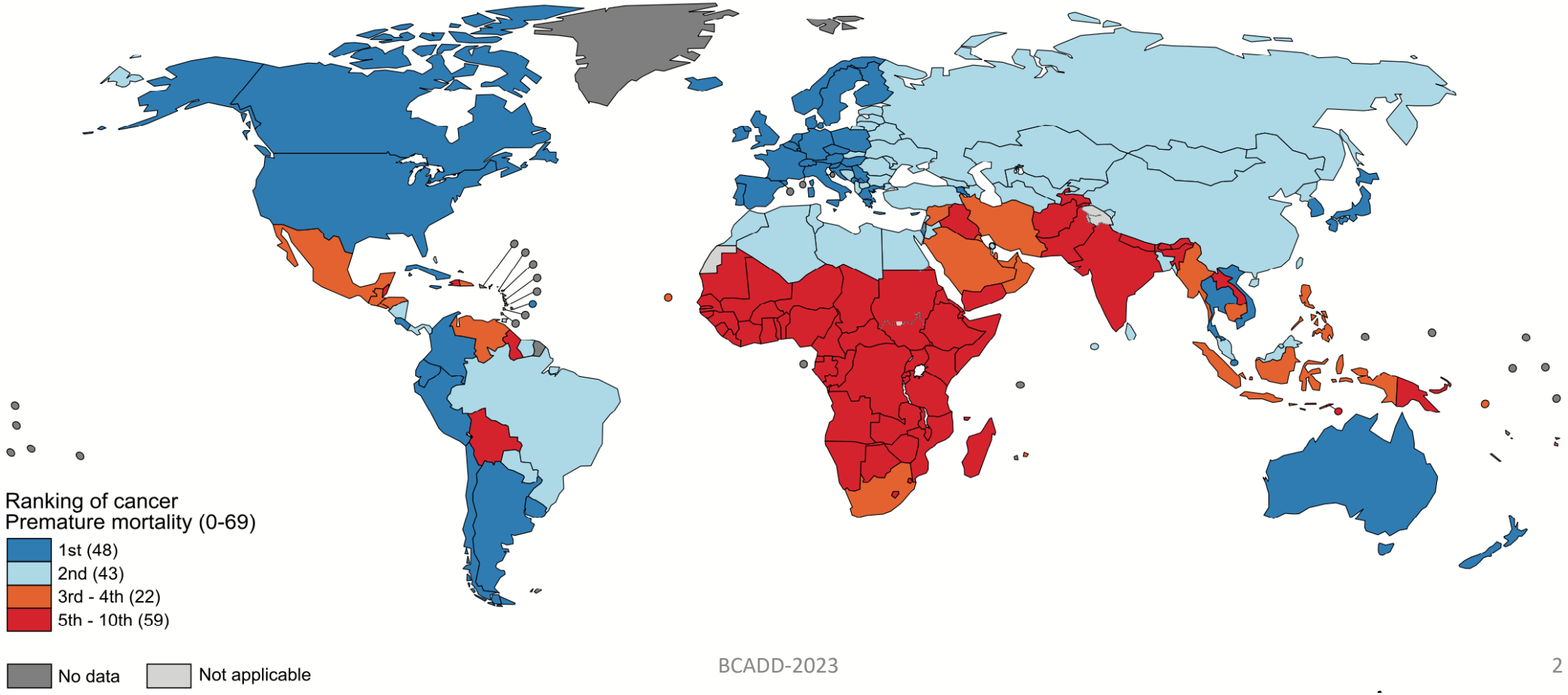
In silico prediction of cell-lines cytotoxicity of drug-like compounds based on their structural formula

Alexey Lagunin

Institute of Biomedical Chemistry, Moscow, Russia;
Pirogov Russian National Research Medical University, Moscow, Russia.

Oncological diseases

Bray, F., et al. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin, 2018, 68, 394–424



Cell-line cytotoxicity studies

- More 1000 cell-lines are used in the studies;
- It accelerates the process of drug discovery;
- It estimates the efficacy of drug-candidates against tumor cell-lines;
- It estimates toxicity in normal cell-lines;
- It allows selecting the most effective and safe compounds;
- In spite of introduction the first panel of cell-lines - NCI60 (several dozens years ago) only several hundreds thousand of compounds have been tested.

Three web-applications for cell-line cytotoxicity prediction

- **CLC-Pred (CLC-Pred 2.0)** - Cell-Line Cytotoxicity Predictor for *in silico* prediction of human cell line cytotoxicity for drug-like compounds;
- **BC CLC-Pred** – Quantitative and qualitative prediction of cytotoxicity for drug-like compounds in 9 Breast Cancer cell-lines;
- **CLC-Pred Synergy** - Prediction of synergistic cytotoxicity of drug pairs in 34 NCI60 cell lines.

CLC-Pred (Cell-Line Cytotoxicity Predictor) was introduced in 2018

- It used PASS (MNA descriptors and Bayesian-like algorithm) for SAR modeling;
- It predicts 278 tumor and 27 non tumor cell lines with mean accuracy of prediction (AUC LOO CV) more than 0.93;
- **59,882** unique structures;
- More 100 citations.

https://www.way2drug.com/cell-line/

International Journal of Molecular Sciences

Open Access Article

CLC-Pred 2.0: A Freely Available Web Application for In Silico Prediction of Human Cell Line Cytotoxicity and Molecular Mechanisms of Action for Druglike Compounds

by Alexey A. Lagunin [1,2,*], Anastasia V. Rudik [1], Pavel V. Pogodin [1], Polina I. Savosina [1], Olga A. Tarasova [1], Alexander V. Dmitriev [1], Sergey M. Ivanov [1,2], Nadezhda Y. Biziukova [1], Dmitry S. Druzhilovskiy [1], Dmitry A. Filimonov [1] and Vladimir V. Poroikov [1,*]

[1] Department of Bioinformatics, Institute of Biomedical Chemistry, 119435 Moscow, Russia

[2] Department of Bioinformatics, Pirogov Russian National Research Medical University, 117997 Moscow, Russia

* Authors to whom correspondence should be addressed.

Int. J. Mol. Sci. **2023**, *24*(2), 1689; **https://doi.org/10.3390/ijms24021689**

