Advances in CLIP technologies for studies of protein-RNA interactions

Flora C. Y. Lee[1,2], Jernej Ule[1,2,3]

[1] Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London, WC1N 3BG, UK

[2] The Francis Crick Institute, 1 Midland Road, London, NW1 1AT, UK

[3] Lead contact: jernej.ule@crick.ac.uk

Summary

RNA binding proteins (RBPs) regulate all aspects in the life cycle of RNA molecules. To elucidate the elements that guide RNA specificity, regulatory mechanisms and functions of RBPs, methods that identify direct endogenous protein-RNA interactions are particularly valuable. UV Crosslinking and Immunoprecipitation (CLIP) purifies short RNA fragments that crosslink to a specific protein, and then identifies these fragments by sequencing. When combined with high-throughput sequencing, CLIP can produce transcriptome-wide maps of RNA crosslink sites. The protocol is comprised of several dozen biochemical steps, and improvements made over the last 15 years have increased its resolution, sensitivity and convenience. Adaptations of CLIP are also emerging in the epitranscriptomic field to map the positions of RNA modifications accurately. Here, we describe the rationale for each step in the protocol and discuss the impact of variations to help users determine the most suitable option.

Main Text

Introduction

RNA binding proteins (RBPs) play a role in diverse mechanisms of RNA regulation, from pre-mRNA splicing and 3' end processing to RNA modification, translation, stability and localisation. Most RBPs are localised at specific cellular locations, where they are presented with a unique composition of potential RNA targets and other RBPs that affect the binding patterns through competitive or cooperative interactions. Over a thousand human proteins have been shown to crosslink to RNA by mass spectrometry studies, including RNA enzymes, and proteins that lack canonical RNA-binding domains (Baltz et al., 2012; Castello et al., 2012). In order to disentangle the diverse roles of these proteins, it is necessary to map their *in vivo* binding sites across the transcriptome.

Several methods can be used to identify the endogenous protein-RNA interactions with variable specificity and sensitivity. The first method developed for this purpose used antibodies against the spliceosomal Sm proteins (lupus autoimmune sera) to identify the small nuclear RNAs, which interact with Sm proteins within the abundant spliceosomal small nuclear ribonucleoproteins (Lerner and Steitz, 1979). This method, later referred to as RIP (for RNP/RNA immunoprecipitation), relies on immunoprecipitation (IP) of an RBP under conditions that preserve ribonucleoprotein complexes (RNPs) (Niranjanakumari et al., 2002). RNPs are preserved either due to mild washing conditions during IP, or by exposing cells to formaldehyde, which crosslinks protein-protein and protein-RNA interactions. Subsequent studies used microarrays for transcriptomic analysis of the purified RNAs, and the resulting method has been referred to as RIP-chip (Keene et al., 2006; Tenenbaum et al., 2000). In 2010, RIP was combined with high-throughput sequencing, and termed RIP-seq (Zhao et al., 2010). While RIP can identify abundant RNAs bound by an RNP, it is not well suited to studies of direct protein-RNA contacts. This is because it preserves protein-protein interactions, and thus can purify multiple RBPs in complex with their bound RNAs. Under some conditions it can also identify interactions that result from *in vitro* re-associations (Mili and Steitz, 2004). Therefore, methods with increased specificity for direct RNA binding sites are needed, especially if one wishes to identify binding sites in lowly abundant RNAs.

To identify the position of direct protein-RNA interactions, it is crucial to use a method that preserves endogenous protein-RNA contacts, while ensuring that only a single specific RBP is purified. Crosslinking and immunoprecipitation (CLIP) was developed for this purpose by exploiting zero-length covalent protein-RNA crosslinking and RNA fragmentation (Ule et al.,

2003). This enables CLIP to purify RNAs bound to a specific RBP under conditions that are stringent enough to prevent co-purification of additional RBPs or free RNAs. Moreover, since only the RNA fragments that are crosslinked to the RBP are isolated, CLIP can identify the position of the direct RNA binding sites.

Initially, CLIP relied on Sanger sequencing to identify 340 sequences corresponding to RNA interactions of splicing factors Nova1 and Nova2 in mouse brain (Ule et al., 2003). 244 of these sequences were intronic or intergenic, confirming that CLIP is sensitive enough to efficiently identify binding sites within low-abundance RNAs. Abundant RNAs such as rRNAs were absent, underlining the specificity of the method. Even though the sequences were only approximately 50nt long, they contained on average four Nova-binding motifs, further confirming the high specificity of CLIP data. Several of the sequences were located next to alternative exons that turned out to be regulated by Nova proteins, thus demonstrating the capacity of CLIP to identify the position of functionally important binding sites.

Since the original study, multiple variant protocols have been derived to improve the conditions of RNA fragmentation, RBP purification and cDNA library preparation (Ule et al., 2005), establish denaturing conditions for RBP purification (Granneman et al., 2009), employ high-throughput sequencing (Licatalosi et al., 2008; Yeo et al., 2009), determine the position of crosslink sites at nucleotide resolution (Hafner et al., 2010; König et al., 2010), and increase the efficiency and convenience of the protocol (Table 1, Table 2). Since RIP or ChIP were originally combined with microarray readout, the addition of '-seq' (e.g. RIP-seq) was needed to specify the use of sequencing as opposed to of microarrays. In contrast, the original CLIP and all the derived variants rely on sequencing. We therefore use the term 'CLIP' to refer generically to all protocols that purify covalently crosslinked protein-RNA complexes and then sequence the bound RNA fragments.

We describe the core steps of CLIP, the rationale behind each available variation, and their likely effects on the sensitivity, resolution, specificity or convenience of different protocols (Fig. 1, Table 2). For comparative purpose, we also provide an overview of the steps that are employed by each of the 28 published protocols (Table S1). We outline the basic approaches to assess the sensitivity and specificity of CLIP data, while a comprehensive summary of computational methods for CLIP data analysis is reviewed in more detail elsewhere (Chakrabarti et al., 2017). We conclude with a discussion of the quality control analyses, and of the opportunities to apply CLIP to new purposes.

Core steps of CLIP

Although most steps of CLIP have undergone several variations, the core concepts behind each of the steps and the order of the steps remain largely the same (Fig. 1, Table 1, Table 2). The variants either modify the way the steps are performed, add or omit some of the steps, in order to increase the efficiency or convenience of the protocol. A central aspect of the experimental design common to all protocols is the use of appropriate negative controls, which is important for interpreting the specificity of the purification procedure. The ideal control is to perform the same purification from cells where the RBP is absent, such as knockout cells, or when using tag-based purification, cells that do not express a tagged protein (Huppertz et al., 2014; Ule et al., 2005). As an alternative, non-specific serum or IgG can be used for IP. It is also valuable to immunoprecipitate the RBP from non-crosslinked cells. If CLIP conditions are well optimised, these controls should not produce any clearly detectable signal during SDS-PAGE analysis, and sequencing of their libraries should result in at least 100-fold fewer unique cDNAs compared to the specific experiments (König et al., 2010).

Covalent crosslinking of protein-RNA contacts

Most variants of CLIP exploit the capacity of ultraviolet (UV) light to promote formation of covalent bonds between RBPs and their direct RNA binding sites (Table S1). Unlike the formaldehyde crosslinking that is used in chromatin immunoprecipitation (ChIP) and some variants of RIP, UV does not crosslink proteins to each other. UV crosslinking requires direct contact between an amino acid and a nucleobase, and therefore ensures that only direct protein-RNA interactions are preserved, and the high strength of the covalent bond allows further stringent purification of individual RBPs and their crosslinked RNA fragments.

The original and most later variants of CLIP exposes cells or triturated tissues to the UV-C wavelength (254nm), which can crosslink RBPs to their bound RNAs without the need for any additional pre-treatment (Ule et al., 2003). Cells are placed on ice during the short period of crosslinking in order to avoid any cellular responses, for instance the induction of UV-induced DNA damage response. The recommended crosslinking procedure for cells in a monolayer takes 40 seconds (using an energy of 150mJ/cm^2) (König et al., 2010); this short period allows a snapshot of the interactions to be captured, thus enabling CLIP to monitor changes in RNP assembly that occur upon a response to extracellular signals or other treatments (Schor et al., 2012). A higher total energy can be employed for tissues or cells in suspension, where multiple rounds of UV exposure with intermittent mixing are needed in order to obtain evenly-distributed crosslinking throughout the sample. Some protocols employ high UV-C crosslinking energies also for cells in a monolayer, since this increases crosslinking efficiency

and thus sensitivity of CLIP, but this can also increase a cellular response to UV-induced damage, and the propensity of multiple RBPs to crosslink on the same RNA fragment, thus risking co-purification of contaminating RBPs. In cases of proteins that do not crosslink well to RNA, digestion optimized (DO)-RIP-seq could also be employed to identify fragments of RNAs that are proximal to the protein of interest (Nicholson et al., 2017).

PAR-CLIP introduces a variation in the crosslinking strategy (Hafner et al., 2010) (Table 1). Cells are pre-incubated with photoactivatable ribonucleosides 4-thiouridine (4SU) or 6-thioguanosine (6SG), which enable protein-RNA crosslinking to be performed with UV-A wavelength (365nm). Mass spectroscopy analyses indicate two thirds of RBPs efficiently crosslink with either the standard UV-C (CL) or with the 365nm (PAR-CL), but twice as many RBPs (24% of the interactome) were identified only by CL compared with 12% for PAR-CL (Castello et al., 2012). So far, only one mass spectrometry study has compared the CL and PAR-CL, and therefore the features of RBPs that confer the differential efficiency of these two crosslinking methods remain unclear. The use of 4SU or 6SG restricts crosslinking to a single base, and therefore the crosslinking efficiency might also depend on the proximity of these bases to the binding site. The PAR-CL protocol is limited to biological systems where the photoactivatable nucleosides can be efficiently incorporated. Incorporation of 4SU through liquid culture for C. elegans or intraperitoneal injection for mouse has enabled studies in model organisms, however the incorporation rates are lower than in HEK cells in culture (Jungkamp et al., 2011; Kim et al., 2014), hence they have drawbacks in sensitivity. Moreover, prolonged preincubation with 6SG (and to a lesser extent 4SU) could cause cellular toxicity, and therefore care needs to be taken to monitor the cellular response to these ribonucleosides (Burger et al., 2013; Huppertz et al., 2014). Application of pulsed 4SU has been used in techniques for tagging and enriching nascent RNAs for sequencing (Windhager et al., 2012), and this concept could be combined with CLIP to study the patterns of co-transcriptional RNP assembly on newly transcribed RNAs.

