easyCLIP Analysis of RNA-Protein Interactions Incorporating Absolute Quantification

Douglas Porter (dfporter@stanford.edu)
Stanford University https://orcid.org/0000-0001-9491-8258
Paul Khavari (khavari@stanford.edu)
Stanford University

Method Article

Keywords: RNA, RNA-binding protein, CLIP, Cancer

Posted Date: March 25th, 2021

DOI: https://doi.org/10.21203/rs.3.pex-1333/v1

Abstract

In general, an RNA-binding protein (RBP) may be considered a protein with an abnormally frequent interaction with RNA, and a "target RNA" for a specific protein may be considered an RNA with an abnormally frequent interaction with that protein. Traditional ultraviolet (UV) cross-linking immunoprecipitation (CLIP)-sequencing analysis methods generally define interactions by methods that are indirectly assessing the latter. However, there has been limited direct assessment of these metrics by determining RNA cross-link rates or the RNA-binding profiles of non-RBPs.

Here, we describe a method to determine RNA cross-link rates and create sequencing libraries for a protein of interest, either of which may be performed on their own. easyCLIP is a relatively short, reliable, and efficient CLIP method, with the capacity to directly visualize RNA libraries and diagnose methodological problems. By combining the sequencing data and cross-link rates, the absolute frequency of given cross-links may be compared between control and experimental proteins, to nominate potential RBPs and distinctive RNA-protein interactions.

Introduction

Currently, there is no general method to estimate absolute RNA-protein interaction frequencies and a quantitative test is needed to assess whether any non-random interaction with an RNA exists. The frequencies of RNA-protein complexes, per-cell and per-interaction partner, would enable the fundamental characterization of RNA-protein interaction networks. Determining the targets of an RBP by conventional approaches, such as enrichment over negative control immunopurification or by clustering of cross-links, are ultimately but indirectly determining if the absolute count of an RNA-protein complex in the cell is abnormally high. Defining RNA-protein interaction events per-cell and per-protein in absolute quantities, in contrast, may provide a framework for describing a global and widely reproducible view of RNA-protein interactions.

Here, we report a refinement of current CLIP protocols, termed easyCLIP. easyCLIP produces conventional CLIP-seq libraries, quantifies RNA cross-links-per-protein and provides visual confirmation of each step.

Reagents

T4 RNA Ligase 1 (ssRNA Ligase), High Concentration NEB M0437M

T4 Polynucleotide Kinase NEB M0201L

SUPERase• In™ RNase Inhibitor (20 U/µL) ThermoFisher AM2696

SuperScript™ IV Reverse Transcriptase ThermoFisher 18090010

PrimeSTAR® Max DNA Polymerase Takara R045B

Proteinase K Solution (20 mg/mL) ThermoFisher AM2546

Dynabeads™ Protein G for Immunoprecipitation ThermoFisher 10004D

Pierce™ Anti-HA Magnetic Beads ThermoFisher 88837

DNA LoBind Tubes, DNA LoBind, 1.5 mL Eppendorf 22431021

NuPAGE LDS Sample Buffer (4X) ThermoFisher NP0008

NuPAGE™ 4-12% Bis-Tris Midi Protein Gels, 26-well ThermoFisher WG1403BOX

Adenosine 5'-Triphosphate (ATP) NEB P0756L

DTT ThermoFisher R0861

UltraPure™ 1 M Tris-HCl Buffer, pH 7.5 ThermoFisher 15567027

NaCl (5 M), RNase-free ThermoFisher AM9759

Triton™ X-100 solution Sigma 93443-500ML

Tween™ 20 Surfact-Amps™ Detergent Solution ThermoFisher 28320

NuPAGE™ Transfer Buffer (20X) ThermoFisher NP00061

Poly(ethylene glycol) Sigma 91893-250ML-F

Precision Plus Protein™ All Blue Prestained Protein Standards Bio-Rad #1610373

SYBR™ Green I Nucleic Acid Gel Stain - 10,000X concentrate in DMSO ThermoFisher S7563

