Supporting Information
Protein-Ligand Scoring with Convolutional Neural Networks

Matthew Ragoza,[†,‡] Joshua Hochuli,[‡,¶] Elisa Idrobo,[§] Jocelyn Sunseri,[‖] and David Ryan Koes[*,‖]

†*Department of Neuroscience, University of Pittsburgh, Pittsburgh, PA 15260*
‡*Department of Computer Science, University of Pittsburgh, Pittsburgh, PA 15260*
¶*Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260*
§*Department of Computer Science, The College of New Jersey, Ewing, NJ 08628*
‖*Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA 15260*

E-mail: dkoes@pitt.edu

Figures

Tables

Table S1: Atom types used to define protein-ligand structures for CNN scoring.

Type	Ligand	Receptor
AliphaticCarbonXSHydrophobe	Y	Y
AliphaticCarbonXSNonHydrophobe	Y	Y
AromaticCarbonXSHydrophobe	Y	Y
AromaticCarbonXSNonHydrophobe	Y	Y
Bromine	Y	N
Calcium	N	Y
Chlorine	Y	N
Fluorine	Y	N
Iodine	Y	N
Iron	N	Y
Magnesium	N	Y
Nitrogen	Y	Y
NitrogenXSAcceptor	Y	Y
NitrogenXSDonor	Y	Y
NitrogenXSDonorAcceptor	Y	Y
Oxygen	Y	N
OxygenXSAcceptor	Y	Y
OxygenXSDonorAcceptor	Y	Y
Phosphorus	Y	Y
Sulfur	Y	Y
SulfurAcceptor	Y	N
Zinc	N	Y

Table S2: Information about the PDB structures chosen to provide structural information for the selected test set targets. The PDB accession code, crystal ligand, affinity of the crystal ligand for the protein, and assay type from which the test set actives and decoys were derived are shown.

MUV ID	PDB ID	Ligand	K_i/IC_{50} (nM)	Assay type
600	1yow	P0E	N/A	cell
692	1yow	P0E	N/A	cell
859	5cxv	0HK	N/A	cell
852	4xe4	NAG	N/A	biochemical
548	3poo	S69	N/A	biochemical
832	1au8	0H8	N/A	biochemical
689	2y6o	1N1	25	biochemical
846	5exm	5ST	N/A	biochemical
466	3v2y	ML5	18-77	cell

ChEMBL ID	PDB ID	Ligand	K_i/IC_{50} (nM)	Assay type
10752	4kik	KSA	N/A	biochemical
11359	1mkd	ZAR	160	biochemical
28	1hvy	D16	290	biochemical
276	2qyk	NPV	1-88	biochemical
10498	2xu1	424	22	biochemical
11534	1ms6	BLN	0.3	biochemical
10378	1csb	EP0	N/A	biochemical
219	4daj	0HK	N/A	biochemical
11279	3ks9	Z99	6800	biochemical
12968	4s0v	SUV	0.35-12	biochemical
20014	1mq4	ADP	N/A	biochemical
11631	3v2y	ML5	18-77	biochemical
18061	5ek0	5P2	N/A	biochemical
12670	4xuf	P30	1.3-8.8	biochemical

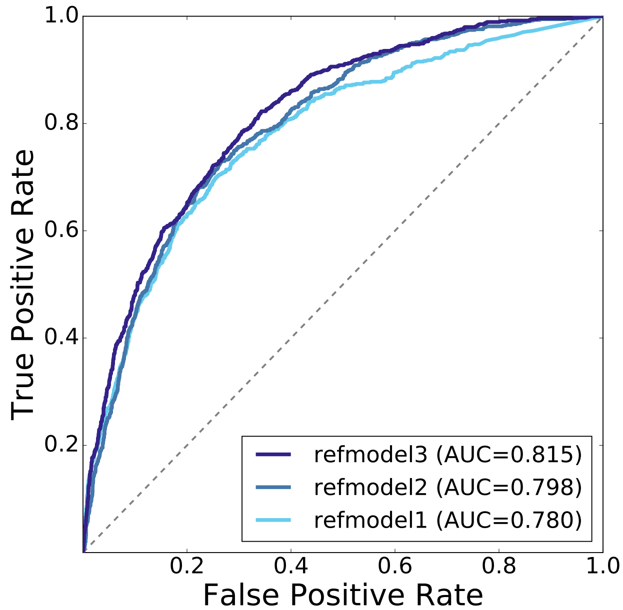

Figure S1: The cross-validation ROC curves for the three reference models used during model optimization.

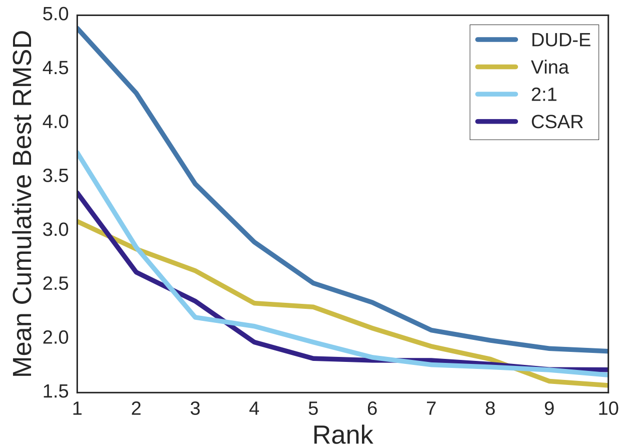

Figure S2: The best RMSD pose identified in the top ranked ligand poses averaged across all PDBbind core subset targets for each scoring method.

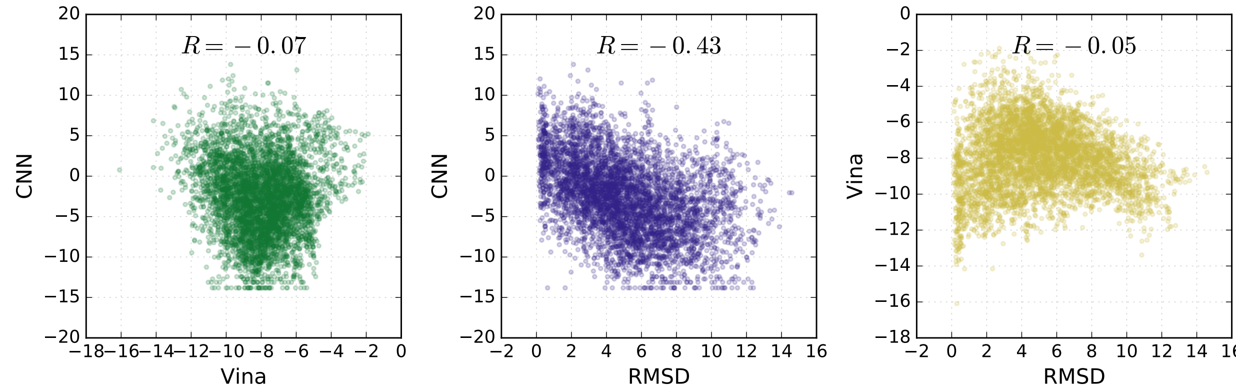

Figure S3: Pearson correlation between Vina, CNN, and RMSD from crystal pose for the CSAR dataset. A logit transformation has been applied to the CNN score, mapping it from probability to linear space, in order to more easily see the relationship. All generated poses are shown, including those with RMSD between 2 and 4 which were omitted from training.

