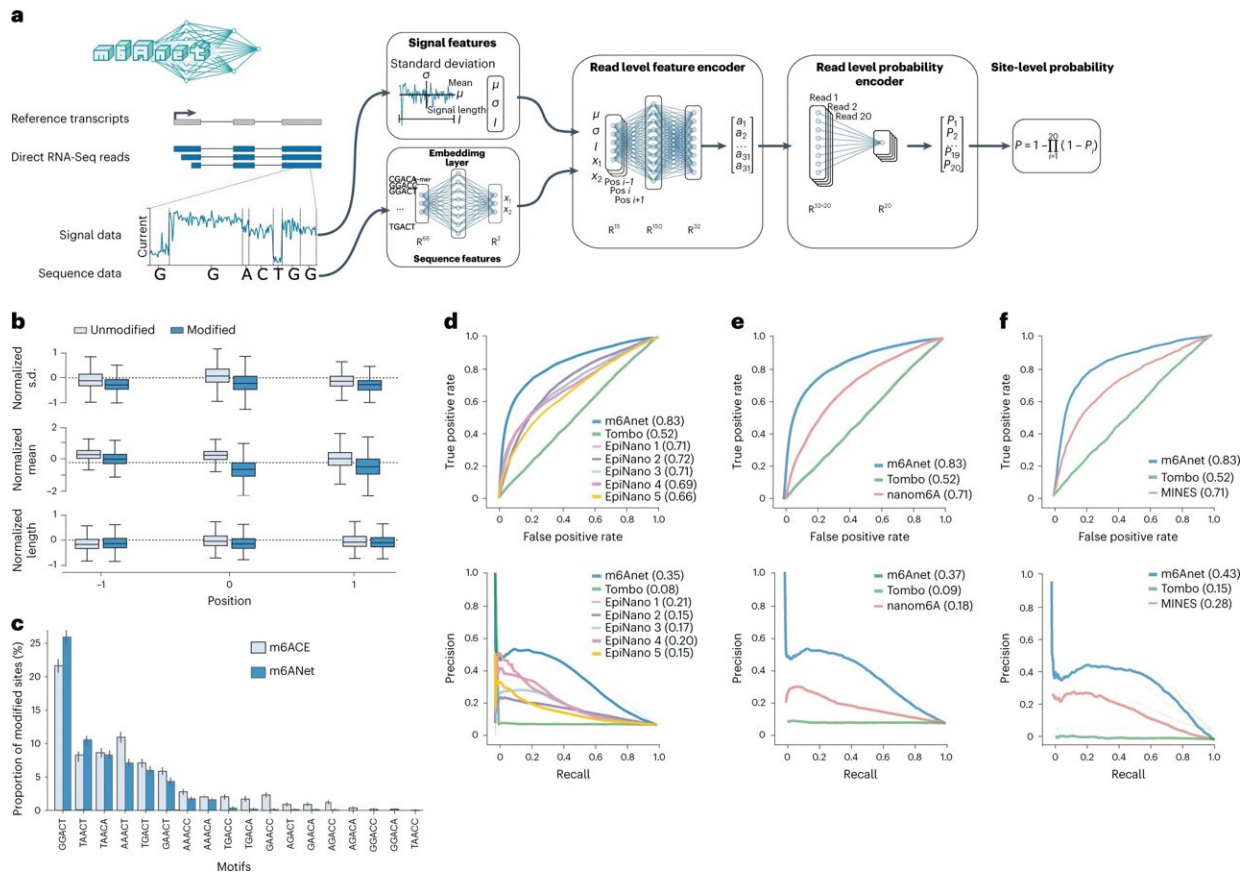


Scientists develop AI-based method to predict RNA modifications

February 9 2023



Schematic of m6Anet and evaluation on detection of m6A in human cell lines. **a**, m6Anet model schematics. **b**, Box plot showing the difference in average features distribution between different m6Anet prediction with $n = 1769$ predicted modified sites and $n = 5031$ predicted unmodified sites for the GGACT fivemer motif. The horizontal lines on the boxes show minima, 25th percentile, median, 75th percentile, and maxima. Points that do not fall within $1.5\times$ of the interquartile range are considered outliers and are not shown on the

