

Virtual Toxicity Panel Screens to aid the Medicinal Chemist

MedChemica
contact@medchemica.com

A. G. Dossetter•, E. Griffen•, A. Leach•+, A. Lin†, J. Stacey†, L. Reid§, S. Montague•.

• Medchemica Ltd, Macclesfield, UK. + Pharmacy and Biomolecular Sciences, Liverpool John Moores University. †Information School, University of Sheffield.

‡Laboratory of Chemoinformatics, Faculty of Chemistry, University of Strasbourg.

§Bioinformatics Institute (A*STAR), 30 Biopolis Street, Matrix, Singapore 138671

Problem

Unforeseen toxicity via secondary pharmacology is a significant risk and when encountered late in a discovery project's life creates major issues and may even terminate it. Chemists need to be alerted to potential risks but to be influenced they must be able to audit the reasons and evidence for the alerts.

Solution

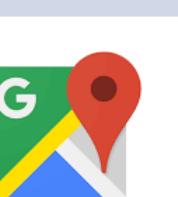
Build transparent models of critical toxicity targets and communicate results in chemical structures rather than just numbers. This is an example of 'Explainable AI' for chemists

Explainable AI

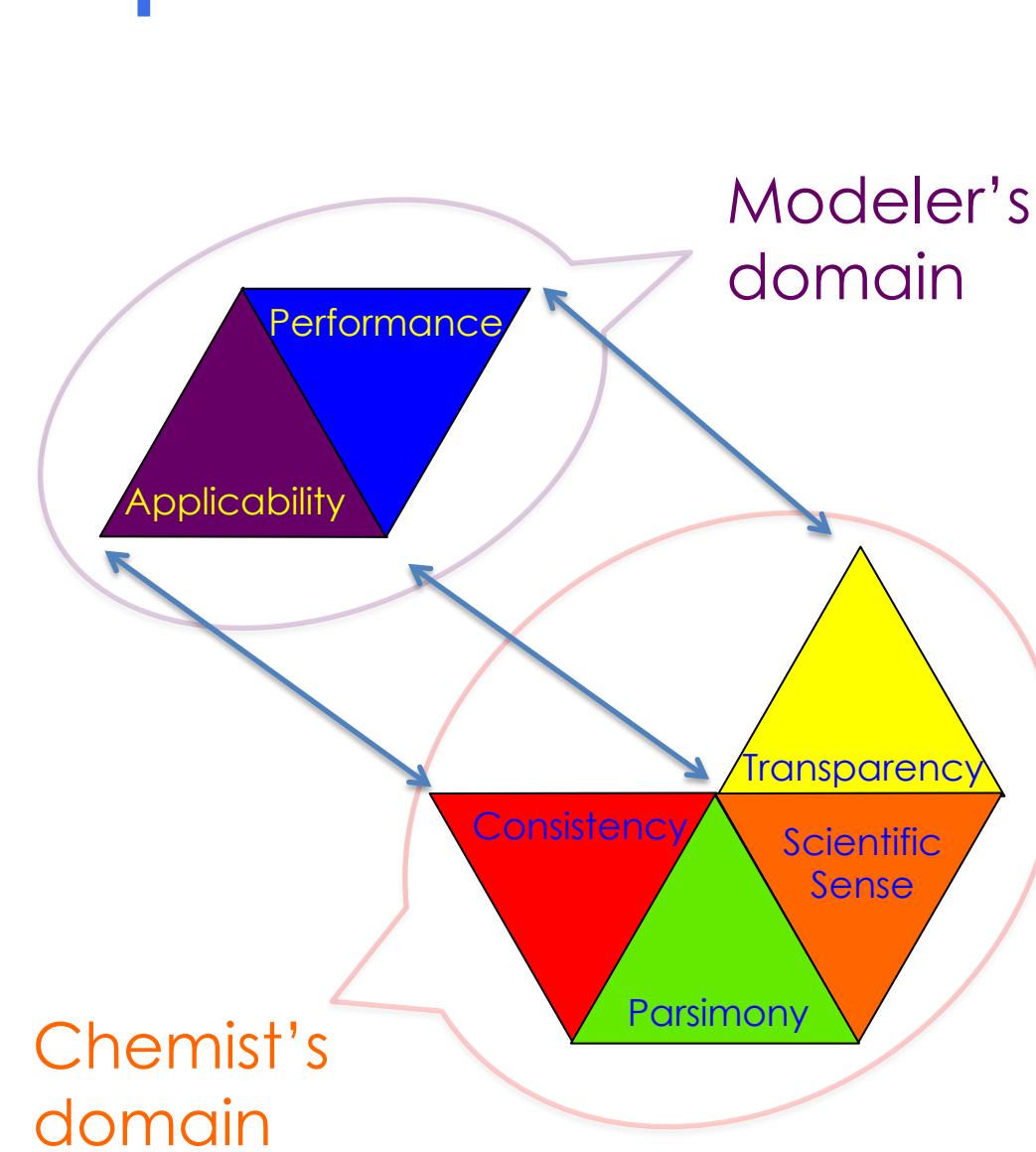
Trying to explain black box models, rather than creating models that are interpretable in the first place, is likely to perpetuate bad practice and can potentially cause great harm to society. [The way forward is to design models that are inherently interpretable](#).

