

Supporting Information

Large-Scale Similarity Search Profiling of ChEMBL Compound Data Sets

Kathrin Heikamp and Jürgen Bajorath

Department of Life Science Informatics, B-IT, LIMES Program Unit Chemical Biology
and Medicinal Chemistry, Rheinische Friedrich-Wilhelms-Universität Bonn,
Dahlmannstr.2, D-53113 Bonn, Germany.

Supplementary Tables S1 and S2

Supplementary Figures S1 – S5

Table S1. Activity classes.

No.	Target ID	Target name	No. cpds.	ECFP4 1NN	MACCS 1NN
1	3	Phosphodiesterase 5A	247	51.6	47.8
2	11910	Phosphodiesterase 7A	75	78.0	48.9
3	19	Estrogen receptor alpha	289	55.0	38.1
4	11359	Phosphodiesterase 4D	152	70.9	46.6
5	165	HERG	499	21.1	13.6
6	4	Voltage-gated T-type calcium channel alpha-1H subunit	51	86.5	74.6
7	6	Dihydrofolate reductase	132	54.6	46.1
8	28	Thymidylate synthase	103	59.2	39.7
9	11536	Ghrelin receptor	493	62.7	41.5
10	8	Tyrosine-protein kinase ABL	170	54.9	37.3
11	197	Platelet-derived growth factor receptor beta	81	70.4	53.6
12	10434	Tyrosine-protein kinase SRC	442	51.1	26.6
13	12670	Tyrosine-protein kinase receptor FLT3	122	43.9	30.7
14	20014	Serine/threonine-protein kinase Aurora-A	124	62.8	42.4
15	10197	Glycogen synthase kinase-3 beta	339	47.1	27.4
16	234	Insulin-like growth factor I receptor	303	76.5	43.7
17	10532	Butyrylcholinesterase	278	51.4	34.1
18	9	Epidermal growth factor receptor erbB1	547	43.3	17.0
19	188	Receptor protein-tyrosine kinase erbB-2	264	57.6	24.6
20	10980	Vascular endothelial growth factor receptor 2	985	35.5	16.7
21	12261	c-Jun N-terminal kinase 1	208	78.0	46.1
22	10188	MAP kinase p38 alpha	736	29.5	16.1
23	100414	Mitogen-activated protein kinase kinase kinase 8	95	90.6	45.8
24	15	Carbonic anhydrase II	379	17.6	15.8
25	11451	Hepatocyte growth factor receptor	182	54.2	27.0
26	13001	Matrix metalloproteinase-2	186	33.1	24.9
27	10140	Tyrosine-protein kinase LCK	350	47.0	20.1
28	11	Thrombin	631	41.9	33.0
29	11942	Urokinase-type plasminogen activator	73	82.8	60.9
30	194	Coagulation factor X	1342	46.1	22.6
31	12725	Matrilysin	55	91.4	87.9
32	235	Leukocyte elastase	265	43.9	29.0
33	193	Coagulation factor IX	94	85.6	60.9
34	12952	Carbonic anhydrase IX	169	27.8	19.9
35	12209	Carbonic anhydrase XII	119	34.5	33.6
36	93	Acetylcholinesterase	422	33.3	17.7
37	10193	Carbonic anhydrase I	61	17.6	17.4
38	10656	Beta amyloid A4 protein	78	76.3	65.8

39	11267	Steryl-sulfatase precursor	74	50.5	47.8
40	65	Cytochrome P450 19A1	336	31.8	21.3
41	174	Estrogen receptor beta	334	49.7	36.8
42	25	Glucocorticoid receptor	485	54.4	37.6
43	11512	Estradiol 17-beta-dehydrogenase 1	127	53.3	35.8
44	36	Progesterone receptor	330	59.6	38.3
45	56	Androgen Receptor	501	54.0	25.3
46	20113	LXR-beta	140	75.3	37.8
47	19904	Calcitonin gene-related peptide type 1 receptor	228	80.2	45.0
48	34	Fibronectin receptor beta	91	93.2	68.0
49	10918	Tyrosine-protein kinase ITK/TSK	164	83.3	72.1
50	12913	Tyrosine-protein kinase TIE-2	104	64.7	38.9
51	11727	Epoxide hydratase	488	41.4	25.0
52	43	Beta-2 adrenergic receptor	150	48.2	34.7
53	51	Serotonin 1a (5-HT1a) receptor	565	34.8	24.8
54	219	Muscarinic acetylcholine receptor M3	252	64.3	46.2
55	61	Muscarinic acetylcholine receptor M1	111	38.1	25.9
56	138	Nociceptin receptor	728	50.8	29.2
57	130	Dopamine D3 receptor	388	49.0	26.6
58	12825	C-X-C chemokine receptor type 3	396	49.4	31.7
59	105	Serotonin 1d (5-HT1d) receptor	67	57.1	26.4
60	280	Adenosine A3 receptor	1075	45.5	28.2
61	125	Alpha-1a adrenergic receptor	153	48.3	33.1
62	72	Dopamine D2 receptor	309	27.0	13.0
63	106	Serotonin 1b (5-HT1b) receptor	61	69.2	50.7
64	127	Histamine H1 receptor	147	52.3	37.5
65	10209	Serotonin 7 (5-HT7) receptor	320	47.8	33.9
66	107	Serotonin 2a (5-HT2a) receptor	361	35.2	20.9
67	121	Serotonin transporter	905	26.5	14.7
68	90	Dopamine D4 receptor	198	33.4	24.3
69	88	Dopamine D1 receptor	176	66.3	57.1
70	108	Serotonin 2c (5-HT2c) receptor	356	30.7	18.7
71	218	Alpha-2c adrenergic receptor	73	65.1	53.7
72	10627	Serotonin 6 (5-HT6) receptor	799	46.3	34.4
73	19905	Melanin-concentrating hormone receptor 1	1793	29.0	16.9
74	10624	Serotonin 5a (5-HT5a) receptor	56	84.3	72.5
75	227	Serotonin 2b (5-HT2b) receptor	82	39.3	41.2
76	52	Alpha-2a adrenergic receptor	74	40.5	28.0
77	100	Norepinephrine transporter	691	41.2	25.9
78	155	Dopamine transporter	257	42.1	25.9
79	103	Alpha-1d adrenergic receptor	66	75.8	59.3
80	10911	Phenylethanolamine N-methyltransferase	109	78.9	60.2
81	11336	Neuropeptide Y receptor type 5	367	60.1	40.7

82	55	Arachidonate 5-lipoxygenase	202	47.7	34.5
83	10102	5-lipoxygenase activating protein	54	100.0	94.7
84	96	Cyclooxygenase-1	88	50.9	35.3
85	101277	Prostaglandin E synthase	63	76.6	45.8
86	20174	G protein-coupled receptor 44	427	56.4	37.9
87	271	Thyroid hormone receptor beta-1	189	67.3	30.6
88	118	Gonadotropin-releasing hormone receptor	625	60.9	43.7
89	11140	Dipeptidyl peptidase IV	983	36.9	17.8
90	117	Somatostatin receptor 2	92	87.4	71.1
91	17045	Cytochrome P450 3A4	156	39.5	23.2
92	104	Monoamine oxidase B	140	38.5	21.5
93	11541	Cytochrome P450 11B2	184	53.4	39.6
94	12949	Cytochrome P450 17A1	95	55.9	33.3
95	126	Cyclooxygenase-2	349	45.0	29.4
96	11489	11-beta-hydroxysteroid dehydrogenase 1	946	25.1	15.2
97	68	Inosine-5'-monophosphate dehydrogenase 2	125	66.0	22.6
98	11225	Renin	550	67.2	37.9
99	10280	Histamine H3 receptor	1174	36.6	20.3
100	11881	GABA receptor beta-3 subunit	63	83.8	76.5
101	79	Steroid 5-alpha-reductase 1	84	62.1	60.6
102	80	FK506-binding protein 1A	71	91.0	71.0
103	87	Cannabinoid CB1 receptor	1253	31.8	14.9
104	259	Cannabinoid CB2 receptor	1027	30.0	15.9
105	12252	Beta-secretase 1	536	54.4	34.1
106	11507	Carboxylesterase 2	62	20.6	46.6
107	136	Delta opioid receptor	362	53.3	29.4
108	129	Mu opioid receptor	564	40.4	25.2
109	10184	Neurokinin 2 receptor	91	66.3	42.1
110	133	Peroxisome proliferator-activated receptor gamma	252	60.3	50.5
111	252	Adenosine A2a receptor	1030	40.7	21.3
112	11682	Glycine transporter 1	174	78.2	47.3
113	134	Vasopressin V1a receptor	188	67.6	52.4
114	112	Vasopressin V2 receptor	124	71.1	46.5
115	116	Oxytocin receptor	161	55.9	31.9
116	278	Adenosine A2b receptor	493	56.0	44.2
117	114	Adenosine A1 receptor	347	31.9	21.1
118	10599	Phosphodiesterase 4B	67	66.9	44.5
119	176	Purinergic receptor P2Y12	523	89.2	81.4
120	11265	Somatostatin receptor 5	130	70.6	49.0
121	10475	Neuropeptide Y receptor type 1	174	72.1	49.1
122	12697	Histone deacetylase 1	568	37.3	20.2
123	124	Corticotropin releasing factor receptor 1	402	65.8	45.0
124	11290	Histamine H4 receptor	148	65.8	56.1

125	250	Neurokinin 1 receptor	492	60.3	17.8
126	137	Kappa opioid receptor	481	63.0	28.4
127	163	Peroxisome proliferator-activated receptor alpha	94	76.9	56.9
128	12227	Peroxisome proliferator-activated receptor delta	117	77.5	56.1
129	12679	C5a anaphylatoxin chemotactic receptor	170	76.3	47.6
130	10472	Cholecystokinin A receptor	80	86.2	55.3
131	262	Prostanoid EP4 receptor	87	86.0	67.9
132	246	Thromboxane A2 receptor	72	83.4	64.8
133	12659	Prostanoid DP receptor	82	92.8	72.8
134	10329	Prostanoid EP3 receptor	327	77.6	64.5
135	146	Glucagon receptor	199	72.0	56.7
136	157	Dihydroorotate dehydrogenase	74	55.9	40.0
137	12666	Serine/threonine-protein kinase AKT	355	75.3	44.1
138	10580	C-C chemokine receptor type 5	623	53.3	25.6
139	282	Calcium sensing receptor	202	77.7	56.3
140	10579	C-C chemokine receptor type 4	142	60.7	39.9
141	10548	C-C chemokine receptor type 3	346	66.5	54.4
142	11575	C-C chemokine receptor type 2	605	65.4	37.0
143	18061	Sodium channel protein type IX alpha subunit	200	69.9	46.9
144	10473	C-X-C chemokine receptor type 4	79	82.3	65.7
145	11156	Bradykinin B1 receptor	452	53.6	24.1
146	237	Leukotriene A4 hydrolase	160	58.1	41.7
147	276	Phosphodiesterase 4A	73	54.1	38.5
148	11534	Cathepsin S	625	57.0	37.3
149	100100	Voltage-gated T-type calcium channel alpha-1G subunit	96	64.7	39.8
150	10695	Serine/threonine-protein kinase AKT2	56	65.6	42.1
151	11362	PI3-kinase p110-alpha subunit	86	84.7	55.3
152	10198	Voltage-gated potassium channel subunit Kv1.5	201	63.6	38.5
153	11365	Cytochrome P450 2D6	130	42.0	28.8
154	10495	Cathepsin K	520	43.0	19.8
155	10260	Vanilloid receptor	500	40.0	20.8
156	175	Equilibrative nucleoside transporter 1	117	71.1	58.9
157	179	Cysteinyl leukotriene receptor 1	81	83.5	54.2
158	184	Glutamate carboxypeptidase II	88	66.8	60.1
159	11024	Matrix metalloproteinase 13	293	47.0	36.4
160	11473	ADAM17	275	76.2	51.3
161	11109	Matrix metalloproteinase 3	108	33.2	27.5
162	10781	Serine/threonine-protein kinase Aurora-B	57	85.1	49.9
163	10498	Cathepsin L	161	64.1	48.3
164	12967	Serine/threonine-protein kinase Chk1	384	52.1	27.4

