

CLIP: A method for identifying protein–RNA interaction sites in living cells

Jernej Ule^a, Kirk Jensen^b, Aldo Mele^a, Robert B. Darnell^{a,*}

^a *Howard Hughes Medical Institute and the Laboratory of Molecular Neuro-Oncology, The Rockefeller University, New York, USA*

^b *School of Molecular and Biomedical Science, and the Centre for the Molecular Genetics of Development, University of Adelaide, Adelaide, SA, Australia*

Accepted 11 July 2005

Abstract

Nucleic-acid binding proteins constitute nearly one-fourth of all functionally annotated human genes. Genome-wide analysis of protein–nucleic acid contacts has not yet been performed for most of these proteins, restricting attempts to establish a comprehensive understanding of protein function. UV cross-linking is a method typically used to determine the position of direct interactions between proteins and nucleic acids. We have developed the cross-linking and immunoprecipitation assay, which exploits the covalent protein–nucleic acid cross-linking to stringently purify a specific protein–RNA complex using immunoprecipitation followed by SDS–PAGE separation. In this way, the vast majority of non-specific contaminating RNA, which can bind to co-immunoprecipitated proteins or beads, can be removed. Here, we present an improved protocol that performs RNA linker ligation before the SDS–PAGE step, and describe its application to the specific purification and amplification of RNA ligands of Nova in neurons.

© 2005 Elsevier Inc. All rights reserved.

Keywords: CLIP; UV cross-linking; Immunoprecipitation; Nova; Protein–RNA binding

1. Introduction

Nucleic-acid (NA) binding proteins constitute 23% of the functionally annotated human genes [1], which reflects the crucial role these proteins play in control of gene expression. To unravel the gene expression networks they regulate it is necessary to systematically identify these protein–NA interactions in intact cells. Initially, *in vitro* techniques such as RNA selection [2–4] were successful in defining the RNA sequence motifs bound by RNA-binding proteins. However, in many cases these motifs are relatively short, such as the YCAY motif identified as the consensus binding sequence for the neuron-specific RNA binding protein Nova [5,6], the UAGG binding motif for hnRNP A [7], the UUCU binding motif for PTB [8], and others. Such sequences occur too frequently in the genome to develop simple bioinformatics means to predict *in vivo* protein binding sites. In other cases,

a high structural and sequence complexity of the *in vitro* defined binding motifs may hinder their identification in the genome, such as the RNA targets identified for the Fragile-X mental retardation protein [9,10]. In addition, the RNA sequence motif defined *in vitro* may fail to reflect physiological interactions, which depend on precise concentration, ionic strength and sub-cellular localization, and may require formation of multi-protein complexes.

Initial studies attempted to identify *in vivo* relevant RNA targets of RBPs by affinity purification with immobilized GST-RBP from cell extract followed by the differential display assay [11,12]. Other studies used immunoprecipitation of protein–RNA complexes from cellular extract followed by various screens such as RT-PCR, library screening, slot blot or microarray hybridization [13–17]. While associations detected by these procedures are often interpreted as reflecting direct *in vivo* interactions, the physiologic problems discussed above have made their interpretation difficult. It is known that FMRP, and perhaps most RNA-binding proteins, are components of multi-protein complexes that

* Corresponding author. Fax: +1 212 327 7109.

E-mail address: darnelr@rockefeller.edu (R.B. Darnell).

include other RNA-binding proteins and a plethora of RNA targets. For instance, 11 RNA-binding proteins have been identified by several groups to be in complex with FMRP, four of which, FXR1, FXR2, NUFIP, 82-FIP, directly associate with FMRP [9]. This observation suggests that an unknown number of RNAs may be present in FMRP immunoprecipitates that associate with proteins other than FMRP. Moreover, the ionic conditions used in these immunoprecipitations (30 mM EDTA) have subsequently been found to abolish interactions between FMRP and KH2 domain target RNAs, so that only a subset of potential RNA partners may have been detected in these experiments [18]. Technical concerns have been discussed bearing on experiments involving co-immunoprecipitation of Hu RNA binding protein–RNA complexes. These co-immunoprecipitation experiments have been shown capable of generating complexes that form artificially *in vitro* via re-association of molecules after cell lysis [19]. Taken together, these observations illustrate that co-immunoprecipitation may not recapitulate the *in vivo* state of ribonucleoprotein complexes.

From these observations, two additional points can be made. First, given the difficulties that may arise in these experiments, the importance of validating the putative protein–RNA interactions by independent means (e.g., genetic, functional studies) cannot be overstated. Second, it is clear that an ideal method for identification of RNAs reflecting the *in vivo* binding of a protein should preserve the protein–RNA interactions in the context of intact cells.

To detect protein–NA interactions in the context of an intact cell, *in vivo* cross-linking provides a robust methodology. Early attempts to use cross-linking *in vivo* used formaldehyde to analyze NA binding sites of DNA and RNA-binding proteins [20,21]. A shortcoming of this approach is that formaldehyde induces large multi-molecular chemical bridges, which can create artifacts, especially in respect to the ability to identify direct versus indirect protein–NA interactions. In addition to inducing protein–NA bridges, formaldehyde also induces protein–protein cross-linking, resulting in purification of a complex of nucleic acids, their binding proteins, and accessory proteins that may also bind nucleic acids. This compounds the difficulty in identifying direct protein–NA interactions and yields a generally weak signal:noise, a particular problem for identification of NAs that are expressed at low levels in cells. Finally, formaldehyde cross-linking technique has not been able to identify the exact position on NA where the protein binds, which is crucial to understand the *in vivo* binding properties of the protein and the mechanistic implications of NA binding. For instance, binding within an intron would suggest a nuclear role of an RNA-binding protein, such as alternative splicing regulation, while binding in a UTR sequence might implicate a cytoplasmic role, such as control of RNA localization or translation.

The UV cross-linking method described here is able to define sites of direct contact between RNA and protein. This method capitalizes on the natural photo-reactivity of the NA bases, especially pyrimidines, and specific amino

acids, such as Cys, Lys, Phe, Trp, and Tyr at 254 nm irradiation [22–24]. UV irradiation does not cross-link proteins to proteins, therefore it only identifies direct protein–NA interactions [25]. Initially, *in vivo* UV cross-linking in combination with SDS–PAGE analysis of protein–NA complexes was used to identify the major proteins bound to specific groups of RNA molecules, such as hnRNAs and mRNAs [26,27], and to analyze protein distribution on specific segments of chromosomal DNA [28]. The position of the UV-induced covalent bond in combination with immunoprecipitation and mass spectroscopy was also used to identify the precise region of the protein that contacts RNA *in vitro* [29,30]. Furthermore, UV cross-linking of protein–DNA complexes in combination with immunoprecipitation and linker ligation has been used for identification of DNA-binding sites of proteins [31–34].

Here, we describe the CLIP (cross-linking and immunoprecipitation) method for isolation of RNA-binding sites of proteins in tissues and cell cultures [35]. CLIP has significant advantages over previous methods. UV cross-linking is performed on live cells, allowing results to reflect an unperturbed *in vivo* environment with relevant intermolecular interactions, salt and ion concentration. Because the covalent bond formed by UV cross-linking is irreversible, the NA can be partially digested to short NA tags. A number of purification steps following the initial immunoprecipitation greatly enhances the signal:noise of NA tags encoding the direct protein binding site [35]. The RNA bound to non-specific co-immunoprecipitated proteins is removed during SDS–PAGE separation of protein–RNA complexes, and free RNA is removed during transfer to nitrocellulose membrane. RNA is ligated to RNA-linkers, which allows PCR-based amplification of CLIP tags. Here, we describe the CLIP protocol in detail, including modifications that aim to increase the specificity of cloned sequences.