PrimeSTAR® Max DNA Polymerase Takara R045B

Equipment

Stratagene UV Stratalinker 2400 Stratagene 2400

Procedure

Buffers

CLIP lysis buffer (Tris pH 7.5 50 mM, NaCl 200 mM, EDTA 1 mM, glycerol 10%, NP40 0.1%, Triton-X 100 1%, N-lauroylsarcosine 0.5%)

High stringency buffer (Tris pH 7.5 15 mM, EDTA 5 mM, Triton X-100 1%, Na-deoxycholate 1%, SDS 0.1%, NaCl 120 mM, KCl 25 mM)

High salt buffer (Tris pH 7.5 15 mM, EDTA 5 mM, Triton X-100 1%, Na-deoxycholate 1%, SDS 0.1%, NaCl 1 M)

Low salt buffer (Tris pH 7.5 15 mM, EDTA 5 mM)

NT2 buffer (Tris pH 7.5 50 mM, NaCl 150 mM, MgCl2 1 mM, Igepal CA-630 0.05%)

Nuclease digestion buffer (per reaction: 50 µL NT2 buffer with enzyme, 10 µL PEG400)

3' PNK buffer (Tris pH 6.8 50 mM, MgCl2 5 mM, PEG400 16.7%, DTT 5 mM, SUPERase•In 0.067 U/µL, T4 PNK 0.5 U/µL)

5' PNK buffer (Tris pH 7.5 50 mM, MgCl2 5 mM, PEG400 16.7%, ATP 1 mM, DTT 5 mM, SUPERase•In 0.33 U/µL, T4 PNK 0.5 U/µL)

3' ligation buffer (Tris pH 7.5 50 mM, MgCl2 10 mM, PEG400 16.7%, DTT 5 mM, SUPERase•In 0.13 U/µL, 3' linker oligo 33 nM, T4 RNA ligase I 1 U/µL)

5' ligation buffer (Tris pH 7.5 50 mM, MgCl2 10 mM, PEG400 16.7%, ATP 0.67 mM, DTT 5 mM, SUPERase•In 0.13 U/µL, 5' linker oligo 167 nM, T4 RNA ligase I 0.5 U/µL)

Library construction for sequencing

UV crosslinking (1h)

1. Remove media, wash cells once with 4°-16° PBS and aspirate PBS.

2. Place dish on ice, remove lid and UV cross-link 0.3J/cm^2

3. Add ~1 ml CLIP lysis buffer per 15 cm plate and scrape to collect.

4. Freeze in dry ice and store at -80°.

Library generation (RNA steps, 1-2 days)

5. Thaw lysates by hand and then keep at 4°.

6. Sonicate lysate for 20-30 seconds at 7 second bursts with microtip sonicator at 10% power

7. Spin lysate at 14 krcf for 10 min at 4°; transfer cleared lysate to new tube

8. Add lysates to anti-HA or antibody-precojugated protein G beads. IP 1 h at 4°.

9. Wash beads 1X High Str, 1X High Salt, 10min each at 4° w/rotation, then 1X Low Salt (1 mL all)

(Note: the above IP and washes can be performed robotically in a KingFisher machine with no visible change.)

10. Perform nuclease digestion: 30°, 15s 1350 rpm, 45s rest, for 3-10 minutes with 0.01-0.1 U/µL RNAse ONE.

11. Remove RNAse mixture and add 1 mL of 4° High Str buffer to deactivate RNAse ONE.

12. Wash beads 2X 1 mL NT2, then step down volume in NT2.

13. Resuspend beads in 36 µl of PNK-3' Mix. Inucbate at 37° for 30min, 15s 1350rpm, 90s rest.

14. Place on magnet and remove PNK reaction (no wash required before ligation).

15. Ligate 2h to overnight in thermomixer at 16° in 36 µl of 3' Ligation Mix. 15s 1350 rpm, 90s rest.

16. Wash 1X H. Salt or H. Str, 2X NT2.

17. Resuspend beads in 70 µL of 5'PNK reaction.

18. Incubate in thermomixer at 37° for 15', 15 s 1000 RPM, 15 s rest.

19. Wash 2X NT2 and step down volume in NT2.

20. Resuspend beads in 70 µL of 5' ligation reaction.

21. Incubate at 16° for 2 h, 15 s 1350 RPM, 90 s rest.

RNA capture (~6 h)

