Methods xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

CRISPR/Cas9-mediated integration enables TAG-eCLIP of endogenously tagged RNA binding proteins

Eric L. Van Nostrand [a,b,c,1], Chelsea Gelboin-Burkhart [a,b,c,1], Ruth Wang [a,b,c], Gabriel A. Pratt [a,b,c,d], Steven M. Blue [a,b,c], Gene W. Yeo [a,b,c,d,e,f,*]

[a] Department of Cellular and Molecular Medicine, University of California at San Diego, La Jolla, CA, USA
[b] Stem Cell Program, University of California at San Diego, La Jolla, CA, USA
[c] Institute for Genomic Medicine, University of California at San Diego, La Jolla, CA, USA
[d] Bioinformatics and Systems Biology Graduate Program, University of California San Diego, La Jolla, CA, USA
[e] Department of Physiology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore
[f] Molecular Engineering Laboratory, A*STAR, Singapore

ARTICLE INFO

Article history:
Received 27 August 2016
Received in revised form 8 December 2016
Accepted 10 December 2016
Available online xxxx

Keywords:
CLIP-seq
eCLIP
RNA binding protein
CRISPR/Cas9
Protein tagging

ABSTRACT

Identification of *in vivo* direct RNA targets for RNA binding proteins (RBPs) provides critical insight into their regulatory activities and mechanisms. Recently, we described a methodology for enhanced crosslinking and immunoprecipitation followed by high-throughput sequencing (eCLIP) using antibodies against endogenous RNA binding proteins. However, in many cases it is desirable to profile targets of an RNA binding protein for which an immunoprecipitation-grade antibody is lacking. Here we describe a scalable method for using CRISPR/Cas9-mediated homologous recombination to insert a peptide tag into the endogenous RNA binding protein locus. Further, we show that TAG-eCLIP performed using tag-specific antibodies can yield the same robust binding profiles after proper control normalization as eCLIP with antibodies against endogenous proteins. Finally, we note that antibodies against commonly used tags can immunoprecipitate significant amounts of antibody-specific RNA, emphasizing the need for paired controls alongside each experiment for normalization. TAG-eCLIP enables eCLIP profiling of new native proteins where no suitable antibody exists, expanding the RBP-RNA interaction landscape.

1. Introduction

Believed previously to be a mere intermediary between DNA and protein, RNA is becoming increasingly appreciated as subject to a variety of post-transcriptional processing steps prior to translation [1]. Analogous to transcription factors and histones that interact with DNA, transcribed RNA is associated with RNA binding proteins (RBPs) which have numerous regulatory functions. These RBPs transport RNAs from the nucleus and throughout the cell, carry out splicing, regulate stabilization, degradation, and translation of RNAs, and form ribonucleoprotein complexes with non-coding RNAs to confer regulatory activity [1]. Recent work indicates that there are likely over a thousand RBPs encoded in the human genome that play a wide range of developmental roles, and mutation or dysfunction of numerous RBPs have been linked to a wide variety of defects including neurodegenerative and autoimmune diseases [1–4].

For an RBP of interest, identifying its binding sites *in vivo* is a critical step towards understanding its functions at the molecular and physiological level. The development of microarray and high-throughput sequencing technologies rapidly led to the development of RNA Immunoprecipitation (RIP) and Crosslinking and Immunoprecipitation (CLIP) methods to profile RNA binding protein target sites transcriptome-wide [5]. Initial RIP methods focused on profiling RBP targets at the transcript level, by pulling down an RBP and its bound RNA for quantification by microarray [6]. Building upon this work, CLIP utilizes crosslinking (typically with UV irradiation) to covalently couple the RBP to its RNA targets. With this irreversible and stable linkage, CLIP allows stringent wash conditions and an RNA fragmentation step to bring target identification from the kilobase transcript-level to clusters that are less than a hundred bases in length [5]. Further work improved

Abbreviations: RBP, RNA binding protein; eCLIP, enhanced crosslinking and immunoprecipitation followed by high-throughput sequencing.

* Corresponding author at: Department of Cellular and Molecular Medicine, University of California at San Diego, La Jolla, CA, USA.
E-mail address: geneyeo@ucsd.edu (G.W. Yeo).

[1] Contributed equally.

http://dx.doi.org/10.1016/j.ymeth.2016.12.007
1046-2023/

crosslinking efficiency through incorporation of the photoactivatable nucleoside analog 4-thiouridine into RNAs during transcription in living cells (PAR-CLIP) [7], and iCLIP described altered library preparation steps to improve efficiency and enable identification of binding sites with single-nucleotide resolution [8]. Recently, we developed an enhanced CLIP (eCLIP) method that builds upon these methods by dramatically improving the efficiency of converting immunoprecipitated RNA into an adapter-ligated and amplified sequencing library, enabling the incorporation of paired input samples to improve signal-to-noise in identifying true binding sites above common artifacts. The robust success of eCLIP enabled profiling of over one hundred RNA binding proteins in K562 and HepG2 cells, and has proven successful in a variety of other cell-types and tissues [9].

However, one major limitation for all RIP and CLIP methods is that they require antibodies for immunoprecipitation. Thus, to profile the targets of an RBP under study, one must first screen through expensive antibodies, oftentimes with irregular success and high levels of background. In many other cases no suitable commercially available antibody yet exists for the RBP of interest, thus requiring custom generation at high cost. To help address this concern, we recently performed a large-scale effort to identify antibodies that could successfully immunoprecipitate RBPs in K562 cells, identifying antibodies for 365 RBPs [10]. Although this was highly successful, hundreds of RBPs remain without antibodies suitable for immunoprecipitation. Additionally, the concern that each antibody may have its own individual off-target or background interactions would be alleviated if all experiments were performed using the same antibody.

