



Identification of m⁶A-Associated RNA Binding Proteins Using an Integrative Computational Framework

Yiqian Zhang^{1*} and Michiaki Hamada^{1,2,3,4*}

¹ Department of Electrical Engineering and Bioscience, Faculty of Science and Engineering, Waseda University, Tokyo, Japan, ² AIST-Waseda University Computational Bio Big-Data Open Innovation Laboratory (CBBDOIL), Tokyo, Japan, ³ Institute for Medical-Oriented Structural Biology, Waseda University, Tokyo, Japan, ⁴ Graduate School of Medicine, Nippon Medical School, Tokyo, Japan

OPEN ACCESS

Edited by:

Jia Meng,
Xi'an Jiaotong-Liverpool University,
China

Reviewed by:

Lin Zhang,
China University of Mining and
Technology, China
Kunqi Chen,
Fujian Medical University, China

*Correspondence:

Yiqian Zhang
z10000507@126.com
Michiaki Hamada
mhamada@waseda.jp

Specialty section:

This article was submitted to
RNA,
a section of the journal
Frontiers in Genetics

Received: 04 November 2020

Accepted: 05 February 2021

Published: 01 March 2021

Citation:

Zhang Y and Hamada M (2021)
Identification of m⁶A-Associated RNA
Binding Proteins Using an Integrative
Computational Framework.
Front. Genet. 12:625797.
doi: 10.3389/fgene.2021.625797

N6-methyladenosine (m⁶A) is an abundant modification on mRNA that plays an important role in regulating essential RNA activities. Several wet lab studies have identified some RNA binding proteins (RBPs) that are related to m⁶A's regulation. The objective of this study was to identify potential m⁶A-associated RBPs using an integrative computational framework. The framework was composed of an enrichment analysis and a classification model. Utilizing RBPs' binding data, we analyzed reproducible m⁶A regions from independent studies using this framework. The enrichment analysis identified known m⁶A-associated RBPs including YTH domain-containing proteins; it also identified RBM3 as a potential m⁶A-associated RBP for mouse. Furthermore, a significant correlation for the identified m⁶A-associated RBPs is observed at the protein expression level rather than the gene expression level. On the other hand, a Random Forest classification model was built for the reproducible m⁶A regions using RBPs' binding data. The RBP-based predictor demonstrated not only competitive performance when compared with sequence-based predictions but also reflected m⁶A's action of repelling against RBPs, which suggested that our framework can infer interaction between m⁶A and m⁶A-associated RBPs beyond sequence level when utilizing RBPs' binding data. In conclusion, we designed an integrative computational framework for the identification of known and potential m⁶A-associated RBPs. We hope the analysis will provide more insights on the studies of m⁶A and RNA modifications.

Keywords: N6-methyladenosine, RNA binding proteins, RNA modification, enrichment analysis, random forest

1. INTRODUCTION

In recent years, RNA modification has emerged as a mode of post-transcriptional gene regulation and has been gaining increasing attention from researchers around the globe. More than 150 types of post-transcriptional modification have been discovered, with N6-methyladenosine (m⁶A) as being one of the most abundant mRNA modification (Roundtree et al., 2017). m⁶A is featured with the DRACH motif (where D = A, G or U; R = A or G; H = A, C or U) and is preferentially located near 3' untranslated regions (3' UTR) (Linder et al., 2015). It has been reported that m⁶A

participates in essential RNA activities including alternative splicing, export, translation, and decay in the nucleus and cytoplasm (Lee et al., 2020).

m⁶A exerts its function through interaction with several RNA binding proteins that can be considered as m⁶A-associated RBPs. There are three main kinds of known m⁶A-associated RBPs that are also known as m⁶A effectors (Shi et al., 2019), they are writer, eraser, and reader. m⁶A writers are methyltransferases like METTL3, METTL14, WTAP, RBM15/15B, while m⁶A erasers are demethylases like FTO, ALKBH5, and m⁶A readers are the proteins that can recognize m⁶A like the YTH domain-containing proteins (YTHDF1/2/3), EIF3 (Lee et al., 2020), FMR1 (Eduvuganti et al., 2017). These m⁶A effectors cooperate with each other to facilitate both temporal and spatial regulation where writers work in the nucleus to introduce the m⁶A modification which is then recognized by various readers in the nucleus and cytoplasm, which can influence activities of their target RNAs.

Furthermore, the roles of m⁶A and m⁶A-associated RBPs in cancer are being a general interest to researchers. The writer METTL3 was early noticed because of its overexpression in acute myeloid leukemia (AML). It was found that m⁶A promotes the translation of oncogenes like c-MYC, BCL2, and PTEN in the human acute myeloid leukemia MOLM-13 cell line (Vu et al., 2017). Because of necessity of METTL3 in the maintain the leukaemic state, it is identified as a potential therapeutic target for AML (Barbieri et al., 2017). Apart from METTL3, a study found that the reader YTHDF2 silenced in HCC cells can provoke inflammation, vascular reconstruction, and metastatic progression (Hou et al., 2019). Besides, m⁶A and m⁶A reader YTHDF1 have been reported to control anti-tumor immunity. YTHDF1 deficient mice had enhanced therapeutic efficacy of PD-1 checkpoint blockade which suggested YTHDF1's potential in anti-cancer immunotherapy (Han et al., 2019). Therefore, the study of m⁶A and m⁶A-associated RBPs enables us to develop a better understanding of gene regulation mechanism and leads to potential therapeutic opportunities.

To unveil m⁶A's regulation mechanism, it is very necessary to study m⁶A-associated RBPs and their target RNAs. High-throughput sequencing technologies like CLIP-seq (Ule et al., 2005) and RIP-Seq (Zhao et al., 2010) make it feasible to study target RNAs of m⁶A effectors at a transcriptome-wide level. Based on the high-throughput sequencing data, a team developed a database for the collection of these target RNAs (Deng et al., 2020), and another team developed a prediction model focused on the targets of m⁶A readers (Zhen et al., 2020). However, computational resources for identification of m⁶A-associated RBPs are still limited. Though there has been a manually-curated database built for the collection of known m⁶A effectors across species (Nie et al., 2020), to identify potential m⁶A-associated RBPs, it needs to develop efficient computational methods. There are some computational methods that have been used to identify m⁶A-associated RBPs. One such method is to build a prediction model based on deep learning and then extract the sequence features (Zhang and Hamada, 2018; Wang and Wang, 2020). However, not all the RBP motifs are available and sequences can not reflect actual binding status, thus limiting their utility in the

identification of m⁶A-associated RBPs. Another group developed an analysis framework to identify cell-specific trans-regulators of m⁶A (An et al., 2020). They identified the association between m⁶A and RBPs but did not take into consideration the interaction such as reading and repelling between them.

