bioRxiv preprint doi: https://doi.org/10.1101/2020.10.03.324715; this version posted October 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

CIGAR-seq, a CRISPR/Cas-based method for unbiased screening of novel mRNA modification regulators

Liang Fang[1,2#], Wen Wang[1,3#], Guipeng Li[1,2#], Li Zhang[1], Jun Li[1], Diwen Gan[1], Jiao Yang[1], Yisen Tang[1], Zewen Ding[1], Min Zhang[1], Wenhao Zhang[1], Daqi Deng[1], Zhengyu Song[1], Qionghua Zhu[1], Huanhuan Cui[1,2], Yuhui Hu[1], Wei Chen[1,2,*]

1. Department of Biology, Southern University of Science and Technology, Shenzhen, Guangdong 518005, China
2. Academy for Advanced Interdisciplinary Studies, Southern University of Science and Technology, Shenzhen, Guangdong 518005, China
3. Harbin Institute of Technology, Harbin, Heilongjiang 150001, China

# These authors contributed equally to this work

* Corresponding author. Tel: +86 755 88018449; E-mail: chenw@sustech.edu.cn

Keywords: CIGAR-seq, CRISPR/Cas-based screening, mRNA modification

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.03.324715; this version posted October 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Cellular RNA is decorated with over 170 types of chemical modifications. Many modifications in mRNA, including m^6A and m^5C, have been associated with critical cellular functions under physiological and/or pathological conditions. To understand the biological functions of these modifications, it is vital to identify the regulators that modulate the modification rate. However, a high-throughput method for unbiased screening of these regulators is so far lacking. Here, we report such a method combining pooled CRISPR screen and reporters with RNA modification readout, termed CRISPR integrated gRNA and reporter sequencing (CIGAR-seq). Using CIGAR-seq, we discovered NSUN6 as a novel mRNA m^5C methyltransferase. Subsequent mRNA bisulfite sequencing in HAP1 cells without or with NSUN6 and/or NSUN2 knockout showed that NSUN6 and NSUN2 worked on non-overlapping subsets of mRNA m^5C sites, and together contributed to almost all the m^5C modification in mRNA. Finally, using m^1A as an example, we demonstrated that CIGAR-seq can be easily adapted for identifying regulators of other mRNA modification.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.03.324715; this version posted October 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Cellular RNAs can be chemically modified in over a hundred different ways, and such modifications have been associated with diverse cellular functions under physiological and/or pathological conditions (Machnicka, Milanowska et al., 2013, Roundtree, Evans et al., 2017). To achieve so called dynamic epitranscriptomic regulation, each modification needs its distinct deposition, removal and recognition factors (termed 'writers', 'erasers' and 'readers', respectively). Yet, comparing to the ever-expanding techniques on detecting RNA modifications (Zhao, Song et al., 2020), the methods to systematically identify writers and erasers of RNA modifications are rather limited. For instance, the first N6-adenosine (m^6A) methyltransferase METTL3 was identified through a combination of in vitro assay, conventional chromatography, electrophoresis and microsequencing (Bokar, Rath-Shambaugh et al., 1994, Bokar, Shambaugh et al., 1997), and the second m^6A methyltransferase METTL14 was discovered through the phylogenetic analysis based on METTL3 (Liu, Yue et al., 2014). In general, the first strategy is less efficient and may have assay-specific bias, while the second strategy relies on the prior knowledge of related molecule(s). So far, unbiased method to screen for novel regulators of RNA modifications is still lacking.

Recently, the rapid development of CRISPR-based gene manipulation provides a new paradigm for high-throughput and genome-wide functional screening. Pooled CRISPR screen outperforms array-based screen by its scalability and low cost, however, was largely restricted to standard readouts, including survival, proliferation and FACS-sortable markers (Hanna & Doench, 2020). Most recently, combining with microscopy-based approaches, CRISPR screen enabled the association of subcellular phenotypes with perturbation of specific gene(s) (Wheeler, Vu et al., 2020, Yan, Stuurman et al., 2020). In studying regulation of gene expression, Perturb-seq, CRISP-seq and CROP-seq, which combine CRISPR-based gene editing with single-cell mRNA sequencing, allowed transcriptome profile to serve as comprehensive molecular readout (Adamson, Norman et al., 2016, Datlinger, Rendeiro et al., 2017, Dixit, Parnas et al., 2016), but often with limited throughput. Until now, pooled CRISPR screen with epitranscriptomic readout has not yet been developed.

One important RNA modification, 5-methylcytosine (m^5C), was first identified in stable and highly abundant tRNA and rRNA (Agris, 2008, Helm, 2006, Schaefer, Pollex et al., 2009). Subsequently, many novel m^5C sites in mRNA were discovered by using next-generation sequencing-based methods,

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.03.324715; this version posted October 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

including mRNA bisulfite sequencing (mRNA-BisSeq) (Schaefer et al., 2009, Squires, Patel et al., 2012), m^5C-RNA immunoprecipitation (RIP) (Edelheit, Schwartz et al., 2013), 5-azacytidine-mediated RNA immunoprecipitation (Aza-IP) (Khoddami & Cairns, 2013) and methylation-individual-nucleotide-resolution crosslinking and immunoprecipitation (miCLIP) (Hussain, Sajini et al., 2013). The m^5C modification has been reported to regulate the structure, stability and translation of mRNAs (Guallar, Bi et al., 2018, Li, Li et al., 2017, Luo, Feng et al., 2016, Schumann, Zhang et al., 2020, Shen, Zhang et al., 2018), and be catalyzed by NOP2/Sun RNA methyltransferase family member 2 (NSUN2) (David, Burgess et al., 2017, Khoddami & Cairns, 2013, Yang, Yang et al., 2017a). However, recent studies have shown that, even after NSUN2 knockout (KO), a significant number of m^5C sites in mRNA remained methylated (Huang, Chen et al., 2019, Trixl & Lusser, 2019), suggesting the existence of additional methyltransferase(s) involved in mRNA m^5C modification. To fully appreciate the function of this modification, it would be important to identify the remaining methyltransferase(s).

Here, we report a method combining pooled CRISPR screen and a reporter with epitrancriptomic readout, termed CRISPR integrated gRNA and reporter sequencing (CIGAR-seq). Using CIGAR-seq with a reporter containing a m^5C modification site, we screening through a gRNA library targeting 829 RNA-binding proteins and identified NSUN6 as a novel m^5C writer of mRNA. mRNA-BisSeq in HAP1 cells without or with NSUN6 and/or NSUN2 knockout showed NSUN6 and NSUN2 worked on non-overlapping subsets of mRNA m^5C sites, and together contributed to almost all the m^5C modification in mRNA. Finally, using m^1A as an example, we demonstrated that CIGAR-seq can be easily adapted for studying other mRNA modification.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.03.324715; this version posted October 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Results

CIGAR-seq: Pooled CRISPR screening with a epitranscriptomic readout

In CIGAR-seq, to integrate pooled CRISPR screening with a epitranscriptomic readout, here more specifically m^5C modification readout, we adopted the previously developed CROP-seq method (Datlinger et al., 2017) and replaced the WPRE cassette on the original vector by an endogenous m^5C site with its flanking region (Fig 1A). Thereby, the mRNA molecules transcribed from this lentiviral vector contain a selection marker followed by an endogenous m^5C site, a U6 promoter and a gRNA sequence. To detect the m^5C level in the gRNA sequence-containing transcripts, total mRNA was firstly subjected to bisulfite treatment followed by reverse transcription. Subsequently, a primer pair flanking the m^5C site and gRNA sequence was used to amplify the region for Sanger or next-generation sequencing (Fig 1A). In this way, the methylation level of the m^5C reporter site can be measured and associated with the gRNA targeting a specific gene.

As a proof-of-concept experiment, a known NSUN2-dependent m^5C site in FAM129B (also known as NIBAN2) gene was cloned into CIGAR-seq vector, together with a control gRNA without any target gene or a gRNA targeting NSUN2 (Fig 1B, upper panel). Seven days after transduction into Cas9-expressing HAP1 cells, m^5C modification level on the reporter transcripts was measured. The result demonstrated that, upon NSUN2 perturbation (Fig EV1A), the m^5C modification rate reduced significantly in the reporter containing NSUN2 targeting gRNAs (Fig 1B, lower panel), whereas the modification remained intact in the reporter containing the control gRNA (Fig 1B, upper panel).

CIGAR-seq identified NSUN6 as a novel mRNA methyltransferase

As suggested by previous studies that large amount of m^5C sites in mRNA remained methylated after NSUN2 knockout (Huang et al., 2019, Trixl & Lusser, 2019), we sought to utilize CIGAR-seq to identify gene(s) that mediate the m^5C modification on NSUN2-independent sites. First, to determine the NSUN2-independent m^5C sites, we established NSUN2 knockout (NSUN2-KO) HAP1 cells (Fig EV4A), and performed mRNA bisulfite sequencing (mRNA-BisSeq) in wildtype as well as NSUN2-KO cells. Following bioinformatic pipeline proposed by Huang et al. (Huang et al., 2019), a set of 208 m^5C sites was identified in wildtype HAP1 cells (Methods), only 90 (43.3%) of which showed significantly reduced m^5C level in NSUN2-KO cells (Fig EV1B).