The third approach to crosslinking is introduced by m5C-miCLIP, which exploits a mutant NSun2 RNA methylase enzyme for transcriptome-wide mapping of 5-methylcytosine (m5C) modification sites (George et al., 2017; Hussain et al., 2013) (Table 1). This mutant enzyme is uncapable of completing the methylation, and instead covalently attaches to the RNA base at the site of modification. This approach is combined with iCLIP, which has been developed to amplify cDNAs that truncate at the crosslink site, thus enabling nucleotide-resolution mapping of the crosslink sites (König et al., 2010) (Table 1). As expected, the crosslink sites identified by m5C-miCLIP are enriched in cytosines, rather than uridines that are most common when using UV-C crosslinking in iCLIP (Sugimoto et al., 2012).

Finally, proteins can be crosslinked to RNA with UV light *in vitro*. This has been exploited by variant protocols aimed at studies of RNA methylation, such as m6A-miCLIP (Chen et al., 2015; George et al., 2017; Ke et al., 2015; Linder et al., 2015) (Table 1). Here, RNA is purified and partially fragmented, then incubated with an antibody recognising the N6-Methyladenosine (m6A) modification. Subsequently, UV-C crosslinking is used to form a covalent bond between the antibody and the modified base. The antibody-RNA complex is then captured on protein A/G magnetic beads, and the sample continues to the on-bead adapter ligation and the rest of the iCLIP protocol. Enrichment of the expected sequence motif at the crosslink sites confirmed the high positional accuracy of the resulting data (Linder et al., 2015).

Cell lysis

In almost all CLIP derived protocols, a stringent buffer containing ionic detergents is used for cell lysis, which disrupts most protein-protein and protein-RNA interactions. This increases the accessibility of RNA and allows unbiased RNase fragmentation, by decreasing the chance that long RNA binding sites remain protected by large RNPs. It also minimises the chance of co-purifying multiple associated RBPs during later immunoprecipitation, thus helping to ensure data specificity.

In addition, with methods where the whole cell lysate is used as the input, the proportions of the different types of RNAs in the resulting data can inform on the cellular distribution of the RBP. For example, predominance of intronic reads can indicate that the RBP primarily binds to nascent RNAs on chromatin, whereas enrichment of exonic and junction-spanning reads indicates that the RBP primarily binds to spliced mRNAs in the cytoplasm. For studies where the interactome of specific subcellular compartmentalisation is of interest, cell lysis can be adapted to accommodate the fractionation of sub-cellular compartments. The first protocol developed for this purpose produced CLIP data from nuclear, cytosolic and polysome fractions (Sanford et al., 2008). More recently the Fr-iCLIP method has been developed, which fractionates the nucleus into chromatin and nucleoplasm before proceeding to iCLIP (Brugiolo et al., 2017) (Table 1).

RNA fragmentation

RNA fragmentation is crucial to avoid co-purifying multiple RBPs that crosslink to the same RNAs, and to provide insight into the position of RNA binding sites, since the RNA fragment

contains the crosslink site. The variation in RNase concentration is unlikely to lead to major changes in the resulting data, and enriched motifs at the crosslink sites are expected to remain the same within a range of RNase concentrations (Van Nostrand et al., 2016). However, analysis of the crosslink sites identified by various PTBP1 iCLIP experiments revealed that variations in RNase concentrations can lead to changes at the ends of the cDNA inserts, which correspond to the sites of RNase cleavage (Haberman et al., 2017). Such constraints at the ends of the cDNA inserts can impact binding site assignment, especially in the case of long binding sites of RBPs, where appropriate optimisation of RNA fragmentation was found to be particularly important.

Overdigestion results in short RNA fragments, and thus a narrow distribution of cDNA sizes. This can introduce constraints at the ends of the cDNA insert due to the preferred pattern of RNase cleavage, and produce short cDNAs that are less likely to map uniquely to the repetitive regions of the genome (Haberman et al., 2017). On the other hand, insufficient RNA digestion can lead to co-purification of additional RBPs that bind to the long RNA fragments together with the immunoprecipitated RBP. This has been exploited in studies which identify RNA-dependent protein interactors of the RBP-of-interest (Botti et al., 2017; Brannan et al., 2016; Flury et al., 2014; Klass et al., 2013). Most RNAs contain a large number of binding sites for many RBPs, hence long RNA fragments could be crosslinked to multiple RBPs at different positions. Thus, by increasing the presence of co-purified RBPs, long RNA fragments could decrease the specificity of the final data. An optimal RNA size range of 30-200 nt can be achieved with a short incubation of the lysate with a low RNase concentration, which can be optimised by using the visualisation of protein-RNA complexes after SDS-PAGE separation upon a titration of RNase conditions (Huppertz et al., 2014; Ule et al., 2005). This optimisation is important especially when starting experiments with a new RBP, or from a new type of cell or tissue, or when testing new reagent stock.

While most protocols perform RNase treatment in the lysate, several protocols employ on-bead RNase treatment. For example, PAR-CLIP and sCLIP protocols digest with RNase in the lysate as well as after IP (Hafner et al., 2010; Kargapolova et al., 2017) (Table 1). Zarnegar and colleagues compared the effects of performing the RNase digestion step either in the lysate, or on-bead after immunoprecipitation. By using the infrared visualisation in irCLIP (Table 1), the amount of adapter-ligated RNA-protein complexes can be examined on the membrane after SDS-PAGE, which showed that the on-bead approach resulted in the highest signal (Zarnegar et al., 2016). However, it is unclear whether this reflects higher efficiency, or an increase in non-specific signal. For example, the presence of non-fragmented RNAs during IP could stabilise large RNPs, leading to formation of multiprotein complexes that would be

harder to perturb with later washing steps. While this possibility remains to be examined, the low density of binding motifs at sites assigned by PTBP1 irCLIP indicates compromised specificity for this experiment (Haberman et al., 2017).

CLIP protocols also differ in the choice of RNase enzymes. The original CLIP used RnaseT1 and RNase A, while the original PAR-CLIP used RnaseT1 and Mnase. These nucleases have sequence preferences in their cleavage patterns, and extensive digestion can lead to biased assignment of binding sites from CLIP protocols (Kishore et al., 2011). The iCLIP protocol introduced the use of RNase I, which is not known to have any nucleotide preferences, and is thus expected to introduce minimal sequence bias at both ends of RNA fragments (König et al., 2010). The irCLIP protocol also introduced the use of S1 nuclease, which leaves a 3'OH instead of a 3' phosphate at the ends of RNA fragments (Zarnegar et al., 2016). This makes the 3' end dephosphorylation step unnecessary, which otherwise needs to precede the adapter ligation step. S1 nuclease is a relatively inefficient enzyme on RNA, and therefore we find its treatment compatible only with the on-bead digestion (data not shown).

It is a common misconception that the RNA fragments in CLIP are a signature of RNase protection. Unlike formaldehyde crosslinking, during which protein-protein interactions are also crosslinked, UV crosslinking is specific to protein-RNA contacts, and therefore does not stabilise large RNPs. Instead, CLIP intentionally uses stringent lysis conditions in order to perturb most native protein-RNA interactions. Thus, the covalent crosslinking normally remains as the only link between the RBP and the RNA, unless the RBP participates in an RNP that is unusually stable. If a signature of RNase protection is desired, it would be possible to lyse the cells under mild conditions that preserve native protein-RNA interactions, perform RNase fragmentation, and then continue to CLIP with more stringent buffers later during immunoprecipitation. Alternatively, ribonuclease-mediated protein footprinting methods such PIP-seq (Silverman et al., 2014) and RIP-iT-Seq (Singh et al., 2014) could be used.

Bead-based purification of the RBP-RNA complex

CLIP allows the purification of RBPs from cells and tissues with stringent immunoprecipitation conditions, including the use of ionic detergents in the lysis and washing buffers, and the use of high salt washes. Purification of endogenous RBPs normally requires that antibodies are available for efficient immunoprecipitation. As an alternative, endogenous RBPs can be epitope-tagged, which can be achieved with the use of genome editing (Van Nostrand et al., 2017a). When using epitope-tagging, however, it is important to confirm that the function, stability and localisation of the tagged RBP remains unperturbed.

The need for stringent purification and quality control varies depending on the type and expression of the RBP being studied. Some RBPs contain many single-stranded RNA binding domains; for example, ELAVL1 contains three RNA recognition motifs domains, and PTBP1 contains four such domains. These domains recognise U-rich motifs, and are thus expected to crosslink efficiently. These RBPs are also typically expressed in high abundance. On the other hand, many RBPs lack canonical binding domains, or recognise the backbone of double-stranded RNA. It has been shown that cysteine, tryptophan, phenylalanine, tyrosine, arginine, lysine and methionine are the most reactive (Shetlar et al., 1984), and thus RBPs lacking these amino acids in close proximity to the RNA base might crosslink poorly. Even a minor co-purification of another RBP that crosslinks with higher efficiency can lead to major loss of specificity, and this problem is exacerbated if the RBP-of-interest is of low abundance. Finally, some RBPs participate in stable RNPs that may not efficiently dissociate under standard CLIP immunoprecipitation conditions, thus increasing their risk of co-purifying multiple RBPs.

In order to minimise the risk of co-purifying multiple RBPs, denaturing strategies are particularly valuable. Several epitopes enable denaturing and sequential purification strategies, which can further reduce the chance of co-purifying non-specific RBPs and RNAs. Denaturing purification was first implemented by CRAC for yeast, and later by CLAP, urea-iCLIP, uvCLAP and dCLIP (Aktaş et al., 2017; Granneman et al., 2009; Huppertz et al., 2014; Rosenberg et al., 2017; Wang et al., 2010) (Table 1). Sequential histidine- and streptavidin-based affinity purification systems are commonly used (Maticzka et al., 2017; Wang et al., 2010), but immunoprecipitation is also possible if the antibody can bind to the denatured epitope. An example of an RBP that crosslinks poorly is the double-stranded RNA binding protein STAU1, which is prone to co-purification with other more abundant and strongly crosslinking RBPs and the RNAs crosslinked to them. Denaturing purification has been implemented with a 3xFlag-STAU1 and an anti-Flag antibody by using two rounds of immunoprecipitation, such that STAU1 was eluted after the first immunoprecipitation with a high concentration of urea, and then diluted to a lower concentration that enables the second round of immunoprecipitation (Huppertz et al., 2014).

