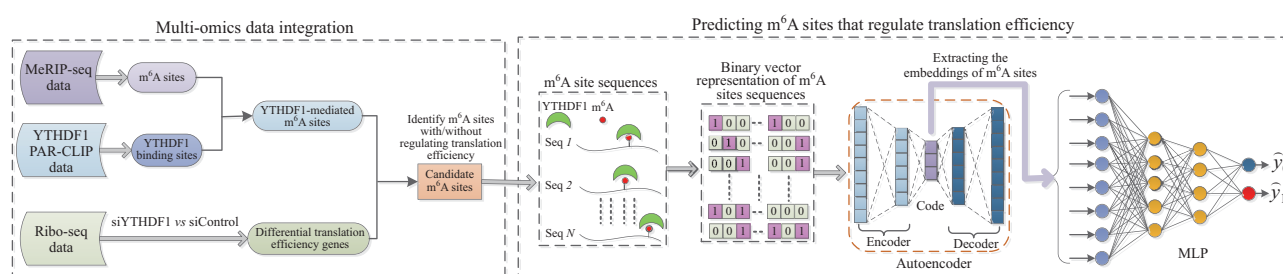




# m<sup>6</sup>ATEpre: Predicting YTHDF1-mediated mRNA Translation Efficiency Regulated by m<sup>6</sup>A Sites via Multi-omics Data Integration\*

ZHANG Teng<sup>1)</sup>, ZHANG Ming<sup>1)</sup>, ZHANG Shao-Wu<sup>2)</sup>, LIU Lian<sup>3)\*\*</sup><sup>1)</sup>School of Computer, Jiangsu University of Science and Technology, Zhenjiang 212100, China;<sup>2)</sup>MOE Key Laboratory of Information Fusion Technology, School of Automation, Northwestern Polytechnical University, Xi'an 710072, China;<sup>3)</sup>School of Artificial Intelligence and Computer Science, Shaanxi Normal University, Xi'an 710119, China)

## Graphical abstract



**Abstract Objective** The most prevalent mRNA modification, N6-methyladenosine (m<sup>6</sup>A) plays an important role in various RNA metabolism, including gene expression and translation. By recruiting different “reader” proteins and their cofactors, m<sup>6</sup>A modification can affect messenger RNA (mRNA) degradation, splicing, nuclear export and translation. However, the selective mechanism by which m<sup>6</sup>A sites regulate mRNA translation through m<sup>6</sup>A reader YTHDF1 binding remains poorly understood, due to a lack of computational methods for identifying context-specific m<sup>6</sup>A sites that regulate translation. To address this, we developed a novel computational framework named m<sup>6</sup>ATEpre, the first tool designed to predict cell-specific m<sup>6</sup>A sites that regulate translation efficiency. **Methods** m<sup>6</sup>ATEpre integrates multi-omics data, introduces a novel feature representation strategy for m<sup>6</sup>A site sequences, and employs an autoencoder to effectively capture embedded feature representations. Specifically, m<sup>6</sup>ATEpre first integrated MeRIP-seq data and PAR-CLIP data through overlapping m<sup>6</sup>A sites with YTHDF1 binding sites and identified YTHDF1-mediated m<sup>6</sup>A sites. Then, m<sup>6</sup>ATEpre detected the translation gene by analyzing the Ribo-seq data under YTHDF1 knockdown vs control condition. Genes whose translation is mediated by YTHDF1 in an m<sup>6</sup>A-dependent manner were identified by a significant decrease in translation efficiency upon YTHDF1 knockdown. Next, we proposed a binary vector indicating the presence or absence of YTHDF1 binding motifs to characterize each m<sup>6</sup>A site sequence. This represents a novel feature representation strategy for m<sup>6</sup>A sites. m<sup>6</sup>ATEpre utilized the autoencoder to extract the potentially important feature representations and constructed a multilayer perceptron neural networks model to predict potential m<sup>6</sup>A sites that regulating translation efficiency. **Results** A comprehensive evaluation of m<sup>6</sup>ATEpre was conducted through a series of experiments. We compared its performance against that of a similar

\* This work was supported by grants from the Natural Science Basic Research Program of Shaanxi, China (2024JCYBQN-0624) and the Fundamental Research Funds for the Central Universities, Shaanxi Normal University (GK202406008).

\*\* Corresponding author.

Tel: 86-29-85310161, E-mail: lian.l@snnu.edu.cn

Received: January 4, 2026 Accepted: March 17, 2026

prediction task model, as well as other classifiers. The results indicate that m<sup>6</sup>ATEpre achieved the best prediction performance. In addition, we analyzed different feature representation strategies and performed ablation experiments to validate the rationality of the model design. The results demonstrate that our proposed feature representation strategy has a greater advantage in improving prediction performance. In the HeLa cell line, bioinformatic analysis of the metagene distribution and sequence minimum free energy of m<sup>6</sup>A sites regulating translation efficiency (m<sup>6</sup>A-reg-TE sites) revealed their specific properties in translation regulation. Functional enrichment analysis indicated that m<sup>6</sup>A-reg-TE genes are associated with specific biological processes and KEGG pathways. By integrating the binding sites of YTHDF1 co-factors with m<sup>6</sup>A-reg-TE sites, we revealed that YTHDF1-mediated and m<sup>6</sup>A-dependent translation efficiency regulation requires the cooperation of multiple translation-regulatory RNA-binding proteins among its co-factors in the HeLa cell line. Furthermore, we extended our predictions to the dataset of the HEK293T cell line. Similarly, bioinformatic analysis of the metagene distribution and functional enrichment revealed the cell-specific characteristic of these predicted m<sup>6</sup>A-reg-TE sites in HEK293T cells. Likewise, integrated analysis of multiple YTHDF1 co-factors and m<sup>6</sup>A-reg-TE sites predicted in the HEK293T cell line reveals their m<sup>6</sup>A-dependent cooperation in regulating translation efficiency.

**Conclusion** m<sup>6</sup>ATEpre is a timely tool that will advance our understanding of the mechanisms of m<sup>6</sup>A regulation in translation efficiency. The source code and datasets used in this work can be downloaded from <https://www.scidb.cn/s/bAZZFr>.

**Key words** m<sup>6</sup>A modification, YTHDF1-mediated translation efficiency, multi-omics data integration, feature representation

**DOI:** 10.3724/j.pibb.2026.0005

**CSTR:** 32369.14.pibb.20260005

N6-methyladenosine (m<sup>6</sup>A), the most prevalent mRNA modification in higher eukaryotes, plays an important role in the post-transcriptional regulation of RNA metabolism<sup>[1-3]</sup> and is associated with various diseases<sup>[4-7]</sup>. The dynamic m<sup>6</sup>A modification can affect messenger RNA (mRNA) degradation, splicing, nuclear export and translation by recruiting different “reader” proteins including YTH family proteins and their cofactor RNA-binding proteins (RBPs)<sup>[8]</sup>. Recent studies have demonstrated that YTH N6-methyladenosine RNA binding protein F2 (YTHDF2) can degrade its target mRNAs in an m<sup>6</sup>A-dependent manner<sup>[9]</sup>. The m<sup>6</sup>A reader protein YTH N6-methyladenosine RNA binding protein F1 (YTHDF1) promotes the translation efficiency of m<sup>6</sup>A-modified mRNAs in HeLa cells by interacting with translation initiation factors and ribosomes<sup>[10]</sup>. In addition, insulin like growth factor 2 mRNA binding protein 1–3 stabilizes mRNA by interacting with the cofactors ELAV like RNA binding protein 1 and matrin 3<sup>[11]</sup>. The m<sup>6</sup>A-modified mRNA targets of these m<sup>6</sup>A readers participate in many biological functions and diseases<sup>[12]</sup>, through the regulatory mechanism of m<sup>6</sup>A-dependent gene expression and translation.

Recently, computational approaches or tools have been developed to investigate m<sup>6</sup>A-participating biological functions and explore m<sup>6</sup>A-mediated regulatory mechanisms. m<sup>6</sup>A-Driver<sup>[13]</sup>, hot-m6A<sup>[14]</sup> and FunDMDeep-m6A<sup>[15]</sup> have been developed to predict m<sup>6</sup>A-driven genes that regulate various biological functions<sup>[16]</sup> based on complex network

methods. Although these computational tools can identify context-specific and functionally significant m<sup>6</sup>A-mediated genes, they cannot directly determine whether these functional m<sup>6</sup>A sites affect mRNA stability, regulate gene expression and change translation efficiency. To further explore the potential m<sup>6</sup>A regulatory mechanism in gene expression, m<sup>6</sup>A-express was first developed for predicting condition-specific m<sup>6</sup>A regulation of gene expression (m<sup>6</sup>A-reg-exp) from limited methylated RNA immunoprecipitation sequencing (MeRIP-Seq) samples<sup>[17]</sup>. m<sup>6</sup>A-express employs a negative binomial regression model under Bayesian hierarchical architecture to predict potential m<sup>6</sup>A-reg-exp genes and achieves higher prediction. However, some m<sup>6</sup>A-reg-exp genes were bound by YTHDF2 or IGF2BP1–3, whose expression is more likely regulated by m<sup>6</sup>A methylation, can't be effectively predicted by m<sup>6</sup>A expression due to high variation from the limited number of samples<sup>[18]</sup>. To solve this problem, m<sup>6</sup>A express-Reader integrates the m<sup>6</sup>A reader binding information into a Bayesian hierarchical negative binomial regression model<sup>[18]</sup>, which improves the prediction performance and biological interpretation of m<sup>6</sup>A regulatory mechanism<sup>[18]</sup>. In addition, m<sup>6</sup>A express-BHM has been developed to predict m<sup>6</sup>A-reg-exp genes in more complex contexts (*e.g.*, cell differentiation at distinct stages or longitudinal methylation data) by employing a Bayesian hierarchical mixture model<sup>[19]</sup>. To understand YTHDF2-mediated mRNA degradation, researchers

adapted a language model, bidirectional encoder representations from transformers (BERT), to pre-train it on a large dataset of m<sup>6</sup>A-containing mRNA sequences and fine-tune the model (named m<sup>6</sup>A-BERT-Deg) on a multi-omics dataset integrating m<sup>6</sup>A sites, YTHDF2 binding and mRNA half-life data<sup>[20]</sup>.

