

TABLE S1. ANNOTATIONS FOR DATASETS

Data source	Drug annotation				Protein annotation		Mutation annotation	Label annotation	
	drug name	Profile		Disease		Profile		Profile	Label Type
		PubChem ID	PubChem Fingerprint	Disease name	ICD11	Uniprot ID	Gene Symbol	Mutation name	
CIViC	Yes	No	No	Yes	No	No	Yes	Yes	classified
PharmGKB	Yes	Yes	Yes	Yes	No	No	Yes	Yes	classified
dbMCS	Yes	Yes	No	No	No	No	Yes	Yes	classified
AIMMS	Yes	No	No	No	No	Yes	No	Yes	$\Delta\Delta G$
RET	Yes	No	No	No	No	No	Yes	Yes	$\Delta\Delta G$
GDSC	Yes	Yes	No	No	No	Yes	Yes	Yes	$\Delta\Delta G$
DepMap	Yes	No	No	No	No	No	Yes	Yes	$\Delta\Delta G$

TABLE S2. OVERVIEW OF DATA REPRESENTED IN MDSDR

Property		No. of counts	No. of counts in source datasets
No. of samples		19669	PharmGKB (4828), CIViC (2381), AIMMS (50), DepMap (193), GDSC (8442), RET (56), dbMCS (3719)
No. of unique drugs		2142	PharmGKB (842), CIViC (623), AIMMS (4), DepMap (5), GDSC (234), RET (4), dbMCS (628)
No. of unique genes		1323	PharmGKB (1054), CIViC (256), AIMMS (3), DepMap (5), GDSC (124), RET (1), dbMCS (112)
No. of unique diseases		843	PharmGKB (666), CIViC (181)
No. of mutation	single-point	12716	PharmGKB (1524), CIViC (820), AIMMS (47), DepMap (173), GDSC (7114), RET (52), dbMCS (2986)
	multi-point	6953	PharmGKB (3304), CIViC (1561), AIMMS (3), DepMap (20), GDSC (1328), RET (4), dbMCS (733)



