[BIOINFO 2023]

Validation of Network-Based Algorithm through Prediction of Drug Efficacy in Type 2 Diabetes

Jiyeon Kim[1], Yujeong Shim[1], Gyuyeon Jang[1], Su Han Cho[1], Yeonbin Jeong[1], Jungeun Kim[1,2]*,

and Ho Kim[1]

[1] *basgenbio Inc., Seoul, Republic of Korea*
[2] *Department of Statistics and Acturial Science, Soongsil University, Seoul, South Korea*
**Corresponding author: jekim@basgenbio.com*

The network algorithm is based on distance calculations within the Protein-Protein Interaction (PPI) network to analyze the relationship between drugs and diseases. This algorithm not only suggests the potential of drugs as treatments for specific diseases but also proposes the repurposing of drugs for previously unknown indications. We verified the reliability of this network algorithm by comparing the distance between type 2 diabetes (T2D) -related genes and T2D drugs to that of other drugs. To construct the PPI network, we utilized the HuRI[5], BioGRID[3,4], and STRING[6] databases. We used interaction information with STRING scores of 0.9 or higher and excluded cases where Entrez IDs were not mapped. Disease modules can be constructed using various methods, such as transcriptome data, literature research, or open databases. In this analysis, we employed the DisGeNet[3,4] database to select 82 genes associated with T2D based on the GDA score. For drug modules, we downloaded curated data provided from ref1 and extracted information on the target genes of the drugs. Within the PPI network, we calculated the shortest distance between drug modules and disease modules. Additionally, we randomly generated modules with the same size of genes which have similar degrees as the drug module, repeating this process 1000 times to create a reference distribution of distances. Subsequently, we normalized the shortest distance using the mean and standard deviation from the reference distribution. The results of the analysis showed that when employing the network algorithm to predict diabetes treatment drugs, the AUROC was 0.84. When comparing the distance between T2D treatment drugs and the T2D disease module to that of unrelated H drugs which are randomly selected 25 hormonal drugs with ATC codes beginning with 'H', a notable difference was observed in the distribution (Fig.4). The network-based algorithm not only offers insights into the efficacy of drugs for diseases but also allows to easily apply for various diseases. Additionally, it suggests the potential for drug repurposing.