The above computational methods or tools achieve high performance in m<sup>6</sup>A regulatory biological function, gene expression and mRNA degradation prediction. However, due to the lack of integrated translational omics data, these methods and tools cannot directly infer potential m<sup>6</sup>A sites or genes involved in regulating translation efficiency in a specific context.

In this paper, we developed the m<sup>6</sup>ATEpre model, the first computational tool for predicting potential m<sup>6</sup>A sites that regulate translation efficiency (m<sup>6</sup>A-reg-TE). m<sup>6</sup>ATEpre integrates multi-omics datasets from the same cell line, including MeRIP-seq data, ribosome profiling sequencing data and YTHDF1 photoactivatable ribonucleoside-enhanced cross-linking and immunoprecipitation data. The integration of multi-omics data is to select potential regulatory m<sup>6</sup>A sites. m<sup>6</sup>ATEpre innovatively utilizes the binary vector of YTHDF1 binding motif as input features, which are obtained from m<sup>6</sup>A sites' sequences. Then, the autoencoder module of m<sup>6</sup>ATEpre is used to effectively extract important feature representations. The multilayer perceptron (MLP) neural networks of m<sup>6</sup>ATEpre is selected to identify m<sup>6</sup>A-regulated translation efficiency sites. The comprehensively evaluation experiments demonstrate that m<sup>6</sup>ATEpre achieves high prediction performance. The bioinformatics analysis for the m<sup>6</sup>A-reg-TE sites in specific cell lines reveals the characterization, potential mechanism and biological function of m<sup>6</sup>A regulated translation efficiency sites or genes.

## 1 Materials and methods

### 1.1 Multi-omics data processing and integration

In this study, MeRIP-seq data was used to identify m<sup>6</sup>A sites, PAR-CLIP data to detect the YTHDF1 binding sites, and Ribo-seq data to select genes with YTHDF1-mediated translation efficiency. Details of the multi-omics data pre-processing are provided in Document S1. By integrating MeRIP-seq data and PAR-CLIP data of overlapping m<sup>6</sup>A sites with YTHDF1 binding sites, we identified YTHDF1-

mediated m<sup>6</sup>A sites. Mapping the m<sup>6</sup>A sites to candidate genes identified from Ribo-seq data. This integrated approach was used to create the final dataset of the HeLa cell line for model training and test. Below, we detail the multi-omics data processing and integration.

We first collected the MeRIP-seq dataset from HeLa cell line, which includes two immunoprecipitation and input control sample control samples. The m<sup>6</sup>A peaks were identified from the collected MeRIP-seq dataset using exomePeak2<sup>[21]</sup>, which is the updated version of exomePeak<sup>[22]</sup>. Although m<sup>6</sup>A peaks refer to m<sup>6</sup>A modifications enriched region, they have low resolution. Therefore, we further employed sequence-based m<sup>6</sup>A sites prediction tools, such as SRAMP<sup>[23]</sup>, to identify high-confidence m<sup>6</sup>A sites at single-base resolution from sequences containing the DRACH motif within each m<sup>6</sup>A peak.

Then we analyzed the YTHDF1 PAR-CLIP data from the HeLa cell line (GSE63591<sup>[10]</sup>) and identified YTHDF1 binding sites. Due to the heterogeneous transcriptome, the isoform-specific m<sup>6</sup>A sites at single base resolution and YTHDF1 binding sites may be unavailable in the presence of multiple isoform transcription of the same gene<sup>[18]</sup>. Thus, we adopted a simple strategy of retaining the longest transcript for each gene to avoid isoform ambiguity. Then we overlapped these single-base resolution m<sup>6</sup>A sites with the YTHDF1 binding sites and identified 3 287 m<sup>6</sup>A sites with YTHDF1 binding. These sites constitute our training and test dataset, and are defined as YTHDF1-mediated m<sup>6</sup>A sites. To identify YTHDF1-mediated m<sup>6</sup>A sites regulating translation efficiency, we further analyzed the Ribo-seq data from YTHDF1 knockdown (siYTHDF1) vs control (siControl) in the HeLa cells (GSE63591<sup>[10]</sup>). Genes whose translation was mediated by YTHDF1 in an m<sup>6</sup>A-dependent manner were identified by a significant decrease in translation efficiency (TE) upon YTHDF1 knockdown<sup>[10]</sup>. The TE under siYTHDF1 and siControl conditions were calculated using equations (1) and (2), respectively. In this study, a threshold of an averaged log fold change ( $\log_2(TE_{\text{siYTHDF1}}/TE_{\text{siControl}})$ ) less than -0.5 is applied by default. This threshold was chosen to optimize the balance between positive and negative sample rates and reduce the false positive rate. More detailed information for

choosing this threshold was discussed in Document S1, Figure S1 and Figure S2. Accordingly, YTHDF1-mediated m<sup>6</sup>A sites located on these candidate genes are selected as positive samples (denoted as m<sup>6</sup>A-reg-TE sites). Conversely, the remaining YTHDF1-mediated m<sup>6</sup>A sites from genes that did not show a significant decrease in translation efficiency upon YTHDF1 knockdown ( $\log_2\left(\frac{TE_{\text{siYTHDF1}}}{TE_{\text{siControl}}}\right) > -0.5$ ) were regarded as negative samples (denoted as non-m<sup>6</sup>A-reg-TE sites). We further extracted 501-bp sequences centered on these m<sup>6</sup>A sites from their corresponding transcripts. To reduce sequence redundancy, we employed the CD-HIT tool<sup>[24]</sup> with a sequence identity cutoff of 90% to cluster the sequences from the combined positive and negative dataset. Finally, we obtained 1 190 positive samples and 1 113 negative samples. The basic statistical analysis for the number of identified m<sup>6</sup>A sites at single-base resolution, m<sup>6</sup>A peaks and candidate genes is shown in Figure S3, and the distribution of differential translation efficiency ( $\log_2\left(\frac{TE_{\text{siYTHDF1}}}{TE_{\text{siControl}}}\right)$ ) for candidate genes is shown in Figure S4. The overall workflow of the data processing is shown in Figure 1.

$$TE_{\text{siYTHDF1}} = \frac{RPF_{\text{siYTHDF1}}}{mRNA_{\text{siYTHDF1}}} \quad (1)$$

$$TE_{\text{siControl}} = \frac{RPF_{\text{siControl}}}{mRNA_{\text{siControl}}} \quad (2)$$

Where  $RPF_{\text{siYTHDF1}}$  and  $RPF_{\text{siControl}}$  are the normalized

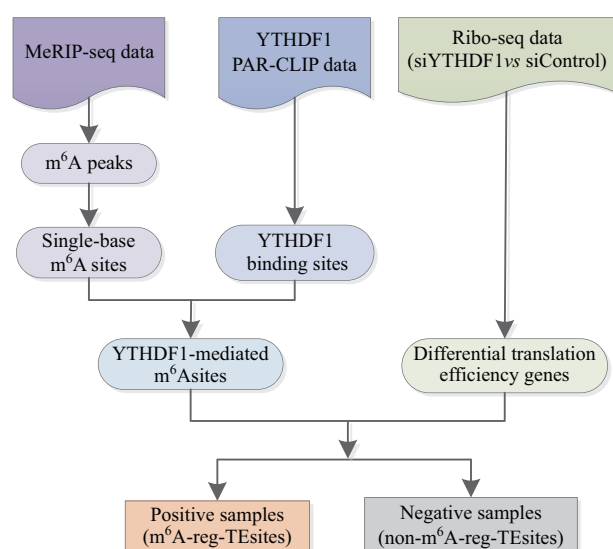


Fig. 1 Multi-omics data processing and integration

ribosome-protected fragment (RPF) counts for each gene under the siYTHDF1 and siControl conditions, respectively;  $mRNA_{\text{siYTHDF1}}$  and  $mRNA_{\text{siControl}}$  are the normalized mRNA reads counts for each gene from paired RNA-seq data under the siYTHDF1 and siControl conditions, respectively. Here, we used reads per kilobase of transcript per million mapped reads (RPKM) as the normalization method both for normalized RPF counts and for mRNA reads counts.

## 1.2 Feature representation of candidate m<sup>6</sup>A sites

To obtain the feature representation of m<sup>6</sup>A sites, we first extracted the 501-bp sequences centered at m<sup>6</sup>A site. Then, we divided the 501-bp sequence of each m<sup>6</sup>A site into 498 sliding windows. Each window spans 4 bp, and the step size between adjacent windows is set to 1 bp by default in the study. Next, each window is scanned for the presence of the subsequence “GGAC” or “GAAC”, which is defined as the YTHDF1 binding motif<sup>[10]</sup>. Previous research has revealed that YTHDF1 binds at exact GRAC (R is G or A) sites instead of other motifs that co-exist in the proximity of m<sup>6</sup>A sites<sup>[10]</sup>, which is why we only considered the two types of subsequences. If the subsequence (“GGAC” or “GAAC”) is detected, the index of the window is marked as 1; otherwise, it is marked as 0. Through this process, we generated a 498-dimensional binary vector indicating the presence or absence of the binding motif. However, this binary vector is both highly sparse and high-dimensional, which could be hard to capture the critical information. Thus, we utilized the autoencoder to reduce the dimensionality and extract the potentially important feature representations. The autoencoder model consists of a single-layer encoder, a bottleneck hidden layer and a single-layer decoder, which is an unsupervised learning model to extract the major features. The 498-dimensional binary vector for each m<sup>6</sup>A site is fed into the encoder layer, which compresses it into a 100-dimensional feature representation using rectified linear unit (ReLU) activation. The bottleneck hidden layer stores the 100-dimensional latent features learned by the autoencoder. The decoder layer aims to reconstruct the original 498-dimensional input from the 100-dimensional hidden representation using sigmoid activation. We trained the autoencoder model for 100 epochs with Adam optimizer and mean squared error loss. Finally, we extracted the 100-dimensional

embeddings from the bottleneck hidden layer to represent the features of m<sup>6</sup>A sites. The overall workflow of feature representation for m<sup>6</sup>A site sequences is shown in Figure 2.

