im6A-TS-CNN: Identifying the N^6-Methyladenine Site in Multiple Tissues by Using the Convolutional Neural Network

Kewei Liu,[1] Lei Cao,[1] Pufeng Du,[2] and Wei Chen[1,3]

[1]School of Life Sciences, Center for Genomics and Computational Biology, North China University of Science and Technology, Tangshan 063210, China; [2]College of Intelligence and Computing, Tianjin University, Tianjin 300350, China; [3]Innovative Institute of Chinese Medicine and Pharmacy, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China

N^6-methyladenosine (m^{6}A) is the most abundant post-transcriptional modification and involves a series of important biological processes. Therefore, accurate detection of the m^{6}A site is very important for revealing its biological functions and impacts on diseases. Although both experimental and computational methods have been proposed for identifying m^{6}A sites, few of them are able to detect m^{6}A sites in different tissues. With the consideration of the spatial specificity of m^{6}A modification, it is necessary to develop methods able to detect the m^{6}A site in different tissues. In this work, by using the convolutional neural network (CNN), we proposed a new method, called im6A-TS-CNN, that can identify m^{6}A sites in brain, liver, kidney, heart, and testis of *Homo sapiens*, *Mus musculus*, and *Rattus norvegicus*. In im6A-TS-CNN, the samples were encoded by using the one-hot encoding scheme. The results from both a 5-fold cross-validation test and independent dataset test demonstrate that im6A-TS-CNN is better than the existing method for the same purpose. The command-line version of im6A-TS-CNN is available at https://github.com/liukeweiaway/DeepM6A_cnn.

INTRODUCTION

As a common and abundant of RNA post-transcriptional modification (PTM), N^6-methyladenosine (m^{6}A) modification plays an important role in almost all processes of cell cycles, such as affecting translation efficiency,[1] cell development,[2] cell viability,[3] etc. m^{6}A is catalyzed by a methyltransferase complex containing METTL3, METTL14, and WTAP. As a kind of dynamic PTM, m^{6}A can be erased by the demethylases FTO and ALKBH5.[4] Recently, more and more studies have revealed that m^{6}A is closely correlated with diseases, such as obesity,[5] thyroid tumor,[6] prostate cancer,[7] zika virus,[8] and acute myelogenous leukemia.[9] However, our knowledge about the functions of m^{6}A modifications is still unintelligible. To deepen our understanding on the functions of m^{6}A, the key step is to know the precise position of m^{6}A in transcriptomes.

There are two main ways to identify m^{6}A sites. One way is using experimental methods, such as Methylated RNA Immunoprecipitation (MeRIP),[10] m^{6}A sequencing (m6A-seq)[11], photo-crosslinking-assisted (PA)-m6A-seq,[12] and m^{6}A-crosslinking immunoprecipitation (CLIP).[13] These experimental methods laid important foundations for the detection of m^{6}A modification sites. Accordingly, some bioinformatics tools that are able to detect m^{6}A sites directly from the reads generated by the experiments were proposed.[14,15] However, as the amount of sequencing data increases, we need to find an effective and efficient way to detect m^{6}A in the transcriptome. Accordingly, sequence information-based computational methods were proposed to identify m^{6}A sites. These methods can be queried in a recent review.[16] With the research on the spatial specificity of gene expression, it has been found that the location of the m^{6}A site is distinct in different tissues and species. Therefore, Dao et al.[17] proposed a tool, called iRNA-m6A that can identify m^{6}A modification sites in different tissues in human, mouse, and rat by using the algorithm of SVM, based on the data of Zhang et al.[18] This method greatly improves the accuracy of predicting the m^{6}A site. However, the performance for predicting the m^{6}A site still has great potential to be improved.