One common solution to the lack of suitable antibodies is to utilize peptide tags which already have high-quality, immunoprecipitation-grade antibodies. Most commonly, the protein of interest, flanked by either N- or C-terminal tags is exogenously expressed and the tag is used to immunoprecipitate the protein of interest along with its interactors [11]. Numerous such tags exist, including the well-characterized V5 and FLAG tags, which have proven successful in a variety of experimental regimes [12,13]. However, over-expression of various DNA- or RNA-binding proteins has sometimes revealed amplified binding to the same targets and other times led to interactions with ectopic or low-affinity sites, complicating interpretation of large-scale over-expression experiments [14,15].

The recent development of CRISPR technologies has made it possible to rapidly and successfully insert these tags into endogenous gene loci [16–18], which enables profiling of RBPs within their normal regulatory context. A recent method to perform endogenous tagging followed by ChIP-seq (CETCh-seq) demonstrated successful use of the CRISPR-Cas9 system to introduce a 3xFLAG tag at the 3′ end of transcription factors [19]. Specifically, ChIP-seq using the FLAG tag yielded substantially similar binding site identification to parallel experiments performed with antibodies targeting native proteins, confirming this approach as a general scheme for profiling DNA binding proteins lacking antibodies.

Here, we describe a scalable methodology for performing and validating CRISPR-mediated tag insertion into RNA binding protein loci. Using two tags (V5 and FLAG), we show that TAG-eCLIP yields the same high-quality target identification as eCLIP with native antibodies. Furthermore, we characterize common non-specific background identified by anti-V5 and anti-FLAG antibodies in wild-type cells, which indicates that such TAG-eCLIP experiments require proper controls for robust analysis. These methods provide further improvements to simpler, more cost-effective RBP target identification in cases where high-quality antibodies do not currently exist.

2. Methods

2.1. Cloning of CRISPR/Cas9 sgRNA vectors

The 100 nt sequence centered on the annotated stop codon was obtained for each desired transcript. sgRNA sequences targeting the 3′ end of the RBP of interest were identified using the Zhang lab CRISPR design tool (available at http://crispr.mit.edu). The sgRNA sequences that were closest to the stop codon, but had maximal score (minimal predicted off-targets), were selected. Two methods were tested for different RBPs: using a single double-strand nuclease Cas9 (pX330-U6-Chimeric_BB-CBh-hSpCas9; Addgene plasmid # 42230, pSpCas9(BB)-2A-GFP (PX458); Addgene plasmid # 48138 and pSpCas9(BB)-2A-Puro (PX459); Addgene plasmid # 62988 were a gift from Feng Zhang), or using a pair of single-strand nickase mutant Cas9 vectors (pX335-U6-Chimeric_BB-CBh-hSpCas9n (D10A); Addgene plasmid # 42335 was a gift from Feng Zhang). For nickase experiments, the pair of sgRNAs that flanked the stop codon with the highest combined score (fewest predicted off-targets) was chosen (Fig. 1B). Cloning was performed by gel extraction of the BbsI-cut backbone, and ligation with phosphorylated oligonucleotides, as previously described [16].

2.2. Cloning of homology-directed repair (HDR) donor vectors

For chosen RBPs, the ~800 nt regions immediately upstream (5′ homology arm) and downstream (3′ homology arm) of the stop codon were computationally identified. The forward primer for the 5′ arm and reverse primer for the 3′ arm were selected using Primer3 (http://bioinfo.ut.ee/primer3-0.4.0/) to be ~700 nt away from the stop codon. This 700–800 nt homology arm size was chosen based on standard recommendations in the field (https://www.addgene.org/crispr/zhang/faq/). The reverse primer for the 5′ arm and forward primer for the 3′ arm were selected by starting at the base flanking the stop codon, and taking the smallest region (20–28 nt long) with a melting temperature >57 °C. Homology tails for Gibson assembly were added in two PCR steps. First, a short extension was added to the 5′ end of the gene-specific primers as follows (see Supplemental Table 1 for gene-specific primers used):

PCR_5_F: CGACGGCCAGTG - gene-specific primer
PCR_5_R: GGCTTACCGAATTC - gene-specific primer (starts at base before stop codon)
PCR_3_F: CTAGATCGGATCC - gene-specific primer (starts at base after stop codon)
PCR_3_R: GCATGCAGTCGA - gene-specific primer.

The first PCR amplification was performed using Phusion polymerase (NEB) on human genomic DNA (gDNA) with 38 cycles of amplification, with 2% DMSO added to aid amplification. After agarose gel extraction (Qiagen) of the specific product, a second PCR was performed (NEB Q5; 6 cycles of amplification at 45 °C followed by 6 cycles at 62 °C) using the following primers to add full homology tails:

2ndPCR_5_L:
GGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTG
2ndPCR_5_R:
CGAGACCGAGGAGAGGGTTAGGGATAGGCTTACCGAATTC
2ndPCR_3_L:
TATCACGTAAGTAGAACATGAAATAACCTAGATCGGATCC
2ndPCR_3_R:
CTGCCTTGGGAAAAGCGCCTCCCCTACCCGCATGCAGTCGA.

This second PCR product was prepared for Gibson assembly by PCR cleanup kit (Qiagen).