



A context-aware deconfounding autoencoder for robust prediction of personalized clinical drug response from cell-line compound screening

In the format provided by the
authors and unedited

A Context-aware Deconfounding Autoencoder for Robust Prediction of Personalized Clinical Drug Response From Cell Line Compound Screening

DI HE¹, QIAO LIU², YOU WU¹, AND LEI XIE^{1,2,3,*}

¹PhD Program in Computer Science, Graduate Center, City University of New York

²Department of Computer Science, Hunter College, City University of New York

³Feil Family Brain & Mind Research Institute, Weill Cornell Medicine, Cornell University

SUPPLEMENTAL TABLES

Table S1. Average ranks of AUROCs across all drugs tested on PDTC dataset by CODE-AE variants. The best performance is underlined.

	Normalization	Non-normalization
CODE-AE-ADV	<u>1.70±1.27</u>	3.80±2.22
CODE-AE-BASE	6.34±3.26	9.02±2.22
CODE-AE-MMD	7.08±3.91	6.32±2.65
CODE-AE-ADV(CONCAT)	6.50±2.76	4.40±2.58
CODE-AE-BASE(CONCAT)	9.20±3.42	7.52±2.52
CODE-AE-MMD(CONCAT)	8.70±2.62	6.90±2.50

Table S2. t-Test results on gene expression differences between responsive and resistant groups per drug-target pair.

Drug	Target	P-value	Avg. Expression (Responsive)	Avg. Expression (Resistant)	Bonferroni	FDR-BH
681640	WEE1	1.10E-22	4.0789	4.6994	1.12E-20	2.44E-22
681640	CHEK1	0.079	3.4623	3.324	1	0.08664516
(5Z)-7-Oxozeaenol	MAP3K7	2.80E-07	3.4173	3.6632	2.86E-05	3.91E-07
AG-014699	PARP2	6.10E-05	4.2797	4.1018	0.006222	7.59E-05
AG-014699	PARP1	0.31	6.0258	5.9679	1	0.329375
AICAR	PRKAA1	8.50E-101	3.5251	4.8481	8.67E-99	5.78E-100
AMG-706	PDGFRB	6.90E-231	3.3645	6.0379	7.04E-229	7.04E-229
AMG-706	KDR	2.20E-46	2.1994	3.2249	2.24E-44	7.01E-46
AMG-706	PDGFRA	9.00E-20	2.9868	4.1865	9.18E-18	1.91E-19
AMG-706	RET	1.10E-06	1.655	1.17	0.0001122	1.46E-06
AMG-706	KIT	0.79	2.4167	2.3898	1	0.79
Axitinib	PDGFRB	4.00E-184	3.7331	6.1445	4.08E-182	1.02E-182
Axitinib	PDGFRA	5.80E-74	2.2025	4.2702	5.92E-72	2.82E-73
Axitinib	KDR	1.40E-24	2.5913	3.3782	1.43E-22	3.25E-24
Axitinib	KIT	0.45	2.4023	2.4853	1	0.459
AZ628	BRAF	2.40E-24	2.8583	3.2854	2.45E-22	5.44E-24
AZD6482	PIK3CB	0.001	3.8311	4.019	0.102	0.00118605
AZD7762	CHEK1	1.70E-185	4.1895	2.2045	1.73E-183	5.78E-184
AZD7762	CHEK2	5.40E-134	4.3098	2.6937	5.51E-132	5.51E-133
AZD8055	MTOR	0.19	4.245	4.1776	1	0.20617021
BI-2536	PLK2	2.60E-143	2.8942	5.4151	2.65E-141	3.31E-142
BI-2536	PLK3	2.90E-41	3.1475	3.8938	2.96E-39	7.99E-41
BI-2536	PLK1	0.00044	4.1665	3.8311	0.04488	0.00053429
BMS-536924	IGF1R	1.80E-72	2.3833	4.0068	1.84E-70	8.35E-72
BMS-754807	IGF1R	9.60E-29	3.8051	2.9307	9.79E-27	2.33E-28
Bosutinib	SRC	2.30E-94	3.4436	4.8906	2.35E-92	1.38E-93
Bosutinib	ABL1	4.80E-77	4.0001	5.1314	4.90E-75	2.45E-76
Bosutinib	TEC	1.10E-65	2.7255	1.1407	1.12E-63	4.16E-65
BX-795	AURKC	4.80E-35	1.163	0.6489	4.90E-33	1.22E-34
BX-795	TBK1	3.70E-08	3.4926	3.822	3.77E-06	5.32E-08
BX-795	IKBKB	2.20E-07	4.7648	4.4565	2.24E-05	3.12E-07
BX-795	AURKB	0.35	4.8741	4.7992	1	0.36428571
BX-795	PDK1	0.37	2.8806	2.9369	1	0.38121212
Camptothecin	TOP1	4.50E-111	4.4576	5.686	4.59E-109	3.82E-110
CEP-701	NTRK3	1.80E-40	2.3181	0.807	1.84E-38	4.83E-40
CEP-701	JAK2	3.10E-12	2.3605	2.8186	3.16E-10	5.55E-12
CEP-701	NTRK2	7.10E-12	3.8394	2.4591	7.24E-10	1.25E-11
CEP-701	FLT3	1.10E-06	1.2452	0.7004	0.0001122	1.46E-06
CEP-701	NTRK1	1.30E-05	0.8474	0.5553	0.001326	1.68E-05
CHIR-99021	GSK3A	0.0033	5.9035	6.0019	0.3366	0.00386897
CHIR-99021	GSK3B	0.56	4.1111	4.0846	1	0.56554455
Embelin	XIAP	5.00E-72	3.5752	4.2196	5.10E-70	2.22E-71
FK866	NAMPT	3.60E-169	4.4288	6.6042	3.67E-167	6.12E-168
GDC0941	PIK3CA	2.20E-69	2.7288	3.5768	2.24E-67	8.98E-69
GDC0941	PIK3CD	9.00E-29	1.6778	2.2801	9.18E-27	2.24E-28
Gefitinib	EGFR	4.80E-47	3.9776	5.7095	4.90E-45	1.58E-46
GSK1904529A	IGF1R	2.90E-100	2.4047	3.9939	2.96E-98	1.85E-99
GSK269962A	ROCK2	3.20E-20	3.2141	3.7583	3.26E-18	6.94E-20
GSK269962A	ROCK1	0.021	3.6288	3.469	1	0.02353846

