nomenclature.⁴ RBPs containing more than one type of RBD are counted multiple times. (b) The domain composition of RBPs is further resolved. The top panel indicates single versus repeated occurrence of the same RBD in RBPs. The middle panel resolves combinations of RBDs within RBPs, and the bottom panel indicates how many RBPs with at least one member of RBD are combined with another enzymatically active protein domain such as nuclease, helicase, or with protein–protein

interaction domains. RNases and helicases without auxiliary RBDs were excluded in the bottom panel.

FIGURE 2.

Outline of photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) and other CLIP methodologies. (a) High-throughput sequencing CLIP (HiTS-CLIP, left panel), individual nucleotide resolution CLIP (iCLIP, middle panel), and individual nucleotide resolution crosslinking affinity purification (iCLAP, right panel) utilize short-wavelength ultraviolet (UV) light of 254 nm (UV-254) to crosslink RNA to interacting RNA-binding proteins (RBPs) in living cells or fresh tissue prior to lysis. (b) PAR-CLIP first incorporates photoactivatable thioribonucleosides into nascent transcripts and then crosslinks by using long-wavelength UV-365. Isolation of RNA-RBP complexes is achieved either by immunoprecipitation (IP) (PAR-CLIP, HiTS-CLIP, and iCLIP) or by streptavidin/IMAC double affinity purification (iCLAP). Similar to iCLAP, the crosslinking and analysis of cDNAs (CRAC)⁵⁶ method (not shown) utilizes double affinity enrichment of an RBP in which the second purification is performed under denaturing conditions. The addition of 3' and 5' adapters differ based on the CLIP method. iCLIP and iCLAP methods circularize an ssDNA intermediate to capture sequences generated from prematurely terminated products of reverse transcription (RT) at the crosslinking sites. In PAR-CLIP, RT through a thioribonucleoside crosslinked RNA, followed by polymerized chain reaction

(PCR) amplification, leads to a characteristic mutation [T to C when using 4-thiouridine (4SU) and G to A when using 6-thioguanosine (6SG)]. Scoring for these mutations leads to the differentiation of clusters of crosslinked reads from clusters originating from non-crosslinked background sequences including abundant cellular RNAs.

FIGURE 3.

Photoactivatable ribonucleoside analogs. (a) Halogenated and thione-containing ribonucleoside analogs have been used for photocrosslinking to probe RNA–RNA as well as RNA–protein interactions. (b) Predicted structures of ribonucleoside analogs upon ultraviolet (UV) 365 crosslinking to aromatic amino acid side chains. The structure upon crosslinking to tyrosine is shown. (c) Analysis of the ratio of 4-thiouridine (4SU) incorporation into cellular RNAs. Total RNA (0.2 OD₂₆₀ units) recovered from HEK293 cells cultured in medium supplemented with 0.1 mM 4SU for 16 h, was digested to mononucleosides using snake venom phosphodiesterase and alkaline phosphatase and analyzed by reverse-phase high-performance liquid chromatography (HPLC). The peaks

corresponding to C, U, G, and A were detected at 280 nm and the peak corresponding to 4SU is detected at 330 nm. An incorporation rate of 4% 4SU was calculated based on the absorption coefficients for U and 4SU. 4SU substitution rates of 1–4% were also observed with other cell lines. 5BrU, 5-bromouridine; 5IU, 5-iodouridine; 5IC, 5-iodocytidine; 6SG, 6-thioguanosine.

(a)

(b)

FIGURE 4.

Photoactivated crosslinking of thioribonucleosides. (a) Photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) of HEK293 cells expressing FLAG/hemagglutinin (HA)-tagged IGF2BP2 containing 4 RRM and 2 K homology (KH) RNA-binding domains (RBDs). The top panel shows the autoradiograph of radiolabeled RNA crosslinked to IGF2BP2 and resolved after immunoprecipitation (IP) by protein-denaturing sodium dodecyl sulfate (SDS) gel electrophoresis. The expected molecular weight of the epitope tagged IGF2BP2 is approximately 75 kDa. The bottom panel shows an anti-HA immunoblot (IB) to control for the amount of tagged IGF2BP2 protein loaded on the SDS gel. (b) Top panel: ultraviolet (UV) 254 crosslinking products of uridine and aromatic side chain amino acids (Tyr, Phe, and Trp). Reverse transcriptase typically stalls at the crosslinked adduct, but can occasionally skip the lesion resulting in a deletion within the cDNA corresponding to the crosslinked position. The bottom panel shows UV-365

crosslinking product for 4-thiouridine (4SU) to aromatic side chain amino acids. The structure of the 4SU crosslink product is distinct from the UV-254 crosslinked photoadduct. Reverse transcriptase also stalls at the lesion, but when it reads through it preferably incorporates dG directly opposite the photoadduct, leading to the characteristic T to C transitions defining the sites of 4SU crosslinking.