165	242	Aldose reductase	116	66.7	72.9
166	10982	Neurokinin 3 receptor	77	79.4	56.1
167	251	Platelet activating factor receptor	51	86.3	74.2
168	12911	Cytochrome P450 2C9	50	59.5	35.2
169	275	Retinoid X receptor alpha	85	92.5	79.2
170	10003	Cathepsin D	69	70.8	53.1
171	12968	Orexin receptor 2	100	67.9	53.3
172	10034	Bradykinin B2 receptor	60	81.2	36.4
173	10056	DNA-dependent protein kinase	146	88.8	81.6
174	10074	Chymase	83	84.3	64.8
175	10087	Deoxycytidine kinase	53	95.9	85.4
176	10131	Caspase-3	238	76.6	47.5
177	11624	Caspase-1	108	73.1	49.6
178	10142	Melanocortin receptor 4	1374	64.4	43.3
179	11006	Melanocortin receptor 5	55	80.5	65.8
180	10144	Bone morphogenetic protein 1	81	97.2	84.0
181	100579	Nicotinic acid receptor 1	170	63.2	43.3
182	11084	Melatonin receptor 1A	77	66.9	50.8
183	13055	Quinone reductase 1	55	61.2	43.5
184	10185	Adenosine kinase	99	75.0	36.9
185	13004	TGF-beta receptor type I	151	66.4	27.4
186	100126	Serine/threonine-protein kinase B-raf	144	68.7	22.6
187	11678	Cyclin-dependent kinase 2	54	83.1	49.9
188	11400	FK506 binding protein 12	209	79.2	54.1
189	11402	Furin	73	91.3	91.0
190	10231	Gamma-secretase subunit APH-1B	53	90.2	60.3
191	13061	Protein-tyrosine phosphatase 1B	241	51.5	32.4
192	10702	Methionine aminopeptidase 2	132	80.4	54.7
193	10608	Acetyl-CoA carboxylase 2	133	84.5	57.1
194	12955	Calpain 1	62	71.9	54.6
195	10378	Cathepsin B	105	57.5	41.3
196	10417	P2X purinoceptor 7	137	70.7	43.0
197	13000	Matrix metalloproteinase-1	173	31.9	22.7
198	10517	Cholesteryl ester transfer protein	188	74.1	47.9
199	11409	Dual specificity mitogen-activated protein kinase kinase 1	118	76.1	66.0
200	10544	Purinergic receptor P2Y1	65	73.8	52.5
201	10547	C-C chemokine receptor type 1	144	78.3	71.7
202	10561	Oligopeptide transporter small intestine isoform	80	26.0	24.8
203	10582	Cytosolic phospholipase A2	136	95.1	76.2
204	11291	Anandamide amidohydrolase	253	56.7	32.9
205	10584	Phospholipase A2 group IIA	84	79.5	72.1
206	10612	3-phosphoinositide dependent protein kinase-1	54	87.4	76.8

207	10653	Proteinase activated receptor 1	238	88.3	55.6
208	10811	Rho-associated protein kinase 1	64	66.8	50.3
209	10839	Serine/threonine-protein kinase PIM1	77	60.5	42.6
210	10752	Inhibitor of nuclear factor kappa B kinase beta subunit	103	65.9	33.6
211	10773	Interleukin-8 receptor B	274	66.4	51.0
212	100077	Cell division cycle 7-related protein kinase	77	89.1	60.8
213	11631	Sphingosine 1-phosphate receptor Edg-1	133	66.1	51.4
214	10845	Phospholipase D1	58	90.4	81.1
215	11307	Histone deacetylase 6	100	41.8	38.3
216	11635	Protein kinase C alpha	157	79.9	72.5
217	12665	Protein kinase C theta	105	93.1	55.2
218	10892	Integrin alpha-4	135	85.3	68.2
219	12214	Tyrosine-protein kinase ZAP-70	57	84.4	57.8
220	10927	Urotensin II receptor	120	69.8	38.9
221	11902	Nerve growth factor receptor Trk-A	57	81.8	68.8
222	12071	Cyclin-dependent kinase 1	52	82.8	48.5
223	11871	Matrix metalloproteinase 12	89	67.8	48.5
224	12592	Matrix metalloproteinase 9	110	39.5	26.6
225	11110	Matrix metalloproteinase 8	56	43.0	27.8
226	11037	Nitric-oxide synthase, brain	62	41.9	33.3
227	12425	Nitric oxide synthase, inducible	170	62.8	39.6
228	11061	Motilin receptor	73	94.3	83.3
229	11082	MAP kinase-activated protein kinase 2	193	65.3	44.8
230	11085	Melatonin receptor 1B	166	62.2	39.4
231	11096	Sodium/hydrogen exchanger 1	97	77.9	88.5
232	11120	Squalene synthetase	66	85.1	69.2
233	12268	Dipeptidyl peptidase II	59	59.3	53.1
234	11442	Liver glycogen phosphorylase	347	70.3	42.7
235	11242	Focal adhesion kinase 1	97	91.9	63.5
236	11269	LXR-alpha	120	71.2	50.8
237	11280	Metabotropic glutamate receptor 5	254	43.7	22.9
238	11279	Metabotropic glutamate receptor 1	188	64.2	33.0
239	11415	Fructose-1,6-bisphosphatase	191	70.7	59.9
240	101400	Smoothed homolog	97	84.3	43.2
241	11488	Estradiol 17-beta-dehydrogenase 3	106	55.5	25.6
242	100643	Indoleamine 2,3-dioxygenase	97	43.1	30.4
243	11653	Heparanase	83	76.9	72.2
244	11663	Poly [ADP-ribose] polymerase-1	394	50.7	21.4
245	11723	Thymidine phosphorylase	61	56.1	57.5
246	11758	Glucagon-like peptide receptor	55	95.8	87.7
247	100098	Serine/threonine-protein kinase WEE1	181	95.2	65.9
248	12622	Telomerase reverse transcriptase	122	62.8	46.5
249	12690	Purine nucleoside phosphorylase	97	83.0	72.3

250	12840	Macrophage colony stimulating factor receptor	117	64.5	30.5
251	12909	Ileal bile acid transporter	133	90.4	89.5
252	20092	Sodium/glucose cotransporter 2	270	76.0	56.2
253	20130	Inhibitor of apoptosis protein 3	66	90.3	86.0
254	100010	Phosphodiesterase 10A	100	91.9	74.3
255	100166	Kinesin-like protein 1	217	40.4	19.0
256	101174	Pituitary adenylate cyclase-activating polypeptide type I receptor	81	100.0	100.0
257	100436	C-C chemokine receptor type 8	101	87.4	47.9
258	100450	Glutaminy-peptide cyclotransferase	91	56.6	53.4
259	100594	Autotaxin	73	48.6	41.6
260	100666	Malonyl-CoA decarboxylase	184	56.8	40.8
261	100862	Metastin receptor	69	93.9	75.5
262	101219	Secreted frizzled-related protein 1	67	99.1	67.1
263	101234	Acyl-CoA desaturase	183	68.9	56.1
264	101395	IgG receptor FcRn large subunit p51	83	100.0	100.0
265	101509	Elongation of very long chain fatty acids protein 6	106	82.3	62.7
266	101502	Egl nine homolog 1	59	76.1	53.5

Activity classes are consecutively numbered, the ChEMBL target ID and target name are provided, and the number of compounds per class is given. For each class, ECFP4 and MACCS 1NN recovery rates (in %) are provided for selection sets equal to the number of ADCs.

Table S2. Average ROC AUC values.

No.	Target ID	Target name	ECFP4	MACCS
			<i>INN</i>	<i>INN</i>
1	3	Phosphodiesterase 5A	0.0690	0.0717
2	11910	Phosphodiesterase 7A	0.0959	0.0903
3	19	Estrogen receptor alpha	0.0832	0.0712
4	11359	Phosphodiesterase 4D	0.0890	0.0721
5	165	HERG	0.0299	0.0362
6	4	Voltage-gated T-type calcium channel alpha-1H subunit	0.0955	0.0952
7	6	Dihydrofolate reductase	0.0861	0.0860
8	28	Thymidylate synthase	0.0842	0.0736
9	11536	Ghrelin receptor	0.0801	0.0676
10	8	Tyrosine-protein kinase ABL	0.0792	0.0755
11	197	Platelet-derived growth factor receptor beta	0.0903	0.0846
12	10434	Tyrosine-protein kinase SRC	0.0741	0.0585
13	12670	Tyrosine-protein kinase receptor FLT3	0.0868	0.0778
14	20014	Serine/threonine-protein kinase Aurora-A	0.0868	0.0763
15	10197	Glycogen synthase kinase-3 beta	0.0649	0.0539
16	234	Insulin-like growth factor I receptor	0.0885	0.0811
17	10532	Butyrylcholinesterase	0.0766	0.0642
18	9	Epidermal growth factor receptor erbB1	0.0761	0.0468
19	188	Receptor protein-tyrosine kinase erbB-2	0.0822	0.0600
20	10980	Vascular endothelial growth factor receptor 2	0.0544	0.0386
21	12261	c-Jun N-terminal kinase 1	0.0916	0.0776
22	10188	MAP kinase p38 alpha	0.0512	0.0378
23	100414	Mitogen-activated protein kinase kinase kinase 8	0.0985	0.0926
24	15	Carbonic anhydrase II	0.0632	0.0574
25	11451	Hepatocyte growth factor receptor	0.0833	0.0605
26	13001	Matrix metalloproteinase-2	0.0750	0.0597
27	10140	Tyrosine-protein kinase LCK	0.0725	0.0608
28	11	Thrombin	0.0734	0.0639
29	11942	Urokinase-type plasminogen activator	0.0962	0.0923
30	194	Coagulation factor X	0.0658	0.0517
31	12725	Matriptase	0.0962	0.0926
32	235	Leukocyte elastase	0.0709	0.0518
33	193	Coagulation factor IX	0.0976	0.0917
34	12952	Carbonic anhydrase IX	0.0796	0.0725
35	12209	Carbonic anhydrase XII	0.0837	0.0803
36	93	Acetylcholinesterase	0.0556	0.0483
37	10193	Carbonic anhydrase I	0.0613	0.0549
38	10656	Beta amyloid A4 protein	0.0872	0.0843

