

Figure 1. Results of tuning Fred 2.2, Glide 2.3 and Rocs 2.3 on the GSK test set with **8 target systems**. Median area under the ROC curve (AUC) values are shown. Error bars are 95% confidence limits on future (i.e., prospective) performance.

GSK : Relative Performance	
Assertion	P-value
Rocs better than Fred	0.45
Fred better than Glide	0.36
Rocs better than Glide	0.29

DUDD : Absolute Performance