2. Modifications of the original CLIP protocol

The CLIP protocol (Fig. 1) that we originally reported [35] has been modified here in one major aspect (Table 1): 3' RNA linker ligation is done on-bead. The 3' and 5' RNA linker ligations were originally performed sequentially, with no intervening purification step, which allowed both competing linker–linker ligations and inadvertent ligation to contaminating RNA molecules that are often present in recombinant proteins, including the commercial RNA ligase. In the current protocol, the first ligation step is performed while RNA is bound to the protein, which ensures that the bulk of the contaminating RNA and free linker is removed during the SDS–PAGE and transfer step (Fig. 1). We have found that this leads to an improved signal:noise. For instance, we split 50 mg UV cross-linked mouse brain tissue between the original and the current CLIP methods. Using the original CLIP method we cloned 53% mouse and 47% bacterial (i.e., contaminant) sequences ($n=90$). Using the current protocol, we obtained 100% mouse sequences (90/90).

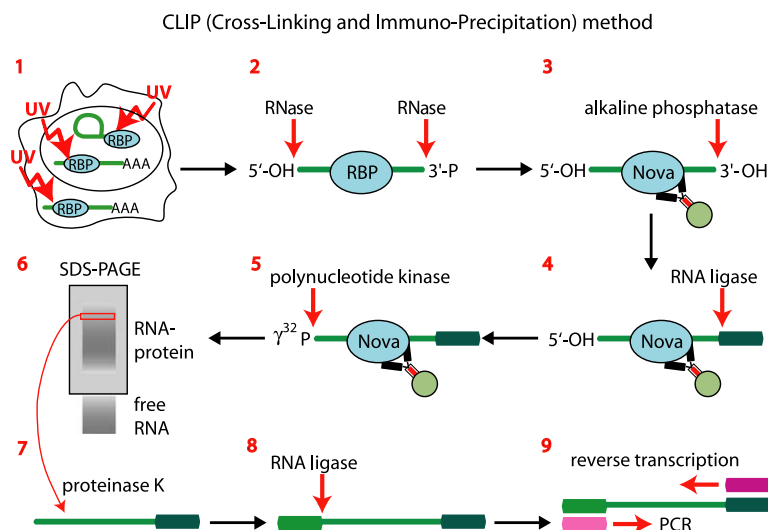


Fig. 1. A schematic representation of CLIP procedure for identification of protein binding sites on RNAs in the context of intact cells. Dissociated cells are UV-irradiated on ice, leading to formation of a covalent bond between protein and RNA. This allows partial digestion of RNA after lysis, followed by stringent protein purification, such as immunoprecipitation, SDS-PAGE electrophoresis and transfer to nitrocellulose. Prior to SDS-PAGE, RNA is dephosphorylated, ligated to 3' RNA linker and radioactively labeled by 5' $\gamma^{32}\text{P}$. After transfer to nitrocellulose, the membrane is exposed to X-ray film, and a thin region of the membrane corresponding to protein-RNA complexes of appropriate size is excised, protein is digested by proteinase K, and 5' RNA linker is ligated to free RNA. The CLIP tags are then amplified by RT-PCR using the DNA primers with sequence complimentary to the RNA linkers.

Table 1
Comparison of the previously reported CLIP protocol [35] with the protocol described here

Original protocol	Current protocol	Reason for change	Benefit of the change	Validation
Ligation of RNA linkers to free RNA (after protein digestion)	On-bead ligation of 3' RNA linker to CLIP RNA tags that are bound to the protein	Ligation to free RNA requires that significant amount of RNA is extracted to ensure that it is the major portion of cloned sequences, rather than the bacterial rRNA contaminants from the recombinant ligase or proteinase K	On-bead ligation of the 3' linker allows purification of the ligated CLIP tags and removal of the bacterial rRNA and free linker using stringent washing of the beads, SDS-PAGE and nitrocellulose transfer	Cloning CLIP tags from the same amount of starting material (25 mg) gives 53% mouse and 47% bacterial sequences using the original method, and 100% mouse sequences using the current method
Sequential ligation of RNA linkers	Several purification steps between 3' and 5' linker ligation	Sequential ligation of linkers allows ligation of 5' linker to 3' linker, which can then be amplified by the RT-PCR	Purification steps between 3' and 5' linker ligation prevent formation of linker-linker products	If starting with small amounts of CLIP RNA, the linker-linker products are amplified using the original, but not the current protocol
Phosphorylation of CLIP RNA tags before ligation of the first RNA linker (5')	Dephosphorylation of CLIP RNA tags before ligation of the first RNA linker (3')	A large excess of 5' linker needs to be added to prevent self-ligation and circularization of phosphorylated CLIP RNA tags	Dephosphorylation of CLIP tag allows use of lower concentration of RNA linker, which enables use of larger reaction volume during on-bead ligation	Dephosphorylation is required for efficient on-bead ligation
Separation of free RNA on urea-PAGE is necessary	Separation of free RNA on urea-PAGE is not necessary	In the current ligation procedure, 5' linker-3' linker product does not form and separation is thus not necessary	Allows cloning of small quantities of RNA that would not be detected when separated on the urea-PAGE	Separation of the amplified RT-PCR products gives a similar result as the separation of free RNA (Fig. 4)

An additional advantage of this modification is that it yields more efficient ligation reactions. Originally, the 5' linker ligation was done before 3' linker ligation, which allowed self-ligation and circularization of the 5' phosphorylated CLIP tag, potentially leading to the loss of a subset of RNA tags. This can be prevented by out competing the self-ligation with a large concentration of 5' linker. However, in the current protocol the first ligation step is performed on-bead, therefore it is not feasible to use such a

high concentration of RNA linkers. Instead, the current protocol dephosphorylates the RNA tags before ligation, ensuring that they cannot circularize. Introduction of the phosphatase step also potentially allows cloning of siRNAs and other short RNA species in the cell. The dephosphorylated RNA tags are first ligated to the 3' linker, which is blocked at the 3' end by puromycin, preventing self-ligation of the RNA tag during the subsequent ligation of the 5' linker.

The original protocol separated linker-ligated RNA on a denaturing acrylamide gel and visualized RNA by radioactivity. This step was necessary to remove the linker-linker multimers that occurred in the sequential ligation of RNA linkers. To be able to detect this relatively small amount of RNA, large amounts of starting material and radioactive ATP had to be used. For instance, in Nova CLIP, we need to use ~300 mg brain tissue and 0.2 mCi [γ - 32 P]ATP to easily detect linker-ligated RNA on urea-PAGE. However, the recovery of small amounts of RNA at this step was suboptimal. In the current protocol, RNA separation is not absolutely necessary.¹ Therefore, the RNA ligation is followed directly by the RT-PCR and urea-PAGE separation of DNA products, which also serves to purify CLIP tags specific to the protein of interest (Fig. 4).

3. CLIP protocol: UV cross-linking and immunoprecipitation

Prior to undertaking CLIP, it is important to optimize the conditions for protein immunoprecipitation. We recommend first determining the conditions that clear the protein from cell extracts, and then increasing salt and detergent stringency as high as possible without sacrificing the immunoprecipitation efficiency. In addition, while the protocol below utilizes a whole-cell lysate, users may want to consider preparation from specific subcellular fractions (nuclei, cytoplasm, etc).

An advantage of the covalent cross-linking of protein to RNA in CLIP is that stringent means of protein purification may be considered. While we have only used immunoprecipitation for protein purification, there is no reason that it cannot be combined with or even substituted by other purification schemes (e.g., use of protein tags, gel filtration, sucrose gradient, etc), with the one caveat that these must not include conditions that would damage nucleic acids (e.g., alkaline hydrolysis of RNA). After optimization of protein purification conditions, stringent purification during CLIP, including rigorous washes following immunoprecipitation, denaturing by heat in SDS-PAGE protein-loading buffer, separation of protein-NA complexes on SDS-PAGE gel and transfer to the nitrocellulose membrane (during which free nucleic acids pass through, while protein-NA complexes are trapped), together yield a source of nucleic acids covalently bound to a specific protein and highly purified away from contaminant nucleic acids (Fig. 1).

Here, we describe a detailed protocol for isolation of protein-binding RNA tags.