39	11267	Steryl-sulfatase precursor	0.0851	0.0832
40	65	Cytochrome P450 19A1	0.0683	0.0576
41	174	Estrogen receptor beta	0.0896	0.0793
42	25	Glucocorticoid receptor	0.0712	0.0572
43	11512	Estradiol 17-beta-dehydrogenase 1	0.0972	0.0798
44	36	Progesterone receptor	0.0782	0.0663
45	56	Androgen Receptor	0.0788	0.0528
46	20113	LXR-beta	0.0911	0.0739
47	19904	Calcitonin gene-related peptide type 1 receptor	0.0925	0.0952
48	34	Fibronectin receptor beta	0.0996	0.0976
49	10918	Tyrosine-protein kinase ITK/TSK	0.0972	0.0967
50	12913	Tyrosine-protein kinase TIE-2	0.0958	0.0837
51	11727	Epoxide hydratase	0.0746	0.0652
52	43	Beta-2 adrenergic receptor	0.0872	0.0740
53	51	Serotonin 1a (5-HT1a) receptor	0.0659	0.0585
54	219	Muscarinic acetylcholine receptor M3	0.0763	0.0748
55	61	Muscarinic acetylcholine receptor M1	0.0631	0.0567
56	138	Nociceptin receptor	0.0707	0.0688
57	130	Dopamine D3 receptor	0.0794	0.0705
58	12825	C-X-C chemokine receptor type 3	0.0747	0.0668
59	105	Serotonin 1d (5-HT1d) receptor	0.0819	0.0772
60	280	Adenosine A3 receptor	0.0688	0.0688
61	125	Alpha-1a adrenergic receptor	0.0738	0.0648
62	72	Dopamine D2 receptor	0.0656	0.0521
63	106	Serotonin 1b (5-HT1b) receptor	0.0876	0.0866
64	127	Histamine H1 receptor	0.0731	0.0700
65	10209	Serotonin 7 (5-HT7) receptor	0.0760	0.0648
66	107	Serotonin 2a (5-HT2a) receptor	0.0539	0.0482
67	121	Serotonin transporter	0.0473	0.0404
68	90	Dopamine D4 receptor	0.0763	0.0786
69	88	Dopamine D1 receptor	0.0834	0.0821
70	108	Serotonin 2c (5-HT2c) receptor	0.0505	0.0455
71	218	Alpha-2c adrenergic receptor	0.0828	0.0840
72	10627	Serotonin 6 (5-HT6) receptor	0.0764	0.0743
73	19905	Melanin-concentrating hormone receptor 1	0.0528	0.0523
74	10624	Serotonin 5a (5-HT5a) receptor	0.0964	0.0960
75	227	Serotonin 2b (5-HT2b) receptor	0.0797	0.0746
76	52	Alpha-2a adrenergic receptor	0.0620	0.0524
77	100	Norepinephrine transporter	0.0672	0.0538
78	155	Dopamine transporter	0.0740	0.0585
79	103	Alpha-1d adrenergic receptor	0.0858	0.0854
80	10911	Phenylethanolamine N-methyltransferase	0.0951	0.0908
81	11336	Neuropeptide Y receptor type 5	0.0783	0.0748

82	55	Arachidonate 5-lipoxygenase	0.0636	0.0535
83	10102	5-lipoxygenase activating protein	0.1000	0.1000
84	96	Cyclooxygenase-1	0.0771	0.0692
85	101277	Prostaglandin E synthase	0.0887	0.0791
86	20174	G protein-coupled receptor 44	0.0808	0.0692
87	271	Thyroid hormone receptor beta-1	0.0943	0.0745
88	118	Gonadotropin-releasing hormone receptor	0.0749	0.0726
89	11140	Dipeptidyl peptidase IV	0.0635	0.0522
90	117	Somatostatin receptor 2	0.0984	0.0940
91	17045	Cytochrome P450 3A4	0.0445	0.0443
92	104	Monoamine oxidase B	0.0684	0.0668
93	11541	Cytochrome P450 11B2	0.0900	0.0769
94	12949	Cytochrome P450 17A1	0.0918	0.0789
95	126	Cyclooxygenase-2	0.0697	0.0627
96	11489	11-beta-hydroxysteroid dehydrogenase 1	0.0510	0.0370
97	68	Inosine-5'-monophosphate dehydrogenase 2	0.0877	0.0594
98	11225	Renin	0.0852	0.0737
99	10280	Histamine H3 receptor	0.0692	0.0597
100	11881	GABA receptor beta-3 subunit	0.0918	0.0917
101	79	Steroid 5-alpha-reductase 1	0.0900	0.0885
102	80	FK506-binding protein 1A	0.0973	0.0958
103	87	Cannabinoid CB1 receptor	0.0589	0.0406
104	259	Cannabinoid CB2 receptor	0.0471	0.0366
105	12252	Beta-secretase 1	0.0769	0.0643
106	11507	Carboxylesterase 2	0.0882	0.0902
107	136	Delta opioid receptor	0.0755	0.0685
108	129	Mu opioid receptor	0.0672	0.0652
109	10184	Neurokinin 2 receptor	0.0761	0.0690
110	133	Peroxisome proliferator-activated receptor gamma	0.0831	0.0728
111	252	Adenosine A2a receptor	0.0679	0.0542
112	11682	Glycine transporter 1	0.0900	0.0869
113	134	Vasopressin V1a receptor	0.0882	0.0770
114	112	Vasopressin V2 receptor	0.0956	0.0869
115	116	Oxytocin receptor	0.0888	0.0810
116	278	Adenosine A2b receptor	0.0803	0.0744
117	114	Adenosine A1 receptor	0.0703	0.0574
118	10599	Phosphodiesterase 4B	0.0937	0.0761
119	176	Purinergic receptor P2Y12	0.0931	0.0922
120	11265	Somatostatin receptor 5	0.0869	0.0780
121	10475	Neuropeptide Y receptor type 1	0.0782	0.0749
122	12697	Histone deacetylase 1	0.0776	0.0531
123	124	Corticotropin releasing factor receptor 1	0.0915	0.0870
124	11290	Histamine H4 receptor	0.0875	0.0901

125	250	Neurokinin 1 receptor	0.0874	0.0575
126	137	Kappa opioid receptor	0.0782	0.0662
127	163	Peroxisome proliferator-activated receptor alpha	0.0951	0.0850
128	12227	Peroxisome proliferator-activated receptor delta	0.0981	0.0885
129	12679	C5a anaphylatoxin chemotactic receptor	0.0916	0.0825
130	10472	Cholecystokinin A receptor	0.0918	0.0877
131	262	Prostanoid EP4 receptor	0.0983	0.0888
132	246	Thromboxane A2 receptor	0.0911	0.0894
133	12659	Prostanoid DP receptor	0.0957	0.0937
134	10329	Prostanoid EP3 receptor	0.0962	0.0917
135	146	Glucagon receptor	0.0907	0.0841
136	157	Dihydroorotate dehydrogenase	0.0919	0.0874
137	12666	Serine/threonine-protein kinase AKT	0.0877	0.0834
138	10580	C-C chemokine receptor type 5	0.0745	0.0589
139	282	Calcium sensing receptor	0.0947	0.0911
140	10579	C-C chemokine receptor type 4	0.0862	0.0786
141	10548	C-C chemokine receptor type 3	0.0898	0.0873
142	11575	C-C chemokine receptor type 2	0.0890	0.0820
143	18061	Sodium channel protein type IX alpha subunit	0.0904	0.0813
144	10473	C-X-C chemokine receptor type 4	0.0955	0.0973
145	11156	Bradykinin B1 receptor	0.0800	0.0628
146	237	Leukotriene A4 hydrolase	0.0930	0.0841
147	276	Phosphodiesterase 4A	0.0893	0.0657
148	11534	Cathepsin S	0.0784	0.0576
149	100100	Voltage-gated T-type calcium channel alpha-1G subunit	0.0948	0.0868
150	10695	Serine/threonine-protein kinase AKT2	0.0822	0.0795
151	11362	PI3-kinase p110-alpha subunit	0.0939	0.0936
152	10198	Voltage-gated potassium channel subunit Kv1.5	0.0844	0.0709
153	11365	Cytochrome P450 2D6	0.0546	0.0556
154	10495	Cathepsin K	0.0722	0.0482
155	10260	Vanilloid receptor	0.0734	0.0558
156	175	Equilibrative nucleoside transporter 1	0.0969	0.0976
157	179	Cysteinyl leukotriene receptor 1	0.0960	0.0858
158	184	Glutamate carboxypeptidase II	0.0955	0.0847
159	11024	Matrix metalloproteinase 13	0.0698	0.0607
160	11473	ADAM17	0.0946	0.0861
161	11109	Matrix metalloproteinase 3	0.0768	0.0688
162	10781	Serine/threonine-protein kinase Aurora-B	0.0876	0.0860
163	10498	Cathepsin L	0.0824	0.0732
164	12967	Serine/threonine-protein kinase Chk1	0.0723	0.0612
165	242	Aldose reductase	0.0828	0.0844

166	10982	Neurokinin 3 receptor	0.0913	0.0879
167	251	Platelet activating factor receptor	0.0964	0.0903
168	12911	Cytochrome P450 2C9	0.0691	0.0617
169	275	Retinoid X receptor alpha	0.0989	0.0930
170	10003	Cathepsin D	0.0870	0.0845
171	12968	Orexin receptor 2	0.0883	0.0861
172	10034	Bradykinin B2 receptor	0.0941	0.0841
173	10056	DNA-dependent protein kinase	0.0996	0.0992
174	10074	Chymase	0.0911	0.0847
175	10087	Deoxycytidine kinase	0.0995	0.0980
176	10131	Caspase-3	0.0960	0.0841
177	11624	Caspase-1	0.0966	0.0853
178	10142	Melanocortin receptor 4	0.0824	0.0805
179	11006	Melanocortin receptor 5	0.0970	0.0920
180	10144	Bone morphogenetic protein 1	0.0999	0.0997
181	100579	Nicotinic acid receptor 1	0.0892	0.0787
182	11084	Melatonin receptor 1A	0.0974	0.0907
183	13055	Quinone reductase 1)	0.0912	0.0853
184	10185	Adenosine kinase	0.0966	0.0950
185	13004	TGF-beta receptor type I	0.0932	0.0677
186	100126	Serine/threonine-protein kinase B-raf	0.0865	0.0718
187	11678	Cyclin-dependent kinase 2	0.0933	0.0848
188	11400	FK506 binding protein 12	0.0942	0.0906
189	11402	Furin	0.0969	0.0968
190	10231	Gamma-secretase subunit APH-1B	0.0999	0.0971
191	13061	Protein-tyrosine phosphatase 1B	0.0647	0.0518
192	10702	Methionine aminopeptidase 2	0.0905	0.0870
193	10608	Acetyl-CoA carboxylase 2	0.0978	0.0913
194	12955	Calpain 1	0.0974	0.0894
195	10378	Cathepsin B	0.0786	0.0670
196	10417	P2X purinoceptor 7	0.0826	0.0746
197	13000	Matrix metalloproteinase-1	0.0784	0.0699
198	10517	Cholesteryl ester transfer protein	0.0852	0.0814
199	11409	Dual specificity mitogen-activated protein kinase kinase 1	0.0911	0.0867
200	10544	Purinergic receptor P2Y1	0.0943	0.0889
201	10547	C-C chemokine receptor type 1	0.0951	0.0908
202	10561	Oligopeptide transporter small intestine isoform	0.0976	0.0951
203	10582	Cytosolic phospholipase A2	0.0978	0.0930
204	11291	Anandamide amidohydrolase	0.0797	0.0616
205	10584	Phospholipase A2 group IIA	0.0922	0.0902
206	10612	3-phosphoinositide dependent protein kinase-1	0.0981	0.0961

