As was mentioned in the main text of the manuscript, the target proteins are divided into four groups including nuclear receptors, GPCRs, ion channels and enzymes. All of the operations of this study were implemented separately for each group. For example, the datasets and implemented source codes for the nuclear receptors group include the following:

Nuclear receptor file: The Known drug-target interactions proposed by Yamanishi et al.

Nucleardescriptors file: The drug compound descriptors extracted by PaDEL-Descriptor software.

Nuclearfeatures file: The target protein features extracted by protr R package.

NuclearFASTA folder: The sequence of all target proteins in nuclear receptors group (in FASTA format).

Nuclearreceptordrugs folder: The mol file of all drug compounds so that have known interaction with at least one target protein in the nuclear receptors group.

Supplementary folder: This folder contains a table of evaluation results and new predicted interactions in each target protein group that are referred to in the main text of the manuscript.

BRNS_negative folder: This folder contains three subfolders. In each subfolder, there are source codes related to each of the feature selection methods mentioned in the main text of the article. Selecting negative samples (non-interacting drug-target pairs) was performed based on the BRNS algorithm.

Random_negative folder: Similar to the BRNS folder, this folder also contains three subfolders. The difference between the source codes in these three subfolders and the subfolders in BRNS is in the negative samples selection method, which was done randomly here.

The order of executing the source codes in these two folders is as follows:

1) **Makenegname.R:** implementation of BRNS algorithm in order to selecting balanced reliable negative samples.
2) **Makesemisubdpfeatures.R:** Build a semi-supervised dataset (only for semi-supervised feature selection method)
3) **feature selection.R:** The source code for feature selection.
4) **Cross validation.R:** The source code for 10-fold cross validation.

The source codes for the other three groups of target proteins can be generalized exactly the same as the source codes presented in the nuclear receptor group according to the available dataset in every group.