

# RNA–protein analysis using a conditional CRISPR nuclease

Ho Young Lee<sup>a,b</sup>, Rachel E. Haurwitz<sup>a,1</sup>, Alex Apffel<sup>c</sup>, Kaihong Zhou<sup>a,b</sup>, Brian Smart<sup>c</sup>, Craig D. Wenger<sup>c</sup>, Stephen Laderman<sup>c</sup>, Laurakay Bruhn<sup>c</sup>, and Jennifer A. Doudna<sup>a,b,d,e,2</sup>

<sup>a</sup>Department of Molecular and Cell Biology, and <sup>b</sup>Howard Hughes Medical Institute, University of California, Berkeley, CA 94720; <sup>c</sup>Agilent Technologies, Inc., Santa Clara, CA 95051; <sup>d</sup>Department of Chemistry, University of California, Berkeley, CA 94720; and <sup>e</sup>Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

Contributed by Jennifer A. Doudna, February 14, 2013 (sent for review January 11, 2013)

**RNA-binding proteins control the fate and function of the transcriptome in all cells. Here we present technology for isolating RNA–protein partners efficiently and accurately using an engineered clustered regularly interspaced short palindromic repeats (CRISPR) endoribonuclease. An inactive version of the Csy4 nuclease binds irreversibly to transcripts engineered with a 16-nt hairpin sequence at their 5' ends. Once immobilized by Csy4 on a solid support, contaminating proteins and other molecules can be removed by extensive washing. Upon addition of imidazole, Csy4 is activated to cleave the RNA, removing the hairpin tag and releasing the native transcript along with its specifically bound protein partners. This conditional Csy4 enzyme enables recovery of specific RNA-binding partners with minimal false-positive contamination. We use this method, coupled with quantitative MS, to identify cell type-specific human pre-microRNA-binding proteins. We also show that this technology is suitable for analyzing diverse size transcripts, and that it is suitable for adaptation to a high-throughput discovery format.**

non-coding RNA | RNA processing | miRNA | mass spectrometry

**R**NA molecules function together with specific binding proteins to regulate cellular pathways at the levels of transcription, posttranscriptional modification, and translation (1–4). For example, microRNAs (miRNA), siRNAs, and piwi-interacting RNAs (piRNAs) regulate more than 30% of mammalian gene expression (5). Small nucleolar RNAs (snoRNAs) govern the sites and efficiencies of RNA chemical modifications in cells (6), whereas long noncoding RNAs (lncRNAs), such as HOTAIR and MALAT1, have been implicated in chromatin remodeling, transcriptional activation, and tumorigenesis (7, 8). UTRs of mRNA transcripts are also known to interact with regulatory proteins to control their expression level as well as their stability. Understanding how these RNAs function in cells and how they may be manipulated for therapeutic purposes is an important goal that spans many areas of biology.

Although new ncRNAs are being discovered at a rapid pace, determination of their biochemical activities is often slow. A major barrier to such functional analysis is the current difficulty of identifying RNA-binding partners that associate with specific transcripts and participate in their biological behavior. Despite the development of various RNA affinity purification methods, the expense and/or technical challenges associated with each approach has precluded its use by nonspecialists or in high-throughput discovery experiments. Current strategies for identifying RNA-binding proteins that associate with specific transcripts involve the use of affinity tags, including biotin, aptamers, and particular protein-binding sequences (9–13). In each case, however, the modest affinity or specificity of tag recognition and the difficulty of selective elution complicate sample analysis. A strategy that will allow simple and rapid identification of proteins that associate selectively with particular transcripts will expand our understanding of RNA biology.

Here we present a highly effective method of RNA–protein complex purification and analysis based on an engineered version of the Csy4 endoribonuclease. Three properties of this enzyme

lend it to selective RNA isolation: small size, high-affinity binding to a short hairpin sequence, and robust site-specific cleavage activity. In addition, we found that a single active site histidine mutation inactivates RNA cleavage activity in a manner that is readily reversible after the addition of imidazole. This activatable version of Csy4 allows a simple procedure in which tagged transcripts, together with their associated binding proteins and/or nucleic acids, can be selectively purified for analysis by MS, Western blot analysis, and next-generation sequencing. We validate this method as a superior strategy for identifying specific RNA-binding proteins, and propose its use in the search for human pre-miRNA-binding partners.

## Results

**Csy4 H29A Is an Inducible Endoribonuclease.** Csy4 is an endoribonuclease that participates in the *Pseudomonas aeruginosa* UCBPP-PA14 CRISPR system by cleaving transcripts bearing a short hairpin sequence to liberate guide RNAs used for adaptive immunity against foreign DNA (14–16). Previous studies have shown that Csy4 selectively recognizes a 16-nt hairpin sequence with exceptionally high affinity (equilibrium dissociation constant,  $K_d = 50$  pM) and is relatively intolerant of sequence changes in the substrate hairpin stem (15). Furthermore, the active site of Csy4 contains an essential histidine residue (H29) that functions as a general base during RNA strand scission (Fig. 1A) (16). Mutation of H29 to alanine inactivates Csy4 without affecting substrate binding affinity or specificity (14, 15). We found that the cleavage activity of Csy4 H29A can be rescued in the presence of imidazole, which substitutes for the imidazole side chain of histidine in the active site (Fig. 1B). We made use of the conditional activity of Csy4 H29A to develop a highly selective affinity purification method for the analysis of RNA–protein complexes.

