[BIOINFO 2025]

A redesigned probabilistic Transformer for cascading multi-omics integration

Wonjin Yang[1,2], Wookyung Yu[1], and Mookyung Cheon[2,*]

[1] *Department of Brain Sciences, DGIST, Daegu 42988, Republic of Korea*
[2] *Dementia Research Group, Korea Brain Research Institute (KBRI), Daegu 41062, Republic of Korea*
**Corresponding author: mkcheon@kbri.re.kr*

The rapid evolution of multi-omics technologies has generated vast, heterogeneous datasets that demand advanced computational strategies for integrative analysis, particularly in elucidating disease mechanisms through sequential data modalities. To address the challenges of integrating diverse omics layers—such as single-cell RNA sequencing (scRNA-seq) and gene-disease associations—we introduce TABICO (Transformer Architecture with Bayesian Inference for Cascading Omics), a redesigned probabilistic Transformer architecture for cascading multi-omics integration. This approach reinterprets Transformer architectures in a Bayesian manner, enabling modular inference across a wide range of multi-omics data. By leveraging attention mechanisms with uncertainty quantification, the framework supports seamless cascading of new omics measurements with prior data, preserving contextual information across modalities while adapting to emerging datasets without retraining from scratch. We apply TABICO to integrate scRNA-seq data and gene-disease associations, yielding interpretable identifications of disease-relevant genes, cell types, and molecular environments, including clinical contexts. TABICO enhances scalability and robustness, addressing limitations in other language model-based methods. Overall, TABICO offers a versatile tool for modular multi-omics analysis, advancing precision medicine through adaptive inference in the era of sequential biological data generation.