In our study we decided to focus on the use of reproducible m⁶A regions for identification of m⁶A-associated RBPs, with consideration of variation among MeRIP-Seq datasets (about 30–60% between studies, even in the same cell type; McIntyre et al., 2020). We aimed to identify m⁶A-associated RBPs from reproducible m⁶A regions using an integrative computational framework. This framework is composed of an enrichment analysis and a classification model. The enrichment analysis allows us to identify RBPs enriched in the m⁶A regions. We were able to identify not only the known m⁶A-associated RBPs like YTH domain-containing proteins, but also a potential m⁶A-associated RBP, RBM3, for mouse. We went on to evaluate the correlation of these m⁶A-associated RBPs with some known m⁶A effectors and compared these to other RBPs. We observed a significant correlation in the protein expression level rather than the gene expression level, which suggested that the m⁶A-associated RBPs participate in potential pathways at the protein-level in gene regulation. On the other hand, we built a Random Forest classification model for the reproducible m⁶A regions using RBPs' binding data in an effort to understand how RBPs contribute to the profiling of m⁶A regions. This RBP-based predictor demonstrated competitive performance when compared with sequence-based methods. Furthermore, the feature importance inferred from this model can be used to reflect m⁶A's action of repelling against RBPs. These results suggested that this framework could enable researchers to infer interaction between m⁶A and m⁶A-associated RBPs beyond sequence level when utilizing RBPs' binding data.

2. MATERIALS AND METHODS

2.1. MeRIP-Seq Data Collection and Processing

To obtain MeRIP-Seq data of which cell lines are also available for RBPs' binding data, we manually searched GEO database (Barrett et al., 2013) and finally collected raw MeRIP-Seq FASTA files from four independent studies using human HEK293T cell line (human embryonic kidney 293 cells) from European Nucleotide Archive with accession numbers SRP090687 (Lichinchi et al., 2016), SRP039397 (Schwartz et al., 2014), SRP007335 (Meyer et al., 2012), and SRP162223. We also collected MeRIP-Seq data from four independent studies using mouse embryonic fibroblasts (MEF) with accession numbers SRP039402 (Schwartz et al., 2014), SRP048596 (Geula et al., 2015), SRP115436 (Zhou et al., 2018), and SRP061617 (Zhou et al., 2015).

We pre-processed MeRIP-Seq data by using FastQC (Andrews et al., 2012) for quality control and Cutadapt (Martin, 2011) for adapter-trimming. Then, we used MoAIMS, a transcriptome-based peak-calling tool, to detect m⁶A regions with steps including mapping, keeping uniquely mapped reads, sorting, and marking duplicates (Zhang and Hamada, 2020). MoAIMS is an

efficient software we developed based on a statistical framework of a mixture negative-binomial distribution. We run MoAIMS with default parameters except that we set `sep_bin_info=F` when analyzing studies with replicates. MoAIMS called enriched regions at 200-bp resolution as default, therefore we obtained

m⁶A-enriched regions with a size of 200 bp for each MeRIP-Seq sample, and then we identified reproducible m⁶A regions using the criteria that regions are called in at least 60% of the replicates in any one study and further in at least three studies.

2.2. The Enrichment Analysis

We retrieved binding site data of RBPs from the POSTAR2 database (Zhu et al., 2019) and identified RBPs enriched in the reproducible m⁶A regions. A permutation test was adopted to assess the significance of RBP's binding in the m⁶A regions. The rest of regions in genes with m⁶A was used as control and then sampled 1,000 times. We kept the ratio of the number of bins in exons to the number of bins spanning exons the same for both m⁶A and control regions to avoid the regions' position being a confounding factor. For each RBP, we calculated the enrichment ratio using the Equation (1) where N_t is the number of m⁶A regions with the RBP and $E(N_c)$ is the average number of control regions with the RBP from 1,000 times of sampling. Then, a p -value was calculated as the proportion of N_c which were equal to or greater than N_t . After that, multiple testing was performed using Benjamini and Hochberg (1995).

$$R = \frac{N_t}{E(N_c)} \quad (1)$$

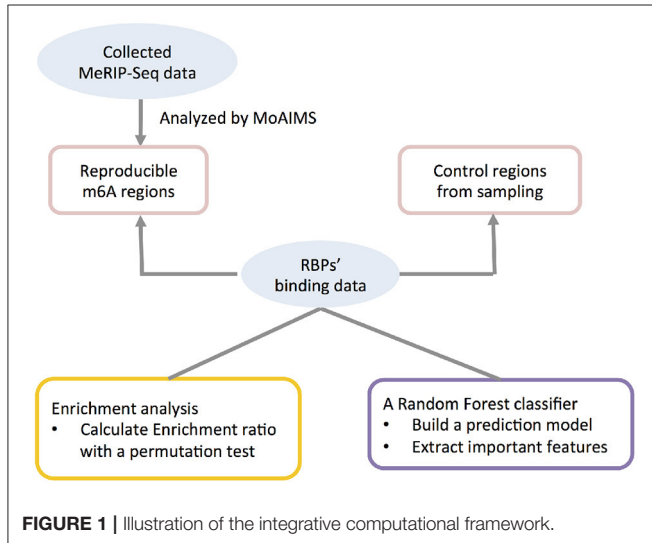


TABLE 1 | RNA binding proteins (RBPs) enriched in reproducible m⁶A regions.

HEK293T	Enrichment ratios*	# m ⁶ A regions with RBPs	p-value**	FDR adjusted p-value
YTHDF2	3.90	6,964	<0.001	<0.003
RBM15	2.73	3,534	<0.001	<0.003
YTHDF3	2.70	52	<0.001	<0.003
YTHDF1	2.49	9,196	<0.001	<0.003
RBM15B	2.32	6,375	<0.001	<0.003
YTHDC1	2.15	7,224	<0.001	<0.003
EIF3D	1.88	593	<0.001	<0.003
NOP58	1.74	159	<0.001	<0.003
HNRNPH1	1.57	47	0.002	0.006
NUDT21	1.48	5,201	<0.001	<0.003
FMR1	1.46	4,443	<0.001	<0.003
DDX3X	1.44	9,470	<0.001	<0.003
EIF3A	1.39	293	<0.001	<0.003
CPSF6	1.34	3,593	<0.001	<0.003
CPSF7	1.31	4,413	<0.001	<0.003
MEF	Enrichment ratio*	# m ⁶ A regions with RBPs	p-value**	FDR adjusted p-value
RBM3	5.81	485	<0.001	<0.001
CREBBP	2.47	24	<0.001	<0.001
SRSF2	2.24	793	<0.001	<0.001
SRSF1	2.13	467	<0.001	<0.001
CPSF6	2.07	94	<0.001	<0.001
CIRBP	1.76	401	<0.001	<0.001

*RBPs are ranked by their enrichment ratios. **P-values were calculated from 1,000 times of permutation. When p-value is zero, it is shown in the table as < 0.001 because it is possible that the p-value is actually <0.001 if times of permutation were increased.

