GENE-X: A Dual-Transformer AI Model for Enhanced Gene Expression Prediction From Chromatin Accessibility Data

Tekari, Mohamed (School: Lycee Pilote Monastir)

Accurate prediction of gene expression from chromatin accessibility data is essential for understanding transcriptional regulation across cell types. While the General Expression Transformer (GET) has shown strong predictive performance using regulatory sequence and chromatin accessibility data, its ability to fully model both short- and long-range genomic interactions remains limited. This project presents GENE-X, an AI-based model designed to advance gene expression prediction through a dual Transformer framework and a contextual attention mechanism that integrates local and distant regulatory features. Publicly available single-cell datasets were used for training, with dimensionality reduction performed via Principal Component Analysis. GENE-X utilizes parallel local and distant attention modules, whose outputs are fused through adaptive weighting to capture hierarchical regulatory dependencies. The model was trained using the AdamW optimizer and evaluated on ATAC-seq dataset . GENE-X showed a 3.05% improvement in R^2 compared to GET, reflecting enhanced predictive accuracy. Additionally, Pearson correlation performance improved by 2.7%, demonstrating the model's potential for more precise regulatory inference. These preliminary results highlight GENE-X’s promising architecture for advancing regulatory inference, demonstrating that it offers a more robust and scalable approach to modeling transcriptional regulation, with potential applications in precision medicine and functional genomics

Awards Won: