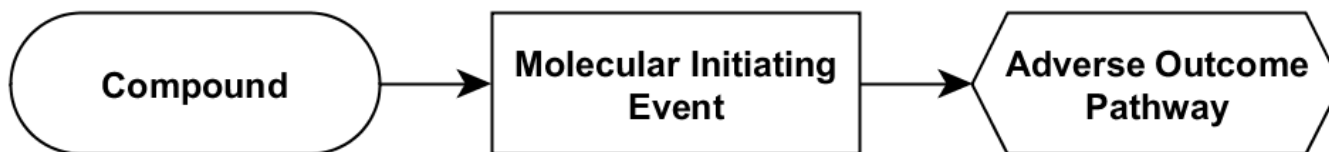


AI, Machine Learning and Big Data in Predictive Toxicology



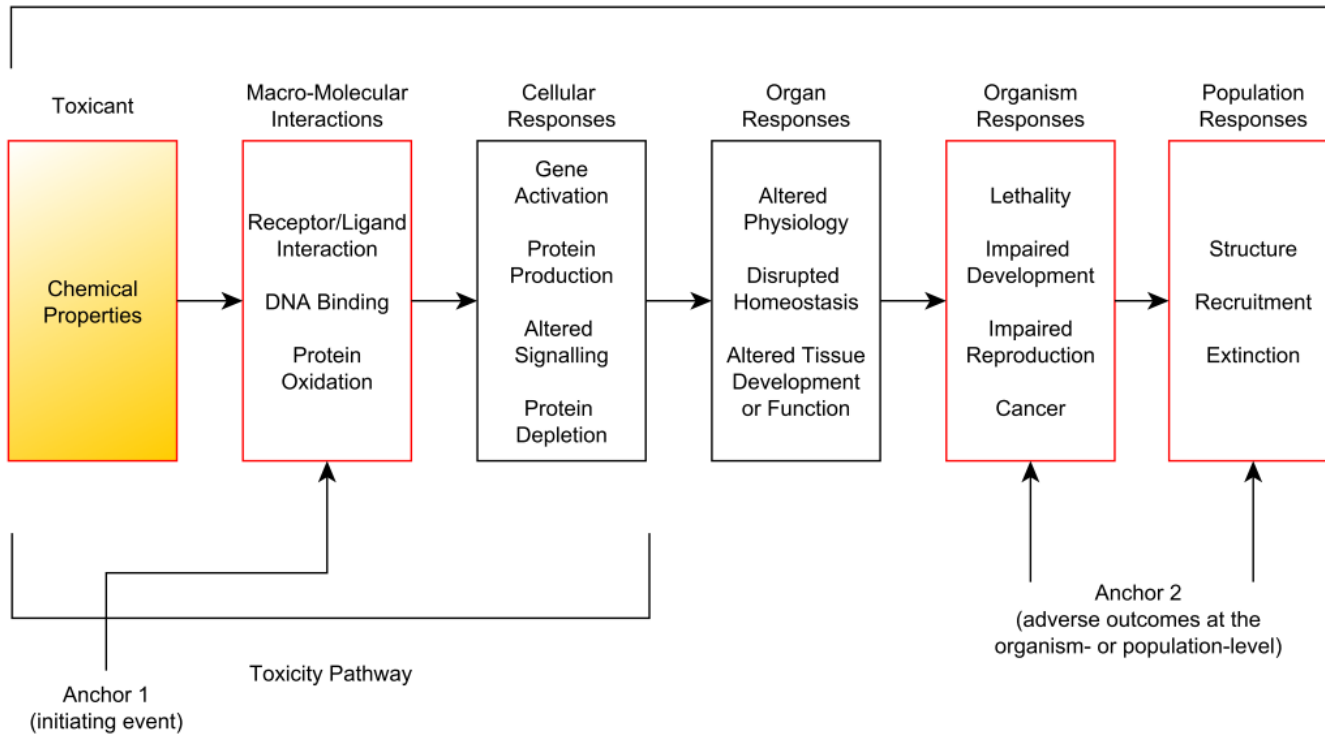
Timothy E H Allen

AJ Wedlake, M Folia, S Piechota, E Gelžinytė,
JM Goodman, S Gutsell, P Kukic, PJ Russell

