

# miCLIP-MaPseq, a Substrate Identification Approach for Radical SAM **RNA Methylating Enzymes**

Vanja Stojković,<sup>†</sup> Tongyue Chu,<sup>†</sup> Gabriel Therizols,<sup>†</sup> David E. Weinberg,<sup>†</sup> and Danica Galonić Fujimori\*,<sup>†,‡</sup>

<sup>†</sup>Department of Cellular and Molecular Pharmacology, University of California, San Francisco, California 94158, United States <sup>‡</sup>Department of Pharmaceutical Chemistry, University of California, 600 16th Street, MC2280 San Francisco, California 94158, United States

Supporting Information

ABSTRACT: Although present across bacteria, the large family of radical SAM RNA methylating enzymes is largely uncharacterized. Escherichia coli RlmN, the founding member of the family, methylates an adenosine in 23S rRNA and several tRNAs to yield 2-methyladenosine (m<sup>2</sup>A). However, varied RNA substrate specificity among RlmN enzymes, combined with the ability of certain family members to generate 8-methyladenosine (m<sup>8</sup>A), makes functional predictions across this family challenging. Here, we present a



method for unbiased substrate identification that exploits highly efficient, mechanism-based cross-linking between the enzyme and its RNA substrates. Additionally, by determining that the thermostable group II intron reverse transcriptase introduces mismatches at the site of the cross-link, we have identified the precise positions of RNA modification using mismatch profiling. These results illustrate the capability of our method to define enzyme-substrate pairs and determine modification sites of the largely uncharacterized radical SAM RNA methylating enzyme family.

#### INTRODUCTION

RNA methylation is the most common post-transcriptional RNA modification and is implicated in diverse biological processes including regulation of translation, antibiotic resistance and stress response. While numerous methyl marks decorate RNA molecules in all domains of life, knowledge of their locations, functions, and corresponding RNA methylating enzymes remains limited. Most RNA methylating enzymes are thought to modify only a single type of RNA molecule; however, due to advances in high-throughput methods, new RNA targets are still being identified.<sup>1</sup> One of the few known multisubstrate enzymes is a widespread bacterial methylating enzyme RlmN, which methylates the C2 carbon on adenosine (m<sup>2</sup>A) in ribosomal RNA (rRNA) and in select transfer (tRNAs).<sup>2,3</sup> A closely related enzyme, Cfr, methylates the same nucleotide in rRNA, albeit with different regiochemistry, generating 8-methyladenosine (m<sup>8</sup>A). Cfr is usually found in pathogenic bacteria, and Cfr-mediated methylation of rRNA causes broad-spectrum antibiotic resistance.4-

RlmN and Cfr are unique among RNA methylating enzymes, as they use a distinct radical mechanism. Unlike RNA methyltransferases, these enzymes are RNA methylsynthases as they incorporate only a methylene fragment, derived from the methyl group of S-adenosyl-L-methionine (SAM), into the methyladenosine product.<sup>13-19</sup> Recent sequence analysis has identified over 2000 proteins annotated as RlmN/Cfr enzymes, yet only a handful have been functionally characterized (Figure 1).<sup>8,11,20,21</sup> Certain species, including several human pathogens within the Clostridia class, possess multiple copies of RlmN/ Cfr enzymes, suggesting substrate specialization and/or modification of additional yet unknown targets.<sup>8,21</sup> Unsuccessful attempts to functionally characterize enzymes within this family further emphasize difficulties in functional annotation of enzymes within the RlmN/Cfr family.  $^{\!\!\!8,21}$  A critical challenge for functional annotation of these enzymes is the absence of high-throughput and sensitive methods to identify their substrates in an unbiased manner. In particular, lack of methods that allow for the transcriptome-wide mapping of m<sup>2</sup>A/m<sup>8</sup>A hampers functional characterization of RlmN/Cfr family members.

Recent improvements in methods that couple immunoprecipitation of modified RNA or chemical treatments of RNA with next-generation sequencing have revolutionized mapping of the location and abundance of a subset of RNA modifications, such as N6-methyladenosine (m<sup>6</sup>A), 5-methylcytosine (m<sup>5</sup>C), and pseudouridine ( $\Psi$ ).<sup>22–29</sup> However, these approaches are limited to RNA modifications for which specific antibodies are available<sup>22,23</sup> or where the inherent chemical reactivity of the installed methyl group allows for selective modification of the methylated nucleotide over unmethylated nucleotide.<sup>24</sup> Strategies based on UV-cross-linking and immunoprecipitation, known as cross-linking and immunoprecipitation (CLIP)-based methods, cross-link a protein or an

Received: March 7, 2018 Published: May 21, 2018



**Figure 1.** Protein sequence similarity network (SSN) of RlmN/Cfr family. *E* value threshold of 10<sup>-75</sup> (>35% sequence identity) was applied to the SSN. Each node represents either one protein or a cluster of proteins that share >70% sequence identity. Nodes containing functionally characterized proteins (circles) are enlarged, with RlmNs shown in red and Cfrs in orange. Nodes representing functionally characterized RlmNs contain enzymes from *Bacillus subtilis*,<sup>8</sup> *Brevibacillus brevis*,<sup>8</sup> *Escherichia coli*,<sup>2</sup> and *Thermus thermophilus*.<sup>9</sup> Cfr nodes contain functionally characterized Cfrs from *Staphylococcus aureus*,<sup>10</sup> *Bacillus amyloliquefaciens*,<sup>11</sup> *Bascillus clausii*,<sup>11</sup> *B. brevis*,<sup>11</sup> *Paenibacillus* sp.,<sup>8</sup> and *Peptoclostridium difficile*.<sup>12</sup>

antibody to RNA molecules to identify interacting RNAs for a protein of interest.<sup>30–35</sup> These approaches were successfully used to generate transcriptome-wide binding maps for several RNA-binding proteins<sup>30–32</sup> and to locate m<sup>6</sup>A and N6,2'-*O*-dimethyladenosine (m<sup>6</sup>Am) throughout the transcriptome.<sup>33</sup> However, to date, none of the above strategies have been employed to map m<sup>2</sup>A and m<sup>8</sup>A RNA modifications.

Utilizing the unique mechanistic features of enzymes in the RlmN/Cfr family, here we describe the development of a novel strategy, where individual-nucleotide-resolution cross-linking and immunoprecipitation (miCLIP) is combined with mutational profiling with sequencing (MaPseq). The method, which we termed miCLIP-MaPseq, enables identification of substrates and sites of modification for any member of RlmN/Cfr family. Our method exploits the highly efficient, mechanism-based cross-linking between a conserved cysteine residue (C355 in E. coli RlmN) and the substrate adenosine in RNA.<sup>15–17</sup> During the catalytic cycle, this enzyme-RNA covalent intermediate is resolved by a second conserved cysteine (C118), forming the methylated RNA product. Mutation of C118 (C118A) stabilizes the covalently linked protein-RNA intermediate, enabling isolation of the enzyme-RNA adduct (Figure 2).<sup>14,16,17,36</sup> In our method, we use a C118A mutant protein to generate stable enzyme-RNA cross-links, which are subsequently enriched to allow for sequencing and identi-



**Figure 2.** Mechanistic scheme for RlmN-mediated methylation of RNA showing key steps. The stable covalent intermediate trapped by mutation of Cys118 into Ala is shown in the inset.

fication of cross-linked RNA substrates. An added feature of our method is identification of the modification site, enabled by the presence of a protein-derived remnant (a protein "scar") on the RNA substrate. Taking advantage of the processive, highfidelity, thermostable group II reverse transcriptase (TGIRT) that introduces a mismatch when it encounters the protein scar on RNA, our method additionally enables identification of the sites of RlmN-mediated methylation. Moreover, presence of the mismatch in the enriched RNAs serves to validate RNA substrates.

