

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

Liang Fang, Wen Wang, Guipeng Li, Li Zhang, Jun Li, Diwen Gan, Jiao Yang, Yisen Tang, Zewen Ding, Min Zhang, Wenhao Zhang, Daqi Deng, Zhengyu Song, Qionghua Zhu, Huanhuan Cui, Yuhui Hu, and Wei Chen

DOI: [10.15252/msb.202010025](https://doi.org/10.15252/msb.202010025)

Corresponding author(s): *Wei Chen (chenw@sustc.edu.cn)*, *Liang Fang (fangl@sustech.edu.cn)*

Review Timeline:

Submission Date:	25th Sep 20
Editorial Decision:	26th Oct 20
Revision Received:	28th Oct 20
Editorial Decision:	3rd Nov 20
Revision Received:	4th Nov 20
Accepted:	4th Nov 20

Editor: Maria Polychronidou

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. Depending on transfer agreements, referee reports obtained elsewhere may or may not be included in this compilation. Referee reports are anonymous unless the Referee chooses to sign their reports.)

Thank you again for submitting your work to Molecular Systems Biology. We have now heard back from the three referees who agreed to evaluate your study. Overall, the reviewers acknowledge that the proposed method seems interesting. However, they raise a series of concerns, which we would ask you to address in a major revision.

I think that the recommendations of the referees are clear and relatively straightforward to address, and I therefore see no need to repeat any of the points listed below. Please let me know in case you would like to discuss in further detail any of the issues raised. All issues raised by the referees would need to be satisfactorily addressed.

On a more editorial level, we would ask you to address the following.

REFeree REPORTS

Reviewer #1:

The authors developed a novel method named CIGAR-seq to screen mRNA modification enzymes. They selected m5C and m1A as examples. Using this novel CIGAR-seq approach, they identified NSUN6 as mRNA m5C methyltransferase., and NSUN6 and NSUN2 have different subsets of mRNA m5C sites. This work is interesting and provides useful method to study RNA modification. I am happy to recommend its publications after addressing the following minor issues.

- 1- Compared with Bisulfite method, CIGAR-seq might favor to mRNA targets with relative high level of m5C sites, thus I can see why the authors find the numbers of m5C sites are lower than that of previous published m5C sites. This should be included in the discussion.
- 2- Two previous results showed Nsun6 does not affect mRNA m5C level. I believe this could be caused by depletion approaches. Others used knock-down, but this work applied knock-out. The authors are encouraged to discuss this potential difference in the discussion section.
- 3- Yang X. Cell 2017 has been listed twice in the reference list.

Reviewer #2:

Fang et al. developed CIGAR-seq, a method that combines pooled CRISPR screen and reporters with RNA modification readout, and applied that to identify a novel m5C methyltransferase NSUN6. The manuscript is very focused and clearly written, providing strong evidence that NSUN6 is likely to be the only other m5C methyltransferase besides NSUN2 that was previously found. I noticed that there are two other groups that independently identified NSUN6 (published in BioRxiv), which would further validate the findings by this and other groups.

<https://www.biorxiv.org/content/10.1101/2020.10.01.320036v1>

<https://www.biorxiv.org/content/10.1101/2020.10.03.324707v1>

<https://www.biorxiv.org/content/10.1101/2020.10.03.324715v1> (this work)

Although I have some minor concerns listed below, I highly recommend acceptance of this work to be published in MSB.

1. While it is great to identify NSUN6 as the top hit, I am curious about a few other hits shown in Fig. 1D. I do not see the raw data from the screen. What are these hits - are they regulators of NSUN6?
2. The authors mentioned in their manuscript: "However, two previous studies did not show significant m5C changes in mRNA after NSUN6 perturbation in HeLa cells (Huang et al., 2019, Yang et al., 2017a)." This makes me wonder if this is due to technical challenges of bisulfite sequencing or biological differences in different cell types. I would like to see some experimental confirmation, and minimally some discussion of speculations.

Reviewer #3:

This is an interesting manuscript that describes an approach for using CRISPR to discover the enzyme that mediates the formation of a nucleotide modification within a given selected sequence. This approach is applied to a variety of sequences that are known to contain m5C in mRNA. For a site that is mediated by the enzyme NSUN2, the new CRISPR approach was able to successfully predict this enzyme. However for another site which contains m5C mediated by an unknown enzyme, this study may be important discovery that this was mediated by NSUN6. This is a new finding because this enzyme has not been previously linked as mediating m5C sites for epitranscriptome. Frankly, this entire manuscript could have been written from the perspective of discovering this new enzyme. Regardless, there are important basic science implications by discovering this enzyme in addition to the new method that has been described here.

