Methods 118–119 (2017) 16–23

Contents lists available at ScienceDirect

Methods

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

DO-RIP-seq to quantify RNA binding sites transcriptome-wide

Cindo O. Nicholson [a,b], Matthew B. Friedersdorf [a,b], Laura S. Bisogno [a], Jack D. Keene [a,b,*]

[a] Department of Molecular Genetics & Microbiology, USA
[b] Duke University Medical Center, Durham, NC 27710, USA

ARTICLE INFO

Article history:
Received 10 August 2016
Received in revised form 21 October 2016
Accepted 7 November 2016
Available online 10 November 2016

Keywords:
DO-RIP-seq
RNA-binding proteins
RNA-binding sites
Whole-transcript targets
Quantitation

ABSTRACT

Post-transcriptional processes orchestrate gene expression through dynamic protein-RNA interactions. These interactions occur at specific sites determined by RNA sequence, secondary structure, or nucleotide modifications. Methods have been developed either to quantify binding of whole transcripts or to identify the binding sites, but there is none proven to quantify binding at both the whole transcript and binding site levels. Here we describe digestion optimized RNA immunoprecipitation with deep sequencing (DO-RIP-seq) as a method that quantitates at the whole transcript target (RIP-Seq-Like or RSL) level and at the binding site level (BSL) using continuous metrics. DO-RIP-seq methodology was developed using the RBP HuR/ELAVL1 as a test case (Nicholson et al., 2016). DO-RIP-seq employs treatment of cell lysates with a nuclease under optimized conditions to yield partially digested RNA fragments bound by RNA binding proteins, followed by immunoprecipitations that capture the digested RNA-protein complexes and assess non-specific or background interactions. Analyses of sequenced cDNA libraries made from the bound RNA fragments yielded two types of enrichment scores; one for RSL binding events and the other for BSL events (Nicholson et al., 2016). These analyses plus the extensive read coverage of DO-RIP-seq allows seamless integration of binding site and whole transcript information. Therefore, DO-RIP-seq is useful for quantifying RBP binding events that are regulated during dynamic biological processes.

1. Introduction

RNA-protein interactions mark sites and regions of post-transcriptional regulation of gene expression. RNA-binding proteins (RBPs) and noncoding RNAs govern the functions of mRNAs by binding to specific RNA sequences, structures or chemical modifications. Knowledge of precise RNA binding sites can illuminate combinatorial interactions between RBPs and non-coding RNA, the determinants of ribonucleoprotein assembly, and functional outcomes at the whole transcript level. Therefore, it is crucial to develop methods that quantitatively identify the relative binding strength of mRNA targets of RBPs in the cell across the transcriptome.

The first transcriptome wide RIP (RNA-binding protein immunoprecipitation) method, RIP-Chip, was originally developed in our lab to identify RNAs targeted by a RBP [1] and has since been modified, refined and applied by numerous labs to a wide variety of problems (reviewed in [2]). Our RIP method calculates probabilities of RBP-RNA association by normalizing the abundance of RNA isolated by immunoprecipitation (IP) of a RBP to the abundance in the negative IP (or background). The probability scores that result can be continuous metrics of RBP-RNA association that quantify changes in the association that result in RNP remodeling during dynamic biological conditions [1,3–6]. However, RIP does not yield the binding sites of the RBP along the target RNAs.

Global RIP-crosslinking procedures whereby RBPs and RNAs are crosslinked in cells were also invented in our lab [7] and applied by Darnell and coworkers to identify mRNA targets [8]. The identification of binding sites was achieved by using biochemical and computational adaptations to these crosslinking immunoprecipitation (CLIP) procedures [9–12]. However, the CLIP method as initially practiced did not integrate background-binding measurements into the identification of binding sites. This is because it was assumed that the stringent conditions used to separate crosslinked protein-RNA complexes would remove all background. It has since been shown that considerable crosslinking background remains in CLIP preparations, but the background can be used to improve "binding site calling" and reduces false positives when properly applied [13]. However, despite the incorporation of background binding measurements, CLIP procedures are unable to quantify binding events largely because of the low efficiency of UV

* Corresponding author at: Duke University Medical Center, Durham, NC 27710, USA.
E-mail addresses: con6@duke.edu (C.O. Nicholson), matthew.friedersdorf@duke.edu (M.B. Friedersdorf), les36@duke.edu (L.S. Bisogno), jack.keene@dm.duke.edu (J.D. Keene).

http://dx.doi.org/10.1016/j.ymeth.2016.11.004
1046-2023/

crosslinking [14]. Nevertheless, RIP and CLIP can provide useful complementary information despite their differences [15–24], and therefore, we sought to develop a procedure that uses aspects of both protocols to quantify whole transcript targets as well as binding sites.

Here we outline the DO-RIP-seq protocol for transcriptome-wide quantification of RNA binding sites for RBPs. DO-RIP-seq melds features of both RIP and CLIP procedures to measure enrichment scores for binding to both whole transcripts and binding sites. This is achieved by treating protein-RNA complexes in cell lysates with nucleases; in this case, micrococcal nuclease under conditions optimized for partial digestion (Fig. 1). This step is followed by immunoprecipitations with antibodies against the RBP of interest, while using non-specific antibodies in parallel to measure background and account for overall transcript abundance. As described here, DO-RIP-seq was developed using the RBP ELAVL1/HuR as a test case, and the experiments yielded probabilistic measures (log of odds scores, LOD) for binding sites in HEK293 cells [25], and for XIST long non-coding RNA in mouse trophoblasts [26]. HuR binding site LOD scores correlated with binding strength and motif preference [25]. Also DO-RIP-seq analysis of whole transcript association distinguished functional groups of messages and enriched gene sets [25]. In addition, we have used DO-RIP-seq to successfully identify the binding sites of other RBPs, e.g. RBM38, CELF1, and TRA2B (*unpublished works in progress*). Therefore, DO-RIP-seq provides continuous metrics of RBP targeting at the resolution of the whole transcript and the binding site, and thus, allows one to connect cumulative binding site metrics to functional outcomes at the whole transcript level. Note that while our presentation of the protocol is exceedingly detailed, the procedure is relatively straightforward, rapid, and user friendly.