3.1. Buffers used

Buffer A: 1× PBS (tissue culture grade; no Mg²⁺, no Ca²⁺), 0.1% SDS, 0.5% deoxycholate, 0.5% NP-40

Buffer B: 5× PBS (tissue culture grade; no Mg²⁺, no Ca²⁺), 0.1% SDS, 0.5% deoxycholate, 0.5% NP-40

Buffer C: 50 mM Tris-Cl, pH 7.4, 10 mM MgCl₂, 0.5% NP-40

Buffer D: 50 mM Tris-Cl, pH 7.4, 20 mM EGTA, 0.5% NP-40

Buffer E: 100 mM Tris-Cl, pH 7.5, 50 mM NaCl, 10 mM EDTA

Buffer F (this buffer must be made fresh): 100 mM Tris-Cl, pH 7.5, 50 mM NaCl, 10 mM EDTA, 7 M urea

3.2. RNA linker sequences

L5: 5'-OH AGG GAG GAC GAU GCG G 3'-OH

L3: 5'-P GTG TCA GTC ACT TCC AGC GG 3'-puro-mycin

We order the protected RNA linkers from Dharmacon and gel purify them on our own. Run 50 μ l of 500 μ M stock of deprotected RNA on 20% polyacrylamide gel, visualize the RNA by UV shadowing, cut out the band and purify RNA as described in step 3.11.

3.3. UV cross-linking

Harvest tissue, add 10 cell volumes of ice-cold PBS and gently triturate tissue five times using a 5 ml pipette with a 100 μ l micropipette tip attached on top. Because UV light can penetrate several cell layers, stringent trituration to the single-cell level is not necessary. Transfer the suspension to a petri dish or any other container with a flat surface; the depth of the suspension should be approximately 1 mm. Irradiate three times for 100–400 mJ/cm² (~15 cm distance from UV source) in Stratalinker (Stratagene model 2400) on ice and mix the suspension between each irradiation.² We use 400 mJ/cm² three times for Nova, but have found that several other RNA binding proteins cross-link better to RNA, so in those cases we could use shorter irradiation (data not shown). Alternatively, if cell culture is used, we have found that one time irradiation for 100–400 mJ/cm² in Stratalinker is sufficient. After irradiation immediately collect and pellet cells by centrifugation at 2500 rpm for 3 min at 4°C, re-suspend the pellet in PBS and distribute 1 ml of suspension to microfuge tubes. Quick spin cells (10 s max speed) at 4°C and remove supernatant. At this point, pellets can be frozen pellets at –80°C until one is ready to continue to step 3.4.

3.4. Preparation of reagents

Use 400 μ l mixed slurry of protein A Dynabeads (Dyna, 100.02) for each microfuge tube of cross-linked lysate. Wash

¹ The first ligation step is performed before the SDS-PAGE step, thereby removing the 3' linker before ligation of the 5' linker.

² The duration of cross-linking should be optimized for each protein separately by using three different conditions and choosing the one where a protein-specific band is detectable on SDS-PAGE gel after an overnight exposure.

the beads three times with 1 ml of 0.1 M Na-phosphate, pH 8.1 and then resuspend in 320 μ l of 0.1 M Na-phosphate pH 8.1, and 80 μ l of polyclonal anti-Nova-specific antibody (the ratio of bead slurry:antibody needs to be optimized for each antibody, and clearly will depend on the titer, affinity, and IP efficiency of each antibody). Allow antibody to bind by rotating the tubes at room temperature for 30–45 min. Beads are then washed three times with 1 ml buffer A, and left in the last wash step until ready to add the cross-linked lysate. Make RNase A dilution #1 (USB 70194Y) at 1:100 in buffer A (final conc. 0.2 U/ μ l; high RNase) and RNase A dilution #2 at 1:5000 in buffer A (0.004 U/ μ l; low RNase).³

3.5. Partial RNase digestion and ultracentrifugation

To obtain RNA CLIP tags of an appropriate size, cross-linked samples are partially digested prior to immunoprecipitation. If the sample was frozen, thaw and resuspend each tube of cross-linked lysate using 600 μ l of buffer A.⁴ Add 30 μ l of RQ1 DNase (Promega, M6101; 30 U)⁵ to each tube and incubate at 37 °C for 5 min, 1000 rpm in Thermomixer R⁶ (Eppendorf). Each experiment should be done in parallel with two different RNase concentrations.⁷ RNase dilution #1 (high RNase) will give the modal size of protein–RNA complexes on SDS–PAGE gel migrating \sim 7 kDa above the expected MW of the protein (Fig. 2B), which is useful for visualization of protein–RNA complexes as discrete bands, but in most cases is not an ideal size for

³ Other RNases can also be used, such as RNase T1 [31]. Because Nova binds YCA Y-rich sequences we prefer to use RNase A (which cuts after C and U), so that we can make sure that the YCA Y enrichment in CLIP tags is not a bias due to RNase T1 (which cuts after G). Instead of RNases, alkaline hydrolysis might also be used.

⁴ Before starting CLIP, lysis conditions need to be adapted to the tissue that is used—often, sonication may be required for complete lysis. In addition, immunoprecipitation conditions needs to be optimized with the goal of finding optimum conditions for complete IP using the most stringent conditions tolerable, which empirically should yield the best possible signal:noise.

⁵ It is important to ensure that the CLIP tags are primarily cleaved by the RNase A that is added to the lysate, rather than the native RNases from the lysate—otherwise it will be hard to control for the correct size of CLIP RNA tags. If the tissue has a high level of native RNases (for instance, if pancreas is used for CLIP), then an RNase inhibitor, such as 10 mM vanadyl-ribonucleoside complex (Gibco-BRL cat#15522-014) could be added—in this case, rather than adding the RNase to the lysate, one could vary the amount of inhibitor added to get low- and high-RNase conditions. RNasin (Promega, N2515) is denatured by SDS and it requires a minimum DTT concentration of 1 mM, therefore it is of no use in the current lysis conditions. But if mild lysis conditions are used, RNasin can be added to inhibit native RNases. RNasin inhibits RNases A, B, and C, but not RNases T1 or H, so if RNasin is added to a mild lysis buffer, RNase T1 should be used for partial RNA digestion.

⁶ We find the Thermomixer useful throughout the CLIP procedure, but in most cases it is not absolutely necessary—the samples could also be manually mixed a few times during the incubation.

⁷ If only one tube of lysate is available, divide it into two and continue with half the amounts.

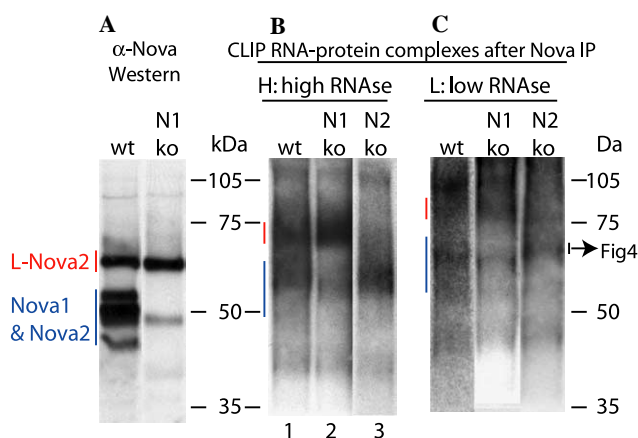


Fig. 2. Analysis of CLIP protein–RNA complexes isolated with α -Nova immunoprecipitation. (A) Western blot analysis of Nova isoforms in wild-type (wt) and *Nova1*^{-/-} (N1 ko) brain is shown to illustrate the mobility of Nova proteins not bound to RNA. (B and C) Analysis of SDS–PAGE separated protein–RNA complexes purified by Nova CLIP after RNA digestion with high (B) or low (C) RNase. The free Nova1 and Nova2 (A) or in complex with RNA (B and C) are marked with blue line, and L-Nova2 with red line.