GW 441756	NTRK1	1.50E-10	0.3287	0.6309	1.53E-08	2.43E-10
JNJ-26854165	MDM2	4.50E-07	5.7871	5.4236	4.59E-05	6.12E-07
JNK Inhibitor VIII	MAPK8	2.50E-46	3.6722	4.5588	2.55E-44	7.73E-46
JQ1	BRD2	1.70E-08	6.9069	7.1648	1.73E-06	2.55E-08
JQ1	BRD4	2.70E-08	4.5332	4.7908	2.75E-06	3.93E-08
JQ1	BRDT	0.004	0.1288	0.255	0.408	0.00463636
JQ1	BRD3	0.019	3.7328	3.8674	1	0.02177528
KU-55933	ATM	0.03	3.5916	3.4025	1	0.03326087
LY317615	PRKCB	4.10E-10	3.6947	2.8402	4.18E-08	6.43E-10
MK-2206	AKT1	2.90E-106	5.5111	6.7304	2.96E-104	2.28E-105
MK-2206	AKT2	5.00E-39	5.5542	6.3624	5.10E-37	1.31E-38
Nilotinib	ABL1	6.90E-87	3.7639	4.9984	7.04E-85	3.70E-86
NU-7441	PRKDC	0.32	5.5597	5.4917	1	0.33649485
Obatoclox Mesylate	BCL2L2	1.50E-138	3.3805	4.8422	1.53E-136	1.70E-137
Obatoclox Mesylate	BCL2L1	3.10E-88	5.5934	6.6164	3.16E-86	1.76E-87
Obatoclox Mesylate	BCL2	1.60E-25	2.7184	3.7067	1.63E-23	3.80E-25
Obatoclox Mesylate	MCL1	1.20E-15	6.8187	7.3133	1.22E-13	2.31E-15
PAC-1	CASP7	1.50E-173	3.3918	5.0627	1.53E-171	3.06E-172
PAC-1	CASP3	1.10E-70	3.709	4.6999	1.12E-68	4.67E-70
PD-0325901	MAP2K1	1.30E-17	4.5549	5.0474	1.33E-15	2.65E-17
PD-0325901	MAP2K2	5.70E-05	6.7388	6.9083	0.005814	7.18E-05
PD-0332991	CDK4	2.40E-43	6.4158	7.1613	2.45E-41	6.80E-43
PD-0332991	CDK6	2.60E-12	3.271	2.4058	2.65E-10	4.74E-12
PD-173074	FGFR1	1.80E-18	4.1757	4.9268	1.84E-16	3.75E-18
PD-173074	FGFR2	5.40E-11	3.1222	3.8407	5.51E-09	8.88E-11
PD-173074	FGFR3	3.10E-09	2.9842	3.7217	3.16E-07	4.72E-09
PF-4708671	RPS6KB1	0.021	3.6593	3.5448	1	0.02353846
PF-562271	PTK2B	3.10E-07	4.3494	4.0449	3.16E-05	4.27E-07
PF-562271	PTK2	0.00085	5.769	5.6216	0.0867	0.00102
PLX4720	BRAF	1.80E-08	2.9273	3.2625	1.84E-06	2.66E-08
RO-3306	CDK1	0.23	4.6398	4.737	1	0.24694737
SB 216763	GSK3A	4.70E-44	5.4215	6.0307	4.79E-42	1.37E-43
SB 216763	GSK3B	2.30E-11	3.7463	4.084	2.35E-09	3.98E-11
SL 0101-1	PIM3	5.10E-61	4.7294	5.7019	5.20E-59	1.86E-60
SL 0101-1	RPS6KA1	3.00E-11	4.5357	4.9659	3.06E-09	5.10E-11
SL 0101-1	PIM1	1.40E-05	4.7857	4.4367	0.001428	1.79E-05
SL 0101-1	AURKB	0.00041	4.2786	3.9083	0.04182	0.00050386
Sorafenib	PDGFRB	2.30E-222	2.6129	5.6893	2.35E-220	1.17E-220
Sorafenib	PDGFRA	1.20E-166	1.0217	3.7956	1.22E-164	1.75E-165
Sorafenib	KDR	5.70E-58	1.6047	2.8772	5.81E-56	2.00E-57
Sorafenib	KIT	2.10E-12	2.8686	2.0071	2.14E-10	3.89E-12
Sorafenib	RAF1	2.00E-10	5.0648	5.4119	2.04E-08	3.19E-10
Tamoxifen	ESR1	2.20E-09	0.8943	1.5082	2.24E-07	3.40E-09
TW 37	MCL1	4.10E-106	6.7557	7.6602	4.18E-104	2.99E-105
TW 37	BCL2	3.80E-55	3.1667	1.9914	3.88E-53	1.29E-54
TW 37	BCL2L1	1.90E-17	6.3134	6.6994	1.94E-15	3.80E-17
Vorinostat	HDAC1	1.90E-66	5.5303	6.6089	1.94E-64	7.45E-66
Vorinostat	HDAC2	1.50E-44	5.1935	6.1185	1.53E-42	4.50E-44
Vorinostat	HDAC6	2.70E-16	5.1075	4.7073	2.75E-14	5.30E-16
Vorinostat	HDAC3	1.30E-15	5.1247	5.4736	1.33E-13	2.46E-15
YK 4-279	DHX9	1.20E-06	6.2476	6.4496	0.0001224	1.57E-06
ZM-447439	AURKB	3.00E-133	5.1265	2.8614	3.06E-131	2.78E-132
ZM-447439	AURKA	5.10E-11	3.732	3.2247	5.20E-09	8.53E-11

