Evolutionary Bioinformatics

m6A-CAPred: domain-characteristics enabled machine learning of cancer-associated N6-methyladenosine (m6A) sites

Journal:	*Evolutionary Bioinformatics*
Manuscript ID	EVB-24-0102
Manuscript Type:	Original Research Article
Date Submitted by the Author:	31-Oct-2024
Complete List of Authors:	Chen, Ze; The Affiliated Hospital of Xuzhou Medical University, Urology Liu, Yu; Nanjing University of Chinese Medicine Cao, Sheng; The Affiliated Hospital of Xuzhou Medical University, Urology Huang, Jia; Xi'an Jiaotong-Liverpool University Wang, Xuan; Xi'an Jiaotong-Liverpool University Zhong, Wei; The Affiliated Hospital of Xuzhou Medical University, Urology Xiao, Yong; The Affiliated Hospital of Xuzhou Medical University, Urology
Keywords:	N6-methyladenosine, Machine-learning, Epitranscriptomic, Cancer-associated, SVM
Abstract:	N6-methyladenosine (m6A) is the most prevalent post-transcriptional modification in eukaryotic cells, playing a crucial role in regulating various biological processes. Dysregulation of m6A status is implicated in multiple human diseases, including cancer. Several prediction frameworks have been proposed for high-accuracy identification of putative m6A sites; however, none have tar-geted direct prediction of cancer-associated (or pro-cancer) m6A residues at the base-resolution level. Here, we report m6A-CAPred, a computational tool for predicting pro-cancer m6A sites learned from a comprehensive dataset of experimentally validated m6A sites. Our findings indi-cate that sequence information alone achieves limited performance. However, by leveraging domain-related knowledge (genome-derived features), m6A-CAPred successfully captures dis-tinct domain characteristics between potentially pro-cancer m6A modifications and normal ones, with an average AUROC of 0.885 tested on an independent dataset. Leveraging the power of machine learning, we then performed transcriptome-wide prediction for large-scale screening of potentially pro-cancer m6A sites. Somatic variants derived from 33 types of TCGA cancer projects were extracted for additional validation, and the results showed that SNP density clearly differ-entiated the predicted pro-cancer and normal m6A sites. The m6A-CAPred web server is freely accessible at: www.rnamd.org/m6A-CAPred.
Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.	

[Genes] Manuscript ID_ genes-3297438 - Declined for Publication.eml

genes MDPI

Article

m6A-CAPred: domain-characteristics enabled machine learning of cancer-associated N6-methyladenosine (m^6A) sites

Zeyu Chen[1,#], Yuqi Liu[2,#], Sheng Cao[1,#], Jiaming Huang[3], Xuan Wang[3], Wei Zhong[4,*] and Yongshuang Xiao[1,*]

[1] Department of Urology, The Affiliated Hospital of Xuzhou Medical University, Xuzhou Medical University, Xuzhou, 221006, China
[2] Department of Public Health, School of Medicine, Nanjing University of Chinese Medicine, Nanjing, 210023, China
[3] Department of Biological Sciences, School of Sciences, Xi'an Jiaotong-Liverpool University, Suzhou, 215123, China
[4] Department of Urology, Suining County People's Hospital, Suining, 221200, China
# Contributed equally to this work
* Correspondence: ys.xiao@xzhmu.edu.cn (Y.X.) or snzhongwei@outlook.com (W.Z.)

N6-methyladenosine (m^6A) is the most prevalent post-transcriptional modification in eukaryotic cells, playing a crucial role in regulating various biological processes. Dysregulation of m^6A status is implicated in multiple human diseases, including cancer. Several prediction frameworks have been proposed for high-accuracy identification of putative m^6A sites; however, none have targeted direct prediction of cancer-associated (or pro-cancer) m^6A residues at the base-resolution level. Here, we report m6A-CAPred, a computational tool for predicting pro-cancer m^6A sites learned from a comprehensive dataset of experimentally validated m^6A sites. Our findings indicate that sequence information alone achieves limited performance. However, by leveraging domain-related knowledge (genome-derived features), m6A-CAPred successfully captures distinct domain characteristics between potentially pro-cancer m^6A modifications and normal ones, with an average AUROC of 0.885 tested on an independent dataset. Leveraging the power of machine learning, we then performed transcriptome-wide prediction for large-scale screening of potentially pro-cancer m^6A sites. Somatic variants derived from 33 types of TCGA cancer projects were extracted for additional validation, and the results showed that SNP density clearly differentiated the predicted pro-cancer and normal m^6A sites. The m6A-CAPred web server is freely accessible at: www.rnamd.org/m6A-CAPred.

Keywords: N6-methyladenosine; Machine-learning; Epitranscriptomic

Citation: To be added by editorial staff during production.

Academic Editor:

Received: date
Revised: date
Accepted: date
Published: date

1. Introduction

Exploration of RNA epigenetics has led to the discovery of more than 170 types of RNA modification (1). Among them, N6-methyladenosine (m^6A), the most prevalent marker in messenger RNA and long noncoding RNA (2), has been identified as an abundant and dynamically regulated modification (3). m^6A was first discovered in poly-A RNA in 1974 (4), and since then, it has been identified in various eukaryotic organisms. Over the past decades, multiple studies have underscored the biological significance of m^6A modification in various aspects, including but not limited to, regulation of mRNA translation (5), RNA-protein interaction (6), microRNA (miRNA) processing (7), DNA damage response, and regulation of RNA stability (8). In the context of cancer, m^6A modification has been demonstrated to be a key factor due to its implications in cancer metastasis (9), abnormal mRNA expression levels (10), and immune evasion (11).

Dysregulation of m^{6}A has been reported to play an essential role in tumor proliferation and invasion (12,13), including liver cancer (14,15), lung cancer (16) and breast cancer (17,18). It is apparent that accurately identifying the locations of cancer-associated (or pro-cancer) m^{6}A modifications is crucial for the study and understanding of the downstream effects of RNA modification in biology.

The first whole transcriptome m^{6}A profiling approach, m^{6}A-seq (MeRIP-seq), was developed in 2012 (19,20) to locate m^{6}A sites. It utilizes antibody-based enrichment of the m^{6}A signal to detect and identify regions containing m^{6}A modification, with a resolution of approximately 150 nucleotides. Recent advancements have also been made in integrating ultraviolet cross-linking, enzymatic activity, and domain fusion to achieve high resolution or even base-resolution m^{6}A detection. These techniques include miCLIP/m^{6}A-CLIP-seq (21,22), m^{6}A-REF-seq (23), and DART-seq (24). Nevertheless, these alternative approaches require more complex experimental procedures compared to m^{6}A-seq and have thus far been utilized in fewer biological contexts.

To date, a number of epitranscriptome databases (1,25-30) and computational approaches (31-38) have been developed for large-scale collection and accurate identification of RNA modifications. For m^{6}A methylation, pioneer studies such as SRAMP (17) and iRNA toolkits (39-41) were developed based on combination of sequence-derived information. WHISTLE (42) was a high-accuracy predictor that firstly integrating domain knowledges (genomic features) into m^{6}A prediction framework. And most recently, the deep learning-based approaches were also proved to be another powerful way for m^{6}A prediction (43-45). These works together have greatly facilitated the in silico identification of modified residues. However, these computational models only report whether a nucleotide is modified or not, without differentiating the potentially functional such as pro-cancer m^{6}A sites.

Here, we present m6A-CAPred, a computational framework for accurately classifying potentially pro-cancer and normal m^{6}A modification sites at the base-resolution level. By learning the domain characteristics of m^{6}A modification revealed from a large array of cancer and normal tissues contexts, m6A-CAPred achieved an average AUROC of 0.894 tested on independent datasets. Based on the proposed model, we then conducted a large-scale prediction on ~430,000 experimentally validated m^{6}A sites to identify potentially cancer-associated m^{6}A residues. Independent validation test showed that SNP density can be clearly differentiated between the predicted pro-cancer and normal m^{6}A sites. To share our findings, we developed a user-friendly web interface for the proposed framework, which comprises the following major components: (i) a database of 111,937 high-confidence m^{6}A sites annotated with cancer context labels, which can be extracted for further analysis or model development, and (ii) a web server for high-accuracy prediction of potentially pro-cancer m^{6}A sites from user-provided data. The m6A-CAPred is freely accessible at: www.rnamd.org/m6A-CAPred.

2. Materials and Methods

2.1 Benchmark dataset

m6A-CAPred was proposed to predict the pro-cancer m^{6}A methylation sites at base-resolution level. The positive dataset (P) and negative dataset (N) were all high-confidence experimentally validated m^{6}A sites collected from m^{6}A-Atlas database (with record time > 2, a total of 111,937 sites) (46). The dataset P (pro-cancer m^{6}A sites) and N (normal m^{6}A sites) were further classified by checking whether they localized in m^{6}A-enriched regions from 25 cancer cell lines and 23 normal tissue samples collected from m6A-TSHub (47) (**Supplementary Sheet S1**). It may be worth noting that, the m^{6}A sites that did not occur in either cancer or normal conditions were identified as background noise and were consequently excluded from further experiments.

Specifically, the difference ratio (DR) was calculated to represent the observed difference of a m^{6}A site between cancer and normal conditions:

$$\mathrm{DR} = \frac{P_{Cancer} - P_{Normal}}{2} \quad (1)$$

Where P_{Cancer} represents the percentage of occurrence of an m^{6}A site in the cancer cell lines, while P_{Normal} represents that of normal-condition samples. The DR ranges from -0.5 to 0.5, with 0.5 indicating the most cancer-associated m^{6}A sites and -0.5 indicating m^{6}A sites totally observed in normal samples. A DR of zero represents no difference between cancer and normal contexts. **Figure 1** shows the overall distribution of DRs of all high-confidence m^{6}A sites, with a majority of m^{6}A sites showing no significant differences between cancer and normal conditions.

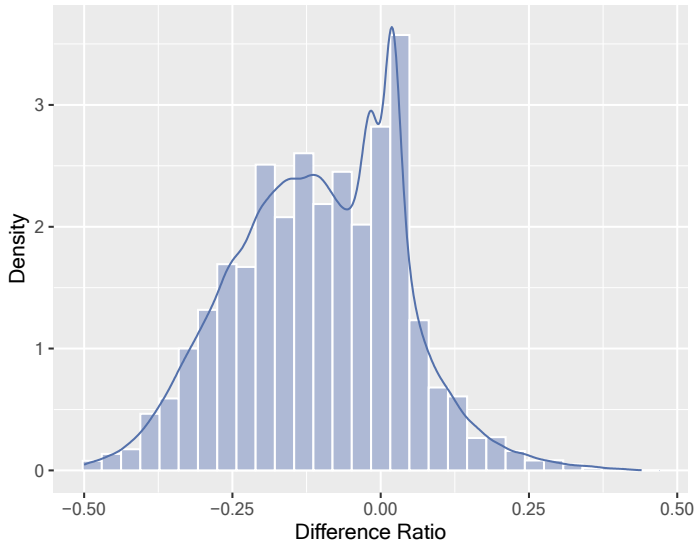


Figure 1. The difference ratio (DR) was calculated to represent the observed difference of a m^{6}A site between cancer and normal conditions. Only a small number of m^{6}A sites show differences under these two contexts and can be further classified into positive and negative training datasets.