Adapter ligation

To prepare cDNA libraries from the CLIP RNA fragments, they must contain common sequences complementary to the primers used in reverse transcription (RT) and PCR. Most CLIP protocols have a similar organisation of the sequenced reads (Fig. 1), therefore we have named the adapters according to their conventional orientation relative to sequencing. The

exceptions are eCLIP and sCLIP (Kargapolova et al., 2017; Van Nostrand et al., 2016), where the orientation is switched (Table 1). The SeqRv adapter is complementary to the RT primer, and is ligated to the 3' end of the RNA. The first version of the CLIP protocol ligates the adapters to purified RNA fragments (Ule et al., 2003), but most later variants perform on-bead RNA ligation, which reduces the amount of contaminating RNAs (Ule et al., 2005). On-bead ligation also allows removal of excess adapters by stringent washes of the beads instead of using denaturing acrylamide gel purification, thus minimising the loss of specific RNAs. An alternative to the ligation of an adapter to the 3' end of the RNA is developed by sCLIP, where the purified RNAs are polyadenylated, followed by the use of modified oligo-dT primers for RT (Kargapolova et al., 2017).

The original CLIP protocols ligated both adapters to the RNA fragments, which is also employed by HITS-CLIP and PAR-CLIP (Table 1, Table 2). This was modified by the iCLIP protocol, which ligates only the SeqRv adapter to the RNA. The SeqFw adapter, which was originally ligated to the 5' end of RNA fragments in previous protocols, is introduced to the 5' end of the RT primer and then brought to the 3' end of the cDNA via circularisation in iCLIP (König et al., 2010). This enables amplification of cDNAs that prematurely truncate at the crosslinked nucleotide. These truncated cDNAs lack the SeqFw adapter in the original CLIP protocols, and are therefore lost. Beyond increasing the sensitivity of the experiment, the amplification of truncated cDNAs has an additional advantage by enabling nucleotide-resolution mapping of the crosslink sites, which are located at the start of the great majority of iCLIP cDNA inserts (Haberman et al., 2017; Sugimoto et al., 2012). Since the development of iCLIP, 17 other published protocols similarly amplify truncated cDNAs, including BrdU CLIP, eCLIP, irCLIP and FLASH (Aktaş et al., 2017; Van Nostrand et al., 2016; Weyn-Vanhentenryck et al., 2014; Zarnegar et al., 2016) (Table 1, Table 2, Table S1). The eCLIP protocol achieves this by ligating the SeqFw adapter to cDNAs with an intermolecular, rather than intramolecular ligation.

Visualisation of the purified complexes on SDS-PAGE

Visualisation of the protein-RNA complex is the central quality control step in CLIP. It serves to optimise RNA fragmentation, and to control for the specificity of purified complexes. Inclusion of this step guarantees the comparative value of CLIP data produced across the different RBPs, laboratories, and experimental settings. This step purifies the protein-RNA complexes with the use of SDS-PAGE and membrane transfer. The SDS-PAGE separation reduces contamination of non-crosslinked RNAs, which normally run at a lower range of the

gel than the fragments crosslinked to the RBP. These are further removed by nitrocellulose membrane transfer, since the membrane has poor RNA-binding capacity. This also helps to remove excess adapters that can remain stuck on the beads after the RNA ligation step. Presentation of the resulting images from SDS-PAGE analysis alongside the publication of data ensures a quality control standard that has been established by the first publication of CLIP (Ule et al., 2003).

In the original protocol, the 5' end of the RNA is radioactively labelled with ^{32}P in order to visualize the protein-RNA complexes after transfer to the membrane (Ule et al., 2003). This serves to control for the specificity of crosslinked RNAs, and to check that the RNA fragmentation conditions are appropriate (Ule et al., 2005). In certain cases, such as for Ago HITS-CLIP, where 5' labelling is inefficient, a radiolabeled SeqRv adapter was ligated to the RNA to enable visualisation (Chi et al., 2009). Two RNA fragmentation conditions are recommended for the initial experiments. The high RNase condition serves to visualise the specificity of purified complex, since it migrates as a clear band slightly higher than the molecular weight (MW) of the immunoprecipitated RBP. To ensure the specificity of CLIP, no other bands should be visible near the expected band, since these bands indicate co-purification of non-specific RBPs. The low RNase conditions, in contrast, serves to purify RNAs for preparation of the cDNA library. This condition should lead to complexes which migrate diffusely above the apparent MW of the immunoprecipitated RBP, since the diverse sizes of RNA fragments variably affects the migration of the RBP. The complexes are then excised from the appropriate region of the membrane according to the described recommendations (Huppertz et al., 2014; Ule et al., 2005).

In the early versions of PAR-CLIP, the membrane transfer step is omitted and the RNA fragments are purified directly from the SDS-PAGE gel (Hafner et al., 2010). However, the developers included the option of performing the nitrocellulose membrane transfer in a more recent PAR-CLIP publication (Garzia et al., 2017). Recently, irCLIP has been developed for non-radioactive labelling of the purified protein-RNA complexes, increasing the convenience of this quality control step (Zarnegar et al., 2016). This is achieved by covalently coupling an infrared dye to the SeqRv adapter, which allows visualisation of the complexes after the SDS-PAGE and transfer, with infrared imaging which can be performed with a LI-COR Odyssey CLx Imager. Since the infrared signal is present in the adapter that needs to be ligated to the RNA fragments, it additionally allows monitoring of on-bead adapter ligation efficiency. Another strategy for non-radioactive visualisation has been developed in sCLIP, where an aliquot of the immunoprecipitated sample is labelled with biotin-ADP and RNA ligase. The labelled and unlabelled fractions then proceed to SDS-PAGE and nitrocellulose transfer,

followed by incubation of the membrane with streptavidin-HRP and ECL, in order to visualise the biotinylated RNA (Kargapolova et al., 2017).

Several protocols omit the visualisation of purified complexes. One example is uvCLAP, which employs denaturing affinity purification, since the additional specificity gained by the denaturing step reduces the need for further purification by SDS-PAGE (Aktaş et al., 2017). The FLASH protocol also skips the SDS-PAGE and membrane transfer steps (Aktaş et al., 2017), but unlike uvCLAP, it does not include denaturing affinity purification, which could compromise the specificity of data. In eCLIP and seCLIP, the SDS-PAGE and membrane transfer are used without labelling the RNA, and the RBP-RNA complexes are cut from the membrane by considering the MW of the RBP as observed on the IP-western performed in parallel and the predicted shift upwards on the gel due to the crosslinked RNA fragments (Van Nostrand et al., 2016, 2017b) (Table 1).

While increasing the convenience, these protocols risk sacrificing the high specificity of the method. The specificity of purification conditions can be affected by many factors, including the cellular material and lysis conditions used, the type of RBP studied, and the stock and storage time of RNase and other reagents. Maintaining specificity is particularly challenging for studies of non-canonical RBPs that might crosslink weakly to RNA, or lowly expressed RBPs, since even a small amount of another co-purified RBP can lead to dominance of its crosslinked RNAs in the resulting libraries. We therefore advise that the SDS-PAGE visualisation is used at least initially to optimize the conditions for each RBP, in order to ensure that complexes are specific and that the RNase fragmentation conditions are appropriate. This allows the users to be confident in the consistent specificity and comparative value of CLIP data. When this step is omitted, additional computational quality control steps are crucial in order to evaluate the specificity of data (Chakrabarti et al., 2017).

Reverse transcription (RT)

After visualising the complexes on the nitrocellulose membrane, the appropriate region of the membrane is excised, and the RBP is digested with proteinase K, which leaves only a short peptide at the crosslink site and releases the RNA fragments into solution. The resulting RNA fragments are then available for RT with a primer that contains a sequence complementary to the SeqRv adapter. In iCLIP, additional sequences have been introduced to the tail of the RT primer. These include the SeqFw adapter, which is oriented in the opposite direction, an experimental barcode and the unique molecular identifier (UMI) (Fig. 1). The SeqFw adapter enables the later intramolecular ligation of the adapter to truncated and readthrough cDNAs;

the experimental barcodes enable multiplexing of different cDNA reactions before proceeding to further steps; and the UMIs (which consist of a sequence of random nucleotides) enable quantification of unique cDNAs in combination with computational analysis that removes artefacts of variable PCR amplification (König et al., 2010). UMIs have also been introduced into the RNA SeqFw adapter in an early HITS-CLIP study (Chi et al., 2009), but explanation of its use for the analysis of sequencing data was only provided in later publications (Darnell et al., 2011; Moore et al., 2014).

CLIP variants often use different RT enzymes and conditions, including the use of Superscript II, III or IV, AffinityScript and TGIRT. The impact of different RT conditions on cDNA truncation and readthrough has been recently investigated (Van Nostrand et al., 2017c). The standard RT conditions primarily lead to truncation of cDNAs at the crosslinked nucleotide. This feature of the RT is well exploited by iCLIP and other protocols that ligate SeqFw adapter to cDNAs after reverse transcription, as they rely on the truncated cDNAs for precise mapping of the crosslinked nucleotide position. However, use of manganese instead of magnesium ions in the buffer can increase the efficiency of readthrough, especially when used in combination with Superscript IV, and this could benefit techniques that rely on readthrough cDNAs, such as the original CLIP or PAR-CLIP. While most enzymes produce similar cDNA truncation, the position of truncation may be offset by one nucleotide when AffinityScript (used in eCLIP), is compared to other enzymes such as Superscript (used in iCLIP and most other protocols) (Van Nostrand et al., 2017c). This needs to be further examined by comparing the position of crosslink sites assigned with eCLIP and iCLIP for multiple different RBPs.

cDNA purification and amplification

In the original CLIP or iCLIP protocols, gel purification is used to purify the RNA fragments or cDNAs, respectively (König et al., 2010; Ule et al., 2003). The primary purpose is to remove free adapters or RT primers, which would otherwise become templates for reverse transcription or PCR. Carry-over of excess adapters can lead to cDNA libraries that are dominated by PCR artefacts which contain only the barcode or adapter sequences. However, gel purification, phenol-chloroform extractions and ethanol precipitations can be laborious, especially in large-scale experiments. In recent years, several independent approaches have been developed to replace the gel purification steps with approaches that increase the convenience and speed of CLIP protocols, as well as minimising loss of material. These approaches can be separated conceptually into two types: the first captures nucleic acids above a certain size range with the use of silica-like beads or columns; the second specifically

captures the cDNAs via an incorporated molecule, such as BrdU in BrdU CLIP, or a biotinylated SeqRv adapter that remains annealed to the cDNA in FAST-iCLIP (Aktaş et al., 2017; Flynn et al., 2015; Kargapolova et al., 2017; Van Nostrand et al., 2016; Weyn-Vanhentenryck et al., 2014; Zarnegar et al., 2016) (Table 1, Table 2). It remains to be seen which of these variant solutions will be most broadly adopted; the ideal solution should be practical, while efficiently capturing all specific cDNAs without any bias in cDNA size or sequence, in order to maximize the sensitivity of CLIP.

While less convenient, gel extraction provides the most precise size selection of RNA fragments or cDNAs of defined length, thus mitigating potential variations in cDNA length distribution in the final library, and ensuring that adapter products are completely removed prior to PCR. To compensate for this, methods such as FAST-iCLIP and eCLIP that omit gel extraction before PCR often employ an additional gel extraction of the final PCR-amplified cDNA libraries (Flynn et al., 2015; Van Nostrand et al., 2016). Loss of material is not a major concern at this step, since many copies of each cDNA are available due to amplification. However, this approach could be prone to amplifying adapter artefacts in situations where the amount of specific cDNA is limiting, for example when studying an RBP that crosslinks poorly.

For PCR amplification of cDNAs, most recent methods use enzymes that are slightly more efficient than the Accuprime enzymes that were used by the original protocols. The switch to the Phusion enzyme allows amplification of the final cDNA library with a decreased number of PCR cycles, in our hands, without much impact on the data quality (data not shown). In general, while a reduced number of PCR cycles required for cDNA amplification is a promising sign, it should be interpreted cautiously, since it can be due to either an increase in sensitivity or a decrease in specificity. For example, reduced PCR cycle numbers could arise from the increased co-purification of non-specific RBPs and their crosslinked RNAs. When the SDS-PAGE quality control analysis is omitted, one cannot distinguish between these two possibilities.

Primary data analysis and sequencing requirements

The first step in analysing sequencing data produced by CLIP is to examine the experimental barcodes to demultiplex the cDNA libraries, which is followed by mapping the data to the genome. For iCLIP and the 17 later protocols that introduce UMIs into cDNAs (Table 2, Table S1), this can be used to quantify unique cDNAs that map to same loci on the genome without ambiguity. The results can then be exploited by using the full position of mapped reads as in HITS-CLIP (Licatalosi et al., 2008), or by identifying the most likely position of the crosslink

site, which can be achieved in three ways. PAR-CLIP examines C to T transitions in reads (Hafner et al., 2010), iCLIP (and 17 other protocols) examines cDNA truncations (König et al., 2010), and crosslinking-induced mutation sites (CIMS) in HITS-CLIP reads examines deletions and other types of mutations (Zhang and Darnell, 2011). Interestingly, analysis of cDNA truncations was found to be most appropriate for the protocol that crosslinks cells with 4SU as in PAR-CLIP, but then uses iCLIP to prepare the cDNA library (4SU-iCLIP) (Haberman et al., 2017).

Single-end sequencing is appropriate for iCLIP and several derived protocols (such as irCLIP), because both the experimental barcode and UMI are present at the start of the trimmed sequencing read (Fig. 1). The start of the cDNA insert contains information for the crosslink site; the end of the cDNA insert corresponds to the position of RNA cleavage, which is useful to assess biases of RNA fragmentation (Haberman et al., 2017), but is otherwise not crucial for data analysis. However, several protocols introduce important information at both sides of the cDNA inserts, which requires paired-end sequencing, or long-read single-end sequencing that covers the whole cDNA insert. This applies to eCLIP and sCLIP, where the cDNA insert is inverted relative to orientation of sequencing; hence the crosslink site needs to be sequenced from the reverse direction (Kargapolova et al., 2017; Van Nostrand et al., 2016). uvCLAP and FLASH protocols also use paired-end sequencing, where a part of the experimental barcode and UMI are introduced by the SeqRv adapter, and are therefore positioned at the end of the cDNA insert (Aktaş et al., 2017; Garzia et al., 2017). However, single-end eCLIP (seCLIP) has been recently described, which reverts to the iCLIP-like read structure compatible with shorter single-end sequencing (Van Nostrand et al., 2017b). Sequencing of long reads is also beneficial for protocols where the full read or internal mutations are used for analysis, such as HITS-CLIP and PAR-CLIP, as it allows to fully quantify the internal mutations in cDNAs.

Analysis of quality and normalisation of CLIP data

In addition to visualising the purified protein-RNA complexes during the CLIP protocol, the quality of the resulting data can also be examined computationally. Many parameters can be considered to compare the effectiveness of CLIP-derived methodologies, with sensitivity and specificity being the two central measures. The simplest measures of sensitivity and specificity are the number of unique cDNAs in the sequencing library, and clusters of significant crosslinking events (‘peaks’), respectively (Chakrabarti et al., 2017).

To monitor the sensitivity of CLIP, the capacity to quantify unique cDNAs with the use of UMIs is particularly valuable. This is because UMIs distinguish unique cDNAs from those that have been duplicated as a result of PCR amplification. In addition, the ratio of unique versus duplicated cDNAs is also a useful measure to assess whether the depth of sequencing was optimal, and thereby inform on the conditions for most cost-effective sequencing.

To monitor the specificity of CLIP, a suitable peak-calling program needs to be chosen according to the CLIP protocol used to produce the data (Chakrabarti et al., 2017). A low number of peaks indicates that the cDNAs are randomly dispersed along the transcripts, or that they are concentrated in a small number of abundant RNAs (such as rRNA). However, these features could reflect true binding preferences of the RBP that is studied, since many RBPs don't recognise specific sequence or structural RNA motifs (Jankowsky and Harris, 2015), and therefore the number of crosslink peaks is only a rough approximation of specificity. For example, proteins such as FUS or SUZ12 have been shown to have low sequence preference, and therefore their crosslink sites are broadly dispersed across nascent RNAs (Beltran et al., 2016; Rogelj et al., 2012) and rarely lead to crosslink peaks. Nevertheless, analysis of crosslink peaks is particularly valuable to compare multiple data sets for the same RBP. For example, data produced by iCLIP of PTBP1 led to a larger number of peaks than data produced by other protocols, even though the number of unique cDNAs in iCLIP is equal or smaller, and this agrees with highest motif enrichment in iCLIP peaks, especially when compared with irCLIP (Haberman et al., 2017).

To normalise the binding patterns relative to RNA abundance, input control libraries that have not undergone immunoprecipitation can be produced (Ule et al., 2005). Here, the lysate of crosslinked cells after treatment with RNase is loaded on the gel and transferred to the membrane. The RNAs that crosslink to all RBPs present in a selected section of the membrane are isolated and their cDNA library is prepared in the same way as for specific RBPs. This has been exploited for an approach to analyse eCLIP data, which filters the sites that are not significantly enriched compared to the size-matched input (SMInput) control (Van Nostrand et al., 2016, 2017d). This approach can help to enrich the high-affinity binding sites relative to low-affinity transient interactions, both of which can be detected by CLIP.

However, neither definition of crosslink clusters, nor the normalisation by SMInput control can ensure the specificity of CLIP data. Presence of non-specific RNAs in CLIP is most often introduced via co-purification of one RBP or a small number of RBPs, along with their crosslinked RNAs. Since these non-specific RNAs were bound by the co-purified RBPs, they can lead to the identification of distinct binding peaks that are strongly enriched compared to the SMInput control, and can have clear motif enrichment. Therefore, the ideal way to validate

the specificity of CLIP data is to experimentally visualise the quality of purified RBP-RNA complexes on SDS-PAGE. Moreover, integrative computational analyses that use orthogonal functional information can be used, such as comparison with motifs known to be bound by immunoprecipitated RBP (Haberman et al., 2017). In addition, the metaprofile of binding sites can be visualised around exons or other RNA landmarks that are regulated by the RBP, and compared with non-regulated exons, which is commonly referred to as RNA maps. These approaches, and other methods and databases for computational analysis of CLIP data are discussed in detail elsewhere (Chakrabarti et al., 2017).

Conclusion and future perspectives

The large number and modularity of steps in CLIP provides many opportunities for innovation, and new purposes to which the method can be applied continue to be discovered. While the initial development of CLIP was led primarily by the desire for stringent purification and quality control standards, the more recent methods prioritise speed and convenience. This improves the capacity for high-throughput studies of many RBPs across many types of conditions, tissues or species. The use of modified UV illuminators and high-performance UV lasers that can crosslink proteins to RNA *in vivo* in seconds can also improve the capacity to monitor the dynamics of protein-RNA complexes at high temporal resolution (van Nues et al., 2017).

While these modifications are generally beneficial, common standards for quality control should be maintained to allow robust comparisons between datasets. We note that current publically available data from CLIP experiments are generated with protocols that employ varying stringencies of RBP purification, and depending on the stability of the RNP complexes, this results in data of variable specificity. Thus, quality measures of specificity will be important to have a clear interpretation of whether the data represents specific isolation of direct binding sites for the RBP-of-interest, or rather just an enrichment of such sites (Chakrabarti et al., 2017). The visualisation of SDS-PAGE-separated protein-RNA complexes and computational tools for quality analysis of sequenced CLIP libraries will be particularly valuable. Comparison with methods that do not rely on protein purification or crosslinking to identify RNAs interacting with RBPs *in vivo*, such as RNA tagging or TRIBE (targets of RNA-binding proteins identified by editing) (Lapointe et al., 2015; McMahon et al., 2016), could also prove valuable in the interpretation of CLIP data.

