3' RNA Adapter Ligation to Input RNA

Book Chapter

In 1 collection

Eric L. Van Nostrand[1,2,3], Thai B. Nguyen[1,2,3], Chelsea Gelboin-Burkhart[1,2,3], Ruth Wang[1,2,3], Steven M. Blue[1,2,3], Gabriel A. Pratt[1,2,3,4], Ashley L. Louie[1,2,3], Gene W. Yeo[1,2,3,4,5,6,7]

[1]Department of Cellular and Molecular Medicine, University of California at San Diego, La Jolla, USA;

[2]Stem Cell Program, University of California at San Diego, La Jolla, USA;

[3]Institute for Genomic Medicine, University of California at San Diego, La Jolla, USA;

[4]Bioinformatics and Systems Biology Graduate Program, University of California at San Diego, La Jolla, USA;

[5]Department of Physiology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore;

[6]Molecular Engineering Laboratory, A*STAR, Singapore, Singapore;

[7]Sanford Consortium for Regenerative Medicine, University of California at San Diego, La Jolla, USA

Sep 02, 2021

1	Works for me

Share

dx.doi.org/10.17504/protocols.io.bph6mj9e

Springer Nature Books

satyavati Kharde

ABSTRACT

Profiling of RNA binding protein targets in vivo provides critical insights into the mechanistic roles they play in regulating RNA processing. The enhanced crosslinking and immunoprecipitation (eCLIP) methodology provides a framework for robust, reproducible identification of transcriptome-wide protein-RNA interactions, with dramatically improved efficiency over previous methods. Here we provide a step-by-step description of the eCLIP method, along with insights into optimal performance of critical steps in the protocol. In particular, we describe improvements to the adaptor strategy that enables single-end enhanced CLIP (seCLIP), which removes the requirement for paired-end sequencing of eCLIP libraries. Further, we describe the observation of contaminating RNA present in standard nitrocellulose membrane suppliers, and present options with significantly reduced contamination for sensitive applications. These notes further refine the eCLIP methodology, simplifying robust RNA binding protein studies for all users.

COLLECTIONS ⓘ

Robust, Cost-Effective Profiling of RNA Binding Protein Targets with Single-end Enhanced Crosslinking and Immunoprecipitation (seCLIP)

KEYWORDS

RNA binding protein, CLIP-seq, eCLIP, seCLIP-seq, seCLIP, CLIP, eCLIP, RNA genomics

GUIDELINES

Introduction

RNA processing has been shown to play pivotal roles in shaping the cellular landscape through regulation of both protein-coding RNAs as well as regulation of processing and function of multiple classes of noncoding RNAs including long intergenic noncoding RNAs (lincRNAs) and small RNAs including microRNAs and piRNAs among others [1,2,3]. These regulatory steps have been shown to play critical roles across a variety of developmental stages, and misregulation of RNA processing has been implicated in many human diseases, including cancer [2,3]. Orchestrating these RNA processing and regulatory roles are RNA binding proteins (RBPs), which play a variety of roles including controlling alternative splicing of mRNA transcripts, targeting RNAs to specific organelles or subcellular localizations with the cell, controlling RNA stability and turnover, and defining the timing and rate of translation [4]. Recent studies have indicated that there are over 1500 RBPs, and this number is continuing to expand with additional studies [5]. Despite these known important roles, however, detailed studies to describe the targets of and regulatory mechanisms have largely focused on a small set of RBPs, with the majority of RBPs remaining poorly characterized.

Recent advancements in next-generation sequencing have made it possible to study RBPs and their RNA targets in an unbiased and transcriptome-wide manner [6,7]. Building upon early RNA ImmunoPrecipitation (RIP) approaches that identified protein binding to entire transcripts, CrossLinking and ImmunoPrecipitation (CLIP) enabled high-resolution profiling of binding sites [8]. In CLIP, RNA-protein interactions are stabilized via ultraviolet crosslinking, a desired protein is immunoprecipitated using a factor-specific antibody, and associated RNA is isolated and converted into DNA library suitable for high-throughput sequencing [8]. Various modifications of CLIP have since been described, including the use of photoactivatable nucleoside analogs (PNAs) to improve crosslinking efficiency (PAR-CLIP) [9] and computational and experimental methods to identify binding with single-nucleotide resolution. Proteinase K treatment of UV-crosslinked protein-RNA complexes leaves at least one amino acid covalently crosslinked to its associated ribonucleotide. Reverse transcriptase enzymes can create deletions at these positions [10] or, more often, terminate elongation due to the inability to read through this coupling, leading

to a substantial fraction of cDNA fragments that terminate at the position of crosslinking. By incorporating a circular ligation step, iCLIP positions this crosslinking site at the start of the sequencing reads to enable identification of binding sites with single-nucleotide resolution [11]. However, widespread usage of these methods has been limited by the low efficiency of converting RNA molecules into adapter-ligated library, which leads to high experimental failure rates and high wasted sequencing (often >90% of reads) due to the presence of substantial PCR duplication [12].

