

## ***In silico* Prediction of Chemical Subcellular Localization via Multi-classification Methods**

Hongbin Yang, Xiao Li, Yingchun Cai, Qin Wang, Weihua Li, Guixia Liu, Yun Tang\*

*Shanghai Key Laboratory of New Drug Design, School of Pharmacy, East China University of  
Science and Technology, Shanghai 200237, China*

**Supplementary Table 1: Performance of models from 5-fold cross validation.**

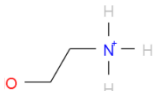
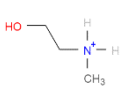
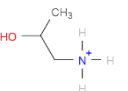
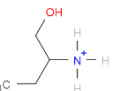
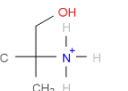
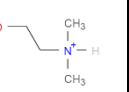
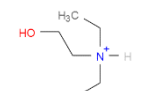
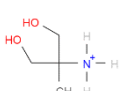
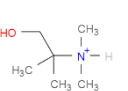
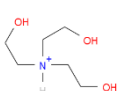
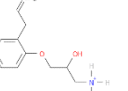
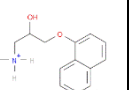
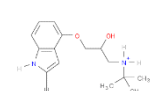
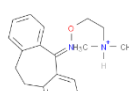
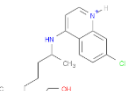
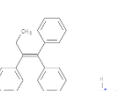
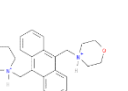
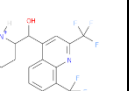
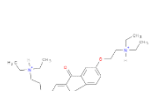
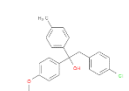
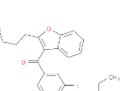
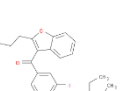
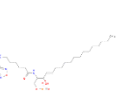

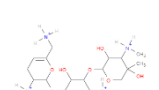
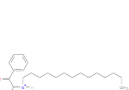
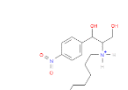


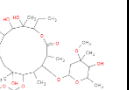
Feature	Method	Q <sub>total</sub>	Q <sub>1</sub>	Q <sub>2</sub>	Q <sub>3</sub>	Q <sub>4</sub>
CDK	kNN	0.709	0.720	0.616	0.778	0.830
CDK	SVM	0.756	0.705	0.811	0.728	0.739
CDK	ANN	0.747	0.720	0.753	0.765	0.761
CDK	RF	0.703	0.561	0.858	0.593	0.682
CDK	NB	0.529	0.621	0.463	0.506	0.557
CDK	CT	0.570	0.530	0.532	0.568	0.716
SubFP	kNN	0.686	0.705	0.653	0.654	0.761
SubFP	SVM	0.719	0.636	0.795	0.605	0.784
SubFP	ANN	0.69	0.674	0.684	0.617	0.795
SubFP	RF	0.599	0.455	0.937	0.099	0.545
SubFP	NB	0.625	0.621	0.647	0.506	0.693
SubFP	CT	0.646	0.576	0.637	0.580	0.830
Estate	kNN	0.703	0.629	0.721	0.691	0.784
Estate	SVM	0.705	0.667	0.768	0.593	0.727
Estate	ANN	0.684	0.629	0.732	0.617	0.727
Estate	RF	0.605	0.462	0.916	0.173	0.545
Estate	NB	0.603	0.583	0.632	0.531	0.636
Estate	CT	0.668	0.644	0.663	0.605	0.773
Pubchem	kNN	0.711	0.697	0.674	0.765	0.761
Pubchem	SVM	0.745	0.727	0.763	0.691	0.784
Pubchem	ANN	0.713	0.682	0.716	0.716	0.75
Pubchem	RF	0.692	0.682	0.747	0.593	0.682
Pubchem	NB	0.560	0.545	0.605	0.519	0.523
Pubchem	CT	0.648	0.568	0.663	0.642	0.739
KRFP	kNN	0.715	0.742	0.668	0.691	0.795
KRFP	SVM	0.753	0.674	0.832	0.679	0.773

