

UCSF

UC San Francisco Previously Published Works

Title

Genetically encoded chemical crosslinking of RNA in vivo.

Permalink

<https://escholarship.org/uc/item/6jn44756>

Journal

Nature chemistry, 15(1)

ISSN

1755-4330

Authors

Sun, Wei
Wang, Nanxi
Liu, Hongjiang
[et al.](#)

Publication Date

2023

DOI

10.1038/s41557-022-01038-4

Peer reviewed



Published in final edited form as:

Nat Chem. 2023 January ; 15(1): 21–32. doi:10.1038/s41557-022-01038-4.

Genetically Encoded Chemical Cross-linking of RNA *in vivo*

Wei Sun^{1,†}, Nanxi Wang^{1,†}, Hongjiang Liu², Bingchen Yu¹, Ling Jin¹, Xingjie Ren², Yin Shen^{2,3}, Lei Wang^{1,*}

¹Department of Pharmaceutical Chemistry and the Cardiovascular Research Institute, University of California San Francisco, San Francisco, California, United States

²Institute for Human Genetics, University of California San Francisco, San Francisco, California, United States

³Department of Neurology, University of California San Francisco, San Francisco, California, United States

Abstract

Protein-RNA interactions regulate RNA fate and function, and defects can lead to various disorders. Such interactions have mainly been studied with nucleoside-based UV cross-linking methods, which lack broad *in vivo* compatibility and the ability to resolve specific amino acids. Here we genetically encoded latent bioreactive unnatural amino acids into proteins to react with bound RNA via proximity-enabled reactivity, and demonstrated genetically encoded chemical cross-linking of proteins with target RNA (GECX-RNA) *in vivo*. Applying GECX-RNA to an RNA chaperon Hfq in *E. coli* identified target RNAs with amino acid specificity. Combining GECX-RNA with immunoprecipitation and high-throughput sequencing on a N6-methyladenosine (m6A) reader protein in mammalian cells allowed *in vivo* identification of unknown m6A on RNA with single-nucleotide resolution throughout transcriptome. GECX-RNA affords resolution at the nucleotide and amino acid level for interrogating protein-RNA interactions *in vivo*. It also enables precise engineering of covalent linkages between a protein and RNA, which will inspire innovative solutions for RNA-related research and therapeutics.

Editorial Summary:

Protein-RNA interactions regulate RNA fate and function, and are generally noncovalent and reversible. Genetically introducing a latent bioreactive amino acid into a protein has now been

* Lei.Wang2@ucsf.edu .

AUTHOR CONTRIBUTIONS

W.S. designed and conducted experiments, analyzed data, and wrote the manuscript; N.W. evolved SFYRS and characterized SFY incorporation and cross-linking of proteins; H.L. conducted data analysis of GRIP-seq; B.Y. synthesized FSY and SFY, tested SFY reaction with NMPs *in vitro* and analyzed the data; L.J. helped with the experiment of dCas13b target RNA cross-linking and enrichment in mammalian cells; X.R. and Y.S. helped with the experiments of GRIP-seq; L.W. conceived, directed, supported the project, and wrote the manuscript.

[†]These authors contributed equally to this work.

COMPETING INTERESTS

The authors declare no competing interests.

CODE AVAILABILITY

Custom code used is available at <https://github.com/Shall-We-Dance/GRIP-seq>.

shown to enable the protein to covalently crosslink to a bound RNA molecule *in vivo*. This method offers innovative avenues for developing protein-RNA research and applications.

INTRODUCTION

RNA-binding proteins (RBPs) regulate almost all aspects of RNA molecules inside cells, from pre-mRNA splicing, 3' tail processing, to RNA modification, translation, degradation, and localization¹. These regulatory roles of RBPs are essential for cells and organisms to maintain normal physiological status. Functional defects of RBPs could be the causes of many disorders, such as neurodegeneration and cancer^{2,3}. Numerous monogenic diseases have mutations enriched in RNA-binding regions, suggesting they arise from altered RNA binding⁴. Inside cells, most RBPs have a specific or multiple subcellular localizations, where they could interact with different sets of target RNA molecules through competing or collaborating with other RBPs⁵. In addition, hundreds of RBPs have recently been uncovered to lack conventional RNA-binding domains (RBDs) and many bind RNA with intrinsically disordered regions⁶. Some RBPs may even be inversely regulated by RNA. To understand the complex regulatory mechanisms and emerging novel aspects of RBPs, it is critical to identify the interactions between RBPs and their endogenous target RNA molecules under physiological conditions, ideally with nucleotide resolution and amino acid resolution.

Interactions of RBP and RNA *in vivo* are generally dynamic, transient, and weak⁷. To preserve RBP-RNA interactions for identification, the most widely used approach is nucleoside-based UV cross-linking⁸, in which nucleoside base produces radicals in response to UV-irradiation to cross-link with proximal amino acid residues⁹. Various technologies based on this mechanism have been developed for improving protein-RNA cross-linking. For example, photoactivatable ribonucleosides 4-thiouridine (4sU) or 6-thioguanosine (6sG) have been introduced into cell culture for incorporation into nascent RNA molecules, which allow RNA to cross-link with protein using UV-A (365 nm) light in higher efficiency^{10,11}. Together with immunoprecipitation (IP) and high-throughput sequencing techniques (CLIP-seq), RNA targets of many RBPs can be determined, which has largely expanded our understandings of RNA regulations^{6,10-15}. However, nucleoside-based UV cross-linking has strong nucleotide bias toward uridine or incorporated photoactivatable ribonucleosides^{9,10,16}, making it difficult to study RBPs lacking uridines in the target RNA regions. The high reactivity of nucleoside radicals tends to lead to more non-productive side reactions, decreasing detection sensitivity. Due to poor tissue penetrance, UV cross-linking cannot be applied to intact nontransparent animals for *in vivo* studies. More importantly, as the cross-linking moiety is generated on RNA and radicals cross-link with amino acid residues nonspecifically, nucleoside-based UV cross-linking makes amino acid resolution for RBPs technically demanding. For instance, RNA targets of different RNA-binding regions of an RBP cannot be reliably resolved. Moreover, nonspecific cross-linking of RNA to protein residues makes it infeasible to rationally design and engineer of proteins-RNA complex with precise covalent linkages.

Recently, latent bioreactive unnatural amino acids (Uaas) have been genetically incorporated into proteins in live cells, which react with specific natural amino acid residues via proximity-enabled reactivity^{17,18}. These latent bioreactive Uaas permit the selective chemical cross-linking of protein with protein both *in vitro* and *in vivo*, which has enabled a broad range of new applications such as pinpointing ligand-receptor binding, capturing elusive protein-protein interactions, and developing covalent protein drugs^{19–21}. To date, latent bioreactive Uaas have been applied to chemically cross-link proteins only, but not any other class of biomolecules.

Here we established a general biocompatible chemical method, genetically encoded chemical cross-linking of proteins with RNA (GECX-RNA) via latent bioreactive Uaas, which allows the site-specific introduction of covalent linkages between proteins and RNAs *in vivo* for the first time. We genetically incorporated two latent bioreactive Uaas, fluorosulfate-L-tyrosine (FSY) and *o*-sulfonyl fluoride-*O*-methyltyrosine (SFY), that were able to react with all four RNA nucleotides via proximity-enabled SuFEx reaction only when the RNA bound to RBP, irreversibly capturing target RNA on RBP in both *E. coli* and mammalian cells. By applying GECX-RNA to the RNA chaperon Hfq in *E. coli*, we demonstrated RNA identification with protein residue specificity *in vivo*. In addition, through genetic incorporation of SFY into the YTH domain in mammalian cells and combining with high-throughput sequencing, we devised a new, antibody-free method for *in vivo* identification of *N*⁶-methyladenosine (m6A) sites on RNA in the transcriptome with single-nucleotide resolution, uncovering previously unknown m6A sites. GECX-RNA now enables the study of protein-RNA interactions *in vivo* with single-nucleotide resolution for RNAs as well as amino acid specificity for proteins. Furthermore, selectively targeting RNA via proximity-enabled reactivity will open new avenues of generating stable covalent protein-RNA complex for research and therapeutic applications.

RESULTS

Developing GECX-RNA to cross-link RNA to RBP via genetically encoding latent bioreactive Uaas

The latent bioreactive Uaa fluorosulfate-L-tyrosine (FSY) has recently been genetically incorporated into proteins in *E. coli* and mammalian cells via a newly evolved orthogonal tRNA^{Pyl}/FSYRS pair²². Through proximity-enabled SuFEx reaction, the incorporated FSY specifically cross-links with proximal Lys, His, and Tyr side chains, forming covalent linkages within or between proteins *in vivo*²². When there is no nucleophiles placed in close proximity, the aryl fluorosulfate of FSY remains intact in proteins and inside cells. Based on such selective reactivity, we reasoned that SuFEx reactions could potentially target the nucleophilic 2'-hydroxyl group of ribose or amine groups of base in proximal nucleotides, thus forming specific covalent linkages between a protein and its bound RNA (Fig. 1a).

To test this hypothesis, we used Cas13 as the model protein to examine whether FSY-incorporated RBPs could cross-link with interacting target RNA. Cas13b is a class 2 type VI RNA-guided RNA-targeting CRISPR-Cas effector^{23–25}. Catalytically inactive Cas13b from *Prevotella sp. P5–125* (dPsCas13b) maintains targeted RNA binding activity and could only bind to targeted RNA through guidance of CRISPR RNA (crRNA)²⁵. Based

on the crystal structure of homologous Cas13a-crRNA-target RNA ternary complex²⁶ (Fig. 1b), we first prepared the catalytically inactive Cas13b (dPsCas13b) mutant by mutating His133 and His1058 of PsCas13b into alanine (Fig. 1b)²⁴. These two His are conserved catalytic residues responsible for RNA backbone cleavage. We thus incorporated FSY separately into position 133 and 1058 of dPsCas13b, reasoning that at these positions FSY side chain should aim at the 2'-OH group of RNA backbone (Fig. 1b). The wildtype dPsCas13b (dCas13-WT) and two FSY-incorporated dPsCas13b mutant proteins (dCas13b-133FSY, dCas13b-1058FSY) were expressed and purified, and then incubated with crRNA (crRNA-1) and target RNA (ssRNA-1). After incubation, electrophoretic mobility shift assay (EMSA) was performed by separating the samples on denaturing Urea-PAGE. RNA with cross-linked dCas13b proteins would run slower than RNA alone on Urea-PAGE. Indeed, we observed protein-RNA cross-linked bands for samples containing dCas13b-133FSY, whereas no such band for samples containing dCas13b-WT or dCas13b-1058FSY (Fig. 1c), suggesting that FSY-incorporated at His133 of dCas13b could cross-link with RNA.

To further validate whether the cross-linked bands were the cross-linking products of FSY-incorporated protein and RNA, we treated the cross-linked product of dCas13b-133FSY with proteinase K and re-analyzed with denaturing Urea-PAGE (Fig. 1c). Concomitant with the disappearance of the cross-linked bands, the target and guide RNA bands reappeared, indicating they were captured by dCas13b-133FSY protein. Apart from the target ssRNA, the excess guide crRNA was also covalently cross-linked by dCas13b-133FSY, which is consistent with the collateral RNA cleavage activity of cas13b *in vitro*²⁷. In addition, we verified the cross-linked band contained the target ssRNA by using 5'-fluorescently labeled target ssRNA (IRD680-ssRNA-1), which showed fluorescence in the cross-linked band (Fig. 1d). Moreover, in the absence of the guide crRNA, no cross-linked protein-RNA bands were detected (Fig. 1d). Since the presence of guide RNA is necessary for cas13b to bind and cleave RNA, this result indicated that the covalent cross-link of RNA to dCas13b-133FSY depends on RNA binding, which places RNA in close proximity to the FSY located at the catalytic site 133.

We next investigated whether FSY could cross-link the 2' hydroxyl group on the ribose agnostic to the base identity. To do so, we used another cleavage feature of Cas13b protein. In addition to the cleavage site responsible for target RNA cleavage, Cas13b protein also contains another cleavage site that site specifically cleaves precursor crRNA (pre-crRNA) into mature crRNA^{23,27,28}. Based on the crystal structure of *Bergeyella zoohelcum* Cas13b-crRNA binary complex²⁸ (Fig. 1e), we reasoned that positively charged amino acids on β -sheets 5 and 6 are involved in pre-crRNA cleavage because they are in close contact with two cleavage nucleotides located directly 3'-downstream of the hairpin repeat region of pre-crRNA (Fig 1f)²⁸. To identify the positively charged amino acids of PsCas13b involved in pre-crRNA cleavage, we mutated K367, K370, R378, and R380 of dPsCas13b into alanine, respectively, based on homology alignment (Fig. S1, Extended Data Fig. 1a). A pre-crRNA containing a 38-nt sequence at 3'-downstream of the hairpin repeat region of pre-crRNA was used as the cleavage target. We found that dPsCas13b-WT was active in pre-crRNA cleavage (Extended Data Fig. 1b). For dPsCas13b alanine mutants, K367A, K370A, and R378A mutants were still active in pre-crRNA cleavage, while R380A mutant abolished

pre-crRNA cleavage (Extended Data Fig. 1b). Considering R380 site is conserved among Cas13b proteins from different species (Fig. S1, Extended Data Fig. 1a) and its homologous R459 is in close contact with the backbone of the two cleavage nucleotides in the crystal structure of *Bergeyella zoohelcum* Cas13b-crRNA binary complex, R380 of dPsCas13b must play an important role in pre-crRNA cleavage. Thus, we mutated R380 of dPsCas13b into FSY, and incubate dPsCas13b-380 mutant proteins separately with four pre-crRNAs that had different nucleotide compositions at cleavage sites of pre-crRNA (Fig. 1g). As expected, no cross-linking was observed for dPsCas13b-380Ala incubations. Gratifyingly, dPsCas13b-380FSY protein was able to cross-link each pre-crRNA with all four types of nucleotides separately at the cleavage site (Fig 1g), indicating an inclusive nucleotide cross-linking reactivity of FSY when targeting the ribose 2'-hydroxyl.

These results together demonstrated that FSY incorporated into RBP could cross-link with RNA bound in proximity *in vitro*, representing the first proximity driven, genetically encoded chemical cross-linking of RNA (GECX-RNA) in an inclusive nucleotide-cross-linking manner.

GECX-RNA enables cross-linking of target RNA to RBP with amino acid specificity in *E. coli*

To test if GECX-RNA was suitable for capturing interactions of RBPs with target RNAs *in vivo*, we first examined if it could be applied for cross-linking the endogenous RNA targets of a bacterial RBP, host factor required for Q β replication (Hfq), in *E. coli*. Hfq is a widely conserved bacterial RNA chaperone²⁹, interacting with hundreds of sRNAs and more than one thousand mRNAs in Gram-negative bacteria, such as *E. coli*^{30–32}. In light of the structure of *E. coli* Hfq binding to target RNA³³, we introduced FSY into the binding interface at sites 25, 30, and 49 of *E. coli* Hfq protein (Fig. 2a), respectively. The *E. coli* Hfq-WT, Hfq-25FSY, Hfq-30FSY, or Hfq-49FSY protein was separately expressed in *E. coli* DH10B strain. After culturing, Western blot analysis of cell lysates showed cross-linking bands in all three samples expressing FSY-incorporated Hfq, but not Hfq-WT (Fig. 2b). These cross-linking bands disappeared or were down-shifted when samples were treated with RNase, indicating that the cross-linking products were Hfq-FSY proteins cross-linked with RNAs (Fig. 2b).

To determine whether the RNAs cross-linked with the Hfq-FSY proteins were endogenous target RNAs of Hfq, we expressed and purified Hfq proteins from *E. coli*, and examined the abundances of rpoS RNA, one known Hfq target RNA, co-purified with different Hfq proteins (Fig. 2c, Fig. S2a). rpoS RNA showed similar up-regulations in both Hfq-WT and Hfq-FSY expressing cells, suggesting that FSY-incorporated Hfq proteins were functional in *E. coli* cells (Fig. S2b). More importantly, RT-qPCR analysis showed that, in comparison with the RNA samples co-purified with Hfq-WT, those co-purified with Hfq-25FSY and Hfq-30FSY were more enriched in rpoS RNA (Fig. 2d). Hfq-49FSY did not enrich rpoS RNA, possibly because site 49 locates differently from sites 25 and 30 on the ring-shaped Hfq homo-hexamer (Fig. 2a) and rpoS RNA binds to Hfq in a position dependent manner. These results demonstrate that FSY-incorporation in Hfq could specifically cross-link and enrich target RNAs in *E. coli* cells.