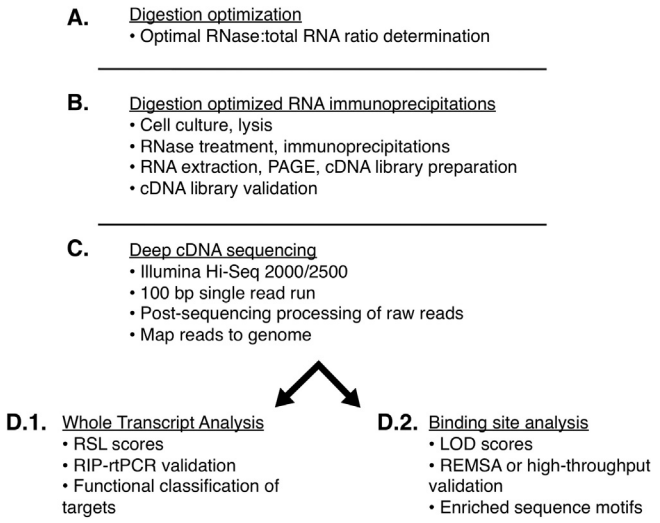

Fig. 1. The DO-RIP-seq workflow. A. Determine the optimal quantitative ratio of RNase to total RNA for obtaining protein-bound RNA fragments for mapping binding sites. B. Digestion optimized RNA immunoprecipitations of protein-bound and non-specifically bound (background) RNA fragments, and preparation of cDNA libraries from the extracted RNA. C. Sequencing of cDNA libraries on Illumina Hi-Seq 2000/2500 using 100 base pair (bp) single read runs. Processing of raw sequenced reads to remove oligonucleotide adapters and unique molecular identifiers, and then mapping to the appropriate species genome. D.1. Whole transcript analysis to calculate enrichment scores (RSL, RIP-seq-like) for expressed genes. RSL scores can be validated by RIP-rtPCR, and used as criteria for functional classification of targets. D.2. Binding site analysis to generate log of odds (LOD) scores which quantifies the probability of a site being bound by the protein relative to the site being in the background. Binding sites can be validated using REMSA or compared to *in vitro* high-throughput data. Also the binding sites can be used to discover enriched sequence motifs.

2. Materials & methods

2.1. Buffers

2.1.1. Polysome lysis buffer

Prepare polysome lysis buffer with the following components in distilled, nuclease-free water and store it at 4 °C:

10 mM HEPES pH 7.0
100 mM KCl
5 mM $MgCl_2$
5 mM $CaCl_2$
0.5% (v/v) IGEPAL CA-630

Add the following components to the polysome lysis buffer when cells are ready for harvesting:

1 mM Dithiothreitol (DTT)
1X cOmplete™ protease inhibitor (Roche)
100 Units/ml RNaseOUT (Thermo Fisher Scientific)

If necessary greater lysis of certain cell types can be achieved by keeping the magnesium and calcium salts out of the lysis buffer. Magnesium is believed to have a stabilizing effect on membranes through electrostatic interactions with the negatively charged groups of the membranes [27]. We have found that leaving both magnesium and calcium out of the lysis buffers increases lysis efficiency for some cell types and this should be empirically determined in each case (not shown). However, these salts should be added to the lysates before treating the lysates with micrococcal nuclease. Micrococcal nuclease requires Ca^{2+} for activity [28], and Mg^{2+} is important for stabilizing RNA structures [29].

2.1.2. NT2 buffer

Prepare NT2 buffer in distilled, nuclease-free water using the following components and store it at 4 °C:

50 mM Tris-HCl pH 7.4
1 mM $MgCl_2$
150 mM NaCl
0.05% (v/v) IGEPAL

2.2. Cell culture and lysate preparation

A single DO-RIP-seq experiment will require enough cell lysate for at least two immunoprecipitations (IPs); one IP with antibodies against the RBP of interest, and another using non-specific antibodies to measure background. The non-specific antibody we prefer is normal serum, for example normal mouse serum (Jackson ImmunoResearch Laboratories, cat. No. 015-000-001, see Section 3.3). Normal serum from mouse is used as a negative control when the antibody used to immunoprecipitate the RBP is from mouse as well. Therefore, antibodies used to immunoprecipitate the RBP and to perform negative control IPs should be from matching species. The number of cells required for DO-RIP-seq experiments will depend on the abundance of the RBP in the lysate. We recommend starting with up to five 15-cm dishes of cells that are 80–90% confluent (approximately 12×10^6 cells per dish for HEK293 cell line) for each IP if possible. In our experience one 15-cm dish of HEK293 cells per IP is sufficient for DO-RIP-seq experiments done with antibodies against endogenous HuR/ELAVL1. While these amounts are ideal, smaller amounts have been used successfully in other cases.

Harvest cells by first removing the culture media from dish of cells, adding 2 ml of cold 1X PBS, and then scraping the cells from