

# Using hiCLIP to identify RNA duplexes that interact with a specific RNA binding protein

Yoichiro Sugimoto<sup>1,2</sup>, Anob M. Chakrabarti<sup>2,3</sup>, Nicholas M. Luscombe<sup>2,4</sup> & Jernej Ule<sup>1,2</sup>

<sup>1</sup>Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London, WC1N 3BG, UK

<sup>2</sup>The Francis Crick Institute, 1 Midland Road, London, NW1 1AT, UK

<sup>3</sup>UCL Genetics Institute, Department of Genetics, Evolution and Environment, University College London, Gower Street, London WC1E 6BT, UK

<sup>4</sup>Okinawa Institute of Science and Technology Graduate University, 1919-1 Tancha, Onna-son, Okinawa 904-0495, Japan

## Author details

- Yoichiro Sugimoto
  - Department of Molecular Neuroscience at the UCL Institute of Neurology, London, UK and The Francis Crick Institute, London, UK
  - yoichiro.sugimoto@crick.ac.uk
- Anob M. Chakrabarti
  - The Francis Crick Institute, London, UK and UCL Genetics Institute, Department of Genetics, Evolution and Environment, London UK
  - anob.chakrabarti@crick.ac.uk
- Nicholas M. Luscombe
  - The Francis Crick Institute, London, UK; UCL Genetics Institute, Department of Genetics, Evolution and Environment, London UK and Okinawa Institute of Science and Technology Graduate University, Okinawa, Japan
  - nicholas.luscombe@crick.ac.uk
- Jernej Ule
  - Department of Molecular Neuroscience at the UCL Institute of Neurology, London, UK and The Francis Crick Institute, London, UK
  - jernej.ule@crick.ac.uk
  - Telephone number: +44(0)2037963137

## ABSTRACT

The structure of RNA molecules is crucial to regulating gene expression<sup>1</sup>, largely through influencing interactions with RNA binding proteins (RBPs). Hybridization individual-nucleotide resolution UV crosslinking and immunoprecipitation (hiCLIP) is a transcriptome-wide method to monitor these interactions by identifying RNA duplexes bound by a specific RBP. The hiCLIP protocol consists of the following steps: *in vivo* crosslinking of RBPs to their bound RNAs; partial RNA digestion and purification of RNA duplexes interacting with the specific RBP using immunoprecipitation; ligation of the two arms of RNA duplexes via a linker; reverse transcription; cDNA library amplification, and finally high-throughput DNA sequencing. Mapping of the sequenced arms to a reference transcriptome identifies the exact locations of duplexes. hiCLIP data can directly identify all types of RNA duplexes bound by RBPs, including those that are challenging to predict computationally, such as inter-molecular and long-range intra-molecular duplexes. Moreover, the use of an additional adaptor that links the two arms of the RNA duplex permits hiCLIP to unambiguously identify the duplexes. Here we describe in detail the procedure for a hiCLIP experiment and the subsequent streamlined data analysis with an R package, “hiclipr” (<https://github.com/luslab/hiclipr>). Preparation of the library for high-throughput DNA sequencing takes approximately seven days, and the basic bioinformatic pipeline takes one day.

## INTRODUCTION

RNA structures play a pivotal role in the regulation of gene expression. Appropriate folding of RNA is essential for the assembly of ribonucleoprotein (RNP) complexes, such as the ribosome and spliceosome, and their catalytic activities. Moreover, RNA structure also modulates the binding of regulatory molecules to mRNAs, such as micro RNAs (miRNAs) and RNA binding proteins (RBPs) that control production of encoded proteins<sup>1,2</sup>. One of the central aspects of RNA folding is the formation of RNA duplexes. Intra-molecular RNA duplexes result from base pairing within an RNA molecule, which can lead to diverse RNA structural motifs, such as the stem-loop and pseudoknot<sup>3</sup>. In addition, inter-molecular RNA duplexes can arise from base pairing between different RNA molecules, including miRNA-mRNA<sup>4</sup>, lncRNA-mRNA<sup>5,6</sup>, and mRNA-mRNA duplexes<sup>7</sup>, among others<sup>8-11</sup>. Due to their importance,

various methods have been developed to assess comprehensively the nature and role of RNA structures and duplexes in a cell with high-throughput DNA sequencing.

### **High-throughput DNA sequencing approaches to study *in vivo* RNA structures**

Two primary approaches use high-throughput DNA sequencing to study RNA structures: (i) methods that employ nucleotide probing, using chemicals with different reactivity for single-stranded or double-stranded nucleotides, and (ii) methods that employ proximity ligation to join the two strands of RNA duplexes<sup>2</sup> (Fig. 1).

The first genome-wide method to study RNA structures, named parallel analysis of RNA structure (PARS), probes single- or double-stranded nucleotides using RNases that preferentially cut either single-stranded or double-stranded regions, and analyzes the probed sites using high-throughput DNA sequencing<sup>12</sup>. However, this RNase-based approach can only be used *in vitro* with purified RNA. In order to study *in vivo* RNA structures, methods using cell-membrane-permeable chemicals to probe single-stranded nucleotides *in vivo* were developed. These methods, including dimethyl sulphate-sequencing (DMS-Seq)<sup>13,14</sup> and *in vivo* click selective 2'-hydroxyl acylation and profiling experiment (icSHAPE)<sup>15,16</sup>, revealed the general characteristics of RNA structures in the cell. However, the data need to be interpreted with caution to gain insights into these structures, since the *in vivo* nucleotide probing efficiency can be influenced not only by single-strandedness, but also by associated proteins, which can confound the analysis<sup>17</sup>. In addition to assessing the single-strandedness of nucleotides, the information can also be used to improve the accuracy of computational predictions of RNA secondary, or potentially even higher-order, structures<sup>13-15,18</sup>. However, these methods share some of the same limitations as the current computational RNA structure prediction tools. They have difficulties resolving long-range intra-molecular RNA-RNA contacts or inter-molecular RNA-RNA contacts, where too many possible alternative structures need to be considered. Furthermore, the prediction of tertiary RNA structures such as pseudoknots and kissing-loops remains challenging for most programs<sup>19</sup>.

An alternative approach to studying RNA structures uses proximity ligation, which directly identifies RNA duplexes. Here, single-stranded RNA is digested, leaving

behind the RNA duplexes; the two duplex arms are then ligated to generate a single hybrid RNA molecule. Thus, by analyzing the hybrid RNAs using high-throughput DNA sequencing, it is possible to examine which regions of RNAs form duplexes. The first high-throughput method using proximity ligation, termed crosslinking, ligation, and sequencing of hybrids (CLASH), was developed to study *in vivo* RNA duplexes recognized by a specific RNA binding protein<sup>20</sup>. CLASH uses a modified version of the CLIP protocol. Like CLIP, CLASH employs UV-C irradiation to crosslink RBPs to the bound RNAs, followed by immunopurification of the RBP-RNA complex. These RNA fragments are then identified using high-throughput DNA sequencing. In CLASH, there is an additional proximity ligation step, which is designed to ligate together the two arms of the isolated RNA duplexes. CLASH has been successfully used to study RNA duplexes formed by snoRNA-rRNA<sup>20</sup> and mRNA-miRNA<sup>21,22</sup> in a genome-wide manner.

While CLASH can identify *in vivo* RNA duplexes recognized by RBPs in a high-throughput manner, it does not examine the structures of unbound RNAs. In order to study all RNA duplexes in the cell, methods crosslinking the two arms of RNA duplexes using a psoralen derivative (4'-aminomethyltrioxalen: AMT) were developed. The psoralen derivative, AMT, is a water-soluble chemical that can form crosslinks between double-stranded uridine and its opposing nucleotide upon irradiation with 365 nm UV light; the crosslinks can be reversed by irradiation with 254 nm UV light<sup>23,24</sup>. The psoralen-crosslinking based methods include sequencing of psoralen cross-linked, ligated, and selected hybrids (SPLASH)<sup>9</sup>, psoralen analysis of RNA interactions and structures (PARIS)<sup>10</sup>, and ligation of interacting RNA followed by high-throughput sequencing (LIGR-Seq)<sup>11</sup>. These methods first crosslink RNA duplexes by incubating the cells with psoralen, purify the crosslinked RNA duplexes, ligate the two arms of RNA duplexes to generate hybrid RNAs, and analyze the hybrid RNAs in a similar way to CLASH. These psoralen-crosslinking based methods have been used to reveal a global view of RNA duplexes<sup>9-11</sup>.

### **Development of hiCLIP**

We built on the insights of the CLASH technique in developing hiCLIP<sup>8</sup>, which also relies on proximity ligation to identify RNA duplexes bound by RBPs. hiCLIP

introduces a unique ligation step, where an additional linker is used to ligate the two arms of an RNA duplex (Fig. 1). Thereby, hiCLIP overcomes the following three limitations of other proximity ligation methods. First, a flexible single-stranded RNA next to the RNA duplexes is likely to be required to ligate the two arms of RNA duplexes, as was shown by the need of at least 8 nucleotides in an RNA for the efficient circularization<sup>8,25</sup>. Such single-stranded RNA overhang may not exist after the partial RNA digestion. The hiCLIP linker adaptor adds a flexible end to one arm of RNA duplexes, and thus enables the efficient ligation of the two arms. Second, potential endogenous ligase activity can make data interpretation more difficult<sup>24</sup>. Use of the linker ensures all hybrid reads arise as a result of the controlled ligation reaction. Third, separating the two arms of hybrid RNAs generated by the direct ligation is computationally challenging after sequencing. Introduction of the identifiable linker between the two duplex arms allows unambiguous identification of these two arms in hybrid reads before mapping. This enables hiCLIP to be applied to a variety of dsRBPs without the need for any prior information regarding the target RNA duplexes.

### **Overview of the hiCLIP procedure**

Here we provide a detailed protocol for hiCLIP and subsequent bioinformatic analysis using a specific RBP and cell line as an example (FLAG-tagged Staufen 1 protein, STAU1, in an HEK293 cell line). However, like CLIP<sup>26</sup> or iCLIP<sup>27</sup>, hiCLIP should be generally applicable to any dsRBP, cell line or tissue with a few modifications considering the factors detailed in **Experimental Design**.

The procedure to generate a high-throughput DNA sequencing compatible library consists of 10 stages (Fig. 2). (Stage 1) The cells are irradiated with UV-C light to form covalent bonds between directly interacting RBPs and RNA. The UV cross-linking is used to stabilize the RNA duplexes on RBPs. This enables extensive purification of the RBP of interest in order to examine only those RNA duplexes in direct contact with the protein. (Stage 2) The cells are then lysed with a detergent-containing buffer, and (Stage 3) RNA is partially digested using RNase I. (Stage 4) The RBP of interest and the crosslinked RNA duplexes are purified with

immunoprecipitation, and (Stage 5) a different type of RNA adaptor is ligated to the 3' end of each arm of the RNA duplex. RNA Adaptor A has an irreversible 3' block (dideoxynucleotide) that prevents additional molecules ligating to it, and later serves as a priming site of the reverse transcription primer. RNA Adaptor B has an enzymatically removable 3' block (phosphorylation), and later serves as a linker when generating hybrid RNAs. With identical concentrations of Adaptors A and B, half of the RNA duplexes where both arms are ligated by adaptors will have one arm ligated by Adaptor A and the other arm ligated by Adaptor B. RNA duplexes ligated by only Adaptor A will be cloned as a non-hybrid RNA whereas RNA duplexes ligated by only Adaptor B will not be cloned. (Stage 6) The 3' block of adaptor B is removed, and (Stage 7) the two arms of the RNA duplexes are ligated via adaptor B. Since the two arms of RNA duplexes are proximal to each other, they are the preferred substrates for the ligation reaction. (Stage 8) The RBP-RNA duplex complex is further purified with a second round of immunoprecipitation, SDS-PAGE, and nitrocellulose membrane transfer. To monitor the quality of experiments, a fraction of the RNA is radio-labeled before SDS-PAGE and visualized on nitrocellulose membrane using autoradiography. (Stage 9) The RNA is then extracted from the crosslinked RBP, (Stage 10) reverse transcribed with a primer containing a unique molecular identifier and the cDNA is PCR amplified to generate a library compatible with high-throughput sequencing, using a protocol similar to iCLIP<sup>27,28</sup>.

The bioinformatic analysis (Stage 11) uses the tailored R package “hiclipr” (<https://github.com/luslab/hiclipr>). First, hybrid reads are identified based on the presence of the linker sequence (Adaptor B), and defining the sequence of each hybrid arm. Each hybrid arm is mapped individually to a reference genome and transcriptome in four phases: (i) first, reads are aligned to annotated ribosomal and transfer RNAs; (ii) any remaining unaligned reads are then mapped to the mitochondrial genome; (iii) next, remaining reads are aligned to a reference transcriptome and (iv) finally, remaining reads are mapped to the reference genome. Alignments from phase 1, 3 and 4 are combined for further analysis. Alignments from phase 2 are usually a tiny proportion of all alignments and are discarded as here we focus on cytoplasmic RNAs. In phases 3 and 4, only uniquely aligning reads are allowed and this alignment strategy maximizes the yield of aligned hybrid reads. PCR

duplicates are removed using a unique molecular identifier introduced by the reverse transcription primer. Each hybrid arm is then annotated with the gene of origin, the transcript biotype (e.g. protein-coding, microRNA, long-non-coding RNA) and the transcript region (e.g. 5' UTR, CDS, 3' UTR).

To define RNA duplexes bound by RBPs, the intragenic hybrids (i.e. hybrids in which both arms align to the same RNA transcript) are further evaluated. If hybrid reads overlap, they are collapsed to a single representative read. Then, the two arms of all reads are computationally annealed to each other to predict the longest RNA duplex contained within. In addition, to estimate the occurrence of RNA duplexes bound by the RBPs, the number of hybrid reads overlapping with each duplex is used to score each duplex.

The pipeline creates a number of plots that describe characteristics of the RBP-bound RNA duplexes including: (i) the transcript regions to which the arms align; (ii) the ratio of intragenic to intergenic hybrids (for those that occur in mature transcripts, rather than rRNA or tRNA); (iii) the loop lengths within 3' UTR and CDS regions of intragenic duplexes (Fig. 3). It also calculates the genomic location of the duplexes, which can be exported to view in a genome browser (Fig. 4). All of these results are contained in the final table created by the pipeline, which can be analyzed further by the user using scripts tailored to their biological question of interest.

### **Applications and advantages of hiCLIP**

hiCLIP is particularly useful for investigating the specificity of RBPs for different types of RNA structures, which is crucial to understand their roles in post-transcriptional regulation. hiCLIP can examine multiple aspects of RNA duplexes: their preferred locations within different regions of RNAs; their sequence and structural characteristics; their formation into intra- versus inter-molecular structures; and the distance between the two arms of the intra-molecular RNA duplexes<sup>8</sup>. Hybrid reads can also be used to study alternative RNA structural conformations in the cell, and how this may contribute to the regulation of RBP-binding<sup>2,8</sup>.

hiCLIP complements methods based on psoralen-crosslinking: hiCLIP focuses on determining RNA duplexes that are bound by specific RBPs, while the psoralen-crosslinking based methods allow identification of RNA duplexes globally (Fig. 1). One potential advantage of hiCLIP over psoralen-crosslinking based methods is the shorter crosslinking time. UV-C crosslinking used in hiCLIP takes less than a minute, whereas psoralen crosslinking (including pre-incubation) requires 40-60 min in PARIS<sup>10</sup> and LIGR-Seq<sup>11</sup>. In SPLASH<sup>9</sup>, psoralen crosslinking requires only 25 min, but here the pre-incubation includes a detergent that can affect cellular functions. Thus, hiCLIP may be more suitable to studying dynamic changes of RNA structure occurring over short time frames and with minimal perturbation of cells.

### **Limitations of hiCLIP**

Similar to CLIP and CLASH (reviewed recently<sup>22,28</sup>), the requirement for immunoprecipitation limits the use of hiCLIP to endogenous RBPs for which high-quality antibodies are available. While it is also possible to express tagged proteins or add a tag to the endogenous protein using genome editing techniques<sup>29</sup>, care must be taken to ensure that the added tag does not interfere with the functions of the RBP.

