[BIOINFO 2024]

A Multimodal Approach for Predicting Drug Metabolism Using CYP2D6 Sequences and Drug Structures

Yeabean Na[1] and Sunyong Yoo[*]

[1]*Department of Intelligent Electronics and Computer Engineering, Chonnam National University*
**Corresponding author: syyoo@jnu.ac.kr*

Cytochrome P450 2D6 (CYP2D6) is involved in the metabolism of up to 25% of the drugs that are in common use in the clinic. Characterized by high polymorphism, CYP2D6 is one of the key pharmacogenes in the field of pharmacogenomics. This genetic variability can lead to significant inter-patient differences in drug metabolism, resulting in differential therapeutic responses and adverse effects among individuals taking the same medication. However, Conducting *in vivo* or *in vitro* experiments for each CYP2D6 variant across various drugs is time-consuming, ethically challenging, and expensive. Given these constraints, *In silico* modeling approaches for predicting drug metabolism profiles of CYP2D6 variants emerge as a critical necessity. Previous study has demonstrated the feasibility of *in silico* approaches to predict metabolizer phenotypes for various CYP2D6 allelic variants based on genotype data. Traditionally, individuals were classified into four metabolizer categories based on an activity score system, which was derived from the sum of individual allele values assigned to CYP2D6 allelic variants. However, the activity scores are primarily based on extensive literature reviewing *in vivo* and *in vitro* activities of commonly used probe drugs, which may not be consistent across all pharmaceuticals metabolized by CYP2D6. This limitation underscores the necessity for drug-specific metabolism prediction approaches. To address the need for drug-specific metabolism prediction, we propose a deep learning approach that combines CYP2D6 genotype data and drug structural information. This approach utilizes a Convolutional Neural Network (CNN) for processing genotype data and a Graph Convolutional Network (GCN) for analyzing drug structures, integrating these diverse data types into a multimodal model for predicting drug metabolism. The developed multimodal model demonstrated better performance in predicting drug metabolism compared to model utilizing solely CYP2D6 genotype data. We anticipate that our model will enhance the prediction of metabolic capacity in previously uncharacterized CYP2D6 variants, potentially contributing to the reduction of adverse drug reactions.