207	10653	Proteinase activated receptor 1	0.0938	0.0876
208	10811	Rho-associated protein kinase 1	0.0885	0.0868
209	10839	Serine/threonine-protein kinase PIM1	0.0840	0.0807
210	10752	Inhibitor of nuclear factor kappa B kinase beta subunit	0.0859	0.0738
211	10773	Interleukin-8 receptor B	0.0822	0.0743
212	100077	Cell division cycle 7-related protein kinase	0.0999	0.0990
213	11631	Sphingosine 1-phosphate receptor Edg-1	0.0907	0.0809
214	10845	Phospholipase D1	0.0973	0.0973
215	11307	Histone deacetylase 6	0.0775	0.0689
216	11635	Protein kinase C alpha	0.0916	0.0923
217	12665	Protein kinase C theta	0.0994	0.0961
218	10892	Integrin alpha-4	0.0989	0.0916
219	12214	Tyrosine-protein kinase ZAP-70	0.0935	0.0867
220	10927	Urotensin II receptor	0.0910	0.0865
221	11902	Nerve growth factor receptor Trk-A	0.0939	0.0897
222	12071	Cyclin-dependent kinase 1	0.0925	0.0849
223	11871	Matrix metalloproteinase 12	0.0849	0.0723
224	12592	Matrix metalloproteinase 9	0.0779	0.0652
225	11110	Matrix metalloproteinase 8	0.0815	0.0696
226	11037	Nitric-oxide synthase, brain	0.0737	0.0711
227	12425	Nitric oxide synthase, inducible	0.0751	0.0665
228	11061	Motilin receptor	0.0946	0.0970
229	11082	MAP kinase-activated protein kinase 2	0.0858	0.0782
230	11085	Melatonin receptor 1B	0.0941	0.0876
231	11096	Sodium/hydrogen exchanger 1	0.0998	0.0987
232	11120	Squalene synthetase	0.0913	0.0850
233	12268	Dipeptidyl peptidase II	0.0807	0.0829
234	11442	Liver glycogen phosphorylase	0.0920	0.0845
235	11242	Focal adhesion kinase 1	0.0996	0.0950
236	11269	LXR-alpha	0.0893	0.0836
237	11280	Metabotropic glutamate receptor 5	0.0779	0.0541
238	11279	Metabotropic glutamate receptor 1	0.0851	0.0781
239	11415	Fructose-1,6-bisphosphatase	0.0894	0.0886
240	101400	Smoothed homolog	0.0964	0.0923
241	11488	Estradiol 17-beta-dehydrogenase 3	0.0896	0.0873
242	100643	Indoleamine 2,3-dioxygenase	0.0712	0.0627
243	11653	Heparanase	0.0913	0.0912
244	11663	Poly [ADP-ribose] polymerase-1	0.0749	0.0585
245	11723	Thymidine phosphorylase	0.0917	0.0921
246	11758	Glucagon-like peptide receptor	0.0986	0.0985
247	100098	Serine/threonine-protein kinase WEE1	0.0998	0.0968
248	12622	Telomerase reverse transcriptase	0.0810	0.0779
249	12690	Purine nucleoside phosphorylase	0.0975	0.0967

250	12840	Macrophage colony stimulating factor receptor	0.0873	0.0740
251	12909	Ileal bile acid transporter	0.0979	0.0982
252	20092	Sodium/glucose cotransporter 2	0.0983	0.0958
253	20130	Inhibitor of apoptosis protein 3	0.0963	0.0955
254	100010	Phosphodiesterase 10A	0.0975	0.0951
255	100166	Kinesin-like protein 1	0.0756	0.0480
256	101174	Pituitary adenylate cyclase-activating polypeptide type I receptor	0.0999	0.0999
257	100436	C-C chemokine receptor type 8	0.0977	0.0904
258	100450	Glutaminy-peptide cyclotransferase	0.0992	0.0960
259	100594	Autotaxin	0.0808	0.0840
260	100666	Malonyl-CoA decarboxylase	0.0958	0.0821
261	100862	Metastin receptor	0.0973	0.0969
262	101219	Secreted frizzled-related protein 1	0.0999	0.0998
263	101234	Acyl-CoA desaturase	0.0905	0.0911
264	101395	IgG receptor FcRn large subunit p51	0.0999	0.0999
265	101509	Elongation of very long chain fatty acids protein 6	0.0977	0.0944
266	101502	Egl nine homolog 1	0.0928	0.0906

For both ECFP4 and MACCS, average ROC AUC scores are listed for 1NN similarity searching on all activity class. The scores were calculated for ROC curves capturing the rankings of the first 100,000 database compounds. As a consequence, the largest possible ROC AUC score is 0.1 and the random score is 0.005.

Supplementary figure legends

Figure S1. Similarity search profile. Average recovery rates (selection set size equal to the number of ADCs) of all 266 activity classes are plotted for MACCS (black) and ECFP4 (red). Index on the x-axis gives the consecutively numbered activity classes according to Table S1. Search strategy: **(a)** 1NN, **(b)** 5NN, **(c)** 10NN.

Figure S2. Comparison of search strategies. Average recovery rates (selection set size equal to the number of ADCs) of all 266 activity classes are plotted for the 1NN (black), 5NN (red), and 10NN (green) search strategies. Index gives the consecutively numbered activity classes according to Table S1. Fingerprints: **(a)** ECFP4, **(b)** MACCS.

Figure S3. Enrichment characteristics. Average recovery rates of all 266 activity classes are plotted for selection set sizes of one or two times the number of ADCs per activity class. Index gives the consecutively numbered activity classes according to Table S1. Fingerprints and search strategies: **(a)** ECFP4/1NN, **(b)** ECFP4/5NN, **(c)** ECFP4/10NN, **(d)** MACCS/1NN, **(e)** MACCS/5NN, **(f)** MACCS/10NN.

Figure S4. Enrichment characteristics in profile subsets. Average recovery rates of a representative subset of 20 activity classes (number 120-139 in Table S1) are reported for selection set sizes of one or two times the number of ADCs per activity class. Index reports the consecutively numbered activity classes. Fingerprints and search strategies: **(a)** ECFP4/5NN, **(b)** ECFP4/10NN, **(c)** MACCS/5NN, **(d)** MACCS/10NN.

Figure S5. Average ROC curves. For exemplary activity classes, averaged ROC curves are shown for 1NN similarity searching using MACCS (black) and ECFP4 (red) that capture the ranking of the first 100,000 database compounds. TPR and FPR stand for true-positive rate and false-positive rate, respectively. **(a)** Class no. 253/high search performance, **(b)** No. 169/high search performance, **(c)** No. 5/low search performance, **(d)** No. 96/low search performance, **(e)** No. 4/preferred activity class, **(f)** No. 42/preferred activity class.