**Selective Ribonucleoprotein Complex Isolation Using Csy4.** Based on the crystal structure of WT Csy4, a serine residue (S50) on the protein surface was mutated to cysteine to enable coupling to biotin. We developed an RNA affinity purification strategy in which RNAs containing a cleavable Csy4 hairpin tag at the 5' end are captured by biotinylated Csy4 H29A/S50C (designated Csy4\* in this study) immobilized on avidin resin (Fig. 2A). To test the efficiency of Csy4\*-based RNA affinity purification using cell extract, we used

Author contributions: H.Y.L., R.E.H., L.B., and J.A.D. designed research; H.Y.L., R.E.H., A.A., and K.Z. performed research; H.Y.L., R.E.H., A.A., B.S., and C.D.W. contributed new reagents/analytic tools; H.Y.L., R.E.H., A.A., S.L., L.B., and J.A.D. analyzed data; and H.Y.L., R.E.H., and J.A.D. wrote the paper.

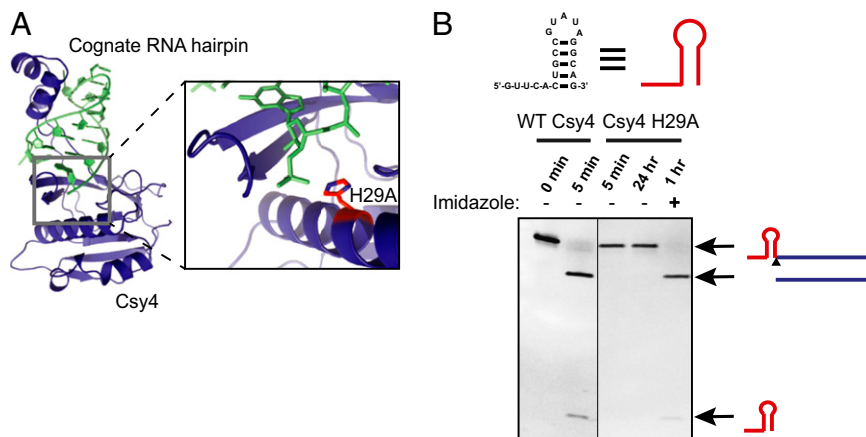
Conflict of interest statement: The Regents of the University of California (R.E.H. and J.A.D.) have filed a related patent, Endoribonuclease Compositions and Methods of Use Thereof.

Freely available online through the PNAS open access option.

<sup>1</sup>Present address: Caribou Biosciences, Inc., Berkeley, CA 94720.

<sup>2</sup>To whom correspondence should be addressed. E-mail: doudna@berkeley.edu.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1302807110/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1302807110/-DCSupplemental).



**Fig. 1.** Csy4 H29A is a conditional endoribonuclease. (A) Crystal structure of WT Csy4 bound to its RNA substrate (Protein Data Bank ID code 4AL5) (Left) and close-up of the Csy4 active site (Right). The active site histidine (H29; shown in red) serves as the general base for cleavage of the phosphate backbone at the base of the RNA hairpin. (B) In vitro cleavage activity of WT and mutant Csy4. Cleavage products were separated by 12% denaturing PAGE and visualized with ethidium bromide staining.

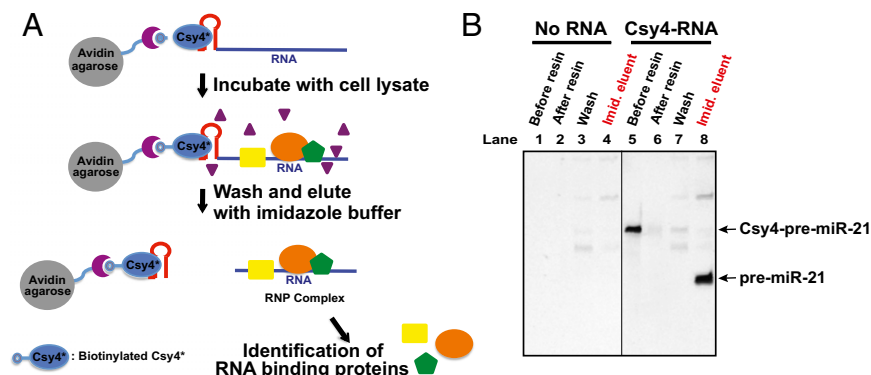
in vitro transcription to generate a pre-miR-21 RNA transcript containing a 5' 20-nt tag (designated the Csy4 hairpin tag in this study) comprising four single-stranded nucleotides followed by the 16-nt hairpin (80-nt total pre-miRNA transcript length). We preincubated the tagged transcript, Csy4\*, and avidin agarose resin, forming a resin-bound Csy4\*–RNA complex. We then incubated either HeLa or NTERA2 (NT2) cell extract with the Csy4\*–RNA-bound resin to capture pre-miRNA (pre-miR)-21-specific RNA-binding proteins. We washed the samples three times with a buffer containing 100–150 mM NaCl to remove nonspecific proteins and cellular RNAs. RNA–protein complexes still associated with the beads were eluted by the addition of 500 mM imidazole overnight at 4 °C. Imidazole activates the site-specific cleavage activity of Csy4\*, thereby releasing RNA–protein complexes by removing the 5' hairpin tag. This procedure showed near-quantitative RNA binding by biotinylated Csy4\* (Fig. 2B, lanes 5 and 6) with little loss of RNA during the washing step (Fig. 2B, lane 7). As expected, only one cleaved RNA species was recovered after the addition of imidazole (Fig. 2B, lane 8).