In recent years, the deep learning algorithms made great contributions to bioinformatics. A large number of computational methods based on deep-learning algorithm, such as Gene2Vec,[19] BERMP,[20] DeepM6ASeq,[21] and iPseU-CNN[22] have been proposed. Inspired by these successful applications of the deep-learning algorithm in identifying RNA modifications, in the present work, we proposed a convolutional neural network (CNN)-based method, called im6A-TS-CNN, to identify m^{6}A sites in different tissues from human, mouse, and rat. Results from a 5-fold cross-validation test and independent dataset test demonstrated that the performance of im6A-TS-CNN is better than or comparable with that of the existing method for the same aim. Moreover, the universality of im6A-TS-CNN was also demonstrated by a cross-species validation test. The framework of im6A-TS-CNN is illustrated in Figure 1.
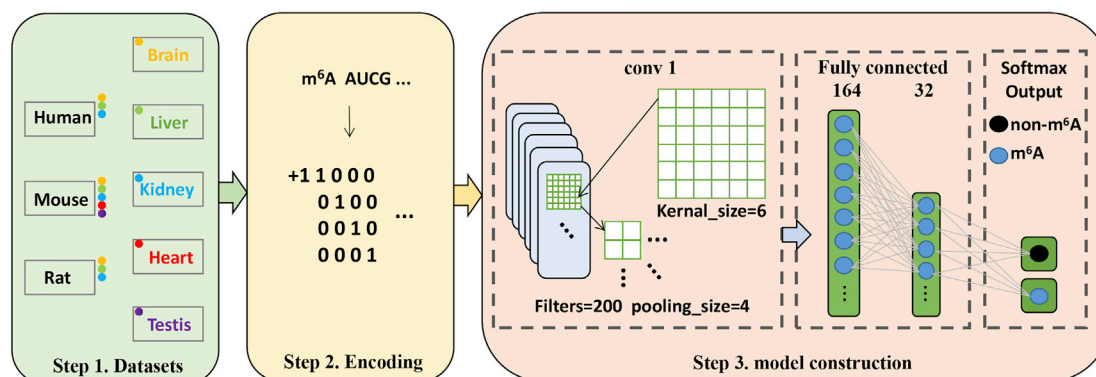
Figure 1. The Framework of the im6A-TS-CNN

The first step is to collect tissue-specific m⁶A data from the human, mouse, and rat. The second step is encoding the sequences by using the one-hot scheme. The third step is model construction.

RESULTS AND DISCUSSION

Model Performance

In this article, the Keras in TensorFlow 2.0 under Python 3.6 was used to perform the predictions. The results from a 5-fold cross-validation test and independent dataset test of the proposed method for identifying the tissue-specific m^{6}A modification sites in the human, mouse, and rat were shown in Table 1. With the comparison of results from a 5-fold cross-validation test and independent test, it was found that the proposed method is stable for identifying the m^{6}A sites.

To measure objectively the performance of the proposed method, the receiver operating characteristic (ROC) curves[23,24] from a 5-fold cross-validation test and independent test were plotted in Figure 2 as well. It was found that most of the areas under the ROC curve (AUCs) are higher than 0.8 in both the 5-fold cross-validation test and independent test, demonstrating the reliability of the proposed method for identifying tissue-specific m^{6}A sites.

Comparison with Existing Method

To further testify the superiority of im6A-TS-CNN, we compared its performance with that of Zhang et al.'s[18] iRNA-m6A model, based on both the 5-fold cross-validation test and independent test. The comparative results in terms of AUC are shown in Table 2. Except for the identification of the m^{6}A sites from the brain of mouse and rat, im6A-TS-CNN outperforms iRNA-m6A for the identification of m^{6}A sites in the other tissues in the human, mouse, and rat. These results demonstrate that im6A-TS-CNN is a powerful tool for identifying tissue-specific m^{6}A sites from different species.

Cross-Species and Cross-Tissue Validation

Since the datasets are from different species and tissues, it is interesting to test whether the model, trained based on the data from a specific tissue in a species, is able to identify m^{6}A from other tissues and species. Accordingly, the cross-species and cross-tissue validation was performed. The AUCs of im6A-TS-CNN for identifying m^{6}A sites from other species and tissues are shown in Figure 3. As shown in Figure 3, it can be concluded that im6A-TS-CNN is also effective for the cross-species and cross-tissue identification of m^{6}A sites, demonstrating the universality of the proposed method.