We recently described enhanced CLIP (eCLIP), which incorporated high-efficiency enzymatic steps to achieve thousand-fold improved library efficiency [12]. The improved efficiency dramatically decreases experimental failure rates and PCR duplication, and enabled quantitative comparison with paired size-matched input to remove common CLIP artifacts. Here, we describe a detailed protocol for seCLIP, a simplified, single-end version of the eCLIP methodology, as well as assorted notes for critical handling steps. By ligating an adapter at the 3' end of the cDNA fragment, eCLIP (similar to iCLIP) utilizes this termination to enrich for read pileups at these sites. Due to the adapter strategy used in our initial eCLIP procedure that positioned the I7 adaptor at this site, eCLIP initially required paired-end sequencing to obtain these crosslink sites at the 5' end of the second read, as well as the unique molecular identifier (UMI, or random-mer) used for PCR duplicate identification. In this protocol, we describe an altered adapter strategy to enable single-end sequencing for eCLIP experiments.

Additionally, during eCLIP experiments on cell types with low total RNA quantity, we observed that a substantial fraction (in some cases more than half) of sequenced reads did not map to the human genome and instead mapped to a single bacterial contamination. We trace this source of contamination to manufacturer-supplied nitrocellulose membranes, and describe alternate sources for this material that alleviate this contamination and a method to assay membranes for eCLIP suitability. With a new 3' linker ligation strategy that allows for single-end sequencing, and the alleviation of contamination found commonly in major suppliers, seCLIP brings modifications to eCLIP that will allow for more cost-efficient sequencing as well as paving the way for future low-input RNA as starting material for CLIP-seq experiments.
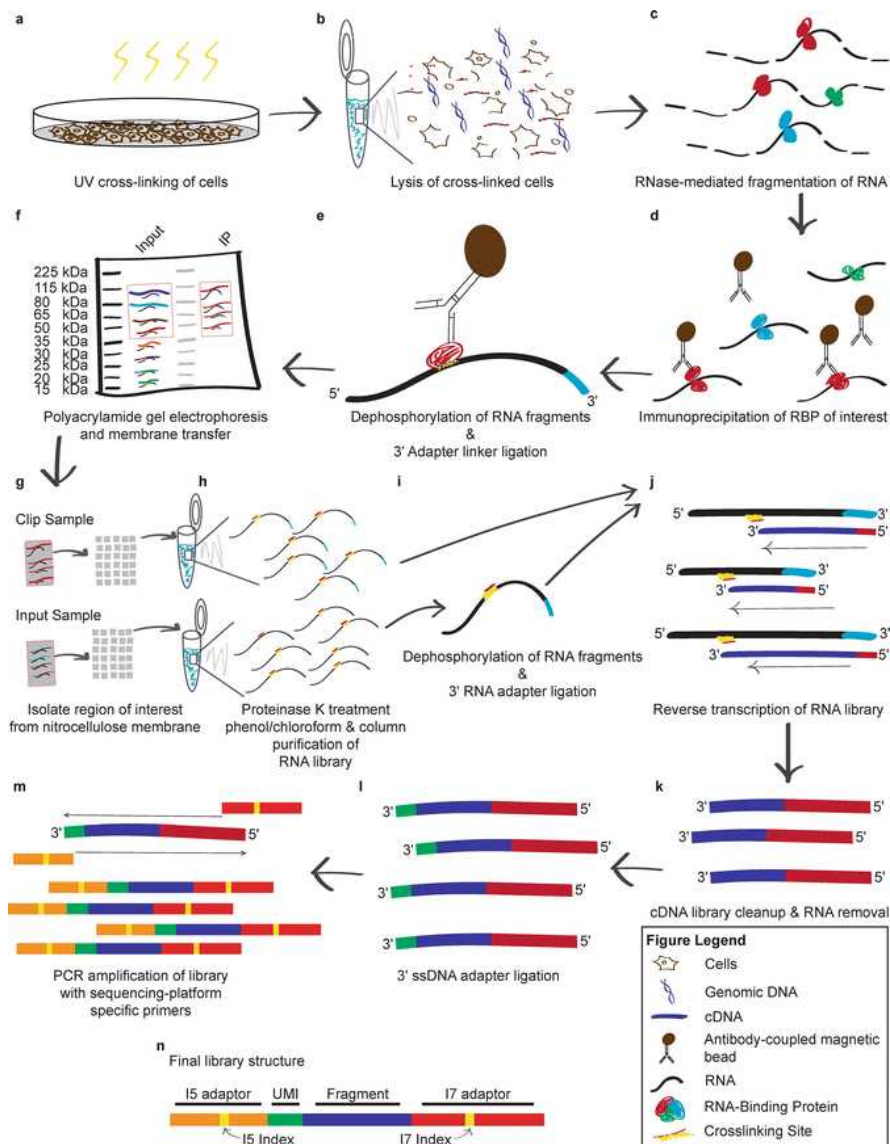
Methods

Fig. 1
Schematic of seCLIP method. (**a**) Crosslinking of cultured cells (Subheadings 3.1.1 and 3.1.2). (**b**) Lysis of crosslinked cells (Subheading 3.2.1). (**c**) RNA fragmentation with RNase (Subheading 3.2.2). (**d**) Immunoprecipitation of RBP -RNA complexes (Subheadings 3.3.1−3.3.5). (**e**) Dephosphorylation of RNA fragments and ligation of 3' RNA adapter (Subheading 3.5.1). (**f**) Polyacrylamide gel electrophoresis and membrane transfer (Subheadings 3.6.1−3.6.5). (**g**) Mince preparative membrane into ~2 mm squares (Subheadings 3.6.6) (**h**) RNA isolation from membrane (Subheadings 3.6.7 and 3.6.8). (**i**) Dephosphorylation of RNA fragments and ligation of 3' RNA adapter for input samples (Subheadings 3.7.1 − 3.8.3). (**j**) Reverse transcription of RNA (Subheadings 3.9.1 and 3.9.2). (**k**) cDNA cleanup (removal of excess primers and RNA) (Subheading 3.9.3). (**l**) Ligation of 3' DNA adapter (on-bead) and cleanup (Subheadings 3.10.1 − 3.10.3). (**m**) PCR amplification of cDNA library and cleanup (Subheadings 3.11.2 − 3.12.2). (**n**) Final Structure of eCLIP library fragment. The unique molecular identifier or random-mer is shown in green and abbreviated as UMI