**Direct Comparison of Csy4\* and Biotin Affinity Purification Methods.** Key challenges in RNA–protein affinity purification are the

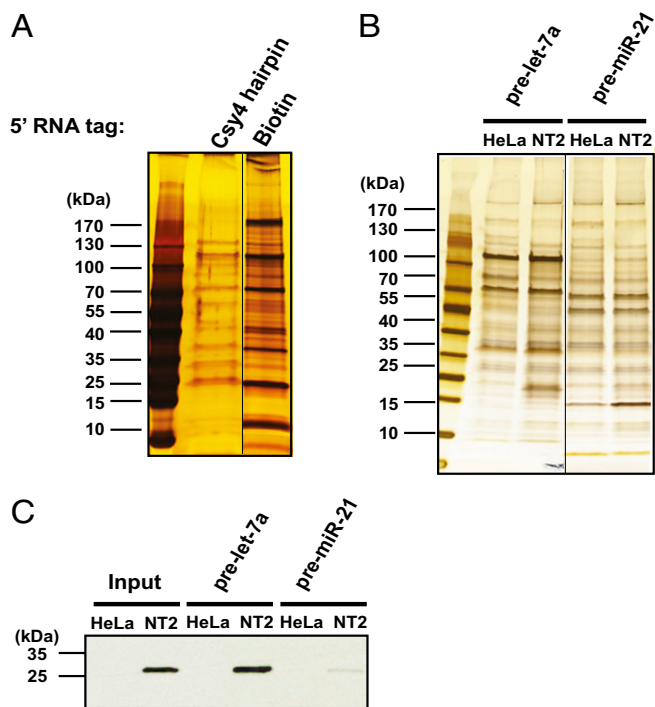
retention of specifically bound proteins and the removal of nonspecific proteins. To test the efficacy of Csy4\* for selective RNA–protein isolation, we prepared a human pre-miRNA, pre-let-7a, with either a 5' Csy4 hairpin tag or a 5' biotin moiety. Biotinylation of RNA is one of the most widely used methods for RNA–protein affinity purification. Both transcripts were incubated in either HeLa or NT2 cell extract as described above, followed by immobilization on beads coupled to Csy4\* or streptavidin, respectively. After washing as described above, Csy4 hairpin-tagged transcripts were eluted with imidazole, and biotinylated transcripts were eluted by boiling the resin in an SDS buffer.

We observed a dramatic difference in the amount of background protein contamination between the two elution samples (Fig. 3A). When hairpin-tagged transcripts were instead eluted by boiling in SDS buffer, higher background levels were observed (Fig. S1). These results indicate that specific elution via enzymatic cleavage is the key to selective RNA–protein complex isolation in Csy4\*-based RNA–protein affinity purification.

**Csy4\*-Based Affinity Purification of a Known pre-miRNA–Protein Binding Partner.** Lin28A controls the processing of pre-let-7a to produce the let-7a miRNA in a developmentally regulated fashion by binding to the loop of the pre-let-7a hairpin (17). To validate



**Fig. 2.** Csy4-based pull-down strategy using a conditional endoribonuclease. (A) Scheme of Csy4-based RNA affinity purification. Avidin agarose, biotinylated Csy4\*, and Csy4 hairpin-tagged transcript are preincubated to make a ternary complex, which is then incubated with cell extract. After washing, the RNA–protein complexes are eluted by cleaving off the Csy4 hairpin tag and releasing RNA–protein complexes on the addition of imidazole. (B) Efficient RNA recovery of Csy4-based RNA purification with HeLa cell extract. Tagged RNA transcripts, biotinylated Csy4\*, and avidin agarose resin were incubated for 2 h at 4 °C and subjected to mock purification. The samples were separated on 12% denaturing PAGE and visualized by staining with SYBR Gold (Invitrogen).



**Fig. 3.** Csy4\*-based RNA affinity purification isolates known RNA-binding proteins. (A) Comparison of Csy4\*-based and biotinylated RNA purification for pre-let-7a. Both transcripts were tagged at the 5' end. Eluted samples were separated via 4–20% SDS/PAGE and visualized by silver staining. (B) Csy4 hairpin RNA-tagged pre-let-7a copurifies with Lin28A. Csy4 hairpin RNA-tagged pre-let-7a and pre-miR-21 were used as bait for Csy4\*-based RNA affinity purification from HeLa and NT2 cellular extracts. Pull-down eluents were analyzed by SDS/PAGE and visualized by silver staining (Left). (C) Western blots of Csy4\*-based pull-down eluents from B using an anti-Lin28A antibody.

the Csy4\* RNA–protein affinity purification strategy, we tested whether a pre-let-7a:Lin28A complex could be isolated using this method. Lin28A is highly expressed in embryonic stem cells and is expressed at trace levels in somatic cells (18). We generated a 5' Csy4 hairpin-tagged pre-let-7a transcript by *in vitro* transcription, and used HeLa and NT2 cell extracts for complex purification. NT2 cells are embryonic carcinoma cells with similar characteristics as embryonic stem cells, including high Lin28A expression levels (19).

After incubation, washing, and imidazole elution as described above, samples were analyzed with silver-stained SDS/PAGE gels and Western blot analysis using a Lin28A-specific antibody (Fig. 3 B and C). This analysis showed that Lin28A was isolated in complex with pre-let-7a RNA from NT2 cell extract, but not from HeLa cell extract. In a control experiment, Csy4 RNA hairpin-tagged pre-miR-21 had little interaction with Lin28A (Fig. 3C), supporting the conclusion that the pre-let-7a–Lin28A interaction is RNA-specific. We also tested Csy4 hairpin-tagged histone mRNA (486 nt), a transcript known to interact with the stem loop binding protein (SLBP) (20). We found that Csy4\* affinity purification can selectively isolate SLBP–histone mRNA complexes from NT2 cell extracts, whereas a control histone mRNA transcript in which the SLBP-binding hairpin of histone mRNA is swapped to the U1A hairpin does not copurify with SLBP (Fig. S2).

