m5UPred: A Web Server for the Prediction of RNA 5-Methyluridine Sites from Sequences

Jie Jiang,1,5 Bowen Song,2,5 Yujiao Tang,1,5 Kunqi Chen,1,4 Zhen Wei,1,5 and Jia Meng1,3,5

Department of Biological Sciences, Xi’an Jiaotong-Liverpool University, Suzhou, Jiangsu, 215123, China;²Department of Mathematical Sciences, Xi’an Jiaotong-Liverpool University, Suzhou, Jiangsu, 215123, China;³AI University Research Centre, Xi’an Jiaotong-Liverpool University, Suzhou, Jiangsu, 215123, China;⁴Institute of Ageing & Chronic Disease, University of Liverpool, L7 8TX, Liverpool, UK;⁵Institute of Systems, Molecular and Integrative Biology, University of Liverpool, L7 8TX, Liverpool, UK

As one of the widely occurring RNA modifications, 5-methyluridine (m5U) has recently been shown to play critical roles in various biological functions and disease pathogenesis, such as under stress response and during breast cancer development. Precise identification of m5U sites on RNA is vital for the understanding of the regulatory mechanisms of RNA life. We present here m5UPred, the first web server for in silico identification of m5U sites from the primary sequences of RNA. Built upon the support vector machine (SVM) algorithm and the biochemical encoding scheme, m5UPred achieved reasonable prediction performance with the area under the receiver operating characteristic curve (AUC) greater than 0.954 by 5-fold cross-validation and independent testing datasets. To critically test and validate the performance of our newly proposed predictor, the experimentally validated m5U sites were further separated by high-throughput sequencing techniques (miCLIP-Seq and FICC-Seq) and cell types (HEK293 and HAPI). When tested on cross-technique and cross-cell-type validation using independent datasets, m5UPred achieved an average AUC of 0.922 and 0.926 under mature mRNA mode, respectively, showing reasonable accuracy and reliability. The m5UPred web server is freely accessible now and it should make a useful tool for the researchers who are interested in m5U RNA modification.

INTRODUCTION

The development of high-resolution transcriptome mapping and quantification technologies has made epigenetic modifications of RNA one of the fastest-growing fields in biological research in the past several years. Over 170 post-transcriptional modifications have been identified, with the majority of them occurring in tRNAs and rRNAs.1 RNA modifications showed imperative roles in varied biological functions (e.g., embryonic stem cell development,2 cancer cell survival, migration,3 and response to environmental exposures4). Over 100 RNA modification enzyme mutations have been associated with human diseases.5 Besides the well-known role of fine-tuning RNA structures and functions to regulate gene expression and protein synthesis by RNA modifications,6 there are many other functions associated with RNA modifications. Some post-transcriptional RNA modifications have shown to be dynamic processes that have regulatory roles similar to post-translational protein modifications in controlling cell-type-specific functions.7

However, as one of the most abundant RNA modifications, the identification and functional characterization of 5-methyluridine (m5U) remain extremely limited in the literature. As a pyrimidine modification, m5U involves methylation at the 5-carbon position of uridine, which may be the first pyrimidine methyltransferases evolved to catalyze the pyrimidine-C5 methylation.8 The enzymes catalyzing the modification of m5U are TrmA in Escherichia coli,9,10 Trm2p in Saccharomyces cerevisiae,11 and TRMT2A and TRMT2B in mammals.12,13 m5U has been reported to participate in the development of breast cancer,14 systemic lupus erythematosus,15 and regulation of stress response and development in plants.16 Accurate identification of m5U sites is crucial to understanding fundamental biological processes and functions in all species. Some wet-lab experimental methods, such as miCLIP-Seq, iCLIP, and FICC-seq, have been developed to identify m5U sites.17,18,19 However, RNA sequencing could be a high-cost and time-consuming process, and the specificity of antibodies used for immunoprecipitation restricts the delivery of accurate sequencing results. Thus, only very limited data have been generated for m5U sites so far. In this study, we would like to propose an in silico technique for the identification of m5U sites based on sequence-derived information using a support vector machine (SVM) algorithm.