PUM2

3'UTR of ELF1

AAATGTTTTTAGATTACTTTTCAACTGTAAATAATGTACATTTAATGTCACAAGAAAA	reads	error
-----ATTACTTTTCAACTGTAAACAATGTACATTT-----	581	1
-----ATTACTTTTCAACTGTAAATAATGTACACTT-----	239	1
-----ATTACTTTTCAACTGTAAATAATGTACATTT-----	113	0
-----ACTTTTCAACTGTAAACAATGTACATTTAAT-----	82	1
-----ATTACTTTTCAACTGTAAATAATGTACATCT-----	67	1

QKI

3'UTR of CTNNB1

TGTAGAACACTAATTCATAATCACTCTAATTAATTGTAATCT	reads	error
-----AACACTAATTCATAATCACTCTAATTAACTG-----	81	1
-----AACACTAATTCATAATCACCTAATTAATTGT-----	39	1
-----AACACTAATTCATAATCACTCTAACTAATTGT-----	29	1
-----AACACTAATTCATAATCACTCTAATTAATTGT-----	21	0
-----AACACTAATTCACAATCACTCTAATTAATTGT-----	15	1
-----AACACCAATTCATAATCACTCTAATTAATTGT-----	13	1

IGF2BP1-3

3'UTR of MRPL9

TGTCTCCAGTACTTGCCCTCATTCTCATCATCCAACTGAACATTTGTATCCC	reads	error
-----CCTCATTCTCATCACTCAAACTG-----	18	1
-----CCTCACTCTCATCATCCAACTG-----	15	1
-----CCTCATTCTCATCATCCAACTG-----	11	0
-----CCTCATTCTCACATCCAACTG-----	9	1
-----CCATTCTCATCATCCAACTG-----	9	1

FIGURE 5.

Examples of photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) clusters of overlapping sequence reads and the corresponding RNA recognition elements (RREs). Clusters of overlapping sequence reads obtained from PAR-CLIP of the RNA-binding proteins (RBPs) PUM2, QKI, and IGF2BP family members. Reads were aligned to the reference sequence of the indicated gene transcripts; only the most frequent reads are shown. The sequence read copy numbers and the number of mismatches to the reference sequence are indicated to the right; they all represent T to C mutations (shown in red letters). Blue boxes indicate the location(s) of the RREs, as defined by Phylo-Gibbs motif analysis (right panels).

FIGURE 6.

Photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) RNA-binding proteins (RBPs) binding sites for CDK1 mRNA. The CDK1 mRNA is 1923 nt long (NM_001786.4). This figure illustrates only the exonic binding sites identified for several nuclear and cytoplasmic-localized RBPs. The subcellular localization of each RBP is shown. The reported subcellular distribution of some RBPs would indicate that they are predominantly nuclear (FUS), versus cytoplasmic (PUM2). However, the nuclear to cytoplasmic ratios of RBPs relative to each other are not known, nor where they would associate with their mRNA targets.

TABLE 1

Overview of the Most Frequently Encountered RBDs and Examples of RBPs Comprising These Domains

