CellPress

Molecular Cell
Technology Preview

Easier, Better, Faster, Stronger: Improved Methods for RNA-Protein Interaction Studies

Nazmul Haque[1] and J. Robert Hogg[1,*]
[1]Biochemistry and Biophysics Center, National Heart, Lung, and Blood Institute, NIH, Bethesda, MD 20892, USA
*Correspondence: j.hogg@nih.gov
http://dx.doi.org/10.1016/j.molcel.2016.05.019

The RNA field has been revolutionized by methods that allow genome-scale identification of RNA-protein interaction sites. Two reports now introduce more efficient approaches, opening the technology to wider adoption (Van Nostrand et al., 2016; Zarnegar et al., 2016).

Methods combining UV-crosslinking and RNA binding protein (RBP) immunoprecipitation with high-throughput sequencing (CLIP-seq) promise to allow comprehensive high-resolution identification of all of the RNA binding sites occupied by any RBP of interest. In CLIP-seq experiments, cells are irradiated with ultraviolet light to covalently trap RNA-protein interactions, followed by immunopurification of a specific RBP (Figure 1). RNAs covalently linked to the target RBP are partially digested to generate short sequence tags suitable for cDNA library preparation and high-throughput sequencing. The resulting datasets offer a transcriptome-wide picture of RBP binding sites at up to single-nucleotide resolution. A series of variations on CLIP-seq have been used to evaluate the functions and mechanisms of dozens of RBPs (Flynn et al., 2015; Hafner et al., 2010; König et al., 2010; Licatalosi et al., 2008), but technical barriers to their universal application persist.

CLIP-seq methods are complex—protocols list 40 or more individual steps requiring several days to complete—and prone to failure. Despite methods calling for large numbers of input cells (typically tens of millions), it is often difficult to obtain sufficient material to generate high-complexity cDNA libraries for sequencing. To begin to address this problem, Zarnegar and colleagues first established a method to track RNA through a CLIP-seq experiment more easily. Whereas previous CLIP-seq protocols use 5′ radiolabeling to monitor RNAs through gel electrophoresis, adaptor ligation, and RNA purification steps, their method, infrared-CLIP (irCLIP), uses an oligonucleotide labeled with an infrared fluorescent dye for 3′-adaptor ligation (Zarnegar et al., 2016). The ligated product can then be quickly and sensitively detected at multiple points in the protocol. The authors use this system to optimize several aspects of the CLIP-seq workflow, including improving the fragmentation of immunopurified RNA and streamlining or eliminating RNA precipitation and purification steps. The benefits accrued from multiple points of optimization result in much lower required starting cell numbers than other CLIP-seq variants. While input requirements will ultimately depend on the abundance of the RBP of interest, its crosslinking efficiency to its cognate RNAs, and other factors, these improvements allowed productive sequencing of cDNA libraries from as few as 20,000 cells.

Among the most notable advances introduced in irCLIP is the use of thermostable group II intron reverse transcriptase (TGIRT) for cDNA synthesis. This enzyme exhibits a number of favorable properties when compared to widely used retroviral reverse transcriptases, including higher processivity, thermostability, and fidelity, as well as the ability to act on highly structured or modified RNA templates (Mohr et al., 2013). Simply adapting the library construction protocol to use the TGIRT enzyme led to an approximate 8-fold increase in cDNA production when compared to the optimized murine leukemia virus-derived Superscript III enzyme (Zarnegar et al., 2016). For difficult templates, this approach may yield even greater dividends.

Independently, van Nostrand and colleagues pursued a parallel path toward democratization of CLIP-based approaches. Their twist on the technique, "enhanced" CLIP (eCLIP), similarly promises tremendous gains in efficiency over previous methods. As with irCLIP, eCLIP involves streamlining of several steps of RNA and cDNA handling, all directed at minimizing loss of precious low-abundance material. Most importantly, eCLIP incorporates improved RNA sequencing (RNA-seq) library preparation methods to vastly increase the efficiency of the adaptor ligation steps required for reverse transcription and deep sequencing (Shishkin et al., 2015). Together, these enhancements lead to as much as a 1,000-fold decrease in the PCR amplification required to generate high-quality libraries for sequencing when compared to previous methods (Van Nostrand et al., 2016). As part of the ENCODE consortium, the authors have already used eCLIP to generate 102 datasets from 73 RBPs, illustrating its scalability and broad applicability.

Rather than the ideal of unbiased, comprehensive identification of RNA-protein interactions, current methods preferentially identify interactions between abundant RBPs and abundant RNAs. A major shortcoming in many existing CLIP-seq methods is the lack of controls to monitor non-specific background or to account for differences in abundance of distinct cellular RNAs in the starting material. Combined with the inefficiencies addressed by irCLIP and eCLIP, a lack of standardized normalization and background subtraction methods for data analysis results in overrepresentation of highly abundant RNAs in CLIP-seq datasets and can frequently lead to false positives. To address this problem, the eCLIP