There are limitations to this method. Most importantly, the modified nucleotide has to be able to incorporate or introduce a signature mutation. This is not always the case. Secondly, in most cases, it's pretty clear which enzyme makes the modified nucleotide. There are some exceptions, and the authors have clearly identified a good example.

Overall, this is nicely done, and the subsequent experiments to globally assess which m5C sites performed by which enzyme were nice. I think this is overall a good study. I have no major comments.

I have the following minor comments:

1. METTL14 is not a methyltransferase. The authors refer to METTL14 as a methyltransferase as proposed by He and colleagues (Liu et al 2013, doi: 10.1038/nchembio.1432). However it is now known that these METTL14 preparations prepared in insect cells were contaminated with insect METTL3 (see DOI: 10.1016/j.molcel.2016.05.041). Thus METTL14 is not an enzyme. The better reference would be Wang et al. 2014 (<https://doi.org/10.1038/ncb2902>). This paper was published at the same time as the earlier study and correctly proposed that METTL14 was part of a complex with METTL3, and the complex mediates m6A formation.
2. There are some spelling errors especially the word reverse is spelled incorrectly several times.

Point-by-point response to the referees' comments**General response to the reviewers:**

We thank the three reviewers for their time and appreciate their constructive comments. During the revision, we have carefully addressed all of them, as shown below marked in red.

Reviewer #1 (Comments to the Author):

The authors developed a novel method named CIGAR-seq to screen mRNA modification enzymes. They selected m5C and m1A as examples. Using this novel CIGAR-seq approach, they identified NSUN6 as mRNA m5C methyltransferase, and NSUN6 and NSUN2 have different subsets of mRNA m5C sites. This work is interesting and provides useful method to study RNA modification. I am happy to recommend its publications after addressing the following minor issues.

1-Compared with Bisulfite method, CIGAR-seq might favor to mRNA targets with relative high level of m5C sites, thus I can see why the authors find the numbers of m5C sites are lower than that of previous published m5C sites. This should be included in the discussion.

Answer:

Indeed, in the first part of our study, to ensure a high sensitivity in identifying the enzyme mediating m⁵C modification, we intentionally selected a m⁵C site of high modification rate to construct the epitranscriptomic reporter.

In the second part of our study, to characterize the NSUN6- and NSUN2-in/dependent m⁵C sites, we set stringent criteria to avoid potential false discovery of m⁵C sites due to incomplete bisulfite conversion. For this purpose, a total of five replicates of mRNA-BisSeq data were generated from the WT HAP1 cells. The high-confident m⁵C sites were determined based on: (1) coverage of the site being at least 20 reads in all five replicates; (2) number of reads containing the unmodified C being at least 2 in all five replicates; (3) the WT methylation level (the minimum methylation level from the five replicates) being at least 0.05 (described in Methods). Whereas such stringent criteria would assure the high specificity of our findings, we would likely also miss some of the true m⁵C sites, particularly those with low modification rate. We added a discussion about this in the revised manuscript (line 183-186).

2-Two previous results showed Nsun6 does not affect mRNA m5C level. I believe this could be caused by depletion approaches. Others used knock-down, but this work applied knock-out. The authors are encouraged to discuss this potential difference in the discussion section.

Answer:

Indeed, we think likewise. 'Huang et al., 2019' used siRNA to knockdown NSUN6 to ~40% of its original level (shown by RT-qPCR and Western blotting, Fig S5 b&c), which might not be sufficient to cause global m5C change. 'Yang et al., 2017a' also used siRNA to knockdown NSUN6 (verified by RT-PCR, Fig S3d), and induced slight, though not significant, global m5C reduction (Fig S3b). In comparison, we used CRISPR/Cas-mediated knockout to completely remove NSUN6, which is much more sensitive in

revealing the gene function. We have discussed the potential cause of this discrepancy in the revised manuscript (line 130).

We also noted the same group, who published 'Huang et al., 2019', now also identified NSUN6 as m⁵C methyltransferase (<https://www.biorxiv.org/content/10.1101/2020.10.03.324707v1>).

3-Yang X. Cell 2017 has been listed twice in the reference list.

Answer:

Thanks for pointing this out. We corrected it accordingly.

Reviewer #2 (Comments to the Author):

Fang et al. developed CIGAR-seq, a method that combines pooled CRISPR screen and reporters with RNA modification readout, and applied that to identify a novel m⁵C methyltransferase NSUN6. The manuscript is very focused and clearly written, providing strong evidence that NSUN6 is likely to be the only other m⁵C methyltransferase besides NSUN2 that was previously found. I noticed that there are two other groups that independently identified NSUN6 (published in BioRxiv), which would further validate the findings by this and other groups.