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.03.324715; this version posted October 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

We then chose a NSUN2-independent site in the 3'UTR of FURIN gene with a high m^5C modification rate as the reporter site for CIGAR-seq. Meanwhile, a gRNA library targeting 829 RNA binding proteins (RBP) was synthesized (Table EV1). To establish the CIGAR-seq vector pool for the genetic screen, the gRNA library was firstly cloned into the vector followed by the insertion of the FURIN m^5C site with its flanking genomic region (Methods). Cas9-expressing HAP1 cells were then transduced with CIGAR-seq virus pool (Fig 1C, Methods). Seven days after transduction, cells were collected and subjected to RNA extraction. Enriched polyA RNA was then bisulfite-treated, reveres-transcribed and PCR-amplified using primers flanking the m^5C site and gRNA sequence to generate next-generation sequencing (NGS) library (Methods). After pair-end sequencing and data processing, 811 genes were detected with at least one gRNA, of which 782 genes had at least two gRNAs (Fig EV1C). The m^5C modification rate of reporter site was calculated for each gRNA. While the median m^5C modification rates of gRNA-associated reporter sites were around 93.5%, a small part of gRNAs showed significantly reduced m^5C rates (Fig EV1D). To prioritize the candidate genes, Stouffer's method was used to calculate the combined P value based on the gRNAs targeting the same gene (Methods). As shown in Fig 1D, it turned out that NSUN6, a member of NOL1/NOP2/sun domain (NSUN) family, was identified as the best hit (Fig 1D). Interestingly, NSUN6 was previously reported to introduce the m^5C in tRNA (Li, Li et al., 2019). However, two previous studies did not show significant m^5C changes in mRNA after NSUN6 perturbation in HeLa cells (Huang et al., 2019, Yang et al., 2017a).

To validate our result, a gRNA targeting NSUN6 was inserted into the CIGAR-seq vector containing the FURIN m^5C site. As shown in Figure 1E, perturbation of NSUN6 (Fig EV2A) indeed reduced the m^5C level in both reporter mRNA and endogenous NSUN2-independent m^5C site in RPSA gene (Fig 1E, left and middle panels). Furthermore, to rule out the potential off-target effect of gRNA, we repressed NSUN6 expression using shRNA (Fig EV2B). Again, the m^5C level at the endogenous site was also reduced in cells with NSUN6 repression (Fig 1E, right panel). Together, these results confirmed NSUN6 as a bona fide mRNA m^5C methyltransferase.

Global profiling of NSUN6-dependent m^5C sites

To globally characterize NSUN6-dependent m^5C sites, we established NSUN6 knockout (NSUN6-KO) HAP1 cells (Fig EV4B) and performed mRNA-BisSeq. Of 208 m^5C sites identified in wildtype HAP1 cells,

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.03.324715; this version posted October 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

65 (31.2%) showed significant reduction at m^5C level in NSUN6-KO cells (Fig 2A). To illustrate the features of sequence flanking NSUN6-dependent m^5C sites in HAP1 cells, motif analysis was performed based on the upstream and downstream 10 nucleotide sequences flanking the m^5C sites. As shown in Figure 2B, NSUN6-dependent m^5C sites were embedded in slightly GC-rich environments with a strongly enriched TCCA motif at 3' of m^5C sites. Previously, a similar 3' TCCA motif was also found at NSUN6 target sites in tRNAs (Li et al., 2019), and has also been proposed as sequence motif around NSUN2-independent sites in another study (Huang et al., 2019). In comparison, the sequence feature around NSUN2-dependent m^5C sites is distinct, which is enrich for 3' NGGG motif (Huang et al., 2019, Yang, Yang et al., 2017b).

Contribution of NSUN6 and NSUN2 to the mRNA m^5C modification

We then evaluated the relative contribution of NSUN6 and NSUN2 to the global mRNA m^5C modification. First, comparing between NSUN2- and NSUN6-dependent m^5C sites, as shown in Figure 3A, these sites were largely non-overlapping, suggesting their non-redundant biological functions. Then, to further examine whether NSUN2 and NSUN6 together are responsible for all mRNA m^5C modifications, NSUN2 and NSUN6 double KO (NSUN2/6-dKO) HAP1 cells were established (Fig EV4C) and subjected to mRNA-BisSeq analysis. As shown in Figure 3B, the modification of m^5C sites depend only on NSUN6 or NSUN2 (62 and 87, respectively) were also abolished in NSUN2/6-dKO cells. While NSUN6-dependent sites were strongly enriched for 3' TCCA motif as shown earlier, NSUN2-dependent sites were enriched for 3' NGGG motif as previously reported (Huang et al., 2019, Yang et al., 2017b) (Fig 3C). Furthermore, we carefully examined the three sites that showed dependence on both NSUN6 and NSUN2, as well as the 56 sites that were independent of both NSUN6 and NSUN2. Comparing to the other three groups, the group of three overlapping sites had very low m^5C level (Fig 3D). In addition, the m^5C sites in ANGEL1 and ZNF707 possessed a 3' TCCA and a 3' AGGG motif, respectively (Fig EV3A&B), suggesting they are very likely a NSUN6- and a NSUN2-dependent site, respectively, but with low m^5C level that led to false negative findings in the mRNA-BisSeq analysis of some but not all the samples. The remaining m^5C site in STRN4 was embedded within a cluster of "pseudo" m^5C sites (Fig EV3C), which was highly likely an artifact due to the incomplete bisulfite conversion as suggested before (Haag, Warda et al., 2015, Huang et al., 2019). Similarly, the group of 56 NSUN2/6-independent sites were also highly enriched for such clusters of pseudo m^5C sites: 52 sites had at least one pseudo m^5C site in vicinity (Table EV2). The remaining four sites all had very low m^5C level.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.03.324715; this version posted October 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

174

175 To explore the modification rate of NSUN6/2-dependent m^5C sites across different tissues, we resorted
176 to mRNA-BisSeq data from a previous study (Huang et al., 2019). As shown in Figure 3E, m^5C
177 modification on 47 NSUN6- and 66 NSUN2-dependent m^5C sites could also be observed in other human
178 tissue(s). While the modification rate of NSUN6-dependent sites was by and large highest in liver,
179 NSUN2-dependent ones did not show such tissue biases (Fig 3E).

180

181 **CIGAR-seq could be used for the study of other mRNA modification**

182 Finally, to explore the potential application of CIGAR-seq in the study of other mRNA modifications, we
183 turned to N-1-methyladenosine (m^1A). As *N*-1-methyladenosine (m^1A) can cause mis-incorporation
184 during cDNA synthesis (Hauenschild, Tserovski et al., 2015), its modification can be detected by direct
185 cDNA sequencing. Similar as the previous NSUN2 proof-of-concept experiment, we chose a well-
186 characterized m^1A site from MALAT1, which is known to be modified by TRMT6/TRMT61A complex
187 (Dominissini, Nachtergaele et al., 2016, Li, Xiong et al., 2017, Safra, Sas-Chen et al., 2017). We cloned
188 the site and its flanking region into CIGAR-seq vector with a control gRNA and two gRNAs targeting
189 TRMT6/TRMT61A complex, respectively (Fig 4A). As shown in Figure 4C, in HAP1 cells with
190 perturbation of either TRMT6 or TRMT61A (Fig 4B), the m^1A modification of the reporter site was
191 completely abolished, whereas in cells transduced with control gRNAs, the modification remains intact.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.03.324715; this version posted October 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Discussion

Combining pooled CRISPR screening strategy and a reporter with epitranscriptomic readout, CIGAR-seq for the first time enables the unbiased screening for novel regulators of mRNA modifications. In this study, we demonstrated its power in identification of NSUN6 as a novel mRNA m^5C methyltransferase. In addition, we also showed its potential application in studying m^1A modification. Integrating additional modification readout strategies into our pipeline, it could be further adapted to investigate other modifications. For instance, we can use CIGAR-seq to search for potential regulators of RNA editing by simply reading the A-G or C-T changes in the cDNA sequence reads derived from A-I or C-U RNA editing reporters. m^6A-iCLIP (Linder, Grozhik et al., 2015) or SELECT (Xiao, Wang et al., 2018) method, which were used to measure the modification rate of individual m^6A site, could also be integrated into our CIGAR-seq in analyzing m^6A regulators. Furthermore, changing reporters to those with other regulatory readout, for example alternative splicing or alternative polyadenylation pattern, the potential application of CIGAR-seq could be easily extended to screen for factors involved in diverse post-transcriptional regulations. There, given the readout is based on directly measuring the reporter-derived RNAs, CIGAR-seq would be in principle superior to current fluorescence-reporter-FACS based screening strategies.