cloning.⁸ RNase dilution #2 (low RNase) will give the modal size of protein–RNA complexes on SDS–PAGE gel migrating \sim 15–20 kDa above the expected MW of the protein (Fig. 2C), which is useful for isolating CLIP tags for cloning. Add 10 μ l of the RNase dilution #1 (high-RNase) or #2 (low RNase) to the two microfuge tubes; incubate at 37 °C for 10 min. Spin lysates in pre-chilled (4 °C) ultra-microcentrifuge with polycarbonate tubes at 60,000 rpm (32.5k, 127814 average RCF in TLA 120.2

⁸ Protein–RNA complexes that contain the RNA of \sim 50–70 nucleotides run \sim 15–20 kDa above the M_r of the non-cross-linked protein (compare Fig. 2B and B/C)—therefore the optimal region for isolation of CLIP tags is usually \sim 15–20 kDa above the M_r of the protein. RNA–protein complexes do not run exactly true to molecular weight, and additional contaminant proteins may be present (see example in Figs. 2 and 3). In such cases, we recommend that users empirically determine the optimal RNase concentration and size of protein–RNA complex to be used by gel purifying a range of protein–RNA complexes and examining the modal RNA sizes present in each fraction (see example in Figs. 2 and 3). The optimal RNase concentration has to be determined individually for each protein by keeping in mind several criteria. First, the RNase concentration will affect the amount of RNA of a particular size that will be isolated—the lower the RNase, the more long CLIP RNA tags will be isolated. Therefore, one should avoid using RNase concentrations that give a modal size of the protein–RNA complex significantly below the ideal size (15–20 kDa above the migration of the isolated protein). Moreover, an excess of RNase may lead to sequence bias; for instance, 70 nt CLIP tags remaining after high RNase A treatment (which cuts after pyrimidines) may be enriched in purines. Finally, the lower the RNase concentration, the wider the distribution of RNA tag sizes will be: therefore, in addition to increasing the amount of long RNA tags bound to the protein of interest, low RNase will also increase the amount of long RNA bound to contaminant proteins. Even though this long RNA can later be removed by size separation of free RNA or RT-PCR products proteins (see “green rectangle” contaminants in Figs. 3 and 4), it may in some cases be preferable to use high-RNase to avoid the isolation of long RNA from contaminant low MW.

rotor) for 20 min at 4 °C.⁹ Carefully recover the supernatant, and set aside 10 µl for immunoblot analysis. First time CLIP is performed, we reserve 20 µl for Total control (this can be used to clone control RNA tags bound to the total pool of RBPs; see the online protocol for cloning these control tags, <http://www.rockefeller.edu/labheads/darnellb/newsrchr.htm>).

3.6. Immunoprecipitation

Add the remaining supernatant to one prepared tube of beads, and rotate under conditions optimized for IP.¹⁰ Remove the cell extract using Dynal MPC-S magnet (cat. 120.20) and save 10 µl for immunoblot analysis (to test the relative depletion of the antigen). Wash beads two times with 1 ml ice-cold buffer A, two times 1 ml buffer B and two times 1 ml buffer C. Throughout the immunoprecipitation protocol, we wash by removing the microfuge tube from the magnet and pipetting two times up and down with 1 ml micropipette—no incubation is necessary between washes. Continue to step 3.7 with the low-RNase sample. With the high-RNase sample, skip steps 3.7 and 3.8, and store sample at 4 °C until step 3.9.

3.7. Phosphatase treatment (on-bead)

Add 80 µl of dephosphorylation reaction (8 µl of 10× dephosphorylation buffer (Roche, 712023), 3 µl Calf Intestinal Alkaline Phosphatase (Roche, 10713023001, 1 µ/µl), 69 µl water) and incubate in Thermomixer R (Eppendorf) at 37 °C for 10 min, with 1000 rpm shaking every 3 min for 15 s. Afterwards wash two times with 1 ml buffer D and two times with 1 ml buffer C.

3.8. 3' RNA linker ligation (on-bead)

Consecutively, add 40 µl of 3' RNA linker¹¹ (4 pmol/µl)¹² and 40 µl of ligation reaction (8 µl 10× T4 RNA ligase buffer (Fermentas), 8 µl BSA (0.2 µg/µl; Fermentas), 8 µl ATP (10 mM; Fermentas; we recommend aliquoting the ATP before first use, because ATP is sensitive to freeze-thaw cycles), 2 µl T4 RNA ligase (Fermentas; EL0021, 10 µ/µl), 14 µl water) to each tube of beads. Incubate at 16 °C overnight in Thermomixer (1000 rpm every 5 min for 10 s). Wash three times with 1 ml buffer C.

⁹ The first time CLIP is performed, it is important to determine the speed where none of the protein is lost in the pellet; a higher speed may help remove the ribosomes, but in case of Nova, that would lead to partial loss of protein.

¹⁰ For Nova, we rotate 1 h at 4 °C. The time of incubation and other immunoprecipitation conditions need to be pre-determined by the user.

¹¹ A good way to specifically label RNA, and test ligation efficiency, is to use a radioactive linker. In this case, a 5'-OH (dephosphorylated) 3' linker is ordered from Dharmacon and prior to ligation, phosphorylated with PNK using [γ -³²P]ATP. Afterwards (in step 3.10), RNA is phosphorylated on 5' end with cold rather than [γ -³²P]ATP. The RNA tags need to be ~50 nucleotides long to allow the access of RNA ligase (according to our experience), so labeling via hot RNA linker may appear differently on the gel than labeling via PNK.

¹² 3' linker has a 5' PO₄ and a 3' puromycin.

3.9. PNK treatment (on-bead)

Add 80 µl of PNK reaction (8 µl 10X PNK buffer (NEB), 2 µl [γ -³²P]ATP, 4 µl T4 PNK enzyme (NEB, M0201L, 10 U/µl), 64 µl water) to each sample and incubate in Thermomixer at 37 °C for 10 min (1000 rpm every 3 min for 15 s). Add 10 µl of 1 mM ATP, and let the reaction go for an additional 5 min. Wash one time with 200 µl buffer C and discard the wash carefully into radioactive waste. Wash three times with 1 ml buffer C and collect washes for liquid radioactive waste.

3.10. SDS-PAGE electrophoresis and transfer to nitrocellulose

Resuspend the beads in 30 µl of buffer C and 30 µl of Novex loading buffer (if the MW of your protein is less than 60 kDa, do not add the reducing agent, otherwise the IgG bands may interfere in the way your protein runs on the gel). Incubate at 70 °C for 10 min at 1000 rpm in Thermomixer and take the supernatant for loading. Load 1 tube per two-wells of a 10-well Novex NuPAGE 10% Bis-Tris gel (Invitrogen, NP0301BOX) and perform electrophoresis (we use Invitrogen XCell SureLock Mini-Cell, EI0001).¹³ Transfer gel (using the Invitrogen XCell II Blot Module, EI9051) to nitrocellulose S&S BA-85 nitrocellulose works best for the later RNA-protein extraction step. After transfer, rinse the nitrocellulose in 1× PBS, wrap nitrocellulose in plastic wrap while still moist and expose to X-ray film. Exposure should be on the order of 10 min to overnight.

3.11. RNA isolation and purification

If the protein binds mRNA and not DNA or small RNAs such as siRNA, then the migration of the protein-RNA complex should be different in the high- and low-RNase samples. The protein-RNA complexes in the over-digested sample lane will migrate as discrete bands ~7 kDa above the expected MW of the protein and can thus serve as size markers, and to determine whether the specific protein-RNA complex of interest has been purified (but do not cut anything from this sample, because its RNA is too small to be useful). The protein-RNA complexes that were digested by low RNase will appear as a diffuse radioactivity migrating above the MW of your protein.¹⁴ Cut three thin (~3 kDa wide) bands, of which one should be approximately 20 kDa above the MW of your protein using a clean

¹³ The Novex NuPage gels are critical, due to the buffer system that uses a pH close to pH 7. The pH of a Laemmli SDS-PAGE gel can get to ~9.5 during the gel run, which can lead to alkaline hydrolysis of the RNA.