**Unique pre-miRNA–Binding Proteins Identified by Csy4\* Affinity Purification.** The cellular levels of many miRNAs are highly regulated depending on developmental stage and tissue type (21, 22). Several specific RNA-binding proteins, including Lin28A, KH-type splicing regulatory protein (KSRP), TAR DNA binding

protein-43 (TDP43), and heteronuclear ribonucleoprotein A1 (hnRNP A1), have been shown to bind precursor miRNAs and to affect miRNA biogenesis (17, 23–25). However, most miRNAs and their precursor transcripts have not been analyzed for specific binding proteins that may regulate their processing and biological functions.

To identify unique specific pre-miRNA–binding proteins using the Csy4\*-based RNA affinity purification, we tested 10 different pre-miRNAs in either HeLa or NT2 cell extract (Figs. 3B and 4A and Fig. S1). Based on an analysis of eluted proteins using silver-stained polyacrylamide gels, we detected unique RNA-binding proteins specific to particular pre-miRNAs (Fig. S1). In addition, some proteins interacted with specific pre-miRNAs in a cell-type-dependent manner, suggesting the possible existence of pre-miRNA–binding proteins similar to Lin28A that influence the processing of pre-miRNAs during development or in different tissue types (Fig. 4A and Fig. S3).

Of note, most of these specific pre-miRNA–binding proteins were detectable only in Csy4\*-based RNA–protein affinity purification samples that were eluted with imidazole, not those eluted by boiling in SDS buffer (Fig. S1). This result demonstrates that Csy4\* cleavage-based elution enables the detection of specific RNA-binding proteins present in low abundance relative to other proteins in the cell extract. We also found that Lin28A bound to several, but not all, of the pre-miRNAs that we tested, including pre-miR-343, pre-miR-200a, and pre-miR-17 (Fig. 4B and Fig. S4), in addition to pre-let-7a. This finding suggests that Lin28A may be involved in regulating a broader range of human miRNAs than has been previously appreciated.

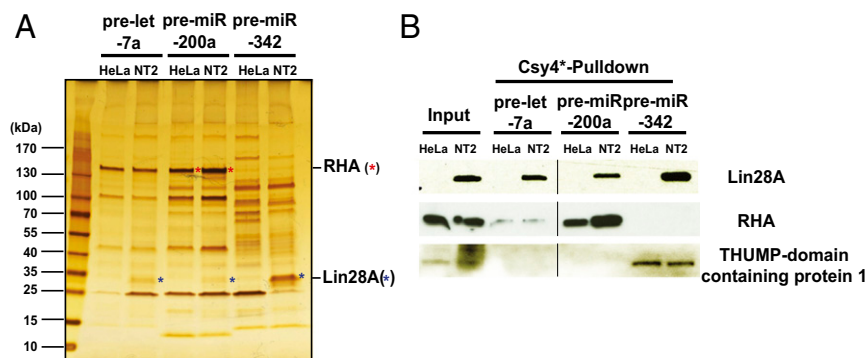
**Proteomic Analysis of Csy4\* Affinity Purification Samples: Development of the MS Method.** After affinity purification of RNA–protein complexes, proteomic analysis is critical for determining the composition of the samples. Typically, after RNA affinity purification, the bands of interest in a protein gel are excised and analyzed by MS. Because Csy4\*-based affinity purification yields samples of superior quality with few nonspecific proteins, we developed an MS analysis method that analyzes whole eluent samples without a gel purification step.

We modified the filter-aided sample preparation methodology described previously (26) to prepare our samples for MS analysis. In this method, molecular weight cutoff (MWCO) filters were used to remove any detergents and salts used for the affinity purification, which allows the flexible use of detergents and buffers without interfering with subsequent MS analyses. In the current protocol, the key features modified with respect to the previously published methods (26, 27) include the use of trifluoroethanol (TFE) in addition to urea for solubilization of the protein samples and the substitution of Tris(2-carboxyethyl)phosphine (TCEP) for DTT in protein reduction. The use of TFE in place of or in addition to urea alone results in improved protein coverage by LC-MS in this experiment.

To prepare samples for MS, the eluent samples from Csy4\* affinity purification were rinsed twice in 8 M urea using MWCO filters, followed by the addition of 40% TFE for efficient denaturation of the target proteins, allowing the effective removal even of detergents with low critical micelle concentrations, such as Triton X-100 (27). After the reduction of disulfide bonds between cysteine residues with TCEP and alkylation of free thiol groups with iodoacetamide to prevent disulfide bond formation during downstream peptide mapping by MS, the proteins were digested in two steps with endoproteinase LysC and then with trypsin. Finally, the digested peptides were rinsed through a MWCO filter, and the flow-through sample was collected for subsequent LC-MS analysis (Fig. S5).