Like any other high-throughput assays, CIGAR-seq has also its own sensitivity and specificity issues, which could be affected by the choice of reporter and CRISPR system. The reporter could affect its performance in two ways. First, the high or low modification level of the reporter site could result in biased performance in detecting positive or negative regulators. For example, in this study, our m^5C site from FURIN genes has a very high modification rate. At this level, it would be much more sensitive in finding the decrease of methylation rate, therefore be much easier to discover the methyltransferase than potential demethylase if any in this case. In contrast, the use of reporter site with low modification rate would not be preferable in identifying methyltransferase. Second, except writer and eraser, most regulators may modulate the modification level through binding to the cis-regulatory elements, which are not necessarily in direct vicinity of the target site. A reporter construct with a limited length might not be able to include all the relevant cis-elements. Consequently, we would fail to identify the regulators with binding sites missed in the reporter. On the other hand, the CIGAR-seq vector itself may contain artificial regulatory sequences affecting the modification of reporter site, which could result in the assay-specific artifacts. Therefore, subsequent careful validation with endogenous sites would be essential

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.03.324715; this version posted October 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

when working with the CIGAR-seq. The choice of CRISPR system could also have an effect. The screening based on CRISPR/Cas9 system, as applied in this study, would have limitations in finding potential regulators that are essential for cell survival and/or proliferation (e.g. METTL3/METTL14) since the gRNAs targeted at those essential genes would be largely depleted in the final sequencing library. This problem could be potentially alleviated by adopting CRISPRi or CRISPRa systems. In the future, with further improvements in CRISPR system and development of more sequencing-based readout with high precision, CIGAR-seq will become a versatile tool for systematic discoveries of players in multiple layer of RNA-based post-transcriptional gene regulation.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.03.324715; this version posted October 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Methods

Experimental methods

Cell culture and gene manipulation

HAP1 cell was obtained from Horizon discovery and cultured in RMPI1640 medium (22400089, Gibco) with 10% FBS (10270106, Gibco) and 1% P/S (15070063, Gibco) at 37°C with 5% CO2. Cas9-expressing HAP1 cell line was established by using lentiCas9-Blast plasmid (#52962, Addgene). To generate NSUN2-KO, NSUN6-KO and NSUN2/6-dKO clonal HAP1 cells, Cas9-expressing HAP1 cells were transduced with CROP-seq (Addgene, #86708) virus expressing following gRNAs: gNSUN2, 5'-GCTGTTCGAGCACTACTACC-3'; gNSUN6, 5'-GACCTTCAAGATGTGTTACT-3'. NSUN6 knockdown mediated by shRNA was performed using pLKO.1-blast plasmid (modified from pLKO.1-puro, #10878, Addgene) with following shRNAs: shControl, 5'-CGTCTGGCTAATAAGGACTCT-3'; shNSUN6-1, 5'-GCAAAGAAATCTTCAGTGGAT-3'; shNSUN6-2, 5'-GCTGGAGATGTTATTTCTGTA-3'.

RT-qPCR

Total RNA was extracted by TRIzol® Reagent (Ambion). First-stand cDNA was synthesized using HiScript III 1st Stand cDNA Synthesis Kit (Vazyme, #R312-02). Quantitative PCR was performed by Hieff qPCR SYBR Green Master Mix (Yeasen, #11201ES08) and the BIO-RAD real-time PCR system. Following primers were used to detect relative gene expression: NSUN6-F, 5'-GGAGCCAAAGAATTTGATGGAACA-3'; NSUN6-R, 5'-ATGCCCATGCCTTTCAGTTC-3'; GAPDH-F, 5'- AGCCACATCGCTCAGACAC-3'; GAPDH-R, 5'- GCCCAATACGACCAAATCC-3'.

CIGAR-seq vector with m^{5}C/m^{1}A reporters and individual gRNA

Sequence flanking m^{5}C site of FAM129B was amplified by forward primer 5'-GCCTGAACGCGTTAAGTCGAC-GGCTGGACACTGCTGGGG-3' and reverse primer 5'-GTAAGTCATTGGTCTTAAAGTCGAC-GGGGAAAGCGAGGCTCG-3' from genomic DNA; m^{5}C site of FURIN by forward primer 5'-GCCTGAACGCGTTAAGTCGAC-CCGGCCCCAGCCAGAGTTC-3' and reverse primer 5'-GTAAGTCATTGGTCTTAAAGTCGAC-TGGTGGAGGCACGGAGCACA-3', and m^{1}A site of MALAT1 by forward primer 5'-GCCTGAACGCGTTAAGTCGAC-CTTCAGTAGGGTCATGAAGGTTTTTCT-3' and reverse primer 5'-GTAAGTCATTGGTCTTAAAGTCGAC-ATACATCAAGGATGTATATAGTTCAAAGATATTGTGC-3'.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.03.324715; this version posted October 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Amplified products were used to replace WPRE cassette in CROP-seq vector (Addgene, #86708) by ClonExpress II One Step Cloning Kit (Vazyme). Afterwards, following gRNAs were inserted at BsmBI sites to knockout individual genes: gControl, 5'-GAGGGATCGTTAGGAAGGG-3'; gNSUN2, 5'-GCTGTTCGAGCACTACTACC-3'; gNSUN6, 5'-GACCTTCAAGATGTGTTACT-3'; gTRMT6, 5'-GGTGCAATGATGGAACGAAT-3'; gTRMT61A, 5'-TTCGGCTCCAAGGTGACGTG-3'.

m^5C detection by bisulfite conversion followed by sanger sequence

Total RNA was extracted by TRIzol® Reagent (Ambion). mRNA was enriched using VAHTS mRNA Capture Beads (Vazyme, #N401). 200 ng mRNA was converted by EZ RNA methylation kit (Zymo Research) according to the manufacturer's protocol with minor modification. More specifically, mRNA was incubated at 70 °C for 10 min, and 60 °C for 1 h. Converted RNA was then reverse transcribed into cDNA using HiScript II Q Select RT SuperMix (Vazyme, #R233-01).

To measure m^5C rate in FURIN m^5C reporter, target site was amplified using vector specific primer pair 5'-TTGTAATTTTTTTTTTTATGAGTGGTTTGGTTTTA-3' and 5'-TTAAAAAATAACTAAAATCTACAACTACCTTATAAATCATTAATCTTAA-3', and sanger-sequenced by primer 5'-TTGTAATTTTTTTTTTTATGAGTGGTTTGGTTTTA-3'. For m^5C detection of endogenous m^5C site in RPSA, target site was amplified using primers 5'-AAATTTTAAGAGGATTTGGGAGAAGTTTTTG-3' and 5'-CAACCCTAAAATCAATAACCACAAAAAACCATA-3', and sanger-sequenced by primer 5'-AAATTTTAAGAGGATTTGGGAGAAGTTTTTG-3'.

m^1A detection based on mis-incorporation during reverse transcription

1 μg total RNA was reverse transcribed by HifairTM Ⅱ 1st Strand cDNA Synthesis Kit (Yeasen, # 11121ES60). The region flanking m^1A site was amplified by plasmid specific primer 5'-TTCACCGTCACCGCCGAC-3' and 5'-CTAATTCACTCCCAACGAAGACAAGATTT-3'. The mismatch site was measured by sanger sequencing using primer 5'-CTAATTCACTCCCAACGAAGACAAGATTT-3'.

mRNA-BisSeq

The quality of 500 ng bisulfite-treated mRNA (see above) was assessed using Agilent RNA 6000 Pico Kit (Agilent, #NC1711873), and then subjected to NGS libraries preparation using VAHTS Stranded

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.03.324715; this version posted October 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

mRNA-seq Library Prep Kit (Vazyme). The library quality was assessed using High Sensitivity DNA Kit (Agilent, #5067-4626). Paired-end sequencing (2x150 bp) was performed with Illumina NovaSeq 6000 System by Haplox genomics center. The raw sequencing data have been deposited to GEO under the accession number GSE157368.

Generation of CIGAR-seq vector pool with a FURIN m^{5}C reporter

A gRNA library containing 4975 gRNA targeting 829 RBP (Table EV1) was synthesized by GENEWIZ and cloned into CROP-seq vector (Addgene, #86708) at BsmBI sites. For measuring the complexity of the gRNA library, the region harboring gRNA sequence was amplified with primer pair 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCTTGTA-TATCCCTTGGAGAACCACCTTGTTG-3' and 5'-CAAGCAGAAGACGGCATACGAGATCCACTCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-CGACTCGGTGCCACTTTTTCAAGTTG-3' for NGS. Afterwards, the FURIN m^{5}C reporter was amplified and used to replace WPRE cassette using ClonExpress II One Step Cloning Kit (Vazyme). During cloning of CIGAR-seq vector pool, electrocompetent Stbl3 cells (Weidi Biotechnology, CAT#: DE1046) was always used.

CIGAR-seq viral package

HEK293T cells were plated onto 15 cm plates at 40% confluence. The next day, cells were transfected with PEI (Polysciences, #23966-2) using 15 μg of CIGAR-seq vector, 15 μg of psPAX2 (Addgene, #12259) and 22.5 μg of pMD2.G (Addgene, #12259). Supernatant containing viral particles were harvested at 48 h and 96 h, and purified with 0.45 μm filter.

