



RESEARCH ARTICLE

Identification of putative reader proteins of 5-methylcytosine and its derivatives in *Caenorhabditis elegans* RNA [version 1; peer review: 1 approved, 2 approved with reservations]

IC Navarro ^{1,2*}, Kin Man Suen ^{1,3*}, Dalila Bensaddek^{4,5}, Arun Tanpure⁶, Angus Lamond ⁴, Shankar Balasubramanian ⁶⁻⁸, Eric A Miska ^{1,2,9}

¹Gurdon Institute, University of Cambridge, Cambridge, CB2 1QN, UK

²Department of Genetics, University of Cambridge, Cambridge, CB2 3EH, UK

³School of Molecular and Cellular Biology, University of Leeds, LC Miall Building, Leeds, LS2 9JT, UK

⁴Laboratory for Quantitative Proteomics, Centre for Gene Regulation and Expression, College of Life Sciences, University of Dundee, Dow Street, Dundee, DD1 5EH, UK

⁵Bioscience Core Labs, King Abdullah University of Science and Technology, Thuwal, 23955-6900, Saudi Arabia

⁶Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, UK

⁷Cancer Research (UK), Cambridge Institute, Li Ka Shing Centre, University of Cambridge, Robinson Way, Cambridge, CB2 0RE, UK

⁸School of Clinical Medicine, University of Cambridge, Cambridge, CB2 0SP, UK

⁹Wellcome Sanger Institute, Wellcome Genome Campus, Cambridge, CB10 1SA, UK

* Equal contributors

V1 First published: 17 Nov 2022, 7:282
<https://doi.org/10.12688/wellcomeopenres.17893.1>
 Latest published: 17 Nov 2022, 7:282
<https://doi.org/10.12688/wellcomeopenres.17893.1>

Abstract

Background: Methylation of carbon-5 of cytosines (m^5C) is a conserved post-transcriptional nucleotide modification of RNA with widespread distribution across organisms. It can be further modified to yield 5-hydroxymethylcytidine (hm^5C), 5-formylcytidine (f^5C), 2'-O-methyl-5-hydroxymethylcytidine (hm^5Cm) and 2'-O-methyl-5-formylcytidine (f^5Cm). How m^5C , and specially its derivatives, contribute to biology mechanistically is poorly understood. We recently showed that m^5C is required for *Caenorhabditis elegans* development and fertility under heat stress. m^5C has been shown to participate in mRNA transport and maintain mRNA stability through its recognition by the reader proteins ALYREF and YBX1, respectively. Hence, identifying readers for RNA modifications can enhance our understanding in the biological roles of these modifications.

Methods: To contribute to the understanding of how m^5C and its oxidative derivatives mediate their functions, we developed RNA baits bearing modified cytosines in diverse structural contexts to pulldown potential readers in *C. elegans*. Potential readers were identified using mass spectrometry. The interaction of two of the putative readers with m^5C was validated using immunoblotting.

Results: Our mass spectrometry analyses revealed unique binding

Open Peer Review

Approval Status ? ✓ ?

	1	2	3
version 1 17 Nov 2022	? view	✓ view	? view

1. **Matthias Ronald Schaefer** , Medical University of Vienna, Vienna, Austria
2. **Carolyn Phillips** , University of Southern California, Los Angeles, USA
3. **Anna Y Zinovyeva** , Kansas State University, Manhattan, USA
Jeffrey Medley, Kansas State University, Manhattan, USA

Any reports and responses or comments on the article can be found at the end of the article.

proteins for each of the modifications. *In silico* analysis for phenotype enrichments suggested that hm⁵Cm unique readers are enriched in proteins involved in RNA processing, while readers for m⁵C, hm⁵C and f⁵C are involved in germline processes. We validated our dataset by demonstrating that the nematode ALYREF homologues ALY-1 and ALY-2 preferentially bind m⁵C *in vitro*. Finally, sequence alignment analysis showed that several of the putative m⁵C readers contain the conserved RNA recognition motif (RRM), including ALY-1 and ALY-2.

Conclusions: The dataset presented here serves as an important scientific resource that will support the discovery of new functions of m⁵C and its derivatives. Furthermore, we demonstrate that ALY-1 and ALY-2 bind to m⁵C in *C. elegans*.

Keywords

RNA, 5-methylcytosine, *Caenorhabditis elegans*, epitranscriptomics, readers

Corresponding author: Eric A Miska (eam29@cam.ac.uk)

Author roles: **Navarro I:** Conceptualization, Formal Analysis, Investigation, Methodology, Project Administration, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Suen KM:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Resources, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Bensaddek D:** Data Curation, Investigation, Methodology, Resources, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Tanpure A:** Investigation, Methodology, Resources, Writing – Review & Editing; **Lamond A:** Conceptualization, Funding Acquisition, Project Administration, Supervision, Writing – Review & Editing; **Balasubramanian S:** Conceptualization, Funding Acquisition, Supervision, Writing – Review & Editing; **Miska EA:** Conceptualization, Funding Acquisition, Resources, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: E.A.M. is a co-founder and director of Storm Therapeutics, Cambridge, UK.

Grant information: This work was supported by Wellcome [104640, <https://doi.org/10.35802/104640>; a Senior Investigator award to E.M.] and [209441, <https://doi.org/10.35802/209441>; a Senior Investigator award to S.B.]; a Cancer Research UK award (C13474/A27826) to E.M.; Conselho Nacional de Desenvolvimento Científico e Tecnológico doctorate scholarship (CNPq, Brazil - 205589/2014-6) to I.C.N. The Miska laboratory is supported by core funding from Wellcome [092096; 203144] and Cancer Research UK (C6946/A24843). The Balasubramanian laboratory is supported by Cancer Research UK core (C9545/A19836) and programme award funding (C9681/A29214), and Herchel Smith Funds. Tagged ALY-1 and ALY-2 strains were generated by the TransgeneOme project and provided by the *Caenorhabditis* Genetics Center – the latter is funded by NIH Office of Research Infrastructure Programs (P40 OD010440). *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

Copyright: © 2022 Navarro I *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Navarro I, Suen KM, Bensaddek D *et al.* **Identification of putative reader proteins of 5-methylcytosine and its derivatives in *Caenorhabditis elegans* RNA [version 1; peer review: 1 approved, 2 approved with reservations]** Wellcome Open Research 2022, 7:282 <https://doi.org/10.12688/wellcomeopenres.17893.1>

First published: 17 Nov 2022, 7:282 <https://doi.org/10.12688/wellcomeopenres.17893.1>

Introduction

Nucleoside chemical modifications are a common feature in RNA molecules. More than 160 post-transcriptional modifications have been reported since the discovery of pseudouridine in 1957 (Boccaletto *et al.*, 2018; Cohn, 1960; Davis & Allen, 1957). However, the molecular and physiological functions of most RNA modifications remain unknown.

Discovered in 1958 (Amos & Korn, 1958), the methylation of carbon-5 of cytosines in RNAs (m⁵C) is now known to be a conserved and widely distributed feature in biological systems. m⁵C is catalysed by tRNA aspartic acid MTase 1 (DNMT2) and RNA methyltransferases from the Nop2/Sun domain family (NSUN1-7 in humans), and has been shown to be involved in tRNA stability, ribosome fidelity and translation efficiency (reviewed in García-Vílchez *et al.*, 2019). Previous studies showed that m⁵C is subject to hydroxylation and oxidation by alpha-ketoglutarate-dependent dioxygenase ABH (ALKBH1), forming 5-hydroxymethylcytidine (hm⁵C) and 5-formylcytidine (f⁵C) (Huber *et al.*, 2015; Kawarada *et al.*, 2017). Further processing can occur through 2'-O methylation by FtsJ RNA Methyltransferase Homolog 1 (FTSJ1), yielding 2'-O-methyl-5-hydroxymethylcytidine (hm⁵Cm) and 2'-O-methyl-5-formylcytidine (f⁵Cm) (Huber *et al.*, 2017; Kawarada *et al.*, 2017; País de Barros *et al.*, 1996). While the biogenesis of these chemical marks has been elucidated (Kawarada *et al.*, 2017; Navarro *et al.*, 2020), their functions remain largely unexplored.