¹⁴ If the protein-RNA complexes in the low-RNase condition are not diffuse and do not migrate higher on the gel than in the high-RNase condition, that means that the protein-RNA complex that is seen on the gel is RNase insensitive. There are several possible explanations for such a complex. The protein may be binding small RNAs, such as for instance siRNA, the size of which may not change a lot even in high RNase. Alternatively, the protein may bind DNA. A third (but unlikely) possibility is that the protein might be directly phosphorylated by the PNK.

scalpel blade, and put the nitrocellulose piece into a microfuge tube. Count radioactivity in a scintillation counter. Make a 4 mg/ml proteinase K (Roche, 1373196) solution in buffer E and pre-incubate it at 37 °C for 20 min to digest any RNases. Add 100 µl proteinase K solution to each tube of isolated nitrocellulose pieces; incubate 20 min at 37 °C at 1000 rpm in Thermomixer (Eppendorf). Add 100 µl of 1× PK/7M urea solution; incubate another 20 min at 37 °C at 1000 rpm. Add 200 µl RNA phenol (Ambion, 9710, water saturated, pH 6.6) and 65 µl of CHCl₃ (chloroform 49:1 with isoamyl alcohol) to the solution and incubate at 37 °C for 20 min at 1000 rpm in Thermomixer. Add 1/10th volume (25 µl) 3 M sodium acetate, pH 5.5 (Ambion, 9740), 0.5 µl glycogen (Ambion, 9510, 5 mg/ml) and 2.5× volume (0.8 ml) of ethanol (2.5 µg glycogen is necessary to precipitate small quantity of RNA, but do not add more, otherwise the subsequent RNA ligase may be inhibited). Precipitate overnight at –20 °C (or 30 min at –80 °C).

3.12. 5' RNA linker ligation

Spin down RNA for 10 min at max speed in microcentrifuge. Wash pellet twice with 150 µl of –20 °C 75% ethanol¹⁵ and air-dry for 5–10 min at room temperature.¹⁶ Carefully resuspend in 6 µl RNase free ddH₂O, transfer to another microfuge tube, and check for recovery of RNA by following Cerenkov counts before centrifugation, during washes and in final water solution.¹⁷ Add the RNA ligation reaction (1 µl of 10× T4 RNA ligase buffer (Fermentas), 1 µl BSA (0.2 µg/µl), 1 µl fresh ATP (10 mM), 0.1 µl T4 RNA ligase (3 U, Fermentas, EL0021, 10 U/µl) and 1 µl of L5 RNA linker @ 20 pmol/µl). Incubate at 16 °C overnight.¹⁸ Add the DNase reaction (70 µl ddH₂O, 10 µl of 10× DNase I buffer, 5 µl RNasin, 5 µl RQ1 DNase) and incubate 37 °C for 20 min¹⁹ Extract RNA (add 100 µl H₂O, 200 µl RNA phenol (Ambion, 9710) and 65 µl CHCl₃, vortex, spin, and take the aqueous layer) and precipitate overnight at –20 °C by adding 1/10th volume (25 µl) 3 M sodium acetate, pH 5.5 (Ambion, 9740), 0.5 µl glycogen (Ambion, 9510), and 0.8 ml ethanol. (<http://www.rockefeller.edu/labheads/darnellr/CLIP.php>).

¹⁵ This washing is essential to remove salts prior to RNA ligation, as RNA ligase is very sensitive to salt concentration.

¹⁶ Or a 1–2 min dry in speedvac at room temperature, but here one needs to be careful not to over-dry, otherwise RNA will not resuspend well.

¹⁷ To analyze the modal size of RNA bands, RNA can be separated on a denaturing acrylamide gel if a protein–RNA complex was detected on the membrane after less than 30 min exposure to X-ray film. The free RNA can be separated by urea–PAGE and it should be hot enough to be detected by overnight exposure to X-ray film. RNA of appropriate size can then be extracted, precipitated and used for 5' linker ligation, which serves as an additional purification step. In case of Nova, we skip this step when we use a small amount of starting material, because there is not enough free RNA to be detected on the RNA gel.

¹⁸ Using these conditions, the 5' linker ligates to ~50% of the RNA tags (data not shown).

¹⁹ DNase step is optional—it ensures that no DNA contaminants are cloned.

4. Nucleic acid amplification and cloning

4.1. DNA primers

P5: 5'-AGGGAGGACGATGCGG-3'
 P3: 5'-CCGCTGGAAGTGACTGACAC-3'
 P5Ban1a: 5'-CAGCCAACAGGCACCAGGGAGGACGATGCGG-3'
 P3Ban1a: 5'-GACTAGCTTGGTGCCGCTGGAAGTGACTGACAC-3'

4.2. RT-PCR

Spin down the NA, wash and dry the pellet and count in scintillation counter to check precipitation efficiency. Re-suspend in 8 µl ddH₂O, add 2 µl of P3 (5 pmol/µl), heat at 65 °C for 5 min in a PCR Thermocycler to denature RNA secondary structure and allow binding of the primer and chill to 4 °C. Add 3 µl of 3 mM dNTPs, 1 µl of 0.1 M DTT, 4 µl of 5× SuperScript RT buffer, 1 µl RNasin and 1 µl SuperScript III (Invitrogen, 18080-044) and incubate at 50 °C for 30 min and 90 °C for 3 min. Use 2 µl of the cDNA and perform the PCR by adding 27 µl Accuprime Pfx Supermix (Invitrogen, 12344-040), 0.5 µl P5 primer (30 pmol/µl), and 0.5 µl P3 primer (30 pmol/µl). PCR settings are: 5 min 95 °C; 30 cycles²⁰ of (20 s for 95 °C, 30 s for 61 °C, and 20 s for 68 °C); and 5 min 68 °C. Desalt the PCR on G-25 column (Amersham), pour a 10% denaturing polyacrylamide gel, run the entire PCR and visualize DNA SYBR Gold (Molecular Probes). Cut out the DNA of 80–90 nt, extract DNA with QIAEX II kit (follow protocol for polyacrylamide gel), and re-suspend in 20 µl water.

4.3. Concatemerization and cloning

After RT-PCR amplification of RNA CLIP tags, a second amplification is undertaken to introduce primers for concatemerization of CLIP tags and cloning. These primers contain a non-palindromic BanI Site (GGYRCC) to avoid sequencing errors [36]. Add 95 µl Accuprime Supermix Pfx, 1 µl P5Ban1a (100 µM) and 1 µl P3Ban1a (100 µM) to 3 µl of Qiaex II purified RT-PCR product and re-amplify by PCR with the following settings: 5 min for 95 °C, 15–20 cycles of: 15 s for 95 °C, 30 s for 60 °C, 30 s for 68 °C; and 5 min for 68 °C. Save 5 µl of the reaction to analyze on 2% agarose gel as “undigested.” Bring volume up to 200 µl with ddH₂O, and extract DNA (add 200 µl phenol/chloroform (Sigma, P3803), vortex well, spin to speed 5 min at room temperature, collect the aqueous fraction, add 200 µl chloroform 49:1 with isoamyl alcohol, vortex, spin 5 min, and collect the aqueous fraction). Precipitate DNA (30 min at –80 °C or overnight at –20 °C) by adding 25 µl of 3 M sodium acetate, pH 5.5 (Ambion, 9740) and 400 µl ethanol (2 volumes). Pellet DNA at max speed (14,000 rpm) 4 °C for 15 min,

²⁰ Depending on the amount of starting RNA, the number of PCR cycles can be modified in the range of 25–35 cycles.

wash with 70% ethanol, air dry 5 min and re-suspend in 80 μ l of 1 \times NEB buffer 4. Add 4 μ l of *Ban*I enzyme (20 U/ μ l) and incubate at 37°C for 3 h. Bring volume up to 200 μ l with water, phenol/chloroform extract, purify, and precipitate the DNA and re-suspend in 67 μ l H₂O. Save 1.5 μ l as unligated, and to the remaining 65.6 μ l add 8 μ l of 10 \times T4 DNA Ligase Buffer (NEB), 1.2 μ l P5*Ban*Ia (100 μ M), and 1.2 μ l P3*Ban*Ia (100 μ M). To quench the short cleaved ends and prevent their re-ligation to tags, incubate the sample in Thermocycler 10 min at 65°C, cool to 4°C and keep on ice. Add 4 μ l of T4 DNA Ligase (NEB; 2000 U/ μ l) and incubate overnight at 16°C. Desalt on G-25 column (Amersham) and run on a 2% low melting (NuSieve) agarose gel. Separate well, cut out and purify the desired ligation concatamer product using Qiaex II. Fill in the *Ban*I overhangs and generate A-overhang using *Taq* Polymerase, clone in pCR2.1-TOPO vector (Invitrogen, K450001) and sequence using M13 primer.