Conclusions

In this article, we proposed a CNN-based method, called i6mA-TS-CNN, for identifying m^{6}A in the brain, liver, kidney, heart, and testis from the human, mouse, and rat. The results from a 5-fold cross-validation test and independent test demonstrate that i6mA-TS-CNN is better than the existing method for identifying tissue-specific m^{6}A. For the convenience of the scientific community, the command-line version of i6mA-TS-CNN, together with its source code and user manual, is provided at https://github.com/liukeweiaway/DeepM6A_cnn. In addition, the high-, normal-, and low-threshold options were provided to control the false-positive rate. The corresponding performance with different options was listed in Table S1. Taken together, we hope that the i6mA-TS-CNN will become a useful tool for identifying m^{6}A sites.

MATERIALS AND METHODS

Datasets

A high-quality dataset is very important for the construction of a computational model. In 2019, Zhang et al.[18] developed a high-throughput, antibody-independent m^{6}A detection method based on the m^{6}A-sensitive RNA endoribonuclease to identify the m^{6}A site in different tissues, namely the brain, liver, kidney, heart, and testis from the human, mouse, and rat. Based on these data, Dao et al.[17] built a high-quality benchmark dataset that can be used to train a computational method for identifying m^{6}A sites, which contains both m^{6}A site- and non-m^{6}A site-containing sequences with the length of 41 nt. The CD-HIT program[25] was used to make sure that the sequence similarity of the dataset was less than 80%. The detailed information of this dataset is provided in Table 3.

One-Hot Encoding

One-hot encoding is a common and effective method.[26] According to such a scheme, in an RNA segment, A is represented as (1,0,0,0), U as

Table 1. The Performance of im6A-TS-CNN for Identifying m⁶A Sites

	5-Fold Cross Validation					Independent Test				
	Sn (%)	Sp (%)	Acc (%)	MCC	AUC	Sn (%)	Sp (%)	Acc (%)	MCC	AUC
h_b	75.35	69.71	72.53	0.4523	0.8029	75.17	70.20	72.69	0.4543	0.8056
h_k	81.70	78.25	79.98	0.6006	0.8781	79.95	78.53	79.24	0.5848	0.8727
h_l	80.18	79.69	79.94	0.5992	0.8811	84.81	75.02	79.92	0.6012	0.8805
m_b	81.50	75.85	78.67	0.5749	0.8705	86.22	70.74	78.48	0.5765	0.8722
m_h	78.37	67.60	72.99	0.4633	0.8115	75.82	71.36	73.59	0.4723	0.8161
m_k	79.91	81.00	80.46	0.6094	0.8842	80.52	81.00	80.76	0.6151	0.8855
m_l	72.39	70.24	71.32	0.4288	0.7953	75.56	67.58	71.57	0.4328	0.7927
m_t	75.21	75.61	75.41	0.5090	0.8380	83.45	68.87	76.16	0.5288	0.8467
r_b	79.04	74.23	76.64	0.5379	0.8469	78.05	75.84	76.95	0.5391	0.8516
r_k	84.15	80.77	82.46	0.6500	0.9017	84.85	80.59	82.72	0.6550	0.9077
r_l	81.56	79.63	80.59	0.6126	0.8830	84.51	75.94	80.22	0.6067	0.8847

h, m and r before the hyphen stand for human, mouse, and rat, respectively; after the hyphen stand for brain, heart, kidney, liver, and testis, respectively.

(0,1,0,0), C as (0,0,1,0), and G as (0,0,0,1). Therefore, an RNA sequence of length l can be converted into a 4-l dimensional vector.