22. Wash 1-2X with NT2 and step down volume in NT1.

23. Optionally freeze 10% for WB.

24. Resuspend beads in 20 µL 1.6X LDS.

25. Heat 75° for 15min.

26. Run samples in NuPAGE gel, 180V 30-50 minutes, transfer to .45 µm nitrocellulose at 400 mA for 27'.

27. Rinse membrane 1xPBS, scan on Odessy CLx (lowest quality/resolution OK)

28. For each sample, add 375 µL PK buffer and 25 µL of Proteinase K (Ambion, 20mg/mL) to a DNA LoBind tube (Eppendorf).

29. Place the nitrocellulose on PBS-soaked filter paper and cut with a scalpel in each hand. We usually cut the minimal region and everything higher. Use a scalpel to transfer the slices of membrane to the Proteinase K mixture.

30. Incubate at 50-55° for 45min in thermomixer. 15s 900rpm, 45s rest.

31. Wash 20 µL of oligo(dT)-dynabeads with BIB and leave in 600 µL BIB.

32. Transfer Proteinase K extract to the oligo(dT) mixture in a DNA LoBind tube. 20 min at 4° with rotation.

33. Wash 1X BIB, transfer to PCR tube, wash 2X NT2, and 3X with 4° PBS. Include RNAse inhibitor.

Reverse transcription (2h)

34. Resuspend dT-beads in 12.4 µL water and 3 µL cDNA synthesis primer.

35. Heat 95° 3' and transfer 13.3 µL sup to new PCR tube.

36. Place tubes at 4° for 5min.

37. Add the rest of the annealing mix (6.7 µL).

38. Incubate 53° for 40 min, 55° for 10 min, 80° for 10 min.

PCR and clean-up (~4 h)

39. Create an 8-reaction master mix of PCR mix (per reaction: 25 µL PrimeSTAR Max DNA polymerase 2X MM (Takara), 1.25 µL 20 µM primer mixture, 1 µL 33X SybrGreen I (ThermoFisher),19.75 µL water) without RT product.

40. Distribute 47 µL of PCR1 mix to 7 tubes and add ≤ 3.3 µL RT product to each.

41. Run PCR program (98C 10s, then cycles of 98C 5 s, 65C 5 s, 72C 45 s) and stop a few reactions into the exponential phase. Reactions should not start to plateau.

43. Purify PCR reactions using a NucleoSpin® Gel and PCR Clean-up kit (MACHEREY-NAGEL). Instead of the kit's wash buffer, use freshly prepared 85% ethanol. Libraries are ready for quantification and sequencing.

Cross-link rate determination

Creation of a protein standard

Create a protein standard by making a two-fold dilution series of the desired epitope and aliquoting into single-use striptubes of ~15 µL (see methods section). For the HA-tag, for example, any purified HA-tagged protein will suffice.

1. Quantify the protein concentration of the standard by coomassie gel with a BSA standard curve, or other suitable protein quantification method.

2. Dilute in protein dilution buffer (0.5X Dulbecco's PBS, 0-5% glycerol, 0.05% Tween-20, 0.15 mg/mL BSA) to 200 ng/µL.

3. Make two-fold dilutions down from 20-200 ng/µL, depending on how protein is available, for a total of 8 concentrations of protein.

4. Aliquot ~14 µL of each concentration into striptubes and freeze until use at -80°.

Creation of RNA standard

An RNA standard to determine fluorescence per molecule number is created by UV cross-linking hnRNP C, purifying a large amount, aliquoting and quantifying by quantitative western blot.

1. Add four replicates of 906-1600 µg of 293T/HCT116 lysate from cross-linked cells to ~20 µL Protein G Dynabeads (ThermoFisher Cat #10003D) coupled with 25 µL (5 µg) anti-hnRNP C (4F4) antibody per replicate.

2. Immunoprecipitate at 4° for ~1 hour.

3. RNAse digest with 0.1 U/µL RNAse ONE for 10 minutes.

4. Perform PNK reaction for 14 minutes at 37°.

5. Ligate overnight (17 hours) with 20 pmol labeled L5, and 2 µL high concentration T4 RNA ligase (NEB).

6. Combine samples and split into ~20 aliquots comprising 2.5% of the beads (~10 ng hnRNP C each, ~400 ng total purified) in ~15 µL 1.6X NuPAGE buffer, freeze in dry ice and keep long term at -80°.