Based on the DR value, the positive (P) and negative (N) datasets for model training and testing were selected by identifying the m^{6}A residues most associated with cancer and normal conditions using the Two-tailed test. The pro-cancer m^{6}A sites (P) were defined as the top 2.5% of right-sided (cancer-associated) m^{6}A sites with a DR > 0.165. Conversely, the normal m6A sites (N) were identified as the top 2.5% of left-sided (normal) m^{6}A sites with a DR < -0.38. Specifically, a limited number of base-resolution m^{6}A sites were selected as positive (cancer-associated, 2,660 sites) and negative (normal, 2,959 sites) datasets. The negative sites were randomly chosen to maintain a 1:1 P-to-N ratio. For performance evaluation, 80% of the dataset was randomly selected as training data, while the remaining 20% was used for independent testing.

To test the SNP density around the predicted pro-cancer m^{6}A sites, we collected a total of 2,264,915 cancer somatic variants from 33 different human cancer types in the Cancer Genome Atlas (TCGA) database (version v35) (48). The detailed information of the SNP datasets can be found in **Supplementary Sheet S2**.

2.2 Sequence-derived feature

In our study, five different sequence encoding methods were employed to compose three combinations, including Nucleotide Chemical Property (NCP), position-specific nucleotide propensity (PSNP), Nucleotide Density (ND), Electron-ion interaction potential (EIIP) and pseudo-EIIP (PseEIIP).

The first encoding method NCP categorizes nucleotides into three groups based on their distinct structural chemical properties (49). Primarily, the number of rings in the nucleotides is considered: adenosine and guanosine contain two rings, while cytidine and uridine consist of only one ring. Secondly, adenosine and cytidine are characterized by the presence of an amino group, whereas guanosine and uridine feature a keto group. Lastly, cytidine and guanosine demonstrate stronger hydrogen bonding than adenosine and uridine. Utilizing these properties, the i-th nucleotide in sequence S can be encoded as a vector Si = (xi, yi, zi):

$$x_i = \begin{cases} 1 \text{ if } s_i \in \{A, G\} \\ 0 \text{ if } s_i \in \{C, U\} \end{cases}, \; y_i = \begin{cases} 1 \text{ if } s_i \in \{A, C\} \\ 0 \text{ if } s_i \in \{G, U\} \end{cases}, \; z_i = \begin{cases} 1 \text{ if } s_i \in \{A, U\} \\ 0 \text{ if } s_i \in \{C, G\} \end{cases} \quad (2)$$

PSNP refers to the variation in the frequency of nucleotides at specific positions in RNA sequences between positive and negative datasets. By calculating the frequency of the appearances of A, C, G, and T at the i-position respectively, two matrices with 4×41 dimensions, namely Z_{plus} and Z_{minus}, were obtained. Z_{plus} was derived from the sequence of all positive data, while Z_{minus} was derived from the sequence of all negative data. In this context, the Position-Specific Nucleotide Propensity (PSNP) matrices were denoted as Z_{PSNP}:

$$Z_{PSNP} = Z_{plus} - Z_{minus} \quad (3)$$

ND represents the distribution and cumulative frequency of nucleotides at each position. The density of the ith nucleotide is calculated as the number of nucleotides of the same type appearing before the $(i + 1)^{th}$ position, divided by i. For the sequence 'AGAUUCA', the density of 'A' is 1 (1/1), 0.67 (2/3), and 0.43 (3/7) at the 1st, 3rd, and 7th positions, respectively.

The EIIP values of nucleotides was originally proposed by Nair et al in 2006 (50). Specifically, each nucleotide is encoded as a numeric value that represents its electron-ion interaction potential (**Supplementary Table S1**). Additionally, the pseudo-EIIP (PseEIIP) was calculated by multiplying the sum of the numeric value of tri-nucleotides by their frequency in a given sequence.

In the following section, we will explore which combination of encoding strategies yields the best performance for model development.

2.3 Genome-derived feature

Genome-derived feature guided by domain characteristics have been encoded as an effective feature type that contributing to the performance of prediction models in classifying modified or unmodified RNA residues (51). In our work, we try to capture the distinct patterns between pro-cancer and normal m^6A modification sites. Specifically, 54 domain (genomic) features were extracted for both pro-cancer and normal base-resolution m^6A sites. These genomic properties including dummy variables (1: overlapped; 0: no overlapped) indicating overlapped regions (such as CDS, 5'UTR), counting of adjacent input site and neighboring A, region length, conservation score (PhastCons (52) and fitCons (53)), RNA secondary structures predicted by RNAfold package (54), and distance to regions' 5'/3' ends. The 'TxDb.Hsapiens.UCSC.hg38.knownGene' annotation file was used to extract the corresponding human genomic regions. Please refer to **Supplementary Table S2** for more details about the genomic features considered in the m6A-CAPred model.

2.4 Machine-learning approach used for model construction

Support Vector Machine (SVM) has been widely used in previous prediction model with satisfactory performance (55-57). We used the R language interface of LIBSVM (58) to build the final m6A-CAPred model, the radial basis function was set as kernel, with other parameters using the default setting.

2.5 Performance evaluation

For performance evaluation, we applied the following evaluation metrics. In general, Receiver Operating Characteristic (ROC) curve (sensitivity against 1-specificity) and the area under the ROC curve (AUROC) were used as the primary performance evaluation metrics. In addition, we also calculated sensitivity (Sn), specificity (Sp), Matthew's Correlation Coefficient (MCC), and overall accuracy (ACC) as additional indicators for evaluating the reliability of the model. A 5-fold cross-validation was applied on training datasets, while the testing datasets was used for independent testing. Only the m^{6}A sites that do not include as training step were selected for independent testing purpose.

$$Sn = \frac{TP}{TP + FN} \tag{4}$$

$$Sp = \frac{TN}{TN + FP} \tag{5}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \tag{6}$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

Among them, TP represents true positive, while TN represents true negative; FP stands for the number of false positive, and FN stands for the number of false negative.

2.6 Website construction

m6A-CAPred web interfaces were constructed by using HyperText Markup Language (HTML), Hypertext Preprocessor (PHP), and Cascading Style Sheets (CSS). All metadata was stored using MySQL tables. To present statistical diagrams, EChars was exploited.

3. Results

3.1. Different sequence encoding approaches used for prediction of pro-cancer m^{6}A sites

To try to capture the distinct patterns of pro-cancer m^{6}A sites, we applied different sequence-based feature extracting approaches for model development and tested their performances. Firstly, we tried to explore whether there are significant differences in the primary RNA sequences between pro-cancer and normal m^{6}A sites. The 41nt RNA sequence centered on each m^{6}A site was extracted and encoded using combined sequence-based approaches. We considered a total of three combination: NCP + PSNP, NCP + ND + EIIP, and EIIP + PseEIIP. The Support Vector Machine (SVM) was applied to represent traditional machine learning framework. The results (**Table 1**) showed that sequence-based information only achieved very limited performances, with the best performance achieved by combination of EIIP and PseEIIP (AUROC of 0.577), suggesting that sequence information alone cannot effectively classify the pro-cancer m^{6}A residues from the normal ones.

Table 1. Performance evaluation using different sequence-based encoding approaches

Encoding approach	Independent test				
	Sn	**Sp**	**ACC**	**MCC**	**AUROC**
NCP + PSNP	58.3%	49.1%	53.7%	0.074	0.550
NCP + ND + EIIP	52.8%	56.0%	54.4%	0.088	0.566
EIIP + PseEIIP	54.1%	55.5%	54.8%	0.096	0.577

Note: we randomly selected 80% of dataset as training dataset and the performance of predictors were evaluated by the rest of 20% of dataset as independent testing data, only data not included in training step was selected for independent testing purpose.

3.2 Classifying pro-cancer m^{6}A sites using genome-derived information

We next tried to train the classifier by adding genome-derived information, with the integrated model combining sequence-based and 54 additional genomic knowledges (**Supplementary Table S2**). We found that genomic features greatly enhanced the sequence-based model, improving by 30.8% to 31.9% (**Table 2**). Specifically, the integrated model (genomic + EIIP + PseEIIP) achieved the best prediction performance with an AUROC of 0.885, followed by genomic + NCP + ND + EIIP (AUROC of 0.876) and genomic + NCP + PSNP (AUROC of 0.869), tested on independent datasets. Our results suggested that domain knowledges may be the key factor to capture the distinct patterns between pro-cancer and normal m^{6}A sites, indicating the reliability of the proposed m6A-CAPredmodel.

Table 2. Performance evaluation using integrated encoding methods

Method	5-fold cross validation					Independent test				
	Sn	**Sp**	**ACC**	**MCC**	**AUROC**	**Sn**	**Sp**	**ACC**	**MCC**	**AUROC**
Integrated model 1*	78.4%	78.5%	78.4%	0.569	0.872	79.1%	79.7%	79.4%	0.588	0.869
Integrated model 2*	78.8%	78.3%	78.6%	0.571	0.869	78.9%	80.1%	79.5%	0.590	0.876
Integrated model 3*	80.2%	80.7%	80.4%	0.608	0.884	80.8%	79.3%	80.1%	0.602	0.885

Note: Integrated model 1*: NCP + PSNP + genomic feature; Integrated model 2*: NCP + ND + EIIP + genomic feature; Integrated model 3*: EIIP + PseEIIP + genomic feature

3.3 SNP density analysis clearly differentiated the predicted pro-cancer and normal m^{6}A sites

Leveraging the proposed machine learning-powered classifier, we then performed a large-scale prediction on a total of 427,586 experimentally validated m^{6}A sites at base-resolution level (46). We applied different cut-off values (0.3 to 0.9) for classifying the potentially pro-cancer m^{6}A residues and calculated the SNP density around pro-cancer and normal m^{6}A sites, respectively (**Table 3**). Specifically, the cancer-related somatic variants were extracted from 33 types of TCGA cancer projects, and the SNP density was calculated within a ±2 bp flanking window of each base-resolution m^{6}A site, with a higher

density value indicating a stronger association with cancer through the disruption of these m^{6}A methylation sites. We found that the SNP density of cancer somatic variants around the predicted pro-cancer m^{6}A sites was significantly higher than that of the normal m^{6}A group across all cutoff values ranging from 0.3 to 0.9 (**Table 3**). These results suggest that the m^{6}A sites classified into the pro-cancer group using m6A-CAPred are generally more associated with cancer, thereby demonstrating the effectiveness of our newly proposed m6A-CAPred framework.