Here, we describe the development of miCLIP-MaPseq and its validation on *E. coli* RlmN, a well-characterized member of the RlmN/Cfr family. *E. coli* RlmN is particularly well suited for method validation as this enzyme modifies both 23S rRNA and several tRNAs. Importantly, since our method relies on mechanism-based cross-linking via a highly conserved Cys residue in these enzymes, miCLIP-MaPseq could provide an efficient and reliable approach to identify the substrates and modification sites for any member of the radical SAM methylating enzyme family.

#### RESULTS

Development of the miCLIP-MaPseq Method for the Detection of RImN/Cfr Substrates. Our method development was focused on the well-characterized E. coli RlmN, a known dual-specificity enzyme.<sup>3</sup> This enzyme modifies both A2503 in 23S rRNA and A37 in a subset of tRNAs, including  $tRNA^{Arg}_{ACG}$ ,  $tRNA^{Asp}_{GUC}$ ,  $tRNA^{Gln}_{UUG}$ ,  $tRNA^{Gln}_{CUG}$ ,  $tRNA^{Gln}_{UUC}$ , and  $tRNA^{His}_{GUG}$ . Taking advantage of the unique mechanism of RNA methylation by RlmN/Cfr enzymes, we developed a novel substrate identification method that relies on immunoprecipitation of a stable covalent enzyme-RNA complex, followed by high-throughput RNA sequencing (Figure 3). To enable mechanism-based substrate trapping, we used an E. coli BW25113 strain that expresses a FLAGtagged C118A RlmN mutant from the endogenous locus of RlmN.<sup>16</sup> Following immunoprecipitation, protein cross-linked to RNA was digested using Proteinase K. Importantly, Proteinase K digestion leaves a peptide scar on the RNA at the site of the RlmN-RNA cross-link formation, which reflects the methylation site. RNA was then size-selected on a denaturing TBE-urea gel (Figure S1a). RNA fragments larger than 300 nucleotides were fragmented using an established zinc chloride fragmentation protocol prior to dephosphorylation with T4 polynucleotide kinase.<sup>37</sup> RNA fragments smaller than 300 nucleotides were dephosphorylated without prior fragmentation. After size selection, RNA was converted to cDNA



Figure 3. Schematic representation of library preparation strategies for identification of substrates and methylation sites of RlmN. Black bars represent stop-site signals, while red bars represent fraction of mismatch at a specific nucleotide.

using TGIRT, followed by PCR amplification of the library and high-throughput sequencing.

Although the mechanism of enzyme-substrate cross-linking implemented in our method is unique to the RlmN/Cfr family, the use of mechanism-based cross-linking between an enzyme and RNA substrates parallels the miCLIP (methylation iCLIP) and Aza-IP (5-azacytidine-IP) methods developed for substrate identification of NSun2, a mechanistically distinct RNA modifying enzyme that introduces an m5C modification on several RNAs.<sup>1,24</sup> More broadly, our strategy expands on existing methods that rely on substrate trapping using catalytic mutant enzymes which have been used to identify non-RNA substrates of various enzymes, such as protein-tyrosine phosphatase.<sup>38,39</sup> An additional advantage of our method is the implementation of TGIRT to generate cDNA. This reverse transcriptase is highly processive and is especially suitable for structured and heavily modified RNAs, such as tRNAs.<sup>40,41</sup> To evaluate the suitability of other reverse transcriptases for miCLIP-MaPseq, we prepared sequencing libraries using either Superscript III or Superscript IV, as they differ from TGIRT in thermostability and processivity. Each library was generated in at least two biological replicates. Lastly, we assessed the specificity of miCLIP-MaPseq by comparing RNAs identified by our mechanism-based trapping method (FLAG-tagged C118A RlmN sample) to RNAs that strongly associate with RlmN in the absence of the covalent trapping (FLAG-tagged WT RlmN sample). In the latter sample, wild-type RlmN was FLAG-tagged and expressed from its endogenous locus. The comparison between these two samples allowed us to unambiguously determine the advantages of our mechanismbased covalent trapping strategy.

miCLIP-MaPseq Enables Detection of *E. coli* RlmN Substrates. The majority (~90%) of miCLIP-MaPseq reads mapped to rRNA, specifically to 23S rRNA, a known substrate of RlmN (Figure 4a,c). The remaining reads mapped consistently to tRNAs (~10%), while only a small number of reads (<1%) mapped to messenger (mRNA) and antisense RNAs (asRNA). Low read counts for the asRNAs and mRNAs suggest that the interaction between *E. coli* C118A RlmN enzyme and its substrates, 23S rRNA and tRNAs, is specific and that asRNAs and mRNAs are not substrates of RlmN. Additionally, we did not detect any of the mRNAs or asRNAs as enriched in the FLAG-tagged C118A RlmN sample, nor were these RNAs detected by mutational profiling (see next section). These findings suggest that mRNAs and asRNAs are not RlmN substrates.

To further assess the specificity of the interaction between the mutant enzyme and its substrates we examined which tRNAs were enriched in the FLAG-tagged C118A RlmN sample. E. coli RlmN is known to modify only a subset of tRNAs that contain adenosine at the position A37.<sup>3,42</sup> Thus, for each tRNA we determined the fold change in its abundance between the FLAG-tagged C118A RlmN sample and the input control sample (Figure 4b). The control sample consisted of rRNA-depleted total RNA isolated from the E. coli BW25113 strain where endogenous RlmN was not FLAG-tagged; rRNA was depleted from the control total RNA to facilitate identification of low abundance RNAs. The control RNA sample was treated in an identical manner to the RNA isolated after immunoprecipitation and Proteinase K treatment of FLAG-tagged enzymes (Figure S1). A tRNA was considered to be enriched in the FLAG-tagged C118A RlmN sample over the control when log2 fold change >2 and adjusted P-value <0.01 (Figure 4b). Our analysis identified tRNA<sup>Asp</sup><sub>GUC</sub>, tRNA<sup>Gln</sup><sub>CUG</sub>,  $tRNA^{Gln}{}_{UUG}\!\!\!\!$   $tRNA^{Glu}{}_{UUC}\!\!\!\!$  and  $tRNA^{His}{}_{GUG}$  to be enriched in the FLAG-tagged C118A RlmN sample. All of the identified tRNAs are known E. coli RlmN substrates. tRNA<sup>Arg</sup><sub>ACG</sub> was the only known tRNA substrate of E. coli RlmN that was not enriched in the FLAG-tagged C118A RlmN sample. Although not enriched relative to the control sample, tRNA<sup>Arg</sup><sub>ACG</sub> is one of the most abundant tRNAs present in the FLAG-tagged C118A RlmN sample and accounts for ~15% of total tRNA reads (Figure 4c). Few low abundant nonsubstrates,  $tRNA^{Ala}_{GGC}$ ,  $tRNA^{Ala}_{UGC}$ ,  $tRNA^{Ser}_{CGA}$ ,  $tRNA^{Ser}_{GGA}$ , and  ${\rm tRNA^{Trp}}_{\rm CCA}$  were enriched as evident from log2 fold value of  $\sim$ 2 and adjusted *P*-value <0.01. These tRNAs contain adenosine at the position 37 and are not known to be modified by RlmN. Importantly, in the absence of the covalent cross-link, we observed a high background signal, high false positive rate of substrate identification despite stringent washing, and no



