Online Supplementary Data

Machine learning-based prediction of drug-drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties

Feixiong Cheng,[1] Zhongming Zhao[1,2,3,4,*]

[1]Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, Tennessee, USA

[2]Department of Cancer Biology, Vanderbilt University School of Medicine, Nashville, Tennessee, USA

[3]Department of Psychiatry, Vanderbilt University School of Medicine, Nashville, Tennessee, USA

[4]Center for Quantitative Sciences, Vanderbilt University Medical Center, Nashville, Tennessee, USA

[*]Address correspondence to:

Zhongming Zhao, Ph.D.

Department of Biomedical Informatics

Vanderbilt University School of Medicine

2525 West End Avenue, Suite 600

Nashville, TN 37203, USA

Phone: (615) 343-9158

Fax: (615) 936-8545

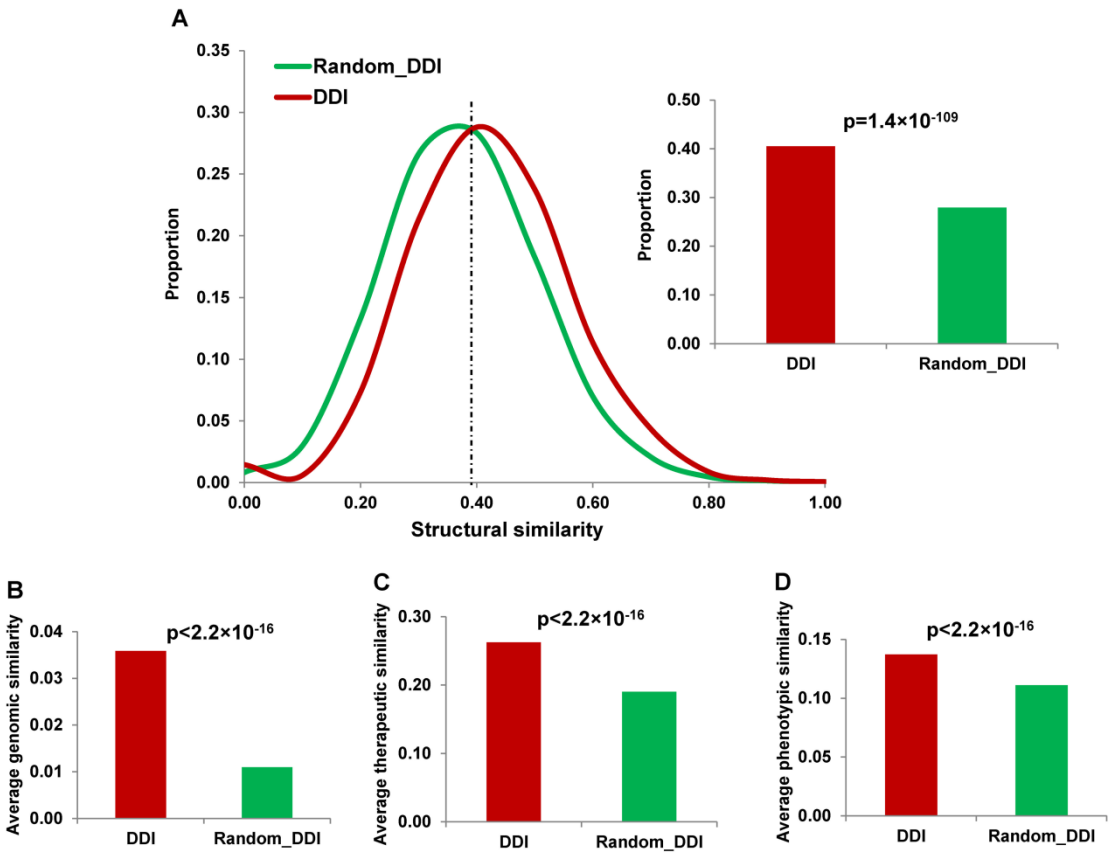
Email: zhongming.zhao@vanderbilt.edu

Supplementary tables and figures

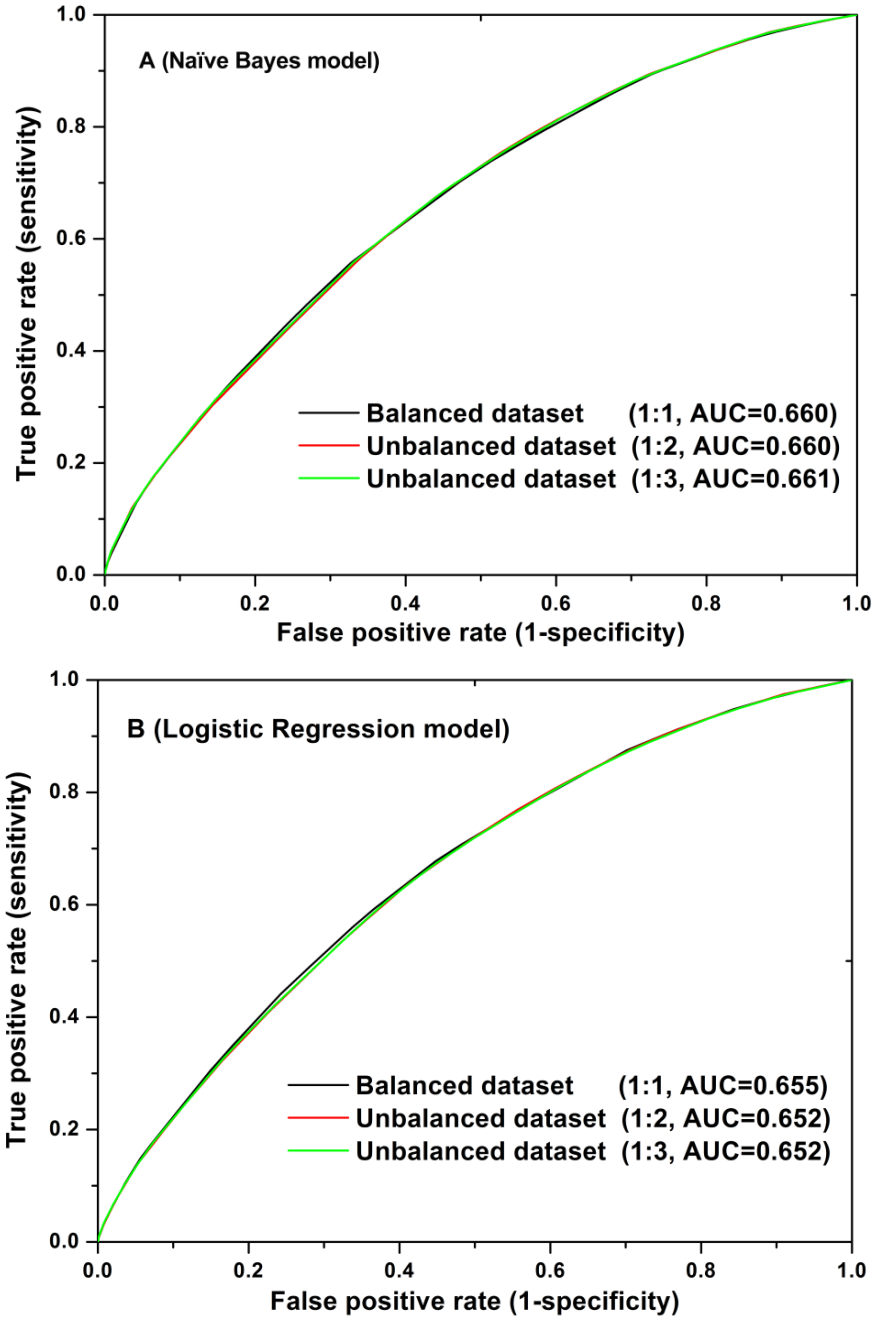
Supplementary table S1. Details of positive drug-drug interaction pairs and negative drug-drug interaction pairs used in this study. (Table S1.xlsx)

Supplementary table S2. The new predicted drug-drug interactions using the predictive heterogeneous network-assisted inference framework. (Table S2.xlsx)