Convolutional Neural Network

In recent years, **Convolutional Neural Network (CNN)** has been widely used to solve biological problems.[22,27,28] The structure of the CNN is shown in Figure 1. It contains a convolutional layer with 200 filters in which the kernel size is 6. After convolution operation,

a max-pooling layer with the size of 4 was added. The convolution layer is mathematically represented and computed as the following:

$$\text{Conv}(R)_{jf} = \text{ReLU}\left(\sum_{s=0}^{S-1} \sum_{n=0}^{N-1} W_{sn}^{f} R_{j+s,n} \right), \quad \text{(Equation 1)}$$

where R represents the RNA segment, f denotes the index of the kernel, and j denotes the index of the output position. In Equation 1, each filter W^f is an S × N weight matrix, where S is the filter size, and
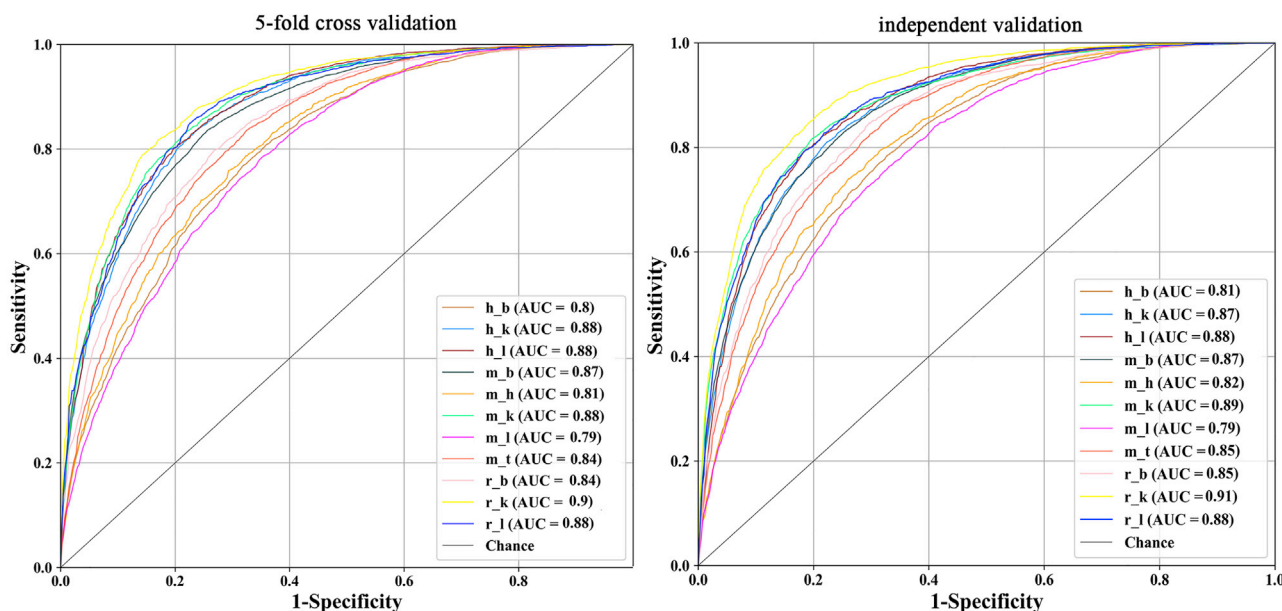

Figure 2. The ROC Curves for Identifying m⁶A in Different Tissues in the Three Species under the 5-Fold Cross-Validation Test and Independent Dataset Test
The value of AUC is given in the right corner of each graph.

Table 2. Comparative Results between im6A-TS-CNN and iRNA-m⁶A under the 5-Fold Cross-Validation Test and Independent Test

	5-Fold Cross Validation (AUC)			Independent Test (AUC)		
	m6A-TS-CNN	iRNA-m⁶A	Difference	im6A-TS-CNN	iRNA-m⁶A	Difference
h_b	0.8029	0.7756	0.0273*	0.8056	0.7845	0.0211*
h_k	0.8781	0.8634	0.0147*	0.8727	0.8565	0.0162*
h_l	0.8811	0.8738	0.0073*	0.8805	0.8681	0.0124*
m_b	0.8705	0.8731	−0.0026	0.8722	0.8613	0.0109*
m_h	0.8115	0.7948	0.0167*	0.8161	0.7878	0.0283*
m_k	0.8842	0.8726	0.0116*	0.8855	0.8697	0.0158*
m_l	0.7953	0.7743	0.0210*	0.7927	0.762	0.0307*
m_t	0.8380	0.8156	0.0224*	0.8467	0.8182	0.0285*
r_b	0.8469	0.8282	0.0187*	0.8516	0.8968	−0.0452
r_k	0.9017	0.8877	0.0140*	0.9077	0.8761	0.0316*
r_l	0.8830	0.8766	0.0064*	0.8847	0.8265	0.0582*

