Supporting Information

Prediction of human cytochrome P450 inhibition using a multi-task deep autoencoder neural network

Xiang Li[1,2], Youjun Xu[3], Luhua Lai[1,2,3] and Jianfeng Pei[3,*]*

1. Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, 100871, China.

2. BNLMS, State Key Laboratory for Structural Chemistry of Unstable and Stable Species, College of Chemistry and Molecular Engineering, Peking University, Beijing, 100871, China.

3. Center for Quantitative Biology, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, 100871, China.

Table S1. Curve classes and corresponding description.

Curve class	Curve Description
1.1	Complete curve, high efficacy
1.2	Complete curve, partial efficacy
1.3	Complete curve, high efficacy, poor fit
1.4	Complete curve, partial efficacy, poor fit
2.1	Partial curve, high efficacy
2.2	Partial curve, partial efficacy
2.3	Partial curve, high efficacy, poor fit
2.4	Partial curve, partial efficacy, poor fit
3	Single point of activity
4	NA

Table S2. The number of inhibitors and non-inhibitors, under different criteria combinations.

Inhibitor number	1A2	2C9	2C19	2D6	3A4
AC50	5655	4996	5680	2643	4882
Score	5690	3925	5639	2335	4994
Curve	4638	3222	5146	1592	3682
AC50∩Score	4292	2855	4145	1634	3344
AC50∩Curve	4128	2679	4058	1383	3042
Score∩Curve	4330	2910	4880	1421	3424
AC50∩Score∩Curve	3826	2374	3794	1228	2784
Non-inhibitor number	1A2	2C9	2C19	2D6	3A4
AC50	6361	7628	6518	10259	6714
Score	6333	7596	6478	10226	6708
Curve	6333	7596	6478	10226	6708
AC50∩Score	6333	7596	6478	10226	6708
AC50∩Curve	6333	7596	6478	10226	6708
Score∩Curve	6333	7596	6478	10226	6708
AC50∩Score∩Curve	6333	7596	6478	10226	6708

Table S3. Top 10 sensitive positively correlated and negatively correlated variables of 1A2 task in multi-task model.

Variable		Sensitivity	Description
Positively Correlated	PubchemFP385	0.590	C(:C)(:C)(:C)
	PubchemFP469	0.571	S-S-C:C
	nssS	0.568	Count of atom-type E-State: -S-
	PubchemFP862	0.536	CC1C(S)CCC1
	PubchemFP737	0.502	Cc1cc(N)ccc1
	PubchemFP363	0.479	C(~F)(~F)
	nF10Ring	0.472	Number of 10-membered fused rings
	PubchemFP841	0.459	CC1CC(S)CC1
	naaaC	0.456	Count of atom-type E-State: ::C:
	nT10Ring	0.455	Number of 10-membered rings (includes counts from fused rings)
Negatively Correlated	nBase	2.111	Number of basic groups
	nG12Ring	1.387	Number of >12-membered rings
	nG12HeteroRing	1.361	Number of >12-membered rings containing heteroatoms (N, O, P, S, or halogens)
	nAcid	1.322	Number of acidic groups
	ndssS	1.147	Count of atom-type E-State: >S=
	nssssNp	1.088	Count of atom-type E-State: >N<+
	SssssNp	1.060	Sum of atom-type E-State: >N<+
	maxssssNp	0.983	Maximum atom-type E-State: >N<+
	minssssNp	0.956	Minimum atom-type E-State: >N<+
	nsssCH	0.917	Count of atom-type E-State: >CH-

Table S4. Top 10 sensitive positively correlated and negatively correlated variables of 1A2 task in single-task model.

Variable		Sensitivity	Description
Positively Correlated	nF10Ring	1.868	Number of 10-membered fused rings
	nT10Ring	1.862	Number of 10-membered rings (includes counts from fused rings)
	ETA_dAlpha_A	1.603	A measure of count of non-hydrogen heteroatoms
	PubchemFP469	1.502	S-S-C:C
	PubchemFP862	1.314	CC1C(S)CCC1
	nssS	1.280	Count of atom-type E-State: -S-
	nT10HeteroRing	1.247	Number of 10-membered rings (includes counts from fused rings) containing heteroatoms (N, O, P, S, or halogens)
	nF10HeteroRing	1.231	Number of 12-membered fused rings containing heteroatoms (N, O, P, S, or halogens)
	nF5Ring	1.227	Number of 5-membered fused rings
	nF9Ring	1.202	Number of 9-membered fused rings
Negatively Correlated	nBase	3.001	Number of basic groups
	nsssCH	2.995	Count of atom-type E-State: >CH-
	nF6Ring	2.881	Number of 6-membered fused rings
	nAcid	2.724	Number of acidic groups
	ndssS	2.058	Count of atom-type E-State: >S=
	ALogp2	2.002	Square of ALogP
	nG12Ring	1.910	Number of >12-membered rings
	nG12HeteroRing	1.898	Number of >12-membered rings containing heteroatoms (N, O, P, S, or halogens)
	PubchemFP348	1.896	C(~C)(~H)(~P)
	C3SP3	1.888	Singly bound carbon bound to three other carbons