The identification of readers, i.e. proteins that specifically, or preferentially, interact with RNA molecules bearing a certain modification, has proven to be an important step towards the understanding of a modification's biological function. For example, the discovery of N⁶-methyladenosine (m⁶A) readers added new layers of complexity to this pathway, revealing context- and stimuli-dependent functions previously not appreciated (Shi *et al.*, 2019). To date, only two reader proteins have been identified for m⁵C. The Aly/REF export factor (ALYREF) was identified in an RNA pulldown experiment and shown to interact with NSUN2-modified CG-rich regions and regions immediately downstream of translation initiation sites. NSUN2-mediated mRNA methylation is interpreted by binding of ALYREF to modulate nuclear-cytoplasm shuttling of transcripts. (Yang *et al.*, 2017). More recently, the Y-box binding protein 1 (YBX1) was shown to recognise m⁵C in mRNAs and enhance its stability through the recruitment of ELAV1 in human urothelial carcinoma cells (Chen *et al.*, 2019). In zebrafish, YBX1 was shown to stabilise m⁵C-modified transcripts via Pabpc1a during the maternal-to-zygotic transition (Yang *et al.*, 2019).

Recently, we used *Caenorhabditis elegans* as a model to engineer the first organism completely devoid of m⁵C in RNA. We mapped m⁵C with single nucleotide resolution in different RNA species and found that m⁵C is required for physiological adaptation and translation efficiency at high temperatures (Navarro *et al.*, 2020). Here, to further expand our knowledge on the m⁵C pathway in this organism, we produced the first list of putative readers of m⁵C and its metabolic derivatives

hm⁵C, hm⁵Cm and f⁵C. The putative readers are enriched in proteins with roles in germline development and RNA surveillance. We show that the *C. elegans* ALYREF homologues ALY-1 and ALY-2 preferentially bind m⁵C *in vitro* and identify other proteins that contain a conserved RNA recognition motif potentially involved in m⁵C binding.

Methods

Genetics

C. elegans strains were grown and maintained on NGM plates (3 g/L NaCl, 17 g/L agar, 2.5 g/L peptone, 1mM CaCl₂, 5 µg/mL cholesterol, 1mM MgSO₄, 25 mM KPO₄) seeded with *Escherichia coli* and transferred to fresh plates regularly to avoid starvation (Brenner, 1974). The strains were kept at 20°C, unless otherwise indicated. HB101 strain *E. coli* was grown in B broth (10 g/L Bacto-tryptone, 5 g/L NaCl) and used as food source (*Caenorhabditis* Genetics Center, University of Minnesota, Twin Cities, MN, USA). Bristol N2 was used as the wild type strain.

Oligonucleotide synthesis

The synthesis of 2'-O-methyl-5-hydroxymethylcytidine (hm⁵Cm), 5-hydroxymethylcytidine (hm⁵C) and 5-formylcytidine (f⁵C) phosphoramidite monomers was performed as described in Tanpure & Balasubramanian, 2017.

RNA oligonucleotides (5'-GCXUCCGAUGXUACGGAG-GCUGAXC-biotin-3', where X = C, m⁵C, hm⁵C, f⁵C or hm⁵Cm) were synthesised in collaboration with ATDBio, Southampton, UK. Purity and integrity of all modified RNA oligonucleotides were confirmed by LC-MS analysis.

Protein extraction

Nematodes were grown until gravid adults in 140 mm NGM plates seeded with concentrated HB101 *E. coli* and then harvested and washed twice in M9 buffer. A final wash was performed in lysis buffer (20 mM HEPES pH 7.5, 150 mM KCl, 1.5 mM MgCl₂, 0.1% IGEPAL, 0.5 mM DTT) and the animals were resuspended in 4 ml of lysis buffer before the addition of a protease inhibitor cocktail (SIGMAFAST protease inhibitor tablets, Sigma). The animals were pelleted by centrifugation at 2,000 rpm for 2 min (Eppendorf 5810) and the supernatant was removed. Samples were frozen in liquid nitrogen drop-wise, using a Pasteur pipette. Frozen droplets were grinded to powder in metallic capsules for 25 sec in a mixer (Retsch MM 400 Mixer Mill). The powder was stored at -80°C until required. The powder was defrosted at 4°C and the lysate was sonicated for 10 cycles of 20 sec, with breaks of 20 sec in between, in a Bioruptor Pico (Diagenode). The sample was centrifuged in an Eppendorf 5430R microcentrifuge at maximum speed for 15 min at 4°C and the protein concentration of the supernatant was determined using a Bio-Rad Bradford protein assay (Bio-Rad 5000006) according to the manufacturer's instructions.

RNA pulldown

RNA pulldown protocol was adapted from Dominissini *et al.*, 2012. Magnetic beads (Dynabeads MyOne Streptavidin C1, Invitrogen) were washed twice in 100 mM Tris-HCl pH 7.5,

10 mM EDTA, 1 M NaCl, and 0.1% Tween-20, followed by two washes in lysis buffer. In order to reduce non-specific protein binding to the RNA bait matrix, lysates were pre-cleared by incubation with washed beads for 1 h at 4°C with constant rotation prior to the pulldown. Finally, the following pulldown mix was prepared: 2 mg of protein lysate, 50 U of RNase inhibitor Superase-IN (Invitrogen), 500 pmol of 25-mer biotinylated RNA bait, up to a final volume of 250 µl in lysis buffer. The mixture was incubated with constant rotation for 2 h at 4°C and then added to 20 µl of streptavidin-conjugated beads for pulldown for another 2 h at 4°C. The beads were washed 3–5 times in 1 ml of lysis buffer and the bound proteins were eluted for mass spectrometry in 50 µl 4% SDS 100 mM TEAB at 95°C for 10 min.

Sample preparation for protein mass spectrometry

Samples were processed and analysed as described in [Bensaddek et al., 2016](#). Briefly, samples were reduced using 25 mM tris-carboxyethylphosphine TCEP for 30 min at room temperature, then alkylated in the dark for 30 min using 50 mM iodoacetamide. Protein concentration was quantified using the EZQ assay (Thermo Fisher). The lysates were diluted with 100 mM triethyl ammonium bicarbonate (TEAB) four-fold for the first digestion with endoprotease Lys-C (Fujifilm Wako Chemicals). The lysates were further diluted 2.5-fold before a second digestion with trypsin. Lys-C and trypsin were used (1:50 (w/w) enzyme:substrate), carried out overnight at 37°C. The digestion was terminated with 1% (v/v) trifluoroacetic acid (TFA) C18 Sep-Pak cartridges (Waters) were used to desalt the peptides. Desalted peptides were dried and dissolved in 5% formic acid (FA).

Reverse-phase liquid chromatography-MS

RP-LC was performed using a Dionex RSLC nano HPLC (Thermo Fisher Scientific). Peptides were injected onto a 75 µm x 2 cm PepMap-C18 pre-column and resolved on a 75 µm x 50 cm RP- C18 EASY-Spray temperature controlled integrated column-emitter (Thermo Fisher Scientific). The mobile phases were: 2% ACN incorporating 0.1% FA (Solvent A) and 80% ACN incorporating 0.1% FA (Solvent B). A gradient from 5% to 35% solvent B was used over 4 h with a constant flow of 200 nL/min. The spray was initiated by applying 2.5 kV to the EASY-Spray emitter and the data were acquired on a Q-Exactive Orbitrap (Thermo Fisher Scientific) under the control of [Xcalibur Software](#) version 2.8-280502/2.8.1.2806 (RRID:SCR_014593) in a data dependent mode selecting the 15 most intense ions for HCD-MS/MS. Raw MS data was processed using [MaxQuant](#) (version 1.3.0.5) (RRID:SCR_014485) ([Cox & Mann, 2008](#)).

Mass spectrometry data analysis

Putative readers were ranked based on the median of the intensity-based absolute quantification (iBAQ) values ([Schwanhäusser et al., 2011](#)). Briefly, for normalisation, iBAQ values were divided by the total sum of intensity of each sample and the median of normalised biological replicates was calculated. A value of 1×10^{-9} (a number 100 times smaller than the smallest value in the dataset) was then added to all median values to calculate fold changes.