h, m and r before the hyphen stand for human, mouse and rat; b, h, k, l, t after the hyphen stand for brain, liver, kidney, heart and testis, respectively.
*inidcates the performance of im6A-TS-CNN is better than iRNA-m6A for identifying m⁶A sites.

N is the input channels. The rectified linear function (ReLU) is expressed as the following:

$$\text{ReLU}(z) = \begin{cases} z \text{ if } z \geq 0 \\ 0 \text{ if } z < 0 \end{cases} \quad \text{(Equation 2)}$$

In order to prevent overfitting, we choose to lose some parameters and set the dropout rate of 0.16. The results were output to a fully connected layer containing 164 neural units and then compressed to 32 neural units. Finally, the softmax function was used to predict whether the RNA segment contains m⁶A sites or not and is expressed as the following:

$$\text{Softmax}(x) = \frac{e^x}{\sum e^x} \quad \text{(Equation 3)}$$

When building the model, the stochastic gradient descent (SGD) was used as the optimizer with a learning rate of 0.001, and the categorical
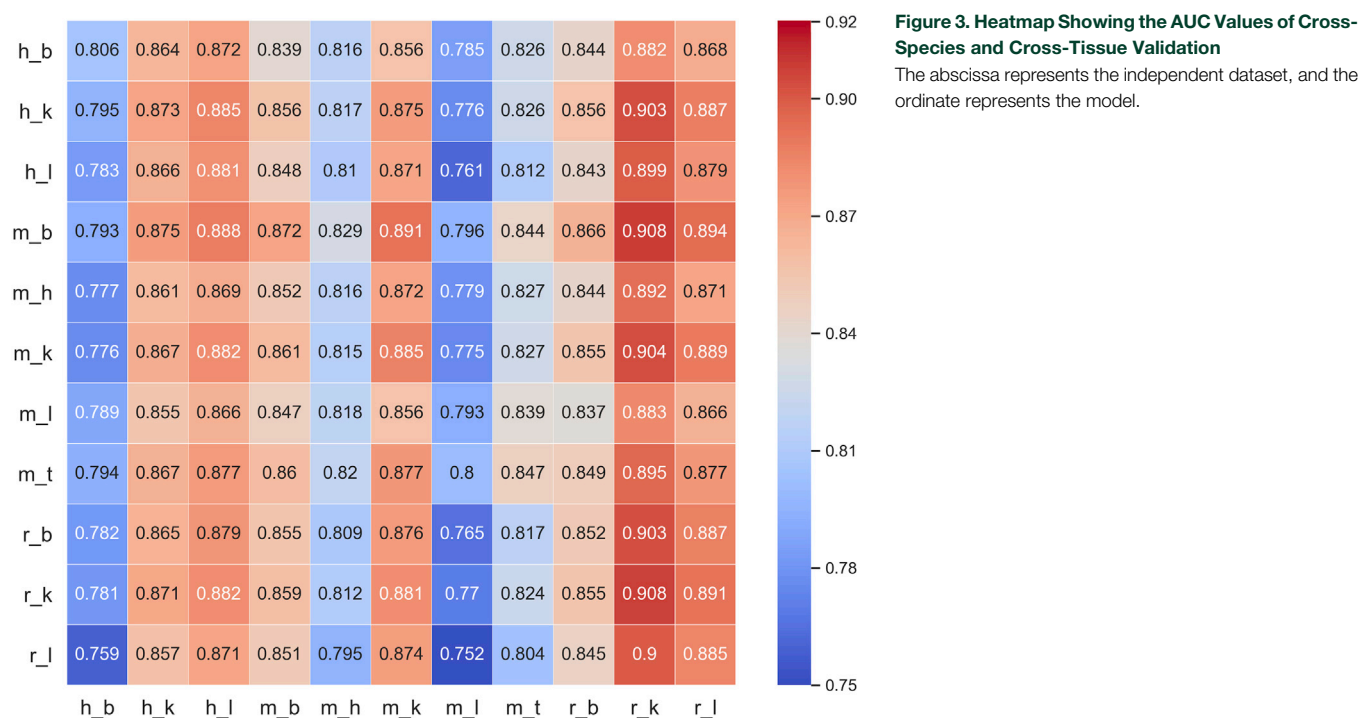

