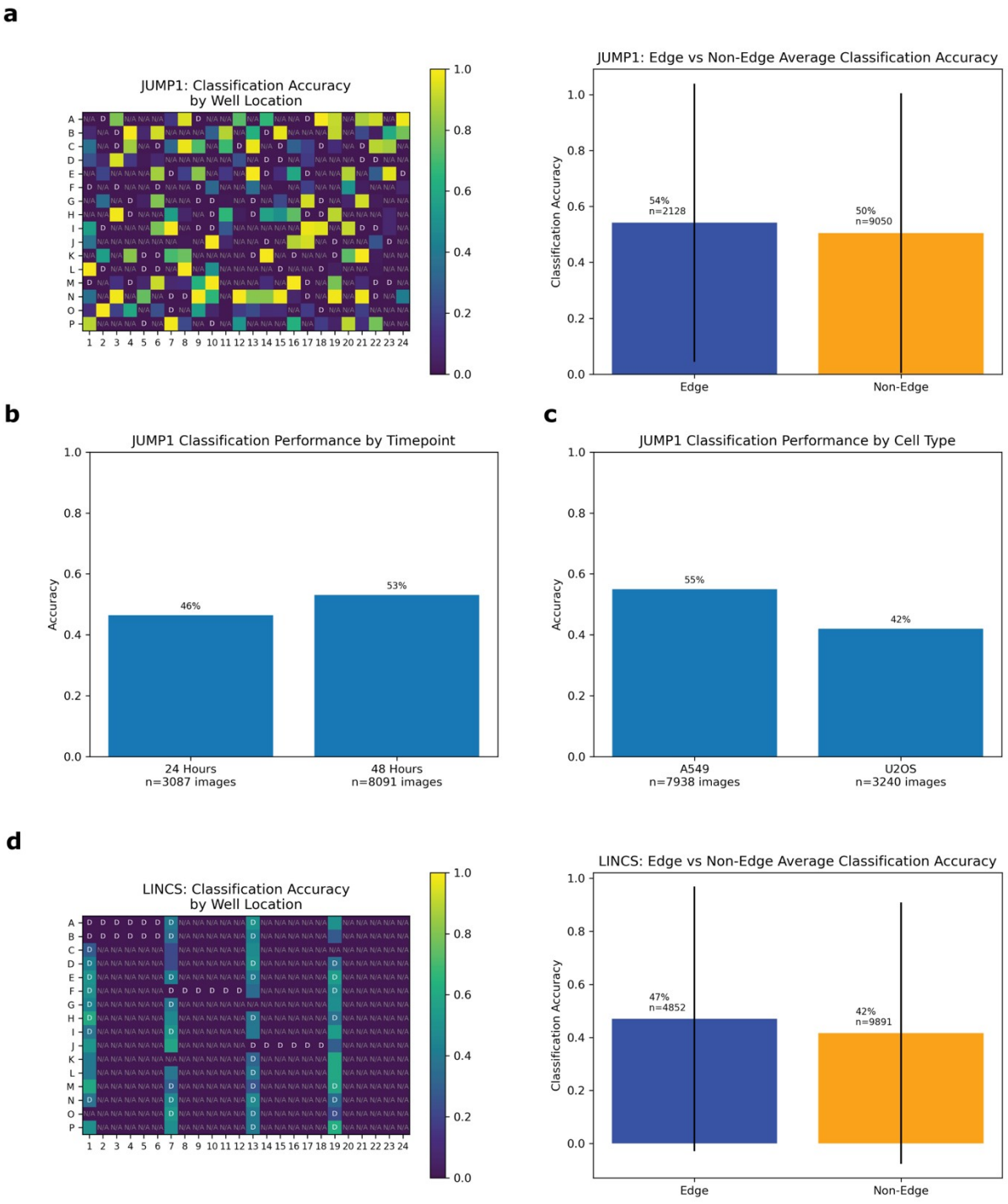
Electronic Supplementary Material (ESI) for Digital Discovery.
This journal is © The Royal Society of Chemistry 2023

Supplemental Information:

Deep Representation Learning Determines Drug Mechanism of Action from Cell Painting Images

Daniel R. Wong[1], David J. Logan[2], Santosh Hariharan[3], Robert Stanton[1], Djork-Arné Clevert[1], Andrew Kiruluta[1]

1. *Machine Learning and Computational Sciences, Pfizer Worldwide Research Development and Medical, 610 Main Street, Cambridge, Massachusetts 02139, United States*
2. *Internal Medicine Research Unit, Pfizer Worldwide Research Development and Medical, 610 Main Street, Cambridge, Massachusetts 02139, United States*
3. *Discovery Sciences, Pfizer Global Research and Development, Groton Laboratories, 280 Shennecossett Rd, Groton, CT 06340*

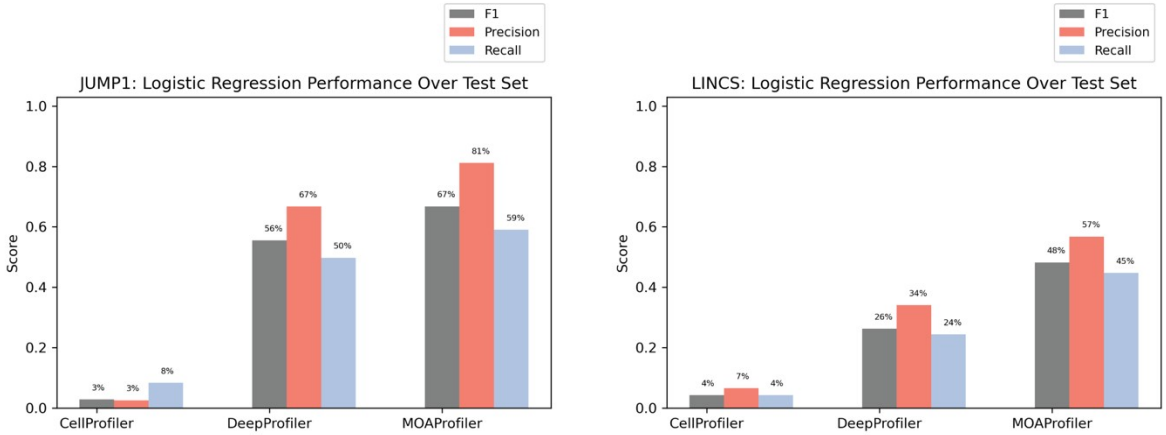


Supplemental Figure 1: Investigating potential experimental biases.