A number of computational methods have been developed to predict epigenetic modifications of RNA, including m6A site predictors WHISTLE,17 SRAMP,18 Gene2vec,19 iRNA-Methyl,20 and M6AMRFS;21 m5C site predictors RNA5CFinder,22 iRNA-m5C,23 and M5C-HPCR;24 pseudouridine site predictors iRNA-PseU,25 PseUI,26 PPUS,27 and PIANO;28 and one-stop platform iMRM for simultaneously identifying various RNA modifications in multiple species.29 However, to the best of our knowledge, there is no m5U site predictor available so far. Thus, in this study we would like to propose the first prediction framework, m5UPred, which can be utilized for high-accuracy identification of m5U site from RNA sequences. A

Received 7 July 2020; accepted 25 September 2020; https://doi.org/10.1016/j.omtn.2020.09.031.
Correspondence: Bowen Song, Department of Mathematical Sciences, Xi’an Jiaotong-Liverpool University, Suzhou, Jiangsu, 215123, China.
E-mail: bowen.song@liverpool.ac.uk
AUC for mature mRNA mode.

When evaluated by the independent testing data, m5UPred showed that the positive m5U sites generated from different techniques and cell types, we further evaluated the performance of m5UPred by cross-technique and cross-cell-type validation (Table 3). When tested by the independent dataset generated from another technique or cell type, m5UPred achieved AUC of 0.882 and 0.922 for cross-technique validation, and AUC of 0.899 and 0.926 for cross-cell-type validation, under full transcript mode and mature mRNA mode, respectively. It shows robustness and reliability of m5UPred.

Web Implementation
To facilitate the access of our model by experimental researchers, a web server has been developed using Hyper Text Markup Language (HTML), Cascading Style Sheets (CSS), and Hypertext Preprocessor (PHP) and is accessible at https://www.xjtlu.edu.cn/biologicalsciences/m5u. It allows users to submit query RNA sequences in FASTA format with over 41 nt in length and a uridine in the center for analysis. The web server will evaluate the possibility of m5U modification in the given sequences, which return all putative m5U sites with download function (Figure 1).

DISCUSSION
In this study, we extracted m5U modification sites from a data source generated by two sequencing methods and two cell types. A high-accuracy predictor was built by using a SVM model to predict the m5U modification sites. Satisfactory prediction performance was observed, with an average AUC of 0.956 for full transcript and 0.954 for mature mRNA when it was evaluated by an independent test set with 20% randomly selected sites from the dataset. The model was further tested by cross-technique and cross-cell-type validation. The results show that the positive m5U sites generated from different techniques and cell types share some common features, and these features can be captured by our m5UPred by extracting sequence-derived information. To help understand the sequence composition, we analyzed full transcript sequences from different cell types and sequencing methods by DREME motif discovery tool. The motifs identified are provided in Table S1 and are consistent with the tRNA T-loop motif of GTTCG/AA proposed by the data source paper. A web server was developed based on our model and made available to the public to assist the prediction of m5U sites by other researchers.

Table 1. Performance Evaluation of m5UPred by Independent Testing Dataset

| Mode       | Algorithms | Sn (%) | Sp (%) | ACC (%) | MCC | AUC  |
|------------|------------|--------|--------|---------|-----|------|
| Full transcript | SVM        | 86.06  | 84.72  | 85.39   | 0.708| 0.933 |
|            | RF         | 85.18  | 84.44  | 84.31   | 0.687| 0.929 |
|            | NB         | 86.06  | 61.53  | 73.80   | 0.491| 0.845 |
|            | GLM        | 83.03  | 80.45  | 81.74   | 0.635| 0.897 |
| Mature mRNA | SVM        | 86.99  | 89.07  | 88.03   | 0.761| 0.951 |
|            | RF         | 81.30  | 97.56  | 89.43   | 0.799| 0.951 |
|            | NB         | 92.36  | 46.91  | 69.63   | 0.441| 0.862 |
|            | GLM        | 85.57  | 81.54  | 83.56   | 0.672| 0.915 |

Table 2. Prediction Performance Using Cross Validation and Independent Testing Dataset

| Dataset          | Testing Method       | Sn (%) | Sp (%) | ACC (%) | MCC | AUC  |
|------------------|----------------------|--------|--------|---------|-----|------|
| Full transcript  | cross validation     | 87.59  | 89.04  | 88.32   | 0.767| 0.956 |
|                  | independent test set | 87.90  | 88.80  | 88.35   | 0.767| 0.956 |
| Mature mRNA      | cross validation     | 88.64  | 91.18  | 89.91   | 0.798| 0.956 |
|                  | independent test set | 87.44  | 91.95  | 89.70   | 0.795| 0.954 |

80% of experimentally validated m5U sites were used for training, while its performance was evaluated by the remaining 20% as independent testing data.