Table 3. The SNP density test of TCGA somatic variants around different m^{6}A groups

Mutation type	Cut-off	# of predicted pro-cancer m^{6}A sites	# of predicted normal m^{6}A sites	SNPs around pro-cancer m^{6}A sites (within ± 2 bp motif)	SNPs around normal m^{6}A sites (within ± 2 bp motif)	*P*-Value
TCGA somatic variant	0.3	169,680	257,906	17,366 (10.23%)	25,890 (10.04%)	*
	0.4	143,216	284,370	15,174 (11.60%)	28,080 (9.87%)	***
	0.5	121,623	305,963	133,35 (10.96%)	199,09 (9.78%)	***
	0.6	103,077	324,509	116,44 (11.30%)	31,603 (9.74%)	***
	0.7	84,787	342,799	9,668 (11.40%)	33,588 (9.80%)	***
	0.8	64,245	363,341	7,257 (11.30%)	35,992 (9.91%)	***
	0.9	36,637	390,949	4,114 (11.23%)	39,148 (10.01%)	***

Note: * stands for *P*-value < 0.05; ** stands for *P*-value < 0.01; and *** stands for *P*-value < 0.001.

3.4 Web interface

We developed an online platform to share our findings and facilitate access to the newly proposed model (**Figure 2**). The online resource comprises two major components: i) a database containing 111,937 experimentally validated m^{6}A sites, annotated with cancer and normal context labels. Users can filter the database by difference ratio, and the returned results present detailed information for each base-resolution m^{6}A site, including chromosome position, experimental sources, profiling technique, gene symbol, gene type, Ensembl ID, and cancer/normal context labels. ii) Users can upload their query m^{6}A sites with genome coordinates to the online web server; the returned results indicate whether the predicted m^{6}A sites can be classified into pro-cancer or normal groups. All results can be downloaded freely.

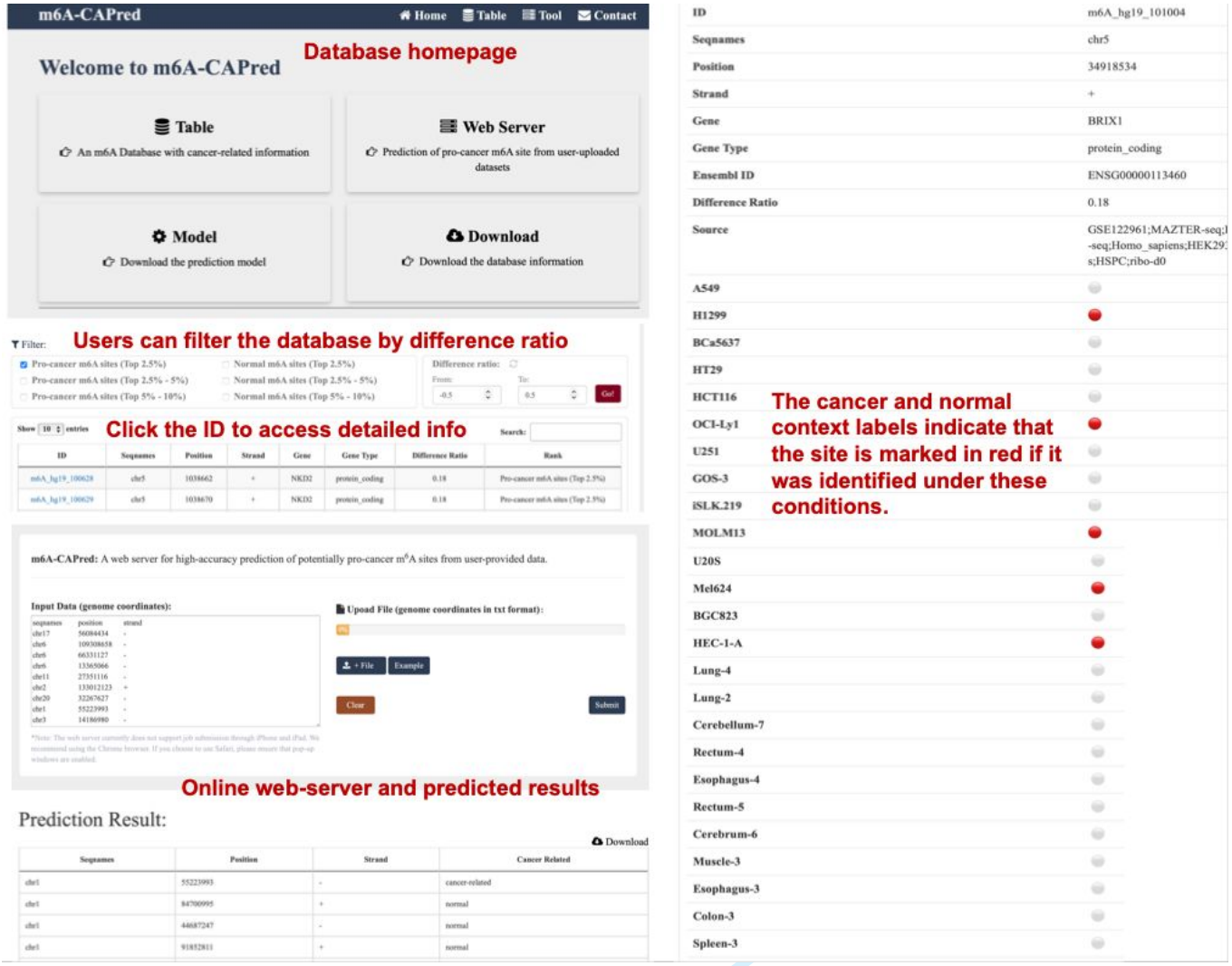


Figure 2. m6A-CAPred online resources. The online resource consists of two components: i) a database featuring 111,937 experimentally validated m^{6}A sites, annotated with difference ratio and context labels. ii) Users can also upload their query m^{6}A sites with genome coordinates to the online web server, which returns results indicating whether the predicted m^{6}A sites are classified as pro-cancer or normal.

4. Discussion

To date, the regulatory roles and disease/cancer associations of N6-methyladenosine (m^{6}A) have been largely elucidated. Despite the development of numerous bioinformatic tools aimed at facilitating m^{6}A prediction, none have specifically targeted the accurate prediction of cancer-associated m^{6}A sites. To address this gap, we developed a predictive framework to distinguish potentially pro-cancer m^{6}A sites from normal ones. Our findings demonstrate that genome-derived information significantly enhances the performance of traditional sequence-based models. The m6A-CAPred web server is freely accessible, providing a valuable resource for researchers interested in m^{6}A modifications related to various cancer types. By accurately identifying pro-cancer m^{6}A sites, m6A-CAPred contributes to a more comprehensive understanding of m^{6}A modification's role in cancer development, which may aid in identifying potential therapeutic targets. Further investigation is essential to fully elucidate the mechanisms underlying m^{6}A modification dysregulation in cancer and to explore the clinical implications of our findings.

Supplementary Materials: Supplemental Information is available online.

Author Contributions: Conceptualization, Y.X. and W.Z.; methodology, Z.C., J.H. and S.C.; software, X.W.; validation, Y.L.; writing—original draft preparation, Z.C. and J.H.; writing—review and editing, Y.X.; supervision, Y.X.; funding acquisition, Y.X. All authors have read and agreed to the published version of the manuscript.

Funding: Project of Key research and development Plan of Xuzhou Science and Technology Bureau (NO.2023103070)

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

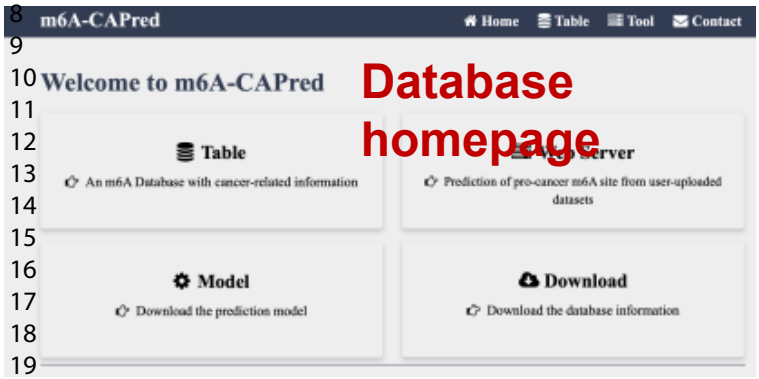
References

1. Boccaletto, P., Machnicka, M.A., Purta, E., Piatkowski, P., Baginski, B., Wirecki, T.K., de Crecy-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**, D303-D307.
2. Meyer, K.D. and Jaffrey, S.R. (2017) Rethinking m(6)A Readers, Writers, and Erasers. *Annu Rev Cell Dev Biol*, **33**, 319-342.
3. Niu, Y., Zhao, X., Wu, Y.S., Li, M.M., Wang, X.J. and Yang, Y.G. (2013) N6-methyl-adenosine (m6A) in RNA: an old modification with a novel epigenetic function. *Genomics Proteomics Bioinformatics*, **11**, 8-17.
4. Desrosiers, R., Friderici, K. and Rottman, F. (1974) Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. *Proc Natl Acad Sci U S A*, **71**, 3971-3975.
5. Meyer, K.D. and Jaffrey, S.R. (2014) The dynamic epitranscriptome: N6-methyladenosine and gene expression control. *Nat Rev Mol Cell Biol*, **15**, 313-326.
6. Liu, N., Dai, Q., Zheng, G., He, C., Parisien, M. and Pan, T. (2015) N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature*, **518**, 560-564.
7. Alarcon, C.R., Lee, H., Goodarzi, H., Halberg, N. and Tavazoie, S.F. (2015) N6-methyladenosine marks primary microRNAs for processing. *Nature*, **519**, 482-485.
8. Xiang, Y., Laurent, B., Hsu, C.H., Nachtergaele, S., Lu, Z., Sheng, W., Xu, C., Chen, H., Ouyang, J., Wang, S. *et al.* (2017) RNA m(6)A methylation regulates the ultraviolet-induced DNA damage response. *Nature*, **543**, 573-576.
9. Yin, H., Zhang, X., Yang, P., Zhang, X., Peng, Y., Li, D., Yu, Y., Wu, Y., Wang, Y., Zhang, J. *et al.* (2021) RNA m6A methylation orchestrates cancer growth and metastasis via macrophage reprogramming. *Nat Commun*, **12**, 1394.
10. Wu, Y., Chen, X., Bao, W., Hong, X., Li, C., Lu, J., Zhang, D. and Zhu, A. (2022) Effect of Humantenine on mRNA m6A Modification and Expression in Human Colon Cancer Cell Line HCT116. *Genes (Basel)*, **13**.
11. Liu, Z., Wang, T., She, Y., Wu, K., Gu, S., Li, L., Dong, C., Chen, C. and Zhou, Y. (2021) N(6)-methyladenosine-modified circIGF2BP3 inhibits CD8(+) T-cell responses to facilitate tumor immune evasion by promoting the deubiquitination of PD-L1 in non-small cell lung cancer. *Mol Cancer*, **20**, 105.
12. Sun, T., Wu, R. and Ming, L. (2019) The role of m6A RNA methylation in cancer. *Biomed Pharmacother*, **112**, 108613.
13. An, Y. and Duan, H. (2022) The role of m6A RNA methylation in cancer metabolism. *Mol Cancer*, **21**, 14.
14. Xu, Q., Ren, N., Ren, L., Yang, Y., Pan, J. and Shang, H. (2024) RNA m6A methylation regulators in liver cancer. *Cancer Cell Int*, **24**, 1.