SDS-PAGE

Samples were prepared and validation pulldowns were performed as described in “Protein extraction” and “RNA pulldown”, respectively. Proteins bound to RNA-baits were eluted in 50 µL of sample buffer (Biorad, 1610737) and boiled for 5 min at 95 °C in a heat block (Eppendorf 5355). Boiled samples were then spun at 13,000 rpm in a microcentrifuge (Eppendorf 5430R). Pulldown eluates (100%) and input controls (5%) were loaded alongside a protein ladder (PageRuler Plus pre-stained protein ladder 10–250 kDa, Thermo Fisher Scientific 26619) in a 4-12% Bis-Tris gel and ran in NuPAGE MOPS SDS running buffer (Invitrogen) at 200V for 50 min in a XCell SureLock electrophoresis system (Thermo Fisher Scientific) connected to a Bio-Rad PowerPac Universal Power Supply (Bio-Rad 1645070).

Western blotting

Proteins resolved in SDS-PAGE were transferred onto a nitrocellulose membrane (Hybond ECL, Amersham) using NuPAGE Transfer Buffer (Invitrogen), for 2 h, at 250 mA, 4°C in a Mini Trans-Blot Cell (Bio-Rad). Membranes were blocked in 3% non-fat dry milk in TBS-T buffer for 30 min at room temperature. Membranes were incubated with primary antibody (mouse monoclonal anti-FLAG M2, Sigma F1804) at 1:1000 dilutions in 3% milk-TBS-T overnight rotating at 4°C. Following three washes in TBST-T for 10 min each, membranes were incubated with goat polyclonal anti-mouse IgG conjugated to HRP (1:10000; Invitrogen, A16078) diluted in milk/TBS-T for 1–2 hours at room temperature. Following three washes in TBS-T, bands were detected using Immobilon Western Chemiluminescent HRP Substrate (Millipore) according to manufacturer’s instructions. Medical X-ray films (Super Rx, Fuji) were exposed to luminescent membrane for the time required and the films were developed on a Compact X4 automatic X-ray film processor (Xograph Imaging Systems Ltd).

Results

To increase the chances to identify *bona-fide* modification readers, we used modified RNA baits with biologically relevant characteristics. Considering that m⁵C is deposited in tRNAs in a structure-dependent manner, we designed baits bearing modifications in three structural contexts: single-strand, double-strand, and loop ([Figure 1A](#)). m⁵C, hm⁵C, hm⁵Cm and f⁵C monomers were synthesised and incorporated into oligonucleotides tethered to triethylene glycol (TEG)-biotin ([Tanpure & Balasubramanian, 2017](#)).

The RNA pulldown experiments were performed in independent biological triplicate. We incubated the RNA baits with pre-cleared whole worm lysates prepared from synchronised populations of gravid adults. To increase the stringency of our screen, we included two controls: beads only and an unmodified RNA bait. Following the pulldown experiments, proteins were eluted from the RNA baits and identified by mass spectrometry ([Suen, 2022](#)).

Proteins were ranked according to their enrichment in modified RNA pulldowns in comparison to the controls. Proteins were

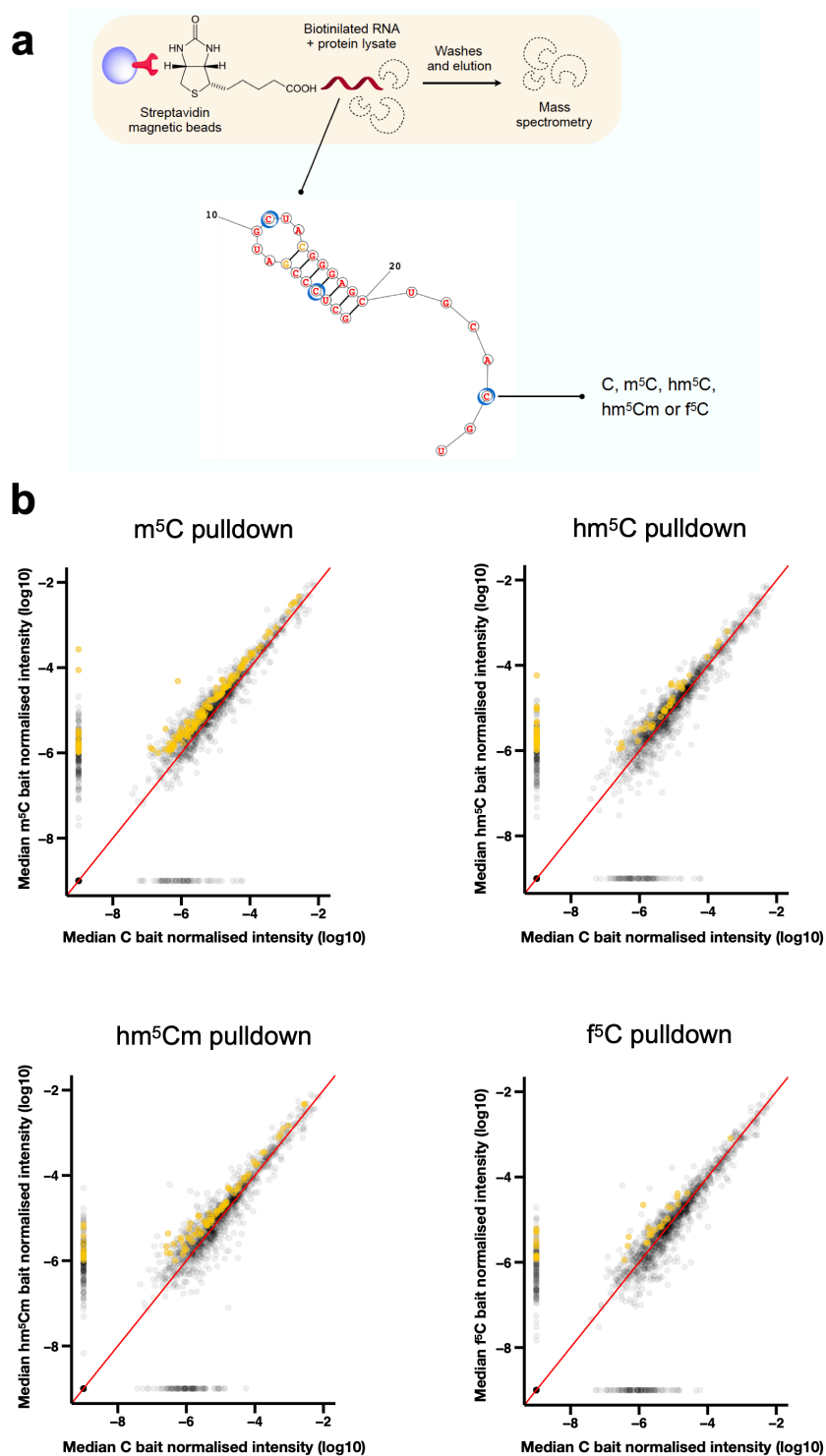


Figure 1. Identification of RNA modification candidate readers. (a) Molecular baits were designed bearing modifications in three structural contexts: single-strand, double-strand, and loop. m⁵C, hm⁵C, hm⁵Cm and f⁵C monomers were synthesised and incorporated into oligonucleotides tethered to TEG-biotin. Blue circles indicate modified positions. Pre-cleared protein lysate from synchronised adult populations was incubated with unmodified or modified biotinylated RNA baits. Streptavidin-conjugated magnetic beads were used to capture proteins that interact with the RNA. Following washes and elution, proteins were processed and identified by RP LC-MS. (b) Scatter plot of proteins bound to modified versus unmodified RNA baits. Median of the normalised iBAQ values were plotted. Proteins uniquely enriched in m⁵C, hm⁵C, hm⁵Cm and f⁵C pulldowns are highlighted in mustard. Proteins were considered candidate readers whenever the fold-change of iBAQ values over the “beads only” and “unmodified C” controls was >2.0 and >1.5, respectively. n = 3 independent biological replicates.

considered candidate readers if they preferentially bound to the modified-RNA bait, i.e., at least 2.0- and 1.5-fold increased as compared with the beads and unmodified C controls, respectively (Figure 1B).