Figure S1a

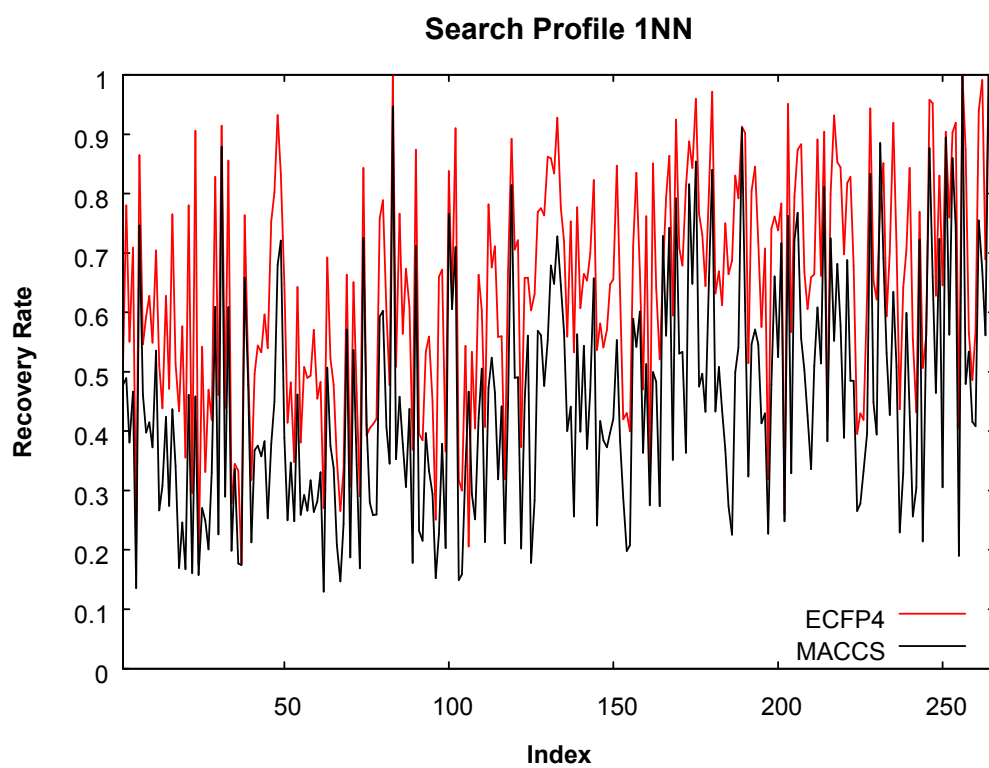


Figure S1b

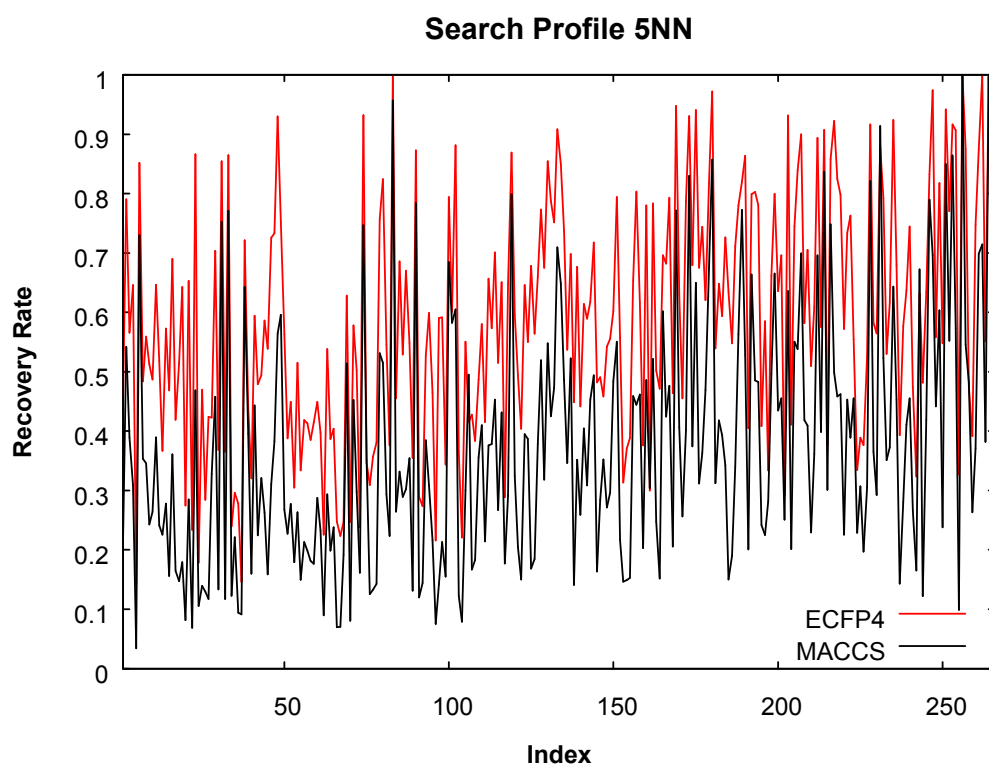


Figure S1c

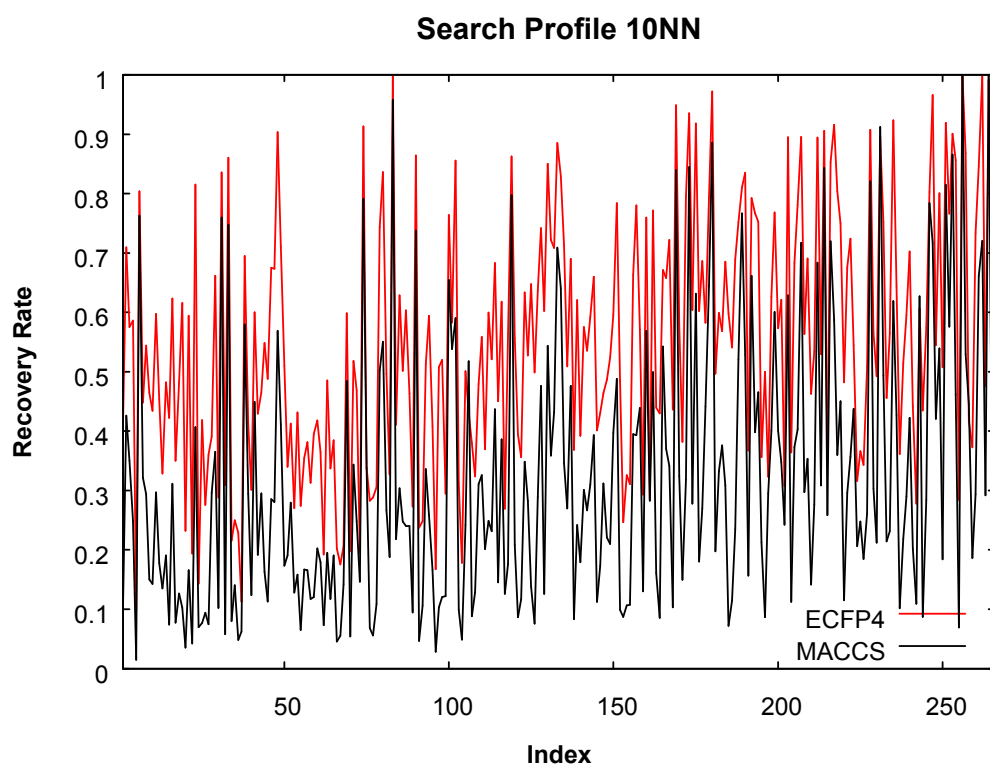


Figure S2a

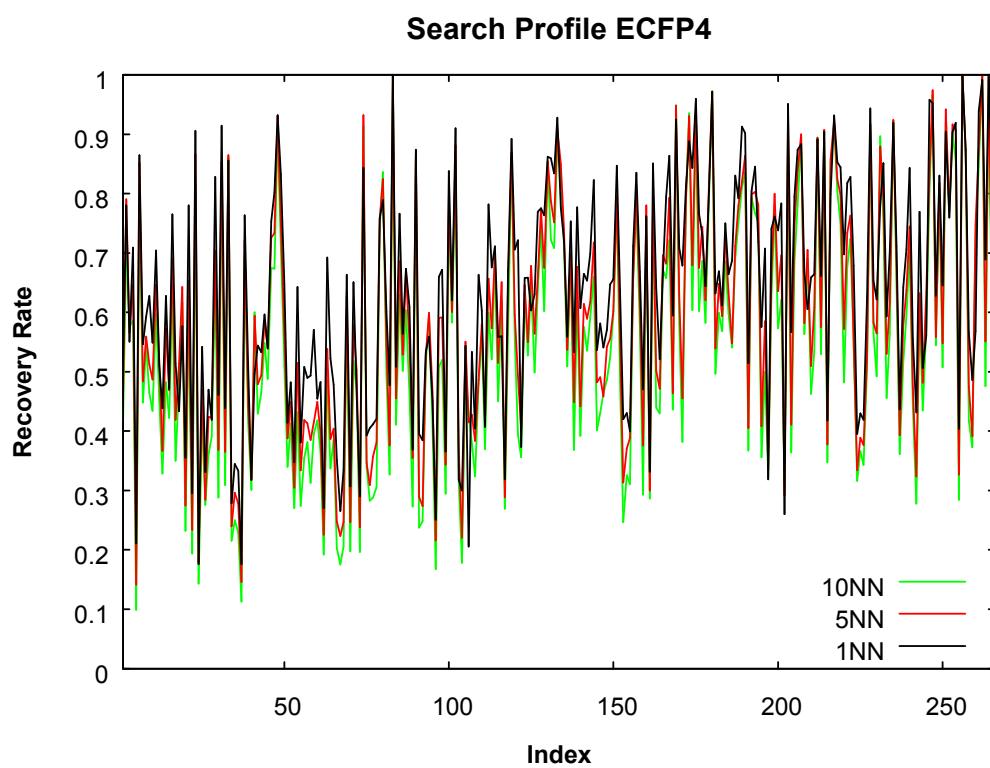


Figure S2b

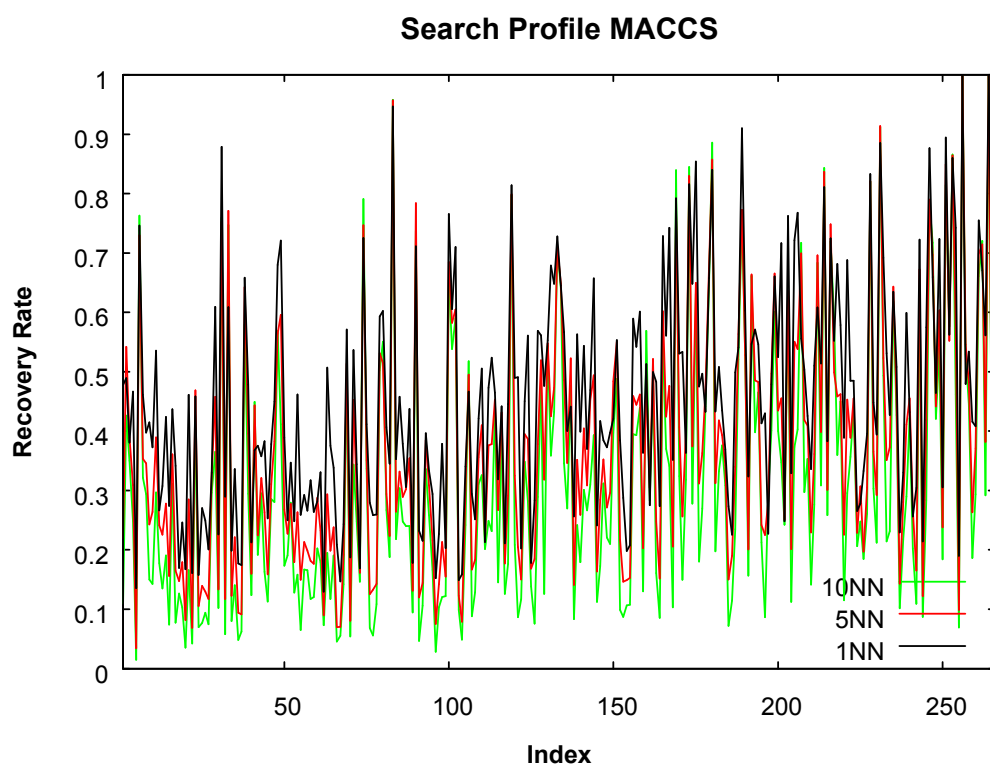


