[BIOINFO 2023]

B cell epitope prediction using graph attention network and ESM-based pretrained protein model embeddings

Sungjin Choi[1], Dongsup Kim[1]

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

Knowledge of B cell epitopes is fundamental for vaccine design, diagnostics, and therapeutics. As the experimental validation for epitopes is time-consuming and costly, many *in silico* tools have been developed to computationally predict the B cell epitope. While most methods have shown poor performance, deep learning methods in recent years were developed to have promising results. Here, we developed EpiGraph which exceeds previous methods, including the recent ones which showed significantly improved performance. Our model outperforms others in two ways. (1) a combination of structure and sequence feature embeddings was obtained from pretrained ESM-IF and ESM-2 models. ESM-IF and ESM-2 could capture structural and evolutionary features of the B cell epitopes, respectively. (2) graph attention network could learn the spatial property of the B cell epitopes with graph homophily. Our model could achieve the best performance on the independent benchmark dataset with 10-fold cross-validation. The result was consistent in different dataset. In the ablation study, we could observe the model superiority with respect to feature engineering and architecture. A User-friendly web server is freely available at http://epigraph.kaist.ac.kr.