Table S3. Classification performance differences between CODE-AE (fine-tuned) over gene expression and propagated mutation data, as well as elastic net over **concatenated pre-trained (CODE-AE) representation** of gene expression and propagated mutation data.

	CODE-AE (gex)		CODE-AE (mut)		Elastic Net (gex + mut)	
	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
Temozolomide (TCGA)	0.7509±0.02	0.7617±0.01	0.5147±0.09	0.7078±0.1	0.4771±0.05	0.6837±0.12
Gemcitabine (TCGA)	0.5945±0.01	0.7512±0.01	0.6104±0.04	0.7546±0.02	0.48±0.04	0.6888±0.11
Cisplatin (TCGA)	0.6515±0.06	0.7429±0.01	0.6133±0.15	0.7354±0.03	0.507±0.01	0.7034±0.09
Sorafenib (TCGA)	0.6305±0.02	0.7048±0.06	0.5235±0.03	0.642±0.13	0.5091±0.02	0.7101±0.06
5-Fluorouracil (TCGA)	0.8675±0.03	0.7398±0.01	0.6318±0.05	0.6846±0.06	0.4818±0.04	0.6717±0.13
Wee1 Inhibitor	0.778±0.06	0.7895±0.02	0.5622±0.01	0.7071±0.06	0.4711±0.06	0.6936±0.09
(5Z)-7-Oxozeaenol	0.7214±0.08	0.7429±0.01	0.5837±0.11	0.6339±0.15	0.5796±0.16	0.7772±0.05
Rucaparib	0.725±0.03	0.7629±0.03	0.551±0.18	0.6122±0.18	0.4694±0.06	0.6768±0.15
AICA Ribonucleotide	0.8062±0.13	0.8118±0.03	0.5857±0.1	0.636±0.15	0.5393±0.08	0.6934±0.08
Motesanib	0.694±0.11	0.7637±0.02	0.5867±0.03	0.7037±0.04	0.4322±0.05	0.4425±0.03
Axitinib	0.84±0.05	0.7623±0.03	0.6467±0.04	0.65±0.11	0.5022±0.0	0.7183±0.04
AZ628	0.754±0.1	0.6792±0.15	0.7878±0.0	0.7457±0.02	0.51±0.02	0.7395±0.01
AZD6482	0.615±0.02	0.7517±0.02	0.5267±0.04	0.6168±0.1	0.5067±0.01	0.681±0.11
AZD7762	0.7036±0.07	0.76±0.06	0.5429±0.03	0.7302±0.04	0.602±0.15	0.563±0.15
AZD8055	0.708±0.05	0.7629±0.02	0.5633±0.06	0.6457±0.11	0.5022±0.0	0.6988±0.08
BI-2536	0.8071±0.12	0.7581±0.07	0.5612±0.01	0.707±0.09	0.5265±0.05	0.7266±0.05
BMS-536924	0.675±0.07	0.7511±0.01	0.5644±0.02	0.6652±0.09	0.52±0.04	0.6857±0.1
BMS-754807	0.786±0.06	0.7433±0.07	0.4944±0.12	0.6021±0.16	0.5378±0.08	0.7398±0.01
Bosutinib	0.764±0.05	0.7579±0.01	0.6922±0.02	0.7368±0.0	0.5089±0.02	0.6842±0.11
BX795	0.862±0.05	0.7655±0.03	0.6689±0.06	0.676±0.05	0.5611±0.12	0.7637±0.05