Figure S3a

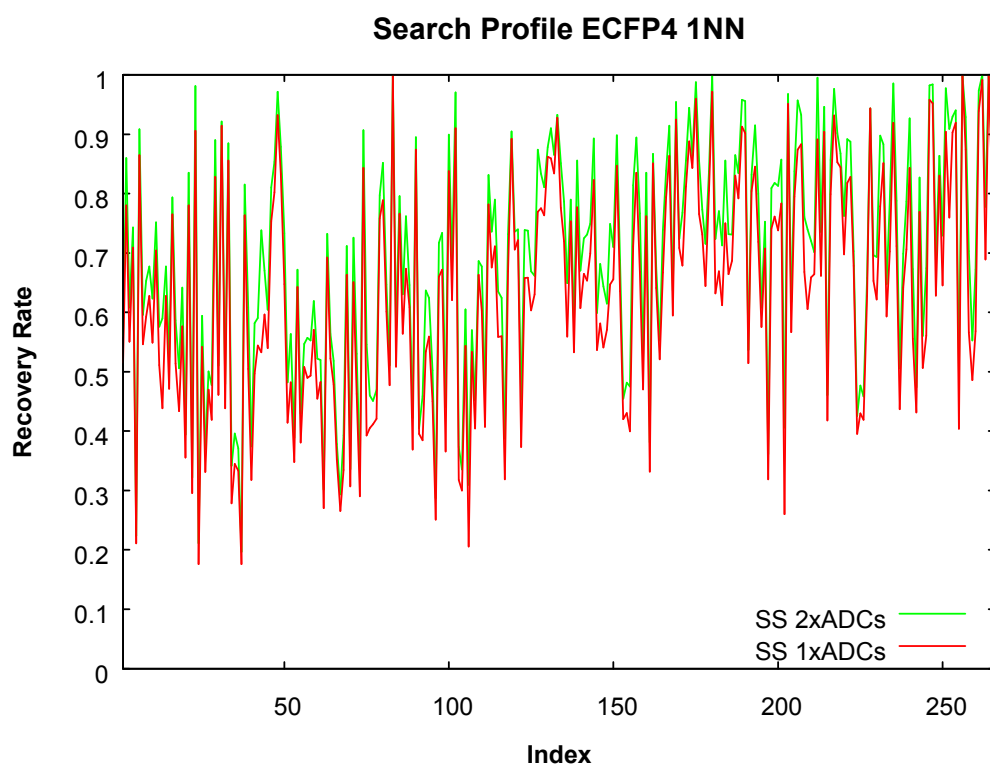


Figure S3b

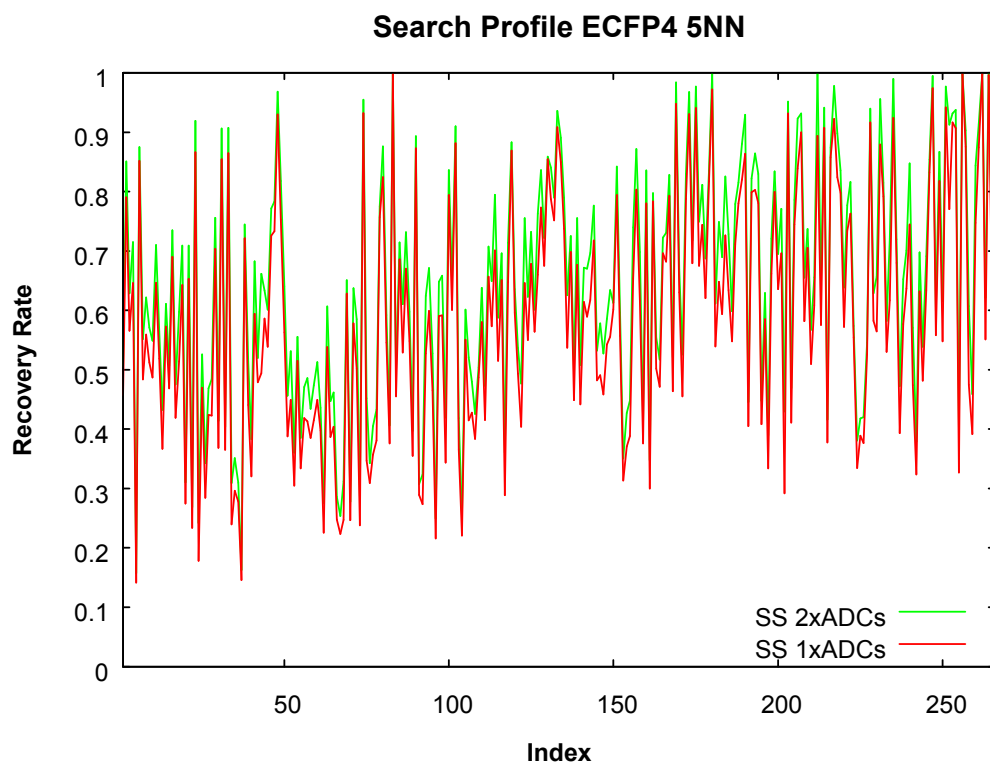


Figure S3c

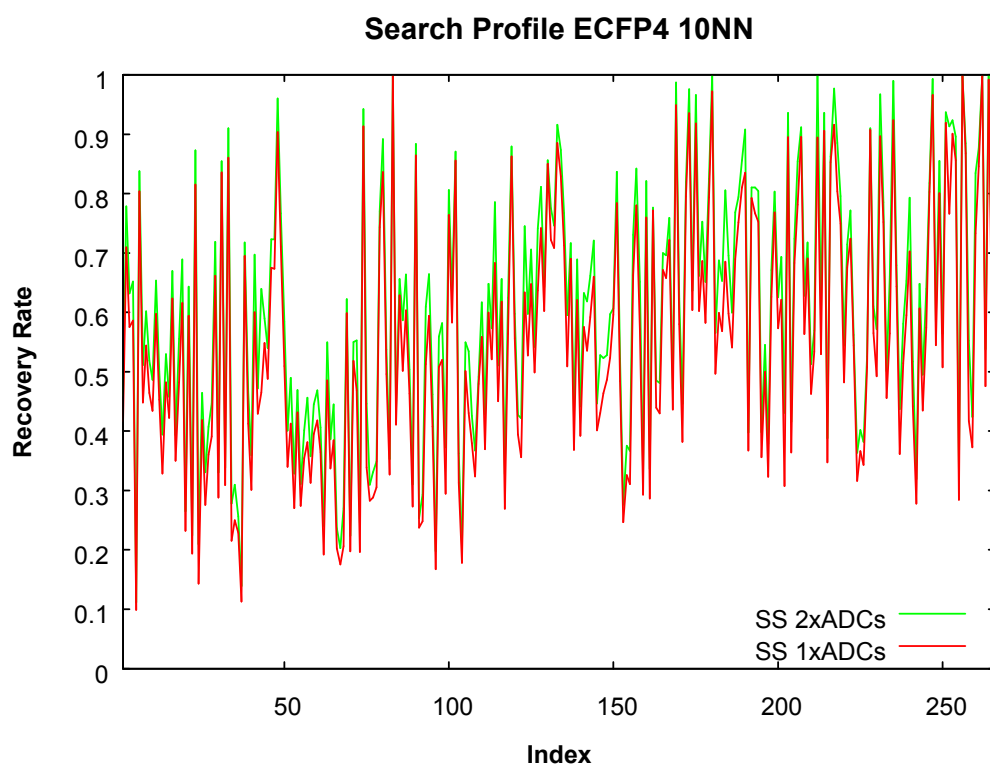


Figure S3d

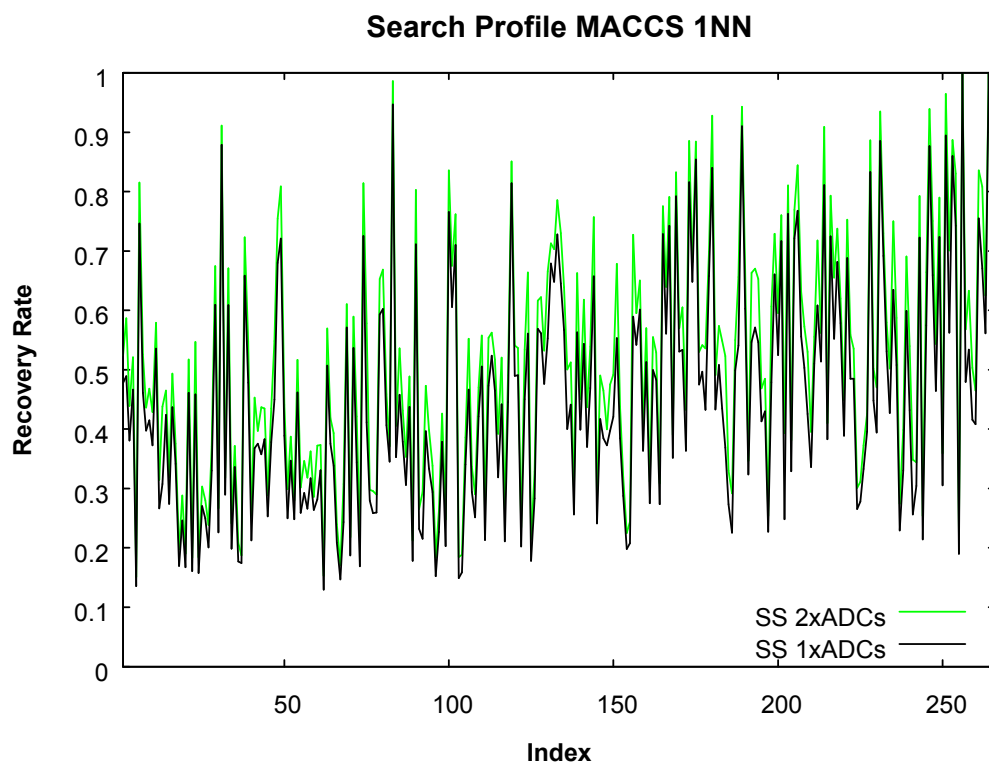


Figure S3e

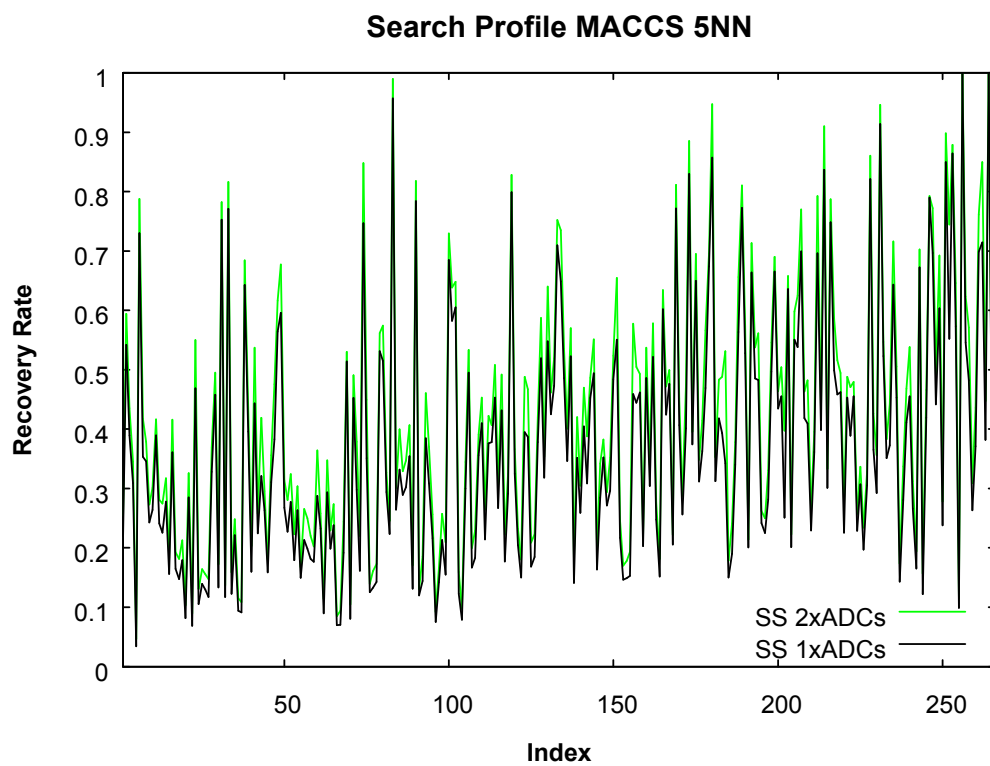