A key potential of GECX-RNA is to capture and identify RNA with amino acid specificity of RBPs, since the chemical cross-linker is site-selectively introduced into RBP and not in RNA. To demonstrate this ability, we combined GECX-RNA with immunoprecipitation (GRIP) to determine RNA cross-linking sites of specific amino acid positions on RBPs (Fig. 2e). In general, RBP with FSY incorporated at a desired site will be expressed in cells to allow RNA cross-linking *in vivo*. Following cell lysis, RBP-FSY together with cross-linked RNA will be purified and digested with proteinase K to remove RBP and release the cross-linked RNAs. The RNAs will be reverse transcribed with gene-specific primer targeting a downstream region of cross-linked positions, which will terminate at the cross-linking site due to the cross-linked FSY residue. After removal of RNA, cDNA will be ligated to an adaptor at the 3' end and then amplified with PCR. The PCR products will be sequenced, from which the Uaa-induced cross-linking sites on target RNA will be identified at the ligation sites of PCR amplicons.

Site 25 of Hfq protein has been proposed to contact with an (AAN)₄ element on rpoS RNA for regulation³⁴, but all evidence is from either *in vitro* experiments or indirect *in vivo* indications. To directly detect the binding nucleotides of Tyr25 of Hfq protein in *E. coli* cells, we applied GRIP on Hfq-25FSY expressed in *E. coli* cells using gene-specific RT and PCR primers for rpoS. As expected, in the final PCR product, Hfq-WT sample had no insertion, while Hfq-25FSY sample had distinct insertion (Fig. S2c), indicating that crosslinking-mediated reverse-transcription-termination on rpoS RNA could be achieved with Hfq-25FSY but not Hfq-WT. Sanger sequencing of PCR products from the Hfq-25FSY sample revealed that reverse transcription was terminated at the (AAN)₄ element or its immediate 3' region (Extended Data Fig. 2). These results indicate that site 25 of Hfq protein contacted with the (AAN)₄ element on rpoS RNA in *E. coli* cells, providing *in vivo* evidence to support previous *in vitro* studies³⁴.

We further applied GRIP with gene-specific RT and PCR primers for another Hfq target mRNA, ptsG^{35,36}. Previously, it has been predicted that the ARN motifs in the UTR region of ptsG interact with Hfq proteins in *E. coli* cells³⁵, but there is no direct evidence of which domain of Hfq protein binds with ptsG mRNA *in vivo*. After Sanger sequencing of PCR products from the Hfq-25FSY sample, similar to the result for rpoS gene, we also identified that reverse transcripts of ptsG mRNAs terminated at the (ARN)₄ element or its immediate 3' region (Fig. S2d), indicating that site 25 of Hfq directly contacts the (ARN)₄ element of the ptsG RNA. These results from rpoS and ptsG mRNAs presented direct *in vivo* experimental evidence of site 25 of Hfq binding with ARN elements on target RNAs, demonstrating the power of GECX-RNA in probing *in vivo* protein-RNA interactions with amino acid specificity.

GECX-RNA enables specific cross-linking of target RNA to RBP in mammalian cells

We next tested whether GECX-RNA could work in mammalian cells and enable cross-linking of RBP with target RNA specifically. dCas13b proteins bind to specific endogenous RNA targets via the guidance of crRNA in mammalian systems^{24,25}, which should serve as an excellent system to test GECX-RNA for cross-linking and specificity. We therefore co-transfected plasmids expressing wild type dPsCas13b (WT) or FSY-incorporated

dPsCas13b (dPsCas13b-133FSY) together with crRNA expressing plasmids into HEK293T cells. Western blot analysis confirmed the successful expression of dPsCas13b-WT and dPsCas13b-133FSY protein (Fig. S3). They were purified through immunoprecipitation and digested by proteinase K to release the bound RNA (Fig. 3a). The RNAs were then reverse transcribed and quantified by qPCR using primers specific for the same gene as the guide crRNA. In the first case where target RNA was ACTB (Fig. 3b), in the negative control without introducing guide crRNA, dPsCas13b-133FSY had no enrichment of target RNA over dPsCas13b-WT, indicating that dPsCas13b-133FSY did not cross-link the target RNA when it was not bound. In contrast, in the presence of crRNA, dPsCas13b-WT still did not enrich target RNA over the control, while dPsCas13b-133FSY showed 4-fold enrichment over dPsCas13b-WT, indicating that at least for this system, cross-linking is necessary for target RNA enrichment. Similar results were also obtained for NEAT1 RNA using two guide crRNAs targeting different regions of NEAT1 RNA, respectively (Fig. 3b). These results demonstrate that GECX-RNA could cross-link and enrich target RNA in mammalian cells, and the cross-linking was dependent on RNA binding with RBP.

Genetically encoding SFY in mammalian cells to expand SuFEx-based cross-linking in cells

FSY has the SuFEx group at the *para* position, which has limited reaction area. To cope with different orientations of protein-RNA interactions, it would be desirable to encode a Uaa containing the SuFEx group at the *meta* position to expand reaction area³⁷. We thus evolved new orthogonal Mm-tRNA^{Pyl}/MmSFYRS and Ma-tRNA^{Pyl}/MaSFYRS pair to genetically incorporate *o*-sulfonyl fluoride-*O*-methyltyrosine (SFY) into proteins in *E. coli* (Fig. 4a). Here we demonstrate the incorporation of SFY into proteins in mammalian cells and the ability of SFY to cross-link proximal nucleophilic amino acid sidechains via SuFEx directly in *E. coli* and mammalian cells.

To test SFY incorporation in mammalian cells, we transfected HEK293 cells with plasmid pcDNA-EGFP-40TAG expressing EGFP gene containing a TAG codon at site Tyr40 and plasmid pNEU-MmSFYRS expressing the Mm-tRNA^{Pyl}/MmSFYRS. Fluorescence confocal microscopy showed that, in the presence of SFY, strong EGFP fluorescence was observed throughout the cells, and cell morphology remained normal (Fig. 4b), indicating SFY was incorporated at the TAG site to produce full-length EGFP. No fluorescence signal was detected when SFY was not added. HEK293 cells expressing pcDNA-EGFP-40TAG and Mm-tRNA^{Pyl}/MmSFYRS or Ma-tRNA^{Pyl}/MaSFYRS were further quantified by flow cytometry (Fig. 4c, Fig. S4). Strong EGFP fluorescence was measured from cells only when SFY was added, and the fluorescence intensity increased with tRNA^{Pyl} copy number. The incorporation efficiency of SFY was comparable with FSY. In addition, we did not observe obvious toxicity of SFY to HEK293T cells (Fig. S5), a valuable property for in cell applications.

To determine which amino acid residues could react with SFY via proximity-enabled reactivity directly in cells, we coexpressed in *E. coli* the Z protein and an affibody (Afb) that specifically binds the Z protein. Based on the crystal structure of Afb-Z complex, we introduced SFY at site 24 of the Z protein and various natural residues at site 7 of

the affibody (Fig. 4d), placing the two residues in close proximity upon Afb-Z binding²². After expression of Afb-24SFY and Z-7X (X = target residue) for 6 h, cells were lysed and analyzed with Western blot under denatured conditions (Fig. 4e). Cross-linking bands corresponding to the adduct of Afb and Z were detected for target residue His, Tyr, and Lys, for both Mm-tRNA^{Pyl}/MmSFYRS and Ma-tRNA^{Pyl}/MaSFYRS. We then purified 6xHis-tagged Z and Afb proteins from cells and analyzed with SDS-PAGE. Consistently, a protein band corresponding to the cross-linked Z with Afb was clearly observed for Afb-7Lys, Afb-7His, and Afb-7Tyr (Fig. 4e). We further tested if SFY could cross-link with these residues in mammalian cells. GST is a dimeric protein, whose structure shows that residue 103 of one monomer is close to residue 107 of the other monomer at the dimer interface (Fig. 4f), which has been used to determine proximity-enabled reactivity³⁸. We incorporated SFY at site 103 of GST and mutated residue 107 to various target residues. HEK293T cells expressing these GST mutants were lysed and Western blotted to detect covalent GST dimer formation (Fig. 4g). Clearly SFY was shown to react with His, Tyr, and Lys placed in proximity in mammalian cells.

We also verified if SFY incorporated into Hfq could covalently capture RNA in *E. coli* cells. *E. coli* DH10B cells expressing Hfq-25SFY or Hfq-49SFY were lysed and analyzed with Urea-PAGE (Fig. 4h). Cross-linking bands were detected, some of which disappeared when samples were treated with RNase, indicating that Hfq-SFY was able to cross-link RNAs in *E. coli*. In addition, to check if SFY could cross-link nucleotides, we incubated 50 mM SFY with 50 mM different nucleoside monophosphates (NMPs: AMP, UMP, CMP, or GMP) at 37 °C for 16 h. Cross-linking adducts of SFY with all four NMPs were detected using MS, confirming SFY could also cross-link all four nucleotides (Fig. S6).

Due to the respective *meta* and *para* positioning of warheads in SFY and FSY, they should complement each other in targeting His, Tyr, and Lys with different side chain orientations in proteins. In addition, sulfonyl fluoride of SFY is chemically more reactive than fluorosulfonate of FSY, which enables SFY to covalently target weaker nucleophiles elusive to FSY, such as the hydroxyl group of glycans (companion manuscript).

An *in vivo* method for detecting m6A in mammalian cells with single-nucleotide resolution

N⁶-methyladenosine (m6A) is a widespread RNA modification that play important roles in the regulations and functions of mRNA³⁹. Identification of the m6A sites in RNA is critical for understanding m6A function. Although many m6A detection methods have been reported^{40–42}, the majority of them lack single nucleotide resolution and rely on the use of m6A-specific antibody, in which the recognition of m6A is *in vitro* in nature. Enlightened by the success of applying GRIP in identifying the cross-linked nucleotides of rpoS RNA in *E. coli* above, we reasoned that GRIP could be adapted for capturing m6A sites on RNA *in vivo* for subsequent identification, which may better preserve m6A physiological status than the *in vitro* methods. Specifically, we proposed to use a reader protein of m6A to recognize m6A sites on RNA, and to incorporate a bioreactive Uaa into the m6A binding site of the reader to cross-link nucleotides neighboring m6A (Fig. 5a). Expression of the reader-Uaa protein in cells would cross-link at m6A sites on RNA, enabling the recognition and capture of m6A *in vivo*. Immunoprecipitation of the reader protein followed with proteinase K

digestion then release the captured RNAs for reverse transcription, adaptor ligation, and sequencing (Fig. 5a). The identified Uaa-cross-linked nucleotides thus reveal m6A site to be immediately adjacent.

We decided to use the YTH domain of human YTHDF1 protein, which is a conserved m6A reader^{43,44}. Based on the crystal structure of YTHDF1 in complex with a 5-mer m6A RNA⁴³, we chose Tyr397, a residue next to the binding pocket of m6A, as the site for incorporating the bioreactive Uaa, to aim the Uaa side chain for targeting nucleotides upstream of m6A (Fig. 5b). Initial incorporation of FSY at site 397 of the YTH domain failed to cross-link any RNA. A more careful analysis of the structure revealed a Lys469 at the *para*-position of Tyr397, which is known to react with FSY. Therefore, we changed to the new SuFEx capable bioreactive Uaa, SFY, which has similar proximity-enabled reactivity as FSY in cross-linking nearby Lys, His, and Tyr. As SFY has sulfonyl fluoride installed at the *meta*-position of the phenyl ring, which should avoid Lys469 contact and reactivity (Fig. 4a, Fig. 5b). As shown below, YTH-397SFY was indeed able to cross-link m6A containing RNAs in mammalian cells.

To detect endogenous m6A sites in mammalian cells throughout the transcriptome, we developed GRIP-seq through combining GRIP for m6A with high-throughput sequencing, enabling identification of m6A sites *in vivo* with single-nucleotide resolution (Fig. 5a). In brief, HEK293T cells expressing YTH-397SFY protein (Fig. S7a) were lysed and treated with RNase to partially digest RNA into short fragments. After GRIP for these cell lysates (Fig. S7a), the purified protein-RNA cross-links were treated with proteinase K to release the cross-linked RNA fragments, which were converted into a cDNA library through adapting the enhanced CLIP protocol⁴⁵ and then subjected to high-throughput sequencing. All m6A sites cross-linked by YTH-397SFY protein *in vivo* would be captured and identified by reverse transcription termination at the upstream cross-linked nucleotides.

We generated four pairs of GRIP-seq libraries. For each pair, we generated one library for the INPUT sample, which represents the RNA fragments from the whole cell lysate, and one library for the IP sample, which represents the RNA fragments cross-linked with the purified YTH proteins. These four pairs included one pair from HEK293 cells expressing YTH-WT protein serving as quality control, and three pairs from the three biological replicates of HEK293 cells expressing YTH-397SFY protein. For each library, around 10 to 35 million reads were obtained (Table S2). After removing adaptors, we first mapped the reads to the transcriptome. For IP libraries, we then used the CLIPPER algorithm⁴⁶ to identify enriched peaks, which would represent RNA regions covering the reverse transcriptional termination sites and the cross-linking sites. While only 16,659 peaks were identified from the YTH-WT IP sample, 118,746, 151,153, and 139,741 peaks were separately identified from the three YTH-397SFY IP samples. Aside from the drastic difference in total peak numbers between YTH-397SFY and YTH-WT IP samples, comparisons of each gene's peak numbers among the three YTH-397SFY IP samples indicated high reproducibility (Pearson's $r > 0.96$, Fig S7b), while the comparison of each gene's peak numbers between YTH-397SFY-2 and YTH-WT IP samples showed low correlation (Pearson's $r = 0.29$, Fig S7b). These results demonstrate that the peaks in YTH-397SFY replicates were specifically introduced through SFY incorporation and cross-linking.

To determine if YTH-397SFY IP samples enriched m6A sites, we first identified the cross-linking-caused reverse-transcription-termination sites in these peaks (see materials and methods). Next, we performed the sequence logo analysis of the sequences surrounding these reverse-transcription-termination sites. In all YTH-397SFY IP samples, the highest enriched motif was DRACH motif, which matched exactly the preferred consensus motif for m6A^{40–42,44} (Fig. 6a, Fig. S7c). On the other hand, such DRACH motif could not be found enriched in the YTH-WT IP sample (Fig S7c). In addition, the metagene profiles for the reverse-transcription-termination sites from the YTH-397SFY IP samples followed the typical distributions of m6A along mRNAs with strong enrichments around the stop codon (Fig. 6b), while those from the YTH-WT IP sample did not. Moreover, examinations of many individual RNAs, for example, JUN mRNA and DICER1 mRNA, showed that peaks from YTH-397SFY IP samples specifically enriched and terminated at previously identified m6A sites⁴⁷ (Extended Data Fig. 3). These data indicated that *in vivo* expression of YTH-397SFY protein specifically cross-linked and enriched m6A modified RNA in mammalian cells.

In our design, the SFY residue in YTH-397SFY protein should cross-link with the nucleotide at the close upstream of m6A (Fig. 5a, 5b). To pinpoint which nucleotide next to m6A was cross-linked by YTH-397SFY, we analyzed the position of the cross-linked nucleotide relative to the DRACH motif in DRACH-containing reads from the enriched peaks. Indeed, if we denoted the middle A (m6A) in DRACH motif as position 0, cross-linking occurred at position –3 in 80.4% of DRACH-containing reads, and at position –4 in 9.3% of DRACH-containing reads (Fig. 6c), demonstrating that GRIP-seq could identify the m6A site with single-nucleotide-resolution. Moreover, analyzing the nucleotide composition of the cross-linked nucleotides revealed that SFY could cross-link with all four RNA nucleotides *in vivo* (Fig. 6d, Fig. S7d), consistent with our *in vitro* experiment data (Fig. 1i, Fig. S6).

Based on these features, we predicted a total of 13,968 m6A sites from the GRIP-seq data (Table S3). To further validate the m6A sites identified in GRIP-seq, we applied individual m6A GRIP procedures for two RNA regions that contain known m6A sites in JUN mRNA and DICER mRNA⁴⁷, employing gene-specific reverse transcription, ligation, amplification, and Sanger sequencing (Fig. S7e). As expected, in final PCR products YTH-WT samples had no insertion, while YTH-397SFY samples showed distinct insertions for both genes (Fig. S7f). After cloning and Sanger sequencing of YTH-397SFY PCR products, the identified cross-linking sites from Sanger sequencing matched the cross-linking sites from the GRIP-seq data (Fig. S7g, S7h), confirming that GRIP-seq was able to correctly identify m6A sites in mammalian cells as designed. Interestingly, apart from the known m6A site⁴⁷, in the amplified region of JUN gene, we also identified one novel m6A site using GRIP with Sanger sequencing capacity (Fig. S7g), which was also identified in the GRIP-seq data.

To further evaluate the capacity of GRIP-seq for identifying novel m6A sites, we compared the m6A sites from GRIP-seq with the known human m6A sites from the m6A-atlas⁴⁷, a comprehensive database for human m6A sites collected from seven published m6A-identification methods. The 6,072 m6A sites from GRIP-seq were known m6A sites that have been annotated in the m6A atlas, further validating GRIP-seq's ability in identifying

m6A. Interestingly, 7,896 m6A sites from GRIP-seq have not been reported by any method in the m6A-atlas. Sequence logo analysis of these novel m6A sites from GRIP-seq showed strong enrichment of DRACH motif (Fig. S7i), and the metagene profile of these novel m6A sites also followed the typical distributions of m6A along mRNAs (Fig. S7j). These results demonstrated that GRIP-seq was able to identify new m6A sites undiscovered before.

