CHEMMEDCHEM
Chemistry & Drug Discovery

Supporting Information

© Copyright Wiley-VCH Verlag GmbH & Co. KGaA, 69451 Weinheim, 2010

Prospective Validation of a Comprehensive In silico hERG Model and its Applications to Commercial Compound and Drug Databases

Munikumar R. Doddareddy, Elisabeth C. Klaasse, Shagufta, Adriaan P. IJzerman, and Andreas Bender*[a]

cmdc_201000024_sm_miscellaneous_information.pdf

Supporting information

Contents

Tables

Table S1a. Classification statistics of LDA and SVM models generated using ECFP_4 and FCFP_4 fingerprints

		ECFP_4						FCFP_4					
		LDA1	**LDA2**	**LDA3**	**SVM1**	**SVM2**	**SVM3**	**LDA1**	**LDA2**	**LDA3**	**SVM1**	**SVM2**	**SVM3**
CV	**TP**	476	638	788	488	657	814	514	669	824	502	669	823
	FP	129	165	199	84	123	144	146	170	207	81	129	149
	TN	1256	1220	1186	1301	1262	1241	1239	1215	1178	1304	1256	1236
	FN	176	197	216	164	178	190	138	166	180	150	166	181
	AUC	0.90	0.89	0.89	0.93	0.93	0.92	0.91	0.91	0.90	0.94	0.94	0.93
	C	0.65	0.65	0.64	0.71	0.71	0.71	0.68	0.67	0.66	0.73	0.72	0.72
	GHa	0.75	0.78	0.79	0.80	0.81	0.83	0.78	0.79	0.80	0.81	0.82	0.83
	GHi	0.89	0.87	0.85	0.91	0.89	0.88	0.89	0.87	0.86	0.91	0.90	0.88
	SE	0.73	0.76	0.78	0.75	0.79	0.81	0.78	0.80	0.82	0.77	0.80	0.82
	SP	0.90	0.88	0.85	0.94	0.91	0.89	0.89	0.87	0.85	0.94	0.91	0.89
	Q	0.85	0.84	0.82	0.88	0.86	0.86	0.86	0.84	0.83	0.89	0.87	0.86
NCV	**TP**	528	681	828	591	776	939	549	722	871	602	773	940
	FP	77	106	138	20	33	47	97	125	149	26	42	64
	TN	1308	1279	1247	1365	1352	1338	1288	1260	1236	1359	1343	1321
	FN	124	154	176	61	59	65	103	113	133	50	62	64
	AUC	0.95	0.94	0.94	0.99	0.99	0.99	0.96	0.95	0.95	0.99	0.99	0.98
	C	0.77	0.74	0.73	0.91	0.91	0.90	0.77	0.77	0.76	0.91	0.90	0.89
	GHa	0.84	0.84	0.84	0.93	0.94	0.94	0.84	0.85	0.86	0.94	0.93	0.93
	GHi	0.93	0.91	0.88	0.97	0.96	0.96	0.93	0.91	0.90	0.97	0.96	0.95
	SE	0.80	0.81	0.82	0.90	0.93	0.93	0.84	0.86	0.86	0.92	0.92	0.93
	SP	0.94	0.92	0.90	0.98	0.97	0.96	0.93	0.91	0.89	0.98	0.97	0.95
	Q	0.90	0.88	0.86	0.96	0.95	0.95	0.90	0.89	0.88	0.96	0.95	0.94

CV: cross validated models(5-fold), NCV: Non cross validated models, TP: True positives, FP: False positives, TN: True negatives, FN: False negatives, AUC: area under curve of false positive rate *vs* true positive rate plot, *C*: Matthews correlation coefficient, GHa: GH score for actives(Blockers), GHi: GH scores for inactives(Non-blockers), SE: sensitivity, SP: specificity, Q: overall accuracy

Dataset 1: Blockers = IC_{50} < 3μM; Dataset 2: Blockers = IC_{50} < 6μM; Dataset 3: Blockers = IC_{50} < 10μ?

Non-blockers: FDA approved drugs and literature compounds with IC_{50} > 30μM

Table S1b. Classification statistics of LDA and SVM models generated using ECFP_6 and FCFP_6 fingerprints(Models from dataset 3 were highlighted)