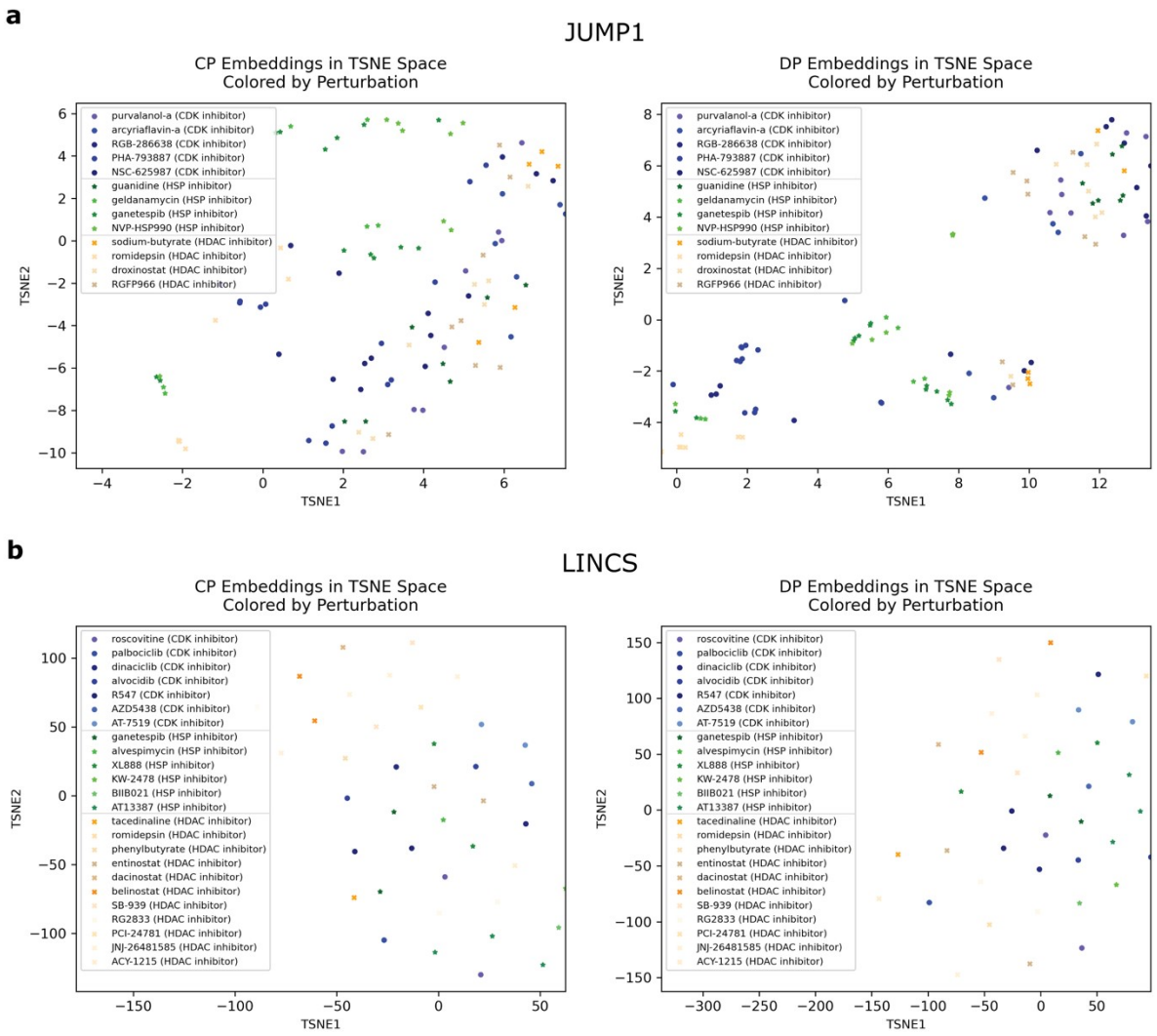
a) Well location effects for JUMP1 Dataset. Left: 384-well plate heatmap (over 23 plates) of model classification accuracies over image fields. An edge well is a well with rows “A” or “P” or columns 1 or 24. “D” indicates negative control DMSO wells (no perturbation). “N/A” indicates wells excluded in our study due to compounds with more than one known MOA, or MOAs with just

one compound. Right: Average classification accuracy over all image fields (not embeddings) residing in edge and non-edge wells.

b) Model classification accuracy by timepoint. Y-axis: image classification accuracy, x-axis: timepoint.
c) Model classification accuracy by cell type. Y-axis: image classification accuracy, x-axis: cell type.
d) Well location effects for LINCS Dataset. Same analysis as in (a) except the LINCS Dataset was over 136 384-well plates. “N/A” indicates wells excluded in our study due to compounds with more than one known MOA, concentration not at 10μM, or MOAs with just one compound.