Camptothecin	0.726±0.02	0.7647±0.04	0.7333±0.04	0.7498±0.01	0.5267±0.05	0.6948±0.08
Lestaurtinib	0.654±0.05	0.7795±0.02	0.5689±0.06	0.6466±0.11	0.4711±0.06	0.6634±0.15
CHIR-99021	0.664±0.06	0.7854±0.01	0.68±0.27	0.6527±0.17	0.5178±0.05	0.5575±0.07
Embelin	0.7411±0.13	0.7433±0.01	0.5837±0.09	0.723±0.05	0.4408±0.12	0.6688±0.16
Daporinad	0.7±0.03	0.7376±0.05	0.6694±0.03	0.7571±0.01	0.5306±0.06	0.7072±0.09
Pictilisib	0.646±0.02	0.7749±0.01	0.5478±0.03	0.6847±0.1	0.5222±0.07	0.5794±0.03
GSK1904529A	0.8±0.02	0.7671±0.03	0.6327±0.0	0.7462±0.01	0.5551±0.11	0.761±0.02
Gefitinib	0.613±0.04	0.7781±0.01	0.5722±0.04	0.6343±0.13	0.51±0.02	0.7395±0.01
GSK269962A	0.706±0.03	0.7903±0.04	0.5356±0.02	0.6822±0.11	0.5244±0.05	0.7202±0.03
GW441756	0.626±0.08	0.76±0.02	0.62±0.2	0.67±0.13	0.4978±0.0	0.6819±0.11
Serdemetan	0.5786±0.05	0.7517±0.02	0.6245±0.05	0.7214±0.01	0.4592±0.09	0.5322±0.07
JNK Inhibitor VIII	0.712±0.05	0.7545±0.05	0.5956±0.08	0.6151±0.16	0.5067±0.01	0.6948±0.08
JQ1	0.59±0.05	0.7553±0.01	0.64±0.11	0.6673±0.14	0.5244±0.05	0.7237±0.03
KU-55933	0.772±0.04	0.7429±0.02	0.5489±0.02	0.5858±0.12	0.51±0.02	0.7395±0.01
Enzastaurin	0.6571±0.06	0.7686±0.02	0.7367±0.0	0.7784±0.01	0.4714±0.06	0.7275±0.05
MK-2206	0.618±0.09	0.7412±0.08	0.5444±0.05	0.7368±0.0	0.4844±0.03	0.6863±0.1
Nilotinib	0.76±0.06	0.7066±0.09	0.8711±0.01	0.8352±0.0	0.4933±0.01	0.6987±0.08
NU7441	0.606±0.03	0.7395±0.03	0.5222±0.05	0.6626±0.11	0.4833±0.15	0.4774±0.14
Obatoclax Mesylate	0.847±0.05	0.7787±0.01	0.8556±0.0	0.8607±0.0	0.4889±0.02	0.6878±0.1
PAC-1	0.6821±0.01	0.7717±0.03	0.5755±0.07	0.7182±0.06	0.4816±0.04	0.6825±0.14
PD0325901	0.673±0.04	0.7594±0.03	0.5489±0.07	0.6305±0.13	0.51±0.02	0.7395±0.01
Palbociclib	0.7±0.1	0.74±0.07	0.6433±0.0	0.7313±0.01	0.5244±0.05	0.7127±0.05
PD173074	0.59±0.04	0.6762±0.1	0.5533±0.03	0.6712±0.1	0.54±0.08	0.7163±0.04
PF-4708671	0.68±0.05	0.795±0.02	0.5533±0.05	0.6951±0.08	0.4978±0.09	0.5655±0.05
PF-562271	0.5893±0.03	0.7622±0.02	0.5408±0.07	0.6874±0.13	0.449±0.1	0.6714±0.16
PLX-4720	0.675±0.01	0.7676±0.04	0.5911±0.0	0.6938±0.09	0.4933±0.01	0.6945±0.08