RNA secondary structure could alter the ability of RBPs' binding to target RNA⁴⁸ and the reactivity of RNA nucleotides⁴⁹. To assess the potential effect of RNA secondary structure on GECX-RNA, we analyzed the predicted structural potential⁵⁰ in RNA regions surrounding m6A sites from GRIP-seq and from the m6A-atlas⁴⁷, respectively. The m6A regions from GRIP-seq displayed a slightly less potential for stable secondary structures than the m6A regions from the m6A-atlas (Fig. S7k). However, most of m6A sites from the m6A-atlas were identified through detecting m6A on purified RNA molecules *in vitro*, while GRIP-Seq detected m6A on native cellular RNAs *in vivo*. *In vitro* purification and detection could disrupt stable *in vivo* RNA secondary structures and allow m6A more accessible for detection. We thus further compared the predicted structural potential in RNA regions surrounding m6A sites from GRIP-seq with those from DART-seq, another method that detects m6A sites *in vivo*^{44,47}. Interestingly, the m6A regions from GRIP-seq showed a much greater potential for secondary structures than those from DART-seq (Fig. S7k). Together, these results suggest that secondary structure folding in m6A regions from GRIP-seq is likely reflecting the *in vivo* binding preference of YTH domain for m6A RNAs.

The proximity driven reactivity of GECX-RNA would enable cross-link with target RNA continuously whenever interaction occurs, allowing enriching the cross-linked product over a long period to improve the capture of interactions on low abundance RNAs. To determine if GRIP-seq was able to detect unknown m6A modifications on low abundance RNAs, we examined the abundance of mRNAs containing m6A sites detected by GRIP-seq. Among the m6A sites identified with GRIP-seq, 6,072 sites were also detected by previous methods and thus termed as “known m6A sites”, while 7,896 sites were detected by GRIP-seq only and termed as “novel m6A sites”. Between the group of genes containing only the known m6A sites and the group of genes containing only the novel m6A sites, we found that the genes containing only the novel m6A sites had slightly yet significantly lower RNA abundances (Fig. 6e). Such low RNA abundances probably caused the neglect of these “novel” m6A sites in other m6A detection methods. Therefore, these results demonstrate that GRIP-seq was capable of capturing protein-RNA interactions on the low abundance RNAs.

DISCUSSION

In summary, through genetic incorporation of latent bioreactive Uaas capable of reacting with RNA in proximity, we developed a novel method, genetically encoded chemical cross-linking of proteins with RNA (GECX-RNA) *in vivo*. GECX-RNA was able to covalently capture target RNA onto RBP when they interacted *in vitro*, in *E. coli*, and in mammalian cells, providing resolution of not only nucleotide but also amino acid residue. By applying GECX-RNA on RNA chaperon Hfq in *E. coli*, we demonstrated RNA cross-linking and identification with amino acid specificity of Hfq. Adapting GECX-RNA in mammalian

cells, we developed a GRIP method for *in vivo* detection of m6A sites on RNA in the transcriptome with single-nucleotide resolution.

GECX-RNA affords several advantages over the common nucleoside-based UV cross-linking for studying protein-RNA interactions. Nucleoside-based UV cross-linking cross-links with uridine predominantly^{9,10,16}, making it unsuitable for RBPs that bind uridine-lacking RNAs, such as poly-A binding proteins^{51,52}. By targeting the 2'-OH group of ribose, GECX-RNA could cross-link all four nucleotides. While UV cross-linking occurs in the short UV irradiation window, GECX-RNA continuously captures RNA whenever they interact, allowing enriching the cross-linked product over a long period to improve detection sensitivity²⁰, which is particularly valuable for detecting dynamic RNA events with weak and transient interactions⁷ or events on low abundance RNAs. UV light offers spatiotemporal control but cannot be applied in nontransparent animal models. In contrast, GECX-RNA reacts spontaneously upon binding, obviating the external light trigger and timing issue. Giving the ability to genetically encode Uaas in animals such as *C. elegans* and mouse⁵³, GECX-RNA should be compatible for *in vivo* use in animals for interrogation in physiological settings.

Compared with existing methods recognizing m6A *in vitro* via antibody, our GECX-RNA based GRIP method represents an antibody-free approach for identifying m6A with single-nucleotide resolution *in vivo*, which should reflect m6A physiological status more closely. Combination of GRIP with high-throughput sequencing¹⁵ (GRIP-seq) enabled precise mapping m6A sites in the transcriptome. The large number of novel m6A sites identified from GRIP-seq demonstrated its ability for discovering unknown target RNA sites. GRIP-seq could also be applied on different m6A recognition proteins to measure/compare their m6A binding behaviors *in vivo*. In addition, the GRIP strategy can be generalized to map other RNA modifications *in vivo* for which a reader or binder exists.

A key advance of GECX-RNA is gaining amino acid resolution for RBPs, which is technically demanding and often unsuccessful for nucleoside-based UV cross-linking methods. GECX-RNA possesses dual resolutions of nucleotide for RNA and of amino acid for RBP, thus dramatically expanding the scope of protein-RNA studies. As a proof-of-principle, we demonstrated RNA identification of Hfq specific for site 25. Indeed, many RBPs possess multiple RNA-binding domains, co-operations among which are critical for the RBP function^{49,50}. A large number of novel RNA-binding regions in unconventional RBPs have recently been identified^{4,6}, and mutations causing Mendelian genetic diseases are found enriched in these RNA-binding regions, emphasizing the need to study these RNA-binding regions with amino acid resolution⁴. RNA-GECX will provide useful tools to investigate these emerging novel aspects of protein-RNA interactions *in vivo*¹⁹.

Possible limitations of GECX-RNA are noted. First, although GECX-RNA could enrich the RBP-RNA cross-links over a long period to enhance the detection sensitivity, such continuously cross-linking between RBP and RNA could potentially perturb the process under study. The potential side effects can be monitored, such as using RNA-seq to monitor the RNA global profile changes, and can be mitigated by decreasing the amount of covalent RBP. Second, structural information of RBP-RNA interactions is highly valuable

for selecting the appropriate sites for Uaa incorporation to achieve covalent targeting of RNA. The Uaa incorporation site should have no direct contacting natural residues that could react with the Uaa. In the absence of structural guidance, screening a large number of sites of the RBP is also feasible by harnessing the genetic encoding nature of GECX-RNA. Fortunately, recent rapid development in the accurate prediction of protein structures and interactions^{56,57} will help make a broad range of proteins amenable to GECX-RNA. Third, when detecting RBP-RNA interaction, RNA secondary structure could be a factor affecting the cross-linking efficiency. In GECX-RNA, the cross-linking reaction is determined by the chemical nature and proximity of the Uaa and the target RNA nucleotide. RNA secondary structure could affect the binding ability of the RBP and thus indirectly the cross-linking efficiency. Using different RBPs preferring various RNA secondary structures as the binder or sensor would yield a more comprehensive coverage as well as RBP-specific mapping of RNA interaction.

Other than cross-linking-based methods, another strategy for detecting RBP-RNA interaction is to fuse RNA-modifying modules, such as nucleoside deaminases, to RBPs-of-interest. The resultant fusion protein can modify RNA nucleotides neighboring with the bound-RNA for subsequent detection^{44,58,59}. Although such strategy could circumvent the need for IP, it could not attain single-nucleotide-resolution in detecting RNA-binding sites, since the modified sites often do not have exactly fixed distances to the RNA-binding sites.

Lastly, we demonstrate here the selective targeting of RNA via proximity-enabled reactivity of genetically encoded latent bioreactive Uaas, which was previously limited to targeting proteins¹⁸. The ability to selectively engineer covalent bonds between proteins has led to a range of innovative applications, such as pinpointing ligand-receptor on live mammalian cells and developing covalent protein drugs^{19,21,60}. Similarly, judicious engineering of a covalent linkage between protein and RNA, enabled by GECX-RNA here, will inspire novel avenues for RNA-related research and therapeutics, such as live transcript imaging, epitranscriptomic modification, interrogation of lncRNA, translational modulation, gene editing, and so on.

METHODS

All oligonucleotides are synthesized at IDT. Sequences of all oligonucleotides in materials and methods are listed in Table S1.

Cloning of pBAD-dCas13 plasmids

pBAD-PsCas13b plasmids—PsCas13b was PCR amplified from pC0046-EF1a-PspCas13b-NES-HIV plasmid (addgene #103862) with primers of pBAD-Nde1-GA-PsCas13b-F and pBAD-Hind3-GA-3HA-R. To generate pBAD-PsCas13b plasmid, the PCR product was cloned into pBAD vector pre-digested with NdeI and HindIII using Gibson Assembly kit (New England Biolabs).

To generate pBAD-dLwCas13b plasmid, residues 133 and 1058 of PsCas13b gene in pBAD-PsCas13b were mutated into alanine codons using site-directed mutagenesis with following primers: primer pair of PspCas13b-H133A-mut-F and PspCas13b-H133-mut-R for H133A

mutation; primer pair of PsCas13b-H1058A-2-mut-F and PsCas13b-H1058-2-mut-R for H1058A mutation.

pBAD-dPsCas13b-TAG-mutants for crRNA-1 and ssRNA-1 cross-linking

To generate pBAD-dPsCas13b-TAG mutant plasmids for crRNA-1 and ssRNA-1 cross-linking, residues 133 and 1058 of dPsCas13b gene in pBAD-dPsCas13b were mutated into an amber stop codon TAG, respectively, using site-directed mutagenesis with following primers:

- primer pair of PspCas13b-H133TAG-mut-F and PspCas13b-H133-mut-R for 133TAG mutation, final plasmid: pBAD-dPsCas13b-133TAG;
- primer pair of PsCas13b-H1058TAG-2-mut-F and PsCas13b-H1058-2-mut-R for H1058TAG mutation, final plasmid pBAD-dPsCas13b-1058TAG.

All pBAD-PsCas13b plasmids contain HA tag and 6His tag at C-terminals.

pBAD-dPsCas13b-mutants for pre-crRNA cleavage and cross-linking assay

To generate pBAD-dPsCas13b mutant plasmids for pre-crRNA cleavage and cross-linking assay, residues 367, 370, 378, and 380 of dPsCas13b gene in pBAD-dPsCas13b were mutated into either an alanine codon or an amber stop codon TAG, respectively, using site-directed mutagenesis with following primers:

- primer pair of Ps13b-K367A-v-F and Ps13b-K367A-v-R for K367A mutation, final plasmid: pBAD-dPsCas13b-367A.
- primer pair of Ps13b-K370A-v-F and Ps13b-K367A-v-R for K370A mutation, final plasmid: pBAD-dPsCas13b-370A.
- primer pair of Ps13b-R378A-v-F and Ps13b-R378A-v-R for R378A mutation, final plasmid: pBAD-dPsCas13b-378A.
- primer pair of Ps13b-R380A-v-F and Ps13b-R380A-v-R for R380A mutation, final plasmid: pBAD-dPsCas13b-380A.
- primer pair of Ps13b-K367U-v-F and Ps13b-K367U-v-R for K367TAG mutation, final plasmid: pBAD-dPsCas13b-367TAG.
- primer pair of Ps13b-K370U-v-F and Ps13b-K367U-v-R for K370TAG mutation, final plasmid: pBAD-dPsCas13b-370TAG.
- primer pair of Ps13b-R378U-v-F and Ps13b-R378U-v-R for R378TAG mutation, final plasmid: pBAD-dPsCas13b-378TAG.
- primer pair of Ps13b-R380U-v-F and Ps13b-R380U-v-R for R380TAG mutation, final plasmid: pBAD-dPsCas13b-380TAG.

All pBAD-PsCas13b plasmids contain HA tag and 6His tag at C-terminals.

dCas13b protein expression and purification

dPsCas13b-WT; dPsCas13b-360A; dPsCas13b-367A; dPsCas13b-370A; dPsCas13b-378A; dPsCas13b-380A—pBAD-dPsCas13b plasmid (dPsCas13b-WT, dPsCas13b-360A, dPsCas13b-367A, dPsCas13b-370A, dPsCas13b-378A, or dPsCas13b-380A) was transformed into DH10B *E. coli* chemical competent cells. The transformants were plated on an LB-Amp100 agar plate and incubated overnight at 37 °C. A single colony was inoculated into 5 mL of 2xYT- Amp100 and cultured overnight at 37 °C. On the following day, 1 mL of overnight cell culture was diluted into 50 mL 2xYT- Amp100 and agitated vigorously at 37 °C. When OD600 reached 0.4~0.6, the cell culture was induced with 0.2% arabinose, then incubated at 18 °C for 18 h. Cell pellets were collected by centrifugation at 4200 g for 30 min at 4 °C and stored at –80 °C.

dPsCas13b-133FSY; dPsCas13b-1058FSY; dPsCas13b-360FSY; dPsCas13b-367FSY; dPsCas13b-370FSY; dPsCas13b-378FSY; dPsCas13b-380FSY—pBAD-dPsCas13b TAG mutant plasmid (pBAD-dPsCas13b-133TAG, pBAD-dPsCas13b-1058TAG, dPsCas13b-360TAG, dPsCas13b-367TAG, dPsCas13b-370TAG, dPsCas13b-378 TAG, or dPsCas13b-380TAG) was co-transformed with pEvol-FSYRS²²(encoding FSY-tRNA synthetase-tRNA system for expression in *E. coli* cells) into DH10B *E. coli* chemical competent cells, respectively. The transformants were plated on an LB-Amp100Cm34 agar plate and incubated overnight at 37 °C. A single colony was inoculated into 5 mL of 2xYT- Amp100Cm34 and cultured overnight at 37 °C. On the following day, 1 mL of overnight cell culture was diluted into 50 mL 2xYT- Amp100Cm34 and agitated vigorously at 37 °C. When OD600 reached 0.4~0.6, the cell culture was induced with 0.2% arabinose and 1 mM FSY (FSY was chemically synthesized as previously reported²²), then incubated at 18 °C for 18 h. Cell pellets were collected by centrifugation at 4200 g for 30 min at 4 °C and stored at –80 °C.

His-tag protein purification

Above cell pellets were resuspended in 14 mL lysis buffer (20 mM Tris-HCl pH 7.4, 500 mM NaCl, 20 mM imidazole, lysozyme 1 mg/mL, and proteinase inhibitors). The cell suspension was lysed at 4 °C for 30 min. Cell lysate was sonicated with Sonic Dismembrator (Fisher Scientific, 30% output, 3 min, 1 sec off, 1 sec on) in an ice-water bath, followed by centrifugation (20,000 g, 30 min, 4 °C). The soluble fractions were collected and incubated with pre-equilibrated Protino®Ni-NTA Agarose resin (400 µL) at 4 °C for 1 h with constant mechanical rotation. The slurry was loaded onto a Poly-Prep® Chromatography Column, washed with 5 mL of wash buffer (20 mM Tris-HCl pH 7.4, 500 mM NaCl, 20 mM imidazole, 2 mM DTT) for 3 times, and eluted with 200 µL of elution buffer (20 mM Tris-HCl pH 7.4, 500 mM NaCl, 500 mM imidazole, 2 mM DTT, 10% glycerol) for 5 times. The eluates were concentrated, and buffer exchanged into 100 µL of protein storage buffer (20 mM Tris-HCl pH 7.4, 500 mM NaCl, 2 mM DTT, 10% glycerol) using Amicon Ultra columns, and stored at –80 °C for future analysis.

RNA transcription and labeling

Templates for crRNA (crRNA-1) were PCR amplified with following primers (T7pro-Ps13b-cr-1-F and Ps13b-crRNA-1-R) to yield dsDNA and then incubated with T7 polymerase at 37°C overnight using the MAXIscript T7 transcription kit (Thermo Fischer Scientific). crRNA-1 was purified using Clean and Concentrator columns (Zymo Research).

Templates for target RNA (ssRNA-1) were PCR amplified with following primers (primer pair of T7pro-ssRNA-1-F and ssRNA-1-R for target RNA ssRNA-1) to yield dsDNA and then incubated with T7 polymerase at 37°C overnight using the MAXIscript T7 transcription kit. Target RNAs (ssRNAs) were purified using Clean and Concentrator columns. 5' end labeling was performed on target RNA (ssRNA-1) using the 5' oligonucleotide kit (VectorLabs, Burlingame, CA) and with a maleimide-IR800 probe (LI-COR Biosciences, Lincoln, NE). Labeled target RNAs (5'-IRD680-ssRNA-1) were purified using Clean and Concentrator columns.