Immunoblot with ~1:3000 αhnRNP C conjugated to AF790 (Santa Cruz Biotechnology, sc-32308 AF790), or other suitable antibody, in PBS blocking buffer (LI-COR) for ~1 hour at room temperature. Use protein standards to quantify the amount of hnRNP C in the L5-ligated band by quantitative western blotting. ~5000 fluorescence units per fmol is a typical result.

Cross-link rate determination using standards

UV Crosslinking (1 h)

1. Remove media, wash cells once with 4°-16° PBS and aspirate PBS.

2. Place dish on ice, remove lid and UV cross-link 0.3J/cm^2

3. Add ~1 ml CLIP lysis buffer per 15 cm plate and scrape to collect.

4. Freeze in dry ice and store at -80°.

IP and RNA modification (1-2 days)

5. Thaw lysates by hand and then keep at 4°.

6. Sonicate lysate for 20-30 seconds at 7 second bursts with microtip sonicator at 10% power

7. Spin lysate at 14 krcf for 10 min at 4°; transfer cleared lysate to new tube

8. Add lysates to anti-HA or antibody-precojugated protein G beads. IP 1 h at 4°.

9. Wash beads 1X High Str, 1X High Salt, 10min each at 4° w/rotation, then 1X Low Salt (1 mL all) (Note: the above IP and washes can be performed robotically in a KingFisher machine with no visible effect.)

10. Perform nuclease digestion: 30°, 15s 1350 rpm, 45s rest, for 3-10 minutes with 0.01-0.1 U/µL RNAse ONE.

11. Remove RNAse mixture and add 1 mL of 4° High Str buffer to deactivate RNAse ONE.

12. Wash beads 2X 1 mL NT2, then step down volume in NT2.

13. Resuspend beads in 70 µL of 5'PNK reaction.

14. Incubate in thermomixer at 37° for 15', 15 s 1000 RPM, 15 s rest.

15. Wash 2X NT2 and step down volume in NT2.

16. Resuspend beads in 70 µL of 5' ligation reaction.

17. Incubate at 16° overnight, 15 s 1350 RPM, 90 s rest.

Gel (~6 h)

18. Wash 1-2X NT2 and step down volume in NT2.

19. Thaw aliquot of RNA standard (cross-linked hnRNP C, frozen in 1.6X LDS), and PCR striptubes of pre-aliquoted protein standard.

20. Add 10 µL of protein standard PCR tubes to a set of striptubes with 10 µL 3.2X LDS using a multichannel pipette.

21. Resuspend beads in 20 µL 1.6X LDS. Heat all samples at 75° for 15min, applying intermittent shaking to samples with beads.

22. Run samples in NuPAGE gel @180V 30-50 minutes, transfer to nitrocellulose (.45 um) at 400 mA for 27'.

23. Place membrane in PBS, scan on Odessy CLx (lowest quality/resolution OK)

24. Perform a western blot against the epitope on the purified proteins and protein standard.

Calculation (1 h)

25. Fit the protein standards to a hyperbolic curve. The fit should be very good.

26. Use the hyperbolic curve to estimate immunopurified protein running in the non-cross-linked protein band. 27. Let this amount be “P”.

28. Divide the fmols of RNA in the RNA standard by the observed L5 signal to obtain fmols RNA per fluorescence unit. Let this amount be “F”.

29. Determine the amount of cross-linked RNA in immunopurified samples by multiplying fluorescence by the fmols/fluorescence of the RNA standard. Let this amount be “R”, calculated as R = (observed fluorescence for RNA) * F.

30. Add the fmols of cross-linked RNA to the fmols of monomeric, un-cross-linked protein to obtain the total amount of protein. Let this amount be “Pt”, calculated as Pt = R + P.

31. Divide the fmols of RNA by the total amount of protein to obtain the fraction cross-linked and ligated. This is calculated as R/Pt.