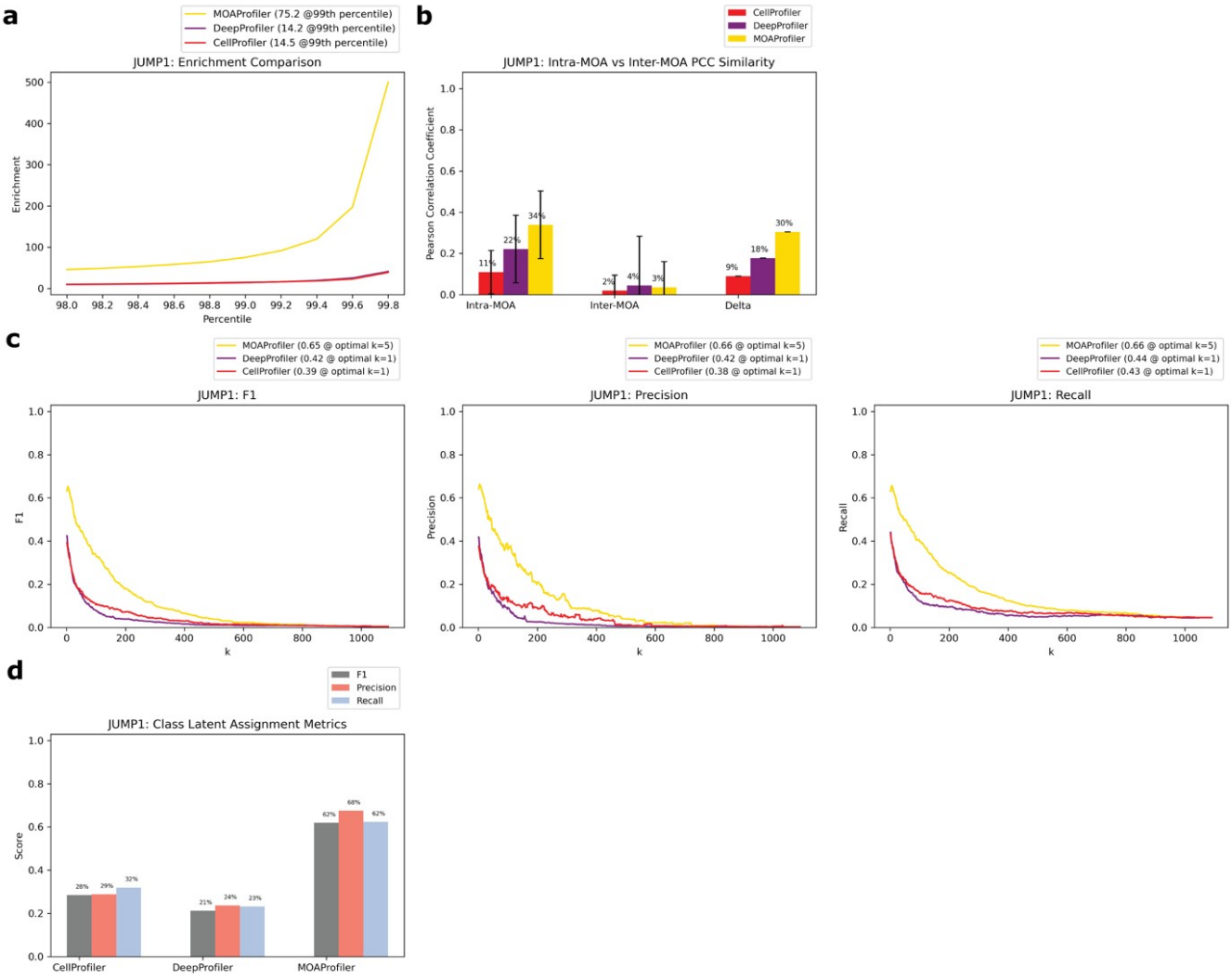


Supplemental Figure 2: Logistic regression performance over the test set. We trained separate logistic regression models for each of CP, DP, and MP. We trained the logistic regression models on the training set embeddings with the task of classifying MOA and assessed the logistic regression models on the test set embeddings (Methods). Left: JUMP1, right: LINCS.



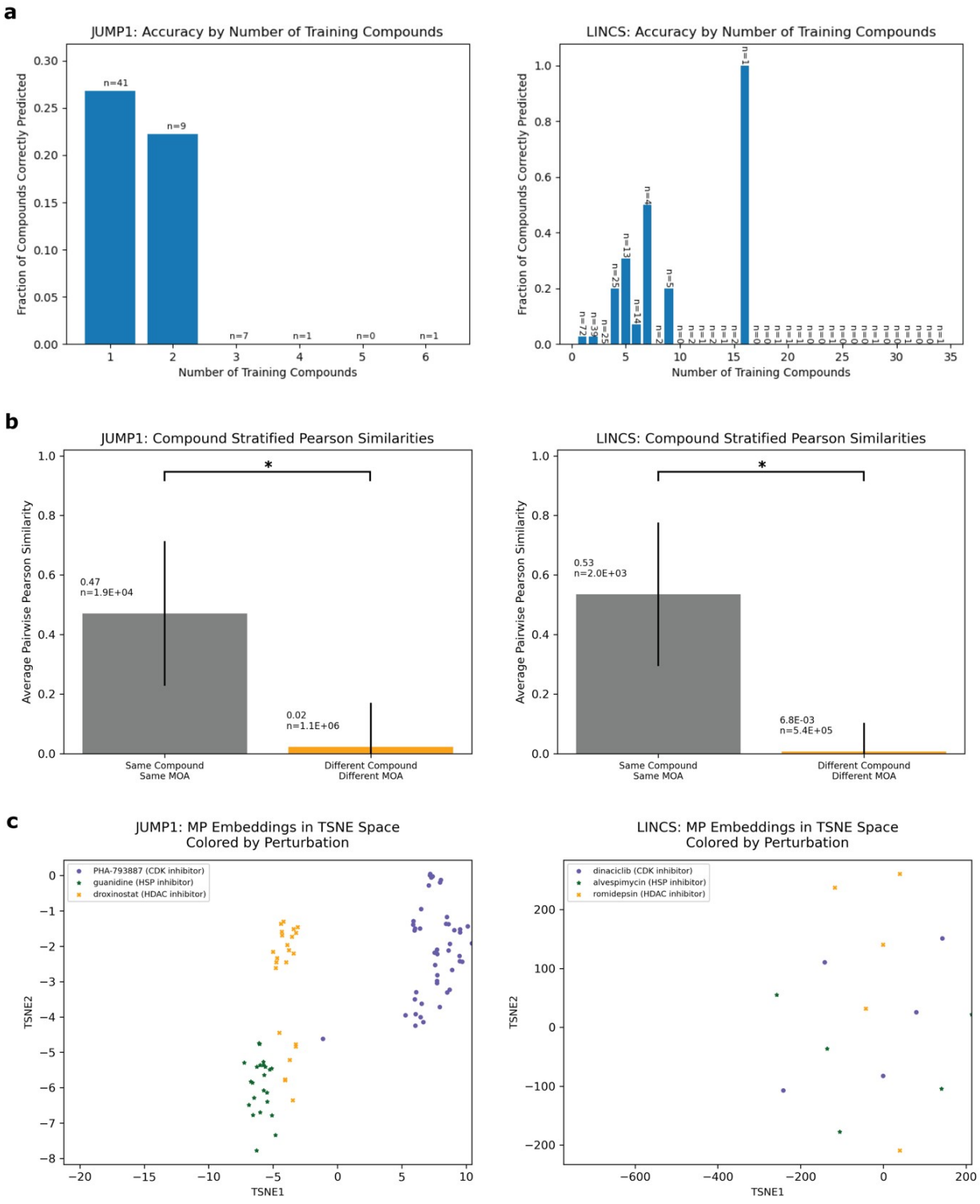
Supplemental Figure 3: CP (left) and DP (right) embeddings in TSNE latent space for example MOAs.

a) JUMP1.
b) LINCS.



Supplemental Figure 4: Comparing heavily-optimized JUMP1 CP embeddings to DMSO-standardized DP and MP embeddings.

a) Enrichment comparison at different percentiles.
b) Intra-MOA vs inter-MOA average pairwise PCC similarity.
c) K-NN embedding metrics.
d) Class latent assignment metrics.