2.3. The Classification Model

We built a Random Forest (RF) classifier to evaluate how much RBPs contribute in discriminating reproducible m⁶A regions. We used the human m⁶A regions with RBPs' binding as the positive data (13,978 in total) and generated 10 sets of control data from the control regions which were set to be an equal data size. We kept the ratio of the number of bins in exons to the number of bins spanning exons the same in both m⁶A and control regions. The binding information (1 for binding, 0 for non-binding) of RBPs was used as the input features. The data was divided into training and test groups at a ratio of 80:20. We implemented the RF classifier using the R package caret (Kuhn, 2008) and randomForest (Liaw and Wiener, 2002) with 5-fold cross validations and “mtry” (the tuning parameters) as 8 (nearly the square root of the number of features). We used the accuracy to measure the performance of the models as shown in the Equation (2) where TP is true positive, TN is true negative, FP is false positive and FN is false negative.

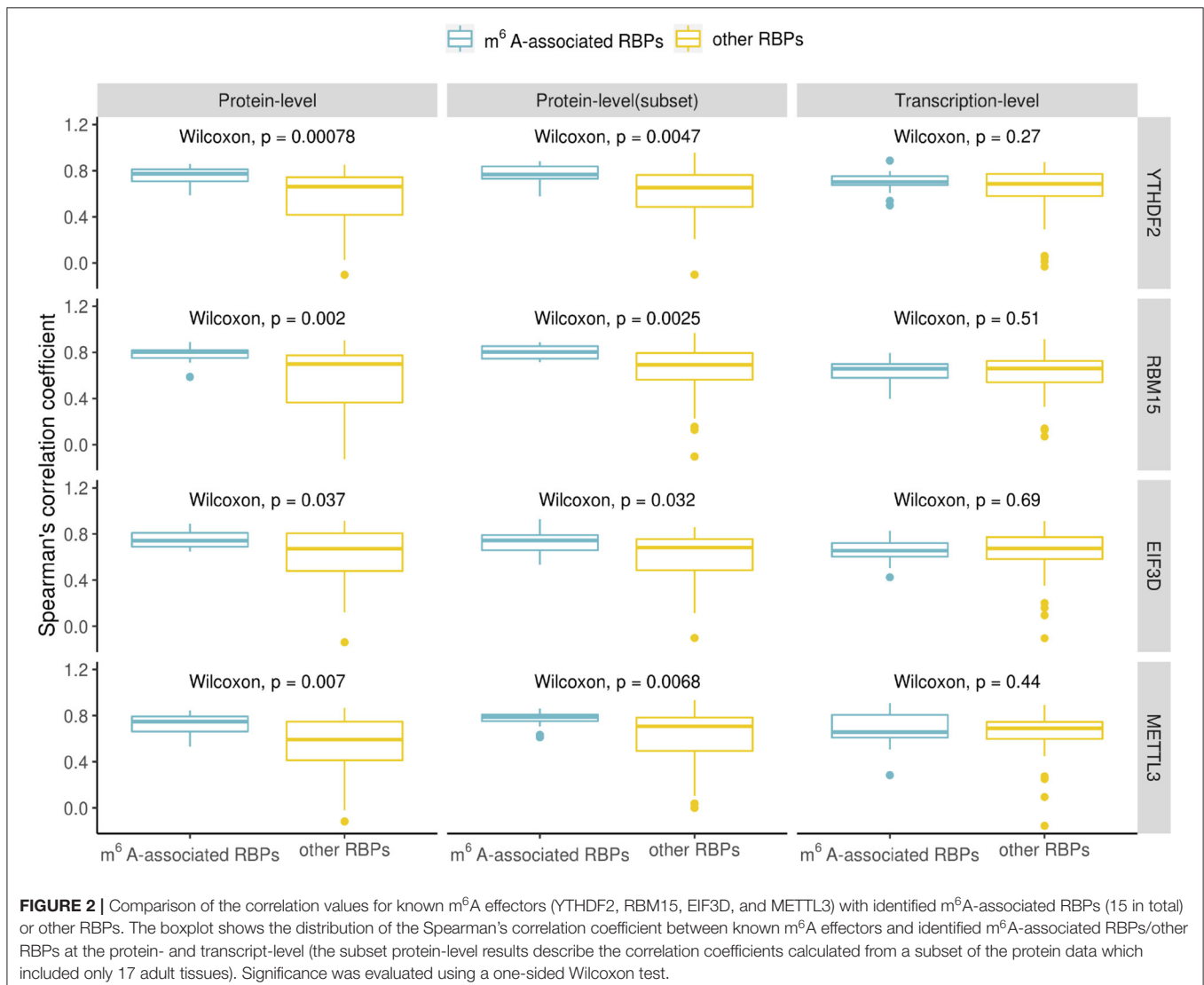
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

A analysis framework including the procedures above was summarized in **Figure 1**.