Notes

1. RNase inhibitor addition to lysis buffer (Section "Lyse Cells" under Cell Lysis and RNA Fragmentation)

Murine RNase inhibitor (NEB) inhibits many endogenous RNase enzymes but does not significantly inhibit RNase I. As such, it can be added either before or after RNase I treatment. For many cell lines (HEK293T, K562) we have observed that this choice does not alter fragmentation or ultimate signal. However, we have observed that for cell types or tissues with moderate to high endogenous RNase activity (e.g., stem cells, differentiated neurons, many tissues), the addition of RNase inhibitor at the initial lysis step is essential to prevent over-fragmentation. We note that the amount of RNase inhibitor may need to be further increased for samples with particularly high RNase activity (e.g., liver or pancreas).

2. Buffer transitions at wash (Section "Wash Beads (Pre-chill All Wash Buffers to 4 °C)", and

Section "Wash Beads (Prechill Buffers to 4 °C)" under Immunoprecipitation Washes and RNA Dephosphorylation, and others)

We often perform intermediate washes with equal mixtures of the previous and current wash buffer, in order to decrease dramatic changes in buffer composition.

3. Modified adaptor strategy for single-end enhanced CLIP (seCLIP) (Section "Prepare RNA Adapter Ligation Master Mix" under Ligate 3' RNA Adapter (On-Bead), Section "Anneal Adapter" under 3' RNA Adapter Ligation to Input RNA, and "Anneal Linker" under 3' Linker Ligate cDNA (On-Bead), and Cleanup)

eCLIP and iCLIP methods utilize the tendency of reverse transcriptase to terminate at the protein-crosslinked RNA nucleotide to enable single-nucleotide resolution of binding sites. In the initial eCLIP adaptor strategy (referred to here as paired-end peCLIP), this position was located at the beginning of the second (paired-end) read in standard Illumina sequencing, which yielded high-quality data but added additional cost. To enable single-end eCLIP (seCLIP), we created a modified adapter strategy that inverted this read structure (Fig. 2a), but used the same highly efficient enzymatic steps (including a 3' RNA adapter ligation to the RNA fragments, reverse transcription using a melting temperature-optimized primer, and a 3' ssDNA ligation to the cDNA fragment). Performing both peCLIP and seCLIP on well-characterized splicing regulator RBFOX2 in HEK293T cells, we observed similar read density profiles for individual example binding sites (Fig. 2b) and comparable results in library yield (quantified as the number of PCR cycles necessary to obtain 100 femtomoles of library (eCT)) with biological replicates averaging 12.3 and 12.7 eCT for peCLIP and seCLIP respectively (Fig. 2c). Transcriptome-wide, we observed high correlation of read density within peaks between peCLIP and seCLIP (R^2 = 0.46 and 0.52), equivalent to those observed between biological replicates of peCLIP (R^2 = 0.52) or seCLIP (R^2 = 0.47) (Fig. 2d). Confirming that seCLIP maintains single-nucleotide resolution, we observed the same stereotypical enrichment for the RBFOX family motif (UGCAUG) enriched at the −6 and −2 positions relative to read start positions (Fig. 2e). Thus, this modified seCLIP adapter strategy enables the same high-quality eCLIP data generation amenable to sequencing with standard single-end, 50 bp chemistry, decreasing cost. Further, the use of standard Illumina barcodes in seCLIP enables pooling with other standard Illumina RNA-seq or other high-throughput sequencing libraries. Open image in new window
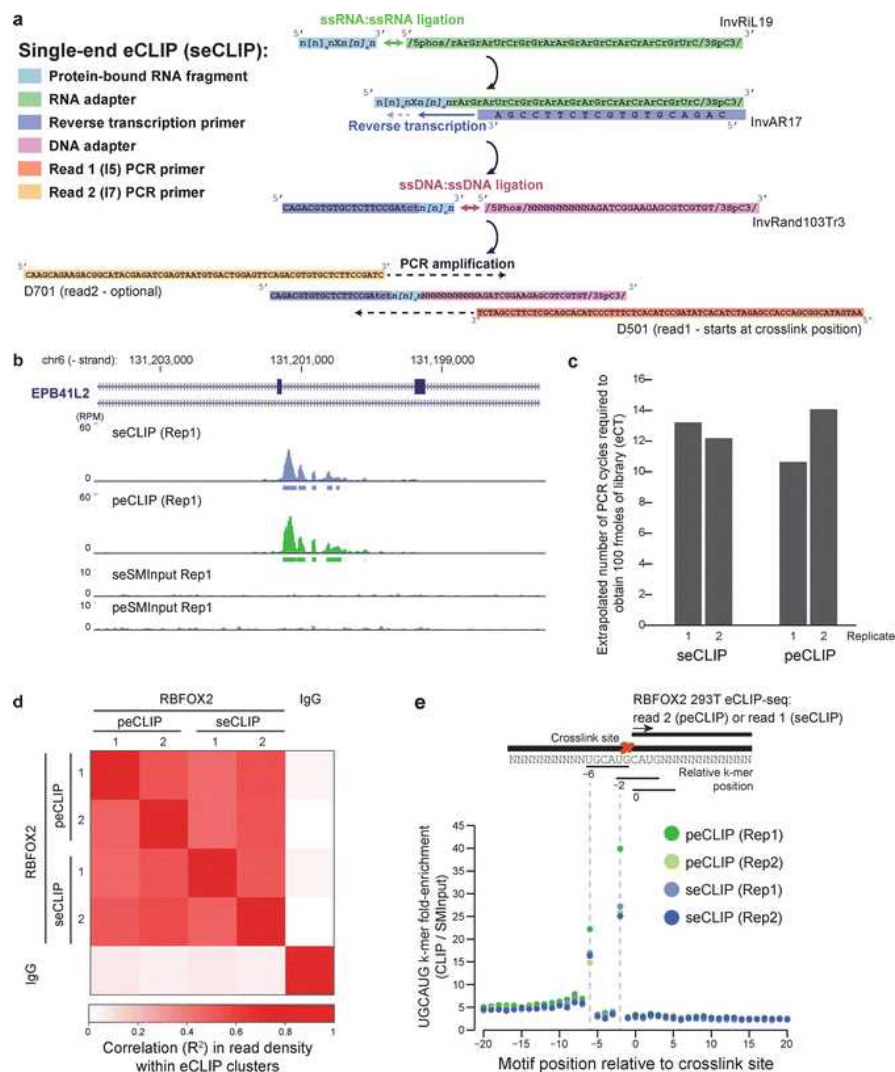
Fig. 2