In addition to studies of endogenous protein-RNA complexes, variants of CLIP have also been put to other purposes (Table 1). This includes studies of intermolecular or intramolecular RNA-RNA contacts (Chi et al., 2009; Imig et al., 2015; Kudla et al., 2011; Sugimoto et al., 2015), as

reviewed in more depth elsewhere (Sugimoto et al., 2017). Moreover, CLIP of poly(A)-binding protein (PABP) can be exploited to study mRNA 3' ends (Hwang et al., 2016); recently, cell-type specific expression of PABP has been engineered in the mouse brain, thus allowing the study of cell type specific transcripts (Hwang et al., 2017). Moreover, iCLIP has been adapted for studies of RNA methylation, as in m5C- and m6A-miCLIP (Hussain et al., 2013; Linder et al., 2015). These methods could be applied also to other modifications with the use of further antibodies and mutant enzymes, thus broadening the use of CLIP for the epitranscriptomic field. Together, the rapidly increasing amount of CLIP data for many RBPs from the ENCODE consortium and other teams, and the orthogonal methods that interrogate the specificity, functions and localisation of these RBPs (Van Nostrand et al., 2017d), will enable the study of how structure and modifications on diverse types of RNAs work together with RBPs to guide RNP assembly, dynamics and function.

Acknowledgements

We wish to thank all members of Ule lab, and in particular Ina Huppertz, Katia Egli, Andrea Elser, Cristina Militti, Chris Sibley, Kathi Zarnack and Anob Chakrabarti for discussions, critical reading and feedback on the manuscript. This work was supported by funding from the European Research Council (617837-Translate) to J.U., a Wellcome Trust Joint Investigator Award (103760/Z/14/Z) to J.U., a Wellcome Trust Four-Year PhD Studentship (105202/Z/14/Z) to F.C.Y.L., and the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC001002), the UK Medical Research Council (FC001002), and the Wellcome Trust (FC001002).

References

Aktaş, T., Avşar Ilık, İ., Maticzka, D., Bhardwaj, V., Pessoa Rodrigues, C., Mittler, G., Manke, T., Backofen, R., and Akhtar, A. (2017). DHX9 suppresses RNA processing defects originating from the Alu invasion of the human genome. Nature *544*, 115–119.

Baltz, A.G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., et al. (2012). The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. Mol. Cell *46*, 674–690.

Beltran, M., Yates, C.M., Skalska, L., Dawson, M., Reis, F.P., Viiri, K., Fisher, C.L., Sibley, C.R., Foster, B.M., Bartke, T., et al. (2016). The interaction of PRC2 with RNA or chromatin is mutually antagonistic. Genome Res. *26*, 896–907.

Botti, V., McNicoll, F., Steiner, M.C., Richter, F.M., Solovyeva, A., Wegener, M., Schwich, O.D., Poser, I., Zarnack, K., Wittig, I., et al. (2017). Cellular differentiation state modulates the mRNA export activity of SR proteins. J. Cell Biol. *216*, 1993–2009.

Brannan, K.W., Jin, W., Huelga, S.C., Banks, C.A.S., Gilmore, J.M., Florens, L., Washburn, M.P., Van Nostrand, E.L., Pratt, G.A., Schwinn, M.K., et al. (2016). SONAR Discovers RNA-Binding Proteins from Analysis of Large-Scale Protein-Protein Interactomes. Mol. Cell *64*, 282–293.

Brugiolo, M., Botti, V., Liu, N., Müller-McNicoll, M., and Neugebauer, K.M. (2017). Fractionation iCLIP detects persistent SR protein binding to conserved, retained introns in chromatin, nucleoplasm and cytoplasm. Nucleic Acids Res. *45*, 10452–10465.

Burger, K., Mühl, B., Kellner, M., Rohrmoser, M., Gruber-Eber, A., Windhager, L., Friedel, C.C., Dölken, L., and Eick, D. (2013). 4-thiouridine inhibits rRNA synthesis and causes a nucleolar stress response. RNA Biol. *10*, 1623–1630.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., et al. (2012). Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. Cell *149*, 1393–1406.

Chakrabarti, A.M., Haberman, N., Praznik, A., Luscombe, N.M., and Ule, J. (2017). Data Science Issues in Understanding Protein-RNA Interactions. BioRxiv doi:10.1101/208124.

Chen, K., Lu, Z., Wang, X., Fu, Y., Luo, G.-Z., Liu, N., Han, D., Dominissini, D., Dai, Q., Pan, T., et al. (2015). High-Resolution N^6-Methyladenosine (m^6A) Map Using Photo-Crosslinking-Assisted m^6A Sequencing. Angew. Chem. Int. Ed. *54*, 1587–1590.

Chi, S.W., Zang, J.B., Mele, A., and Darnell, R.B. (2009). Argonaute HITS-CLIP decodes microRNA–mRNA interaction maps. Nature *460*, 479–486.

Darnell, J.C., Van Driesche, S.J., Zhang, C., Hung, K.Y.S., Mele, A., Fraser, C.E., Stone, E.F., Chen, C., Fak, J.J., Chi, S.W., et al. (2011). FMRP Stalls Ribosomal Translocation on mRNAs Linked to Synaptic Function and Autism. Cell *146*, 247–261.

611 Flury, V., Restuccia, U., Bachi, A., and Mühlemann, O. (2014). Characterization of
612 Phosphorylation- and RNA-Dependent UPF1 Interactors by Quantitative Proteomics. J.
613 Proteome Res. *13*, 3038–3053.

614 Flynn, R.A., Martin, L., Spitale, R.C., Do, B.T., Sagan, S.M., Zarnegar, B., Qu, K., Khavari,
615 P.A., Quake, S.R., Sarnow, P., et al. (2015). Dissecting noncoding and pathogen RNA–
616 protein interactomes. RNA *21*, 135–143.

617 Garzia, A., Meyer, C., Morozov, P., Sajek, M., and Tuschl, T. (2017). Optimization of PAR-
618 CLIP for transcriptome-wide identification of binding sites of RNA-binding proteins.
619 Methods San Diego Calif *118–119*, 24–40.

620 George, H., Ule, J., and Hussain, S. (2017). Illustrating the Epitranscriptome at Nucleotide
621 Resolution Using Methylation-iCLIP (miCLIP). In RNA Methylation, (Humana Press, New
622 York, NY), pp. 91–106.

623 Granneman, S., Kudla, G., Petfalski, E., and Tollervey, D. (2009). Identification of protein
624 binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput
625 analysis of cDNAs. Proc. Natl. Acad. Sci. *106*, 9613–9618.

626 Haberman, N., Huppertz, I., Attig, J., König, J., Wang, Z., Hauer, C., Hentze, M.W., Kulozik,
627 A.E., Le Hir, H., Curk, T., et al. (2017). Insights into the design and interpretation of iCLIP
628 experiments. Genome Biol. *18*, 7.

629 Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer,
630 A., Ascano Jr., M., Jungkamp, A.-C., Munschauer, M., et al. (2010). Transcriptome-wide
631 Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. Cell *141*,
632 129–141.

633 Huppertz, I., Attig, J., D'Ambrogio, A., Easton, L.E., Sibley, C.R., Sugimoto, Y., Tajnik, M.,
634 König, J., and Ule, J. (2014). iCLIP: Protein–RNA interactions at nucleotide resolution.
635 Methods San Diego Calif *65*, 274–287.

636 Hussain, S., Sajini, A.A., Blanco, S., Dietmann, S., Lombard, P., Sugimoto, Y., Paramor, M.,
637 Gleeson, J.G., Odom, D.T., Ule, J., et al. (2013). NSun2-Mediated Cytosine-5 Methylation of
638 Vault Noncoding RNA Determines Its Processing into Regulatory Small RNAs. Cell Rep. *4*,
639 255–261.

640 Hwang, H.-W., Park, C.Y., Goodarzi, H., Fak, J.J., Mele, A., Moore, M.J., Saito, Y., and
641 Darnell, R.B. (2016). PAPERCLIP Identifies MicroRNA Targets and a Role of CstF64/64tau
642 in Promoting Non-canonical poly(A) Site Usage. Cell Rep. *15*, 423–435.

643 Hwang, H.-W., Saito, Y., Park, C.Y., Blachère, N.E., Tajima, Y., Fak, J.J., Zucker-Scharff, I.,
644 and Darnell, R.B. (2017). cTag-PAPERCLIP Reveals Alternative Polyadenylation Promotes
645 Cell-Type Specific Protein Diversity and Shifts Araf Isoforms with Microglia Activation.
646 Neuron *95*, 1334–1349.e5.

647 Imig, J., Brunschweiger, A., Brümmer, A., Guennewig, B., Mittal, N., Kishore, S., Tsikrika,
648 P., Gerber, A.P., Zavolan, M., and Hall, J. (2015). miR-CLIP capture of a miRNA targetome
649 uncovers a lincRNA H19–miR-106a interaction. Nat. Chem. Biol. *11*, 107.

650 Jankowsky, E., and Harris, M.E. (2015). Specificity and nonspecificity in RNA-protein
651 interactions. Nat. Rev. Mol. Cell Biol. *16*, 533–544.

652 Jungkamp, A.-C., Stoeckius, M., Mecenas, D., Grün, D., Mastrobuoni, G., Kempa, S., and
653 Rajewsky, N. (2011). In Vivo and Transcriptome-wide Identification of RNA Binding
654 Protein Target Sites. Mol. Cell *44*, 828–840.

655 Kargapolova, Y., Levin, M., Lackner, K., and Danckwardt, S. (2017). sCLIP—an integrated
656 platform to study RNA–protein interactomes in biomedical research: identification of
657 CSTF2tau in alternative processing of small nuclear RNAs. Nucleic Acids Res. *45*, 6074–
658 6086.

659 Ke, S., Alemu, E.A., Mertens, C., Gantman, E.C., Fak, J.J., Mele, A., Haripal, B., Zucker-
660 Scharff, I., Moore, M.J., Park, C.Y., et al. (2015). A majority of m6A residues are in the last
661 exons, allowing the potential for 3′ UTR regulation. Genes Dev. *29*, 2037–2053.

662 Keene, J.D., Komisarow, J.M., and Friedersdorf, M.B. (2006). RIP-Chip: the isolation and
663 identification of mRNAs, microRNAs and protein components of ribonucleoprotein
664 complexes from cell extracts : Article : Nature Protocols. Nat Protoc. *1*, 302–307.