<https://www.biorxiv.org/content/10.1101/2020.10.01.320036v1>

<https://www.biorxiv.org/content/10.1101/2020.10.03.324707v1>

<https://www.biorxiv.org/content/10.1101/2020.10.03.324715v1> (this work)

Although I have some minor concerns listed below, I highly recommend acceptance of this work to be published in MSB.

1. While it is great to identify NSUN6 as the top hit, I am curious about a few other hits shown in Fig. 1D. I do not see the raw data from the screen. What are these hits - are they regulators of NSUN6?

Answer:

We have supplemented a list of candidate genes in Dataset EV2. Among all the RBPs we screened, only NSUN6 showed very significant P-value ($-\log_{10}(\text{P-value}) = 4.3$, two orders of magnitude lower than the second-ranked gene), and all six gRNAs targeting NSUN6 reduced m⁵C level effectively. Besides NSUN6, we selected ZCCHC11 (3rd-ranked) and EIF3J (5th-ranked) with marginally significant P-value ($-\log_{10}(\text{P-value}) = 1.8$ and 1.5, respectively) and multiple gRNAs (six and five, respectively) showing effect on inhibiting m⁵C modification for validation. However, knockout of either ZCCHC11 or EIF3J did not reduce the m⁵C modification at all. We described this result in the revised manuscript (line 143-146).

Please note that we did not choose 2nd- and 4th-ranked gene for validation, because both of them only had two out of six gRNAs showing effect on decreasing m⁵C level.

2. The authors mentioned in their manuscript: "However, two previous studies did not show significant m5C changes in mRNA after NSUN6 perturbation in HeLa cells (Huang et al., 2019, Yang et al., 2017a)." This makes me wonder if this is due to technical challenges of bisulfite sequencing or biological differences in different cell types. I would like to see some experimental confirmation, and minimally some discussion of speculations.

Answer:

We think the discrepancy between our study and other two is mainly due to the different gene silencing methods. Please see the address to Reviewer #1.

Reviewer #3 (Comments to the Author):

This is an interesting manuscript that describes an approach for using CRISPR to discover the enzyme that mediates the formation of a nucleotide modification within a given selected sequence. This approach is applied to a variety of sequences that are known to contain m5C in mRNA. For a site that is mediated by the enzyme NSUN2, the new CRISPR approach was able to successfully predict this enzyme. However, for another site which contains m5C mediated by an unknown enzyme, this study may be important discovery that this was mediated by NSUN6. This is a new finding because this enzyme has not been previously linked as mediating m5C sites for epitranscriptome. Frankly, this entire manuscript could have been written from the perspective of discovering this new enzyme. Regardless, there are important basic science implications by discovering this enzyme in addition to the new method that has been described here.

There are limitations to this method. Most importantly, the modified nucleotide has to be able to incorporate or introduce a signature mutation. This is not always the case. Secondly, in most cases, it's pretty clear which enzyme makes the modified nucleotide. There are some exceptions, and the authors have clearly identified a good example.

Overall, this is nicely done, and the subsequent experiments to globally assess which m5C sites performed by which enzyme were nice. I think this is overall a good study. I have no major comments.

I have the following minor comments:

1. METTL14 is not a methyltransferase. The authors refer to METTL14 as a methyltransferase as proposed by He and colleagues (Liu et al 2013, doi: 10.1038/nchembio.1432). However it is now known that these METTL14 preparations prepared in insect cells were contaminated with insect METTL3 (see DOI: 10.1016/j.molcel.2016.05.041). Thus METTL14 is not an enzyme. The better reference would be Wang et al. 2014 (<https://doi.org/10.1038/ncb2902>). This paper was published at the same time as the earlier study and correctly proposed that METTL14 was part of a complex with METTL3, and the complex mediates m6A formation.

Answer:

Thanks for pointing out our mistake. We have rewritten this part and cited the correct reference (Line 42-43).

2. There are some spelling errors especially the word reverse is spelled incorrectly several times.

Answer:

Thanks for pointing this out. We corrected them accordingly.

Thank you for sending us your revised manuscript. We think that the performed revisions have satisfactorily addressed the issues raised by the reviewers. I am therefore glad to inform you that we can soon formally accept the study for publication, pending some minor issues listed below:

2nd Authors' Response to Reviewers**4th Nov 2020**

The authors have made all the requested editorial changes.

Accepted**4th Nov 2020**

Thank you for performing these last few requested edits. We are now satisfied with the modifications made and I am pleased to inform you that your paper has been accepted for publication.