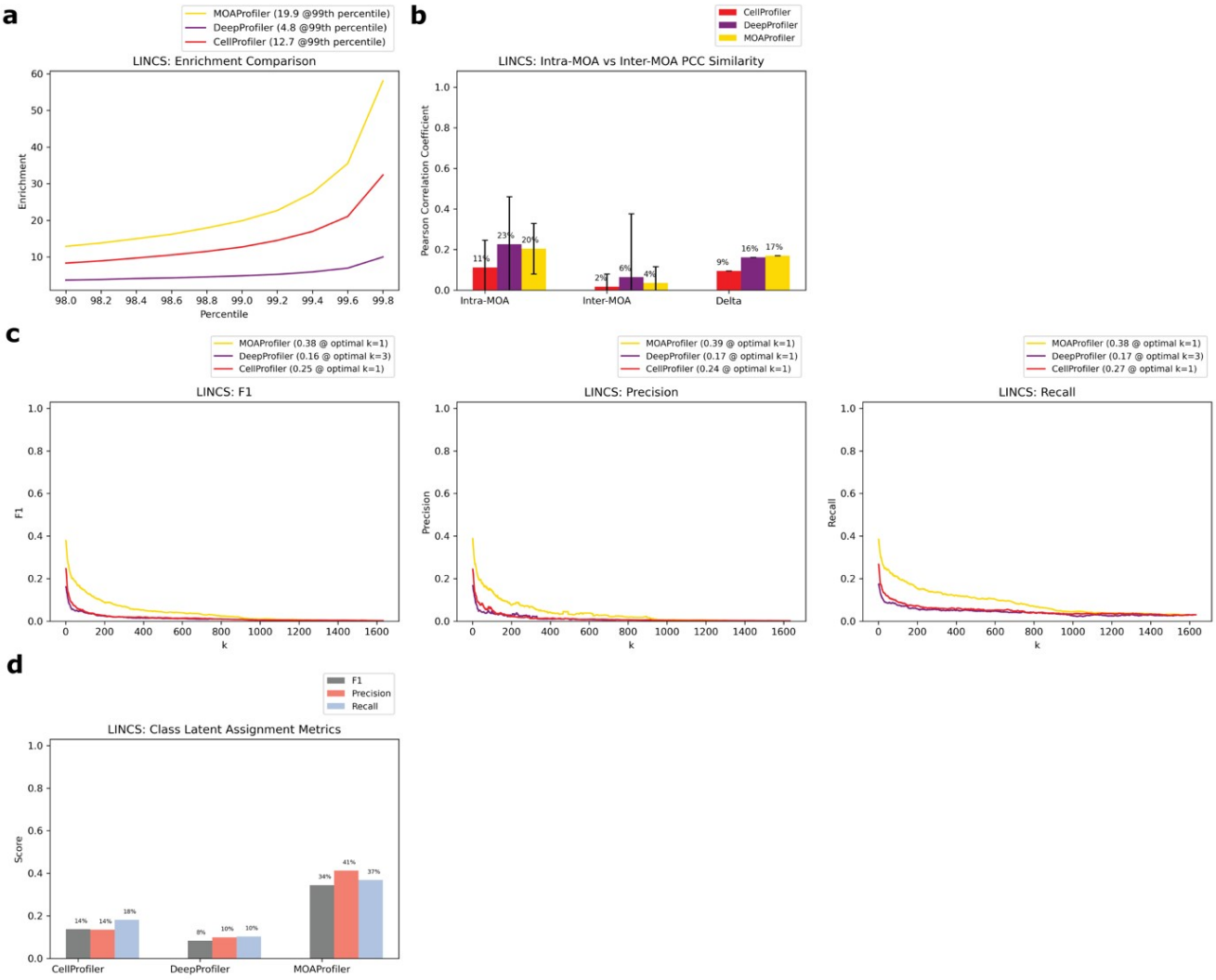
Supplemental Figure 5: Compound-holdout MOA specificity and embedding similarity.

a) Held-out compound accuracy for MP vs number of compounds in the training set for each MOA. X-axis: number of training compounds, y-axis: fraction of held-out compounds with correct MOA prediction, n=number of held-out compounds. Left: JUMP1, right: LINCS. E.g. left panel: 41 held-

out compounds had just one other compound in the training set with the same MOA. Of the 41 held-out compounds, 27% of them had their MOA correctly classified.

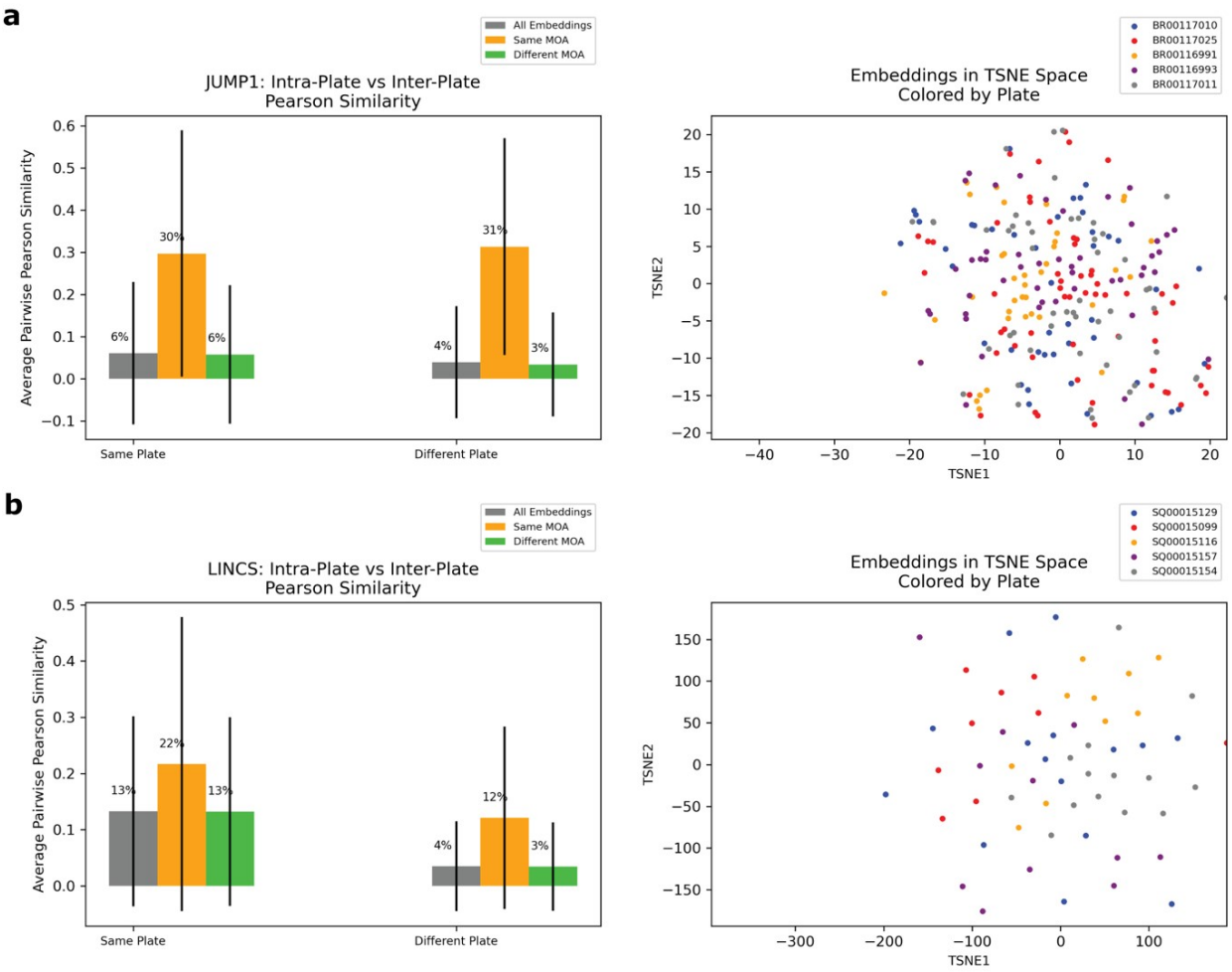
b) Average of pairwise PCCs for two groups of embeddings: same-compound same-MOA and different-compound different-MOA. For each group, we calculated PCCs for each possible pair of wells. Significance (*) indicates $p \ll 0.0001$ for a two-sided z test. Error bars span one standard deviation in each direction. Left: JUMP1, right: LINCS.

