

# Supplemental: Taking a Respite from Representation Learning for Molecular Property Prediction

Jianyuan Deng<sup>1</sup>, Zhibo Yang<sup>2</sup>, Hehe Wang<sup>3</sup>, Iwao Ojima<sup>3</sup>, Dimitris Samaras<sup>2</sup>, and Fusheng Wang<sup>1,2,\*</sup>

<sup>1</sup>Stony Brook University, Department of Biomedical Informatics, Stony Brook, 11790, United States

<sup>2</sup>Stony Brook University, Department of Computer Science, Stony Brook, 11790, United States

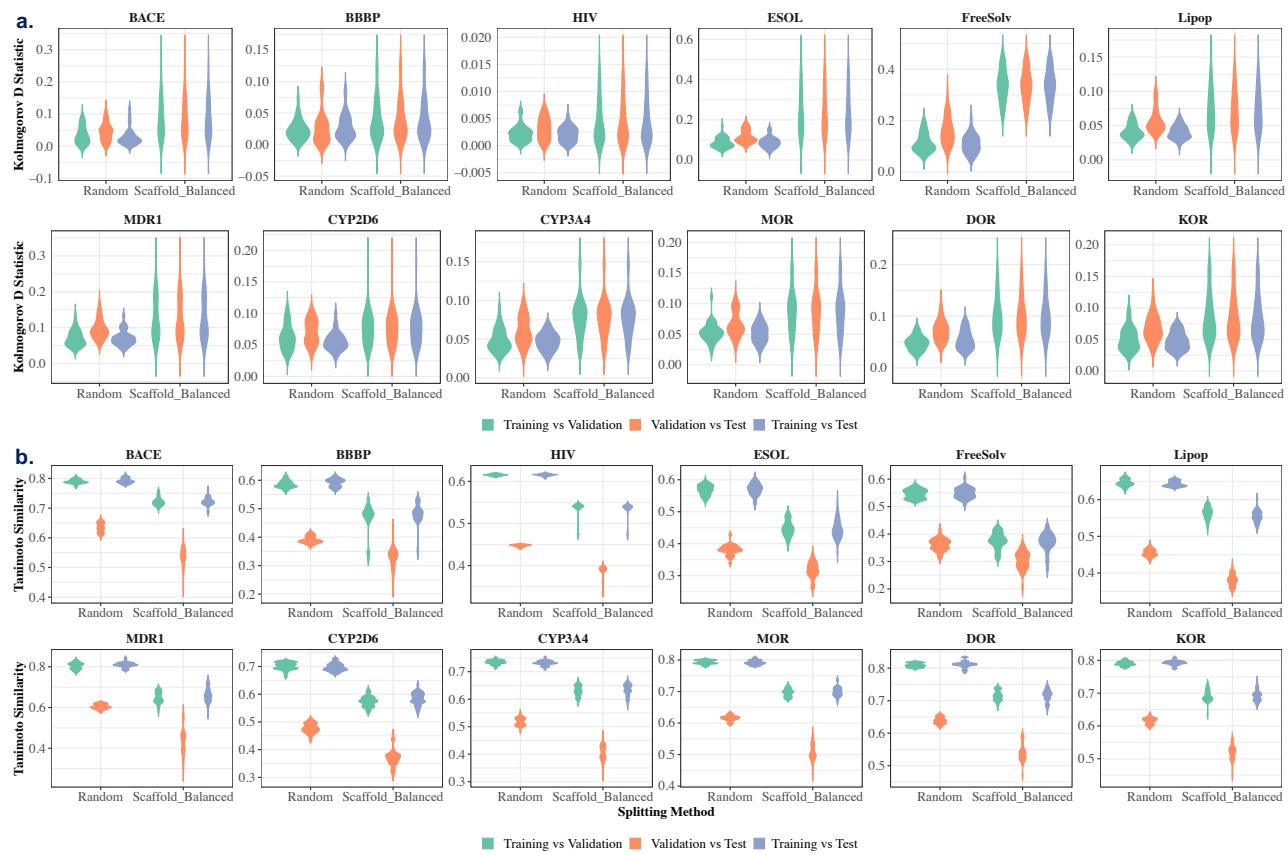
<sup>3</sup>Stony Brook University, Department of Chemistry, Stony Brook, 11790, United States

Dataset		BACE			BBBP			HIV		
Model		RF	MOLBERT	GROVER <sup>T</sup>	RF	MOLBERT	GROVER <sup>T</sup>	RF	MOLBERT	GROVER <sup>T</sup>
AUROC		23	1	6	20	6	4	11	18	1
AUPRC		20	4	6	19	7	4	21	8	1
PPV		20	5	5	14	7	9	19	8	3
NPV		23	4	3	14	10	6	10	20	0
Dataset		ESOL			FreeSolv			Lipop		
Model		RF	MOLBERT	GROVER <sup>T</sup>	RF	MOLBERT	GROVER <sup>T</sup>	RF	MOLBERT	GROVER <sup>T</sup>
RMSE		30	0	0	12	0	18	30	0	0
MAE		30	0	0	13	0	17	30	0	0
R2		30	0	0	12	0	18	30	0	0
PEARSON_R		30	0	0	10	0	20	29	0	1

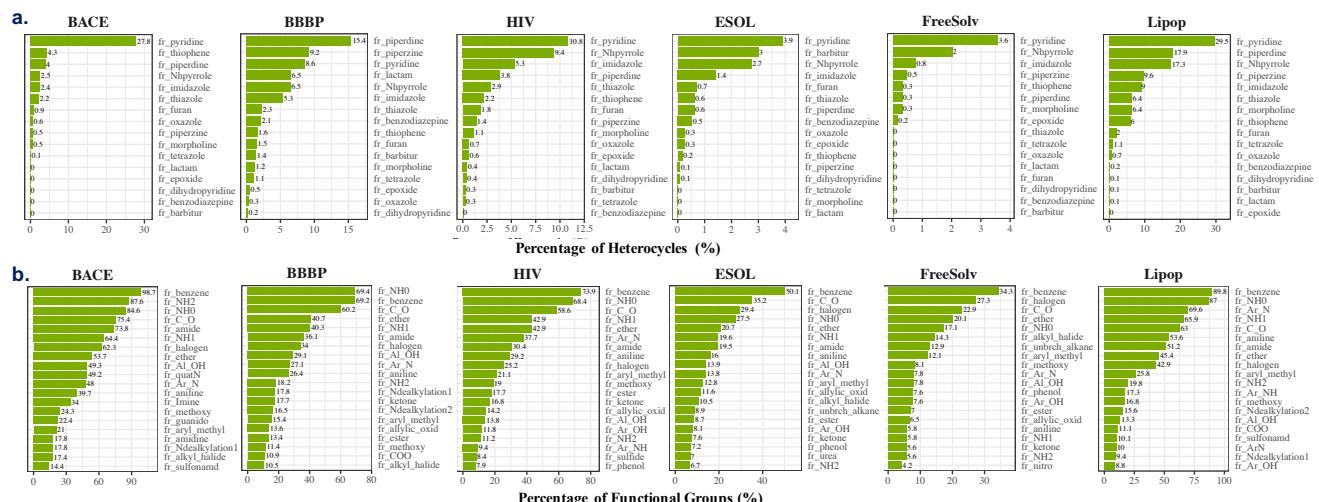
**Table 1.** Number of Single Fold where a Model Achieves the Best Performance using under Scaffold Split.

