Title

Halo-seq: An RNA Proximity Labeling Method for the Isolation and Analysis of Subcellular RNA Populations

Journal

Current Protocols, 2(5)

ISSN

2691-1299

Authors

Lo, Hei-Yong G
Engel, Krysta L
Goering, Raeann
et al.

Publication Date

2022-05-01

DOI

10.1002/cpz1.424

Peer reviewed

HHS Public Access

Author manuscript

Curr Protoc. Author manuscript; available in PMC 2023 May 01.

Published in final edited form as:

Curr Protoc. 2022 May ; 2(5): e424. doi:10.1002/cpz1.424.

Halo-seq: an RNA proximity labeling method for the isolation and analysis of subcellular RNA populations

Hei-Yong G. Lo[1,2], **Krysta L. Engel**[1], **Raeann Goering**[1,2], **Ying Li**[3], **Robert C. Spitale**[4,5,*], **J. Matthew Taliaferro**[1,2,*]

[1]Department of Biochemistry and Molecular Genetics, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

[2]RNA Bioscience Initiative, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

[3]Department of Chemistry, Hong Kong University

[4]Department of Pharmaceutical Sciences, University of California Irvine, Irvine, CA, USA

[5]Department of Chemistry, University of California Irvine, Irvine, CA, USA

Abstract

Across a variety of species and cell types, the subcellular localization of specific RNA molecules promotes localized cellular activity. The misregulation of this RNA targeting can result in developmental defects, and mutations in proteins that regulate this process are associated with multiple diseases. For the vast majority of localized RNAs, however, the mechanisms that underlie their subcellular targeting are unknown, partly due to the difficulty associated with profiling and quantifying subcellular RNA populations. To address this challenge, we developed Halo-seq, a proximity labeling technique that can label and profile the local RNA content at virtually any subcellular location. Halo-seq relies on a HaloTag-fusion protein localized to a subcellular space of interest. Through the use of a radical-producing Halo ligand, RNAs that are near the HaloTag-fusion are specifically labeled, with spatial and temporal control. Labeled RNA is then specifically biotinylated *in vitro* via a “Click” reaction, facilitating its purification from the bulk RNA sample with streptavidin beads. The content of the biotinylated RNA sample is then profiled using high-throughput sequencing. In this manuscript, we describe the experimental and computational procedures for Halo-seq, including important benchmarking and quality control steps. By allowing the flexible profiling of a variety of subcellular RNA populations, we envision Halo-seq facilitating the discovery and further study of RNA localization regulatory mechanisms.

Basic Protocol 1: Visualization of HaloTag fusion protein localization

Basic Protocol 2: In situ Copper-catalyzed cycloaddition of fluorophore via “Click” Reaction

Basic Protocol 3: In vivo RNA alkynylation and RNA extraction

Basic Protocol 4: *In vitro* Copper-catalyzed cycloaddition of biotin via “Click” Reaction

*Address correspondence to: rspitale@uci.edu and matthew.taliaferro@cuanschutz.edu.

CONFLICT OF INTEREST STATEMENT

The authors do not declare any conflicts of interest.

Keywords

RNA localization; proximity labeling; Click chemistry

INTRODUCTION

Essentially all cells contain within them spatially defined regions with specific functions. These functions are often associated with the presence of particular proteins. An efficient sorting process is therefore needed to get proteins to their required destination in order to support localized cellular activities. For many genes, this fundamental problem is solved by transporting the corresponding RNA rather than the protein to the site of interest. The RNA can then be translated on-site to immediately produce a correctly localized protein. For the vast majority of localized RNAs, however, the mechanisms that underlie their trafficking, and the phenotypic effects of their mislocalization, remain unknown.

Many studies have demonstrated that thousands of RNAs are asymmetrically distributed in a variety of cell types (Cajigas et al. 2012; Taliaferro et al. 2016; Moor et al. 2017; Lécuyer et al. 2007). Generally, these studies have interrogated cell types whose polarity, size, and morphology make them well-suited to some kind of subcellular fractionation. For example, the extended morphology of neurons facilitate their mechanical fractionation into cell body and neurite fractions (Zivraj et al. 2010; Arora et al. 2021; Goering et al. 2020). Similarly, the regular polarity of epithelial monolayers make them good candidates for techniques like laser capture microdissection (Moor et al. 2017; Cassella and Ephrussi 2021). Many cell types, however, do not have these properties, and the study of RNA localization within them has, therefore, been challenging.

Proximity labeling techniques have the promise of facilitating the isolation of biomolecules from precisely defined subcellular locations. These approaches generally make use of a genetically tagged protein with a known, well-defined localization (Hung et al. 2016). Molecules in close proximity to this bait protein (typically within 20–200 nm) are specifically labeled through the activity of the protein tag, allowing their subsequent purification. In the past, such techniques have been used primarily with a focus on the identification of localized proteins (Hung et al. 2016; Mair et al. 2019; Choi-Rhee, Schulman, and Cronan 2004; Roux et al. 2012). However, a handful of approaches have recently been developed that apply the logic of proximity labeling also to RNAs (Fazal et al. 2019; Padrón, Iwasaki, and Ingolia 2019; P. Wang et al. 2019; Engel et al. 2021). In each case, RNA molecules near a bait protein are biotinylated, facilitating their purification and quantification. Using these techniques, the primary question of RNA localization, i.e. “which RNAs are localized to a given subcellular location?”, is now approachable regardless of the subcellular location and cell type of interest.

In this protocol article, we lay out the steps for performing one of these proximity labeling techniques, Halo-seq (Halo-mediated localization Sequencing; Figure 1) (Engel et al. 2021) to identify the RNAs present at essentially any subcellular location. Halo-seq begins with the identification of a protein that is a specific marker of the subcellular location of interest. This protein is then genetically fused to a HaloTag domain (Los et al. 2008). HaloTags are small domains that specifically and covalently bind Halo ligands. If a Halo ligand is added to cells that contain a specifically localized Halo fusion protein, the ligand will therefore be similarly spatially restricted. The Halo ligand used during Halo-seq is dibromofluorescein (DBF). DBF is a small molecule fluorophore that produces highly reactive oxygen species when exposed to green light (Li et al. 2017, 2018). Due to their reactivity, these oxygen species are restricted to within approximately 100 nm of their DBF source. They oxidize nearby biomolecules, including RNA and proteins, making them susceptible to nucleophilic attack by an added alkyne-containing amine, propargylamine. RNA and protein molecules near the HaloTag fusion protein can, therefore, be specifically alkynylated. This alkynylation is temporally controlled with exposure to green light. Importantly, these steps occur while the cell is alive and intact, maximizing chances that alkynylated biomolecules are spatially coincident with the region of interest. Total RNA is then extracted from cells. To isolate localized RNA, alkynylated RNA is biotinylated in an *in vitro* copper “Click” reaction, allowing its enrichment with streptavidin beads. The content of the biotinylated RNA sample is then profiled using high-throughput sequencing, and localized RNAs are identified as those enriched in the biotinylated sample compared to a control sample (Figure 1). The identity of this control is discussed in detail in the protocol. Importantly, Halo-seq is not biased towards any transcript type, preserves isoform information, and is temporally and spatially controlled.

Here, we first detail how to accurately assess the localization and specificity of the HaloTag fusion (Basic Protocol 1) and use an *in situ* Click reaction to directly visualize alkynylated RNAs/proteins in their native subcellular location (Basic Protocol 2). These two protocols are critical to determine key factors in the spatial restriction of labeling. In Basic Protocol 3, we detail how to utilize Halo-Seq for proximity labeling and subsequent purification of labeled RNA in HeLa cells, though we have successfully used this technique for other cell types as well. This protocol describes the in-cell labeling reaction that results in the alkynylation of RNAs at a specific subcellular location. Then, in Basic Protocol 4, we describe the steps for biotinylating the alkynylated RNA *in vitro*, with the cycloaddition of biotin-azide. The biotinylated RNA is then visualized on an RNA dot blot, which is an important quality control check for both the in-cell labeling and the *in vitro* Click reaction (Basic Protocol 5). The pool of biotinylated RNA is then enriched via pulldown with streptavidin beads (Basic Protocol 6). This enriched RNA can then be profiled using high-throughput RNA sequencing. Finally, we detail how to go from raw high-throughput sequencing reads to deriving a list of transcripts localized to the region of interest, using commands in Unix and R (Basic Protocol 7). An overview of the entire workflow is shown in Figure 1.

STRATEGIC PLANNING

The goal of Halo-Seq is to label, visualize, extract, enrich, and sequence the RNA at a subcellular location of interest. In theory, this can be done in any model where there is expression and correct localization of a bait protein (i.e. the HaloTag fusion protein). We suggest starting with a basic cell culture model with rapidly proliferating cells to identify critical parameters that are required for successful labeling and sequencing. The Basic Protocols described here utilize HeLa cells expressing a doxycycline-inducible HaloTag fusion protein. Although not required, we suggest using a genome-integrated, inducible expression system to minimize cell-to-cell variation in expression and to allow for temporal control of expression (see Critical Parameters for more details).

The Basic Protocols require a cell line genetically engineered to express the HaloTag-protein fusion. For Basic Protocols 1 and 2, only ~4×10^6 cells are required, which is enough to seed 4 wells of a 12-well plate, each with ~1×10^6 cells. For Basic Protocol 3, where the user is setting up an experiment with the end goal of performing high-throughput sequencing, two conditions are required (see “Discussion on Controls and Comparisons” in Understanding Results), each in triplicate. Therefore, the user should expect to seed six 10-cm dishes at a minimum. At ~3×10^6 cells per dish, this equates to having ~20×10^6 cells prior to starting Basic Protocol 3. This number may differ depending on the cell type used or compartment labeled, since differing amounts of RNA extracted from Basic Protocol 3 may be needed (see Troubleshooting for more details).

BASIC PROTOCOL 1: VISUALIZATION OF HALOTAG FUSION PROTEIN LOCALIZATION

Halo-Seq relies on the labeling of RNA near a HaloTag fusion protein. Prior to labeling RNA (Basic Protocols 2 and 3), it is therefore critical to ensure that the fusion protein is correctly and specifically localized to the subcellular compartment of interest. This protocol allows verification of the proper localization of the HaloTag fusion using fluorescent Halo ligands and fluorescence microscopy. The example data shown in this protocol were derived from HeLa cells expressing a doxycycline-inducible HaloTag fusion. However, other cell lines and expression systems may be used. After completing this protocol, the user should know whether their HaloTag fusion is localized as expected, as well as the parameters required to induce expression and localization of the HaloTag fusion. These parameters will be used in Basic Protocols 2 and 3.

Materials:

Hardware and Instruments: 1.5-mL Eppendorf tubes (Eppendorf, cat. no. 022363204)

12-well Tissue Culture plate (Fablab, cat. no. FL7111)

Fluorescence microscope

Microscope slides (Globe Scientific, cat. no. 1380-20)

PDL Coverslips (NeuVitro, cat. no. GG-12-15-PDL)

Tissue culture plates and flasks

Tissue culture hood

Tweezers (Electron Microscopy Sciences, cat. no. 78326-51)

Reagents, Solutions, Starting Materials, Cell lines: Cells expressing a HaloTag protein fusion with the proper subcellular localization (see Strategic Planning)

Cell culture media (specific to cells being used)

DAPI (Sigma Aldrich, cat. no. D9542-1MG)

Hank's Balanced Salt Solution without calcium or phenol red (VWR, cat. no. VWRL0121-0500)

Doxycycline hydrochloride (Fisher Scientific, AAJ6042203v)

Formaldehyde (37% by weight) (Fisher Scientific, cat. no. BP531-500)

Fluoromount G (Southern Biotech, cat. no. 0100-01)

Janelia Fluor 549 or 646 (Torics, cat. no. 6147)

PBS (Invitrogen, cat. no. AM9625)

Protocol steps:

1. Place PDL coverslips in two separate wells of a 12-well tissue culture plate. Seed cells expressing Halo-protein fusion on the coverslips (~0.2–0.5 × 10^6 cells per well) with sufficient media to cover the cells (1 mL per well). Cells should be 80–95% confluent on the day of Halo ligand addition (step 4).

2. Induce expression of HaloTag fusion protein. For doxycycline-inducible HaloTag systems, we recommend starting by adding doxycycline to 1 μg / mL for at least 24 hr (often 48 hr) to drive sufficient expression of the fusion protein.

   The results of this protocol will determine the doxycycline concentration and duration required to drive sufficient expression and localization of the HaloTag-fusion for RNA labeling.

3. After induction of expression, wash cells with 1 mL of PBS per well and incubate at room temperature for 1 minute.

4. Remove PBS and add fluorescent Janelia Fluor halo ligand to cells (1 mL per well; 1:2000 dilution in HBSS from 100 μM stock). Incubate for 15 minutes in a cell culture incubator.

   We have had better results when adding the Halo ligand before fixation rather than after fixation.

5. Remove ligand-containing solution and wash cells with 1 mL of cell culture media per well. Incubate cells for 10 minutes. Repeat wash with cell culture media.

6. Remove media and wash once with 1 mL of PBS per well for 10 minutes.

7. Freshly prepare 3.7% formaldehyde fixation buffer (diluted from 37% stock in PBS). Fix cells with 1 mL of 3.7% formaldehyde fixation buffer per well and incubate at room temperature for 15 minutes.

8. Remove the fixation buffer and wash cells with 1 mL of PBS for 10 minutes at room temperature. Remove PBS and repeat wash.

9. (Optional) If co-staining with another antibody, proceed to primary antibody staining here.

10. Incubate cells with 500 μL of DAPI solution (100 ng/mL in PBS) per well for 10 minutes at room temperature in the dark.

11. Wash each well with 1 mL of PBS for 10 minutes.

12. Mount the coverslips on a microscope slide.

    a. Place 6 μL of fluoromount G on a microscope slide.

    b. Using tweezers, carefully remove the coverslip from the 12-well.

    c. With cells facing down, place the coverslip on the fluoromount G.

    d. Allow the mounted coverslip to dry for 10 minutes at room temperature. Store slides in 4°C in the dark.

       For best imaging results, image within 24 hours. Fluorescence may decrease even if kept at 4°C in the dark. We have kept slides for up to one week.

13. Image the cells.

    Imaging conditions will be dependent on individual lab set ups and the protein fusion of interest. For specific visualization of localization, we suggest imaging with no less than 60x oil immersion objectives. Example results can be found in Figure 2A.

    The user should now know whether the fusion protein is specifically localized to the desired subcellular location and, if applicable, the concentration and duration of doxycycline treatment required to express and localize the fusion protein. If the localization of the protein is not localized to the subcellular location of interest, see Troubleshooting. The derived concentration and duration of doxycycline treatment will be used again in Basic Protocol 2 and 3.

BASIC PROTOCOL 2: *IN SITU* COPPER-CATALYZED CYCLOADDITION OF FLUOROPHORE VIA "CLICK" REACTION

Although the results of Basic protocol 1 are informative as to the location of the HaloTag fusion, they do not necessarily report on the location of alkynylated molecules that result from a labeling experiment. In this protocol, and after Halo-DBF-mediated alkynylation, the location of alkynylated molecules (both RNA and protein) is visualized by performing a Click reaction *in situ* in fixed cells. In this reaction, instead of biotin-azide (as in a standard Halo-seq protocol), a fluorescent azide is used. Alkynylated molecules are therefore fluorescently labeled, and their subcellular location can be observed using fluorescence microscopy. Importantly, to be confident that alkynylated molecules are being visualized, this fluorescence must be dependent upon the addition of Halo-DBF to cells. Experimental cells will have the addition of Halo-DBF (allowing for alkynyation of RNA/protein around the HaloTag fusion protein; +Halo-DBF), and control cells will lack Halo-DBF (−Halo-DBF). Following this protocol, the user will know the subcellular location of alkynylated molecules following Halo-DBF-mediated labeling as well as the labeling time required to alkynylate them. This is key information for Basic Protocol 3.