Figure 4. Known substrates of E. coli RlmN are enriched by miCLIP-MaPseq. (a) Percentage of normalized miCLIP-MaPseq reads in noncoding and protein-coding RNAs isolated after immunoprecipitation of FLAG-tagged C118A RlmN. Shown are common miCLIP-MaPseq targets averaged across three biological replicates. Error represents standard deviation of the mean. (b) Enrichment of each tRNA in RNA isolated after immunoprecipitation of FLAG-tagged C118A RlmN. Known tRNA substrates (blue bars) are enriched over other E. coli tRNAs (orange bars). Gray bars represent nonsubstrate tRNAs where log2 fold change >1. Red dashed line indicates which tRNAs are enriched in the sample (log2-fold change  $\geq 2$ ). Adjusted P value <0.01 is indicated by \*. (c) Distribution of known tRNA substrates in RNA isolated after immunoprecipitation of FLAG-tagged C118A RlmN, determined from raw reads. (d) Enrichment of each tRNA in RNA isolated after immunoprecipitation of FLAG-tagged WT RlmN. Higher nonspecific enrichment is observed by this strategy, compared to FLAG-tagged C118A RlmN. TGIRT was used for all library preparations.

enrichment of known tRNA substrates (FLAG-tagged WT RlmN sample; Figure 4d).

Together, our results demonstrate that stable, covalent crosslinking between the enzyme and its substrates ensures low background signal. Additionally, our data indicate that the miCLIP aspect of our method is not sufficient for substrate identification, as it can lead to identification of false positives, especially in low abundant RNAs. To overcome this challenge, we combine miCLIP with mutational profiling (see next section) to validate substrates and to identify modification sites.

Lastly, we examined whether RNA enrichment is affected by the type of reverse transcriptase used in miCLIP-MaPseq. While the enrichment results were overall similar between samples prepared using Superscript IV and those prepared using TGIRT, a few differences were noted (Figure S2). For instance, use of Superscript IV allowed us to identify tRNA<sup>Arg</sup><sub>ACG</sub> as a substrate but failed to identify tRNA<sup>Asg</sup><sub>GUC</sub> as a substrate when log2 fold value >1 and adjusted *P*-value <0.01 (Figure S2). These differences in tRNA enrichment might be attributed to the different ability of each enzyme to process RNA structure, sequence, and modification patterns.

Validation of Substrates and Identification of Modification Sites by Mutational Profiling. Accurate detection of the modification site by our method requires precise identification of the protein scar location on RNA. Reversetranscribing protein scar as a mismatch instead of an insertion or a deletion (indel) would be advantageous as mismatches do not suffer from positional ambiguity when aligned across a homopolymeric stretch. Modification-based mismatching in TGIRT-mediated cDNA synthesis, referred to here as mutational profiling (MaP), has been successfully used for detection of endogenous N1-methyladenosine (m<sup>1</sup>A) and 3-methylcytosine (m<sup>3</sup>C) sites in eukaryotic tRNAs,<sup>43,44</sup> as well as chemical probing of RNA structure.<sup>41</sup> We thus tested whether TGIRT could efficiently convert protein scars into single nucleotide mismatches in RNA enriched by our mutant trapping method. Analysis of mutational profiles of sequences prepared by TGIRT demonstrates that almost all sites predicted to contain a protein scar are identified as single nucleotide mismatches instead of indels (99%, Table S1). This high efficiency of mismatch incorporation demonstrates the utility of TGIRTbased reverse transcription for the unambiguous mapping of modification sites cross-linked to the C118A RlmN enzyme.

To assess the efficiency of modification-site identification by TGIRT in miCLIP-MaPseq, we analyzed mismatch frequencies at positions A37 in tRNAs and A2503 in 23S rRNA, which are known to be modified by RlmN (Figure 5). Mismatches were analyzed in two RNA samples: (1) RNA isolated after immunoprecipitation of FLAG-tagged C118A RlmN and (2) rRNA-depleted total RNA isolated from *E. coli* BW25113, which was used as the control. Our analysis shows that TGIRT introduced mismatches at the site of the protein scar with



**Figure 5.** miCLIP-MaPseq detects location of the modification site by *E. coli* RlmN. Nucleotide composition of TGIRT-generated mismatches at position A37 in tRNAs and A2503 in 23S rRNA. C118A: RNA isolated after immunoprecipitation of FLAG-tagged C118A RlmN. Control: rRNA-depleted total RNA isolated from *E. coli* BW25113 strain. Mismatch analysis for 23S rRNA in control sample is not included due to low read counts resulting from rRNA depletion (indicated as \*).<sup>45</sup> Detailed analysis of nucleotide mismatches is presented in Table S2.



**Figure 6.** Lower processivity of Superscript III (a) leads to more frequent hard stops as compared to TGIRT (b). Shown are two representative substrate tRNAs,  $tRNA^{His}_{GUG}$  and  $tRNA^{Gln}_{CUG}$ . Protein scar at nucleotide A37 causes a hard stop at the nucleotide adjacent to the A37 in Superscript III reads (Table S5). Locations of Superscript III-generated stop-site signals are indicated. Signals were normalized to a number of reads for both tRNAs.

frequencies ranging from 20% for tRNA<sup>Arg</sup><sub>ACG</sub> to as high as 80% for tRNA<sup>Glu</sup>UUC, demonstrating that mutational profiling can indeed be used to identify modification sites of RlmN (Figure 5 and Table S2). While A-T mutations were the most prevalent, other mismatches were also observed at the site of the protein scar (Figure 5). All mismatches were considered in our analysis (Table S2). We also analyzed mismatch frequencies for tRNA<sup>Ala</sup><sub>GGC</sub>, tRNA<sup>Ala</sup><sub>UGC</sub>, tRNA<sup>Ser</sup><sub>GGA</sub>, and tRNA<sup>Trp</sup><sub>CCA</sub> (Figure 5), all of which contain adenosine at position 37 but are not known to be modified by RlmN.<sup>3,45</sup> Only tRNA<sup>Trp</sup><sub>CCA</sub> is known to be modified at A37, containing 2-methylthio-N6-isopentenyl adenosine.<sup>45</sup> While tRNA<sup>Ala</sup><sub>GGC</sub>, tRNA<sup>Ala</sup><sub>UGC</sub>, tRNA<sup>Ser</sup><sub>GGA</sub>, and tRNA<sup>Trp</sup><sub>CCA</sub> were not abundant in the FLAG-tagged C118A RlmN sample, they were enriched in this sample (log<sub>2</sub> fold change  $\geq 2$ ; Figure 4b). Importantly, low (<5%) mismatch incorporation at A37 in these four tRNAs confirms that these tRNAs are not modified by RlmN.

To further investigate if any of these tRNAs may be substrates, we examined whether E. coli RlmN can methylate tRNA<sup>Ala</sup><sub>GGC</sub> in vitro. Our analysis shows that *E. coli* RlmN does not methylate tRNA<sup>Ala</sup><sub>GGC</sub>, further confirming that this tRNA is not a substrate (Figure S3).<sup>42</sup> Together, our results demonstrate that TGIRT-mediated mismatching is an efficient strategy to validate RlmN substrates and to identify sites on RNA modified by this enzyme with single-nucleotide precision. While TGIRT reverse transcribes protein scar with high efficiency (>80%) in most tRNA substrates, the efficiency of read-through at A37 decreases in the presence of adjacent bulky modifications (Table S3). Nonetheless, high mismatch incorporation at position 37 combined with the ability to correctly reverse transcribe most endogenous RNA modifications present in our samples makes TGIRT ideally suited for our method (Figure S4).

Importantly, our data indicate that mutational profiling can be used to validate RNA substrates of RlmN/Cfr enzymes. For example, while tRNA<sup>Arg</sup><sub>ACG</sub> was not enriched in the RNA from the FLAG-tagged C118A RlmN sample over that from the control sample (Figure 4b), sequencing analysis at position A37 indicates ~20% mismatch. The ~4–5 fold increase in the mismatch incorporation at A37 over the nonsubstrate tRNAs confirms tRNA<sup>Arg</sup><sub>ACG</sub> as a substrate of *E. coli* RlmN (Table S2). Given the simplicity of mismatch detection, these findings suggest that mutational profiling can be used to validate substrates.

Finally, we investigated if Superscript IV can be used as an alternative reverse transcriptase to identify RlmN-mediated methylation sites in RNA. Similar to what we observed using TGIRT, Superscript IV preferentially introduces mismatches, relative to other mutational events, when it encounters a protein scar (Figure S5a). Despite having read-through capabilities, Superscript IV, under the tested conditions, does not introduce mismatches to the same extent as TGIRT at the site of the protein scar across all substrates (Figure S5b and Table S4). Thus, TGIRT is better suited for miCLIP-MaPseq.

Identification of the Modification Sites by RT Stop-Site Profiling. As an alternative approach, we investigated whether the site of RlmN-mediated methylation can be determined using less processive reverse transcriptase that would terminate at cross-linking sites. Thus, we selected Superscript III for these experiments since this enzyme has lower processivity than TGIRT and Superscript IV. RT stopsites were determined by analyzing the 5'-ends of the reads mapped to the genome, referred to here as stop-site profiling. As expected, Superscript III yielded high termination rates at the cross-link sites under the conditions tested (Figure 6a), an observation consistent with previous studies.<sup>32</sup> Interestingly, we observed that Superscript III can terminate either at the site of the protein scar or at the adjacent nucleotide (Figure 6a and Table S5). However, the major drawback of this approach are stop-sites that result from other post-transcriptional modifica-

#### Journal of the American Chemical Society

tions (Figure 6a), making identification of cross-link sites by stop-site profiling challenging. This is particularly relevant in the case of tRNAs, which are extensively modified. In contrast, TGIRT has high processivity under the tested conditions and readily reads through the majority of modified nucleotides (Figure 6b). Together, our findings indicate that TGIRT is best suited for substrate identification of RNA-methylating radical SAM enzymes and that mutational profiling rather than stopsite profiling is ideally suited for identification of cross-link sites.

# DISCUSSION

Here, we present a robust and simple method, miCLIP-MaPseq, to identify RNA substrates and map modification sites of enzymes that belong to the radical SAM RNA methylating enzyme family. Building on our mechanistic studies, the method utilizes a catalytic mutant of RlmN that generates a stable, covalent adduct between the enzyme and the methylated adenosine in substrate RNAs that can be easily purified and enriched by immunoprecipitation. One of the major advantages of mechanism-based cross-linking is its high cross-linking efficiency relative to UV-cross-linking and, unlike UV light, does not cause random mutations in RNA.<sup>46,47</sup> Importantly, proteolytic removal of the enzyme from the enzyme–RNA adduct generates a scar on RNA, which provides a chemical signature necessary for precise identification of the modification site.

An important feature of the miCLIP-MaPseq strategy is the use of TGIRT as reverse transcriptase. TGIRT can transcribe through the protein scar and generate a scar-induced mismatch. Furthermore, use of TGIRT minimizes mismatches due to the presence of most endogenous modifications. This strategy greatly simplifies identification of modification sites specific to RlmN. We also showed that Superscript IV can be used in lieu of TGIRT to generate cross-link-induced mismatches at modified sites, albeit to a lesser extent. Together, our results indicate that miCLIP-MaPseq using TIGRT provides the most direct and accurate way to identify RNA substrates of enzymes in the RlmN/Cfr family and to map their modified nucleotides at the transcriptome level.

Future modifications to the protocol will include use of spike-in controls to monitor global changes in RNA content and inclusion of unique molecular identifiers (UMIs) for precise estimations of enrichments. Additionally, RNA fragmentation prior to immunoprecipitation will allow for the improved coverage of the sequence containing the modified nucleotide and could increase sensitivity of substrate calling by mutational analysis. These modifications may provide the added benefit of removing the confounding factors that stem from variations in length of the substrates, potentially biasing enrichment calculations.

In summary, the lack of methods for global substrate identification for enzymes in the radical SAM RNA methylating enzyme family has thus far limited their characterization and precluded our understanding of the prevalence of m<sup>2</sup>A and m<sup>8</sup>A marks across RNA. Additionally, it has hampered our ability to correctly annotate genes in this family and to harness genomic information to understand their function. Our strategy makes initial inroads toward the goal of functional annotation of any enzyme in the RlmN/Cfr family by identifying the locations of their m<sup>2</sup>A or m<sup>8</sup>A modifications. Beyond the RlmN/Cfr family, miCLIP-MaPseq could be potentially applied to any RNA modifying enzyme that generates a covalent adduct between enzyme and RNA.