While these high-throughput methods have provided general insights into the characteristics of RNA structures, many challenges remain. The required high complexity of cDNA libraries and sequencing depth may restrict in-depth analysis of full RNA structures to the most highly expressed miRNAs, mRNAs and other non-coding RNAs, such as rRNAs, snRNAs and snoRNAs. In the case of hiCLIP, the relatively long procedure includes potentially inefficient steps, including UV-cross-linking and the generation of hybrid RNAs. The efficiency of UV-cross-linking varies depending on the RBP, but is estimated to be approximately 1-5% for most RBPs<sup>30</sup>, while the proportion of hybrid reads in the total sequence reads are 1-2%<sup>8</sup>. Therefore a large amount of starting material is currently required. This could be improved by incorporating knowledge obtained from recent efforts to optimize the iCLIP procedure<sup>31,32</sup>.

Since the high-throughput methods to study RNA structures are relatively new they have only been applied in limited biological contexts. CLASH and similar methods

have been applied to yeast spliceosome related proteins (Nop1, Nop56, Nop58, and Brr2)<sup>20</sup> and Ago2 proteins<sup>21,33,34</sup>, while hiCLIP has been applied to the human STAU1 protein<sup>8</sup>. The psoralen-crosslinking based methods were only reported recently and were used to examine general characteristics of RNA structures. Comparative analyses between these studies may yield further insights into the characteristics of RNA structures and specificity of dsRBPs. Applying hiCLIP or other high-throughput methods to a wide variety of RBPs in various biological conditions will provide more comprehensive insights into the specificity of particular RBPs for different types of RNA structure.

### **Experimental design**

**Adaptor preparation.** The use of two adaptors (Adaptor A and Adaptor B) is a key aspect of the hiCLIP protocol. For the generation of hybrid RNAs, Adaptor B, which is specially designed for hiCLIP, must be made from RNA as the enzyme used in this protocol (T4 RNA ligase) does not efficiently ligate RNA to DNA. The sequence of Adaptor A is identical that used in iCLIP. It is also RNA in order to achieve a similar ligation efficiency to Adaptor B. Of note, the adaptors described here can be used for the study of any RBPs and materials from both cultured cells and tissue without any modifications.

**In vivo UV crosslinking.** hiCLIP uses UV-C irradiation of the cell to crosslink RNA duplexes directly bound by RBPs. This covalent crosslinking allows later stringent purification to eliminate any associated RBPs and RNA that have not been crosslinked to the RBPs. Depending on the types of materials (e.g. cultured cells or tissues) or analyzed RBPs, the crosslinking step may require optimization to maximize the crosslinking efficiency. Generally, a relatively short exposure of UV-C light (from 100 mJ/cm<sup>2</sup> to 150 mJ/cm<sup>2</sup>) is used for cultured cells<sup>8,27,35</sup>, while longer and repeated exposure (e.g. 4 times at 100 mJ/cm<sup>2</sup> or successive irradiation at 400 mJ/cm<sup>2</sup> and 200 mJ/cm<sup>2</sup>) is used for tissues<sup>28,36</sup>. Past studies have shown that pre-incubating cells with photo-activatable nucleosides to incorporate them into RNAs, followed by UV-A irradiation can improve the crosslinking efficiency for certain RBPs<sup>37</sup>, compared to the standard UV-C crosslinking described here.

**Choice of cell lysis buffer.** The crosslinked protein-RNA complexes are extracted from the cell by lysing the cell with detergent-containing buffer. The cell lysis buffer should be chosen according to the cell fractions to be extracted, the stability of RNA duplexes, and the immunoprecipitation efficiency (see also the Immunoprecipitation section below). The use of different detergents allows the purification of RBPs and crosslinked RNA complexes in the different cell fractions<sup>38-40</sup> (manufacturers of antibodies such as abcam offers a useful guide to choose appropriate detergents in their website; e.g. <http://www.abcam.com/ps/pdf/protocols/WB-beginner.pdf>). In the **Procedure**, cells are lysed with buffer containing 1% (vol/vol) of non-ionic detergent, Triton-X 100, without mechanical disruption. This relatively mild treatment does not completely disrupt nuclei, allowing collection of the cytoplasmic lysate to study the RNA duplexes present in the cytoplasm. A more stringent buffer containing ionic detergents combined with mechanical disruption would be required to lyse the nuclei completely to study RNA duplexes from whole cells.

The buffer should not disrupt RNA duplexes before the ligation of the two arms. We therefore avoid the use of EDTA or denaturing reagents in the buffer at the cell lysis stage. Nevertheless, CLASH identified snoRNA-rRNA and miRNA-mRNA duplexes in spite of using a buffer containing 6M guanidine hydrochloride before the hybridization of the duplexes<sup>20,21</sup>, which is known to denature proteins due to its chaotropic properties. It remains to be seen if duplexes are fully resistant to the 6M guanidine hydrochloride, or if the duplexes were resistant because they were ligated by the endogenous ligase before this step<sup>33</sup>.

**Partial RNA digestion.** For proximity ligation, RNAs surrounding the duplexes need to be digested by the use of RNase or other methods. An appropriate level of RNA digestion is crucial for the success of hiCLIP experiments. The RNA length should be short enough that the sequence can identify the exact region of the duplex recognized by RBPs, and the full length of hybrid reads can be sequenced, but long enough that the RNA source can be uniquely identified. The experimental conditions are optimised with the aim to produce hybrid reads with arm lengths between 16 nts and 40 nts.

We use RNase I since it can cleave RNA at all four bases, and preferentially digests single-stranded RNA, thus leaving RNA duplexes intact. To identify the optimal digestion conditions, we test serial concentrations of RNase I (typically a range from 0.5 U ml<sup>-1</sup> to 20 U ml<sup>-1</sup>) and monitor the degree of digestion by autoradiography after nitrocellulose membrane transfer. In ideal digestion conditions, a diffuse signal of RNA-protein complexes should be visible above the molecular weight of the RBP of interest<sup>35</sup> (see Fig. 5).

**Immunoprecipitation.** Both the efficiency and specificity of immunoprecipitation are important for hiCLIP. The efficiency can be monitored using western blotting by comparing the signal of the input and immunoprecipitated fractions. The absence of other RBP or non-crosslinked RNA contaminations can be monitored by autoradiography (see Control experiments and Fig. 5). Optimization may be required for successful immunoprecipitation; relevant considerations have been discussed previously<sup>28</sup>. Since dsRBPs often do not have a high crosslinking efficiency, it is crucial for hiCLIP that all RBPs associating with the target RBP are removed. In the case of many RBPs, these associating RBPs are eliminated with the “harsh” immunoprecipitation condition of the standard iCLIP (e.g. washing with the buffer containing an ionic detergent such as SDS and high salt). However, in the case of some RBPs such as STAU1, two rounds of immunoprecipitation are required to eliminate the associated RBPs completely<sup>35</sup>. This includes the first standard round of immunoprecipitation, extraction of the RBP of interest from the beads by protein denaturation, followed by a second round of immunoprecipitation.

When employing two rounds of immunoprecipitation, the antibody used for the first round of immunoprecipitation is covalently coupled to Dynabeads Epoxy beads (used in the Dynabeads antibody coupling kit). This obviates contamination of the second round of immunoprecipitation by denatured antibodies from the first round. For the second round of immunoprecipitation, we attach the antibody to Dynabeads Protein G, because our initial optimization experiments indicated that they have less non-specific binding of protein and/or RNA than Dynabeads Epoxy beads (data not shown, YS).

**SDS-PAGE and nitrocellulose membrane transfer.** It is essential to purify only crosslinked RNAs. Short non-crosslinked RNAs and the free radioactive ATP used for RNA labeling are removed during the SDS-PAGE, since they migrate close to the front of the gel that is removed before the transfer. The remaining non-crosslinked RNAs are removed during the nitrocellulose membrane transfer, since nitrocellulose membrane binds to proteins but not nucleic acids<sup>41</sup>. A large amount of radioactivity is lost after nitrocellulose membrane transfer, indicating that this step is essential to eliminate non-crosslinked RNAs.

In the Procedure, we use radio-labelling of RNA for isotopic detection of protein-RNA complexes on the nitrocellulose membrane. Alternatively, protein-RNA complexes can be visualized by ligating a infrared-dye-conjugated adaptor, as has been recently developed for irCLIP<sup>32</sup>. Radiolabeling is highly sensitive and has a high signal-to-noise ratio, but experiments handling radioactive materials have to be done in a controlled environment with due care. So, if a sufficient amount of protein-RNA complexes is purified, non-isotopic labeling is less onerous.

**Reverse transcription and multiplexing of libraries.** The reverse transcription primer adds a unique molecular identifier and an experimental barcode sequence to the cDNA. The hiCLIP data analysis pipeline analyzes the unique molecular identifiers to distinguish PCR duplicates from unique cDNAs that have the same sequence, and thus refines the quantification in each experiment. This analysis is also useful to evaluate the depth of high-throughput DNA sequencing. If most cDNAs have been sequenced only once, it indicates that the library has a high cDNA complexity, and could be re-sequenced for greater sequence depth.

To reduce the cost of sequencing, the experimental barcode allows multiplexing several libraries and sequencing them in the same lane, since the samples can be deconvoluted from the sequence reads using this barcode. In the original hiCLIP paper, we multiplexed three libraries at this step<sup>8</sup>. Should the sequencing results indicate an inadequate depth of sequencing, the library can later be re-sequenced.

When multiplexing libraries, we designed each experimental barcode sequence in the reverse transcription primer to differ at least by two nucleotides so that each barcode could still be unambiguously identified despite sequencing errors. Another consideration is the timing of multiplexing. In principle, the libraries can be multiplexed immediately after reverse transcription since the experimental barcode has already been introduced into the cDNAs. However, at this step, precise quantification of the libraries is difficult. In light of this, we multiplex the libraries after PCR amplification and quantification of each library.

**Library quantification.** Libraries for high-throughput DNA sequencing can be quantified by qPCR (SYBR Green based / TaqMan based method), fluorescence measurement using fluorophores binding to DNA (e.g. Quant-it DNA assay kit by Thermo Fisher Scientific), or Agilent Bioanalyzer / Agilent TapeStation<sup>28</sup>. We recommend the TaqMan based qPCR method for hiCLIP library quantification, as it allows highly sensitive and precise quantification of the molar concentration regardless of the DNA length distribution in the library.

**DNA sequencing.** In general, sequencing of hiCLIP libraries requires longer sequence reads than CLIP and iCLIP, as it is important to sequence the full hybrid containing both duplex arms. The length of hybrid reads depends on the length of the original RNA fragments, which is determined by the degree of RNase digestion. We initially sequenced 196 nt long reads using MiSeq Gene & Small Genome Sequencer, but longer sequence reads could be produced without any problems.

**Control experiments.** Appropriate control experiments are necessary to monitor the quality of the hiCLIP experiment. To monitor the protein-RNA complex purification specificity, we recommend control samples that omit UV-C crosslinking and samples that lack expression of the target RBP (e.g. knockout or knockdown for endogenous RBPs, or omitting induction in the case of inducible expression). These controls are especially important for the preliminary experiments to optimize the purification conditions. For the full hiCLIP experiment, an additional control omitting the second round of ligation can be used to confirm that hybrid reads require both rounds of ligation.

**Preliminary experiments.** Before applying hiCLIP to new RBPs, we generally perform preliminary experiments to optimize the UV crosslinking, lysis buffer, RNase digestion, and immunoprecipitation steps by considering the factors listed above. In these experiments, we perform the hiCLIP procedure up to and including the autoradiograph steps, but skip the adaptor ligation and inter-molecular ligation steps (Steps 23 – 48 and 57 – 73). The autoradiograph is used to evaluate each condition (Fig. 5).

**Bioinformatic analysis.** The hiCLIP data analysis R package “hiclipr” presented here is a more user-friendly version of the pipeline described before<sup>8</sup>. It is set up to analyze data from human samples. We have updated the annotations used to Ensembl GRCh37, release 75, which is currently the static archived GRCh37 release. The internal datasets would need to be modified for other species or assemblies. We have also expanded the hybrid read selection method. Only hybrid reads separated by a single linker were considered in the original study. However, we found that a small proportion of reads contain runs of more than one interrupting linkers (~8% of all reads containing at least one linker) presumably due to the incomplete 3' blocking of the linker adaptor. These are now included as hybrid reads in the new package.

In the Procedure, we describe the basic analysis to detect and annotate RBP-bound RNA duplexes. More advanced analyses would require further scripts tailored to the biological questions of interest.

## **MATERIALS**

### **REAGENTS**

- Appropriate cell line or tissue. In the Procedure, we use FLAG-STAU1-Flp-In 293 T-REx cells: The cell line is produced from Flp-In 293 T-REx cells (Thermo Fisher Scientific, cat. no. R780-07) by transfecting pcDNA5/FRT/TO plasmid with FLAG-STAU1 (GenBank accession number of the STAU1 coding sequence: NM\_017454.2) by following the manufacturer's instructions (details are described in ref. 8).

CRITICAL. We regularly check the absence of mycoplasma contamination using a commercially available kit such as LookOut

Mycoplasma PCR Detection Kit (Sigma-Aldrich, cat. no. MP0035) or MycoAlert Mycoplasma Detection Kit (Lonza, cat. no. LT07-118).

CAUTION The authenticity of the cell line and absence of mycoplasma infection should be regularly tested.

- DMEM (e.g. Thermo Fisher Scientific, cat. no. 10569010)
- FBS (Thermo Fisher Scientific, cat. no. 10500064)
- PBS (e.g. Thermo Fisher Scientific, cat. no. 10010056)
- Nuclease-free water (e.g. Thermo Fisher Scientific, cat. no. 10977049)
- Doxycycline (Sigma-Aldrich, cat. no. D9891)  
[Caution] Doxycycline is harmful if swallowed, and causes skin irritation.
- Blasticidin S HCl (10 mg ml<sup>-1</sup>; Thermo Fisher Scientific, cat. no. A1113903)  
[Caution] Blasticidin S HCl is fatal if swallowed, and harmful in contact with skin
- Hygromycin B (100 mg ml<sup>-1</sup>; InvivoGen; standard Hygromycin B has been discontinued, and replaced by Hygromycin B Gold, cat. no. ant-hg-1)  
[Caution] Hygromycin B is fatal if swallowed or in contact with skin
- 5' DNA adenylation kit (NEB, cat no. E2610). Supplied with Mth RNA ligase (50 pmol μl<sup>-1</sup>), 5' DNA adenylation reaction buffer (10×), and ATP (1 mM)
- Acid-phenol:chloroform (pH 4.5; Thermo Fisher Scientific, cat. no. AM9722)  
[Caution] Phenol is corrosive. Phenol must be used in a fume hood with gloves and other necessary protections
- Linear acrylamide (5 mg ml<sup>-1</sup>; Thermo Fisher Scientific, cat. no. AM9520)
- Sodium acetate (3 M, pH 5.5; Thermo Fisher Scientific, cat. no. AM9740)
- 15% Novex TBE-Urea gels, 10 well (Thermo Fisher Scientific, cat. no. EC6885BOX)
- Ethanol (Sigma-Aldrich, cat. no. 24102)  
[Caution] Ethanol is flammable.
- SYBR green II (Thermo Fisher Scientific, cat. no. S7564)
- SYBR safe (Thermo Fisher Scientific, cat. no. S33102)
- Dynabeads antibody coupling kit (Thermo Fisher Scientific, cat. no. 14311D)
- Tris-HCl (1 M, pH 6.5 / 7.4 / 7.8; dissolved e.g. Fisher Scientific, cat. no. BP152-1 in nuclease-free water and adjusted the pH by adding e.g. Sigma-Aldrich, cat. no. 31087)
- NaCl (5M; dissolved Sigma-Aldrich, cat. no. S7653 in nuclease-free water)

- EDTA (100 mM; diluted from Thermo Fisher Scientific, cat. no. 15575020)
- MgCl<sub>2</sub> (1 M; e.g. Thermo Fisher Scientific, cat. no. AM9530G)
- Protease Inhibitor Cocktail Set III (Calbiochem, cat. no. 539134)
- Igepal CA-630 (Sigma-Aldrich, cat. no. I8896)
- SDS (Sigma-Aldrich, cat. no. L3771)
- Sodium deoxycholate (Sigma-Aldrich, cat. no. D6750)
- Triton X-100 (e.g. Sigma-Aldrich, cat. no. T8787)
- DTT (e.g. Sigma-Aldrich, cat. no. DTT-RO ROCHE)
- RNase I (100 U μl<sup>-1</sup>; Thermo Fisher Scientific, cat. no. AM2294)
- RNasin Ribonuclease Inhibitor (20 – 40 U μl<sup>-1</sup>; Promega, cat. no. N2115)
- RNasin Plus RNase inhibitor (40 U μl<sup>-1</sup>; Promega, cat. no. N2615)
- SUPERaseIn RNase inhibitor (20 U μL<sup>-1</sup>; Thermo Fisher Scientific, cat. no. AM2694)
- TURBO DNase (2 U μl<sup>-1</sup>; Thermo Fisher Scientific, cat. no. AM2238)
- ANTI-RNase (15-30 U μl<sup>-1</sup>; Thermo Fisher Scientific, cat. no. AM2690)
- T4 PNK (10 U μl<sup>-1</sup>; NEB, cat. no. M0201). Supplied with PNK buffer (10×)
- T4 RNA ligase 2, truncated K227Q (200 U μl<sup>-1</sup>; NEB, cat. no. M0351)
- PEG400 (Sigma-Aldrich, cat. no. 81170)
- T4 RNA Ligase 1 (ssRNA Ligase) (NEB, cat. no. M0204). Supplied with T4 RNA ligase reaction 1 buffer (10×; B0216S)

[Critical] Must use the ligation buffer that does not contain ATP

- Urea (Sigma-Aldrich, cat. no. U5378)
- Dynabeads Protein G (Thermo Fisher Scientific, cat. no. 10004D)
- Anti-FLAG M2 antibody (1 mg ml<sup>-1</sup>; Sigma-Aldrich, cat. no. F1804)
- <sup>32</sup>P-γ-ATP (10 mCi ml<sup>-1</sup>; PerkinElmer, cat. no. NEG002A250UC)

[Caution] <sup>32</sup>P-γ-ATP is radioactive. Radioactive materials should be used by following the governmental and local regulations, and be handled with appropriate protections and care.