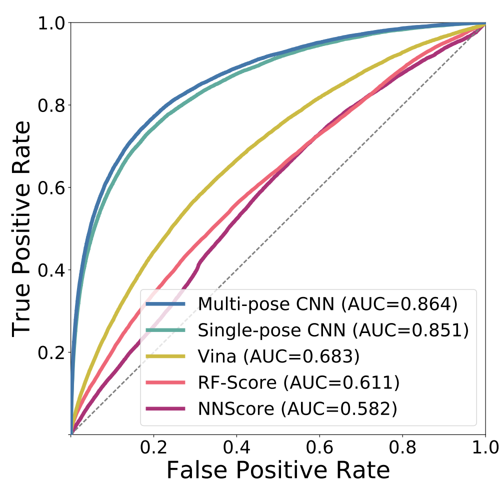

Figure S4: ROC curves for cross-validation virtual screening performance across the full DUD-E benchmark. Single-pose scoring distinguishes between active and inactive compounds using the top ranked pose identified by Vina while multi-pose scoring selects among all docked poses using the maximum CNN score.

Table S3: Cross-validation multipose DUD-E AUC for Vina, RF-Score (RF-S), NNScore (NNS), and CNN models trained using either only DUD-E docked poses or a combination, at a 2:1 ratio, of DUD-E poses and CSAR poses.

Target	DUD E	2 1 D/C	Vina	RF S	NNS	Target	DUD E	2 1 D/C	Vina	RF S	NNS
aa2ar	**0.941**	0 874	0 655	0 494	0 568	igf1r	**0.971**	0 904	0 836	0 572	0 624
abl1	**0.933**	0 892	0 752	0 693	0 626	inha	**0.812**	0 789	0 715	0 658	0 539
ace	**0.799**	0 715	0 564	0 668	0 761	ital	**0.943**	0 774	0 600	0 608	0 481
aces	**0.861**	0 749	0 777	0 681	0 645	jak2	**0.990**	0 941	0 773	0 509	0 629
ada	**0.890**	0 844	0 574	0 606	0 583	kif11	0 793	0 655	**0.833**	0 727	0 543
ada17	**0.939**	0 818	0 683	0 659	0 693	kit	**0.968**	0 822	0 779	0 656	0 563
adrb1	**0.876**	0 871	0 743	0 679	0 584	kith	0 528	**0.851**	0 732	0 547	0 673
adrb2	**0.862**	0 758	0 722	0 630	0 615	kpcb	**0.856**	0 839	0 755	0 679	0 486
akt1	**0.982**	0 867	0 742	0 635	0 491	lck	**0.920**	0 879	0 795	0 488	0 530
akt2	**0.986**	0 845	0 779	0 563	0 649	lkha4	**0.941**	0 887	0 823	0 634	0 350
aldr	0 676	0 644	**0.726**	0 593	0 506	mapk2	**0.891**	0 809	0 885	0 640	0 540
ampc	0 632	**0.736**	0 608	0 349	0 566	mcr	0 799	**0.831**	0 603	0 722	0 721
andr	**0.734**	0 708	0 639	0 635	0 590	met	**0.969**	0 883	0 798	0 744	0 585
aofb	0 623	0 546	**0.780**	0 556	0 568	mk01	**0.928**	0 856	0 852	0 682	0 615
bace1	**0.808**	0 728	0 724	0 786	0 628	mk10	**0.922**	0 788	0 745	0 559	0 570
braf	**0.988**	0 920	0 841	0 632	0 556	mk14	**0.950**	0 848	0 737	0 650	0 508
cah2	0 621	0 755	0 589	**0.813**	0 447	mmp13	**0.966**	0 874	0 645	0 667	0 638
casp3	**0.871**	0 689	0 697	0 538	0 636	mp2k1	**0.825**	0 767	0 550	0 562	0 679
cdk2	**0.842**	0 781	0 716	0 560	0 587	nos1	**0.733**	0 628	0 588	0 527	0 601
comt	0 791	**0.946**	0 631	0 262	0 254	nram	**0.868**	0 692	0 541	0 572	0 600
cp2c9	**0.884**	0 791	0 622	0 576	0 534	pa2ga	0 886	**0.903**	0 611	0 615	0 409
cp3a4	**0.901**	0 775	0 597	0 579	0 515	parp1	0 847	**0.891**	0 846	0 588	0 621
cxcr4	0 707	0 533	0 597	0 691	**0.799**	pde5a	**0.934**	0 902	0 620	0 578	0 660
def	**0.885**	0 845	0 764	0 591	0 562	pgh1	**0.751**	0 740	0 636	0 566	0 451
dhi1	0 600	0 646	**0.771**	0 621	0 683	pgh2	**0.840**	0 832	0 735	0 570	0 456
dpp4	**0.739**	0 666	0 625	0 603	0 674	plk1	**0.939**	0 862	0 650	0 592	0 475
drd3	**0.773**	0 714	0 754	0 671	0 604	pnph	**0.933**	0 743	0 882	0 521	0 514
dyr	**0.866**	0 832	0 769	0 553	0 681	ppara	**0.874**	0 752	0 866	0 623	0 539
egfr	**0.966**	0 908	0 629	0 541	0 547	ppard	**0.871**	0 825	0 761	0 544	0 506
esr1	**0.930**	0 906	0 807	0 640	0 504	pparg	**0.916**	0 802	0 795	0 713	0 528
esr2	**0.924**	0 897	0 790	0 659	0 504	prgr	**0.854**	0 820	0 677	0 680	0 646
fa10	**0.905**	0 861	0 777	0 805	0 555	ptn1	**0.856**	0 847	0 831	0 638	0 564
fa7	0 943	**0.958**	0 910	0 852	0 663	pur2	**0.946**	0 907	0 911	0 734	0 833
fabp4	**0.905**	0 781	0 774	0 673	0 302	pygm	**0.782**	0 716	0 600	0 617	0 667
fak1	**0.986**	0 958	0 801	0 708	0 490	pyrd	**0.920**	0 830	0 768	0 773	0 489
fkb1a	0 684	0 659	**0.770**	0 649	0 573	reni	**0.923**	0 727	0 664	0 699	0 649
fnta	**0.909**	0 805	0 658	0 746	0 570	rock1	**0.925**	0 872	0 717	0 566	0 547
fpps	0 980	0 098	0 286	0 211	**0.999**	rxra	0 678	**0.880**	0 702	0 469	0 530
gcr	**0.902**	0 848	0 612	0 590	0 644	sahh	**0.977**	0 909	0 795	0 547	0 714
glcm	0 610	**0.782**	0 492	0 503	0 597	src	**0.950**	0 903	0 646	0 532	0 552
gria2	0 781	**0.791**	0 746	0 560	0 619	tgfr1	**0.998**	0 985	0 852	0 587	0 486
grik1	0 656	**0.687**	0 593	0 452	0 679	thb	0 832	**0.890**	0 750	0 610	0 392
hdac2	**0.936**	0 813	0 850	0 814	0 572	thrb	**0.924**	0 827	0 766	0 810	0 795
hdac8	**0.950**	0 782	0 824	0 756	0 534	try1	**0.953**	0 861	0 799	0 777	0 719
hivint	**0.870**	0 643	0 707	0 488	0 469	tryb1	**0.902**	0 845	0 707	0 715	0 660
hivpr	**0.891**	0 677	0 721	0 763	0 578	tysy	**0.979**	0 905	0 862	0 675	0 644
hivrt	**0.734**	0 727	0 675	0 485	0 575	urok	**0.962**	0 809	0 771	0 677	0 610
hmdh	0 897	**0.911**	0 787	0 712	0 689	vgfr2	**0.967**	0 859	0 751	0 717	0 496
hs90a	**0.909**	0 836	0 272	0 372	0 445	wee1	0 986	**0.989**	0 830	0 798	0 591
hxk4	**0.875**	0 749	0 569	0 631	0 627	xiap	0 830	**0.849**	0 730	0 731	0 801

Table S4: Cross-validation multipose DUD-E 0.5% ROC enrichment for Vina, RF-Score (RF-S), NNScore (NNS), and CNN models trained using either only DUD-E docked poses or a combination, at a 2:1 ratio, of DUD-E poses and CSAR poses.