RO-3306	0.65±0.02	0.7877±0.02	0.6311±0.03	0.6917±0.09	0.5133±0.03	0.7049±0.06
SB216763	0.638±0.09	0.746±0.02	0.5611±0.11	0.7368±0.0	0.4856±0.11	0.4908±0.09
SL0101	0.826±0.03	0.7693±0.01	0.6±0.13	0.6476±0.14	0.5044±0.01	0.6806±0.11
Sorafenib	0.6982±0.03	0.6231±0.17	0.5±0.06	0.6635±0.11	0.498±0.0	0.6893±0.12
Tamoxifen	0.636±0.04	0.7559±0.06	0.5578±0.0	0.7068±0.06	0.4644±0.07	0.6631±0.15
TW 37	0.6643±0.07	0.7333±0.0	0.5714±0.08	0.6885±0.1	0.4286±0.14	0.6671±0.17
Vorinostat	0.693±0.04	0.7008±0.04	0.8267±0.01	0.8599±0.0	0.5444±0.09	0.7443±0.01
YK-4-279	0.6482±0.07	0.74±0.01	0.5388±0.07	0.6862±0.08	0.6122±0.1	0.6906±0.13
ZM447439	0.714±0.08	0.8122±0.01	0.5556±0.02	0.6827±0.07	0.51±0.02	0.7395±0.01