Dataset		BACE			BBBP			HIV		
Model		RF	MOLBERT	GROVER <sup>T</sup>	RF	MOLBERT	GROVER <sup>T</sup>	RF	MOLBERT	GROVER <sup>T</sup>
AUROC		3,644	23	393	3,189	408	463	1,404	2,635	21
AUPRC		3,162	330	568	2,817	903	340	3,201	848	11
PPV		3,022	386	652	2,435	727	898	3,152	750	158
NPV		3,521	361	178	2,220	1,031	809	1,067	2,993	0
Dataset		ESOL			FreeSolv			Lipop		
Model		RF	MOLBERT	GROVER <sup>T</sup>	RF	MOLBERT	GROVER <sup>T</sup>	RF	MOLBERT	GROVER <sup>T</sup>
RMSE		4,060	0	0	1,450	0	2,610	4,060	0	0
MAE		4,060	0	0	1,655	0	2,405	4,060	0	0
R2		4,060	0	0	1,506	0	2,554	4,060	0	0
PEARSON_R		4,060	0	0	912	0	3,148	4,060	0	0

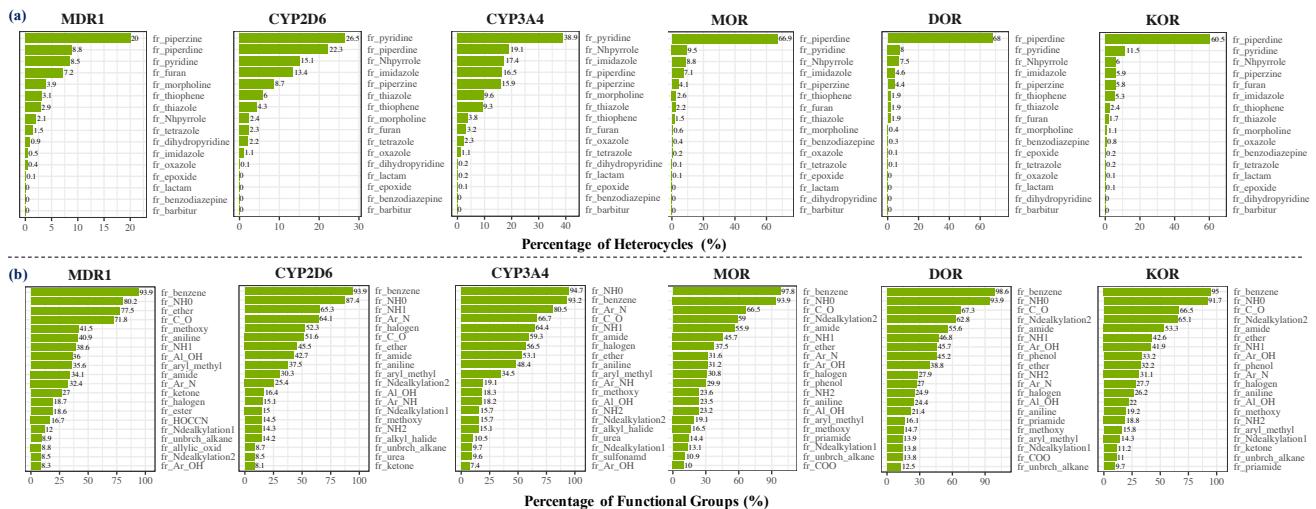
**Table 2.** Number of 3-Fold Combinations where a Model Achieves the Best Performance under Scaffold Split.



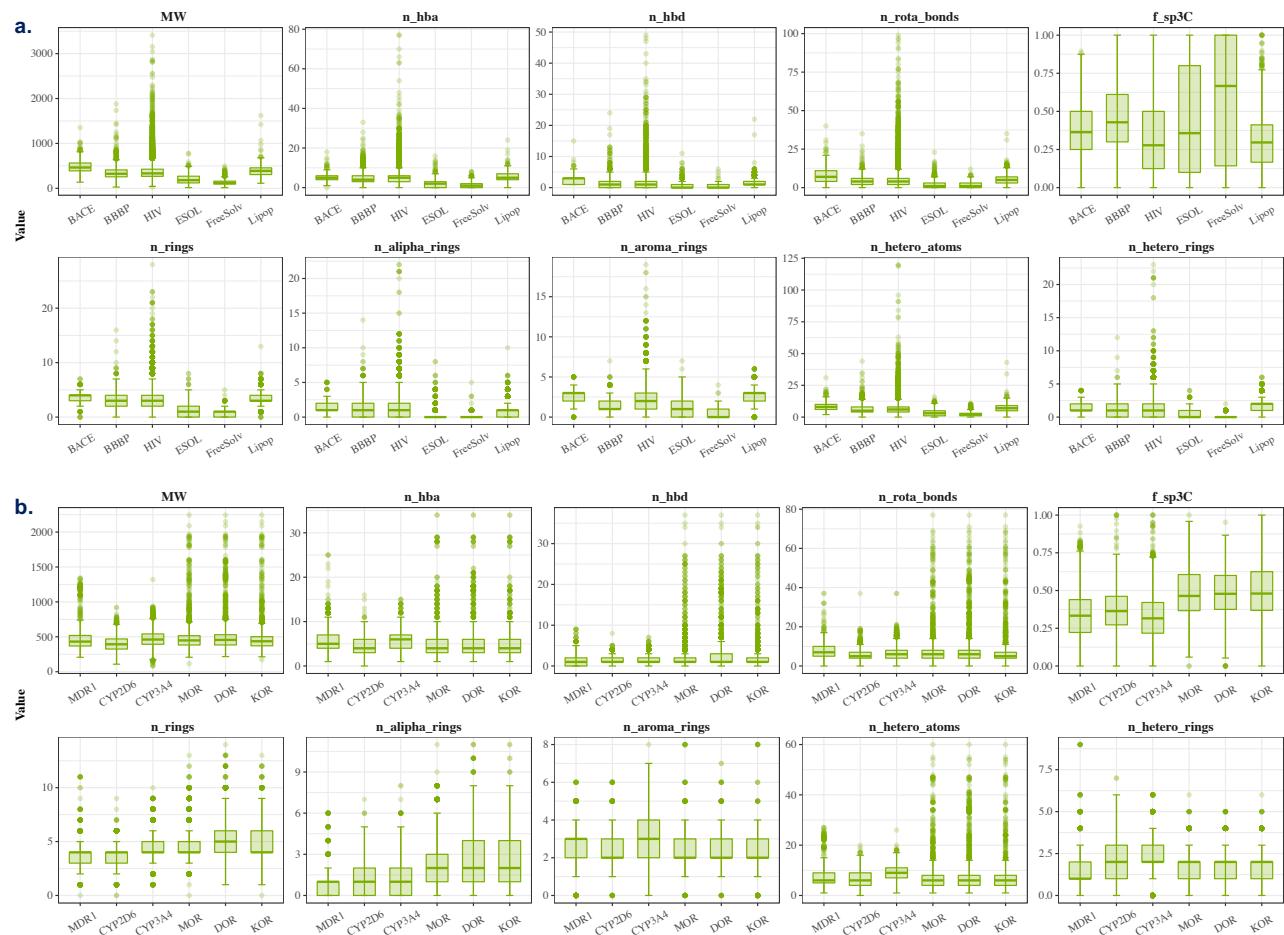
**Figure 1.** Distribution of **a.** Kolmogorov D Statistic and **b.** Tanimoto Similarity among Training, Validation and Test Sets.