Genetic screen for novel m^{5}C regulators

2x10^8 HAP1 cells were infected with CIGAR-seq viral particles (MOI = 0.3) and treated with 1 μg/ml of Puromycin for 24 h post infection. Puromycin resistant cells were cultured for additional seven days, and then 2x10^8 HAP1 cells were collected for RNA extraction. 100 ng of bisulfite-treated mRNA (see above) was quantified using Agilent RNA 6000 Pico Kit (Agilent, #NC1711873) and reverse transcribed into cDNA using HiScript II Q Select RT SuperMix (Vazyme, #R233-01). Finally, using the cDNA, CIGAR-seq NGS library was amplified with forward primer 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.03.324715; this version posted October 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

GGGTTGGTTTAGGAGATATTTGAGGG-3' and reverse primer 5'-CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-AACAATCCTAATACTCAAAAAAAAAAACACCA-3'. Paired-end sequencing (2x150 bp) was performed with Illumina NovaSeq 6000 System by Haplox genomics center. The raw sequencing data have been deposited to GEO under the accession number GSE157368.

Western blotting

Transfected HAP1 cells were collected and lysed by RIPA buffer (150 mM NaCl, 50 mM Tris, 1% EDTA, 1% NP40, 0.1% SDS). Lysate was incubated at 4 °C for 30 min, then sonicated with 10 cycles (30 s On /30 s Off), and then centrifuged at 15,000 g for 15 min at 4 °C. The total protein concentration was measured by BCA (Beyotime, #P0011). 60 μg total protein was loaded and separated on the 10% SDS-polyacrylamide gel. The protein on the gel was transfected to the polyvinylidene difluoride membranes (Immobilon-P, #IPVH00010). The membrane was incubated with primary antibody and horseradish peroxidase–conjugated secondary antibody, and then proteins were detected using the Pierce™ ECL Western Blotting Substrate (Thermo, #32209) by BIO-RAD ChemiDoc^{TM} XRS+ system. The following antibodies were used for western blotting: NSUN2 (Proteintech, #20854-1-AP), NSUN6 (Proteintech, #17240-1-AP) and GAPDH (TransGen Biotech, #N10404).

Computational Methods

CIGRA-seq data analysis

CIGRA-seq NGS data consists of paired end reads. Read1 contains the sequence of m^{5}C reporter site while read2 consists of the gRNA sequence. Raw fastq data were first trimmed using fastp (Chen, Zhou et al., 2018) to remove low-quality bases (-A -w 12 --length_required 30 -q 30). Then the clean read pairs were parsed using a custom script based on pysam package. Specifically, gRNA sequence in read2 was extracted by regex module using regular expression ((CAACTTAACTCTTAAAC[ATCG]{20}CA){s<=1}) . m^{5}C reporter sequence was extracted in the similar way ((GTTATTT[TC]{1}TTTAAGG){s<=1}). At most 1 substitution was allowed during the pattern searching. Read pairs with both reads containing the matched pattern sequences and the m^{5}C sites being C or T were kept for further analysis. Then for each gRNA sequence, the number of supported reads with reporter site being C (m^{5}C) or T were calculated, and the number of C reads divided by the

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.03.324715; this version posted October 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

sum of C and T reads represented the m^5C level. Only the extracted gRNA sequences that match exactly with the RBP gRNA sequences (Table EV1) were kept for further analysis.

To identify the high-confident candidate genes that regulate m^5C level, information of multiple gRNAs of the same genes were combined using the Stouffer's method. gRNAs with no more than 20 supported reads were filtered out. Genes with only one gRNA detected were filtered out. Then, given a gene i, the m^5C level of reporter site correspondence to gRNA j is $X_{i,j.}$ m^5C level was converted to Z-score and p value $P_{i,j}$ was calculated under normal distribution assumption. Then a combined p value for each gene P_i was obtain using the weighted version of Stouffer's method, with the logarithmic scale of read count as weight for each gRNA. Finally, P values of multiple tests were adjusted with Benjamini & Hochberg's method.

mRNA-BisSeq data analysis

mRNA-BisSeq data generated in this study were analyzed following the RNA-m^5C pipeline (Huang et al., 2019) (https://github.com/SYSU-zhanglab/RNA-m5C). Reference genomes (GRCh38) and gene annotation GTF file was downloaded from Ensemble (http://www.ensembl.org/info/data/ftp/index.html). Briefly, raw paired-end reads were trimmed using cutadapt (Martin, 2011) (-a AGATCGGAAGAGCACACGTC -A AGATCGGAAGAGCGTCGTGT -j 12 -e 0.25 -q 30 --trim-n) and then Trimmomatic(Bolger, Lohse et al., 2014) (SLIDINGWINDOW:4:25 AVGQUAL:30 MINLEN:36). Clean read pairs were aligned to both C-to-T and G-to-A converted reference genomes by HISAT2 (Kim, Paggi et al., 2019). Unmapped and multiple mapped reads were then aligned to C-to-T converted transcriptome by Bowtie2 (Langmead & Salzberg, 2012), and the transcript coordinates were liftovered to the genomic coordinates. Reads from HISAT2 and Bowite2 mapping were merged and filtered using the same criteria as in RNA-m^5C pipeline. Bam file was transformed into pileup file (--trim-head 6 --trim-tail 6). Putative m^5C sites were called using script m^5C_caller_multiple.py inside RNA-m^5C pipeline (with parameters -P 8 -c 20 -C 2 -r 0.05 -p 0.05 --method binomial). Default parameters of RNA-m^5C scripts were used unless otherwise specified.

NSUN6-dependent m^5C sites

First, to determine a set of high-confident m^5C sites in HAP1 cells, five replicates of mRNA-BisSeq data generated from the WT HAP1 cells were used. The criteria to determine the high-confident m^5C sites were: (1) coverage of the site being at least 20 reads in all five replicates; (2) number of reads containing

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.03.324715; this version posted October 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

384 the unmodified C being at least 2 in all five replicates; (3) the WT methylation level (the minimum
385 methylation level from the five replicates) being at least 0.05. Then, to determine the NSUN6-dependent
386 m^5C sites, m^5C level of the sites were at least 0.05 in WT HAP1 cells and less than 0.02 or 10% of the
387 WT m^5C level in NSUN6-KO HAP1 cells. NSUN2-dependent sites were defined based on the same
388 criteria.
389
390 **Features of the m^5C sites**
391 The upstream and downstream 10 bp sequences flanking the m^5C sites were extracted from the genome.
392 Motif analysis was performed Using ggseqlogo (Wagih, 2017) R package.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.03.324715; this version posted October 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Acknowledgments

This work was supported by the Shenzhen-Hong Kong Institute of Brain Science-Shenzhen Fundamental Research Institutions (Grant No. 2019SHIBS0002), Shenzhen Science and Technology Program (Grant No. KQTD20180411143432337, JCYJ20190809154407564 and JCYJ20180504165804015) and the National Natural Science Foundation of China (Grant No. 31701237, 31900431 and 31970601). The authors acknowledge the Center for Computational Science and Engineering of SUSTech for the support on computational resource and acknowledge the SUSTech Core Research Facilities and Guixin Ruan for technical support.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.03.324715; this version posted October 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Author contributions

W.C. and L.F. developed the concept of the project. W.W., L.F., L.Z., D.G., J.Y. and Y.T. designed and performed experiments. G.L. performed bioinformatic analysis. W.Z., M.Z., D.D., Z.S., Q.Z. and Z.D. assisted in performing experiments. W.C., L.F., G.L., Y.H., W.W., J.L., H.C. and W.S. reviewed and discussed results. W.C., L.F., G.L. and W.W. wrote the manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.03.324715; this version posted October 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

