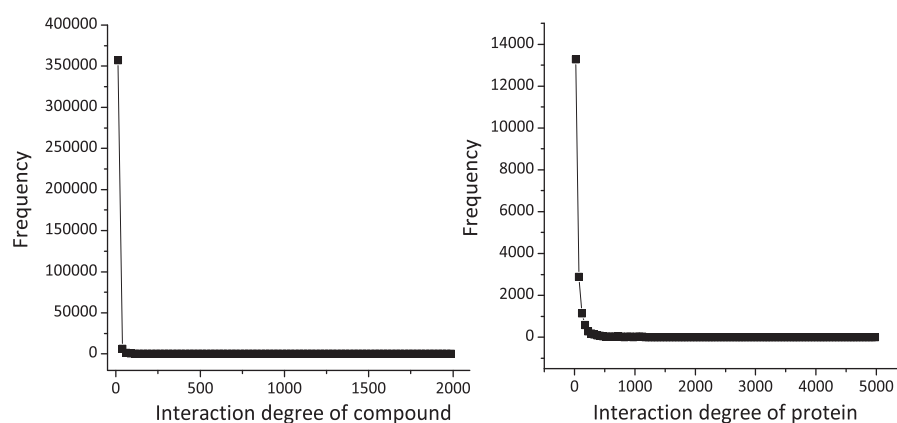


## Supplementary Files: Screening negative compound-protein samples

Liu *et al.*

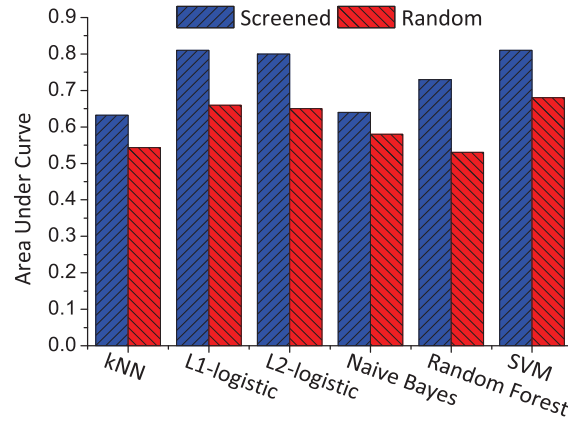
No Institute Given



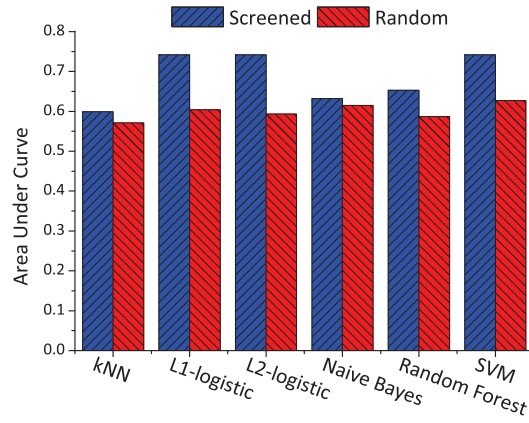
**Fig. S1.** Frequency distribution of validated/predicted interactions vs. interaction degree of involved compounds/proteins.

**Table S1.** AUC/Recall/Precision values of six classical classifiers on screened and randomly generated negative samples of *C. elegans* (pairwise cross-validation).

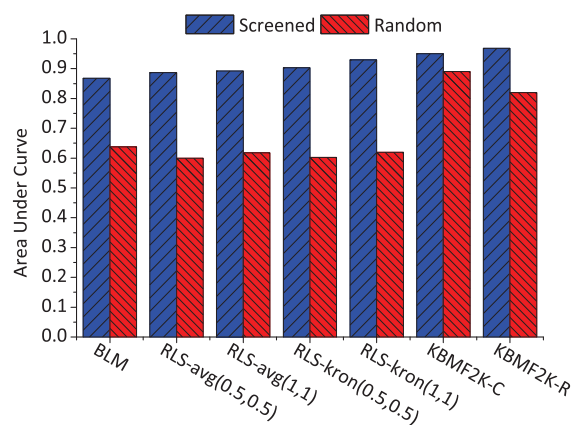
Measure	Negative sample ratio	Naive Bayes		kNN		Random Forest		L1-logistic		L2-logistic		SVM	
		screened	random	screened	random	screened	random	screened	random	screened	random	screened	random
AUC	1	<b>0.645</b>	<b>0.621</b>	0.858	0.628	0.902	0.659	0.882	0.712	0.892	0.693	0.894	0.702
	3	0.633	0.613	0.892	0.668	0.926	0.672	0.895	0.712	0.896	0.698	0.901	0.706
	5	0.632	0.614	<b>0.897</b>	<b>0.682</b>	<b>0.928</b>	<b>0.694</b>	<b>0.902</b>	<b>0.715</b>	<b>0.906</b>	<b>0.702</b>	<b>0.907</b>	<b>0.713</b>
Precision	1	<b>0.613</b>	<b>0.601</b>	<b>0.801</b>	<b>0.573</b>	0.821	0.618	0.872	<b>0.748</b>	<b>0.890</b>	<b>0.763</b>	0.785	<b>0.600</b>
	3	0.351	0.335	0.787	0.468	<b>0.836</b>	0.580	<b>0.863</b>	0.680	0.875	0.689	0.837	0.438
	5	0.247	0.235	0.774	0.524	0.830	<b>0.626</b>	0.857	0.648	0.863	0.667	<b>0.896</b>	0.225
Recall	1	<b>0.465</b>	<b>0.422</b>	<b>0.827</b>	<b>0.560</b>	<b>0.844</b>	<b>0.672</b>	<b>0.849</b>	<b>0.704</b>	<b>0.877</b>	<b>0.729</b>	<b>0.818</b>	<b>0.503</b>
	3	0.454	0.372	0.743	0.323	0.705	0.340	0.648	0.293	0.681	0.330	0.576	0.107
	5	0.442	0.366	0.690	0.205	0.639	0.208	0.566	0.176	0.582	0.195	0.519	0.024