Modified adaptor strategy for single-end enhanced CLIP (seCLIP). (**a**) Schematic of adaptor sequences used in seCLIP. (**b**) Read density (shown as reads per million; RPM) observed in original, paired-end eCLIP (peCLIP) and seCLIP for RBFOX2 in HEK293T cells at EPB41L2 exon 13-14. Boxes below tracks indicate significantly enriched peaks after input normalization. (**c**) Required amplification observed using seCLIP and peCLIP adaptor strategies for two biological replicates profiling RBFOX2 in HEK293T. (**d**) Heatmap indicates correlation across experiments for read fold-enrichment in CLIP versus input, considering peaks identified in indicated experiments (y-axis). (**e**) Plot indicates enrichment for RBFOX2 binding motif UGCAUG at indicated positions around read start positions

4. Decreasing contamination introduced by nitrocellulose membrane sources (Section "Transfer to Membranes" under Western Blotting and RNA Isolation from Membrane).

In 102 K562 and HepG2 eCLIP experiments, an average of 84% of reads uniquely or multiply mapped to the human genome, respectively [11]. However, in preliminary experiments in cell types with decreased RNA yield after membrane transfer (motor neurons (MN) and neural progenitor cells (NPCs) derived from human embryonic stem cells), we observed in some cases more than 90% of sequenced reads were not mapped to the human genome. Using SOAP-denovo [13] to de novo assemble the unmapped reads, we assembled multiple contigs that were queried against the NR database and showed >99% identity to *Acinetobacter johnsonii XBB1* (CP010350.1). Re-mapping these eCLIP datasets revealed millions of reads in many datasets mapping throughout CP010350.1, confirming this specific species as a major contamination source (Fig. 3a). We noted that reads mapping to CP010350.1 had proper CLIP adapter structure, indicating that contamination was likely occurring prior to the 3' linker ligation.

Fig. 3

Genome of Acinetobacter johnsonii XBB1 in nitrocellulose membranes detected by RT−qPCR. (**a**) Sequencing reads from two eCLIP input libraries mapped to Acinetobacter johnsonii XBB1 (CP010350.1). (**b**) Bars indicate Cq from RT−qPCR performed using CP010350.1-specific qPCR primers on eCLIP RNA (Subheading 3.7.3) and supernatant samples (Subheading 3.4.1) from indicated cell types. Lower Cq reflects higher CP010350.1 signal. Numbers indicate replicate experiments. (**c**) Bars indicate RT−qPCR Cq for CP010350.1 from RNA isolated from nitrocellulose membranes via indicated method, PVDF membranes (RNA isolation by Trizol), paper (RNA isolation by Trizol), and H_2O (Trizol extraction). (**d**) Bars indicate RT−qPCR Ct for CP010350.1 from RNA extracted from size-matched nitrocellulose membranes, in technical replicates. Symbols below indicate samples that were either RNase or DNase treated, and those with and without RT enzyme added. (**e**) Bars indicate RT−qPCR Ct for CP010350.1 from RNA extracted from nitrocellulose membrane samples from five sources in technical replicates as follows: (A) commercial source A, (B) ThermoFisher iBlot (IB23001 lot 2NR26016-01), (C) GE Amersham Protran Premium (13600117 lot G6552142), (D) GE Amersham Protran Premium (1060008 lot G9931040), and (O) original commercial source. Error bars indicate standard deviation from RT−qPCR triplicate measurements

In order to modify eCLIP to ameliorate this issue, we set out to identify the source of this contamination by performing RT-qPCR using primers designed against regions of CP010350.1 with high read density. To first confirm that this contamination was not present in initial samples, we extracted RNA (Trizol LS) from supernatant remaining after immunoprecipitation (Subheading 3.4.1) in addition to standard post-membrane transfer and isolation (Subheading 3.7.3) and performed RT-qPCR for bacterial RNA. We observed more than tenfold increased bacterial RNA signal in the membrane-isolated RNA as compared to supernatant RNA from the equivalent number of cells, indicating that the contamination was not present during tissue culture and was introduced during the IP/Western stage (Fig. 3b). Similar RT-qPCR assays performed after RNA isolation on various buffers or enzyme mixes used failed to identify significant contamination (data not shown).