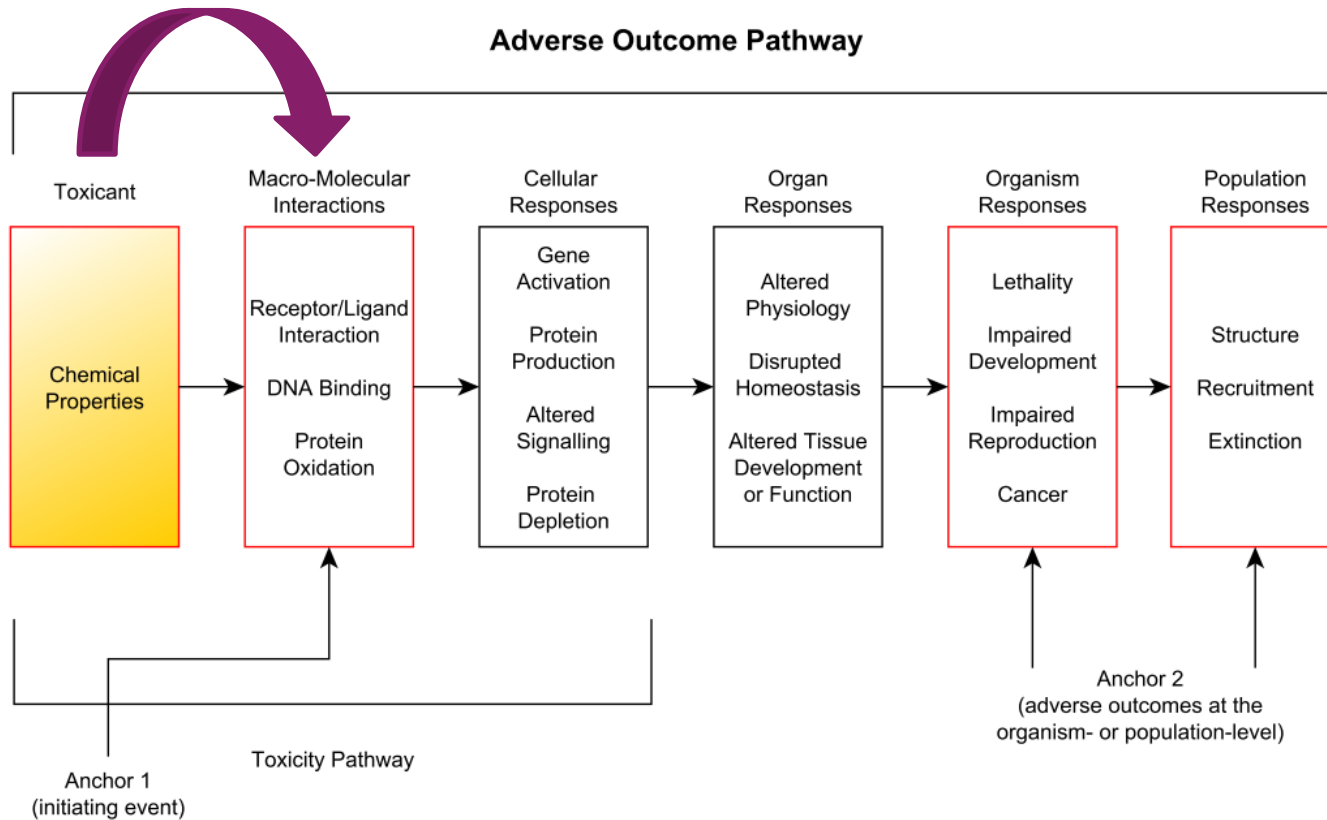
11th March 2020

Adverse Outcome Pathway

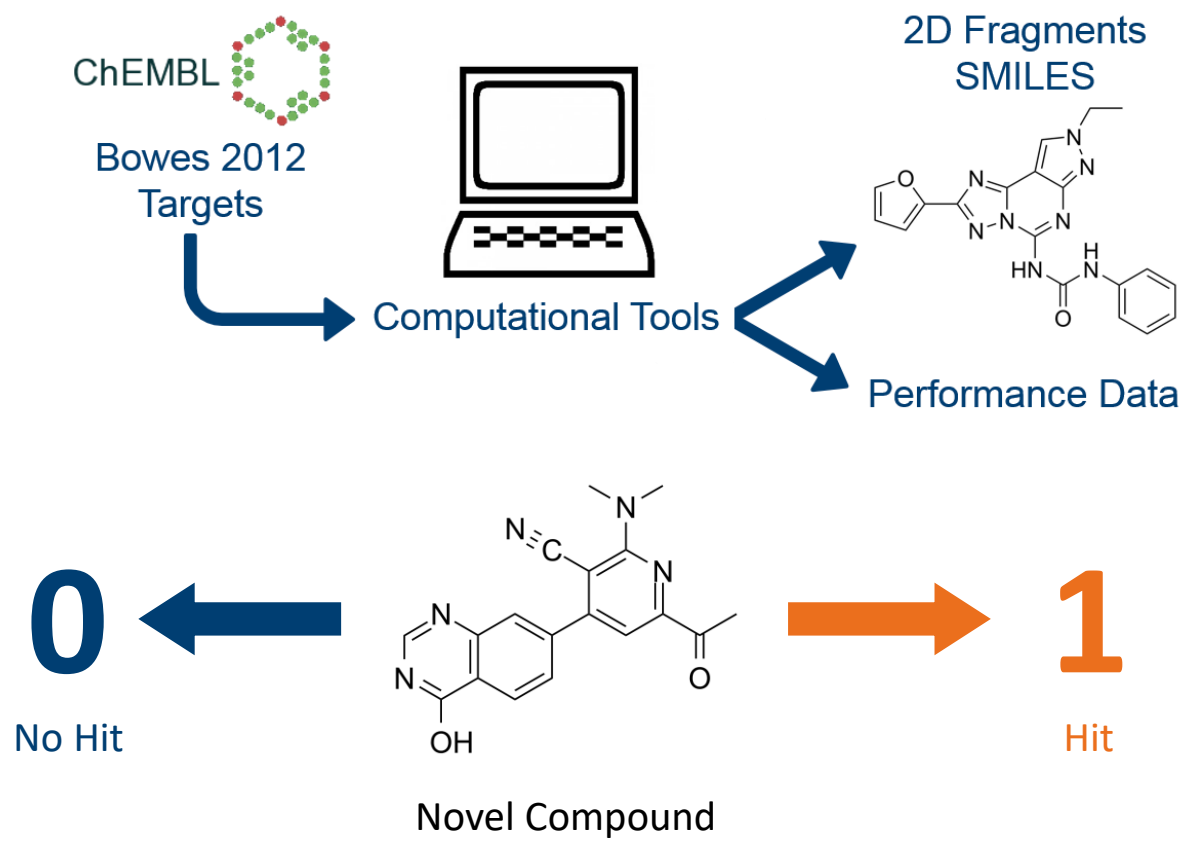
Adverse Outcome Pathway



Adverse Outcome Pathway



Structural Alerts

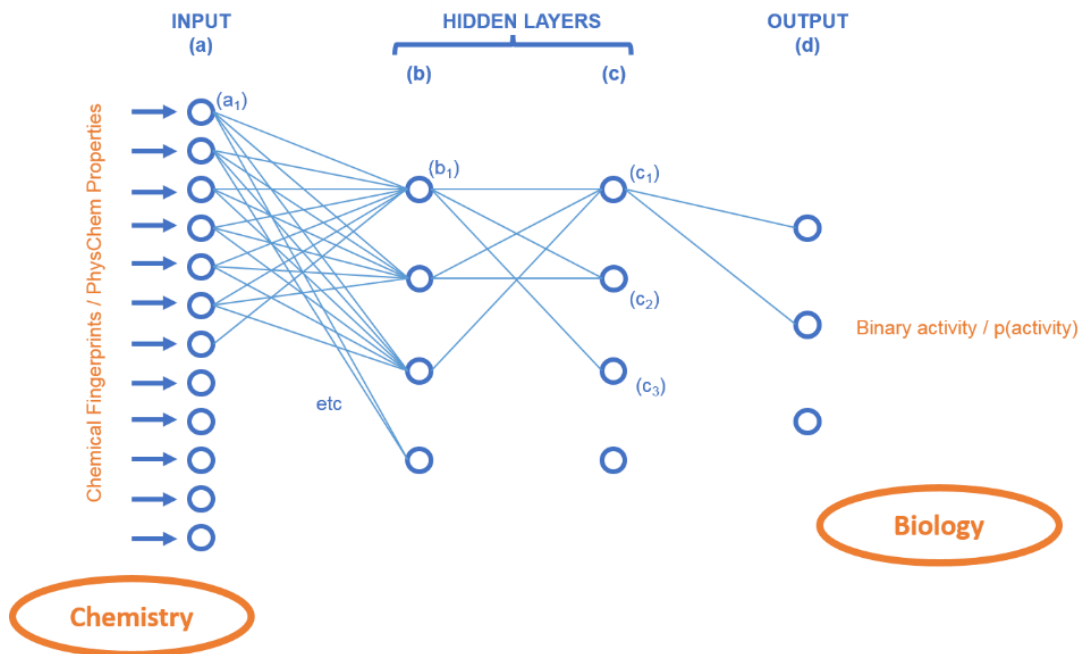


Allen, T.E.H. *et al.* (2018) *Toxicol. Sci.*, 165; 213.
Wedlake, A.J. *et al.* (2019) *Chem. Res. Toxicol.*, 33; 388.

Expanded Target List

Acetylcholinesterase	Serotonin 3a (5-HT3a) receptor	Vascular endothelial growth factor receptor 1	Serine/threonine-protein kinase PAK 4
Adenosine A2a receptor	Serotonin transporter	Vascular endothelial growth factor receptor 3	Phosphodiesterase 4A
Alpha-2a adrenergic receptor	Tyrosine-protein kinase LCK	Tyrosine-protein kinase FYN	Phosphodiesterase 5A
Androgen receptor	Vasopressin V1a receptor	Glycogen synthase kinase-3 beta	PIP ₂ 3-kinase catalytic subunit α
Beta-1 adrenergic receptor	Type-1 angiotensin II receptor	Histone deacetylase 3	Peroxisome proliferator-activated receptor γ
Beta-2 adrenergic receptor	RAC-alpha serine/threonine-protein kinase	Insulin-like growth factor 1 receptor	Protein Tyr phosphatase non-receptor type 1
Delta opioid receptor	Beta-secretase 1	Insulin receptor	Protein Tyr phosphatase non-receptor type 11
Dopamine D1 receptor	Cholinesterase	Vascular endothelial growth factor receptor 2	Protein Tyr phosphatase non-receptor type 2
Dopamine D2 receptor	Caspase-1	Leukotriene B4 receptor 1	RAF proto-oncogene Ser/Thr-protein kinase
Dopamine transporter	Caspase-3	Tyrosine-protein kinase Lyn	Retinoic acid receptor alpha
Endothelin receptor ET-A	Caspase-8	Mitogen-activated protein kinase 1	Retinoic acid receptor beta
Glucocorticoid receptor	Muscarinic acetylcholine receptor M5	Mitogen-activated protein kinase 9	Rho-ass. coiled-coil-containing protein kinase I
hERG	Inhibitor of nuclear factor κ -B kinase subunit α	MAP kinase-activated protein kinase 2	Ribosomal protein S6 kinase alpha-5
Histamine H1 receptor	Macrophage colony-stimulating fac. 1 receptor	Hepatocyte growth factor receptor	NAD-dependent protein deacetylase sirtuin-2
Mu opioid receptor	Casein kinase I isoform delta	Matrix metalloproteinase-13	NAD-dependent protein deacetylase sirtuin-3
Muscarinic acetylcholine receptor M1	Endothelin B receptor	Matrix metalloproteinase-2	Proto-oncogene tyrosine-protein kinase Src
Muscarinic acetylcholine receptor M2	Neutrophil elastase	Matrix metalloproteinase-3	Substance-K receptor
Muscarinic acetylcholine receptor M3	Ephrin type-A receptor 2	Matrix metalloproteinase-9	Thromboxane A2 receptor
Norepinephrine transporter	Fibroblast growth factor receptor 1	Serine/threonine-protein kinase NEK2	Tyrosine-protein kinase receptor TEK
Serotonin 2a (5-HT2a) receptor	Peptidyl-prolyl cis-trans isomerase	P2Y purinoceptor 1	