### 1.3 Predicting m<sup>6</sup>A sites regulating translation efficiency

To predict the potential m<sup>6</sup>A sites regulating translation efficiency (m<sup>6</sup>A-reg-TE sites), we constructed an MLP model shown in Figure 2. The input to the MLP is the 100-dimensional embedding of the binary vector for YTHDF1 binding motif, generated by an autoencoder from each m<sup>6</sup>A site sequence. The MLP architecture consists of one input layer, two hidden layers with ReLU activation

functions and one output layer with the softmax function that produces prediction scores in the range [0, 1]. In the MLP model, the dimensions of the first and second hidden layers are set to 64 and 32 by default. This prediction is used for the binary classification task: determining whether the predicted m<sup>6</sup>A sites regulate translation efficiency. During the MLP model training process, the loss function used is the negative log-likelihood loss. In the optimization process of MLP, the Adam optimizer is selected as the function optimizer and the learning rate is set to 0.001. The MLP model is trained for 200 epochs by default on each fold of the five-folds cross-validation.

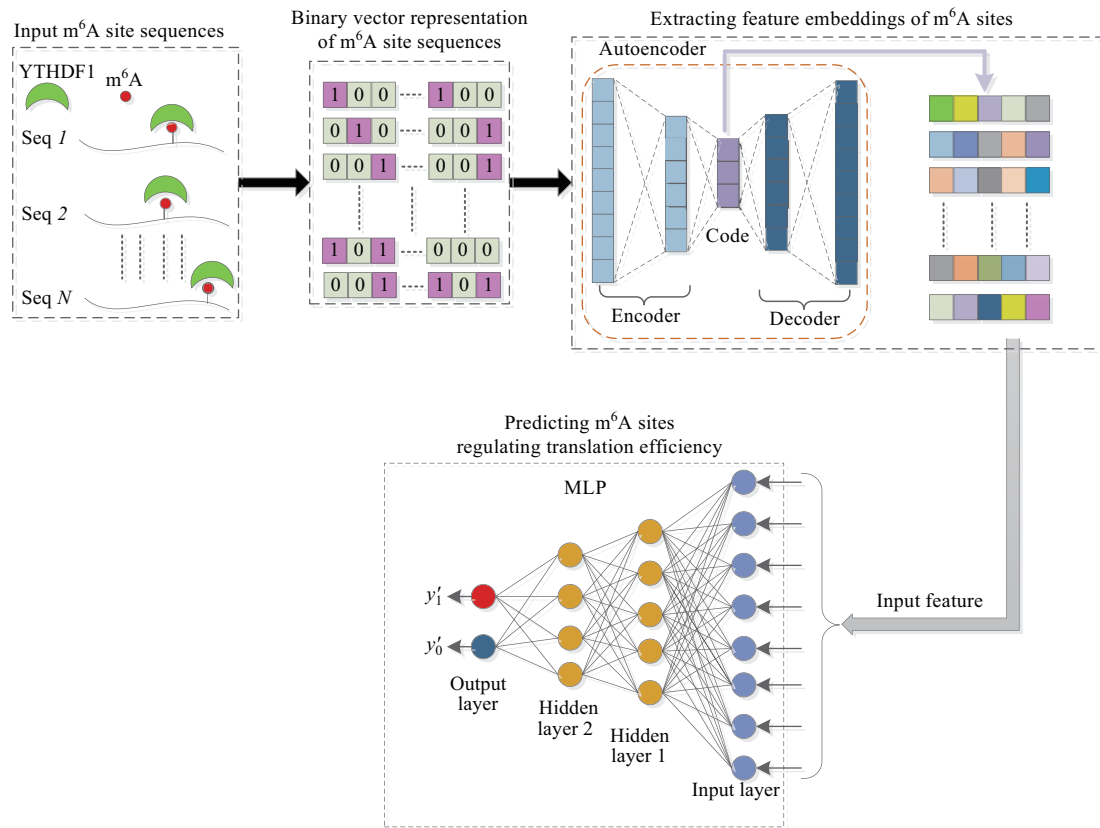


Fig. 2 The overall workflow of predicting the m<sup>6</sup>A sites regulating translation efficiency ( named as m<sup>6</sup>ATEpre )

### 1.4 Performance metrics

To comprehensively evaluate the prediction performance of our proposed computational framework and compare it with other models, we used a five-fold cross-validation (5-CV) method to evaluate the performance of models under six different evaluation metrics. These evaluation metrics include *Accuracy*, *F1-score*, *Precision*, *Recall*, the area under the curve (*AUC*) and the area under precision and

recall (*PRAUC*). The evaluation metrics are defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$F1 - score = \frac{2TP}{2TP + FP + FN} \quad (4)$$

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

where  $TP$ ,  $TN$ ,  $FP$ , and  $FN$  are the number of true positive samples, true negative samples, false positive samples and false negative samples, respectively.

## 2 Results

### 2.1 Performance evaluation and comparison

To our best knowledge, currently there are no computational methods specifically developed for predicting m<sup>6</sup>A sites that regulate translation efficiency. However, an m<sup>6</sup>A sequence-based language model, m<sup>6</sup>A-BERT-Deg<sup>[20]</sup>, has been developed to predict the m<sup>6</sup>As' regulatory impact on YTHDF2-mediated mRNA degradation, which is similar to the prediction task of m<sup>6</sup>ATEpre. Therefore, we compared our proposed m<sup>6</sup>ATEpre with m<sup>6</sup>A-BERT-Deg. To achieve a relatively fair comparison,

we first used the training data (80% of all input data) to pre-train the m<sup>6</sup>A-BERT-Deg model, and then evaluated its prediction performance on the test data (20% of all input data). The training dataset is used to train the m<sup>6</sup>ATEpre model, and the test set is used to evaluate the performance of m<sup>6</sup>ATEpre under different evaluation metrics. As shown in Figure 3, m<sup>6</sup>ATEpre achieves performance gains of over 0.33 in all evaluation metrics relative to m<sup>6</sup>A-BERT-Deg, demonstrating significantly improved prediction performance compared to the m<sup>6</sup>A-BERT-Deg model. This result indicates that m<sup>6</sup>A-BERT-Deg may have difficulty capturing more specific context features related to YTHDF1-mediated translation only depending on m<sup>6</sup>A site sequences. Additionally, the full leveraging of m<sup>6</sup>A-BERT-Deg's capabilities was likely limited by the small training samples<sup>[20]</sup>.

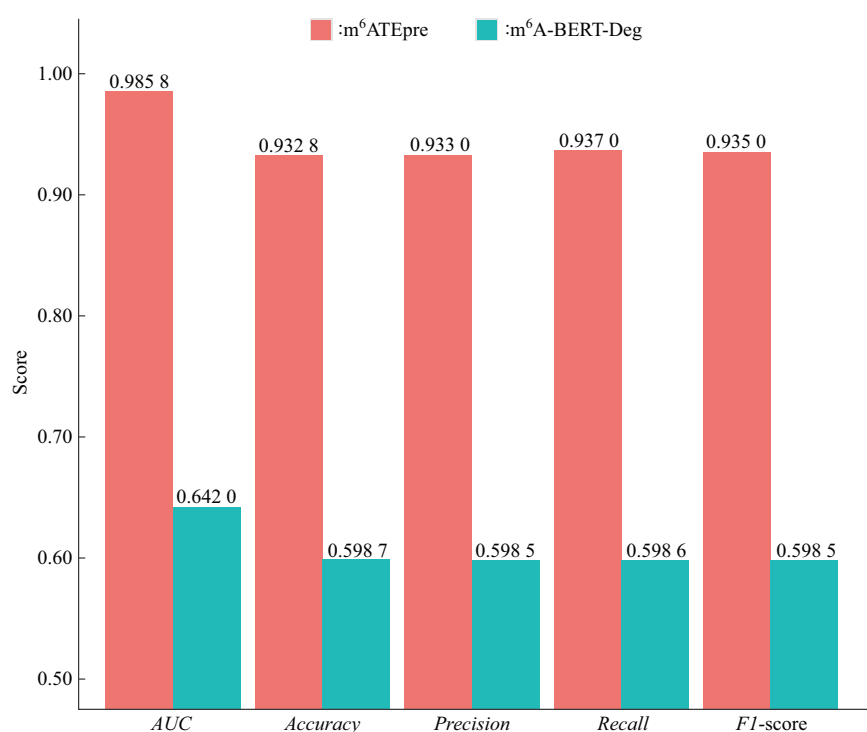


Fig. 3 The performance comparison of m<sup>6</sup>ATEpre and m<sup>6</sup>A-BERT-Deg

In m<sup>6</sup>ATEpre computational framework, we used the MLP as the prediction model to identify the potential positive samples. The choice of prediction model or classifier may affect the predictive power in identifying m<sup>6</sup>A-mediated translation regulation. Therefore, in this study, we also evaluated the prediction performance of different predicting models including MLP, Random Forest (RF) and XGboost *via*

the 5-CV method. In addition, we also performed a comparison with a Transformer-based sequence classification model (named VectorTransformer), which excels at processing sequential data and capturing long-range dependency<sup>[25-26]</sup>. The input features for MLP, RF, XGboost and VectorTransformer are the embedding representations of m<sup>6</sup>A sites obtained by autoencoder. As shown in

Figure 4, except XGboost, m<sup>6</sup>ATEpre with MLP, RF and VectorTransformer classifiers all achieved the higher prediction performance, whose evaluation metrics are all more than 0.9. But m<sup>6</sup>ATEpre with

MLP achieved the highest prediction performance compared to the other three models. Therefore, the MLP is selected as the prediction model in the m<sup>6</sup>ATEpre computational framework.