3. RESULTS

3.1. Identification of m⁶A-Associated RBPs Enriched in Reproducible m⁶A Regions

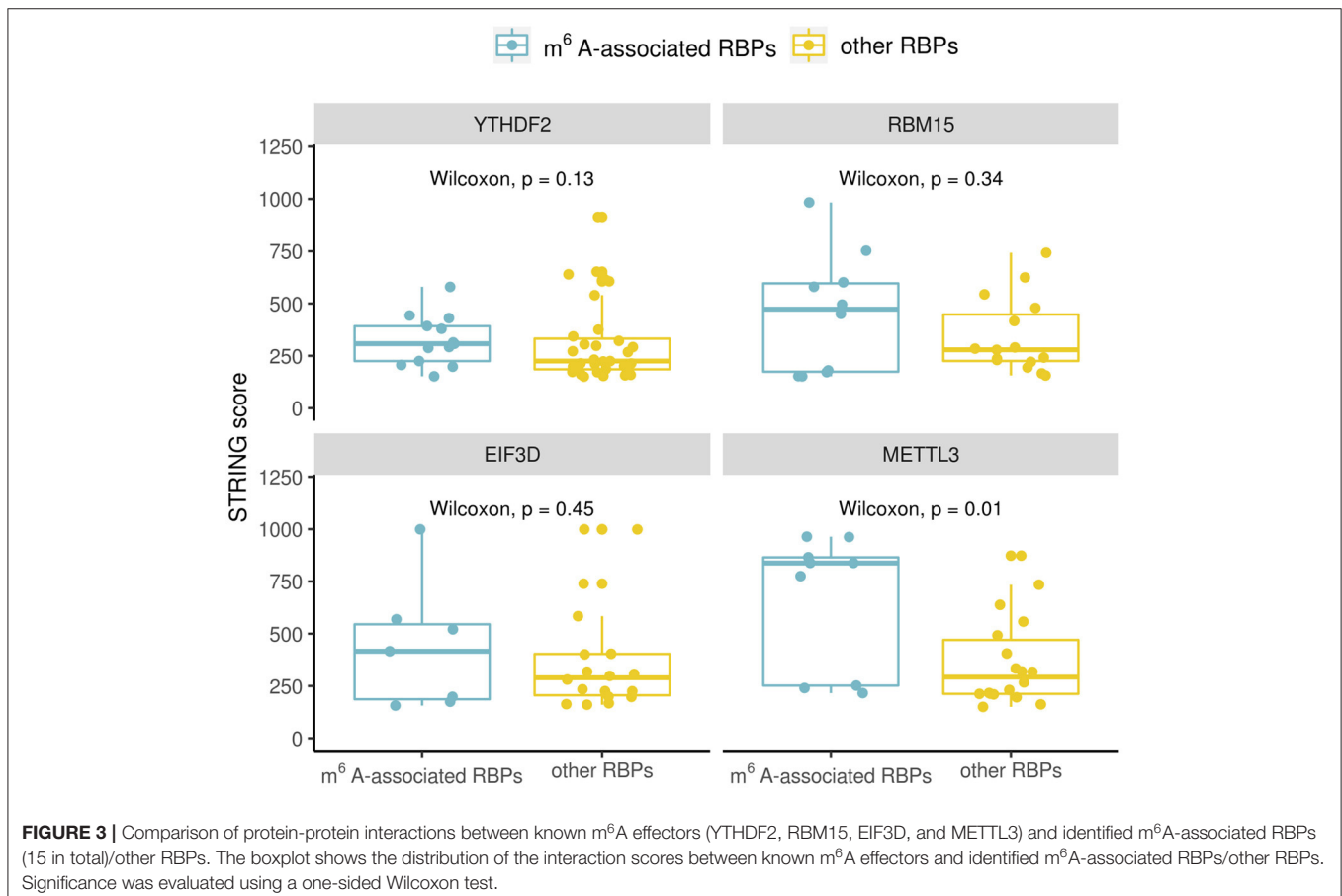
Because of the considerable variation in the m⁶A datasets (McIntyre et al., 2020), we generated *reproducible* m⁶A regions by collecting MeRIP-Seq data from nine samples of human HEK293T cell line of four independent studies and six samples of mouse MEF cell line of four independent studies. The details of detection of these m⁶A regions are provided in section 2. With a relatively strict criteria, we finally obtained 14,803 reproducible m⁶A regions for HEK293T cell line and 5,576 reproducible m⁶A regions for MEF cell line.



To identify RBPs enriched in m⁶A regions, 71 RBPs for HEK293T/HEK293 and nine RBPs for MEF were retrieved from the POSTAR2 database. For each RBP, we calculated an enrichment score and assessed its significance using a permutation test as described in section 2. When setting the threshold for the enrichment ratio to ≥ 1.3 and FDR (false discovery rate) adjusted *p*-value to ≤ 0.05 , we obtained enriched RBPs listed in **Table 1**. For HEK293T, we identified several known m⁶A readers including YTH family proteins, FMR1, EIF3, and m⁶A writers RBM15/15B, a component of the WTAP-METTL3 complex (Patil et al., 2016; Lee et al., 2020). For MEF, we found a common RBP, CPSF6, which is enriched for both human and mouse. CPSF6 is a polyadenylation cleavage factor and has been reported to be associated with VIRMA, which mediates preferential m⁶A methylation in the 3' UTR and near stop codon and participates alternative polyadenylation (APA) in human (Yue et al., 2018). Another study found YTHDC1's association with CPSF6 during mouse oocyte development (Kasowitz et al., 2018). In addition, we noticed that RBM3 was highly enriched in m⁶A regions of MEF. RBM3 is an important regulator of circadian gene expression by controlling APA (Liu et al., 2013), therefore we suggest that RBM3 could be associated with m⁶A in the APA regulation process. The full list of enrichment ratios for each of the RBPs is provided in **Supplementary Tables 1, 2**. Besides, for each enriched RBP

(overlap with more than 100 m⁶A regions), we also listed the RBPs that more than 60% of the enriched RBP is overlapped with for HEK293T in **Supplementary Table 3**. As expected, YTHDF1 and DDX3X were shown to have the highest overlapping percentage as they have a considerable overlap with m⁶A regions.

The RBPs in **Table 1** are considered as m⁶A-associated RBPs, therefore we wondered how they are correlated with known m⁶A effectors when compared with other RBPs at both the transcription and the protein expression level. We performed a correlation analysis for all the human RBPs. To do the correlation analysis at the transcription level, we downloaded Illumina Body Map (HBM) (Asmann et al., 2012; Barbosa-Morais et al., 2012; Derrien et al., 2012) from ArrayExpress (Athar et al., 2019) with the accession number E-MTAB-513, which provides gene expression data for 16 human tissues. For the correlation analysis at the protein level, we downloaded mass spectrometry data from Human Proteome Map (HPM) (Kim et al., 2014) for 30 human tissues/cell lines. We checked some known m⁶A effectors including YTHDF2, RBM15, EIF3D which ranked at the top of **Table 1** and METTL3 of which binding data is not available but is a well-known m⁶A writer, and compared their correlation with the identified m⁶A-associated RBPs (15 in total) or with the rest of RBPs (56 in total). Correlation was calculated using the Spearman's correlation coefficient. We observed a similar trend in all the investigated known m⁶A effectors which showed that



the identified m⁶A-associated RBPs are more correlated with them at the protein-level than the transcription level (Figure 2). Because the protein data included more tissues/cell lines than the transcription data, we chose to compare a subset of 17 adult tissues to check the correlation values for avoiding any biased introduced by different dataset sizes. The higher correlation at the protein level was still observed in this subset evaluation as shown in Figure 2. Some studies have reported cases of gene regulation with dependency between m⁶A-associated RBPs such as METTL3 and YTHDF2 (Chen et al., 2018; Kasowitz et al., 2018). This observation supports the hypothesis that

m⁶A-associated RBPs are more likely to participate in potential pathways at the protein level. Then, we went on to confirm to if these higher correlation values are the result of protein-protein interactions. To do this we retrieved the protein-protein interaction data from STRING (von Mering et al., 2005). The available interaction scores do not show significant difference between m⁶A-associated RBPs and other RBPs except for METTL3 (Figure 3). Because the protein-protein interaction data is still limited, from the available data it is suggested that the higher correlation at the protein-level is marginally related to protein-protein interaction. m⁶A modification is a dynamic