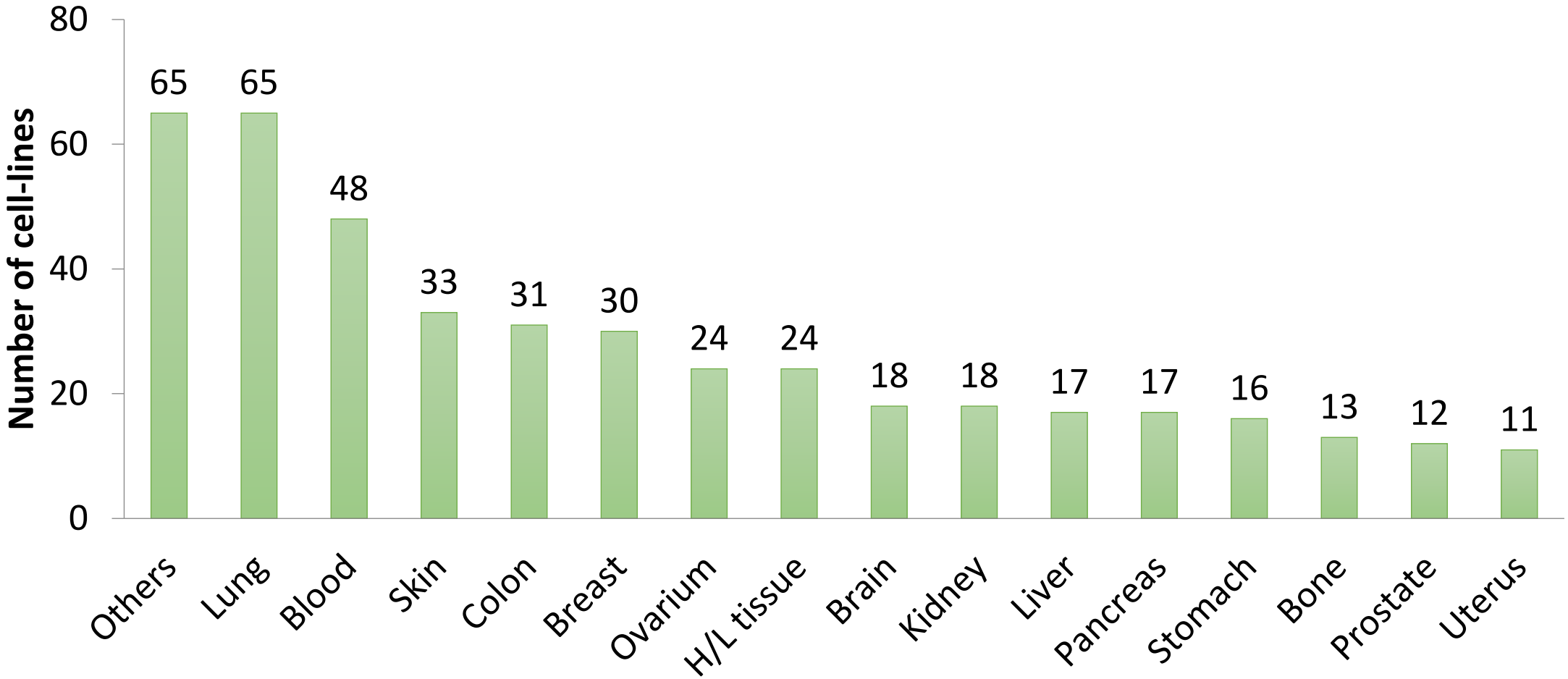
Received: 1 December 2022 / Revised: 4 January 2023 / Accepted: 12 January 2023 / Published: 14 January 2023

(This article belongs to the Section **Molecular Toxicology**)

ChEMBL (v. 31) and PubChem (February 2022) data cell-line cytotoxicity

- **128,545** unique structures of compounds tested against 1162 cell-lines;
- **379,767** experimental values (IG50, IC50, and % inhibition);
- **438** human cell-lines (391 tumor and 47 non tumor cell-lines) with the accuracy of prediction (AUC LOO CV) higher than 0.8 were selected after the PASS training;
- 10,000 nM (or 50%) is a threshold between active and inactive compounds.