c) TSNE visualization of well embeddings of three example MOAs. Circles = CDK inhibitor, stars = HSP inhibitor, x marks = HDAC inhibitor. Different compounds with the same MOA were given similar but different colors.



Supplemental Figure 6: Comparing heavily-optimized LINCS CP embeddings to DMSO-standardized DP and MP embeddings.

a) Enrichment comparison at different percentiles.
b) Intra-MOA vs inter-MOA average pairwise PCC similarity.
c) K-NN embedding metrics.
d) Class latent assignment metrics.

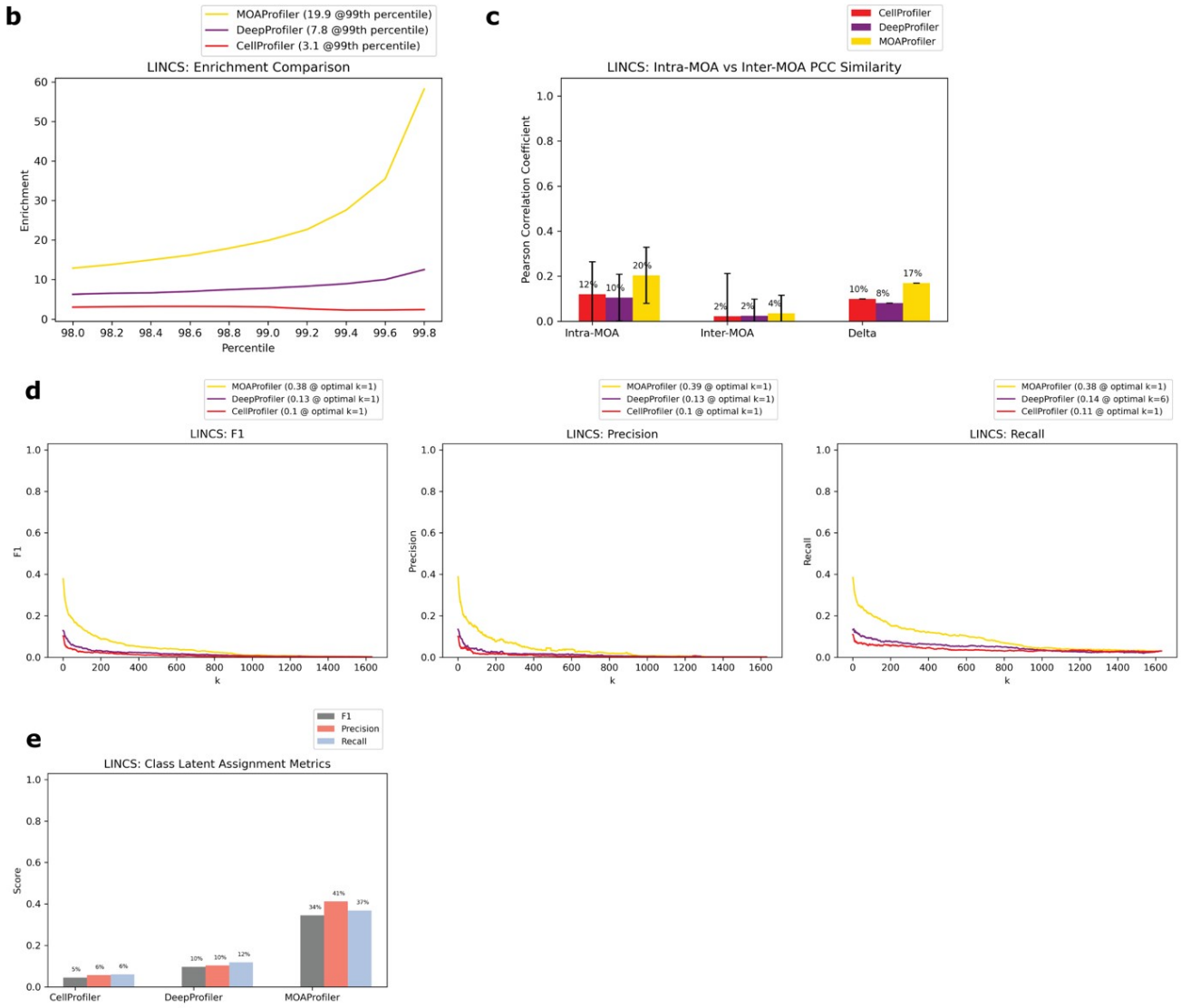


Supplemental Figure 7: Quantifying batch effects on MP-derived embeddings. Left: average pairwise PCCs over the test set for three groups: 1) embeddings from different plates (grey), 2) embeddings from different plates with the same MOA (orange), and 3) embeddings from different plates with different MOAs (green). Error bars span one standard deviation in each direction. Right: TSNE visualization of all test-set MP embeddings from five randomly chosen plates. Embeddings colored by plate.

a) JUMP1.
b) LINCS.