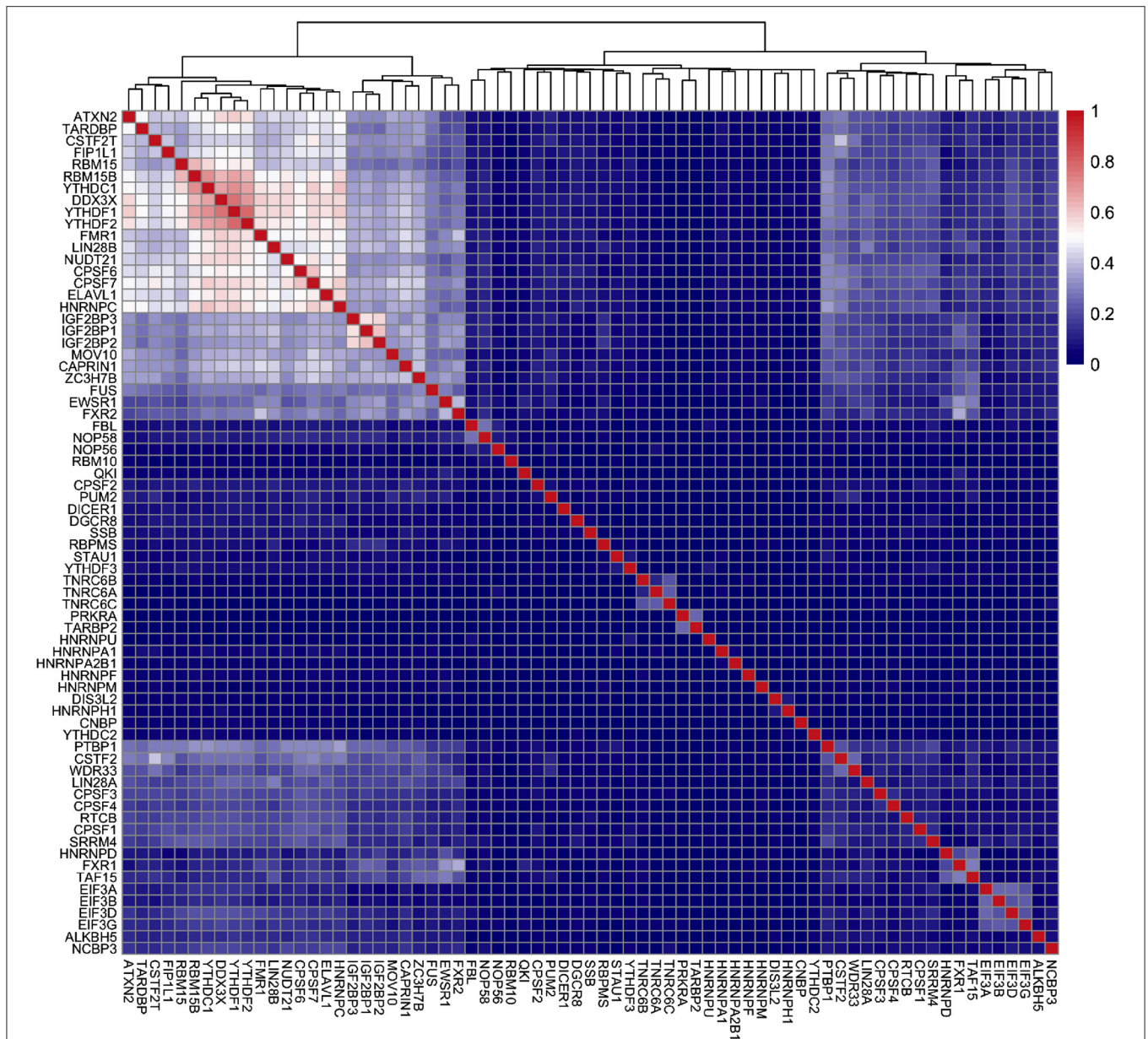


FIGURE 4 | Clustering of RNA binding proteins (RBPs) in the m⁶A regions of HEK293T cell line. X-axis and Y-axis represent the names of the RBPs. The color scale indicates the cosine similarity between the RBPs.

process involving both temporal and spatial regulation between m⁶A effectors, therefore it is expected to have further studies to unveil the regulation mechanism of these proteins.

3.2. Identification of m⁶A-Associated RBPs Contributing to the Classification of m⁶A Regions

After we identified RBPs enriched in the reproducible m⁶A regions, we wanted to develop a more comprehensive

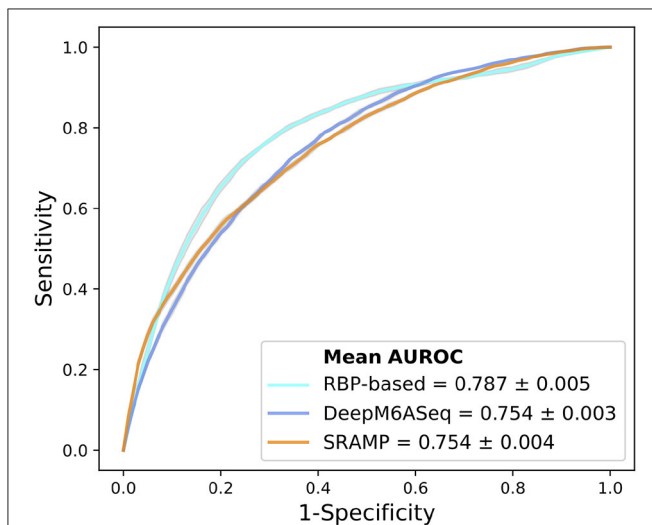


FIGURE 5 | Comparison of AUROC between the RBPs (RNA binding proteins)-based predictor, DeepM6ASeq, and SRAMP in mature mRNA mode for the classification of HEK293T m⁶A regions. The plot represents average ROC from ten times of sampling control regions for each predictor.