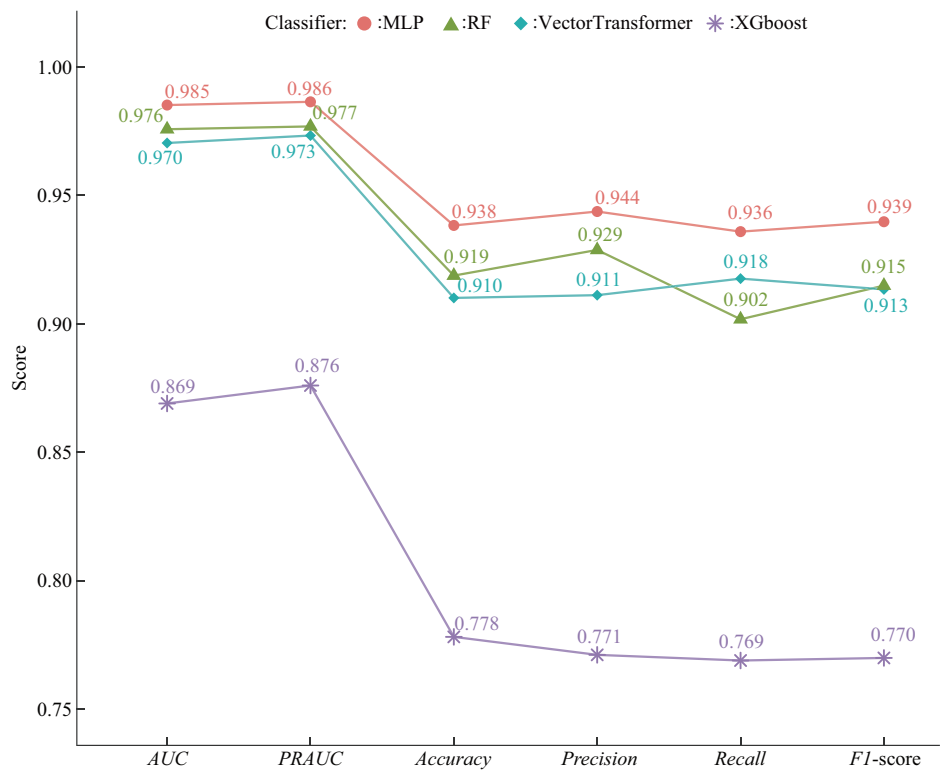


Fig. 4 Performance comparison of different classifiers

To verify the efficacy of our proposed feature representations for m<sup>6</sup>A sites in m<sup>6</sup>ATEpre, we further compared them with other feature representation strategies for predicting m<sup>6</sup>A sites involved in translation regulation. Feature representation is a key factor determining the model's predictive ability. In this study, we directly used the *k*-mer (*k*=3, 4, and 5) frequency to represent the input features of the 501-bp m<sup>6</sup>A sites' sequences. In addition, we also applied the word embedding model Word2Vec<sup>[27]</sup> to generate the low-dimensional embeddings for each m<sup>6</sup>A site's sequence. These feature representations including *k*-mer (*k*=3, 4, and 5) frequency and low-dimensional embedding *via* Word2Vec are directly used as input features for the MLP classifier with same parameters. Finally, we evaluated the prediction performance using different feature representation strategies for m<sup>6</sup>A sites' sequences, including m<sup>6</sup>ATEpre's feature

representation, *k*-mer frequency and word embeddings *via* the 5-CV method. As shown in Figure 5, the five-fold cross-validation results indicate that the prediction performance of both *k*-mer (*k*=3, 4, and 5) frequency and word embeddings is significantly lower than that of the m<sup>6</sup>ATEpre feature representation (the embedding of binary vector for YTHDF1 binding motif). For *k*-mer frequency, the prediction performance improves with higher *k* values across all six evaluated metrics, as shown in Figure 5. In addition, the prediction performance of word embeddings is the lowest according to the four evaluated metrics, including *AUC*, *PRAUC*, *Accuracy* and *Precision*. This result further demonstrates that adapting the autoencoder to obtain lower-dimensional embeddings has a stronger advantage than other feature representation strategies in predicting m<sup>6</sup>A-regulated translation efficiency.

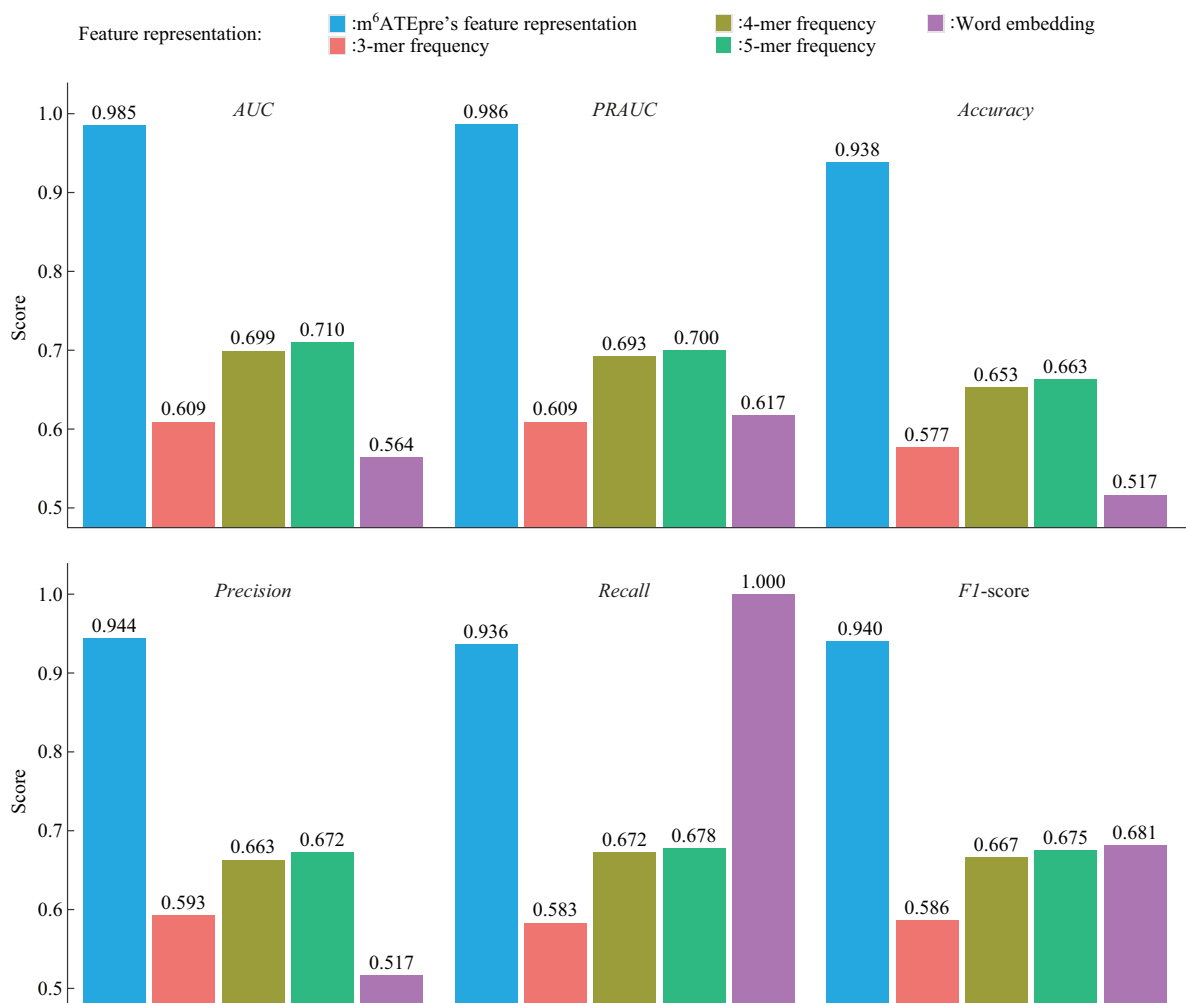


Fig. 5 The prediction performance comparison for different feature representation methods

## 2.2 Ablation experiment

In this study, the main function of the autoencoder module in m<sup>6</sup>ATEpre is to reduce the dimensionality of the input features and extract the important embedding features. To demonstrate the advantage of the autoencoder module for predicting m<sup>6</sup>A sites with regulating translation efficiency, we first utilized the uniform manifold approximation and projection (UMAP) technique to visualize the high-dimensional input features with or without autoencoder processing for both positive and negative samples. As shown in Figure 6a, it may be hard to distinguish positive and negative samples just from the raw input features without autoencoder processing, which are the raw binary vector of YTHDF1 binding motif presentation. However, after the feature extraction by the autoencoder, the distinction between positive and negative samples becomes much clearer, as shown in Figure 6b.

Therefore, we performed an ablation experiment to verify the contribution of the autoencoder module in m<sup>6</sup>ATEpre. Specifically, we removed it from m<sup>6</sup>ATEpre and assessed the predictive performance on m<sup>6</sup>A sites associated with regulating translation efficiency. The experimental results in Figure 7 indicate that the prediction performance decreases significantly for the m<sup>6</sup>ATEpre model without the autoencoder module. As shown in Figure 7, the values of AUC, PRAUC, Accuracy, Precision, Recall and F1-score are 0.470, 0.438, 0.437, 0.427, 0.407 and 0.416, respectively, which are all significantly lower than the corresponding values of m<sup>6</sup>ATEpre with the autoencoder module. This result indicates that using the autoencoder module to obtain the lower-dimensional embeddings of the input features is highly effective in predicting potential m<sup>6</sup>A sites regulating translation efficiency.