Figure S3f

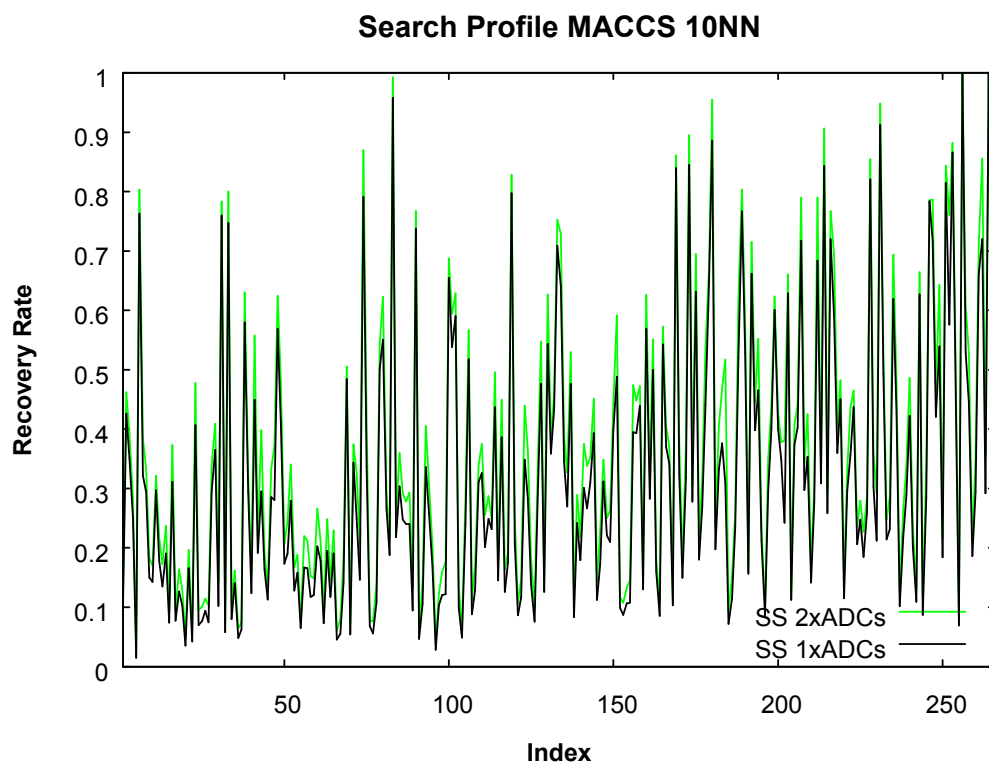


Figure S4a

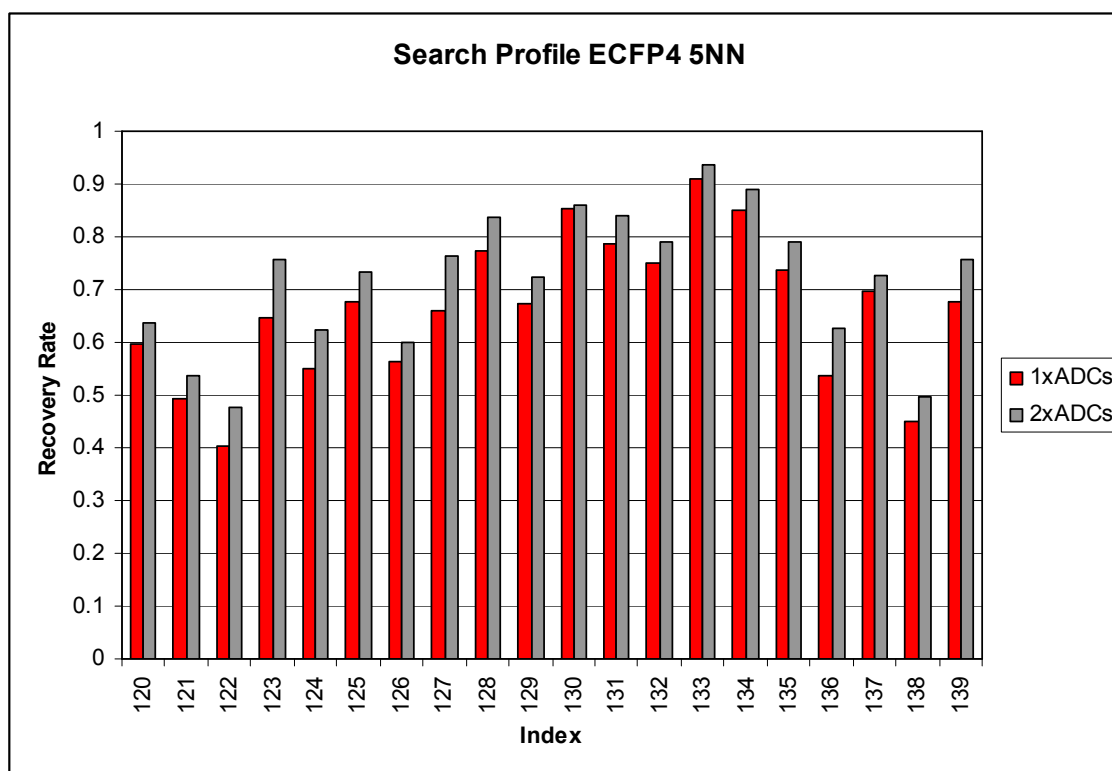


Figure S4b

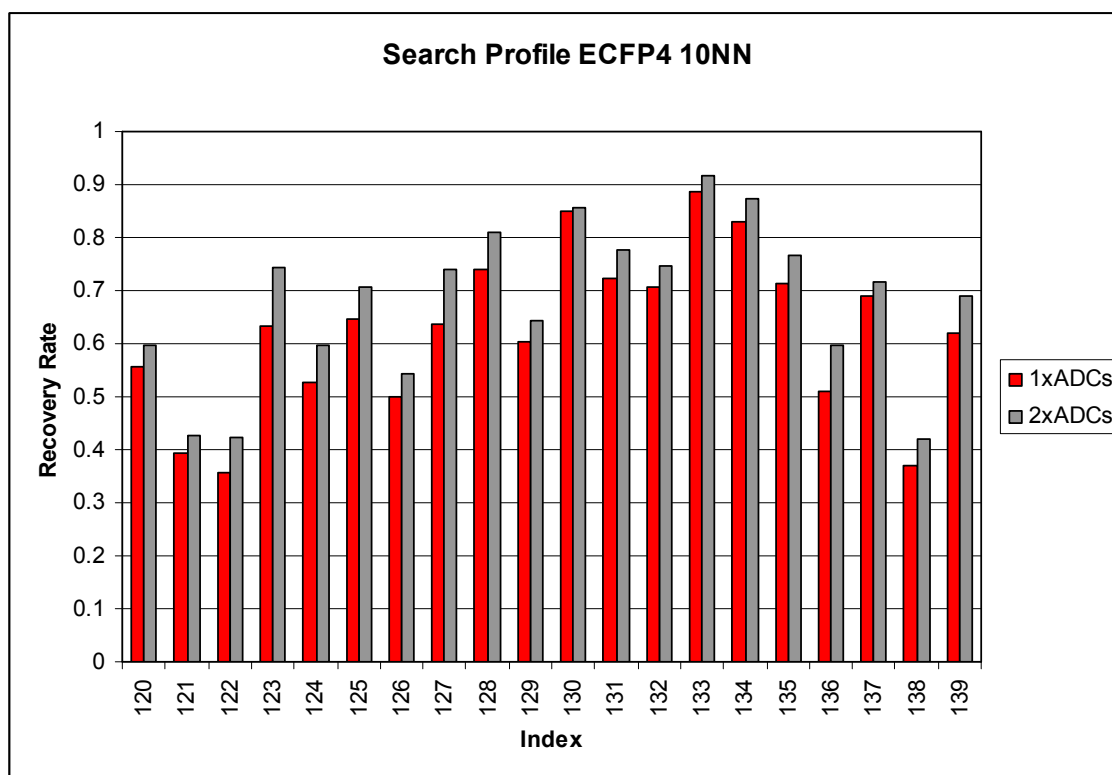


Figure S4c

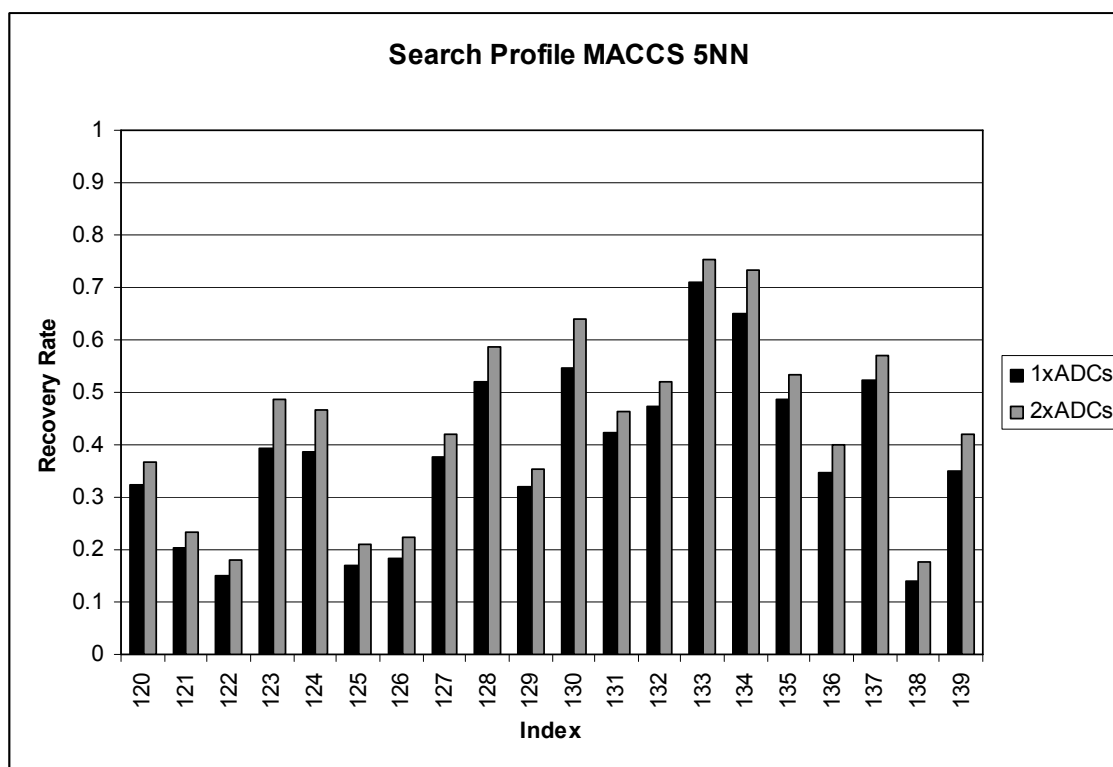


Figure S4d

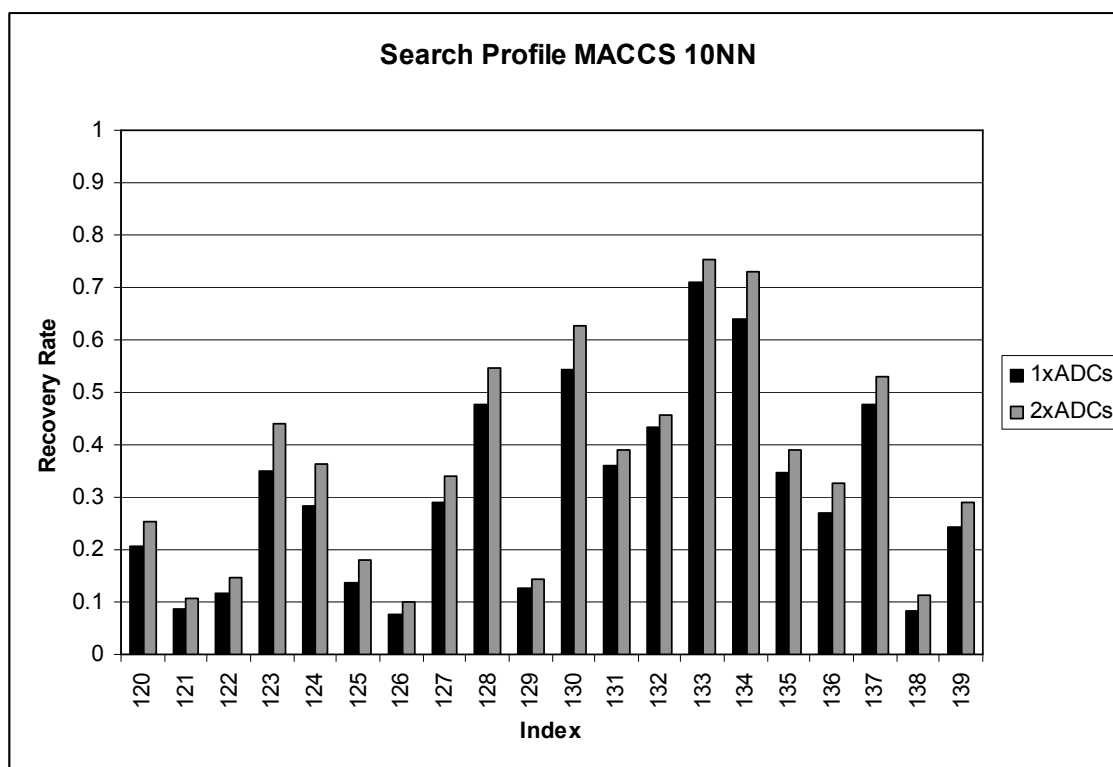


Figure S5a