15. Wang, J., Yu, H., Dong, W., Zhang, C., Hu, M., Ma, W., Jiang, X., Li, H., Yang, P. and Xiang, D. (2023) N6-Methyladenosine-Mediated Up-Regulation of FZD10 Regulates Liver Cancer Stem Cells' Properties and Lenvatinib Resistance Through WNT/beta-Catenin and Hippo Signaling Pathways. *Gastroenterology*, **164**, 990-1005.
16. Zhang, Q. and Xu, K. (2023) The role of regulators of RNA m(6)A methylation in lung cancer. *Genes Dis*, **10**, 495-504.
17. Zhou, Y., Zeng, P., Li, Y.H., Zhang, Z. and Cui, Q. (2016) SRAMP: prediction of mammalian N6-methyladenosine (m6A) sites based on sequence-derived features. *Nucleic Acids Res*, **44**, e91.
18. Zhang, C., Zhi, W.I., Lu, H., Samanta, D., Chen, I., Gabrielson, E. and Semenza, G.L. (2016) Hypoxia-inducible factors regulate pluripotency factor expression by ZNF217- and ALKBH5-mediated modulation of RNA methylation in breast cancer cells. *Oncotarget*, **7**, 64527-64542.
19. 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.
20. 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.
21. Linder, B., Grozhik, A.V., Olarerin-George, A.O., Meydan, C., Mason, C.E. and Jaffrey, S.R. (2015) Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat Methods*, **12**, 767-772.
22. Ke, S., Pandya-Jones, A., Saito, Y., Fak, J.J., Vagbo, C.B., Geula, S., Hanna, J.H., Black, D.L., Darnell, J.E., Jr. and Darnell, R.B. (2017) m(6)A mRNA modifications are deposited in nascent pre-mRNA and are not required for splicing but do specify cytoplasmic turnover. *Genes Dev*, **31**, 990-1006.
23. 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.
24. Meyer, K.D. (2019) DART-seq: an antibody-free method for global m(6)A detection. *Nat Methods*, **16**, 1275-1280.
25. Song, B., Chen, K., Tang, Y., Wei, Z., Su, J., de Magalhaes, J.P., Rigden, D.J. and Meng, J. (2021) ConsRM: collection and large-scale prediction of the evolutionarily conserved RNA methylation sites, with implications for the functional epitranscriptome. *Brief Bioinform*, **22**.
26. Xuan, J.J., Sun, W.J., Lin, P.H., Zhou, K.R., Liu, S., Zheng, L.L., Qu, L.H. and Yang, J.H. (2018) RMBase v2.0: deciphering the map of RNA modifications from epitranscriptome sequencing data. *Nucleic Acids Res*, **46**, D327-D334.
27. Bao, X., Zhang, Y., Li, H., Teng, Y., Ma, L., Chen, Z., Luo, X., Zheng, J., Zhao, A., Ren, J. *et al.* (2023) RM2Target: a comprehensive database for targets of writers, erasers and readers of RNA modifications. *Nucleic Acids Res*, **51**, D269-D279.
28. Luo, X., Li, H., Liang, J., Zhao, Q., Xie, Y., Ren, J. and Zuo, Z. (2021) RMVar: an updated database of functional variants involved in RNA modifications. *Nucleic Acids Res*, **49**, D1405-D1412.
29. Wang, X., Zhang, Y., Chen, K., Liang, Z., Ma, J., Xia, R., de Magalhaes, J.P., Rigden, D.J., Meng, J. and Song, B. (2024) m7GHub V2.0: an updated database for decoding the N7-methylguanosine (m7G) epitranscriptome. *Nucleic Acids Res*, **52**, D203-D212.
30. Song, B., Wang, X., Liang, Z., Ma, J., Huang, D., Wang, Y., de Magalhaes, J.P., Rigden, D.J., Meng, J., Liu, G. *et al.* (2023) RMDisease V2.0: an updated database of genetic variants that affect RNA modifications with disease and trait implication. *Nucleic Acids Res*, **51**, D1388-D1396.
31. Qiu, W.R., Jiang, S.Y., Xu, Z.C., Xiao, X. and Chou, K.C. (2017) iRNAm5C-PseDNC: identifying RNA 5-methylcytosine sites by incorporating physical-chemical properties into pseudo dinucleotide composition. *Oncotarget*, **8**, 41178-41188.
32. Chen, W., Song, X., Lv, H. and Lin, H. (2019) iRNA-m2G: Identifying N(2)-methylguanosine Sites Based on Sequence-Derived Information. *Mol Ther Nucleic Acids*, **18**, 253-258.

33. Zhai, J., Song, J., Cheng, Q., Tang, Y. and Ma, C. (2018) PEA: an integrated R toolkit for plant epitranscriptome analysis. *Bioinformatics*, **34**, 3747-3749.
34. Jiang, J., Song, B., Tang, Y., Chen, K., Wei, Z. and Meng, J. (2020) m5UPred: A Web Server for the Prediction of RNA 5-Methyluridine Sites from Sequences. *Mol Ther Nucleic Acids*, **22**, 742-747.
35. Körtel, N., Rücklé, C., Zhou, Y., Busch, A., Hoch-Kraft, P., Sutandy, F.X.R., Haase, J., Pradhan, M., Musheev, M., Ostareck, D. *et al.* (2021) Deep and accurate detection of m6A RNA modifications using miCLIP2 and m6Aboost machine learning. *Nucleic Acids Research*.
36. Xiong, Y., He, X., Zhao, D., Tian, T., Hong, L., Jiang, T. and Zeng, J. (2021) Modeling multi-species RNA modification through multi-task curriculum learning. *Nucleic Acids Research*.
37. Huang, J., Wang, X., Xia, R., Yang, D., Liu, J., Lv, Q., Yu, X., Meng, J., Chen, K., Song, B. *et al.* (2024) Domain-knowledge enabled ensemble learning of 5-formylcytosine (f5C) modification sites. *Computational and Structural Biotechnology Journal*, **23**, 3175-3185.
38. Liang, Z., Zhang, L., Chen, H., Huang, D. and Song, B. (2022) m6A-Maize: Weakly supervised prediction of m(6)A-carrying transcripts and m(6)A-affecting mutations in maize (Zea mays). *Methods*, **203**, 226-232.
39. Chen, W., Ding, H., Zhou, X., Lin, H. and Chou, K.C. (2018) iRNA(m6A)-PseDNC: Identifying N(6)-methyladenosine sites using pseudo dinucleotide composition. *Anal Biochem*, **561-562**, 59-65.
40. Liu, K. and Chen, W. (2020) iMRM: a platform for simultaneously identifying multiple kinds of RNA modifications. *Bioinformatics*, **36**, 3336-3342.
41. Chen, W., Feng, P., Ding, H., Lin, H. and Chou, K.C. (2015) iRNA-Methyl: Identifying N(6)-methyladenosine sites using pseudo nucleotide composition. *Anal Biochem*, **490**, 26-33.
42. Liu, L., Song, B., Chen, K., Zhang, Y., de Magalhaes, J.P., Rigden, D.J., Lei, X. and Wei, Z. (2022) WHISTLE server: A high-accuracy genomic coordinate-based machine learning platform for RNA modification prediction. *Methods*, **203**, 378-382.
43. 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.
44. Chen, Z., Zhao, P., Li, F., Wang, Y., Smith, A.I., Webb, G.I., Akutsu, T., Baggag, A., Bensmail, H. and Song, J. (2020) Comprehensive review and assessment of computational methods for predicting RNA post-transcriptional modification sites from RNA sequences. *Brief Bioinform*, **21**, 1676-1696.
45. Huang, D., Song, B., Wei, J., Su, J., Coenen, F. and Meng, J. (2021) Weakly supervised learning of RNA modifications from low-resolution epitranscriptome data. *Bioinformatics*, **37**, i222-i230.
46. Tang, Y., Chen, K., Song, B., Ma, J., Wu, X., Xu, Q., Wei, Z., Su, J., Liu, G., Rong, R. *et al.* (2021) m6A-Atlas: a comprehensive knowledgebase for unraveling the N6-methyladenosine (m6A) epitranscriptome. *Nucleic Acids Res*, **49**, D134-D143.
47. Song, B., Huang, D., Zhang, Y., Wei, Z., Su, J., Pedro de Magalhaes, J., Rigden, D.J., Meng, J. and Chen, K. (2023) m6A-TSHub: Unveiling the Context-specific m(6)A Methylation and m(6)A-affecting Mutations in 23 Human Tissues. *Genomics Proteomics Bioinformatics*, **21**, 678-694.
48. Tomczak, K., Czerwińska, P. and Wiznerowicz, M. (2015) The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemporary oncology*, **19**, A68.
49. Yang, H., Lv, H., Ding, H., Chen, W. and Lin, H. (2018) iRNA-2OM: A Sequence-Based Predictor for Identifying 2'-O-Methylation Sites in Homo sapiens. *J Comput Biol*, **25**, 1266-1277.
50. Nair, A.S. and Sreenadhan, S.P. (2006) A coding measure scheme employing electron-ion interaction pseudopotential (EIIP). *Bioinformation*, **1**, 197-202.

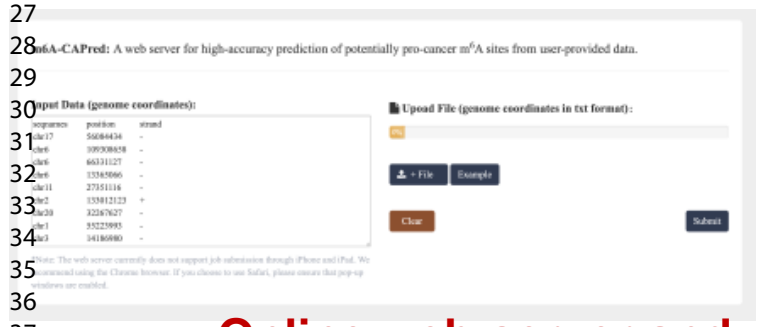
51. Song, B., Tang, Y., Chen, K., Wei, Z., Rong, R., Lu, Z., Su, J., de Magalhaes, J.P., Rigden, D.J. and Meng, J. (2020) m7GHub: deciphering the location, regulation and pathogenesis of internal mRNA N7-methylguanosine (m7G) sites in human. *Bioinformatics*, **36**, 3528-3536.
52. Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S. *et al.* (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res*, **15**, 1034-1050.
53. Gulko, B., Hubisz, M.J., Gronau, I. and Siepel, A. (2015) A method for calculating probabilities of fitness consequences for point mutations across the human genome. *Nat Genet*, **47**, 276-283.
54. Lorenz, R., Bernhart, S.H., Honer Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F. and Hofacker, I.L. (2011) ViennaRNA Package 2.0. *Algorithms Mol Biol*, **6**, 26.
55. Liu, H., Yue, D., Chen, Y., Gao, S.J. and Huang, Y. (2010) Improving performance of mammalian microRNA target prediction. *BMC bioinformatics*, **11**, 476.
56. Chen, W., Tang, H. and Lin, H. (2017) MethyRNA: a web server for identification of N6-methyladenosine sites. *Journal of Biomolecular Structure and Dynamics*, **35**, 683-687.
57. Wong, Y.-H., Lee, T.-Y., Liang, H.-K., Huang, C.-M., Wang, T.-Y., Yang, Y.-H., Chu, C.-H., Huang, H.-D., Ko, M.-T. and Hwang, J.-K. (2007) KinasePhos 2.0: a web server for identifying protein kinase-specific phosphorylation sites based on sequences and coupling patterns. *Nucleic Acids Research*, **35**, W588-W594.
58. Chang, C.-C. and Lin, C.-J. (2011) LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, **2**, 1-27.