406 **Reference**

407

408 Adamson B, Norman TM, Jost M, Cho MY, Nunez JK, Chen Y, Villalta JE, Gilbert LA, Horlbeck MA,
409 Hein MY, Pak RA, Gray AN, Gross CA, Dixit A, Parnas O, Regev A, Weissman JS (2016) A Multiplexed
410 Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein
411 Response. Cell 167: 1867-1882 e21
412 Agris PF (2008) Bringing order to translation: the contributions of transfer RNA anticodon-domain
413 modifications. EMBO Rep 9: 629-35
414 Bokar JA, Rath-Shambaugh ME, Ludwiczak R, Narayan P, Rottman F (1994) Characterization and
415 partial purification of mRNA N6-adenosine methyltransferase from HeLa cell nuclei. Internal mRNA
416 methylation requires a multisubunit complex. J Biol Chem 269: 17697-704
417 Bokar JA, Shambaugh ME, Polayes D, Matera AG, Rottman FM (1997) Purification and cDNA cloning
418 of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. RNA 3: 1233-
419 47
420 Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data.
421 Bioinformatics 30: 2114-20
422 Chen S, Zhou Y, Chen Y, Gu J (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics
423 34: i884-i890
424 Datlinger P, Rendeiro AF, Schmidl C, Krausgruber T, Traxler P, Klughammer J, Schuster LC, Kuchler
425 A, Alpar D, Bock C (2017) Pooled CRISPR screening with single-cell transcriptome readout. Nat
426 Methods 14: 297-301
427 David R, Burgess A, Parker B, Li J, Pulsford K, Sibbritt T, Preiss T, Searle IR (2017) Transcriptome-
428 Wide Mapping of RNA 5-Methylcytosine in Arabidopsis mRNAs and Noncoding RNAs. Plant Cell 29:
429 445-460
430 Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, Marjanovic ND, Dionne D, Burks T,
431 Raychowdhury R, Adamson B, Norman TM, Lander ES, Weissman JS, Friedman N, Regev A (2016)
432 Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic
433 Screens. Cell 167: 1853-1866 e17
434 Dominissini D, Nachtergaele S, Moshitch-Moshkovitz S, Peer E, Kol N, Ben-Haim MS, Dai Q, Di Segni
435 A, Salmon-Divon M, Clark WC, Zheng G, Pan T, Solomon O, Eyal E, Hershkovitz V, Han D, Dore LC,
436 Amariglio N, Rechavi G, He C (2016) The dynamic N(1)-methyladenosine methylome in eukaryotic
437 messenger RNA. Nature 530: 441-6
438 Edelheit S, Schwartz S, Mumbach MR, Wurtzel O, Sorek R (2013) Transcriptome-wide mapping of 5-
439 methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs.
440 PLoS Genet 9: e1003602
441 Guallar D, Bi X, Pardavila JA, Huang X, Saenz C, Shi X, Zhou H, Faiola F, Ding J, Haruehanroengra P,
442 Yang F, Li D, Sanchez-Priego C, Saunders A, Pan F, Valdes VJ, Kelley K, Blanco MG, Chen L, Wang
443 H et al. (2018) RNA-dependent chromatin targeting of TET2 for endogenous retrovirus control in
444 pluripotent stem cells. Nat Genet 50: 443-451
445 Haag S, Warda AS, Kretschmer J, Gunnigmann MA, Hobartner C, Bohnsack MT (2015) NSUN6 is a
446 human RNA methyltransferase that catalyzes formation of m5C72 in specific tRNAs. RNA 21: 1532-43
447 Hanna RE, Doench JG (2020) Design and analysis of CRISPR-Cas experiments. Nat Biotechnol
448 Hauenschild R, Tserovski L, Schmid K, Thuring K, Winz ML, Sharma S, Entian KD, Wacheul L,
449 Lafontaine DL, Anderson J, Alfonzo J, Hildebrandt A, Jaschke A, Motorin Y, Helm M (2015) The reverse
450 transcription signature of N-1-methyladenosine in RNA-Seq is sequence dependent. Nucleic Acids Res
451 43: 9950-64
452 Helm M (2006) Post-transcriptional nucleotide modification and alternative folding of RNA. Nucleic Acids
453 Res 34: 721-33
454 Huang T, Chen W, Liu J, Gu N, Zhang R (2019) Genome-wide identification of mRNA 5-methylcytosine
455 in mammals. Nat Struct Mol Biol 26: 380-388
456 Hussain S, Sajini AA, Blanco S, Dietmann S, Lombard P, Sugimoto Y, Paramor M, Gleeson JG, Odom
457 DT, Ule J, Frye M (2013) NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines
458 its processing into regulatory small RNAs. Cell Rep 4: 255-61
459 Khoddami V, Cairns BR (2013) Identification of direct targets and modified bases of RNA cytosine
460 methyltransferases. Nat Biotechnol 31: 458-64
461 Kim D, Paggi JM, Park C, Bennett C, Salzberg SL (2019) Graph-based genome alignment and
462 genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 37: 907-915
463 Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9: 357-9

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.03.324715; this version posted October 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Li J, Li H, Long T, Dong H, Wang ED, Liu RJ (2019) Archaeal NSUN6 catalyzes m5C72 modification on a wide-range of specific tRNAs. Nucleic Acids Res 47: 2041-2055

Li Q, Li X, Tang H, Jiang B, Dou Y, Gorospe M, Wang W (2017) NSUN2-Mediated m5C Methylation and METTL3/METTL14-Mediated m6A Methylation Cooperatively Enhance p21 Translation. J Cell Biochem 118: 2587-2598

Li X, Xiong X, Zhang M, Wang K, Chen Y, Zhou J, Mao Y, Lv J, Yi D, Chen XW, Wang C, Qian SB, Yi C (2017) Base-Resolution Mapping Reveals Distinct m(1)A Methylome in Nuclear- and Mitochondrial-Encoded Transcripts. Mol Cell 68: 993-1005 e9

Linder B, Grozhik AV, Olarerin-George AO, Meydan C, Mason CE, Jaffrey SR (2015) Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. Nat Methods 12: 767-72

Liu J, Yue Y, Han D, Wang X, Fu Y, Zhang L, Jia G, Yu M, Lu Z, Deng X, Dai Q, Chen W, He C (2014) A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. Nat Chem Biol 10: 93-5

Luo Y, Feng J, Xu Q, Wang W, Wang X (2016) NSun2 Deficiency Protects Endothelium From Inflammation via mRNA Methylation of ICAM-1. Circ Res 118: 944-56

Machnicka MA, Milanowska K, Osman Oglou O, Purta E, Kurkowska M, Olchowik A, Januszewski W, Kalinowski S, Dunin-Horkawicz S, Rother KM, Helm M, Bujnicki JM, Grosjean H (2013) MODOMICS: a database of RNA modification pathways--2013 update. Nucleic Acids Res 41: D262-7

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. 2011 17: 3 %J EMBnet.journal

Roundtree IA, Evans ME, Pan T, He C (2017) Dynamic RNA Modifications in Gene Expression Regulation. Cell 169: 1187-1200

Safra M, Sas-Chen A, Nir R, Winkler R, Nachshon A, Bar-Yaacov D, Erlacher M, Rossmanith W, Stern-Ginossar N, Schwartz S (2017) The m1A landscape on cytosolic and mitochondrial mRNA at single-base resolution. Nature 551: 251-255

Schaefer M, Pollex T, Hanna K, Lyko F (2009) RNA cytosine methylation analysis by bisulfite sequencing. Nucleic Acids Res 37: e12

Schumann U, Zhang HN, Sibbritt T, Pan A, Horvath A, Gross S, Clark SJ, Yang L, Preiss T (2020) Multiple links between 5-methylcytosine content of mRNA and translation. BMC Biol 18: 40

Shen Q, Zhang Q, Shi Y, Shi Q, Jiang Y, Gu Y, Li Z, Li X, Zhao K, Wang C, Li N, Cao X (2018) Tet2 promotes pathogen infection-induced myelopoiesis through mRNA oxidation. Nature 554: 123-127

Squires JE, Patel HR, Nousch M, Sibbritt T, Humphreys DT, Parker BJ, Suter CM, Preiss T (2012) Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. Nucleic Acids Res 40: 5023-33

Trixl L, Lusser A (2019) Getting a hold on cytosine methylation in mRNA. Nat Struct Mol Biol 26: 339-340

Wagih O (2017) ggseqlogo: a versatile R package for drawing sequence logos. Bioinformatics 33: 3645-3647

Wheeler EC, Vu AQ, Einstein JM, DiSalvo M, Ahmed N, Van Nostrand EL, Shishkin AA, Jin W, Allbritton NL, Yeo GW (2020) Pooled CRISPR screens with imaging on microraft arrays reveals stress granule-regulatory factors. Nat Methods 17: 636-642

Xiao Y, Wang Y, Tang Q, Wei L, Zhang X, Jia G (2018) An Elongation- and Ligation-Based qPCR Amplification Method for the Radiolabeling-Free Detection of Locus-Specific N(6) -Methyladenosine Modification. Angew Chem Int Ed Engl 57: 15995-16000

Yan X, Stuurman N, Ribeiro SA, Tanenbaum ME, Horlbeck MA, Liem CR, Jost M, Weissman JS, Vale RD (2020) High-content Imaging-based Pooled CRISPR Screens in Mammalian Cells. bioRxiv

Yang X, Yang Y, Sun BF, Chen YS, Xu JW, Lai WY, Li A, Wang X, Bhattarai DP, Xiao W, Sun HY, Zhu Q, Ma HL, Adhikari S, Sun M, Hao YJ, Zhang B, Huang CM, Huang N, Jiang GB et al. (2017a) 5-methylcytosine promotes mRNA export - NSUN2 as the methyltransferase and ALYREF as an m(5)C reader. Cell Res 27: 606-625

Yang X, Yang Y, Sun BF, Chen YS, Xu JW, Lai WY, Li A, Wang X, Bhattarai DP, Xiao W, Sun HY, Zhu Q, Ma HL, Adhikari S, Sun M, Hao YJ, Zhang B, Huang CM, Huang N, Jiang GB et al. (2017b) 5-methylcytosine promotes mRNA export-NSUN2 as the methyltransferase and ALYREF as an m(5)C reader. Cell Res 27: 606-625

Zhao LY, Song J, Liu Y, Song CX, Yi C (2020) Mapping the epigenetic modifications of DNA and RNA. Protein Cell

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.03.324715; this version posted October 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure legends

Figure 1. CIGAR-seq identified NSUN6 as a novel mRNA methyltransferase. A, An illustration of CIGAR-seq method in studying m^5C modification. The WPRE cassette on the original CROP-seq vector was replaced by an endogenous m^5C site with its flanking region. To measure the m^5C level of the reporter site in gRNA sequence-containing transcripts, mRNA was subjected to bisulfite treatment followed by reverse transcription. Then, a primer pair flanking the m^5C site and gRNA sequence was used to amplify the region for subsequent Sanger or next-generation sequencing. **B**, Validation of CIGAR-seq method using a m^5C reporter site derived from FAM129B gene with a control gRNA without target genes and a gRNA targeting NSUN2. Upon NSUN2 knockout, the m^5C modification is diminished in the FAM129B reporter mRNA, whereas the modification remains intact in control knockout cells. **C**, The application of CIGAR-seq in screening for regulators of m^5C sites. Cas9-expressing HAP1 cells were transduced with viral particles that express Cigar vectors combining NSUN2-independent m^5C reporter sites derived from FURIN gene and a gRNA library targeting 829 RBPs. Seven days after transduction, enriched polyA RNA was bisulfite-treated, reveres-transcribed and PCR-amplified using primers flanking the m^5C site and gRNA sequence to generate NGS library. **D**, The rank of gene whose knockout reduced m^5C modification rate of the reporter site (Methods). NSUN6 was the top hit. **E**, Validation of NSUN6 as a mRNA m^5C methyltransferase. Knockout as well as knockdown NSUN6 reduced the m^5C level in both FURIN m^5C reporter transcripts and endogenous NSUN2-independent m^5C sites in RPSA gene.