Target	DUD E	2 1 D/C	Vina	RF S	NNS
aa2ar	**29.461**	11 203	2 490	2 588	0 415
abl1	**96.133**	41 989	17 680	26 794	2 198
ace	**10.714**	1 429	2 143	7 692	6 186
aces	10 619	4 867	**26.106**	1 099	2 208
ada	**19.355**	12 903	0 000	3 738	0 000
ada17	**73.308**	28 571	10 902	2 802	1 880
adrb1	20 408	**22.041**	3 265	0 717	2 429
adrb2	**25.217**	1 747	7 826	1 455	6 061
akt1	**147.440**	18 430	9 556	1 198	0 000
akt2	**94.017**	22 222	17 094	0 000	1 709
aldr	**15.094**	0 000	6 289	1 163	0 000
ampc	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
andr	2 230	2 974	**16.357**	0 308	1 413
aofb	1 639	1 639	**14.754**	0 000	0 000
bace1	9 187	10 601	4 947	**29.909**	2 827
braf	**139.474**	27 632	26 316	16 078	1 316
cah2	4 499	2 857	0 000	**36.481**	3 318
casp3	**14.141**	4 040	1 010	0 952	3 030
cdk2	**34.672**	19 450	8 879	0 000	2 114
comt	4 878	**48.780**	9 756	0 000	2 198
cp2c9	**46.667**	8 333	5 000	10 145	0 000
cp3a4	**39.521**	13 174	2 395	4 082	0 000
cxcr4	0 000	0 000	0 000	0 000	**5.000**
def	**56.863**	27 451	15 686	9 091	0 000
dhi1	1 212	0 606	2 424	**7.629**	2 424
dpp4	**7.505**	1 876	1 501	1 358	3 752
drd3	4 641	1 684	**8.017**	5 062	1 235
dyr	**41.739**	33 043	4 348	0 905	0 866
egfr	**103.882**	36 969	3 697	0 663	2 947
esr1	21 636	**47.895**	16 359	0 000	1 042
esr2	10 989	**71.233**	12 088	0 000	0 545
fa10	38 433	**39.179**	6 343	23 166	1 117
fa7	21 053	31 579	17 544	**34.146**	0 000
fabp4	**59.574**	0 000	4 255	0 000	0 000
fak1	**156.000**	138 000	2 000	0 000	0 000
fkb1a	0 000	0 000	**9.009**	0 000	0 000
fnta	**51.777**	11 506	1 354	8 006	2 024
fpps	61 176	0 000	0 000	0 000	**193.023**
gcr	9 302	**10.078**	9 302	3 883	0 775
glcm	0 000	**14.815**	0 000	0 000	3 704
gria2	1 266	13 924	**17.722**	3 529	15 190
grik1	2 000	0 000	**4.000**	3 704	2 000
hdac2	**130.055**	2 186	18 579	2 771	1 081
hdac8	**161.905**	3 571	29 762	3 371	1 176
hivint	6 000	**8.000**	4 000	0 000	0 000
hivpr	11 632	1 498	5 629	**43.467**	0 373
hivrt	**8.358**	5 952	2 985	0 000	7 977
hmdh	14 118	**50.588**	5 882	10 000	2 353
hs90a	4 545	**20.455**	0 000	0 000	4 545
hxk4	**63.043**	26 087	0 000	0 000	0 000

Target	DUD E	2 1	Vina	RF S	NNS
igf1r	**132.432**	54 054	10 811	23 077	1 351
inha	0 000	9 302	**18.605**	0 000	0 000
ital	**48.175**	13 139	0 000	1 389	1 460
jak2	**168.224**	95 327	7 477	13 223	1 869
kif11	20 690	6 897	8 621	**28.800**	0 000
kit	**57.831**	19 277	3 614	11 236	2 410
kith	0 000	0 000	**24.561**	3 175	0 000
kpcb	7 463	5 970	**8.955**	0 000	0 000
lck	**83.055**	49 642	7 637	5 996	0 476
lkha4	**54.118**	40 000	8 235	0 000	0 000
mapk2	**57.426**	7 921	17 822	0 000	0 000
mcr	4 255	4 255	10 638	1 942	**14.583**
met	**121.687**	66 265	4 819	35 724	1 205
mk01	**43.038**	10 127	2 532	7 143	0 000
mk10	**83.495**	40 385	9 709	5 217	0 000
mk14	**74.177**	32 180	4 853	13 938	1 038
mmp13	**89.510**	22 727	3 846	7 724	2 763
mp2k1	**55.000**	1 667	0 000	0 000	6 667
nos1	**16.000**	2 000	0 000	0 000	2 000
nram	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
pa2ga	**22.449**	10 204	2 041	0 000	0 000
parp1	8 268	3 150	**20.079**	2 812	0 394
pde5a	63 317	**64.322**	6 030	1 357	1 471
pgh1	11 458	**15.544**	2 083	1 786	0 000
pgh2	21 911	**28.637**	20 513	2 271	0 920
plk1	**64.151**	54 717	3 774	5 996	7 547
pnph	**27.451**	0 000	19 608	0 000	3 306
ppara	**12.869**	6 971	6 971	0 990	1 072
ppard	9 167	**14.167**	0 000	0 000	0 000
pparg	**26.860**	11 570	7 025	4 671	1 653
prgr	6 826	**11.604**	9 556	0 000	3 311
ptn1	**55.385**	30 769	47 692	15 603	7 246
pur2	0 000	4 000	8 000	0 000	**24.000**
pygm	0 000	0 000	0 000	0 000	**12.987**
pyrd	**59.459**	50 450	10 811	9 839	0 000
reni	3 883	3 883	**5.825**	1 754	1 869
rock1	**68.000**	34 000	2 000	0 000	2 000
rxra	0 000	1 527	**47.328**	0 000	0 000
sahh	0 000	22 222	**25.397**	0 000	0 000
src	**98.084**	50 192	5 364	1 112	0 758
tgfr1	**181.955**	147 368	9 023	1 235	0 000
thb	5 825	**29.126**	23 301	0 000	0 000
thrb	**52.928**	8 677	3 471	9 815	6 508
try1	**75.724**	10 690	3 118	16 129	4 009
tryb1	6 757	4 054	**13.514**	7 701	1 351
tysy	**75.229**	42 202	27 523	1 681	1 835
urok	**77.778**	7 407	9 877	3 352	0 000
vgfr2	**88.020**	41 565	11 247	8 153	0 000
wee1	106 931	**124.752**	15 842	0 000	0 000
xiap	0 000	0 000	**8.081**	0 000	6 061

Table S5: Cross-validation multipose DUD-E 1.0% ROC enrichment for Vina, RF-Score (RF-S), NNScore (NNS), and CNN models trained using either only DUD-E docked poses or a combination, at a 2:1 ratio, of DUD-E poses and CSAR poses.

