[BIOINFO 2024]

Comprehensive discovery of RNA modification sites in the human transcriptome

Gihyeon Kang[1], Hyeonseo Hwang[1], Hyeonseong Jeon[2,3], Hee Ryung Chang[1], Heejin Choi[1], Nagyeong Yeo[1], Jungmin Lim[1], Jaeung Yun[1,3], Daehyun Baek[1,2,3]

[1]*School of Biological Sciences, Seoul National University*
[2]*Interdisciplinary Program in Bioinformatics, Seoul National University*
[3]*Genome4me Inc.*
**Corresponding author: baek@snu.ac.kr*

RNA modification (RM) is a key regulator of numerous biological processes including gene expression regulation. However, elucidating the functional roles and regulatory mechanisms of RM has been impeded by an absence of accurate, quantitative, and high-resolution detection method. Recently, Nanopore direct RNA sequencing has emerged as a promising solution for identifying RM at a single-base resolution, as modified and unmodified ribonucleotides produce different electric signals during sequencing. However, existing Nanopore-based methods can identify RM only in a small subset of sequence contexts, also lacking the detection accuracy enough to accurately measure the degree of modification for a given site. To address these challenges, we constructed a massive-scale RM library encompassing all possible sequence contexts by ligating random oligonucleotides. By training on this dataset, we developed a deep learning model that identifies m6A sites and measures their degrees of modification with an unprecedented accuracy across any sequence context. Using this model, we discovered an extensive group of non-DRACH and weakly modified m6A sites across the human transcriptome. Our method serves as a scalable framework for transcriptome-level profiling of diverse types of RM, providing a powerful opportunity to comprehensively investigate the human epitranscriptome landscape.