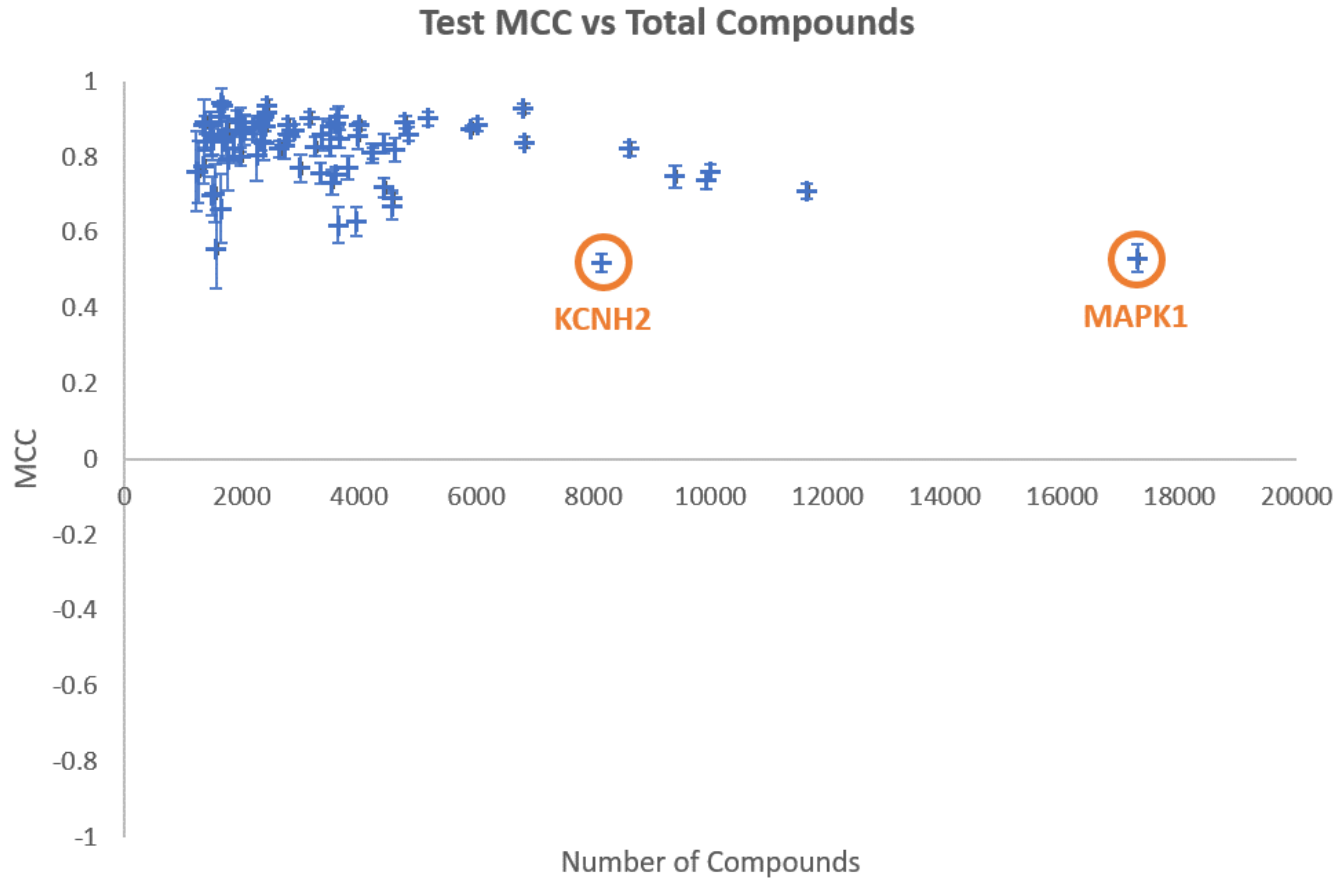
Neural Networks



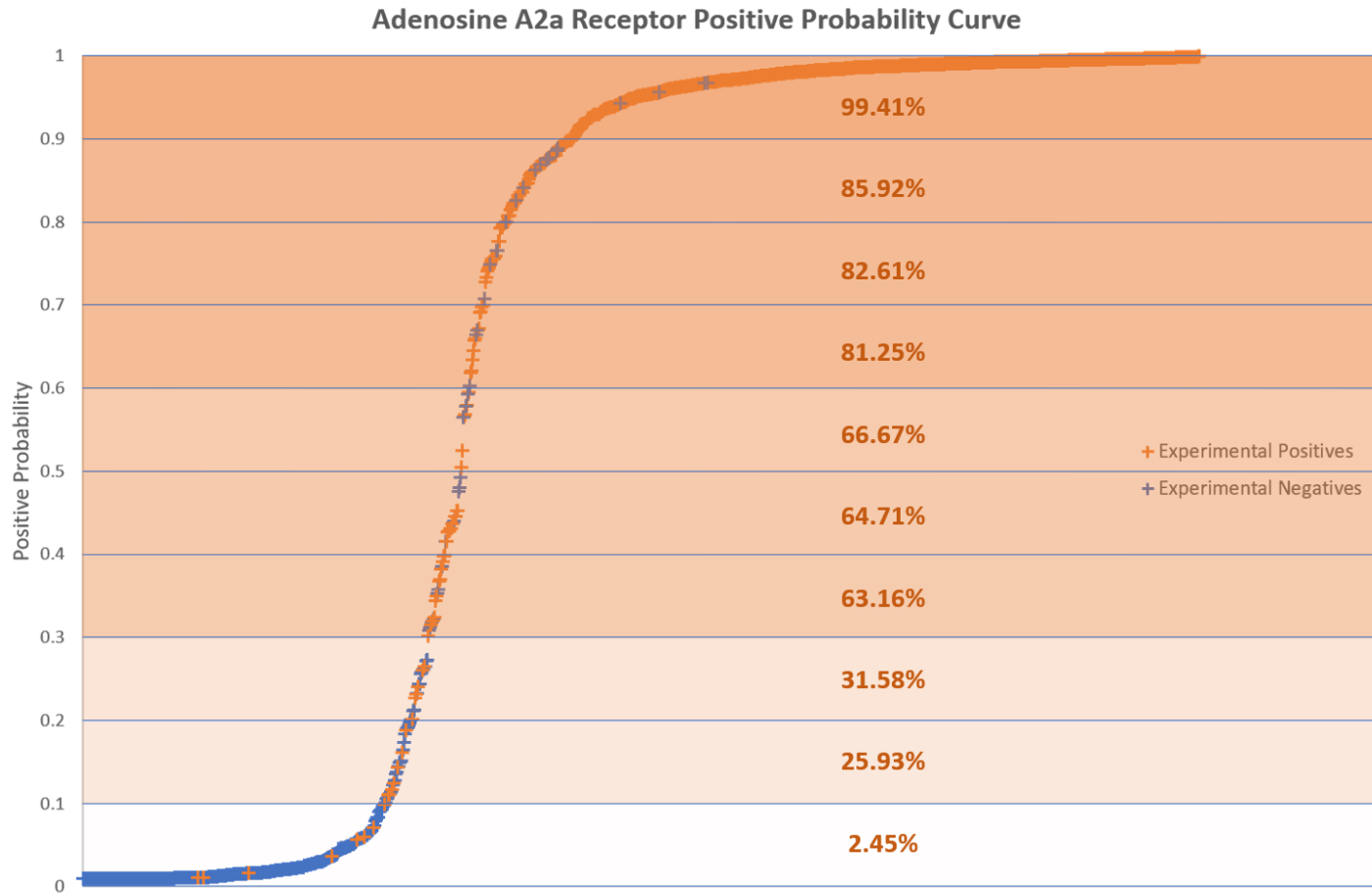
Average Model Performance

	Training Data					Validation Data					Test Data				
	SE	SP	ACC	MCC	ROC-AUC	SE	SP	ACC	MCC	ROC-AUC	SE	SP	ACC	MCC	ROC-AUC
AVERAGE	92.1	96.5	95.8	0.901	0.99	86.9	93.2	92.5	0.822	0.96	86.2	92.9	92.2	0.814	0.96
SD	8.8	4.2	3.1	0.069	0.02	11.7	5.9	4.1	0.091	0.04	12.1	6.5	4.2	0.093	0.04

MCC vs Total Compounds



Positive Probability Curve



Comparative Model Performance

Target Gene	SE	Δ SA	Δ RF	SP	Δ SA	Δ RF	ACC	Δ SA	Δ RF	MCC	Δ SA	Δ RF
AChE	88.4	6.6	-3.2	85.3	-4.6	7.5	87.0	1.6	1.6	0.737	0.024	0.030
ADORA2A	97.6	2.8	-0.5	93.2	2.0	4.2	96.1	2.5	1.1	0.912	0.054	0.024
ADRA2A	91.3	11.2	2.0	93.9	-1.2	-1.2	92.7	4.3	0.2	0.853	0.084	0.004
AR	66.5	-0.8	1.1	99.1	1.4	1.2	90.5	0.8	1.2	0.749	0.026	0.036
ADRB1	92.7	7.0	0.3	89.5	-2.5	1.5	91.2	2.5	0.9	0.823	0.046	0.017
ADRB2	72.9	-2.3	-3.3	89.9	2.2	1.4	81.6	0.0	-0.9	0.639	0.004	-0.014
OPRD1	97.1	1.1	-1.2	81.0	-0.3	4.5	92.4	0.7	0.4	0.813	0.018	0.011
DRD1	77.4	0.9	-3.7	96.6	1.5	4.3	89.2	1.4	1.2	0.773	0.030	0.028
DRD2	98.3	1.9	-1.0	84.8	5.7	7.8	96.1	2.5	0.5	0.855	0.091	0.021
SLC6A3	89.9	1.3	-3.1	94.5	1.9	4.9	91.8	1.5	0.2	0.837	0.032	0.010
EDNRA	93.8	-0.6	-3.4	95.6	1.7	4.8	94.6	0.4	0.4	0.893	0.010	0.009
NR3C1	74.5	2.3	0.7	96.9	0.1	0.6	90.1	0.7	0.6	0.760	0.018	0.015
KCNH2	84.4	15.7	-8.6	70.9	-11.5	17.4	79.1	5.0	1.6	0.558	0.059	0.036
HRH1	95.2	8.0	-0.6	88.4	-5.3	0.7	92.0	1.7	0.0	0.840	0.032	-0.001