The observation that these reads were only present in input but not CLIP samples (despite input often having more than 100-fold more RNA recovery and library yield [12]), implicated the difference in RNA adaptor ligation of CLIP (3' RNA adaptor ligation on-bead, before the protein electrophoresis step) versus input (3' RNA adaptor ligation after RNA isolation off of membranes) RNA. Surprisingly, we found that RNA extraction (Trizol) of nitrocellulose membrane alone yielded RT-qPCR signal similar to our contaminated libraries, in contrast to the lower signal observed after RNA isolation from PVDF membranes, Whatman and other lab paper, or negative controls (Fig. 3c). We observed identical results with freshly ordered membrane stock (data not shown). To further explore the nature of the contamination, we synthesized cDNA from either RNase or DNase-treated membrane samples and repeated the RT-qPCR assay. This indicated that the contamination was likely RNA, as the RT-qPCR signal was sensitive to both RNase and the no-RT control, but not DNase (Fig. 3d). The strand-specific signal observed in reads similarly implicated RNA contamination (Fig. 3a).

As the nitrocellulose transfer provides key specificity for isolating RNA crosslinked to protein, we set out to identify optimized alternative sources that had decreased RNA background. We obtained four additional nitrocellulose membrane sources (A), (B) ThermoFisher iBlot2 (IB23001 lot 2NR26016-01), (C) GE Amersham Protran Premium (13600117 lot G6552142), and (D) GE Amersham Protran Premium (1060008 lot G9931040), in addition to our original commercial source (O). For each, we performed RNA isolation followed by the bacterial RNA RT-qPCR, and

observed that whereas O and A showed similar CP010350.1 contamination, B, C, and D did not (Fig. 3e). When we prepared libraries according to our standard protocol for eCLIP input samples, we observed that library yields reflected these results, with B, C, and D showing the least amount of overall contamination (data not shown). Importantly, we observed no difference in RNA or library yield when we prepared standard eCLIP input libraries for two protein size ranges of roughly equal membrane size (10−50 kDa and 50−225 kDa) for multiple membrane types. Thus, these results indicate that testing of nitrocellulose membranes enables optimization of eCLIP by removing substantial background contamination. Our results identify (B) ThermoFisher iBlot and (C) GE Amersham Protran Premium (1060008) membranes as options which show a dramatic decrease in contamination for sensitive eCLIP experiments without altering true library yield. Although yielding equally low contamination, (D) GE Amersham Protran Premium (13600117) is "trial-size" packaging and generally commercially unavailable for large-scale use. Other sources can be tested using the RT-qPCR method described here to determine whether they are of sufficiently low background for use in eCLIP.

Acknowledgments

References

1. Morris KV, Mattick JS (2014) The rise of regulatory RNA. Nat Rev Genet 15(6):423−437. doi: 10.1038/nrg3722CrossRefPubMedPubMedCentralGoogle Scholar
2. Quinn JJ, Chang HY (2016) Unique features of long non-coding RNA biogenesis and function. Nat Rev Genet 17(1):47−62. doi: 10.1038/nrg.2015.10CrossRefPubMedGoogle Scholar
3. Jonas S, Izaurralde E (2015) Towards a molecular understanding of microRNA-mediated gene silencing. Nat Rev Genet 16(7):421−433. doi: 10.1038/nrg3965CrossRefPubMedGoogle Scholar
4. Glisovic T, Bachorik JL, Yong J, Dreyfuss G (2008) RNA-binding proteins and post-transcriptional gene regulation. FEBS Lett 582(14):1977−1986. doi: 10.1016/j.febslet.2008.03.004CrossRefPubMedPubMedCentralGoogle Scholar
5. Gerstberger S, Hafner M, Tuschl T (2014) A census of human RNA-binding proteins. Nat Rev Genet 15(12):829−845. doi: 10.1038/nrg3813CrossRefPubMedGoogle Scholar
6. Licatalosi DD, Darnell RB (2010) RNA processing and its regulation: global insights into biological networks. Nat Rev Genet 11(1):75−87. doi: 10.1038/nrg2673CrossRefPubMedPubMedCentralGoogle Scholar
7. Konig J, Zarnack K, Luscombe NM, Ule J (2011) Protein-RNA interactions: new genomic technologies and perspectives. Nat Rev Genet 13(2):77−83. doi: 10.1038/nrg3141CrossRefGoogle Scholar
8. Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB (2003) CLIP identifies Nova-regulated RNA networks in the brain. Science 302(5648):1212−1215. doi: 10.1126/science.1090095CrossRefPubMedGoogle Scholar
9. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M Jr, Jungkamp AC, Munschauer M, Ulrich A, Wardle GS, Dewell S, Zavolan M, Tuschl T (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 141(1):129−141. doi: 10.1016/j.cell.2010.03.009CrossRefPubMedPubMedCentralGoogle Scholar
10. Zhang C, Darnell RB (2011) Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. Nat Biotechnol 29(7):607−614. doi: 10.1038/nbt.1873CrossRefPubMedPubMedCentralGoogle Scholar
11. Konig J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, Turner DJ, Luscombe NM, Ule J (2010) iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. Nat Struct Mol Biol 17(7):909−915. doi: 10.1038/nsmb.1838CrossRefPubMedPubMedCentralGoogle Scholar
12. Van Nostrand EL, Pratt GA, Shishkin AA, Gelboin-Burkhart C, Fang MY, Sundararaman B, Blue SM, Nguyen TB, Surka C, Elkins K, Stanton R, Rigo F, Guttman M, Yeo GW (2016) Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat Methods 13(6):508−514. doi: 10.1038/nmeth.3810CrossRefPubMedPubMedCentralGoogle Scholar
13. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW, Wang J (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler.