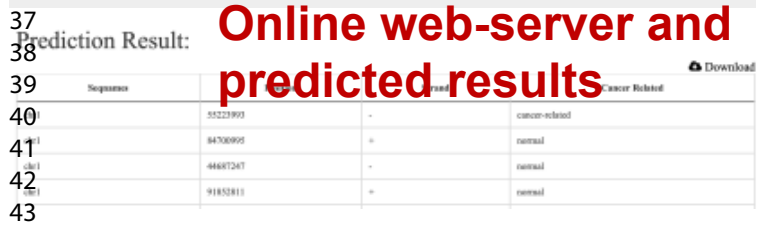


Database homepage

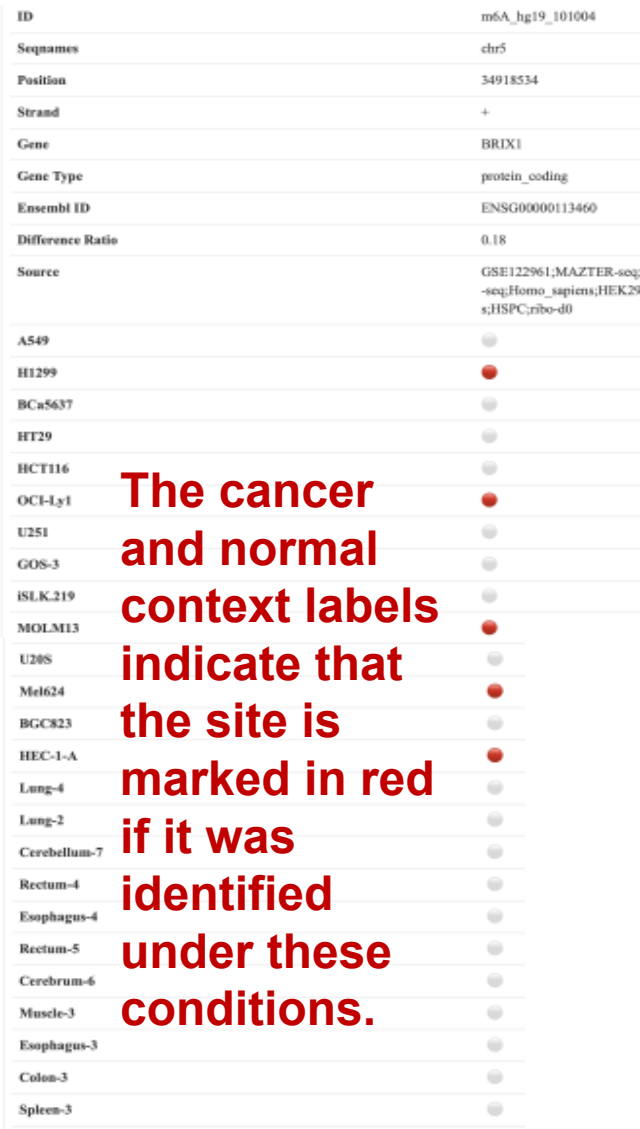
Users can filter the database by difference ratio

Click the ID to access detailed info





Online web-server and predicted results



The cancer and normal context labels indicate that the site is marked in red if it was identified under these conditions.

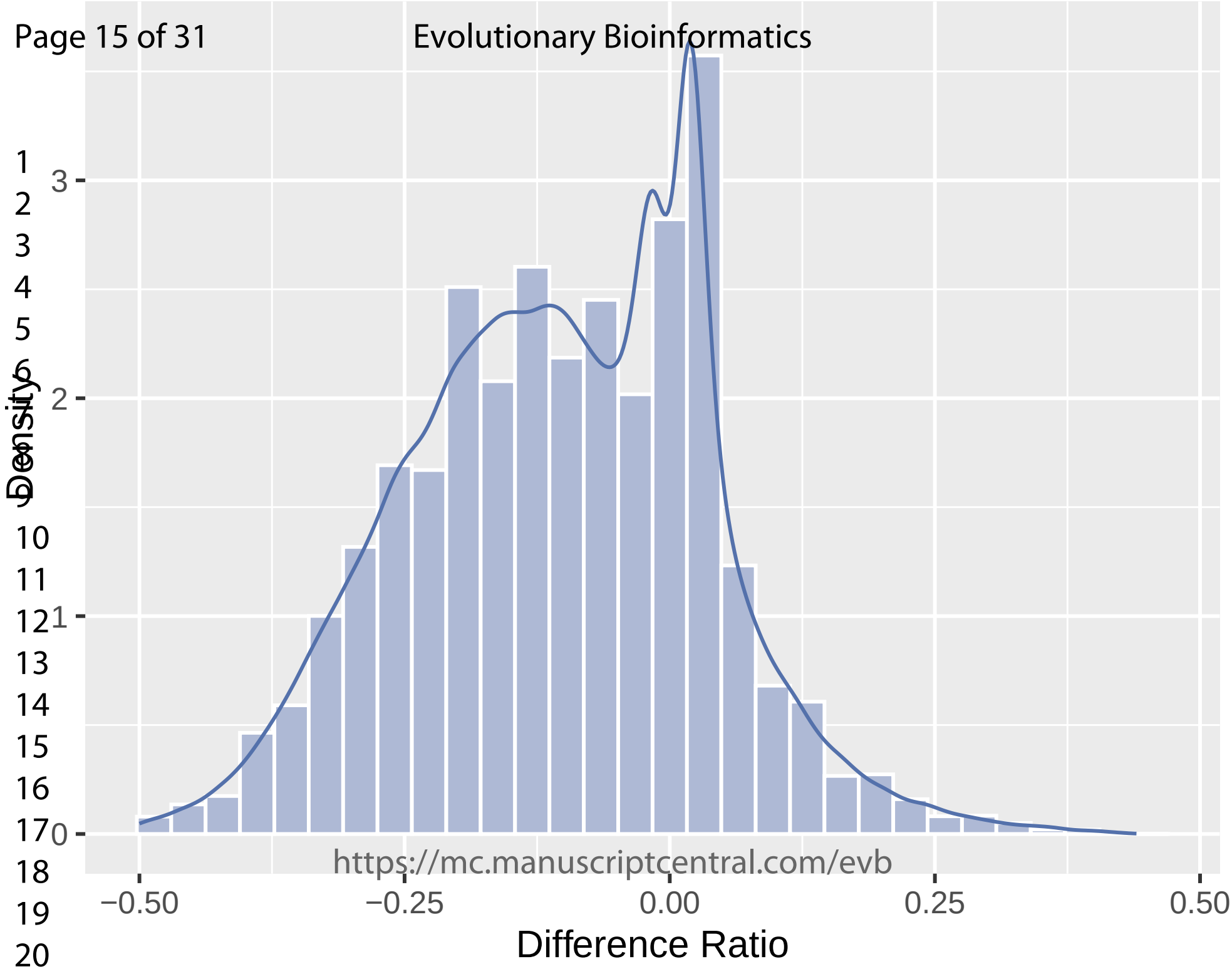

Density
3
2
1
0
−0.50
−0.25
0.00
0.25
0.50
Difference Ratio

Table S1. m6A profiling samples in cancer and normal conditions using m6A-MeRIP-seq

Experiment ID	Paper name	Journal	Pubmed ID	CRA/GSE
1	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
2	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
3	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
4	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
5	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
6	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
7	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
8	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
9	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
10	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
11	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
12	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
13	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
14	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
15	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
16	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
17	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
18	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
19	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
20	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
21	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
22	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
23	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
24	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
25	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
26	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
27	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
28	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
29	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
30	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
31	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
32	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
33	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
34	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
35	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
36	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
37	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
38	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
39	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
40	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
41	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
42	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
43	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
44	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
45	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315

46	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
47	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
48	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
49	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
50	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
51	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
52	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
53	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
54	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
55	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
56	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
57	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
58	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
59	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
60	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
61	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
62	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
63	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
64	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
65	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
66	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
67	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
68	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
69	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
70	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
71	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
72	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
73	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
74	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
75	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
76	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
77	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
78	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
79	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
80	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
81	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
82	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
83	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
84	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
85	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
86	Landscape and Regulation of m6A	Mol Cell	31676230	CRA001315
87	Altered m6A Modification of Spe	Mol Cell	31810760	GSE138730
88	Altered m6A Modification of Spe	Mol Cell	31810760	GSE138730
89	Altered m6A Modification of Spe	Mol Cell	31810760	GSE138730
90	Altered m6A Modification of Spe	Mol Cell	31810760	GSE138730
91	Altered m6A Modification of Spe	Mol Cell	31810760	GSE138730
92	Altered m6A Modification of Spe	Mol Cell	31810760	GSE138730