OPRM1	94.8	1.2	-1.1	94.5	1.7	3.2	94.7	1.4	0.6	0.889	0.030	0.014
CHRM1	96.6	6.4	0.9	83.3	-2.9	0.7	91.7	2.9	0.8	0.821	0.062	0.019
CHRM2	93.9	2.9	0.0	94.5	1.1	3.2	94.2	1.9	1.8	0.883	0.039	0.035
CHRM3	91.9	3.4	-3.2	93.8	1.8	5.8	92.8	2.8	0.8	0.854	0.053	0.017

Comparative Model Performance

Target Gene	SE	Δ SA	Δ RF	SP	Δ SA	Δ RF	ACC	Δ SA	Δ RF	MCC	Δ SA	Δ RF
SLC6A2	94.9	4.2	-0.2	92.5	-1.0	1.0	93.9	2.0	0.3	0.875	0.039	0.008
HTR2A	99.1	2.4	-0.4	88.2	-0.4	6.1	96.7	1.8	1.0	0.901	0.051	0.030
HTR3A	89.4	5.7	-1.0	98.2	-0.4	0.4	95.8	1.3	0.0	0.893	0.034	0.000
SLC6A4	98.4	2.7	-0.3	89.0	2.1	6.3	96.3	2.5	1.2	0.892	0.070	0.037
LCK	95.5	3.7	0.2	79.8	-2.6	-0.9	92.3	2.5	0.0	0.763	0.054	-0.002
AVPR1A	93.9	1.9	0.6	99.3	1.5	4.8	97.2	1.6	3.2	0.941	0.034	0.068
AGTR1	87.3	0.0	3.4	99.3	1.0	1.9	94.5	0.6	2.6	0.888	0.014	0.054
AKT1	95.4	1.0	-1.4	91.3	4.4	6.6	94.1	2.1	1.1	0.864	0.049	0.028
BACE1	92.0	-1.1	-5.7	93.5	5.7	14.1	92.5	0.9	0.1	0.827	0.028	0.016
BCHE	85.6	7.1	-0.9	93.6	0.6	4.7	90.4	3.3	2.5	0.799	0.067	0.048
CASP1	69.1	5.3	-1.9	94.7	-0.1	-0.1	86.5	1.5	-0.8	0.680	0.039	-0.018
CASP3	84.8	3.4	3.7	94.9	-1.1	-1.1	91.0	0.6	0.7	0.809	0.013	0.015
CASP8	86.8	-4.9	-2.5	95.4	-2.1	-2.8	92.1	-4.1	-4.1	0.832	-0.060	-0.059
CHRM5	87.4	3.6	-4.2	95.3	1.5	4.0	92.3	2.3	0.9	0.835	0.048	0.015
CHUK	88.8	5.7	-3.3	97.4	0.4	0.0	95.2	1.7	-0.9	0.871	0.047	-0.024
CSF1R	94.3	5.9	-2.4	97.0	0.8	3.1	95.5	3.7	0.0	0.910	0.070	0.001
CSNK1D	91.4	12.4	3.8	94.8	-1.6	2.0	93.4	4.4	2.8	0.864	0.085	0.056
EDNRB	96.1	2.0	-0.5	94.4	-0.4	-0.7	95.1	0.6	-0.6	0.899	0.013	-0.012