- NuPAGE LDS sample buffer (4×; Thermo Fisher Scientific, cat. no. NP0007)
- NuPAGE sample reducing agent (Thermo Fisher Scientific, cat. no. NP0004)
- 4-12% NuPAGE Bis-Tris gel, 1.0 mm, 10-well (Thermo Fisher Scientific, cat. no. NP0321BOX)

[Critical] Choose 1.0 mm and 10-well gels since thicker gels are less efficient when transferring protein-RNA complexes to nitrocellulose

membranes and wider wells give better separation of protein-RNA complexes during electrophoresis

- MOPS running buffer (20×; Thermo Fisher Scientific, cat. no. NP0001)
- Methanol (e.g. Thermo Fisher Scientific, cat. no. A412-1)
- NuPAGE antioxidant (Thermo Fisher Scientific, cat. no. NP0005)
- PageRuler Plus prestained protein ladder (Thermo Fisher Scientific, cat. no. SM1811)
- Protran BA85 nitrocellulose membrane (Whatman, cat. no. BA85)
- NuPAGE Transfer Buffer (20×; Thermo Fisher Scientific, cat. no. NP00061)
- Carestream Kodak BioMax XAR film (Sigma-Aldrich, cat. no. Z358487)
- Proteinase K (Roche cat. no. 03115828001)
- Superscript III RT (200 U  $\mu\text{l}^{-1}$ ; Thermo Fisher Scientific, cat. no. 18080-044). Supplied with First-Strand Buffer (5×) and DTT (0.1 M)
- dNTP mix (10 mM; e.g. Thermo Fisher Scientific, cat. no. 18427-013)
- NaOH (1 M; Diluted from Sigma-Aldrich, cat. no. 221465)

[Caution] NaOH causes severe burn in contact with skin.

NaOH must be used in a fume hood with gloves and other necessary protections

- HEPES-NaOH (1 M; Thermo Fisher Scientific, cat. no. BP299-500)
- Novex TBE-Urea sample buffer (2×; Thermo Fisher Scientific, cat. no. LC6876)
- 6% Novex TBE-Urea gels, 10 well (Thermo Fisher Scientific, cat. no. EC6865BOX)
- CirLigase II ssDNA ligase (100 U  $\mu\text{l}^{-1}$ ; Epicentre, cat. no. CL9025K). Supplied with CirLigase II reaction buffer (10×) and  $\text{MnCl}_2$  (50 mM)
- FastDigest BamHI (Thermo Fisher Scientific, cat. no. FD0055). Supplied with FastDigest buffer (10×)
- Accuprime Supermix I enzyme (Thermo Fisher Scientific, cat. no. 12342010)
- 6% Novex TBE gels (e.g. Thermo Fisher Scientific, cat. no. EC6265BOX)
- Sample loading dye (5×; Bio-Rad, cat. no. 1660401)
- AMPure XP (Beckman Coulter, cat. no. A63880)
- Platinum Quantitative PCR SuperMix-UDG w/ROX (Thermo Fisher Scientific, cat. no. 11743500)

- Oligonucleotide (See Table 1; custom order from Integrated DNA Technologies except DLP oligo and qPCR primer 1 and 2, which are custom order from Sigma-Aldrich)

## **EQUIPMENT**

- 15 cm cell culture dishes (e.g. Sigma-Aldrich, cat. no. CLS430599)
- Cell lifters (e.g. Corning, cat. no. 3008)
- 1.5 ml microfuge tubes (Nonstick, RNase-free; Thermo Fisher Scientific, cat. no. AM12450)
- 50 ml conical centrifuge tubes (e.g. Thermo Fisher Scientific, cat. no. 14-432-22)
- 0.2 ml PCR tubes (e.g. Corning, cat. no. AXY-PCR-02-C)
- 2 ml Phase Lock Gel Heavy tube (VWR, 713-2536)
- 1 cm Glass filters (Whatman, cat. no. 1823010)
- Costar SpinX column (Corning, cat. no. 8161)
- 96-well plate for qPCR (e.g. Thermo Fisher Scientific, cat. no. 4346906)
- Optical adhesive film (e.g. Thermo Fisher Scientific, cat. no. 4311971)
- Rotating wheels for 50 ml conical tubes and 1.5 ml microfuge (e.g. Stuart, cat. no. SB3)
- Squisher-Single (To crush gel pieces; Zymo Research, cat. no. H1001-50)
- 0.45  $\mu\text{m}$  syringe filter (e.g. Millipore, cat. no. SLHV033RS)
- 50 ml disposable syringe (e.g. VWR, cat. no. 97048-662)
- 21G needle (e.g. Greiner Bio-One, cat. no. 450072)
- Refrigerated centrifuge for 50 ml conical centrifuge tubes (e.g. Eppendorf, cat. no. 5810R)
- Tabletop refrigerated centrifuge for 1.5 ml microfuge tubes (Maximum RCF should be  $\geq 21,040g$ ; e.g. Eppendorf, cat. no. 5424R)
- Thermal cycler (e.g. Bio-Rad, cat. no. 1851148)
- qPCR machine (compatible with the following method: detector: FAM-TAMRA and passive reference: ROX; e.g. Thermo Fisher Scientific, cat. no. 4485697)
- NanoDrop (e.g. Thermo Fisher Scientific, cat. no. ND-2000)
- Thermomixer Comfort (Eppendorf, cat. no. 5355 000.011) or Thermomixer Compact (Eppendorf, cat. no. 5350 000.013)
- UV transilluminator (e.g. Bio-Rad, cat. no. 1708195)
- Vortex (e.g. Scientific Industries, cat. no. SI-0236)

- Cling Film (e.g. Fisher Scientific, cat. no. 12872233)
- Scintillation counter (To monitor radioactivity; e.g. Fisher Scientific, Mini 900 Scintillation Monitors)
- Fluor-Coated TLC Plate (10 × 10 cm) (Thermo Fisher Scientific, cat. no. AM10110)
- Handheld 254 nm UV light (for UV shadowing; e.g. UVP, cat. no. 95-0007-05)
- UV crosslinker (e.g. Stratagene, Stratalinker 2400)
- Tray that can be put in the UV crosslinker (Dimensions: 409mm × 300mm × 42 mm; to place cell culture dishes on ice during UV-C crosslinking; e.g. Kartell, 5714)
- XCell SureLock Mini-Cell with XCell II blot module kit (For electrophoresis and membrane transfer. Power supply unit is also needed; Thermo Fisher Scientific, cat. no. EI0002)
- Magnetic stand (e.g. DynaMag-2 Magnet, Thermo Fisher Scientific, cat. no. 12321D)
- MiSeq Gene & Small Genome Sequencer (illumina)
- MiSeq Reagent Kit that is compatible to sequence more than 200 bases in a single read (e.g. illumina, cat. no. MS-102-2002)
- 64-bit computer running either Linux or Mac OS X (tested on 10.11); 16 GB RAM (> 32 GB preferred); 4 cores (> 8 preferred)
- Installation of the “hiclipr” R package (from <https://github.com/luslab/hiclipr>) and dependencies as detailed in Box 1.

## REAGENT SETUP

**TE buffer.** 100 mM Tris-HCl (pH 7.4), 10 mM EDTA in nuclease-free water. To prepare 1 l of TE buffer, mix 100 ml of 1 M Tris-HCl (pH 7.4), 100 ml of 100 mM EDTA, and 800 ml of nuclease-free water. This buffer can be stored at room temperature (20–25°C) for up to several months unless otherwise specified.

**CLIP lysis buffer.** 50 mM Tris-HCl (pH 7.4), 100 mM NaCl, 1% (vol/vol) Igepal CA-630, 0.1% (wt/vol) SDS, 0.5% (wt/vol) sodium deoxycholate in nuclease-free water. To prepare 1 l of CLIP lysis buffer, mix 50 ml of 1 M Tris-HCl (pH 7.4), 20 ml of 5M NaCl, 10 ml of Igepal CA-630, 10 ml of 10% (wt/vol) SDS, and 50 ml of 10% (wt/vol) sodium deoxycholate, and add nuclease-free water so that the final volume

will be 1 l. This buffer is filtered using a 0.22  $\mu\text{m}$  vacuum filter unit, and can be stored at 4°C for up to several months.

**High-salt wash.** High-salt wash contains 50 mM Tris-HCl (pH 7.4), 1 M NaCl, 1 mM EDTA, 1% (vol/vol) Igepal CA-630, 0.1% (wt/vol) SDS, and 0.5% (wt/vol) sodium deoxycholate in nuclease-free water. In order to prepare 1 l of high-salt wash, mix 50 ml of 1 M Tris-HCl (pH 7.4), 200 ml of 5M NaCl, 10 ml of 100 mM EDTA, 10 ml of Igepal CA-630, 10 ml of 10% (wt/vol) SDS, and 50 ml of 10% (wt/vol) sodium deoxycholate, and add nuclease-free water so that the final volume will be 1 l. This buffer is filtrated using a 0.22  $\mu\text{m}$  vacuum filter unit, and can be stored at 4°C for up to several months.

**PG cell lysis buffer.** PG cell lysis buffer contains 20 m Tris-HCL (pH 7.4), 140 mM NaCl, 5 mM  $\text{MgCl}_2$ , and 1% (vol/vol) Triton-X 100 in nuclease-free water. In order to prepare 1 l of PG cell lysis buffer, mix 20 ml of 1 M Tris-HCl (pH 7.4), 28 ml of 5M NaCl, 5 ml of 1 M  $\text{MgCl}_2$ , and 10 ml of Triton-X 100, and add nuclease-free water so that the final volume will be 1 l. This buffer is filtrated using a 0.22  $\mu\text{m}$  vacuum filter unit, and can be stored at 4°C for up to several months.

**DTT solution (1M).** Dissolve 1.54 g of DTT in 10 ml of nuclease-free water. Aliquot 100  $\mu\text{l}$  or 1 ml per tube. This buffer can be stored at -20 °C for up to several months, but avoid repeated freeze-thaw cycles.

**PG cell lysis buffer complete.** Mix PG cell lysis buffer with 1 mM DTT, 1/1,000 (vol/vol) anti-RNase, 4/1,000 (vol/vol) Protease Inhibitor Cocktail Set III, and 1/10,000 (vol/vol) TURBO DNase. This buffer should be freshly prepared before use, and kept on ice.

**Urea cracking buffer (1.25 $\times$ ).** Urea cracking buffer (1.25 $\times$ ) contains 66.6 mM Tris-HCl, 8M urea, 1.33% SDS, and 31.25% (vol/vol) PBS in nuclease-free water. In order to prepare 10 ml of urea cracking buffer (1.25 $\times$ ), mix 666  $\mu\text{L}$  of 1 M Tris-HCl (pH 7.4), 1.33 ml of 10% (wt/vol) SDS, 3.125 ml of PBS, and 4.8 g of urea, and add nuclease-free water so that the final volume will be 10 ml. Aliquot 1 ml per tube. This

buffer can be stored at -20 °C for up to several months, but avoid repeated freeze-thaw cycles.

**T-20 IP buffer.** T-20 IP buffer contains 50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 0.1 mM EDTA, and 0.5% (vol/vol) Tween-20 in nuclease-free water. In order to prepare 1 l of T-20 IP buffer, mix 50 ml of 1 M Tris-HCl (pH 7.4), 30 ml of 5M NaCl, 1 ml of 100 mM EDTA, and 5 ml of Tween-20, and add nuclease-free water so that the final volume will be 1 l. This buffer is filtrated using a 0.22  $\mu$ m vacuum filter unit, and can be stored at 4°C for up to several months.

**PNK buffer.** PNK buffer contains 20 mM Tris-HCl (pH 7.4), 10 mM MgCl<sub>2</sub>, 0.2% (vol/vol) Tween-20 in nuclease-free water. In order to prepare 1 l of PNK buffer, mix 20 ml of 1 M Tris-HCl (pH 7.4), 10 ml of 1M MgCl<sub>2</sub>, and 2 ml of Tween-20, and add nuclease-free water so that the final volume will be 1 l. This buffer is filtrated using a 0.22  $\mu$ m vacuum filter unit, and can be stored at 4°C for up to several months.

**PNK pH 6.5 buffer (5×).** PNK pH 6.5 buffer (5×) contains 350 mM Tris-HCl (pH 6.5), 50 mM MgCl<sub>2</sub>, and 25 mM DTT in nuclease-free water. In order to prepare 1 ml of PNK pH 6.5 buffer (5×), mix 350  $\mu$ l of 1M Tris-HCl (pH 6.5), 50  $\mu$ l of 1M MgCl<sub>2</sub>, 25  $\mu$ l of 1M DTT, and 575  $\mu$ l of nuclease-free water. Aliquot 100  $\mu$ l per tube. This buffer can be stored at -20 °C for up to 1 year, but avoid repeated freeze-thaw cycles.

**Ligation buffer (4×).** Ligation buffer (4×) contains 200 mM Tris-HCl (pH 7.8), 40 mM MgCl<sub>2</sub>, and 40 mM DTT in nuclease-free water. In order to prepare 1 ml of ligation buffer (4×), mix 200  $\mu$ l of 1 M Tris-HCl (pH 7.8), 40  $\mu$ l of 1 M MgCl<sub>2</sub>, 40  $\mu$ l of 1 M DTT, and 720  $\mu$ l of nuclease-free water. Aliquot 100  $\mu$ l per tube. This buffer can be stored at -20 °C for up to 1 year, but avoid repeated freeze-thaw cycles.

**PK buffer.** PK buffer contains 100 mM Tris-HCl (pH 7.4), 50 mM NaCl, and 10 mM EDTA in nuclease-free water. In order to prepare 1 l of PK buffer, mix 100 ml of 1 M Tris-HCl (pH 7.4), 50 ml of NaCl, 100 ml of 100 mM EDTA, and 750 ml of nuclease-free water. This buffer is filtrated using a 0.22  $\mu$ m vacuum filter unit, and can be stored at 4°C for up to several months.