Templates for pre-crRNAs were PCR amplified using PCR amplicon of crRNA-1 as template with following primers (primer pair of T7pro-ssRNA-1-F and Ps13b-pre-crRNA-AAA-R for pre-crRNA-AAA; primer pair of T7pro-ssRNA-1-F and Ps13b-pre-crRNA-UUU-R for pre-crRNA-UUU; primer pair of T7pro-ssRNA-1-F and Ps13b-pre-crRNA-CCC-R for pre-crRNA-CCC; primer pair of T7pro-ssRNA-1-F and Ps13b-pre-crRNA-GGG-R for pre-crRNA-GGG) to yield dsDNA and then incubated with T7 polymerase at 37°C overnight using the MAXIscript T7 transcription kit (Thermo Fischer Scientific). pre-crRNAs were purified using Clean and Concentrator columns (Zymo Research).

dCas13b-crRNA-ssRNA in vitro cross-linking assay

dCas13b cross-linking assays were performed with 10 nM of unlabeled or 5'-IRD-680 labeled ssRNA target, 100 nM purified dPsCas13b proteins, and 30 nM crRNA in 40 mM Tris-HCl, 60 mM NaCl, 10 mM EDTA, 10 µg/mL heparin, pH 7.4. Incubations were performed at 37°C overnight. After incubation, the samples were then denatured with laemmli sample buffer (1% LDS, 50 mM DTT) at 95°C for 5 minutes. Samples were analyzed by denaturing gel electrophoresis on 10% 8 M Urea TBE PAGE. Gels were imaged by scanning fluorescent signal after SybrGold staining (for ssRNA-1), or by direct scanning with an Odyssey scanner (LI-COR Biosciences) (for 5'-IRD680-ssRNA-1).

pre-crRNA cleavage assay

pre-crRNA cleavage assays were performed with 0.2 µM of pre-crRNA-AAA and 1 µM of purified dPsCas13b proteins in 40 mM Tris-HCl, 60 mM NaCl, 10 mM EDTA, 5 mM MgCl₂, pH 7.4. Incubations were performed at 37°C for 45 min. After incubation, to quenched the reactions, the samples were immediately denatured with laemmli sample buffer (1% LDS, 50 mM DTT) at 95°C for 5 min. Samples were analyzed by denaturing gel electrophoresis on 10% 8 M Urea TBE PAGE. Gels were imaged by scanning fluorescent signal after SybrGold staining.

dCas13b-380 mutant proteins and pre-crRNA in vitro cross-linking assay

pre-crRNA cross-linking assays were performed with 40 nM of different pre-crRNAs (pre-crRNA-AAA, pre-crRNA-UUU, pre-crRNA-CCC, or pre-crRNA-GGG) and 200 nM of different purified dPsCas13b-380 mutant proteins (380A or 380FSY) in 40 mM Tris-HCl, 60 mM NaCl, 10 mM EDTA, pH 7.4. Incubations were performed at 37°C overnight. After incubation, electrophoretic mobility shift assay (EMSA) was performed on the samples. In brief, the samples were denatured with laemmli sample buffer (1% LDS, 50 mM DTT) at 95°C for 5 min. Samples were analyzed by denaturing gel electrophoresis on 10% 8 M Urea TBE PAGE. Gels were imaged by scanning fluorescent signal after SybrGold staining.

Cloning of pBAD-Hfq plasmids

To generate pBAD-Hfq-WT plasmid, the Hfq encoding gene was amplified by colony PCR with primer pair of Hfq-Nde1-F and Hfq-6H-Hind3-R, digested with Nde I and Hind III, and ligated into the pBAD vector pre-treated with the same restriction enzymes.

To generate pBAD-Hfq-TAG mutant plasmids, residue 25, 30 and 49 of Hfq gene in pBAD-Hfq-WT were mutated into an amber stop codon TAG, respectively, using site-directed mutagenesis with following primers:

- primer pair of Hfq-Tyr25TAG-F and Hfq-Tyr25TAG-R for 25TAG mutation, final plasmid: pBAD-Hfq-25TAG;
- primer pair of Hfq-Ile30TAG-F and Hfq-Ile30TAG-R for 30TAG mutation, final plasmid: pBAD-Hfq-30TAG;
- primer pair of Hfq-Thr49TAG-F and Hfq-Thr49TAG-R for 49TAG mutation, final plasmid: pBAD-Hfq-49TAG.

All pBAD-Hfq plasmids contain 6His tag at C-terminals.

Expression of exogenous Hfq proteins in *E. coli* cells (Hfq-WT and Hfq-FSY samples)

pBAD-Hfq-WT was transformed into DH10B *E. coli* chemical competent cells. The transformants were plated on an LB-Amp100 agar plate and incubated overnight at 37 °C. A single colony was inoculated into 5 mL of LB- Amp100 and cultured overnight at 37 °C. On the following day, 1 mL of overnight cell culture was diluted into 15 mL LB- Amp100 and agitated vigorously at 37 °C. When OD600 reached 0.4~0.6, the cell culture was induced with 0.2% arabinose, then incubated at 37 °C for 16 h. Cell pellets were collected by centrifugation at 4200 g for 30 min at 4 °C and stored at -80 °C.

pBAD-Hfq TAG mutant plasmids (pBAD-Hfq-25TAG, pBAD-Hfq-30TAG, pBAD-Hfq-49TAG) was co-transformed with pEvol-FSYRS²² into DH10B *E. coli* chemical competent cells, respectively. The transformants were plated on an LB-Amp100Cm34 agar plate and incubated overnight at 37 °C. A single colony was inoculated into 5 mL of LB-Amp100Cm34 and cultured overnight at 37 °C. On the following day, 1 mL of overnight cell culture was diluted into 15 mL LB- Amp100Cm34 and agitated vigorously at 37 °C. When OD600 reached 0.4~0.6, the cell culture was induced with 0.2% arabinose and 1 mM FSY,

then incubated at 37 °C for 16 h. Cell pellets were collected by centrifugation at 4200 g for 30 min at 4 °C and stored at –80 °C.

RNase treatment and detection for exogenous Hfq-expressing *E. coli* cells (Hfq-WT and Hfq-FSY samples)

For exogenous Hfq-expressing *E. coli* cell pellets (Hfq-WT and Hfq-FSY samples), 100 µL PBS were added to resuspended the cell pellets. Then 200 µL 0.5mm glass beads were added. The samples were then put into dry ice, and then lyophilized for 3hr. Dried cells were disrupted by vortexing for 5 min, at intervals of 1 min to avoid warming the sample. The disrupted samples were then resuspended in 1ml PBS. For RNase treated samples, 10 µL of resuspended disrupted sample was aliquoted, added with 1U/µL of RNase A (Qiagen) and proteinase inhibitors, shaking-incubated @ 37 °C for 1hr, then boiled with laemmli buffer and loaded on SDS-PAGE. For samples without RNase treatment, 10 µL of resuspended disrupted sample was directly boiled with laemmli buffer and loaded on SDS-PAGE. The SDS-PAGE was then separated via electrophoresis and immunoblotted with 1:10000 anti-his monoclonal antibody (Proteintech #HRP-66005) to detect exogenously expressed Hfq proteins.

Purification and quantification for Hfq-RNA binding

For exogenous Hfq-expressing *E. coli* cell pellets, 100 µL PBS were added to resuspended the cell pellets. Then 200 µL 0.5 mm glass beads were added. The samples were then put into dry ice, and then lyophilized for 3hr. Dried cells were disrupted by vortexing for 5 min, at intervals of 1 min to avoid warming the sample. The disrupted samples were then resuspended in 1.4 ml of 6M GuHCl, 400mM NaCl, 1xPBS, 10mM imidazole, 0.2% TritonX100, 0.5mM DTT, and centrifuged at 10 °C, 15000g for 10min. Two sets of 10 µL of the supernatants were preserved as cell lysate samples for RNA extraction and western blot detection. The supernatant was added with 20 µL of HisPur magnetic beads (thermo fisher) and rotated at room temperature for 1.5 hr. Then the beads were washed twice with 6 M GuHCl, 400 mM NaCl, 1xPBS, 10 mM imidazole, 0.2% TritonX100, 0.5 mM DTT, washed once with 1x 150 mM NaCl, PBS at 4 °C, washed once with 1 × PBS at 4 °C, shaking-incubated with 50 µL TurboDNase (2U) at 37 °C for 15min, washed once with ddH₂O at room temperature. 1/10 of the beads was then aliquoted as purified samples preserved for western blot detection. The rest of beads was then shaking-incubated with 50 µL of 5mg/ml proteinase K, 2M urea at 37 °C for 1h.

Preserved cell lysate samples and beads for western blot detection were boiled with laemmli buffer, separated on SDS-PAGE, and immunoblotted with 1:10000 anti-his monoclonal antibody (Proteintech #HRP-66005) to detect Hfq proteins.

RNA from proteinase-treated beads (purified samples) and preserved cell lysate samples were purified using QuickRNA micro prep kits (Zymo Research). Purified RNA was reverse transcribed to cDNA using SuperScript IV First-Strand Synthesis System (Thermofisher). Enrichments of target RNA was quantified with quantitative PCR (qPCR) using ChamQ Universal SYBR qPCR Master Mix (Vazyme). All qPCR reactions were performed in 10-µl reactions with 3 technical replicates in 96-well format and read out using a LightCycler 96

Instrument (Roche). Enrichment was quantified for samples compared with their matched control samples without Hfq exogenous expression. qPCR primers are listed in Table S1.

GRIP for detecting in vivo RNA cross-linking sites of FSY-incorporated Hfq proteins

GECX-RNA with immunoprecipitation (GRIP) was performed on RNA from purified Hfq-WT and Hfq-25FSY samples (Materials and Methods, section “Purification and quantification for Hfq-RNA binding. In general, RNA from purified Hfq samples was reverse-transcribed with gene-specific RT primers targeting different cross-linking genes and regions (rpoS-XL-RT and ptsG-xL-RT, listed in Table S1) with SuperScript IV First-Strand Synthesis System (ThermoFisher). The cDNA was treated with ExoSAP-IT to remove free primers, and then treated with NaOH to degrade RNA molecules. After clean-up with DynaBeads MyONE Silane (ThermoFisher), a 5' linker (Rand3Tr3 adaptor, Table S1) was ligated to cDNA molecules by T4 RNA ligase in on-beads solution with high concentration of PEG8000 at room temperature for 16hr. The ligated product was cleaned up again with DynaBeads MyONE Silane, and then amplified with primers targeting gene-specific regions and the 5' linker (PCR primer pair for the eCLIP region from rpoS RNA are primers pBADf-rpoS-xL-pF and pBADr-eCLIP-Rand103tr3-pR; PCR primer pair for the eCLIP region from ptsG RNA are primers pBADf-ptsG-xL-pF and pBADr-eCLIP-Rand103tr3-pR). The PCR product was separated on agarose gel. The insertion bands were cut out, purified and cloned into pBAD vector, transformed into DH10B competent cells, and plated onto LB-Amp100 agar plate and incubated overnight at 37 °C. Plasmids were then extracted from colonies and sequenced. The sequenced inserts from plasmids were aligned to target RNA (rpoS RNA or ptsG RNA). The ligation sites of 5' linker represent the cross-linking sites of Hfq-25FSY proteins on target RNA molecules.

Cloning of pcDNA3.1-dPsCas13b plasmids

To generate plasmids suitable for expressing dPsCas13b proteins in mammalian cells, dPsCas13b-WT insert and dPsCas13b-H133TAG insert were amplified from pBAD-dPsCas13b and pBAD-dPsCas13b-H133TAG using primer pair of pcDNA31-Hind3-PsCas13b-F and pcDNA31-BamH1-PsCas13b-R, and cloned into pcDNA3.1 vector pre-digested with BamHI and HindIII using Gibson Assembly kit (New England Biolabs).

Cloning of pC0046-crRNA plasmids

To generate plasmids suitable for expressing crRNA of dPsCas13b in mammalian cells, different crRNA inserts were PCR amplified with different pairs of primers (primers for crACTB are pc43-Ps13b-crACTB1-F and pc43-Ps13b-cr-R; primers for crNEAT1-1 are pc43-Ps13b-NEAT1-1-F and pc43-Ps13b-cr-R; primers for crNEAT1-2 are pc43-Ps13b-crNEAT1-2-F and pc43-Ps13b-cr-R), and cloned into pC0043-PspCas13b crRNA backbone (addgene #103854) vector pre-digested with KpnI and BbsI using ClonExpress II one step cloning kit (Vazyme).

Purification and quantification for dPsCas13b-RNA binding in mammalian cells with RNA immunoprecipitation

For RNA immunoprecipitation experiments, HEK293T cells were plated in six-well plates and transfected with 1 μ g of pcDNA3.1-dPsCas13b plasmids and 1 μ g of pC0043-PspCas13b crRNA plasmids, with an additional 1 μ g of pMP-FSYRS plasmid²²(encoding FSY-tRNA synthetase-tRNA system for expression in mammalian cells) and 1 mM FSY for conditions involving Cas13b-133FSY protein expression. Forty-eight hours after transfection, cells were washed twice with ice-cold PBS, and centrifuged to collect as cell pellets. Cells were lysed with 200 μ L of 1 \times RIPA Buffer (ThermoFisher) supplemented with proteinase inhibitors and RNase inhibitors. Cells were lysed on ice for 10 min and then passed through 26G-needles for 10 times to achieve full lysis. Lysates were then pelleted by centrifugation at 16,000 g for 10 min at 4 $^{\circ}$ C, and the supernatants containing cleared lysates were used for pulldown with magnetic beads.

To conjugate antibodies to magnetic beads, 100 μ L per sample of Dynabeads Protein G for Immunoprecipitation (Thermo Fisher Scientific) were pelleted by application of a magnet, and the supernatant was removed. Beads were resuspended in 200 μ L of wash buffer (PBS, 0.02% Tween 20) and 5 μ g of anti-HA antibody (Thermo Fisher Scientific #26183) was added. The sample was incubated for 10 min at room temperature on a rotator for antibody-beads conjugation. After incubation, beads were pelleted using a magnet, supernatant was removed, and beads were washed twice with wash buffer, and resuspended in 200 μ L 1 \times RIPA with proteinase inhibitors and RNase inhibitor. 200 μ L of sample lysate were added to beads and rotated overnight at 4 $^{\circ}$ C.

After incubation with sample lysate, beads were pelleted, washed four times with 1 \times RIPA, 0.02% Tween 20, and then washed with DNase buffer (350 mM Tris-HCl (pH 6.5); 50 mM MgCl₂; 5 mM DTT). Beads were resuspended in DNase buffer and TURBO DNase was added to a final concentration of 0.1 U/ μ L. DNase was shaking-incubated for 30 min at 37 $^{\circ}$ C. Proteins were then digested by shaking-incubation with 50 μ L of 5mg/ml proteinase K, 2M urea at 37 $^{\circ}$ C for 1hr.

RNA was purified using QuickRNA micro prep kits. Purified RNA was reverse transcribed to cDNA using SuperScript IV First-Strand Synthesis System. Enrichments of target RNA was quantified with qPCR using ChamQ Universal SYBR qPCR Master Mix. All qPCR reactions were performed in 10- μ L reactions with 3 technical replicates in 96-well format and read out using a LightCycler 96 Instrument. Enrichment was quantified for samples compared with their matched control cells without crRNA transfection. qPCR primers used are listed in Table S1.

Cloning of pNEU-MmSFYRS-4xU6M15 plasmid

The MmSFYRS gene was amplified with primers HR-MmPyIRS-NheI-F/HR-MmPyIRS-NotI-R and ligated into pNEU-XYRS-4xU6M15 (derived from pNEU-hMbPyIRS-4xU6M15, a gift from Irene Coin, Addgene plasmid # 105830) which was linearized with *NheI/NotI* to generate pNEU-MmSFYRS-4xU6-M15.

Cloning of pNEU-MaSFYRS-NxU6-MaPylT (N = 1 to 4) plasmids

The MaSFYRS and Ma-PylT expression cassettes were cloned into pNEU-XYRS-4xU6M15. Specifically, the U6 promoter was amplified from pNEU-XYRS-4xU6M15 with primers U6-F1/U6-R1, and the evolved Ma-PylT(6) was amplified from pEvol-MaSFYRS with primers Ma-PylT(6)-F2/Ma-PylT(6)-R2. The resulting fragments were joined together by overlapping PCR with primers U6-F1/Ma-PylT(6)-R2 and then amplified again with primers HR-pNEU-tRNA-XhoI-F/HR-pNEU-tRNA-SalI-R to generate a monomeric U6-MaPylT expression cassette containing *XbaI*-*XhoI* and *SalI* restriction sites. The first monomeric U6-MaPylT expression cassette was ligated into pNEU-XYRS-4xU6M15 vector which was linearized with *XhoI/SalI* to generate pNEU-XYRS-1xU6-MaPylT. Then the MaSFYRS was amplified from pEvol-MaSFYRS with primers HR-Ma-SFYRS-NheI-F/HR-Ma-SFYRS-NotI-R and ligated into pNEU-XYRS-1xU6-MaPylT vector which was linearized with *NheI/NotI* to generate pNEU-MaSFYRS-1xU6-MaPylT. The second U6-MaPylT cassette was digested with *XbaI/SalI* and ligated into pNEU-MaSFYRS-1xU6-MaPylT vector that was linearized with *XbaI/XhoI* to generate pNEU-MaSFYRS-2xU6-MaPylT. Two more U6-MaPylT cassettes were tandemly introduced into the pNEU-MaSFYRS vector following the same procedure to construct the pNEU-MaSFYRS-4xU6-MaPylT.