MATERIALS TEXT

1. Crosslinking of Cultured Cells
1. 1× DPBS.
2. 254 nM UV crosslinker.
3. Cell scraper.
4. Liquid Nitrogen.

2. seCLIP
1. **Lysis buffer**: 50 mM Tris−HCl pH 7.4, 100 mM NaCl, 1% NP-40 (Igepal CA630), 0.1% SDS, 0.5% sodium deoxycholate (protect from light), 1:200 Protease Inhibitor Cocktail III (add fresh), in RNase/DNase-free H_2O.
2. Protease Inhibitor Cocktail III.
3. DNase.
4. RNase I.
5. RNase Inhibitor.
6. Dynabeads M-280 sheep anti-rabbit or Protein A/G magnetic beads.
7. **High salt wash buffer**: 50 mM Tris−HCl pH 7.4, 1 M NaCl, 1 mM EDTA, 1% NP-40, 0.1% SDS, 0.5% sodium deoxycholate (protect from light), in RNase/DNase-free H_2O.
8. **Wash buffer**: 20 mM Tris−HCl pH 7.4, 10 mM $MgCl_2$, 0.2% Tween-20, in RNase/DNase-free H_2O.
9. **1× TAP Buffer**: 10 mM Tris pH 7.5, 5 mM $MgCl_2$, 100 mM KCl, 0.02% Triton X-100, in RNase/DNase-free H_2O.
10. Thermosensitive Alkaline Phosphatase (TAP) (1 unit/μL).
11. **5× PNK pH 6.5 buffer**: 350 mM Tris−HCl pH 6.5, 50 mM $MgCl_2$, in RNase/DNase-free H_2O.
12. 0.1 M DTT.
13. T4 PNK.
14. **1× RNA LigaseBuffer**: 50 mM Tris−HCl pH 7.5, 10 mM $MgCl_2$, in RNase/DNase-free H_2O.
15. 10× Ligase Buffer without DTT.
16. 0.1 M ATP.
17. 100% DMSO.
18. 50% PEG 8000.
19. T4 RNA ligase 1 high concentration.
20. 4−12% Bis-Tris Gel.
21. NuPAGE 4× LDS Sample Buffer.
22. NuPAGE MOPS SDS Running Buffer 20×.
23. NuPAGE Transfer Buffer 20×.
24. PVDF membrane.
25. Nitrocellulose membrane:

A	B	C	D
(a) iBlot 2 Transfer Stacks	ThermoFisher	IB23001	lot #2NR26016-01
or			
(b) Amersham Protran Premium	GE	1060008	lot #G9931040

In 102 K562 and HepG2 eCLIP experiments, an average of 84% of reads uniquely or multiply mapped to the human genome, respectively [11]. However, in preliminary experiments in cell types with decreased RNA yield after membrane transfer (motor neurons (MN) and neural progenitor cells (NPCs) derived from human embryonic stem cells), we observed in some cases more than 90% of sequenced reads were not mapped to the human genome. Using SOAP-denovo [13] to de novo assemble the unmapped reads, we assembled multiple contigs that were queried against the NR database and showed >99% identity to *Acinetobacter johnsonii XBB1* (CP010350.1). Re-mapping these eCLIP datasets revealed millions of reads in many datasets mapping throughout CP010350.1, confirming this specific species as a major contamination source (Fig. 3a). We noted that reads mapping to CP010350.1 had proper CLIP adapter structure, indicating that contamination was likely occurring prior to the 3' linker ligation.
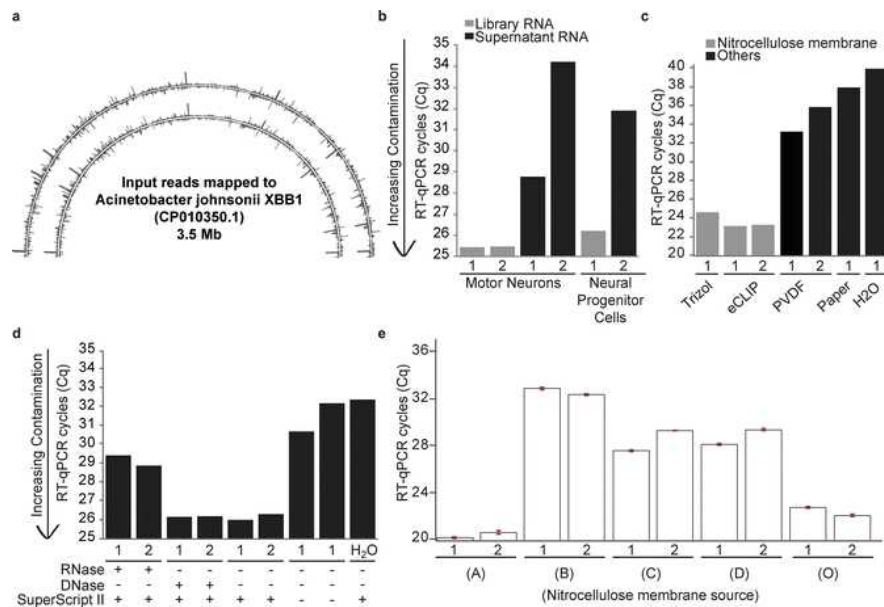
Fig. 3