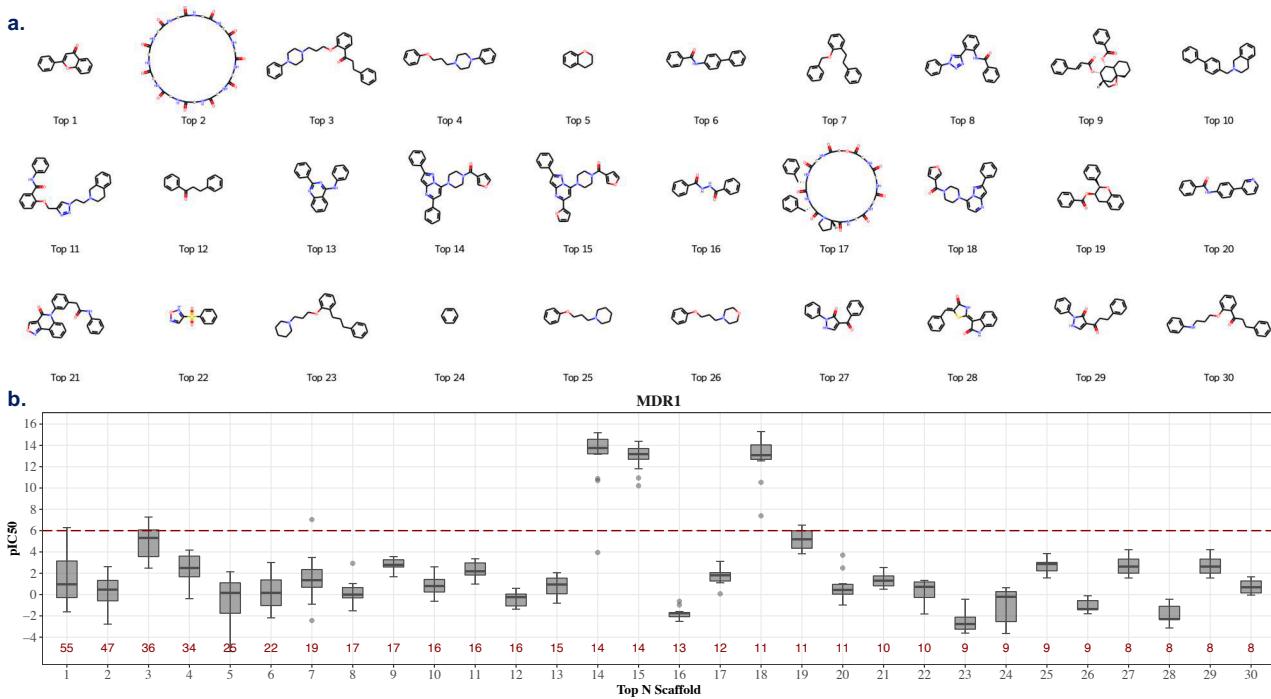
**Figure 2.** Top Fragments Prevalence for the Benchmark Datasets. **a.** Top heterocycles **b.** Top functional groups.



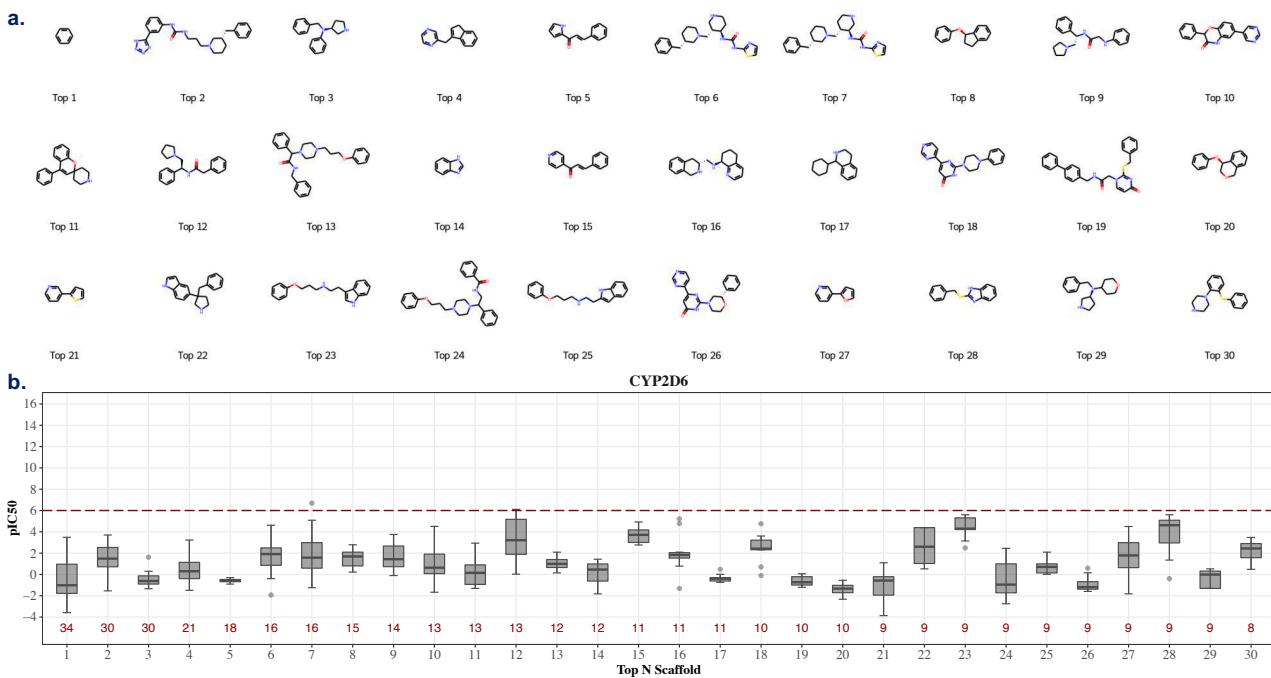
**Figure 3.** Top Fragments Prevalence for the Opioids-related Datasets. **a.** Top heterocycles **b.** Top functional groups.



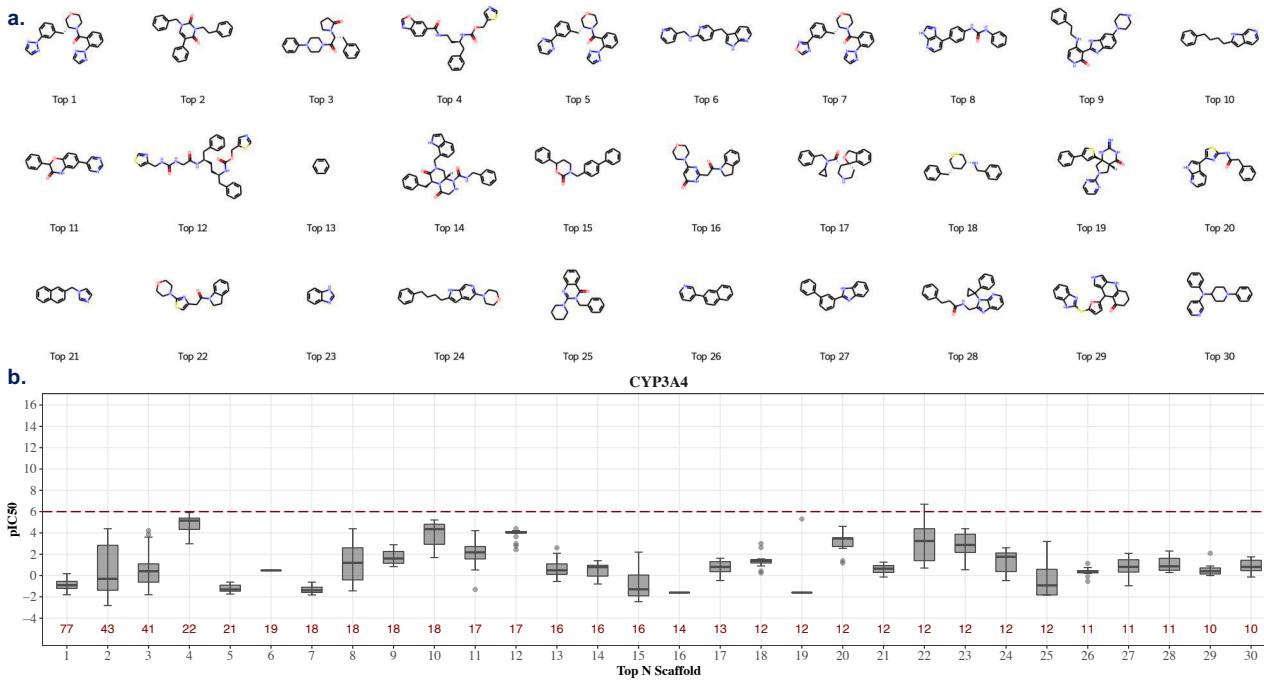
**Figure 4.** Distribution of Other Structural Traits for **a.** MoleculeNet Benchmark Datasets and **b.** Opioids-related Datasets.