To narrow down the number of candidate readers, we focused on unique binders, i.e. proteins that bound specifically to a single modified RNA bait. Using these criteria, we found 143, 62, 81 and 33 putative unique readers for m⁵C, hm⁵C, hm⁵Cm and f⁵C, respectively (Figure 2A). We used the WormBase Enrichment Suite to perform phenotype enrichment analysis on the lists of unique candidates (Angeles-Albores *et al.*, 2016). This tool analyses a list of genes according to phenotypes that have been reported by researchers to WormBase upon knockout or RNAi, thus suggesting a potential involvement of such genes in specific biological processes. m⁵C unique binders are the most abundant proteins when compared to its oxidised derivatives (Huber *et al.*, 2015; Huber *et al.*, 2017), likely due to higher levels of this modification in RNA. This subset is enriched mostly for germline development phenotypes, and a few RNA surveillance processes. hm⁵Cm-unique binders show specific enrichment for RNA processing phenotypes, whereas hm⁵C and f⁵C show exclusively germline-related phenotypes (Figure 2B).

To validate our dataset, we verified whether ALYREF homologues could be found among our ranked m⁵C reader candidates. Yang *et al.* identified amino acid residues on ALYREF required for binding to m⁵C (Yang *et al.*, 2019). Protein alignments show that some of these residues are conserved in ALY-1 and ALY-2 (Figure 3A). Indeed, ALY-1 and ALY-2 are 1.8-fold enriched across all triplicates of m⁵C pulldowns. We used transgenic strains expressing either ALY-1 (OP502), or ALY-2 (OP217), tagged at the C-terminus with TY1::eGFP::3xFLAG, to confirm these findings (Sarav *et al.*, 2012). We performed pulldown experiments under the same conditions as before and confirmed using western blotting that both ALY-1 and ALY-2 preferentially bind m⁵C *in vitro* (Figure 3B; Suen, 2022).

Some m⁶A readers use common RNA binding domains to identify modified RNA (Shi *et al.*, 2019). Amino acid residues shown to be important for m⁵C binding in ALYREF also reside in the RNA recognition motif (RMM) domain, which is a known RNA-binding domain (Yang *et al.*, 2017). The *C. elegans* orthologues ALY-1 and -2 are predicted to have a similar protein fold in the same region (Figure 3C).

Having shown that ALY-1 and ALY-2 preferentially bind m⁵C-modified RNA, we wondered if other RRM-containing proteins were also identified as m⁵C unique binders. We found that three other proteins with one or more RRMs, RNP-4, SQD-1 and PAB-2, bind to m⁵C baits preferentially. RRM domains consist of a four-stranded antiparallel β -sheet with two α -helices packed against it (Maris *et al.*, 2005). Sequence alignments of the RRM domains in ALY-1, ALY-2, RNP-4, SQD-1, PAB-2 and ALYREF show that a high degree of

conservation exists across these proteins in the two ribonucleoprotein (RNP) consensus sequences that are important for nucleic acid binding. Interestingly, a phenylalanine residue outside of the two RNPs is highly conserved in all six proteins (Figure 3D).

In summary, we have identified putative readers for m⁵C, hm⁵C, hm⁵Cm and f⁵C in *C. elegans* using proteomics. To verify the relevance of this proteomic dataset, we confirmed that ALY-1 and ALY-2 interact with m⁵C orthogonally. Finally, we identified three candidate proteins that share a conserved RRM domain with ALY-1/2 (RNP-4, SQD-1 and PAB-2), highlighting these as potential *bona fide* m⁵C readers.

Discussion

This work presents a list of candidate readers of the m⁵C oxidative derivatives hm⁵C, hm⁵Cm and f⁵C. Stable isotope labelling experiments suggest that hm⁵Cm is a stable modification (Huber *et al.*, 2017). In addition, it has been shown that the ratios between hm⁵C, hm⁵Cm and f⁵Cm in the tRNA Leu-CAA pool varies between organisms and tissues, suggesting a dynamic regulation of these marks in different physiological contexts (Kawarada *et al.*, 2017). This could suggest that these modifications carry out cellular functions themselves, rather than being metabolic by-products of a demethylation pathway. Considering the reported function of f⁵C in mitochondrial protein synthesis (Takemoto *et al.*, 2009), it is expected that the oxidation of m⁵C in tRNAs might affect their role in cytoplasmic translation. The identification of readers provides a first step to elucidate these functions. Importantly, pathogenic mutations have been mapped to NSUN2, NSUN3 and FTSJ1, suggesting that imbalances in these modifications might contribute to the pathogenesis of human diseases (Abbasi-Moheb *et al.*, 2012; Freude *et al.*, 2004; Guy *et al.*, 2015; Khan *et al.*, 2012; Komara *et al.*, 2015; Ramser *et al.*, 2004; Takano *et al.*, 2008; Van Haute *et al.*, 2016).

The use of synthetic modified RNAs as a bait for identification of proteins that are either specifically, or preferentially, interacting with ribonucleoside modifications has been applied with success in the past (Chen *et al.*, 2019; Dominissini *et al.*, 2012; Yang *et al.*, 2017; Yang *et al.*, 2019). Rather than focusing on sequence context, we developed baits bearing modifications in diverse structural contexts. In this work, we identified two proteins showing higher binding affinity to modified RNA: ALY-1 and ALY-2.

The genome of *C. elegans* encodes three members of the Aly/REF family. It has been shown that simultaneous downregulation of these three genes by RNAi does not compromise either viability, or development. In addition, no defects in mRNA export were observed upon simultaneous knockdown of ALY-1, ALY-2 and ALY-3, suggesting that these proteins mediate alternative processes in the nematode, or that their role in transcript export is redundant with other genes (Longman *et al.*, 2003). This is in agreement with reports showing that in *Drosophila* ALY proteins are not