Cross-linking of MBP-Z24SFY and Afb_{4A}-7X in live *e. coli* cells

The pET-Duet-Afb_{4A}-7X-MBP-Z24TAG (X= A, C, S, T, H, Y, or K)²² was co-transformed with pEvol-MmSFYRS and pEvol-MaSFYRS² respectively into BL21(DE3) *E. coli* chemical competent cells. The transformants were plated on an LB-Amp100Cm34 agar plate and incubated overnight at 37 °C. A single colony was inoculated into 5 mL of 2xYT-Amp100Cm34 and cultured overnight at 37 °C. On the following day, 1 mL of overnight cell culture was diluted into 50 mL 2xYT-Amp100Cm34 and agitated vigorously at 37 °C. When OD₆₀₀ reached 0.4~0.6, the cell culture was induced with 0.5 mM IPTG and 0.2% arabinose in the presence of 1 mM SFY, and then incubated at 37 °C for 6 h. 1 mL of cell pellets were collected by centrifugation at 21000 g for 5 min at 4 °C and directly applied for immunoblot analysis. The rest of cell pellets were collected by centrifugation at 4200g for 30min at 4 °C. The cross-linking products of MBP-Z24SFY and Afb_{4A}-7X (X= H, Y, or K) with affinity chromatography as described previously²².

Cross-linking of GST-103SFY-107X in live mammalian cells

One day before transfection, 3×10⁵ HEK293T cells were seeded in a Greiner 6-well cell culture dish containing 2 mL of DMEM media with 10% FBS, and incubated at 37 °C in a CO₂ incubator. 1 µg of pcDNA-GST-103TAG-107X (X= A, H, Y or K)³ and 1 µg of pNEU-MmSFYRS-4xU6M15 were co-transfected into target cells using 5 µL of lipofectamine 2000 following the manufacturer's instructions. Six hours post transfection, the media were replaced with complete DMEM media with or without 1 mM SFY. The cells were incubated at 37 °C for additional 48 h, collected, and applied for immunoblot analysis.

Fluorescence confocal microscopy

One day before transfection, 3×10^5 HEK293T cells were seeded in a Greiner 6-well cell culture dish containing 2 mL of DMEM media with 10% FBS, and incubated at 37 °C in a CO₂ incubator. Plasmids pcDNA-EGFP-40TAG (1 µg) and pNEU-MmSFYRS-4xU6M15 (1 µg) were co-transfected into target cells using 5 µL of lipofectamine 2000 following the manufacturer's instructions. Six hours post transfection, the media were replaced with complete DMEM media with or without 1 mM SFY. The cells were incubated at 37 °C for additional 24–48 h and imaged with Nikon Eclipse Ti confocal microscope.

FACS analysis of SFY incorporation

One day before transfection, 3×10^5 HEK293T cells were seeded in a Greiner 6 well-cell culture dish containing 2 mL of DMEM media with 10% FBS, and incubated at 37 °C in a CO₂ incubator. Plasmids pcDNA-EGFP-40TAG (1 µg) and pNEU-MaSFYRS-NxU6-MaPyIT (N=1 to 4) (1 µg) were co-transfected into target cells using 5 µL of lipofectamine 2000 following the manufacturer's instructions. Six hours post transfection, the media containing transfection complex were replaced with fresh DMEM media with 10% FBS in the presence or absence of 1 mM SFY. After incubation at 37 °C for 24–48 h, transfected cells were trypsinized and collected by centrifugation (1500 rpm, 5 min, r.t.). The cells were resuspended in 500 µL of FACS buffer (1×PBS, 2% FBS, 1 mM EDTA, 0.1% sodium azide, 0.28 µM DAPI) and analyzed by BD LSRFortessa™ cell analyzer. The FlowJo (Tree Star Inc., Ashland, OR, USA) was used for the data analysis.

Cell viability assay

2×10^4 cells/well of HEK293T cells were seeded in a 96-well plate. On the next day, the media were replaced with fresh DMEM media supplemented with 0, 0.0625, 0.125, 0.25, 0.5, or 1 mM of SFY. The SFY-treated and control cells were cultured for an additional 24–48 h at 37 °C and then analyzed with CellTiter-Blue® Cell Viability Assay following the manufacturer's instructions.

RNase treatment and detection for exogenous Hfq-expressing *E. coli* cells (Hfq-SFY samples)

The procedure is the same as the RNase treatment and detection for exogenous Hfq-expressing *E. coli* cells (Hfq-WT and Hfq-FSY samples), with the following modifications:

For the transformations, pBAD-Hfq TAG mutant plasmids (pBAD-Hfq-25TAG, pBAD-Hfq-49TAG) was co-transformed with pEvol-MmSFYRS into DH10B *E. coli* chemical competent cells, respectively.

For the exogenous expression of Hfq-SFY proteins, the cell culture was induced with 0.2% arabinose and 1 mM SFY.

In vitro incubations of NMPs and SFY

50 mM SFY (HCl salt) and 50 mM NMP was incubated in DI H₂O. 50 mM NaOH was added to neutralize the HCl salt. The mixture was incubated at 37 °C for 48 h. Then the reaction mixture was diluted for 50 times in H₂O/acetonitrile (50/50, v/v, with 0.1 %

trifluoroacetic acid) and subjected to mass spectrum analysis using positive mode. Mass spectrum analysis was performed on SCIEX MDS, 3200 Q TRAP system.

The molecular weight (MW) of adduct products between SFY and NMP was calculated following this equation: MW (adduct product) = MW (SFY) + MW (NMP) – MW (HF). The calculated MWs are listed in Table S4.

Cloning of YTH domain from human YTHDF1 protein

To generate plasmids expressing YTH domain from human YTHDF1 protein with TwinStrep tag and HA tag at C-terminal in mammalian cells, three PCR products were prepared. Insert with YTHDF1 domain was amplified with primer pair of pc31-Hd3-YTHDF1-F and YTHDF1-2xstrep-R using cDNA reverse-transcribed from total RNA of HEK293T cells as template. Insert with TwinStrep tag was amplified with primer pair of 2xstrep-tag_Hs-F and 2xstrep-tag_Hs-R. pcDNA3.1 vector backbone was amplified with primer pair of pc31-HA-strep-F and pc31-Nde1-R using empty pcDNA3.1 vector as template. The final plasmid pcDNA3.1-HsYTHDF1-WT expressing wildtype YTHDF1 domain with TwinStrep tag and HA tag at C-terminal was cloned by ligating these three PCR products together using ClonExpress II one step cloning kit (Vazyme).

To generate pcDNA3.1-HsYTHDF1-397TAG mutant plasmid, residue 397 of YTHDF1 gene in pcDNA3.1-HsYTHDF1-WT were mutated into an amber stop codon TAG using site-directed mutagenesis with following primers: YTHDF1-Y397TAG-F and YTHDF1-Y397TAG-R.

Library preparation for GRIP-seq

HEK293T cells were plated in 15-cm plates and transfected with 15 µg of pcDNA3.1-HsYTHDF1 plasmids, with an additional 15 µg of pNEU-SFYRS plasmid (encoding SFY-synthetase-tRNA system for expression in mammalian cells) and 1 mM SFY for conditions involving YTHDF1-397SFY protein expression. Forty-eight hours after transfection, cells were washed twice with ice-cold PBS, and centrifuged to collect as cell pellets. The library preparation procedure for GRIP-seq was similar to the protocol from eCLIP⁴⁵. In brief, the cell pellets were lysed in 1 mL of eCLIP lysis buffer⁴⁵, partially digested with RNase I (Invitrogen). 20 µL of the cell lysate was stored as “INPUT” sample for subsequent direct library preparation (similar as in eCLIP protocol⁴⁵). The remainder of the cell lysate (~1 mL) was immunoprecipitated using 200 µL of pre-washed strep-tactin-XT magnetic beads (Iba-lifesciences) targeting 2xStrep-tag sequence fused at C-terminal of YTH proteins, and stringently washed (twice with high-salt denaturing buffer (PBS buffer with 6 M Urea, 1 M NaCl, 1 mM DTT) and twice with PBS buffer). After dephosphorylation with FastAP (ThermoFisher) and T4 PNK (NEB), a barcoded RNA adaptor (1:1 mixed RNA_X1A and RNA_X1B adaptors, Table S1) was ligated to the 3' end (T4 RNA Ligase, NEB) of cross-linked and co-purified RNA. Ligations were performed on-bead. Next, Samples were run on protein gels and transferred to nitrocellulose membranes. On the membranes, the regions containing YTH protein-RNA cross-links were excised (membrane regions 75 kDa above the YTH protein) and treated with proteinase K to release the cross-linked RNA. RNA was then reverse-transcribed with SuperScript IV reverse transcriptase (ThermoFisher) and

AR17 primer (Table S1), and treated with ExoSAP-IT (ThermoFisher) to remove excess oligonucleotides. A second DNA adaptor (Rand3Tr3 adaptor, Table S1) was then ligated to the 3' end of the cDNA fragment (T4 RNA Ligase, NEB). After cleanup (Dynabeads MyOne Silane, ThermoFisher), an aliquot of each sample was first subjected to qPCR for determining the proper number of PCR cycles. Then, the remainder was amplified (Phanta Max Super-Fidelity DNA Polymerase, Vazyme) with a pair of PCR primer for final library amplification (P1A-0N-F and P1A-0N-R, “N” represents the specific index for different sample, Table S1) and size selected via agarose gel electrophoresis. Samples were sequenced on the Illumina NovaSeq S4 platform with paired-end 2×100 format.

Data analysis for GRIP-seq

Read processing—After standard illumina Hiseq demultiplexing, GRIP-seq libraries were first processed with Fastp tool⁶² to remove PCR duplications and cut illumina adaptors, and then processed with Fastp tool⁶² to remove the GRIP-seq adaptors and retrieve the inserted RNA sequences according to the following GRIP-seq final library structure.

Library structure with X1A_adaptor: (Read1) NNNNNCCTATAT-INSERT-
NNNNNNNNNN (Read2)

Library structure with X1B_adaptor: (Read1) NNNNNTGCTATT-INSERT-
NNNNNNNNNN (Read2)

Note: “N” in library structures representing random nucleotides.

Read mapping

Reads were mapped with STAR⁶³ to the human genome (hg19) by default setting.

Identification of m6A clusters and reverse-transcription-termination sites

After mapping, CLIPper⁴⁶ was applied on the mapped reads 2 (reads 2 is the read starting right after the cross-linking site (Fig. 5a)) with options “--FDR 0.01 --poisson-cutoff 1e-10 --minreads 5 --binomial 0.01” to identify the read clusters. After cluster identification, the precise reverse-transcription-termination sites of clusters were identified by scanning all sites within the cluster using the following criteria: If the read number covering the scanned site was > 1.5-fold bigger and had > 40 reads than the read number covering the neighboring site, this site will be designated as reverse-transcription-termination site.

Metagene and motif analyses

After reverse-transcription-termination site was identified, the sequences spanning a region 10-nt up- and downstream of termination sites were extracted and used as input for motif discovery using MEME⁶⁴. Metagene analysis was performed with reads mapped within m6A clusters using metaPlotR⁶⁵.

Analysis of cross-linking site positions relative to the m6A motif

Reads 2 overlapping with regions containing motif DRACH from motif analysis were extracted. The numbers of reads 2 starting right after each position relative to DRACH motif (the middle A in the motif was designated as position 0) were calculated and plotted.

Analysis of nucleotide composition at cross-linking sites

For reads in m6A clusters, the cross-linking sites were designated as the nucleotides 1-nt upstream of read 2 starting positions.

Identification of m6A sites

After the position of cross-linking site relative to m6A motif was revealed, the precise m6A sites were assigned according to the distance to the revers-transcription-termination sites.

Secondary structure analysis around m6A sites

The coordinates of published m6A sites were from m6A-atlas database⁴⁷. The coordinates of m6A sites from DART-seq were m6A sites from “HEK293T, DART-seq, control sample” in m6A-atlas database⁴⁷. For each m6A site, a sliding window of 30 nucleotides with a step of 3 nucleotides was used to calculate RNA minimum fold free energy (MFE) spanning the regions 120-nt up- and downstream of m6A sites. For each window, MFE was calculated by ViennaRNA⁶⁶, using default parameters. For m6A sites from different datasets, a mean MFE in each window was calculated by averaging MFE values of the windows in the same position.

Analysis of GRIP identified novel m6A sites

To calculate how many m6A sites were novel, sites identified by GRIP were compared with known Human m6A sites from m6A-Atlas database⁴⁷. Those sites which have exactly same chromosome coordinates in the database were considered as known ones, while others were considered novel.

GRIP for individual in vivo m6A detection

HEK293T cells were plated in 15-cm plates and transfected with 15 µg of pcDNA3.1-HsYTHDF1 plasmids, with an additional 15 µg of pNEU-SFYRS plasmid (encoding SFY-tRNA synthetase-tRNA system for expression in mammalian cells) and 1 mM SFY for conditions involving YTHDF1–397SFY protein expression. Forty-eight hours after transfection, cells were washed twice with ice-cold PBS, and centrifuged to collect as cell pellets. Cells were lysed with 1.5 mL of 1× RIPA Buffer supplemented with proteinase inhibitors and RNase inhibitor. Cells were lysed on ice for 10 min and then passed through 26G-needles for 20 times to achieve full lysis. Lysates were then pelleted by centrifugation at 16,000 g for 10 min at 4 °C, and the supernatants containing cleared lysates were used for pulldown with magnetic beads.

For strep-tactin-XT magnetic beads (Iba-lifesciences), 200 µL per sample of beads were pelleted by application of a magnet, and the supernatant was removed. Beads were washed twice with wash buffer (PBS buffer with 6 M Urea, 1 M NaCl, 1 mM DTT), and

resuspended in 11.25 mL of wash buffer (PBS buffer with 6M Urea, 1 M NaCl, 1 mM DTT). 750 μ L of sample lysate were added to beads and rotated overnight at 4 °C.

After incubation with sample lysate, beads were pelleted, washed three times with 6M Urea, 1 M NaCl, PBS buffer, 1 mM DTT, wash once with PBS buffer with 1 M NaCl, wash once with PBS buffer, and then washed with Dnase buffer (350 mM Tris-HCl (pH 6.5); 50 mM MgCl₂; 5 mM DTT). Beads were resuspended in Dnase buffer and TURBO Dnase was added to a final concentration of 0.1 U/ μ L. Dnase was shaking-incubated for 30 min at 37 °C. Proteins were then digested by shaking-incubation with 50 μ L of 5 mg/mL proteinase K, 2 M urea at 37 °C for 1 h. RNA was purified using QuickRNA micro prep kits.

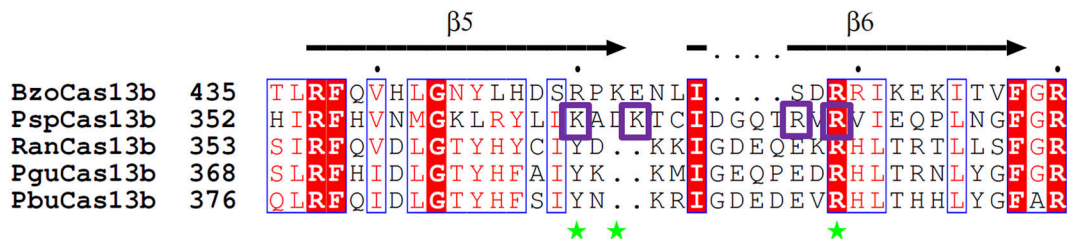
RNA samples were reverse-transcribed with gene-specific RT primers targeting different cross-linking genes and regions (DICER1-m6A-1-RT, and JUN-m6A-1-RT, as listed in Table S1) with SuperScript IV First-Strand Synthesis System. The cDNA was treated with ExoSAP-IT to remove free primers, and then treated with NaOH to degrade RNA molecules. After clean-up with DynaBeads MyONE Silane, a 5' linker (Rand3Tr3 adaptor, Table S1) was ligated to cDNA molecules by T4 RNA ligase in on-beads solution with high concentration of PEG8000 at room temperature for 16 h. The ligated product was cleaned up again with DynaBeads MyONE Silane, and then amplified with primers targeting gene-specific regions and the 5' linker (PCR primer pair for the GRIP region of DICER1 RNA are primers pBADf-DICER1-m6A-1-pF and pBADr-eCLIP-Rand103tr3-pR; and PCR primer pair for the GRIP region of JUN RNA are primers pBADf-JUN-m6A-1-pF and pBADr-eCLIP-Rand103tr3-pR). The PCR product was separated on agarose gel. The insertion bands were cut out, purified and cloned into pBAD vector, transformed into DH10B competent cells, and plated onto LB-Amp100 agar plate and incubated overnight at 37 °C. Plasmids were then extracted from colonies and sequenced. The sequenced inserts from plasmids were aligned to target RNA regions (DICER1, or JUN), the ligation sites of 5' linker represent the cross-linking sites of YTHDF1–397SFY proteins on target RNA molecules, thus also representing m6A sites on target RNA molecules.