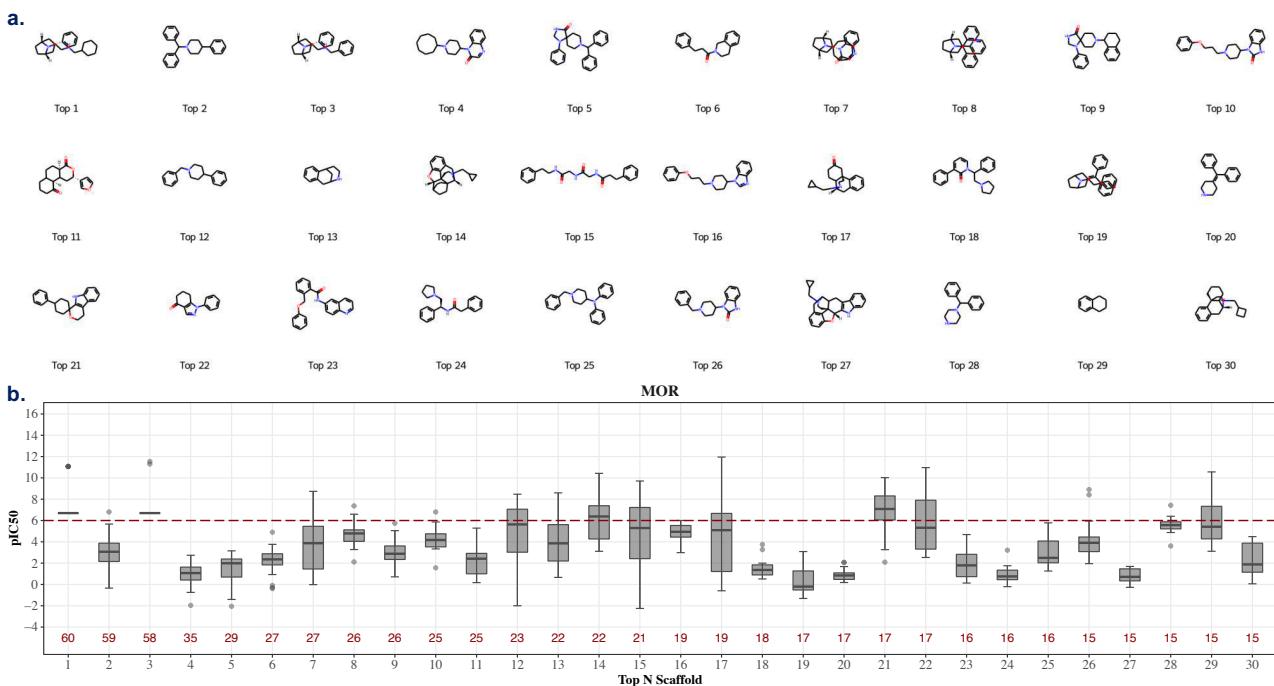
**Figure 5.** Scaffolds and Label Distribution in the MDR1 Dataset. **a.** Top 30 scaffolds in the MDR1 dataset. **b.** pIC50 distribution for molecules with the top30 scaffolds (red number shows how many molecules are equipped with the scaffold).



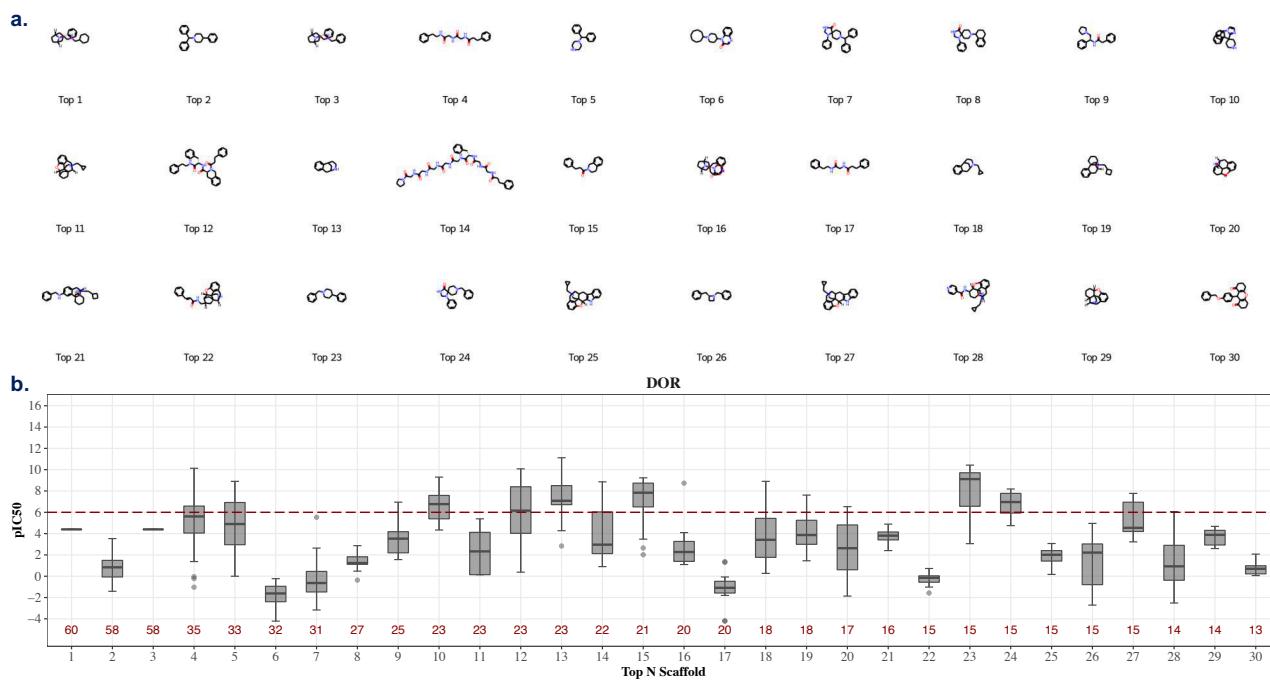
**Figure 6.** Scaffolds and Label Distribution in the CYP2D6 Dataset. **a.** Top 30 scaffolds in the CYP2D6 dataset. **b.** pIC50 distribution for molecules with the top30 scaffolds (red number shows how many molecules are equipped with the scaffold).



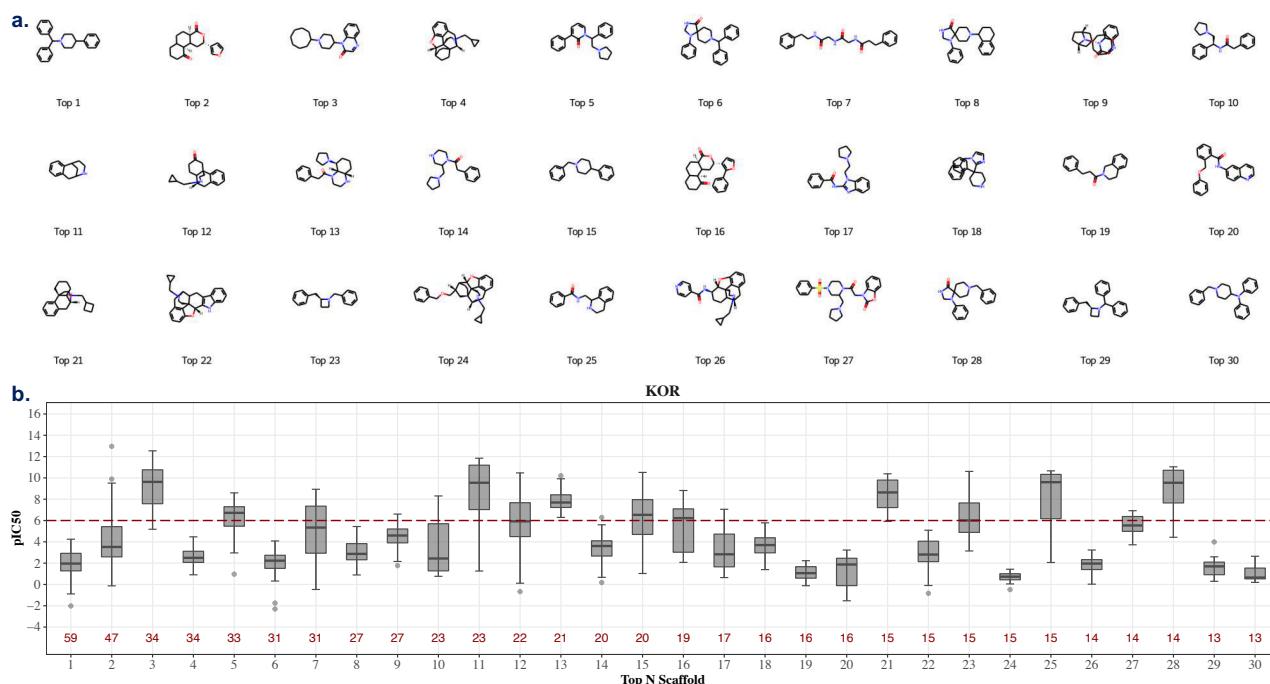
**Figure 7.** Scaffolds and Label Distribution in the CYP3A4 Dataset. **a.** Top 30 scaffolds in the CYP3A4 dataset. **b.** pIC<sub>50</sub> distribution for molecules with the top30 scaffolds (red number shows how many molecules are equipped with the scaffold).