Figure 2. Global profiling of NSUN6-dependent m^5C sites. A, mRNA bisulfite sequencing revealed NSUN6-dependent m^5C sites in HAP1 cells. Of 208 m^5C sites identified in wildtype cells, 65 showed significantly reduced modification in NSUN6 knocked out cells. X and Y axis represented the modification rate in wildtype and NSUN6 knocked out HAP1 cells, respectively. **B**, The sequence features of NSUN6-dependent m^5C sites in HAP1 cells. A strong 3' TCCA motif was found in NSUN6-dependent sites.

Figure 3. Comparison between NSUN6- and NSUN2-dependent m^5C modification sites. A, The largely non-overlap between NSUN2- and NSUN6-dependent m^5C sites. **B**, Heatmap showing the m^5C modification rate in wildtype, NSUN2-KO, NSUN6-KO and NSUN2/6-dKO HAP1 cells for NSUN6-

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.03.324715; this version posted October 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(upper panel) and NSUN2-dependent sites (lower panel), respectively. **C**, The sequence features of NSUN6-only- (upper panel) and NSUN2-only-dependent m^5C sites (lower panel) in HAP1 cells. While NSUN6-dependent sites were strongly enriched for 3' TCCA motif, NSUN2-dependent sites were enriched for 3' NGGG motif. **D**, The modification rate of 4 groups of m^5C sites that showed different dependence. Comparing to the other three groups, the group of three overlapping sites showed very low m5C level. **E**, Modification rate of NSUN6- and NSUN2-dependent m5C sites across different tissues.

Figure 4. Exemplar application of CIGAR-seq in the study of m^1A modification. A, An illustration of CIGAR-seq vector designed for m^1A modification. A known TRMT6/61A complex-dependent m^1C site in MALAT1 gene was cloned into CIGAR-seq vector, together with control gRNA as well as gRNAs targeting TRMT6 and TRMT61A, respectively. **B&C**, Upon perturbation of either TRMT6 or TRMT61A in HAP1 cells, the m^1A modification was completed abolished in m^1A reporter site, whereas the modification remains intact in control knockout cells.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.03.324715; this version posted October 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure Legends of supplementary figures

Figure EV1. NSUN2 Knock out and the effect on the reporter as well as endogenous m^5C sites. **A**, The NSUN2 KO efficiency in Cas9-expressing HAP1 cells. Seven days after viral transduction, NSUN2 was efficiently mutated. **B**, Scatter plot demonstrating the effect of NSUN2 knocked out in HAP1 cells. 208 m^5C sites were identified in wildtype HAP1 cells, 90 (43.3%) of which showed significantly reduced m^5C level in NSUN2-KO cells. **C**, Distribution of genes with different number of gRNAs detected. **D**, Distribution of m^5C level of the reporter site associated with individual gRNA.

Figure EV2. NSUN6 was efficiently knocked out or knocked down. A, The KO efficiency of NSUN6 in Cas9-expressing HAP1 cells. Seven days after viral transduction, NSUN6 was efficiently mutated. **B**, The knockdown efficiency of NSUN6 in HAP1 cells.

Figure EV3. mRNA-BisSeq profiles of the three m^5C modification sites depend on both NSUN6 and NSUN2. IGV plots showing the m^5C sites in ANGEL1, ZNF707 and STRN4 genes. The m^5C sites in ANGEL1 (**A**) and ZNF707 (**B**) possessed a 3' TCCA and a 3' AGGG motif, respectively, while the m^5C site in STRN4 was among a cluster of "pseudo" m^5C sites (**C**).

Figure EV4. KO of NSUN2 and/or NSUN6 in HAP1 cells. Western Blot demonstrating the effect NSUN2 (**A**), NSUN6 (**B**) as well as NSUN2/6 double knockout (**C**). NSUN2/6 double knockout (NSUN2/6-dKO) HAP1 cells were established based on clonal NSUN6-KO cells.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.03.324715; this version posted October 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

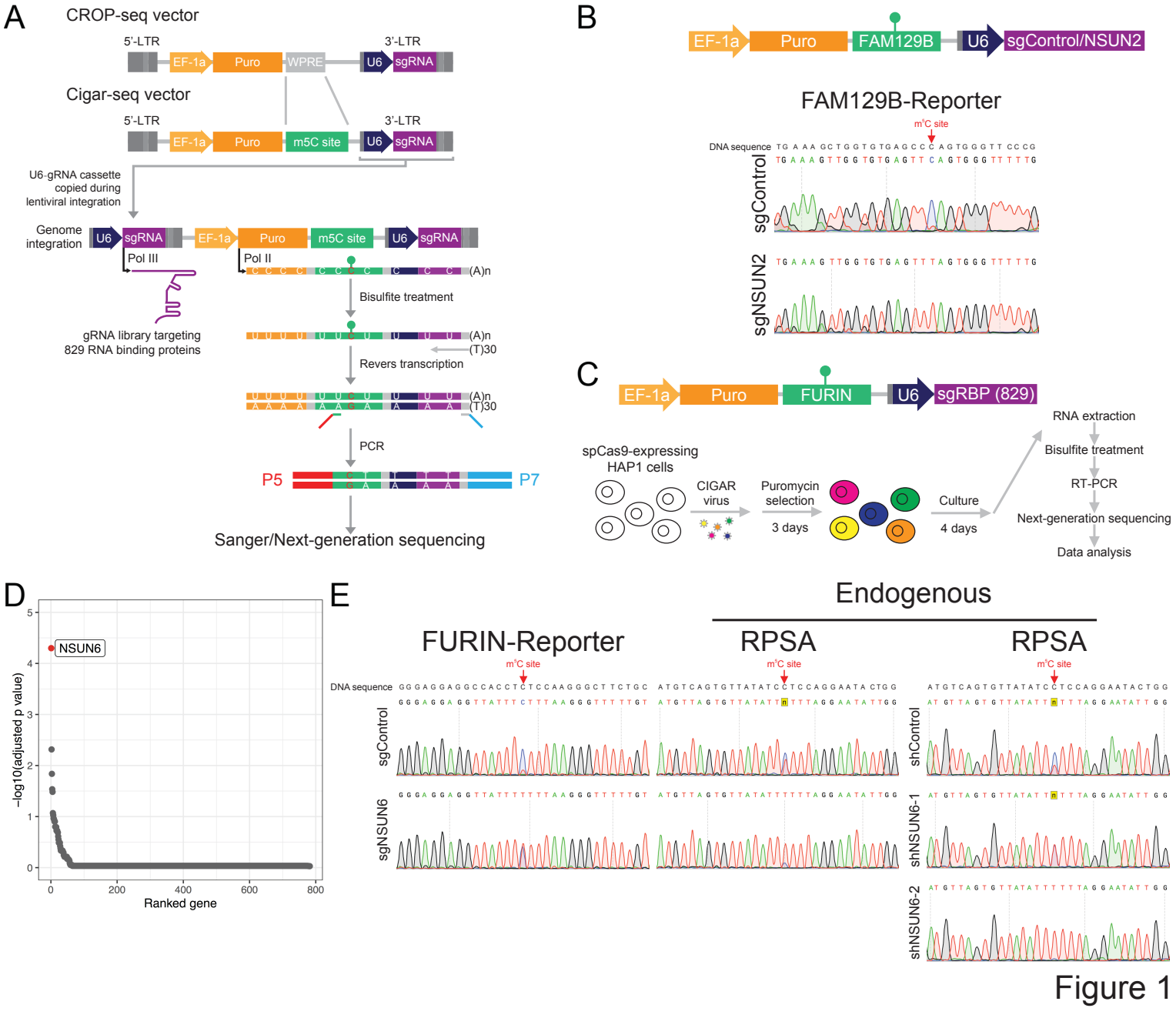
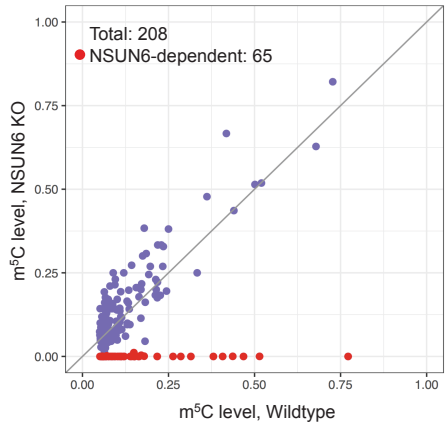