### MATERIALS AND METHODS

miCLIP Analysis and Fragmentation. A C-terminal DYKDDDDK octapeptide (FLAG) sequence was fused to the genomic version of rlmN or rlmN containing C118A mutation as described previously.<sup>16</sup>

LB medium (1 L) was inoculated with 10 mL from an overnight culture of E. coli BW25113 either expressing the FLAG-tagged WT RlmN or FLAG-tagged C118A RlmN. The cells were harvested after 90 min for WT RlmN or after 150 min for C118A RlmN. Approximately 1.5 g of cells was resuspended in lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% (v/v) Triton X-100) and lysed via sonication. The resulting lysate was first treated with RQ1 RNase-free DNase I (Promega) at 37 °C for 15 min followed by immunoprecipitation using anti-FLAG M2 affinity gel (Sigma). Prior to elution of the enzyme, the resin was washed five times with stringent-TBS wash buffer (50 mM Tris pH 7.5, 500 mM NaCl) in order to remove RNA species nonspecifically bound to the enzyme. Protein was digested with Proteinase K (Biorad) at 37 °C for 2 h. Four RNA size fractions were recovered from the 10% Novex TBE-urea gel (Invitrogen; Figure S1a). The fraction containing RNA species larger than 300 nt (fraction A in Figure S1a) was fragmented to 50–200 nt using 1× RNA fragmentation reagent (Zn<sup>2+</sup> based, Ambion). Specifically, RNA from fraction A was first denatured for 2 min at 95 °C and then fragmented at 95 °C for 1–2 min depending on the amount of starting RNA. The reaction was stopped with  $1 \times$  stop solution (Ambion) and quickly placed on ice. Fragmented RNA was run on a 10% Novex TBE-urea gel (Invitrogen) for 50 min at 150 V. After staining with Sybr Gold (ThermoFisher), RNA fragments of 50-200 nt in size were eluted from the gel, and ethanol precipitated in the presence of GlycoBlue (Ambion).

**RNA Isolation from the** *E. coli* **BW25113 Strain.** *E. coli* **BW25113** strain was grown at 37 °C until  $OD_{600} \sim 0.4-0.6$ . For extraction of total RNA, cells were resuspended in ice-cold buffer (0.3 M NaOAc pH 4.5, 10 mM EDTA) followed by direct phenol extraction using cold acid phenol/chloroform (5:1, pH 4.5). RNA was extracted from the resulting pellet by dissolving the pellet in ice-cold buffer (10 mM NaOAc pH 4.5, 800 mM LiCl). Ribosomal RNA was further depleted from the sample via three consecutive precipitations with 0.2 volumes of 2-propanol, followed by a fourth precipitation with 1 volume of 2-propanol. RNA that was obtained in the third (fraction A) and fourth (fractions B–D) precipitation steps was size selected on a 10% Novex TBE–urea gel (Invitrogen) as indicated in Figure S1b.

Reverse Transcription with Superscript III and Superscript IV. Prior to reverse transcription, size-selected RNA was dephosphorylated with 20 U of T4 polynucleotide kinase (NEB) in the presence of 20 U of Superase-In (Invitrogen) and T4 Polynucleotide Kinase Reaction Buffer (70 mM Tris-HCl pH 7.6, 10 mM MgCl<sub>2</sub>, 5 mM DTT) at 37 °C for 1 h. A single-stranded preadenylated linker containing a spacer on 3'-end (5'-rApp-CTG TAG GCA CCA TCA ATC-3SpC3; IDT) was ligated to the 3'-end of the RNA to serve as a priming site for reverse transcription. Reaction conditions were as follows: 1 µg of preadenylated linker, 50 mM Tris-HCl pH 7.5, 10 mM MgCl<sub>2</sub>, 5 mM DTT, 15% PEG 8000, and 300 U T4 RNA ligase 2 truncated (NEB). The reaction was allowed to proceed at 37 °C for 2 h, after which RNA was precipitated with ethanol in the presence of GlycoBlue (Ambion). Precipitated RNA was resuspended in 10 mM Tris-HCl pH 7.0, after which RT primer containing an eight-atom hexaethylene glycol spacer (/5'Phos/GAT CGT CGG ACT GTA GAA CTC TGA ACC TGT CG/iSp18/C AAG CAG AAG ACG GCA TAC GAG ATA TTG ATG GTG CCT ACA G-3'; IDT) was added to a final concentration of 0.4  $\mu$ M. The sample was first incubated at 65 °C for 5 min, followed by incubation at 35 °C for 5 min. Annealed RNA was then reverse transcribed under the following conditions: 50 mM Tris-HCl pH 8.3, 4 mM MgCl<sub>2</sub>, 10 mM DTT, 50 mM KCl, 0.5 mM dNTPs, 20 U Superase-In, and 200 U Superscript IV (or Superscript III). Reactions were incubated at 52 °C for 12 min (Superscript III; Invitrogen) or for 22 min (Superscript IV; Invitrogen), after which enzyme was inactivated by addition of 1 M NaOH to a final concentration of 90 mM followed by incubation at 95

#### Journal of the American Chemical Society

°C for 5 min. The reaction was then neutralized by addition of 1 M HCl to a final concentration of 85 mM. The resulting cDNA was circularized using 100 U CircLigase II (Epicenter) in the presence of 33 mM Tris-Acetate pH 7.5, 66 mM potassium acetate and 5 mM DTT, at 60 °C for 1 h. Upon completion of the reaction, CircLigase II was inactivated for 10 min at 80 °C. Circularized cDNA was purified on 10% Novex TBE-urea gel (Invitrogen) and ethanol precipitated. The purified cDNA was amplified with Phusion-HF (ThermoScientific) using O231 primer (5'-CAA GCA GAA GAC GGC ATA CGA-3') and barcode primer (5'-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACG ATC GGA AGA GCA CAC GTC TGA ACT CCA GTC AC [barcode] CGA CAG GTT CAG AGT TC-3') for 12-18 cycles of 98  $^\circ$ C for 10 s, 60  $^\circ$ C for 10 s and 72  $^\circ$ C for 12 s. The PCR products were purified on an 8% Novex TBE gel (Invitrogen) and quantified using qPCR. Quality of the RNA was checked on the Agilent 2100 Bioanalyzer. Libraries were sequenced with oNTI202 primer (5'-CGA CAG GTT CAG AGT TCT ACA GTC CGA CGA TC-3') using 50-nt single-end reads on the HiSeq4000 (Illumina). Each library for FLAG-tagged C118A RlmN sample and FLAG-tagged WT RlmN sample was generated in two biological replicates. Library for control sample (rRNA-depleted total RNA isolated from WT E. coli) was generated in two biological replicates.