	h_b	h_k	h_l	m_b	m_h	m_k	m_l	m_t	r_b	r_k	r_l
h_b	0.806	0.864	0.872	0.839	0.816	0.856	0.785	0.826	0.844	0.882	0.868
h_k	0.795	0.873	0.885	0.856	0.817	0.875	0.776	0.826	0.856	0.903	0.887
h_l	0.783	0.866	0.881	0.848	0.81	0.871	0.761	0.812	0.843	0.899	0.879
m_b	0.793	0.875	0.888	0.872	0.829	0.891	0.796	0.844	0.866	0.908	0.894
m_h	0.777	0.861	0.869	0.852	0.816	0.872	0.779	0.827	0.844	0.892	0.871
m_k	0.776	0.867	0.882	0.861	0.815	0.885	0.775	0.827	0.855	0.904	0.889
m_l	0.789	0.855	0.866	0.847	0.818	0.856	0.793	0.839	0.837	0.883	0.866
m_t	0.794	0.867	0.877	0.86	0.82	0.877	0.8	0.847	0.849	0.895	0.877
r_b	0.782	0.865	0.879	0.855	0.809	0.876	0.765	0.817	0.852	0.903	0.887
r_k	0.781	0.871	0.882	0.859	0.812	0.881	0.77	0.824	0.855	0.908	0.891
r_l	0.759	0.857	0.871	0.851	0.795	0.874	0.752	0.804	0.845	0.9	0.885

Figure 3. Heatmap Showing the AUC Values of Cross-Species and Cross-Tissue Validation
The abscissa represents the independent dataset, and the ordinate represents the model.

Table 3. The Information of Benchmark Datasets for Predicting RNA m⁶A Sites

Name	Training		Testing	
	Positive	Negative	Positive	Negative
h_b	4,605	4,605	4,604	4,604
h_k	2,634	2,634	2,634	2,634
h_l	4,574	4,574	4,573	4,573
m_b	8,025	8,025	8,025	8,025
m_h	4,133	4,133	4,133	4,133
m_k	3,953	3,953	3,952	3,952
m_l	2,201	2,201	2,200	2,200
m_t	4,707	4,707	4,706	4,706
r_b	2,352	2,352	2,351	2,351
r_k	1,762	1,762	1,762	1,762
r_l	3,433	3,433	3,432	3,432

h, m and r before the hyphen stand for human, mouse and rat; b, h, k, l, t after the hyphen stand for brain, liver, kidney, heart and testis, respectively.

cross entropy was used as the loss function. In the training process, a total of 2,000 epochs were carried out by using the early stopping method with the patience of 50 and min_delta of 0.001.

Evaluation Metrics

In order to evaluate the model, we use the sensitivity (Sn), specificity (Sp), accuracy (Acc), and Matthews correlation coefficient (MCC), which are defined as the following,[26,29,30] to evaluate the performance of the following model:

$$
\begin{cases}
Sn = 1 - \dfrac{N_-^+}{N^+} \\[2mm]
Sp = 1 - \dfrac{N_+^-}{N^-} \\[2mm]
Acc = 1 - \dfrac{N_-^+ + N_+^-}{N^+ N^-} \\[4mm]
Mcc = \dfrac{1 - \dfrac{N_-^+ + N_+^-}{N^+ N^-}}{\sqrt{\left(1 + \dfrac{N_+^- - N_-^+}{N^+}\right)\left(1 + \dfrac{N_-^+ - N_+^-}{N^-}\right)}}
\end{cases}
, \qquad \text{(Equation 4)}
$$

where N^+ is the total number of the RNA sequence containing modification site, N_-^+ is the number of false-negative samples, N^- is the total number of the RNA sequence that did not contain any modification site, and N_+^- is teh number of false-positive samples.