32. Multiply by 2 to adjust for ~50% ligation efficiency. So the cross-link rate is 2*R/Pt, or 2*R/(R + P).

Combination of cross-link rate with sequencing date

For a given location or RNA, the number of reads in a CLIP library at that location or RNA as a fraction of all reads is the fraction of cross-links at that RNA or location. This fraction may be multiplied by the cross-links per protein to determine the cross-links per protein at the given location or RNA.

Troubleshooting

RNAse:

Verified by fluorescence on nitrocellulose (NC): under-digested sample runs at the top of the gel, and over-digested sample has no smear upward in the minimal region. It is easy to over-digest RNA so that it has a very low mapping rate, but relatively difficult to over-digest so much that the 5' ligation does not occur - this could be caused by the RNAse and ligase having similar requirements for substrate length. RNAse levels may work better for library constructions for many proteins at ~3 minutes digestion at 0.01 U/ul RNAse.

5’ and 3’ ligations:

Assessed by protein shifts and total fluorescence on NC. If both adapters are labeled, the success of the ligations is directly visualized by the amount of fluorescence. Single vs dual protein shifting is visible with some proteins, and is the most sensitive (and by far the easiest) of the three methods we have outlined for ligation efficiency estimates.

Nitrocellulose to reverse transcription

The extraction from NC and oligonucleotide(dT) purification can be assayed by dot-blotting the oligonucleotide(dT) eluate on nylon. Running an hnRNP C sample (or any other good positive control protein - CELF1, hnRNP D, FUBP1) will ensure that you have enough signal for dot blot tests despite sample loss or dilution. The amount of fluorescence extracted from a lane as determined by dot blotting the extract (before poly(dT) purification) is generally greater than the amount of fluorescence on the gel (that is, the apparent efficiency is often over 100%), due to increased brightness per molecule when dot blotting the extract. To determine the success of the poly(dT)-bead purification and elution, save some of the eluate from the poly(dT) beads to dot blot - this corresponds to the input to the RT reaction. Purified RNA must contain the L3 adapter because the L3 adapter contains the poly(A) tail, so fluorescence from the L5 adapter corresponds to completed library.

Reverse transcription (RT)

Verified by comparison of qPCR amplification with input to RT. PCR reactions amplify more quickly the more complex the input sample. The L5 adapter fluorescence input to RT corresponds to completed library. If L5 fluorescence input is in the ~1000s, the input is in the femtomolar range, which should be visibly in the exponential PCR phase before cycle 14. It should never be necessary to amplify more than ~16-18 cycles. Completed libraries before RT can also be visualized by the RNA shift method.

Amount of input RNA

Running the 5'-ligated purified protein alongside a quantified standard to translate fluorescence values into RNA molecule numbers allow for a determination of the total amount of input RNA. On the imaging machines at Stanford, and using our stock of adapters, 1 fmol of ligated L5 adapter gave about 5,000 fluorescence units, and signal >10 fmols was generally enough for library construction.

Time Taken

Library construction for sequencing (~4 days)

UV crosslinking 1 h

Library generation 1-2 days

RNA capture ~6 h

Reverse transcription 2h

PCR and clean-up ~4 h

Cross-link rate determination

Creation of a protein standard ~6 h

Creation of RNA standard ~2 days

Cross-link rate determination using standards ~2 days

Anticipated Results

Proteins typically have cross-link rates above 0.01%, and RNA-binding proteins above 0.1% and below 20%.

An immunopurification of a modestly abundant protein from a 15 cm petri plate of ~80% confluent 293T cells is likely to yield around 1-10 pmol protein, which would generate 10-100 fmol RNA at a cross-link rate of 1%, of which 1-10 fmol would be dual ligated (the largest hit probably being from the 3' ligation, which can be improved by increasing the L3 adapter concentration - this protocol uses a lower L3 adapter concentration than optimal). A conservative estimate of 1% efficiency for all remaining steps would generate 0.1-1 fmol library, or approximately 60-600 million unique reads. A general rule-of-thumb for publishable CLIP data is 1 million unique mapped reads, or much less than 60-600 million. Typically, for ~1 fmol libraries, the largest difficultly is having many empty sequencing reads (which is not an issue for larger libraries); these can be reduced by size purification or deeper sequencing.