**PK buffer with 7M urea.** PK buffer with 7M urea contains 100 mM Tris-HCl (pH 7.4), 50 mM NaCl, 10 mM EDTA, and 7M urea in nuclease-free water. In order to prepare 10 ml of PK buffer with 7M urea, mix 1 ml of 1 M Tris-HCl (pH 7.4), 100  $\mu$ l of NaCl, 1 ml of 100 mM EDTA, and 4.2 g of urea, and add nuclease-free water so that the final volume will be 10 ml. Aliquot 1 ml per tube. This buffer can be stored at -20 °C for up to several months, but avoid repeated freeze-thaw cycles.

Box 1: Downloading and installing software; Timing variable; ~1 hour

The hiCLIP analysis pipeline is available as an R (3.2.4) package “hicliplr” designed to be run in a UNIX environment (e.g. Linux or Mac OS X).

1. Download the “hicliplr” package (hicliplr\_0.3.tar.gz) from

<https://github.com/luslab/hicliplr>.

2. Within R, ensure the package dependencies are installed. Each command is prefixed here with the R command prompt “>” which does not need to be entered.

```
> install.packages(c("ggplot2", "rentrez", "data.table"))
> source("https://bioconductor.org/biocLite.R")
> biocLite(c("BiocGenerics", "Biostrings", "BSgenome.Hsapiens.UCSC.hg19",
"GenomicAlignments", "GenomicRanges", "IRanges", "rtracklayer",
"ShortRead"))
```

3. Install the “hicliplr” package

```
> install.packages("path/to/hicliplr.tar.gz", repos = NULL, type = "source")
```

4. The hiCLIP analysis pipeline and “hicliplr” also requires the following programs to be installed according to the developer’s instructions:

Bowtie (1.1.2)<sup>42</sup> from <http://bowtie-bio.sourceforge.net/index.shtml>

Samtools (1.3)<sup>43</sup> from <https://github.com/samtools/samtools>

Sambamba (0.6.3)<sup>44</sup> from <https://github.com/lomereiter/sambamba>

Cutadapt (1.9.1)<sup>45</sup> from <http://cutadapt.readthedocs.io/en/stable/>

RNAhybrid (2.1.2)<sup>46</sup> from <http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/>

CRITICAL. Ensure the executable files are named: “bowtie”, “samtools”,

“sambamba”, “cutadapt” and “RNAhybrid” respectively.

5. Add these programs to the “PATH” environment variable using the appropriate method for the operating system.

END OF BOX 1

## PROCEDURE

[Preparation of adaptors; Timing 2h (plus overnight and ~5h the following day)]

1. Set up two adenylation reaction mixes as below: one for RNA adaptor A and one for RNA adaptor B.

Component	Amount ( $\mu$ l)	Final concentration
RNA adaptor A or RNA adaptor B (100 $\mu$ M)	10	5 $\mu$ M
5' DNA adenylation reaction buffer (10 $\times$ )	20	1 $\times$
ATP (1 mM)	20	0.1 mM
Nuclease-free water	130	
Mth RNA ligase (50 pmol $\mu$ l <sup>-1</sup> )	20	5 $\mu$ M
Total	200	

[Critical Step] We purchase RNA adaptors with a 5' phosphate, and adenylyate the 5' end with the protocol described from Step1 to Step 22. We usually adenylyate 1 nmol of RNA adaptor A or RNA adaptor B in a single experiment.

2. Split the reaction mix into ten 200  $\mu$ l PCR tubes (20  $\mu$ l per tube), and incubate for 1 h at 65°C in a thermal cycler.
3. For each adaptor, pool samples in a 1.5 ml microfuge tube and add 200  $\mu$ l of TE buffer (the final volume is 400  $\mu$ l).
4. Add 400  $\mu$ l of acid-phenol:chloroform (pH 4.5) to each sample, and transfer to a 2 ml Phase Lock Gel Heavy tube.
5. Invert the samples 10 times, and then centrifuge at room temperature for 5 min at 15,700g to separate the aqueous and organic phases.

6. For each sample, transfer the aqueous phase to a new 1.5 ml microfuge tube, and add 1  $\mu$ l of linear acrylamide and 40  $\mu$ l of sodium acetate (3M, pH 5.5) to it.
7. Add 1 ml of ethanol to each sample, and vortex.
8. Incubate the samples overnight at -20 °C to precipitate RNA.
9. Next day, centrifuge the samples at 4 °C for 15 min at > 20,000g to pellet RNA, and then discard the supernatant.
10. Add 1 ml of 80% (vol/vol) ethanol to wash the RNA pellet, vortex, and then centrifuge at 4 °C for > 5 min at > 20,000g to pellet RNA again.
11. Discard the supernatant and air-dry the RNA pellet for 3 min.
12. Resuspend the RNA pellet in 6  $\mu$ l of nuclease-free water.  
[Pause Point] The purified RNA adaptors can be stored for at least a month at -80 °C.
13. Add 6  $\mu$ l of Novex TBE-Urea sample buffer (2 $\times$ ) to each sample. Also as a control, prepare two samples containing 100 ng of non-adenylated RNA Adaptor in 12  $\mu$ L of Novex TBE-Urea sample buffer (one for Adaptor A and one for Adaptor B).
14. Incubate the samples for 5 min at 80 °C to denature RNA, and then place on ice.
15. Load the samples and control samples on a 15% Novex TBE-Urea Gels (10 well), and perform electrophoresis for 70 min at 180 V using a XCell SureLock Mini-Cell by following the manufacturer's instruction. As a control, non-adenylated adaptors can be loaded together with the samples to determine the expected migration in the gel.  
[Critical step] In order to facilitate the excision of bands, load samples in alternate wells.

16. Visualize the RNA adaptor in the gel using UV shadowing, and purify the adaptor from the gel. In detail, cover a Fluor-Coated TLC Plate with saran wrap and place the gel on top of it. Then irradiate the gel using a handheld 254 nm UV light. The gel region containing RNA will be darker<sup>47</sup>. Excise the gel regions corresponding to the adenylated adaptor, and put the gel pieces into a new 1.5 ml microfuge tube.

[Critical step] In order to minimize the damage to RNA adaptors, UV light exposure time should be as short as possible.

[Critical step] Only the band corresponding to the adenylated adaptors should be excised. The band from adenylated adaptor migrates 1nt higher than the non-adenylated adaptor. Using the conditions described here, the adenylation efficiency should be > 90% (Fig. 6).

? Troubleshooting

17. Add 400  $\mu$ l of TE buffer and 1  $\mu$ l of RNasin (20-40 U  $\mu$ l<sup>-1</sup>) to the gel pieces, and crush them using a Squisher-Single.
18. Incubate the samples at 37°C for 1 h at 1,100 rpm shaking in a Thermomixer. Place them on dry ice for 2 min, and then incubate again at 37°C for 1 h at 1,100 rpm shaking in a Thermomixer.
19. For each sample, place two 1 cm glass filters in a Costar SpinX column, and transfer the liquid portion of the sample into the column.
20. Centrifuge the columns at 15,700g for 1 min at room temperature to remove any remaining gel pieces, and transfer the filtered sample to a new 1.5 ml microfuge tube.
21. Perform acid phenol-chloroform-isoamyl alcohol (PCI) extraction and ethanol precipitation as described in Step 4-11.

22. Resuspend each RNA pellet in 25  $\mu\text{l}$  of nuclease-free water, and quantify the wt/vol concentration of the sample using NanoDrop or similar. Add nuclease-free water to each sample to a final concentration of 10  $\mu\text{M}$ .

[Critical Step] In hiCLIP, it is important to have equimolar amounts of RNA adaptor A and B in the first round of ligation (Step 52-53), since unequal concentrations may result in a smaller proportion of hybrid reads in the final library. Thus, the molar concentration of the adaptors must be carefully adjusted using the wt/vol concentration measured by NanoDrop (or a similar device) and the expected molecular weight of the adaptors (around 6,500 and 7,100 for RNA Adaptor A and B respectively).

[Pause Point] The purified RNA adaptors can be stored for at least a month at -80 °C.

[Preparation of cells and UV crosslinking; Timing ~2 days]

23. Culture FLAG-STAU1-Flp-In 293 T-REx cells on 15 cm dishes in DMEM containing 10% (vol/vol) FBS, blasticidin S HCl (3  $\mu\text{g } \mu\text{l}^{-1}$ ), and hygromycin B (200  $\mu\text{g } \mu\text{l}^{-1}$ ) to 60-70% confluent. Add doxycycline (final concentration 250 ng  $\text{ml}^{-1}$ ) to the medium in order to induce the FLAG-STAU1 expression. We generally use cells from five 15 cm dishes for each hiCLIP or control sample. For example, two hiCLIP samples and two negative control samples (without second ligation and without FLAG tagged STAU1 expression) would require twenty 15 cm dishes.
24. 16 h after induction, transfer the dishes to ice and remove the medium from them.
25. For each dish, wash the cells with 10 ml of ice-cold PBS, and then add 10 ml of ice-cold PBS.
26. Place the dishes on an ice-filled tray and remove the lids. Put them in a UV crosslinker (Stratlinker 2400). We generally crosslink cells on 3 dishes at once.
27. Irradiate the cells once with UV light (150  $\text{mJ}/\text{cm}^2$  at 254 nm) to crosslink proteins and RNAs.

28. Scrape the cells from the dishes using cell lifters, and pool the cell suspension in a 50 ml conical tube (The cells from 5 dishes are pooled in a single tube).
29. Centrifuge the tubes at 514g for 5 min at 4 °C to pellet the cells, and discard the supernatant.
30. Snap-freeze the cell pellets by placing the tubes on dry ice, and store at -80 °C.  
[Pause point] Frozen cells can be stored at -80 °C for at least a month.

[Preparation of antibody-coupled beads; Timing 4 h]

31. Covalently couple 300  $\mu$ g of anti-FLAG M2 antibody with 10 mg of Dynabeads using Dynabeads antibody coupling kit by following the manufacturer's instruction. After the coupling, we resuspend 10 mg of beads in 1 ml of the buffer supplied with the kit.
32. Transfer 100  $\mu$ l per sample of antibody-coupled beads to a new 1.5 ml microfuge tube, and discard the supernatant using a magnetic stand. We prepare the beads for up to 4 samples in one 1.5 ml microfuge tube.
33. Wash the antibody-coupled beads once with 1 ml of CLIP lysis buffer, and twice with 1 ml of high-salt wash. After each wash, discard the supernatant using a magnetic stand. After the final wash, resuspend the beads in 100  $\mu$ l per sample of PG cell lysis buffer, and then store the beads at 4 °C.  
[Pause point] Antibody-coupled beads can be stored at 4 °C for a week.

[Cell lysis and partial RNA digestion; Timing 2 h]

34. Add 30 ml of ice-cold PG cell lysis buffer complete to the cell pellet in a 50 ml conical tube. For a single hiCLIP sample, we generally use the cell pellets in a 50 ml conical tube (i.e. cells from five 15 cm dishes).  
[Critical step] After cell lysis, care should be taken to avoid RNA degradation. The equipment should be kept RNase-free. Keep samples on ice, and use ice-cold buffer unless otherwise specified.

35. For each lysate from a 50 ml conical tube, pass it through a syringe needle twice using a 21G needle and 50 ml disposable syringe, and collect the sample in a new 50 ml conical tube.

36. Centrifuge the tubes at 14,000g for 5 min at 4 °C, and transfer the supernatant to a new 50 ml conical tube. Repeat this clarification step once more.

37. Add 20 ml of ice-cold PG cell lysis buffer to the lysate to bring the final volume of each sample to 50 ml, filter the sample using a 0.45  $\mu\text{m}$  syringe filter and a 50 ml disposable syringe, and collect in a new 50 ml conical tube.

[Critical Step] The lysate is diluted to facilitate filtration. The filter should be replaced if there is clogging. We generally use 2 filters to process 50 ml of sample.

? Troubleshooting

38. Add 2  $\mu\text{l}$  or 1  $\mu\text{l}$  of RNase I (original concentration 100,000 U  $\text{ml}^{-1}$ ) for a final concentration of 4 U  $\text{ml}^{-1}$  (high-RNase condition) or 2 U  $\text{ml}^{-1}$  (low-RNase condition), respectively to each sample.

[Critical step] Digest RNA to a suitable length for sequencing. We generally perform hiCLIP experiments with multiple RNase I concentrations as described here. The optimal condition should be tested for new experimental setups (see “partial RNA digestion” section of **Experimental design**).

39. Incubate the samples for 5 min at 37 °C in water bath agitating by hand. After incubation, immediately transfer the samples to ice, and add 20  $\mu\text{l}$  of SUPERaseIn RNase inhibitor (10 U  $\mu\text{l}^{-1}$ ). After mixing, leave the sample on ice for 5 min.

[Critical step] SUPERaseIn RNase inhibitor is used to inhibit RNase I activity.

[First round of immunoprecipitation; Timing 3-4 h]

40. For each hiCLIP sample, transfer 100  $\mu\text{l}$  of the antibody-coupled beads to a new 1.5ml microfuge tube, discard the supernatant using a magnetic stand, and

resuspend the beads in 1 ml of PG cell lysis buffer.

[Critical step] The amount of beads used for immunoprecipitation may need to be adjusted depending on the antibody and the target protein.

41. Add the antibody-coupled beads in 1 ml of PG cell lysis buffer to each cell lysate from step 39.
42. Incubate the beads/lysate mix for 2-3 h at 4 °C on a rotating wheel.
43. Centrifuge the samples at 800g for 5 min at 4 °C to pellet the beads, and discard the supernatant.
44. Add 500  $\mu$ l of PG cell lysis buffer to resuspend the beads, and transfer them to a new 1.5 ml microfuge tube.
45. Discard the supernatant using a magnetic stand.
46. Wash the beads twice with 1 ml of PG cell lysis buffer. After each wash, discard the supernatant using a magnetic stand. For the second wash, incubate the mixture at 4 °C for 1 min on a rotating wheel.
47. Add 1 ml of PG cell lysis buffer with 2  $\mu$ l of TURBO DNase (2 U  $\mu$ l<sup>-1</sup>), 1  $\mu$ l of RNasin Plus RNase inhibitor (40 U  $\mu$ l<sup>-1</sup>) and, 1  $\mu$ l of SUPERaseIn RNase inhibitor (20 U  $\mu$ l<sup>-1</sup>) to each sample. Incubate the samples at 37 °C for 3 min at 1,100 rpm, shaking in a Thermomixer.  
[Critical step] This step ensures that any residual DNA is removed.
48. Wash the beads twice with 1 ml of PG cell lysis buffer, and three times with PNK buffer. After each wash, discard the supernatant using a magnetic stand. For the second wash with PG cell lysis buffer, incubate the mixture at 4 °C for 1 min with rotating.

[3' end dephosphorylation; Timing 1 h]

49. For each sample, resuspend the beads in 20  $\mu$ l of PNK reaction mix as tabulated below, and incubate at 37 °C for 20 min at 1,100 rpm shaking in a Thermomixer:

Component	Amount ( $\mu$ l)	Final concentration
PNK pH 6.5 buffer (5 $\times$ )	4	1 $\times$
Nuclease-free water	14.5	
RNasin Plus RNase inhibitor (40 U $\mu$ l <sup>-1</sup> )	0.5	1 U $\mu$ L <sup>-1</sup>
SUPERaseIn RNase inhibitor (20 U $\mu$ l <sup>-1</sup> )	0.5	0.5 U $\mu$ L <sup>-1</sup>
T4 PNK (10 U $\mu$ l <sup>-1</sup> )	0.5	0.25 U $\mu$ L <sup>-1</sup>
Total	20	

50. Wash the beads once with 1 ml of PNK buffer, twice with 1 ml of PG cell lysis buffer, and three times with 1 ml of PNK buffer. For the washes with PG cell lysis buffer, incubate at 4 °C for 5 min on a rotating wheel. After each wash, discard the supernatant using a magnetic stand.