Target	DUD E	2 1	Vina	RF S	NNS	Target	DUD E	2 1	Vina	RF S	NNS
aa2ar	**24.066**	9 544	3 112	2 033	0 622	igf1r	**79.730**	38 514	11 486	13 292	2 703
abl1	**68.508**	28 177	9 945	13 397	2 747	inha	2 326	**13.953**	9 302	0 000	0 000
ace	**7.500**	1 429	2 143	5 983	5 155	ital	**40.876**	10 219	0 000	1 156	0 730
aces	12 389	3 540	**20.133**	5 719	1 545	jak2	**85.981**	58 879	10 280	8 264	2 804
ada	**29.032**	19 355	0 000	3 738	0 000	kif11	11 207	3 448	11 207	**17.600**	0 000
ada17	**50.188**	18 797	8 083	2 700	3 947	kit	**48.795**	15 663	3 012	6 180	2 410
adrb1	**18.776**	14 286	5 306	0 358	2 024	kith	0 000	0 000	**22.807**	1 587	0 000
adrb2	**23.913**	4 367	6 522	0 727	3 463	kpcb	5 224	8 209	**10.448**	0 000	2 222
akt1	**84.642**	13 993	7 509	1 796	0 000	lck	**53.699**	32 458	7 160	3 640	0 714
akt2	**66.667**	13 675	15 385	0 813	1 709	lkha4	**36.471**	26 471	7 059	0 000	0 000
aldr	**11.321**	1 887	3 145	1 163	0 000	mapk2	**35.644**	6 931	13 861	1 818	1 980
ampc	**2.083**	0 000	0 000	0 000	0 000	mcr	2 128	5 319	8 511	5 133	**12.500**
andr	2 602	2 974	**15.613**	1 079	2 473	met	**74.096**	36 747	6 627	22 662	1 205
aofb	1 639	0 820	**12.295**	0 676	0 000	mk01	**45.570**	15 190	3 797	4 762	1 266
bace1	7 067	9 187	5 300	**21.271**	2 120	mk10	**55.340**	28 846	6 796	5 217	0 962
braf	**78.289**	24 342	17 763	14 430	0 658	mk14	**47.834**	22 491	4 679	8 248	0 519
cah2	4 294	4 082	0 000	**22.265**	2 133	mmp13	**57.867**	17 308	3 671	6 808	2 073
casp3	**13.636**	2 525	1 515	1 905	4 040	mp2k1	**43.333**	1 667	0 833	0 769	5 833
cdk2	**25.793**	15 222	6 131	0 759	1 480	nos1	**10.000**	2 000	1 000	0 000	2 000
comt	2 439	**29.268**	4 878	0 000	2 198	nram	**1.020**	0 000	0 000	0 000	0 000
cp2c9	**34.167**	5 833	2 500	5 797	0 000	pa2ga	**21.429**	15 306	1 020	0 000	0 000
cp3a4	**28.743**	8 982	1 796	3 429	0 000	parp1	8 661	4 331	**13.583**	2 758	1 181
cxcr4	**5.000**	0 000	0 000	0 000	**5.000**	pde5a	**43.467**	42 462	4 523	0 905	1 225
def	**32.353**	23 529	11 765	4 545	1 961	pgh1	8 333	**11.399**	3 125	1 475	1 005
dhi1	2 424	1 515	3 030	**6.118**	1 818	pgh2	17 949	**19.400**	12 821	1 691	1 379
dpp4	**6.379**	2 251	1 313	1 097	3 377	plk1	**47.170**	36 792	3 774	5 988	5 660
drd3	4 641	2 105	**5.907**	3 982	1 235	pnph	**31.373**	0 980	13 725	0 000	2 479
dyr	**28.696**	22 174	3 043	1 527	1 299	ppara	**11.796**	5 362	8 847	0 743	2 145
egfr	**62.477**	26 063	2 403	2 641	2 947	ppard	7 500	**11.667**	1 667	1 158	0 417
esr1	22 427	**36.842**	11 609	0 000	1 302	pparg	**21.901**	7 851	6 198	3 633	1 240
esr2	11 538	**43.288**	11 264	0 000	0 272	prgr	6 826	8 191	**8.532**	2 439	3 974
fa10	**34.328**	26 866	5 784	15 725	0 931	ptn1	**31.538**	23 077	30 000	11 115	4 348
fa7	20 175	**39.474**	14 912	23 266	0 000	pur2	4 000	2 000	12 000	0 000	**14.000**
fabp4	**29.787**	0 000	4 255	0 000	0 000	pygm	1 299	0 000	0 000	0 000	**10.390**
fak1	**87.000**	73 000	5 000	1 835	0 000	pyrd	**42.342**	31 532	8 108	11 864	0 000
fkb1a	0 000	1 802	**7.207**	1 653	0 000	reni	**8.738**	1 942	3 883	4 386	0 935
fnta	**38.579**	8 122	1 861	8 551	1 518	rock1	**43.000**	23 000	4 000	0 901	1 000
fpps	45 882	0 000	0 000	0 000	**98.837**	rxra	0 763	3 053	**24.427**	0 704	0 000
gcr	**12.791**	10 078	6 977	2 215	2 326	sahh	**31.746**	25 397	22 222	0 000	0 000
glcm	1 852	**18.519**	0 000	0 000	1 852	src	**61.494**	32 950	3 448	1 676	1 705
gria2	2 532	**13.291**	12 658	4 118	10 127	tgfr1	**95.489**	81 955	6 767	0 617	0 000
grik1	4 000	2 000	**5.000**	1 852	4 000	thb	8 738	**33.010**	11 650	1 575	0 000
hdac2	**69.399**	5 464	16 393	2 552	2 162	thrb	**39.262**	10 846	4 555	7 977	6 508
hdac8	**83.929**	4 167	26 190	3 371	0 588	try1	**56.793**	12 027	2 895	10 148	3 563
hivint	4 000	**5.000**	2 000	0 000	0 000	tryb1	**8.108**	4 730	**8.108**	7 528	0 676
hivpr	12 195	1 124	6 004	**25.149**	0 187	tysy	**65.138**	32 110	18 349	4 855	2 752
hivrt	**9.851**	5 655	2 388	0 252	5 413	urok	**53.704**	9 877	8 642	2 235	0 000
hmdh	15 882	**37.059**	4 706	5 556	1 765	vgfr2	**60.636**	26 161	8 068	6 334	0 000
hs90a	6 818	**11.364**	0 000	0 000	2 273	wee1	**77.228**	76 238	7 921	1 550	0 000
hxk4	**32.609**	21 739	2 174	0 000	0 000	xiap	2 020	5 051	**11.111**	0 000	4 040

Table S6: Cross-validation multipose DUD-E 2.0% ROC enrichment for Vina, RF-Score (RF-S), NNScore (NNS), and CNN models trained using either only DUD-E docked poses or a combination, at a 2:1 ratio, of DUD-E poses and CSAR poses.