a

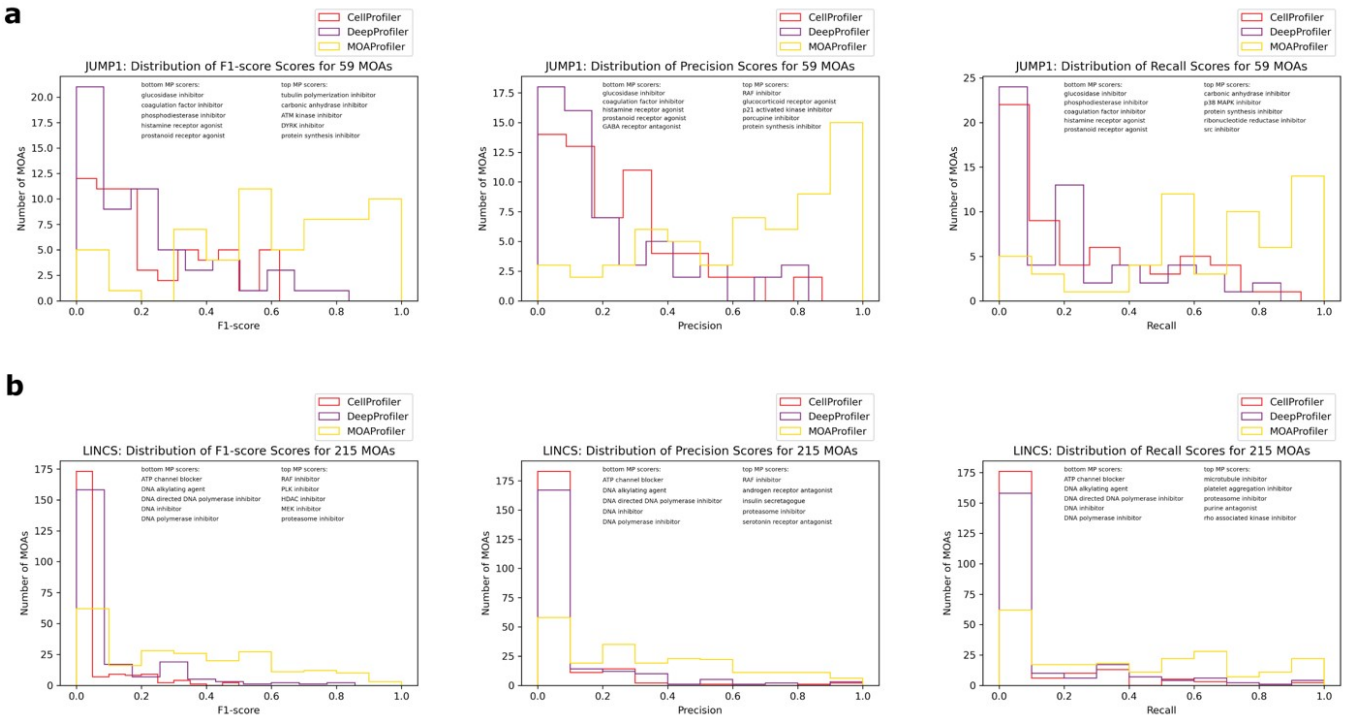
MOA Profiler	Deep Profiler
Does not require single cell locations	Requires single cell locations
Training classification at field level	Training classification at single-cell level
gamma decay learning rate scheduler	cosine learning rate scheduler
Strong supervision: classify MOAs, extract embeddings for MOA determination	Weak supervision: classify compounds, extract embeddings for MOA determination
No label smoothing	Label smoothing



Supplemental Figure 8: Comparing MP to DP trained on the LINCS dataset. Otherwise, embedding analyses was the same as Figure 5.

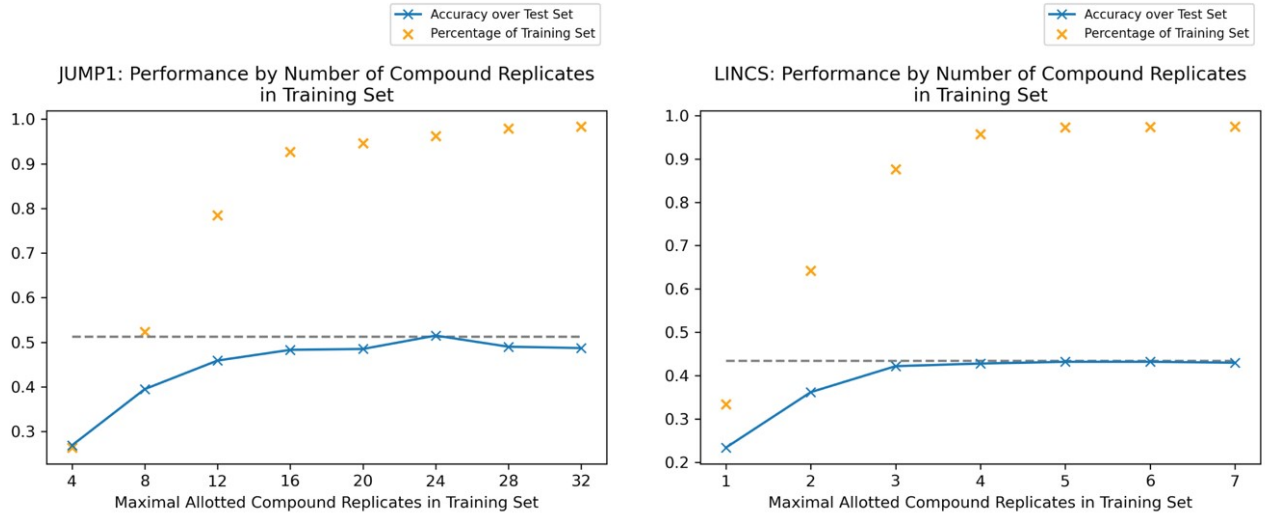
a) Key differences between MP and DP methods for MOA determination.
b) Enrichment comparison at different percentiles.
c) Intra-MOA vs Inter-MOA average PCC similarity.
d) k-NN embedding metrics.

e) Class latent assignment metrics.



Supplemental Figure 9: Distribution of class latent assignment scores over the test set for each MOA. X-axis: class latent assignment score, y-axis: count of MOAs achieving the score. Listed: worst and best performing MOAs by score for MP.

a) JUMP1.
b) LINCS. X-axis: class latent assignment score, y-axis: count of MOAs achieving the score. Listed: worst and best performing MOAs by score for MP.



Supplemental Figure 10: Performance with smaller training sets. JUMP (left), LINCS (right). Keeping the validation and test sets the same, we systematically limited the number of compound-replicate wells present in the training set and trained new models from these smaller sets (Methods). X-axis: maximal number of compound-replicate wells allowed in the training set, blue: accuracy across held-out test set, orange: percentage of images compared to the full training set. Dashed horizontal line indicates the performance over the test set when keeping the full training set.

Supplemental Table 1: JUMP1 Compound holdout by MOA. “0” indicates wrong prediction, “1” indicates correct prediction.