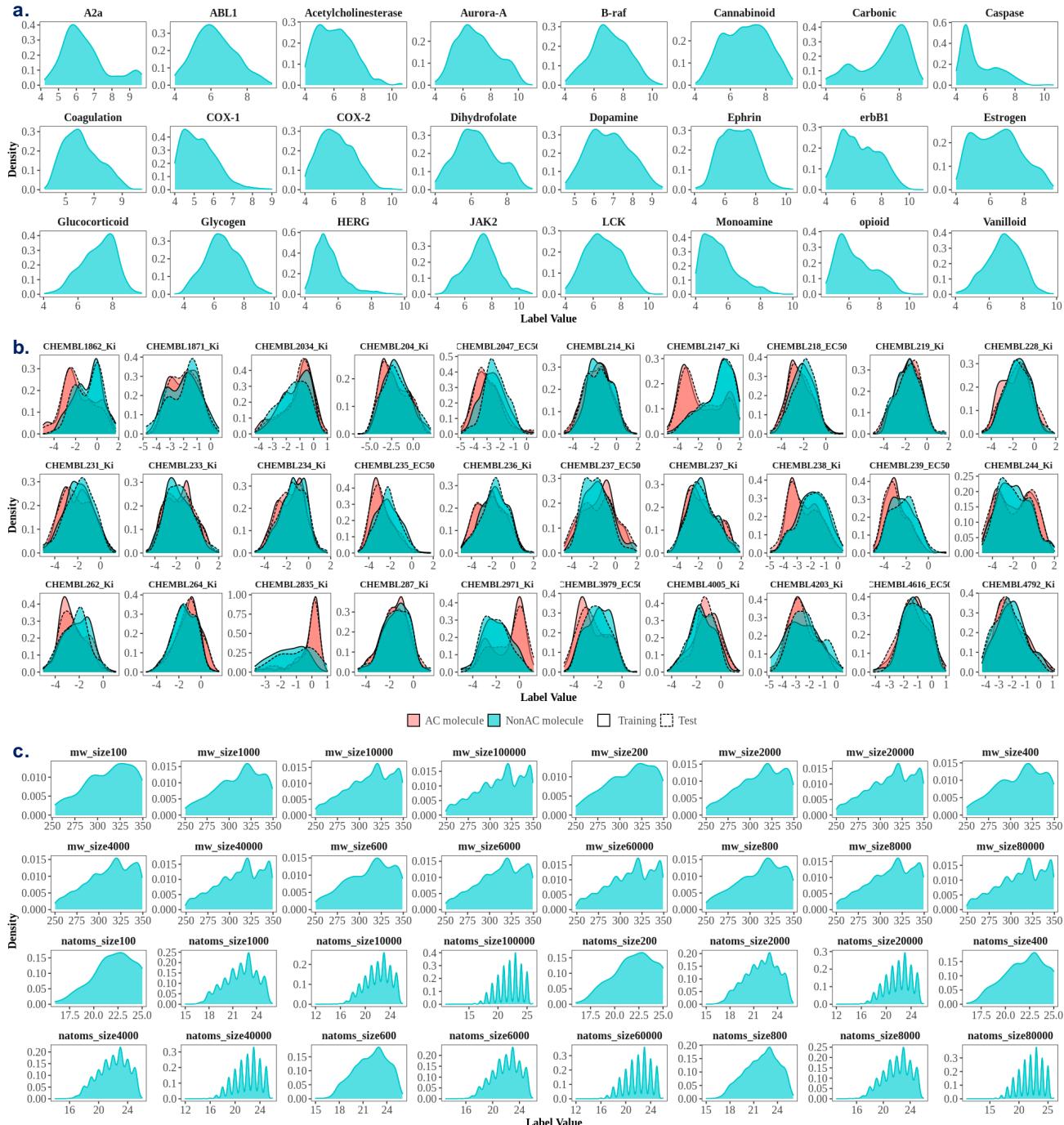
**Figure 8.** Scaffolds and Label Distribution in the MOR Dataset. **a.** Top 30 scaffolds in the MOR dataset. **b.** pIC<sub>50</sub> distribution for molecules with the top30 scaffolds (red number shows how many molecules are equipped with the scaffold).



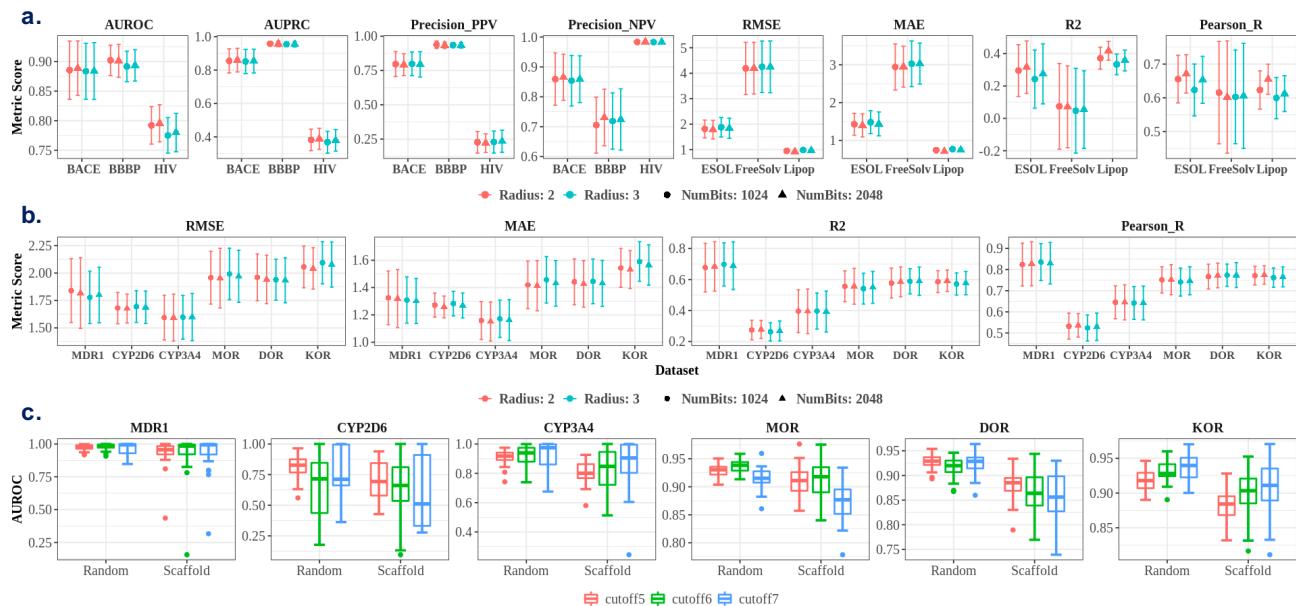
**Figure 9.** Scaffolds and Label Distribution in the DOR Dataset. **a.** Top 30 scaffolds in the DOR dataset. **b.** pIC50 distribution for molecules with the top30 scaffolds (red number shows how many molecules are equipped with the scaffold).