Figure 1

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.03.324715; this version posted October 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

A



B

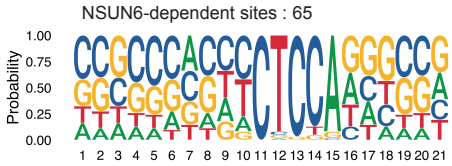


Figure 2

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.03.324715; this version posted October 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

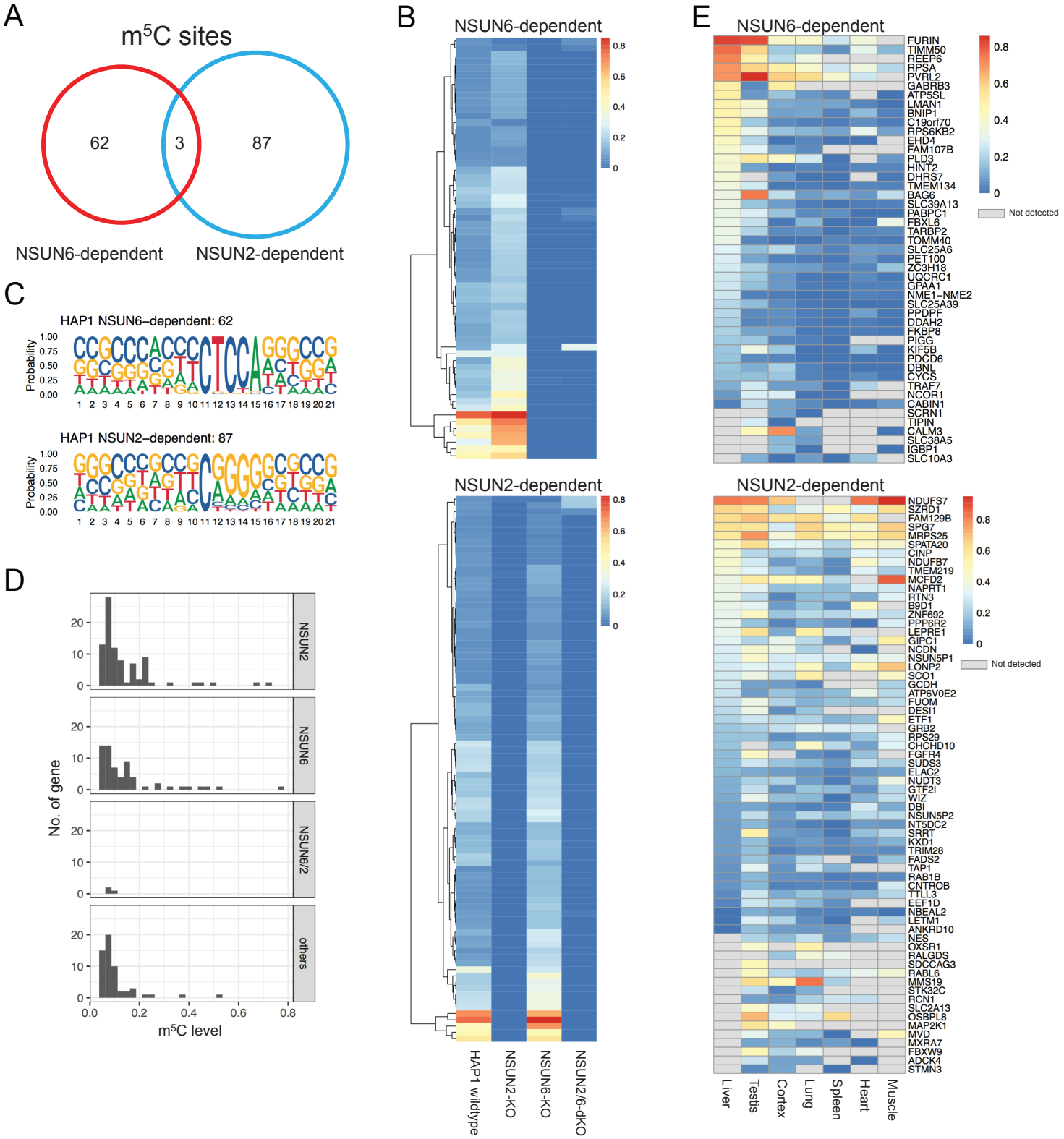


Figure 3

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.03.324715; this version posted October 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

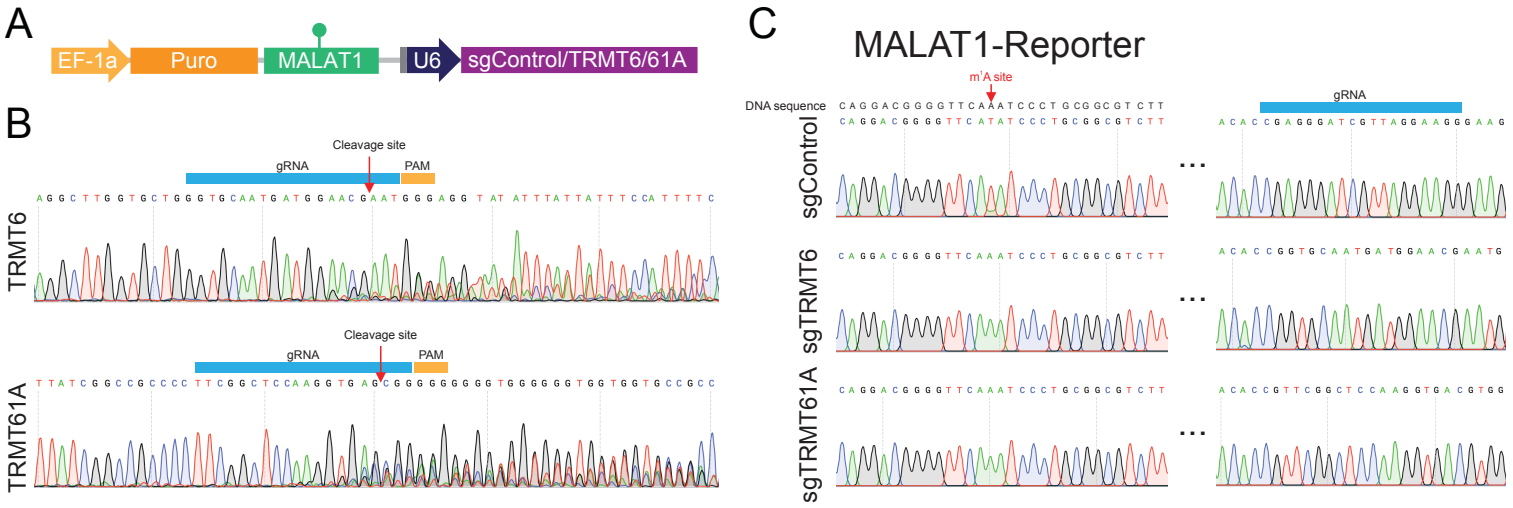


Figure 4

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.03.324715; this version posted October 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