Statistics and Reproducibility

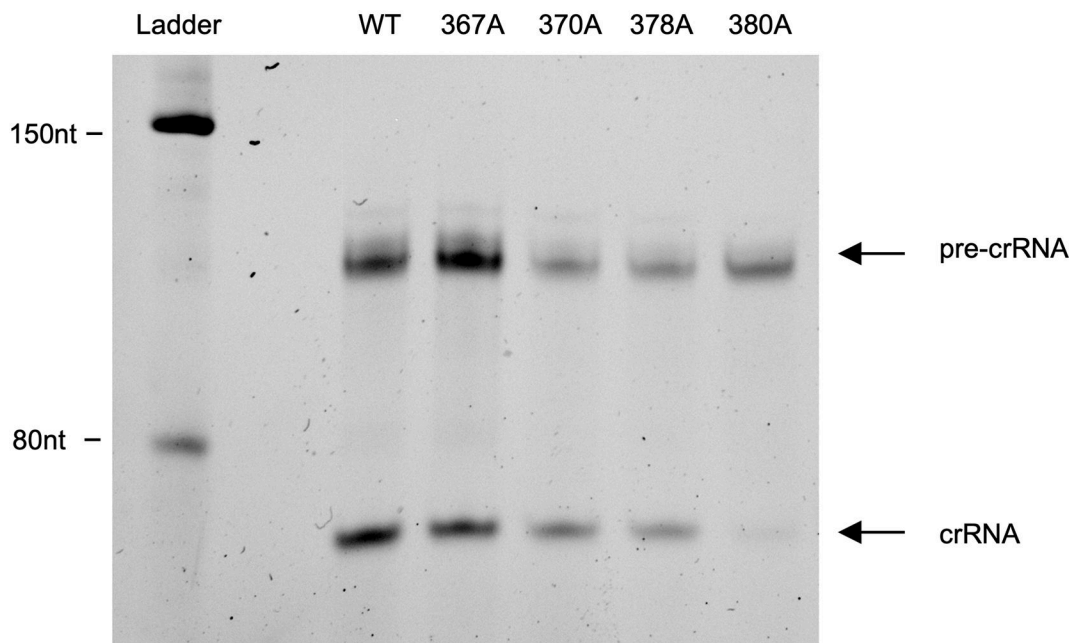
The experiments in the following figures were repeated for at least three times independently: Fig 1c, 1d, 1g, 2b, 4b, 4e, 4g–h, Extended Data Fig 1b, Fig S2a, S2c, S3, S7a, and S7f.

Extended Data

a



b

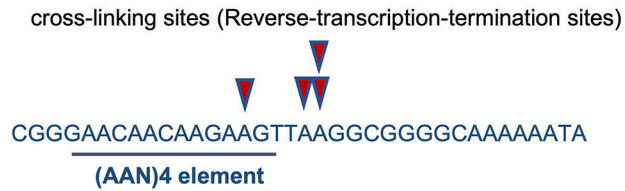


Extended Data Fig. 1. Identification of the positively charged residues of PsCas13b involved in pre-crRNA cleavage.

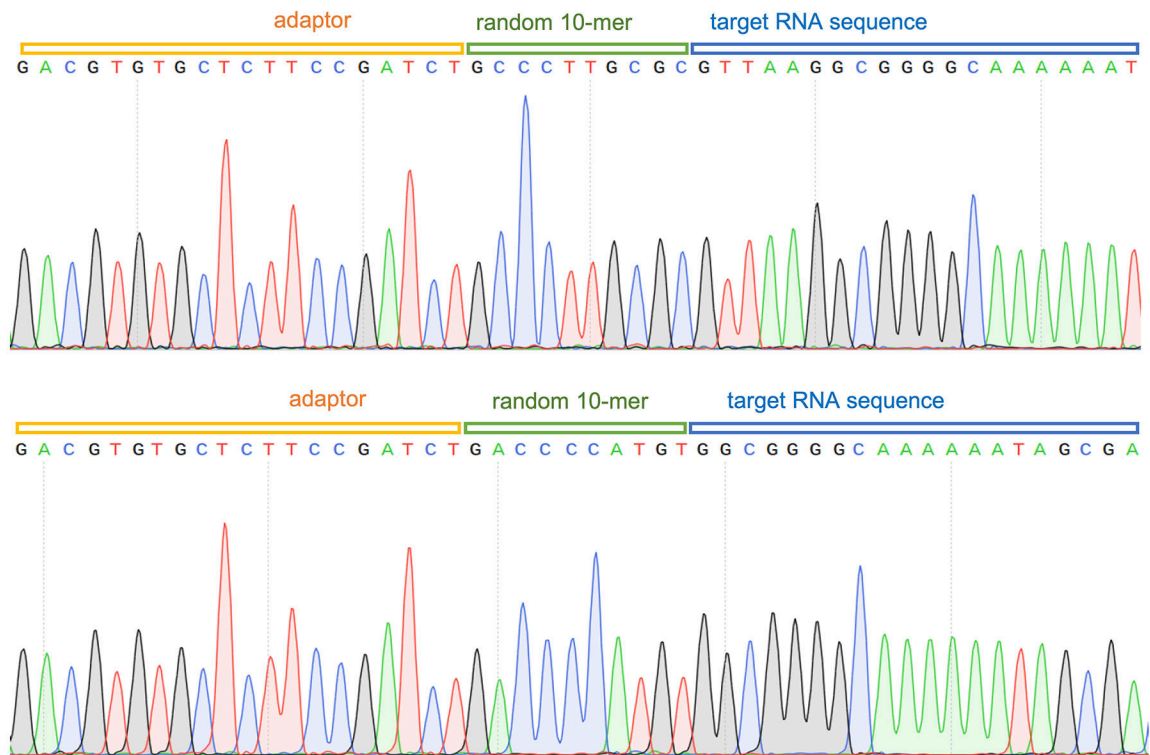
a) Multiple sequence alignment of Cas13b proteins from different species (Bzo: *Bergeyella zoohelcum*, Psp: *Prevotella sp.* P5–125, Pgu: *Porphyromonas gingivalis*, Pbu: *Prevotella buccae*, and Ran: *Riemerella anatipestifer*) for β -sheets 5 and 6 involved in pre-crRNA cleavage. The secondary structure of BzoCas13b is shown above the sequence²⁸. Identical and similar residues are highlighted in red and white boxes, respectively. Positive charged catalytic residues in BzoCas13b involved in the pre-crRNA cleavage on β -sheets 5 and 6 (450R, 452K, 459R) are marked with green stars on the bottom. Positive charged residues in PspCas13b located on β -sheets 5 and 6 (367K, 370K, 378R, 380R) are marked with purple squares. Multiple sequence alignment of full-length Cas13b proteins from different species is shown in Fig. S1. **b)** Denaturing urea-PAGE demonstrating the pre-crRNA cleavage by dPsCas13b-WT and dPsCas13b-Ala-mutants speculatively involved in the pre-crRNA processing. dPsCas13b-WT and dPsCas13b-Ala-mutants were incubated with pre-crRNA

and then separated on denaturing urea-PAGE. The Urea-gel was stained with SybrGold for fluorescent detection of RNA.

rpoS RNA 5' UTR region

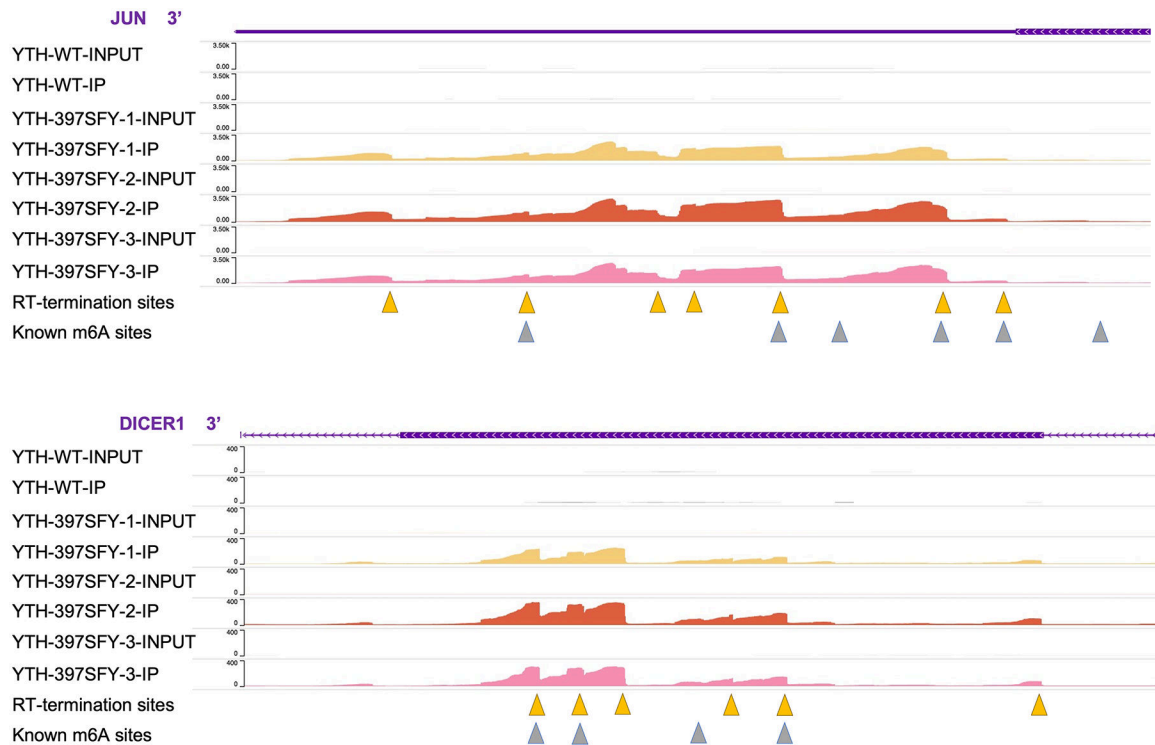


Sequenced clone examples



Extended Data Fig. 2. GRIP results demonstrate that site 25 of Hfq directly binds with (AAN)4 elements of rpoS RNA.

GRIP identified the binding sites of Tyr25 of Hfq on rpoS RNA in *E. coli* cells. Red triangles indicate cross-linking sites identified from GRIP for rpoS RNA from Hfq-25FSY expressing *E. coli* cells. Two examples of Sanger sequencing of clones from Hfq-25FSY sample were shown below.



Extended Data Fig. 3. Examples of GRIP-seq data for m6A identification.

Genome browser tracks of GRIP-seq data in *JUN* and *DICER1* mRNA regions. Reverse-transcription-termination sites (RT-termination sites) from GRIP-seq were marked as yellow triangles. Known m6A sites from published datasets were marked as grey triangles.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

L.W. acknowledges the support of the NIH (R01GM118384 and R01CA258300). Y.S. acknowledges the support of the NIH (R01AG057497 and R01EY027789).

DATA AVAILABILITY

All GRIP-seq data are available in SRA database with Accession number: PRJNA797913.

All other data generated or analyzed in this study are available within the Article and its Supplementary Information and Source Data. Source data are provided with this paper.

REFERENCES

1. Gerstberger S, Hafner M & Tuschl T A census of human RNA-binding proteins. *Nat. Rev. Genet* 15, 829–845 (2014). [PubMed: 25365966]
2. Castello A, Fischer B, Hentze MW & Preiss T RNA-binding proteins in Mendelian disease. *Trends Genet.* 29, 318–327 (2013). [PubMed: 23415593]

3. Nussbacher JK, Batra R, Lagier-Tourenne C & Yeo GW RNA-binding proteins in neurodegeneration: Seq and you shall receive. *Trends Neurosci.* 38, 226–236 (2015). [PubMed: 25765321]
4. Castello A et al. Comprehensive Identification of RNA-Binding Domains in Human Cells. *Mol. Cell* 60, 696–710 (2016).
5. Benhalevy D, Anastasakis DG & Hafner M Proximity-CLIP provides a snapshot of protein-occupied RNA elements in subcellular compartments. *Nat. Methods* 15, 1074–1082 (2018). [PubMed: 30478324]
6. Hentze MW, Castello A, Schwarzl T & Preiss T A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol* 19, 327–341 (2018). [PubMed: 29339797]
7. Müller-Mcnicoll M & Neugebauer KM How cells get the message: Dynamic assembly and function of mRNA-protein complexes. *Nat. Rev. Genet* 14, 275–287 (2013). [PubMed: 23478349]
8. Wagenmakers AJM, Reinders RJ & van Venrooij WJ Cross-linking of mRNA to Proteins by Irradiation of Intact Cells with Ultraviolet Light. *Eur. J. Biochem* 112, 323–330 (1980). [PubMed: 7460925]
9. Saito I & Matsuura T Chemical Aspects of UV-Induced Cross-Linking of Proteins to Nucleic Acids. Photoreactions with Lysine and Tryptophan. *Acc. Chem. Res* 18, 134–141 (1985).
10. Hafner M et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 141, 129–141 (2010). [PubMed: 20371350]
11. Baltz AG et al. The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. *Mol. Cell* 46, 674–690 (2012). [PubMed: 22681889]
12. Licatalosi DD et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* 456, 464–469 (2008). [PubMed: 18978773]
13. König J et al. ICLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol* 17, 909–915 (2010). [PubMed: 20601959]
14. Castello A et al. Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell* 149, 1393–1406 (2012). [PubMed: 22658674]
15. Lee FCY & Ule J Advances in CLIP Technologies for Studies of Protein-RNA Interactions. *Mol. Cell* 69, 354–369 (2018). [PubMed: 29395060]
16. Sugimoto Y et al. Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome Biol.* 13, R67 (2012). [PubMed: 22863408]
17. Xiang Z et al. Adding an unnatural covalent bond to proteins through proximity-enhanced bioreactivity. *Nat. Methods* 10, 885–888 (2013). [PubMed: 23913257]
18. Wang L Genetically encoding new bioreactivity. *N. Biotechnol* 38, 16–25 (2017). [PubMed: 27721014]
19. Coin I et al. Genetically encoded chemical probes in cells reveal the binding path of urocortin-i to CRF class B GPCR. *Cell* 155, 1258–1269 (2013). [PubMed: 24290358]
20. Yang B et al. Spontaneous and specific chemical cross-linking in live cells to capture and identify protein interactions. *Nat. Commun* 8, 2240 (2017). [PubMed: 29269770]
21. Li Q et al. Developing Covalent Protein Drugs via Proximity-Enabled Reactive Therapeutics. *Cell* 182, 85–97 (2020). [PubMed: 32579975]
22. Wang N et al. Genetically encoding fluorosulfate-1-tyrosine to react with lysine, histidine, and tyrosine via SuFEx in proteins in vivo. *J. Am. Chem. Soc* 140, 4995–4999 (2018). [PubMed: 29601199]
23. Abudayyeh OO et al. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* 353, aaf5573 (2016). [PubMed: 27256883]
24. Cox DBT et al. RNA editing with CRISPR-Cas13. *Science* 358, 1019–1027 (2017). [PubMed: 29070703]
25. Yang LZ et al. Dynamic Imaging of RNA in Living Cells by CRISPR-Cas13 Systems. *Mol. Cell* 76, 981–997 (2019). [PubMed: 31757757]
26. Liu L et al. Two Distant Catalytic Sites Are Responsible for C2c2 RNase Activities. *Cell* 168, 121–134 (2017). [PubMed: 28086085]