Table S5. Top 10 sensitive positively correlated and negatively correlated variables of 2C9 task in multi-task model.

Variable		Sensitivity	Description
Positively Correlated	nssS	0.752	Count of atom-type E-State: -S-
	LipinskiFailures	0.679	Number failures of the Lipinski's Rule Of 5
	nAtomLAC	0.658	Number of atoms in the longest aliphatic chain
	maxdS	0.588	Maximum atom-type E-State: =S
	PubchemFP865	0.550	CC1C(Br)CCC1
	ndS	0.544	Count of atom-type E-State: =S
	PubchemFP407	0.543	O(~C)(~P)
	SC.6	0.536	Simple cluster, order 6
	PubchemFP770	0.500	Nc1c(N)cccc1
	PubchemFP469	0.477	S-S-C:C
Negatively Correlated	nBase	2.796	Number of basic groups
	nG12Ring	1.882	Number of >12-membered rings
	nG12HeteroRing	1.845	Number of >12-membered rings containing heteroatoms (N, O, P, S, or halogens)
	nssssNp	1.657	Count of atom-type E-State: >N<+
	SssssNp	1.604	Sum of atom-type E-State: >N<+
	ndssS	1.512	Count of atom-type E-State: >S=
	maxssssNp	1.498	Maximum atom-type E-State: >N<+
	minssssNp	1.476	Minimum atom-type E-State: >N<+
	PubchemFP47	1.249	>= 2 I
	nHBint8	1.143	Count of E-State descriptors of strength for potential Hydrogen Bonds of path length 8

Table S6. Top 10 sensitive positively correlated and negatively correlated variables of 2C9 task in single-task model.

Variable		Sensitivity	Description
Positively Correlated	nssS	1.543	Count of atom-type E-State: -S-
	PubchemFP469	1.372	S-S-C:C
	SC.6	1.166	Simple cluster, order 6
	PubchemFP648	1.110	O=N-C:C-N
	LipinskiFailures	1.085	Number failures of the Lipinski's Rule Of 5
	nAtomLAC	1.068	Number of atoms in the longest aliphatic chain
	PubchemFP730	1.039	Nc1ccc(Br)cc1
	PubchemFP880	1.035	BrC1C(Br)CCC1
	PubchemFP793	1.003	NC1CCC(Br)CC1
	PubchemFP360	0.972	C(~Cl)(~Cl)
Negatively Correlated	nBase	4.564	Number of basic groups
	nssssNp	3.784	Count of atom-type E-State: >N<+
	minssssNp	3.731	Minimum atom-type E-State: >N<+
	maxssssNp	3.680	Maximum atom-type E-State: >N<+
	SssssNp	3.653	Sum of atom-type E-State: >N<+
	nG12Ring	3.212	Number of >12-membered rings
	nG12HeteroRing	3.196	Number of >12-membered rings containing heteroatoms (N, O, P, S, or halogens)
	nHBint8	2.961	Count of E-State descriptors of strength for potential Hydrogen Bonds of path length 8
	PubchemFP117	2.615	>= 1 saturated or aromatic nitrogen-containing ring size 3
	PubchemFP343	2.519	C(~C)(~Cl)(~H)

Table S7. Top 10 sensitive positively correlated and negatively correlated variables of 2C19 task in multi-task model.

