[BIOINFO 2025]

Deep learning-based modeling of 3D genome dynamics

Kyukwang Kim[1] and Inkyung Jung[1,*]

[1]*Department of Biological Sciences, Korea Advanced Institute of Science and Technology*
**Corresponding author: ijung@kaist.ac.kr*

Dynamics of 3D chromatin structure and the resulting gene regulation is known to play a crucial role during cellular differentiation. However, due to experimental limitations, the 3D chromatin structure state at specific time points or cell type stages is observable, which makes it difficult to obtain sufficient information about the sequential changes in 3D chromatin structure throughout the differentiation process. To resolve this issue in a computational approach, we developed a deep learning-based method to generate Hi-C contact maps for intermediate stages of differentiation. Our approach firstly uses a Variational Autoencoder (VAE) to convert 2D Hi-C contact maps into 1D feature representations. We then use these feature vectors as tokens for a causal attention transformer model (a decoder-only model) to predict non-linear genome dynamics. Finally, the VAE decodes these predictions back into the Hi-C contact maps. The model was trained using Hi-C data from the differentiation of stem cells into cardiomyocytes. The model was tested by generating features for the intermediate cardiac progenitor stage, when the stem cell and cardiomyocyte data, which serve as the start and end time points, were given. After 50 epochs of training, the MSE loss decreased by 2.55-fold on the training set and 2.08-fold on the test set. The predicted features for 394 windows on chromosome 17 had an average cosine distance of 0.11 compared to the original data. Our results present an approach that utilizes tokenization and transformer-based sequence modeling to predict the non-linear and complex dynamics of 3D genomic changes and identify locus-specific characteristics by analyzing the deviations from these predictions.