We randomly selected 80% of experimentally validated m5U sites as training dataset and the performance of predictors was evaluated by the remaining 20% of m5U sites as independent testing data.

*SVM achieved best performance among all classifier with AUC of 0.933 and 0.951 for full transcript mode and mature mRNA mode.

A user-friendly web server has also been developed and made free access publicly at https://www.xjtlu.edu.cn/biologicalsciences/m5u. We anticipate that our newly proposed model, m5UPred, could make the best use of limited experimentally detected data and facilitate the research of m5U site modification by providing an alternative computational approach.

RESULTS
Determine the Best Machine Learning Algorithm Used for m5U Site Prediction
In order to determine the best classifier for constructing m5UPred, the performances of different classifiers were tested on the independent dataset (see Table 1). SVM achieved area under the receiver operating characteristic curve (AUC) of 0.933 for full transcript mode and 0.951 for mature mRNA mode, which was the best among all classifiers and used to build m5UPred.

Performance Evaluation of m5UPred by Benchmark and Independent Testing Dataset
The prediction performance of m5UPred was evaluated by 5-fold cross-validation and an independent testing dataset, respectively (see Table 2). Using SVM as the final classifier, m5UPred was developed and tested using the same datasets that were previously applied for algorithm selection. In addition, the tune length of SVM in caret package was set to 9, with a random grid search for optimization. When evaluated by the independent testing data, m5UPred showed good generalization capability and achieved 0.767 Matthews correlation coefficient (MCC) with 88.35% accuracy and 0.956 AUC for full transcript mode and 0.795 MCC with 89.70% accuracy and 0.954 AUC for mature mRNA mode.

Performance Evaluation of m5UPred by Cross-Technique and Cross-Cell-Type Validation
It was shown previously that positive samples of RNA modification captured by different techniques may have different overall patterns. To test the performance and robustness of our newly proposed predictor in finding m5U sites generated from different techniques and cell types, we further evaluated the performance of m5UPred by cross-technique and cross-cell-type validation (Table 3). When tested by the independent dataset generated from another technique or cell type, m5UPred achieved AUC of 0.882 and 0.922 for cross-technique validation, and AUC of 0.899 and 0.926 for cross-cell-type validation, under full transcript mode and mature mRNA mode, respectively. It shows robustness and reliability of m5UPred.

We randomly selected 80% of experimentally validated m5U sites as training dataset and the performance of predictors was evaluated by the remaining 20% as independent testing data.

*SVM achieved best performance among all classifier with AUC of 0.933 and 0.951 for full transcript mode and mature mRNA mode.

A user-friendly web server has also been developed and made free access publicly at https://www.xjtlu.edu.cn/biologicalsciences/m5u. We anticipate that our newly proposed model, m5UPred, could make the best use of limited experimentally detected data and facilitate the research of m5U site modification by providing an alternative computational approach.

RESULTS
Determine the Best Machine Learning Algorithm Used for m5U Site Prediction
In order to determine the best classifier for constructing m5UPred, the performances of different classifiers were tested on the independent dataset (see Table 1). SVM achieved area under the receiver operating characteristic curve (AUC) of 0.933 for full transcript mode and 0.951 for mature mRNA mode, which was the best among all classifiers and used to build m5UPred.

Performance Evaluation of m5UPred by Benchmark and Independent Testing Dataset
The prediction performance of m5UPred was evaluated by 5-fold cross-validation and an independent testing dataset, respectively (see Table 2). Using SVM as the final classifier, m5UPred was developed and tested using the same datasets that were previously applied for algorithm selection. In addition, the tune length of SVM in caret package was set to 9, with a random grid search for optimization. When evaluated by the independent testing data, m5UPred showed good generalization capability and achieved 0.767 Matthews correlation coefficient (MCC) with 88.35% accuracy and 0.956 AUC for full transcript mode and 0.795 MCC with 89.70% accuracy and 0.954 AUC for mature mRNA mode.

