[BIOINFO 2024]

Alignment-free Nucleic Acid Binding Site Prediction using Equivariant Graph Neural Network and ESM Embbedings

Joohyun Cho[1], Dongsup Kim[1,*]

[1]*Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology*
**Corresponding author: kds@kaist.ac.kr*

Understanding the interactions between proteins and nucleic acids is crucial for revealing the regulatory processes within cells and developing innovative pharmaceuticals. However, accurately pinpointing the protein residues that bind to nucleic acids continues to be a significant challenge. Current state-of-the-art methods make predictions using the multiple sequence alignment (MSA) of the protein, which is highly time-consuming. Here, we present ANAbind (**A**lignment-free EGNN-based **n**ucleic **a**cid **bind**ing site predictor), a nucleic acid binding site prediction method based on the E(n)-equivariant graph neural networks (EGNN). We conducted a rigorous assessment of our model on the two benchmark datasets consisted of native structures and Alphafold2 predicted structures. We find that ANAbind achieves performances competitive with, or better than, the alignment-based state-of-the-art methods. Also, an ablation study on ESM2 and ESM-IF features demonstrated that structural feature plays more critical role in performance on RNA binding site prediction than they do in DNA binding site prediction. Compare to the previous methods, ANAbind is alignment-free, which greatly reduces the computational time.