Target	DUD E	2 1	Vina	RF S	NNS	Target	DUD E	2 1	Vina	RF S	NNS
aa2ar	**20.643**	9 336	3 527	1 386	0 415	igf1r	**43.243**	25 000	9 797	7 692	2 027
abl1	**37.845**	19 337	6 906	7 863	1 648	inha	5 814	**12.791**	5 814	0 000	0 000
ace	**5.000**	0 714	1 607	4 647	4 467	ital	**28.467**	8 029	0 365	2 778	0 365
aces	11 283	3 540	**14.270**	6 339	1 104	jak2	**45.327**	30 841	7 009	4 545	3 271
ada	**24.194**	16 667	1 075	5 407	0 000	kif11	6 034	3 448	9 914	**12.821**	0 431
ada17	**33.459**	13 346	6 485	3 976	4 229	kit	**31.627**	10 843	4 518	3 972	1 506
adrb1	**16.122**	12 245	5 714	1 804	1 417	kith	1 754	2 632	**14.912**	1 587	0 877
adrb2	**15.870**	5 240	4 565	0 364	2 165	kpcb	5 597	7 463	**10.075**	0 697	2 222
akt1	**45.222**	11 775	5 631	3 443	0 512	lck	**33.174**	21 241	6 205	2 355	0 952
akt2	**39.316**	8 974	11 538	0 813	2 564	lkha4	**22.059**	18 824	4 412	1 562	0 294
aldr	**6.604**	2 516	3 459	1 453	0 314	mapk2	**22.772**	7 426	13 861	3 182	1 485
ampc	**4.167**	1 042	1 042	0 000	1 042	mcr	2 660	4 787	6 383	5 825	**9.375**
andr	2 788	3 903	**11.338**	1 249	1 413	met	**40.663**	20 181	6 325	16 201	1 205
aofb	2 459	0 410	**8.197**	0 338	0 820	mk01	**26.582**	15 190	5 696	4 167	3 165
bace1	5 477	7 067	4 770	**13.661**	1 413	mk10	**31.068**	17 308	3 883	3 043	1 442
braf	**42.434**	19 408	11 184	10 602	1 645	mk14	**30.763**	15 571	4 159	5 382	0 519
cah2	2 863	5 204	0 307	**12.479**	1 540	mmp13	**35.577**	12 413	3 147	5 499	1 727
casp3	**13.384**	3 030	2 778	1 905	3 283	mp2k1	**23.750**	2 500	0 417	0 769	2 917
cdk2	**16.808**	11 099	5 603	0 948	1 374	nos1	**8.000**	2 500	2 500	0 909	2 000
comt	2 439	**21.951**	2 439	0 000	1 648	nram	**1.531**	0 000	0 000	0 030	**1.531**
cp2c9	**22.083**	6 250	2 500	4 348	0 417	pa2ga	16 837	**19.388**	0 510	2 427	1 531
cp3a4	**20.958**	8 383	1 796	3 294	0 599	parp1	7 283	7 579	**8.465**	2 636	1 083
cxcr4	5 000	0 000	0 000	0 000	**8.750**	pde5a	**26.005**	25 879	2 261	1 131	1 103
def	**20.098**	17 647	7 843	2 384	1 961	pgh1	7 552	**8.549**	2 865	1 705	0 503
dhi1	2 576	1 061	**5.758**	5 352	1 515	pgh2	**13.170**	12 587	8 392	1 579	0 805
dpp4	**5.159**	2 064	1 126	1 019	3 189	plk1	**31.132**	22 170	4 245	5 556	4 245
drd3	4 430	2 000	**5.063**	4 867	1 440	pnph	**22.549**	5 882	8 824	0 000	2 066
dyr	**17.826**	14 130	3 913	1 145	1 948	ppara	**10.590**	4 960	8 579	0 866	1 475
egfr	**37.431**	18 669	2 495	3 020	3 039	ppard	8 125	**9.583**	2 083	2 098	0 625
esr1	19 129	**23.816**	8 179	0 000	0 911	pparg	**19.215**	7 231	5 269	3 573	1 240
esr2	12 225	**26.575**	9 890	0 000	0 817	prgr	6 314	6 997	**8.532**	3 040	4 305
fa10	**24.440**	17 817	5 410	10 846	0 838	ptn1	**18.846**	16 923	**18.846**	10 708	2 899
fa7	17 105	**28.070**	14 035	15 041	1 230	pur2	9 000	8 000	12 000	0 000	**13.000**
fabp4	**17.021**	1 064	6 383	2 922	0 000	pygm	4 545	0 000	0 649	0 000	**9.740**
fak1	**46.000**	43 000	7 000	2 294	0 000	pyrd	**26.577**	17 117	4 955	8 653	0 000
fkb1a	0 901	0 901	**5.856**	1 653	0 450	reni	**14.078**	3 398	4 854	2 632	0 935
fnta	**24.112**	7 530	2 961	6 239	1 433	rock1	**24.000**	16 000	4 000	0 901	0 500
fpps	34 118	1 176	0 000	0 000	**49.419**	rxra	0 382	3 053	**13.359**	1 761	0 382
gcr	10 016	**11.240**	3 876	1 582	2 907	sahh	**37.302**	26 190	16 667	0 000	4 762
glcm	1 852	**14.815**	0 000	0 000	2 778	src	**36.877**	22 414	3 352	1 897	2 083
gria2	5 696	7 595	**8.544**	4 118	5 696	tgfr1	**48.496**	46 617	6 015	0 617	0 000
grik1	**3.500**	**3.500**	**3.500**	0 926	2 000	thb	11 165	**24.757**	9 709	1 181	0 971
hdac2	**36.885**	6 011	11 202	4 301	2 703	thrb	**26.681**	9 653	5 098	5 875	5 531
hdac8	**42.560**	4 464	18 155	3 371	0 882	try1	**33.296**	9 354	2 673	7 428	3 898
hivint	5 000	**5.500**	1 000	0 000	0 000	tryb1	**8.446**	6 419	5 743	6 938	1 351
hivpr	11 445	1 404	5 910	**15.027**	0 560	tysy	**40.367**	21 101	14 220	3 782	2 294
hivrt	**8.358**	6 845	2 388	0 658	4 274	urok	**30.864**	8 025	7 716	1 676	0 926
hmdh	15 588	**25.294**	4 118	3 611	1 765	vgfr2	**36.430**	16 993	6 479	5 765	0 244
hs90a	**9.659**	8 523	0 000	0 000	1 136	wee1	**44.554**	42 574	5 941	5 259	0 495
hxk4	**23.370**	15 217	2 717	0 521	0 000	xiap	3 030	6 566	**7.071**	0 000	4 040

Table S7: Cross-validation multipose DUD-E 5.0% ROC enrichment for Vina, RF-Score (RF-S), NNScore (NNS), and CNN models trained using either only DUD-E docked poses or a combination, at a 2:1 ratio, of DUD-E poses and CSAR poses.

Target	DUD E	2 1	Vina	RF S	NNS
aa2ar	**13.527**	8 133	2 988	0 961	0 373
abl1	**17.127**	12 265	5 193	5 055	1 099
ace	4 786	1 143	1 357	3 333	**5.704**
aces	**8.319**	3 451	7 611	4 173	1 015
ada	**11.613**	9 247	1 075	3 172	0 215
ada17	**16.316**	7 594	4 586	3 659	3 308
adrb1	**10.531**	8 653	4 245	1 912	0 729
adrb2	**9.826**	4 454	4 435	0 872	1 385
akt1	**18.908**	8 737	4 027	2 391	0 956
akt2	**18.632**	7 350	7 350	1 463	2 222
aldr	**4.780**	2 013	3 396	1 047	0 503
ampc	**4.167**	1 667	1 667	1 154	0 833
andr	3 792	3 346	**6.097**	1 576	1 696
aofb	2 623	0 492	**5.410**	2 027	0 984
bace1	4 240	4 452	3 322	**7.799**	0 989
braf	**18.947**	10 789	7 237	5 562	1 184
cah2	2 822	5 551	0 695	**6.085**	1 232
casp3	**8.990**	2 424	2 727	1 727	3 131
cdk2	**9.725**	6 681	4 482	1 022	1 184
comt	3 415	**17.073**	1 951	0 000	0 879
cp2c9	**11.833**	6 833	2 000	2 754	0 667
cp3a4	**12.455**	5 868	2 036	2 143	0 479
cxcr4	3 000	2 000	1 000	1 021	**6.000**
def	**9.412**	7 451	6 471	1 818	1 765
dhi1	2 242	1 152	**4.364**	3 106	2 182
dpp4	**4.353**	1 951	1 126	1 259	2 477
drd3	4 008	1 979	**4.093**	3 603	1 276
dyr	**9.913**	9 652	4 000	0 840	2 511
egfr	**17.671**	11 054	2 181	2 653	2 578
esr1	11 715	**12.316**	5 541	0 047	0 625
esr2	11 429	**12.603**	7 527	0 321	0 654
fa10	**13.022**	9 403	4 254	7 199	0 968
fa7	11 228	**14.386**	12 456	7 313	1 148
fabp4	**12.766**	1 702	7 234	2 182	0 426
fak1	**18.600**	17 600	4 800	2 098	0 400
fkb1a	2 342	1 982	**3.784**	1 488	0 541
fnta	**12.623**	5 854	2 910	3 515	1 282
fpps	18 118	0 471	0 000	0 000	**19.767**
gcr	**9.535**	7 674	3 876	1 409	2 481
glcm	2 593	**8.148**	0 000	0 000	1 481
gria2	**6.962**	6 203	5 823	2 588	3 038
grik1	**3.800**	3 000	3 200	0 926	3 000
hdac2	**16.284**	6 339	8 743	4 654	1 622
hdac8	**17.262**	5 000	9 405	3 068	0 824
hivint	**8.800**	4 000	2 800	0 189	0 200
hivpr	**8.480**	1 798	4 728	7 438	0 896
hivrt	**5.672**	5 357	2 269	0 856	2 336
hmdh	10 353	**11.882**	3 647	2 222	2 118
hs90a	**9.318**	5 455	0 455	0 194	0 455
hxk4	**11.522**	6 739	1 304	0 208	0 435

