Supplementary information

DrugRep: an automatic virtual screening server for drug repurposing

Jian-hong Gan[1], Ji-xiang Liu[1,2], Yang Liu[1], Shu-wen Chen[1,3], Wen-tao Dai[2,4], Zhi-Xiong Xiao[1], Yang Cao[1,5*]

[1] Center of Growth, Metabolism and Aging, Key Laboratory of Bio-Resource and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu 610065, China.

[2] Shanghai Engineering Research Center of Pharmaceutical Translation, Shanghai 201203, China.

[3] State Key Laboratory of Biotherapy and Cancer Center, Sichuan University, Chengdu 610065, China.

[4] NHC Key Lab of Reproduction Regulation (Shanghai Institute for Biomedical and Pharmaceutical Technologies), Fudan University, Shanghai 200032, China.

[5] Animal Disease Prevention and Food Safety Key Laboratory of Sichuan Province, Microbiology and Metabolic Engineering Key Laboratory of Sichuan Province, Chengdu 610065, China.

* Corresponding author

Yang Cao (cao@scu.edu.cn)

Table of content

Table S1. Predicted cavities on DUD

DUD ID	Top N cavity	DUD ID	Top N cavity
ACE	1	HIVRT	7
ACHE	1	HMGA	2
ADA	1	HSP90	1
ALR2	1	Inha	1
Ampc	2	MR	1
AR	1	NA	4
CDK2	1	P38	1
COMT	1	PARP	1
COX-1	1	PDE5	1
COX-2	1	PDGFrb	1
DHFR	1	PNP	5
EGFR	3	PPARg	1
ER_agonist	1	PR	1
ER_antagonist	1	RXRa	1
FGFr1	3	SAHH	19
Fxa	1	SRC	5
GART	1	Thrombin	1
GPB	6	TK	1
GR	2	Trpsin	1
HIVPR	1	VEGFr2	2

Table S2. AUC and Enrichment factor on DUD

	DrugRep			X-score		
DUD ID	**EF1**	**EF5**	**AUC**	**EF1**	**EF5**	**AUC**
ACE	5.95	2.02	0.37	9.91	3.24	0.57
ACHE	5.6	4.86	0.67	9.34	2.24	0.57
ADA	0	0	0.5	0	1.01	0.41
ALR2	3.57	6.92	0.72	3.57	0.77	0.55
Ampc	0	0	0.29	0	0	0.4
AR	19.79	11.36	0.8	7.42	4.54	0.79
CDK2	12.19	6.9	0.64	2.71	5.52	0.53
COMT	0	0	0.34	17.38	3.62	0.5
COX-1	14.96	5.57	0.6	7.48	4.77	0.58
COX-2	24.49	10.89	0.79	10.96	6.34	0.85
DHFR	10.7	7.66	0.85	4.86	2.78	0.64
EGFR	1.47	2.23	0.66	2.52	2.02	0.57
ER_agonist	2.91	1.19	0.49	1.46	2.09	0.7
ER_antagonist	0	0	0.31	2.54	3.05	0.6

FGFr1	1.65	0.5	0.28	0	0	0.48
Fxa	8.21	5.06	0.66	2.74	2.19	0.57
GART	0	0	0.78	0	0	0.5
GPB	0	1.53	0.37	5.75	2.3	0.36
GR	7.5	2.55	0.55	20.01	8.93	0.74
HIVPR	6.45	4.19	0.71	9.67	4.84	0.72
HIVRT	13.61	5.97	0.68	13.61	5.05	0.6
HMGA	0	0	0.43	0	3.42	0.64
HSP90	0	0	0.6	0	0	0.59
Inha	17.19	5.8	0.53	16.04	5.1	0.42
MR	12.38	2.63	0.54	0	6.57	0.89
NA	0	0	0.37	7.84	2.83	0.63
P38	2.2	2.69	0.63	3.52	1.81	0.58
PARP	14.13	6.22	0.54	8.48	4.52	0.49
PDE5	8.94	3.84	0.61	7.82	2.26	0.53
PDGFrb	7	2.35	0.45	5.25	2.35	0.4
PNP	0	0.79	0.68	7.89	2.76	0.52
PPARg	4.58	3.28	0.69	4.58	3.75	0.65
PR	0	1.46	0.42	17.96	4.39	0.5
RXRa	0	4.93	0.86	28.84	12.82	0.98
SAHH	26.84	10.89	0.85	0	0	0.42
SRC	1.25	2.26	0.62	3.13	3.14	0.7
Thrombin	2.7	0.83	0.46	5.4	1.38	0.5
TK	0	0	0.47	0	0	0.3
Trpsin	1.94	2.44	0.62	3.88	6.09	0.77
VEGFr2	6.8	4.08	0.58	9.07	3.85	0.55
Average	6.13	3.35	0.58	6.54	3.31	0.58