**Identification of RNA- and Cell Line-Specific pre-miRNA–Binding Proteins.** Using the purification and analytic methodology described above,





**Fig. 4.** Identification of RNA- and cell line-specific pre-miRNA-binding proteins using Csy4\*-based RNA affinity purification. (A) In vitro-transcribed, Csy4 hairpin-tagged pre-let-7a, pre-miR-200a, and pre-miR-342 were used as bait for Csy4\*-based affinity purification from HeLa and NT2 cell extracts. Equal amounts of purified samples were separated via 4–20% SDS/PAGE and visualized by silver staining. (B) Purified samples from A were separated by SDS/PAGE and probed for the presence of Lin28A, ATP-dependent RHA, and THUMP domain-containing protein 1 by Western blot analysis.

we tested whether RNA- and cell line-specific binding proteins could be identified for three human pre-miRNAs: pre-let-7a, pre-miR-200a, and pre-miR-342. These RNAs are precursors to mature miRNAs whose activities are implicated in the progression of cancers in somatic and embryonic tissues (28–31). Csy4\*-purified samples of pre-let-7a, pre-miR-200a, and pre-miR-342 complexes using HeLa (somatic) and NT2 (embryonic) cell extracts were analyzed by MS. The majority of the proteins detected were

annotated RNA-binding proteins (Table 1). Some of the proteins found were associated with only one of the three pre-miRNAs; for example, the THUMP domain-containing protein was found in complex with pre-miR-342, but not with pre-let-7a or pre-miR-200a. Other proteins were observed in only one of the two cell extracts tested; for example, Lin28A and 60S ribosomal protein L13 were found in samples isolated using NT2 cells and HeLa cells, respectively (Table 1). Western blot analysis for proteins

**Table 1.** MS analysis of hairpin-tagged pre-miRNA affinity purification eluent samples

Accession no. (UniProt database)	Name	pre-let-7a		pre-miR-200a		pre-miR-342	
		HeLa (8)	NT2 (11)	HeLa (10)	NT2 (10)	HeLa (14)	NT2 (7)
Q08211	ATP-dependent RNA helicase A (RHA)	o		o	o		
P05388	60S acidic ribosomal protein P0	o	o			o	
P05386	60S acidic ribosomal protein P1		o			o	
P05387	60S acidic ribosomal protein P2		o	o	o	o	o
P26373	60S ribosomal protein L13	o				o	
P18124	60S ribosomal protein L7					o	
O60318	80-kDa MCM3-associated protein	o					
P62633	Cellular nucleic acid-binding protein			o	o	o	o
P12107	Collagen alpha-1(XI) chain		o				
Q07021	Complement component 1 Q subcomponent-binding protein, mitochondrial			o	o		
Q02413	Desmoglein-1		o				
P16989	DNA-binding protein A			o	o	o	
P09651	Heterogeneous nuclear ribonucleoprotein A1					o	
Q14103	Heterogeneous nuclear ribonucleoprotein D0			o		o	o
Q12905	Interleukin enhancer-binding factor 2	o			o		
Q12906	Interleukin enhancer-binding factor 3	o		o	o		
Q8N371	JmjC domain-containing protein 5		o				
Q9P2V4	Leucine-rich repeat, Ig-like domain and transmembrane domain-containing protein1	o	o				
Q9H9Z2	Lin-28 homolog A		o		o		o
P05455	Lupus La protein			o	o	o	
P19338	Nucleolin	o		o	o	o	o
Q13415	Origin recognition complex subunit 1					o	o
Q8NC51	Plasminogen activator inhibitor 1 RNA-binding protein		o				
Q9H2U1	Probable ATP-dependent RNA helicase DHX36			o			
Q9NXG2	THUMP domain-containing protein 1					o	o
P62988	Ubiquitin		o				
O94763	Unconventional pre-foldin RPB5 interactor		o			o	

Pre-let-7a, pre-miR-200a, and pre-miR-342 transcripts were used as RNA baits, and HeLa and NT2 cell extracts were used for each RNA pull-down experiment. For each experiment, two biological replicate samples were used, and three technical MS replicate runs were performed for each sample, for a total of six MS runs for each experiment. Proteins with a false discovery rate <1% and detected in at least three of six MS runs are listed. The parenthetical number indicates the number of proteins detected by MS.

including Lin28A, ATP-dependent RNA helicase A (RHA), and THUMP domain-containing protein verified the MS results (Fig. 4B and Table 1). These findings demonstrate the specificity of the method described herein, and provide evidence of an abundance of RNA-binding proteins that selectively regulate human pre-miRNA processing. We believe that this method can be broadly applied to identify RNA-binding proteins of transcripts of interest.

## Discussion

Several different affinity purification strategies have been developed to identify proteins that bind to specific RNA molecules (9–12). Unfortunately, however, none of the existing methods has proven suitable for widespread or high-throughput use owing to their associated costs and/or technical challenges. In particular, the modest affinity or specificity of transcript recognition, lack of selective sample elution, and lack of quantitative MS procedures have hampered the widespread adoption of methodologies including biotinylation, aptamers, and bacteriophage coat protein recognition sites.

A previous study used a variant hepatitis delta virus (HDV) ribozyme with imidazole-dependent activity for non-denaturing RNA purification (32). The length of the tag in this method, which includes two signal recognition particle RNAs, is quite long (~180 nt) and the affinity between signal recognition particle RNA and the *T.maritima* SRP Ffh M-domain (TmaM) protein to which it binds is much lower than that of the Csy4\*–RNA hairpin interaction. The use of an activatable version of the Csy4 endoribonuclease for RNA–protein affinity purification circumvents these technical and practical challenges to meet a critical need in transcriptome analysis. The Csy4 hairpin tag is only 16 nt, with a 5-bp stem and a 5-nt loop forming the hairpin. It is unlikely that this hairpin folding would be interrupted by other sequences in the tagged transcript; however, insertion of single-stranded RNA (ssRNA)-flanking regions adjacent to the Csy4 hairpin may help ensure Csy4 hairpin tag folding.

The ability of Csy4\* to bind with exceptionally high affinity to Csy4 hairpin-tagged RNAs and to be activated only in the presence of imidazole enables selective transcript isolation with little background contamination. Because of the intrinsic purity of samples isolated using this approach, gel purification is not required before MS analysis. In addition, the expression and purification of Csy4\* is very robust (5–10 mg/L). Thus, the method is sufficiently simple and inexpensive to be suitable for routine analysis of individual RNA samples.

