

Title: A robust computational quest: Discovering potential hits to improve the treatment of pyrazinamide-resistant *Mycobacterium tuberculosis*

Author Names : Md. Shazzad Hossain Shaon, Tasmin Karim, Md. Mamun Ali, Kawsar Ahmed, Francis M. Bui, Li Chen & Mohammad Ali Moni

Abstract: RNA 5-methyluridine (m5U) sites play a significant role in understanding RNA modifications, which influence numerous biological processes such as gene expression and cellular functioning. Consequently, the identification of m5U sites can play a vital role in the integrity, structure, and function of RNA molecules. Therefore, this study introduces GRUpred-m5U, a novel deep learning based framework based on a gated recurrent unit in mature RNA and full transcript RNA datasets. We used three descriptor groups: nucleic acid composition, pseudo nucleic acid composition, and physicochemical properties, which include five feature extraction methods ENAC, Kmer, DPCP, DPCP type 2, and PseDNC. Initially, we aggregated all the feature extraction methods and created a new merged set. Three hybrid models were developed employing deep-learning methods and evaluated through 10-fold cross-validation with seven evaluation metrics. After a comprehensive evaluation, the GRUpred-m5U model outperformed the other applied models, obtaining 98.41% and 96.70% accuracy on the two datasets, respectively. To our knowledge, the proposed model outperformed all the existing state-of-the-art technology. The proposed supervised machine learning model was evaluated using unsupervised machine learning techniques such as principal component analysis (PCA), and it was observed that the proposed method provided a valid performance for identifying m5U. Considering its multi-layered construction, the GRUpred-m5U model has tremendous potential for future applications in the biological industry. The model, which consisted of neurons processing complicated input, excelled at pattern recognition and produced reliable results. Despite its greater size, the model obtained accurate results, essential in detecting m5U.

Keywords: RNA 5-methyluridine, RNA modifications, Physicochemical properties, Deep-learning, Principal component analysis, Transcript RNA

<https://www.nature.com/articles/s41598-024-76148-9#:~:text=The%20current%20study%20has%20proposed,DPCP%20type%201%2C%20and%20Kmer.>