Materials:

Hardware and Instruments: 2x Green LED flood light. (Specifications: AC: 85–265V, light source: 144pcs SMD2835 LED, power: 100W; manufacturer: T- SUNRISE, cat. no. B01N1S6D8K)

1.5-mL Eppendorf tubes (Eppendorf, cat. no. 022363204)

12-well Tissue Culture plate (Fablab, cat. no. FL7111)

Fluorescence microscope

Microscope slides (Globe Scientific, cat. no. 1380-20)

PDL Coverslips (NeuVitro, cat. no. GG-12-15-PDL)

Tissue culture plates and flasks

Tissue culture hood

Tweezers (Electron Microscopy Sciences, cat. no. 78326-51)

Reagents, Solutions, Starting Materials, Cell lines: 100% Ethanol (VWR, cat. no. EM-EX0276-3S)

(+)-Sodium L-ascorbate (Sigma Aldrich, cat. no. 11140-50G)

Blocking buffer (see Reagents & Solutions)

Cells expressing a HaloTag fusion protein with the proper subcellular localization (Strategic Planning and Basic Protocol 1)

Cell culture media specific to cells being used

Click blocking buffer (see **Reagents & Solutions**)

Click buffer A (see **Reagents & Solutions**)

Copper(II) sulfate pentahydrate (Fisher Scientific, cat. no. AC197722500)

Cy5 picolyl azide (Click Chemistry Tools, cat. no. 1177-1)

DAPI buffer (see Reagents & Solutions)

Halo-DBF (5mM stocks, dissolved in anhydrous DMSO and stored in the dark in −20°C). *Currently, the Halo-DBF ligand is not commercially available. Please follow the published procedure on its synthesis* (Li et al. 2017) *and contact Robert Spitale* (rspitale@uci.edu) *with any questions or for samples. We are currently working with potential vendors to supply the Halo-DBF ligand*.

DMSO (EMD millipore, cat. no MX1458-6)

Doxycycline hydrochloride (Fisher Scientific, AAJ6042203)

Fluoromount G (Southern Biotech, cat. no. 0100-01)

Hank's Balanced Salt Solution without calcium or phenol red (VWR, cat. no. VWRL0121-0500)

PBS without calcium, magnesium (Cytiva, cat. no. SH30256.01)

Proparyglamine (Sigma Aldrich, cat. no. P50900-1G)

THPTA (tris-hydroxypropyltriazolylmethylamine) (Click Chemistry Tools, cat. no. 1010-100)

Protocol steps:

1. Place PDL coverslips in two wells of a 12-well tissue culture plate. Seed cells expressing Halo-protein fusion on coverslips (~0.2–0.5 × 10^6 cells per well) with sufficient media to cover the cells (1 mL per well). Cells should be 80–95% confluent on the day of Halo ligand addition (step 4).
2. Induce expression of HaloTag fusion protein using the conditions for induction and localization from Basic Protocol 1 for all cells.
3. On the day of labeling, wash cells with 1 mL of PBS per well and incubate at room temperature for 1 minute.
4. Incubate one well (experimental sample) with 1 mL ligand buffer per well (1 μM Halo-DBF in HBSS) in a 37°C cell culture incubator. For the control sample, omit Halo-DBF and, instead, incubate cells with HBSS.

   From this point on, all samples should be kept in the dark when possible.

All steps should be performed in parallel for control (−Halo-DBF) and experimental (+Halo-DBF) cells.

5. Remove ligand buffer (or HBSS for controls) and wash with 1 mL of cell media per well for 10 minutes in a 37°C cell culture incubator.

6. Repeat wash with 1 mL of cell media per well for another 10 minutes in a 37°C cell culture incubator.

7. Incubate all cells with 1 mL of propargylamine buffer (1 mM propargylamine in HBSS) for 5 minutes in a 37°C incubator.

8. Irradiate cells with green light on the highest setting for 10 minutes in an incubator (Figure 2B).

   The results of this Basic Protocol will determine whether 10 minutes is enough for sufficient alkynylation in the user's cells. The optimal amount of labeling time is the shortest time that results in detectable alkynylation. In our experience, 10 minutes has been generally sufficient, so it is a good starting point.

   See Table 1 for recommended labeling times for some reference compartments.

9. Remove media and wash with 1mL of PBS per well for 5 minutes. Repeat wash two additional times.

10. Fix cells in each well with 1 mL of freshly prepared 3.7% formaldehyde (diluted from 37% in PBS) for 15 minutes at room temperature.

11. Remove fixation buffer and wash cells in each well with 1 mL of PBS for 10 minutes at room temperature. Remove PBS and repeat wash.

12. Prepare enough Click Blocking Buffer (1 mL per well) (**see Reagents and Solutions**).

13. Block each well with 1 mL of Click Blocking buffer at room temperature for 30 minutes.

14. Remove Click Blocking buffer and wash with 1 mL of PBS per well for 5 minutes at room temperature. Repeat wash two additional times.

15. Prepare fresh enough Click buffer A (100 μL per well) (see **Reagents and Solutions**).

16. Remove PBS and incubate cells with 100 μL of Click buffer A per well for 1 hr at 37°C in the dark.

17. Remove Click buffer A and wash with 1 mL of PBS at room temperature for 5 minutes. Repeat wash two additional times to completely remove Click buffer A.

18. Incubate cells with 500 μL of DAPI solution (100 ng/mL) per well for 10 minutes at room temperature in the dark.

19. Wash each well with 1 mL of PBS for 10 minutes.

20. Mount the coverslips on a microscope slide.

a. Place 6 μL of fluoromount G on a microscope slide.

b. Using tweezers, carefully remove the coverslip from the well.

c. With cells facing down, place the coverslip on the fluoromount G.

d. Allow the mounted coverslip to dry for 10 minutes at room temperature. Store slides in 4°C in the dark.

For best imaging results, image within 24 hours. Fluorescence may decrease even if kept at 4°C in the dark. We have kept slides for up to one week.

21. Image the cells..

Imaging conditions will be dependent on individual lab set ups and the protein fusion of interest. For proper visualization of localization, we suggest imaging with no less than 60x oil immersion objectives.

Be sure to use the same transmission and exposure time for each image in a given fluorescent channel for comparable images. Example results for this experiment can be found in Figure 2C.

The user should now know whether the labeling time of 10 minutes is sufficient to label RNA/protein around their compartment of interest. If the labeled RNA/protein is not visible or not specific, see Troubleshooting. The duration of the labeling time may need to be altered accordingly. The labeling time defined here will be used in Basic Protocol 3, step 9

BASIC PROTOCOL 3: *IN VIVO* RNA ALKYNYLATION AND RNA EXTRACTION

In this protocol, the user will induce expression of a HaloTag fusion protein and allow for its proper localization, and will then proceed with RNA alkynylation and extraction. In all subsequent Basic Protocols, users should have an experimental (labeled) sample and a control (unlabeled) sample. The experimental sample is differentiated from the control by the addition of Halo-DBF (Figure 1, see below). This allows confirmation that any observed biotinylation is due to the Halo-DBF-dependent alkynylation reaction. Since the final result is profiling using high-throughput sequencing, we recommend having at least three replicates for each condition to be tested. Prior to starting Basic Protocol 3, users should have already ensured that the fusion protein has the intended protein localization (Basic Protocol 1) and determined the proper labeling time for the specific compartment (Basic Protocol 2).

As in Basic Protocols 1 and 2, Basic Protocol 3 assumes the expression of a stably integrated, doxycyline-inducible fusion protein, although other transgenic expression systems are also likely usable. The experimental sample is then treated with the light sensitive ligand Halo-DBF, which will covalently attach to the HaloTag. This treatment is omitted from the control sample. After washing out excess Halo-DBF, the cells are treated with the alkyne-containing molecule propargylamine. Upon excitation with green

light, the propargylamine will nucleophilically attack Halo-DBF-proximal RNA molecules, resulting in their alkynylation. Subsequently, the cells are lysed, and total RNA is extracted using a standard Trizol RNA extraction. The labeled RNA is then ready for an *in vitro* copper-mediated cycloaddition of biotin-azide (Basic Protocol 4).

Materials:

Hardware and Instruments: 2x Green LED flood light. (Specifications: AC: 85–265V, light source: 144pcs SMD2835 LED, power: 100W; manufacturer: T- SUNRISE, cat. no. B01N1S6D8K)

1.5-mL Eppendorf tubes (Eppendorf, cat. no. 022363204)

Cell lifters (Fisher scientific, cat. no. 07-200-364)

Cell culture microscope

Luer Lock Syringes (1mL; Air-Tite, cat. no. ML-1)

Syringe (20G, 1.5 in; BD, cat. no. 305176)

Tissue culture hood

Tissue culture plates and flasks

Reagents, Solutions, Starting Materials, Cell lines: Cells expressing a HaloTag protein fusion with the proper subcellular localization (Basic Protocol 1 and Strategic Planning)

Cell culture media specific to cells of interest

Halo-DBF (5mM stocks, dissolved in anhydrous DMSO and stored in the dark in −20°C). *Currently, the Halo-DBF ligand is not commercially available. Please follow the published procedure on its synthesis* (Li et al. 2017) *and contact Robert Spitale* (rspitale@uci.edu) *with any questions or for samples. We are currently working with potential vendors to supply the Halo-DBF ligand.*

DMSO (EMD millipore, cat. no MX1458-6)

DNase I (New England Biolabs, cat. no. M0303S)

Doxycycline hydrochloride (Fisher Scientific, cat. no. AAJ6042203)

EtOH, 100% (VWR, cat. no. EM-EX0276-3S)

Hank’s Balanced Salt Solution without calcium or phenol red (VWR, cat. no. VWRL0121-0500)

Isopropyl alcohol (VWR, cat. no. bdh1133-4LP)

Nuclease-free water (Invitrogen, cat. no. AM9937)

PBS without calcium, magnesium (Cytiva, cat. no. SH30256.01)

Proparyglamine (Sigma Aldrich, cat. no. P50900-1G)

Trizol (Invitrogen, cat. no. 15596018)

Quick-RNA Miniprep Kit (Zymogen Research, cat. no. R1055)

Quick-RNA MidiPrep Kit (Zymogen Research, cat. no. R1056)

Protocol steps:

DBF stock solutions should be protected from water.

All incubation steps are performed in a humidified 37°C, 5% CO2 incubator unless otherwise specified.

RNA samples should be kept on ice unless otherwise specified.

1. Seed HaloTag fusion protein-expressing cells in 10-cm dishes such that they will be 80–95% confluent on the day of labeling (~3×10^6 cells per plate). Each condition (e.g. experimental (+Halo-DBF) vs control (−Halo-DBF)) should have 3 separate technical replicates, resulting in a total of 6 plates per experiment.

   Choosing the size of the culture dish will depend on the amount of total RNA desired from the labeling reaction. For example, for HeLa cells, one 10-cm dish is sufficient to recover 200–300 μg of total RNA. (See Troubleshooting section for more details).

2. Induce expression of the HaloTag-fusion protein in all cells using the conditions for induction and localization determined in Basic Protocol 1.

3. On the day of labeling, wash each 10-cm dish with 10 mL of PBS and incubate for 1 minute.

4. Incubate each experimental plate (+Halo-DBF) with 12 mL of ligand buffer (1 μM Halo-DBF in 10mL HBSS) and each control plate (−Halo-DBF) with 12 mL of HBSS in a 37°C cell incubator for 15 minutes.

   From this point on, all samples should be kept in the dark when possible.

   All steps should be performed in parallel for control and experimental cells.

5. Aspirate the ligand buffer and wash unbound Halo-DBF away by adding 10–12 mL of cell culture media. Do the same for the control wells. Incubate for 10 min at 37°C.

6. Repeat wash with an additional 10 mL of cell media and incubate for 10 min in a 37°C incubator.

   Note on labeling space: Plan ahead to make sure that there is sufficient space in the incubator for the irradiation of cells in the light fixture. In general, three

(3) 10-cm dishes will fit at once in one light fixture set up (see Figure 2B). If needed, stagger labeling times by stalling at this step.

7. Prepare 12 mL of propargylamine (PA) buffer (1 mM PA in HBSS) per dish.

8. Incubate each dish with 12 mL of PA buffer for 5 min in a 37°C incubator.

9. Irradiate cells with green light on the highest setting in an incubator for the experimentally determined amount of time in Basic Protocol 2. Place one light above and below the plates, in a sandwich configuration (Figure 2B), with the top light resting on the top of the dish with the lid off.

   Note on labeling time: The optimal labeling time should have been identified in Basic Protocol 2. We have found that, in general, 10 minutes is sufficient to label most subcellular compartments. See Table 1 for recommended labeling times.

10. Remove PA buffer from each plate and add 1 mL Trizol to each 10-cm plate. Incubate the cells in Trizol for at least one minute.

11. Scrape cell debris with a cell lifter and collect samples into separate, labeled 1.5-mL Eppendorf tubes.

12. Break up cell debris by triturating Trizol solution with a 20G needle 20 times.

    PAUSE STEP: The sample can be stored in Trizol at −20°C for up to a year.

13. Purify RNA following the Trizol manufacturer protocol with the following amendments:

    a. For greater purity, add an additional 70% EtOH wash step.

    b. Incubate on ice for 10 minutes for each wash step.

    c. Resuspend each RNA sample in 30 μL of RNase-free water.

       PAUSE STEP: RNA samples should be stored in −80°C for short and long-term storage.

14. Treat each RNA sample with DNase I for 30 minutes at 37°C following the manufacturer's instructions.

15. Recover the RNA using the Quick-RNA Miniprep Kit. Elute the RNA in an appropriate amount of RNase-free water in preparation for the Click-mediated biotinylation reaction (desired RNA concentration range is 800–1500 ng/μL).

    Ideally, you should have at least 100–200μg of RNA per condition (experimental and control). RNA should have a 260/280 of ≥2.0 and a 260/230 of ≥2.0.

    PAUSE STEP: RNA samples should be stored in −80°C for short and long-term storage.

BASIC PROTOCOL 4: *IN VITRO* COPPER-CATALYZED CYCLOADDITION OF BIOTIN VIA "CLICK" REACTION

Here, the user will biotinylate the alkynylated RNA generated in Basic Protocol 3. All 6 samples collected, regardless of condition (treatment or control), will be processed identically. The copper-catalyzed cycloaddition of biotin to labeled RNA (so-called "Click" reaction (Hein, Liu, and Wang 2008)) is critical for the enrichment of localized RNA. The "Click" reaction results in the fusion of molecules that contain alkyne and azide moieties. In this case, localized RNA has been alkynylated and will be reacted with biotin-azide. If successful, this reaction results in the biotinylation of alkynylated RNA while minimizing the background biotinylation of unalkynylated RNA. By the end of this protocol, the user should have biotinylated RNA for each sample tested. Some of the biotinylated RNA will subsequently be used to verify biotinylation (Basic Protocol 5), some of the RNA will be saved for high-throughput sequencing to control for transcriptome-wide differences between samples (Basic Protocol 7), and the rest of the biotinylated RNA will be enriched for via streptavidin bead pulldown (Basic Protocol 6).

Click chemistry is sensitive to numerous factors, including the concentration of copper, concentration of biotin, pH and temperature of the reaction, light, volume of the reaction, and overall reaction time (Hein, Liu, and Wang 2008). Therefore, great care should be placed in processing the experimental and control samples in parallel and identically. To reduce variability in technical replicates (3 per condition), we recommend doing all technical replicates at once. If infeasible, we recommend randomizing samples.