93	Altered m6A Modification of Spe	Mol Cell	31810760	GSE138730
94	Altered m6A Modification of Spe	Mol Cell	31810760	GSE138730
95	Topology of the human and mous	Nature	22575960	GSE37002
96	Topology of the human and mous	Nature	22575960	GSE37002
97	Topology of the human and mous	Nature	22575960	GSE37002
98	Topology of the human and mous	Nature	22575960	GSE37002
99	Topology of the human and mous	Nature	22575960	GSE37002
100	Topology of the human and mous	Nature	22575960	GSE37002
101	Perturbation of m6A writers revea	Cell report	24981863	GSE55572
102	Perturbation of m6A writers revea	Cell report	24981863	GSE55572
103	The m(6)A Methyltransferase ME	Mol Cell	27117702	GSE76367
104	The m(6)A Methyltransferase ME	Mol Cell	27117702	GSE76367
105	The m(6)A Methyltransferase ME	Mol Cell	27117702	GSE76367
106	The m(6)A Methyltransferase ME	Mol Cell	27117702	GSE76367
107	FTO Plays an Oncogenic Role in	Cancer Cell	28017614	GSE76414
108	FTO Plays an Oncogenic Role in	Cancer Cell	28017614	GSE76414
109	FTO Plays an Oncogenic Role in	Cancer Cell	28017614	GSE76414
110	FTO Plays an Oncogenic Role in	Cancer Cell	28017614	GSE76414
111	m6A RNA Methylation Regulates	Cell Rep	28297667	GSE94808
112	m6A RNA Methylation Regulates	Cell Rep	28297667	GSE94808
113	Promoter-bound METTL3 mainta	Nature	29186125	GSE94613
114	Promoter-bound METTL3 mainta	Nature	29186125	GSE94613
115	Promoter-bound METTL3 mainta	Nature	29186125	GSE94613
116	Promoter-bound METTL3 mainta	Nature	29186125	GSE94613
117	Promoter-bound METTL3 mainta	Nature	29186125	GSE94613
118	Promoter-bound METTL3 mainta	Nature	29186125	GSE94613
119	Promoter-bound METTL3 mainta	Nature	29186125	GSE94613
120	Promoter-bound METTL3 mainta	Nature	29186125	GSE94613
121	R-2HG Exhibits Anti-tumor Activ	Cell	29249359	GSE87190
122	R-2HG Exhibits Anti-tumor Activ	Cell	29249359	GSE87190
123	R-2HG Exhibits Anti-tumor Activ	Cell	29249359	GSE87190
124	R-2HG Exhibits Anti-tumor Activ	Cell	29249359	GSE87190
125	Recognition of RNA N6-methyla	Nat Cell Biol	29476152	GSE90642
126	Recognition of RNA N6-methyla	Nat Cell Biol	29476152	GSE90642
127	Recognition of RNA N6-methyla	Nat Cell Biol	29476152	GSE90642
128	Recognition of RNA N6-methyla	Nat Cell Biol	29476152	GSE90642
129	N6-Methyladenosine methyltrans	Nat Chem Biol	30531910	GSE102336
130	N6-Methyladenosine methyltrans	Nat Chem Biol	30531910	GSE102336
131	N6-Methyladenosine methyltrans	Nat Chem Biol	30531910	GSE102336
132	N6-Methyladenosine methyltrans	Nat Chem Biol	30531910	GSE102336
133	Histone H3 trimethylation at lysin	Nature	30867593	GSE110320
134	Histone H3 trimethylation at lysin	Nature	30867593	GSE110320
135	Histone H3 trimethylation at lysin	Nature	30867593	GSE110320
136	Histone H3 trimethylation at lysin	Nature	30867593	GSE110320
137	Histone H3 trimethylation at lysin	Nature	30867593	GSE110320
138	Histone H3 trimethylation at lysin	Nature	30867593	GSE110320
139	Limits in the detection of m6A ch	biorxiv	32313079	GSE130892

140	Limits in the detection of m6A ch	biorxiv	32313079	GSE130892
141	Limits in the detection of m6A ch	biorxiv	32313079	GSE130892
142	Limits in the detection of m6A ch	biorxiv	32313079	GSE130892
143	Limits in the detection of m6A ch	biorxiv	32313079	GSE130892
144	Limits in the detection of m6A ch	biorxiv	32313079	GSE130892
145	PCIF1 catalyzes m6Am mRNA m	Mol Cell	31279659	GSE122803
146	PCIF1 catalyzes m6Am mRNA m	Mol Cell	31279659	GSE122803
147	PCIF1 catalyzes m6Am mRNA m	Mol Cell	31279659	GSE122803
148	PCIF1 catalyzes m6Am mRNA m	Mol Cell	31279659	GSE122803
149	PCIF1 catalyzes m6Am mRNA m	Mol Cell	31279659	GSE122803
150	PCIF1 catalyzes m6Am mRNA m	Mol Cell	31279659	GSE122803
151	m6A mRNA methylation regulate	Nat Metabolism		GSE132306
152	m6A mRNA methylation regulate	Nat Metabolism		GSE132306
153	m6A mRNA methylation regulate	Nat Metabolism		GSE132306
154	m6A mRNA methylation regulate	Nat Metabolism		GSE132306
155	m6A mRNA methylation regulate	Nat Metabolism		GSE132306
156	m6A mRNA methylation regulate	Nat Metabolism		GSE132306
157	m6A mRNA methylation regulate	Nat Cell Biol	30154548	GSE93911
158	m6A mRNA methylation regulate	Nat Cell Biol	30154548	GSE93911
159	N6-methyladenosine modification	PLoS Pathog	29659627	GSE104621
160	N6-methyladenosine modification	PLoS Pathog	29659627	GSE104621
161	N6-methyladenosine modification	PLoS Pathog	29659627	GSE104621
162	N6-methyladenosine modification	PLoS Pathog	29659627	GSE104621
163	Long noncoding RNA GAS5 inhi	Mol Cancer	31619268	GSE129716
164	Long noncoding RNA GAS5 inhi	Mol Cancer	31619268	GSE129716
165	Long noncoding RNA GAS5 inhi	Mol Cancer	31619268	GSE129716
166	Long noncoding RNA GAS5 inhi	Mol Cancer	31619268	GSE129716
167	N 6-Methylation of Adenosine of	Cancer Res	30967398	GSE119963
168	N 6-Methylation of Adenosine of	Cancer Res	30967398	GSE119963
169	METTL3-mediated N6-methylade	Mol Cancer	31607270	GSE133132
170	METTL3-mediated N6-methylade	Mol Cancer	31607270	GSE133132
171	YTHDF2 reduction fuels inflamm	Mol Cancer	31735169	GSE120860
172	YTHDF2 reduction fuels inflamm	Mol Cancer	31735169	GSE120860
173	YTHDF2 reduction fuels inflamm	Mol Cancer	31735169	GSE120860
174	YTHDF2 reduction fuels inflamm	Mol Cancer	31735169	GSE120860
175	ALKBH5 suppresses malignancy	Mol Cancer	32772918	GSE149510
176	ALKBH5 suppresses malignancy	Mol Cancer	32772918	GSE149510
177	ALKBH5 suppresses malignancy	Mol Cancer	32772918	GSE149510
178	ALKBH5 suppresses malignancy	Mol Cancer	32772918	GSE149510
179	Leukemogenic Chromatin Alterat	Cell Stem Cell	32402251	GSE128575
180	Leukemogenic Chromatin Alterat	Cell Stem Cell	32402251	GSE128575
181	Leukemogenic Chromatin Alterat	Cell Stem Cell	32402251	GSE128575
182	Leukemogenic Chromatin Alterat	Cell Stem Cell	32402251	GSE128575
183	The m6A methyltransferase MET	*oncogene*	30659266	PRJNA498900
184	The m6A methyltransferase MET	*oncogene*	30659266	PRJNA498900
185	The m6A methyltransferase MET	*oncogene*	30659266	PRJNA498900
186	The Role of m 6 A/m-RNA Meth	Neuron	30048615	GSE113798

187	The Role of m 6 A/m-RNA Methy	Neuron	30048615	GSE113798
188	The Role of m 6 A/m-RNA Methy	Neuron	30048615	GSE113798
189	The Role of m 6 A/m-RNA Methy	Neuron	30048615	GSE113798
190	The Role of m 6 A/m-RNA Methy	Neuron	30048615	GSE113798
191	The Role of m 6 A/m-RNA Methy	Neuron	30048615	GSE113798
192	m6A mRNA methylation regulate	Nat Metabolism		GSE120024
193	m6A mRNA methylation regulate	Nat Metabolism		GSE120024
194	m6A mRNA methylation regulate	Nat Metabolism		GSE120024
195	m6A mRNA methylation regulate	Nat Metabolism		GSE120024
196	m6A mRNA methylation regulate	Nat Metabolism		GSE120024
197	m6A mRNA methylation regulate	Nat Metabolism		GSE120024
198	m6A mRNA methylation regulate	Nat Metabolism		GSE120024
199	m6A mRNA methylation regulate	Nat Metabolism		GSE120024
200	m6A mRNA methylation regulate	Nat Metabolism		GSE120024
201	m6A mRNA methylation regulate	Nat Metabolism		GSE120024
202	m6A mRNA methylation regulate	Nat Metabolism		GSE120024
203	m6A mRNA methylation regulate	Nat Metabolism		GSE120024
204	m6A mRNA methylation regulate	Nat Metabolism		GSE120024
205	m6A mRNA methylation regulate	Nat Metabolism		GSE120024
206	m 6 A mRNA methylation regulat	Nat Cell Biol	30154548	GSE93911
207	m 6 A mRNA methylation regulat	Nat Cell Biol	30154548	GSE93911
208	m 6 A mRNA methylation regulat	Nat Cell Biol	30154548	GSE93911
209	m 6 A mRNA methylation regulat	Nat Cell Biol	30154548	GSE93911
210	m 6 A mRNA methylation regulat	Nat Cell Biol	30154548	GSE93911
211	m 6 A mRNA methylation regulat	Nat Cell Biol	30154548	GSE93911
212	m 6 A mRNA methylation regulat	Nat Cell Biol	30154548	GSE93911
213	m 6 A mRNA methylation regulat	Nat Cell Biol	30154548	GSE93911
214	RADAR: differential analysis of M	*Genome Biol*	31870409	GSE119168
215	RADAR: differential analysis of M	*Genome Biol*	31870409	GSE119168
216	RADAR: differential analysis of M	*Genome Biol*	31870409	GSE119168
217	RADAR: differential analysis of M	*Genome Biol*	31870409	GSE119168
218	RADAR: differential analysis of M	*Genome Biol*	31870409	GSE119168
219	RADAR: differential analysis of M	*Genome Biol*	31870409	GSE119168
220	RADAR: differential analysis of M	*Genome Biol*	31870409	GSE119168
221	RADAR: differential analysis of M	*Genome Biol*	31870409	GSE119168
222	RADAR: differential analysis of M	*Genome Biol*	31870409	GSE119168
223	RADAR: differential analysis of M	*Genome Biol*	31870409	GSE119168
224	RADAR: differential analysis of M	*Genome Biol*	31870409	GSE119168
225	RADAR: differential analysis of M	*Genome Biol*	31870409	GSE119168
226	RADAR: differential analysis of M	*Genome Biol*	31870409	GSE119168
227	RADAR: differential analysis of M	*Genome Biol*	31870409	GSE119168
228	Dynamic landscape and evolution	*Nucleic Acids Res*	32406913	GSE122744
229	Dynamic landscape and evolution	*Nucleic Acids Res*	32406913	GSE122744
230	Dynamic landscape and evolution	*Nucleic Acids Res*	32406913	GSE122744
231	Dynamic landscape and evolution	*Nucleic Acids Res*	32406913	GSE122744
232	Dynamic landscape and evolution	*Nucleic Acids Res*	32406913	GSE122744
233	Dynamic landscape and evolution	*Nucleic Acids Res*	32406913	GSE122744