ChEMBL data.
Distribution of 438 cell lines by organs and tissue

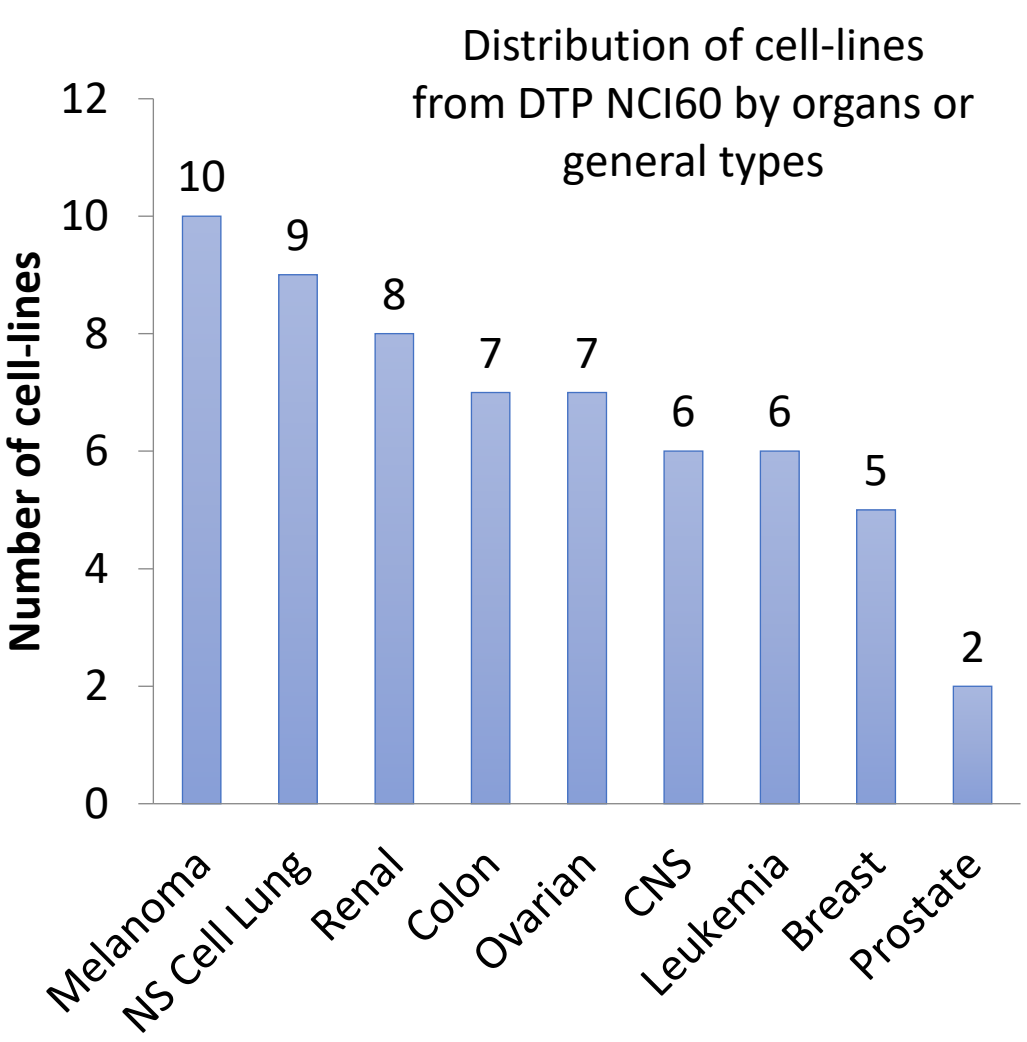


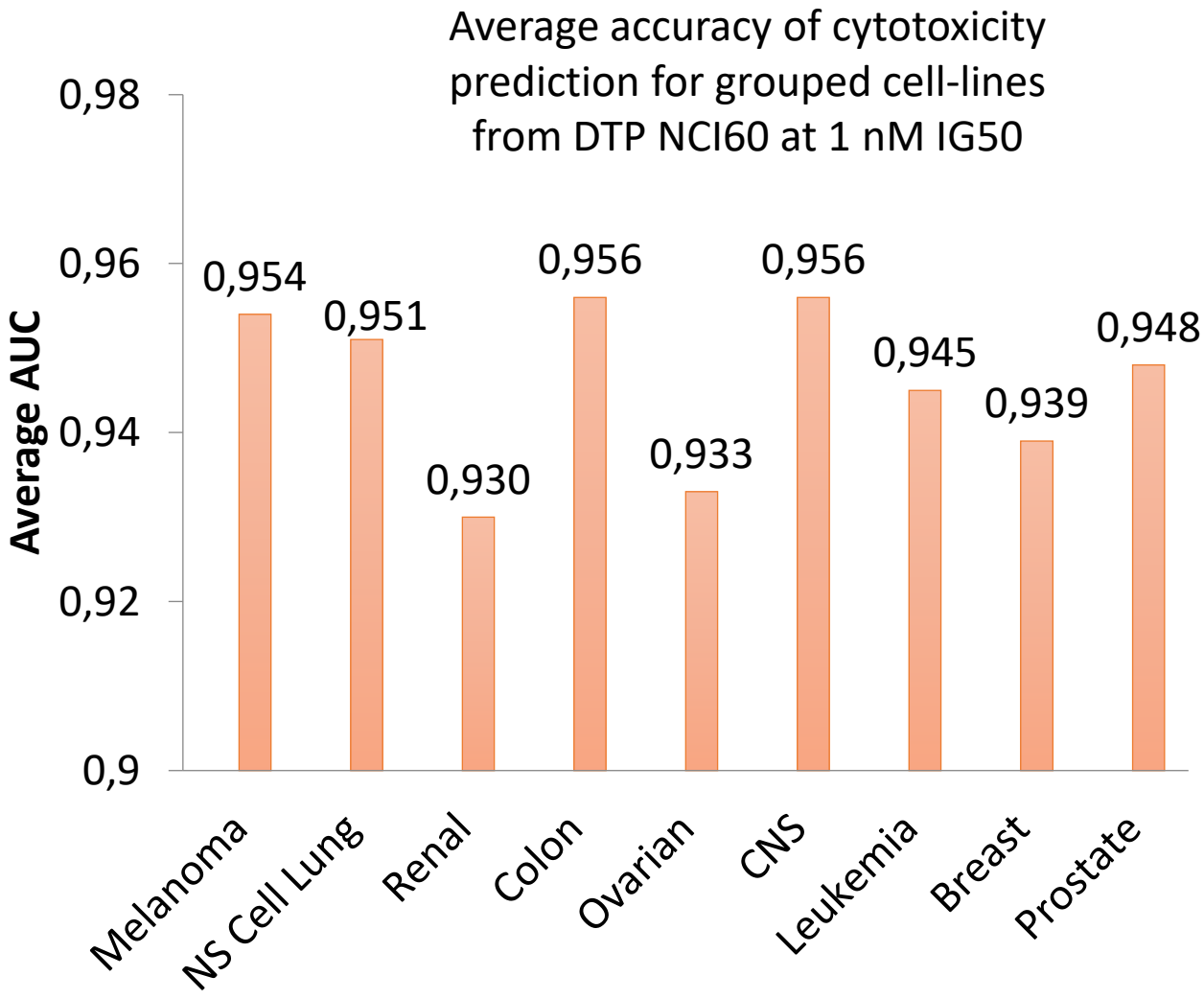
Developmental Therapeutics Program (DTP) NCI60 data

- 22,726 unique structures tested against NCI60 cell lines.
- 1,262,878 experimental GI50 values measured by the same protocol.

