[BIOINFO 2024]

Nanopore Direct Sequencing of RNA-DNA Hybrid Molecules

Sojeong Lee[†,1,2], Yoojung Han[†,1,3], and Hyeshik Chang[1,2,*]

[1] *Center for RNA Research, Institute for Basic Science (IBS), Seoul National University*
[2] *School of Biological science, Seoul National University*
[3] *Interdiciplinary Program in Bioinformatics, Seoul National University*
**Corresponding author: hyeshik@snu.ac.kr*
[†]Equal contribution

Nanopore direct RNA sequencing (DRS) enables transcriptome-wide, single-molecule analysis of full-length RNA by detecting ionic current signals that reveal base sequences as well as modifications like m6A and 2′-O-methylation. However, these signals can vary significantly due to factors like electrolyte concentrations and the quality of nanoelectronics components. To improve detection accuracy, pooling multiple samples in a single run is necessary, but existing multiplexing solutions are outdated and limited. Therefore, a new approach to DRS multiplexing is needed. We developed a basecaller that directly translates DRS ionic current signals into DNA sequences, bypassing the reliance on limited sets of known barcode patterns. Since DNA and RNA produce distinct nanopore signals, we trained a new basecaller from scratch in three stages: manual mapping of mean current levels using designed k-mer report DNA; preliminary neural network training on signal segments generated by a hidden Markov model; and refinement using DNA fragments from digested *E.coli* genomic DNA. Beyond decoding barcode signals for multiplexing, our basecaller enables simultaneous analysis of DNA and RNA within single hybrid molecules, applicable to RNA-DNA hybrids like R-loops. This allows examination of the regulatory roles of RNA and DNA methylation within single R-loop complexes. The approach extends to applications requiring concurrent RNA and DNA sequencing, such as studies of chromatin-associated RNAs, viral RNA-DNA interactions, RNA modification analysis with complementary DNA, and single-cell DNA-RNA co-sequencing.