Comparative Model Performance

Target Gene	SE	ΔSA	ΔRF	SP	ΔSA	ΔRF	ACC	ΔSA	ΔRF	MCC	ΔSA	ΔRF
ELANE	90.9	-0.4	-5.2	93.8	0.2	5.8	92.1	-0.1	-0.8	0.839	-0.001	-0.012
EPHA2	87.2	0.0	-0.7	99.6	1.1	1.8	95.4	0.7	1.0	0.899	0.018	0.023
FGFR1	96.8	5.9	0.8	92.0	-2.8	1.4	95.1	2.8	1.0	0.892	0.054	0.021
FKBP1A	88.1	-3.0	-6.9	97.4	-0.4	1.5	94.8	-1.1	-0.9	0.869	-0.028	-0.025
FLT1	91.6	1.6	-3.0	99.0	0.9	2.5	96.2	1.2	0.4	0.919	0.024	0.008
FLT4	91.9	10.7	0.0	96.5	-0.4	1.2	94.5	4.3	0.6	0.888	0.086	0.014
FYN	77.8	5.1	-3.0	98.2	0.4	1.5	92.7	1.6	0.3	0.809	0.044	0.006
GSK3B	96.8	12.2	-1.1	78.3	-5.2	2.5	90.4	6.2	0.1	0.785	0.121	0.001
HDAC3	94.1	2.3	0.3	94.8	1.5	2.7	94.5	2.0	1.6	0.890	0.039	0.032
IGF1R	94.6	-1.0	-2.1	95.5	1.3	5.1	94.9	-0.2	0.3	0.889	-0.003	0.011
INSR	91.7	1.7	-3.1	98.9	1.2	1.9	95.5	1.4	-0.5	0.912	0.028	-0.007
KDR	97.4	2.8	-0.3	73.1	-9.7	9.2	93.5	0.8	1.2	0.748	0.005	0.058
LTB4R	91.4	4.3	2.2	98.8	0.0	1.2	96.8	1.2	1.5	0.918	0.030	0.038
LYN	89.1	10.9	-1.8	98.0	2.7	2.7	95.4	5.2	1.4	0.888	0.127	0.030
MAPK1	43.5	-4.4	2.3	99.2	4.0	-0.2	79.3	1.0	0.6	0.557	0.043	0.013
MAPK9	95.5	3.5	-2.9	97.7	2.2	7.9	96.5	2.9	2.0	0.931	0.058	0.040
MAPKAPK2	86.9	5.1	-3.3	94.1	1.4	2.5	91.0	3.0	0.0	0.816	0.062	-0.001
MET	97.7	4.3	-1.0	91.9	3.2	8.4	95.9	4.0	1.9	0.904	0.091	0.045

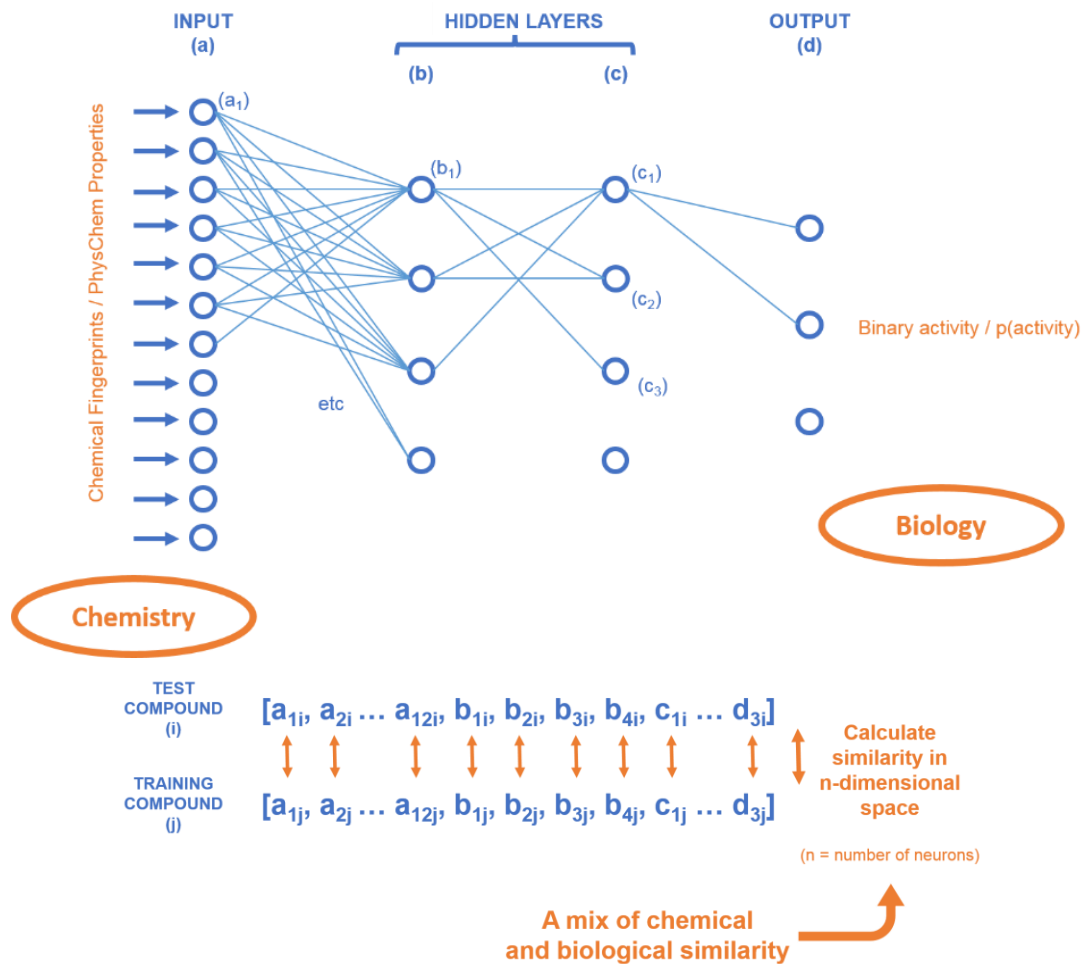
Comparative Model Performance

Target Gene	SE	ΔSA	ΔRF	SP	ΔSA	ΔRF	ACC	ΔSA	ΔRF	MCC	ΔSA	ΔRF
MMP13	94.9	1.5	-2.1	94.7	3.3	10.6	94.8	2.1	2.3	0.887	0.045	0.054
MMP2	95.5	1.7	0.0	89.2	-3.6	6.0	93.2	-0.2	2.1	0.852	-0.006	0.048
MMP3	94.7	1.8	-1.7	92.4	1.2	8.8	93.8	1.5	2.1	0.868	0.033	0.047
MMP9	81.6	-0.2	-5.0	88.8	-1.5	6.9	84.6	-0.8	0.0	0.696	-0.016	0.011
NEK2	77.0	9.4	-2.7	98.5	0.7	1.0	94.0	2.6	0.3	0.813	0.085	0.007
P2RY1	92.7	-2.4	-2.4	100.0	1.5	3.4	97.7	0.3	1.5	0.947	0.007	0.035
PAK4	89.4	7.0	-4.7	99.3	0.8	1.1	96.9	2.2	-0.3	0.914	0.064	-0.009
PDE4A	90.8	2.9	-1.1	94.9	-0.4	0.5	93.1	1.0	-0.3	0.859	0.020	-0.005
PDE5A	90.1	3.4	-4.8	96.5	2.8	7.3	93.0	3.1	0.6	0.862	0.062	0.016
PIK3CA	98.9	1.5	-0.4	93.3	-1.5	4.6	97.2	0.6	1.2	0.934	0.014	0.027
PPARG	69.5	-0.8	-2.9	96.1	3.4	2.0	86.0	1.8	0.2	0.702	0.041	0.006
PTPN1	76.6	9.5	-3.9	89.7	-2.6	3.1	84.7	2.0	0.4	0.673	0.046	0.005
PTPN11	64.8	26.2	14.8	91.9	-3.4	-3.4	85.7	3.4	0.8	0.584	0.153	0.052
PTPN2	67.9	2.5	2.5	96.0	-1.7	2.3	90.1	-0.7	2.4	0.687	-0.022	0.069
RAF1	99.7	4.2	0.0	95.2	-1.1	1.5	97.7	1.8	0.7	0.954	0.038	0.013
RARA	63.1	1.9	1.9	99.4	1.8	0.0	95.2	1.9	0.3	0.742	0.092	0.013
RARB	85.9	11.3	5.6	99.9	1.1	0.0	98.8	1.9	0.5	0.913	0.137	0.033
ROCK1	94.3	5.4	-2.5	92.0	-3.6	1.1	93.2	1.2	-0.9	0.864	0.021	-0.018