https://dtp.cancer.gov/

DTP NCI60 data

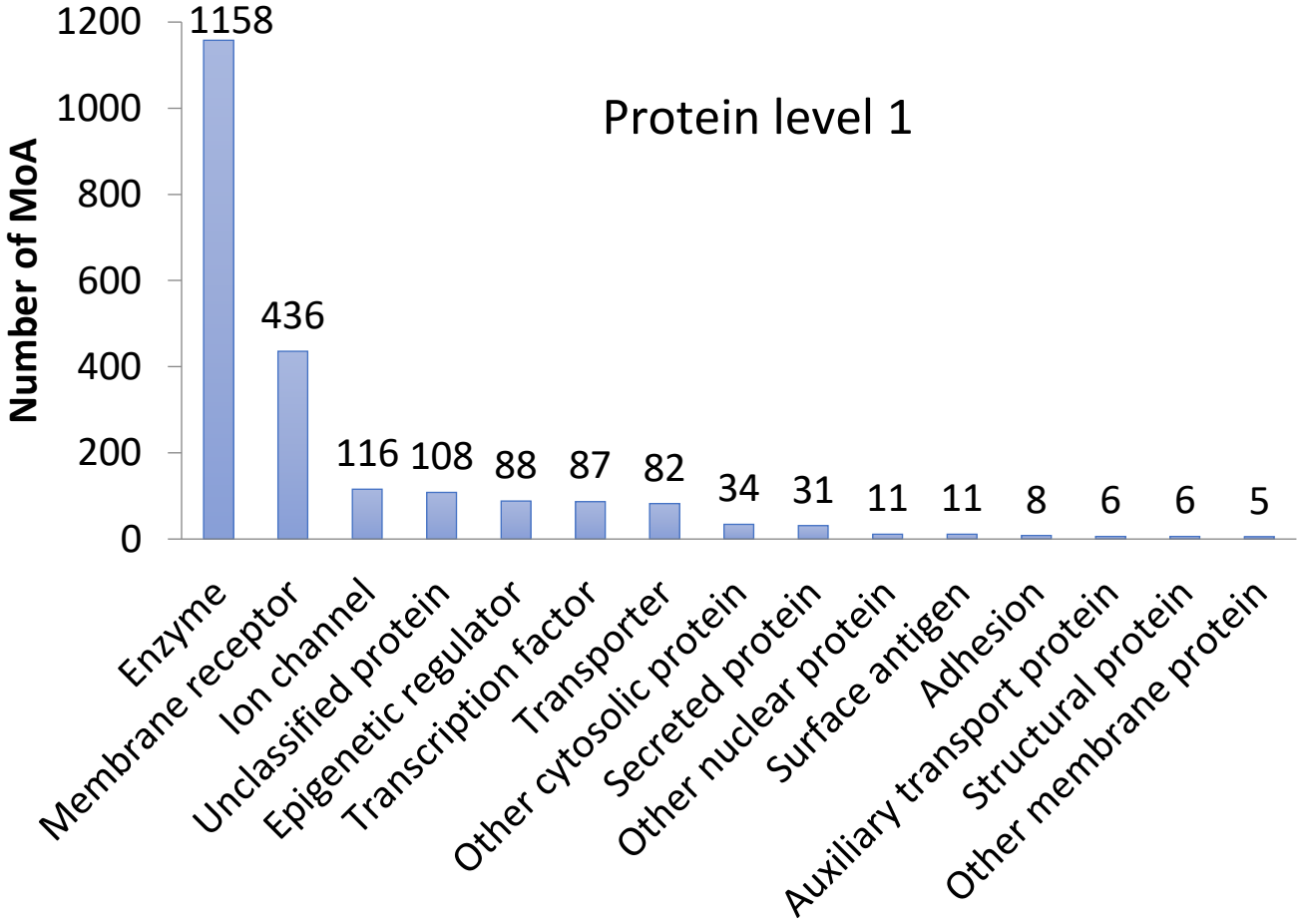


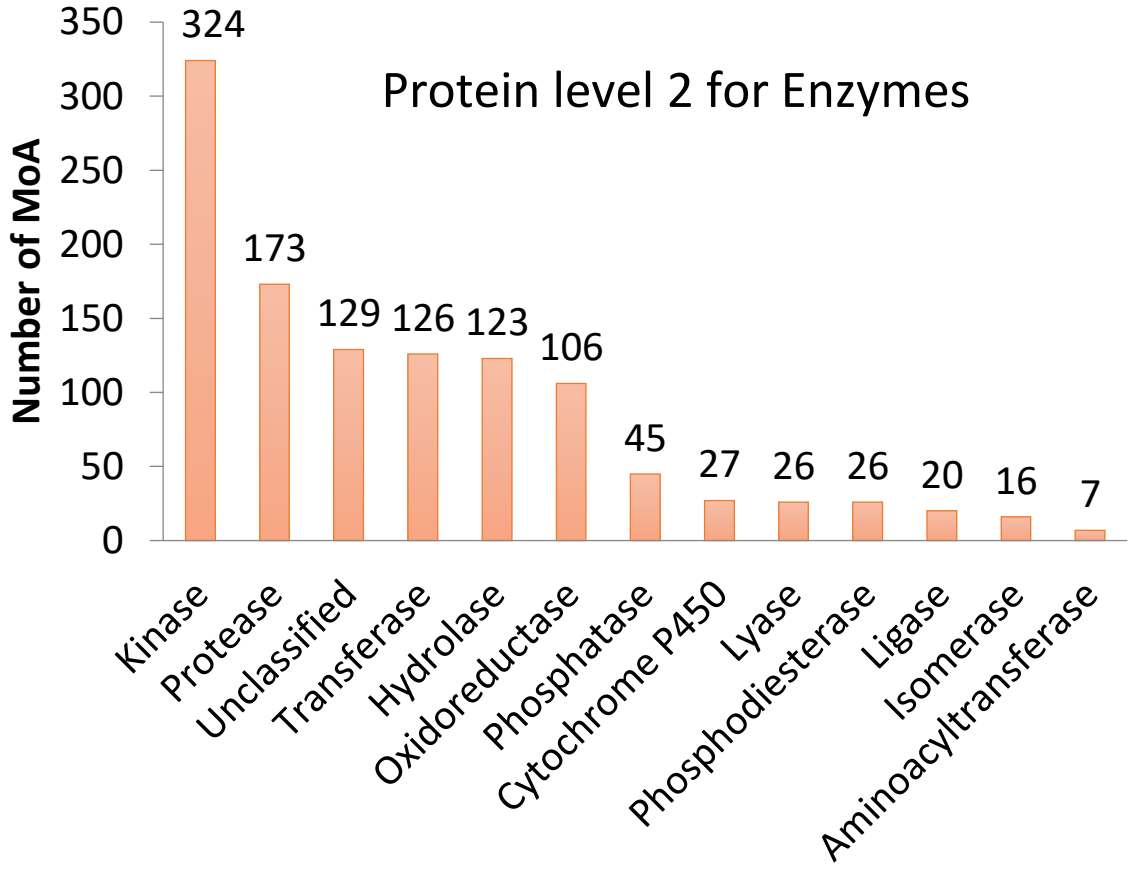


ChEMBL and PubChem data on mechanisms of action

- **656,011** unique structures;
- **957,545** records with experimental values;
- **2,170** molecular mechanisms of action with the accuracy of prediction (AUC LOO CV) higher than 0.8 were selected after the PASS training.

Number of predicted MoA classifying by ChEMBL protein classification





CLC-Pred prediction results of general cell line cytotoxicity for erlotinib

SMILES

Use Files

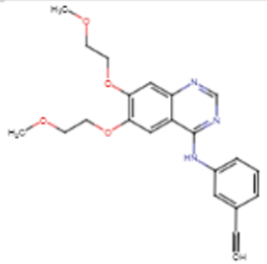
Use drug name

Marvin Molecular Editor

Erlotinib

Make prediction

Examples: Ampicillin



Cell-Line | TDP NCI-60 (1nM) | TDP NCI-60 (10nM) | TDP NCI-60 (100nM) | Target

Copy | Excel | CSV | PDF | Print

Pa	Pi	Cell-line	Description	Tissue/Organ	Type	IAP*
0.825	0.004	A-431	Epidermoid carcinoma	Skin	Carcinoma	0.937
0.714	0.004	NCI-H358	Bronchioalveolar Carcinoma	Lung; Bronchiole	Carcinoma	0.805
0.685	0.005	SK-LU-1	Adenocarcinoma	Lung	Carcinoma	0.810
0.502	0.024	EKVX	Non-small cell lung carcinoma	Lung	Carcinoma	0.841
0.499	0.013	CL97	Lung adenocarcinoma	Lung	Adenocarcinoma	0.872
0.470	0.004	SK-HEP1	Hepatocellular carcinoma	Liver	Carcinoma	0.912
0.461	0.032	ATH-8	HTLV-I-infected human T-cell line	Peripheral blood	Normal	0.964
0.451	0.029	OS-RC-2	Clear cell renal cell carcinoma	Kidney	Carcinoma	0.812

Pi | Search C

Showing 1 to 10 of 86 entries

Previous 1 2 3 4 5 ... 9 Next

* - IAP: Invariant Accuracy of Prediction (equal to AUC value) was calculated by leave-one-out cross-validation procedure

CLC-Pred 2.0 results of cytotoxicity prediction based on TDP NCI60 data for erlotinib at (a) at 1 nM, (b) at 10 nM.

(a)

Cell-Line | TDP NCI-60 (1nM) | TDP NCI-60 (10nM) | TDP NCI-60 (100nM) | Target

Copy | Excel | CSV | PDF | Print

Pa	Pi	DTP NCI60 cell-line	Tissue of origin	Sex	Epithelial	Histology	Ploidy	p53	IAP*
0.191	0.145	RE:TK-10	Renal	M	yes	Renal Spindle cell carcinoma	4n, Tetraploid (92)	MT	0.912

(b)

Cell-Line | TDP NCI-60 (1nM) | TDP NCI-60 (10nM) | TDP NCI-60 (100nM) | Target

Copy | Excel | CSV | PDF | Print

Pa	Pi	DTP NCI60 cell-line	Tissue of origin	Sex	Epithelial	Histology	Ploidy	p53	IAP*
0.361	0.060	RE:TK-10	Renal	M	yes	Renal Spindle cell carcinoma	4n, Tetraploid (92)	MT	0.912
0.264	0.150	RE:UO-31	Renal	F	yes	Renal cell carcinoma-vpd	2n+/-, Near-diploid 46+/- (35-57)	WT	0.854
0.235	0.143	OV:SK-OV-3	Ovarian	F	yes	Adenocarcinoma-vpd	4n+/-, Near-tetraploid 92+/- (81-103)	NA	0.909

CLC-Pred 2.0 results of cytotoxicity prediction based on TDP NCI60 data for erlotinib at 100 nM.