plot. **c**, Comparison of the proportion of modified sites predicted as modified by m6Anet and by m6ACE on the DRACH fivemer motifs. The bar plot center represents the proportion of modified sites for each fivemer motif while the error bar represents the estimated 95% confidence interval around the center values with a total of $n = 5,579$ for m6ACE modified positions, $n = 4,784$ for m6Anet-predicted modified positions and $n = 121,853$ for m6ACE unmodified positions and $n = 122,648$ for m6Anet-predicted unmodified positions. **d**, ROC curve (top) and PR curve (bottom) of m6Anet against all five EpiNano models and Tombo. **e**, ROC curve (top) and PR curve (bottom) of m6Anet against nanom6A and Tombo. **f**, ROC curve (top) and PR curve (bottom) of m6Anet against MINES and Tombo. Credit: *Nature Methods* (2022). DOI: 10.1038/s41592-022-01666-1

A team of researchers from the Agency for Science, Technology and Research (A*STAR) and the National University of Singapore (NUS) has developed a software method that accurately predicts chemical modifications of RNA molecules from genomic data. Their method, called m6Anet, was published in *Nature Methods*.

Within the RNA, different types of chemical molecules added to the RNA determine how the RNA molecule functions. However, these RNA changes are often invisible to standard approaches used by scientists to read RNA. Presently, more than 160 RNA modifications have been discovered, of which the most prevalent RNA modification—N6-Methyladenosine (m6A)—is associated with human diseases such as cancer.

In the past, identifying RNA modifications required time-consuming and laborious bench experiments that were not accessible to most laboratories. Furthermore, previous methods failed to detect m6A at the single-molecule resolution, which is critical for understanding the biological mechanisms involving m6A.

The team overcame these limitations by leveraging direct Nanopore RNA sequencing, an [emerging technology](#) that sequences a raw RNA molecule together with its RNA modifications. In this study, they developed m6Anet, a software that trains [deep neural networks](#) with abundant direct Nanopore RNA sequencing data and Multiple-Instance Learning (MIL) approach, to accurately detect the presence of m6A.

"In traditional machine learning, we often have one label for each example we want to classify. For example, each image is either a cat or not a cat, and the algorithm learns to differentiate cat images from other images based on their labels. The issue with detecting m6A is that we have an overwhelming amount of data with unclear labels. Imagine having a large photo album with a cat photo hidden among millions of other photos, and attempting to identify that particular photo without having any labels to base your search upon. Fortunately, this has been studied in machine learning literature before and is known as the MIL problem," explained Christopher Hendra, current Ph.D. student at A*STAR's Genome Institute of Singapore (GIS) and NUS Institute of Data Science, and the first author of the study.

In this study, the team demonstrated that m6Anet can predict the presence of m6A with high accuracy at a single-molecule resolution from a single sample across species.

"Our AI model has only seen data from a human sample, but it is able to accurately identify RNA modifications even in samples from species that the model has not seen before," said Dr. Jonathan Göke, Group Leader of the Laboratory of Computational Transcriptomics at A*STAR's GIS and senior author of the study. "The ability to identify RNA modifications in different biological samples can be used to understand their role in many different applications such as in cancer research or plant genomics."

"It is very satisfying to see how theoretically-grounded and well-studied machine-learning techniques such as the MIL can be leveraged to offer an elegant solution to this challenging problem. Witnessing the software being adopted so rapidly by the scientific community is a reward for our efforts!" said Associate Professor Alexandre Thiery, Department of Statistics and Data Science, NUS Faculty of Science, who co-led the study.

Prof Patrick Tan, Executive Director of A*STAR's GIS, said, "Accurately and efficiently identifying RNA modifications had been a long-standing challenge, and m6Anet helps to address these limitations. To benefit the wider [scientific community](#), this AI method, along with results from the study, have been made public for other scientists to accelerate their research."

More information: Christopher Hendra et al, Detection of m6A from direct RNA sequencing using a multiple instance learning framework, *Nature Methods* (2022). [DOI: 10.1038/s41592-022-01666-1](https://doi.org/10.1038/s41592-022-01666-1)

Source code: github.com/GoekeLab/m6anet

Software documentation: m6anet.readthedocs.io/en/latest/

Provided by Agency for Science, Technology and Research (A*STAR), Singapore

Citation: Scientists develop AI-based method to predict RNA modifications (2023, February 9) retrieved 26 April 2024 from <https://phys.org/news/2023-02-scientists-ai-based-method-rna-modifications.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private

study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.