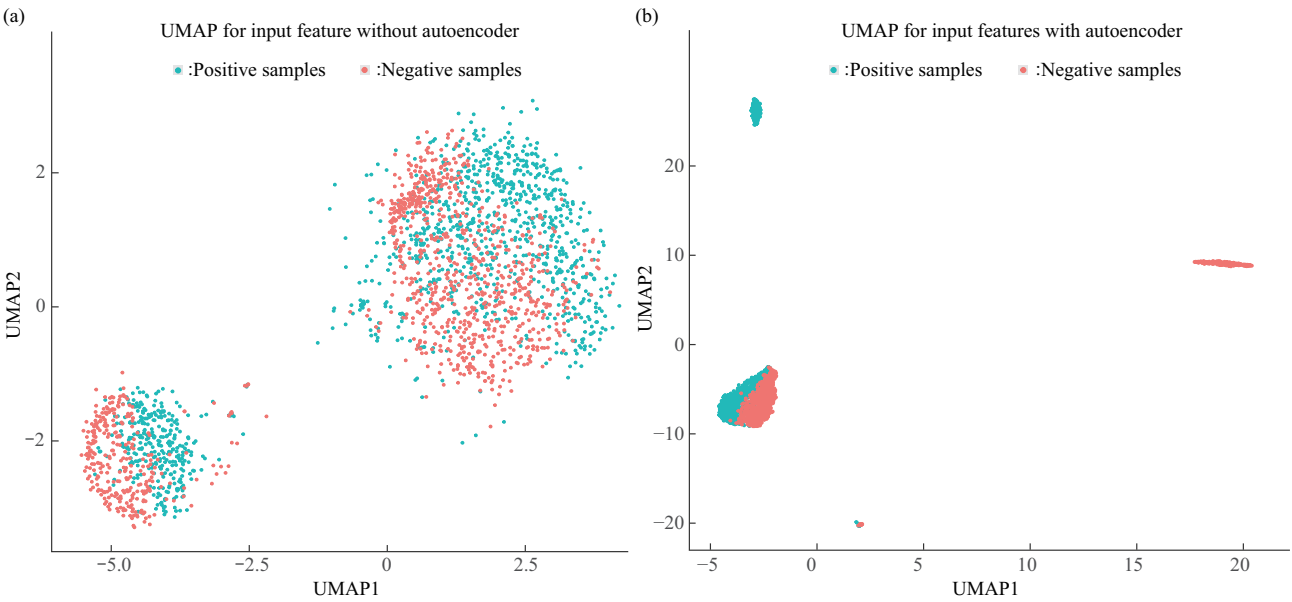


**Fig. 6 UMAP dimensionality reduction for input features with or without autoencoder processing**  
(a) UMAP dimensionality reduction on raw input features without autoencoder. (b) UMAP dimensionality reduction for the input features with autoencoder processing.

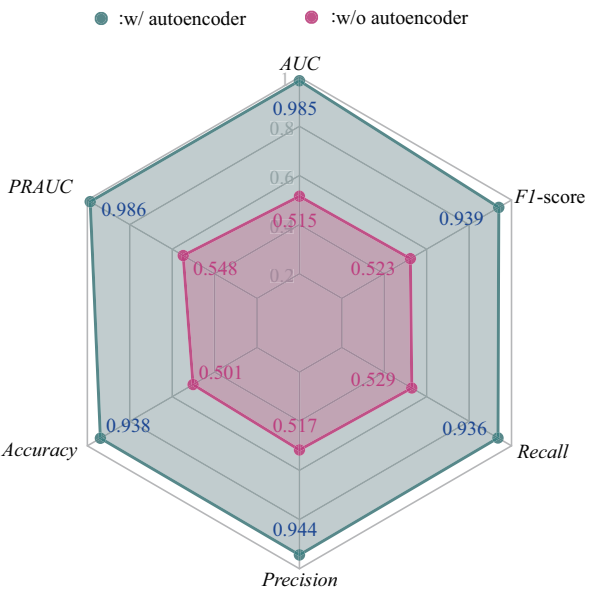


**Fig. 7 The prediction performance of m<sup>6</sup>ATEpre with and without autoencoder processing**

**2.3 Characterization of m<sup>6</sup>A sites regulating translation efficiency**

To explore the potential pattern of m<sup>6</sup>A-regulated translation efficiency, we first analyzed the metagene distribution of m<sup>6</sup>A sites regulating and not regulating translation efficiency. As shown in Figure 8a, both

YTHDF1 binding sites and the identified m<sup>6</sup>A sites are all enriched in stop codon regions, consistent with the metagene distribution of canonical m<sup>6</sup>A sites. However, the distribution of YTHDF1 binding sites is more consistent with that of m<sup>6</sup>A sites regulating translation efficiency (m<sup>6</sup>A-reg-TE) than with that of m<sup>6</sup>A sites not regulating translation efficiency (non-m<sup>6</sup>A-reg-TE). It may indicate that YTHDF1-mediated regulation of mRNA translation is dependent on m<sup>6</sup>A methylation. Then, we further calculated the minimum free energy (MFE) by RNAfold<sup>[28]</sup> for these sequences of m<sup>6</sup>A sites to explore whether m<sup>6</sup>A sites regulating translation efficiency could help maintain mRNA stability. As shown in Figure 8b, the MFEs of m<sup>6</sup>A-reg-TE sites are significantly lower than those of non-m<sup>6</sup>A-reg-TE sites ( $P=5.68\times10^{-9}$ ;  $t$ -test). This result indicates that m<sup>6</sup>A sites regulating translation efficiency (m<sup>6</sup>A-reg-TE) could maintain the stability of the mRNA secondary structure. To investigate the potential mechanism of YTHDF1-mediated, m<sup>6</sup>A-dependent regulation of mRNA translation, we further collected the binding sites of RNA binding proteins (RBPs) in the HeLa cell line from the POSTAR3 database<sup>[29]</sup> and overlapped them with m<sup>6</sup>A site sequences to explore the potential co-factors of YTHDF1. Specifically, we first overlapped the YTHDF1 binding sites with the RBP's binding sites

to select the YTHDF1's co-factors. In Table S1, 27 RBPs were identified as YTHDF1's co-factors including 8 translation regulators (YTHDF2, METTL14, METTL3, ALKBH1, AGO2, UPF1, ELAVL1 and TIA1). We then mapped YTHDF1 and the eight translation regulators onto a protein-protein interaction (PPI) network downloaded from the STRING database<sup>[30]</sup>. As shown in Figure 8c, the eight co-factors strongly interact with YTHDF1 as well as with each other. We overlapped these predicted m<sup>6</sup>A

sites with the binding sites of these 27 YTHDF1 co-factors. 59 m<sup>6</sup>A-reg-TE sites (positive samples) overlap with 134 co-factors' binding sites, but only 33 non-m<sup>6</sup>A-reg-TE sites (negative samples) overlap with 66 co-factors' binding sites. Fisher's exact test revealed a significant enrichment of YTHDF1 cofactors (involved in regulating translation) on m<sup>6</sup>A-reg-TE sites compared to non-m<sup>6</sup>A-reg-TE sites ( $P=0.0094$ , odds ratio=1.71).

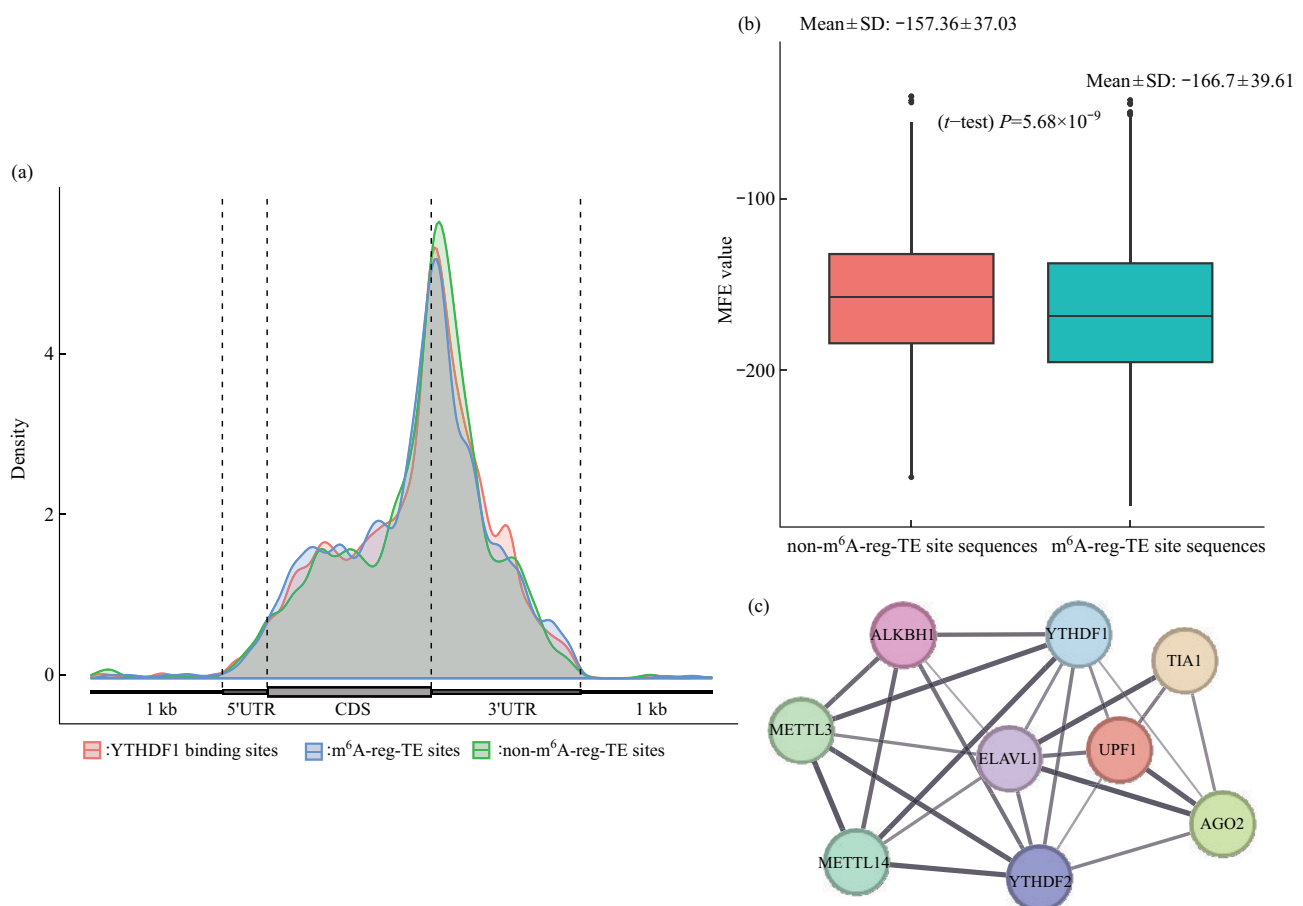


**Fig. 8 The characteristics of m<sup>6</sup>A sites regulating translation efficiency ( m<sup>6</sup>A-reg-TE )**

(a) The metagene distribution of YTHDF1 binding sites, m<sup>6</sup>A-reg-TE sites and non-m<sup>6</sup>A-reg-TE sites. (b) Boxplot showing the distribution of the minimum free energy (MFE) for both m<sup>6</sup>A-reg-TE site sequences and non-m<sup>6</sup>A-reg-TE site sequences. (c) Protein-protein interaction (PPI) network of YTHDF1 co-factors enriched in translation regulation process. Nodes represent individual proteins, including: m<sup>6</sup>A methyltransferase methyltransferase 3 (METTL3); methyltransferase 14 (METTL14), ALKBH1 (AlkB homolog 1), m<sup>6</sup>A readers (YTH N<sup>6</sup>-methyladenosine RNA binding protein F1 (YTHDF1); YTH N<sup>6</sup>-methyladenosine RNA binding protein F2 (YTHDF2)), and RNA-binding proteins involved in post-transcriptional regulation (ELAV-like RNA binding protein 1 (ELAVL1); TIA1 cytotoxic granule associated RNA binding protein (TIA1); UPF1 RNA helicase and ATPase (UPF1); Argonaute RISC catalytic component 2 (AGO2)). Edges indicate the interactions between proteins. m<sup>6</sup>A-reg-TE: m<sup>6</sup>A sites regulating translation efficiency; SD: standard deviation.