Cell-Line | TDP NCI-60 (1nM) | TDP NCI-60 (10nM) | TDP NCI-60 (100nM) | Target

Copy | Excel | CSV | PDF | Print

Pa	Pi	DTP NCI60 cell-line	Tissue of origin	Sex	Epithelial	Histology	Ploidy	p53	IAP*
0.344	0.098	OV:IGROV1	Ovarian	F	yes	Cystoadenocarcinoma-pd	4n+/-, Near-tetraploid 92+/- (81-103)	MT	0.871
0.322	0.087	BR:T-47D	Breast	F	yes	infiltrating ductal carcinoma	2n+, Hyperdiploid (47-57)	MT	0.874
0.293	0.189	LC:HOP-92	Non-Small Cell Lung	M	yes	Large cell-ud	4n+/-, Near-tetraploid 92+/- (81-103)	MT	0.829
0.267	0.177	RE:UO-31	Renal	F	yes	Renal cell carcinoma-vpd	2n+/-, Near-diploid 46+/- (35-57)	WT	0.853

CLC-Pred 2.0 with prediction results of molecular mechanisms of action for erlotinib

Cell-Line | TDP NCI-60 (1nM) | TDP NCI-60 (10nM) | TDP NCI-60 (100nM) | Target

Copy | Excel | CSV | PDF | Print

Pa	Pi	Mechanism of action	UniProt	ChEMBL Protein family		IAP*
				Level 1	Level 2	
0.882	0.003	Epidermal growth factor receptor antagonist	P00533	Enzyme	Kinase	0.975
0.710	0.001	Cyclin-G-associated kinase inhibitor	O14976	Enzyme	Kinase	0.968
0.704	0.003	Receptor tyrosine-protein kinase erbB-2 antagonist	P04626	Enzyme	Kinase	0.986
0.623	0.005	Vascular endothelial growth factor receptor 2 antagonist	P35968	Enzyme	Kinase	0.974
0.438	0.044	Bile salt export pump inhibitor	O95342	Transporter	Primary active transporter	0.838
		Platelet-derived growth factor receptor				
	Pi	Search	Search			

Showing 1 to 10 of 70 entries

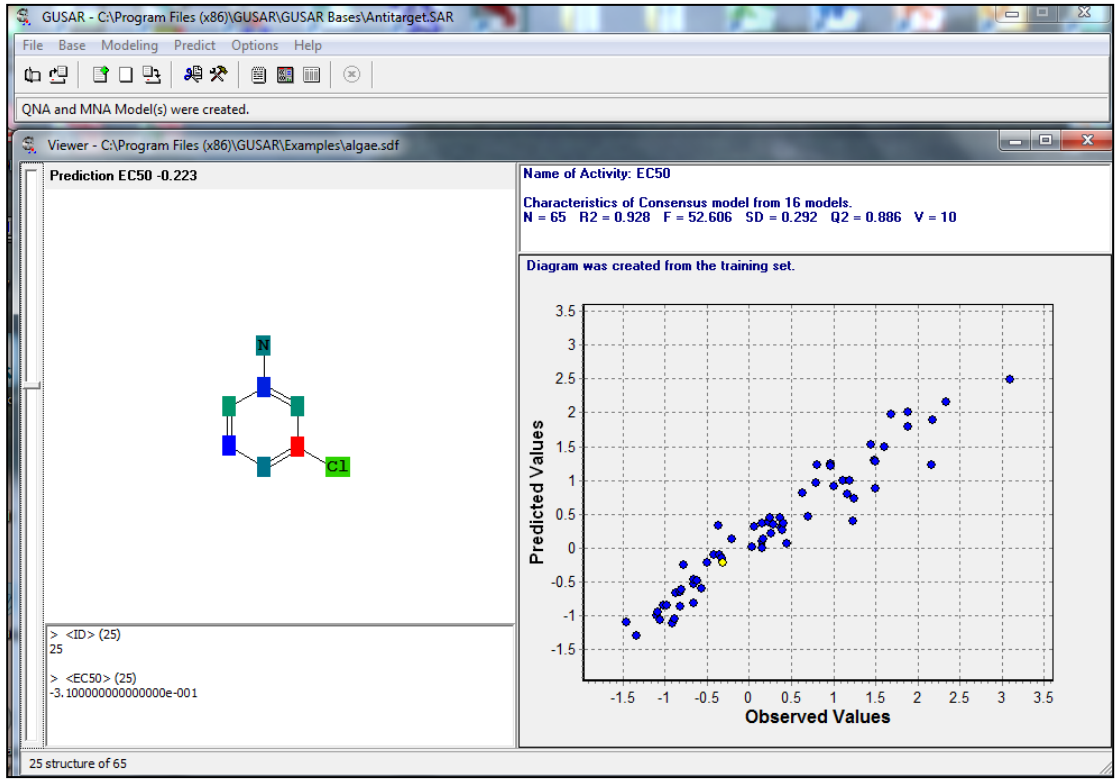
Previous 1 2 3 4 5 6 7 Next

* - IAP: Invariant Accuracy of Prediction (equal to AUC value) was calculated by leave-one-out cross-validation procedure

ChEMBL data for Breast Cancer cell-lines

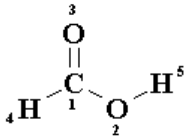
	GI50			IC50		
	pGI50	Active	Not active	pIC50	Active	Not active
T47D	[3.894:10.7]	369	1140	[1.567:10.096]	440	1263
ZR-75-1	[3.99:8.698]	100	83	[4.125:9.619]	73	40
MX1	–	–	–	[3.57:8.72]	118	151
Hs-578T	[3.88:9.95]	105	364	[1.367:8.866]	41	142
MCF7-DOX	–	–	–	[4.0:8.638]	11	27
MCF7	[3.585:11.30]	1635	3348	[0.022:13.69]	5213	19998
Bcap37	–	–	–	[3.05:6.568]	5	267
MCF7R	–	–	–	[4.17:8.79]	67	11
BT-20	[4.167:8.097]	7	21	[1.816:8.523]	32	145

GUSAR - General Unrestricted Structure-Activity Relationships



Filimonov D.A., et al. (2009). *SAR and QSAR Environ. Res.*, 20 (7-8), 679-709

QNA: Quantitative Neighborhoods of Atoms descriptors



a)

$$C = \begin{vmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{vmatrix}$$

b)

$$\mathrm{Exp}\left(-\tfrac{1}{2}C\right) = \begin{vmatrix} 1.40 & -0.59 & -0.57 & -0.57 & 0.14 \\ -0.59 & 1.27 & 0.14 & 0.14 & -0.54 \\ -0.57 & 0.14 & 1.13 & 0.13 & -0.02 \\ -0.57 & 0.14 & 0.13 & 1.13 & -0.02 \\ 0.14 & -0.54 & -0.02 & -0.02 & 1.13 \end{vmatrix}$$

c)

d)

	EA	IP	*A*	*B*	P	Q
C	1.263	11.26	6.262	0.316	-0.00218	-0.1820
O	1.461	13.62	7.541	0.287	0.02944	0.3019
O	1.461	13.62	7.541	0.287	0.06199	0.5297
H	0.754	13.60	7.177	0.279	0.05812	0.4706
H	0.754	13.60	7.177	0.279	0.05304	0.3533

(a) structural formula;
(b) connectivity matrix;
(c) exponent of the connectivity matrix;
(d) electron affinities (EA), ionization potentials (IP), parameters A and B, P and Q values for each of the atoms of *formic acid* molecule.