understanding of how RBPs' binding contributes to the profile of m⁶A regions. To do this, we performed a further analysis on the human RBPs. First, we investigated the overall profile of the binding information of RBPs (0 for non-binding and 1 for binding) in the reproducible m⁶A regions. We calculated the pairwise distance between RBPs using cosine similarity and performed clustering (**Figure 4**). The result of the clustering analysis demonstrated the co-occurrence of YTH family proteins and RBM15B which all ranked in the top of the enrichment analysis. Then, we built a Random Forest classifier which incorporated the binding information for each of the RBPs as features. The details of models are described in section 2. The classifier achieved an average accuracy of 0.736 and AUROC (Area Under Receiver Operating Characteristic) of 0.788 as shown in **Figure 5**. We also compared the RBP-based classifier with two sequence-based predictors SRAMP (Zhou et al., 2016) in mature mRNA mode and DeepM6ASeq which showed an accuracy of 0.660 and 0.686, respectively and AUROC of 0.754 for both (**Figure 5**). We plotted top 10 most important features as shown in **Figure 6** and among them found the enriched m⁶A-associated RBPs such as the readers YTHDF1/2, YTHDC1, the writers RBM15/15B. Besides, it is noticed that ELAVL1 also has contribution to the classification of m⁶A regions to some extent. ELAVL1 is reported to have action of being repelled by m⁶A in general that can lead to RNA decay (Wang et al., 2014; Lee et al., 2020). The repelling action of m⁶A against ELAVL1 is consistent with the enrichment results, which show that its enrichment ratio is 0.816. In summary, the RBP-based classifier not only demonstrated competitive performance in the prediction of reproducible m⁶A regions but also helped to infer interaction between m⁶A and m⁶A-associated RBPs beyond sequence level when combined with the results of the enrichment analysis.

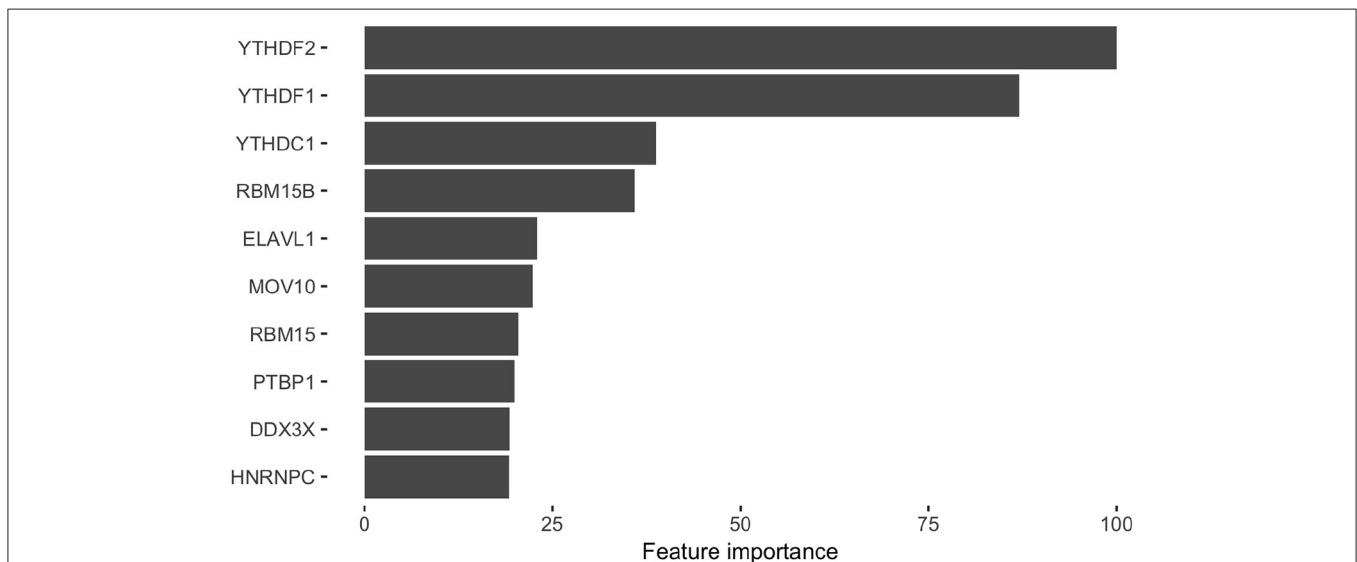


FIGURE 6 | Top 10 RNA binding proteins (RBPs) identified from the classification of the HEK293T reproducible m⁶A regions. The bar graph shows the top 10 RBPs extracted from the classifier for the m⁶A regions. X-axis represents the name of RBPs and Y-axis represents the average importance score from ten times of sampling control regions.

4. DISCUSSION

Utilizing the binding information of RBPs, this computational framework enabled us to identify potential m⁶A-associated RBPs and infer their interaction with m⁶A. This analysis serves as a first step, and future analyses may include some improvements and expansions. First, this framework was designed and tested on a limited number of cell types and organisms. With the increasing amount of data available for m⁶A and RBPs in more cell lines and tissues, this framework could be tested on much larger datasets and may provide valuable insights into the m⁶A regulatory network. Especially, this framework is promising in the application of cancer research. Several studies have identified function of m⁶A effectors like METTL3/14, YTHDF1/2, and IGF2BP1 in multiple cancer types (Cui et al., 2017; Li et al., 2018; Chen et al., 2019; Han et al., 2019; Müller et al., 2019). This framework is expected to provide clues for potential m⁶A effectors and the interaction among them in cancer research. In addition, this framework could be applied to other RNA modifications such as N1-methyladenosine (m1A) (Dominissini et al., 2016), and 5-methylcytosine (m5C) (Amort et al., 2017), which have also been identified as critical RNA modification. Such analyses could help improve experimental design in wet lab applications and help researchers narrow their focus. Third, apart from RBPs, other genomic features like transcription factors and histone modification are worth inspecting for studying the m⁶A regulation networks at multiple layers. These applications highlight the future utility of this framework and its value in the current research climate.

5. CONCLUSION

We designed an integrative computational framework for identification of m⁶A-associated RBPs in reproducible m⁶A regions. This computational framework is composed of an enrichment analysis and a classification model. Using the enrichment analysis, we were able to identify known m⁶A-associated RBPs and several potential ones including RBM3 from mouse. These identified m⁶A-associated RBPs show a significant degree of correlation at their protein level, although this is not seen in their transcriptional profile, which suggests that these m⁶A-associated RBPs participate in potential pathways at the protein-level in gene regulation. On the other hand, we built a classification model for m⁶A regions using a Random

Forest algorithm that uses RBPs' binding information as its input features. The RBP-based predictor not only demonstrated comparable performance to sequence-based predictions but also helped infer interaction between m⁶A and m⁶A-associated RBPs like actions of reading and repelling beyond sequence level. We hope that this analysis framework can assist biologists in their study of RNA modifications.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

YZ conceived this study, implemented the methods, performed the experiments, and wrote the draft. MH supervised this study and revised the manuscript critically. YZ and MH contributed to the construction of methods and analysis/interpretation of the data. All authors read and approved the final manuscript.