Table S3. AUC value of the enrichment testing of 7 targets on DUD-E

DUD-E ID	Masters L [1]	DrugRep
ACE	0.54	0.52
CDK2	0.70	0.71
ESR1	0.76	0.62
FA10	0.64	0.79
FABP4	0.83	0.81
GLCM	0.44	0.43
PLK1	0.60	0.60
AVE.	**0.64**	**0.64**

We performed enrichment tests using on another popular dataset of DUD-E. For comparison, we benchmarked DrugRep using the 7 targets (ACE, CSK2, ESR1, FA10, FABP4, GLCM, PLK1), which are also computed by Masters L [1]. Masters L used co-crystallized ligands to determine the docking boxes, which are supposed to be

the most accurate approach. As shown in Table S3, both methods achieve the same average AUC of 0.64, and the differences are less than 0.02 in 5 of the 7 targets. Again, this test indicates that our predicted docking boxes are accurate and our screening process is feasible.

Table S4. Enrichment factor 1 on MUV

ID	targets	Rigid-LS-align	Flexi-LS-align	LiSiCA	FP2	FP4	MACCS	Morgan Fingerprint	FitDock	LigMate
466	**S1P1rec.**	7.30	4.55	**11.50**	6.45	9.50	8.55	8.90	7.00	9.00
548	**PKA**	6.50	6.40	17.50	16.30	8.40	2.50	27.05	9.45	**31.05**
600	**SF1 Inh.**	10.90	11.75	**15.50**	14.55	4.10	2.15	13.05	9.30	14.05
644	**Rho-Kinase2**	6.90	5.80	15.10	**19.15**	9.50	3.90	15.80	4.00	15.95
652	**HIVRT-RNase**	3.85	2.90	7.65	4.75	4.65	5.15	4.20	**7.70**	5.75
689	**Ephrec.A4**	1.20	0.65	5.60	5.25	**6.50**	2.10	3.60	3.75	3.45
692	**SF1 Ago.**	2.20	2.90	**4.50**	2.10	1.25	0.00	0.00	1.10	0.00
712	**HSP 90**	7.80	7.60	**12.85**	7.45	2.35	8.45	9.80	7.30	11.25
713	**ER-α-Coact.Bind.Inh.**	5.00	5.00	4.65	6.85	2.60	6.10	9.80	4.10	**10.15**
733	**ER-β-Coact.Bind.Inh.**	3.60	3.20	**8.30**	4.60	6.55	1.75	4.45	1.25	3.25
737	**ER-α-Coact.Bind.Pot.**	2.65	2.15	4.35	8.40	2.70	2.20	4.65	5.60	**8.85**
810	**FAK**	4.95	4.50	9.30	7.20	11.35	4.70	**13.25**	5.10	10.40
832	**CathepsinG**	23.90	23.15	37.40	**52.65**	29.95	35.80	38.55	14.10	50.55
846	**FXIa**	21.10	21.45	25.75	31.60	22.70	24.20	30.40	22.50	**38.80**
852	**FXIIa**	19.60	22.10	26.95	32.20	26.00	25.15	34.25	9.55	**39.55**
858	**D1rec.**	12.55	12.00	6.80	8.85	10.45	6.40	**15.55**	3.50	9.10
859	**M1rec.**	3.35	3.30	0.00	**3.95**	2.10	3.90	2.25	1.50	1.25
	Avg.	8.43	8.20	12.57	13.66	9.45	8.41	13.86	6.87	**15.44**

The maximum value of each line is bold.

