Trade off predictivity and explainability for ML-powered predictive toxicology: an in-depth investigation with Tox21 datasets

Leihong Wu[1], Ruili Huang[2], Igor V. Tetko,[3] Zhonghua Xia,[3] Joshua Xu[1], Weida Tong[1*]

1 Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, FDA. 3900 NCTR Rd., Jefferson, Arkansas, 72079, USA

2 Division of Preclinical Innovation, National Center for Advancing Translational Sciences, National Institutes of Health, 9800 Medical Center Drive, Rockville, Maryland 20850, USA

3 Institute of Structural Biology, Helmholtz Zentrum München-Research Center for Environmental Health (GmbH), Ingolstädter Landstraße 1, 85764, Neuherberg, Germany

Table s1. Averaged/collapsed AUC for 65 Tox21 assays

ASSAY	Positive	Negative	AUC	AUC_TEST
tox21-er-luc-bg1-4e2-agonist-p4 (1)	104	6364	0.93	0.90
tox21-pr-bla-agonist-p1 (1)	103	6465	0.90	0.95
tox21-erb-bla-p1 (1)	85	6296	0.88	0.92
tox21-ms-ache-p2 (-1)	93	6586	0.87	0.81
tox21-mitotox-p1 (-1)	745	4551	0.87	0.85
tox21-ahr-p1 (1)	692	5321	0.86	0.89
tox21-pr-bla-antagonist-p1 (-1)	667	4919	0.86	0.89
tox21-hdac-p1 (-1)	251	6287	0.86	0.89
tox21-ar-bla-agonist-p1 (1)	216	6149	0.85	0.88
tox21-pxr-p1 (1)	1288	4259	0.85	0.85
tox21-luc-biochem-p1 (-1)	592	5702	0.85	0.87
tox21-ar-mda-kb2-luc-agonist-p3 (1)	167	6190	0.85	0.88
tox21-car-agonist-p1 (1)	671	5288	0.84	0.86
tox21-car-antagonist-p1 (-1)	810	4443	0.84	0.88
tox21-ar-mda-kb2-luc-antagonist-p2 (-1)	650	4577	0.84	0.83
tox21-rar-agonist-p1 (1)	265	5196	0.84	0.89
tox21-ache-p1 (-1)	601	5540	0.83	0.87
tox21-err-p1 (1)	818	4784	0.83	0.84
tox21-pgc-err-p1 (-1)	303	4542	0.83	0.85
tox21-er-luc-bg1-4e2-antagonist-p1 (-1)	334	5611	0.81	0.86
tox21-ar-bla-antagonist-p1 (-1)	386	5265	0.81	0.84
tox21-gr-hela-bla-agonist-p1 (1)	156	6224	0.81	0.90
tox21-rar-antagonist-p2 (-1)	495	4155	0.81	0.88
tox21-ache-p3 (-1)	292	6068	0.81	0.84
tox21-gr-hela-bla-antagonist-p1 (-1)	281	5270	0.81	0.86
tox21-ache-p5 (-1)	190	6396	0.81	0.82
tox21-erb-bla-antagonist-p1 (-1)	342	5073	0.81	0.77
tox21-aromatase-p1 (-1)	402	5081	0.80	0.83
tox21-shh-3t3-gli3-antagonist-p1 (-1)	633	4133	0.80	0.86
tox21-err-p1 (-1)	373	4828	0.80	0.81
tox21-ar-mda-kb2-luc-agonist-p1 (1)	256	6573	0.79	0.86
tox21-er-bla-agonist-p2 (1)	328	6161	0.79	0.81
tox21-pgc-err-p1 (1)	290	5427	0.79	0.84
tox21-ror-cho-antagonist-p1 (-1)	291	4349	0.79	0.65
tox21-er-luc-bg1-4e2-antagonist-p2 (-1)	277	5380	0.78	0.84
tox21-p53-bla-p1 (1)	415	6031	0.78	0.80
tox21-ar-mda-kb2-luc-antagonist-p1 (-1)	299	5339	0.77	0.73
tox21-are-bla-p1 (1)	750	4377	0.77	0.82
tox21-h2ax-cho-p2 (1)	259	5985	0.77	0.76
tox21-pparg-bla-antagonist-p1 (-1)	295	5101	0.77	0.81
tox21-fxr-bla-antagonist-p1 (-1)	163	5364	0.77	0.81
tox21-rxr-bla-agonist-p1 (1)	254	4982	0.77	0.81

tox21-er-bla-antagonist-p1 (-1)	251	5394	0.77	0.80
tox21-ap1-agonist-p1 (1)	317	5620	0.77	0.77
tox21-sbe-bla-antagonist-p1 (-1)	263	5788	0.76	0.80
tox21-gh3-tre-agonist-p1 (1)	94	6404	0.76	0.71
tox21-casp3-hepg2-p1 (1)	160	6235	0.76	0.80
tox21-tshr-antagonist-p1 (-1)	138	6190	0.75	0.75
tox21-elg1-luc-agonist-p1 (1)	229	6487	0.75	0.82
tox21-gh3-tre-antagonist-p1 (-1)	255	4739	0.75	0.78
tox21-tshr-agonist-p1 (1)	216	5967	0.75	0.78
tox21-vdr-bla-agonist-p1 (1)	28	5939	0.74	0.67
tox21-vdr-bla-antagonist-p1 (-1)	48	5437	0.74	0.71
tox21-pparg-bla-agonist-p1 (1)	172	5915	0.73	0.75
tox21-hre-bla-agonist-p1 (1)	33	6297	0.73	0.71
tox21-er-luc-bg1-4e2-agonist-p2 (1)	601	4945	0.73	0.81
tox21-esre-bla-p1 (1)	56	5835	0.73	0.80
tox21-trhr-hek293-p1 (-1)	48	6543	0.73	0.72
tox21-sbe-bla-agonist-p1 (1)	25	6702	0.72	0.77
tox21-hse-bla-p1 (1)	313	5561	0.72	0.76
tox21-trhr-hek293-p1 (1)	41	6608	0.72	0.58
tox21-fxr-bla-agonist-p2 (1)	90	6032	0.71	0.75
tox21-ppard-bla-agonist-p1 (1)	77	5780	0.71	0.73
tox21-casp3-cho-p1 (1)	80	6517	0.71	0.78
tox21-ppard-bla-antagonist-p1 (-1)	56	5568	0.70	0.73

Table s2. Full Dataset of the Chemicals and Endpoints used in this study

(separately attached)