FUNDING

Publication costs are funded by Waseda University [basic research budget]. This study was supported by the Ministry of Education, Culture, Sports, Science and Technology (KAKENHI) [grant numbers JP17K20032, JP16H05879, JP16H01318, and JP16H02484 to MH]. The funding bodies did not play any role in the design of the study or collection, analysis and interpretation of data or in writing the manuscript.

ACKNOWLEDGMENTS

Computation was performed by server facilities at AIST-Waseda University Computational Bio Big-Data Open Innovation Laboratory (CBBB-OIL).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.625797/full#supplementary-material>

REFERENCES

- Amort, T., Rieder, D., Wille, A., Khokhlova-Cubberley, D., Riml, C., Trixl, L., et al. (2017). Distinct 5-methylcytosine profiles in poly(A) RNA from mouse embryonic stem cells and brain. *Genome Biol.* 18:1. doi: 10.1186/s13059-016-1139-1
- An, S., Huang, W., Huang, X., Cun, Y., Cheng, W., Sun, X., et al. (2020). Integrative network analysis identifies cell-specific trans regulators of m6A. *Nucleic Acids Res.* 48, 1715–1729. doi: 10.1093/nar/gkz1206
- Andrews, S., Krueger, F., Segonds-Pichon, A., Biggins, L., Krueger, C., and Wingett, S. (2012). *FastQC*. Babraham Institute.
- Asmann, Y. W., Necela, B. M., Kalari, K. R., Hossain, A., Baker, T. R., Carr, J. M., et al. (2012). Detection of redundant fusion transcripts as biomarkers or disease-specific therapeutic targets in breast cancer. *Cancer Res.* 72, 1921–1928. doi: 10.1158/0008-5472.CAN-11-3142
- Athar, A., Füllgrabe, A., George, N., Iqbal, H., Huerta, L., Ali, A., et al. (2019). ArrayExpress update - from bulk to single-cell expression data. *Nucleic Acids Res.* 47, D711–D715. doi: 10.1093/nar/gky964
- Barbieri, I., Tzelepis, K., Pandolfini, L., Shi, J., Millán-Zambrano, G., Robson, S. C., et al. (2017). Promoter-bound METTL3 maintains myeloid leukaemia by m6A-dependent translation control. *Nature* 552, 126–131. doi: 10.1038/nature24678