Target	DUD E	2 1	Vina	RF S	NNS
igf1r	**18.243**	13 243	7 703	4 103	2 297
inha	5 116	**8.372**	4 186	3 971	0 000
ital	**15.036**	5 839	0 730	1 806	0 292
jak2	**19.065**	14 766	5 607	2 645	2 617
kif11	3 621	1 724	**7.759**	6 997	0 862
kit	**16.867**	6 386	3 614	3 258	0 843
kith	2 807	3 860	**8.772**	2 222	3 158
kpcb	5 970	6 119	**6.866**	2 723	1 185
lck	**15.322**	10 835	5 060	1 713	1 000
lkha4	**13.176**	11 647	5 176	2 485	0 118
mapk2	**11.881**	5 941	10 297	1 273	1 188
mcr	3 617	4 255	3 617	3 689	**6.667**
met	**16.988**	10 000	5 422	7 643	1 205
mk01	**15.190**	10 633	7 595	3 893	3 544
mk10	**12.816**	10 000	3 495	2 087	1 346
mk14	**15.078**	8 616	3 640	3 713	1 073
mmp13	**17.028**	8 462	2 727	3 792	1 796
mp2k1	**10.833**	3 000	0 833	1 538	2 167
nos1	**6.600**	1 200	2 000	0 695	2 200
nram	**2.653**	2 449	0 204	0 561	2 449
pa2ga	10 612	**11.837**	0 204	2 136	0 816
parp1	6 575	**8.386**	5 752	2 285	1 181
pde5a	**13.568**	12 161	1 407	1 644	1 471
pgh1	**5.417**	5 181	3 021	2 103	0 302
pgh2	8 298	**9.099**	5 874	1 648	0 414
plk1	**14.906**	10 189	4 340	3 932	1 698
pnph	**11.961**	4 314	7 843	1 052	1 322
ppara	**7.882**	4 021	7 453	0 910	1 233
ppard	6 250	**7.083**	2 667	1 699	0 583
pparg	**12.190**	5 620	4 917	3 329	1 198
prgr	**5.666**	4 710	5 392	2 530	3 245
ptn1	**10.000**	**10.000**	9 538	5 716	2 174
pur2	**14.000**	7 600	10 400	0 741	7 200
pygm	**6.753**	3 896	1 558	0 250	5 714
pyrd	**13.333**	8 649	3 063	6 102	0 360
reni	**9.515**	3 689	3 689	2 076	0 935
rock1	**11.800**	10 000	2 400	1 081	0 800
rxra	1 221	4 885	**5.649**	0 845	0 611
sahh	**17.460**	16 825	9 524	0 650	5 079
src	**16.628**	12 529	2 490	1 696	1 667
tgfr1	**19.850**	19 398	4 211	1 496	0 451
thb	10 097	**12.816**	5 825	2 047	0 971
thrb	**13.883**	7 505	3 905	4 867	6 117
try1	**16.080**	7 572	3 608	5 182	3 653
tryb1	**9.054**	5 541	4 054	5 505	1 486
tysy	**18.349**	12 477	11 193	4 034	2 569
urok	**15.926**	7 160	6 049	1 676	0 988
vgfr2	**17.457**	9 389	5 281	4 033	0 538
wee1	**19.010**	**19.010**	3 762	11 008	0 396
xiap	3 030	**5.657**	4 444	2 638	3 636

Table S8: Multipose ChEMBL AUC for Vina, RF-Score (RF-S), NNScore (NNS), and CNN models trained using either only DUD-E docked poses or a combination, at a 2:1 ratio, of DUD-E poses and CSAR poses.

Target	DUD-E	2:1 D/C	Vina	RF-Score	NNScore
10378	**0.704**	0.464	0.533	0.663	0.461
10498	0.729	0.481	0.578	**0.744**	0.521
10752	**0.814**	0.753	0.733	0.530	0.383
11279	**0.828**	0.501	0.536	0.485	0.386
11359	**0.796**	0.711	0.672	0.726	0.558
11534	0.758	0.545	0.569	**0.759**	0.540
11631	**0.796**	0.727	0.668	0.764	0.487
12670	**0.880**	0.749	0.718	0.674	0.483
12968	0.495	0.492	0.645	**0.784**	0.602
18061	0.722	0.662	**0.740**	0.712	0.493
20014	**0.917**	0.767	0.791	0.647	0.477
219	0.674	**0.753**	0.516	0.583	0.532
276	**0.879**	0.738	0.852	0.803	0.572
28	**0.917**	0.653	0.762	0.544	0.282

Table S9: Multipose ChEMBL 0.5% ROC enrichment for Vina, RF-Score (RF-S), NNScore (NNS), and CNN models trained using either only DUD-E docked poses or a combination, at a 2:1 ratio, of DUD-E poses and CSAR poses.

Target	DUD-E	2:1 D/C	Vina	RF-Score	NNScore
10378	20.000	0.000	6.000	**34.000**	0.000
10498	34.000	2.000	4.000	**38.000**	2.000
10752	**54.000**	12.000	18.000	10.000	0.000
11279	**6.000**	0.000	0.000	0.000	2.000
11359	**46.000**	4.000	6.000	25.565	0.000
11534	26.000	0.000	2.000	**40.000**	8.000
11631	**26.000**	4.000	4.000	4.000	0.000
12670	**72.000**	16.000	10.000	16.180	0.000
12968	0.000	0.000	**8.000**	0.000	0.000
18061	8.000	4.000	14.000	**18.000**	0.000
20014	**100.000**	34.000	12.000	6.000	6.000
219	8.081	**10.101**	**10.101**	5.263	2.632
276	**90.000**	12.000	16.000	25.055	0.000
28	**80.000**	8.000	24.000	2.000	0.000

Table S10: Multipose ChEMBL 1.0% ROC enrichment for Vina, RF-Score (RF-S), NNScore (NNS), and CNN models trained using either only DUD-E docked poses or a combination, at a 2:1 ratio, of DUD-E poses and CSAR poses.