665 Kim, K.K., Yang, Y., Zhu, J., Adelstein, R.S., and Kawamoto, S. (2014). Rbfox3 controls the
666 biogenesis of a subset of microRNAs. Nat. Struct. Mol. Biol. *21*, 901.

667 Kishore, S., Jaskiewicz, L., Burger, L., Hausser, J., Khorshid, M., and Zavolan, M. (2011). A
668 quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins.
669 Nat. Methods *8*, 559–564.

670 Klass, D.M., Scheibe, M., Butter, F., Hogan, G.J., Mann, M., and Brown, P.O. (2013).
671 Quantitative proteomic analysis reveals concurrent RNA–protein interactions and identifies
672 new RNA-binding proteins in Saccharomyces cerevisiae. Genome Res. *23*, 1028–1038.

673 König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe,
674 N.M., and Ule, J. (2010). iCLIP reveals the function of hnRNP particles in splicing at
675 individual nucleotide resolution. Nat. Struct. Mol. Biol. *17*, 909–915.

676 Kudla, G., Granneman, S., Hahn, D., Beggs, J.D., and Tollervey, D. (2011). Cross-linking,
677 ligation, and sequencing of hybrids reveals RNA–RNA interactions in yeast. Proc. Natl.
678 Acad. Sci. *108*, 10010–10015.

679 Lapointe, C.P., Wilinski, D., Saunders, H.A.J., and Wickens, M. (2015). Protein-RNA
680 networks revealed through covalent RNA marks. Nat. Methods *12*, 1163–1170.

681 Lerner, M.R., and Steitz, J.A. (1979). Antibodies to small nuclear RNAs complexed with
682 proteins are produced by patients with systemic lupus erythematosus. Proc. Natl. Acad. Sci.
683 U. S. A. *76*, 5495–5499.

684 Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer,
685 A.C., Blume, J.E., Wang, X., et al. (2008). HITS-CLIP yields genome-wide insights into
686 brain alternative RNA processing. Nature *456*, 464–469.

Linder, B., Grozhik, A.V., Olarerin-George, A.O., Meydan, C., Mason, C.E., and Jaffrey, S.R. (2015). Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. Nat. Methods *12*, 767–772.

Maticzka, D., Ilik, I.A., Aktas, T., Backofen, R., and Akhtar, A. (2017). uvCLAP: a fast, non-radioactive method to identify in vivo targets of RNA-binding proteins. BioRxiv doi: 10.1101/158410.

McMahon, A.C., Rahman, R., Jin, H., Shen, J.L., Fieldsend, A., Luo, W., and Rosbash, M. (2016). TRIBE: Hijacking an RNA-Editing Enzyme to Identify Cell-Specific Targets of RNA-Binding Proteins. Cell *165*, 742–753.

Mili, S., and Steitz, J.A. (2004). Evidence for reassociation of RNA-binding proteins after cell lysis: Implications for the interpretation of immunoprecipitation analyses. RNA *10*, 1692–1694.

Moore, M.J., Zhang, C., Gantman, E.C., Mele, A., Darnell, J.C., and Darnell, R.B. (2014). Mapping Argonaute and conventional RNA-binding protein interactions with RNA at single-nucleotide resolution using HITS-CLIP and CIMS analysis. Nat. Protoc. *9*, 263–293.

Nicholson, C.O., Friedersdorf, M., and Keene, J.D. (2017). Quantifying RNA binding sites transcriptome-wide using DO-RIP-seq. RNA *23*, 32–46.

Niranjanakumari, S., Lasda, E., Brazas, R., and Garcia-Blanco, M.A. (2002). Reversible cross-linking combined with immunoprecipitation to study RNA–protein interactions in vivo. Methods *26*, 182–190.

Nues, R. van, Schweikert, G., Leau, E. de, Selega, A., Langford, A., Franklin, R., Iosub, I., Wadsworth, P., Sanguinetti, G., and Granneman, S. (2017). Kinetic CRAC uncovers a role for Nab3 in determining gene expression profiles during stress. Nat. Commun. *8*, 12.

Rogelj, B., Easton, L.E., Bogu, G.K., Stanton, L.W., Rot, G., Curk, T., Zupan, B., Sugimoto, Y., Modic, M., Haberman, N., et al. (2012). Widespread binding of FUS along nascent RNA regulates alternative splicing in the brain. Sci. Rep. *2*, 603.

Rosenberg, M., Blum, R., Kesner, B., Maier, V.K., Szanto, A., and Lee, J.T. (2017). Denaturing CLIP, dCLIP, Pipeline Identifies Discrete RNA Footprints on Chromatin-Associated Proteins and Reveals that CBX7 Targets 3′ UTRs to Regulate mRNA Expression. Cell Syst. *5*, 368–385.e15.

Sanford, J.R., Coutinho, P., Hackett, J.A., Wang, X., Ranahan, W., and Caceres, J.F. (2008). Identification of Nuclear and Cytoplasmic mRNA Targets for the Shuttling Protein SF2/ASF. PLoS ONE *3*, e3369.

Schor, I.E., Llères, D., Risso, G.J., Pawellek, A., Ule, J., Lamond, A.I., and Kornblihtt, A.R. (2012). Perturbation of Chromatin Structure Globally Affects Localization and Recruitment of Splicing Factors. PLOS ONE *7*, e48084.

Shetlar, M.D., Carbone, J., Steady, E., and Hom, K. (1984). Photochemical Addition of Amino Acids and Peptides to Polyuridylic Acid. Photochem. Photobiol. *39*, 141–144.

Silverman, I.M., Li, F., Alexander, A., Goff, L., Trapnell, C., Rinn, J.L., and Gregory, B.D. (2014). RNase-mediated protein footprint sequencing reveals protein-binding sites throughout the human transcriptome. Genome Biol. *15*, R3.

Singh, G., Ricci, E.P., and Moore, M.J. (2014). RIPiT-Seq: A high-throughput approach for footprinting RNA:protein complexes. Methods *65*, 320–332.

Sugimoto, Y., König, J., Hussain, S., Zupan, B., Curk, T., Frye, M., and Ule, J. (2012). Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. Genome Biol *13*, R67.

Sugimoto, Y., Vigilante, A., Darbo, E., Zirra, A., Militti, C., D'Ambrogio, A., Luscombe, N.M., and Ule, J. (2015). hiCLIP reveals the in vivo atlas of mRNA secondary structures recognized by Staufen 1. Nature *519*, 491–494.

Sugimoto, Y., Chakrabarti, A.M., Luscombe, N.M., and Ule, J. (2017). Using hiCLIP to identify RNA duplexes that interact with a specific RNA-binding protein. Nat. Protoc. *12*, 611–637.

Tenenbaum, S.A., Carson, C.C., Lager, P.J., and Keene, J.D. (2000). Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. Proc. Natl. Acad. Sci. *97*, 14085–14090.

Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R.B. (2003). CLIP Identifies Nova-Regulated RNA Networks in the Brain. Science *302*, 1212–1215.

Ule, J., Jensen, K., Mele, A., and Darnell, R.B. (2005). CLIP: A method for identifying protein–RNA interaction sites in living cells. Methods *37*, 376–386.

Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K., et al. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat. Methods *13*, 508–514.

Van Nostrand, E.L., Gelboin-Burkhart, C., Wang, R., Pratt, G.A., Blue, S.M., and Yeo, G.W. (2017a). CRISPR/Cas9-mediated integration enables TAG-eCLIP of endogenously tagged RNA binding proteins. Methods *118–119*, 50–59.

Van Nostrand, E.L., Nguyen, T.B., Gelboin-Burkhart, C., Wang, R., Blue, S.M., Pratt, G.A., Louie, A.L., and Yeo, G.W. (2017b). Robust, Cost-Effective Profiling of RNA Binding Protein Targets with Single-end Enhanced Crosslinking and Immunoprecipitation (seCLIP). In MRNA Processing, Y. Shi, ed. (New York, NY: Springer New York), pp. 177–200.

Van Nostrand, E.L., Shishkin, A.A., Pratt, G.A., Nguyen, T.B., and Yeo, G.W. (2017c). Variation in single-nucleotide sensitivity of eCLIP derived from reverse transcription conditions. Methods *126*, 29–37.

Van Nostrand, E.L., Freese, P., Pratt, G.A., Wang, X., Wei, X., Blue, S.M., Dominguez, D., Cody, N.A.L., Olson, S., Sundararaman, B., et al. (2017d). A Large-Scale Binding and Functional Map of Human RNA Binding Proteins. BioRxiv doi: 10.1101/179648.

763 Wang, Z., Kayikci, M., Briese, M., Zarnack, K., Luscombe, N.M., Rot, G., Zupan, B., Curk,
764 T., and Ule, J. (2010). iCLIP Predicts the Dual Splicing Effects of TIA-RNA Interactions.
765 PLOS Biol. *8*, e1000530.

766 Weyn-Vanhentenryck, S.M., Mele, A., Yan, Q., Sun, S., Farny, N., Zhang, Z., Xue, C.,
767 Herre, M., Silver, P.A., Zhang, M.Q., et al. (2014). HITS-CLIP and Integrative Modeling
768 Define the Rbfox Splicing-Regulatory Network Linked to Brain Development and Autism.
769 Cell Rep. *6*, 1139–1152.

770 Windhager, L., Bonfert, T., Burger, K., Ruzsics, Z., Krebs, S., Kaufmann, S., Malterer, G.,
771 L'Hernault, A., Schilhabel, M., Schreiber, S., et al. (2012). Ultrashort and progressive 4sU-
772 tagging reveals key characteristics of RNA processing at nucleotide resolution. Genome Res.
773 *22*, 2031–2042.

774 Yeo, G.W., Coufal, N.G., Liang, T.Y., Peng, G.E., Fu, X.-D., and Gage, F.H. (2009). An
775 RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in
776 stem cells. Nat. Struct. Mol. Biol. *16*, 130–137.

777 Zarnegar, B.J., Flynn, R.A., Shen, Y., Do, B.T., Chang, H.Y., and Khavari, P.A. (2016).
778 irCLIP platform for efficient characterization of protein-RNA interactions. Nat. Methods *13*,
779 489–492.

780 Zhang, C., and Darnell, R.B. (2011). Mapping in vivo protein-RNA interactions at single-
781 nucleotide resolution from HITS-CLIP data. Nat. Biotechnol. *29*, 607–614.