- Barbosa-Morais, N. L., Irimia, M., Pan, Q., Xiong, H. Y., Gueroussov, S., Lee, L. J., et al. (2012). The evolutionary landscape of alternative splicing in vertebrate species. *Science* 338, 1587–1593. doi: 10.1126/science.1230612
- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2013). NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* 41, D991–D995. doi: 10.1093/nar/gks1193
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Chen, M., Wei, L., Law, C. T., Tsang, F. H., Shen, J., Cheng, C. L., et al. (2018). RNA N6-methyladenosine methyltransferase-like 3 promotes liver cancer progression through YTHDF2-dependent posttranscriptional silencing of SOCS2. *Hepatology* 67, 2254–2270. doi: 10.1002/hep.29683
- Chen, X. Y., Zhang, J., and Zhu, J. S. (2019). The role of m6A RNA methylation in human cancer. *Mol. Cancer* 18:103. doi: 10.1186/s12943-019-1033-z
- Cui, Q., Shi, H., Ye, P., Li, L., Qu, Q., Sun, G., et al. (2017). m6A RNA methylation regulates the self-renewal and tumorigenesis of glioblastoma stem cells. *Cell Rep.* 18, 2622–2634. doi: 10.1016/j.celrep.2017.02.059
- Deng, S., Zhang, H., Zhu, K., Li, X., Ye, Y., Li, R., et al. (2020). M6A2Target: a comprehensive database for targets of m6A writers, erasers and readers. *Brief. Bioinform.* bbaa055. doi: 10.1093/bib/bbaa055
- Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 22, 1775–1789. doi: 10.1101/gr.132159.111
- Dominissini, D., Nachtergaale, S., Moshitch-Moshkovitz, S., Peer, E., Kol, N., Ben-Haim, M. S., et al. (2016). The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA. *Nature* 530, 441–446. doi: 10.1038/nature16998
- Edupuganti, R. R., Geiger, S., Lindeboom, R. G. H., Shi, H., Hsu, P. J., Lu, Z., et al. (2017). N6-methyladenosine (m6A) recruits and repels proteins to regulate mRNA homeostasis. *Nat. Struct. Mol. Biol.* 24, 870–878. doi: 10.1038/nsmb.3462
- Geula, S., Moshitch-Moshkovitz, S., Dominissini, D., Mansour, A. A., Kol, N., Salmon-Divon, M., et al. (2015). Stem cells. m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. *Science* 347, 1002–1006. doi: 10.1126/science.1261417
- Han, D., Liu, J., Chen, C., Dong, L., Liu, Y., Chang, R., et al. (2019). Anti-tumour immunity controlled through mRNA m6A methylation and YTHDF1 in dendritic cells. *Nature* 566, 270–274. doi: 10.1038/s41586-019-0916-x
- Hou, J., Zhang, H., Liu, J., Zhao, Z., Wang, J., Lu, Z., et al. (2019). YTHDF2 reduction fuels inflammation and vascular abnormalization in hepatocellular carcinoma. *Mol. Cancer* 18:163. doi: 10.1186/s12943-019-1082-3
- Kasowitz, S. D., Ma, J., Anderson, S. J., Leu, N. A., Xu, Y., Gregory, B. D., et al. (2018). Nuclear m6A reader YTHDC1 regulates alternative polyadenylation and splicing during mouse oocyte development. *PLoS Genet.* 14:e1007412. doi: 10.1371/journal.pgen.1007412
- Kim, M. S., Pinto, S. M., Getnet, D., Nirujogi, R. S., Manda, S. S., Chaerkady, R., et al. (2014). A draft map of the human proteome. *Nature* 509, 575–581. doi: 10.1038/nature13302
- Kuhn, M. (2008). Building predictive models in R using the caret package. *J. Stat. Softw.* 28, 1–26. doi: 10.18637/jss.v028.i05
- Lee, Y., Choe, J., Park, O. H., and Kim, Y. K. (2020). Molecular mechanisms driving mRNA degradation by m6A modification. *Trends Genet.* 36, 177–188. doi: 10.1016/j.tig.2019.12.007
- Li, Z., Qian, P., Shao, W., Shi, H., He, X. C., Gogol, M., et al. (2018). Suppression of m6A reader Ythdf2 promotes hematopoietic stem cell expansion. *Cell Res.* 28, 904–917. doi: 10.1038/s41422-018-0072-0
- Liaw, A., and Wiener, M. (2002). Classification and regression by randomforest. *R News* 2, 18–22.
- Lichinchi, G., Zhao, B. S., Wu, Y., Lu, Z., Qin, Y., He, C., and Rana, T. M. (2016). Dynamics of human and viral RNA methylation during Zika virus infection. *Cell Host Microbe* 20, 666–673. doi: 10.1016/j.chom.2016.10.002
- Linder, B., Grozhik, A. V., Olarerin-George, A. O., Meydan, C., Mason, C. E., and Jaffrey, S. R. (2015). Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat. Methods* 12, 767–772. doi: 10.1038/nmeth.3453
- Liu, Y., Hu, W., Murakawa, Y., Yin, J., Wang, G., Landthaler, M., et al. (2013). Cold-induced RNA-binding proteins regulate circadian gene expression by controlling alternative polyadenylation. *Sci. Rep.* 3:2054. doi: 10.1038/srep02054
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* 17, 10–12. doi: 10.14806/ej.17.1.200
- McIntyre, A. B. R., Gokhale, N. S., Cerchietti, L., Jaffrey, S. R., Horner, S. M., and Mason, C. E. (2020). Limits in the detection of m6A changes using MeRIP/m6A-seq. *Sci. Rep.* 10:6590. doi: 10.1038/s41598-020-63355-3
- Meyer, K. D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C. E., and Jaffrey, S. R. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* 149, 1635–1646. doi: 10.1016/j.cell.2012.05.003
- Müller, S., Glaß, M., Singh, A. K., Haase, J., Bley, N., Fuchs, T., et al. (2019). IGF2BP1 promotes SRF-dependent transcription in cancer in a m6A- and miRNA-dependent manner. *Nucleic Acids Res.* 47, 375–390. doi: 10.1093/nar/gky1012
- Nie, F., Feng, P., Song, X., Wu, M., Tang, Q., and Chen, W. (2020). RNAWRE: a resource of writers, readers and erasers of RNA modifications. *Database* 2020:baaa049. doi: 10.1093/database/baaa049
- Patil, D. P., Chen, C. K., Pickering, B. F., Chow, A., Jackson, C., Guttman, M., et al. (2016). m(6)A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* 537, 369–373. doi: 10.1038/nature19342
- Roundtree, I. A., Evans, M. E., Pan, T., and He, C. (2017). Dynamic RNA modifications in gene expression regulation. *Cell* 169, 1187–1200. doi: 10.1016/j.cell.2017.05.045
- Schwartz, S., Mumbach, M. R., Jovanovic, M., Wang, T., Maciag, K., Bushkin, G. G., et al. (2014). Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5' sites. *Cell Rep.* 8, 284–296. doi: 10.1016/j.celrep.2014.05.048
- Shi, H., Wei, J., and He, C. (2019). Where, when, and how: context-dependent functions of RNA methylation writers, readers, and erasers. *Mol. Cell* 74, 640–650. doi: 10.1016/j.molcel.2019.04.025
- Ule, J., Jensen, K., Mele, A., and Darnell, R. B. (2005). CLIP: a method for identifying protein-RNA interaction sites in living cells. *Methods* 37, 376–386. doi: 10.1016/j.ymeth.2005.07.018
- von Mering, C., Jensen, L. J., Snel, B., Hooper, S. D., Krupp, M., Foglierini, M., et al. (2005). STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res.* 33, D433–D437. doi: 10.1093/nar/gki005
- Vu, L. P., Pickering, B. F., Cheng, Y., Zaccara, S., Nguyen, D., Minuesa, G., et al. (2017). The N6-methyladenosine (m6A)-forming enzyme METTL3 controls myeloid differentiation of normal hematopoietic and leukemia cells. *Nat. Med.* 23, 1369–1376. doi: 10.1038/nm.4416
- Wang, J., and Wang, L. (2020). Deep analysis of RNA N6-adenosine methylation (m6A) patterns in human cells. *NAR Genomics Bioinform.* 2:lqaa007. doi: 10.1093/nargab/lqaa007
- Wang, Y., Li, Y., Toth, J. I., Petroski, M. D., Zhang, Z., and Zhao, J. C. (2014). N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat. Cell Biol.* 16, 191–198. doi: 10.1038/ncb2902
- Yue, Y., Liu, J., Cui, X., Cao, J., Luo, G., Zhang, Z., et al. (2018). VIRMA mediates preferential m6A mRNA methylation in 3'UTR and near stop codon and associates with alternative polyadenylation. *Cell Discov.* 4:10. doi: 10.1038/s41421-018-0019-0
- Zhang, Y., and Hamada, M. (2018). DeepM6ASeq: prediction and characterization of m6A-containing sequences using deep learning. *BMC Bioinformatics* 19(Suppl. 19):524. doi: 10.1186/s12859-018-2516-4
- Zhang, Y., and Hamada, M. (2020). MoAIMS: efficient software for detection of enriched regions of MeRIP-Seq. *BMC Bioinformatics* 21:103. doi: 10.1186/s12859-020-3430-0
- Zhao, J., Ohsumi, T. K., Kung, J. T., Ogawa, Y., Grau, D. J., Sarma, K., et al. (2010). Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol. Cell* 40, 939–953. doi: 10.1016/j.molcel.2010.12.011
- Zhen, D., Wu, Y., Zhang, Y., Chen, K., Song, B., Xu, H., et al. (2020). m6A Reader: epitranscriptome target prediction and functional characterization of N6-methyladenosine (m6A) readers. *Front. Cell. Dev. Biol.* 8:741. doi: 10.3389/fcell.2020.00741
- Zhou, J., Wan, J., Gao, X., Zhang, X., Jaffrey, S. R., and Qian, S. B. (2015). Dynamic m(6)A mRNA methylation directs translational control of heat shock response. *Nature* 526, 591–594. doi: 10.1038/nature15377

- Zhou, J., Wan, J., Shu, X. E., Mao, Y., Liu, X. M., Yuan, X., et al. (2018). N6-methyladenosine guides mRNA alternative translation during integrated stress response. *Mol. Cell* 69, 636–647. doi: 10.1016/j.molcel.2018.01.019
- Zhou, Y., Zeng, P., Li, Y. H., Zhang, Z., and Cui, Q. (2016). SRAMP: prediction of mammalian N6-methyladenosine (m6A) sites based on sequence-derived features. *Nucleic Acids Res.* 44:e91. doi: 10.1093/nar/gkw104
- Zhu, Y., Xu, G., Yang, Y. T., Xu, Z., Chen, X., Shi, B., et al. (2019). POSTAR2: deciphering the post-transcriptional regulatory logics. *Nucleic Acids Res.* 47, D203–D211. doi: 10.1093/nar/gky830

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Zhang and Hamada. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.