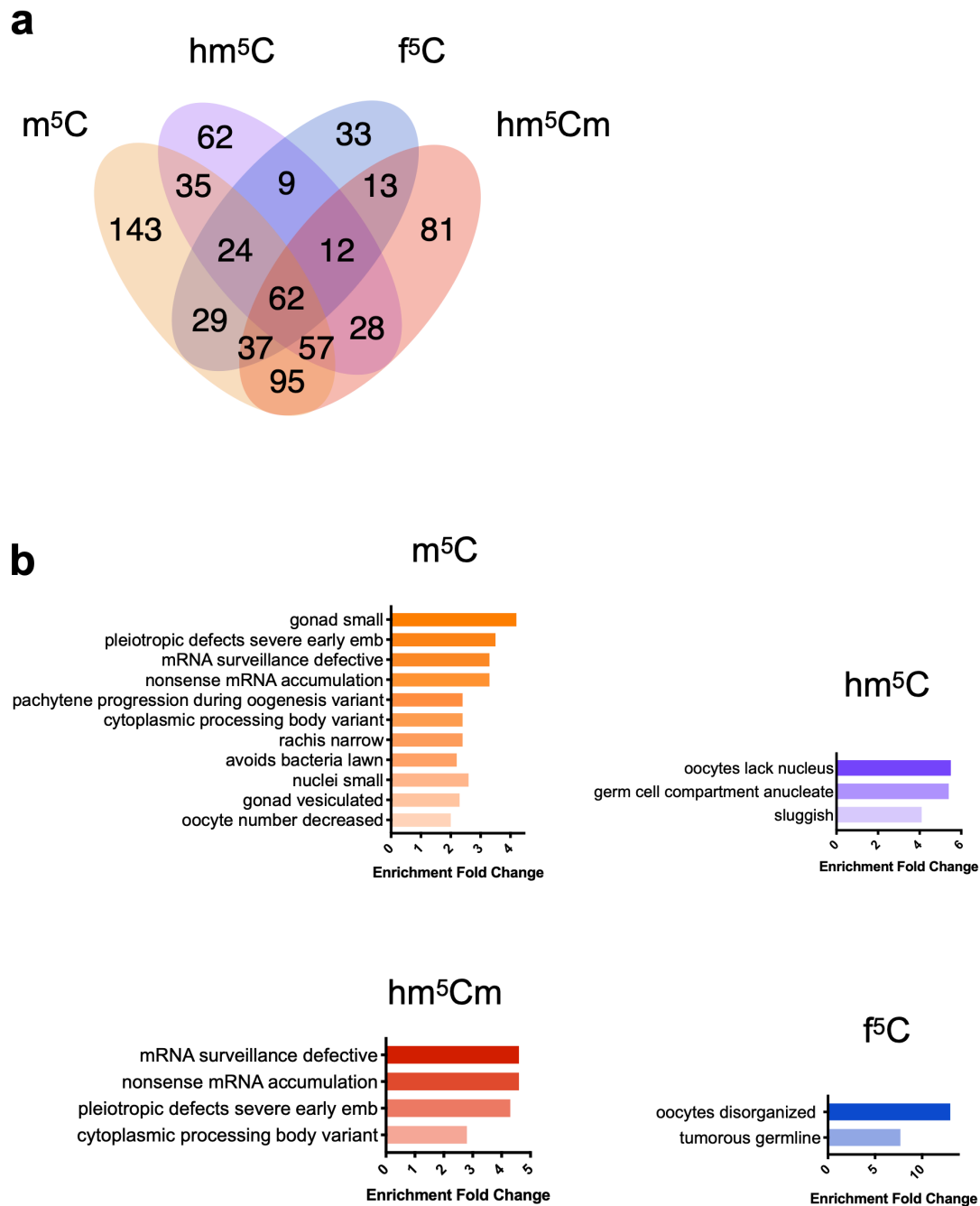


Figure 2. Phenotype enrichment for proteins identified as unique candidate readers. (a) Venn diagram showing the number of enriched proteins detected by mass spectrometry in pull-downs using RNA with different m⁵C, hm⁵C, hm⁵Cm or f⁵C. (b) Phenotype enrichment analysis of unique binders of m⁵C, hm⁵C, hm⁵Cm or f⁵C. Enrichment analysis was performed with the Wormbase Enrichment Suite. Colour code: m⁵C in orange, hm⁵C in purple, hm⁵Cm in red and f⁵C in blue.

required for general mRNA transport (Gatfield & Izaurralde, 2002). In contrast, ALY-1 and ALY-2 have been implicated in nuclear retention of a specific mRNA. In *C. elegans*, sex is initially specified by the ratio of X chromosomes to autosomal chromosomes, *i.e.* XX animals are hermaphrodites and XO are males. Female fate requires the genes *tra-1* and

tra-2, while male cell fate requires inhibition of *tra-2* activity. Kuersten *et al.* showed that binding of NXF-2 and ALY-1/2 inhibits nuclear export of *tra-2*. In support of these findings, RNAi against ALY proteins led to 5–7% female occurrence in the progeny (Kuersten *et al.*, 2004). It remains to be determined whether m⁵C acts as an intermediate in

this process, for example by enhancing binding affinity of ALY proteins to *tra-2* mRNA. Notably, our previous report demonstrated that m⁵C is absent or occurs very rarely in the coding transcripts in *C. elegans*, suggesting that ALY-1/2 could be recognising m⁵C in other molecules (Navarro *et al.*, 2020). In addition, the existence of multiple orthologs of ALY genes in *C. elegans* raises the possibility that these proteins could act redundantly or bind differential sets of modified residues in RNA.

Conclusions

Our study presents the first comprehensive investigation of putative readers of m⁵C and its derivatives in a whole organism. This dataset represents an important resource for the discovery of new functions of the RNA m⁵C pathway.

Data availability

Underlying data

Mendeley Data: Underlying data for ‘Identification of putative reader proteins of 5-methylcytosine and its derivatives in

Caenorhabditis elegans RNA’. <https://doi.org/10.17632/k6z3c5xftr.3> (Suen, 2022)

This project contains the following underlying data:

- Data file: proteinGroups.txt
- Data file: iBAQ values
- Supplementary file: Original Western blots

Data are available under the terms of the [Creative Commons Zero “No rights reserved” data waiver](#) (CC0 1.0 Public domain dedication).

Accession numbers

PRIDE Project: Putative reader proteins of RNA 5-methylcytosine and its derivatives in *Caenorhabditis elegans*. Accession number PXD035761; <https://identifiers.org/pride.project:PXD035761>

Acknowledgements

We would like to thank the Gurdon Institute Media Kitchen for their support providing reagents and media.

References

- Abbasi-Moheb L, Mertel S, Gonsior M, *et al.*: **Mutations in *NSUN2* cause autosomal-recessive intellectual disability.** *Am J Hum Genet.* 2012; **90**(5): 847–855.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Amos H, Korn M: **5-Methyl cytosine in the RNA of *Escherichia coli*.** *Biochim Biophys Acta.* 1958; **29**(2): 444–445.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Angeles-Albores D, Lee RY, Chan J, *et al.*: **Tissue enrichment analysis for *C. elegans* genomics.** *BMC Bioinformatics.* 2016; **17**(1): 366.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bensaddek D, Narayan V, Nicolas A, *et al.*: **Micro-proteomics with iterative data analysis: Proteome analysis in *C. elegans* at the single worm level.** *Proteomics.* 2016; **16**(3): 381–392.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Boccaletto P, Machnicka MA, Purta E, *et al.*: **MODOMICS: a database of RNA modification pathways. 2017 update.** *Nucleic Acids Res.* 2018; **46**(D1): 303–307.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Brenner S: **The genetics of *Caenorhabditis elegans*.** *Genetics.* 1974; **77**(1): 71–94.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chen X, Li A, Sun BF, *et al.*: **5-methylcytosine promotes pathogenesis of bladder cancer through stabilizing mRNAs.** *Nat Cell Biol.* 2019; **21**(8): 978–990.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Cohn WE: **Pseudouridine, a carbon-carbon linked ribonucleoside in ribonucleic acids: isolation, structure, and chemical characteristics.** *J Biol Chem.* 1960; **235**: 1488–1498.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Cox J, Mann M: **MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification.** *Nat Biotechnol.* 2008; **26**(12): 1367–1372.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Davis F, Allen FW: **Ribonucleic acids from yeast which contain a fifth nucleotide.** *J Biol Chem.* 1957; **227**(2): 907–915.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Dominissini D, Moshitch-Moshkovitz S, Schwartz S, *et al.*: **Topology of the human and mouse m⁶A RNA methylomes revealed by m⁶A-seq.** *Nature.* 2012; **485**(7397): 201–206.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Freude K, Hoffmann K, Jensen LR, *et al.*: **Mutations in the *FTSJ1* Gene Coding for a Novel S-Adenosylmethionine-Binding Protein Cause Nonsyndromic X-Linked Mental Retardation.** *Am J Hum Genet.* 2004; **75**(2): 305–309.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- García-Vílchez R, Sevilla A, Blanco S: **Post-transcriptional regulation by cytosine-5 methylation of RNA.** *Biochim Biophys Acta Gene Regul Mech.* 2019; **1862**(3): 240–252.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Gatfield D, Izaurrealde E: **REF1/Aly and the additional exon junction complex proteins are dispensable for nuclear mRNA export.** *J Cell Biol.* 2002; **159**(4): 579–588.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Guy MP, Shaw M, Weiner CL, *et al.*: **Defects in tRNA Anticodon Loop 2'-O-Methylation Are Implicated in Nonsyndromic X-Linked Intellectual Disability due to Mutations in *FTSJ1*.** *Hum Mutat.* 2015; **36**(12): 1176–1187.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Huber SM, van Delft P, Mendil L, *et al.*: **Formation and Abundance of 5-Hydroxymethylcytosine in RNA.** *Chembiochem.* 2015; **16**(5): 752–755.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Huber SM, van Delft P, Tanpure A, *et al.*: **2'-O-methyl-5-hydroxymethylcytidine: A second oxidative derivative of 5-methylcytidine in RNA.** *J Am Chem Soc.* 2017; **139**(5): 1766–1769.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kawarada L, Suzuki T, Ohira T, *et al.*: **ALKBH1 is an RNA dioxygenase responsible for cytoplasmic and mitochondrial tRNA modifications.** *Nucleic Acids Res.* 2017; **45**(12): 7401–7415.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Khan MA, Rafiq MA, Noor A, *et al.*: **Mutation in *NSUN2*, which encodes an RNA methyltransferase, causes autosomal-recessive intellectual disability.** *Am J Hum Genet.* 2012; **90**(5): 856–863.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Komara M, Al-Shamsi AM, Ben-Salem S, *et al.*: **A Novel Single-Nucleotide Deletion (c.1020delA) in *NSUN2* Causes Intellectual Disability in an Emirati Child.** *J Mol Neurosci.* 2015; **57**(3): 393–399.
[PubMed Abstract](#) | [Publisher Full Text](#)