Variable		Sensitivity	Description
Positively Correlated	nssS	0.630	Count of atom-type E-State: -S-
	nAtomLAC	0.589	Number of atoms in the longest aliphatic chain
	PubchemFP407	0.502	O(~C)(~P)
	PubchemFP865	0.471	CC1C(Br)CCC1
	LipinskiFailures	0.469	Number failures of the Lipinski's Rule Of 5
	maxdS	0.448	Maximum atom-type E-State: =S
	nF5Ring	0.434	Number of 5-membered fused rings
	PubchemFP469	0.415	S-S-C:C
	nX	0.411	Number of halogen atoms (F, Cl, Br, I, At, Uus)
	ndS	0.399	Count of atom-type E-State: =S
Negatively Correlated	nBase	2.454	Number of basic groups
	nG12Ring	1.709	Number of >12-membered rings
	nG12HeteroRing	1.678	Number of >12-membered rings containing heteroatoms (N, O, P, S, or halogens)
	nssssNp	1.491	Count of atom-type E-State: >N<+
	SssssNp	1.437	Sum of atom-type E-State: >N<+
	ndssS	1.368	Count of atom-type E-State: >S=
	maxssssNp	1.317	Maximum atom-type E-State: >N<+
	minssssNp	1.296	Minimum atom-type E-State: >N<+
	PubchemFP47	1.143	>= 2 I
	nAcid	1.113	Number of acidic groups

Table S8. Top 10 sensitive positively correlated and negatively correlated variables of 2C19 task in single-task model.

Variable		Sensitivity	Description
Positively Correlated	nssS	1.356	Count of atom-type E-State: -S-
	PubchemFP407	1.313	O(~C)(~P)
	PubchemFP469	1.149	S-S-C:C
	PubchemFP865	1.088	CC1C(Br)CCC1
	SC.6	1.086	Simple cluster, order 6
	maxdS	1.043	Maximum atom-type E-State: =S
	PubchemFP182	1.040	>= 1 unsaturated non-aromatic carbon-only ring size 6
	SdS	0.961	Sum of atom-type E-State: =S
	PubchemFP770	0.955	Nc1c(N)cccc1
	PubchemFP875	0.941	NC1C(N)CCC1
Negatively Correlated	nBase	4.074	Number of basic groups
	ndssS	3.191	Count of atom-type E-State: >S=
	nG12Ring	2.958	Number of >12-membered rings
	nG12HeteroRing	2.940	Number of >12-membered rings containing heteroatoms (N, O, P, S, or halogens)
	nssssNp	2.855	Count of atom-type E-State: >N<+
	SssssNp	2.617	Sum of atom-type E-State: >N<+
	maxssssNp	2.323	Maximum atom-type E-State: >N<+
	minssssNp	2.296	Minimum atom-type E-State: >N<+
	nF7Ring	2.292	Number of 7-membered fused rings
	minddC	2.285	Minimum atom-type E-State: =C=

Table S9. Top 10 sensitive positively correlated and negatively correlated variables of 2D6 task in multi-task model.

Variable		Sensitivity	Description
Positively Correlated	nBase	1.126	Number of basic groups
	PubchemFP367	0.953	C(~H)(~O)(~O)
	nAtomLAC	0.784	Number of atoms in the longest aliphatic chain
	maxsssCH	0.660	Maximum atom-type E-State: >CH-
	maxsI	0.656	Maximum atom-type E-State: -I
	PubchemFP661	0.633	O-C-O-C-C
	PubchemFP298	0.582	C-I
	PubchemFP46	0.581	>= 1 I
	PubchemFP350	0.575	C(~C)(~I)
	PubchemFP38	0.545	>= 4 S
Negatively Correlated	nG12Ring	1.599	Number of >12-membered rings
	nG12HeteroRing	1.580	Number of >12-membered rings containing heteroatoms (N, O, P, S, or halogens)
	nssssNp	1.195	Count of atom-type E-State: >N<+
	PubchemFP410	1.126	P(~C)(~C)
	SssssNp	1.111	Sum of atom-type E-State: >N<+
	PubchemFP47	1.095	>= 2 I
	ndssS	1.079	Count of atom-type E-State: >S=
	PubchemFP262	1.057	>= 4 hetero-aromatic rings
	PubchemFP408	1.052	O(~H)(~S)
	PubchemFP845	0.933	OC1CC(O)CC1

Table S10. Top 10 sensitive positively correlated and negatively correlated variables of 2D6 task in single-task model.