27. Smargon AA et al. Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28. *Mol. Cell* 65, 618–630 (2017). [PubMed: 28065598]
28. Zhang B et al. Structural insights into Cas13b-guided CRISPR RNA maturation and recognition. *Cell Res.* 28, 1198–1201 (2018). [PubMed: 30425321]
29. Wilusz CJ & Wilusz J Eukaryotic Lsm proteins: Lessons from bacteria. *Nat. Struct. Mol. Biol* 12, 1031–1306 (2005). [PubMed: 16327775]
30. Bilusic I, Popitsch N, Rescheneder P, Schroeder R & Lybecker M Revisiting the coding potential of the *E. coli* genome through Hfq co-immunoprecipitation. *RNA Biol.* 11, 641–654 (2014). [PubMed: 24922322]
31. Holmqvist E et al. Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking in vivo. *EMBO J.* 35, 991–1011 (2016). [PubMed: 27044921]
32. Chao Y, Papenfort K, Reinhardt R, Sharma CM & Vogel J An atlas of Hfq-bound transcripts reveals 3' UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J.* 31, 4005–4019 (2012). [PubMed: 22922465]
33. Wang W, Wang L, Wu J, Gong Q & Shi Y Hfq-bridged ternary complex is important for translation activation of *rpoS* by DsrA. *Nucleic Acids Res.* 41, 5938–5948 (2013). [PubMed: 23605038]
34. Peng Y, Curtis JE, Fang X & Woodson SA Structural model of an mRNA in complex with the bacterial chaperone Hfq. *Proc. Natl. Acad. Sci. U. S. A* 111, 17134–17139 (2014). [PubMed: 25404287]
35. Tree JJ, Granneman S, McAteer SP, Tollervey D & Gally DL Identification of Bacteriophage-Encoded Anti-sRNAs in Pathogenic *Escherichia coli*. *Mol. Cell* 55, 199–213 (2014). [PubMed: 24910100]
36. Schu DJ, Zhang A, Gottesman S & Storz G Alternative Hfq- sRNA interaction modes dictate alternative mRNA recognition. *EMBO J.* 34, 2557–2573 (2015). [PubMed: 26373314]
37. Hoppmann C & Wang L Proximity-enabled bioreactivity to generate covalent peptide inhibitors of p53-Mdm4. *Chem. Commun* 52, 5140–5143 (2016).
38. Liu J et al. Genetically Encoding Photocaged Quinone Methide to Multitarget Protein Residues Covalently in Vivo. *J. Am. Chem. Soc* 141, 9458–9462 (2019). [PubMed: 31184146]
39. Nachtergaele S & He C Chemical modifications in the life of an mRNA transcript. *Annu. Rev. Genet* 52, 349–372 (2018). [PubMed: 30230927]
40. Meyer KD et al. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* 149, 1635–1646 (2012). [PubMed: 22608085]
41. Dominissini D et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* 485, 201–206 (2012). [PubMed: 22575960]
42. Linder B et al. Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat. Methods* 12, 767–772 (2015). [PubMed: 26121403]
43. Xu C et al. Structural basis for the discriminative recognition of N6-Methyladenosine RNA by the human YT521-B homology domain family of proteins. *J. Biol. Chem* 290, 24902–24913 (2015). [PubMed: 26318451]
44. Meyer KD DART-seq: an antibody-free method for global m6A detection. *Nat. Methods* 16, 1275–1280 (2019). [PubMed: 31548708]
45. Van Nostrand EL et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* 13, 508–514 (2016). [PubMed: 27018577]
46. Lovci MT et al. Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nat. Struct. Mol. Biol* 20, 1434–1442 (2013). [PubMed: 24213538]
47. Tang Y et al. m6A-Atlas: a comprehensive knowledgebase for unraveling the N6-methyladenosine (m6A) epitranscriptome. *Nucleic Acids Res.* 49, D134–D143 (2020).
48. Sanchez de Groot N et al. RNA structure drives interaction with proteins. *Nat. Commun* 10, 3246 (2019). [PubMed: 31324771]
49. Siegfried NA, Busan S, Rice GM, Nelson JAE & Weeks KM RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods* 11, 959–965 (2014). [PubMed: 25028896]

50. Gruber AR, Lorenz R, Bernhart SH, Neuböck R & Hofacker IL The Vienna RNA websuite. *Nucleic Acids Res.* 36, W70–W74 (2008). [PubMed: 18424795]
51. Hwang H-W et al. PAPERCLIP Identifies MicroRNA Targets and a Role of CstF64/64tau in Promoting Non-canonical poly(A) Site Usage. *Cell Rep.* 15, 423–435 (2016). [PubMed: 27050522]
52. Kini HK, Silverman IM, Ji X, Gregory BD & Liebhaber SA Cytoplasmic poly(A) binding protein-1 binds to genomically encoded sequences within mammalian mRNAs. *RNA* 22, 61–74 (2016). [PubMed: 26554031]
53. Wang L Engineering the Genetic Code in Cells and Animals: Biological Considerations and Impacts. *Acc. Chem. Res* 50, 2767–2776 (2017). [PubMed: 28984438]
54. Mackereth CD & Sattler M Dynamics in multi-domain protein recognition of RNA. *Curr. Opin. Struct. Biol* 22, 287–296 (2012). [PubMed: 22516180]
55. Lunde BM, Moore C & Varani G RNA-binding proteins: Modular design for efficient function. *Nat. Rev. Mol. Cell Biol* 8, 479–490 (2007). [PubMed: 17473849]
56. Jumper J et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589 (2021). [PubMed: 34265844]
57. Baek M et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373, 871–876 (2021). [PubMed: 34282049]
58. McMahon AC et al. TRIBE: Hijacking an RNA-Editing Enzyme to Identify Cell-Specific Targets of RNA-Binding Proteins. *Cell* 165, 742–753 (2016). [PubMed: 27040499]
59. Brannan KW et al. Robust single-cell discovery of RNA targets of RNA-binding proteins and ribosomes. *Nat. Methods* 18, 507–519 (2021). [PubMed: 33963355]
60. Cao L & Wang L New covalent bonding ability for proteins. *Protein Sci.* doi: 10.1002/pro.4228. (2021).
61. Uhlén M et al. Tissue-based map of the human proteome. *Science* 347, 1260419 (2015). [PubMed: 25613900]
62. Chen S, Zhou Y, Chen Y & Gu J Fastp: An ultra-fast all-in-one FASTQ preprocessor. in *Bioinformatics* vol. 34 i884–i890 (2018). [PubMed: 30423086]
63. Dobin A et al. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21 (2013). [PubMed: 23104886]
64. Bailey TL et al. MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208 (2009). [PubMed: 19458158]
65. Olarerin-George AO & Jaffrey SR MetaPlotR: A Perl/R pipeline for plotting metagenes of nucleotide modifications and other transcriptomic sites. *Bioinformatics* 33, 1563–1564 (2017). [PubMed: 28158328]
66. Lorenz R et al. ViennaRNA Package 2.0. *Algorithms Mol. Biol* 6, 26 (2011). [PubMed: 22115189]

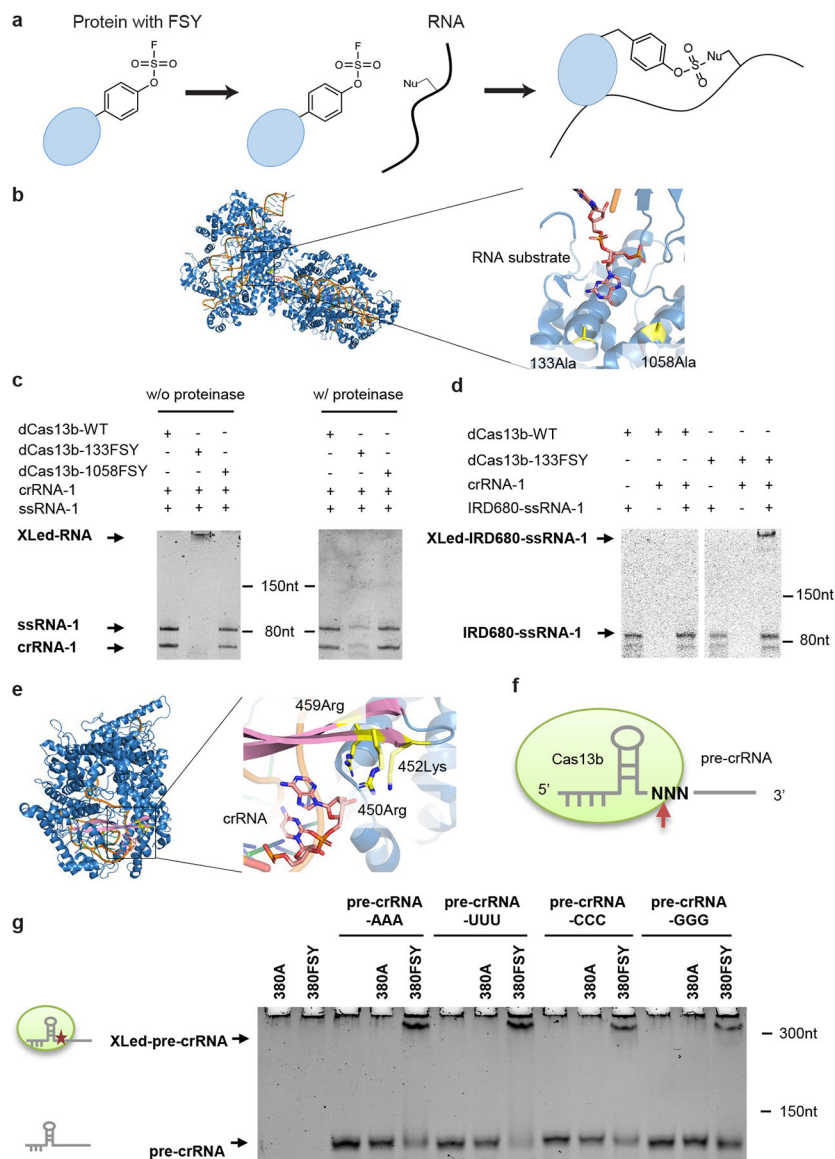


Fig. 1. GECX-RNA enables FSY-incorporated dPsCas13b to cross-link target RNA *in vitro*. **a)** Scheme showing proximity-enabled SuFEx reaction between FSY and a nucleophilic group of RNA, which can be 2'-OH on ribose or amino group on base. The binding of protein with RNA places FSY in close proximity to the target nucleophile in the RNA, driving the formation of a specific covalent linkage. **b)** Structure of Cas13-crRNA-target RNA ternary complex showing sites 133 and 1058 (yellow stick) chosen for FSY incorporations in dPsCas13b protein (PDB: 5XWP). **c)** EMSA on denaturing Urea-PAGE gel demonstrating dPsCas13b-133FSY cross-linked with the target RNA (ssRNA-1) with guidance of crRNA (crRNA-1). After incubation, samples were treated with or without proteinase K followed with separation on denaturing Urea-PAGE. The Urea-gels were stained with SybrGold for fluorescent detection of RNA. **d)** EMSA on denaturing Urea-PAGE gel demonstrating cross-linking of target RNA (IRD680-ssRNA-1) required guidance of crRNA. dPsCas13b-WT or 133FSY protein was incubated with different combinations of

crRNA-1 and target RNA fluorescently labeled with IRD680 at 5' end (IRD680-ssRNA-1). After incubation, samples were separated on denaturing Urea-PAGE. The gel was imaged by scanning IRD680 signal. **e)** Structure of BzoCas13b-crRNA binary complex showing positively charged amino acids (yellow stick) located on β -sheets 5 and 6 (magenta colored) involved in pre-crRNA cleavage. crRNA was shown as salmon-color stick (PDB: 6AAY). **f)** Scheme of Cas13b processing pre-crRNA at the phosphodiester bond connecting two nucleotides located directly 3'-downstream of the hairpin repeat region. Red arrow indicates the cleavage site. **g)** EMSA on denaturing urea-PAGE demonstrating FSY cross-linking of all four RNA nucleotides. dPsCas13b-380A or dPsCas13b-380FSY was incubated with pre-crRNAs containing different nucleotide compositions at cleavage site. Nucleotide sequences at cleavage sites (as NNN shown in **f)**) were placed as AAA, UUU, CCC, or GGG in the indicated pre-crRNAs. After incubation, samples were separated on denaturing Urea-PAGE. The Urea-gels were stained with SybrGold for fluorescent detection of RNA.

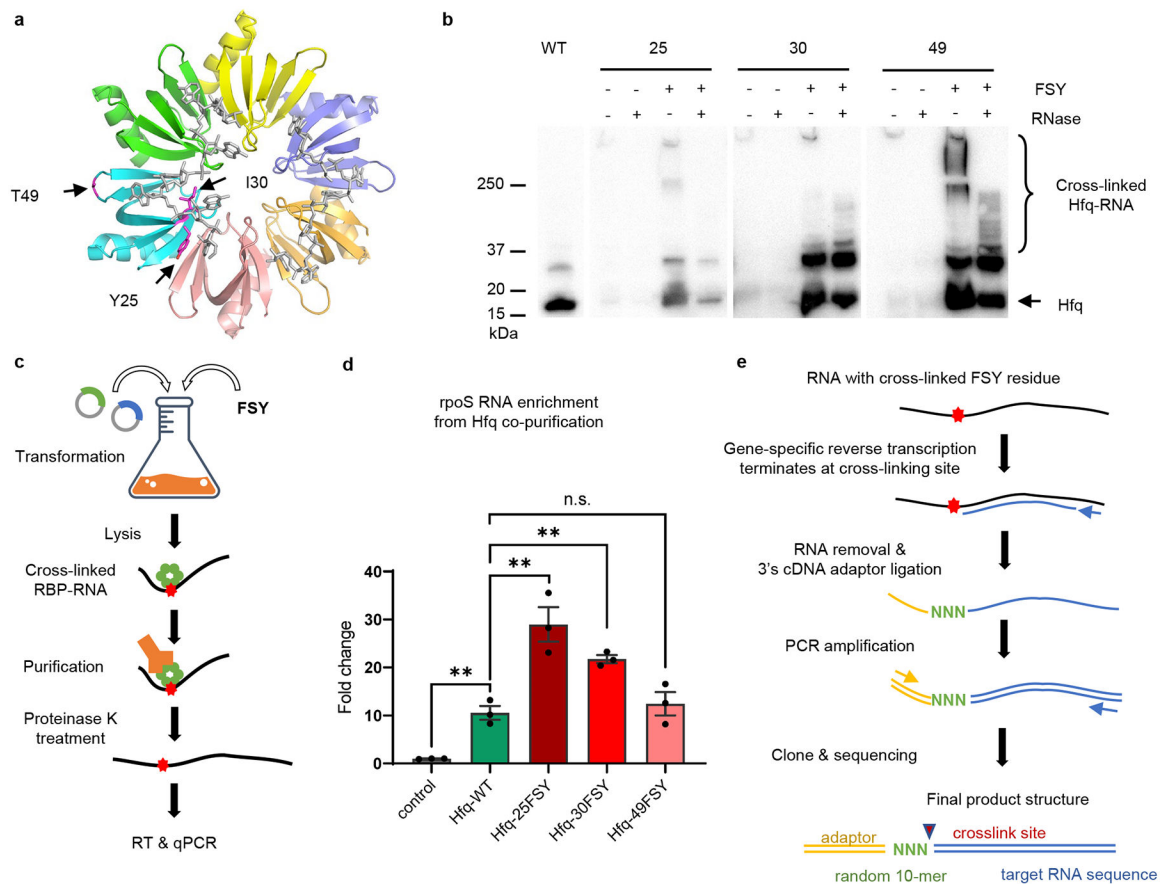


Fig. 2. GECX-RNA enables FSY-incorporated Hfq proteins to cross-link target RNA in *E. coli*.

a) Structure of *E. coli* Hfq bound to target RNA. Hfq is a ring-shaped homo-hexamers, with each monomer colored differently. RNA is in grey. The three chosen sites (Y25, I30, and T49) for FSY incorporation were shown in pink stick on one monomer (PDB: 4HT8).

b) Western blot analysis demonstrating FSY-incorporated Hfq proteins cross-linked with RNA molecules in *E. coli* cells. Hfq-FSY proteins were expressed in *E. coli* DH10B strain. Cell lysate samples were treated with or without RNase before loading, and an anti-His antibody was used to detect the 6xHis tag appended at the C-terminus of expressed Hfq.

c) Scheme of reverse transcription (RT) and quantitative-PCR (qPCR) of RNA cross-linked by Hfq. **d)** RT-qPCR analyses of Hfq co-purified RNA demonstrate that FSY-incorporated Hfq proteins cross-linked and enriched target RNA *rpoS* in *E. coli* cells. Hfq-WT and Hfq-FSY proteins were separately purified from *E. coli* cells, and RT-qPCR analysis was performed on co-purified RNA samples. Enrichment fold changes were calculated based on normalizations to input-RNA samples using *mpb* gene as reference. Control sample was cells without exogenous Hfq expression. Fold-changes of target RNAs in Hfq-FSY samples compared to Hfq-WT samples were shown. Bar heights represent mean; Error bars represent s.e.m.; n = 3 independent biological replicates; ** p < 0.01; n.s., not significant; multiple t test. P-values (control vs WT: 0.0026; WT vs 25FSY: 0.0089; WT vs 30FSY: 0.0025; WT vs 49FSY: 0.5371). **e)** Scheme of GRIP. After proteinase K treatment, co-purified Hfq-cross-linked RNA were reversely transcribed by gene-specific RT primer, followed by RNA removal and ligation of a 3's cDNA adaptor containing a random-10mer at the ligation

site. After ligation, PCR was performed with a primer pair, one targeting gene-specific region and the other targeting 3' s cDNA adaptor region. Sequencing of the PCR product could identify the ligation sites, indicating RT terminating sites and the cross-linking sites (red triangle).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