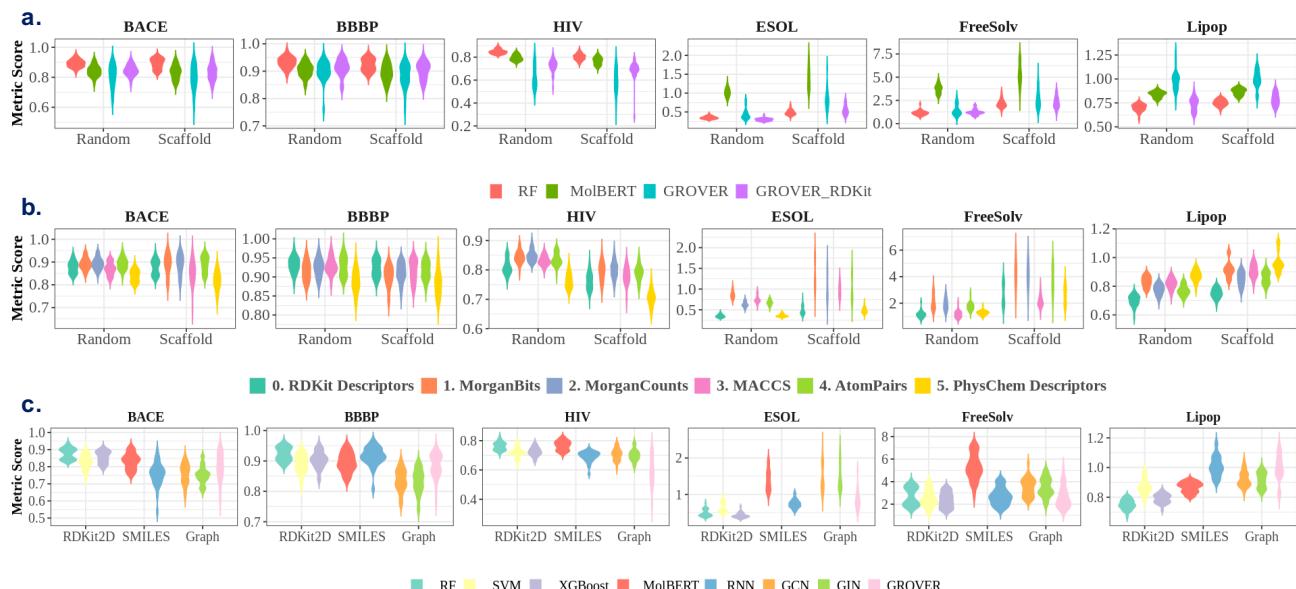
**Figure 10.** Scaffolds and Label Distribution in the KOR Dataset. **a.** Top 30 scaffolds in the KOR dataset. **b.** pIC50 distribution for molecules with the top30 scaffolds (red number shows how many molecules are equipped with the scaffold).



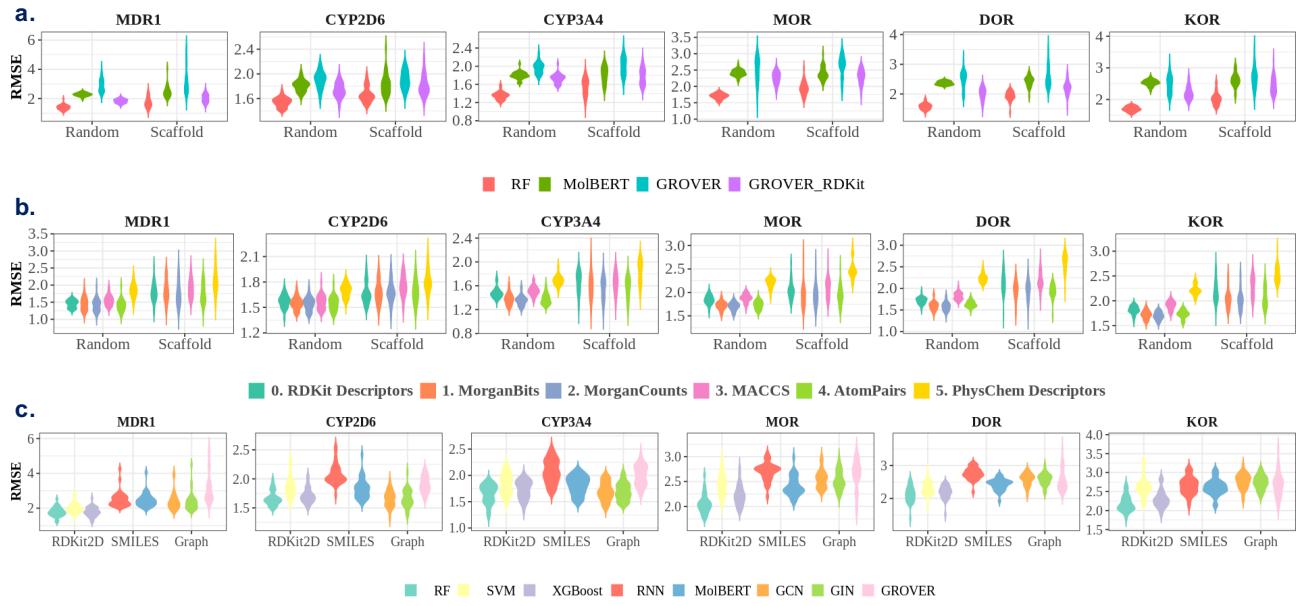
**Figure 11.** Label Distribution for **a.** Activity datasets proposed by Cortés-Ciriano et al, **b.** Activity datasets proposed by Tilborg et al and **c.** Descriptor Datasets.



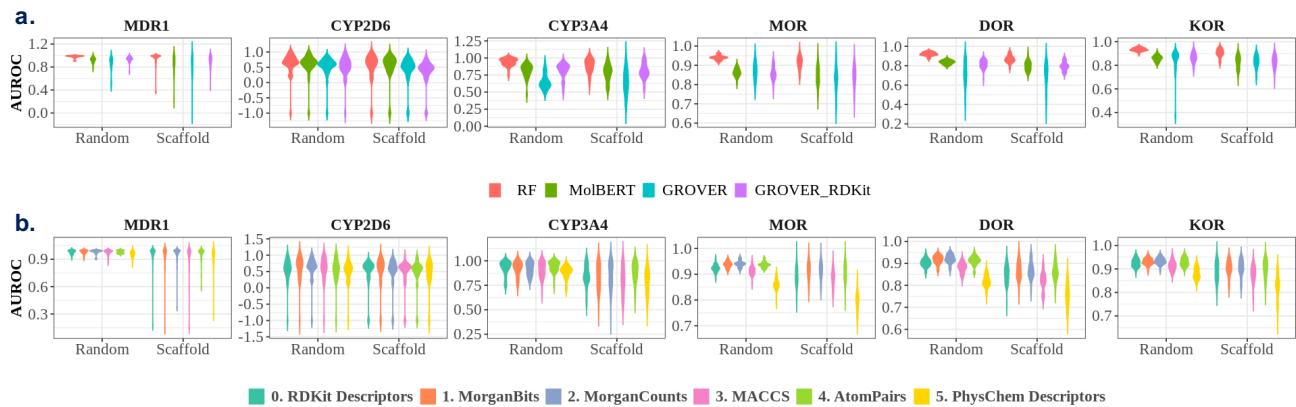
**Figure 12.** Prediction Performance RF on MorganBits Fingerprint in **a.** MoleculeNet benchmark datasets and **b.** opioids-related datasets at regression setting. **c.** opioids-related datasets at classification setting with different cutoff values (radius: 2; numbits: 2048).