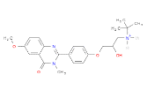
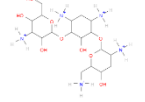
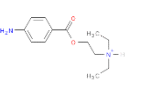
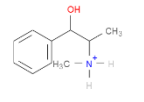
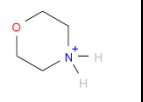
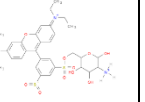
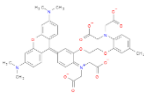
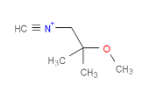
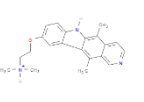
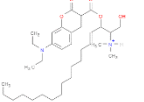
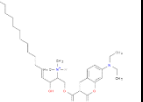
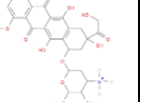
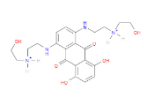
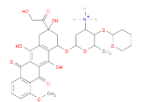
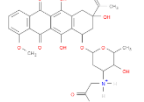
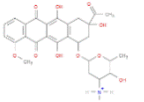
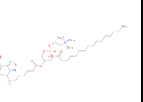
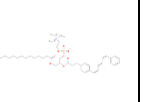
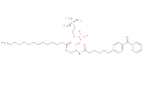
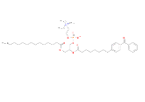
KRFP	ANN	0.725	0.697	0.726	0.667	0.818
KRFP	RF	0.656	0.492	0.884	0.407	0.636
KRFP	NB	0.654	0.598	0.695	0.568	0.727
KRFP	CT	0.668	0.621	0.637	0.691	0.784
MACCS	kNN	0.715	0.697	0.668	0.728	0.83
MACCS	SVM	0.762	0.720	0.811	0.679	0.795
MACCS	ANN	0.711	0.652	0.747	0.642	0.784
MACCS	RF	0.727	0.682	0.832	0.58	0.705
MACCS	NB	0.597	0.598	0.616	0.481	0.659
MACCS	CT	0.664	0.621	0.637	0.679	0.773
Full_Des	kNN	0.755	0.699	0.750	0.765	0.841
Full_Des	SVM	0.761	0.729	0.823	0.654	0.773
Full_Des	NB	0.609	0.526	0.771	0.412	0.500
Full_Des	RF	0.660	0.609	0.708	0.605	0.682
Full_Des	ANN	0.731	0.707	0.776	0.667	0.727
Full_Des	CT	0.573	0.632	0.557	0.444	0.636
CT_Des	kNN	0.707	0.682	0.753	0.593	0.750
CT_Des	SVM	0.613	0.515	0.763	0.593	0.557
CT_Des	NB	0.599	0.507	0.653	0.482	0.625
CT_Des	RF	0.581	0.455	0.837	0.296	0.477
CT_Des	ANN	0.715	0.720	0.711	0.630	0.796
CT_Des	CT	0.617	0.652	0.653	0.432	0.659
F20_Des	kNN	0.637	0.667	0.647	0.556	0.648
F20_Des	SVM	0.495	0.349	0.790	0.161	0.386
F20_Des	NB	0.430	0.409	0.384	0.309	0.671
F20_Des	RF	0.587	0.530	0.753	0.358	0.523
F20_Des	ANN	0.640	0.545	0.768	0.432	0.693
F20_Des	CT	0.576	0.621	0.584	0.444	0.614
F100_Des	kNN	0.566	0.568	0.579	0.457	0.636
F100_Des	SVM	0.493	0.349	0.763	0.612	0.523
F100_Des	NB	0.448	0.364	0.421	0.469	0.614
F100_Des	RF	0.635	0.644	0.690	0.494	0.636
F100_Des	ANN	0.686	0.667	0.721	0.593	0.727
F100_Des	CT	0.614	0.599	0.642	0.556	0.614
GS_Des	kNN	0.705	0.644	0.711	0.716	0.773
GS_Des	SVM	0.672	0.576	0.832	0.457	0.671
GS_Des	NB	0.646	0.561	0.616	0.728	0.761
GS_Des	RF	0.662	0.614	0.784	0.395	0.716
GS_Des	ANN	0.712	0.689	0.711	0.679	0.773
GS_Des	CT	0.580	0.621	0.590	0.407	0.659

---

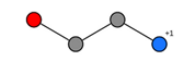

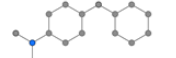


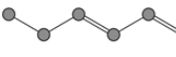

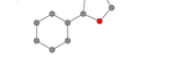
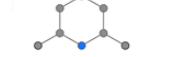
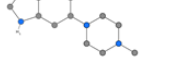
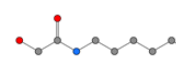
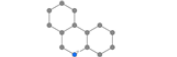


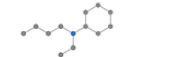



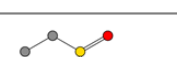




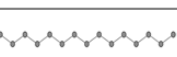

**Supplemental Table 2: Compounds that contain the key substructures.**

The caption 1 means the compound is 2-hydroxyethan-1-aminium derivatives (containing substructure 1) and the caption 2 means the compound is N-ethyl-2-hydroxy-N-methylethan-1-aminium derivatives (containing substructure 2).