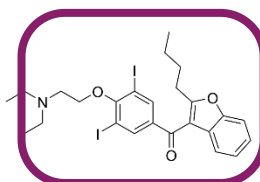
Comparative Model Performance

Target Gene	SE	Δ SA	Δ RF	SP	Δ SA	Δ RF	ACC	Δ SA	Δ RF	MCC	Δ SA	Δ RF
RPS6KA5	73.7	12.3	0.0	100.0	1.5	2.7	95.3	3.4	2.2	0.835	0.135	0.080
SIRT2	70.8	2.3	5.6	95.2	-0.6	-1.6	89.8	0.0	0.0	0.692	0.003	0.006
SIRT3	76.7	-2.4	-4.7	98.5	0.8	0.8	95.4	0.3	0.0	0.802	0.010	-0.005
SRC	94.7	4.6	-2.7	88.7	0.7	8.1	92.4	3.1	1.5	0.839	0.064	0.029
TACR2	87.6	-1.6	-4.4	100.0	2.4	2.4	96.0	1.1	0.2	0.910	0.029	0.008
TBXA2R	88.8	-0.4	-2.1	94.9	-0.4	1.4	93.0	-0.4	0.3	0.838	-0.010	0.003
TEK	89.9	2.5	-4.1	97.8	2.6	1.9	94.5	2.6	-0.6	0.887	0.053	-0.013

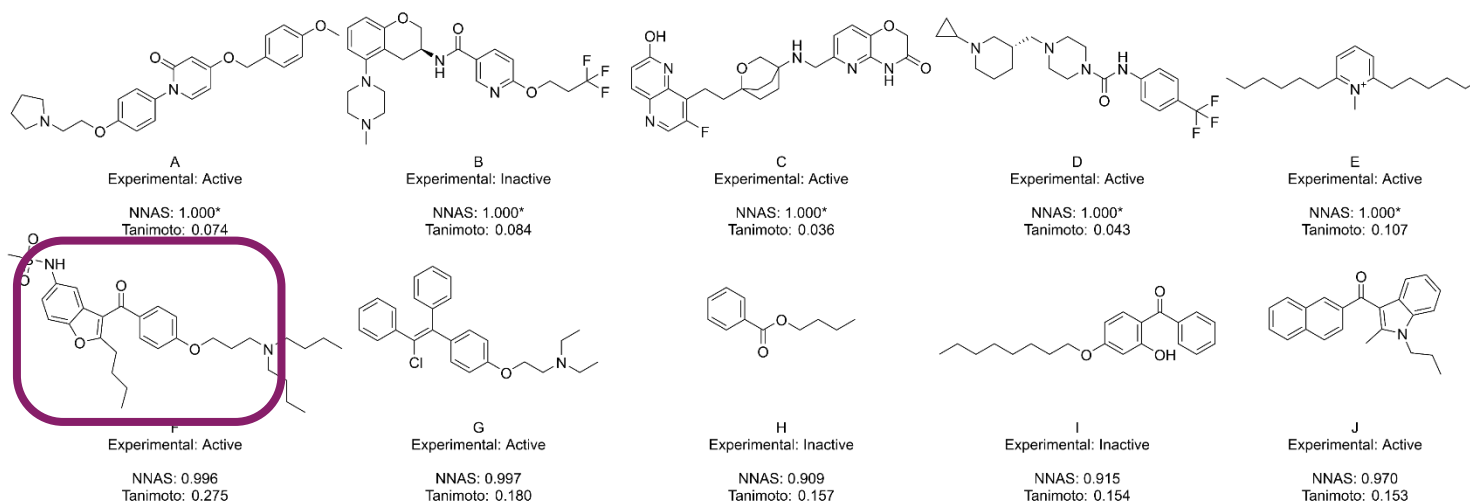
Neural Network Activation Similarity



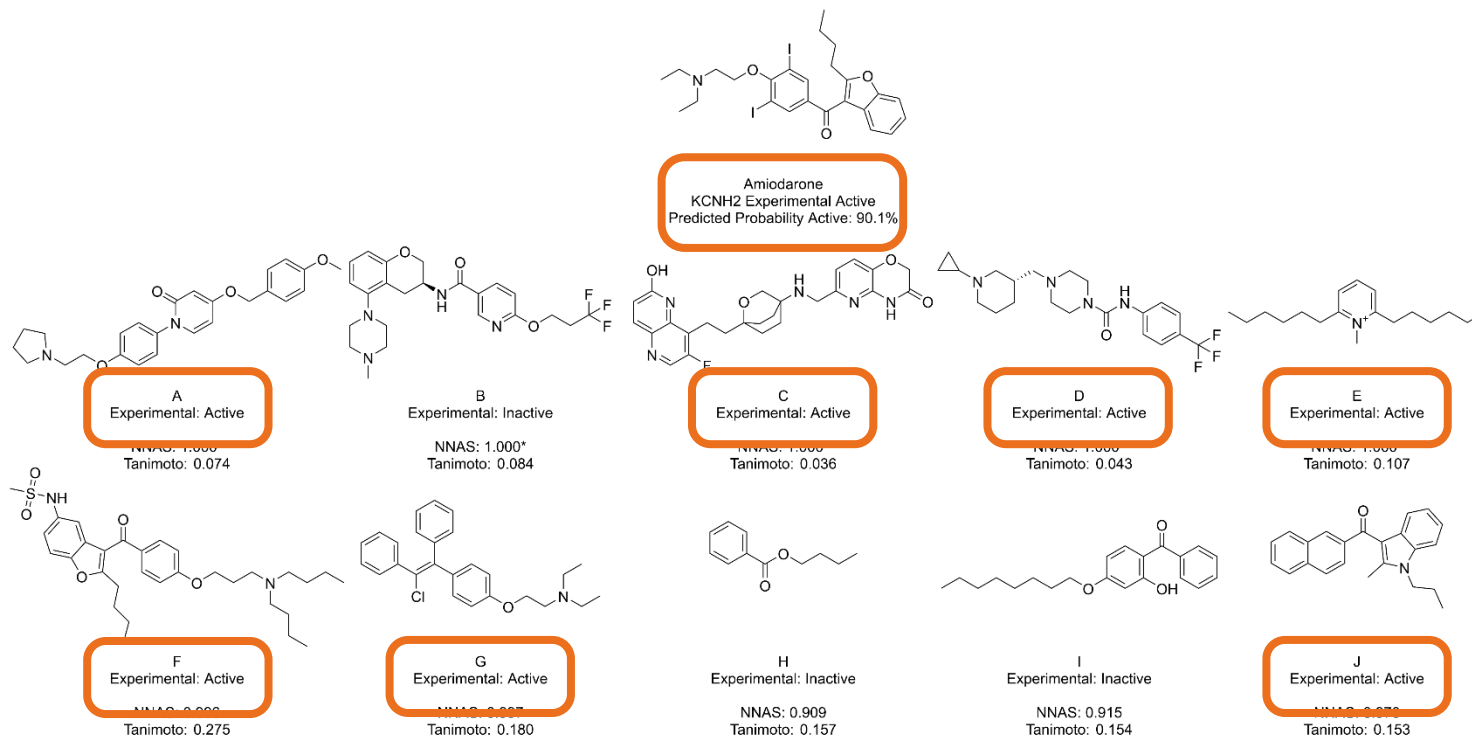
Amiodarone (hERG)