Thermostable Group II Intron RT Template-Switching. Template-switching reactions were performed using a TGIRT Template Switching RNA-seq kit (InGex), and protocols were followed as described previously.<sup>48,49</sup> Briefly, prior to the reverse transcription step, size-selected RNA was dephosphorylated with 20 U of T4 polynucleotide kinase (NEB) in the presence of 20 U of Superase-In (Invitrogen) and T4 Polynucleotide kinase reaction buffer (70 mM Tris-HCl pH 7.6, 10 mM MgCl<sub>2</sub>, 5 mM DTT) at 37 °C for 1 h. For the reverse transcription step, we used template-primer substrate consisting of 34-nt RNA oligonucleotide (5'-AGA UCG GAA GAG CAC ACG UCU GAA CUC CAG UCA C/3'SpC3/) that contains Illumina Read 2 sequence, aligned to a complementary 35-nt DNA primer, which contains a single nucleotide 3'-overhang (InGex). In a typical reaction, up to 50 ng of RNA was mixed with the template primer substrate (100 nM) and 500 nM TGIRT-III in the presence of reaction buffer (450 mM NaCl, 5 mM MgCl<sub>2</sub>, 5 mM DTT, 20 mM Tris-HCl, pH 7.5). Reaction was preincubated for 30 min at room temperature prior to addition of 25 mM dNTPs to a final concentration of 1 mM, followed by incubation at 60 °C for 1 h. The reaction was quenched by addition of 5 M NaOH to a final concentration of 0.25 M and kept at 65 °C for 15 min, after which the reaction was neutralized by addition of 5 M HCl. The cDNA was analyzed on an 8% Novex TBE gel (Invitrogen) and ethanol precipitated in the presence of 0.3 M sodium acetate, pH 5.2, and GlycoBlue coprecipitant. An R1R DNA linker containing a small RNA Illumina sequencing primer site (5'-/5Phos/GAT CGT CGG ACT GTA GAA CTC TGA ACG TGT AG/3SpC3/) was first adenylated at the 5'-end using a 5' DNA adenylation kit (NEB), after which it was ligated to cDNA using Thermostable 5'-AppDNA/RNA Ligase (NEB). Thermostable ligation was carried out in the presence of NEBuffer 1 (10 mM Bis-Tris-Propane-HCl pH 7.0, 10 mM MgCl<sub>2</sub>, 1 mM DTT), 5 mM MnCl<sub>2</sub>, 2 mM adenylated R1R DNA linker, for 2 h at 65 °C. Ligated cDNA was purified using MinElute PCR purification kit (Qiagen). The purified cDNA was then amplified with Phusion HF (ThermoScientific) using Illumina multiplex primer (5'-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACG TTC AGA GTT CTA CAG TCC GAC GAT C-3') and specific barcode primer (5-CAA GCA GAA GAC GGC ATA CGA GAT [barcode] GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T-3') for 15-21 cycles of 98  $^{\circ}\text{C}$  for 5 s, 60  $^{\circ}\text{C}$  for 10 s, and 72  $^{\circ}\text{C}$  for 12 s. Amplified cDNA was purified on an 8% Novex TBE gel (Invitrogen) and quantified using qPCR. Quality of the RNA was checked on an Agilent 2100 bioanalyzer prior to sequencing on an Illumina HiSeq4000 using 50-nt single-end reads. Libraries for FLAG-tagged C118A RlmN sample and FLAG-tagged WT RlmN sample were generated in three biological replicates. The library for control sample (rRNA-depleted total RNA isolated from WT E. coli) was generated in two biological replicates.

Sequencing Read Mapping and Analysis. Sequencing data were uploaded to the Galaxy web platform (version 1.0.1), and the public server at usegalaxy.org was used to analyze the data.<sup>50</sup> Prior to mapping reads, reads were processed with FASTQ Groomer<sup>51</sup> followed by adapter removal using Clip tool available through Galaxy platform. Sequences greater than 15 bp were aligned to the *E. coli* BW25113 genome using Bowtie2 (Galaxy Version 2.2.6.2) with default options.<sup>52,53</sup> The *E. coli* BW25113 genome was obtained through the Ensembl Bacteria Genome Database (EMBL-EBI) on Jan 23, 2017. For each sample (FLAG-tagged C118A RlmN, FLAG-tagged WT RlmN, or control) library variation, we collected a combined total of ~6–10 million mapped reads between TGIRT replicates and S0–100 million for Superscript IV replicates.

Raw counts per gene were determined using a HTSeq-count script, which is part of the HTSeq python module within Galaxy web platform.<sup>50,54</sup> The intersection-nonempty mode was selected to handle reads overlapping more than one feature. For enrichment analysis of reads mapped to tRNA genes we used DESeq2 module, with default parameters, available within Galaxy web platform.<sup>55</sup> DESeq2 takes into account the variability between the replicates and normalizes read counts to account for differences in sequencing depth between samples, reporting fold change values between the sample and the control. We used an arbitrary cutoff of 4-fold increase in abundance and adjusted P value of <0.01 as our threshold for identifying substrates in samples were TGIRT was used as reverse transcriptase. DESeq2 adjusted P-values are adjusted for multiple-comparison testing and are used to lower the false positive detection.

To determine the 5' end of the reads (stop-sites) we used script "make\_wiggle" to convert sorted and indexed BAM files to wiggle files. This script was developed by the Weissman lab at UCSF and is readily available through Plastid.<sup>56</sup>

Stop-sites and mismatches were both visualized using the Integrated Genomic Viewer (IGV).<sup>57</sup> The percentage of mismatches was calculated by cumulative analysis of all biological replicates. The nucleotide composition at position A37 in specific tRNAs and A2503 in 23S rRNA was determined from IGV and was normalized across the samples.

**Protein Sequence Similarity Network.** Prior to generating sequence similarity networks (SSNs), we retrieved RlmN/Cfr sequences by BLAST (NCBI database) using Cfr sequence from *S. aureus* as a query. The *E* value was set to 1e-10. The sequences that shared <90% sequence coverage and those that shared >85% sequence identity to the query were removed using CD-HIT.<sup>58</sup> This resulted in 2970 sequences.

SSNs were constructed using the Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST)<sup>59</sup> and were visualized by Cytoscape  $3.3.^{60}$  Network contains 1,396 nodes where each node contains sequences that share >70% sequence identity. Majority of nodes contain <15 sequences. The edge indicates that the two nodes share significant similarity with an E-value less than the selected cutoff.

Data and Software availability. Illumina sequencing raw data (FASTQ files) are available under the NCBI SRA submission SRP145395.

#### ASSOCIATED CONTENT

#### **Supporting Information**

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.8b02618.

TBE-urea analysis of RNA (Figure S1), enrichment of tRNAs in libraries prepared with Superscript IV (Figure S2), in vitro methylation of in vitro transcribed tRNA<sup>Gln</sup><sub>UUG</sub> and tRNA<sup>Ala</sup><sub>GGC</sub> (Figure S3), raw read-profiles for tRNA<sup>Gln</sup><sub>UUG</sub> and tRNA<sup>His</sup><sub>GUG</sub> (Figure S4), nucleotide composition of Superscript IV-generated mismatches at position A37 in tRNAs and A2503 in 23S rRNA (Figure S5 and Table S4), distribution of TGIRT-transcribed mutations (Table S1) and mismatches (Table S2), analysis of efficiency of read-

through at position A37 in known tRNA substrates (Table S3), and analysis of Superscript III stops in known substrate RNAs (Table S5) (PDF)

# AUTHOR INFORMATION

Corresponding Author

\*danica.fujimori@ucsf.edu

#### ORCID 0

Danica Galonić Fujimori: 0000-0002-4066-9417

#### Notes

The authors declare no competing financial interest.