Localized in lysosome						
						
No.	1	2	3	4	5	6
Subs.	(1)	(1)	(1)	(1)	(1)	(1)
						
No.	7	8	9	10	11	12
Subs	(1)	(1), (2)	(1)	(1), (2)	(1)	(1)
						
No.	13	14	15	16	17	18
Subs	(1)	(1)	(1), (2)	(1)	(1), (2)	(1)
						
No.	19	20	21	22	23	24
Subs	(1), (2)	(1), (2)	(1)	(1), (2)	(1)	(1), (2)
						
No.	25	26	27	28	29	30
Subs	(1)	(1)	(1)	(1)	(1)	(1)

						
No.	31	32	33	34	35	36
Subs	(1)	(1)	(1), (2)	(1)	(1)	(1)
Localized in other organelles						
						
No.	37	38	39	40	41	42
Subs	(1)	(1)	(1)	(1)	(1)	(1)
Loc.	Mito.	Mito.	Mito.	Mito.	Mito.	Nucleus
						
No.	43	44	45	46	47	48
Subs	(1)	(1)	(1)	(1)	(1)	(1)
Loc	Nucleus	Nucleus	Nucleus	Nucleus	Plasma mem.	Plasma mem.
						
No.	49	50				
Subs	(1)	(1)				
Loc.	Plasma mem.	Plasma mem.				

## Supplemental Figure S1

				
No. 1 Lysosome	No.2 Lysosome	No.3 Mitochondria	No.4 Mitochondria	No.5 Mitochondria
<chem>OCC[N+]</chem>	<chem>CC[NH+](C)CCO</chem>	<chem>CN(C)c1ccc(cc1)Cc1ccccc1</chem>	<chem>[P+](c1ccccc1)(c1ccccc1)</chem>	<chem>n1c(CC)ccc1C</chem>
				
No.6 Mitochondria	No.7 Nucleus	No.8 Nucleus	No.9 Nucleus	No.10 Nucleus
<chem>C(C)C(=CC(=O))</chem>	<chem>CCl</chem>	<chem>c1ccc(cc1)c1ccc(o1)</chem>	<chem>n1c(C)cccc1C</chem>	<chem>c1[nH]c2cc(ccc2n1)N1CCN(CC1)C</chem>
				
No.11 Nucleus	No.12 Nucleus	No.13 Nucleus	No.14 Nucleus	No.15 Plasma Membrane
<chem>OCC(=O)NCCCCC</chem>	<chem>C[n+]1cc2ccccc2c2ccccc12</chem>	<chem>N=C(c1ccc(cc1))N</chem>	<chem>c1csc(C)n1</chem>	<chem>CCCCN(CCC)c1ccccc1</chem>
				
No.16 Plasma Membrane	No.17 Plasma Membrane	No.18 Plasma Membrane	No.19 Plasma Membrane	No.20 Plasma Membrane
<chem>c1ccc2cc(N)ccc2c1</chem>	<chem>CCCCc1ccc2c(Cc3ccccc23)c1</chem>	<chem>c1csc(c1)</chem>	<chem>CCS(=O)(=O)</chem>	<chem>C=Cc1cc[n+](CCC[N+](C)(C)C)c1</chem>
				
No.21 Lysosome	No.22 Mitochondria	No.23 Mitochondria	No.24 Mitochondria	No.25 Plasma membrane
<chem>NC1(CCCCC1)C</chem>	<chem>Oc1ccc(Cl)cc1</chem>	<chem>C(=O)Nc1ccc(Cl)cc1</chem>	<chem>[O-]C(=O)CCCCCCCCCCCC</chem>	<chem>S(=O)(=O)c1ccc(cc1)N1CCC(=N1)</chem>

**Figure S1.** The significant substructures drawn by SMARTSviewer. The numeric orders are corresponding to the **Table 6** and **Table 7**.