[First round of ligation (adaptor ligation); Timing 30 min (plus overnight and 30 min the following day)]

51. For each sample, resuspend the beads in 20  $\mu$ l of first round ligation reaction mix as tabulated below, and incubate at 16 °C overnight at 1,100 rpm shaking in a Thermomixer.

Component	Amount ( $\mu$ l)	Final concentration
Ligation buffer (4 $\times$ )	4	0.8 $\times$
Nuclease-free water	6	
PEG400	4	20% (vol/vol)
Pre-adenylated RNA adaptor A (10 $\mu$ M; from Step 23)	2	1 $\mu$ M
Pre-adenylated RNA adaptor B (10 $\mu$ M; from Step 23)	2	1 $\mu$ M
RNasin Plus RNase inhibitor (40 U $\mu$ l <sup>-1</sup> )	0.5	1 U $\mu$ L <sup>-1</sup>
SUPERaseIn RNase inhibitor (20 U $\mu$ l <sup>-1</sup> )	0.5	0.5 U $\mu$ L <sup>-1</sup>
T4 RNA ligase 2, truncated K227Q (200 U $\mu$ l <sup>-1</sup> )	1	10 U $\mu$ L <sup>-1</sup>
Total	20	

[Critical] We found that T4 RNA ligase 2, truncated K227Q is more specific and efficient at ligating the adenylated RNA adaptor to RNA than T4 RNA ligase 1

(data not shown, YS).

52. Wash the beads once with 500  $\mu$ l of PNK buffer, twice with 1 ml of PG cell lysis buffer, and once with 1 ml of PNK buffer. For the washes with PG cell lysis buffer, incubate the mixture at 4 °C for 1 min on a rotating wheel. After each wash, discard the supernatant using a magnetic stand.

[5' end phosphorylation of RNA and removal of adaptor B's phosphate blocking;  
Timing 1 h]

53. For each sample, resuspend the beads in 20  $\mu$ l of PNK reaction mix as tabulated below, and incubate at 37°C for 30 min 1,100 rpm shaking in a Thermomixer.

Component	Amount ( $\mu$ l)	Final concentration
ATP (10 mM)	2	1 mM
PNK buffer (10 $\times$ )	2	1 $\times$
Nuclease-free water	14.5	
RNasin Plus RNase inhibitor (40 U $\mu$ l <sup>-1</sup> )	0.25	0.5 U $\mu$ L <sup>-1</sup>
SUPERaseIn RNase inhibitor (20 U $\mu$ l <sup>-1</sup> )	0.25	0.25 U $\mu$ L <sup>-1</sup>
T4 PNK (10 U $\mu$ L <sup>-1</sup> )	1	0.5 U $\mu$ L <sup>-1</sup>
Total	20	

54. Wash the beads once with 1 ml of PNK buffer, twice with 1 ml of PG cell lysis buffer, and once with 1 ml of PNK buffer. After each wash, discard the supernatant using a magnetic stand. For the second wash with PG cell lysis buffer, incubate the mixture at 4 °C for 5 min on a rotating wheel.

[Second round of ligation (hybrid generation); Timing 30 min (plus overnight and 30 min the following day)]

55. For each sample, resuspend the beads in 100  $\mu$ l of second round ligation reaction mix as tabulated below, and incubate overnight at 16°C at 1,100 rpm shaking in a Thermomixer.

Component	Amount ( $\mu$ l)	Final concentration
ATP (10 mM)	10	1 mM

T4 RNA ligase reaction buffer (10×)	10	1×
Nuclease-free water	72.5	
RNasin Plus RNase inhibitor (40 U $\mu\text{l}^{-1}$ )	2	0.8 U $\mu\text{l}^{-1}$
SUPERaseIn RNase inhibitor (20 U $\mu\text{l}^{-1}$ )	0.5	0.1 U $\mu\text{l}^{-1}$
T4 RNA ligase 1 (10 U $\mu\text{l}^{-1}$ )	5	0.5 U $\mu\text{l}^{-1}$
Total	100	

[Critical step] The use of RNA ligase 1 is crucial since this ligase can ligate RNA without 5' adenylation.

56. Wash the beads once with 800  $\mu\text{l}$  of PNK buffer, once with 1 ml of high salt wash, and once with 1 ml of PNK buffer. After each wash, discard the supernatant using a magnetic stand, and after the final wash, resuspend the beads in 1 ml of PNK buffer.

[Pause point] The sample can be kept at 4 °C overnight.

[Preparation of antibody-attached beads; Timing 2-3 h]

57. Transfer 100  $\mu\text{l}$  per sample of Dynabeads Protein G to a new 1.5 ml microfuge tube, and discard the supernatant using a magnetic stand. We prepare the beads for up to 4 samples in one 1.5 ml microfuge tube.
58. Wash the beads twice with 1 ml of CLIP lysis buffer. After each wash, discard the supernatant using a magnetic stand.
59. Resuspend the beads in 100  $\mu\text{l}$  per sample of CLIP lysis buffer with 10  $\mu\text{g}$  per sample M2 Flag antibody, and incubate at room temperature for 1-2 h on a rotating wheel.
60. Wash the beads twice with 1 ml of CLIP lysis buffer and once with 1 ml of T-20 IP buffer. After each wash, discard the supernatant using a magnetic stand. After the final wash, resuspend the beads in 100  $\mu\text{l}$  per sample of T-20 IP buffer.

[Second round of immunoprecipitation; Timing 3 h]

61. For each sample, discard the supernatant from the beads from Step 56. Add 80  $\mu\text{l}$  of urea cracking buffer (1.25 $\times$ ), and incubate at 65 °C for 3 min at 1,100 rpm shaking in a Thermomixer to elute protein-RNA complexes from the beads.
62. Collect the supernatant containing protein-RNA complexes to a new 1.5 ml microfuge tube, and mix it with 920  $\mu\text{l}$  of ice-cold T20-IP buffer with 1  $\mu\text{l}$  of ANTI-RNase (15-30 U  $\mu\text{l}^{-1}$ ), 10  $\mu\text{l}$  of Protease Inhibitor Cocktail Set III, and 1  $\mu\text{l}$  of SUPERaseIn RNase inhibitor (20 U  $\mu\text{l}^{-1}$ ).
63. Place the sample on a magnetic stand, and transfer the supernatant to a new 1.5 ml microfuge tube.  
[Critical Step] This step is required to remove any residual beads.
64. Add 100  $\mu\text{l}$  of antibody attached beads from Step 60 to each sample, and incubate at 4 °C for 2 h on a rotating wheel.
65. Wash the beads twice with 1 ml of high salt wash, and once with 1 ml of PNK buffer. After each wash, discard the supernatant using a magnetic stand, and after the final wash, resuspend the beads in 1 ml of PNK buffer.

[Radiolabelling of RNA; Timing 30 min]

66. For each sample, transfer 200  $\mu\text{l}$  of the beads (20% of all beads) from Step 65 to a new 1.5 ml microfuge tube, and discard the supernatant. Resuspend the beads in 4  $\mu\text{l}$  of hot-PNK reaction mix as tabulated below, and incubate at 37 °C for 5 min at 1,100 rpm shaking in a Thermomixer.  
[Critical steps] Radiolabelling of RNA is very efficient. In order to minimize the radioactivity of samples, we only radiolabel 20% of a sample, and mix it together with the remaining 80% at Step 68.

Component	Amount ( $\mu\text{l}$ )	Final concentration
$^{32}\text{P}$ - $\gamma$ -ATP (10 mCi $\text{ml}^{-1}$ )	0.4	1 mCi $\text{ml}^{-1}$
PNK buffer (10 $\times$ )	0.4	1 $\times$
Nuclease-free water	3	

T4 PNK (10 U $\mu\text{l}^{-1}$ )	0.2	0.5 U $\mu\text{l}^{-1}$
Total	4	

67. Discard the supernatant, and add 20  $\mu\text{l}$  of NuPAGE LDS sample buffer (1 $\times$ ) with NuPAGE sample reducing agent (1 $\times$ ) to the radiolabelled beads.

[SDS-PAGE, nitrocellulose membrane transfer, and autoradiography; Timing 3h]

68. For each sample, discard the supernatant from 800  $\mu\text{l}$  of the remaining beads (80% of all beads) from Step 65. Add the radiolabelled beads in 20  $\mu\text{l}$  of the buffer from Step 67 to it, and mix.
69. Incubate the samples at 70°C for 5 min at 1,100 rpm shaking in a Thermomixer, and transfer the supernatant to new 1.5 ml microfuge tubes using a magnetic stand.
70. Load the supernatant on a 4-12% NuPAGE Bis-Tris gel and perform electrophoresis for 45 min at 180 V using a XCell SureLock Mini-Cell with MOPS running buffer (1 $\times$ ) and 500  $\mu\text{l}$  of NuPAGE antioxidant by following the manufacturer's instruction. We also run 5  $\mu\text{l}$  of PageRuler Plus prestained protein ladder on the last lane as a protein size marker.  
[Critical step] In order to facilitate excision of membrane, load samples in alternate wells.
71. Remove the gel cassettes from the tank and open the cassettes. Cut out and discard the part of the gel containing the dye front.  
[Caution] This contains free  $^{32}\text{P}$ - $\gamma$ -ATP, which is highly radioactive. Discarding this part can minimize the radioactivity of materials handled at later steps.
72. Transfer the protein-RNA complexes from the gel to a Protan BA85 nitrocellulose membrane at 30V for 1 h using a XCell SureLock Mini-Cell and XCell IITM blot module kit with NuPAGE transfer buffer (1 $\times$ ) and methanol (10%; vol/vol) by following the manufacturer's instructions.

[RNA isolation; Timing ~4 h (plus overnight and 1h the following day)]

73. Take out the membrane from the apparatus. Cover the membrane with saran wrap, and expose it to a Carestream Kodak BioMax XAR Film at -80 °C for 1-2 h (place a fluorescent sticker next to the membrane to facilitate subsequent alignment of the film and membrane).

[Critical Step] The exposure time to obtain sufficient signal on the films varies depending on the radioactivity of the samples. From our experience of iCLIP, the required exposure time varies from 15 min to overnight.

[Critical Step] The visualization of SDS-PAGE separated protein-RNA complexes is one of the crucial quality control steps. It can be used to assess the RNase and purification conditions for both the preliminary and the sample preparation experiments (Fig. 5). The length of crosslinked RNA can be estimated by the position of the signal on the autoradiograph compared to the expected position of the non-crosslinked target protein, which can be determined by western blotting. Roughly, a 2 kDa shift indicates that the target protein is crosslinked to 7 nts of RNA<sup>35</sup>. On the other hand, the absence of RNA that is not crosslinked to the target RBP can be confirmed by: the absence of the additional band for the lane where the sample is loaded; and the absence of signal for the lanes where the control samples are loaded (omitting UV crosslinking or omitting the induction of the target protein).

[Pause point] The samples on the membranes can be stored at -80 °C for a week.

? Trouble shooting

74. For each sample, cut the region of membrane containing the protein-RNA complexes using the autoradiograph as a guide. The region to be cut is indicated in Fig. 5 by the black dotted lines.

[Critical Step] The diffuse signal above the expected molecular weight of the RBP should be excised from the membrane. The length of RNA can be estimated by the location of the signal related to the molecular weight of the RBP as described at Step 73. Here we purify a broad size range of crosslinked RNA. The precise size selection is performed at Step 93.

75. Cut the excised membranes into small pieces, and place them into a 1.5 ml microfuge tube.
76. Add 200  $\mu$ l PK buffer with 10  $\mu$ l proteinase K to the membrane pieces, and incubate them at 37°C for 20 min at 1,100 rpm shaking in a Thermomixer.
77. Add 200  $\mu$ l of PK buffer with 7 M urea to each sample, and incubate for 20 min at 37°C and 1,100 rpm in a Thermomixer.
78. Perform acid-PCI extraction and ethanol precipitation to purify the extracted RNA as described in Step 4-11.  
 [Critical step] Most of the RNA should be extracted from the membrane to the solution after proteinase K treatment. Check the extraction efficiency by monitoring radioactivity of the materials using a scintillation counter.  
 ? Troubleshooting
79. Resuspend the RNA pellet in 6.25  $\mu$ l of nuclease-free water.  
 [Pause Point] The purified RNA can be stored for at least a month at -80 °C.
- [Reverse transcription and cDNA purification; Timing 4 h (plus overnight)]
80. For each sample, add 1  $\mu$ L of reverse transcription primer (0.5  $\mu$ M) and 1  $\mu$ L of dNTPs (10 mM).  
 [Critical step] In order to allow multiplexing of samples, use reverse transcription primers with different experimental barcodes for different samples (see “reverse transcription and multiplexing libraries” section of **Experimental design**).
81. Incubate the samples for 5 min at 70 °C in a thermal cycler to denature RNA, and hold at 25 °C until ready to proceed to the next step .
82. Add 11.75  $\mu$ l of RT reaction mix as tabulated below to each sample.

Component	Amount ( $\mu$ l)	Final concentration (in 20 $\mu$ L)
-----------	-------------------	-------------------------------------

First-strand buffer (5×)	4	1×
DTT (100 mM)	1	5 mM
Nuclease-free water	5.75	
Superscript III RT (200 U $\mu\text{l}^{-1}$ )	0.5	5 U $\mu\text{l}^{-1}$
RNasin Plus RNase inhibitor (40 U $\mu\text{l}^{-1}$ )	0.5	1 U $\mu\text{l}^{-1}$
Total	11.75	

83. Perform the reverse transcription reaction in a thermal cycler by incubating the samples at 25 °C for 5 min, 42 °C for 20 min, 50 °C for 40 min, 80 °C for 5 min, and holding at 4 °C.
84. Add 2.2  $\mu\text{l}$  of 1 M NaOH to each sample, and incubate at 98 °C for 20 min to hydrolyse the template RNA.
85. Add 25  $\mu\text{l}$  of 1M HEPES-NaOH to each sample to neutralize the buffer.
86. Add 352.8  $\mu\text{l}$  of TE buffer, 1  $\mu\text{l}$  of linear acrylamide, and 40  $\mu\text{l}$  of sodium acetate (3M, pH 5.5), to each sample, and mix. Perform ethanol precipitation as described in Step 7-11.
87. Resuspend the cDNA pellet in 6  $\mu\text{l}$  of nuclease-free water.  
[Pause point] The purified cDNA can be stored for at least a month at -20°C.
88. Add 6  $\mu\text{l}$  of TBE-urea loading buffer (2×) to the cDNA. As a size marker, mix 100 ng of Low Molecular Weight DNA Ladder in 12  $\mu\text{l}$  of TBE-urea loading buffer (1×).
89. Incubate the samples for 5 min at 80°C to denature DNA, and place on ice.
90. Load the samples on a 6% Novex TBE-Urea gel (10 well) and the marker sample on the last lane, and perform electrophoresis for 40 min at 180 V using a XCell SureLock Mini-Cell with 900 ml of TBE buffer (1×) by following the manufacturer's instructions.  
[Critical step] In order to facilitate gel excision, load samples in alternate wells..

91. Take out the gel from the cassette. Cut off the lane of the gel containing the size marker, and stain it by incubating for 10 min shaking in 10 ml TBE buffer with 2  $\mu$ l SYBR green II. Wash once with TBE buffer.
92. Visualize the size marker using a UV transilluminator, and print the image.
93. For each sample, excise a region of the gel corresponding to 100–300 nt single-stranded DNA using the printed image of the size marker from Step 92, and place the gel piece with 400  $\mu$ l of TE buffer into a new 1.5 ml microfuge tube. [Critical step] The adaptor sequence accounts for 71 nt of the cDNA. If each arm of hybrid reads is longer than 15 nt, the resulting cDNA should be longer than 100 nt. This is why only regions above 100 nt should be excised. Further details of the gel purification procedure have been described previously as part of the iCLIP procedure<sup>48</sup>.
94. Purify and pellet the cDNA from the gel piece as described in step 18-21. [Critical step] The cDNA sample should not be radioactive from this step on, since the radiolabelled RNA is hydrolysed and removed by the gel electrophoresis. Confirm the absence of radioactive signal from the samples using a scintillation counter. [Pause point] The purified cDNA can be stored for at least a month at -20 °C.
- ? Troubleshooting

[Circularization and linearization of cDNA; Timing 3 h (plus overnight)]

95. For each sample, resuspend the cDNA in 8  $\mu$ l of circularization reaction mix as tabulated below, and incubate for 1 h at 60°C in a thermal cycler.

Component	Amount ( $\mu$ l)	Final concentration
MnCl <sub>2</sub> (50 mM)	0.4	2.5 mM
CircLigase II reaction buffer (10×)	0.8	1×
Nuclease-free water	6.5	
CircLigase II ssDNA ligase (100 U $\mu$ l <sup>-1</sup> )	0.3	3.75 U $\mu$ l <sup>-1</sup>
Total	8	

96. Add 30  $\mu$ l of oligo annealing reaction mix as tabulated below to each sample.

Component	Amount ( $\mu$ l)	Final concentration
FastDigest buffer (10 $\times$ )	3	1 $\times$
Cut oligo (10 $\mu$ M)	1	0.3 $\mu$ M
Nuclease-free water	26	
Total	30	

97. Anneal the cut oligo to the cDNA in a thermal cycler using the following programme:

Cycle number	Condition
1	95 $^{\circ}$ C, 2 min
2-72	Successive cycles of 20 s, starting from 95 $^{\circ}$ C and decreasing the temperature by 1 $^{\circ}$ C each cycle down to 25 $^{\circ}$ C
73	25 $^{\circ}$ C, hold

98. Add 2  $\mu$ l of FastDigest BamHI to each sample, and incubate at 37 $^{\circ}$ C for 30 min, and then 80  $^{\circ}$ C for 10 min in a thermal cycler.

99. Add 360  $\mu$ l of TE, 1  $\mu$ l of linear acrylamide, and 40  $\mu$ l of sodium acetate (3M, pH 5.5) to each sample, and mix. Perform ethanol precipitation as described in Step 7-11.

100. Resuspend the cDNA pellet in 11  $\mu$ l of nuclease-free water.

[Pause point] The cDNA sample can be stored for at least a month at -20  $^{\circ}$ C.

[PCR amplification of cDNA library: cycle number optimization; Timing 2 h]

101. For each sample, split 1.5  $\mu$ l of the sample from Step 100 to three new 1.5 ml microfuge tubes (0.5  $\mu$ l each). The remaining 9.5  $\mu$ l of the sample is used at Step 108 to produce a library for sequencing.

[Critical step] The number of PCR cycles required to achieve optimal amplification of the cDNA library varies depending on the samples and

experimental conditions. Therefore, we always perform this pilot experiment to determine the appropriate cycle number.

102. Add 9.5  $\mu$ l of PCR reaction mix as tabulated below to each sample.

Component	Amount ( $\mu$ l)	Final concentration (in 10 $\mu$ L)
P5Solexa/P3Solexa primer mix (10 $\mu$ M each)	0.25	0.25 $\mu$ M
AccuPrime SuperMix I	5	
Nuclease-free water	4.25	
Total	9.5	

103. Perform the PCR amplification in a thermal cycler using the following program:

Cycle number	Denature	Anneal	Extend	Final
1	94°C, 2 min			
2-26 (or 2-31)	94°C, 15 s	65°C, 30 s	68°C, 30 s	
27 (or 32)				68°C, 3 min; 25°C, hold

[Critical Step] We test 25 and 30 cycles of PCR amplification to identify the optimal cycle number.

[Critical Step] All the post-PCR experiments should be performed in a dedicated area. Since most high-throughput DNA sequencing based methods share the adaptor sequences, it is imperative to avoid cross-contamination of post-PCR products between samples.

104. For each sample, mix 8  $\mu$ l of PCR product with 2  $\mu$ l of sample loading dye (5 $\times$ ) in a 1.5 ml microfuge tube. As a size marker, mix 100 ng of Low Molecular Weight DNA Ladder in 10  $\mu$ l of sample loading dye (1 $\times$ ).

105. Load the samples on a 6% Novex TBE gel and the marker sample on the last lane, and perform electrophoresis for 25 min at 180 V using a XCell SureLock Mini-Cell with 900 ml of TBE buffer (1 $\times$ ) by following the manufacturer's

instructions.

106. Take out the gel from the cassette, and stain it by incubating for 10 min shaking in 10 ml TBE buffer with 1  $\mu$ l SYBR safe. Wash once with TBE buffer.

107. Visualize DNA in the gel using a UV transilluminator, and determine the optimal PCR cycle number for library preparation. The ideal PCR cycle number produces a band within the expected DNA size range and minimal primer dimers (Fig. 7).

[Critical Step] It is also important to confirm that the library is at a linear amplification stage during PCR and at a sufficient concentration for sequencing<sup>28</sup>.

We generally aim for the amplified DNA to have approximately similar signal intensity as the size marker on the gel. The PCR gel can also be used to confirm the absence of background signal from non-crosslinked RNA. The PCR band should only exist for the hiCLIP sample and not control samples.

? Troubleshooting

[PCR amplification of cDNA library; Timing 2 h]

108. For each sample from Step 101, transfer 5  $\mu$ l of the sample to a new 1.5 ml microfuge tube.

109. Add 95  $\mu$ l of PCR reaction mix as tabulated below to each sample. If it exceeds the maximum volume recommended for the thermal cycler, split it across multiple tubes so that the volume in a single tube is appropriate.

Component	Amount ( $\mu$ l)	Final concentration (in 100 $\mu$ L)
P5Solexa/P3Solexa primer mix (10 $\mu$ M each)	2.5	0.25 $\mu$ M
AccuPrime SuperMix I	50	
Nuclease-free water	42.5	
Total	95	

110. Perform PCR amplification as described in Step 103 with the cycle number determined in Step 107.

111. To confirm the appropriate amplification of the library, transfer 8  $\mu$ l of the post-PCR sample to a new 1.5  $\mu$ l microfuge tube. Mix it with 2  $\mu$ l of sample loading dye (5 $\times$ ). As a size marker, mix 100 ng of low molecular weight DNA ladder in 10  $\mu$ l of sample loading dye (1 $\times$ ).

112. Load the samples on a 6% Novex TBE gel and the marker sample on the last lane, and perform electrophoresis for 25 min at 180 V using a XCell SureLock Mini-Cell with 900 ml of TBE buffer (1 $\times$ ) by following the manufacturer's instructions. Visualize the samples as described in Step 105-107. Confirm that the band is within the expected DNA size range, and the signal from primer dimers is minimal. (See Step 107 and Fig. 7)

? Troubleshooting

113. Purify the remaining 92  $\mu$ l of post-PCR sample from Step 110 using AMPure XP by following the manufacturer's instructions. After the purification, resuspend the sample in 20  $\mu$ l of nuclease-free water.

[Pause point] The post-PCR library can be stored for at least 3 months at -20  $^{\circ}$ C.

[Library quantification and high-throughput DNA sequencing; Timing 2 h (plus 24 h)]

114. For each sample, prepare 2  $\mu$ l of serial dilutions (1/100, 1/1,000, 1/10,000, and 1/100,000) in a 96-well plate for qPCR. Also prepare 2  $\mu$ l of reference libraries with known concentration at 100 pM, 10 pM, 1 pM and 0.1 pM.

[Critical Step] The reference libraries must contain adaptor sequences that can be amplified by qPCR primer 1 and 2 as well as the sequence that can anneal to DLP oligo.

115. Add 18  $\mu$ L of qPCR reaction mix below to each qPCR sample. Note that the DLP oligo anneals to the adaptor region of the hiCLIP library, and thus can quantify hiCLIP library for any RBP generated using the adaptors described here.

Component	Amount ( $\mu$ l)	Final concentration (in 20 $\mu$ L)
DLP oligo (10 $\mu$ M)	0.5	0.25 $\mu$ M
qPCR primer 1 (10 $\mu$ M)	0.6	0.3 $\mu$ M
qPCR primer 2 (10 $\mu$ M)	0.6	0.3 $\mu$ M
Platinum Quantitative PCR SuperMix-UDG w/ROX	10	
Nuclease-free water	6.3	
Total	18	

116. Seal the plate with MicroAmp optical adhesive film.

117. Perform qPCR in a qPCR machine following the manufacturer's instructions with the following method (Detector: FAM-TAMRA; Passive reference: ROX) and the following program:

Cycle number	UDG incubation*	Denature	Anneal	Extend
1	54°C, 2 min			
2		94°C, 10 min		
3-42		94°C, 15 s	62°C, 1 min	72°C, 30s

\* Uracil-DNA glycosylase (UDG) incubation step is included to eliminate contaminating DNA from the previous qPCR experiments.

118. Plot the concentrations of reference libraries versus the Ct values on a log scale, and fit a straight line. The formula of the line can be used to calculate the concentration from the Ct value. Average the concentrations of the samples estimated from the serial dilution, and use the average concentration as the sample concentration.

119. We multiplex all the libraries where the signal can be observed by the gel at Step 107 (Fig. 7; hiCLIP replicates together with the control sample that omits the second round of ligation). Sequence the library using a MiSeq Gene & Small Genome Sequencer and MiSeq Reagent Kit following the manufacturer's instruction. We sequenced 196 bases for the hiCLIP library of STAU1<sup>8</sup>.  
? Troubleshooting

[Bioinformatic analysis: Setting the environment; TIMING seconds]

[Critical] All of the commands for the bioinformatic analysis (steps 120 – 139) are intended to be run in R. Each command is prefixed here with either the terminal command prompt “\$” or the R command prompt “>” which does not need to be entered. Each R function is associated with a help file that can be consulted if needed.

120. Open R. If using RConsole or RStudio for Mac OS X, open from the terminal rather than clicking on the icon to allow R to access the system’s PATH environment variable.

(A) RConsole

```
$ open -a R
```

(B) RStudio

```
$ open -a RStudio
```

121. Load the “hiclipr” library. Create the directory structure within the specified working directory and set the working directory (in this example “~/hiCLIP”).

```
> library(hiclipr)
> CreateDirectories(folder = "~/hiCLIP")
```

[Bioinformatic analysis: Pre-processing the reads; Timing: 30 min]

122. Download the raw sequencing reads (FASTQ files) from the sequencer to the “data” folder. This example uses “LUs27\_196.fq.gz” from the original hiCLIP paper<sup>8</sup>, which can be downloaded from the European Nucleotide Archive (experiment accession: ERR605257).

```
> download.file(url =
"ftp://ftp.sra.ebi.ac.uk/vol1/ERA358/ERA358132/fastq/LUs27_196.fq.gz",
destfile = "data/LUs27_196.fq.gz")
```

123. Extract the UMI (positions 1-2 and 7-9 at the 5' end of the read) and save as tab-separated values for collapsing PCR duplicates downstream. The output FASTQ file is renamed with the suffix “\_reordered”. Both output files are saved in the “results/processed\_data” folder.

```
> ExtractRandomBarcodes(fq.file = "LUs27_196.fq.gz", barcode.filename =  
"LUs27_196_barcode.tsv")
```

124. Use the experimental barcode (positions 3-6 at the 5' end of the read) to demultiplex the experiments, specifying the experiment names. The demultiplexed reads are renamed according to the experiment names and saved in the “results/processed\_data” folder. In this example, “LigPlusLow” corresponds to low RNase and “LigPlusHigh” to high RNase conditions, while LigMinus corresponds to the control sample (where the second round of ligation is omitted).

```
> DemultiplexReads(fq.file = "LUs27_196_reordered.fq.gz", LigPlusLow =  
"AATA", LigPlusHigh = "GGTT", LigMinus = "GGCG")
```

125. Identify and select the hybrid reads specifying the adapters. The output FASTQ file is renamed with the suffix “\_hybrid” and saved in the “results/processed\_data” folder.

```
> SelectHybrid(fq.file = "LigPlusLow.fq.gz", adapterA =  
"AGATCGGAAGAGCGGTTCAG", adapterB = "CTGTAGGCACCATACAATG")  
> SelectHybrid(fq.file = "LigPlusHigh.fq.gz", adapterA =  
"AGATCGGAAGAGCGGTTCAG", adapterB = "CTGTAGGCACCATACAATG")
```

[Bioinformatic analysis: Align the hybrid reads; Timing: variable, a few hours]

126. Create the Bowtie index files in the “ref/bowtie” folder for the four-stage alignment strategy. This only needs to be performed first time the pipeline is run. The function will download the relevant FASTA files and create four indices: a ribosomal and transfer RNA index (prefix: rRNA\_tRNA); a mitochondrial and pre-rRNA index (prefix: mtDNA\_prerRNA); a transcriptome index using the longest RNA transcript for each gene (prefix: longestRNA); a genome index (prefix: GRCh37)

```
> CreateBowtieIndices()
```

NOTE: Ensembl GRCh37 release 75 has been used.

127. Apply the four-phase alignment strategy, specifying the aligned BAM filename and number of processors to be used. The BAM alignment file will be output to the “results/mapped” folder.

```
> AlignReadsBowtie(fq.file = "LigPlusLow_hybrid.fq", bam.file =  
"LigPlusLow.bam", processors = 2)  
> AlignReadsBowtie(fq.file = "LigPlusHigh_hybrid.fq", bam.file =  
"LigPlusHigh.bam", processors = 2)
```

[Caution] The number of processors specified should be less than half the total available as `AlignReadsBowtie` runs two processes in parallel.

[Critical step] For alignment phase 1 (to rRNA and tRNA) and 2 (to mtDNA and pre-rRNA), reads are allowed to align to multiple locations (with primary alignments carried forward) due to their high sequence redundancy. This also prevents them from impacting upon alignment phase 3 (to the transcriptome), where much of the analysis is focused and only uniquely aligned reads are considered to ensure validity of the results.

[Bioinformatic analysis: Post-processing; Timing: 30 min]

128. Load the BAM files in turn and ensure only valid hybrid reads are kept (i. e. both hybrid arms have been aligned)

```
> LigPlusLow.hybrids <- LoadBAM(bam.file = "LigPlusLow.bam")  
> LigPlusHigh.hybrids <- LoadBAM(bam.file = "LigPlusHigh.bam")
```

129. Collapse PCR duplicates (defined as hybrid reads that have the same UMI and the same seqname, start and strand SAM fields). The previously created barcode file is passed to the function. The frequency table of PCR duplicates is output to the console.

```
> LigPlusLow.collapsed <- EvaluateBarcodes(bam.dt = LigPlusLow.hybrids,  
barcodeindex = "LUs27_196_barcode.tsv")  
> LigPlusHigh.collapsed <- EvaluateBarcodes(bam.dt =  
LigPlusHigh.hybrids, barcodeindex = "LUs27_196_barcode.tsv")
```

130. Annotate the hybrids with the transcript biotype and transcript region (e.g. 5' UTR, CDS, 3' UTR). Then combine the two for downstream analysis (as in the original paper<sup>8</sup>).

```
> LigPlusLow.annotated <- AnnotateHybrid(LigPlusLow.collapsed)
> LigPlusHigh.annotated <- AnnotateHybrid(LigPlusHigh.collapsed)
> annotated.hybrids <- rbind(LigPlusLow.annotated,
LigPlusHigh.annotated)
```

131. Select hybrids that are mapped to transcripts (i. e. not ribosomal or transfer RNA). There is the option to select mature or immature transcripts; if immature transcripts are selected, introns are included (in genomic coordinates). There is also the option to select intragenic or intergenic hybrids. Intragenic hybrids are used for downstream analysis.

```
> intragenic.hybrids <- SelectGenicHybrid(annotated.hybrids, type =
"intra", class = "mature")
```

132. FOR INTRAGENIC HYBRIDS ONLY: Create a set of non-redundant hybrids from the intragenic hybrids. When more than two hybrids cover the same area (with a specified minimum width and coverage), they are collapsed with the coverage providing a level of confidence in these hybrids reads. They are merged with the remaining hybrids to create the non-redundant set of hybrids.

```
> islands <- FindIslands(intragenic.hybrids, minimum.coverage = 2,
minimum.width = 8)
> nonislands <- FindNonIslandHybrids(intragenic.hybrids, islands)
> nonredundant.hybrids <- CreateNonRedundantHybrids(islands,
nonislands)
```

133. FOR INTRAGENIC HYBRIDS ONLY: Identify the duplexes from the non-redundant hybrids. The duplex sequences are obtained and RNAhybrid<sup>46</sup> used to identify the duplexes.

```
> hybrid.sequences <- GetSequence(nonredundant.hybrids)
> RunRNAHybrid(hybrid.sequences, "rnahybrid.out")
> nonredundant.duplexes <- FindDuplex(hybrid.sequences,
"rnahybrid.out")
```

[Caution] The RNAHybrid output file (in this example “rnaybrid.out”) is written to the “results/RNAhybrid” directory. RunRNAhybrid will not work if the file already exists.