Genome of Acinetobacter johnsonii XBB1 in nitrocellulose membranes detected by RT−qPCR. (**a**) Sequencing reads from two eCLIP input libraries mapped to Acinetobacter johnsonii XBB1 (CP010350.1). (**b**) Bars indicate Cq from RT−qPCR performed using CP010350.1-specific qPCR primers on eCLIP RNA and supernatant samples from indicated cell types. Lower Cq reflects higher CP010350.1 signal. Numbers indicate replicate experiments. (**c**) Bars indicate RT−qPCR Cq for CP010350.1 from RNA isolated from nitrocellulose membranes via indicated method, PVDF membranes (RNA isolation by Trizol), paper (RNA isolation by Trizol), and H_2O (Trizol extraction). (**d**) Bars indicate RT−qPCR Ct from RNA extracted from size-matched nitrocellulose membranes, in technical replicates. Symbols below indicate samples that were either RNase or DNase treated, and those with and without RT enzyme added. (**e**) Bars indicate RT−qPCR Ct for CP010350.1 from RNA extracted from nitrocellulose membrane samples from five sources in technical replicates as follows: (A) commercial source A, (B) ThermoFisher iBlot (IB23001 lot 2NR26016-01), (C) GE Amersham Protran Premium (13600117 lot G6552142), (D) GE Amersham Protran Premium (1060008 lot G9931040), and (O) original commercial source. Error bars indicate standard deviation from RT−qPCR triplicate measurements

In order to modify eCLIP to ameliorate this issue, we set out to identify the source of this contamination by performing RT-qPCR using primers designed against regions of CP010350.1 with high read density. To first confirm that this contamination was not present in initial samples, we extracted RNA (Trizol LS) from supernatant remaining after immunoprecipitation in addition to standard post-membrane transfer and isolation and performed RT-qPCR for bacterial RNA. We observed more than tenfold increased bacterial RNA signal in the membrane-isolated RNA as compared to supernatant RNA from the equivalent number of cells, indicating that the contamination was not present during tissue culture and was introduced during the IP/Western stage (Fig. 3b). Similar RT-qPCR assays performed after RNA isolation on various buffers or enzyme mixes used failed to identify significant contamination (data not shown).

The observation that these reads were only present in input but not CLIP samples (despite input often having more than 100-fold more RNA recovery and library yield [12]), implicated the difference in RNA adaptor ligation of CLIP (3' RNA adaptor ligation on-bead, before the protein electrophoresis step) versus input (3' RNA adaptor ligation after RNA isolation off of membranes) RNA. Surprisingly, we found that RNA extraction (Trizol) of nitrocellulose membrane alone yielded RT-qPCR signal similar to our contaminated libraries, in contrast to the lower signal observed after RNA isolation from PVDF membranes, Whatman and other lab paper, or negative controls (Fig. 3c). We observed identical results with freshly ordered membrane stock (data not shown). To further explore the nature of the contamination, we synthesized cDNA from either RNase or DNase-treated membrane samples and repeated the RT-qPCR assay. This indicated that the contamination was likely RNA, as the RT-qPCR signal was sensitive to both RNase and the no-RT control, but not DNase (Fig. 3d). The strand-specific signal observed in reads similarly implicated RNA contamination (Fig. 3a).

As the nitrocellulose transfer provides key specificity for isolating RNA crosslinked to protein, we set out to identify optimized alternative sources that had decreased RNA background. We obtained four additional nitrocellulose membrane sources (A), (B) ThermoFisher iBlot2 (IB23001 lot 2NR26016-01), (C) GE Amersham Protran Premium (13600117 lot G6552142), and (D) GE Amersham Protran Premium (1060008 lot G9931040), in addition to our original commercial source (O). For each, we performed RNA isolation followed by the bacterial RNA RT-qPCR, and observed that whereas O and A showed similar CP010350.1

09/02/2021

contamination, B, C, and D did not (Fig. 3e). When we prepared libraries according to our standard protocol for eCLIP input samples, we observed that library yields reflected these results, with B, C, and D showing the least amount of overall contamination (data not shown). Importantly, we observed no difference in RNA or library yield when we prepared standard eCLIP input libraries for two protein size ranges of roughly equal membrane size (10–50 kDa and 50–225 kDa) for multiple membrane types. Thus, these results indicate that testing of nitrocellulose membranes enables optimization of eCLIP by removing substantial background contamination. Our results identify (B) ThermoFisher iBlot and (C) GE Amersham Protran Premium (1060008) membranes as options which show a dramatic decrease in contamination for sensitive eCLIP experiments without altering true library yield. Although yielding equally low contamination, (D) GE Amersham Protran Premium (13600117) is "trial-size" packaging and generally commercially unavailable for large-scale use. Other sources can be tested using the RT-qPCR method described here to determine whether they are of sufficiently low background for use in eCLIP.