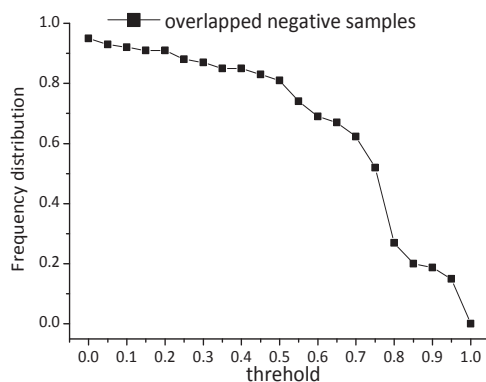
**Fig.S2.** Histogram of the AUC values achieved by six classical classifiers on screened and randomly-selected negative samples of *human* (block-wise cross-validation).



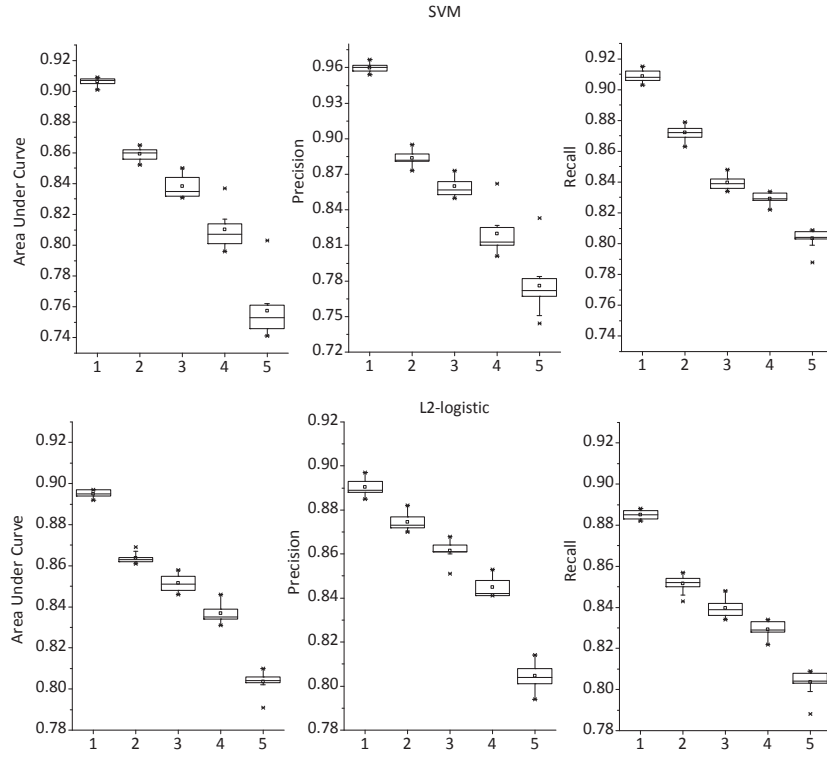
**Fig.S3.** Histogram of the AUC values achieved by six classical classifiers on screened and randomly-selected negative samples of *C. elegans* (block-wise cross-validation).



**Fig. S4.** Histogram of the AUC values achieved by existing predictive methods on screened and randomly-selected negative samples of *C. elegans*.



**Fig. S5.** Frequency distribution of the overlapped interactions between the negative samples extracted from Davis *et al.* (2011) drug bioactivity data and our screened negative samples, with respect to different thresholds of the confidence scores  $d_{kj}$ .