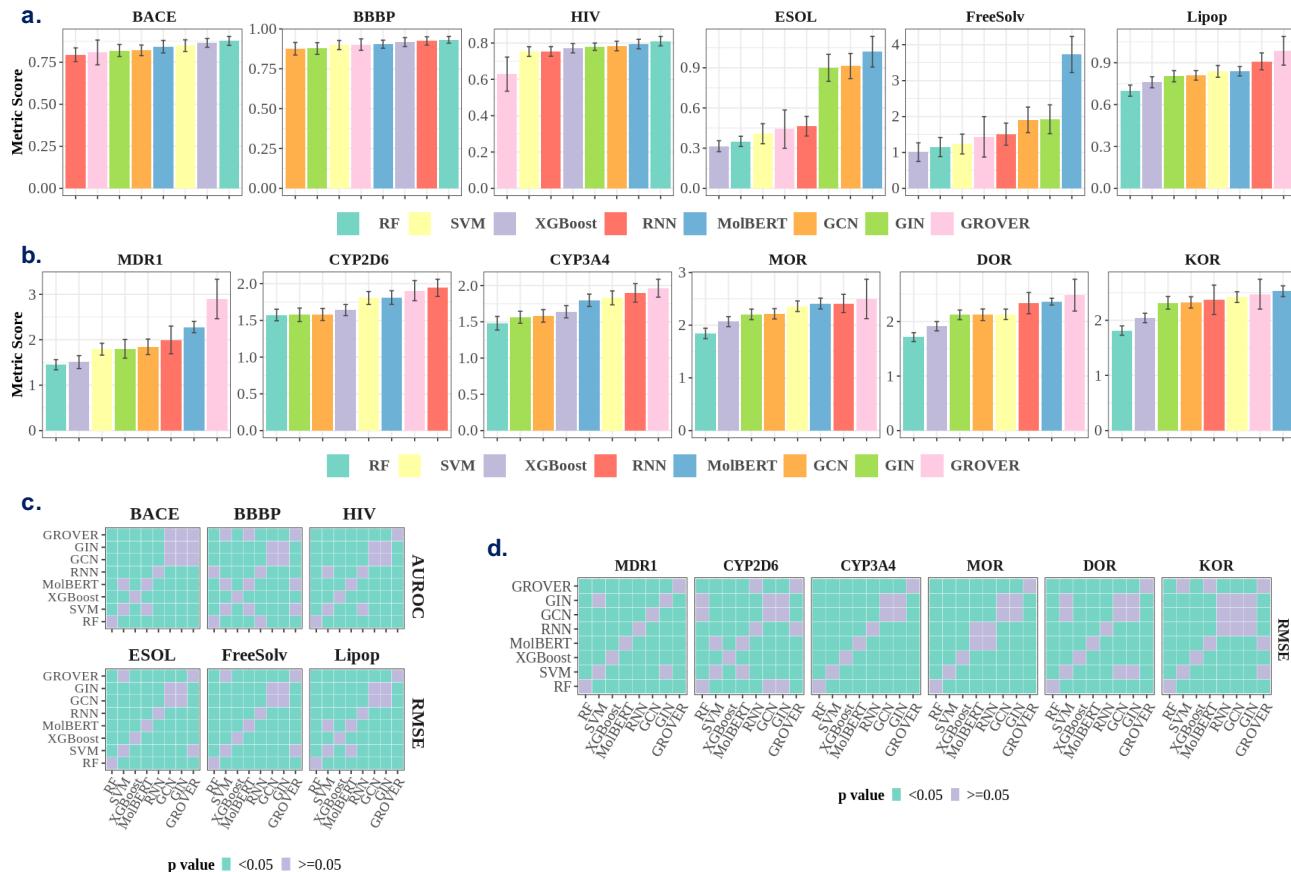
**Figure 13.** Performance Metrics Distribution in MoleculeNet Benchmark Datasets for **a.** RF on RDKit2D descriptors, MolBERT, GROVER and GROVER\_RDKit using recommended metrics **b.** RF on fixed representations and **c.** RF, SVM & XGBoost on RDKit2D descriptors, RNN & MolBERT and GCN & GROVER.



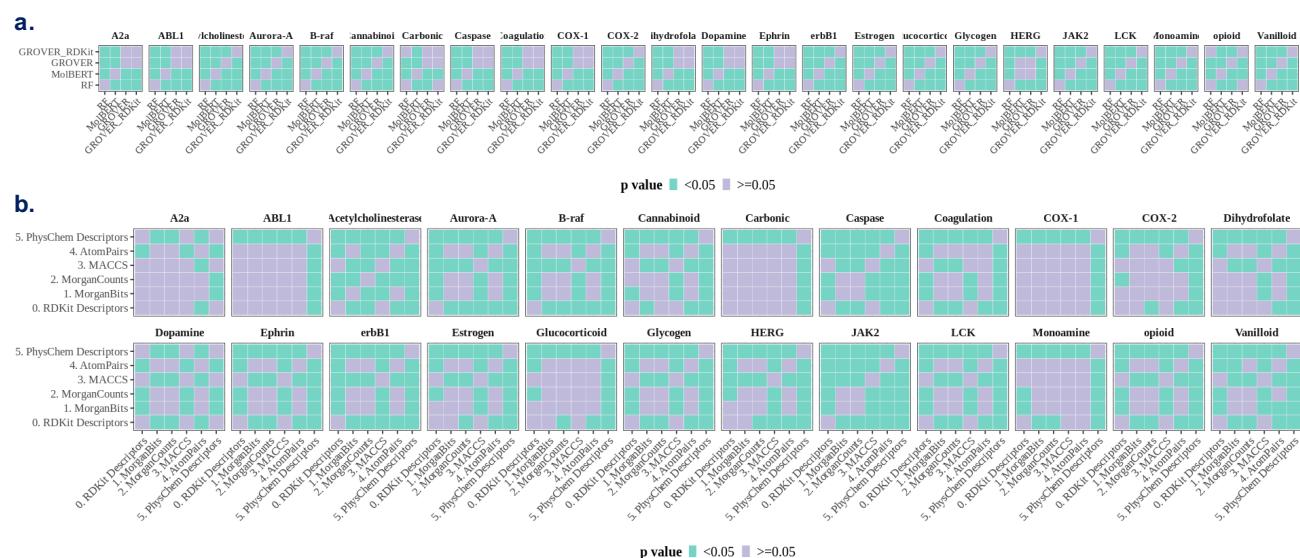
**Figure 14.** Performance Metrics Distribution in Opioids-related Datasets at Regression Setting for **a.** RF on RDKit2D descriptors, MolBERT, GROVER and GROVER\_RDKit using recommended metrics **b.** RF on fixed representations and **c.** RF, SVM & XGBoost on RDKit2D descriptors, RNN & MolBERT and GCN, GIN & GROVER.