5. Results

Detection of protein–RNA separated by SDS–PAGE gel is a critical purification and quality control step in CLIP, as it enables selective purification of RNA tags that correspond to the direct protein-binding site, allows comparison of experimental to control samples, and confirms efficient cross-linking of protein to RNA. We have used one postnatal (day 7) brain (~150 mg tissue) as a starting material to immunoprecipitate Nova-RNA to routinely detect signal on Kodak Biomax MR X-ray film after 30 min exposure (Figs. 2B and C). A smaller amount of starting material (10 mg tissue) can still generate high quality CLIP tag data, but requires an overnight exposure of nitrocellulose to X-ray film to detect a clear signal.

To ensure that the isolated tags correspond to the protein of interest, and not to other contaminating proteins, several controls should be considered. First, an experiment with no UV cross-linking should be done—in the case of Nova, no radioactivity is detected after purification from this experiment, confirming that the UV-induced covalent bond between protein and RNA is necessary for purification in the CLIP procedure [35]. An even more important control is an experiment using a biologic material that lacks the protein that is to be immunoprecipitated. For tissue-based experiments, we have used knockout mouse tissue as a control. If a transfected tagged protein is being purified, a vector control could be used. To analyze these control experiments, the high RNase experiment is used (Figs. 2B and H: high RNase). In this experiment, the proteins are bound to short RNA tags and therefore migrate only ~7–15 kDa higher than the expected MW of the protein and are visible as relatively discrete bands on the gel (Fig. 2B, see also Fig. 1 in [35]).

As an example of the importance of size selecting CLIP'ed protein–RNA complexes for purification and cloning of RNA tags, we show representative data from a Nova CLIP experiment. This CLIP gives a complex series

of targets, in part because different Nova isoforms immunoprecipitate in the experiment. Nova1 has two splice isoforms of ~50–52 kDa in addition to a minor 45 kDa isoform, and Nova2 has an isoform of ~49 kDa and a higher molecular weight isoform migrating at ~66 kDa (L-Nova2) ([37] and data not shown) (Fig. 2A, where the 45–52 kDa Nova products are marked in blue, and L-Nova2 in red). When Nova protein–RNA CLIP products are run on SDS–PAGE, three major complexes are seen under conditions of high RNase treatment (Fig. 2B). An upper band corresponding to L-Nova2 (red) and a middle band corresponding to Nova1 and Nova2 (blue) are present. These can be unambiguously identified, as the upper Nova2 band is missing in Nova2^{-/-} CLIP samples (Fig. 2B, lane 3), and the Nova1 bands are missing (albeit with a residual signal from Nova2) in Nova1^{-/-} CLIP samples (Fig. 2B, lane 2). The M_r of these bands, even under these conditions of high concentrations of RNase, is larger than that of proteins on standard Western blots due to the presence of cross-linked RNA (we estimate this RNA to have a modal size of 20–50 nt, or ~7–16 kDa). However, as the different proteins are bound to the same modal size RNAs, the distance between the bands is preserved (Figs. 2B and C). A third band is present on the CLIP gels, running below the size of Nova, at M_r ~42–45 kDa, which we believe corresponds to an unidentified ~35 kDa protein present in the Nova CLIP samples (assuming it is bound to ~20–25 nt RNA). This band is present in both Nova1^{-/-} and Nova2^{-/-} brain, suggesting that it may correspond to an unrelated protein that coimmunoprecipitates with Nova, or a cross-reacting RNA binding protein that binds to the antiserum.

Using low RNase A generates longer RNA tags which vary in size from ~50–200 nt (Fig. 3, lanes marked “L”). This results in more diffuse migration of protein–RNA complexes on the gel (L: low-RNase, Fig. 2C).²¹ To analyze the size of RNA tags bound to proteins running at different molecular sizes, five thin bands of M_r of ~50, 60, 70, 75, and 90 kDa were excised from both the high- (H) and low-RNase (L) samples isolated from wild-type brain (Fig. 3A). After removing protein with proteinase K, free RNA tags were separated on a denaturing acrylamide gel and imaged by autoradiography (Fig. 3B). As expected, 20–60 nt RNAs are predominant in high RNase samples (Fig. 3B, lanes marked “H”),²² and 50–180 nucleotide RNAs are predominant in low RNase samples (Fig. 3B, lanes marked “L”).

We correlated the size of the RNA bands with their origin in the Nova-RNA SDS–PAGE co-IP gels. The protein–RNA band #1 (~50 kDa) was excised below the size of Nova–RNA complexes (Fig. 3A). The RNA isolated from this band has a modal size of ~50 nt (Fig. 3B, #1, green

²¹ The absence of discrete bands in the low-RNase sample also suggests that Nova does not bind DNA or small RNAs such as siRNAs.

²² RNA smaller than 40 nt does not precipitate very well in our hands, especially because we are working with very small concentration of RNA. Therefore, the amount of RNA smaller than 40 nt may not be properly represented by the RNA gel.

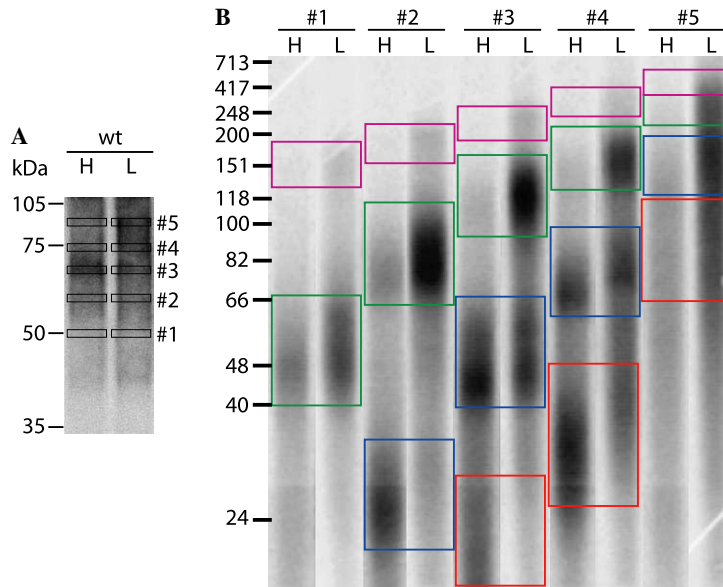


Fig. 3. Size separation of free RNA isolated from Nova CLIP protein–RNA complexes. (A) Five thin regions (marked by rectangles) from the SDS–PAGE separated protein–RNA complexes were used for isolation of RNA. (B) After digestion of protein, radioactively labeled free RNA was separated by 4% denaturing poly-acrylamide gel and visualized by exposure to X-ray film. The lengths of the RNA ladder are shown on the left; wt, wildtype postnatal day 7 (P7) brain; N1 ko, *Nova1*^{-/-} P1 brain (we had to collect P1, rather than P7 brain due to the early postnatal lethality of *Nova1*^{-/-} mice); N2 ko, *Nova2*^{-/-} P7 brain.

rectangle). We interpret this RNA to most likely be CLIP-tags derived from the 35-kDa protein for the following reason. The predicted MW of a 50 nt RNA is 16.2 kDa; if the RNA migrates close to its predicted M_r , then when cross-linked to a 35-kDa protein it would form a 50-kDa complex. In addition, a faint RNA band of ≥ 150 nt is present (Fig. 3B, #1, purple rectangle), which has a predicted MW of 50 kDa. This RNA may correspond to small amounts of free RNA separated by the SDS–PAGE and giving background on the nitrocellulose membrane (although the bulk of free RNA passes through the nitrocellulose).

