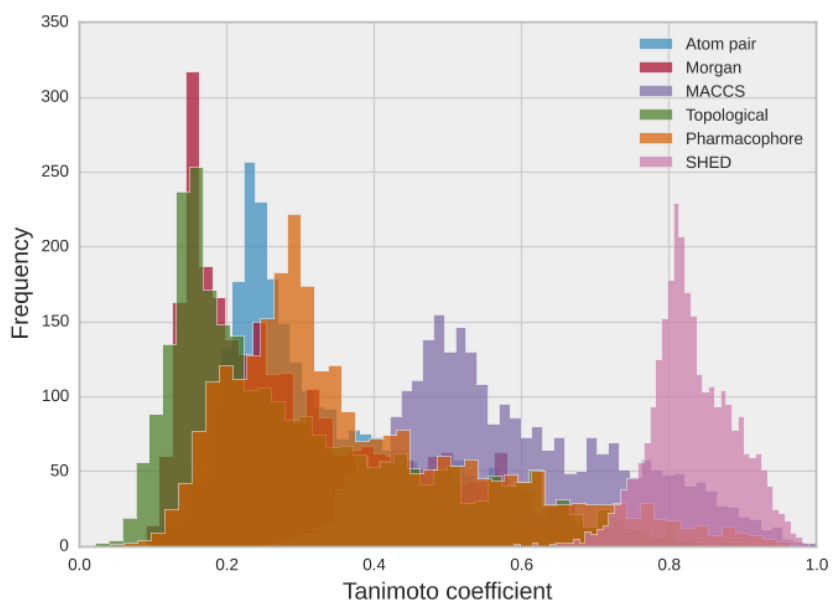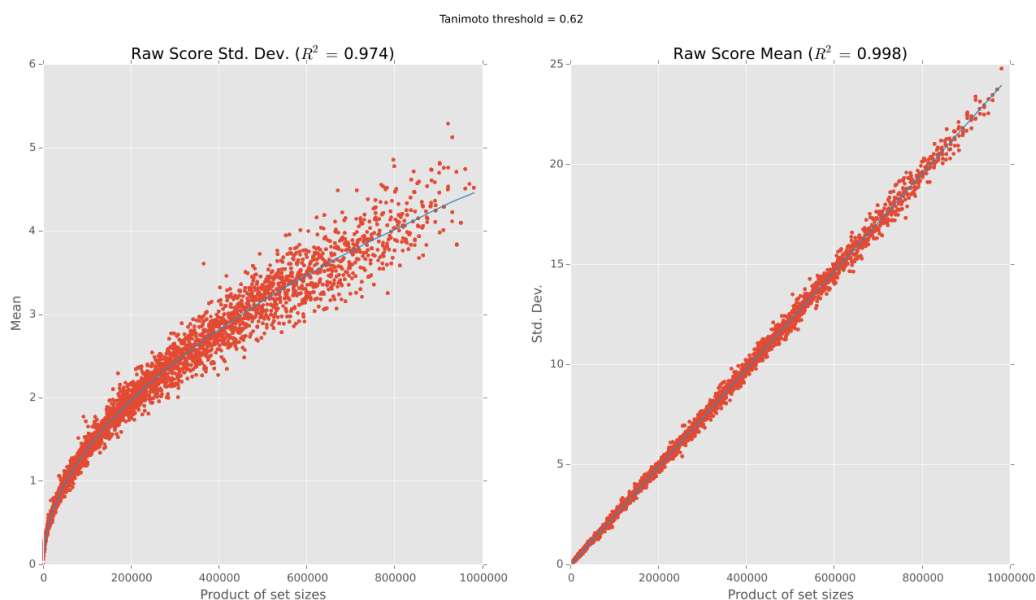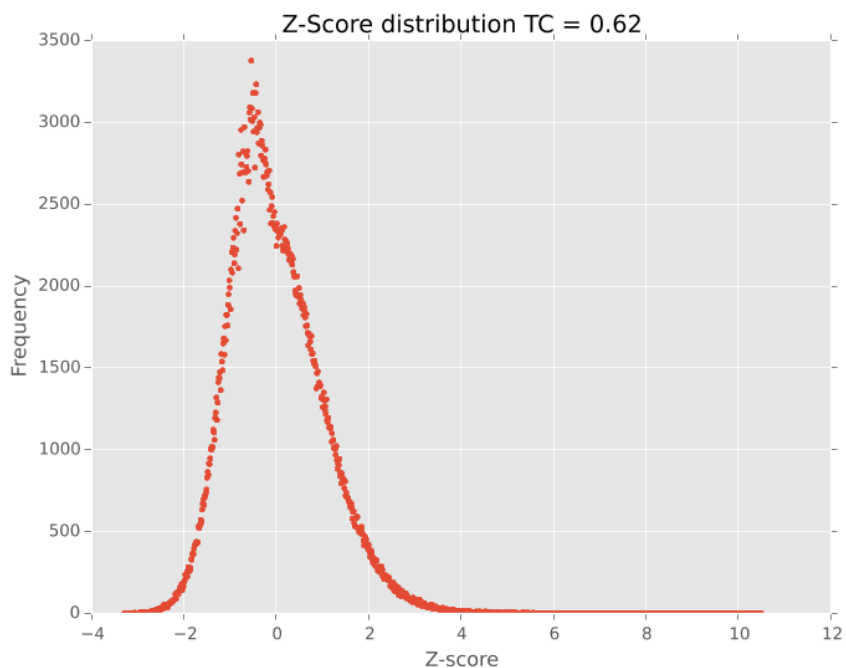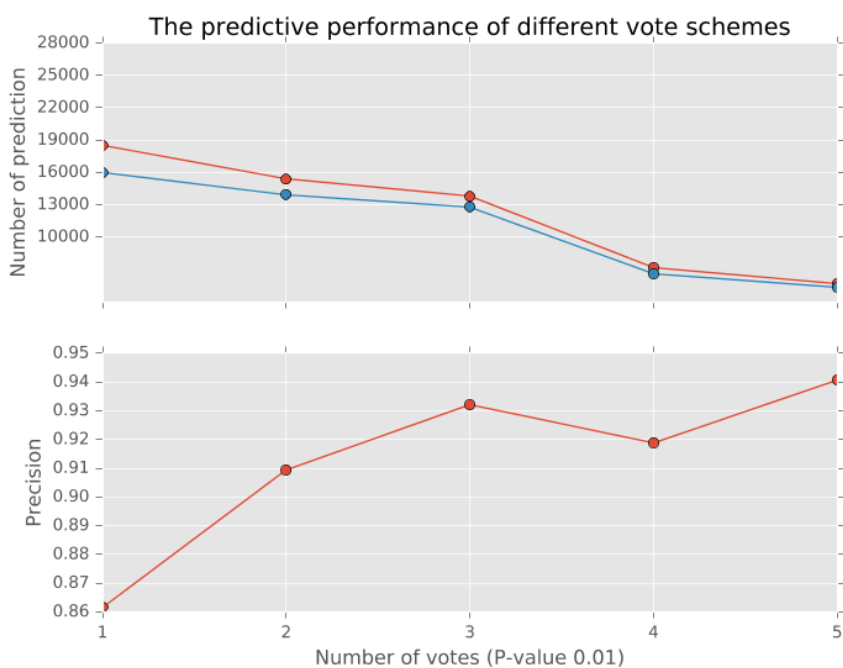# Additional Figures

Figure S1. The average similarity of five fingerprints (Atom pair, Morgan, MACCS, Topological and Pharmacophore, which are implemented in RDKit package [http://rdkit.org/].) and SHED descriptor. The similarity criteria for SHED is the normalized Euclidean distance (see the main manuscript) and for the other five fingerprints are Tanimoto coefficient.

Figure S2. Statistical model fits for Morgan based SEA on the random background data set create from ChEMBL 19.

Figure S3. Z-score distribution (Morgan fingerprint) of the random background data set created from ChEMBL 19 database.

Figure S4. The predictive performance of different vote schemes with significant level P-value ≤ 0.01. The upper plot illustrates the total number of positive (in red) and true positive prediction (in blue), and the lower plot is the corresponding precision.