Characteristics of (Q)SAR models created by GUSAR

Name		QSAR				SAR			
		N	R^2	Q^2	SD	N	Sens	Spec	Bal.Acc.
T47D	GI50	1497	0.711	0.613	0.570	1497	0.853	0.865	0.859
	IC50	1683	0.767	0.673	0.562	1684	0.835	0.819	0.827
ZR-75-1	GI50	183	0.804	0.723	0.462	183	0.890	0.892	0.891
	IC50	113	0.784	0.719	0.673	113	0.904	0.875	0.890
MX-1	IC50	268	0.856	0.786	0.457	268	0.863	0.841	0.852
Hs-578T	GI50	–	–	–	–	466	0.825	0.826	0.826
	IC50	–	–	–	–	177	0.744	0.891	0.817
MCF7-DOX	IC50	38	0.904	0.825	0.617	38	1.000	0.963	0.981
MCF7	GI50	4953	0.640	0.581	0.749	4953	0.866	0.809	0.837
	IC50	25077	0.521	0.496	0.812	25077	0.771	0.821	0.796
Bcap37	IC50	–	–	–	–	272	1.000	0.974	0.987
MCF7R	IC50	78	0.797	0.732	0.567	78	1.000	0.909	0.955
BT-20	GI50	28	0.912	0.821	0.464	28	1.000	1.000	1.000
	IC50	–	–	–	–	176	0.844	0.854	0.849

Accuracy of prediction during 5-fold cross-validation

Name		N	QSAR		SAR
			R^2	RMSE	Bal.Accuracy
T47D	GI50	1497	0.498	0.638	0.82
	IC50	1683	0.518	0.674	0.77
ZR-75-1	GI50	183	0.56	0.576	0.86
	IC50	113	0.655	0.733	0.82
MX-1	IC50	268	0.627	0.587	0.80
Hs-578T	GI50	466	0.357	0.702	0.76
	IC50	177	0.046	1.336	0.76
MCF7-DOX	IC50	38	0.557	0.954	0.92
MCF7	GI50	4953	0.588	0.738	0.84
	IC50	25077	0.671	0.649	0.88
Bcap37	IC50	272	0.387	0.336	0.97
MCF7R	IC50	78	0.672	0.625	0.98
BT-20	GI50	28	0.641	0.612	0.85
	IC50	176	0.287	1.122	0.77