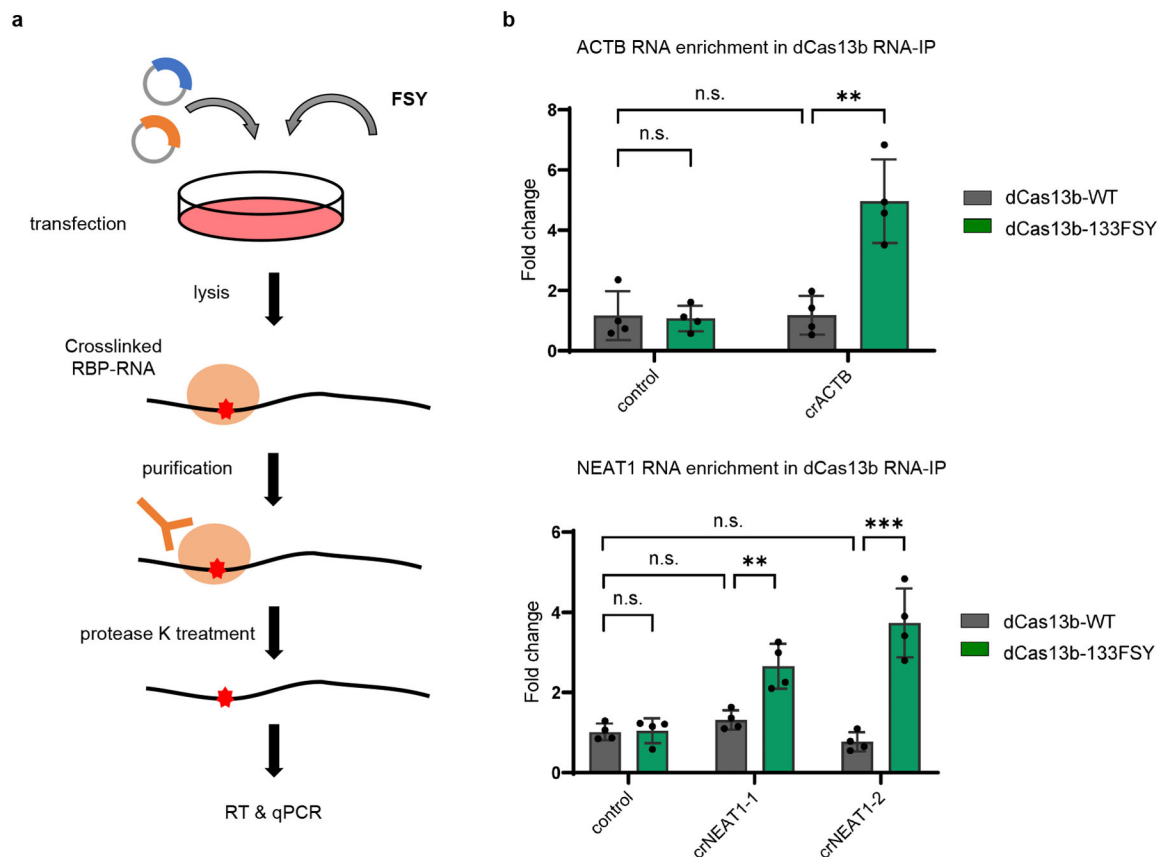


Fig. 3. GECX-RNA enables FSY-incorporated dPsCas13b proteins to cross-link target RNA in mammalian cells.

a) Scheme showing the procedures for quantification of RNA co-purified with dPsCas13b from mammalian cells. **b)** RT-qPCR analysis of dPsCas13b co-purified RNA showed that dPsCas13b-133FSY enriched more target RNA molecules than dPsCas13b-WT with the guidance of crRNA. Control samples had no crRNA plasmid transfected, while crACTB, crNEAT1-1, and crNEAT1-2 samples were transfected with distinct crRNA plasmids targeting ACTB mRNA or NEAT1 RNA. Bar chart showed the fold-changes of target RNAs in crRNA transfected samples compared to control samples (normalized to GAPDH RNA abundance). Bar heights represent mean; Error bars represent s.e.m.; $n = 4$ independent biological replicates; ** $p < 0.01$; *** $p < 0.001$; n.s., not significant; multiple t test (p-values: WT-control vs 133FSY-control: 1.00; WT-control vs WT-crACTB: 0.93; WT-crACTB vs 133FSY vs crACTB: 0.0032; WT-control vs WT-crNEAT1-1: 0.10; WT-control vs WT-crNEAT1-2: 0.15; WT-crNEAT1-1 vs 133FSY- crNEAT1-1: 0.0024; WT-crNEAT1-1 vs 133FSY- crNEAT1-1: 0.00015).

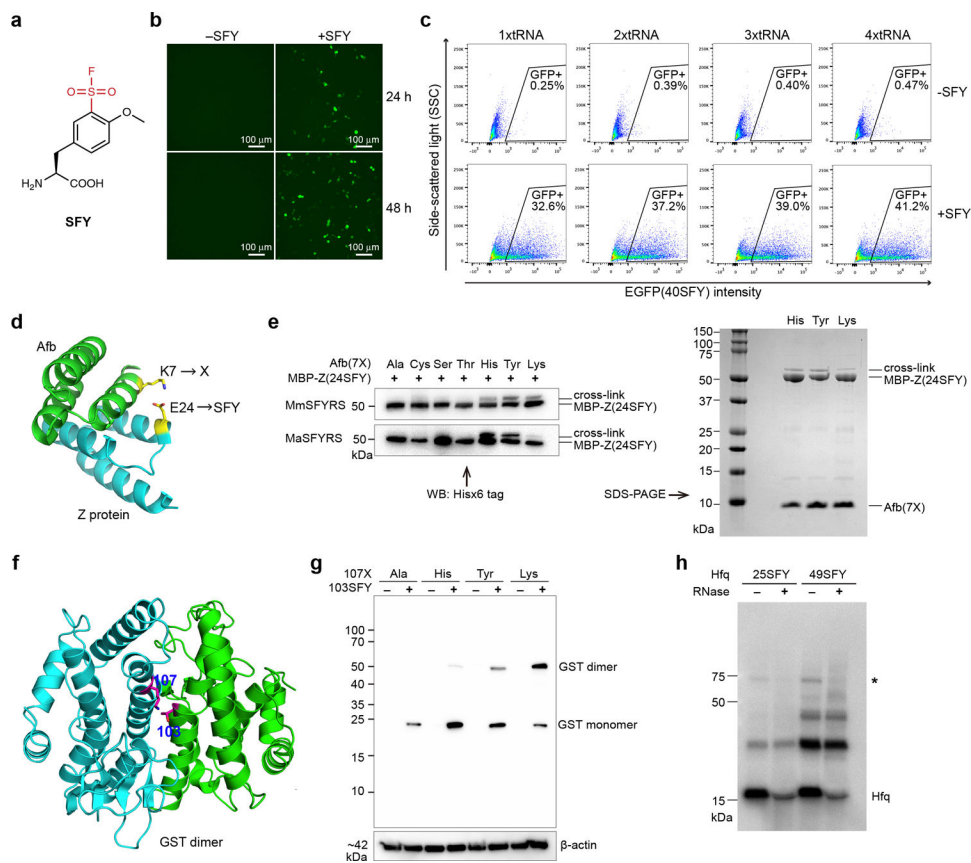


Fig. 4. Genetically encoding SFY allows cross-linking of His, Tyr, Lys residues in protein and of RNA in cells.

a) Structure of SFY. **b)** Fluorescence confocal images HEK293 cells expressing EGFP(40TAG) gene and the Mm-tRNA^{Pyl}/MmSFYRS with and without 1 mM SFY. **c)** Flow cytometric analysis of SFY incorporation into EGFP(40TAG) in HEK293 cells using Ma-tRNA^{Pyl}/MaSFYRS. **d)** Structure of Afb-Z complex (PDB: 1LP1) showing two proximal sites for SFY and target residue X incorporation. **e)** Analysis of cross-linking of Afb(24SFY) with MBP-Z(7X) in *E. coli* cells. Left: Western blot of *E. coli* cell lysate; Right: SDS-PAGE of proteins purified from *E. coli* via Ni²⁺ affinity chromatography. Maltose binding protein (MBP) was fused to the N-terminus of Z protein to better separate Z from Afb in size. **f)** Crystal structure of *E. coli* GST (PDB: 1A0F) showing site 103 and 107 at the dimer interface. **g)** Western blot analysis of lysate of HEK293T cells expressing GST(103SFY-107X). X is the target residue indicated. **h)** Western blot analysis *E. coli* cells expressing Hfq with SFY incorporated at site 25 or 49. Cell lysate samples were treated with or without RNase before loading, and an anti-His antibody was used to detect the 6xHis tag appended at the C-terminus of expressed Hfq. Star indicates a cross-linked band.

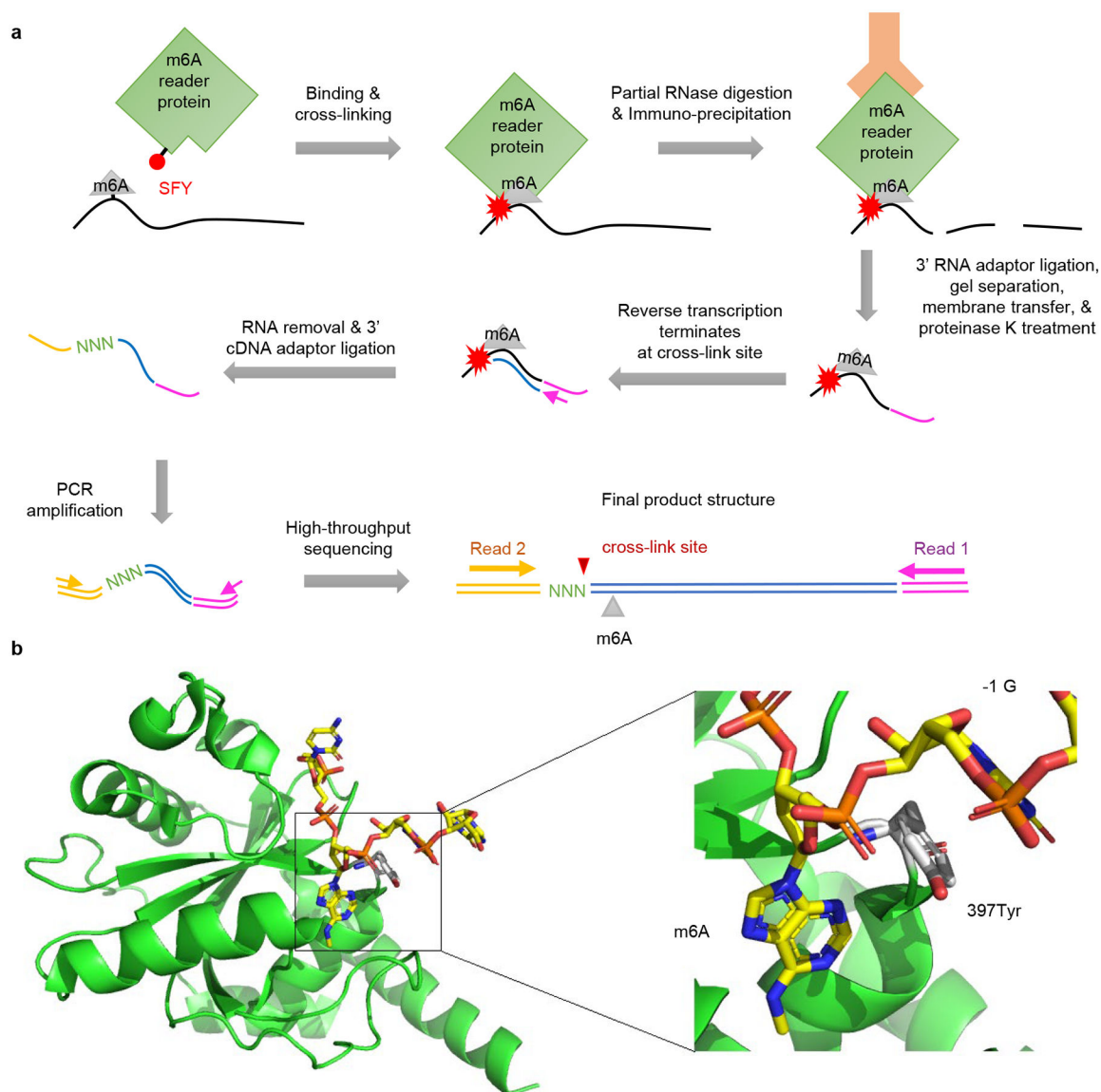


Fig. 5. Design of GRIP-seq for *in vivo* detection of m6A on RNA with single-nucleotide resolution.

a) Scheme showing the principle of using GRIP-seq to detect RNA modifications *in vivo*, using m6A as an example. A reader protein recognizing the RNA modification is expressed in cells, with a latent bioreactive Uaa (SFY) incorporated near the recognition site to cross-link bound RNA for identification. This is followed by partial RNase digestion and an immunoprecipitation enriching reader-proteins and their cross-linked RNA fragments. After dephosphorylation and 3' adaptor ligation with RNA fragments, the cross-linked protein-RNA are separated by SDS-PAGE and transferred to a nitrocellulose membrane. The membrane regions above the read-protein (75 kDa above) are excised and treated with proteinase K to release the cross-linked RNA fragments. The released RNA fragments are further prepared into libraries for pair-end high-throughput sequencing. In the final libraries, read 2 begins with a random-mer sequence (random 10mer, added with 3' cDNA adaptor ligation) followed by the sequence corresponding to the 3' end of reverse-

transcribed cDNA, the junction of which indicates the cross-link sites causing the reverse-transcription termination (See materials and methods). **b**) Structure of YTH domain (from human YTHDF1) binding with m⁶A nucleotide (PDB: 4RCJ). Tyr397, the site chosen for incorporation of SFY is shown in grey stick. RNA is colored in yellow and YTH protein in green.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

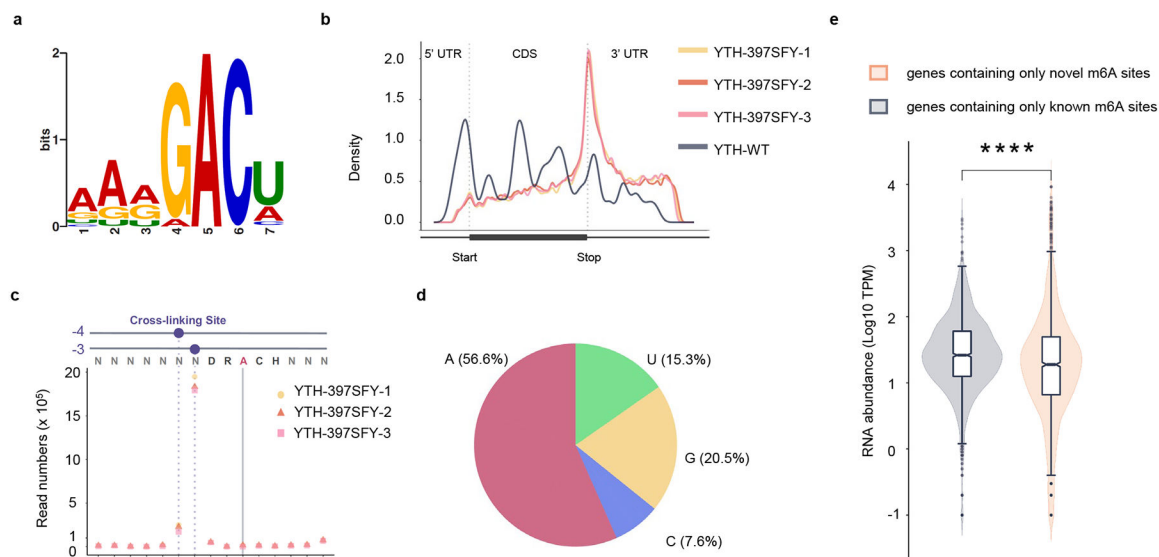


Fig. 6. GRIP-seq *in vivo* detected m6A on RNA with single-nucleotide resolution in mammalian cells.

a) The most enriched motif found in GRIP-seq data of YTH-397SFY-IP samples. The enriched DRACH motif was identical to the published m6A consensus motif. **b)** Reverse-transcription-termination sites identified from YTH-397SFY IP samples showed metagene distribution profiles typical for m6A **c)** Plot showing the cross-links enriched at the upstream of the DRACH motif. X-axis indicated the position relative to m6A (0 position) in the DRACH motif. Y-axis indicated the read numbers (representing RNA molecules) of cross-links at the corresponding positions from YTH-397SFY IP samples. **d)** Pie chart showing the nucleotide composition at the cross-linking sites. **e)** Violin plot (with box plot inside) presenting the distribution of RNA abundance of the two gene groups: genes containing only novel m6A sites (salmon, $N = 1,699$), and genes containing only known m6A sites (grey, $N = 1,826$). Y-axis: TPM (Transcript per million reads) values in log₁₀ scale, representing the RNA abundance. TPM values of each gene in HEK293T cells were from Protein Atlas database⁶¹. Boxes extend from the 25% to 75% of values, with center line denoting median; whiskers denote values within 1.5 interquartile range of the 25th and 75th percentile. Two-sided Wilcoxon rank-sum test for the RNA abundance of the two gene groups: **** $p < 0.0001$ (p value: 4.645×10^{-15}).