Csy4\* affinity purification is versatile; although we have demonstrated its use for 5'-Csy4 hairpin-tagged pre-miRNAs, it has potential application in other situations as well. For example, aCsy4 hairpin tag could be inserted in the middle of the transcripts as ssRNA nucleotides flanking the Csy4 hairpin contribute negligibly to Csy4 binding and cleavage activity (15). In addition, the use of Csy4 affinity purification to purify native RNP in cells could be an interesting application. Whereas commonly used biotinylated transcripts for RNA affinity purification have limited access for in vivo application, given that limited transfection efficiency and high cost of obtaining sufficient quantity of biotinylated transcripts for transfection, Csy4-tagged transcripts could be readily expressed in cells, and native RNP could be purified using the Csy4\* affinity purification method after cell lysis.

Moreover, the method that we present here can be readily adapted for high-throughput analysis of multiple RNAs in parallel. Our method provides a simple and effective strategy for transcriptome analysis that will enable the study of many previously intractable systems to reveal the composition of RNA–protein complexes and their functions in a wide variety of biological systems.

## Materials and Methods

**Preparation of Biotinylated Csy4 H29A/S50C.** Csy4 H29A/S50C was prepared using the same procedure as for WT Csy4 (14). To 1 mg of Csy4 H29A/S50C in 455  $\mu$ L of buffer [100 mM Hepes (pH 7.5), 150 mM KCl, 1 mM TCEP, 5% (vol/vol) glycerol], we added 45  $\mu$ L of EZ-Link maleimide-PEG2-biotin (10 mg/mL stock in 1 $\times$  PBS; Pierce) and incubated at room temperature overnight. The biotinylated Csy4 H29A/S50C was dialyzed overnight at 4  $^{\circ}$ C to remove free maleimide-PEG2-biotin. Biotinylated Csy4 H29A/S50C was removed from the dialysis membrane, aliquoted, and flash-frozen in liquid nitrogen.

**Preparation of Hairpin-Tagged Transcripts.** DNA templates containing the T7 promoter, RNA tag, target transcript, and HDV ribozyme were used for in vitro transcription to prepare tagged RNAs. For in vitro transcription, 5  $\mu$ g of DNA template was incubated with 5 mM each NTP (ATP, GTP, CTP, and UTP), 25 mM MgCl<sub>2</sub>, 1 mM DTT, 0.01% Triton X-100, 2 mM spermidine, 0.1 U of inorganic pyrophosphatase, 10  $\mu$ g of T7 RNA polymerase, and 30 mM Tris (pH 8.1) in a 100- $\mu$ L reaction. After in vitro transcription, RNAs were purified by denaturing PAGE. The following RNA sequences were used in this study (with the Csy4 hairpin tag sequences underscored): Csy4–pre-let-7a, 5'-GUUCACUGCCGUAUAGGCAGUGAGGUAGUAGGUUGUAUAGUUUAGGGU-CACACCCACCAUGGGAGAUAAUAUACAACUACUGUCUUACC-3'; Csy4–pre-miR-21, 5'-GUUCACUGCCGUAUAGGCAGUAGCUUAUCAGACUGAUGUUGAC-UGUUGAAUCUCAUGGCAACACCCAGUCGUAUGGGGUGUC-3'; Csy4–pre-miR-200a, 5'-GUUCACUGCCGUAUAGGCAGAUUUACCGGACAGUCUGGAUUUCCCA-GCUUGACUCUAAACUGUCUGGUAACGAUGU-3'; Csy4–pre-miR-342, 5'-GUUCACUGCCGUAUAGGCAGAGGGGUGCUAUCUGUGAUUAGGGGACAUGGUU-AAUGGAAUUGUCUCACACAGAAAUCGCACCCGUC-3'.

**In Vitro Cleavage Assays.** For these assays, 5  $\mu$ M WT Csy4 or Csy4 H29A was added to 0.5  $\mu$ M Csy4 hairpin-tagged RNA in the reaction buffer [20 mM Tris (pH 7.5), 100 mM NaCl, and 1 mM TCEP], and the mixture was incubated at room temperature. The reaction was quenched at each time point with 1.2 volumes of 2 $\times$  loading buffer [95% (wt/vol) formamide, 18 mM EDTA, 0.025% SDS, 0.1% xylene cyanol, and 0.1% bromophenol blue] and boiled at 70  $^{\circ}$ C for 10 min before being loaded onto 12% (wt/vol) denaturing polyacrylamide gel.

**Preparation of Cell Extract.** First, 1 L of HeLa cells were harvested in log phase (4–5  $\times$  10<sup>5</sup> cells/mL), washed twice with PBS, and frozen. Then the HeLa cell pellet was thawed, and 2.5 mL of hypotonic lysis buffer [10 mM Hepes (pH 7.6), 10 mM potassium acetate, 1.5 mM magnesium acetate, and 2 mM TCEP], adjusted to pH 7.6 with potassium hydroxide, was added, along with 2 $\times$  protease inhibitor and 2 $\times$  phosphatase inhibitor (Roche). Samples were swollen on ice for 10 min, and then dounced with ~20 strokes with a type B (tight) pestle. Samples were spun at 640  $\times$  g for 5 min and then at 10,400  $\times$  g for 20 min. Additional 5- or 10-min spins were used to remove any precipitation as necessary. Because we used frozen cells and a tight pestle to dounce cells, the resulting extract contained not only cytoplasmic extract, but also nuclear extract, as validated by Western blot analysis with topoisomerase II-beta Ab.