YOU MUST COMPLETE ALL CELLS WITH A PINK BACKGROUND ↓

PLEASE NOTE THAT THIS CHECKLIST WILL BE PUBLISHED ALONGSIDE YOUR PAPER

Corresponding Author Name: Wei Chen

Journal Submitted to: MSB

Manuscript Number: MSB-20-10025

Reporting Checklist For Life Sciences Articles (Rev. June 2017)

This checklist is used to ensure good reporting standards and to improve the reproducibility of published results. These guidelines are consistent with the Principles and Guidelines for Reporting Preclinical Research issued by the NIH in 2014. Please follow the journal's authorship guidelines in preparing your manuscript.

A- Figures

1. Data

The data shown in figures should satisfy the following conditions:

- the data were obtained and processed according to the field's best practice and are presented to reflect the results of the experiments in an accurate and unbiased manner.
- Figure panels include only data points, measurements or observations that can be compared to each other in a scientifically meaningful way.
- graphs include clearly labeled error bars for independent experiments and sample sizes. Unless justified, error bars should not be shown for technical replicates.
- if $n < 5$, the individual data points from each experiment should be plotted and any statistical test employed should be justified
- Source Data should be included to report the data underlying graphs. Please follow the guidelines set out in the author ship guidelines on Data Presentation.

2. Captions

Each figure caption should contain the following information, for each panel where they are relevant:

- a specification of the experimental system investigated (eg cell line, species name).
- the assay(s) and method(s) used to carry out the reported observations and measurements
- an explicit mention of the biological and chemical entity(ies) that are being measured.
- an explicit mention of the biological and chemical entity(ies) that are altered/varied/perturbed in a controlled manner.
- the exact sample size (n) for each experimental group/condition, given as a number, not a range;
- a description of the sample collection allowing the reader to understand whether the samples represent technical or biological replicates (including how many animals, litters, cultures, etc.).
- a statement of how many times the experiment shown was independently replicated in the laboratory.
- definitions of statistical methods and measures:
 - common tests, such as t-test (please specify whether paired vs. unpaired), simple χ^2 tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods section;
 - are tests one-sided or two-sided?
 - are there adjustments for multiple comparisons?
 - exact statistical test results, e.g., P values = x but not P values < x;
 - definition of 'center values' as median or average;
 - definition of error bars as s.d. or s.e.m.

Any descriptions too long for the figure legend should be included in the methods section and/or with the source data.

In the pink boxes below, please ensure that the answers to the following questions are reported in the manuscript itself. Every question should be answered. If the question is not relevant to your research, please write NA (non applicable). We encourage you to include a specific subsection in the methods section for statistics, reagents, animal models and human subjects.

B- Statistics and general methods

Please fill out these boxes ↓ (Do not worry if you cannot see all your text once you press return)

1.a. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size?	To determine a set of high-confident m5C 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 m5C sites were: (1) coverage of the site being at least 20 reads in all five replicates; (2) number of reads containing the unmodified C being at least 2 in all five replicates; (3) the WT methylation level (the minimum methylation level from the five replicates) being at least 0.05.
1.b. For animal studies, include a statement about sample size estimate even if no statistical methods were used.	NA
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established?	No sample was excluded.
3. Were any steps taken to minimize the effects of subjective bias when allocating animals/samples to treatment (e.g. randomization procedure)? If yes, please describe.	NA
For animal studies, include a statement about randomization even if no randomization was used.	NA
4.a. Were any steps taken to minimize the effects of subjective bias during group allocation or/and when assessing results (e.g. blinding of the investigator)? If yes please describe.	NA
4.b. For animal studies, include a statement about blinding even if no blinding was done	NA
5. For every figure, are statistical tests justified as appropriate?	Yes
Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it.	NA

USEFUL LINKS FOR COMPLETING THIS FORM

<http://www.antibodypedia.com>
<http://1degreebio.org>
<http://www.equator-network.org/reporting-guidelines/improving-bioscience-research-reporting>
<http://grants.nih.gov/grants/olaw/olaw.htm>
<http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Useofanimals/index.htm>
<http://ClinicalTrials.gov>
<http://www.consort-statement.org>
<http://www.consort-statement.org/checklists/view/32-consort/66-title>
<http://www.equator-network.org/reporting-guidelines/reporting-recommendations-for-tumour-research>
<http://datadrivad.org>
<http://figshare.com>
<http://www.ncbi.nlm.nih.gov/gap>
<http://www.ebi.ac.uk/ega>
<http://biomodels.net/>
<http://biomodels.net/miriam/>
<http://ijb.biochem.sun.ac.za>
http://oba.od.nih.gov/biosecurity/biosecurity_documents.html
<http://www.selectagents.gov/>