CRR/SRR	Cell line and batch
CRR073021	Lung-4-4-Input
CRR073020	Lung-4-4-IP
CRR042297	Lung-4-2-Input
CRR042296	Lung-4-2-IP
CRR073019	Lung-2-4-Input-human
CRR073018	Lung-2-4-IP-human
CRR055534	Lung-2-1-Input
CRR055533	Lung-2-1-IP
CRR073017	Cerebellum-7-4-Input
CRR073016	Cerebellum-7-4-IP
CRR055564	Rectum-4-2-Input
CRR055563	Rectum-4-2-IP
CRR055562	Esophagus-4-2-Input
CRR055561	Esophagus-4-2-IP
CRR055560	Rectum-5-3-Input
CRR055559	Rectum-5-3-IP
CRR055554	Cerebrum-6-3-Input
CRR055553	Cerebrum-6-3-IP
CRR055550	Muscle-3-2-Input
CRR055549	Muscle-3-2-IP
CRR055548	Esophagus-3-2-Input
CRR055547	Esophagus-3-2-IP
CRR055546	Colon-3-2-Input
CRR055545	Colon-3-2-IP
CRR055542	Spleen-3-2-Input
CRR055541	Spleen-3-2-IP
CRR055540	Urinary_bladder-2-1-Input
CRR055539	Urinary_bladder-2-1-IP
CRR055538	Tongue-2-1-Input
CRR055537	Tongue-2-1-IP
CRR055535	Spleen-2-1-Input
CRR055536	Spleen-2-1-IP
CRR055530	Spleen-1-1-Input
CRR055529	Spleen-1-1-IP
CRR055528	Heart-1-1-Input
CRR055527	Heart-1-1-IP
CRR055526	Adipose-1-1-Input
CRR055525	Adipose-1-1-IP
CRR042321	Urinary_bladder-5-3-Input
CRR042320	Urinary_bladder-5-3-IP
CRR042319	Urinary_bladder-4-2-Input
CRR042318	Urinary_bladder-4-2-IP
CRR042317	Trachea-5-3-Input
CRR042316	Trachea-5-3-IP
CRR042315	Thyroid_gland-5-3-Input

CRR042314	Thyroid_gland-5-3-IP
CRR042313	Thyroid_gland-4-2-Input
CRR042312	Thyroid_gland-4-2-IP
CRR042311	Testis-4-2-Input
CRR042310	Testis-4-2-IP
CRR042309	Stomach-5-3-Input
CRR042308	Stomach-5-3-IP
CRR042307	Stomach-4-2-Input
CRR042306	Stomach-4-2-IP
CRR042305	Skin-1-1-Input
CRR042304	Skin-1-1-IP
CRR042303	Skin-4-2-Input
CRR042302	Skin-4-2-IP
CRR042301	Prostate-4-2-Input
CRR042300	Prostate-4-2-IP
CRR042299	Muscle-5-3-Input
CRR042298	Muscle-5-3-IP
CRR042295	Liver-4-2-Input
CRR042294	Liver-4-2-IP
CRR042293	Hypothalamus-5-3-Input
CRR042292	Hypothalamus-5-3-IP
CRR042291	Heart-4-2-Input
CRR042290	Heart-4-2-IP
CRR042287	Cerebrum-5-3-Input
CRR042286	Cerebrum-5-3-IP
CRR042285	Cerebellum-5-3-Input
CRR042284	Cerebellum-5-3-IP
CRR042283	Brainstem-5-3-Input
CRR042282	Brainstem-5-3-IP
CRR042281	Aorta-4-2-Input
CRR042280	Aorta-4-2-IP
CRR042279	Adrenal_gland-1-1-Input
CRR042278	Adrenal_gland-1-1-IP
CRR073004	U251-Input
CRR073005	U251-IP
CRR072998	HT29-Input
CRR072999	HT29-IP
CRR072990	GOS-3-1-Input
CRR072991	GOS-3-1-IP
CRR072992	GOS-3-2-Input
CRR072993	GOS-3-2-IP
SRR10259052	Huh7
SRR10259053	Huh7
SRR10259054	Huh7
SRR10259055	Huh7
SRR10259056	Huh7
SRR10259057	Huh7

SRR10259058	Huh7
SRR10259059	Huh7
SRR456551	HepG2
SRR456552	HepG2
SRR456553	HepG2
SRR456555	HepG2
SRR456556	HepG2
SRR456557	HepG2
SRR1182633	A549
SRR1182634	A549
SRR3057328	A549
SRR3057327	A549
SRR3057334	H1299
SRR3057333	H1299
SRR3066066	MONO-MAC-6
SRR3066067	MONO-MAC-6
SRR3066068	MONO-MAC-6
SRR3066069	MONO-MAC-6
SRR5248992	PBT003
SRR5248996	PBT003
SRR5239086	MOLM13
SRR5239087	MOLM13
SRR5239088	MOLM13
SRR5239089	MOLM13
SRR5239098	MOLM13
SRR5239099	MOLM13
SRR5239100	MOLM13
SRR5239101	MOLM13
SRR4288705	NOMO-1
SRR4288706	NOMO-1
SRR4288709	MA9.3ITD
SRR4288710	MA9.3ITD
SRR5060388	HepG2
SRR5861462	HepG2
SRR5060389	HepG2
SRR5861463	HepG2
SRR5907119	HepG2
SRR5907120	HepG2
SRR5907121	HepG2
SRR5907122	HepG2
SRR6686554	HepG2
SRR6686555	HepG2
SRR6686556	HepG2
SRR6686557	HepG2
SRR6686558	HepG2
SRR6686559	HepG2
SRR9029568	OCI-Ly1

SRR9029569	OCI-Ly1
SRR9029570	OCI-Ly1
SRR9029571	OCI-Ly1
SRR9029572	OCI-Ly1
SRR9029573	OCI-Ly1
SRR8234036	Mel624
SRR8234037	Mel624
SRR8234038	Mel624
SRR8234039	Mel624
SRR8234040	Mel624
SRR8234041	Mel624
SRR9211553	EndoC-bH1
SRR9211554	EndoC-bH1
SRR9211555	EndoC-bH1
SRR9211562	EndoC-bH1
SRR9211563	EndoC-bH1
SRR9211564	EndoC-bH1
SRR5194801	HEC-1-A
SRR5194802	HEC-1-A
SRR6132499	iSLK.219
SRR6132500	iSLK.219
SRR6132503	iSLK.219
SRR6132504	iSLK.219
SRR8889196	HCT116
SRR8889197	HCT116
SRR8889198	HCT116
SRR8889199	HCT116
SRR7829546	PEO1
SRR7829548	PEO1
SRR9336432	BGC823
SRR9336434	BGC823
SRR7965996	SMMC7721
SRR7965997	SMMC7721
SRR7965998	SMMC7721
SRR7965999	SMMC7721
SRR11626649	HCCLM3
SRR11626650	HCCLM3
SRR11626651	HCCLM3
SRR11626652	HCCLM3
SRR8755805	THP1
SRR8755806	THP1
SRR8755811	THP1
SRR8755812	THP1
SRR8118687	BCa5637
SRR8118688	BCa5637
SRR8118689	BCa5637
SRR7075085	B_lymphocyte

SRR7075089	B_lymphocyte
SRR7075093	B_lymphocyte
SRR7075097	B_lymphocyte
SRR7075101	B_lymphocyte
SRR7075105	B_lymphocyte
SRR7851591	Islets
SRR7851592	Islets
SRR7851593	Islets
SRR7851594	Islets
SRR7851595	Islets
SRR7851596	Islets
SRR7851597	Islets
SRR7851606	Islets
SRR7851607	Islets
SRR7851608	Islets
SRR7851609	Islets
SRR7851610	Islets
SRR7851611	Islets
SRR7851612	Islets
SRR5194775	Endometrial
SRR5194776	Endometrial
SRR5194779	Endometrial
SRR5194780	Endometrial
SRR5194783	Endometrial
SRR5194784	Endometrial
SRR5194787	Endometrial
SRR5194788	Endometrial
SRR7763577	Ovary
SRR7763578	Ovary
SRR7763579	Ovary
SRR7763580	Ovary
SRR7763581	Ovary
SRR7763582	Ovary
SRR7763583	Ovary
SRR7763564	Ovary
SRR7763565	Ovary
SRR7763566	Ovary
SRR7763567	Ovary
SRR7763568	Ovary
SRR7763569	Ovary
SRR7763570	Ovary
SRR8209856	Kidney
SRR8209857	Kidney
SRR8209858	Kidney
SRR8209859	Kidney
SRR8209860	Kidney
SRR8209861	Kidney

Table S2. Sources of genetic variants

Database	Species	Tumor Type	SNP number
TCGA (v35)	Human	TCGA-BRCA	82,280
TCGA (v35)	Human	TCGA-THCA	5,129
TCGA (v35)	Human	TCGA-UCEC	561,179
TCGA (v35)	Human	TCGA-DLBC	6,309
TCGA (v35)	Human	TCGA-COAD	186,914
TCGA (v35)	Human	TCGA-CESC	66,316
TCGA (v35)	Human	TCGA-BLCA	112,098
TCGA (v35)	Human	TCGA-CHOL	3,321
TCGA (v35)	Human	TCGA-ESCA	27,404
TCGA (v35)	Human	TCGA-ACC	7,657
TCGA (v35)	Human	TCGA-KICH	2,171
TCGA (v35)	Human	TCGA-HNSC	83,690
TCGA (v35)	Human	TCGA-LIHC	40,094
TCGA (v35)	Human	TCGA-MESO	2,510
TCGA (v35)	Human	TCGA-LAML	3,559
TCGA (v35)	Human	TCGA-KIRP	16,530
TCGA (v35)	Human	TCGA-KIRC	18,495
TCGA (v35)	Human	TCGA-GBM	47,187
TCGA (v35)	Human	TCGA-LGG	30,129
TCGA (v35)	Human	TCGA-SARC	16,651
TCGA (v35)	Human	TCGA-PCPG	1,801
TCGA (v35)	Human	TCGA-READ	51,570
TCGA (v35)	Human	TCGA-PAAD	24,214
TCGA (v35)	Human	TCGA-LUAD	171,843
TCGA (v35)	Human	TCGA-PRAD	23,207
TCGA (v35)	Human	TCGA-OV	32,673
TCGA (v35)	Human	TCGA-LUSC	150,852
TCGA (v35)	Human	TCGA-TGCT	2,465
TCGA (v35)	Human	TCGA-THYM	2,041
TCGA (v35)	Human	TCGA-UVM	1,338
TCGA (v35)	Human	TCGA-SKCM	323,031
TCGA (v35)	Human	TCGA-UCS	8,291
TCGA (v35)	Human	TCGA-STAD	151,966
Total: 33 cancer types			2,264,915