**General Procedure for Csy4\*-Based Pull-Down.** In this procedure, 50 pmol of Csy4 hairpin-tagged transcript in 125  $\mu$ L of DEPC water was boiled at 70  $^{\circ}$ C for 10 min and then chilled on ice for 2 min. Then 125  $\mu$ L of 2 $\times$  RNA folding buffer [40 mM Tris (pH 7.5), 200 mM NaCl, and 6 mM MgCl<sub>2</sub>] was added, followed by incubation at 37  $^{\circ}$ C for 10 min. After this, 100 pmol of biotinylated Csy4\* was added, and the resulting mixture was incubated for 30–60 min at 4  $^{\circ}$ C with gentle rocking. Then 25  $\mu$ L of avidin agarose resin (Thermo Scientific) was washed twice with 300  $\mu$ L of 1 $\times$  binding buffer [20 mM Tris (pH 7.5), 100 mM NaCl, 0.1% Triton X-100, 5% (vol/vol) glycerol, and 1 mM TCEP], with centrifugation at 1,000  $\times$  g for 1 min after each wash. The washed resin was mixed with the preincubated RNA and biotinylated Csy4\* and incubated for 1 h at 4  $^{\circ}$ C. Samples were spun for 1 min at 1,000  $\times$  g, and the resulting supernatant was removed, followed by the addition of 150–200  $\mu$ L of fresh 1 $\times$  binding buffer with 1 $\times$  protease inhibitor (Roche), 1 $\times$  phosphatase inhibitor (Roche), and 5  $\mu$ L RNasin (Promega). Approximately 1 mg of cell extract was added, to make a total volume of 250  $\mu$ L, and the sample was incubated for 1–2 h at 4  $^{\circ}$ C. The resin was washed three times with 500  $\mu$ L of washing buffer [20 mM Tris (pH 7.5), 100–200 mM NaCl, 0.1% Triton X-100, 5% (vol/vol) glycerol, and 1 mM TCEP]. For each wash, the resin was incubated with washing buffer for 3–5 min at 4  $^{\circ}$ C with gentle rocking.

Finally, RNA–protein complexes were eluted by adding 50  $\mu$ L of elution buffer [20 mM Tris (pH 7.5), 100 mM NaCl, 5% (vol/vol) glycerol, 1 mM

TCEP, and 500 mM imidazole] and incubating overnight at 4 °C. After the resin was spun at 1,000 × g for 1 min, supernatant was obtained as an eluent and analyzed by silver staining, Western blot analysis, and MS.

**MS Sample Preparation.** To minimize the presence of detergents in the samples analyzed by MS, we washed the sample-containing resin two or three times with buffer lacking detergent [20 mM Tris (pH 7.5), 100 mM NaCl, 5% (vol/vol) glycerol, and 1 mM TCEP] before the elution of imidazole. Samples were eluted as described above. The Csy4\* pull-down eluent sample (250 μL) was added to a filter tube with 30-kDa MWCO filter (Millipore) and then centrifuged at 14,000 × g for 45 min at 20 °C, with the flow-through discarded. Then 200 μL of 8 M urea/0.1 M ammonium bicarbonate (ABC), pH 8.5, in water supplemented with 50 mM TCEP was added, followed by centrifugation at 14,000 × g for 45 min at 20 °C, with the flow-through discarded. Then 200 μL of 50 mM TCEP (71.7 mg/5 mL) in 40% (vol/vol) TFE/0.1 M ABC (pH 8.5) was added to the sample, followed by centrifugation at 14,000 × g for 45 min at 20 °C, with the flow-through discarded. This wash was repeated once; then 200 μL of the iodoacetamide (IAM) solution [50 mM IAM, 40% (vol/vol) TFE, and 0.1 M ABC] was added, and the mixture was incubated for 60 min at room temperature in the dark. The filter tube containing the sample was centrifuged at 14,000 × g for 60 min at 20 °C, and the flow-through was discarded. A total of 200 μL of 40% (vol/vol) TFE/0.1 M ABC was added to each well, followed by centrifugation at 14,000 × g for 60 min at 20 °C, with the flow-through discarded. This wash was repeated once. Then 50 μL of 40% (vol/vol) TFE/0.1 M ABC containing endoproteinase LysC at a 1:50 (enzyme:substrate) ratio was added to the sample (~1 μg per well), followed by overnight incubation at room temperature under a hood and covered. Then 200 μL of the trypsin solution (1 μg/200 μL trypsin in 0.1 M ABC) was added at a 1:50 (enzyme:substrate) ratio,

covered tightly, and incubated for 4 h at 37 °C. (Note that the addition of this volume buffer reduced the concentration of TFE to <10%, under which conditions trypsin is active.)

After digestion, the filter tube containing the sample was centrifuged at 14,000 × g for 1 h into a polypropylene receiving tube. Then 100 μL of 70% (vol/vol) acetonitrile (ACN)/0.1% formic acid (FA) was added, followed by centrifugation at 14,000 × g for 60 min at 20 °C into the tube (total volume now 350 μL). The received sample was dried in a SpeedVac (Savant) overnight and then stored in a –20 °C freezer until use in MS analyses. For MS analysis, samples were dissolved in 20 μL of Milli-Q water (purified using Milli-Q water purification system) with 0.1% FA and then placed in an HPLC autosampler.