The protein–RNA band #2 (Fig. 2C) was excised from the center of the band of Nova1–RNA complexes from the high-RNase sample (~ 58 – 60 kDa) (Fig. 3A). This protein–RNA band yields RNA of ~ 25 nt (Fig. 3B, #2, lane H, blue rectangle; ~ 8 kDa), which we interpret as bound by the 49–52 kDa Nova1 and Nova2 protein isoforms (Figs. 2A and B). In addition, an RNA band of ~ 80 nt (~ 26 kDa) is present (Fig. 3B, #2, green rectangle), which we interpret as bound by the ~ 35 kDa protein.²³ These data and subsequent analyses (below) demonstrate that size selection of thin bands of CLIP’ed protein–RNA complexes (Fig. 3A), together with the use of different RNase concentrations and size selection of RNA (or PCR amplified products), can yield significant additional purification.

The protein–RNA band #3 (~ 70 kDa) was excised from the center of the band of Nova1–RNA complexes (and the minor Nova2–RNA) from the low-RNase sample (~ 68 –

70 kDa). This band yields the RNA of ~ 40 – 60 nt band (Fig. 3B, #3, blue rectangle) which we interpret as bound by the 49–52 kDa Nova1 and Nova2 protein isoforms (Fig. 2B). In the high-RNase sample (Fig. 3, #3, red rectangle) we interpret the short (< 24 nt) RNAs as L-Nova2 CLIP tags, because the protein–RNA complexes were isolated ~ 3 kDa above the M_r of L-Nova2 (66 kDa). In addition, longer RNAs of 100–150 nt (Fig. 3, green rectangle), and ≥ 200 nt (Fig. 3, purple rectangle), are seen, which we interpret as 35 kDa protein CLIP tags and free RNA, respectively. This again illustrates the potential advantage of excising thin bands of RNA–protein complexes from the nitrocellulose membrane. Each such band subsequently led to identification of multiple (up to four) discrete bands on the RNA gel, allowing separation of free RNA and the RNA bound to the different proteins. In addition, analysis of band #3 shows that either the high-RNase and low-RNase samples could be used to clone Nova1 CLIP tags of ~ 60 nt length.

The RNA tags bound to L-Nova2 in band #4 are ~ 35 nt long (Fig. 3, #4, red rectangle), which corresponds to M_r of 11 kDa. These tags are too short for efficient CLIP analysis. Therefore, a band at 90 kDa was also excised, corresponding to L-Nova2 bound to RNAs of ~ 80 nt (Fig. 3, #5, red rectangle). In general, protein–RNA complexes migrating approximately 15–20 kDa higher than the expected M_r of the free protein are used for isolating corresponding RNA tags, as these should correspond to cross-linked RNA of 50–70 nt, although in some proteins, the ideal size of an RNA tag may be different. For example, some proteins, such as FMRP, may have a more complex binding site, and CLIP tags may need to be longer to encompass the whole binding site [10,18]. In other cases, such as proteins that bind siRNAs, tags as short as 21 nucleotides are preferred,

²³ The precipitation of small RNAs is often inefficient, which may explain why the ~ 80 nt band appears stronger than the ~ 25 nt band, even though the Nova1–RNA band on the SDS–PAGE gel is stronger than the band of the 35-kDa protein bound to RNA.

therefore the protein–RNA complexes migrating approximately 6 kDa higher than the expected MW of the free protein should be used for isolating corresponding RNA tags.

We used RT-PCR to amplify the CLIP tags bound to Nova1/Nova2 from a thin (~3 kDa wide) region of the SDS–PAGE separated protein–RNA complexes at approximately 70 kDa, which is ~20 kDa above the expected M_r of Nova1 (arrow in Fig. 2C). The free RNA in protein–RNA complexes of this MW from low RNase samples showed the presence of two RNA species of 40–60 and 100–120 kDa (#3, L). After ligation of RNA linkers and PCR, the size of the product should increase for 36 nucleotides (20 for 3' linker and 16 for 5' linker), therefore the amplified RNA with two RNA linkers should migrate at a modal size of ~90 and ~150 nt. Analysis of RT-PCR amplified CLIP tags appear as two bands migrating at ~90 kDa and ~150 nt, which corresponds to the predicted sizes of ligated CLIP tags (Fig. 4). The ~150 nt band corresponds to the CLIP tags of the ~35 kDa protein, therefore the intensity of this band is the same in both *Nova1*^{-/-} and *Nova2*^{-/-} samples. On the other hand, the ~90 nt band is diminished when the RNA is isolated from the *Nova1*^{-/-} brain (Fig. 4, lane 2; the remaining signal may correspond to the small Nova2 isoform bound to RNA). These results suggest that the ~90 nt product corresponds primarily to Nova1-bound RNA, and agrees with the prediction that the RNA of ~55 nt shifts the migration of Nova1 to ~70 kDa. Therefore, we can purify the ~90 nt PCR

products isolated from the *Nova2*^{-/-} brain to specifically identify Nova1-binding CLIP tags. In conclusion, if several proteins are co-immunoprecipitated, it may be helpful to cut thin bands from the SDS–PAGE separated protein–RNA complexes to allow confirmation that CLIP tags amplified from the isolated RNA correspond to the protein of interest, rather than another co-immunoprecipitated protein.

6. Concluding remarks

Gene expression is regulated by proteins that interact with specific sequences on DNA and RNA in the context of macro-molecular complexes, precise sub-cellular location and developmental timing. The use of UV light for fixation of these interactions is the only method currently established to specifically analyze direct protein–RNA contacts in vivo, because it forms very precise protein–RNA covalent bonds, while it does not form protein–protein cross-links. The cross-linking procedure itself is fast (several minutes) and is performed at 4°C to prevent biologic events such as UV-induced activation of DNA damage response pathways. The CLIP method exploits the irreversibility of the covalent bond formed by UV cross-link to purify protein–RNA complexes and separate them on SDS–PAGE gel, which allows isolation of RNA tags that bind a specific protein.

Even though the steady state of different RNA molecules can vary greatly in the cells, CLIP ensures that high-abundance non-specific sequences (such as ribosomal RNA) do not out-compete the low abundance specific protein-binding sequences. For instance, CLIP can isolate intronic RNA sequences that normally occur at a very low steady state level in the cell [35]. On the other hand, among different RNA targets of a single protein, CLIP will yield RNAs in proportion to their abundance, and this may need to be considered in interpreting results. For this purpose, the relative frequency of CLIP tags in different positions on the pre-mRNAs and mature RNAs has to be normalized.²⁴

We often see a single nucleotide substitution within the YCAY motif of the Nova CLIP tags, suggesting that the amino acid that remains bound to RNA after protein digestion often leads to a mutation. Previous studies have also shown that such remaining amino acid can interfere with the reverse transcription, and also occasionally cause a stop in reverse transcription [29]. Thus, if a large number of CLIP tags are cloned, the positions of single nucleotide substitutions may potentially be used to determine the nucleotide that was cross-linked to the RNA-binding protein.²⁵

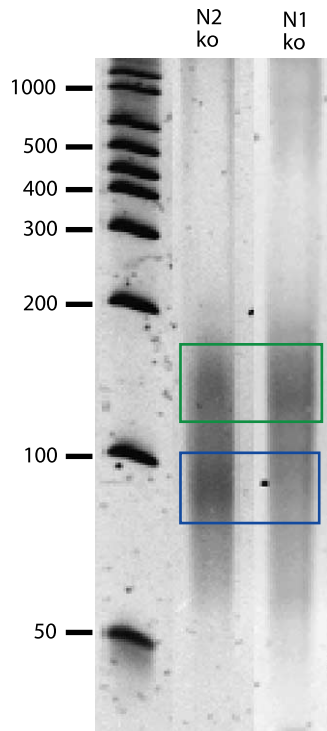


Fig. 4. Size separation of CLIP tags amplified by RT-PCR from the Nova CLIP protein–RNA complexes (from the region indicated by an arrow in Fig. 2C) isolated from *Nova2*^{-/-} (N2 ko) and *Nova1*^{-/-} (N1 ko) brain. After purification of RNA and ligation of both RNA linkers, the RT-PCR amplified CLIP tags were separated by 4% denaturing poly-acrylamide gel and visualized by staining with SYBR Gold. The lengths of the DNA ladder are shown on the left.