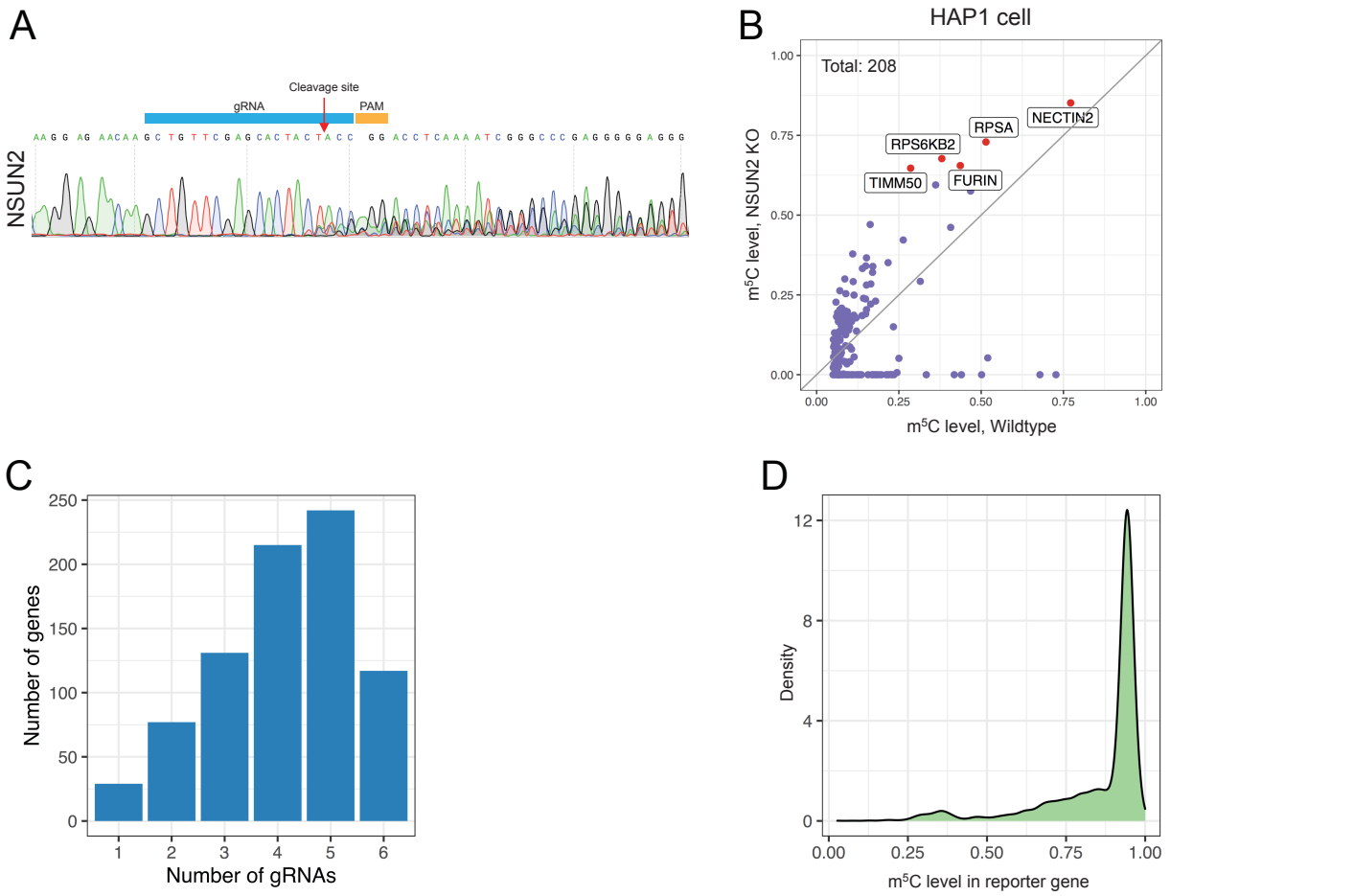


Figure EV1

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.03.324715; this version posted October 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

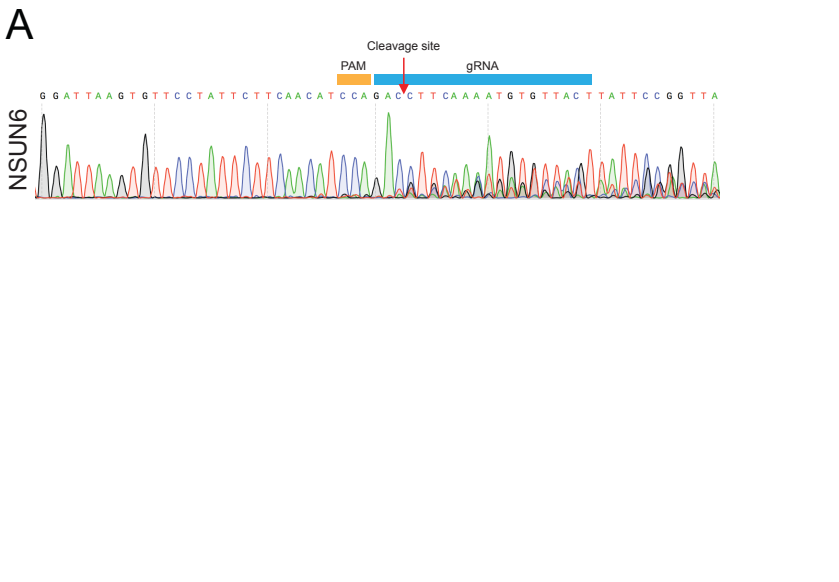


Figure EV2

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.03.324715; this version posted October 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

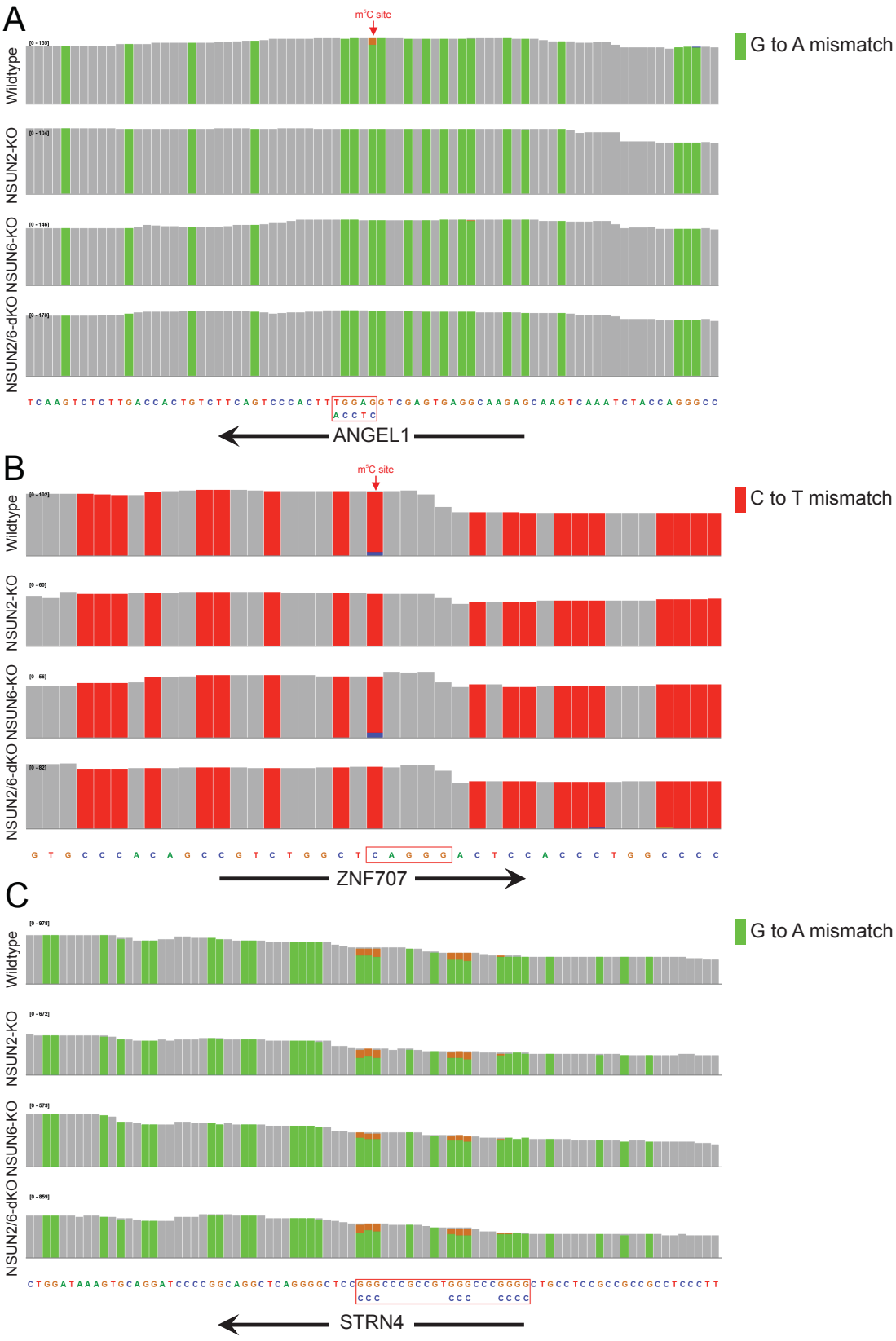


Figure EV3

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.03.324715; this version posted October 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

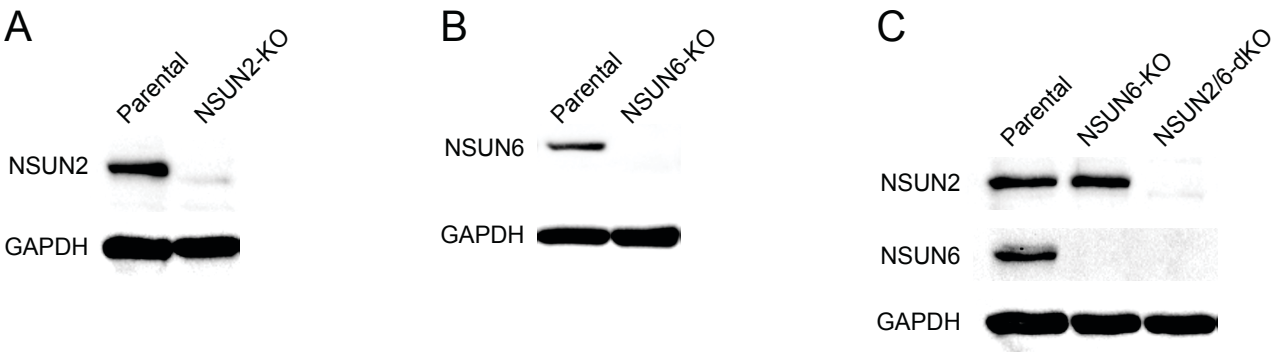


Figure EV4