Domain	Domain Description	Gene Symbol	Gene Alias	Reference
RRM	RRMs are found in a variety of RBPs, including various hnRNP proteins, proteins implicated in regulation of alternative splicing, and protein components of snRNPs.	<i>CELF1</i>	<i>CUGBP1</i>	5
zf-C2H2_jaz	Mammalian members of this group occur multiple times along the protein, joined by flexible linkers, and are referred to as JAZ—dsRNA-binding ZF protein—zinc fingers.	<i>ZNF346</i>	<i>JAZ, Zfp346</i>	6
zf-CCCH	Zinc finger C-x8-C-x5-C-x3-H type (and similar). Found in many known RBPs.	<i>ZFP36</i>	<i>TTP; TIS11; NUP475; RNF162A</i>	7,8
KH	K homology RBD, type I. KH binds single-stranded RNA. There are two different KH domains that belong to different protein folds, but they share a single KH motif.	<i>HNRNPK</i>	<i>CSBP; TUNP; HNRPK</i>	9
G-patch	Found in a number of RBPs, and is also found in proteins that contain RBDs. This domain has seven highly conserved glycines.	<i>RBM5</i>	<i>LUCA15, LUCA</i>	10
LSM	The like SM (LSM) domain contains Sm proteins as well as other related LSM (like Sm) proteins.	<i>LSM1-7</i>		11
SAP	The SAP (after SAF-A/B, Acinus, and PIAS) motif is a putative DNA/RBD found in diverse nuclear and cytoplasmic proteins.	<i>HNRNPU</i>		12
zf-RanBP	Zn-finger found in Ran-binding protein and others.	<i>TAF15</i>	<i>Npl3, RBP56, TAF2N, TAFII68, hTAFII68</i>	13
zf-CCHC	The zinc knuckle is a zinc-binding motif composed of the following CX2CX4HX4C where X can be any amino acid.	<i>ZCCHC11</i>	<i>PAPD3, TUT4</i>	14
HABP4_PAIRBP1	This family includes the hyaluronan-binding protein 4 (HABP4) family of hyaluronan-binding proteins, and the PAI-1 mRBP, PAIRBP1.	<i>SERBP1</i>	<i>CGI-55, CHD3IP, HABP4L, PAIRBP1, PAIRBP1</i>	15
dsrm	Double-stranded RNA-binding motif. Found in a variety of proteins including dsRNA-dependent protein kinase protein kinase R (PKR), RNA helicases, Drosophila staufen protein, <i>E. coli</i> RNase III, RNases H1, and dsRNA-dependent adenosine deaminases.	<i>STAU1</i>	<i>STAU</i>	16
TUDOR	Often thought to act as protein–protein interaction domain and some protein members of this domain class recognize dimethylated arginines. For some family members RNA binding has also been observed.	<i>SMN1</i>	<i>SMN, SMN2, SMNT</i>	17
MIF4G	MIF4G is named after middle domain of eukaryotic initiation factor 4G (eIF4G). The domain is rich in α helices and may contain multiple α -helical repeats.	<i>EIF4G</i>	<i>EIF-4G1, EIF4F, EIF4G, EIF4GI, P220</i>	18
PAZ	This domain is named PAZ after the proteins Piwi Argonaut and Zwillig. PAZ can bind the characteristic two-base 3' overhangs of siRNAs.	<i>EIF2C1-4</i>	<i>AGO1-4, Argonaute 1-4</i>	19,20
RBM1CTR	This C-terminal region is found in RBM1-like RNA-binding hnRNPs.	<i>RBMX</i>	<i>HNRPG, RBMXP1, RBMXRT, RNMX hnRNP-G</i>	21
S1	The S1 domain occurs in a wide range of RNA-associated proteins.	<i>DHX8</i>	<i>DDX8, HRH1, PRP22, PRPF22</i>	22

Domain	Domain Description	Gene Symbol	Gene Alias	Reference
Piwi	This domain is found in the protein Piwi and its relatives. The function of this domain is the dsRNA guided hydrolysis of ssRNA.	<i>EIF2C1-4</i>	<i>AGO1-4, Argonaute 1-4</i>	19,20
cold shock domain (CSD)	RBD that functions as a RNA chaperone in bacteria and is involved in regulating translation in eukaryotes.	<i>LIN28B</i>	<i>CSDD2</i>	23

RBDs, RNA-binding domains; RBPs, RNA-binding proteins; RNPs, ribonucleoprotein complexes.

TABLE 2

Large-Scale Methods for the Discovery of RBP/RNP RNA Targets

Method	Goal	Description	Selection of RBPs Analyzed
Non-crosslinking methods			
RIP-Chip	Target gene identification	IP of epitope tagged or endogenous RBP/RNP to isolate-associated RNAs. Isolated RNAs are analyzed by microarray. RBP/RNP targets are defined by calculating and scoring according to enrichment of IP'd transcripts over control expression values.	For comprehensive list, see Table 1 in Ref 33.
RNA IP and high throughput sequencing (RIP-Seq)	Target gene identification	Same procedure as RNA IP followed by microarray analysis (RIP-Chip); IP'd RNA is quantified by RNAseq. RBP/RNP targets are defined, similar to RIP-Chip, by calculating enrichment scores over control.	TDP-43, ³⁴ LIN-28, ³⁵ Polycomb proteins ³²
Photocrosslinking methods			
CLIP	Target site identification; definition of RRE	Ultraviolet (UV) 254 photocrosslinking of RNA to RBPs in live cells or tissues prior to lysis. Crosslinked RNA is trimmed by RNases and RNA–protein complexes are fractionated by Sodium dodecyl sulfate polyacrylamide gel electrophoresis. Crosslinked RNAs are isolated, 3' and 5' adapters for RT and polymerized chain reaction ligated. The cDNA library is inserted into plasmid for bacterial transformation, cloning, and sequencing.	Nova, ³⁶ SF2, ³⁷ CUGBP1, ³⁸ Rsr, ³⁹ MSY2, ⁴⁰ SAM68, ⁴¹ hnRNP A1 ⁴²
HiTS-CLIP	Target site identification; definition of RRE	Similar to standard CLIP, however, cDNAs are deep sequenced (454 or Solexa). Overlapping sequence reads are clustered and RBP/RNP targets are defined based on enrichment over negative controls.	TDP-43, ^{43,44} Nova, ⁴⁵ AGO2, ^{46,47} Alg-1, ⁴⁸ SFRS1, ⁴⁹ FOX2, ⁵⁰ PTB, ⁵¹ Khdl ⁵²
iCLIP	Target site identification; definition of RRE	Crosslinking procedures similar to CLIP. Partial cDNAs, generated from stalled RT of crosslinked RNAs, are circularized and ultimately deep sequenced. RBP/RNP targets are similarly defined as in HiTS-CLIP. Sites of crosslinking are interpreted as the (–1) position a sequence read maps to (indicating the putative RT stall sites).	hnRNP C, ⁵³ TIA1, TIAL1 ⁵⁴
iCLAP	Target site identification; definition of RRE	Similar to iCLIP except that the RBP is Strep- and polyhistidine epitope tagged. Streptavidin beads are used in the first purification step. Immobilized metal-ion affinity chromatography (IMAC) under denaturing conditions is performed as a secondary purification. Isolated RNAs are converted into cDNA and sequenced similar to iCLIP.	TIA1, TIAL1 ⁵⁴
CRAC	Target site identification	Similar in concept to CLAP, where a tandem affinity purification protocol is used. RBPs are engineered to contain C-terminal 6X-histidine, tobacco etch virus (TEV) protease site, and Protein A tags. Immunoglobulin G beads are used as a first purification step, followed by TEV protease treatment to elute crosslinked RNA–protein complexes. IMAC is performed under denaturing conditions as a secondary purification. Individual	Prp43, ⁵⁵ U3 snoRNA-binding sites of Nop1, Nop56, Nop58, And Rrp956