Way2Drug
Understanding Chemical-Biological Interactions

PREDICTIVE SERVICES

BC CLC-Pred: Breast cancer cell-line cytotoxicity prediction

» Home | » Training Set | » Products/Services | » Interpretation | » Contacts

https://www.way2drug.com/bc/

SMILES

Use Files

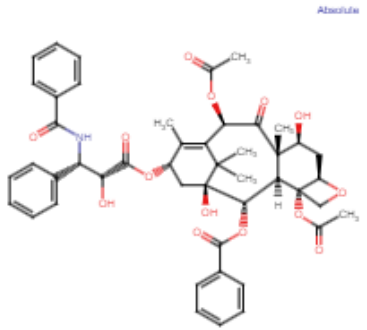
Use drug name

Marvin Molecular Editor

Paclitaxel

Make prediction

Examples:Ampicillin

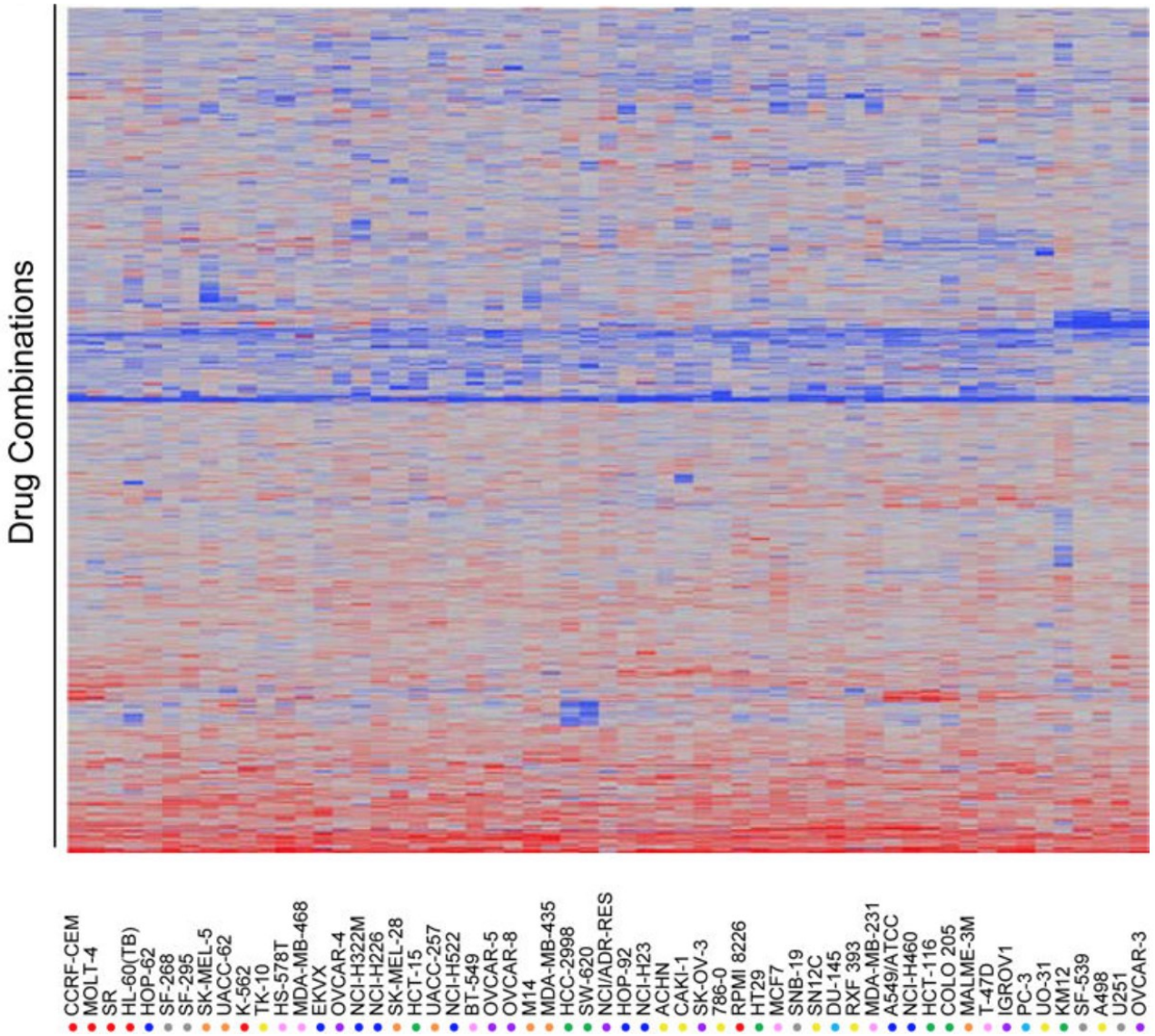


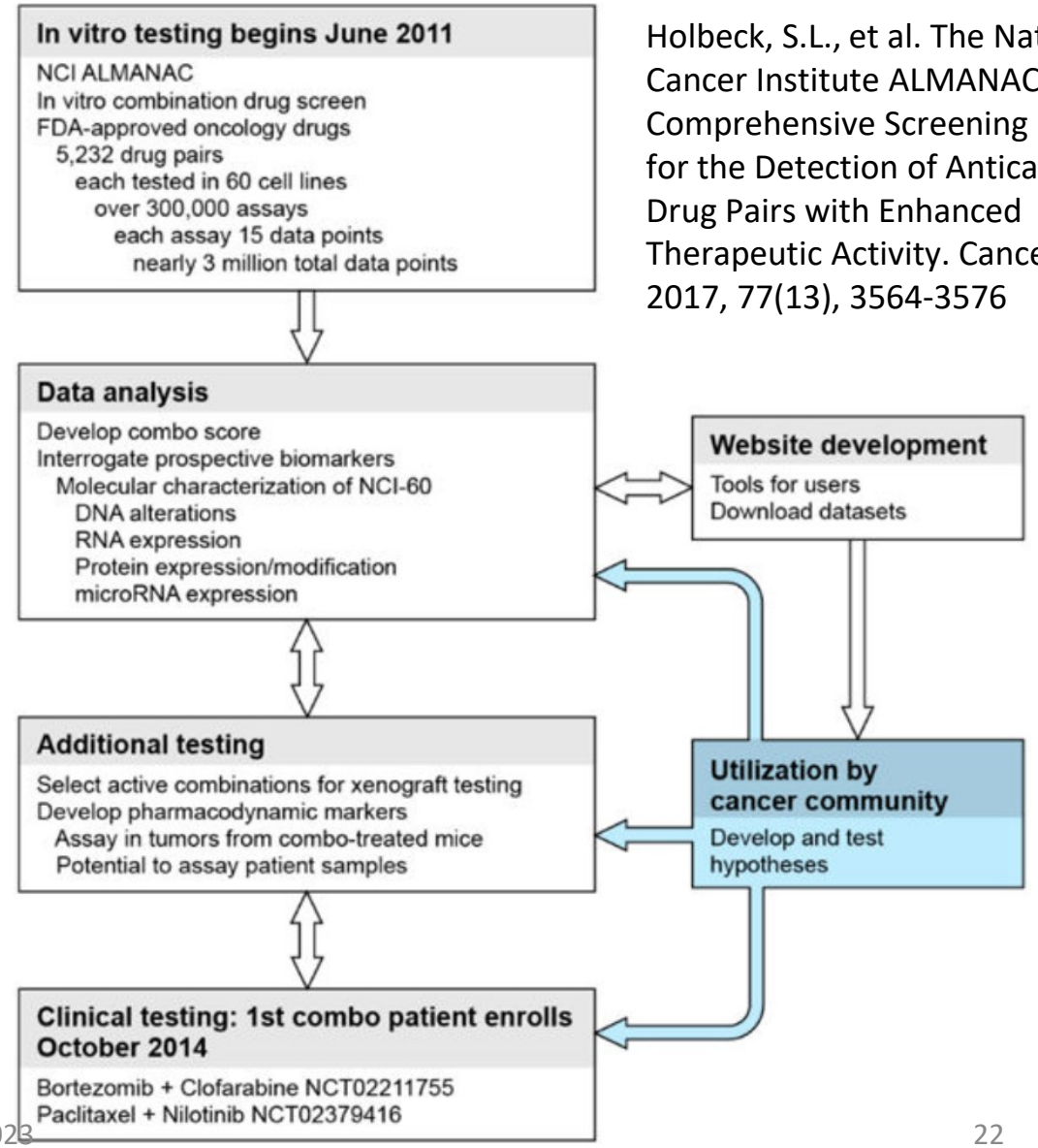
PDF | Copy | CSV | Excel

Search:

Name	Value	Applicability domain
Bcap37 (IC50) Class	active	in AD
BT-20 (GI50) Class	inactive	in AD
BT-20 (IC50) Class	active	in AD
BT-20 (pGI50)	5.3027	in AD
Hs-578T (GI50) Class	active	in AD
Hs-578T (IC50) Class	active	in AD
MCF7 (GI50) Class	active	in AD
MCF7 (IC50) Class	active	in AD
MCF7 (pGI50)	7.7512	in AD

NCI ALMANAC database



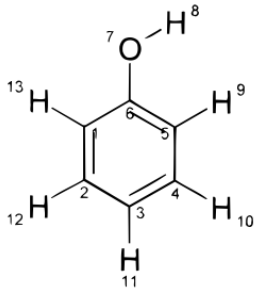


Holbeck, S.L., et al. The National Cancer Institute ALMANAC: A Comprehensive Screening Resource for the Detection of Anticancer Drug Pairs with Enhanced Therapeutic Activity. Cancer Res. 2017, 77(13), 3564-3576

MNA Descriptors

In PASS

MNA (Multilevel Neighborhoods of Atoms)

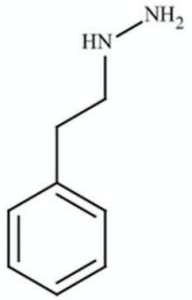


Phenol as a set of MNA descriptors of zero, first and second levels (MNA/0, MNA/1, MNA/2)

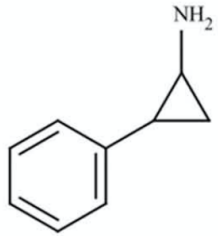
atom	MNA/0	MNA/1	MNA/2
1	C	C(CC-H)	C(C(CC-H)C(CC-O)-H(C))
2	C	C(CC-H)	C(C(CC-H)C(CC-H)-H(C))
3	C	C(CC-H)	C(C(CC-H)C(CC-H)-H(C))
4	C	C(CC-H)	C(C(CC-H)C(CC-H)-H(C))
5	C	C(CC-H)	C(C(CC-H)C(CC-O)-H(C))
6	C	C(CC-O)	C(C(CC-H)C(CC-H)-O(C-H))
7	-O	-O(C-H)	-O(C(CC-O)-H(-O))
8	-H	-H(-O)	-H(-O(C-H))
9	-H	-H(C)	-H(C(CC-H))
10	-H	-H(C)	-H(C(CC-H))
11	-H	-H(C)	-H(C(CC-H))
12	-H	-H(C)	-H(C(CC-H))
13	-H	-H(C)	-H(C(CC-H))

Filimonov, D., et al. Chemical Similarity Assessment through Multilevel Neighborhoods of Atoms: Definition and Comparison with the Other Descriptors. J. Chem. Inf. Comput. Sci., 1999, 39, 666-670.

