Decoding RNA 5' Cap Diversity through Reverse Transcriptase Fingerprints

Flaminia Mancini1, 2, Gibran Pedraza-Vázquez1 , Roman Bushuiev1 , Nikolas Tolar3 , Hana Cahova1 1 Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences (IOCB Prague), Chemical Biology of Nucleic Acids, Prague, Czech Republic, 2Charles University, Faculty of Science, Department of Cell Biology, Prague, Czech Republic, 3University of South Bohemia in České Budějovice, Faculty of Science, České Budějovice, Czech Republic

Background: For decades, the N7-methylguanosine (m^7G) cap was believed to be the sole 5' cap structure in viral and eukaryotic mRNAs, playing a central role in regulating RNA biogenesis, stability, translation, and decay. However, the discovery of non-canonical caps derived from metabolites and cofactors, including NAD, FAD, CoA and dinucleoside polyphosphates, has expanded the repertoire of RNA 5'-end modifications and highlighted the need for methods able to determine cap identity on defined RNA molecules. Method: Reverse transcription across RNA modifications can generate characteristic patterns of misincorporation, arrest and nucleotide skipping. Inspired by RT fingerprinting approaches developed for internal RNA modifications, we applied this concept, to our knowledge for the first time, to the identification of 5' RNA caps. In vitro-transcribed RNAs carrying defined caps, including m^7Gp_3A, trimethylguanosine (TMG), NAD, Ap_4A and γ-monomethylphosphate, were analyzed using five selected commercially available RTs and Oxford Nanopore sequencing. Their cap-specific fingerprints were used to build a quantitative classification algorithm. Results: Distinct 5' caps generated reproducible RT-dependent fingerprints. Nanopore long reads provide transcript and isoform identity, allowing cap composition to be resolved for defined RNA species. In endogenous RNAs, the method identified the expected m^7G on multiple mRNAs and γ-monomethylphosphate on U6 snRNA. Analysis of canonical U1-1 snRNA revealed TMG as the predominant cap (~90%), together with minor populations carrying alternative caps, including m^7G and NAD. Conclusion: Our approach combines 5'-cap identification and quantification with transcript- and isoform-level resolution, enabling the detection of cap heterogeneity within individual RNA species. We are applying this framework across four human cell models—HEK293T, BEAS-2B, MCF-7 and A549—representing distinct cellular and disease contexts, and extending it to biologically relevant long non-coding RNAs to investigate previously unexplored transcript-specific capping patterns.