		ECFP_6						FCFP_6					
		LDA1	**LDA2**	**LDA3**	**SVM1**	**SVM2**	**SVM3**	**LDA1**	**LDA2**	**LDA3**	**SVM1**	**SVM2**	**SVM3**
CV	**TP**	481	637	**793**	494	651	**811**	475	641	**805**	483	667	**821**
	FP	143	174	**201**	93	133	**151**	151	184	**212**	79	131	**152**
	TN	1244	1213	**1186**	1294	1254	**1236**	1236	1203	**1175**	1308	1256	**1235**
	FN	171	198	**211**	158	184	**193**	177	194	**199**	169	168	**183**
	AUC	0.90	0.90	**0.89**	0.93	0.93	**0.93**	0.91	0.90	**0.90**	0.94	0.93	**0.93**
	C	0.64	0.64	**0.64**	0.71	0.69	**0.70**	0.62	0.64	**0.64**	0.71	0.71	**0.71**
	GHa	0.75	0.77	**0.79**	0.80	0.80	**0.83**	0.74	0.77	**0.79**	0.80	0.82	**0.83**
	GHi	0.88	0.87	**0.85**	0.91	0.89	**0.88**	0.88	0.86	**0.85**	0.91	0.90	**0.88**
	SE	0.73	0.76	**0.79**	0.75	0.78	**0.81**	0.72	0.76	**0.80**	0.74	0.79	**0.81**
	SP	0.90	0.87	**0.85**	0.93	0.90	**0.89**	0.89	0.87	**0.85**	0.94	0.91	**0.89**
	Q	0.84	0.83	**0.82**	0.87	0.86	**0.86**	0.83	0.83	**0.82**	0.88	0.87	**0.86**
NCV	**TP**	532	677	**827**	597	780	**940**	534	696	**862**	596	771	**939**
	FP	83	114	**141**	19	31	**53**	100	127	**156**	17	33	**47**
	TN	1304	1273	**1246**	1368	1356	**1334**	1287	1260	**1231**	1370	1354	**1340**
	FN	120	158	**177**	55	55	**64**	118	139	**142**	56	64	**65**
	AUC	0.96	0.95	**0.94**	0.99	0.99	**0.99**	0.96	0.95	**0.95**	0.99	0.99	**0.99**
	C	0.76	0.74	**0.73**	0.91	0.92	**0.90**	0.75	0.74	**0.74**	0.91	0.90	**0.90**
	GHa	0.84	0.83	**0.84**	0.94	0.94	**0.94**	0.83	0.83	**0.85**	0.94	0.94	**0.94**
	GHi	0.92	0.90	**0.89**	0.97	0.97	**0.96**	0.92	0.90	**0.89**	0.97	0.96	**0.96**
	SE	0.81	0.81	**0.82**	0.91	0.93	**0.93**	0.81	0.83	**0.85**	0.91	0.92	**0.93**
	SP	0.94	0.92	**0.90**	0.99	0.98	**0.96**	0.93	0.91	**0.89**	0.99	0.97	**0.97**
	Q	0.90	0.87	**0.87**	0.96	0.96	**0.95**	0.89	0.88	**0.87**	0.96	0.95	**0.95**

CV: cross validated models(5-fold), NCV: Non cross validated models, TP: True positives, FP: False positives, TN: True negatives, FN: False negatives, AUC: area under curve of false positive rate *vs* true positive rate plot, *C*: Matthews correlation coefficient, GHa: GH score for actives(Blockers), GHi: GH scores for inactives(Non-blockers), SE: sensitivity, SP: specificity, Q: overall accuracy

Dataset 1: Blockers = IC_{50} < 3μM; Dataset 2: Blockers = IC_{50} < 6μM; Dataset 3: Blockers = IC_{50} < 10μ?

Non-blockers: FDA approved drugs and literature compounds with IC_{50} > 30μM

Table S2a. Classification statistics of individual RLB and PC SVM models using FCFP_6 fingerprints.

		FCFP_6	
		RLB	**PC**
CV	**TP**	517	215
	FP	47	46
	TN	1292	1261
	FN	88	197
	AUC	0.97	0.88
	C	0.84	0.58
	GHa	0.89	0.67
	GHi	0.95	0.91
	SE	0.85	0.52
	SP	0.96	0.96
	Q	0.93	0.86
NCV	**TP**	569	334
	FP	16	14
	TN	1323	1293
	FN	36	78
	AUC	0.99	0.98
	C	0.94	0.85
	GHa	0.96	0.88
	GHi	0.98	0.96
	SE	0.94	0.81
	SP	0.99	0.99
	Q	0.97	0.95

CV: cross-validated models(5-fold), NCV: Non cross-validated models, TP: True positives, FP: False positives, TN: True negatives, FN: False negatives, AUC: area under curve of false positive rate *vs* true positive rate plot, *C*: Matthews correlation coefficient, GHa: GH score for actives(Blockers), GHi: GH scores for inactives(Non-blockers), SE: sensitivity, SP: specificity, Q: overall accuracy; Dataset 3: Blockers = IC_{50} < 10µ? , Non-blockers: FDA-approved drugs and literature compounds with IC_{50} > 30µM.

RLB blockers : 605; Non-blockers : 1339

PC blockers : 412; Non-blockers 1307

Table S2b. Predictions of all the compounds of the dataset by mixed, RLB and PC models generated using SVM and FCFP_6 fingerprints

	Predicted active	%Predicted active	Predicted inactive	%Predicted inactive	% Total
Mixed model	1026	92%	1468	96%	94%
RLB model	701	63%	1495	98%	83%
PC model	526	47%	1497	98%	76%
Total Blockers : 1112; Total Non-blockers : 1532					