PoSMNA (Pair of Substances MNA) in **PASS DDI**



MNA/2 descriptors for the combination of phenelzine and tranylcypromine



C(C(CC-H)C(CC-H)-H(C))
C(C(CC-H)C(CC-H)-C(C-H-H-C))
C(C(CC-H)C(CC-C)-H(C))
-C(C(CC-C)-H(-C)-H(-C)-C(-H-H-C-N))
-C(-H(-C)-H(-C)-C(C-H-H-C)-N(-H-C-N))
-N(-H(-N)-H(-N)-N(-H-C-N))
-N(-H(-N)-C(-H-H-C-N)-N(-H-H-N))

C(C(CCC-H)C(CC-H-H)-H(C)-N(C-H-H))
C(C(CCC-H)C(CC-H-N)-H(C)-H(C))
C(C(CCC-H)C(CC-H)C(CC-H))
C(C(CCC)C(CC-H-H)C(CC-H-N)-H(C))
C(C(CCC)C(CC-H)-H(C))
C(C(CC-H)C(CC-H)-H(C))
-N(C(CC-H-N)-H(-N)-H(-N))

MNA for Phenelzine

MNA for Tranylcypromine

Combining all possible pairs of MNA descriptors of each molecule

C(C(CC-H)C(CC-H)-H(C)) C(C(CCC-H)C(CC-H-H)-H(C)-N(C-H-H))
C(C(CC-H)C(CC-H)-H(C)) C(C(CCC-H)C(CC-H-N)-H(C)-H(C))
C(C(CC-H)C(CC-H)-H(C)) C(C(CCC-H)C(CC-H)C(CC-H)

-N(-H(-N)-C(-H-H-C-N)-N(-H-H-N)) C(C(CCC)C(CC-H)-H(C))
-N(-H(-N)-C(-H-H-C-N)-N(-H-H-N)) C(C(CC-H)C(CC-H)-H(C)
-N(-H(-N)-C(-H-H-C-N)-N(-H-H-N)) -N(C(CC-H-N)-H(-N)-H(-N))

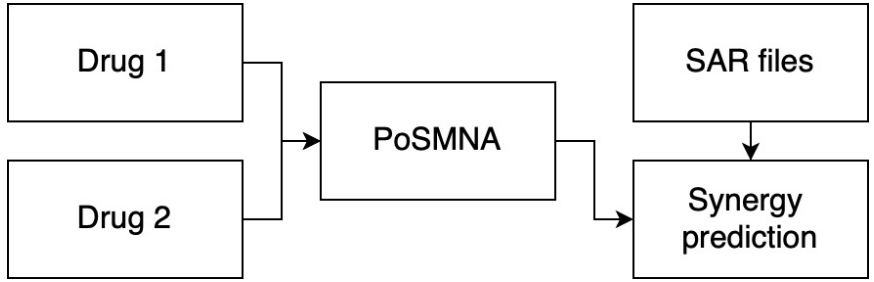
PoSMNA for the pair of Phenelzine and Tranylcypromine

Dmitriev AV et al., In Silico Prediction of Drug-Drug Interactions Mediated by Cytochrome P450 Isoforms. Pharmaceutics. 2021, 13(4):538

Accuracy of prediction calculated by LOO CV compounds out strategy

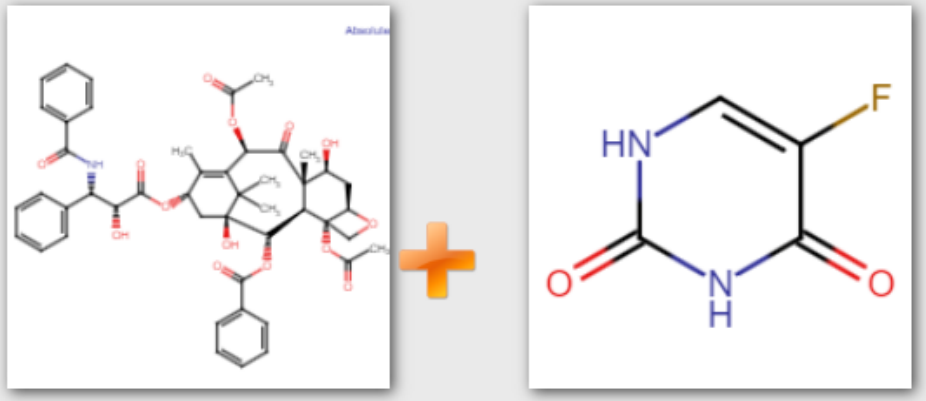
Number of synergistic drug pairs	AUC	Activity Type
110	0.713	Breast Cancer_BT-549
113	0.774	Breast Cancer_HS 578T
77	0.739	CNS Cancer_SF-268
164	0.723	CNS Cancer_SF-539
99	0.750	CNS Cancer_SNB-75
98	0.754	CNS Cancer_U251
130	0.856	Colon Cancer_COLO 205
101	0.817	Colon Cancer_HCC-2998
107	0.727	Colon Cancer_HCT-116
84	0.826	Colon Cancer_HCT-15
108	0.781	Colon Cancer_HT29
73	0.768	Colon Cancer_KM12
91	0.747	Colon Cancer_SW-620
165	0.799	Leukemia_K-562
288	0.735	Leukemia_RPMI-8226
107	0.760	Melanoma_LOX IMVI
105	0.757	Melanoma_M14
126	0.803	Melanoma_MDA-MB-435
191	0.726	Melanoma_SK-MEL-5
92	0.767	Melanoma_UACC-257
117	0.701	Melanoma_UACC-62
81	0.741	Non-Small Cell Lung Cancer_EKVX
78	0.741	Non-Small Cell Lung Cancer_NCI-H226
119	0.703	Non-Small Cell Lung Cancer_NCI-H23
74	0.717	Non-Small Cell Lung Cancer_NCI-H322M
131	0.763	Non-Small Cell Lung Cancer_NCI-H522
58	0.711	Ovarian Cancer_IGROV1
126	0.800	Ovarian Cancer_OVCAR-3
70	0.708	Ovarian Cancer_OVCAR-5
118	0.737	Ovarian Cancer_OVCAR-8
142	0.731	Prostate Cancer_DU-145
65	0.814	Prostate Cancer_PC-3
129	0.737	Renal Cancer_RXF 393
75	0.713	Renal Cancer_UO-31

CLC-Pred synergy:
http://www.way2drug.com/clc-pred-syn/



Way2Drug Home Activities About Contacts

CLC-Pred Synergy – Prediction of synergistic cytotoxicity of drug pairs on 34 NCI60 cell lines. The user guide can be found in the section "**About**".



Drugs name	Paclitaxel
SMILES	[C@H]1(OC(=

Drugs name	Fluorouracil
SMILES	Fc1c(=O)[nH]c

PREDICT SYNERGY

Copy Excel CSV PDF Print

Pa	Pi	Cell-line synergy	Tissue of origin	Histology	AUC LOO CV*
0.744	0.026	HS_578T	Breast	Carcinosarcoma	0.774
0.406	0.303	H322M	Lung	Carcinoma	0.717
0.355	0.126	MDA-MB-435	Breast	Melanoma	0.803

Show 10 entries

Showing 1 to 3 of 3 entries

Previous 1 Next

AUC LOO CV* - the accuracy of prediction (AUC) calculated by leave-one-out compound out (excluded all drug pairs with any drug from the pair) cross validation.

Acknowledgments

Laboratory of structure-function drug design, Institute of Biomedical Chemistry:

A.Rudik, P. Pogodin, P. Savosina, O. Tarasova, A. Dmitriev, S. Ivanov, N. Biziukova, D. Druzhilovskiy, D. Filimonov, V. Poroikov

Bioinformatics department, Pirogov Russian National Research Medical University:

A. Dublin, A. Sezganov

Russian Science Foundation:

Grant 19-15-00396, https://rscf.ru/project/19-15-00396/