²⁴ In addition to bioinformatical and microarray approaches to normalize the data, one could also compare the CLIP data obtained with different RNA-binding proteins. Furthermore, a control experiment is described in detail online (<http://www.rockefeller.edu/labheads/darnell/CLIP.php>), which isolates RNA tags bound to the total pool of RNA-binding proteins in the cell.

²⁵ From the technical standpoint, care needs to be taken not to increase the efficiency of UV cross-link to the extent that multiple cross-links between protein and RNA are formed, because this might impair productive reverse transcription.

The CLIP method has been instrumental in confirming that Nova binds YCAY-rich sequences *in vivo*, that it binds either intronic or exonic sequences to regulated alternative exon usage, and that in addition it also binds to mature RNAs, in particular 3' UTR sequences, whereby Nova may contribute to cytoplasmic regulation of these RNAs [35]. While we have not discussed validation experiments here, we found that ~30% of a subset of Nova CLIP tags that suggested a role in alternative splicing correctly predicted a splicing regulatory role of Nova by analyzing *Nova2*^{-/-} brain [35]. We also showed that several of these tags were capable of binding Nova2 *in vitro* and mediating Nova-dependent splicing in a heterologous context [35,38]. The remaining CLIP tags suggest that Nova may be playing roles other than alternative splicing regulation, and this hypothesis remains to be tested. In any case, the importance of CLIP tag validation using genetic and functional assays cannot be underestimated in following up CLIP experiments.

Acknowledgments

We thank members of the lab for their help and insight throughout this work. Supported by the NIH (R01 NS34389 and NS40955 to R.B.D.) and the Howard Hughes Medical Institute. R.B.D. is an Investigator of the Howard Hughes Medical Institute.

References

- [1] J.C. Venter, et al., *Science* 291 (2001) 1304–1351.
- [2] C. Tuerk, L. Gold, *Science* 249 (1990) 505–510.
- [3] A.D. Ellington, J.W. Szostak, *Nature* 346 (1990) 818–822.
- [4] K.B. Jensen, L. Green, S. MacDougal-Waugh, C. Tuerk, *J. Mol. Biol.* 235 (1994) 237–247.
- [5] R.J. Buckanovich, R.B. Darnell, *Mol. Cell. Biol.* 17 (1997) 3194–3201.
- [6] K.B. Jensen, K. Musunuru, H.A. Lewis, S.K. Burley, R.B. Darnell, *Proc. Natl. Acad. Sci. USA* 97 (2000) 5740–5745.
- [7] C.G. Burd, G. Dreyfuss, *EMBO J.* 13 (1994) 1197–1204.
- [8] I. Perez, C.H. Lin, J.G. McAfee, J.G. Patton, *RNA* 3 (1997) 764–778.
- [9] J.C. Darnell, O. Mostovetsky, R.B. Darnell, *Genes Brain Behav.* (2005).
- [10] J.C. Darnell, K.B. Jensen, P. Jin, V. Brown, S.T. Warren, R.B. Darnell, *Cell* 107 (2001) 489–499.
- [11] P. Trifillis, N. Day, M. Kiledjian, *RNA* 5 (1999) 1071–1082.
- [12] M. Itoh, I. Haga, Q.H. Li, J. Fujisawa, *Nucleic Acids Res.* 30 (2002) 5452–5464.
- [13] S.A. Brooks, W.F. Rigby, *Nucleic Acids Res.* 28 (2000) E49.
- [14] S.A. Tenenbaum, C.C. Carson, P.J. Lager, J.D. Keene, *Proc. Natl. Acad. Sci. USA* 97 (2000) 14085–14090.
- [15] V. Brown, P. Jin, S. Ceman, J.C. Darnell, W.T. O'Donnell, S.A. Tenenbaum, X. Jin, Y. Feng, K.D. Wilkinson, J.D. Keene, R.B. Darnell, S.T. Warren, *Cell* 107 (2001) 477–487.
- [16] H. Hieronymus, P.A. Silver, *Nat. Genet.* 33 (2003) 155–161.
- [17] S.A. Waggoner, S.A. Liebhaber, *Mol. Cell. Biol.* 23 (2003) 7055–7067.
- [18] J.C. Darnell, C.E. Fraser, O. Mostovetsky, G. Stefani, T.A. Jones, S.R. Eddy, R.B. Darnell, *Genes Dev.* 19 (2005) 903–918.
- [19] S. Mili, J.A. Steitz, *RNA* 10 (2004) 1692–1694.
- [20] M.J. Solomon, P.L. Larsen, A. Varshavsky, *Cell* 53 (1988) 937–947.
- [21] S. Niranjankumari, E. Lasda, R. Brazas, M.A. Garcia-Blanco, *Methods* 26 (2002) 182–190.
- [22] R. Brimacombe, W. Stiege, A. Kyriatsoulis, P. Maly, *Methods Enzymol.* 164 (1988) 287–309.
- [23] J.W. Hockensmith, W.L. Kubasek, W.R. Vorachek, P.H. von Hippel, *J. Biol. Chem.* 261 (1986) 3512–3518.
- [24] M.D. Shetlar, J. Christensen, K. Hom, *Photochem. Photobiol.* 39 (1984) 125–133.
- [25] J.R. Greenberg, *Nucleic Acids Res.* 6 (1979) 715–732.
- [26] A.J. Wagenmakers, R.J. Reinders, W.J. van Venrooij, *Eur. J. Biochem.* 112 (1980) 323–330.
- [27] G. Dreyfuss, Y.D. Choi, S.A. Adam, *Mol. Cell. Biol.* 4 (1984) 1104–1114.
- [28] D.S. Gilmour, J.T. Lis, *Proc. Natl. Acad. Sci. USA* 81 (1984) 4275–4279.
- [29] H. Urlaub, K. Hartmuth, R. Luhrmann, *Methods* 26 (2002) 170–181.
- [30] B.M. Merrill, K.L. Stone, F. Cobianchi, S.H. Wilson, K.R. Williams, *J. Biol. Chem.* 263 (1988) 3307–3313.
- [31] G.P. Pfeifer, S.D. Steigerwald, P.R. Mueller, B. Wold, A.D. Riggs, *Science* 246 (1989) 810–813.
- [32] P.R. Mueller, B. Wold, *Science* 246 (1989) 780–786.
- [33] R. Santoro, S. Wolf, H.P. Saluz, *Biochem. Biophys. Res. Commun.* 256 (1999) 68–74.
- [34] P.J. Solano, B. Mugat, D. Martin, F. Girard, J.M. Huibant, C. Ferraz, B. Jacq, J. Demaille, F. Maschat, *Development* 130 (2003) 1243–1254.
- [35] J. Ule, K.B. Jensen, M. Ruggiu, A. Mele, A. Ule, R.B. Darnell, *Science* 302 (2003) 1212–1215.
- [36] S. Pfeffer, A. Sewer, M. Lagos-Quintana, R. Sheridan, C. Sander, F.A. Grasser, L.F. van Dyk, C.K. Ho, S. Shuman, M. Chien, J.J. Russo, J. Ju, G. Randall, B.D. Lindenbach, C.M. Rice, V. Simon, D.D. Ho, M. Zavolan, T. Tuschl, *Nat. Methods* 2 (2005) 269–276.
- [37] Y.Y. Yang, G.L. Yin, R.B. Darnell, *Proc. Natl. Acad. Sci. USA* 95 (1998) 13254–13259.
- [38] B.K. Dredge, G. Stefani, C.C. Engelhard, R.B. Darnell, *EMBO J.* (2005) 27.