Performance Evaluation of m5UPred by Cross-Technique and Cross-Cell-Type Validation
It was shown previously that positive samples of RNA modification captured by different techniques may have different overall pat-
It is worth noting that although m5U sites were generated under different conditions, they are still from a single source; this may pose an over-fitting problem, and the performance of the model may be over-estimated. We calculated the false discovery rate (FDR) and false omission rate (FOR) at different thresholds and the true positive rate (TPR). The result confirmed our previous statement about over-estimation, and the result is provided in the Supplemental Information. We will keep updating our model when more data are available. When comparing the cross-link peaks from two sequencing methods of the raw data, only around 35% of sites with over 5 crosslink peaks are identified by both methods; therefore, the true-positive m5U sites will need to be further confirmed by more experimental data. We noticed that tRNA modifications take the dominant position in the dataset we used for training, which was also confirmed by the data source paper. Considering the biological structures and functional difference of varied RNA molecules, further study will need to be performed to optimize the model using optimized window size, secondary structures, modification motif, and genome information to further improve the model robustness and generalization ability.

**MATERIALS AND METHODS**

**Training and Testing Datasets**

**Benchmark and Independent Testing Dataset Used for m5U Site Prediction**

The positive samples (m5U sites) were obtained from the recently published single-nucleotide resolution m5U sequencing data, and the sequencing results were generated by FICC-seq and miCLIP-seq technologies on two cell lines, HEK293 and HAP1 (Table 4). Data were downloaded from Gene Expression Omnibus (GEO), with the GEO accession number GEO: GSE109183. The result confirmed our previous statement about over-estimation, and the result is provided in the Supplemental Information. We will keep updating our model when more data are available. When comparing the cross-link peaks from two sequencing methods of the raw data, only around 35% of sites with over 5 crosslink peaks are identified by both methods; therefore, the true-positive m5U sites will need to be further confirmed by more experimental data. We noticed that tRNA modifications take the dominant position in the dataset we used for training, which was also confirmed by the data source paper. Considering the biological structures and functional difference of varied RNA molecules, further study will need to be performed to optimize the model using optimized window size, secondary structures, modification motif, and genome information to further improve the model robustness and generalization ability.

**Table 3. Cross-Technique and Cross-Cell-Type Validation**

| Mode                | Testing Method | Evaluation Metric | Cross-Technique Validation | Cross-Cell-Type Validation |
|---------------------|----------------|-------------------|-----------------------------|----------------------------|
|                     |               | Sn (%)            | miCLIP-Seq                  | HEK293                     |
|                     |               |                   | 86.70                       | 89.67                      |
| cross validation    |               | Sp (%)            | 86.83                       | 91.37                      |
| Full transcript     | MEI            | ACC (%)           | 86.76                       | 89.05                      |
|                     |               | MCC               | 0.735                       | 0.612                      |
|                     |               | AUC               | 0.946                       | 0.677                      |
| independent dataset |               | Sn (%)            | 75.36                       | 56.48                      |
|                     |               | Sp (%)            | 89.23                       | 90.10                      |
|                     | MEI            | ACC (%)           | 82.29                       | 73.29                      |
|                     |               | MCC               | 0.652                       | 0.495                      |
|                     |               | AUC               | 0.910                       | 0.853                      |
| cross validation    | mature mRNA   | Sn (%)            | 88.34                       | 94.14                      |
|                     |               | Sp (%)            | 90.52                       | 98.04                      |
|                     | MEI            | ACC (%)           | 89.43                       | 96.09                      |
|                     |               | MCC               | 0.789                       | 0.922                      |
|                     |               | AUC               | 0.962                       | 0.992                      |
| mature mRNA         | independent dataset | Sn (%)        | 90.07                       | 38.81                      |
|                     |               | Sp (%)            | 90.86                       | 97.51                      |
|                     | MEI            | ACC (%)           | 90.46                       | 68.16                      |
|                     |               | MCC               | 0.809                       | 0.449                      |
|                     |               | AUC               | 0.970                       | 0.873                      |