MOA	CP	DP	MP	# of training compounds
calcium-activated potassium channel activator	0	0	0	1
bacterial cell wall synthesis inhibitor	0	0	0	1
benzodiazepine receptor agonist	0	0	0	1
tubulin polymerization inhibitor	1	1	1	1
carbonic anhydrase inhibitor	1	0	1	1
coagulation factor inhibitor	0	0	0	1
endothelin receptor antagonist	0	0	1	1
glucocorticoid receptor agonist	1	1	1	1
glucosidase inhibitor	0	0	0	1
histamine receptor agonist	0	0	0	1
aldose reductase inhibitor	0	0	0	1
histamine receptor antagonist	0	0	0	1
matrix metalloprotease inhibitor	1	1	0	1
p21 activated kinase inhibitor	0	0	0	1
p38 MAPK inhibitor	0	1	1	1
porcupine inhibitor	0	0	0	1
prostanoid receptor agonist	0	0	0	1
purinergic receptor antagonist	1	0	1	1

ribonucleotide reductase inhibitor	1	1	1	1
smoothened receptor antagonist	0	0	0	1
somatostatin receptor agonist	0	0	0	1
lysophosphatidic acid receptor antagonist	0	0	0	1
alcohol dehydrogenase inhibitor	0	0	0	1
ubiquitin specific protease inhibitor	0	0	0	1
acetylcholine receptor antagonist	0	1	1	1
DNA synthesis inhibitor	0	0	0	1
HMGCR inhibitor	1	0	1	1
ICAM1 expression inhibitor	0	0	0	1
IGF-1 inhibitor	0	0	0	1
Bruton's tyrosine kinase (BTK) inhibitor	0	0	0	1
DYRK inhibitor	0	0	0	1
Aurora kinase inhibitor	0	1	0	1
JAK inhibitor	0	0	0	1
PPAR receptor antagonist	1	0	1	1
RAF inhibitor	0	0	0	1
ATM kinase inhibitor	0	0	0	1
RAGE receptor antagonist	0	0	0	1
SIRT inhibitor	0	0	0	1
acetylcholine receptor agonist	0	1	1	1
JNK inhibitor	0	0	0	1
G protein signaling inhibitor	0	0	0	1
CDC inhibitor	0	1	1	2
protein synthesis inhibitor	0	0	0	2

potassium channel blocker	0	0	0	2
ALK tyrosine kinase receptor inhibitor	1	1	1	2
adrenergic receptor antagonist	0	0	0	2
histone lysine methyltransferase inhibitor	0	0	0	2
glutamate receptor antagonist	0	0	0	2
free fatty acid receptor agonist	0	0	0	2
GABA receptor antagonist	0	0	0	2
phosphodiesterase inhibitor	0	0	0	3
HSP inhibitor	0	0	0	3
cyclooxygenase inhibitor	0	0	0	3
sodium channel blocker	0	0	0	3
sphingosine 1 phosphate receptor agonist	0	0	0	3
src inhibitor	1	1	0	3
HDAC inhibitor	0	0	0	3
CDK inhibitor	0	0	0	4
calcium channel blocker	0	0	0	6

Supplemental Table 2: LINCS compound holdout by MOA. “0” indicates wrong prediction, “1” indicates correct prediction.

MOA	CP	DP	MP	# of training compounds
elastase inhibitor	0	0	0	1
aldose reductase inhibitor	0	0	0	1
alcohol dehydrogenase inhibitor	0	0	0	1
hypoxia inducible factor inhibitor	0	0	0	1
immunostimulant	0	0	0	1
adrenergic inhibitor	0	0	0	1
adenylyl cyclase activator	0	0	0	1
immunosuppressant	0	0	0	1
inosine monophosphate dehydrogenase inhibitor	0	0	1	1
acetylcholine release enhancer	0	0	0	1
isocitrate dehydrogenase inhibitor	0	0	0	1
leukotriene synthesis inhibitor	0	0	0	1
lipase inhibitor	0	0	0	1
lipid peroxidase inhibitor	0	0	0	1
TRPV antagonist	0	0	0	1
cathepsin inhibitor	0	0	0	1
RNA synthesis inhibitor	0	0	0	1
membrane integrity inhibitor	1	0	0	1
histamine receptor modulator	0	0	0	1
membrane permeability inhibitor	0	0	0	1
hepatocyte growth factor receptor inhibitor	0	0	0	1
growth hormone secretagogue receptor agonist	0	0	0	1
carnitine palmitoyltransferase inhibitor	0	0	0	1
cannabinoid receptor inverse agonist	0	0	0	1
cannabinoid receptor antagonist	0	0	0	1
calmodulin antagonist	0	0	0	1
estrogen receptor antagonist	0	0	0	1
c-Met inhibitor	0	0	0	1
beta-secretase inhibitor	0	0	0	1
beta lactamase inhibitor	0	0	0	1
fatty acid synthase inhibitor	1	0	0	1
focal adhesion kinase inhibitor	0	0	0	1

bacterial antifolate	0	0	0	1
gap junction modulator	0	0	0	1
gonadotropin releasing factor hormone receptor agonist	0	0	0	1
antithyroid agent	0	0	0	1
antioxidant	0	0	0	1
antimalarial agent	0	0	0	1
anthelmintic agent	0	0	0	1
guanylate cyclase stimulant	0	0	0	1
Pim kinase inhibitor	0	0	0	1
cholesteryl ester transfer protein inhibitor	0	0	0	1
sigma receptor agonist	0	0	0	1
Bcr-Abl kinase inhibitor	0	0	0	1
tricyclic antidepressant	0	0	0	1
tryptophan hydroxylase inhibitor	0	0	0	1
CFTR channel agonist	0	0	0	1
tumor necrosis factor production inhibitor	0	0	0	1
sulfonylurea	0	0	0	1
tumor necrosis factor release inhibitor	0	0	0	1
serotoninV¢¬Ä¬ìnorepinephrine reuptake inhibitor (SNRI)	0	0	0	1
serotonin transporter (SERT) inhibitor	0	0	0	1
GABA receptor antagonist	0	0	0	1
serotonin reuptake inhibitor	0	0	0	1
renin inhibitor	0	0	0	1
purinergic receptor antagonist	0	0	0	1
protein tyrosine kinase inhibitor	0	0	0	1
purine antagonist	0	0	1	1
progesterone receptor antagonist	0	0	0	1
mucolytic agent	0	0	0	1
neprilysin inhibitor	0	0	0	1
neurotrophic agent	0	0	0	1
P glycoprotein inhibitor	0	0	0	1
O6-alkylguanine-DNA alkyltransferase inhibitor	0	0	0	1
tyrosinase inhibitor	0	0	0	1
nitric oxide stimulant	0	0	0	1
nitric oxide donor	0	0	0	1
L-type calcium channel blocker	0	0	0	1
JNK inhibitor	0	0	0	1