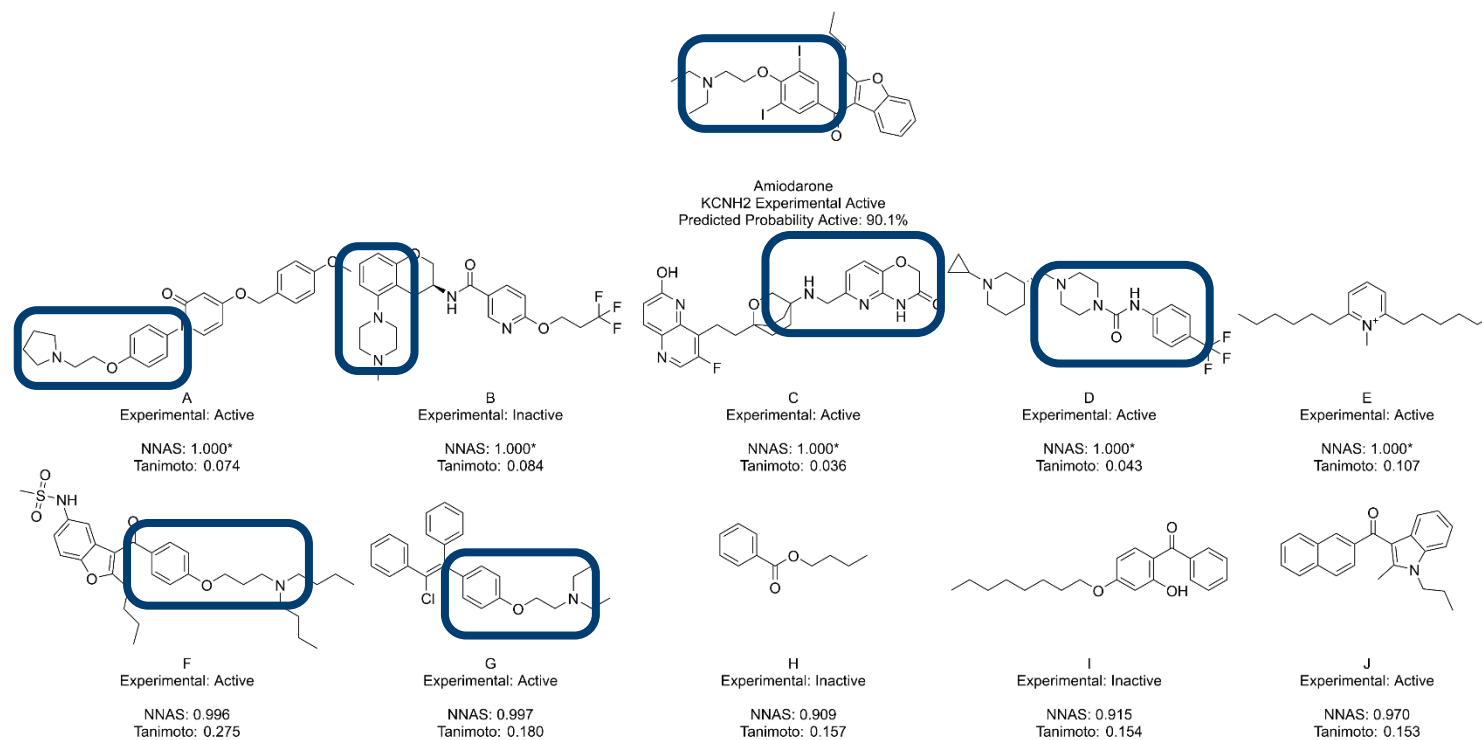
Amiodarone
KCNH2 Experimental Active
Predicted Probability Active: 90.1%



Amiodarone (hERG)



Amiodarone (hERG)

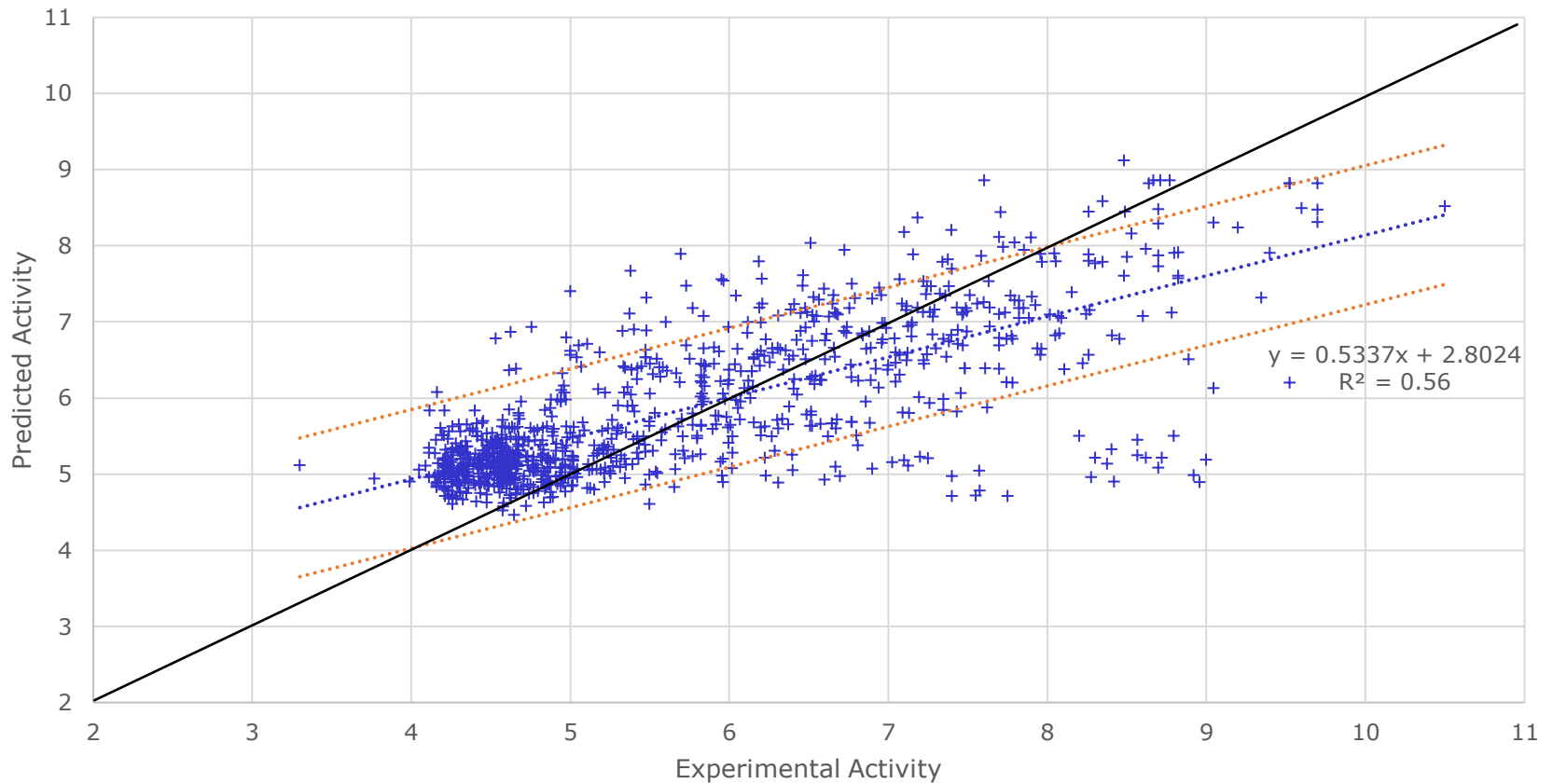


Quantitative Predictions

- Dose-Response Relationships and Risk assessment Procedures require Quantitative Information
- Adjustment of AR dataset to contain only quantitative activity vales ($p(\text{Activity})$, 4880 values)
- Change of loss function to MSE
- Single output node with linear activation function
- Models evaluated using MSE and RMSE