[Bioinformatic analysis: Plots; Timing: variable; a few minutes]

CRITICAL. All the plots are saved in the “plots” folder.

134. Plot the regions to which the hybrid arms map.

```
> hybrids <- rbind(LigPlusLow.hybrids, LigPlusHigh.hybrids)
> annotated <- AnnotateHybrid(hybrids)
> PlotHybridRegion(annotated, filename = "hybrid_region.png")
```

[Caution] Our strategy for processing the unique molecular identifier is not suitable for evaluating the cDNA of origin of reads aligned to multiple locations (i.e. phase 1), because it considers the alignment location in conjunction with the unique molecular identifier. This could lead to underestimating the number of PCR duplicates for this phase. However, for this plot, we assume the rate of PCR duplication is reasonably uniform across regions and so the relative proportion of regions to which the hybrid arms map are unlikely to be affected.

135. Plot the ratio of intragenic to intergenic reads.

```
> genic.hybrids <- SelectGenicHybrid(annotated.hybrids, type = "all",
class = "mature")
> PlotIntraInterRatio(genic.hybrids, filename =
"intra_inter_ratio.png")
```

136. Compare intragenic and intergenic regions

```
> PlotGenicRegion(genic.hybrids, type = "bar", filename =
"genic_region.png")
```

137. Plot the mate for hybrid arms mapped to a particular region (in this example, 3' UTR).

```
> PlotRegionMate(intragenic.hybrids, region = "UTR3", filename =
"region_mate.png")
```

138. Plot the loop length distribution for non-redundant duplexes. These can be weighting according to coverage of the islands/duplexes.

```
> PlotLoopLength(nonredundant.duplexes, weighting = TRUE, ratio =  
FALSE, filename = "loop_length.png")
```

[Bioinformatic analysis: Export intragenic duplexes for visualization; Timing: 30 mins]

139. Convert the non-redundant duplexes from transcriptomic to genomic coordinates for visualization in a genome browser. The exported BED file is written to the “results” folder.

```
> ExportBED(nonredundant.duplexes, filename =  
"nonredundant_duplexes.bed")
```

## **TROUBLESHOOTING**

Troubleshooting advice can be found in Table 2

## **TIMING**

- Step 1-22, Preparation of adaptors: 2h, overnight RNA precipitation, ~5h the following day
- Step 23-30, Preparation of cells and UV crosslinking: ~2 days (depending of the growth of the cell)
- Step 31-33, Preparation of antibody-coupled beads: 4 h
- Step 34-39, Cell lysis and partial RNA digestion: 2 h
- Step 40-48, First round of immunoprecipitation: 3-4 h
- Step 49-50, 3' end dephosphorylation: 1 h
- Step 51-52, First round of ligation (adaptor ligation): 30 min, overnight ligation, 30 min the following day
- Step 53-54, 5' end phosphorylation of RNA and removal of adaptor B's phosphate blocking: 1 h
- Step 55-56, Second round of ligation (hybrid generation): 30 min, overnight ligation, 30 min the following day
- Step 57-60, Preparation of antibody-attached beads: 2-3 h

- Step 61-65, Second round of immunoprecipitation: 3 h
- Step 66-67, Radiolabelling of RNA: 30 min
- Step 68-72, SDS-PAGE and nitrocellulose membrane transfer: 3h
- Step 73-79, RNA isolation: ~4 h, overnight precipitation, 1h the following day
- Step 80-94, Reverse transcription and cDNA purification: 3 h
- Step 95-100, Circulization and linearization of cDNA: 3 h
- Step 101-107, PCR amplification of cDNA library (the cycle number optimisation): 2 h
- Step 108-113, PCR amplification of cDNA library: 2 h
- Step 114-119, Library quantification and high-throughput DNA sequencing: 2 h, ~24 h for the sequencing
- Step 120-139, Bioinformatic analysis: ~1 day

## **ANTICIPATED RESULTS**

hiCLIP generates hybrid RNAs from RNA duplexes bound by the RBP of interest; the two arms of the duplexes are ligated via a linker. Those hybrid RNAs crosslinked to the RBP are purified and converted into a high-throughput DNA sequencing compatible library. The high-throughput DNA sequencing data can be used to investigate RNA duplexes as described in the overview.

The results of hiCLIP will depend on the RBPs and biological conditions studied. Here we summarize the key statistics when we applied hiCLIP to FLAG-tagged STAU1 of Flp-In 293 T-REx cells as an example. Using cells growing on ten 15 cm dishes, approximately 35,000 hybrid reads were obtained. The proportion of hybrid reads among all the reads where the RNA source can be identified was 1 – 2 %.

The output of the bioinformatic analysis of this example dataset is shown in Fig. 3. This includes the following analyses: the source of hybrid RNAs, the proportion of intragenic and intergenic hybrids and the distance between the RNA duplex arms in CDS and 3' UTRs. In case of STAU1 target RNA duplexes, intra-molecular RNA duplexes dominated over inter-molecular RNA duplexes and a longer distance between intra-molecular RNA duplex arms in the 3' UTRs was observed compared to

in the CDS. Although we primarily focused on intra-molecular RNA duplexes due to their dominance in this dataset, the same approach should be applicable to inter-molecular RNA duplexes. A BED file is also created to visualize the RNA duplexes in standard genome browsers such as the University of California Santa Cruz (UCSC) genome browser<sup>49</sup> and the Integrative Genomics Viewer (IGV)<sup>50</sup> (Fig. 4).

The “hiclipr” package was developed to allow researchers to analyze their own hiCLIP data easily. To this end, the original pipeline has been updated and refined and there are now fewer dependencies. Thus, the results obtained by the “hiclipr” package are very similar to the original study, but there are some differences. These are accounted for largely by: (i) the use of a more recent Ensembl release and (ii) the updated method of hybrid read selection building on the original as described in Experimental design.

hiCLIP offers a universal approach to study RNA duplexes recognized by a specific RBP. The experimental and computational procedure described here can be used to study any RBPs bound to RNA duplexes and biological conditions, and thus advance understanding of the role of RNA structures and RBPs in post-transcriptional regulation.

## **Figure legends**

**Figure 1 | Overview of high-throughput methods to study *in vivo* RNA structures**

**Figure 2 | Schematic overview of the hiCLIP procedure**

**Figure 3 | Anticipated results and comparison with the original pipeline**

The left-hand plots are from the new pipeline while the right-hand plots are from the original paper. The top panel (A) shows the summary of regions to which the hybrid arms align. The middle panel (B) shows the proportion of intragenic and intergenic hybrid reads. The bottom panel (C) shows the density distribution of loop lengths of duplexes in the CDS and 3' UTR regions. The differences in these three plots can be accounted for by the use of a newer Ensembl annotation release and an updated method of hybrid read selection. The outputs of the original pipeline are reproduced from Sugimoto et al. (2015)<sup>8</sup>.

**Figure 4 | Example genome browser view**

The genome browser view for XBP1 is shown using the UCSC genome browser. External data tracks can be added as desired by the user; here we have selected

conservation and single nucleotide polymorphism tracks. The difference between the hybrids and the duplexes contained within them is also shown.

### **Figure 5 | Quality control of hiCLIP by autoradiograph**

The checkpoints for examining the autoradiographs, and their implications, are shown together with representative gel images. The gel images are reproduced from Sugimoto et al. (2015)<sup>8</sup>.

### **Figure 6 | Adaptor preparation (adenylation and purification)**

RNA adaptor A and RNA adaptor B are adenylated, and analyzed by PAGE. The RNA adaptors are visualized by SYBR Green II here, but for the actual purification experiments, we used UV-shadowing. The bands corresponding to the adenylated adaptors are indicated by red arrows. The bands are excised and the adenylated adaptors extracted from the gel.

### **Figure 7 | PCR cycle number optimization for cDNA library amplification**

The gel images of PCR cycle number optimization experiments are shown. The left panel summarizes the checkpoints for preliminary experiments to determine the appropriate PCR cycle number together with a representative gel image and its implications. Based on the result of the preliminary experiment, a PCR cycle number of 24 was chosen. The right panel shows a representative gel image of PCR amplified cDNA libraries suitable for high-throughput DNA sequencing.

### **Table 1 | Oligonucleotides**

5Phos, 3Phos, dd, r, N, 6FAM and TAMRA indicate 5' phosphorylation, 3' phosphorylation, dideoxynucleotide, ribonucleotide, random nucleotide, 5' 6-FAM and 3' TAMRA modification respectively. All the oligonucleotides are generic and should be applicable for the study of any RBPs.

### **Table 2 | Troubleshooting**

## **AUTHOR CONTRIBUTION**

Y.S. and J.U. conceived the hiCLIP protocol; Y.S. and A.C. wrote and described the software for data analysis; A.C. developed the “hiclipr” package; Y.S. A.C., N.L., and J.U. wrote the manuscript; J.U. and N.L. supervised the project.

## **Acknowledgments**

We would like to thank Federico Agostini for helpful advice on the “hiclipr” package; Cristina Militti for valuable comments on the manuscript; Flora Lee and Igor Ruiz de los Mozos for testing the “hiclipr” package; and all the members of the Ule and Luscombe labs for providing fruitful discussions throughout the study. The hiCLIP project was supported by funding from European Research Council (617837-

Translate) to J.U., a Wellcome Trust Joint Investigator Award to J.U. and N.L. (103760/Z/14/Z), the NAKAJIMA FOUNDATION FELLOWSHIP and MRC Centenary Early Career Award to Y.S, a Wellcome Trust PhD Training Fellowship for Clinicians to A.C., 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).

### **Competing financial interests**

The authors declare no competing financial interests.

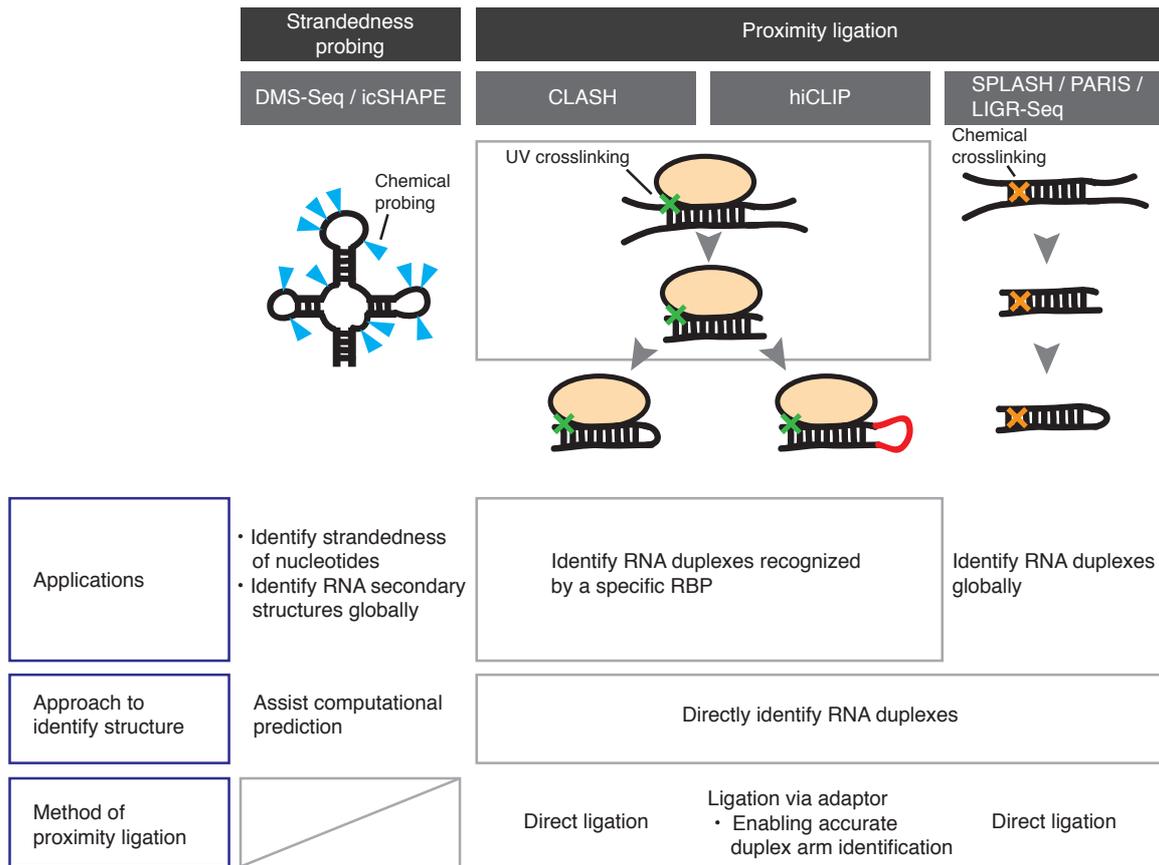
## References

1. Wan, Y., Kertesz, M., Spitale, R.C., Segal, E. & Chang, H.Y. Understanding the transcriptome through RNA structure. *Nature reviews. Genetics* **12**, 641-655 (2011).
2. Lu, Z. & Chang, H.Y. Decoding the RNA structurome. *Current opinion in structural biology* **36**, 142-148 (2016).
3. Tian, B., Bevilacqua, P.C., Diegelman-Parente, A. & Mathews, M.B. The double-stranded-RNA-binding motif: interference and much more. *Nature reviews. Molecular cell biology* **5**, 1013-1023 (2004).
4. He, L. & Hannon, G.J. MicroRNAs: small RNAs with a big role in gene regulation. *Nature reviews. Genetics* **5**, 522-531 (2004).
5. Gong, C. & Maquat, L.E. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature* **470**, 284-288 (2011).
6. Kretz, M., *et al.* Control of somatic tissue differentiation by the long non-coding RNA TINCR. *Nature* **493**, 231-235 (2013).
7. Gong, C., Tang, Y. & Maquat, L.E. mRNA-mRNA duplexes that autoelicit Staufen1-mediated mRNA decay. *Nature structural & molecular biology* **20**, 1214-1220 (2013).
8. Sugimoto, Y., *et al.* hiCLIP reveals the in vivo atlas of mRNA secondary structures recognized by Staufen 1. *Nature* **519**, 491-494 (2015).
9. Aw, J.G., *et al.* In Vivo Mapping of Eukaryotic RNA Interactomes Reveals Principles of Higher-Order Organization and Regulation. *Molecular cell* **62**, 603-617 (2016).
10. Lu, Z., *et al.* RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. *Cell* **165**, 1267-1279 (2016).
11. Sharma, E., Sterne-Weiler, T., O'Hanlon, D. & Blencowe, B.J. Global Mapping of Human RNA-RNA Interactions. *Molecular cell* **62**, 618-626 (2016).
12. Kertesz, M., *et al.* Genome-wide measurement of RNA secondary structure in yeast. *Nature* **467**, 103-107 (2010).
13. Rouskin, S., Zubradt, M., Washietl, S., Kellis, M. & Weissman, J.S. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* **505**, 701-705 (2014).
14. Ding, Y., *et al.* In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* **505**, 696-700 (2014).
15. Spitale, R.C., *et al.* Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* **519**, 486-490 (2015).
16. Flynn, R.A., *et al.* Transcriptome-wide interrogation of RNA secondary structure in living cells with icSHAPE. *Nature protocols* **11**, 273-290 (2016).
17. Kwok, C.K., Ding, Y., Tang, Y., Assmann, S.M. & Bevilacqua, P.C. Determination of in vivo RNA structure in low-abundance transcripts. *Nature communications* **4**, 2971 (2013).
18. Weeks, K.M. & Mugaer, D.M. Exploring RNA structural codes with SHAPE chemistry. *Accounts of chemical research* **44**, 1280-1291 (2011).
19. Zuker, M. & Stiegler, P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic acids research* **9**, 133-148 (1981).