#### ACKNOWLEDGMENTS

We thank Dr. Anthony Shiver and Dr. Carol Gross from UCSF for providing necessary *E. coli* strains, Dr. Lindsey Pack for providing technical assistance with library generation, and Dr. Meghan McKeon Zubradt and Dr. Joshua Dunn for assistance with stop-sites analysis. We also thank Dr. Eric Chow at the UCSF Center for Advance Technology for sequencing assistance and Dr. Hani Goodarzi, Dr. Mary McMahon, Dr. Michel Tassetto, and Kaitlyn Tsai for helpful discussions and comments on the manuscript. This research was supported by UCSF Program for Breakthrough Biomedical Research (PBBR) Postdoctoral Grant (to V.S.), NIAID R01AI095393 (to D.G.F.), UCSF Program for Breakthrough Biomedical Research funded in part by the Sandler Foundation (to D.E.W.), and NIH Director's Early Independence Award DP5OD017895 (to D.E.W.).

# REFERENCES

(1) Hussain, S.; Sajini, A. A.; Blanco, S.; Dietmann, S.; Lombard, P.; Sugimoto, Y.; Paramor, M.; Gleeson, J. G.; Odom, D. T.; Ule, J.; Frye, M. *Cell Rep.* **2013**, *4* (2), 255–261.

(2) Toh, S. M.; Xiong, L.; Bae, T.; Mankin, A. S. RNA 2008, 14 (1), 98-106.

(3) Benitez-Paez, A.; Villarroya, M.; Armengod, M. E. RNA **2012**, *18* (10), 1783–95.

(4) Long, K. S.; Poehlsgaard, J.; Kehrenberg, C.; Schwarz, S.; Vester, B. Antimicrob. Agents Chemother. 2006, 50 (7), 2500-5.

(5) Smith, L. K.; Mankin, A. S. Antimicrob. Agents Chemother. 2008, 52 (5), 1703–1712.

(6) Polikanov, Y. S.; Starosta, A. L.; Juette, M. F.; Altman, R. B.; Terry, D. S.; Lu, W.; Burnett, B. J.; Dinos, G.; Reynolds, K. A.; Blanchard, S. C.; Steitz, T. A.; Wilson, D. N. *Mol. Cell* **2015**, *58* (5), 832–44.

(7) McCusker, K. P.; Fujimori, D. G. ACS Chem. Biol. 2012, 7 (1), 64-72.

(8) Atkinson, G. C.; Hansen, L. H.; Tenson, T.; Rasmussen, A.; Kirpekar, F.; Vester, B. Antimicrob. Agents Chemother. **2013**, 57 (8), 4019–26.

(9) Mengel-Jorgensen, J.; Jensen, S. S.; Rasmussen, A.; Poehlsgaard, J.; Iversen, J. J.; Kirpekar, F. J. Biol. Chem. **2006**, 281 (31), 22108–17.

(10) Giessing, A. M. B.; Jensen, S. S.; Rasmussen, A.; Hansen, L. H.; Gondela, A.; Long, K.; Vester, B.; Kirpekar, F. RNA **2009**, *15* (2), 327–336.

(11) Hansen, L. H.; Planellas, M. H.; Long, K. S.; Vester, B. Antimicrob. Agents Chemother. 2012, 56 (7), 3563-7.

(12) Marin, M.; Martin, A.; Alcala, L.; Cercenado, E.; Iglesias, C.; Reigadas, E.; Bouza, E. *Antimicrob. Agents Chemother.* **2015**, 59 (1), 586–9.

(13) Boal, A. K.; Grove, T. L.; McLaughlin, M. I.; Yennawar, N. H.; Booker, S. J.; Rosenzweig, A. C. *Science* **2011**, 332 (6033), 1089–92.

(14) Grove, T. L.; Benner, J. S.; Radle, M. I.; Ahlum, J. H.; Landgraf, B. J.; Krebs, C.; Booker, S. J. Science **2011**, 332 (6029), 604-7.

(15) Grove, T. L.; Livada, J.; Schwalm, E. L.; Green, M. T.; Booker, S. J.; Silakov, A. *Nat. Chem. Biol.* **2013**, *9* (7), 422–7.

- (16) McCusker, K. P.; Medzihradszky, K. F.; Shiver, A. L.; Nichols, R. J.; Yan, F.; Maltby, D. A.; Gross, C. A.; Fujimori, D. G. *J. Am. Chem. Soc.* **2012**, *134* (43), 18074–81.
- (17) Silakov, A.; Grove, T. L.; Radle, M. I.; Bauerle, M. R.; Green, M. T.; Rosenzweig, A. C.; Boal, A. K.; Booker, S. J. *J. Am. Chem. Soc.* **2014**, *136* (23), 8221–8.
- (18) Yan, F.; Fujimori, D. G. Proc. Natl. Acad. Sci. U. S. A. 2011, 108 (10), 3930-4.
- (19) Yan, F.; LaMarre, J. M.; Rohrich, R.; Wiesner, J.; Jomaa, H.; Mankin, A. S.; Fujimori, D. G. J. Am. Chem. Soc. **2010**, 132 (11), 3953-64.
- (20) Kaminska, K. H.; Purta, E.; Hansen, L. H.; Bujnicki, J. M.; Vester, B.; Long, K. S. *Nucleic Acids Res.* **2010**, 38 (5), 1652–63.

(21) Stojković, V.; Noda-Garcia, L.; Tawfik, D. S.; Fujimori, D. G. *Nucleic Acids Res.* **2016**, 44 (18), 8897–8907.

(22) Dominissini, D.; Moshitch-Moshkovitz, S.; Schwartz, S.; Salmon-Divon, M.; Ungar, L.; Osenberg, S.; Cesarkas, K.; Jacob-Hirsch, J.; Amariglio, N.; Kupiec, M.; Sorek, R.; Rechavi, G. *Nature* **2012**, 485 (7397), 201–6.

(23) Meyer, K. D.; Saletore, Y.; Zumbo, P.; Elemento, O.; Mason, C. E.; Jaffrey, S. R. *Cell* **2012**, *149* (7), 1635–46.

(24) Khoddami, V.; Cairns, B. R. Nat. Biotechnol. 2013, 31 (5), 458–64.

(25) Carlile, T. M.; Rojas-Duran, M. F.; Zinshteyn, B.; Shin, H.; Bartoli, K. M.; Gilbert, W. V. *Nature* **2014**, *515* (7525), 143–6.

(26) Schwartz, S.; Agarwala, S. D.; Mumbach, M. R.; Jovanovic, M.; Mertins, P.; Shishkin, A.; Tabach, Y.; Mikkelsen, T. S.; Satija, R.; Ruvkun, G.; Carr, S. A.; Lander, E. S.; Fink, G. R.; Regev, A. *Cell* **2013**, *155* (6), 1409–21.