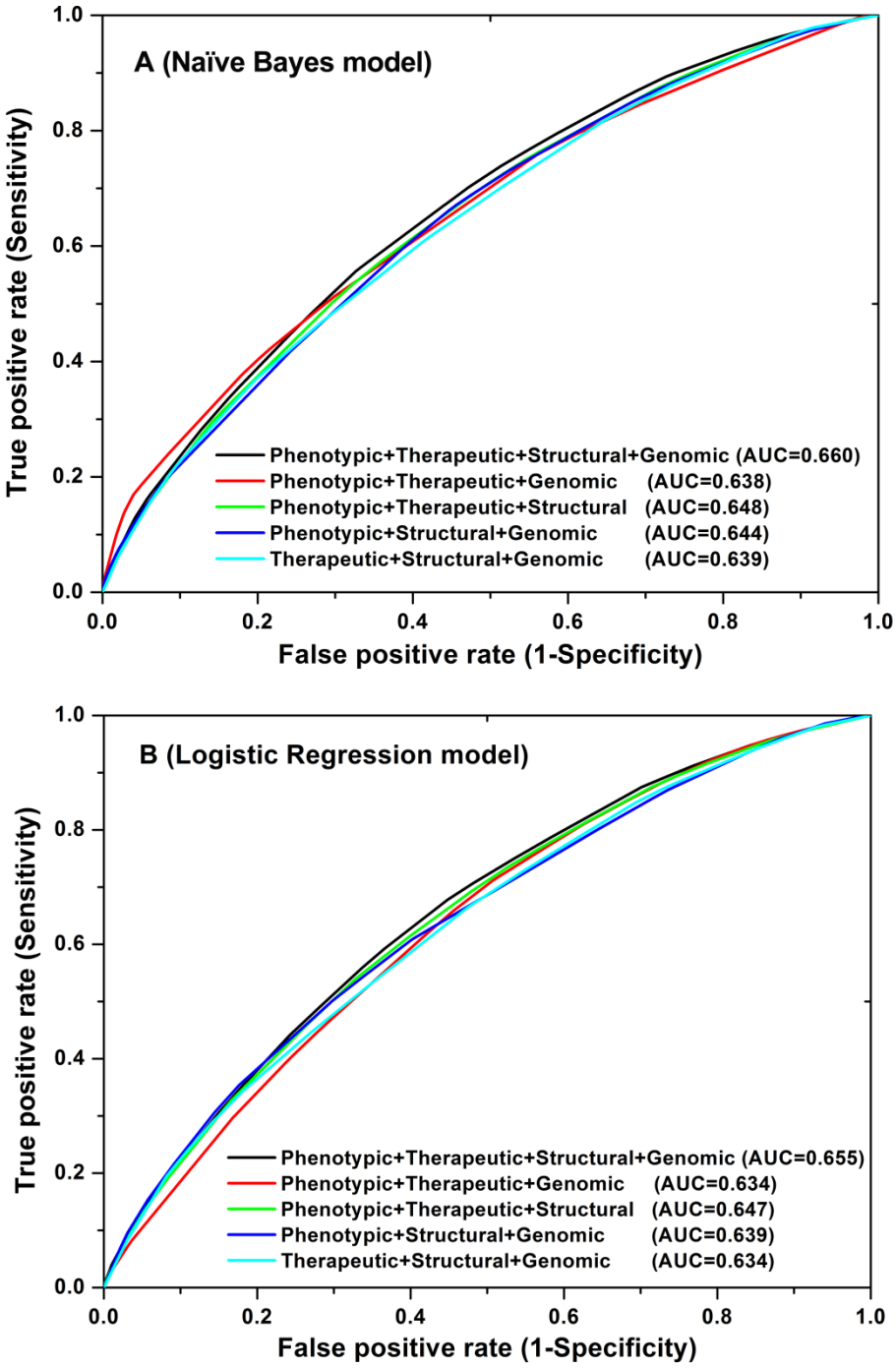
Supplementary table S3. The known drug-target interactions, known drug-drug interaction and new predicted drug-drug interactions for antipsychotic drugs. (Table S3.xlsx)



Supplementary figure S1. The distribution of four similarities in positive drug-drug interaction (DDI) pairs and random DDI pairs (online supplementary table S1). These four similarities include structural similarity, genomic similarity, therapeutic similarity, and phenotypic similarity. We found that the similarities among positive DDI pairs are significantly higher than those among random DDI pairs, which confirms the hypothesis of the Heterogeneous Network-Assisted Inference (HNAI) framework. The p-value of **A** was calculated using Fisher's exact test. The p-values of **B**, **C**, and **D** were calculated using Wilcoxon's test.



Supplementary figure S2. The receiver operating characteristic (ROC) curves of 5-fold cross-validation when using the balanced dataset (the ratio of the number of positive drug-drug interactions (DDIs) and negative DDIs is equal 1, online supplementary table S1), and two unbalanced datasets (the ratio of the number of positive DDIs and negative DDIs is equal 1/2 and 1/3, respectively). We found that the performance of the balanced dataset is marginally higher than the unbalanced datasets.



Supplementary figure S3. The receiver operating characteristic (ROC) curves of 5-fold cross-validation when removing each drug-drug pair feature (phenotypic, therapeutic, structural, and genomic similarity), respectively. We found that both the naïve Bayes model and the logistic regression model yielded poor performances when structural or phenotypic features were removed. Thus, the structural or phenotypic features are important for model performance compared to the other two features.