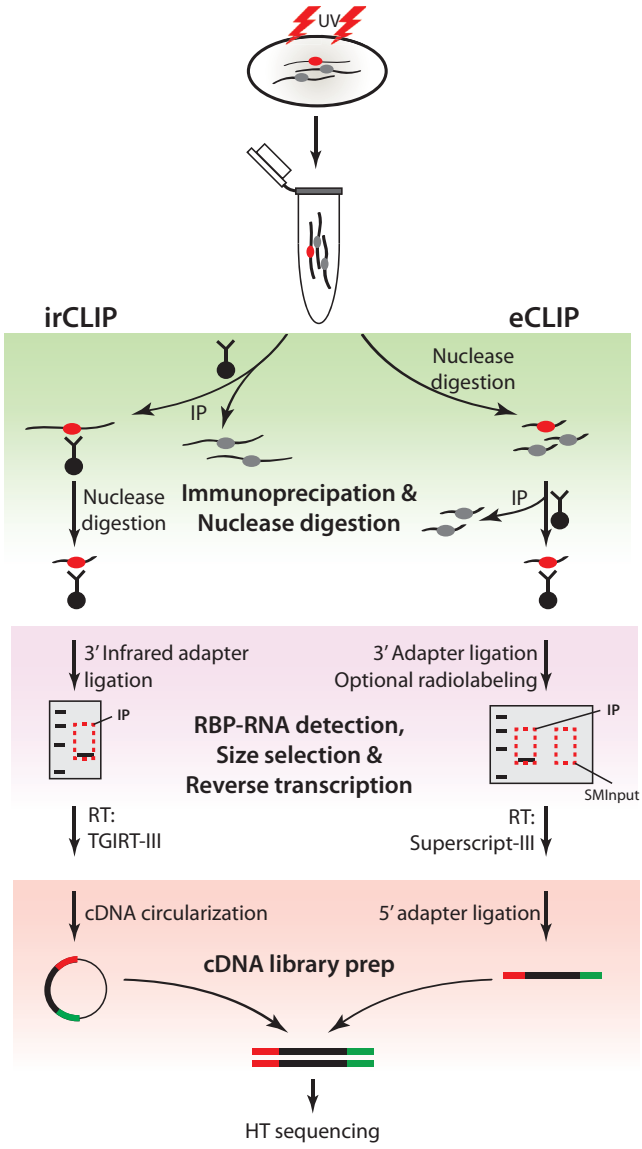

Figure 1. Schematic of CLIP-Seq Workflow and Major Modifications Introduced in irCLIP and eCLIP

Top, cells are UV irradiated to covalently link RNA-protein complexes, followed by lysis and RBP immunopurification. In-lysate nuclease digestion precedes RBP purification in eCLIP, while on-bead nuclease digestion is performed on immunopurified complexes in irCLIP. Ligation of an IR dye-labeled 3′ adaptor in irCLIP allows rapid detection of RNA-protein complexes, rather than immunoblotting or radiolabeling, as in eCLIP. Purified material is resolved by SDS-PAGE, transferred to nitrocellulose membranes, and size-selected in both methods, but eCLIP introduces purification and sequencing of size-matched RNA to allow normalization to input RNA levels (SMInput). In irCLIP, TGIRT-III reverse transcriptase is used to enhance cDNA synthesis, and a second adaptor-ligation step is omitted by circularization of cDNA. In contrast, cDNA generated in eCLIP is ligated to a second adaptor using optimized ligation methods. The products are then amplified by PCR and analyzed by high-throughput sequencing.

pipeline includes controls for normalization to input RNA abundance to aid in accurate interpretation of CLIP-seq data. Specifically, RNA from crude input extracts is fragmented and size-selected in parallel with immunopurified RNA. This input sample can then be used to test for significant enrichment of mRNA regions in CLIP-seq experiments relative to input samples. This addition has the potential to reduce false positives, aid in identifying interactions between RBPs and low-abundance RNAs, and improve reproducibility.

Notably, eCLIP and irCLIP take largely complementary approaches to CLIP-seq optimization. In addition to the immediate benefits available to RNA biologists, this suggests that future gains could be made by combining the lessons learned in each study. Together, they open the door to more routine use of CLIP-seq to study a wider range of RNA-protein interactions, in biological systems that go beyond the common transformed cell lines thus far used for most CLIP-seq experiments.

REFERENCES

Flynn, R.A., Martin, L., Spitale, R.C., Do, B.T., Sagan, S.M., Zarnegar, B., Qu, K., Khavari, P.A., Quake, S.R., Sarnow, P., and Chang, H.Y. (2015). RNA *21*, 135–143.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Jr., Ascano, M., Jr., Jungkamp, A.C., Munschauer, M., et al. (2010). Cell *141*, 129–141.

König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M., and Ule, J. (2010). Nat. Struct. Mol. Biol. *17*, 909–915.

Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X., et al. (2008). Nature *456*, 464–469.

Mohr, S., Ghanem, E., Smith, W., Sheeter, D., Qin, Y., King, O., Polioudakis, D., Iyer, V.R., Hunicke-Smith, S., Swamy, S., et al. (2013). RNA *19*, 958–970.

Shishkin, A.A., Giannoukos, G., Kucukural, A., Ciulla, D., Busby, M., Surka, C., Chen, J., Bhattacharyya, R.P., Rudy, R.F., Patel, M.M., et al. (2015). Nat. Methods *12*, 323–325.

Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K., et al. (2016). Nat. Methods. http://dx.doi.org/10.1038/nmeth.3810.

Zarnegar, B.J., Flynn, R.A., Shen, Y., Do, B.T., Chang, H.Y., and Khavari, P.A. (2016). Nat. Methods. http://dx.doi.org/10.1038/nmeth.3840.