20. Kudla, G., Granneman, S., Hahn, D., Beggs, J.D. & Tollervey, D. Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 10010-10015 (2011).
21. Helwak, A., Kudla, G., Dudnakova, T. & Tollervey, D. Mapping the Human miRNA Interactome by CLASH Reveals Frequent Noncanonical Binding. *Cell* **153**, 654-665 (2013).
22. Helwak, A. & Tollervey, D. Mapping the miRNA interactome by cross-linking ligation and sequencing of hybrids (CLASH). *Nature protocols* **9**, 711-728 (2014).
23. Calvet, J.P. & Pederson, T. Heterogeneous nuclear RNA double-stranded regions probed in living HeLa cells by crosslinking with the psoralen derivative aminomethyltrioxsalen. *Proceedings of the National Academy of Sciences of the United States of America* **76**, 755-759 (1979).
24. Cimino, G.D., Gamper, H.B., Isaacs, S.T. & Hearst, J.E. Psoralens as photoactive probes of nucleic acid structure and function: organic chemistry, photochemistry, and biochemistry. *Annual review of biochemistry* **54**, 1151-1193 (1985).
25. Kaufmann, G., Klein, T. & Littauer, U.Z. T4 RNA ligase: substrate chain length requirements. *FEBS letters* **46**, 271-275 (1974).
26. Ule, J., *et al.* CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302**, 1212-1215 (2003).
27. Konig, J., *et al.* iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nature structural & molecular biology* **17**, 909-915 (2010).
28. Moore, M.J., *et al.* Mapping Argonaute and conventional RNA-binding protein interactions with RNA at single-nucleotide resolution using HITS-CLIP and CIMS analysis. *Nature protocols* **9**, 263-293 (2014).
29. Sander, J.D. & Joung, J.K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol* **32**, 347-355 (2014).
30. Darnell, R.B. HITS-CLIP: panoramic views of protein-RNA regulation in living cells. *Wiley interdisciplinary reviews. RNA* **1**, 266-286 (2010).
31. Van Nostrand, E.L., *et al.* Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nature methods* **13**, 508-514 (2016).
32. Zarnegar, B.J., *et al.* irCLIP platform for efficient characterization of protein-RNA interactions. *Nature methods* **13**, 489-492 (2016).
33. Grosswendt, S., *et al.* Unambiguous identification of miRNA:target site interactions by different types of ligation reactions. *Molecular cell* **54**, 1042-1054 (2014).
34. Moore, M.J., *et al.* miRNA-target chimeras reveal miRNA 3'-end pairing as a major determinant of Argonaute target specificity. *Nature communications* **6**, 8864 (2015).
35. Huppertz, I., *et al.* iCLIP: protein-RNA interactions at nucleotide resolution. *Methods* **65**, 274-287 (2014).
36. Tollervey, J.R., *et al.* Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nat Neurosci* **14**, 452-458 (2011).
37. Castello, A., *et al.* Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393-1406 (2012).

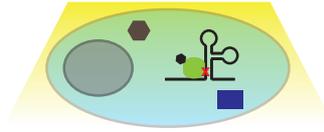
38. Harlow, E. & Lane, D. *Using antibodies : a laboratory manual*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1999).
39. Liu, J., Haorah, J. & Xiong, H. Western Blotting Technique in Biomedical Research. in *Current Laboratory Methods in Neuroscience Research* 187-200 (Springer, 2014).
40. Hymer, W.C. & Kuff, E.L. Isolation of Nuclei from Mammalian Tissues through the Use of Triton X-100. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* **12**, 359-363 (1964).
41. Stockley, P.G. Filter-binding assays. *Methods Mol Biol* **543**, 1-14 (2009).
42. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**, 1-10 (2009).
43. Li, H., *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
44. Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032-2034 (2015).
45. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**(2011).
46. Rehmsmeier, M., Steffen, P., Hochsmann, M. & Giegerich, R. Fast and effective prediction of microRNA/target duplexes. *RNA* **10**, 1507-1517 (2004).
47. Farrell, R.E. *RNA methodologies : a laboratory guide for isolation and characterization*, (Academic Press, San Diego, 1993).
48. Konig, J., *et al.* iCLIP--transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution. *Journal of visualized experiments : JoVE* (2011).
49. Kent, W.J., *et al.* The human genome browser at UCSC. *Genome Res* **12**, 996-1006 (2002).
50. Robinson, J.T., *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26 (2011).

Figure 1 | Overview of high-throughput methods to study *in vivo* RNA structures

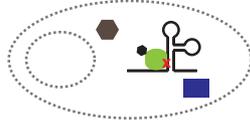


**Figure 2 | Schematic overview of the hiCLIP procedure**

Stage 1. In vivo UV cross-linking  
(Step 23-30)



Stage 2. Cell lysis  
(Step 34-37)



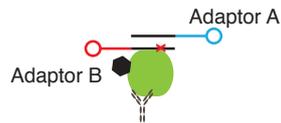
Stage 3. Partial RNA digestion  
(Step 38-39)



Stage 4. First round of immunoprecipitation  
(Step 40-48)



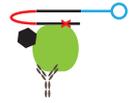
Stage 5. First round of ligation  
(adaptor ligation;  
Step 49-52)



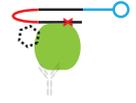
Stage 6. Removal of 3' block  
on Adaptor B  
(Step 53-54)



Stage 7. Second round of ligation  
(hybrid generation;  
Step 55-56)



Stage 8. Second round of immunoprecipitation,  
SDS-PAGE, membrane transfer,  
and autoradiography  
(Step 57-72)



Stage 9. RNA extraction  
(Step 73-79)



Stage 10. cDNA library generation,  
and high-throughput DNA sequencing  
(Step 80-119)

Stage 11. Bioinformatic analysis  
(Step 120-139)

Figure 3 | Anticipated results and comparison with the original pipeline

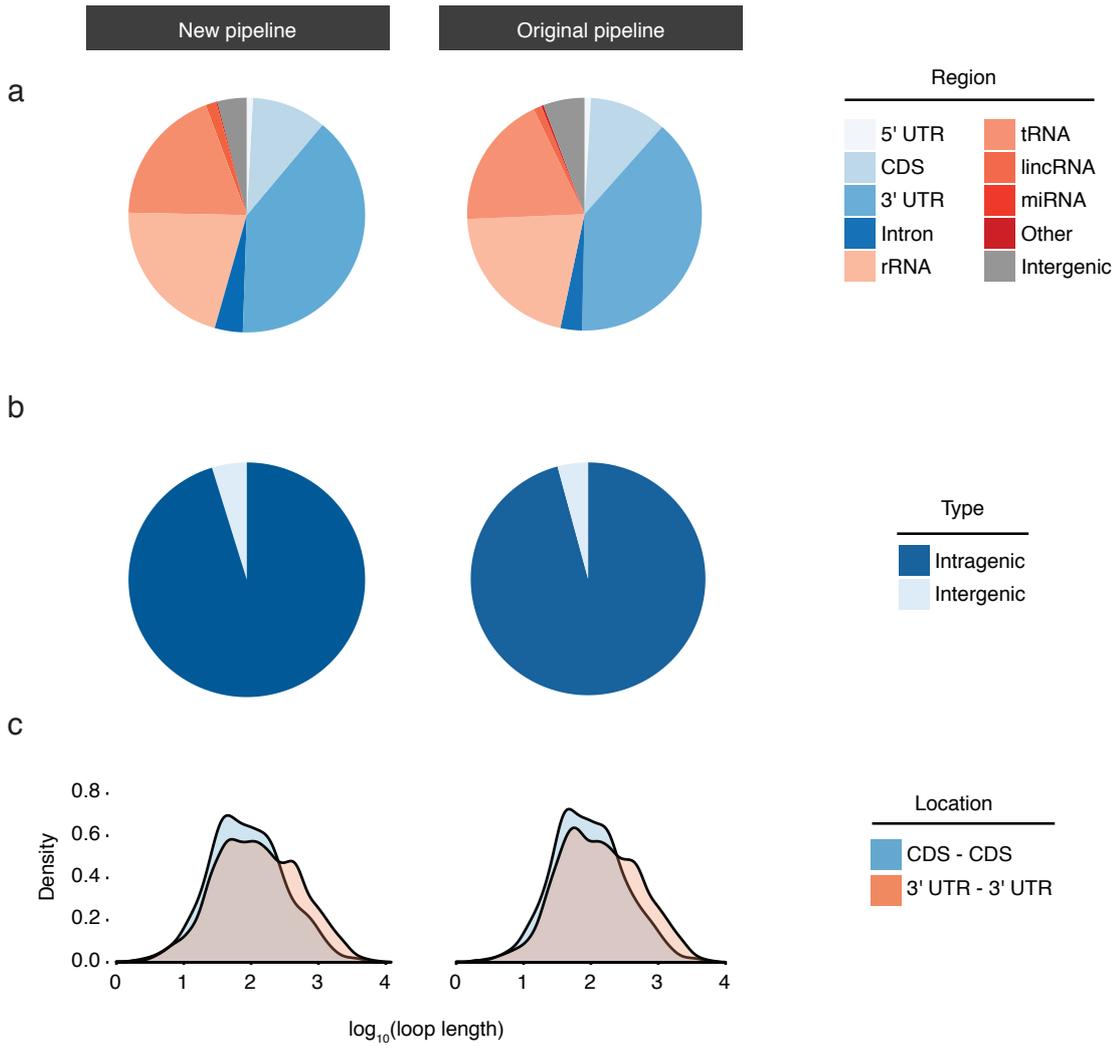


Figure 4 | Example genome browser view

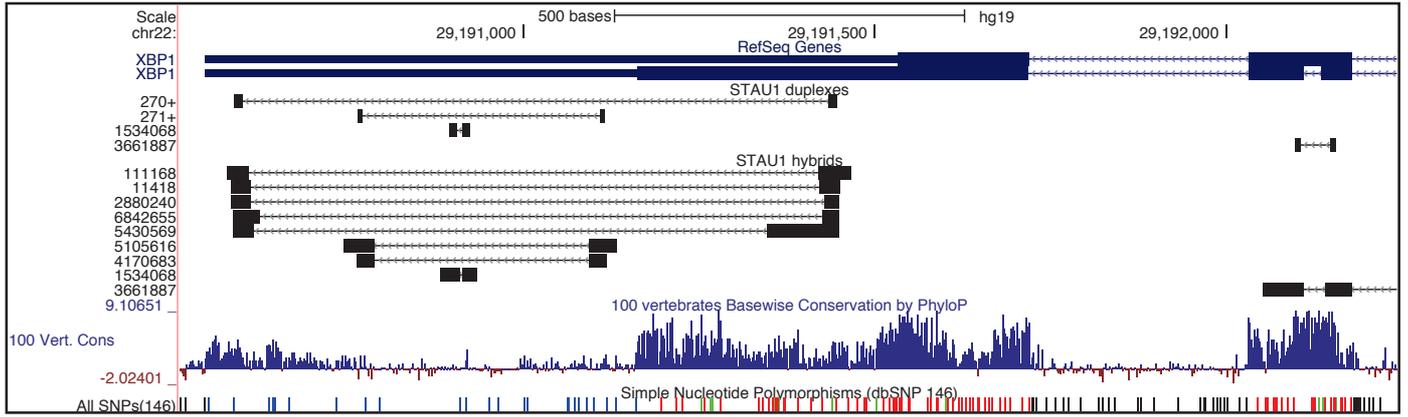


Figure 5 | Quality control of hiCLIP by autoradiograph

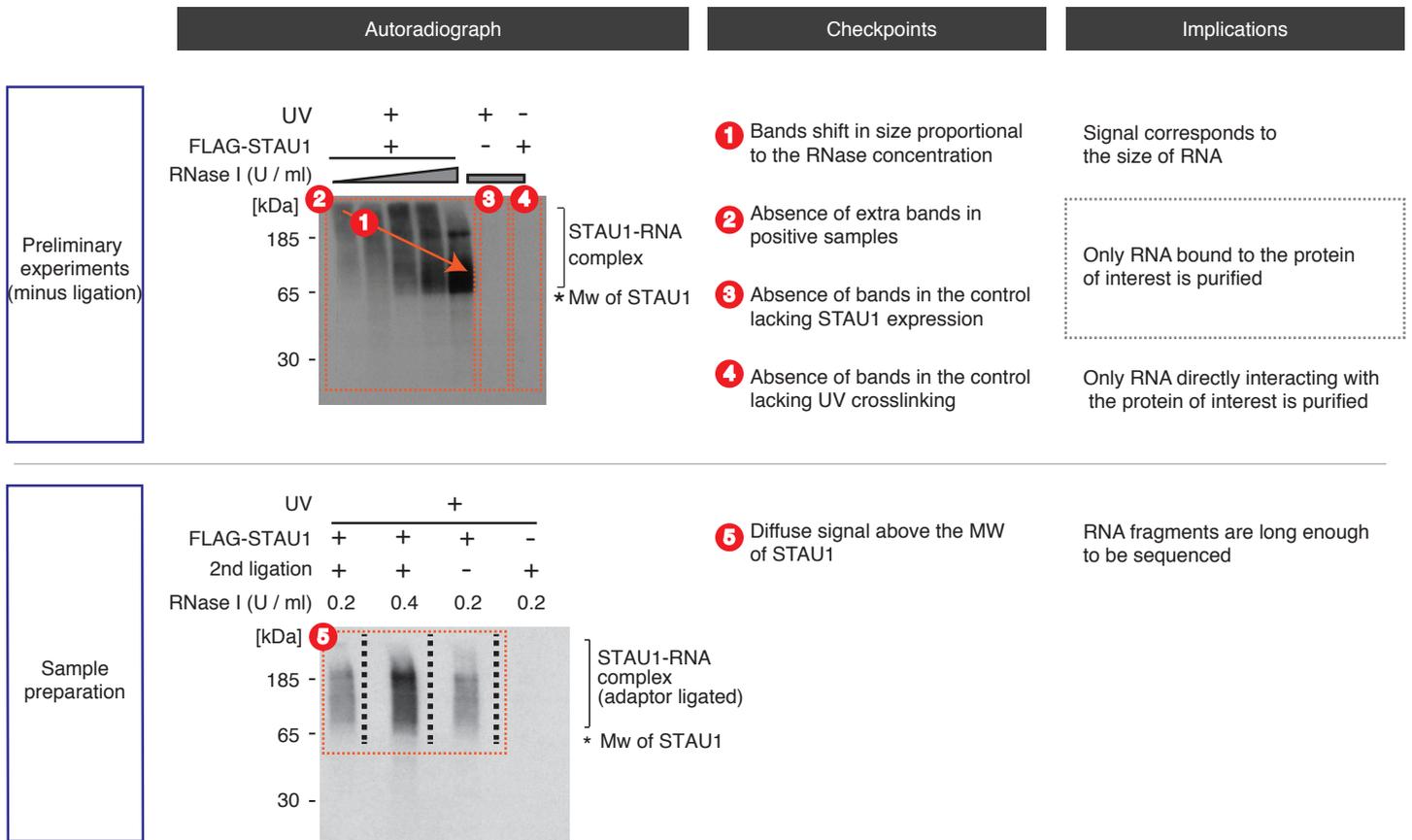


Figure 6 | Adaptor preparation (adenylation and purification)

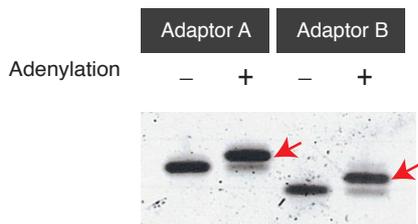


Figure 7 | PCR cycle number optimization for cDNA library amplification