Target	DUD-E	2:1 D/C	Vina	RF-Score	NNScore
10378	13.000	0.000	4.000	**17.000**	0.000
10498	20.000	2.000	3.000	**22.000**	2.000
10752	**33.000**	13.000	12.000	7.000	2.000
11279	**7.000**	0.000	0.000	0.000	2.000
11359	**33.000**	6.000	7.000	19.000	0.000
11534	16.000	0.000	1.000	**25.475**	6.000
11631	**19.000**	3.000	6.000	4.000	1.000
12670	**44.000**	13.000	10.000	12.000	0.000
12968	1.000	1.000	**4.000**	3.000	0.000
18061	9.000	6.000	**12.000**	10.000	0.000
20014	**55.000**	21.000	8.000	4.000	6.000
219	**8.081**	**8.081**	7.071	7.895	5.263
276	**55.000**	11.000	14.000	17.000	0.000
28	**44.000**	4.000	20.000	1.360	0.000

Table S11: Multipose ChEMBL 2.0% ROC enrichment for Vina, RF-Score (RF-S), NNScore (NNS), and CNN models trained using either only DUD-E docked poses or a combination, at a 2:1 ratio, of DUD-E poses and CSAR poses.

Target	DUD-E	2:1 D/C	Vina	RF-Score	NNScore
10378	**10.000**	0.500	3.000	9.500	0.500
10498	13.500	1.000	3.000	**15.000**	1.500
10752	**18.500**	9.500	7.000	7.000	1.500
11279	**6.500**	0.000	0.000	0.000	1.000
11359	**22.000**	5.500	5.500	11.500	0.000
11534	10.500	1.000	1.500	**16.000**	4.000
11631	**12.500**	2.500	4.000	3.500	1.000
12670	**25.000**	11.000	8.000	8.000	0.500
12968	0.500	1.500	**3.000**	2.000	0.000
18061	5.500	4.500	**8.000**	6.500	0.000
20014	**32.000**	12.500	6.000	4.000	3.500
219	4.545	**8.081**	4.040	6.579	3.947
276	**30.500**	6.000	11.000	10.617	0.500
28	**26.500**	3.000	13.000	2.000	0.000

Table S12: Multipose ChEMBL 5.0% ROC enrichment for Vina, RF-Score (RF-S), NNScore (NNS), and CNN models trained using either only DUD-E docked poses or a combination, at a 2:1 ratio, of DUD-E poses and CSAR poses.

Target	DUD-E	2:1 D/C	Vina	RF-Score	NNScore
10378	**6.000**	1.000	2.600	4.600	1.200
10498	**7.200**	1.800	3.000	6.600	1.000
10752	**9.200**	6.800	5.600	3.800	1.000
11279	**6.600**	0.600	0.000	0.000	1.200
11359	**10.200**	5.000	3.800	6.600	0.200
11534	5.600	1.000	1.600	**8.109**	2.000
11631	**7.200**	2.800	4.200	3.800	1.000
12670	**13.200**	7.400	5.600	5.400	1.000
12968	0.200	0.800	**4.600**	2.064	0.200
18061	**5.600**	2.800	5.400	4.617	0.000
20014	**14.800**	7.800	5.600	2.400	1.800
219	3.838	**5.253**	3.030	4.737	2.632
276	**13.800**	4.000	8.200	7.200	1.400
28	**12.800**	3.000	8.200	1.400	0.000

Table S13: Multipose MUV AUC for Vina, RF-Score (RF-S), NNScore (NNS), and CNN models trained using either only DUD-E docked poses or a combination, at a 2:1 ratio, of DUD-E poses and CSAR poses.

Target	DUD-E	2:1 D/C	Vina	RF-S	NNS
466	**0.663**	0.492	0.593	0.592	0.417
548	**0.791**	0.697	0.460	0.345	0.365
600	0.559	0.489	0.578	**0.640**	0.528
689	0.514	0.377	0.596	0.562	**0.601**
692	**0.480**	0.447	0.413	0.429	0.457
832	0.402	0.495	**0.610**	0.610	0.339
846	0.384	0.504	**0.655**	0.458	0.412
852	0.348	0.501	**0.515**	0.405	0.374
859	0.560	0.488	0.517	**0.571**	0.474

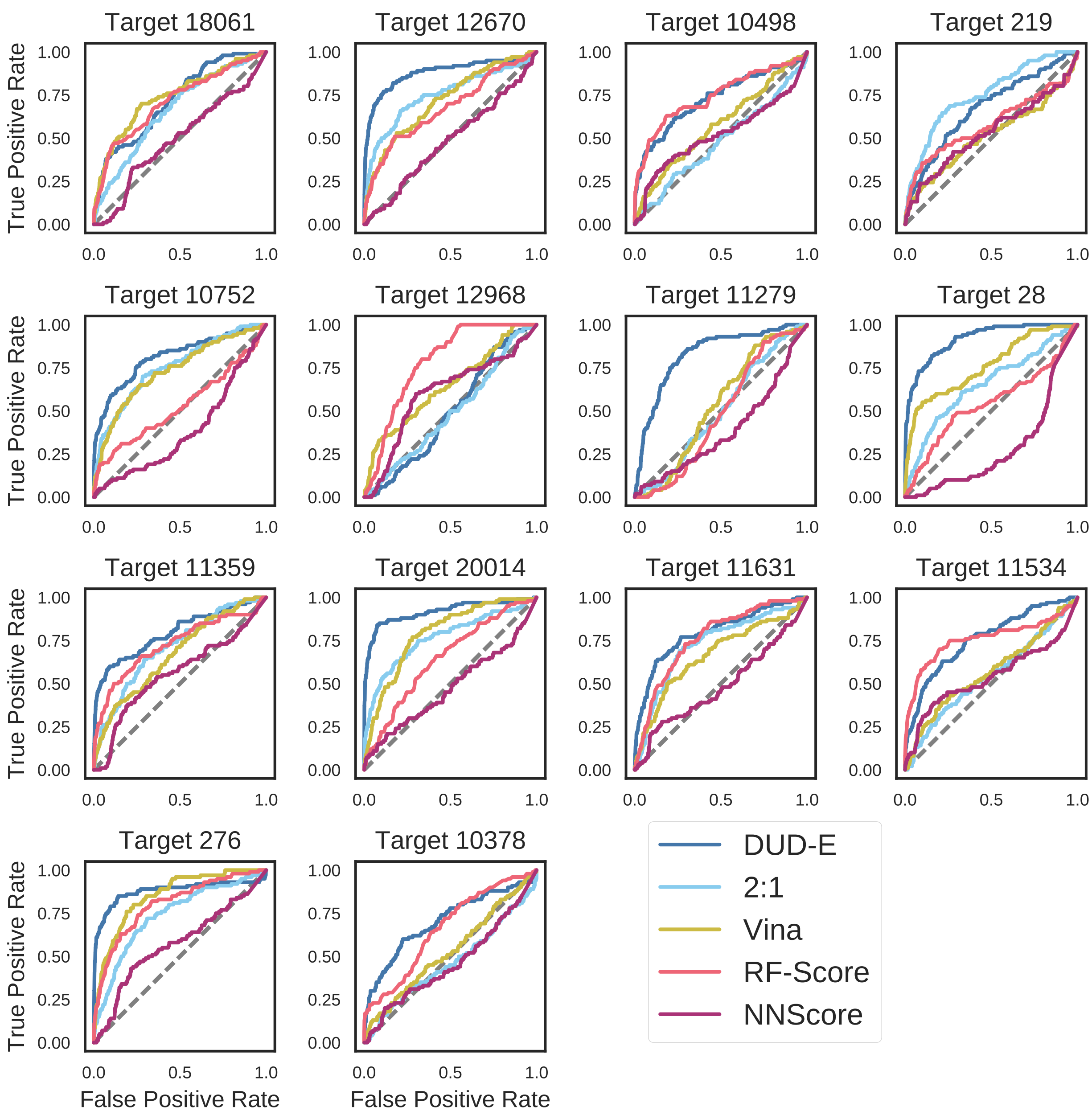

Figure S5: Per-target ROC curves using DUD-E CNN, 2:1 DUD-E/CSAR CNN, Vina, RF-Score, and NNScore for the ChEMBL test set.