- Kuersten S, Segal SP, Verheyden J, *et al.*: **NXF-2, REF-1, and REF-2 Affect the Choice of Nuclear Export Pathway for *tra-2* mRNA in *C. elegans*.** *Mol Cell*. 2004; **14**(5): 599–610.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Longman D, Johnstone IL, Cáceres JF: **The Ref/Aly proteins are dispensable for mRNA export and development in *Caenorhabditis elegans*.** *RNA*. 2003; **9**(7): 881–891.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Maris C, Dominguez C, Allain FHT: **The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression.** *FEBS J*. 2005; **272**(9): 2118–2131.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Navarro IC, Tuorto F, Jordan D, *et al.*: **Translational adaptation to heat stress is mediated by RNA 5-methylcytosine in *Caenorhabditis elegans*.** *EMBO J*. 2020; **40**: e105496.
[Publisher Full Text](#)
- Pais de Barros JP, Keith G, El Adlouni C, *et al.*: **2'-O-methyl-5-formylcytidine (F^oCm), a new modified nucleotide at the 'wobble' of two cytoplasmic tRNAs^{Leu} (NAA) from bovine liver.** *Nucleic Acids Res*. 1996; **24**(8): 1489–1496.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ramser J, Winnepenninckx B, Lenski C, *et al.*: **A splice site mutation in the methyltransferase gene *FTSJ1* in Xp11.23 is associated with non-syndromic mental retardation in a large Belgian family (MRX9).** *J Med Genet*. 2004; **41**(9): 679–683.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Sarov M, Murray JJ, Schanze K, *et al.*: **A genome-scale resource for in vivo tag-based protein function exploration in *C. elegans*.** *Cell*. 2012; **150**(4): 855–866.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Schwanhäusser B, Busse D, Li N, *et al.*: **Global quantification of mammalian gene expression control.** *Nature*. 2011; **473**(7347): 337–342.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Shi H, Wei J, He C: **Where, When, and How: Context-Dependent Functions of RNA Methylation Writers, Readers, and Erasers.** *Mol Cell*. 2019; **74**(4): 640–650.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Suen K: **Identification of putative reader proteins of 5-methylcytosine and its derivatives in *Caenorhabditis elegans* RNA.** *Mendeley Data*. [Data]. 2022. <http://www.doi.org/10.17632/k6z3c5xftr.1>
- Takano K, Nakagawa E, Inoue K, *et al.*: **A loss-of-function mutation in the *FTSJ1* gene causes nonsyndromic x-linked mental retardation in a Japanese family.** *Am J Med Genet B Neuropsychiatr Genet*. 2008; **147B**(4): 479–484.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Takemoto C, Spremulli LL, Benkowski LA, *et al.*: **Unconventional decoding of the AUA codon as methionine by mitochondrial tRNA^{Met} with the anticodon f^oCAU as revealed with a mitochondrial *in vitro* translation system.** *Nucleic Acids Res*. 2009; **37**(5): 1616–1627.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Tanpure AA, Balasubramanian S: **Synthesis and Multiple Incorporations of 2'-O-Methyl-5-hydroxymethylcytidine, 5-Hydroxymethylcytidine and 5-Formylcytidine Monomers into RNA Oligonucleotides.** *ChemBiochem*. 2017; **18**(22): 2236–2241.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Van Haute L, Dietmann S, Kremer L, *et al.*: **Deficient methylation and formylation of mt-tRNA^{Met} wobble cytosine in a patient carrying mutations in NSUN3.** *Nat Commun*. 2016; **7**: 12039.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Yang X, Yang Y, Sun BF, *et al.*: **5-methylcytosine promotes mRNA export — NSUN2 as the methyltransferase and ALYREF as an m^oC reader.** *Cell Res*. 2017; **27**(5): 606–625.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Yang Y, Wang L, Han X, *et al.*: **RNA 5-Methylcytosine Facilitates the Maternal-to-Zygotic Transition by Preventing Maternal mRNA Decay.** *Mol Cell*. 2019; **75**(6): 1188–1202.e11.
[PubMed Abstract](#) | [Publisher Full Text](#)

Open Peer Review

Current Peer Review Status: ? ✓ ?

Version 1

Reviewer Report 18 July 2023

<https://doi.org/10.21956/wellcomeopenres.19826.r59586>

© 2023 Zinovyeva A et al. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Anna Y Zinovyeva

Kansas State University, Manhattan, Kansas, USA

Jeffrey Medley

Division of Biology, Kansas State University, Manhattan, Kansas, USA

RNA modifications have gained much interest in the recent years as they become associated with important biological functions. 5-methylcytosine (m⁵C) is an abundant and conserved post-transcriptional modification found on a variety of RNA molecules. Thus far, much effort has been placed on identifying m⁵C modification sites and unraveling the biological significance of these modifications through “writer” loss-of-function experiments. More is currently known about the “writers” and “erasers” of the m⁵C modifications than the “readers”, which could introduce specificity and unique biological effects through combinatorial modification plus target context.

In “Identification of putative reader proteins of 5-methylcytosine and its derivatives in *Caenorhabditis elegans* RNA”, Navarro *et al.* takes advantage of *C. elegans*, with its excellent genetic tractability, to identify potential readers of m⁵C modification and its derivatives. Identifying proteins that are responsible for reading specific RNA modifications is an important and valuable first step towards understanding the functional significance of distinct RNA modifications. This study aims to do this by identifying *C. elegans* proteins that preferentially interact with synthetic RNA baits containing different chemical modifications. The mass spec identification of bait RNA-associated proteins is a great first-order approach and is thus well-conceived, with appropriate unmodified RNA oligonucleotides and beads-only controls. However, addressing two major areas would be needed to elevate this manuscript to its intended form as a community resource.

1. The first key issue that needs to be addressed is demonstrating to the readers that the provided lists of putative modified oligo interactors are a solid starting point for follow up work. This can be achieved through addressing the following comments:

Additional mass-spec data analysis should be presented to increase the scientific readers' confidence in the identified lists of putative m⁵C, etc. “readers”. Usage of the intensity-based absolute quantification (iBAQ) *median* across three (3) replicates could be wildly misleading if there

is a high degree of variability among the replicates and if pairing of the data exists. Use of a median essentially reduces the analysis to a single replicate comparison between the experimental conditions. One way to bolster the confidence in the data would be to provide a principal component analysis (or analyses) of the proteomics data to demonstrate data consistency between control and m⁵C (and derivatives) experimental conditions.

Figure 1 legend and manuscript text: Please define what you mean by “Proteins uniquely enriched in pulldowns”. Does that mean 1. proteins that were enriched 1.5 or 2-fold over control in each experiment, 2. proteins that were unique to modification type but not enriched over control, or 3. proteins that were 1.5 or 2-fold enriched over control and unique to m⁵C pulldown vs. other modifications? The reader assumption is that it’s likely the latter, however, Figure 1B highlights “uniquely enriched proteins” that appear to fail the 1.5 or 2-fold enrichment threshold. Please address this discrepancy. In addition, use of 1.5- or 2-fold lines on Figure 1B plots, as well as color-coding fold enrichment vs. uniqueness in the plots would further improve the clarity of the data analysis. Labeling highly changed/high confidence candidate readers on the plots (as well as ALY-1, ALY-2, etc.) would also provide context of these proteins’ enrichment for the reader. As shown, Figure 1B provides limited and unclear information.

The expectation inherent to the described experimental approach is that the “reader” proteins are uniquely and specifically able to discriminate between m⁵C RNA modification and its derivatives hm⁵C, hm⁵Cm, and f⁵Cm. A discussion addressing the justifications for this expectation is important to set the stage for the reader to understand the appropriateness of the “uniquely binding” criteria of the co-precipitating proteins. A protein appears to be disqualified from being considered a candidate “reader” if it is not unique to any given modification derivative. Is there a precedent of this assumption and could this assumption perhaps erroneously eliminate “readers” capable of reading m⁵C and some of its derivatives? Further discussion on this is needed to ensure the manuscript audience can evaluate the validity of the approach used to qualify any given putative interactor as a potential “reader”.