Variable		Sensitivity	Description
Positively Correlated	nBase	3.000	Number of basic groups
	nF9HeteroRing	1.466	Number of 9-membered fused rings containing heteroatoms (N, O, P, S, or halogens)
	nT9HeteroRing	1.464	Number of 9-membered rings (includes counts from fused rings)
	nHBint7	1.439	Count of E-State descriptors of strength for potential Hydrogen Bonds of path length 7
	PubchemFP46	1.378	>= 1 I
	PubchemFP298	1.354	C-I
	PubchemFP350	1.346	C(~C)(~I)
	maxsssCH	1.303	Maximum atom-type E-State: >CH-
	nF9Ring	1.211	Number of 9-membered fused rings
	nT9Ring	1.175	Number of 9-membered rings (includes counts from fused rings)
Negatively Correlated	nG12HeteroRing	2.980	Number of >12-membered rings containing heteroatoms (N, O, P, S, or halogens)
	nG12Ring	2.971	Number of >12-membered rings
	PubchemFP262	2.535	>= 4 hetero-aromatic rings
	PubchemFP408	2.299	O(~H)(~S)
	PubchemFP331	2.272	C(~Br)(:N)
	PubchemFP47	2.215	>= 2 I
	PubchemFP481	2.174	O-S-C:C
	PubchemFP745	2.125	Sc1cc(S)ccc1
	PubchemFP808	2.087	SC1CC(S)CCC1
	ndssS	2.066	Count of atom-type E-State: >S=

Table S11. Top 10 sensitive positively correlated and negatively correlated variables of 3A4 task in multi-task model.

Variable		Sensitivity	Description
Positively Correlated	nAtomLAC	0.732	Number of atoms in the longest aliphatic chain
	PubchemFP661	0.690	O-C-O-C-C
	PubchemFP563	0.588	O-C-O-C
	PubchemFP367	0.582	C(~H)(~O)(~O)
	PubchemFP407	0.537	O(~C)(~P)
	PubchemFP865	0.522	CC1C(Br)CCC1
	PubchemFP360	0.514	C(~Cl)(~Cl)
	ATS8m	0.498	Broto-Moreau autocorrelation - lag 8 / weighted by mass
	ATS8p	0.491	Broto-Moreau autocorrelation - lag 8 / weighted by polarizabilities
	PubchemFP226	0.485	>= 2 unsaturated non-aromatic heteroatom-containing ring size 7
Negatively Correlated	nBase	1.672	Number of basic groups
	nssssNp	1.374	Count of atom-type E-State: >N<+
	SssssNp	1.336	Sum of atom-type E-State: >N<+
	nG12Ring	1.286	Number of >12-membered rings
	nG12HeteroRing	1.256	Number of >12-membered rings containing heteroatoms (N, O, P, S, or halogens)
	maxssssNp	1.206	Maximum atom-type E-State: >N<+
	minssssNp	1.185	Minimum atom-type E-State: >N<+
	ndssS	1.016	Count of atom-type E-State: >S=
	PubchemFP47	1.012	>= 2 I
	PubchemFP408	0.988	O(~H)(~S)

Table S12. Top 10 sensitive positively correlated and negatively correlated variables of 3A4 task in single-task model.

Variable		Sensitivity	Description
Positively Correlated	PubchemFP865	1.572	CC1C(Br)CCC1
	nHCHnX	1.546	Count of atom-type H E-State: CHnX
	nF5Ring	1.529	Number of 5-membered fused rings
	MDEC.14	1.524	Molecular distance edge between all primary and quaternary carbons
	SdS	1.461	Sum of atom-type E-State: =S
	C3SP2	1.434	Doubly bound carbon bound to three other carbons
	PubchemFP123	1.425	>= 2 saturated or aromatic carbon-only ring size 3
	PubchemFP329	1.398	C(~Br)(~H)
	PubchemFP360	1.328	C(~Cl)(~Cl)
	MDEC.34	1.266	Molecular distance edge between all tertiary and quaternary carbons
Negatively Correlated	nBase	3.597	Number of basic groups
	SssssNp	3.533	Count of atom-type E-State: >N<+
	nssssNp	3.182	Sum of atom-type E-State: >N<+
	maxssssNp	3.178	Maximum atom-type E-State: >N<+
	minssssNp	3.108	Minimum atom-type E-State: >N<+
	nAcid	2.942	Number of acidic groups
	PubchemFP408	2.307	O(~H)(~S)
	nF7Ring	2.247	Number of 7-membered fused rings
	PubchemFP47	2.185	>= 2 I
	nT4HeteroRing	1.912	Number of 4-membered rings (includes counts from fused rings)