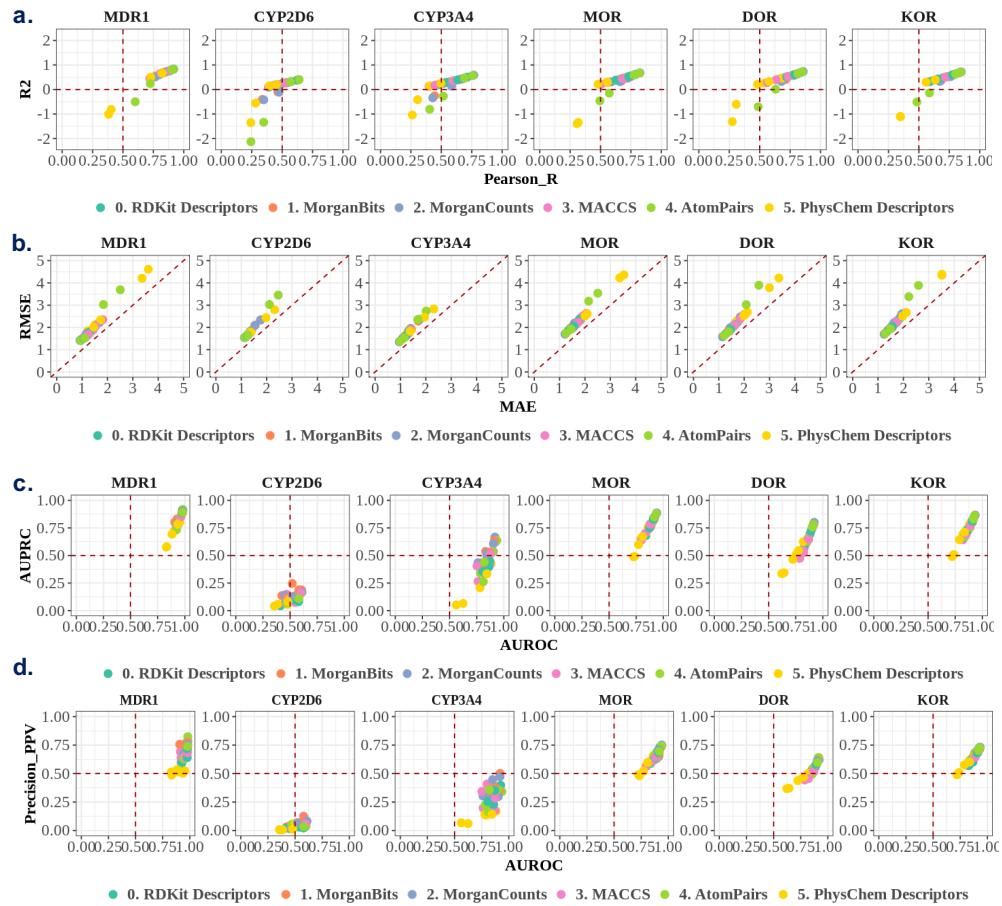
**Figure 15.** Performance Metrics Distribution in Opioids-related Datasets at Classification Setting for **a.** RF on RDKit2D descriptors, MolBERT, GROVER and GROVER\_RDKit using recommended metrics and **b.** RF on fixed representations.



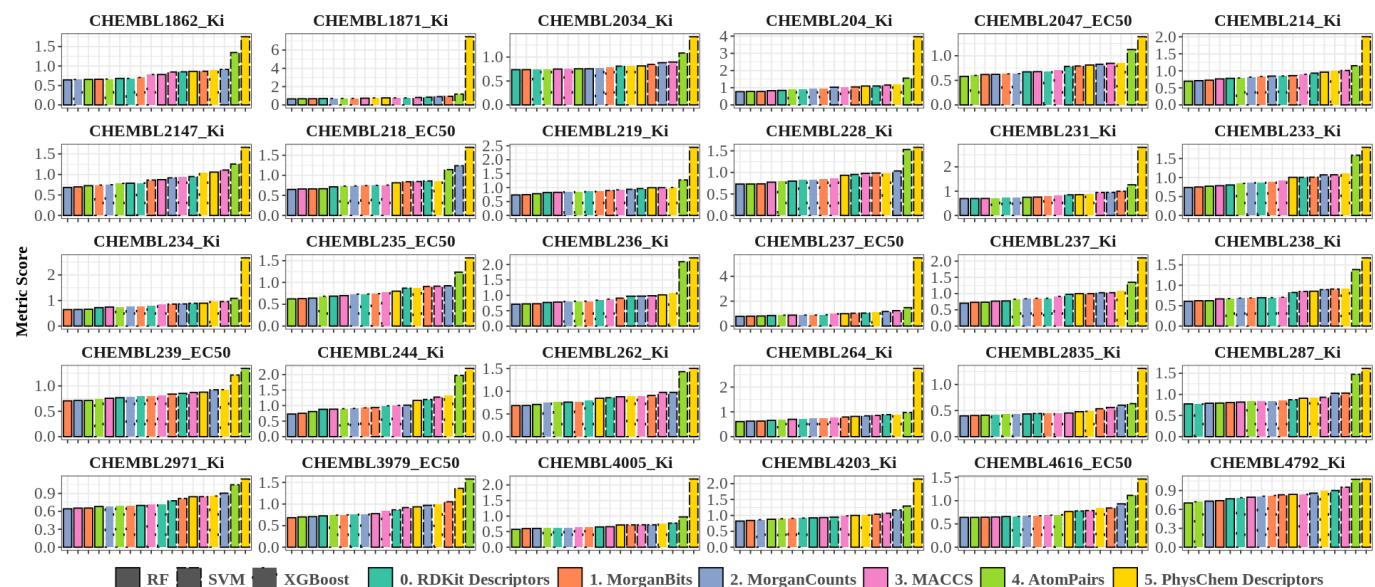
**Figure 16.** Prediction Performance under Random Split **a.** Performance of RF, SVM & XGBoost on RDKit2D descriptors, RNN & MolBERT and GCN, GIN & GROVER in MoleculeNet Benchmark Datasets **b.** Performance of RF, SVM & XGBoost on RDKit2D descriptors, RNN & MolBERT and GCN, GIN & GROVER in Opioids-related Datasets at Regression Setting **c.** Mann-Whitney *U* significance for plot **a** **d.** Mann-Whitney *U* significance for plot **b**



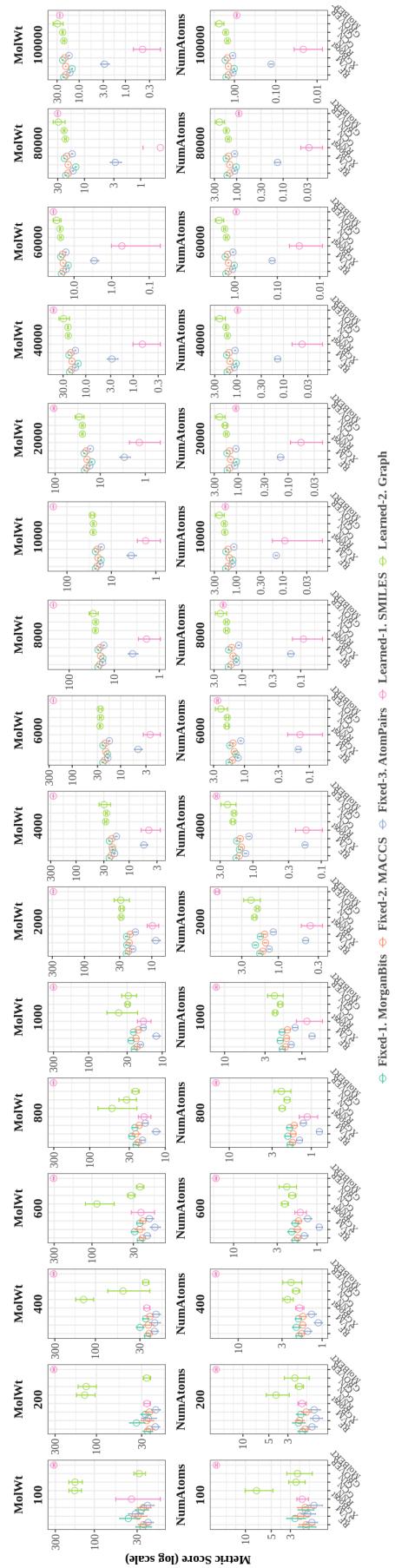
**Figure 17.** Mann-Whitney *U* Significance for Prediction Performance in Bender Datasets under Random Split **a.** Comparison among RF on RDKit2D descriptors, MolBERT, GROVER and GROVER\_RDKit **b.** Comparison among fixed representations



**Figure 18.** Relationship between Common Metrics based on Performance of RF on Fixed Representations in Opioids-related Datasets **a.** R2 vs Pearson\_R **b.** RMSE vs MAE **c.** AUPRC vs AUROC **d.** PPV vs AUROC



**Figure 19.** Prediction Performance in Activity Datasets by Tilborg et al.



Fixed-1. MorganBits ⬤ Fixed-2. MACCS ⬤ Fixed-3. AtomPairs ⬤ Learned-1. SMILES ⬤ Learned-2. Graph ⬤

**Figure 20.** Comparison of prediction performance in descriptors datasets at varying dataset sizes.