In a similar vein of thought, the manuscript does not address the justifications for using the specific oligos and the three modification sites within each oligonucleotide. Thus, the manuscript would greatly benefit from a discussion of what previous data drove the design of the RNA oligonucleotides used in this study, as well as the justification for using the three (3) specific locations for m⁵C within the oligos. Some discussion about how the baits were designed and potential limitations would be helpful to understand the significance of the findings. Given that many m⁵C modification occurs on tRNAs, were the baits designed to mimic the tRNA structure (since the structural context is important)? Is there evidence that RNAs contain these specific modifications at multiple sites? If so, is it known whether modification “readers” can differentiate between molecules containing multiple modifications and single modifications? If single-modified baits were used, how might that impact the identified readers? It seems plausible that some “readers” might differentiate between single-stranded, double-stranded and loop modifications. An expanded discussion of some of these points would help clarify the significance and limitations of this study, necessary to identify potential avenues for further analysis.

2. In Figure 3A, the residues required for RNA binding are highlighted, with the authors referring to Yang *et al.*, 2017¹: “residues in ALYREF required for m⁵C binding previously identified by Yang *et al.* are highlighted in red”. However, these residues do not appear to be defined in that study. According to Yang *et al.*, 2017¹, the same residues are described as being “the relatively conserved

amino acids used for constructing mutants" (and have not each been validated as important for RNA binding as suggested in this study, with one potential exception of K171). The authors should also clarify the relevance of the Yang *et al.*, 2019² citation, as it appears to have been referenced in error.

In Figure 3A and 3D, numbering of amino acid positions would help orient manuscript readers to positions of functional domains and understand the take-home message. In Figure 3A, the highlighted residues appear to be different from the RNP1 and RNP2 motifs highlighted in Figure 3D. What is the basis for RNP1 and RNP2 being highlighted in Figure 3D as RNA binding domains? This should be clarified in both the manuscript text and figure legend. Further, these reviewers are not convinced that the provided alignment shows a "high degree of conservation", as shown in Figure 3D. Usage of protein isoforms identical for the shown region can skew the Muscle analysis and more information is needed for why the authors deem the highlighted regions (RNP1 and 2) in Figure 3D as "important for nucleic acid binding". Finally, the language should be softened to propose the highlighted "RNP" regions as potentially important, with qualifications that the highlighted regions have not been demonstrated to bind nucleic acids. Justifications for highlighting the proposed regions should still be provided.

ALY-1/2 pulldowns/Western findings would be bolstered by providing all replicates in Supplemental data. In addition, language within the manuscript should be softened to avoid stating that pulldowns and Western unambiguously confirm ALY-1 and ALY-2 as m⁵C readers.

Additional comments:

These reviewers appreciate the list of putative interactive proteins provided as a resource in The Supplemental Table 1. However, the practical usability of that resource would be dramatically improved by including the *C. elegans* gene names. This would also allow the readers to independently assess the nature of the putative interacting proteins beyond the provided GO term analysis.

Figure 3C: The legend should state that only the RRM domains of the proteins are shown. This will prevent manuscript readership from misinterpreting the shown predicted structures as structurally conserved across the entire protein length.

Based on GO-term analysis, the authors claim that hm5C and f5C modifications show "exclusively germline related phenotypes". However, one of the phenotypes for hm5C was sluggish, which seems unlikely to result from a germline defect.

Please state in the Methods section how much protein was used per each pulldown experiment going into mass spectrometry. As written, it appears that only 2mg of protein has been used, an inadequate amount for most mass spectrometry experiments. Alternatively, please reference the literature that supports that 2mg of protein is sufficient for RP-LC mass spectrometry.

References

1. Yang X, Yang Y, Sun BF, Chen YS, et al.: 5-methylcytosine promotes mRNA export - NSUN2 as the methyltransferase and ALYREF as an m⁵C reader. *Cell Res.* 2017; **27** (5): 606-625 [PubMed Abstract](#) | [Publisher Full Text](#)
2. Yang Y, Wang L, Han X, Yang WL, et al.: RNA 5-Methylcytosine Facilitates the Maternal-to-Zygotic

Transition by Preventing Maternal mRNA Decay. *Mol Cell*. 2019; **75** (6): 1188-1202.e11 [PubMed Abstract](#) | [Publisher Full Text](#)

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Not applicable

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: RNA biology, *C. elegans*, RNA pulldown techniques

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however we have significant reservations, as outlined above.

Reviewer Report 11 July 2023

<https://doi.org/10.21956/wellcomeopenres.19826.r61033>

© 2023 Phillips C. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Carolyn Phillips 

University of Southern California, Los Angeles, California, USA

m⁵C is a conserved RNA modification that has been detected in tRNAs, rRNAs, and mRNAs from other systems that has roles such as promoting tRNA stability and translational fidelity. In *C. elegans*, m⁵C is non-essential but nonetheless prevalent in rRNAs and tRNAs, as shown by the same research group in previous work (Navarro et al EMBO 2020). mRNA modification appears to be less common. Loss of m⁵C from *C. elegans* leads to developmental and fertility defects as well as a decrease in translational efficiency. Here, Navarro et al seek to identify readers of the m⁵C

modification in *C. elegans*. Only two m⁵C readers have been identified to date in any organism, so identification of more candidate proteins is an important question. To this end, Navarro *et al*, synthesize RNA baits modified with m⁵C and derivatives, incubate the baits with RNA lysate, and following pulldown, perform mass spec for interacting proteins. A significant number of candidate readers were identified, including orthologs of ALYREF, a previously known m⁵C reader, which were then further validated by m⁵C RNA pulldown and western blot. Further analysis will be needed to parse the false positives from the true m⁵C readers and to determine the biological function of each protein.

Comments –

- RNA baits were designed in three structural contexts, single-strand, double-strand and loop – but it is not discussed in the results whether the identified m⁵C candidate readers were found using one or all three of these structures. And were unmodified baits also used in all three structural contexts?
- The text indicates that candidates were limited to those that bound only one of the modified RNA baits, but it is not discussed whether the known readers are unable to bind the m⁵C variants. Is it possible that interesting candidates would have been eliminated at this step? At a minimum, the full list of mass spec interactors, including those found binding multiple m⁵C variants should be provided.
- Figure 3A and text “Protein alignments show that some of these residues are conserved in ALY-1 and ALY-2” – Of the residues highlighted in the red box, several are weakly conserved or not conserved at all. It is not discussed whether the loss of these critical residues will likely affect the ability of these proteins to bind the m⁵C modification.
- In the last sentence of the results, the word “bona fide” should be removed since these interactions have not yet been validated.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: C. elegans RNA silencing

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 09 December 2022

<https://doi.org/10.21956/wellcomeopenres.19826.r53385>

© 2022 Schaefer M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Matthias Ronald Schaefer 

Centre for Anatomy and Cell Biology, Division of Cell-and Developmental Biology, Medical University of Vienna, Vienna, Austria

Besides mapping chemical modifications in RNA, the vast majority of epitranscriptomics studies is currently reporting on associations between the function of a gene (writer, reader, eraser) and a multitude of observable phenotypes. These approaches are usually very simplistic and classically involve gene knockdowns or even knockouts, followed by the description of phenotypic changes, many of which seem to be rather pleiotropic resulting in little insight which of the RNA substrates of a particular writer, reader or eraser protein has contributed to the observed changes. Interestingly, even though mapping studies have provided positional clues about the occurrence of particular modifications, there seems to be little incentive in following up on assigning functional impact of specifically modified nucleotides in specific RNAs on a particular trait and/or phenotype.