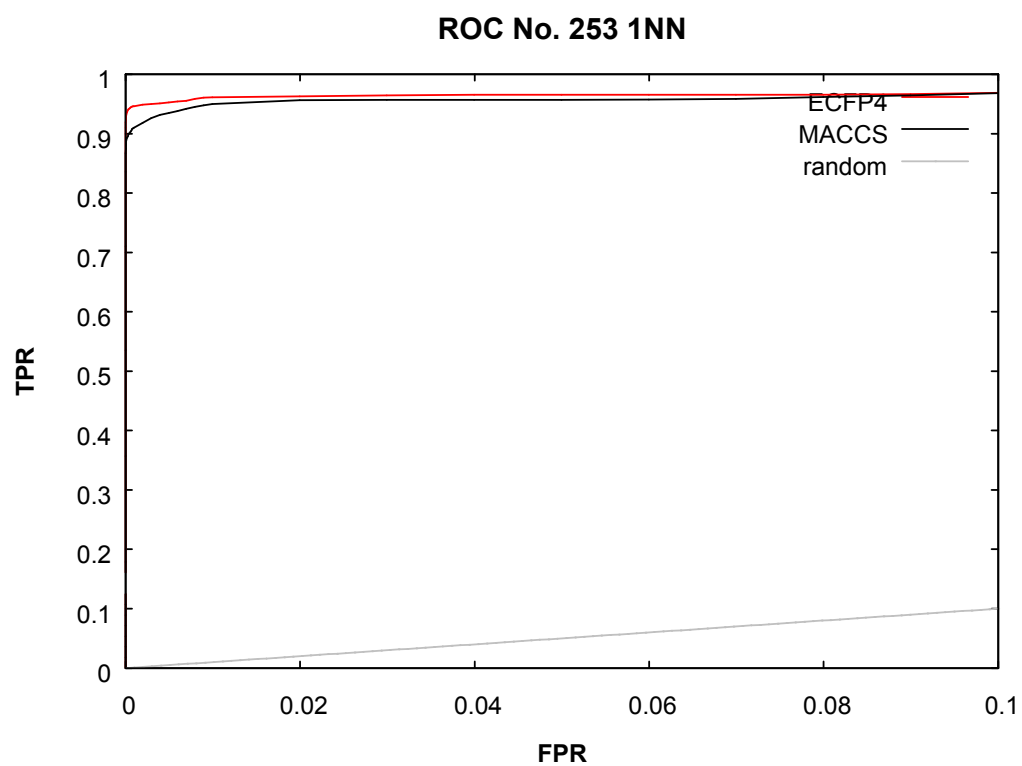


Figure S5b

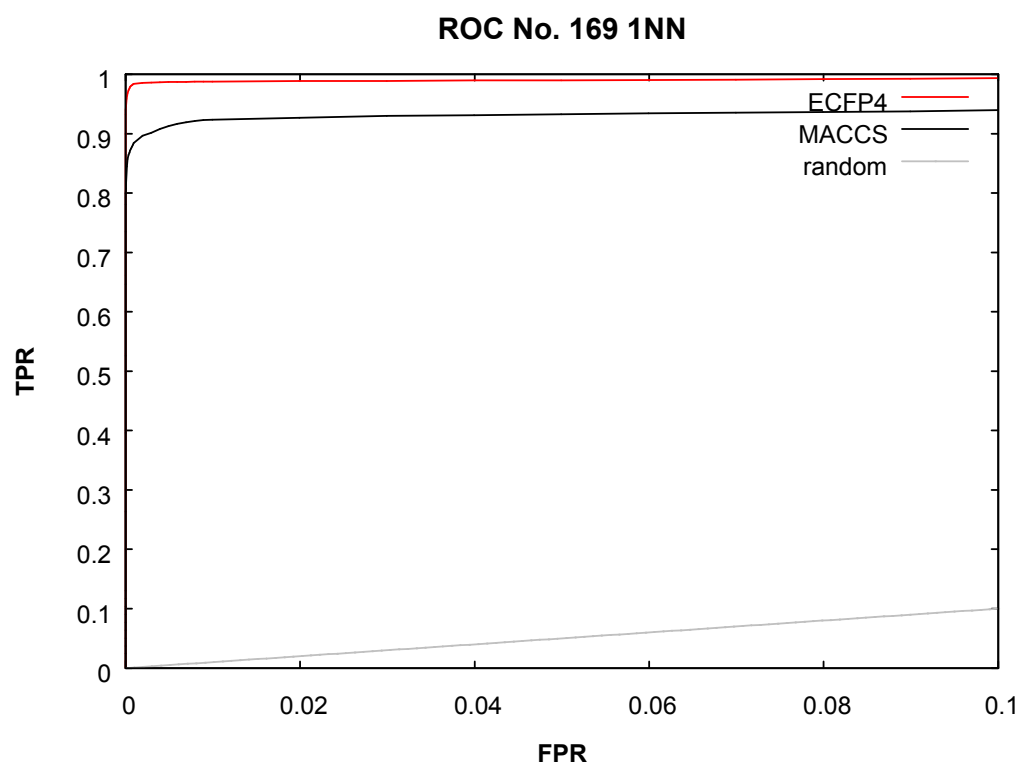


Figure S5c

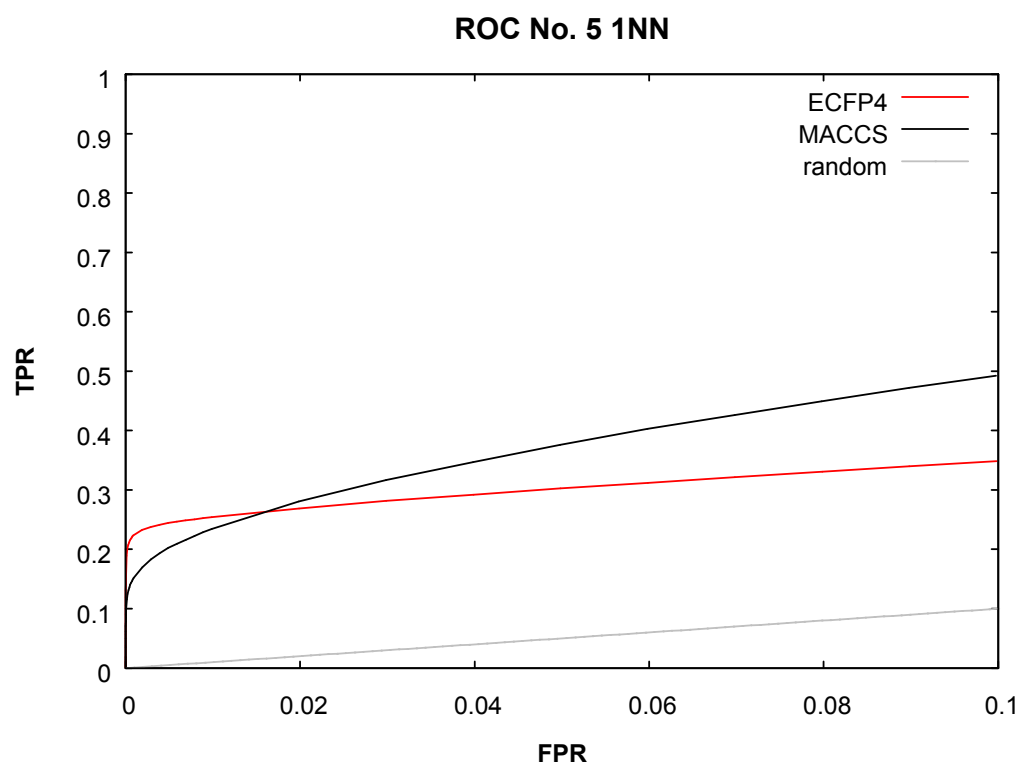


Figure S5d

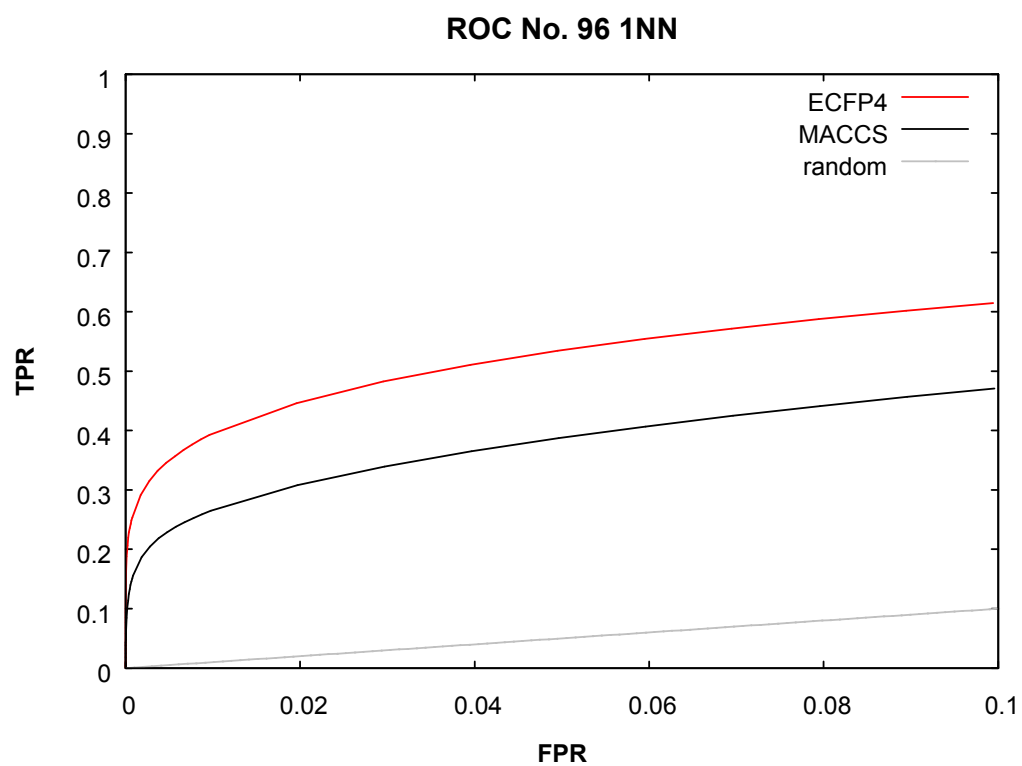


Figure S5e

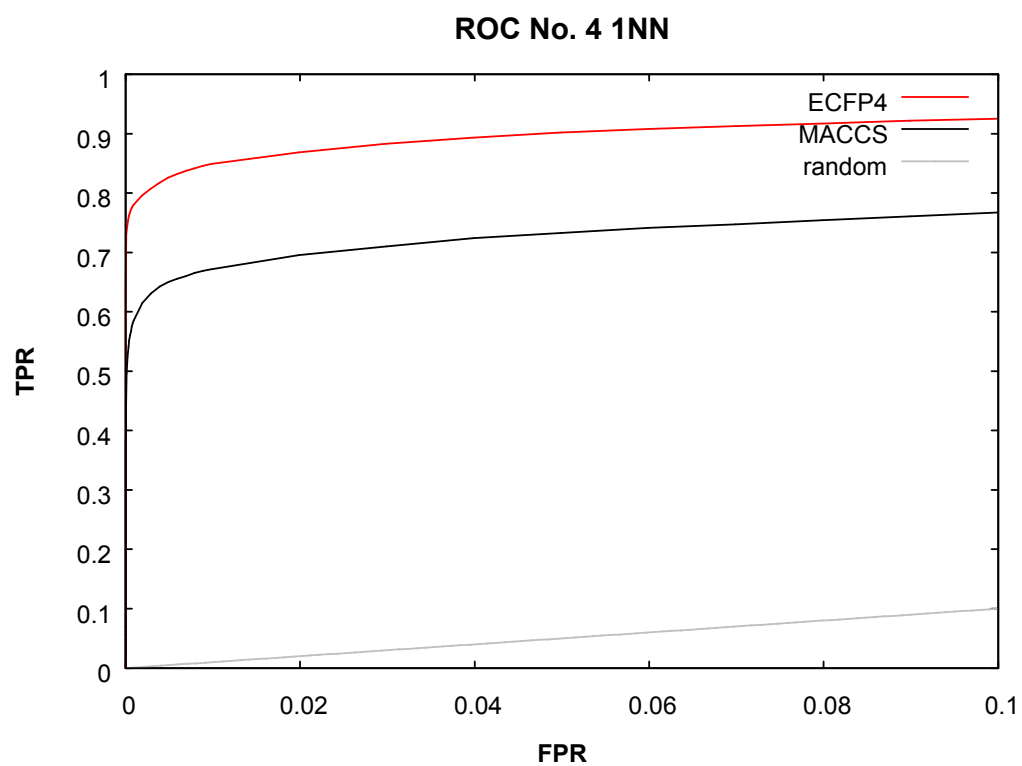


Figure S5f