782 Zhao, J., Ohsumi, T.K., Kung, J.T., Ogawa, Y., Grau, D.J., Sarma, K., Song, J.J., Kingston,
783 R.E., Borowsky, M., and Lee, J.T. (2010). Genome-wide Identification of Polycomb-
784 Associated RNAs by RIP-seq. Mol. Cell *40*, 939–953.

785

Figure Legends

Figure 1: The core steps of iCLIP and other variants of CLIP.

The majority of currently available CLIP protocols (18 out of 28, Table S1) amplify truncated cDNAs to identify the protein-RNA crosslink sites. Therefore, this schematic follows the core steps of iCLIP, a variant that was developed to amplify truncated cDNAs. The structure of RNA fragments, cDNA inserts, and sequenced reads is marked along with colour-coded adapters, unique molecular identifiers (UMI), experimental barcodes and primers. The adapters are named as SeqRv and SeqFw, according to their conventional orientations relative to the final sequenced reads. Where indicated, variations introduced by other CLIP protocols are illustrated.

Table 1: List of CLIP and related protocols

Protocols are ordered by the year of publication to reflect their historical development. Updated publications introducing important variations to the same method are grouped with the initial publication. Protocols that are not aimed at studying the specificity of an RBP, but that apply the CLIP technology to a new purpose, are listed at the end.

Table 2: The core steps of CLIP and their variations

The 11 core steps of the CLIP are listed, as well as the primary variations made in each step over the last 15 years, along with the names and publication dates of protocols that first introduced each variation. The number of CLIP protocols from the list in Table 1 and the number of developer labs that adopted each variant is shown, with the full list behind these numbers available in Table S1. A description and explanation of the rationale behind each variation is provided**.**

Table S1. Related to Figure 1 and Table 2: Adopted variations in published CLIP protocols

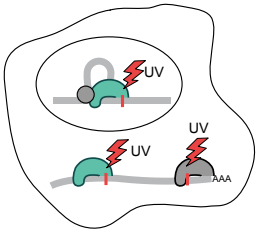
For all CLIP variants listed in Table 1, the variations adopted by each specific protocol are annotated. Boxes in black apply when none of the variants of the corresponding step are implemented by a protocol.

Preparation of crosslinked cell lysate

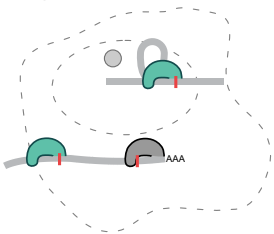
Purification of specific crosslinked RNA fragments

cDNA library preparation and sequencing

1. Covalent protein-RNA crosslinking

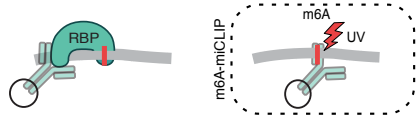


2. Cell lysis

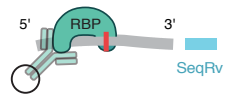


3. RNA fragmentation

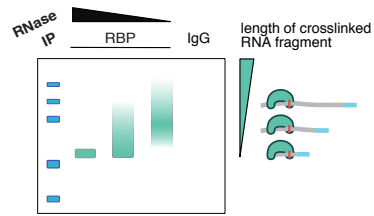
4. Purification of protein-RNA complexes



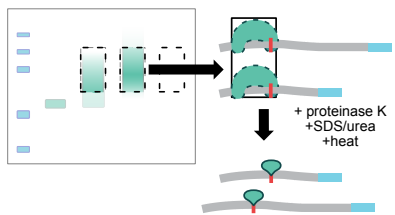
5. Ligation of SeqRv adapter to fragmented RNA



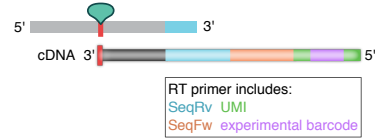
6. Quality control



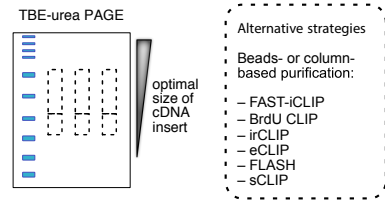
7. RNA extraction



8. Reverse transcription

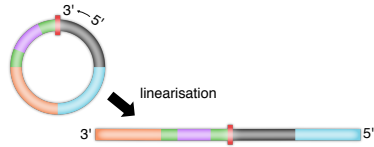


9. cDNA purification



10. SeqFw adapter ligation

intramolecular ligation (iCLIP and 15 derived protocols)



11. cDNA amplification and sequencing of multiplexed libraries

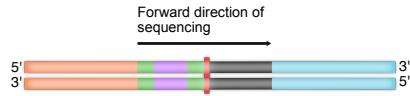


Table1

Acronym	Full Name	Citation of Protocol
CLIP and related protocols		
RIP	RNA immunoprecipitation	Lerner and Steitz, 1979
CLIP	(UV) Crosslinking and immunoprecipitation	Ule et al., 2003 Ule et al., 2005
Fractionation CLIP	CLIP from nucleus, cytosol and polysomes	Sanford et al., 2008
HITS-CLIP	High-throughput sequencing of RNA isolated by CLIP	Licatalosi et al., 2008 Chi et al., 2009
CLIP-seq	CLIP coupled with high-throughput sequencing	Yeo et al., 2009
CRAC	UV cross-linking and analysis of cDNAs	Granneman et al., 2009
PAR-CLIP	Photoactivable ribonucleoside-enhanced CLIP	Hafner et al., 2010 Garzia et al., 2017
iCLIP	Individual-nucleotide resolution CLIP	König et al., 2010
CLAP	Crosslinking and affinity purification	Wang et al., 2010
4SU-iCLIP	4SU-mediated crosslinking followed by iCLIP	Huppertz et al., 2014
urea-iCLIP	iCLIP with denaturing purification	Huppertz et al., 2014
BrdU CLIP	Bromodeoxyuridine UV CLIP	Weyn-Vanhentenryck et al., 2014
FAST-iCLIP	Fully automated and standardized iCLIP	Flynn et al., 2015
irCLIP	Infrared-CLIP	Zarnegar et al., 2016
eCLIP	Enhanced CLIP	Van Nostrand et al., 2016
seCLIP	Single-end eCLIP	Van Nostrand et al., 2017c
uvCLAP	UV crosslinking and affinity purification	Aktaş et al., 2017
FLASH	Fast ligation of RNA after some sort of affinity purification for high-throughput sequencing	Aktaş et al., 2017
Fr-iCLIP	Fractionation iCLIP	Brugiolo et al., 2017
sCLIP	Simplified CLIP	Kargapolova et al., 2017
dCLIP	Denaturing CLIP	Rosenberg et al., 2017
Further applications of CLIP		
CLASH	Cross-linking, ligation, and sequencing of hybrids	Kudla et al., 2011
hiCLIP	RNA hybrid and iCLIP	Sugimoto et al., 2015
PAPERCLIP	Poly(A) binding protein-mediated mRNA 3 end retrieval by CLIP	Hwang et al., 2016
cTag-PAPERCLIP	"Conditionally" tagged-PAPERCLIP	Hwang et al., 2017
m5C-miCLIP	Cytosine-5 methylation iCLIP	Hussain et al., 2013
m6A-miCLIP	N6-methyladenosine iCLIP	Linder et al., 2015

