[BIOINFO 2025]

Reconstructing gene expression effects for chemical perturbation with biologically informed generative model

Jiyoon Park[1], Jaeseung Song[1], and Wonhee Jang[1,*]

[1]*Department of Life Sciences, Dongguk University*
**Corresponding author: wany@dongguk.edu*

The complexity and heterogeneity of disease microenvironment vary by each individual, which makes it challenging to precisely design the best therapeutic strategy for each patient. Recent advances in artificial intelligence have enabled the prediction of single-cell-level gene expression changes upon combinatorial perturbation, facilitating the identification of optimal drug combinations. In detail, generative models have shown promising performance in predicting cellular responses in gene expression level upon unobserved perturbations, including new drugs in varying range of doses, and genetic modifications. In this study, we propose a novel generative framework that amplifies the biology-awareness of the encoding layer. We train an autoencoder with adversarial training between two neural networks, an encoder and a discriminator to learn a disentangled latent representation. In this process, the encoder is trained to create a basal state latent vector with the perturbation signal removed from gene expression, while the discriminator is optimized to predict the original perturbation. Our model aims to predict effects from unseen drugs, independent of their chemical structures, by embedding both differential gene expression signatures representing drug-induced transcriptomic shifts and textual drug representations using a natural language processing module. The gene embedding space of our module integratively represents scRNA-seq data with protein-protein interaction networks to extract the complicated biological features. Throughout these processes, our framework is enabled to generate the expected genome-wide expression patterns for unseen drugs, which also provides the potential drug-target interactions, significantly improving model interpretability. We believe that our framework can accelerate the *in silico* drug screening process with a highly interpretable manner.