In addition, we also used the ROC curve[31] and the area under the ROC curve (AUC) to evaluate the proposed model.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.omtn.2020.07.034.

REFERENCES

1. Wang, X., Zhao, B.S., Roundtree, I.A., Lu, Z., Han, D., Ma, H., Weng, X., Chen, K., Shi, H., and He, C. (2015). N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. Cell 161, 1388–1399.

2. 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.

3. Bokar, J.A. (2005). The biosynthesis and functional roles of methylated nucleosides in eukaryotic mRNA. In Fine-Tuning of RNA Functions by Modification and Editing. Topics in Current Genetics, H. Grosjean, ed. (Springer), pp. 141–177.

4. Wu, R., Jiang, D., Wang, Y., and Wang, X. (2016). N (6)-Methyladenosine (m(6)A) Methylation in mRNA with A Dynamic and Reversible Epigenetic Modification. Mol. Biotechnol. 58, 450–459.

5. Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., Yi, C., Lindahl, T., Pan, T., Yang, Y.G., and He, C. (2011). N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat. Chem. Biol. 7, 885–887.

6. Heiliger, K.J., Hess, J., Vitagliano, D., Salerno, P., Braselmann, H., Salvatore, G., Ugolini, C., Summerer, I., Bogdanova, T., Unger, K., et al. (2012). Novel candidate genes of thyroid tumourigenesis identified in Trk-T1 transgenic mice. Endocr. Relat. Cancer 19, 409–421.

7. Machiela, M.J., Lindström, S., Allen, N.E., Haiman, C.A., Albanes, D., Barricarte, A., Berndt, S.I., Bueno-de-Mesquita, H.B., Chanock, S., Gaziano, J.M., et al. (2012). Association of type 2 diabetes susceptibility variants with advanced prostate cancer risk in the Breast and Prostate Cancer Cohort Consortium. Am. J. Epidemiol. 176, 1121–1129.

8. 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.

9. Bansal, H., Yihua, Q., Iyer, S.P., Ganapathy, S., Proia, D.A., Penalva, L.O., Uren, P.J., Suresh, U., Carew, J.S., Karnad, A.B., et al. (2014). WTAP is a novel oncogenic protein in acute myeloid leukemia. Leukemia 28, 1171–1174.

10. 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.

11. Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M., et al. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature 485, 201–206.

12. Chen, K., Lu, Z., Wang, X., Fu, Y., Luo, G.-Z., Liu, N., Han, D., Dominissini, D., Dai, Q., Pan, T., and He, C. (2015). High-resolution N(6) -methyladenosine (m(6) A) map using photo-crosslinking-assisted m(6) A sequencing. Angew. Chem. Int. Ed. Engl. 54, 1587–1590.

13. Ke, S., Alemu, E.A., Mertens, C., Gantman, E.C., Fak, J.J., Mele, A., Haripal, B., Zucker-Scharff, I., Moore, M.J., Park, C.Y., et al. (2015). A majority of m6A residues

are in the last exons, allowing the potential for 3′ UTR regulation. Genes Dev. *29*, 2037–2053.

14. Chen, K., Wei, Z., Zhang, Q., Wu, X., Rong, R., Lu, Z., Su, J., de Magalhães, J.P., Rigden, D.J., and Meng, J. (2019). WHISTLE: a high-accuracy map of the human N6-methyladenosine (m6A) epitranscriptome predicted using a machine learning approach. Nucleic Acids Res. *47*, e41.

15. Meng, J., Lu, Z., Liu, H., Zhang, L., Zhang, S., Chen, Y., Rao, M.K., and Huang, Y. (2014). A protocol for RNA methylation differential analysis with MeRIP-Seq data and exomePeak R/Bioconductor package. Methods *69*, 274–281.