**Fig. S6.** Statistical boxplots of the AUC, precision and recall measures of the SVM and L2-logistic classifiers trained on screened negative samples, which were partitioned into 5 bins according to the confidence scores  $d_{kj}$  in descending order.

**Table S2.** Top 125 predicted target proteins of *Donepezil* and related functional annotations involved in signaling pathways, together with the p-value derived by gene set enrichment analysis.

Annotations	Proteins
Target proteins	ADORA1, ADORA2A, ADORA2B, ADRA1A, ADRA1B, ADRA1D, ADRA2A, ADRB1, ADRB2, AGTR2, AVPR2, BDKRB2, CCKBR, CHRM1, CHRM2, CHRM3, CHRM4, CNR1, CYSLTR1, DRD2, EDNRB, F2, FPR1, GNRHR, HRH2, HRH3, HTR1A, HTR1D, HTR2A, HTR2C, MTNR1A, NPY2R, NR3C1, NTSR2, OPRK1, OPRM1, P2RY2, PTGER1, PTGFR, PTGIR, TAAR1, TACR1, TBXA2R, TSPO, ABCB1, ABCB11, ABCB4, ABCC1, ABCC10, ABCC2, ABCC3, ABCC4, ABCC6, ABCC8, ABCC9, ACE, ACHE, ALB, ALOX5, ANXA1, BCHE, CALM1, CCR5, CFTR, COMT, CXCR1, CYP11A1, CYP11B1, CYP17A1, CYP19A1, CYP1A1, CYP1A2, CYP1B1, CYP24A1, CYP26A1, CYP27B1, CYP2A6, CYP2B6, CYP2C18, CYP2C19, CYP2C8, CYP2C9, CYP2E1, CYP2J2, CYP3A4, CYP3A5, CYP3A7, CYP4B1, CYP51A1, DPP4, EGFR, ESR1, GPR162, GPR18, GPR44, JUN, KCNH2, MPO, NPSR1, P2RY12, PDE6H, PGR, POMC, PPARG, PROS1, PTGER3, PTGIS, PTGS1, PTGS2, REN, RHO, SHBG, SLC15A1, SLC22A1, SLC22A2, SLC22A4, SLC22A5, SLC22A6, SLC22A8, SLC2A2, SLC6A4, TAP1, TBXA1, TOP2A, TPO
Neuroactive ligand-receptor interaction pathway (p-value $\leq$ 4.7e-32)	ADORA1, ADORA2A, ADORA2B, ADRA1A, ADRA1B, ADRA1D, ADRA2A, ADRB1, ADRB2, AGTR2, AVPR2, BDKRB2, CCKBR, CHRM1, CHRM2, CHRM3, CHRM4, CNR1, CYSLTR1, DRD2, EDNRB, F2, FPR1, GNRHR, HRH2, HRH3, HTR1A, HTR1D, HTR2A, HTR2C, MTNR1A, NPY2R, NR3C1, NTSR2, OPRK1, OPRM1, P2RY2, PTGER1, PTGFR, PTGIR, TAAR1, TACR1, TBXA2R, TSPO
hypertension (p-value $\leq$ 1.5E-11)	ADORA1, ADORA2A, ADRA1A, ADRA1B, ADRA2A, ADRB1, ADRB2, AGTR2, BDKRB2, EDNRB, F2, HTR2A, NR3C1, P2RY2, ABCB1, ACE, CCR5, COMT, CYP11B1, CYP1B1, CYP2C19, CYP2C8, CYP2C9, CYP2J2, CYP3A4, CYP3A5, CYP4B1, ESR1, PPARG, PTGIS, PTGS2, REN, SLC22A2, SLC6A4
Alzheimer's (p-value $\leq$ 3.1e-4)	ADRA2A, HTR2A, ABCB1, ACE, CCR5, COMT, ESR1, PPARG, PTGS2, REN, SLC6A4, CHRM1, HTR2C, ACHE, ALB, ALOX5, BCHE, CYP17A1, CYP19A1, CYP1A1, MPO
Parkinson's (p-value $\leq$ 3.0E-4)	ADRA2A, HTR2A, ABCB1, ACE, CCR5, COMT, ESR1, PTGS2, SLC6A4, CYP1A1, ADRA1A, CYP1B1, CCKBR, CNR1, DRD2, CYP2E1, POMC
atherosclerosis (p-value $\leq$ 1.3E-3)	ACE, CCR5, ESR1, PTGS2, CYP2E1, PPARG, BCHE, CYP17A1, CYP19A1, ADRB2, AGTR2, BDKRB2, F2, NR3C1, CYP2J2, ABCC6, CALM1, P2RY12, PTGS1