(27) Delatte, B.; Wang, F.; Ngoc, L. V.; Collignon, E.; Bonvin, E.; Deplus, R.; Calonne, E.; Hassabi, B.; Putmans, P.; Awe, S.; Wetzel, C.; Kreher, J.; Soin, R.; Creppe, C.; Limbach, P. A.; Gueydan, C.; Kruys, V.; Brehm, A.; Minakhina, S.; Defrance, M.; Steward, R.; Fuks, F. *Science* **2016**, 351 (6270), 282–5.

(28) Li, X.; Zhu, P.; Ma, S.; Song, J.; Bai, J.; Sun, F.; Yi, C. Nat. Chem. Biol. 2015, 11 (8), 592-7.

(29) Lovejoy, A. F.; Riordan, D. P.; Brown, P. O. PLoS One 2014, 9 (10), e110799.

(30) Ule, J.; Jensen, K. B.; Ruggiu, M.; Mele, A.; Ule, A.; Darnell, R. B. *Science* **2003**, 302 (5648), 1212-5.

(31) Hafner, M.; Lianoglou, S.; Tuschl, T.; Betel, D. Methods 2012, 58 (2), 94–105.

(32) Konig, J.; Zarnack, K.; Rot, G.; Curk, T.; Kayikci, M.; Zupan, B.; Turner, D. J.; Luscombe, N. M.; Ule, J. *Nat. Struct. Mol. Biol.* **2010**, *17* (7), 909–15.

(33) Linder, B.; Grozhik, A. V.; Olarerin-George, A. O.; Meydan, C.; Mason, C. E.; Jaffrey, S. R. *Nat. Methods* **2015**, *12* (8), 767–72.

(34) Haag, S.; Kretschmer, J.; Sloan, K. E.; Bohnsack, M. T. *Methods Mol. Biol.* **2017**, 1562, 269–281.

- (35) Zhang, C. L.; Darnell, R. B. Nat. Biotechnol. 2011, 29 (7), 607–U86.
- (36) Schwalm, E. L.; Grove, T. L.; Booker, S. J.; Boal, A. K. Science **2016**, 352 (6283), 309–12.
- (37) Dominissini, D.; Moshitch-Moshkovitz, S.; Salmon-Divon, M.; Amariglio, N.; Rechavi, G. *Nat. Protoc.* **2013**, *8* (1), 176–89.

(38) Flint, A. J.; Tiganis, T.; Barford, D.; Tonks, N. K. Proc. Natl. Acad. Sci. U. S. A. 1997, 94 (5), 1680-5.

- (39) Blanchetot, C.; Chagnon, M.; Dube, N.; Halle, M.; Tremblay, M. L. *Methods* **2005**, 35 (1), 44–53.
- (40) Zheng, G.; Qin, Y.; Clark, W. C.; Dai, Q.; Yi, C.; He, C.; Lambowitz, A. M.; Pan, T. *Nat. Methods* **2015**, *12* (9), 835–837.
- (41) Zubradt, M.; Gupta, P.; Persad, S.; Lambowitz, A. M.; Weissman, J. S.; Rouskin, S. *Nat. Methods* **2017**, *14* (1), 75–82.

<sup>(42)</sup> Fitzsimmons, C. M.; Fujimori, D. G. PLoS One 2016, 11 (11), e0167298.

# Journal of the American Chemical Society

(43) Mohr, S.; Ghanem, E.; Smith, W.; Sheeter, D.; Qin, Y.; King, O.; Polioudakis, D.; Iyer, V. R.; Hunicke-Smith, S.; Swamy, S.; Kuersten, S.; Lambowitz, A. M. *RNA* **2013**, *19* (7), 958–70.

(44) Katibah, G. E.; Qin, Y.; Sidote, D. J.; Yao, J.; Lambowitz, A. M.; Collins, K. Proc. Natl. Acad. Sci. U. S. A. 2014, 111 (33), 12025-12030.

(45) Boccaletto, P.; Machnicka, M. A.; Purta, E.; Piątkowski, P.; Bagiński, B.; Wirecki, T. K.; de Crécy-Lagard, V.; Ross, R.; Limbach, P. A.; Kotter, A.; Helm, M. *Nucleic Acids Res.* **2018**, *46* (D1), D303– D307.

(46) Sugimoto, Y.; Konig, J.; Hussain, S.; Zupan, B.; Curk, T.; Frye, M.; Ule, J. *Genome Biol.* **2012**, *13* (8), R67.

(47) Kishore, S.; Jaskiewicz, L.; Burger, L.; Hausser, J.; Khorshid, M.; Zavolan, M. Nat. Methods 2011, 8 (7), 559–64.

(48) Qin, Y.; Yao, J.; Wu, D. C.; Nottingham, R. M.; Mohr, S.; Hunicke-Smith, S.; Lambowitz, A. M. *RNA* **2016**, 22 (1), 111–28.

(49) Nottingham, R. M.; Wu, D. C.; Qin, Y.; Yao, J.; Hunicke-Smith, S.; Lambowitz, A. M. *RNA* **2016**, *22* (4), 597–613.

(50) Afgan, E.; Baker, D.; van den Beek, M.; Blankenberg, D.; Bouvier, D.; Cech, M.; Chilton, J.; Clements, D.; Coraor, N.; Eberhard, C.; Gruning, B.; Guerler, A.; Hillman-Jackson, J.; Von Kuster, G.; Rasche, E.; Soranzo, N.; Turaga, N.; Taylor, J.; Nekrutenko, A.; Goecks, J. *Nucleic Acids Res.* **2016**, *44* (W1), W3– W10.

(51) Blankenberg, D.; Gordon, A.; Von Kuster, G.; Coraor, N.; Taylor, J.; Nekrutenko, A. *Bioinformatics* **2010**, *26* (14), 1783–5.

(52) Langmead, B.; Salzberg, S. L. Nat. Methods **2012**, 9 (4), 357–9. (53) Langmead, B.; Trapnell, C.; Pop, M.; Salzberg, S. L. Genome Biol. **2009**, 10 (3), R25.

(54) Anders, S.; Pyl, P. T.; Huber, W. Bioinformatics 2015, 31 (2), 166-9.

(55) Love, M. I.; Huber, W.; Anders, S. Genome Biol. 2014, 15 (12), 550.

(56) Dunn, J. G.; Weissman, J. S. *BMC Genomics* **2016**, *17* (1), 958. (57) Thorvaldsdottir, H.; Robinson, J. T.; Mesirov, J. P. Briefings Bioinf. **2013**, *14* (2), 178–92.

(58) Huang, Y.; Niu, B.; Gao, Y.; Fu, L.; Li, W. Bioinformatics 2010, 26 (5), 680-2.

(59) Gerlt, J. A.; Bouvier, J. T.; Davidson, D. B.; Imker, H. J.; Sadkhin, B.; Slater, D. R.; Whalen, K. L. *Biochim. Biophys. Acta, Proteins Proteomics* **2015**, *1854* (8), 1019–37.

(60) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. *Genome Res.* **2003**, *13* (11), 2498–504. Article