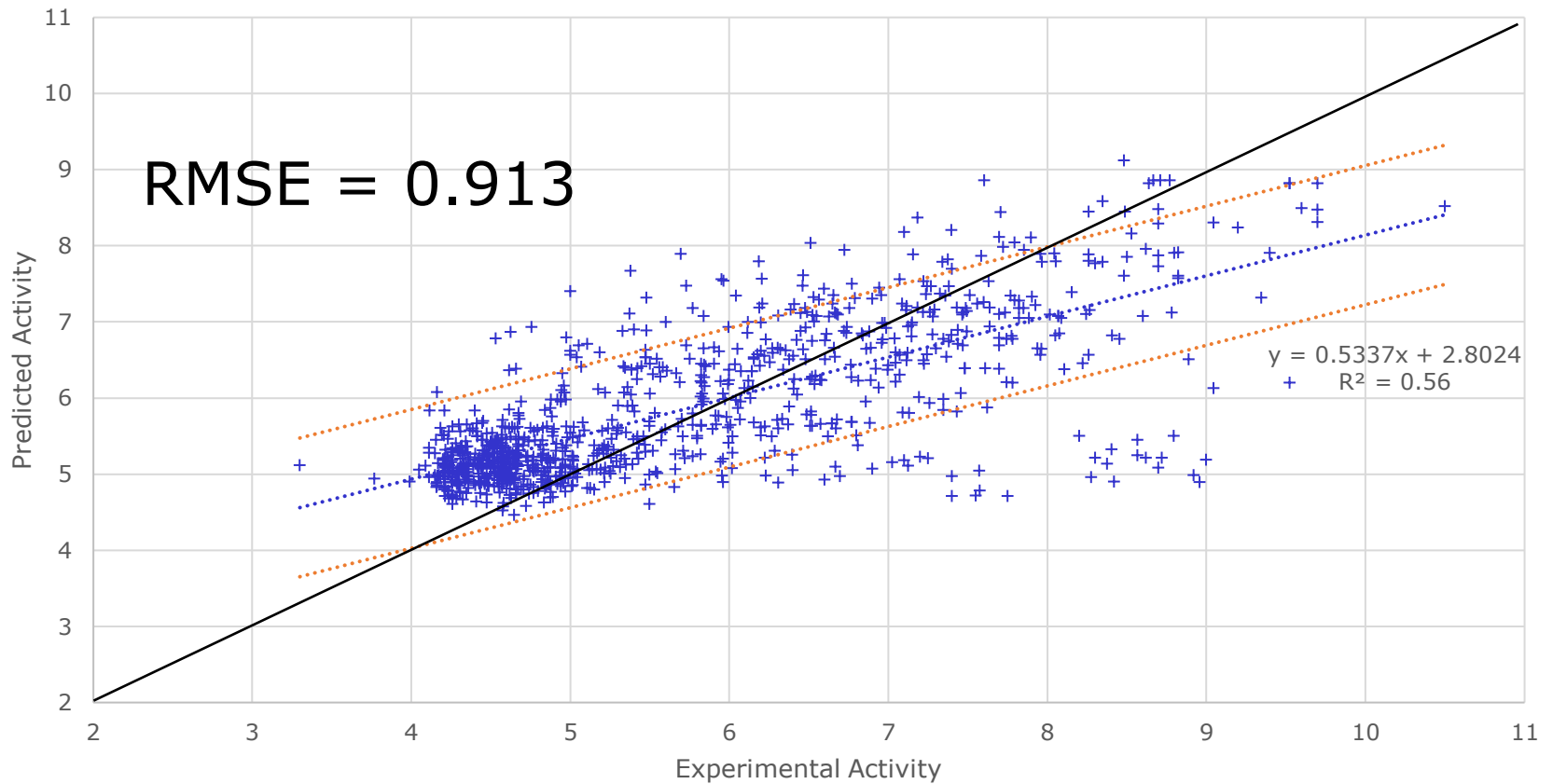
Quantitative Predictions

Quantitative Predictions at AR



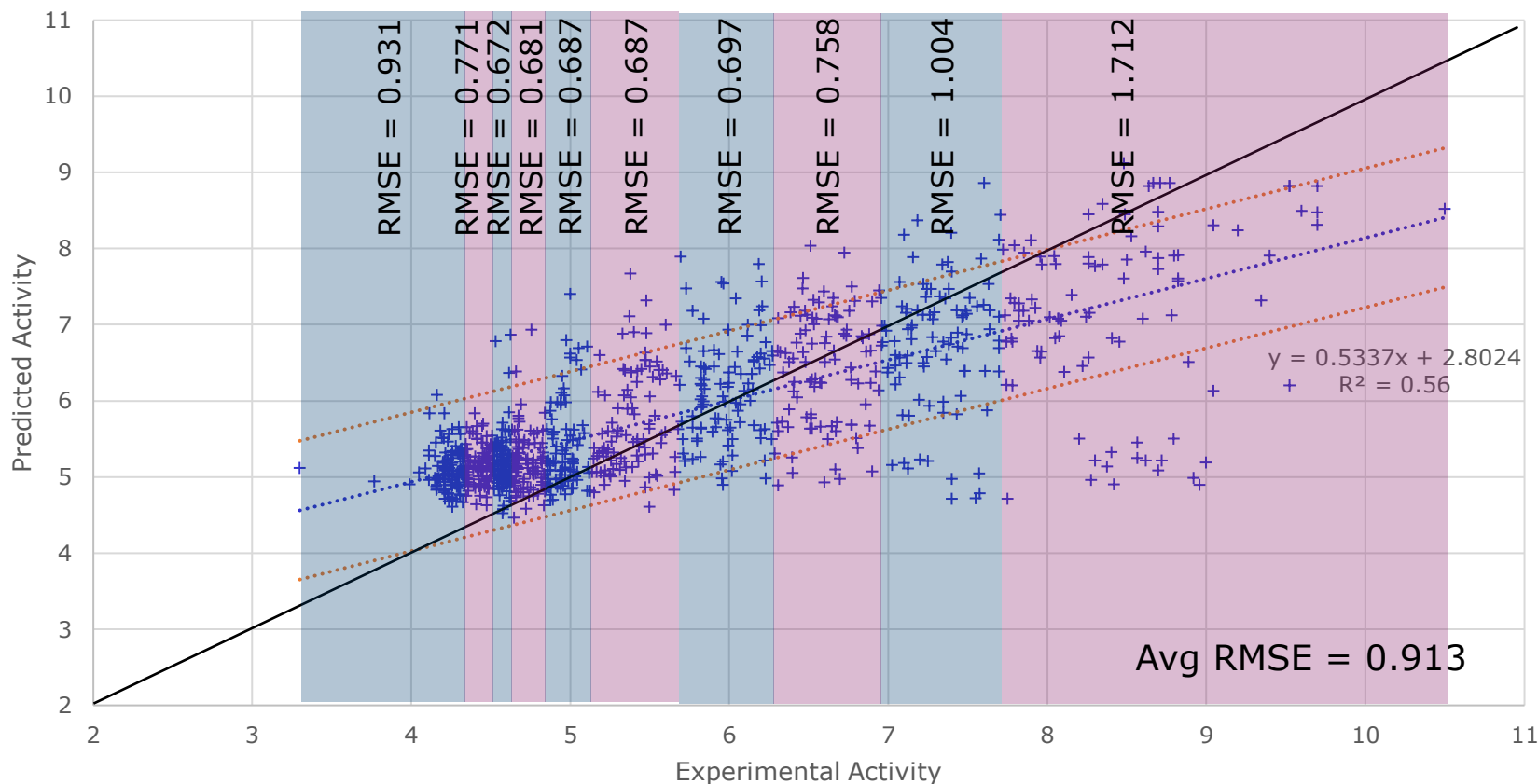
Quantitative Predictions

Quantitative Predictions at AR



Quantitative Predictions

Quantitative Predictions at AR



Conclusions

- Neural Networks are a class of Machine Learning Algorithms that can provide both binary and quantitative predictions
- Structural Alerts, Random Forests and Neural Networks have been used to try and predict binary activity at Human MIEs
- A combination of these models and understanding of their workings is key to highest performance and model use in toxicology decision making
- Quantitative predictions help push this methodology closer to use in risk assessment, rather than just hazard identification

Acknowledgements



- Professor Jonathan Goodman
- Professor Anne Willis
- Unilever
- Dr Paul Russell, Dr Steve Gutsell & colleagues at SEAC, Unilever
- Andrew Wedlake
- Maria Folia & Dr Sam Piechota
- The Centre for Molecular Informatics
- The MRC Toxicology Unit
- St. John's College