Materials:

Hardware and Instruments: Thermocycler

0.2-mL PCR tubes (Light Labs, cat. no. A-4003-A)

1.5-mL Eppendorf tubes (Eppendorf, cat. no. 022363204)

Reagents, Solutions, Starting Materials, Cell lines: Click buffer B (see Reagents & Solutions)

DNase-treated purified RNA from Basic Protocol 3

Nuclease-free Water (Invitrogen, cat. no. AM9937)

Quick-RNA Miniprep Kit (Zymogen Research, cat. no. R1055)

Quick-RNA MidiPrep Kit (Zymogen Research, cat. no. R1056)

Quick-RNA MicroPrep Kit (Zymogen Research, cat. no. R1050)

Protocol steps :

1. Prepare a separate Click buffer B mix for each sample, using the example table provided below (volumes provided for 1, 10, and 24 reactions; scale

accordingly). Prepare fresh at room temperature in the following order: Tris-Cl, Biotin-picolyl azide, Sodium ascorbate, THPTA, and Copper.

Once the copper is added, Click buffer B is considered active. Withholding the copper from Click buffer B is advised if more time is needed to set up multiple reactions.

Each Click reaction has a final volume of 50 µL, and contains 16.8 uL of Click Buffer B and 10 µg of RNA. So, as an example, to biotinylate 100 µg of RNA for one sample, prepare 168 µL of Click buffer B, for 10 separate reactions.

Click buffer B ingredients	Stock (mM)	Final Concentration (mM)	Volume for 1 reaction (µl)	Volume for 10 reactions (µl) +10%	Volume for 24 reactions (µl) +10%	Amount of Click buffer B added to each 50-µL reaction:
Tris-Cl	100	10	5	55	132	16.8
Biotin-Picolyl-Azide (in DMSO)	30	2	3.3	36.3	87.1	
NaAsc	500	10	1	11	26.4	
THPTA	40	2	2.5	22	87.1	
CuSO4	1	0.1	5	55	132	

2. Dilute 100 µg of alkynylated RNA from Basic Protocol 3 to 332 µL using nuclease-free water (final concentration 301 ng/µL). Add 168 µL of Click buffer B to each diluted RNA sample to create a Click mastermix. A separate mastermix must be made for each sample.

   Note on amount of RNA to use in Click reaction: The amount of input RNA to be biotinylated depends on the downstream application. For example, if you are only testing alkynylation and/or biotinylation efficiency (Basic Protocol 5), 25 µg of RNA will suffice, and the reaction can be scaled down accordingly. Samples for high-throughput sequencing, on the other hand, require around 100 µg to achieve the amount of RNA needed for streptavidin pulldown (Table 1).

3. Briefly vortex the Click mastermix. Redistribute the mastermix into ten 0.2-mL PCR strip tubes, with each tube containing 50 µL. Repeat for all samples and their respective Click mastermixes.

4. Incubate the PCR tubes for 30 minutes at 25°C in the dark using a thermocycler.

   See Table 1 for recommended time to incubate the reaction.

5. Once completed, immediately place the PCR tubes on ice to stall the reaction.

   Proceed immediately to the next step. The click reaction is still active at 0°C, but slowed.

6. Extract RNA using either a Zymo QuickRNA midiprep, miniprep, or microprep RNA column, depending on the amount of RNA in the reaction. Elute RNA to

a final concentration ≥ 1 μg/μL. For a 100 μg RNA sample, this corresponds to using a miniprep column and eluting in 80 μL nuclease-free water.

Other precipitation methods may work, but extra care should be put into removing the excess biotin. QuickRNA columns efficiently remove unreacted excess biotin-azide.

7. Check quality of RNA. Ideally, the user should have 80 μg-100 μg of RNA for each sample at a concentration of ≥ 1 μg/μL. The purity of the RNA should be high, with 260/280 ≥ 2.0 and 260/230 ≥ 2.0. The biotinylated RNA is now ready for validation (Basic Protocol 5) or enrichment with streptavidin beads (Basic Protocol 6).

   PAUSE STEP: RNA can be stored in −80°C for short and long-term storage.

8. Aliquot 1–2 μg of RNA and store the RNA at −80°C. This sample will serve as the pre-enrichment input control for RNA-seq comparisons (Basic Protocol 6, Step 16).

BASIC PROTOCOL 5: ASSESSMENT OFRNA BIOTINYLATION WITH RNA DOT BLOT

Here, the user will assay some of the biotinylated RNA in the sample after the *in vitro* Click reaction from Basic Protocol 4. While formally optional, this step is important for optimization and quality control purposes to ensure that proximity-dependent RNA biotinylation has occurred. To assay the amount of biotinylated RNA, equal amounts (in both volume and mass) of RNA from experimental (+Halo-DBF) and control (−Halo-DBF) samples are spotted on a nitrocellulose membrane and probed with streptavidin-HRP. Biotinylated RNA can then be visualized through the chemiluminescent activity of HRP. To ensure that equal amounts of RNA are assayed in the experimental and control samples, total RNA should be visualized using methylene blue. Following Basic Protocol 5, the user should know whether the *in cell* alkynylation (Basic Protocol 3) and *in vitro* biotinylation (Basic Protocol 4) were successful.

Materials:

Hardware and Instruments: 0.2-mL PCR tubes (Light Labs, cat. no. A-4003-A)

1.5-mL Eppendorf tubes (Eppendorf, cat. no. 022363204)

Block box (Mini-Blotting Containers, 9.0 × 6.4 × 2.1cm, Research Products International, cat. no. 248716B)

Rocker (Thermo Scientific™ Compact Digital Rocker, cat. no. 11-676-333)

Positively-charged Nitrocellulose Membrane (Fisher Scientific, cat. no. 45-000-850)

Sapphire Biomolecular Imager (Azure Biosystems)

Stratalinker with 254 nm UV bulbs (Spectronics corporation, Spectrolinker XL-1000 UV Crosslinker)

Tweezers (Electron Microscopy Sciences, cat. no. 78326-51)

Reagents, Solutions, Starting Materials, Cell lines: Control and treated RNA samples from Basic Protocol 4

Bovine Serum Albumin (BSA; Fisher Scientific, cat. no. 1265925GM)

Chemiluminescence detection reagents (Advansta, cat. no. K-12043-C20)

Methylene Blue (VWR, cat. no. 470301-814)

Nuclease-free water (Invitrogen, cat. no. AM9937)

PBST (PBS with 1% (v/v) Triton X100, see **Reagents and Solutions**) (VWR, cat. no. 80503-490)

SCC 20x (Fisher Scientific, cat. no. 15557044)

Streptavidin-HRP probes (Abcam, cat. no. ab7403)

Whatman 6720-5002 50 mm In-Line Filters, PTFE, 0.2µm, 10/pk (Cytiva, cat. no. 6720-5002)

Protocol steps:

1. Prepare 5 µg of RNA from each sample in 5 µL (diluting with nuclease-free H2O if necessary) in separate 0.2-mL PCR tubes. Each tube should have 5 µL of RNA at a concentration of 1µg/µL in nuclease-free water.
2. Prepare 20 mL of fresh 2X SCC by taking 2 mL of the stock (20x) and diluting in 18 mL of nuclease-free water. Pour 2X SCC into a container (e.g. block box).
3. Cut an appropriately sized positively-charged nitrocellulose membrane (for two RNA samples, approximately 2 cm × 4 cm). The labeled and unlabeled samples will be dotted on the same blot (Figure 2D).
4. With tweezers, submerge nitrocellulose membrane in 2X SSC for 1 minute.
5. Remove nitrocellulose membrane from the SSC and rest on a dry surface (e.g. Whatman paper). Allow the membrane to dry for 15 min. Wash blot box with deionized water for future use.

   Over-drying the membrane will cause poor RNA bonding. Under-drying the membrane will cause the RNA dot to smear.
6. Spot RNA samples on the membrane with a micropipetter. Each spot must be of the same size (5 µL) and should contain the same amount of RNA (5 µg), in order to accurately compare the signal between samples.

Because the samples will appear identical, it is useful to differentiate the samples by marking a corner of the membrane with a permanent marker.

7. Allow the nitrocellulose membrane to absorb the RNA. Place nitrocellulose membrane on Whatman paper at room temperature for 30 minutes.

8. Carefully transfer the membrane to Stratalinker. Crosslink samples at 100 μJ/cm^2 to secure the RNA onto the membrane.

9. Stain the nitrocellulose membrane with methylene blue.

    a. Prepare 20 mL of 1% (m/v) methylene blue solution (in deionized water) and pour into the washed blot box.

    b. To visualize the total RNA dotted onto the membrane, gently soak the blot in 1% methylene blue solution for 1 minute.

10. Destain the nitrocellulose membrane.

    a. Dispose of methylene blue solution from the blot box.

    b. Run the blot box (with the nitrocellulose membrane in it) under running deionized water, removing any residual methylene blue from the blot box.

    c. Destain membrane in the blot box by washing with ~10 mL of deionized water on a rocker at 10 rpm for 10 minutes. Make sure the membrane is submerged in deionized water throughout the wash.

    d. Replace deionized water every 5–10 minutes. Check blot periodically and repeat washes until the deionized water is light blue. This can take up to 30 minutes.

11. Image the stained RNA when the dotted RNA is clearly visible and the background is washed off (Figure 2D).

    Ideally, the intensity of the dots should be similar, indicating that the same amount of RNA was dotted.

12. Prepare Streptavidin blocking buffer (5% (w/v) BSA in PBST; see **Reagents and Solutions**).

13. Block the membrane in Streptavidin blocking buffer by submerging the membrane for at least 30 minutes at room temperature. Place on rocker at 10 rpm.

14. Wash in PBST for 10 minutes on rocker at 10 rpm. Repeat wash.

15. Assemble 20 mL od Streptavidin buffer (3% BSA in PBST; see **Reagents and solutions**).

16. Probe the dot blot with Streptavidin-HRP.

    a. Add streptavidin-HRP (at 1:2000 dilution from stock) to the Streptavidin buffer.

b. Add the Streptavidin buffer to the blot box. Make sure there is sufficient buffer to cover the membrane.

c. Close the blot box to prevent evaporation.

d. Incubate overnight on 10 rpm rotation on a Rocker at 4°C.

Probing with streptavidin-HRP can also be done at room temperature for a minimum of 3 hours, but positive signal will be reduced and background signal may be higher if done at room temperature.

17. Wash blot with PBST for 5 minutes on a rocker at 10 rpm. Repeat wash twice.

18. Remove PBST and image with chemiluminescence reagents following the manufacturer's instructions. Do not allow nitrocellulose membrane to dry out.

Ideally, there will be strong chemiluminescent signal in the experimental (+ Halo-DBF) sample and little to no signal in the control (−Halo-DBF) sample ("Clicked input" samples in Figure 2D). See Troubleshooting for more details.

BASIC PROTOCOL 6: ENRICHMENT OF BIOTINYLATED RNA USING STREPTAVIDIN BEADS AND RNA-SEQ LIBRARY PREPARATION

Following the successful verification of cycloaddition of biotin to the experimental (+Halo-DBF) and control (−Halo-DBF) RNA samples (Basic Protocol 5), the user will now enrich for biotinylated (i.e. localized) RNA in the experimental sample and the control sample (Figure 1). All steps in this protocol will be done in parallel for all experimental and control samples. The biotinylated RNA will be pulled down with streptavidin beads, followed by RNA extraction from the beads. Before the pulldown is performed, a sample of the biotinylated RNA (approximately 1–2 μg) should have been set aside (as mentioned in Basic Protocol 4; step 8). This sample will be used as a comparison when analyzing RNA-seq data. Following the pulldown, all samples are prepared for high-throughput RNA sequencing.

The amount of streptavidin beads used will depend on the amount of biotinylated RNA in the sample, which depends on the size of the subcellular compartment interrogated. Using an excess of beads can lead to nonspecific RNA binding, while using too few beads will lead to biotinylated RNA remaining in the supernatant of the binding reaction. We have indicated ratios we have used (μL streptavidin beads : μg RNA) for different subcellular compartments in Table 1

Materials:

Hardware and Instruments: 0.2-mL PCR Magnetic rack (if performing magnetic separations in PCR tube) (Permagen Labware, cat. no MSRLV08)

0.2-mL PCR tubes (Light Labs, cat. no. A-4003-A)

1.5-mL Eppendorf tubes (Eppendorf, cat. no. 022363204)

DynaMag -2 (if performing magnetic separations in 1.5-mL tube) (Invitrogen, cat. no 12321D)

Eppendorf tube spinner (Labnet, Labroller II)

RNAse-free low-retention Eppendorf tubes (Research Products International, cat. no. 145491)

Standard tube rotator

Qubit Fluorometer

Reagents, Solutions, Starting Materials, Cell lines: Binding & Washing buffer (see Reagents & Solutions)

Direct-Zol Microprep kit (Zymogen Research, cat. no. R2062)

NaCl (Invitrogen, cat. no. AM9760G)

Nuclease-free water (Invitrogen, cat. no. AM9937)

PBS (Invitrogen, cat. no. AM9625)

Pierce Streptavidin Magnetic beads (Fisher Scientific, cat. no. PI88817)

Control and treatment RNA samples from Basic Protocol 4

SuperaseIn RNAse inhibitor (Thermo Fisher Scientific, AM2696)

Trizol (Invitrogen, cat. no. 15596018)

Qubit RNA HS Assay kit (Thermo Fisher Scientific, cat. no. Q32855)

Qubit dsDNA HS Assay kit (Thermo Fisher Scientific, cat. no. Q32854)

RNAseq library preparation kit of the user's choice (e.g. Kapa RNA hyperprep kit, cat. no. KK8541)

Protocol steps:

1. Take 50 µg of purified, biotinylated RNA from each of the experimental (+Halo-DBF) and the control (−Halo-DBF) samples (Basic Protocol 4) and place them in separate, low-retention 1.5-mL Eppendorf tubes. Dilute all RNA samples to 1 µg/µL. Treat each sample equally in parallel for all steps. Keep on ice.

   The total amount of RNA that will go into each pulldown may depend on the compartment. See Table 1 for more details.

2. To each sample, add NaCl to a final concentration of 2.5 mM and RNase inhibitor to 1X. Keep on ice.

3. Prewash 50 µL of magnetic streptavidin beads per sample (i.e. for six samples, prewash 300 µL). This assumes a 1:1 ratio of µL beads: µg RNA. If using a different ratio, change the amount of beads to wash accordingly. To prewash the beads:

    a. Transfer the beads to a single RNAse-free low-retention 1.5-mL Eppendorf tube

    b. Add 1 mL of Binding & Washing (B&W) buffer (see Reagents and Solutions) and incubate for 5 minutes at room temperature.

    c. Repeat the wash two more times, by collecting the beads on the Dynamag magnet, removing the old B&W buffer, and resuspending them in 1 mL of fresh B&W buffer.

    The concentration of beads is 1 mg/mL. The maximum ratio of beads:RNA is 1 µL streptavidin-beads per 1 µg of RNA. The ratio may need to be changed depending on the amount of RNA that is expected to be labeled in the subcellular compartment. See Table 1 for recommended bead ratios based on tested compartments.

4. Collect beads on the Dynamag magnetic block and resuspend them in enough B&W buffer to have 50 µL of beads per sample (for six samples, 300 µL). Add 50 µL of beads to each RNA sample prepared in step 2.

5. Incubate the beads and RNA mixture at 4°C for 2 hours on a tube spinner at 20 rpm.

6. Remove samples from the tube spinner and place on the Dynamag magnet for 1 minute to clear the beads from the supernatant. Transfer the supernatant to a separate, nuclease-free tube. Keep supernatant on ice.

    Supernatant can be dotted on a nitrocellulose membrane (similar to Basic Protocol 5) to verify efficiency of pulldown. See optional Step 15.

7. Wash the beads with 1 mL of B&W buffer for 5 minutes on rotation at room temperature. Repeat two more times, by collecting beads on the Dynamag magnet rack and washing with fresh B&W buffer.

8. Transfer the final wash, including beads, into a new 1.5-ml low retention tube. Collect beads on the Dynamag magnet and remove all of the remaining buffer. Remove the Eppendorf from the magnetic rack.