It is accepted that RNA modifications are “written” onto RNA sequences by a staggering number of dedicated enzymes. While non-coding RNAs are often stoichiometrically modified with a great number of different RNA modifications (i.e., tRNAs), mRNAs, while being modified with a few different RNA modifications, carry only sub-stoichiometric levels of a particular modification at specific nucleotides. It follows that modified nucleotides in mRNA are not only rare within a transcript (1-3 modified nucleotides per 1000 nucleotides when averaged over all mRNAs for the most abundant mRNA modification, m6A) but also stoichiometrically low (e.g., about 20% m5C can be detected at a given position based on RNA bisulfite sequencing). This is an important and seemingly overlooked conundrum in the field, since it begs the question how a few nucleotides within thousands of nucleotides, and those at every 5th to 10th mRNA molecule should account for physiologically meaningful impact on the biology of a cell.

A solution for this presents the concept of reader proteins. This concept has been put forward by epitranscriptomic influencers borrowing heavily from the notion of epigenetic readers. If there are only a few RNA modifications within a sea of transcripts, then proteins binding preferentially to this modified nucleotides (readers) might make the difference. Is this concept correct? Do experimentally validated data support the assumption that a few RNA modifications will make a

difference through interpretation of particular proteins?

In the last years, especially the m6A-focused field has reported on the identification of various m6A readers. A few attempts also exist in the public record that identified readers of other RNA modifications, including m5C. Notably, much of the current identification and characterization of these RNA modification readers relied on in vitro pull-down experiments using either defined (synthetic) RNA baits (rather than those identified by epitranscriptomic mapping strategies) combined with mass spectrometry or RNA-IPs (and derivatives of the method) to identify RNA substrates of specific reader proteins. From these reports one can deduce that all readers of specific RNA modifications also bind to unmodified RNA sequences. This begs the question if readers bind with lower Kd to modified RNAs than to unmodified RNAs? Curiously, very little has been published reporting on actual interaction parameters for specific readers and their substrates, which is common when studying enzymes or protein-substrate interactions. Instead, fixed end-point assays with readouts based on RNA sequencing or western blotting are currently substituting for deriving quantitative and kinetic parameters from the identified interactions, which would be needed to better understand the biology of putative readers of RNA modifications.

This manuscript by Navarro et al., 2022 is therefore welcome for attempting to identify reader proteins for m5C as potential mediators of RNAs modified with this chemical modification. Assuming that *C. elegans* contains readers for the m5C modification and its oxidized derivatives, the authors use an in vitro RNA pull-down approach with differently modified RNA baits to identify proteins differentially binding to m5C, hm5C, hm5m and f5C. Setting rather tight cut-offs for enrichment of controls, they report on numbers and GO annotations of the identified proteins. Since they found the orthologues of a previously identified human m5C reader protein (ALYREF) in their data, they repeated the pull-down experiment with m5C-modified oligos and identified both worm orthologues of ALYREF to bind preferentially to m5C-modified oligos.

This reviewer accepts that Wellcome Open Research publishes articles reporting any original research that has been funded (or co-funded) by [Wellcome](#), without editorial bias including papers reporting single findings, replications studies, and null results or negative findings. This reviewer will therefore only assess the validity of the article content, rather than the novelty or interest levels.

Comments:

1. This manuscript cannot be disconnected from Navarro et al., 2020¹, which reported on only 188 positions in the worm transcriptome as putative m⁵C sites. Their final conclusion about m⁵C in *C. elegans*: "In summary, we found no evidence of a widespread distribution of m⁵C in coding transcripts. To investigate whether the presence of common characteristics could support a subset of the aforementioned 188 positions as bona fide methylated sites, we performed gene ontology, motif search, genomic localisation and secondary structure analyses on these transcripts and sites; however, no significant shared features were found. While our data do not completely rule out the existence of m5C methylation in mRNAs, it demonstrates that this mark cannot be detected in high stoichiometry in *C. elegans*, as observed in tRNAs and rRNAs." These published results hints towards the fact, that putative readers of m5C in worms will read this modification likely in tRNA and rRNA but not in mRNA. This reviewer suggests to place a reminder of their findings into the introduction of the manuscript, since the current version will be read by many scientists, eager to put

meaning into exceptions, as proof that *C. elegans* contains m⁵C in mRNAs which affect its adaptability to high temperatures.

2. The RNA bait is depicted as a secondary structure in Figure 1a. Please specify in the M&M, how you achieved formation of such structures before performing pull-down experiments with protein extracts.
3. PAGE 4: "To increase the chances to identify bona-fide modification readers..." Since synthetic and not naturally occurring modified cytosine-containing RNA oligos were used, this reviewer would not call the "to-be-identified readers" bona fide (meaning genuine, true).
4. Figure 1b: To this reviewer, mustard is not the color of choice to highlight enriched proteins. Please make this readable. Please, indicate what the vertical and horizontal lines of proteins should signify. Based on the Figure legend, this reviewer would expect them to be all in mustard since they are clearly enriched of unmodified oligo. However, these identities are marked by all colors.
5. PAGE 6: "Yang *et al.* identified amino acid residues on ALYREF required for binding to m⁵C (Yang *et al.*, 2019²). Protein alignments show that some of these residues are conserved in ALY-1 and ALY-2 (Figure 3A)." Yang *et al.*, reported on exactly one (1) amino acid (K171) that seemed to be required for binding of ALYREF to m⁵C-modified RNA oligos. Please, correct this sentence. Figure 3a shows that exactly this lysine (human K171) is a glutamine (Q) in both worm *aly-1* and *-2*. Even though a few other residues in the shown sequence are conserved, none really matters since lysine within the sequence of AMKQ in human ALYREF is the only residue to be important as published by Yang *et al.*, 2019. Spoiler alert for readers of this review: in S4G of Yang *et al.*, 2019, exactly at the level of ALYREF (with the K171A mutation), the binding to the 5mC oligo is obscured by a blind spot (air bubble?), which makes it questionable if that the signal is really reduced.
6. Figure 3a: "*" represents identical amino acids while ":" and "." represent amino acids with strongly and weakly similar properties, respectively." Please indicate, which protein substitution/similarity index was used to support this statement.
7. Figure 3b: "Anti-FLAG western blotting (left) showing enrichment of ALY-1 and ALY-2 following pulldown using m⁵C-modified RNA oligos and relative quantification (right). Mean ± SD, quantification performed using ImageJ, n = 2 independent biological replicates." Provided raw blots at <https://data.mendeley.com/datasets/k6z3c5xft/3> do not show replicates of pull-down followed by *aly-2*-FLAG identification. Please provide, or state that the data come from a single experiment. A comment for interested readers: quantification of overexposed inputs is notoriously difficult and likely not possible. Hence, either a titration curve on input protein or lower exposure would convince this reviewer of the stated enrichment factors using density measurements.
8. "Amino acid residues shown to be important for m⁵C binding in ALYREF also reside in the RNA recognition motif (RMM) domain, which is a known RNA-binding domain (Yang *et al.*, 2017³)." same overstatement as before. Yang *et al.*, 2017 only reported on K171, which lies at the border of the RRM and is not conserved in any of the proteins shown in Figure 3.

Please, rewrite this statement to avoid that AI-based manuscript writing picks up on it and propagates this inconsistency.

9. Given that any orthologues of ALYREF were “only” 1.9x enriched to bind to m5C over C-oligo, what about the other binders listed in Supp Table 1? Provided that m5C is likely present in tRNA and rRNAs in *C. elegans*, how many of the identified proteins in the experiments bind to tRNAs or rRNA?

References

1. Navarro IC, Tuorto F, Jordan D, Legrand C, et al.: Translational adaptation to heat stress is mediated by RNA 5-methylcytosine in *Caenorhabditis elegans*. *EMBO J.* 2021; **40** (6): e105496 [PubMed Abstract](#) | [Publisher Full Text](#)
2. Yang Y, Wang L, Han X, Yang WL, et al.: RNA 5-Methylcytosine Facilitates the Maternal-to-Zygotic Transition by Preventing Maternal mRNA Decay. *Mol Cell.* 2019; **75** (6): 1188-1202.e11 [PubMed Abstract](#) | [Publisher Full Text](#)
3. Yang X, Yang Y, Sun BF, Chen YS, et al.: 5-methylcytosine promotes mRNA export - NSUN2 as the methyltransferase and ALYREF as an m5C reader. *Cell Res.* 2017; **27** (5): 606-625 [PubMed Abstract](#) | [Publisher Full Text](#)

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Not applicable

Are all the source data underlying the results available to ensure full reproducibility?

No

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: RNA biology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.