Is there an estimate of variation within each group of data?	NA
Is the variance similar between the groups that are being statistically compared?	NA

C- Reagents

6. To show that antibodies were profiled for use in the system under study (assay and species), provide a citation, catalog number and/or clone number, supplementary information or reference to an antibody validation profile. e.g., Antibodypedia (see link list at top right), 1DegreeBio (see link list at top right).	The catalog numbers were provided in the Materials and Methods.
7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination.	HAP1 and HEK-293T cell was obtained from Horizon discovery and ATCC, respectively, and tested negative for mycoplasma contamination. No authentication was done.

* for all hyperlinks, please see the table at the top right of the document

D- Animal Models

8. Report species, strain, gender, age of animals and genetic modification status where applicable. Please detail housing and husbandry conditions and the source of animals.	NA
9. For experiments involving live vertebrates, include a statement of compliance with ethical regulations and identify the committee(s) approving the experiments.	NA
10. We recommend consulting the ARRIVE guidelines (see link list at top right) (PLoS Biol. 8(6), e1000412, 2010) to ensure that other relevant aspects of animal studies are adequately reported. See author guidelines, under 'Reporting Guidelines'. See also: NIH (see link list at top right) and MRC (see link list at top right) recommendations. Please confirm compliance.	NA

E- Human Subjects

11. Identify the committee(s) approving the study protocol.	NA
12. Include a statement confirming that informed consent was obtained from all subjects and that the experiments conformed to the principles set out in the WMA Declaration of Helsinki and the Department of Health and Human Services Belmont Report.	NA
13. For publication of patient photos, include a statement confirming that consent to publish was obtained.	NA
14. Report any restrictions on the availability (and/or on the use) of human data or samples.	NA
15. Report the clinical trial registration number (at ClinicalTrials.gov or equivalent), where applicable.	NA
16. For phase II and III randomized controlled trials, please refer to the CONSORT flow diagram (see link list at top right) and submit the CONSORT checklist (see link list at top right) with your submission. See author guidelines, under 'Reporting Guidelines'. Please confirm you have submitted this list.	NA
17. For tumor marker prognostic studies, we recommend that you follow the REMARK reporting guidelines (see link list at top right). See author guidelines, under 'Reporting Guidelines'. Please confirm you have followed these guidelines.	NA

F- Data Accessibility

18. Provide a "Data Availability" section at the end of the Materials & Methods, listing the accession codes for data generated in this study and deposited in a public database (e.g. RNA-Seq data: Gene Expression Omnibus GSE39462, Proteomics data: PRIDE PXD000208 etc.) Please refer to our author guidelines for 'Data Deposition'. Data deposition in a public repository is mandatory for: a. Protein, DNA and RNA sequences b. Macromolecular structures c. Crystallographic data for small molecules d. Functional genomics data e. Proteomics and molecular interactions	All next-generation sequencing data were submitted to Gene Expression Omnibus under the accession number GSE157368 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE157368).
19. Deposition is strongly recommended for any datasets that are central and integral to the study; please consider the journal's data policy. If no structured public repository exists for a given data type, we encourage the provision of datasets in the manuscript as a Supplementary Document (see author guidelines under 'Expanded View' or in unstructured repositories such as Dryad (see link list at top right) or Figshare (see link list at top right)).	All next-generation sequencing data were submitted to Gene Expression Omnibus under the accession number GSE157368 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE157368).
20. Access to human clinical and genomic datasets should be provided with as few restrictions as possible while respecting ethical obligations to the patients and relevant medical and legal issues. If practically possible and compatible with the individual consent agreement used in the study, such data should be deposited in one of the major public access-controlled repositories such as dbGAP (see link list at top right) or EGA (see link list at top right).	NA
21. Computational models that are central and integral to a study should be shared without restrictions and provided in a machine-readable form. The relevant accession numbers or links should be provided. When possible, standardized format (SBML, CellML) should be used instead of scripts (e.g. MATLAB). Authors are strongly encouraged to follow the MIRIAM guidelines (see link list at top right) and deposit their model in a public database such as Biomodels (see link list at top right) or JWS Online (see link list at top right). If computer source code is provided with the paper, it should be deposited in a public repository or included in supplementary information.	NA

G- Dual use research of concern

22. Could your study fall under dual use research restrictions? Please check biosecurity documents (see link list at top right) and list of select agents and toxins (APHIS/CDC) (see link list at top right). According to our biosecurity guidelines, provide a statement only if it could.	NA
---	----