16. Chen, Z., Zhao, P., Li, F., Wang, Y., Smith, A.I., Webb, G.I., Akutsu, T., Baggag, A., Bensmail, H., and Song, J. (2019). Comprehensive review and assessment of computational methods for predicting RNA post-transcriptional modification sites from RNA sequences. Brief. Bioinform. Published online November 11, 2019. https://doi.org/10.1093/bib/bbz112.

17. Dao, F.-Y., Lv, H., Yang, Y.H., Zulfiqar, H., Gao, H., and Lin, H. (2020). Computational identification of N6-methyladenosine sites in multiple tissues of mammals. Comput. Struct. Biotechnol. J. *18*, 1084–1091.

18. Zhang, Z., Chen, L.-Q., Zhao, Y.-L., Yang, C.-G., Roundtree, I.A., Zhang, Z., Ren, J., Xie, W., He, C., and Luo, G.-Z. (2019). Single-base mapping of m(6)A by an antibody-independent method. Sci. Adv. *5*, eaax0250.

19. Zou, Q., Xing, P., Wei, L., and Liu, B. (2019). Gene2vec: gene subsequence embedding for prediction of mammalian N^6-methyladenosine sites from mRNA. RNA *25*, 205–218.

20. Huang, Y., He, N., Chen, Y., Chen, Z., and Li, L. (2018). BERMP: a cross-species classifier for predicting m6A sites by integrating a deep learning algorithm and a random forest approach. Int. J. Biol. Sci. *14*, 1669–1677.

21. Zhang, Y., and Hamada, M. (2018). DeepM6ASeq: prediction and characterization of m6A-containing sequences using deep learning. BMC Bioinformatics *19* (*Suppl 19*), 524.

22. Tahir, M., Tayara, H., and Chong, K.T. (2019). iPseU-CNN: Identifying RNA Pseudouridine Sites Using Convolutional Neural Networks. Mol. Ther. Nucleic Acids *16*, 463–470.

23. Hasan, M.M., Manavalan, B., Shoombuatong, W., Khatun, M.S., and Kurata, H. (2020). i4mC-Mouse: Improved identification of DNA N4-methylcytosine sites in the mouse genome using multiple encoding schemes. Comput. Struct. Biotechnol. J. *18*, 906–912.

24. Hasan, M.M., Manavalan, B., Shoombuatong, W., Khatun, M.S., and Kurata, H. (2020). i6mA-Fuse: improved and robust prediction of DNA 6 mA sites in the Rosaceae genome by fusing multiple feature representation. Plant Mol. Biol. *103*, 225–234.

25. Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics *22*, 1658–1659.

26. Manavalan, B., Basith, S., Shin, T.H., Wei, L., and Lee, G. (2019). Meta-4mCpred: A Sequence-Based Meta-Predictor for Accurate DNA 4mC Site Prediction Using Effective Feature Representation. Mol. Ther. Nucleic Acids *16*, 733–744.

27. Aoki, G., and Sakakibara, Y. (2018). Convolutional neural networks for classification of alignments of non-coding RNA sequences. Bioinformatics *34*, i237–i244.

28. Ren, F., Yang, C., Qiu, Q., Zeng, N., Cai, C., Hou, C., and Zou, Q. (2019). Exploiting Discriminative Regions of Brain Slices Based on 2D CNNs for Alzheimer's Disease Classification. IEEE Access *7*, 181423–181433.

29. Liu, K., and Chen, W. (2020). iMRM: a platform for simultaneously identifying multiple kinds of RNA modifications. Bioinformatics *36*, 3336–3342.

30. Manavalan, B., Basith, S., Shin, T.H., Wei, L., and Lee, G. (2019). AtbPpred: A Robust Sequence-Based Prediction of Anti-Tubercular Peptides Using Extremely Randomized Trees. Comput. Struct. Biotechnol. J. *17*, 972–981.

31. Hsieh, F., and Turnbull, B.W. (1996). Non- & semi-parametric estimation of the receiver operating characteristics (ROC) curve. Ann. Stat. *24*, 25–40.