9. Resuspend the beads in 50 µl of PBS.

10. Add 150 µl of Trizol and pipet to resuspend.

11. Incubate this mixture for 10 minutes at 37°C to dissociate RNA from beads.

12. Place sample on magnet for 1 minute to collect beads. Collect RNA-containing supernatant to a new 1.5-mL Eppendorf placed on ice.

    If beads are not removed, they will clog the column in subsequent steps and prevent RNA isolation.

13. Recover RNA from the RNA-containing supernatant using the Direct-zol RNA microprep kit according to the manufacturer's instructions. Do not place beads on column. Elute in the appropriate amount of nuclease-free water (typically 10–50 µl of nuclease-free H2O). The desired RNA concentration is 100 ng/µL for experimental (+ Halo-DBF) samples.

14. Quantify the recovered RNA using an RNA HS Qubit assay following the manufacturer's instructions.

    Pause step: RNA can be safely stored at −80°C here prior to cDNA library preparation.

15. (Optional): Set up an RNA dot blot (as in Basic Protocol 5). Dot the sample prior to enrichment (input; 1 µg of RNA from supernatant isolated at Basic Protocol 4, Step 8), dot 1–10% of sample after the enrichment (pulldown), and dot the leftover supernatant to visualize enrichment of biotinylated RNA (from supernatant isolated from Basic Protocol 6, Step 6). See Figure 2D.

    The resulting dot blot should show an enrichment of biotinylated RNA in the experimental pulldown compared to the experimental input, as well as overall more biotinylated RNA in the experimental samples compared to control samples. See Understanding Results for more details.

16. Prepare cDNA libraries of the input (pre-enrichment RNA) and streptavidin-enriched RNA samples in preparation for high-throughput sequencing. Aim for an average RNA fragment of 350 nt. The choice of RNA library construction strategy (e.g. rRNA-depletion, polyA-enrichment, etc.) is up to the user. If possible, begin the preparation of all libraries with equivalent amounts of RNA, and amplify the libraries using the same number of PCR cycles.

17. Submit samples for high-throughput sequencing. Typically, 20 to 40 million read pairs are sequenced for each sample. We typically use paired-end sequencing.

BASIC PROTOCOL 7: COMPUTATIONAL ANALYSIS OF HALO-SEQ DATA

The goal of Halo-Seq is to identify transcripts enriched at a subcellular region of interest compared either to an unlabeled control sample or other controls (see Understanding Results). To identify those transcripts, the user will start from raw RNA sequencing reads. There are several different computational pipelines that can accomplish this task. Here, we describe one such pipeline that makes use of the transcript quantification software Salmon (Patro et al. 2017) and the differential gene expression software DESeq2 (Love, Huber, and Anders 2014; Soneson, Love, and Robinson 2015). This basic protocol relies on the user having basic familiarity with the Unix command line the R programming language. Example code will be provided but will need to be altered to fit the user's specifications.

Adapters contained within sequencing reads are first removed using Cutadapt (M. Martin 2011). Transcript abundances are then be determined using Salmon and collated to gene abundances using tximport (Soneson, Love, and Robinson 2015). Finally, genes whose

RNAs are differentially abundant in localized and control RNA samples are identified using DESeq2.

In this protocol, we will process a total of six raw sequencing files (streptavidin-enriched samples from control (−Halo-DBF) and experimental (+Halo-DBF) samples, with 3 replicates each). The conditions are named experimental_X_pulldown or control_X_pulldown, where X represents the replicate number and “pulldown” represents enriched RNA. Other comparisons can be made (see Discussion on controls and comparisons in Understanding Results)

Materials:

Hardware and Instruments: Access to command line interface

Raw sequencing reads (gzipped fastq files) for each sample (Basic Protocol 6), generally ending with **.fq.gz**. We typically use paired-end sequencing.

Software: Cutadapt v1.18 or later

R v3.3 or later

Salmon v0.13 or later

Required R packages: biomaRt (v2.48.3 or later), DESeq2 (v1.32.0 or later), tidyverse (v1.3.1 or later), tximport (v1.20.0 or later)

Protocol steps:

1. Trim the adapters in each read using Cutadapt (M. Martin 2011). An example run is provided below. If using paired-end reads, the parameters -a and -A tell the command what adapters are expected on the 3′ ends of the forward and reverse reads, respectively. If, after trimming, the length of a read is less than the minimum length parameter (25 in the example provided), it will be discarded. Raw files are named as “SampleX_Read_Y_rawreads.fq.gz”, where X is the sample number and Y denotes the read (1 or 2). Parameters -o and -p designate the name of the output files. To avoid overwriting the raw files, rename each read with a suffix “outputfile.trimmed” (e.g. Sample1_Read_1_outputfile.trimmed.fq.gz). If an error occurs during the run, a log of the error will be saved in a text file denoted after the “>” (e.g. Sample1_cutadapt_output.txt). After trimming the adapters off the reads, they are ready for use in transcript quantification.

Example code:

```
cutadapt -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCA
-A AGATCGGAAGAGCGTCGTGTAGGGAAAGA
--minimum-length 25
-o Sample1_Read_1_outputfile.trimmed.fq.gz
```

```
-p Sample1_Read_2_outputfile.trimmed.fq.gz
Sample1_Read_1_rawreads.fq.gz
Sample1_Read_2_rawreads.fq.gz > Sample1_cutadapt_output.txt
```

2. Download the fasta file of transcript sequences appropriate to the species used for sample generation. This protocol used human cells, so the fasta file gencode.v39.transcripts.fa was downloaded from Gencode (https://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_39/gencode.v39.transcripts.fa.gz) and decompressed.

   Salmon (Patro et al. 2017) requires an index of transcript sequences to quantify the raw reads trimmed in Step 1. This index is made from a fasta file of all transcripts that could exist in the sample (i.e. the transcriptome). Depending on the species, these sequences are available from a variety of sources, including Ensembl (www.ensembl.org).

3. Build a salmon index using the fasta file of transcript sequences (downloaded in Step 2) using the example code below for guidance. “Salmon index” calls the command. The option -t tells the program where the transcripts are stored (from the example code, replace “..pathfile/gencode.hg38comprehensive.cdnaall.fa” with the pathname to the fasta file downloaded in step 2). The output is designated with parameter -i (e.g. transcripts.idx). A successful run will generate a salmon index titled “transcripts.idx” in the folder where the command was run.

   Example code:

   ```
   salmon index -t ..pathfile/gencode.v39.transcripts.fa -i
   transcripts.idx -k 31
   ```

4. Quantify transcript abundances from the trimmed reads generated in Step 1 using Salmon, following the example code below. Salmon will automatically detect the library type as long as --libType A is included in the run. For this protocol, two options were used: --seqBias (to correct for sequence-specific biases in the files) and --gcBias (to correct for fragment-level GC biases in the files). For a full list of options and details related to them, please refer to the Salmon documentation (Patro et al. 2017). The read files are called in parameters −1 and −2 (replace ..pathname/SampleX_Read_Y_outputfile.trimmed.fq.gz” with the pathname to the sample generated in Step 1). The output is designated with parameter -o. For each sample, transcript abundances will be stored in a file titled ‘quant.sf’. If an error occurs during the run, a log of the error will be saved in a text file denoted after the “>” (e.g. Sample1_report.txt).

   Example code:

   ```
   salmon quant
   --libType A
   --seqBias
   ```

```
--gcBias
−1 ..pathname/Sample1_Read_1_outputfile.trimmed.fq.gz
−2 ..pathname/Sample1_Read_2_outputfile.trimmed.fq.gz
-o Sample1
--index transcripts.idx > Sample1_report.txt
```

When checking your output report file, pay close attention to your mapping rate. The mapping rate can be found in pathname/Sample1/logs/salmon_quant.log. Mapping rates should be >60% (see Troubleshooting for more details).

5. Open R and install/load the following packages: tidyverse, tximport, biomaRt, and DESeq2.

Example code:

```
library(tidyverse)
library(tximport)
library(biomaRt)
library(DESeq2)
```

6. Using biomaRt, create a table combining the following desired attributes: Ensembl transcript IDs, Ensembl gene IDs, gene names, and length. The final object is a dataframe titled “ens2gene” with all the desired attributes.

Example code:

```
#Download the human genomic dataset with biomart
mart <- biomaRt::useMart(“ENSEMBL_MART_ENSEMBL”, dataset =
“hsapiens_gene_ensembl”, host=‘useast.ensembl.org’)


#Extract the attributes of interest
t2g <- biomaRt::getBM(attributes = c(‘ensembl_transcript_id’,
‘ensembl_gene_id’, ‘external_gene_name’,
‘refseq_mrna’,”start_position”,”end_position”), mart = mart)


#Create a column with the length of transcripts, titled ‘length’
t2g<-mutate(t2g,length = t2g$end_position-t2g$start_position)


#Create a table with gene names in addition to transcript names
t2g <- dplyr::rename(t2g, target_id= ensembl_transcript_id,
ext_gene = external_gene_name)
ens2gene <- t2g[,c(2,3)]
colnames(ens2gene)[2] <- ‘Gene’
ens2gene <- unique(ens2gene)
```

7. Load salmon-derived transcript quantifications for each sample. “Sample_id” should refer to the file name of the specific samples used in the comparison. Replace “path/to/quant/files” to local pathname to files. Here, we are selecting the sequencing files for the enriched RNA from the experimental/+Halo-DBF condition (“experimental”) and the control/-Halo-DBF condition (“control”).

Example code:

```
#Define the file names of samples
sample_id <- c('experimental_1_pulldown',
'experimental_2_pulldown', experimental_3_pulldown',
'control_1_pulldown', 'control_2_pulldown','control_3_pulldown')


#Create a directory with the correct file path and sample names of
the quant.sf files.
salm_dirs <- sapply(sample_id, function(id) file.path("path/to/
quant/files",paste(id, ".sf",sep = "")))
salm_dirs
tx2gene <- t2g[,c(1,2,3)]
colnames(tx2gene) <- c('TXNAME', 'GENEID','ext_gene')
```

8. Use tximport to calculate gene-level abundance estimates. This results in a dataframe with gene-level abundances for each sample.

Example code:

```
#Create a txi object from the quant.sf files.
txi <- tximport(salm_dirs, type = 'salmon', tx2gene
= tx2gene, dropInfReps = TRUE, countsFromAbundance =
'lengthScaledTPM',txOut=FALSE)
```

9. Use DESeq2 to compare gene abundances across conditions. Here, we combine the ens2gene dataframe we derived in Step 6 to create a final dataframe with log2 fold abundance changes between the defined conditions and their associated p values for every gene.

Example code:

```
#Make the txi object readable by converting it to a dataframe with
sample names as column names. Link the file names with a certain
experimental condition.
samples <- data.frame(row.names = c('experimental_1_pulldown',
             'experimental_2_pulldown',
             'experimental_3_pulldown',
             'control_1_pulldown',
```

```
        'control_2_pulldown',
        'control_3_pulldown'),
    condition = c('Experimental', 'Experimental','Experimental',
        'Control', 'Control','Control'))


#Run DESeq to identify relative abundance of transcripts in
localized compartment compared to controls.
ddsTxi <- DESeqDataSetFromTximport(txi, colData = samples, design
= ~ condition)
tpms_samples<-data.frame(txi$abundance)%>%rownames_to_column(.,
var = 'ensembl_gene_id')
ddsTxi <- ddsTxi[rowSums(counts(ddsTxi)) > 25, ]
dds <- DESeq(ddsTxi)


#Make the DESeq output readable as a dataframe. Show log2 fold
change of condition "experimental" over condition "control". A
positive log2FC means it is enriched in the "experimental"
condition (i.e. RNA around subcellular location of interest).
res <- data.frame(results(dds,contrast
=c('condition',"Experimental","Control"))) %>%
  rownames_to_column(., var = 'Gene') %>%
  dplyr::select(., Gene, log2FoldChange, padj) %>%
  dplyr::rename(., log2FC= log2FoldChange, padj.M = padj) %>%
  dplyr::rename(., ensembl_gene_id = Gene) %>%
  inner_join(., ens2gene, by = 'ensembl_gene_id')
```

REAGENTS AND SOLUTIONS:

Streptavidin buffer (1 mL needed per sample)

0.1% (v/v/) Triton (VWR, cat. no. 80503-4900)

3% (w/v) BSA (Fisher Scientific, cat. no. 1265925GM)

in PBS (Invitrogen, cat. no. AM9625)

Prepare fresh. Store on ice until use.

Streptavidin blocking buffer (1 mL needed per sample)

0.1% (v/v/) Triton (VWR, cat. no. 80503-4900)

5% (w/v) BSA (Fisher Scientific, cat. no. 1265925GM)

in PBS (Invitrogen, cat. no. AM9625)

Prepare fresh. Store on ice until use.

Binding & Washing buffer (3 mL for bead washing + 3.5mL needed per sample)

5 mM Tris-Cl pH 8.0 (Thermo Fisher Scientific, cat. no. AAJ22638K2)

0.5 mM EDTA (Invitrogen, cat. no. AM9261)

1 M NaCl (Invitrogen, cat. no. AM9760G)

0.1% (v/v) Tween20 (Fisher Scientific, cat. no. BP337-500)

0.001% SuperaseIn (Invitrogen, cat. no. AM2696)

Prepare fresh and keep on ice.

Click Blocking buffer (1 mL needed per sample)

1 mg/mL BSA (Fisher Scientific, cat. no. 1265925GM)

75 nM NaCl (Invitrogen, cat. no. AM9760G)

0.025% (m/v) sodium azide (Fisher Scientific, cat. no. AC190380050)

0.1% (v/v) triton (VWR, cat. no. 80503-490)

in PBS (Invitrogen, cat. no. AM9625)

Prepare fresh at room temperature.

Click buffer A (combine reagents in order) (100 μL needed per sample)

10 mM Tris-Cl buffer (pH 8.0) (Invitrogen, cat. no. AM9851)

10 μM Cy5 picolyl azide (Click Chemistry Tools, cat. no. 1167-100)

10 mM sodium ascorbate (prepared fresh) (Sigma Aldrich, cat. no. 11140-50G)

2 mM THPTA (tris-hydroxypropyltriazolylmethylamine) (Click Chemistry Tools, cat. no. 1010-100)

100 μM $CuSO_4$ (Fisher Scientific, cat. no. AC197722500)

Prepare fresh at room temperature.

Copper is added last because the reaction is active once added.

Click buffer B (add reagents in order) (16.8 μL needed per 50 μL click reaction / 10 μg of clicked RNA)

2 mM Biotin picolyl azide (Click Chemistry Tools, cat. no. 1167-100)

10 mM Sodium ascorbate (prepared fresh) (Sigma Aldrich, cat. no. 11140-50G)

10 mM Tris-Cl buffer (pH 8.0) (Invitrogen, cat. no. AM9851)

2 mM THPTA (tris-hydroxypropyltriazolylmethylamine) (Click Chemistry Tools, cat. no. 1010-100)

0.1 mM $CuSO_4$ (Fisher Scientific, cat. no. AC197722500)

Prepare fresh at room temperature.

We recommend making separate stocks for ease of assembly: Tris-Cl (100mM; stored at room temperature), BPA (30mM aliquots; stored in −20°C); NaAsc (500mM; made

fresh right before use); THPTA (40mM; stored in −20°C); CuSO4 (1mM; stored at room temperature). All stocks (except NaAsc) are stable for at least up to one year if stored properly.

DAPI buffer (500 μL needed per sample)

100 ng / mL DAPI (Sigma Aldrich, cat. no. D9542-1MG)

in PBS

Prepare fresh at room temperature.

PBST (variable based on wash steps or sample number)

0.1% (v/v) Triton (VWR, cat. no. 80503-4900)

in PBS (Invitrogen, cat. no. AM9625)

Store at room temperature for up to six months.