vesicular monoamine transporter inhibitor	0	0	0	1
platelet aggregation inhibitor	0	0	0	1
thyrotropin releasing hormone receptor agonist	0	0	0	1
dihydroorotate dehydrogenase inhibitor	0	0	0	2
diuretic	0	0	0	2
voltage-gated sodium channel blocker	0	0	1	2
vitamin K antagonist	0	0	0	2
farnesyltransferase inhibitor	0	0	0	2
diacylglycerol O acyltransferase inhibitor	0	0	0	2
mineralocorticoid receptor antagonist	0	0	0	2
thromboxane synthase inhibitor	0	0	0	2
thromboxane receptor antagonist	0	0	0	2
glucose dependent insulinotropic receptor agonist	0	0	0	2
glycine transporter inhibitor	0	0	0	2
thrombin inhibitor	0	0	0	2
sodium/potassium/chloride transporter inhibitor	0	0	0	2
sodium/glucose cotransporter inhibitor	0	0	0	2
microtubule inhibitor	0	0	0	2
histamine receptor agonist	0	0	0	2
protein synthesis stimulant	0	0	0	2
insulin sensitizer	0	0	0	2
platelet activating factor receptor antagonist	0	0	0	2
lipoxygenase inhibitor	0	0	0	2
thymidylate synthase inhibitor	0	0	0	2
matrix metalloprotease inhibitor	0	0	0	2
5 alpha reductase inhibitor	0	0	0	2
benzodiazepine receptor antagonist	0	0	0	2
adenosine receptor agonist	0	0	0	2
calcineurin inhibitor	0	0	0	2
T-type calcium channel blocker	0	0	0	2
TGF beta receptor inhibitor	0	0	0	2
calcitonin antagonist	0	0	0	2

DNA methyltransferase inhibitor	0	0	0	2
DNA directed DNA polymerase inhibitor	0	0	0	2
XIAP inhibitor	0	0	0	2
cannabinoid receptor agonist	0	0	0	2
BCL inhibitor	0	0	0	2
ATP channel blocker	0	0	0	2
acetylcholinesterase inhibitor	0	0	0	2
ACAT inhibitor	0	0	0	2
Bruton's tyrosine kinase (BTK) inhibitor	0	0	0	2
MDM inhibitor	0	0	0	2
opioid receptor agonist	0	0	0	3
VEGFR inhibitor	0	0	0	3
leukotriene receptor antagonist	0	0	0	3
cholesterol inhibitor	0	0	0	3
NFkB pathway inhibitor	0	0	0	3
chelating agent	0	0	0	3
potassium channel activator	0	0	0	3
prostanoid receptor agonist	0	0	0	3
prostanoid receptor antagonist	0	0	0	3
HIV integrase inhibitor	0	0	0	3
rho associated kinase inhibitor	0	0	0	3
GABA receptor modulator	0	0	0	3
sigma receptor antagonist	0	0	0	3
ALK tyrosine kinase receptor inhibitor	0	0	0	3
vasopressin receptor antagonist	0	0	0	3
opioid receptor antagonist	0	0	0	3
angiogenesis inhibitor	0	0	0	3
xanthine oxidase inhibitor	0	0	0	3
free radical scavenger	0	0	0	3
fungal squalene epoxidase inhibitor	0	0	0	3
dihydrofolate reductase inhibitor	0	0	0	3
endothelin receptor antagonist	0	0	0	3
glutamate receptor agonist	0	0	0	3
glutamate receptor modulator	0	0	0	3
bacterial 30S ribosomal subunit inhibitor	0	0	0	3
PLK inhibitor	0	1	1	4
DNA alkylating agent	0	0	0	4
PKC inhibitor	0	0	0	4
DNA polymerase inhibitor	0	0	0	4

dopamine reuptake inhibitor	0	0	0	4
nitric oxide synthase inhibitor	0	0	0	4
non-nucleoside reverse transcriptase inhibitor	0	0	0	4
selective serotonin reuptake inhibitor (SSRI)	0	0	0	4
ribonucleotide reductase inhibitor	0	0	0	4
protein synthesis inhibitor	0	0	1	4
PPAR receptor agonist	0	0	0	4
CCK receptor antagonist	0	0	0	4
PARP inhibitor	0	0	1	4
glucosidase inhibitor	0	0	0	4
coagulation factor inhibitor	0	0	0	4
tyrosine kinase inhibitor	0	0	0	4
adenosine receptor antagonist	0	0	0	4
cytochrome P450 inhibitor	0	0	0	4
proteasome inhibitor	0	0	1	4
aromatase inhibitor	0	0	0	4
insulin secretagogue	0	0	0	4
integrin antagonist	0	0	0	4
carbonic anhydrase inhibitor	0	0	0	4
ATPase inhibitor	0	0	1	4
dipeptidyl peptidase inhibitor	0	0	0	4
tachykinin antagonist	0	0	0	5
AKT inhibitor	0	0	1	5
HMGCR inhibitor	0	0	1	5
vitamin D receptor agonist	0	0	1	5
Aurora kinase inhibitor	0	0	1	5
DNA synthesis inhibitor	0	0	0	5
androgen receptor modulator	0	0	0	5
benzodiazepine receptor agonist	0	0	0	5
angiotensin receptor antagonist	0	0	0	5
RAF inhibitor	0	0	0	5
monoamine oxidase inhibitor	0	0	0	5
bacterial 50S ribosomal subunit inhibitor	0	0	0	5
nucleoside reverse transcriptase inhibitor	0	0	0	5
androgen receptor antagonist	0	0	0	6
glutamate receptor antagonist	0	0	0	6
RNA polymerase inhibitor	0	0	0	6
CC chemokine receptor antagonist	1	0	0	6
sterol demethylase inhibitor	0	0	0	6

potassium channel blocker	0	0	0	6
progesterone receptor agonist	0	0	0	6
retinoid receptor agonist	0	0	0	6
HSP inhibitor	0	0	1	6
JAK inhibitor	0	0	0	6
p38 MAPK inhibitor	0	0	0	6
gamma secretase inhibitor	0	0	0	6
estrogen receptor agonist	0	0	0	6
HIV protease inhibitor	0	0	0	6
CDK inhibitor	0	0	1	7
mTOR inhibitor	0	0	1	7
DNA inhibitor	0	0	0	7
PI3K inhibitor	0	0	0	7
acetylcholine receptor agonist	0	0	0	8
HCV inhibitor	0	0	0	8
tubulin polymerization inhibitor	0	0	0	9
angiotensin converting enzyme inhibitor	0	0	0	9
bacterial DNA gyrase inhibitor	0	0	0	9
serotonin receptor agonist	0	0	0	9
MEK inhibitor	0	0	1	9
HDAC inhibitor	0	0	0	11
dopamine receptor agonist	0	0	0	11
sodium channel blocker	0	0	0	12
bacterial cell wall synthesis inhibitor	0	0	0	13
serotonin receptor antagonist	0	0	0	13
EGFR inhibitor	0	0	0	14
topoisomerase inhibitor	0	0	0	15
calcium channel blocker	0	0	0	15
glucocorticoid receptor agonist	1	1	1	16
acetylcholine receptor antagonist	0	0	0	19
adrenergic receptor agonist	0	0	0	20
dopamine receptor antagonist	0	0	0	21
histamine receptor antagonist	0	0	0	24
cyclooxygenase inhibitor	0	0	0	28
adrenergic receptor antagonist	0	0	0	31
phosphodiesterase inhibitor	0	0	0	34