SUPPLEMENTAL FIGURES

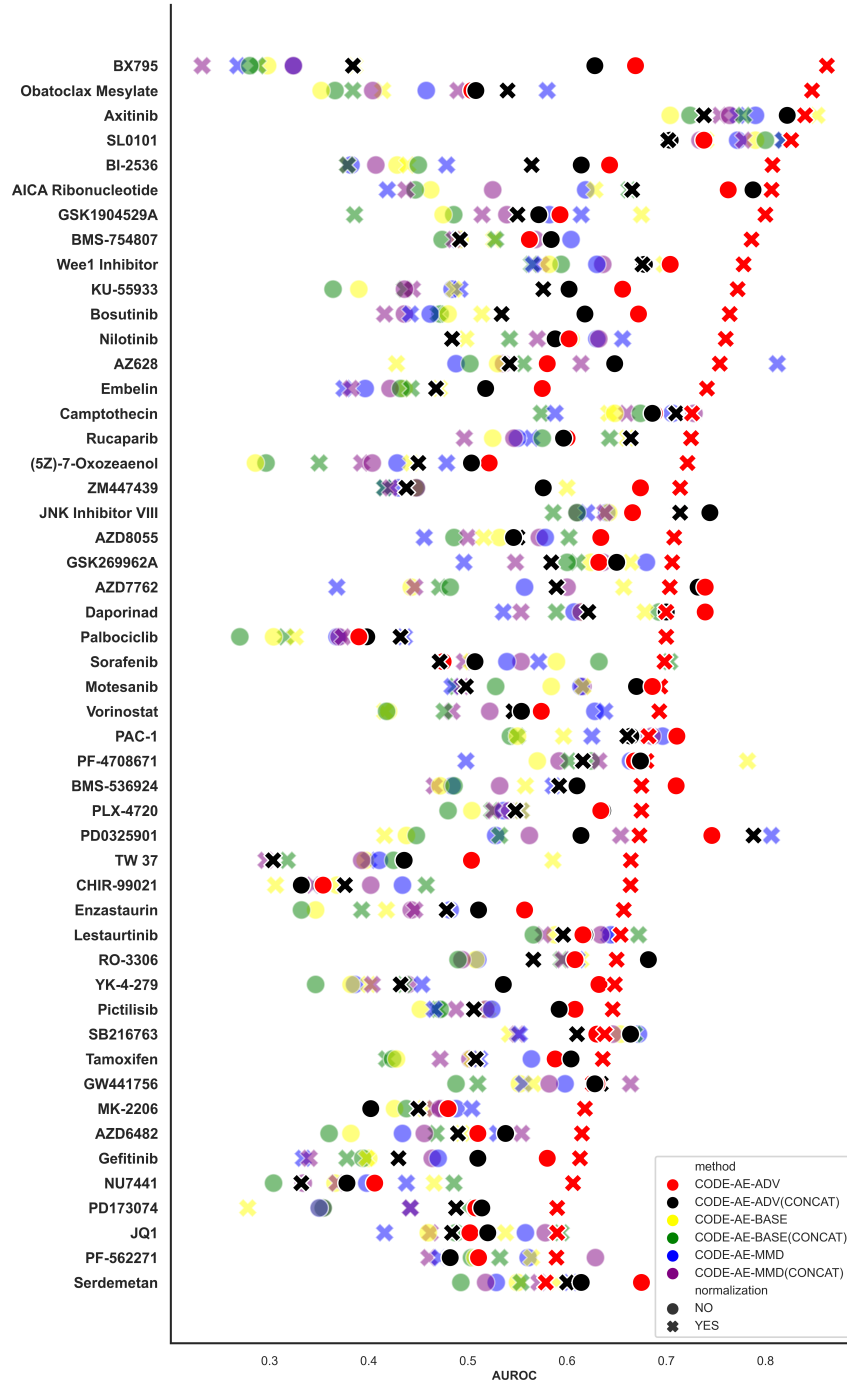


Fig. S1. PDTC drug response classification performance comparison of CODE-AE variants