**MS Analysis.** All analyses were run on an Agilent HPLC-Chip/MS System, composed of a micro-autosampler with a thermostat (set to 4 °C), a capillary and nanoflow pump with micro-degasser, and the Chip-Cube that interfaces LC modules and the MS instrument. HPLC-grade water (0.1% FA) and ACN (0.1% FA) were used as mobile phases A and B, respectively. ACN was obtained from Merck (Germany), water from a Milli-Q water purification system, and the FA and TFA were obtained from Sigma-Aldrich. Separations were conducted on a Polaris-HR-Chip 3C18 (150 mm × 50 μm, 80A 3 μm C18 chip with 360 nL trap column). Sample analysis used a 45-min gradient operating on the nanopump (Table S1). The capillary pump provides a constant flow of 2 μL/min for delivery of samples from the autosampler to the HPLC-Chip interface.

Mass detection was performed with an Agilent 6530 Accurate-Mass Q-TOF LC/MS System operated in positive-ion mode. An Agilent MassHunter Workstation software was used for data acquisition and processing, and an Agilent Spectrum Mill MS Proteomics Workbench was used for database searches with Swiss-Prot human, with the peptide global false discovery rate set at 1%.

- Mansfield KD, Keene JD (2009) The ribonome: A dominant force in co-ordinating gene expression. *Biol Cell* 101(3):169–181.
- Castello A, et al. (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* 149(6):1393–1406.
- Scheibe M, Butter F, Hafner M, Tuschl T, Mann M (2012) Quantitative mass spectrometry and PAR-CLIP to identify RNA–protein interactions. *Nucleic Acids Res* 40(19):9897–9902.
- Wang X, et al. (2008) Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature* 454(7200):126–130.
- Lewis BP, Burge CB, Bartel DP (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120(1):15–20.
- Matera AG, Terns RM, Terns MP (2007) Non-coding RNAs: Lessons from the small nuclear and small nucleolar RNAs. *Nat Rev Mol Cell Biol* 8(3):209–220.
- Huarte M, Rinn JL (2010) Large non-coding RNAs: Missing links in cancer? *Hum Mol Genet* 19(2):R152–R161.
- Tsai MC, et al. (2010) Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 329(5992):689–693.
- Youngman EM, Green R (2005) Affinity purification of in vivo-assembled ribosomes for in vitro biochemical analysis. *Methods* 36(3):305–312.
- Bachler M, Schroeder R, von Ahnen U (1999) StreptoTag: A novel method for the isolation of RNA-binding proteins. *RNA* 5(11):1509–1516.
- Hartmuth K, et al. (2002) Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. *Proc Natl Acad Sci USA* 99(26):16719–16724.
- Srisawat C, Engelke DR (2001) Streptavidin aptamers: Affinity tags for the study of RNAs and ribonucleoproteins. *RNA* 7(4):632–641.
- Hogg JR, Collins K (2007) RNA-based affinity purification reveals 75K RNPs with distinct composition and regulation. *RNA* 13(6):868–880.
- Haurwitz RE, Jinek M, Wiedenheft B, Zhou K, Doudna JA (2010) Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* 329(5997):1355–1358.
- Sternberg SH, Haurwitz RE, Doudna JA (2012) Mechanism of substrate selection by a highly specific CRISPR endoribonuclease. *RNA* 18(4):661–672.
- Haurwitz RE, Sternberg SH, Doudna JA (2012) Csy4 relies on an unusual catalytic dyad to position and cleave CRISPR RNA. *EMBO J* 31(12):2824–2832.
- Viswanathan SR, Daley GQ (2010) Lin28: A microRNA regulator with a macro role. *Cell* 140(4):445–449.
- Thomson JM, et al. (2006) Extensive post-transcriptional regulation of microRNAs and its implications for cancer. *Genes Dev* 20(16):2202–2207.
- Fong YW, et al. (2011) A DNA repair complex functions as an Oct4/Sox2 coactivator in embryonic stem cells. *Cell* 147(1):120–131.
- Whitfield ML, et al. (2004) SLBP is associated with histone mRNA on polyribosomes as a component of the histone mRNP. *Nucleic Acids Res* 32(16):4833–4842.
- Ro S, Park C, Young D, Sanders KM, Yan W (2007) Tissue-dependent paired expression of miRNAs. *Nucleic Acids Res* 35(17):5944–5953.
- Bartel DP (2004) MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* 116(2):281–297.
- Michlewski G, Cáceres JF (2010) Antagonistic role of hnRNP A1 and KSRP in the regulation of let-7a biogenesis. *Nat Struct Mol Biol* 17(8):1011–1018.
- Trabucchi M, et al. (2009) The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. *Nature* 459(7249):1010–1014.
- Kawahara Y, Mieda-Sato A (2012) TDP-43 promotes microRNA biogenesis as a component of the Drosha and Dicer complexes. *Proc Natl Acad Sci USA* 109(9):3347–3352.
- Wiśniewski JR, Zougman A, Nagaraj N, Mann M (2009) Universal sample preparation method for proteome analysis. *Nat Methods* 6(5):359–362.
- Wiśniewski JR, Zielinska DF, Mann M (2011) Comparison of ultrafiltration units for proteomic and N-glycoproteomic analysis by the filter-aided sample preparation method. *Anal Biochem* 410(2):307–309.
- Esquela-Kerscher A, Slack FJ (2006) Oncomirs: MicroRNAs with a role in cancer. *Nat Rev Cancer* 6(4):259–269.
- Johnson SM, et al. (2005) RAS is regulated by the let-7 microRNA family. *Cell* 120(5):635–647.
- Cittelly DM, et al. (2010) Down-regulation of miR-342 is associated with tamoxifen-resistant breast tumors. *Mol Cancer* 9:317.
- Park S-M, Gaur AB, Lengyel E, Peter ME (2008) The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. *Genes Dev* 22(7):894–907.
- Kieft JS, Batey RT (2004) A general method for rapid and nondenaturing purification of RNAs. *RNA* 10(6):988–995.