Table S5. Value under the curve on MUV

ID	targets	Rigid-LS-align	Flexi-LS-align	LiSiCA	FP2	FP4	MACCS	Morgan Fingerprint	FitDock	LigMate
466	**S1P1rec.**	0.64	0.65	0.62	0.61	0.59	0.60	0.57	**0.66**	0.62
548	**PKA**	0.66	0.65	0.79	0.74	0.73	0.62	**0.83**	0.61	0.81

600	**SF1 Inh.**	0.67	0.67	0.64	0.63	0.63	0.60	**0.70**	0.66	0.69
644	**Rho-Kinase2**	0.55	0.54	0.72	0.66	0.69	0.63	**0.76**	0.58	0.73
652	**HIVRT-RNase**	0.63	0.63	0.61	0.52	0.54	0.47	0.58	0.60	**0.64**
689	**Ephrec.A4**	0.63	0.63	**0.65**	0.60	0.59	0.58	0.64	0.59	0.60
692	**SF1 Ago.**	0.59	0.60	**0.61**	0.48	0.55	0.48	0.52	0.59	0.55
712	**HSP 90**	0.59	0.58	0.68	0.51	0.54	0.57	0.63	0.63	**0.73**
713	**ER-α-Coact.Bind.Inh.**	**0.60**	0.58	0.55	0.52	0.58	0.44	0.58	0.49	0.55
733	**ER-β-Coact.Bind.Inh.**	0.59	0.59	0.59	0.60	**0.63**	0.52	0.53	0.48	0.58
737	**ER-α-Coact.Bind.Pot.**	**0.71**	0.70	0.67	0.64	**0.71**	0.68	0.69	0.59	**0.71**
810	**FAK**	0.55	0.55	0.64	0.60	0.63	0.59	0.64	0.60	**0.70**
832	**CathepsinG**	0.78	0.78	0.86	0.83	0.84	0.83	0.77	0.65	**0.89**
846	**FXIa**	0.68	0.68	0.84	0.73	0.81	0.70	0.85	0.69	**0.87**
852	**FXIIa**	0.75	0.76	0.81	0.78	0.79	0.78	0.79	0.63	**0.89**
858	**D1rec.**	0.60	0.60	0.65	0.59	0.62	0.55	0.61	0.59	**0.67**
859	**M1rec.**	0.56	0.55	0.48	0.55	0.53	0.58	**0.58**	0.52	0.57
	Avg.	0.63	0.63	0.67	0.62	0.65	0.60	0.66	0.60	**0.69**

The maximum value of each line is bold.

Table S6. Screened drugs by Ligand-based Screen for COX-2

DrugBank ID	FP2	FP4	Morgan fingerprint	Rigid-LS-align	Flexi-LS-align	LigMate	FitDock
DB00482	0.941	0.866	0.844	0.866	0.857	0.971	0.971
DB00580		0.588	0.4	0.705		0.136	0.845
DB08439		0.473	0.297	0.731	0.71		0.688
DB01628			0.327	0.727	0.696	0.041	0.027
DB00533			0.317	0.727	0.665		0.148

These values represent similarity scores.

Null value indicates that the method does not screen the corresponding drug.

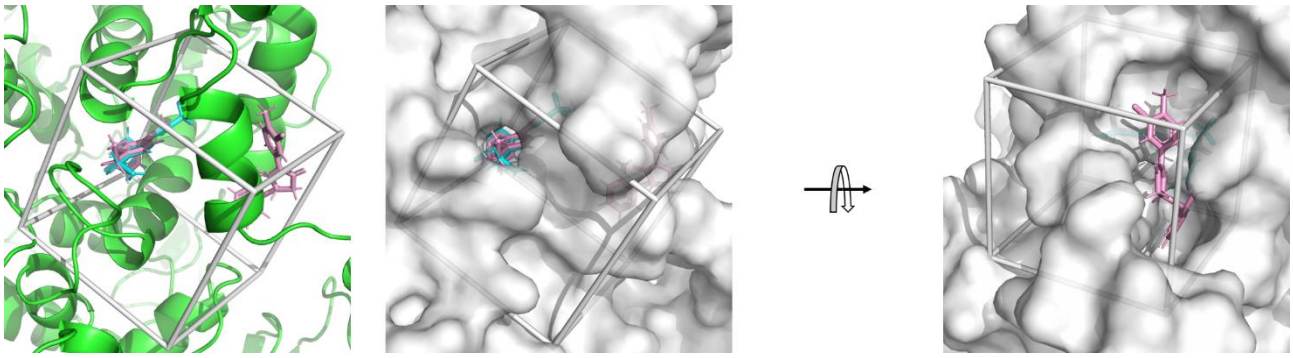

Figure S1. Predicted docking box (gray) and docking for COX2. The crystal ligand is marked in cyan. The ligands from DUD dataset are marked in pink. The results showed that the docking sites of some ligands deviate from the real binding sites due to the large docking box.

References

1. Masters L, Eagon S, Heying M. Evaluation of consensus scoring methods for AutoDock Vina, smina and idock. Journal of Molecular Graphics and Modelling 2020; 96: 107532.