Information

Somatic mutations in Breast Invasive Carcinoma
Somatic mutations in Thyroid carcinoma
Somatic mutations in Uterine Corpus Endometrial Carcinoma
Somatic mutations in Lymphoid Neoplasm Diffuse Large B-cell Lymphoma
Somatic mutations in Colon adenocarcinoma
Somatic mutations in Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma
Somatic mutations in Bladder Urothelial Carcinoma
Somatic mutations in Cholangiocarcinoma
Somatic mutations in Esophageal Carcinoma
Somatic mutations in Adrenocortical carcinoma
Somatic mutations in Kidney Chromophobe
Somatic mutations in Head and Neck Squamous Cell Carcinoma
Somatic mutations in Liver hepatocellular carcinoma
Somatic mutations in Mesothelioma
Somatic mutations in Acute Myeloid Leukemia
Somatic mutations in Kidney renal papillary cell carcinoma
Somatic mutations in Kidney renal clear cell carcinoma
Somatic mutations in Glioblastoma multiforme
Somatic mutations in Brain Lower Grade Glioma
Somatic mutations in Sarcoma
Somatic mutations in Pheochromocytoma and Paraganglioma
Somatic mutations in Rectum adenocarcinoma
Somatic mutations in Pancreatic adenocarcinoma
Somatic mutations in Lung adenocarcinoma
Somatic mutations in Prostate adenocarcinoma
Somatic mutations in Ovarian serous cystadenocarcinoma
Somatic mutations in Lung squamous cell carcinoma
Somatic mutations in Testicular Germ Cell Tumors
Somatic mutations in Thymoma
Somatic mutations in Uveal Melanoma
Somatic mutations in Skin Cutaneous Melanoma
Somatic mutations in Uterine Carcinosarcoma
Somatic mutations in Stomach adenocarcinoma

Table S1. EIIP Value for Each Nucleotide

Nucleotide	EIIP
A	0.1260
U	0.1335
G	0.0806
C	0.1340

Table S2 Domain knowledges considered in m6A-CAPred

ID	Name	Description	Note
1	UTR5	5' UTR	Dummy variables indicating whether the site is overlapped to the topological region on the major RNA transcript.
2	UTR3	3' UTR	
3	cds	Coding sequence	
4	Stop_codons	stop codons flanked by 100bp	
5	Start_codons	start codons flanked by 100bp	
6	TSS	downstream 100bp of TSS	
7	TSS_A	downstream 100bp of TSS on A	
8	exon_stop	exons containing stop codons	
9	alternative_exon	alternative exons	
10	constitutive_exon	constitutive exons	
11	internal_exon	Internal exons	
12	long_exon	long exons (exon length >= 400bp)	
13	last_exon	5' last_exon	
16	intron	intron	
17	length_UTR3	3'UTR length	The region length in bp
18	length_UTR5	5'UTR length	
19	length_cds	coding sequence length	
20	length_tx_full	full transcript length	
21	length_gene_full	full gene length	
22	clust_f1000	count of neighboring input site at 1001 bp	Clustering information
23	clust_f100	count of neighboring input site at 101 bp	
24	clust_A_f1000	count of neighboring A within in 2001 nt window	
25	clust_A_f100	count of neighboring A within 201 nt window	
26	dist_nearest_p2000	distance to the closest neighboring input site at 2001 bp	
27	dist_nearest_p200	distance to the closest neighboring input site at 201 bp	
28	PC_1bp	phastCons scores of the nucleotide	Scores related to evolutionary conservation
29	PC_101bp	average phastCons scores within the flanking 101 bp	
30	FC_1bp	fitCons scores of the nucleotide	
31	FC_101bp	average fitCons scores within the flanking 101 bp region	

32	struct_hybridize	predicted RNA hybridized region	RNA secondary structures
33	struct_loop	predicted RNA loop region	
34	sncRNA	sncRNA	Attributes of the genes or transcripts
35	lncRNA	lncRNA	
36	HK_genes	housekeeping genes	
37	miR_targeted_genes	miRNA targeted genes	
38	HNRNPC_eCLIP	eCLIP data of HNRNPC RNA binding sites	
39	TargetScan	predicted miRNA targeted sites by TargetScan	
40	Verified_miRtargets	miRNA targeted sites verified by experiment	
41	METTL3_TREW	overlapped with binding regions of METTL3	
42	METTL14_TREW	overlapped with binding regions of METTL14	
43	WTAP_TREW	overlapped with binding regions of WTAP	
44	METTL16_CLIP	overlapped with binding regions of METTL16	
45	ALKBH5_PARCLIP	overlapped with binding regions of ALKBH5	
46	FTO_CLIP	overlapped with binding regions of FTO	
47	isoform_num	number of isoforms	Genomic properties
48	exon_num	number of exons	
49	GC_cont_genes	GC composition of genes	
50	GC_cont_101bp_abs	GC composition of 101 bp	
51	pos_UTR5	relative position on 5'UTR	Relative position on the region
52	pos_UTR3	relative position on 3'UTR	
53	pos_cds	relative position on coding sequence	
54	pos_exons	relative position on exon	

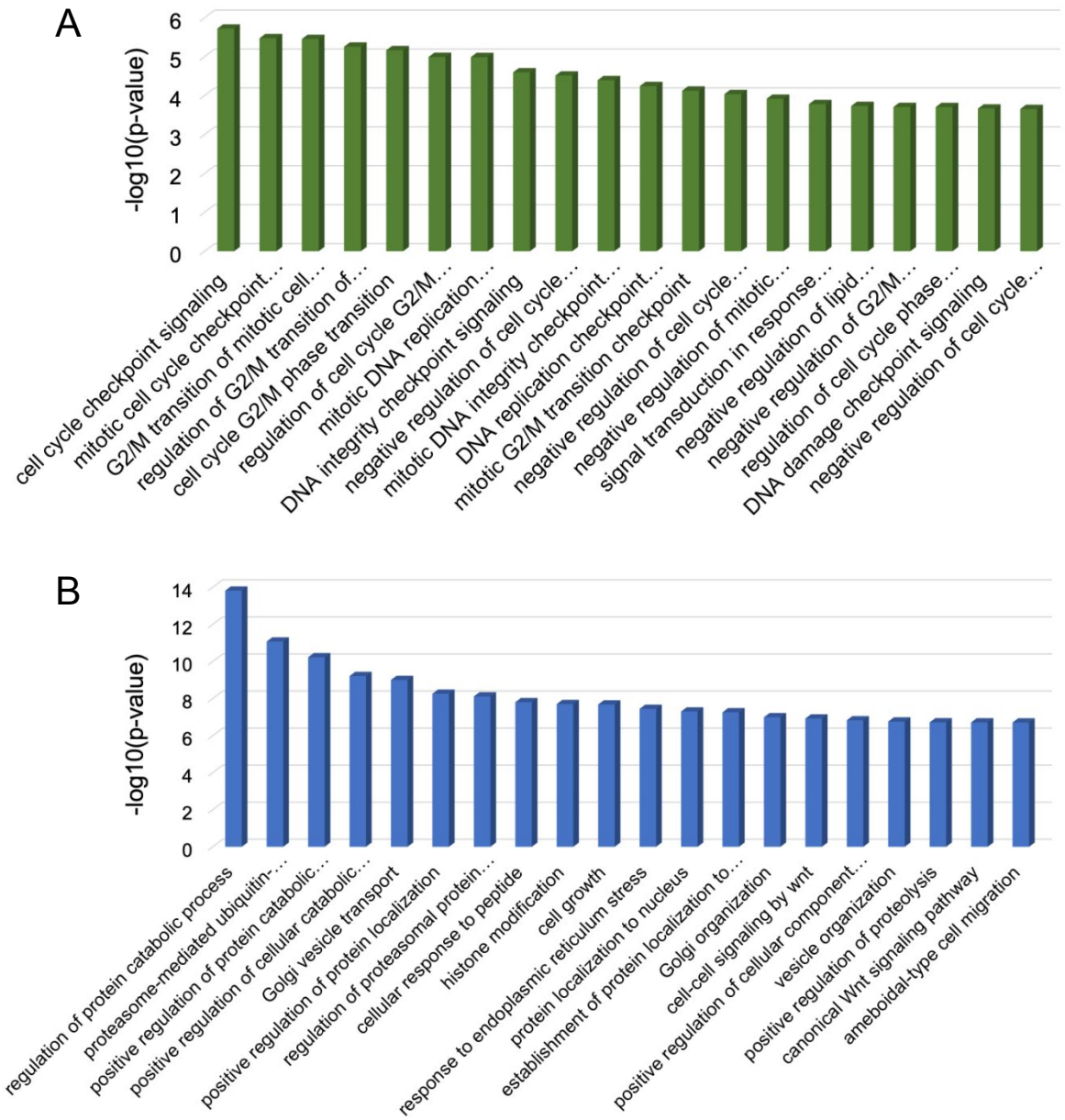


Figure S1. Gene Ontology Enrichment analysis. (A) The top 20 biological processes enriched with pro-cancer m6A sites. **(B)** The top biological processed obtained for anti-cancer group.

[Genes] Manuscript ID: genes-3297438 - Declined for Publication

GO

zoeyrice.wang@mdpi.com 代表 Genes Editorial Office<genes@mdpi.com>

收件人: Yongshuang Xiao; Yuqi Liu

周三 2024/10/23 9:21

抄送: 你; Yuqi Liu; Sheng Cao; Jiaming Huang; Xuan Wang; Bowen Song; Wei Zhong; **+ 另外 2 人**

Dear Dr. Xiao,

Thank you for submitting the following manuscript to Genes:

Manuscript ID: genes-3297438
Type of manuscript: Article
Title: m6A-CAPred: domain-characteristics enabled machine learning of cancer-associated N6-methyladenosine (m6A) sites
Authors: Zeyu Chen, Yuqi Liu, Sheng Cao, Jiaming Huang, Xuan Wang, Bowen Song, Wei Zhong, Yongshuang Xiao *
Received: 21 Oct 2024
E-mails: xyfyczy71@outlook.com, yuqi.liu@njucm.edu.cn, caosheng8099@outlook.com, jiaming.huang21@student.xjtlu.edu.cn, xuan.wang17@student.xjtlu.edu.cn, bowen.song@njucm.edu.cn, snzhongwei@outlook.com, ys.xiao@xzhmu.edu.cn
Bioinformatics
https://www.mdpi.com/journal/genes/sections/bioinformatics
Bioinformatics of RNA Modifications and Epitranscriptome
https://www.mdpi.com/journal/genes/special_issues/9431069KM8
https://susy.mdpi.com/user/manuscripts/review_info/4580fb9c680662ccf9279c42cb9b239d

We regret to inform you that we will not be processing your submission further. Submissions sent for peer-review are selected based on discipline, novelty and general significance, in addition to the usual criteria for publication in scholarly journals. Therefore, our decision does not

591x436mm (57 x 57 DPI)

We regret to inform you that we will not be processing your submission further. Submissions sent for peer-review are selected based on discipline, novelty and general significance, in addition to the usual criteria for publication in scholarly journals. Therefore, our decision does not necessarily reflect the quality of your work.

We wish you every success if you choose to pursue publication elsewhere.

Kind regards,
Genes Editorial Office
genes@mdpi.com

540x290mm (57 x 57 DPI)