COMMENTARY

BACKGROUND INFORMATION

RNA localization contributes to multiple biological functions across a wide variety of tissues and organisms. It is important for mating type switching in yeast (Bertrand et al. 1998), proper development and patterning of *Drosophila melanogaster* embryos (Hachet and Ephrussi 2004; Lécuyer et al. 2007), and neuronal function (Vogelaar et al. 2009; Das, Singer, and Yoon 2019; Yoon et al. 2016). However, although thousands of RNAs are asymmetrically localized, for the vast majority, the mechanisms that underlie their transport are unknown.

RNA localization is commonly thought to contribute to cell function by facilitating localized translation (Lécuyer et al. 2007; K. C. Martin and Ephrussi 2009). Essentially all cells have spatially defined regions with specific functions. These regions are defined by their local protein content and, therefore, maintaining local protein concentrations is critical. Although proteins can be transported to their site of function, transporting, instead, the RNA molecules that encode them, can provide certain advantages. First, since multiple molecules of protein can be made from a single localized RNA, many correctly localized protein molecules can be made from a single transport step. Second, since RNA molecules are often transported in a translationally repressed state (Mardakheh et al. 2015; Moissoglu et al. 2019), their protein production can be quickly activated in response to a perceived stimulus.

For a handful of RNAs, there is considerable knowledge about the mechanistic details of their transport. These RNAs often contain a specific RNA element, usually tens to hundreds of nucleotides in length and located in their 3′ UTR, that is required for their proper transport (Engel et al. 2020). This sequence element, often called a “zipcode”, is bound by RNA-binding proteins (RBPs) that then recruit the machinery required for transport. Other RNAs are localized cotranslationally, relying on the recognition of nascent peptides by locomotive proteins (Sepulveda et al. 2018; Safieddine et al. 2021). For most localized RNAs, however, the identity of the *cis*-elements and *trans*-factors that mediate their transport

are completely unknown. This is in part due to the limited experimental techniques available for the study of RNA localization. Tools are required to identify *what* RNAs are present in various subcellular compartments, before the important questions of *how* these RNAs are getting there and *why* their localization is important can be answered.

Techniques to study RNA localization—Historically, RNA localization regulatory elements have often been identified through laborious experiments involving reporter transcripts. In these approaches, progressively smaller and more targeted pieces of a single localized RNA are fused to a reporter transcript. The ability of these sequences to drive the localization of the reporter is then assayed. Although these experiments form the core of what is currently known about the regulation of RNA localization, they are extremely time consuming, and their results are often difficult to generalize beyond the localized transcript that was dissected. Techniques that allow the identification of numerous RNA transcripts enriched at particular subcellular locations can advance our understanding of the landscape of RNA localization about how localized RNAs reach their destination.

More recently, several techniques have been developed that allow the isolation and characterization of subcellular transcriptomes using high-throughput sequencing (Taliaferro 2022, 2019). These hold the promise of potentially being able to search for patterns that are enriched among groups of transcripts that are present at the same location. Perhaps the most widely utilized of these techniques involve the mechanical separation of cellular processes from the rest of the cell body (Gumy et al. 2011; Zivraj et al. 2010; Taliaferro et al. 2016; Zappulo et al. 2017; Arora et al. 2021; Mili, Moissoglu, and Macara 2008). Although these techniques have yielded valuable insights, they are generally restricted to cells with elaborate, extended morphologies, like neurons. RNA localization has been well-studied in neuronal cell types, but it is also important in cells that lack shapes that lend themselves to this mechanical fractionation technique (Engel et al. 2020; Moor et al. 2017; E. T. Wang et al. 2012).

Newer techniques have been developed to address these limitations and allow the profiling of subcellular transcriptomes in a variety of cell types. Many of these techniques are proximity labeling techniques, where a localized protein of interest is fused to a domain that facilitates the labeling of nearby biomolecules. Originally most often used to label and query the proteomics of subcellular locations (Hung et al. 2016; Mair et al. 2019; To et al. 2016), these techniques produce reactive species, which allow for the spatially restricted (usually <200 nm) tagging of proteins. These approaches have been repurposed in recent years to focus on their ability to label localized RNA molecules as well as proteins. APEX-seq, CAP-seq, and Halo-seq all utilize spatially restricted reactive species to label localized RNA populations (Fazal et al. 2019; Kaewsapsak et al. 2017; Li et al. 2018; P. Wang et al. 2019; Engel et al. 2021). These labels are then used as handles for the subsequent purification and profiling of the localized RNA.

In this protocol, we lay out the procedure for one of these techniques, Halo-seq. This technique allows for labeling and purification of RNA around essentially any subcellular structure without relying on its biochemical properties, offers temporal control of the labeling, and is compatible with multiple downstream applications.

Halo-seq isolates subcellular transcriptomes—Halo-seq relies on the use of HaloTag domains (England, Luo, and Cai 2015). These small (33 kDa) domains specifically and covalently bind Halo ligands (Los et al. 2008). In Halo-seq, a HaloTag domain is genetically fused to a protein that is specifically localized to the subcellular compartment of interest. When a Halo ligand is added, in this case Halo-DBF, it is therefore similarly spatially restricted. Halo-DBF is a fluorophore that emits highly reactive oxygen species when irradiated with green light (Li et al. 2017, 2018). The high reactivity of these species limits their diffusion away from their Halo-DBF source to approximately 100 nm. When these reactive species encounter RNA, they result in the oxidation of RNA bases, making them substrates for nucleophilic attack. In Halo-seq, this nucleophilic attack is done by the alkyne-containing molecule propargylamine. RNA molecules in close proximity to Halo-DBF (and, therefore, the HaloTag fusion protein) are thereby alkynylated. Following total RNA isolation, this makes them substrates for *in vitro* biotinylation through Click chemistry. This biotinylation step is often quite sensitive to perturbation and requires the most amount of optimization (see Critical Parameters). Biotinylated RNA is then purified using streptavidin beads and profiled using high-throughput sequencing.

All of the currently available RNA proximity labeling techniques (APEX-seq, CAP-seq, Halo-seq) rely on the production of reactive species for RNA labeling. However, we have demonstrated that of these, Halo-seq is the most efficient at producing these reactive species and, therefore, labeling RNA (Engel et al. 2021). This may be due, at least in part, to the fact that while APEX-seq and CAP-seq rely on enzymatic activity to produce reactive species (which, after being produced, may damage the enzyme), Halo-seq does so in a nonenzymatic manner.

CRITICAL PARAMETERS

Localizing the HaloTag protein to a subcellular location of interest—The location of the HaloTag within the fusion protein (i.e. whether it is an N-terminal or C-terminal fusion) can affect its localization. Once an endogenous “bait” protein has been chosen, it can be useful to search the literature for information about the localization of its fusion to other domains (e.g. GFP). We have found that, generally, the localization of HaloTag fusions is similar to the localization of other fusions. Regardless, the correct localization of the fusion protein must be verified before beginning a Halo-seq experiment. This can be done using fluorescent Halo ligands (Basic Protocol 1). If a fusion protein is not correctly localized, it is often worth fusing the HaloTag to the opposite terminus of the protein and trying again. Additionally, if a transmembrane protein is being targeted, it is important to ensure that the HaloTag domain is positioned on the desired side of the membrane.

Overall abundance of HaloTag—An important parameter in Halo-seq is that the HaloTag-protein fusion is properly localized to the cell compartment of interest. Notably, the overall abundance of the HaloTag-protein fusion may play a role in this. While we have found that, for most compartments, the overexpression of the HaloTag-protein fusion does not drastically affect its localization, for certain smaller compartments, overexpression of the HaloTag fusion may result in oversaturation of the compartment of interest, resulting

in labeling outside of the intended target. For those smaller compartments, tagging the endogenous gene product using, for example, genome engineering, may be necessary.

In addition to localization, constitutive overexpression of the fusion protein might introduce unintended biological effects, skewing sequencing results. When using an overexpression model, we suggest inducible transgenes due to their lower, more controllable expression (Khandelia, Yap, and Makeyev 2011; Engel et al. 2021). We advise against transient overexpression models, since the amount of protein expressed can be extremely variable from cell to cell. Alternatively, endogenous proteins may be tagged using genome editing techniques. Although this requires considerably more effort, tagging endogenous proteins often removes worries inherent to transgene overexpression.

Labeling time, Sensitivity, Specificity—The time required for Halo-mediated labeling is comparable to other RNA proximity labeling techniques, ranging from 5 to 15 minutes. The labeling radius of the HaloTag fusion has been experimentally shown to be very precise, between 50–200nm from the HaloTag protein (Li et al. 2017). However, given that RNAs and proteins diffuse on a scale of seconds to minutes, the labeling time should be tailored to the compartment to maximize sensitivity and specificity (Basic Protocol 2). Longer labeling times give more labeling but may also result in decreased spatial specificity (Engel et al. 2021). Labeling too briefly may not sufficiently capture the transcripts at the region of interest. Since each subcellular compartment is likely to be different, it is up to the user to optimize these conditions (a table of recommended labeling times can be found in Table 1). A combination of RNA dot blots (Basic Protocol 5) and *in situ* fluorescent Click imaging (Basic Protocol 2) should be used to determine the optimal labeling time. RNA dot blots should be compared between compartments, and signal intensity should correspond to the compartment size (Basic Protocol 5). Generally, to maximize specificity and sensitivity, it is best to use the minimum labeling time that results in sufficient signal from biotinylated RNA (see *Streptavidin enrichment of biotinylated RNA* in Understanding Results).

Conditions for the in vitro biotinylation of RNA—The *in vitro* copper-mediated cycloaddition of biotin picolyl azide is sensitive to time, temperature, and the concentration of copper in the reaction. We have found that 30 minutes at 25°C is the optimal amount of time for the reaction for most compartments (a table of tested compartments can be found in Table 1). However, if there is too much signal in the control sample in which Halo-DBF was omitted, decreasing the time of the reaction may be necessary.

In our experience, changing the amount of copper in the reaction has a substantial effect. For example, if the copper concentration is too high, even the negative control sample without alkynylated RNA will be biotinylated (Figure 3). Therefore, it is important to find a concentration of copper where the experimental Halo-DBF-containing samples are biotinylated, while the control Halo-DBF-lacking samples, are not. Finally, the reaction starts with the addition of the copper, so when processing a large number of samples, add the copper component last to the master mix.

In general, copper ions can negatively impact RNA integrity. Therefore, the Click reaction is a balance between maintaining RNA integrity and the sufficient biotinylation of alkynylated RNA. Copper-less cycloaddition reactions may also work, but we have not tested these.

TROUBLESHOOTING

HaloTag fusion localization—As discussed above, if the HaloTag fusion protein is not localizing to the expected subcellular location, it is worth fusing the HaloTag to another region of the protein. We have observed that N-terminal and C-terminal HaloTag fusions to the same protein may be differentially localized. Additionally, overexpression of the fusion protein may overwhelm cellular transport systems, leading to non-specific localization of the fusion protein. For this reason, we favor integrated transgenes over transiently expressed transgenes since the copy number of integrated transgenes is generally lower. Weaker promoters may also be useful in reducing expression to more manageable levels.

In vitro RNA biotinylation—If no biotinylation is observed for both experimental and control samples, then the problem lies either with the in-cell alkynylation or the *in vitro* biotinylation. It is often useful to have a positive control sample in which there is expected to be a large amount of biotinylated RNA. This can be done using a highly expressed, broadly cytoplasmic HaloTag fusion. Similarly, increasing the amount of time in which cells are exposed to green light will increase alkynylation levels ((Engel et al. 2021)).

If similar levels of biotinylation are observed in the experimental and control samples, this is likely due to nonspecific biotinylation of unalkynylated RNA. Optimizing the *in vitro* biotinylation reaction, paying special attention to the amounts of both copper and biotin azide used in the reaction, can often overcome this.

A more unlikely scenario involves problems with the RNA dot blot experiment itself. As a positive control to test the ability of the experiment to detect biotinylated nucleic acids, it is recommended to spot a small amount of a biotinylated oligonucleotide (DNA or RNA) on the membrane (can be obtained from a range of commercial oligonucleotide synthesis companies, including IDT) and attempt to visualize it with streptavidin-HRP.

Streptavidin pulldown—If promising results were observed in the RNA dot blot yet similar amounts of RNA were precipitated in the streptavidin pulldown experiments using the experimental and control samples, then there may be an excess of free biotin in the RNA samples. It is very important to remove unreacted biotin azide following the *in vitro* biotinylation reaction. We have found that ethanol precipitation of the biotinylation reaction leaves behind a significant amount of unreacted biotin. We have observed much better efficiency of free biotin removal using commercial silica spin columns (Zymo QuickRNA).

Other common problems with the protocols, their causes, and potential solutions are listed in Table 2. For ease, the Table is divided into segments based on which Basic Protocol they pertain to. Regarding Basic Protocol 7, we note that most error codes are informative, and solutions can be found online; we cannot provide a comprehensive list of all potential errors and thus only listed one that is common.

In addition, and as discussed throughout the article, different compartments require different conditions for sequencing. In Table 1, we list some of the compartments we have sequenced and the conditions we have used for the in cell labeling time, *in vitro* click conditions, amount of RNA required for the streptavidin bead pulldown, and bead ratios, to help guide users.

UNDERSTANDING RESULTS

HaloTag fusion protein localization—Generally, the first experiment to be done when performing Halo-seq is the verification that the designed HaloTag fusion protein is localized to the subcellular compartment of interest. This can be done in a straightforward way using fluorescent Halo ligands (Basic Protocol 1; Figure 2A).

Identifying whether a HaloTag fusion is present at a certain subcellular location can be done visually for easily identifiable subcellular locations (e.g. nucleus, cytoplasm, nucleolus). However, some compartments might require co-staining with a primary antibody. Fortunately, the fluorescent Halo ligand covalently binds to the HaloTag, which allows co-staining with a primary antibody against a different protein in the subcellular compartment of interest.

When interpreting the localization of the HaloTag, it is important to make note of not just *if* the fusion protein is present at the desired location, but *what fraction* of the total fusion protein is correctly localized. This can be done by comparing fluorescence intensities at the compartment of interest to total fluorescence in the cell using an image quantification software, such as ImageJ. Since all molecules of the HaloTag fusion have the ability to label RNA, this fraction should ideally be 100%, to obtain the most informative Halo-seq results. In practice, however, this is almost never achievable, but for a successfully designed fusion, the vast majority of it should be at the desired location.

In vivo labeling of RNA—Although visualizing the location of the HaloTag fusion protein is useful, it does not directly tell users where the Halo-DBF-dependent alkynylation is happening. Alkynylated molecules can be visualized *in situ* with an in-cell Click reaction using a fluorescent azide (Basic Protocol 2). With this experiment, when interpreting the overall fluorescence signal intensity, it is important to keep in mind that the reactive oxygen species emitted by the Halo-DBF label both RNA and protein. The observed fluorescent signal is, therefore, a combination of both alkynylated RNA and protein (Figure 2C). Given the stoichiometric differences between RNA and protein, it can also be assumed that most of the alkylated signal is protein.

The copper concentration used in the *in situ* Click reaction may differ significantly from the amount used in the *in vitro* RNA biotinylation reaction. As with the RNA biotinylation reaction, it is not unusual to see background signal in negative control samples (i.e. samples that were not treated with Halo-DBF) prior to optimization of the Click reaction (Figure 3). In an ideal experiment, the background should be minimal, and the signal should be specific to the areas where the HaloTag fusion is localized.

It is important to take into consideration the specific conditions of each Click reaction, because subcellular compartments vary in their sizes and contents.

RNA dot blot of biotinylated RNA—The next critical step is the assaying of RNA biotinylation through the use of an RNA dot blot. Here, it is critical to have both an experimental sample from cells that were treated with Halo-DBF and a control sample in which Halo-DBF was omitted. The ideal result in this step is the presence of strong biotinylation signal in the experimental sample and almost no biotinylation signal in the control sample (as quantified by ImageJ)(Figure 2D). Methylene blue is used as a loading control to ensure that approximately equal amounts of experimental and control RNA are assayed (Figure 2D). The amount of observed biotinylation will be related to the size of the subcellular compartment that was probed and the amount of RNA it contained. If there is no signal for either sample, then there was likely a technical issue either during the in-cell alkynylation step or the *in vitro* biotinylation. If there is substantial signal in both samples, then the *in vitro* biotinylation reaction may need to be optimized. As stated in the Critical Parameters section, the biotinylation reaction is very sensitive to the amount of each reagent in the reaction. Using an excess of copper can result in the reaction becoming quite promiscuous and the consequent biotinylation of nonalkynylated RNA (Figure 3).