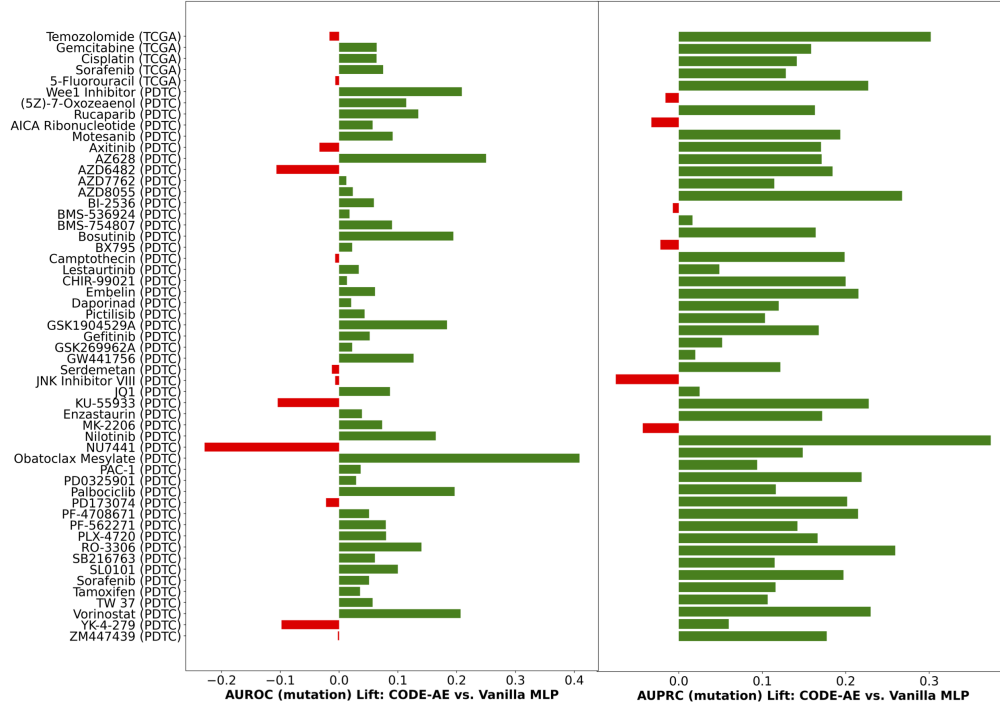


Fig. S2. Classification performance improvements achieved by CODE-AE over vanilla MLP when applied over propagated mutation data

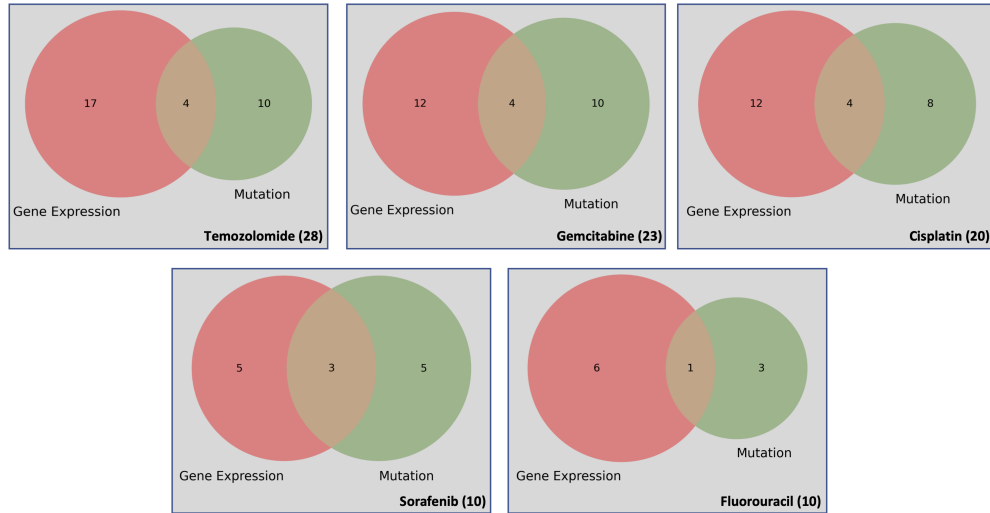


Fig. S3. True positives predicted from CODE-AE over gene expressions and propagated mutations, respectively, in the TCGA patient chemotherapy response prediction task. *Note: right bottom corner of each sub Venn diagram shows the drug name and the number of positive ground truth in the test set.*