Table S14: Multipose MUV 0.5% ROC enrichment for Vina, RF-Score (RF-S), NNScore (NNS), and CNN models trained using either only DUD-E docked poses or a combination, at a 2:1 ratio, of DUD-E poses and CSAR poses.

Target	DUD-E	2:1 D/C	Vina	RF-S	NNS
466	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
548	0.000	**6.667**	0.000	0.000	0.000
600	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
689	**6.667**	0.000	0.000	0.000	0.000
692	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
832	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
846	**6.667**	0.000	0.000	0.000	0.000
852	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
859	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**

Table S15: Multipose MUV 1.0% ROC enrichment for Vina, RF-Score (RF-S), NNScore (NNS), and CNN models trained using either only DUD-E docked poses or a combination, at a 2:1 ratio, of DUD-E poses and CSAR poses.

Target	DUD-E	2:1	Vina	RF-S	NNS
466	0.000	3.333	3.333	**6.667**	0.000
548	**6.667**	**6.667**	0.000	0.000	0.000
600	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
689	**3.333**	0.000	**3.333**	**3.333**	**3.333**
692	0.000	0.000	0.000	**3.333**	0.000
832	0.000	0.000	**3.333**	0.000	0.000
846	**3.333**	0.000	0.000	0.000	0.000
852	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
859	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**

Table S16: Multipose MUV 2.0% ROC enrichment for Vina, RF-Score (RF-S), NNScore (NNS), and CNN models trained using either only DUD-E docked poses or a combination, at a 2:1 ratio, of DUD-E poses and CSAR poses.

Target	DUD-E	2:1	Vina	RF-S	NNS
466	1.667	1.667	**3.333**	**3.333**	0.000
548	**6.667**	**6.667**	0.000	0.000	0.000
600	0.000	0.000	**5.000**	0.000	0.000
689	**1.667**	0.000	**1.667**	**1.667**	**1.667**
692	0.000	0.000	0.000	**1.667**	0.000
832	0.000	0.000	**5.000**	0.000	1.667
846	**1.667**	**1.667**	0.000	0.000	0.000
852	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
859	0.000	0.000	**1.667**	**1.667**	0.000

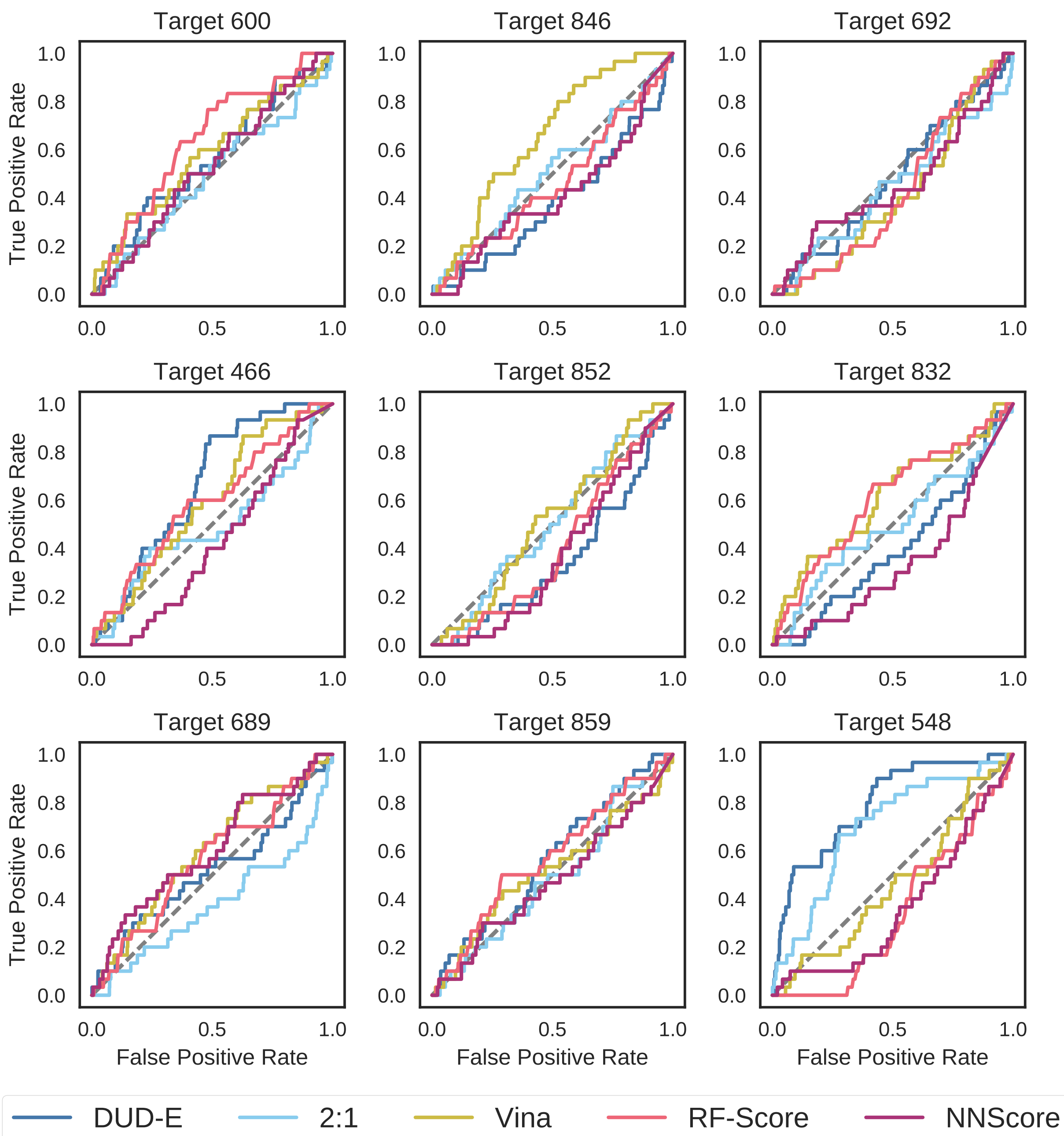

Figure S6: Per-target ROC curves using DUD-E CNN, 2:1 DUD-E/CSAR CNN, Vina, RF-Score and NNScore for the MUV test set.

Table S17: Multipose MUV 5.0% ROC enrichment for Vina, RF-Score (RF-S), NNScore (NNS), and CNN models trained using either only DUD-E docked poses or a combination, at a 2:1 ratio, of DUD-E poses and CSAR poses.

Target	DUD-E	2:1	Vina	RF-S	NNS
466	1.333	0.667	1.333	**2.000**	0.000
548	**6.667**	2.667	0.000	0.000	1.333
600	1.333	0.000	**2.667**	1.333	0.000
689	**2.000**	0.000	**2.000**	1.333	**2.000**
692	0.000	0.000	0.000	**0.667**	**0.667**
832	0.000	0.000	**3.333**	2.148	0.667
846	0.667	**1.333**	0.667	0.667	0.000
852	0.000	0.000	**0.667**	0.000	0.000
859	**2.000**	0.667	1.333	1.333	1.333