We next sought to understand the biological functions associated with m<sup>6</sup>A-reg-TE sites in HeLa cells. We first performed the biological process (BP) enrichment analysis using ClusterProfiler R package<sup>[31]</sup>. The top 10 enriched Gene Ontology (GO)

biological process terms for m<sup>6</sup>A-reg-TE genes were significantly different from those for non-m<sup>6</sup>A-reg-TE genes, as shown in Figure 9a. The m<sup>6</sup>A-reg-TE genes participate in small GTPase-mediated signal transduction, uterine embryonic development, TGF- $\beta$

regulation of signaling pathways and so on (Figure 9a). However, the non-m<sup>6</sup>A-reg-TE genes mainly participate in the regulation of phosphorylation process, phosphate metabolic process, transferase and kinase activity regulation and response to ER stress (Figure 9a). We performed the enrichment analysis of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways in these YTHDF1-mediated genes. We found that YTHDF1-mediated m<sup>6</sup>A-reg-TE genes were significantly enriched in various cellular

processes related pathways (*e. g.*, phospholipase D signaling pathway, mTOR signaling pathway, Rap1 signaling pathway, and endocytosis) and enriched in some cancers (*e. g.*, colorectal cancer and chronic myeloid leukemia), as shown in Figure 9b. However, the YTHDF1-mediated non-m<sup>6</sup>A-reg-TE genes were only significantly enriched in TGF- $\beta$  signaling pathway (Figure 9b). This biological function analysis indicates a marked difference between m<sup>6</sup>A-reg-TE genes and non-m<sup>6</sup>A-reg-TE genes.

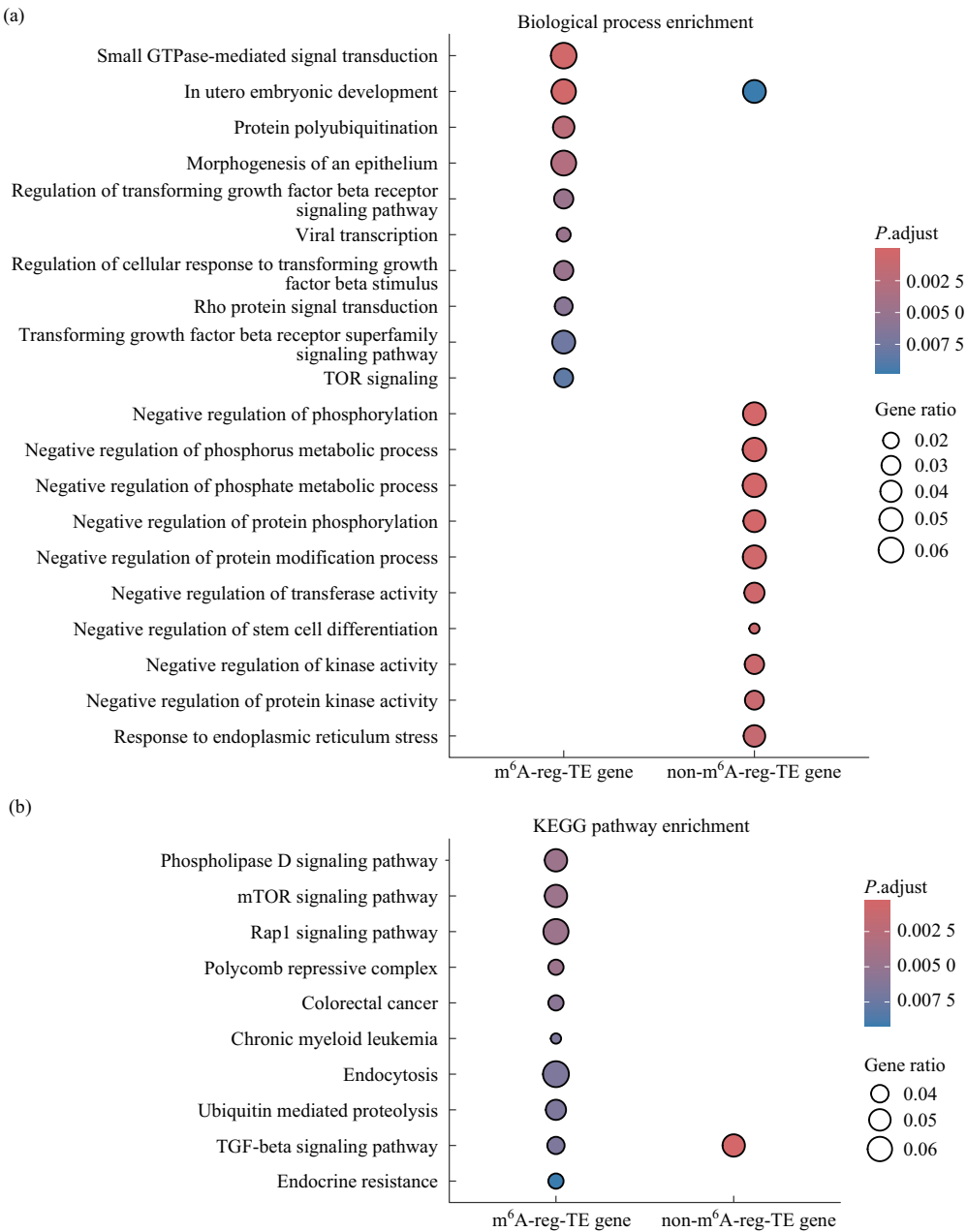


**Fig. 9** Function enrichment analysis for both m<sup>6</sup>A-reg-TE gene and the non-m<sup>6</sup>A-reg-TE gene in the HeLa cell line  
(a) Biological process enrichment for m<sup>6</sup>A-reg-TE genes and non-m<sup>6</sup>A-reg-TE genes in the HeLa cell line. (b) The KEGG pathway enrichment for m<sup>6</sup>A-reg-TE genes and non-m<sup>6</sup>A-reg-TE genes.

## 2.4 Predicting YTHDF1-mediated mRNA translation in HEK293T cells using m<sup>6</sup>ATEpre

In this part, we applied m<sup>6</sup>ATEpre to predict YTHDF1-mediated mRNA translation in the HEK293T cell line. We first collected the m<sup>6</sup>A sites at single-base resolution in HEK293T cell line from m<sup>6</sup>A-AtlasV2 database<sup>[32]</sup> and the YTHDF1 binding sites from the iCLIP dataset (GSE78030) in HEK293T cell line<sup>[33]</sup>. Next, we selected the m<sup>6</sup>A sites located within YTHDF1 binding regions and extracted 501-bp sequences centered on m<sup>6</sup>A sites. We then employed the CD-HIT tool to remove similar sequences and retained 2 211 candidate m<sup>6</sup>A site sequences. These sequences were transformed into binary vectors representing the presence of the YTHDF1 binding motif, which served as the input features for m<sup>6</sup>ATEpre. Based on the above processing, we predicted 962 potential m<sup>6</sup>A sites regulating translation efficiency (m<sup>6</sup>A-reg-TE sites) of 470 genes and 1 249 m<sup>6</sup>A sites not regulating translation efficiency (non-m<sup>6</sup>A-reg-TE sites) of 484 genes.

Similar to the analysis of the results in HeLa cell line, we also firstly used the Guitar package<sup>[34]</sup> to analyze the metagene distribution of these predicted

m<sup>6</sup>A-reg-TE sites and non-m<sup>6</sup>A-reg-TE sites. As shown in Figure 10a, these predicted m<sup>6</sup>A-reg-TE sites in HEK293T cell line are more likely to be enriched in the 3'-untranslated region (3'UTR), whereas the non-m<sup>6</sup>A-reg-TE sites tend to be enriched in the coding DNA sequence (CDS). According to the pie chart in Figure S5a, m<sup>6</sup>A-reg-TE sites are predominantly enriched in the 3'UTR (53.3%) over the CDS (42.3%). In contrast, 49.2% non-m<sup>6</sup>A-reg-TE sites are enriched in 3'UTR and 48.4% the in CDS (Figure S5b). It indicates that in HEK293T cell line, the predicted m<sup>6</sup>A-reg-TE sites are more likely to be enriched in the 3'UTR (Fisher's exact test;  $P$ -value=0.026 8; odds ratio=1.185), while the predicted non-m<sup>6</sup>A-reg-TE sites are more likely to be enriched in the CDS (Fisher exact test;  $P$ =8.618 $\times$ 10<sup>-5</sup>; odds ratio=1.39). To explore the potential mechanism of m<sup>6</sup>A-dependent translation regulation mediated by YTHDF1, we obtained the binding sites of RBPs for the HEK293T cell line from the POSTAR3 database to identify potential YTHDF1 cofactors. For these collected binding sites of RBPs, we first mapped them to the binding regions of YTHDF1 and selected 32 potential YTHDF1 co-factors, which are shown in Table S2. We then mapped the predicted m<sup>6</sup>A-reg-TE