As previously shown in Table 2, the experimentally validated m5U sites were further separated by high-throughput sequencing techniques and cell types, under full transcript and mature mRNA mode, respectively. Independent tests by the other technique or cell type are employed to evaluate the performance additional to 5-fold cross validation. When dataset miCLIP_F was used for training, its performance was tested by 5-fold cross-validation and an independent dataset of FICC_F. Similarly, while dataset FICC_F was used for training, miCLIP_F served as an independent test set. The same testing scheme was used for datasets separated by cell types as well. The performance of different sequencing techniques and cell types are then averaged. The performance evaluation of m5UPred by cross-technique and cross-cell-type validation using different machine learning classifiers is listed in Table S1.
The performance of our predictor was evaluated under two modes, namely full transcript mode and mature mRNA mode. For the full transcript mode, the positive and negative m\textsuperscript{5}U sites located in both exonic and intronic regions are all considered to generate the data, whereas, in the mature mRNA mode, only positive and negative m\textsuperscript{5}U sites located on mature mRNA transcripts are employed. Positive m\textsuperscript{5}U sites in different RNA families are listed in the Supplemental Information. For performance evaluation, 80% of the dataset was randomly selected as a benchmark training dataset, while the remaining 20% was used as independent testing data.

### Dataset Separated by High-Throughput Sequencing Techniques and Cell Types

Besides randomly selected training and testing data from experimentally validated m\textsuperscript{5}U sites, we applied two more strictly evaluation methods to validate the performance of the proposed predictor. The experimentally validated m\textsuperscript{5}U sites were further divided by high-throughput sequencing technique and cell type (Table 5), from which the cross-technique and cross-cell-type testing were applied for performance evaluation under full transcript and mature mRNA modes, respectively. The performance was also evaluated by 5-fold cross-validation and an independent test set. But rather than randomly selecting 20% of the total data as a test set, we generated test sets by different techniques and cell types. Concretely, for cross-technique validation, when dataset miCLIP\textsubscript{F} or miCLIP\textsubscript{M} were used for training, their performance was tested by an independent dataset of FICC\textsubscript{F} or FICC\textsubscript{M}. Then, we switched the train and test set and used dataset FICC\textsubscript{F} or FICC\textsubscript{M} for training and miCLIP\textsubscript{F} or miCLIP\textsubscript{M} for independent testing. The validation scheme was applied for the dataset separated by cell types as well.

### Feature Extraction

High classification accuracy of many machine learning algorithms largely relies on sequence encoding strategies for feature extraction of RNA sequences. To achieve the best performance, the chemical properties of nucleotides and their distribution information are employed for sequence feature extraction in this study. This strategy was adapted from previous work by Bari et al.,\textsuperscript{34} which was originally used for splicing site prediction from DNA sequence and lately is being widely used in encoding RNA sequences for RNA modification prediction.\textsuperscript{35–37}

#### Nucleotide Density

The nucleotide density represents the distribution and frequency information of nucleotides at each position. The density ($d_i$) of a nucleotide $N$ at $i$th position can be calculated by the number ($n$) of $N$ occurred before ($i + 1$)th position divided by $i$: $d_i = n/i$. So, for a sequence “AUAGUCAUAA,” the density of A is 1, 0.67, 0.43, 0.44, and 0.50 at the 1st, 3rd, 7th, 9th, and 10th positions, respectively. Similarly, U is 0.50, 0.40, and 0.38 at positions 2, 5, and 8; C is 0.17 at position 6; and G is 0.25 at position 4 (Table 6).

#### Nucleotide Chemical Property

The encoding scheme of nucleotide chemical property was designed based on different chemical structures of four RNA nucleotides. Four building blocks of RNA, namely adenosine (A), uridine (U), guanosine (G), and cytosine (C), are categorized into different groups depending on the number of ring structures (two for A and G and one for C and T), existence of an amino group (A and C) or keto group (G

### Table 4. Base-Resolution Dataset Used for m\textsuperscript{5}U Site Prediction

| GEO Accession | Technique | Cell Line | Full No. of Sites | Mature No. of Sites | Source |
|---------------|-----------|-----------|-------------------|---------------------|--------|
| GSE109183     | miCLIP-Seq | HEK293    | 3,696             | 1,232               |        |
|               | FICC-Seq  | HAP1      |                   |                     |        |

80% of the data were randomly selected for training, while the remaining 20% were retained as an independent testing dataset for performance evaluation. All positive and negative sequences used in this project can be freely downloaded from [https://www.xjtlu.edu.cn/biologicalsciences/m5u](https://www.xjtlu.edu.cn/biologicalsciences/m5u) and are provided in the Supplemental Information.
and T), and strong (C and G) or weak (A and T) hydrogen bonds. Based on this information, together with the nucleotide density information, nucleotide N at i-th position from sequence S (with length l) can be represented by the formula $N_i = \{x_i, y_i, z_i, d_i\}$ ($i = 1, 2, 3, \ldots l$) which satisfies the following equations:

$$x_i = \begin{cases} 1 & \text{if } s_i \in \{A, G\} \\ 0 & \text{if } s_i \in \{C, U\} \end{cases}, y_i = \begin{cases} 1 & \text{if } s_i \in \{A, C\} \\ 0 & \text{if } s_i \in \{G, U\} \end{cases}, z_i = \begin{cases} 1 & \text{if } s_i \in \{A, U\} \\ 0 & \text{if } s_i \in \{C, G\} \end{cases}$$

(1)

Concretely, A, C, G, and U can be encoded as vectors (1,1,1,1), (0,1,0,0), (1,0,0,1), and (0,0,1,0), respectively. So, each of the nucleotides in the RNA sequences will be transferred into four numeric values; thus, each of our 41 nt RNA sequences will be encoded into a 164-dimension vector.

**Evaluation Matric**

Five metrics were employed to evaluate the performance of our model, namely Sn (Sensitivity), Sp (Specificity), MCC, ACC (overall accuracy), and AUC, with the equations below:

$$Sn = \frac{TP}{TP + FN}$$

(2)

$$Sp = \frac{TN}{TN + FP}$$

(3)

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}}$$

(4)

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

(5)

in which TP, TN, FP, and FN represent the number of true positives, true negatives, false positives, and false negatives, respectively. A threshold of 0.5 was applied to calculate Sn and Sp. R package ROCR\(^{33}\) was used to evaluate the performance of the model.

**Choice of Machine Learning Classifier**

SVM, random forest (RF), generalized linear model (GLM), and Naïve Bayes (NB) are the most popular machine learning classifiers in RNA modification prediction and have been widely used for different site predictions.\(^{24,27,35}\) We evaluated the performance of these algorithms by an independent testing dataset, since the evaluation by cross-validation may over-estimate the performance of models.\(^{39}\) The R package caret was used to construct machine learning models, and all parameters were set by default for primitive evaluation. The results are shown in Table 1. SVM achieved the best results among all four algorithms, which was selected as our final algorithm to build our m\(^5\)U site prediction model.

**SVM**

SVM is one of the most powerful yet flexible algorithms in bioinformatics research and has been used for many applications (e.g., mammalian microRNA target prediction,\(^{40}\) predicting the subcellular location of proteins,\(^{41}\) and RNA modification site prediction\(^{42,43}\)). The final decision of SVM is determined by a representation of different classes in a hyperplane, which can be used for varied classification and regression tasks. In this study, we used the R package caret to construct our SVM model with a non-linear radial basis function. The prediction performance was evaluated by 5-fold cross-validation and an independent test set. Initially, all parameters were set by default to test the performance of the model, and then tune length was set to 9, with a random grid search for optimization. Random grid search has been proved by a previous study with better efficiency than manually set grid searches.\(^{44}\)

**SUPPLEMENTAL INFORMATION**

Supplemental Information can be found online at https://doi.org/10.1016/j.omtn.2020.09.031.

**AUTHOR CONTRIBUTIONS**

B.S. conceived the idea; K.C. and B.S. collected and processed the data; J.J. and B.S. constructed the predictive model and designed the experiments; Y.T. built the website; and J.J. and B.S. drafted the manuscript. All authors read, critically revised and approved the final manuscript.

**CONFLICTS OF INTEREST**

The authors declare no competing interests.

**ACKNOWLEDGMENTS**

This work was supported by the National Natural Science Foundation of China (31671373) and the XJTLU Key Program Special Fund (KSF-E-51). This work is partially supported by the AI University Research Centre through XJTLU Key Programme Special Fund (KSF-P-02).

**REFERENCES**

1. Boccaletto, P., Machnicka, M.A., Purta, E., Piatkowski, P., Baginski, B., Wirecki, T.K., de Crécy-Lagard, V., Ross, R., Limbach, P.A., Kotter, A., et al. (2018). MODOMICS: a database of RNA modification pathways. 2017 update. Nucleic Acids Res. 46 (D1), D303–D307.