Core steps and their variations in CLIP protocols	Number of protocols (developer labs)	First protocol developing the variation	Description and rationale for each step and its variations
1. Covalent protein-RNA crosslinking			
UV-C crosslinking of intact cells or tissues	23 (11)	CLIP (2003)	UV-C crosslinking (254nm) can be applied on any type of sample, including postmortem human tissues, and its efficiency is generally similar to the use of UV-A with 4SU.
UV-A crosslinking of cells after incubation with photoactivatable ribonucleosides	3 (2)	PAR-CLIP (2010)	UV-A crosslinking (365nm) requires preincubation of cells with 4SU or 6SG. It can lead to preferential identification of those protein-RNA contact sites that contain U or G, and long preincubation with 4SU or 6SG can lead to cellular stress (Huppertz et al., 2014). It increases efficiency for some RBPs, and is likely to be particularly valuable for studies of RBP interactions with nascent RNAs.
Mutation-induced crosslinking	1 (1)	m5C-miCLIP (2013)	This method employs a mutant RNA methylase, NSun2, which forms a covalent bond with its m5C methylated base.
In vitro UV-C crosslinking of antibody to purified RNA	1 (1)	m6A-miCLIP (2015)	RNAs are purified from cells and fragmented. The RNA fragments are then incubated with m6A specific antibody. Captured RNA fragments are crosslinked to the antibody with UV-C.
2. Cell lysis			
Total cell	25 (11)	CLIP (2003)	RBP is purified from total cellular lysate, which enables to simultaneously examine all types of RNAs bound by an RBP in all cellular compartments.
Fractionated cells	3 (3)	Fractionation CLIP (2008)	RBP is purified from cellular subcompartments. The basic approach is to fractionate crosslinked cells into nuclear and cytosol fractions, and here polysomes are studied in addition.
		fr-iCLIP (2017)	Here, nucleoplasm and chromatin are studied in addition to cytosol.
3. RNA fragmentation			
RNase digestion in lysate	23 (10)	CLIP (2003)	RNA fragmentation in the lysate ensures that RNA-dependent RNP complexes dissociate before incubation with the beads, thus avoiding co-purification of additional RBPs.
Controlled RNA fragmentation by optimising limited RNAse digestion	21 (9)	CLIP (2005)	The procedure for optimising limited RNAse digestion is presented by using gel shift analysis of protein-RNA complexes separated by SDS-PAGE. This is important to A) Ensures that the final cDNAs are long enough to enable unique genomic mapping. B) Overdigestion introduces sequence constraints and biases due to preferred RNase cleavage patterns (Haberman et al., 2017). C) Avoid insufficient RNase digestion, which could keep larger RNPs intact, thus leading to co-purification of non-specific RBPs and RNAs.
Use of RNase I	12 (6)	iCLIP (2010)	Most RNases preferentially cleave after one or two specific nucleotides. RNase I is capable of cleaving at all nucleotides, and thus has less sequence specificity than other RNases. This minimises the sequence bias of RNA fragmentation, thus decreasing the sequence constraints at cDNA ends (Haberman et al., 2017).
On-bead RNase digestion	5 (4)	PAR-CLIP (2010)	In addition to RNase digestion in lysate, a second round of RNase digestion is performed on beads. This leads protein-RNA complexes migrating as a sharp band on SDS-PAGE, indicative of RNA overdigestion that can lead to short reads which may not map uniquely to the genome.
		irCLIP (2016)	On-bead RNase digestion allows the use of nuclease S1, a less efficient enzyme that is not compatible with in-lysate digestion. Nuclease S1 leaves a 3'OH group on RNA fragments, which is convenient by avoiding the need for an additional phosphatase step. However, the on-beads digestion might be less efficient in dissociating large RNP complexes.
In vitro fragmentation of purified RNA	1 (1)	m6A-miCLIP (2015)	Purified RNAs are fragmented by zinc(III)-mediated RNA cleavage.
4. Purification of protein-RNA complexes			
Immunoprecipitation under mild conditions		RIP (1979)	RNA immunoprecipitation, in its original version, is performed without RNAse, and under conditions that are mild enough to preserve protein binding to the RNA targets without any crosslinking. This serves to identify RNAs enriched in the immunoprecipitation, rather than to define the position of binding sites.
Immunoprecipitation under stringent conditions	21 (10)	CLIP (2003)	Stringent washing with high salt buffers and ionic detergents preserves only the crosslinked protein-RNA contacts, followed by SDS-PAGE and membrane transfer to further separate any remaining co-purified proteins that are of different MW. Nitrocellulose membrane does not bind well to nucleic acids, thus allowing to further remove any remaining free RNAs.
Denaturing purification with the use of epitope tags	6 (4)	CRAC (2009)	Uses two-step affinity purification of tagged proteins in yeast under denaturing conditions to completely remove any interacting RBPs and free RNAs that are not crosslinked to the protein of interest.
		CLAP (2010)	Like CRAC, but uses two-step affinity purification of tagged proteins in mammalian cells under denaturing conditions. Relies on 8xHis- and two Strep-tag II peptides. While ensuring specificity, the method requires expression of tagged proteins, which may not fully reflect the binding pattern of untagged endogenous proteins.
		urea-iCLIP (2014)	Like CLAP, but using a 3xFlag-tag, such that the RBP is eluted after the first immunoprecipitation with denaturing conditions (eg. high SDS or urea and heat), which is then followed by a second immunoprecipitation.
		uvCLAP (2017)	Like CLAP, but replacing the 8xHis- and Strep-tag with 3xFlag-tag and histidine-biotin-histidine-tagging.
		dCLIP (2017)	RBP is fused with a biotinylation tag, which enables it to be biotinylated in cell lines expressing the bacterial biotin ligase BirA. The RBP is then purified with streptavidin beads and subjected to multiple denaturing 8M urea and 2% SDS washes.
5. Ligation of SeqRv adapter to fragmented RNA			
Ligation to purified RNA	4 (3)	CLIP (2003)	In the original protocol, adapters are ligated to RNA after membrane transfer and digestion of the protein. This requires an additional gel purification to remove the adapter, which leads to some loss of RNA. This protocol can also be prone to amplifying non-specific bacterial or yeast RNAs that can be introduced as contaminants during PAGE or transfer of the protein-RNA complexes, and adapter-adapter concatamer artefacts. The protocol can be of use in rare cases where on-bead ligation is inefficient.
On-bead ligation	23 (10)	CLIP (2005)	On-bead ligation allows removal of the adapter by washing the beads, and free adapters are further removed by SDS-PAGE and transfer. Thus no additional step is needed to remove the adapter. The on-beads ligation is efficient when used with magnetic beads, as long as the RNA fragments are >15nt and the relative volume of beads vs. ligation reaction is appropriate. Its efficiency needs to be tested when changing the type of beads used.

Barcoded seqRv adapter	4 (3)	eCLIP (2016), uvCLAP (2017)	Allows multiplexing of experiments immediately after IP, which can save time and reduce experimental variation between samples. However, this loses the capacity to examine the specificity of purified protein-RNA complexes during the membrane visualisation step. It also requires PE sequencing or long sequencing reads, in order to ensure that the full cDNA together with the barcode in the SeqRv adapter are sequenced.
Polyadenylation of purified RNAs	1 (1)	sCLIP (2017)	Instead of ligating SeqRv, the purified RNA fragments are polyadenylated, and the poly(A) tail is then used as the template for annealing the RT primer.
6. Quality control			
Visualisation of PAGE-separated protein-RNA complexes	24 (11)	CLIP (2003)	Allows visualisation and validation of the specificity of the protein-RNA complexes, to confirm absence of non-specific co-purified RBPs or RNAs, and to demonstrate that RNase conditions are well optimised. The original protocol used radioactive 5' end labelling of RNA fragments for this purpose.
Non-radioactive visualisation of protein-RNA complexes	2 (2)	irCLIP (2016)	Infrared signal is introduced via a dye-coupled SeqRv adapter, which allows visualisation of protein-RNA complexes without the use of radioactivity, while also monitoring ligation efficiency.
		sCLIP (2017)	An aliquot of the immunoprecipitation is labelled by conjugating biotin-ADP to the 3' end of the crosslinked RNAs. Subsequently this is visualised with streptavidin-HRP chemiluminescence.
7. RNA extraction			
Proteinase digestion	28 (13)	CLIP (2003)	Proteinase K (PK) is used to cleave the protein crosslinked to RNA under denaturing conditions. This releases the RNA into solution, along with a small peptide that remains on the RNA at the crosslink site.
Use of SDS buffer	5 (4)	CRAC (2009)	Both urea and SDS denature proteins and enhance PK activity, but urea can be unstable upon prolonged storage, and therefore SDS is proposed to be used instead.
8. Reverse transcription			
Conversion of RNA fragments into cDNAs	28 (13)	CLIP (2003)	A primer complementary to the SeqRv adapter is used to convert RNA fragments into cDNAs.
Introduction of experimental barcodes and unique molecular identifiers (UMIs) into cDNAs	15 (7)	iCLIP (2010)	UMIs (also referred to as random barcodes, or randomers) allow to quantify the number of unique cDNAs that map to the same position in the genome, thus differentiating them from PCR amplicons of the same cDNA molecule, taking full advantage of highthroughput sequencing to the quantify cross-linking at specific nucleotides.
9. cDNA purification			
Denaturing acrylamide gel purification of ligated RNAs	3 (2)	CLIP (2003)	This step has been used by protocols that ligated both seqRv and seqFw adapters to the purified RNA in order to remove the adapters. After RT-PCR, the cDNA undergoes further size selection.
TBE-Urea acrylamide gel	9 (4)	iCLIP (2010)	Excess RT primers are removed with gel purification, which also serves to select specific cDNA size ranges as an additional quality control. This is followed by ethanol precipitation. Under optimal conditions, recovery is ~90%, but the method requires some experience to avoid carrying over salts or any other reagents that could inhibit PCR.
BrdU capture	3 (1)	BrdU CLIP (2014)	Br-dUTP replaces dTTP in the reverse transcription reaction, enabling purification of cDNAs by two rounds of immunoprecipitation with an anti-BrdU antibody. During the second round, cDNAs are circularised and linearised as in iCLIP, and then eluted by heating.
Streptavidin beads purification of cDNA	2 (1)	Fast-iCLIP (2015)	Streptavidin purification of cDNA is enabled via biotinylated SeqRv that has been ligated to RNA, and remains attached to cDNAs. After circularisation, cDNA is eluted from the streptavidin beads and column purified. This increases the convenience and speed of the protocol. After PCR amplification, cDNAs are then size-selected with acrylamide gel.
		irCLIP (2016)	Similar to Fast-iCLIP, but after circularisation, cDNA is incubated with isopropanol and AMPure beads for further purification and size selection of cDNA.
Silane beads purification of cDNA	2 (1)	eCLIP (2016)	After enzymatic degradation of free RT primers with Exo-SAP, cDNA is purified with silane beads, and after PCR amplification, further purified with native agarose gel.
Column purification of cDNA	2 (2)	FLASH (2017)	After RT, cDNA is column purified. There are no further purification steps after circularisation.
		sCLIP (2017)	After RT, the protocol employs second strand cDNA synthesis, *in vitro* transcription, and adapter ligation to the antisense RNA, each coupled with column-based purification steps.
10. SeqFw adapter ligation			
SeqFw adapter is ligated to the 5' end of RNA fragments	9 (5)	CLIP (2003)	SeqFw adapter is required to amplify cDNAs. Since it is ligated to the 5' ends of RNA fragments, the full RNA fragment needs to be reverse transcribed in order to create amplifiable cDNAs in CLIP. Therefore, only cDNAs that read through the crosslinked nucleotide can be amplified by PCR, leading to loss of truncated cDNAs.
SeqFw adapter is ligated to 3' end of cDNAs to enable amplification of truncated cDNAs	18 (8)	iCLIP (2010)	Ligation of SeqFw to cDNA is achieved by introducing the seqFw sequence into cDNAs via the RT primer, followed by its efficient intramolecular ligation to the 3' end of cDNAs with the use of circligase. The circular cDNA is then linearised through a BamHI site in the RT primer, which then enables amplification and sequencing of both 'read-through' and 'truncated' cDNAs. When combined with analysis of clustered cDNA starts, this allows to map the position of the high-occupancy cross-linking with nucleotide resolution.
		Fast-iCLIP (2015)	As iCLIP, except that RT primer contains two carbon spacers between the SeqFw and SeqRv sequences, which allow termination of the PCR enzyme, thus removing the need for BamHI digestion.
		eCLIP (2016)	The cDNA circularisation is replaced by an intermolecular ligation of SeqFw with the use of RNA ligase.
11. cDNA amplification and sequencing of multiplexed libraries			
cDNA cloning and Sanger sequencing	4 (3)	CLIP (2003)	Individual cDNAs are cloned for Sanger sequencing.
High-throughput sequencing of multiplexed cDNA libraries	25 (12)	HITS-CLIP (2008)	Overhangs are added to PCR primers, which include experimental barcodes for each sample and sequencing adapters, which allows multiplexing of the cDNA libraries and high-throughput sequencing.
		Fast-iCLIP (2015)	Phusion enzyme is used for PCR, and qPCR is used to determine the optimal number of PCR cycles.