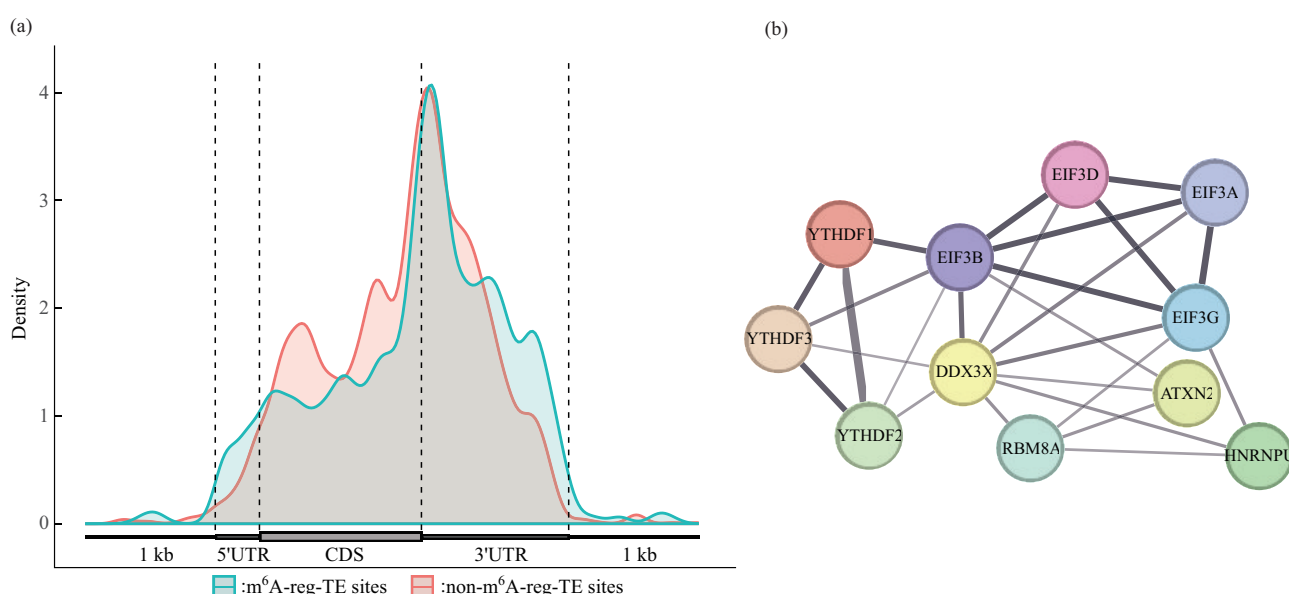


**Fig. 10 The predicted m<sup>6</sup>A-reg-TE sites in the HEK293T cell line by m<sup>6</sup>ATEpre**

(a) The metagene distribution of m<sup>6</sup>A-reg-TE sites and non-m<sup>6</sup>A-reg-TE sites in the HEK293T cell line. (b) Protein-protein interaction (PPI) network of YTHDF1 co-factors associated with translation regulation in the HEK293T cell line. Nodes represent individual proteins, including: m<sup>6</sup>A reader proteins (YTH N<sup>6</sup>-methyladenosine RNA binding protein 1/2/3 (YTHDF1/2/3)), eukaryotic translation initiation factors (eukaryotic translation initiation factor 3 subunit A (EIF3A); eukaryotic translation initiation factor 3 subunit B (EIF3B); eukaryotic translation initiation factor 3 subunit D (EIF3D); eukaryotic translation initiation factor 3 subunit G (EIF3G)), and RNA-binding proteins (DEAD-box helicase 3 X-linked (DDX3X); Ataxin 2 (ATXN2); RNA binding motif protein 8A (RBM8A); heterogeneous nuclear ribonucleoprotein U (HNRNPU)). Edges indicate physical or functional interactions between proteins.

sites to the binding sites of YTHDF1 co-factors and found that 441 of them overlapped. Ten of these YTHDF1 co-factors (YTHDF2, YTHDF3, ATXN2, EIF3A, EIF3B, EIF3D, EIF3G, HNRNPU, RBM8A, and DDX3X) are involved in translation regulation. The PPI network (Figure 10b) reveals close interactions among these factors, and notably, EIF3A, EIF3B, EIF3D and EIF3G are all subunits of the eukaryotic translation factor 3 (eIF3) complex. This result reveals that YTHDF1, potentially together with translation initiation factors, binds to m<sup>6</sup>A sites to regulate the translation efficiency of target genes, which is consistent with previous research<sup>[10]</sup>. Lastly, we also performed biological functional analysis on the predicted m<sup>6</sup>A-reg-TE genes in HEK293T cell line. The biological process enrichment analysis suggests that the predicted m<sup>6</sup>A-reg-TE genes are significantly enriched in transport related biological processes, such as nucleocytoplasmic transport and nuclear transport (Figure S6). The KEGG pathway enrichment analysis indicates that these genes are significantly enriched in the nucleocytoplasmic transport pathway. The biological function enrichment of m<sup>6</sup>A-reg-TE genes differed between the HEK293T and HeLa cell lines, suggesting that their functions may be cell line specific.

### 3 Conclusion

Identifying the potential m<sup>6</sup>A sites that regulate translation can help us further reveal the regulatory mechanism of m<sup>6</sup>A modification. In this study, we innovatively propose a computational method named m<sup>6</sup>ATEpre to predict potential m<sup>6</sup>A regulatory translation efficiency sites. m<sup>6</sup>ATEpre first integrates multi-omics data, including MeRIP-seq data, YTHDF1-binding PAR-CLIP data and Ribo-seq data, to obtain a high-quality training dataset comprising positive and negative m<sup>6</sup>A site sequences from the HeLa cell line. We then utilize the binary vectors of YTHDF1-binding motifs, derived from the m<sup>6</sup>A site sequences, to represent the features of the positive and negative samples. To address the issues of sparsity and high-dimensionality of input features, m<sup>6</sup>ATEpre incorporates an autoencoder module for effective feature extraction and representation. This model subsequently employs an MLP to identify potential m<sup>6</sup>A-regulated translation efficiency sites.

We conducted a comprehensive evaluation of

m<sup>6</sup>ATEpre through a series of experiments. Firstly, we compared its performance against that of m<sup>6</sup>A-BERT-Deg (a similar task prediction model), as well as other classifiers or models, such as Random Forest, XGboost and a transformer-based model. The results indicate that m<sup>6</sup>ATEpre achieves the best prediction performance. Furthermore, we analyzed different feature representation strategies and performed ablation studies to validate the model design. The results demonstrate that our proposed feature representation strategy has a significant advantage in improving prediction performance. Moreover, we investigated the potential patterns and the regulatory mechanism of m<sup>6</sup>A sites regulating translation (m<sup>6</sup>A-reg-TE sites). Specifically, in the HeLa cell line, we performed bioinformatic characterization of m<sup>6</sup>A-reg-TE sites by analyzing their metagene distribution and sequence MFE, which reveals the specific properties of m<sup>6</sup>A sites involved in translation regulation. To explore the underlying regulatory mechanism, we integrated the binding sites of YTHDF1 co-factors with m<sup>6</sup>A-reg-TE sites in the HeLa cell line. Our analysis revealed that YTHDF1-mediated, m<sup>6</sup>A-dependent regulation of translation efficiency requires the cooperation of multiple translation-regulatory RNA-binding proteins as the YTHDF1 co-factors. Finally, the functional enrichment analysis indicates that m<sup>6</sup>A-reg-TE genes are associated with more specific biological processes and KEGG pathways.

To demonstrate the utility of m<sup>6</sup>ATEpre in predicting cell-specific YTHDF1-mediated translation, we conducted predictions within the HEK293T cell line. The metagene distribution analysis and biological function enrichment analysis for these predicted m<sup>6</sup>A-reg-TE sites in HEK293T cells are quite different from those in the HeLa cell line, indicating that m<sup>6</sup>A-reg-TE sites may have cell-specific characteristics. Likewise, the integrated analysis of the binding sites of multiple YTHDF1 co-factors and m<sup>6</sup>A-reg-TE sites reveals the cooperation of YTHDF1 and its co-factors in regulating translation efficiency in an m<sup>6</sup>A-dependent manner.

Although m<sup>6</sup>ATEpre has achieved strong predictive performance in the HeLa cell line, it still has some limitations. Due to the lack of translation efficiency-related datasets or experimental validation in the HEK293T cell line, the predicted m<sup>6</sup>A-reg-TE sites are difficult to verify. However, these predictions

in HEK293T cell line could be verified by analyzing the Ribo-seq data under the siYTHDF1 vs siControl conditions, or through biological experiments in the future. Furthermore, the current focus on YTHDF1-mediated translation regulation means that m<sup>6</sup>ATEpre may overlook other regulatory mechanisms and fail to account for the full complexity of the regulatory process. Additionally, m<sup>6</sup>ATEpre lacks the interpretability regarding how the raw input features contribute to the prediction of m<sup>6</sup>A-reg-TE sites. Therefore, a highly interpretable model and the inclusion of regulatory-mechanism-related features are necessary for future research on m<sup>6</sup>A-regulated translation. Despite some limitations of our proposed computational framework, the underlying design concept of m<sup>6</sup>ATEpre could be extended to other translation-related m<sup>6</sup>A regulators, such as YTHDF3, YTHDC2, or METTL3. Such extension could substantially increase its utility to the epitranscriptomics community and help us reveal the complexity of m<sup>6</sup>A regulating translation efficiency.

**Supplementary** Available online (<http://www.pibb.ac.cn>, <http://www.cnki.net>):

PIBB\_20260005\_Document S1.pdf

PIBB\_20260005\_Table S1.pdf

PIBB\_20260005\_Table S2.pdf

PIBB\_20260005\_Figure S1.pdf

PIBB\_20260005\_Figure S2.pdf

PIBB\_20260005\_Figure S3.pdf

PIBB\_20260005\_Figure S4.pdf

PIBB\_20260005\_Figure S5.pdf

PIBB\_20260005\_Figure S6.pdf

**Data Availability** The associated data of this paper can be accessed and obtained in the Science Data Bank database (<https://www.scidb.cn/s/bAZZFfr>).