Method	Goal	Description	Selection of RBPs Analyzed
Photoactivatable ribonucleoside-enhanced CLIP	Target site identification; definition of RRE	<p>RNAs were analyzed by northern blot or sequenced after amplification with gene-specific primers.</p> <p>Incubation of live cells with photoactivatable ribonucleosides (4-thiouridine and 6-thioguanosine) that are incorporated into nascent transcripts. UV-365 photocrosslinking of RNA to proteins prior to lysis. IP of epitope tagged or endogenous RBP/RNP to isolate-associated RNAs. cDNA library preparation of isolated RNAs. Putative RBP/RNP target sites are scored by frequency of crosslinking evidence, seen as a characteristic T2C (or G2A) nucleotide mutation.</p>	<p>ELAVL1,^{57,58} QKI,⁵⁹ IGF2BP1–3,⁵⁹ AGO1–4,⁵⁹ PUM2,⁵⁹ TNRC6A–C,⁵⁹ FUS,⁶⁰ EWSR1,⁶⁰ TAF15⁶⁰ In analysis: BICC1, CUGBP1, DHX9, DICER, DND1, EXPORTIN5, FMR1, FXR1, FXR2, ELAVL2–4, LIN28B, MBNL1, MOV10, NCL, P54, RBM20, RBPMS, SND1, TARBP2, TDP-43, TIA1, TSN, TSNAX, TTP, WT1</p>

CLIP, crosslinking and immunoprecipitation; HiTS-CLIP, High-throughput sequencing CLIP; iCLAP, individual nucleotide resolution crosslinking affinity purification; iCLIP, individual nucleotide resolution CLIP; IP, immunoprecipitation; RBP, RNA-binding proteins; RNPs, ribonucleoprotein complexes; RRE, RNA recognition element; RT, reverse transcription.

Summary of Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation (PAR-CLIP) Experiments with Indicated RNA-Binding Proteins (RBPs)

TABLE 3

RBPs	Extracted, Filtered Raw Reads	Mapped	mRNA	Clustered Sequence Reads	mRNA Clusters	Fraction Crosslinked
QKI	11,996,168	4,846,699	738,882	152,546	6193	0.50
PUM2	9,135,298	2,388,700	616,227	176,492	7523	0.70
AGO 1–4	17,489,030	9,718,111	2,310,159	857,163	17,319	0.76
IGF2BP 1–3	56,516,935	28,478,253	9,928,049	4,816,921	128,536	0.58

Adaptor extracted and low-entropy filtered sequences are defined as raw reads. Mapped reads are sequences that unambiguously map to the genome with 1 error (substitution, insertion deletion). mRNA refers to the total number of reads that annotate as mRNA. Clustered Sequence Reads are the number of overlapping mRNA sequence reads clustering with 5 uniquely mapping, redundant sequence reads per cluster. mRNA Clusters refers the number of discrete groups of overlapping, uniquely mapping mRNA sequence reads. Fraction Crosslinked refers to the fraction of Clustered Sequence Reads containing the characteristic T to C mutation.