26. 5% milk + TBST (1× TBS pH 7.4 + 0.05% Tween-20).
27. Rabbit TrueBlot HRP secondary antibody.
28. ECL Western Blotting detection assay.
29. Proteinase K.
30. Urea.
31. Acid Phenol/Chloroform/Isoamyalcohol pH 4.5.
32. Phase lock heavy 2 mL Tubes.
33. 100% Ethanol.
34. RNA Clean & Concentrator-5 Kit.
35. Dynabeads MyOne Silane.
36. RLT Buffer.
37. 5 M NaCl.
38. 10× Ligase Buffer with DTT.
39. 10× AffinityScript reverse transcriptase buffer.
40. AffinityScript reverse transcriptase.
41. dNTPs (25 mM each).
42. Exo-SAP-IT.
43. 0.5 M EDTA.
44. 1 M NaOH.
45. 1 M HCl.
46. 5 mM Tris−HCl pH 7.5.
47. 10 mM Tris−HCl pH 7.5.
48. Q5 or other high fidelity PCR Master Mix.
49. qPCR Master Mix.
50. Agencourt AMPure XP beads.
51. MinElute gel purification Kit.
52. D1000 DNA tape/reagent.

3. Contamination Assay

1. TRIzol®Reagent.

2. TRIzol®LS Reagent.

3. SuperScript II (200 unit/µL).

4. Primer Sequences

seCLIP:

1. InvRiL19: /5Phos/rArGrArArUrCrGrGrArArGrArGrCrArCrArCrGrUrC/3SpC3/
(Order 100 nmole RNA oligo, standard desalting; storage stock 200 µM; working stock 40 µM; final concentration 1 µM (input), 4 µM (CLIP)).

2. InvRand3Tr3: /5Phos/NNNNNNNNNNNAGATCGGAAGAGCGTCGTGT/3SpC3/
(Order 100 nmole DNA oligo, standard desalting; storage stock 200 µM; working stock 80 µM; final concentration 3 µM).

3. InvAR17: CAGACGTGTGCTCTTCCGA (25 nmole DNA oligo, standard desalting; storage stock 200 µM; working stock 20 µM; final concentration 0.5 µM).

4. D5x_qPCR: AATGATACGGCGACCACCGAGATCTACACTATAGCCTACACTCTTTCCCTACACGACGCTCTTCCGATCT.

5. D7x_qPCR: CAAGCAGAAGACGGCATACGAGATCGAGTAATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC.

XBB1 contamination primers:

1. XBB1_qPCR _F: GAGGCGGCAAATATCCTGTG.

2. XBB1_qPCR_R: GTTTCACTTCCCCTCGTTCG.

Anneal Adapter	3m

1   Take 📋**5 µl RNA** (from <u>Dephosphorylation of Input RNA</u>).

2   Add 📋**1.5 µl 100% DMSO** and 📋**0.5 µl InvRiL19 adapter** (*see***Note 3**).

3   🔲   3m

Incubate at ❄ **65 °C** for ⏱**00:02:00** . Place ❄ **On ice > ⏱00:01:00** .

Prepare Ligation Master Mix; 13.5 µL Per Sample	1h 15m

4

A	B
10× Ligase Buffer (with DTT)	2.0 µL
0.1 M ATP	0.2 µL
Murine RNase Inhibitor	0.2 µL
100% DMSO	0.3 µL
50% PEG 8000	8.0 µL
RNA Ligase high conc	1.3 µL
H2O	1.5 µL

5   Add 📋**13.5 µl Ligation master mix** to each sample, mix, incubate at ❄ **Room temperature** for ⏱**01:15:00** .   1h 30m

Flick to mix every ~ ⏱**00:15:00** .

Silane Cleanup Input RNA	26m

6   **Prepare beads:**

6.1   Magnetically separate 📋**20 µl MyONE Silane beads** per sample, remove the supernatant.

6.2   🧪

Wash 1× with 🟦**900 μl RLT buffer** and resuspend beads in 🟦**61.6 μl RLT buffer** .

7   Bind RNA:

7.1

Add beads in 🟦**61.6 μl RLT buffer** to sample, mix and add 🟦**61.6 μl 100% EtOH** .

7.2   18m

Pipette mix, leave pipette tip in a tube, pipette mix every ~ 🕒**00:03:00** −5 min for 🕒**00:15:00** .

8   Wash beads:

8.1   Magnetically separate, remove the supernatant.

8.2   Add 🟦**1 mL 75% EtOH** , pipette resuspend and move to a **new tube.**

8.3   1m

After 🕒**00:00:30** , magnetically separate, remove the supernatant, and wash 2× with **75% EtOH** (

🕒**00:00:30**).

8.4   Magnetically separate, remove residual liquid with fine tip → Air-dry 🕒**00:05:00** .   5m

9   Elute RNA:

9.1   Resuspend in 🟦**10 μl H2O** , let it sit for 🕒**00:05:00** .   5m

9.2   Magnetically separate, transfer the supernatant to a new tube.