## References

- [1] Esteva-Socias M, Aguilo F. METTL3 as a master regulator of translation in cancer: mechanisms and implications. *NAR Cancer*, 2024, **6**(1): zcae009
- [2] Xiao S, Cao S, Huang Q, *et al.* The RNA N6-methyladenosine modification landscape of human fetal tissues. *Nat Cell Biol*, 2019, **21**(5): 651-661
- [3] Akhtar J, Lugoboni M, Junion G. m<sup>6</sup>A RNA modification in transcription regulation. *Transcription*, 2021, **12**(5): 266-276
- [4] Yang C, Hu Y, Zhou B, *et al.* The role of m<sup>6</sup>A modification in physiology and disease. *Cell Death Dis*, 2020, **11**(11): 960
- [5] Song B, Wang X, Liang Z, *et al.* RMDisease V<sub>2.0</sub>: an updated database of genetic variants that affect RNA modifications with disease and trait implication. *Nucleic Acids Res*, 2023, **51**(D1): D1388-D1396
- [6] Ma J, Liu H, Mao Y, *et al.* LRTCLS: low-rank tensor completion with Laplacian smoothing regularization for unveiling the post-transcriptional machinery of N6-methylation (m6A) -mediated diseases. *BriefBioinform*, 2022, **23**(5): bbac325
- [7] Zhang T, Liu L. m<sup>6</sup>ADP-GCNPUS: m<sup>6</sup>A-disease prediction via graph convolutional network and positive-unlabeled learning with self-adaptive sampling. *Interdiscip Sci*, 2025. DOI: 10.1007/s12539-025-00760-0
- [8] Boulas K, Greer E L. Biological roles of adenine methylation in RNA. *Nat Rev Genet*, 2023, **24**(3): 143-160
- [9] Wang X, Lu Z, Gomez A, *et al.* N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature*, 2014, **505**(7481): 117-120
- [10] Wang X, Zhao B S, Roundtree I A, *et al.* N 6-methyladenosine modulates messenger RNA translation efficiency. *Cell*, 2015, **161**(6): 1388-1399
- [11] Huang H, Weng H, Sun W, *et al.* Recognition of RNA N<sup>6</sup>-methyladenosine by IGF2BP proteins enhances mRNA stability and translation. *Nat Cell Biol*, 2018, **20**(3): 285-295
- [12] Jiang X, Liu B, Nie Z, *et al.* The role of m6A modification in the biological functions and diseases. *Sig Transduct Target Ther*, 2021, **6**(1): 74
- [13] Zhang S Y, Zhang S W, Liu L, *et al.* m6A-Driver: identifying context-specific mRNA m6A methylation-driven gene interaction networks. *PLoS Comput Biol*, 2016, **12**(12): e1005287
- [14] Zhang S Y, Zhang S W, Fan X N, *et al.* Global analysis of N6-methyladenosine functions and its disease association using deep learning and network-based methods. *PLoS Comput Biol*, 2019, **15**(1): e1006663
- [15] Zhang S Y, Zhang S W, Fan X N, *et al.* FunDMDeep-m6A: identification and prioritization of functional differential m6A methylation genes. *Bioinformatics*, 2019, **35**(14): i90-i98
- [16] Zhang T, Gao S, Zhang S W, *et al.* m<sup>6</sup>Aexpress-enet: predicting the regulatory expression m<sup>6</sup>A sites by an enet-regularization negative binomial regression model. *Methods*, 2024, **226**: 61-70
- [17] Zhang T, Zhang S W, Zhang S Y, *et al.* m6A-express: uncovering complex and condition-specific m6A regulation of gene expression. *Nucleic Acids Res*, 2021, **49**(20): e116
- [18] Zhang T, Zhang S W, Zhang S Y, *et al.* m<sup>6</sup>Aexpress-Reader: prediction of m<sup>6</sup>A regulated expression genes by integrating m<sup>6</sup>A sites and reader binding information in specific- context. *Methods*, 2022, **203**: 167-178
- [19] Zhang T, Zhang S W, Feng J, *et al.* m 6 Aexpress-BHM predicting m6A regulation of gene expression in multiple-groups context by a Bayesian hierarchical mixture model. *Brief Bioinform*, 2022, **23**(4): bbac295
- [20] Zhang T H, Jo S, Zhang M, *et al.* Understanding YTHDF2-mediated mRNA degradation by m<sup>6</sup>A-BERT-deg. *ArXiv*, 2024:

- arXiv: 2401.08004v1
- [21] Duan D, Tang W, Wang R, *et al.* Evaluation of epitranscriptome-wide N6-methyladenosine differential analysis methods. *Brief Bioinform*, 2023, **24**(3): bbad139
- [22] Meng J, Lu Z, Liu H, *et al.* A protocol for RNA methylation differential analysis with MeRIP-Seq data and exomePeak R/Bioconductor package. *Methods*, 2014, **69**(3): 274-281
- [23] Zhou Y, Zeng P, Li Y H, *et al.* SRAMP: prediction of mammalian N6-methyladenosine (m6A) sites based on sequence-derived features. *Nucleic Acids Res*, 2016, **44**(10): e91
- [24] Fu L, Niu B, Zhu Z, *et al.* CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 2012, **28**(23): 3150-3152
- [25] Zhou Z, Ye L, Wang J, *et al.* HiVT: hierarchical vector transformer for multi-agent motion prediction//Efros A, Ramanan D. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, 2022: 8813-8823
- [26] Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need//Guyon I, Luxburg U V, Bengio S, *et al.* Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 2017: 6000-6010
- [27] Zou Q, Xing P, Wei L, *et al.* Gene2vec: gene subsequence embedding for prediction of mammalian N<sup>6</sup>-methyladenosine sites from mRNA. *RNA*, 2019, **25**(2): 205-218
- [28] Hofacker I L. RNA secondary structure analysis using the Vienna RNA package. *Curr Protoc Bioinformatics*, 2009, **26**(1): 12.2.1-12.2.16
- [29] Zhao W, Zhang S, Zhu Y, *et al.* POSTAR3: an updated platform for exploring post-transcriptional regulation coordinated by RNA-binding proteins. *Nucleic Acids Res*, 2022, **50**(D1): D287-D294
- [30] Szklarczyk D, Nastou K, Koutrouli M, *et al.* The STRING database in 2025: protein networks with directionality of regulation. *Nucleic Acids Res*, 2025, **53**(D1): D730-D737
- [31] Wu T, Hu E, Xu S, *et al.* clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation (Camb)*, 2021, **2**(3): 100141
- [32] Liang Z, Ye H, Ma J, *et al.* m6A-Atlas v2.0: updated resources for unraveling the N6-methyladenosine (m6A) epitranscriptome among multiple species. *Nucleic Acids Res*, 2024, **52**(D1): D194-D202
- [33] Patil D P, Chen C K, Pickering B F, *et al.* m(6)a RNA methylation promotes XIST-mediated transcriptional repression. *Nature*, 2016, **537**(7620): 369-373
- [34] Zhang L, Chen S, Chen X, *et al.* Guitar2: an improved version of the Guitar for sketching the transcriptomic view of RNA-related biological features. *Methods*, 2022, **203**: 40-45

# m<sup>6</sup>ATEpre: 基于多组学数据整合的YTHDF1介导调控翻译效率的m<sup>6</sup>A修饰位点预测\*

张 腾<sup>1)</sup> 张 明<sup>1)</sup> 张绍武<sup>2)</sup> 刘 恋<sup>3)\*\*</sup>

(<sup>1)</sup> 江苏科技大学计算机学院, 镇江 212100; (<sup>2)</sup> 西北工业大学自动化学院, 信息融合技术教育部重点实验室, 西安 710072;

(<sup>3)</sup> 陕西师范大学人工智能与计算机学院, 西安 710119)

**摘要 目的** N6-甲基腺苷 (m<sup>6</sup>A) 作为最丰富的 mRNA 表观修饰, 在基因表达、mRNA 翻译等各种 mRNA 代谢过程中起到重要作用。然而, 由于缺乏有效的计算方法, 目前还无法从计算角度识别哪些 m<sup>6</sup>A 甲基化修饰位点能够通过 m<sup>6</sup>A RNA 结合蛋白 1 (YTHDF1) 介导的生物学机理调控 mRNA 翻译效率。鉴于此, 本文设计了全新的计算方法 (m<sup>6</sup>ATEpre), 预测特定细胞系中调控翻译效率的 m<sup>6</sup>A 甲基化修饰位点。**方法** m<sup>6</sup>ATEpre 整合多组学数据, 如 MeRIP-seq 数据、PAR-CLIP 数据和 Ribo-seq 数据, 针对含 m<sup>6</sup>A 修饰位点的序列提出了全新的特征表示策略, 并采用自动编码器有效获取嵌入特征信息。**结果** 在 HeLa 细胞系测序数据集的各种实验结果表明, 相比其他方法, m<sup>6</sup>ATEpre 在调控翻译效率的 m<sup>6</sup>A 位点预测方面取得了较高的预测性能。生物信息分析表明, 在 HeLa 细胞系中调控翻译效率的 m<sup>6</sup>A 位点具有特异性, 并揭示了 YTHDF1 介导的 m<sup>6</sup>A 甲基化调控翻译效率的潜在机制。此外, 在 HEK293T 细胞系测序数据集上进行预测和分析, 发现调控翻译效率的 m<sup>6</sup>A 位点具有细胞特异性。**结论** m<sup>6</sup>ATEpre 是一个及时有效的计算工具, 有助于对 m<sup>6</sup>A 甲基化调控翻译效率机制的理解。m<sup>6</sup>ATEpre 的代码和相关数据可以从 <https://www.scidb.cn/s/bAZZFr> 获取。

**关键词** m<sup>6</sup>A 甲基化修饰, YTHDF1 介导的翻译效率调控, 多组学数据整合, 特征表示

**中图分类号** TP18

**DOI:** 10.3724/j.pibb.2026.0005

**CSTR:** 32369.14.pibb.20260005

\* 陕西省自然科学基金基础研究计划 (2024JCYBQN-0624) 和陕西师范大学中央高校基本科研业务费 (GK202406008) 资助项目。

\*\* 通讯联系人。

Tel: 029-85310161, E-mail: [lian.l@snnu.edu.cn](mailto:lian.l@snnu.edu.cn)

收稿日期: 2026-01-04, 接受日期: 2026-03-17