02-3

Omics-based Generation of Drug Candidate Molecules with Desired Phenotypes by Machine Learning

Kazuma Kaitoh[1] kkaitoh@bio.kyutech.ac.jp
Yoshihiro Yamanishi[1] yamani@bio.kyutech.ac.jp

[1] Department of Bioscience and Bioinformatics, Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology, 680-4, Kawazu, Iizuka, Fukuoka,820-8502, Japan

Keywords: Structure generator, Deep generative model, Machine learning, Omics data

The efficient identification of small molecules with desired phenotypes such as bioactivities is a challenging task in the drug-discovery process. In recent years, machine learning techniques have been used for de novo drug design. In particular, molecular generative models are able to learn the characteristics of specific real-world data and generate the chemical structures of new molecules with similar properties[1]. However, the biological phenomena have not been taken into account in molecular generation.

In this study, we present a novel machine learning method for *de novo* drug design using omics data such as gene expression profiles. We investigated the correlation between chemically induced gene expression profiles (reflecting cellular responses to compound treatment) and target perturbation gene expression profiles (reflecting cellular responses to gene knock-down or gene over-expression of target proteins) in terms of interactions between compounds and target proteins. Then, we proposed novel machine learning methods to generate the chemical structures of new molecules with desired gene expression profiles in the framework of variational auto-encoder (VAE) [2] using a new chemical representation, which does not depend on the SMILES strings. The novelty of the proposed method comes from the bridge between the chemical space and the biological space, the high chemical validity of the newly generated structures, the generation of drug candidate compounds from omics data with high accuracy, and the applicability to target proteins without information on ligands. We confirmed that the molecules that were newly generated using our structure generator were valid from the viewpoint of medicinal chemistry, comparing the structures between the newly generated molecules and known ligands. It was observed that the newly generated molecules were more similar to known ligands than those generated by the previous method with similar goals[3]. Our proposed method for omics-based molecular generation is expected to be useful for de novo drug design in practice.

[1] Gómez-Bombarelli, R.; *et al.*, Automatic Chemical Design Using a Data-driven Continuous Representation of Molecules. *ACS Cent. Sci.*, **2018**, *4*, 268-276.
[2] Hinton, G. E.; Salakhutdinov, R.R. Reducing the Dimensionality of Data with Neural Networks. *Science*, **2006**, *313*, 504–507.
[3] Méndez-Lucio, O.; *et al.*, De Novo Generation of Hit-like Molecules from Gene Expression Signatures Using Artificial Intelligence. *Nat. Commun.*, **2020**, *11*, 10.