Streptavidin enrichment of biotinylated RNA—Before and after the streptavidin pulldown, the RNA is quantified. If the biotinylation and pulldown were successful, then a larger proportion of the experimental RNA sample should be pulled down than of the control RNA sample. Generally, we see between 2 and 20 times more RNA in the experimental pulldown than in the control pulldown (quantified by Qubit). As with the amount of biotinylation, this will vary with the size and RNA content of the interrogated subcellular compartment. If similar amounts of RNA are observed in the experimental and control samples, it is unlikely that it is worth carrying the sample forward to library preparation and sequencing.

Discussion on controls and comparisons—The major goal of Halo-seq is to identify RNAs that are enriched at the subcellular location of interest. As with almost any differential expression analysis from high-throughput RNA sequencing data, this requires the comparison of at least two conditions, with each condition sampled ideally at least in triplicate. However, there is some flexibility in the identity of the conditions that are compared (Figure 4).

Perhaps the most straightforward strategy to identify spatially enriched transcripts is to compare RNA samples from before (input) and after (streptavidin-purified) the enrichment of biotinylated RNA by streptavidin bead pulldown (Figure 4A). This strategy has the advantage of directly comparing two RNA populations from the same cell population. However, it has a disadvantage in that it cannot distinguish truly localized RNAs from those that end up in the streptavidin-purified sample for purely technical reasons. Still, we have found this approach to successfully recapitulate the known localization patterns of dozens of RNAs (Engel et al. 2021), and its simplicity is appreciated.

Alternatively, this potential technical confounder can be controlled for by comparing biotinylated RNA from two Halo-seq conditions, one where the HaloTag fusions is bound to Halo-DBF and the other where the HaloTag is not bound to Halo-DBF (Figure 4B). This is the comparison used in this protocol. This comparison has the advantage of minimizing technical artifacts by comparing two RNA populations that have undergone roughly the same experimental procedures, with the biggest difference between them being whether the Halo-DBF ligand was added. This comparison not only minimizes technical artifacts, but like the previous comparison, also compares RNA populations from the same cell line, reducing biological artifacts. The major difficulty here is that the control (−Halo-DBF) pulled-down sample relies on the background efficiency of beads and, therefore, yields little RNA, making the need for the starting material higher.

Yet another method of comparison can be made, namely, comparing enriched RNAs from two separate subcellular locations (Figure 4C). For example, one experiment may have the HaloTag fusion at a specific location, while the other uses a broadly cytoplasmic HaloTag fusion. This comparison would require the design of a separate cell line and repeating the procedure with new samples. This comparison has the advantage of minimizing technical artifacts by comparing two RNA populations that have undergone roughly the same experimental procedures with the biggest difference between them being the location of the HaloTag fusion. On the other hand, is the approach necessarily compares RNA samples from two different cell populations. We have observed that comparing RNA-seq samples across highly related cell populations (e.g. across transgenic lines derived from the same parental cell line) can reveal many genes that are differentially expressed between them. As such, it is difficult to know whether the RNAs that are differentially abundant in the two biotinylated RNA samples are truly differentially localized or were merely differentially expressed in the original cell samples analyzed. To distinguish between these possibilities, one additional control can be added. This involves sequencing both the input and streptavidin-purified RNA samples from both transgenic cell lines. By comparing ratios of enrichment (streptavidin-purified / input) values across cell lines, contributions from genes that are differentially expressed between the lines themselves can be minimized. However, because this requires RNA-seq analysis of twice the number of samples, it is considerably more expensive, and comparing ratios of expression values across conditions is not as straightforward as quantifying RNA localization using a single expression value ratio.

Interpreting RNA-seq results—As with almost all RNA-seq differential expression analyses, the reported differences in abundances inform on proportional —not absolute— RNA amounts. For example, if transcripts of a given gene are found to be 2-fold enriched at a given location, this does not necessarily mean that there are twice the number of RNA molecules for that gene at that location. Because the total RNA contents of different subcellular compartments can vary dramatically, the reported RNA-seq results do not make statements about absolute numbers of molecules. Rather, a 2-fold enrichment of a transcript X in the experimental condition would mean that the representation of transcript X in the pool of all sequenced transcripts is double for the experimental condition than the control (e.g. 4% of all sequenced transcripts vs 2%). A given RNA species may only have a few

molecules at a particular subcellular location, yet it may still be enriched at that location relative to its concentration in the bulk RNA sample.

When choosing targets for verification, we often choose transcripts that are significantly enriched ($p \leq 0.01$) and have a log2FC ≥ 1.0 over the control. Additionally, while we pay attention to individual transcripts, we also pay particular attention to classes of transcripts or associated transcripts that are enriched together. For example, one proposed role of local RNA is local translation, so finding transcripts encoding for similarly localized proteins is promising. Finally, some transcripts have been characterized to be present at certain compartments. We suggest using those transcripts as a positive control, since they should be more enriched in the experimental condition than the control.

All of this is important to keep in mind when designing and interpreting RNA localization follow up experiments. We prefer using single molecule RNA fluorescence *in situ* hybridization (smFISH), because unlike sequencing, it directly reports on the number of RNA molecules at a location. Additionally, it has the benefit of being completely orthogonal to Halo-Seq. When choosing transcripts to test with smFISH, note that the longer the transcript, the more binding spots there are for smFISH probes, so it is often best to choose longer transcripts. Additionally, it is often best to choose transcripts that are reasonably well expressed. We have had the greatest success validating transcripts with TPM (transcripts per million) expression values of at least 10.

Enrichment of mitochondrial transcripts—We have found that our Halo-seq protocol enriches mitochondrial transcripts across multiple compartments. Because we see this with multiple HaloTag fusion proteins that are not mitochondrially localized, we can only conclude that this is an artifact of our technique. We must, therefore, advise caution when interpreting enrichments of mitochondrial transcripts, and recommend validating any results concerning their localization with smFISH.

TIME CONSIDERATIONS

The first two Basic Protocols are important for establishing the proper conditions and troubleshooting the rest of the protocol. Both visualizing the HaloTag fusion protein (Basic Protocol 1) and visualizing the alkynylated RNAs/proteins (Basic Protocol 2) take 2 days. Expressing and localizing the HaloTag fusion takes 2 days, and the labeling, fixation, and imaging can all be performed on the second day. Therefore, validation of the HaloTag fusion-expressing cell line can take as little as 2 days (if Basic Protocols 1 and 2 are performed concurrently).

Once the cell lines are validated and labeling times are determined, the rest of the procedure of isolating RNA for sequencing and making a cDNA library can be completed in 5 days. The cells would be seeded on Day 0. On Day 2, the cells would be labeled and the RNA, extracted (Basic Protocol 3). On the same day, the RNA can be biotinylated (Basic Protocol 4) and assayed using an RNA dot blot (Basic Protocol 5). On Day 3, the biotinylated RNA can be visualized via dot blot (Basic Protocol 5) and if successful, the biotinylated RNA can then be enriched with a streptavidin bead pulldown (Basic Protocol 6). On Day 4, a high-throughput sequencing library can be prepared (Basic Protocol 6). Once the sequencing

data is returned, the analysis (Basic Protocol 7) can be completed in one day (Day 5). The total time for each Basic Protocol is listed below:

Basic Protocol 1: Total time: 2 days

Basic Protocol 2: Total time: 2 days

Basic Protocol 3: Total time: 2 days; labeling time & RNA extraction: 2 hours 45 minutes.

Basic Protocol 4: Total time: 2 hours.

Basic Protocol 5: Total time: 15 hours.

Basic Protocol 6: Total time: 10 hours.

Basic Protocol 7: Total time depends on the computational systems utilized, minimum of 1 day.

ACKNOWLEDGEMENTS

Funding for the development of Halo-seq was provided by the NIH (R35-GM133885 to JMT, DP2 GM119164 to RCS, NIH-T32-GM008730 to HYGL and RG), the W.M. Keck Foundation (JMT), and the RNA Bioscience Initiative at the University of Colorado Anschutz Medical Campus (JMT, HYGL, and RG). RCS is a Pew Biomedical Scholar. Portions of figures were made using Biorender.

DATA AVAILABILITY STATEMENT

Data demonstrating the ability of Halo-seq to characterize subcellular transcriptomes can be found in the original report describing the method (Engel et al. 2021). High-throughput sequencing data from Halo-seq experiments can be found in the Gene Expression Omnibus (GSE172281).

LITERATURE CITED

Arora Ankita, Goering Raeann, Lo Hei-Yong, and Taliaferro J. Matthew. 2021. “Mechanical Fractionation of Cultured Neuronal Cells into Cell Body and Neurite Fractions.” Bio-Protocol 11 (11). 10.21769/bioprotoc.4048.

Bertrand E, Chartrand P, Schaefer M, Shenoy SM, Singer RH, and Long RM. 1998. “Localization of ASH1 mRNA Particles in Living Yeast.” Molecular Cell 2 (4): 437–45. [PubMed: 9809065]

Cajigas Iván J., Tushev Georgi, Will Tristan J., Susanne tom Dieck Nicole Fuerst, and Schuman Erin M.. 2012. “The Local Transcriptome in the Synaptic Neuropil Revealed by Deep Sequencing and High-Resolution Imaging.” Neuron 74 (3): 453–66. [PubMed: 22578497]

Cassella Lucia, and Ephrussi Anne. 2021. “Subcellular Spatial Transcriptomics Identifies Three Mechanistically Different Classes of Localizing RNAs.” bioRxiv. 10.1101/2021.12.05.471303.

Choi-Rhee Eunjoo, Schulman Howard, and Cronan John E.. 2004. “Promiscuous Protein Biotinylation by Escherichia Coli Biotin Protein Ligase.” Protein Science: A Publication of the Protein Society 13 (11): 3043–50. [PubMed: 15459338]

Das Sulagna, Singer Robert H., and Yoon Young J.. 2019. “The Travels of mRNAs in Neurons: Do They Know Where They Are Going?” Current Opinion in Neurobiology 57 (August): 110–16. [PubMed: 30784978]

Engel Krysta L., Arora Ankita, Goering Raeann, Lo Hei-Yong G., and Taliaferro J. Matthew. 2020. “Mechanisms and Consequences of Subcellular RNA Localization across Diverse Cell Types.” Traffic 21 (6): 404–18. [PubMed: 32291836]

Engel Krysta L., Lo Hei-Yong G., Goering Raeann, Li Ying, Spitale Robert C., and Taliaferro J. Matthew. 2021. “Analysis of Subcellular Transcriptomes by RNA Proximity Labeling with Halo-Seq.” Nucleic Acids Research, December. 10.1093/nar/gkab1185.

England Christopher G., Luo Haiming, and Cai Weibo. 2015. “HaloTag Technology: A Versatile Platform for Biomedical Applications.” Bioconjugate Chemistry 26 (6): 975–86. [PubMed: 25974629]

Fazal Furqan M., Han Shuo, Parker Kevin R., Kaewsapsak Pornchai, Xu Jin, Boettiger Alistair N., Chang Howard Y., and Ting Alice Y.. 2019. “Atlas of Subcellular RNA Localization Revealed by APEX-Seq.” Cell 178 (2): 473–90.e26. [PubMed: 31230715]

Goering Raeann, Hudish Laura I., Guzman Bryan B., Raj Nisha, Bassell Gary J., Russ Holger A., Dominguez Daniel, and Taliaferro J. Matthew. 2020. “FMRP Promotes RNA Localization to Neuronal Projections through Interactions between Its RGG Domain and G-Quadruplex RNA Sequences.” eLife 9 (June). 10.7554/eLife.52621.

Gumy Laura F., Yeo Giles S. H., Yi-Chun Loraine Tung Krishna H. Zivraj, Willis Dianna, Coppola Giovanni, Lam Brian Y. H., Twiss Jeffery L., Holt Christine E., and Fawcett James W.. 2011. “Transcriptome Analysis of Embryonic and Adult Sensory Axons Reveals Changes in mRNA Repertoire Localization.” RNA 17 (1): 85–98. [PubMed: 21098654]

Hachet Olivier, and Ephrussi Anne. 2004. “Splicing of Oskar RNA in the Nucleus Is Coupled to Its Cytoplasmic Localization.” Nature 428 (6986): 959–63. [PubMed: 15118729]

Hein CD, Liu XM, and Wang D. 2008. “Click Chemistry, a Powerful Tool for Pharmaceutical Sciences.” Pharmaceutical Research 25 (10): 2216–30. [PubMed: 18509602]

Hung Victoria, Udeshi Namrata D., Lam Stephanie S., Loh Ken H., Cox Kurt J., Pedram Kayvon, Carr Steven A., and Ting Alice Y.. 2016. “Spatially Resolved Proteomic Mapping in Living Cells with the Engineered Peroxidase APEX2.” Nature Protocols 11 (3): 456–75. [PubMed: 26866790]

Kaewsapsak Pornchai, David Michael Shechner William Mallard, Rinn John L., and Ting Alice Y.. 2017. “Live-Cell Mapping of Organelle-Associated RNAs via Proximity Biotinylation Combined with Protein-RNA Crosslinking.” eLife 6 (December). 10.7554/eLife.29224.

Khandelia Piyush, Yap Karen, and Makeyev Eugene V.. 2011. “Streamlined Platform for Short Hairpin RNA Interference and Transgenesis in Cultured Mammalian Cells.” Proceedings of the National Academy of Sciences of the United States of America 108 (31): 12799–804. [PubMed: 21768390]

Lécuyer Eric, Yoshida Hideki, Parthasarathy Neela, Alm Christina, Babak Tomas, Cerovina Tanja, Hughes Timothy R., Tomancak Pavel, and Krause Henry M.. 2007. “Global Analysis of mRNA Localization Reveals a Prominent Role in Organizing Cellular Architecture and Function.” Cell 131 (1): 174–87. [PubMed: 17923096]

Li Ying, Aggarwal Mahima B., Ke Ke, Nguyen Kim, and Spitale Robert C.. 2018. “Improved Analysis of RNA Localization by Spatially Restricted Oxidation of RNA-Protein Complexes.” Biochemistry 57 (10): 1577–81. [PubMed: 29474061]

Li Ying, Aggarwal Mahima B., Nguyen Kim, Ke Ke, and Spitale Robert C.. 2017. “Assaying RNA Localization in Situ with Spatially-Restricted Nucleobase Oxidation.” ACS Chemical Biology, acschembio.7b00519.

Los GV, Encell LP, McDougall MG, Hartzell DD, Karassina N, Zimprich C, Wood MG, et al. 2008. “HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis.” ACS Chemical Biology 3 (6): 373–82. [PubMed: 18533659]

Love Michael I., Huber Wolfgang, and Anders Simon. 2014. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2.” Genome Biology 15 (12): 550. [PubMed: 25516281]

Mair Andrea, Xu Shou-Ling, Branon Tess C., Ting Alice Y., and Bergmann Dominique C.. 2019. “Proximity Labeling of Protein Complexes and Cell-Type-Specific Organellar Proteomes in Arabidopsis Enabled by TurboID.” eLife 8 (September). 10.7554/eLife.47864.

Mardakheh FK, Paul A, Kumper S, Sadok A, Paterson H, McCarthy A, Yuan Y, and Marshall CJ. 2015. “Global Analysis of mRNA, Translation, and Protein Localization: Local Translation Is a Key Regulator of Cell Protrusions.” Developmental Cell 35 (3): 344–57. [PubMed: 26555054]

Martin Kelsey C., and Ephrussi Anne. 2009. “mRNA Localization: Gene Expression in the Spatial Dimension.” Cell 136 (4): 719–30. [PubMed: 19239891]

Martin Marcel. 2011. “Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads.” EMBnet Journal 17 (1): 10–12.

Mili Stavroula, Moissoglu Konstadinos, and Macara Ian G.. 2008. “Genome-Wide Screen Reveals APC-Associated RNAs Enriched in Cell Protrusions.” Nature 453 (7191): 115–19. [PubMed: 18451862]

Moissoglu Konstadinos, Yasuda Kyota, Wang Tianhong, Chrisafis George, and Mili Stavroula. 2019. “Translational Regulation of Protrusion-Localized RNAs Involves Silencing and Clustering after Transport.” eLife 8 (July): e44752. [PubMed: 31290739]

Moor Andreas E., Golan Matan, Massasa Efi E., Lemze Doron, Weizman Tomer, Shenhav Rom, Baydatch Shaked, et al. 2017. “Global mRNA Polarization Regulates Translation Efficiency in the Intestinal Epithelium.” Science, eaan2399.

Padrón Alejandro, Iwasaki Shintaro, and Ingolia Nicholas T.. 2019. “Proximity RNA Labeling by APEX-Seq Reveals the Organization of Translation Initiation Complexes and Repressive RNA Granules.” Molecular Cell 75 (4): 875–87.e5. [PubMed: 31442426]

Patro Rob, Duggal Geet, Love Michael I., Irizarry Rafael A., and Kingsford Carl. 2017. “Salmon Provides Fast and Bias-Aware Quantification of Transcript Expression.” Nature Methods 14 (4): 417–19. [PubMed: 28263959]

Roux KJ, Kim DI, Raida M, and Burke B. 2012. “A Promiscuous Biotin Ligase Fusion Protein Identifies Proximal and Interacting Proteins in Mammalian Cells.” The Journal of Cell Biology 196 (6): 801–10. [PubMed: 22412018]

Safieddine Adham, Coleno Emeline, Salloum Soha, Imbert Arthur, Traboulsi Abdel-Meneem, Oh Sung Kwon Frederic Lionneton, et al. 2021. “A Choreography of Centrosomal mRNAs Reveals a Conserved Localization Mechanism Involving Active Polysome Transport.” Nature Communications 12 (1): 1352.

Sepulveda Guadalupe, Antkowiak Mark, Ingrid Brust-Mascher Karan Mahe, Ou Tingyoung, Castro Noemi M., Christensen Lana N., et al. 2018. “Co-Translational Protein Targeting Facilitates Centrosomal Recruitment of PCNT during Centrosome Maturation in Vertebrates.” eLife 7 (April). 10.7554/eLife.34959.

Soneson Charlotte, Love Michael I., and Robinson Mark D.. 2015. “Differential Analyses for RNA-Seq: Transcript-Level Estimates Improve Gene-Level Inferences.” F1000Research 4 (December): 1521. [PubMed: 26925227]

Taliaferro J. Matthew. 2019. “Classical and Emerging Techniques to Identify and Quantify Localized RNAs.” Wiley Interdisciplinary Reviews. RNA, May, e1542. [PubMed: 31044542]

Taliaferro J. Matthew. 2022. “Transcriptome-Scale Methods for Uncovering Subcellular RNA Localization Mechanisms.” Biochimica et Biophysica Acta, Molecular Cell Research 1869 (4): 119202. [PubMed: 34998919]

Taliaferro J. Matthew, Vidaki Marina, Oliveira Ruan, Olson Sara, Zhan Lijun, Saxena Tanvi, Wang Eric T., et al. 2016. “Distal Alternative Last Exons Localize mRNAs to Neural Projections.” Molecular Cell 61 (6): 821–33. [PubMed: 26907613]

To Tsz-Leung, Medzihradszky Katalin F., Burlingame Alma L., DeGrado William F., Jo Hyunil, and Shu Xiaokun. 2016. “Photoactivatable Protein Labeling by Singlet Oxygen Mediated Reactions.” Bioorganic & Medicinal Chemistry Letters 26 (14): 3359–63. [PubMed: 27220724]

Vogelaar Christina F., Gervasi Noreen M., Gumy Laura F., Story David J., Ruma Raha-Chowdhury Kin-Mei Leung, Holt Christine E., and Fawcett James W.. 2009. “Axonal mRNAs: Characterisation and Role in the Growth and Regeneration of Dorsal Root Ganglion Axons and Growth Cones.” Molecular and Cellular Neurosciences 42 (2): 102–15. [PubMed: 19520167]

Wang Eric T., Cody Neal A. L., Jog Sonali, Biancolella Michela, Wang Thomas T., Treacy Daniel J., Luo Shujun, et al. 2012. “Transcriptome-Wide Regulation of Pre-mRNA Splicing and mRNA Localization by Muscleblind Proteins.” Cell 150 (4): 710–24. [PubMed: 22901804]

Wang Pengchong, Tang Wei, Li Zeyao, Zou Zhongyu, Zhou Ying, Li Ran, Xiong Tianyu, Wang Jianbin, and Zou Peng. 2019. “Mapping Spatial Transcriptome with Light-Activated Proximity-Dependent RNA Labeling.” Nature Chemical Biology 15 (11): 1110–19. [PubMed: 31591565]

Yoon Young J., Wu Bin, Buxbaum Adina R., Das Sulagna, Tsai Albert, English Brian P., Grimm Jonathan B., Lavis Luke D., and Singer Robert H.. 2016. “Glutamate-Induced RNA Localization

and Translation in Neurons." Proceedings of the National Academy of Sciences of the United States of America 113 (44): E6877–86. [PubMed: 27791158]

Zappulo Alessandra, Van Den Bruck David, Mattioli Camilla Ciolli, Franke Vedran, Imami Koshi, McShane Erik, Moreno-Estelles Mireia, et al. 2017. "RNA Localization Is a Key Determinant of Neurite-Enriched Proteome." Nature Communications 8 (1). 10.1038/s41467-017-00690-6.

Zivraj Krishna H., Tung Yi Chun Loraine, Piper Michael, Gumy Laura, Fawcett James W., Yeo Giles S. H., and Holt Christine E.. 2010. "Subcellular Profiling Reveals Distinct and Developmentally Regulated Repertoire of Growth Cone mRNAs." The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 30 (46): 15464–78. [PubMed: 21084603]

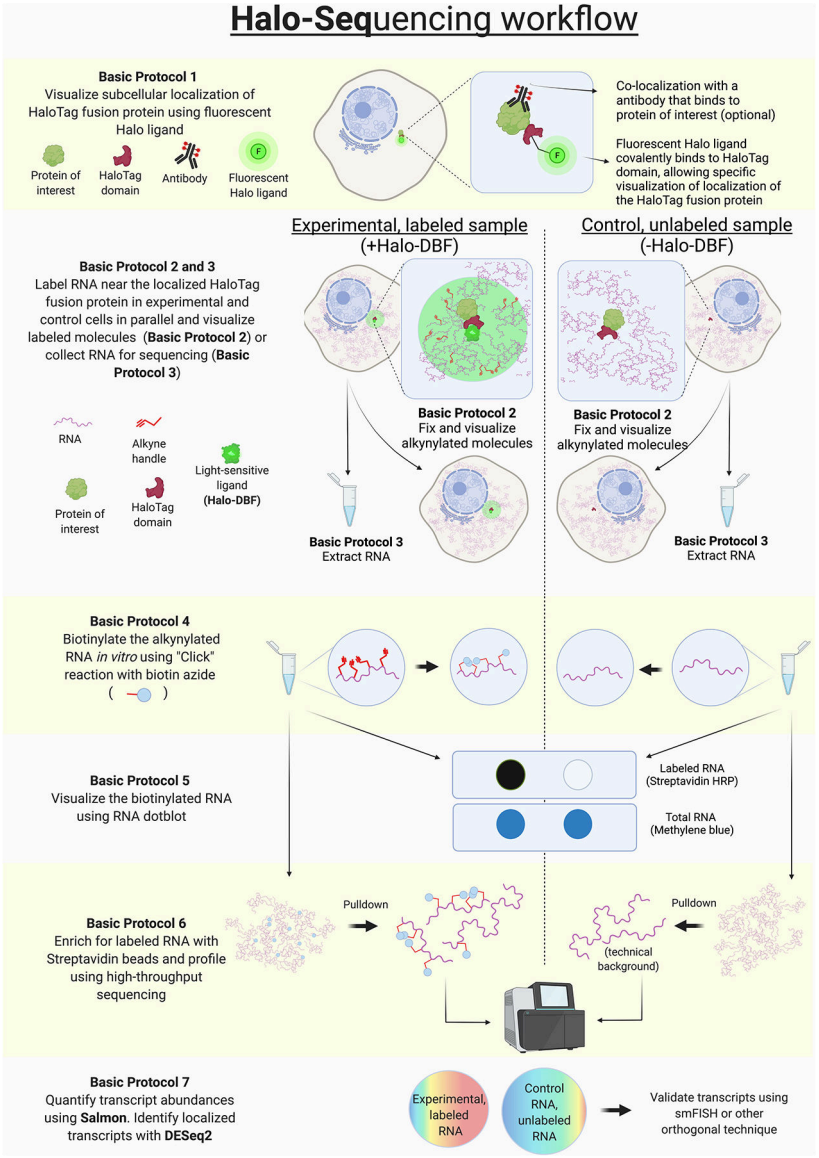

Figure 1. Overview of the protocols described here for Halo-Sequencing.

After establishing the proper conditions(Basic Protocol 1 and 2), the RNA is labeled *in vivo* around the localized HaloTag fusion (Basic Protocol 3). Then, the extracted RNA is biotinylated (Basic Protocol 4 and 3).: Local RNA is then enriched using a streptavidin bead pulldown and sequenced (Basic Protocol 6). Finally, relative transcript abundances are determined computationally (Basic Protocol 7).

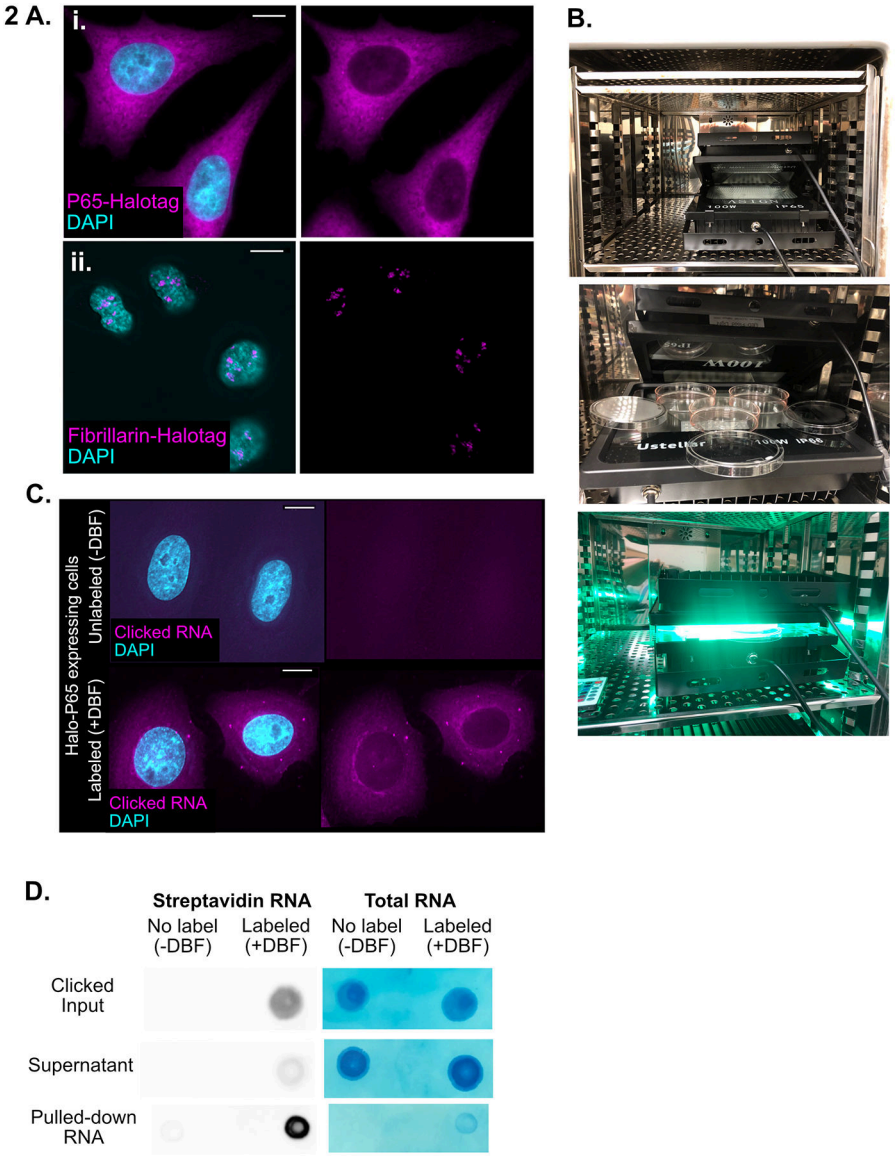

Figure 2. Visualization of Halo-Tag localization and in vitro biotinylated RNA in different cellular compartments.

(**A)** Max projected images of fixed HeLa cells expressing either the cytoplasmically localized P65-HaloTag fusion protein (i) or nucleolus localized Fibrillarin-HaloTag fusion protein (ii). Correct localization of either fusion protein is visualized with a fluorescent Halo-ligand. Nuclei visualized with DAPI. (**B)** In-cell labeling rig set-up. Cells grown on plates are sandwiched with light fixtures and exposed to green light for the labeling process. (**C)** Max-projected fixed HeLa cells expressing the Halo-P65 fusion protein are either labeled (+DBF) or not labeled (−DBF). Alkynylated RNA/protein is visualized following an *in vivo* cycloaddition of Cy5-azide. Nuclei visualized with DAPI. All scale bars represent 10 μm. (**D)** An RNA dot blot of a single pair of labeled (+DBF) and unlabeled (-DBF) samples. Streptavidin-HRP antibodies label biotinylated RNA. Streptavidin signal is only visualized in the labeled and biotinylated +DBF samples and not in unlabeled controls. Streptavidin-bead pull down enriches for RNA, as visualized by increased streptavidin signal

in the pulled-down RNA (10% (v/v) of total pulled-down sample blotted). Equal amounts of RNA were loaded in the clicked input and supernatant condition (visualized by methylene blue staining of total RNA).

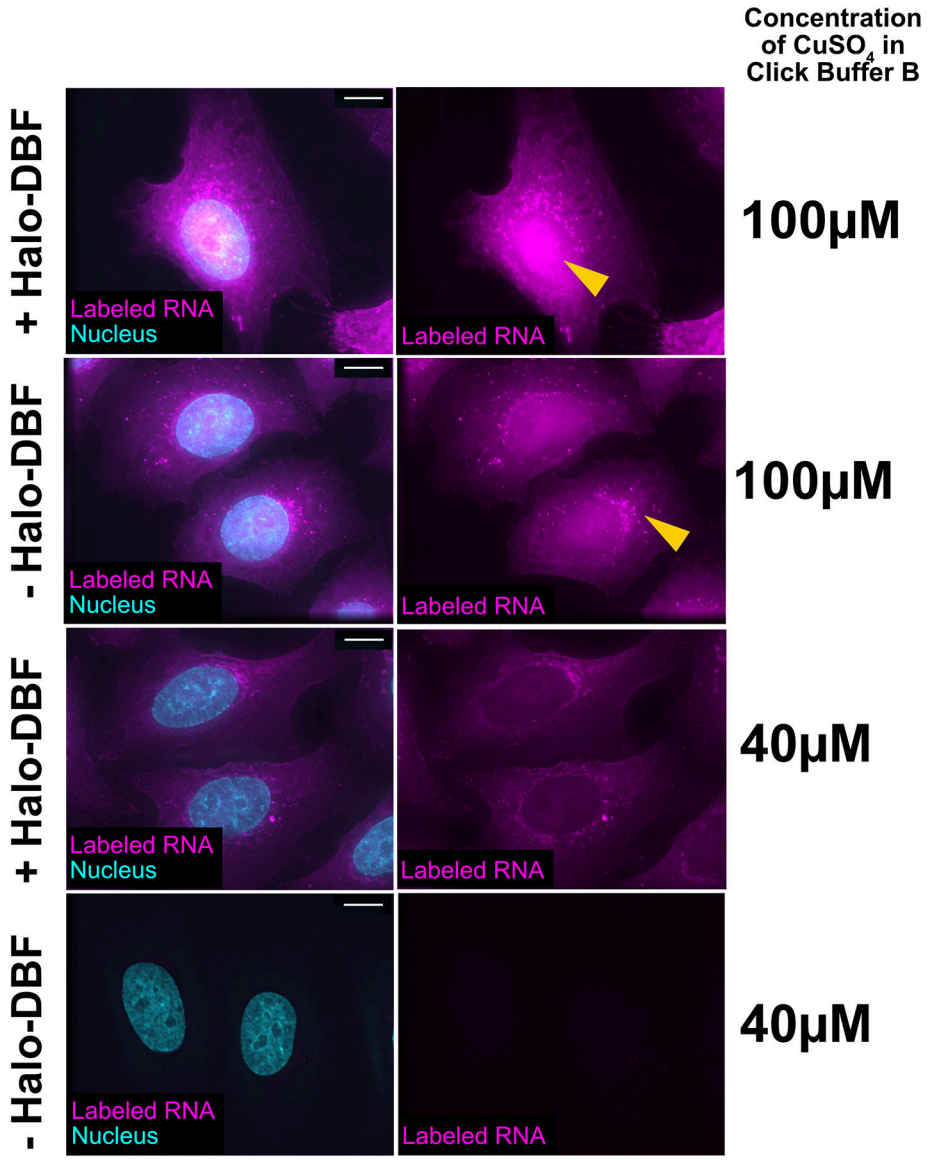

Figure 3. Fixed HeLa cells expressing the Halo-P65 fusion protein.

Halo-P65 cells treated with DBF (+Halo-DBF) allow for proximity labeling of RNA/protein whereas labeling should not occur in control (−Halo-DBF) cells. Alkynylated RNA/protein is visualized following *in vivo* cycloaddition of Cy5 azide. Copper concentration greatly affects the amount of “clicked” product. Background labeling of biomolecules (yellow arrows) can occur when too much copper is added to the reaction (100μM; top four panels). Labeling is considered nonspecific because it is present where the Halo-P65 fusion protein is not (nucleus), and is present even in the cytoplasm of control cells without DBF (−Halo-DBF). In contrast, in conditions with a lower copper concentration (lower four panels) the labeling is both restricted to the cytoplasm and specific to Halo-DBF-treated cells. Images were taken with the same exposure and intensity. All scale bars represent 10 μm.

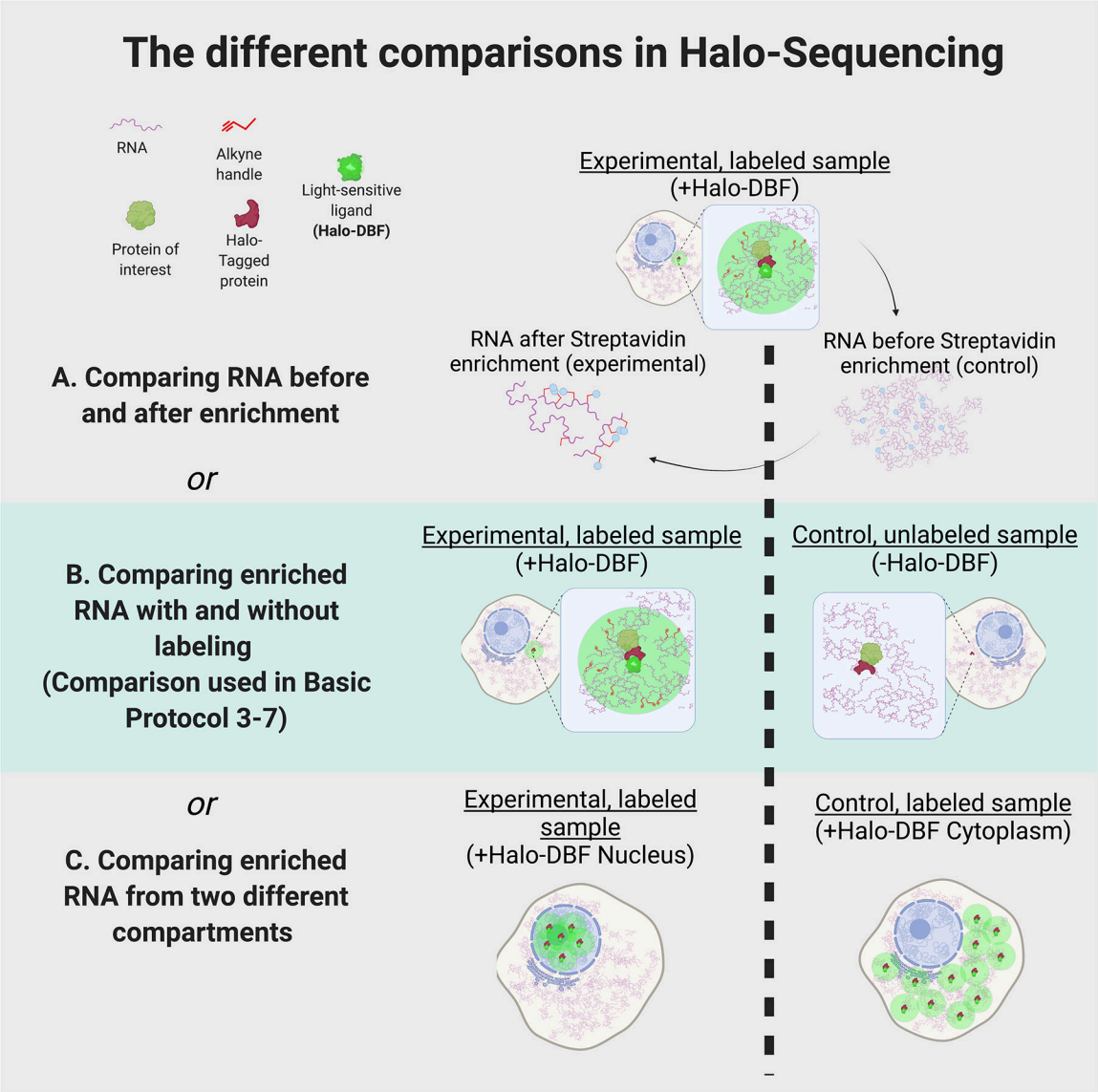

Figure 4.

An overview of the different comparisons that can be used in Halo-Sequencing. (**A**) The simplest comparison is comparing RNA before and after enrichment via streptavidin bead pulldown. (**B**) The comparison utilized in the procedure outlined in this article is comparing RNA after the streptavidin bead pulldown from samples with labeling compared to RNA after streptavidin bead pulldown from samples without labeling. (**C**) Users can also compare RNA labeled and enriched from two separate compartments. In this comparison, labeled RNA is isolated from the area of interest and compared to RNA isolated from a broader compartment (such as the cytoplasm).

Table 1.

Recommended Halo-Seq conditions for tested compartments

Compartment	***In vivo* labeling time (Protocol 2 & 3)**	***In vitro* Copper incubation time (Protocol 4)**	**Amount of RNA needed for pulldown (Protocol 6)**	**µL Bead: µg RNA Ratio (Protocol 6)**
Cytoplasm (Halo-P65)	5–10 minutes	30 minutes	30–100µg	1:1
Cytoplasm (Halo-NES)	5–10 minutes	30 minutes	30–100µg	1:1
Nucleus (Halo-H2B)	5–10 minutes	30 minutes	50–100µg	1:2
Nucleolus (Halo-Fibrillarin)	5–10 minutes	30 minutes	50–100µg	1:1
Centrosome (Halo-PCNT)	15 minutes	30 minutes	150–200µg	1:10

Table 2.

Troubleshooting

Protocol	Problem	Possible cause(s)	Solution
Basic Protocol 1	HaloTag is mislocalized.	Halo-fusion protein is designed incorrectly *or* there is expression of the Halo-fusion protein is too high.	Localization of HaloTag is discussed in the Critical Parameters section of this protocol. Briefly, factors that affect localization of protein include masking of localization domains with the HaloTag or expression levels. First, try decreasing the expression of the Halo-fusion protein expression by switching to an inducible model or decreasing the amount of induction by doxycycline. If that does not fix the problem, it could be that the HaloTag is masking a localization domain in the protein. Re-design the transgene.
	No HaloTag is visualized	Concentration or duration of doxycycline to induce expression of HaloTag fusion protein is insufficient *or* fluorescent Halo ligand is not entering the cell	For inducible systems, the HaloTag fusion protein has to be synthesized and localized. Increase either the amount of doxycycline or the duration of the incubation, to allow full synthesis and localization of the fusion protein. It is also possible that the fluorescent Halo ligand is not penetrating into the cell. We have found that for HeLa cells, 10 minutes of incubation of halo ligand at 1:2000 dilution from 100 μM stock is sufficient to saturate HaloTag binding sites, but users can try to increase the concentration of the Halo ligand or increase the incubation time.
Basic Protocol 2	There is too much labeling in the in-cell Click reaction.	Non-optimized Click conditions.	The in-cell Click reaction may require different parameters than the *in vitro* Click reaction. To decrease signal, decrease copper or fluorophore-azide (Figure 3). In general, the size of the compartment is correlated with the copper required (i.e. the smaller the compartment, the less copper is often required). Concentrations of copper up to 100 times less than recommended in the protocol (100 μM) may be required.
	There is too little labeling in the in-cell Click reaction	Non-optimized Click conditions *or* The labeling time is too short	The in-cell Click reaction may require different parameters than the *in vitro* Click reaction. To increase signal, increase copper or fluorophore-azide (Figure 3). In general, the size of the compartment is correlated with the copper required (i.e. the smaller the compartment, the less copper is often required). The labeling time can also affect the amount of RNA/protein labeled around the halo-fusion. Generally, increasing the labeling time should increase the total amount of RNA/protein labeled.
	There is no labeling in the in-cell Click reaction at all	Halo-DBF is degraded	Halo-DBF is water-, light-, and temperature-sensitive. While normally Halo-DBF easily penetrates cell walls, when improperly stored, Halo-DBF can degrade and lose its ability to enter the cell. This will result in drastically reduced labeling. When not in use, always keep Halo-DBF stored at −20°C in DMSO in the dark to prevent this from happening.
Basic Protocols 3–5	Insufficient RNA to go into the streptavidin bead pulldown	Not enough starting material *or* Poor RNA extraction	The recommended amount of RNA to be pulled down for various compartments can be found in Table 2. If you do not have enough RNA to proceed with the bead pulldown, first try scaling up the total amount of cells labeled in Basic Protocol 3. This may require multiple run throughs of the Basic Protocols in order to acquire the experimental and control samples in triplicate. It is also possible that the RNA extraction from cells is poor. This can be at the Trizol RNA extraction step or subsequent washes. For the Trizol RNA extraction, adding more chloroform to the Trizol RNA extraction (300 μL to each 1 mL of Trizol) may increase the RNA yield. For columns, take the RNA eluate and run it through the column once more, for a total of two times, to increase the maximum amount of RNA in the eluate.
	No enrichment of biotin signal in labeled sample on RNA dot blot	Non-optimized click conditions. *or* Poor labeling	Click: The copper-catalyzed click reaction requires careful fine-tuning of each reactant. The most influential reactants are the concentration of copper and the concentration of the biotin picolyl azide. If the biotin signal of the labeled sample is too low, increase the amount of copper and/or biotin picolyl azide until there is signal in the labeled samples and little-to-no signal in the labeled samples. Labeling: One of the most valuable readouts of a successful labeling reaction is the RNA dot blot. If there is no signal in the labeled sample, it could suggest the labeling was inefficient. Try to increase the labeling time (i.e. the time cells are exposed to green light). This is particularly important for small compartments or compartments with little RNA.
	Too much biotin signal of non-labeled sample.	Non-optimized click conditions.	Click: The copper-catalyzed click reaction requires careful fine-tuning of each reactant. The most influential reactants are the concentration of copper and the concentration of the biotin picolyl azide. Adding too much copper will cause biotinylation of

Protocol	Problem	Possible cause(s)	Solution
			all RNA, regardless of alkynylation status. If the biotin signal of the control is too high, reduce the amount of copper and/or biotin picolyl azide until there is no signal in the unlabeled samples and signal in the labeled samples.
Basic Protocol 6	Poor enrichment of RNA following pulldown (<2.0x enrichment in labeled vs unlabeled).	Non-optimized bead conditions.	There is typically a background amount of RNA in pulled-down samples regardless of labeling. This background is a result of both background biotinylation and background binding of streptavidin beads to unlabeled RNA. If the compartment labeled is too small (i.e. there is not enough RNA labeled), then it may not be sufficient to overcome this background signal. To combat this, reduce the ratio of beads:RNA. We have had success between µL beads: µg RNA from 1:1 to 1:10.
	Not enough RNA recovered from the pulldown for sequencing, but no biotinylated RNA is present in supernatant.	Not enough starting material for pulldown.	The amount of RNA the user expects to enrich after the pulldown is proportional to the concentration of the RNA around the compartment. In general, we have found that the larger the compartment, the more RNA the user can expect to pulldown. Example compartments and the amount of RNA going into the pulldown is shown in Table 2. Given that only 100 ng of RNA is required for most cDNA library prep kits, most compartments need only 100 µg of RNA going into the pulldown. For the unlabeled samples, the amount of RNA recovered is exceptionally low - expect 0.3–0.7% recovery.
	Not enough RNA recovered from the pulldown for sequencing and supernatant still contains biotinylated RNA.	Bead pulldown is inefficient	Saving the supernatant (Basic Protocol 6, step 6) is important to know the general efficiency of the streptavidin bead pulldown. If there is still biotinylated signal in the supernatant following pulldown (visualized by dot blot), then the amount or quality of the beads is likely the culprit. First try increasing the ratio of beads:RNA to 1:1. If that does not solve the issue, then the quality of the beads may be an issue.
	RNA following pulldown does not produce a library	RNA quality is poor *or* Library prep is not optimized	First, check to see the quality of the RNA following the bead pulldown. Ideally, the RNA should have a 260/280 ratio >2.0 and 260/230 ratio >2.0. Second, check to make sure the library prep steps are optimized. The two critical parameters in the KAPA HyperPrep kit we use is the fragmentation time (3 minutes) and the number of PCR cycles (14). Try either changing the fragmentation time or increasing the number of PCR cycles on freshly isolated RNA (that has not been through the protocol) to optimize the conditions required to make a library.
	There are adaptor dimers in the final library prep	Size selection is not optimized	Adaptor dimers can be reduced in any library prep by changing the size selection during the final clean up steps. For our library prep library amplification clean up, we use a 0.8% bead-based clean up as the final step to ensure minimal adaptor dimers for our ~350 nt cDNA libraries.
Basic Protocol 7	Salmon directory could not be found during tximport collation of expression quantifications	Not referencing the correct quant.sf files	The function used to create the salmon directory (salm_dirs) concatenates the file path (“path/to/quant/files”), sample ID, and file name extension (“.sf”). Make sure that all quant.sf files are in the same folder, and thus share the same path. To check to make sure the path names are correct, check the identity of salm_dirs to see the full path name referenced for each file. salm_dirs <- sapply(sample_id, function(id) file.path(“path/to/quant/files”,paste(id, “.sf”,sep = ““))) salm_dirs