- Cynthia Rudin Nature Machine Intelligence (2019), 206–215.

Black Box	Interpretable
Failure cost is low	Failure cost high
Real time response critical	Immature science
Interactive = self correcting	Highly skilled, critical users
Business-2-consumer	Business-2-Business
User agnostic of process	Transparent and auditable



Aspects of Models



Chemists won't make decisions without understanding

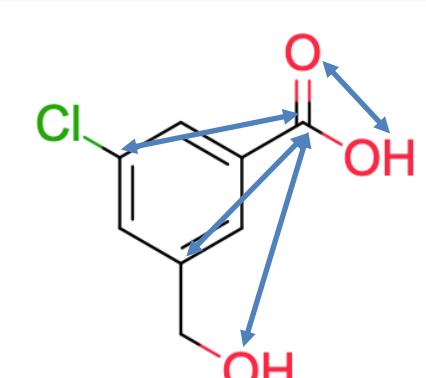
Language of medicinal chemists = structures / clear pharmacophores

Machine Learning method	Description
MMPA transformations	Example pairs
kNN + Morgan fp	Structures of Nearest Neighbours
Random Forest + pharmacophore fp	Compound highlighted with important features
Graph analytics	Connections between compound families
Graph Convolutional Neural Network (GCNN)	Graph node feature importance – a work in progress

Approach

Advanced Pharmacophore Features

Feature	Definition
Basic Group	Atom or group most likely protonated at pH 7.4
Acidic Group	Atom or group most likely deprotonated at pH 7.4, includes N and C acids
Acceptor	Definitions derived from Taylor & Cosgrove
Donor	Definitions derived from Taylor & Cosgrove
Hydrophobic	C4 or greater cyclic or acyclic alkyl group
Aromatic Attachment	connection of any group to an aromatic atom excluding connections within rings
Aliphatic Attachment	connection of any atom to an aliphatic group not in a ring.
Halo	F, Cl, Br, I



Gobbi, A.; Poppinger, D. Biotechnology and Bioengineering 1998, 61 (1), 47–54.
Reutlinger, M.; Koch, C. P.; Reker, D.; Todoroff, N.; Schneider, P.; Rodrigues, T.; Schneider, G. Mol. Inf. 2013, 32 (2), 133–138.
Taylor, R.; Cole, J. C.; Cosgrove, D. A.; Gardner, E. J.; Gillet, V. J.; Korb, O. J. Comput Aided Mol Des 2012, 26 (4), 451–472.

Acid & Base definitions are SMARTS including C, N, heteroaromatic acids, bases excluding weak aniline bases, including amidines, guanidine's - MedChemica definitions.

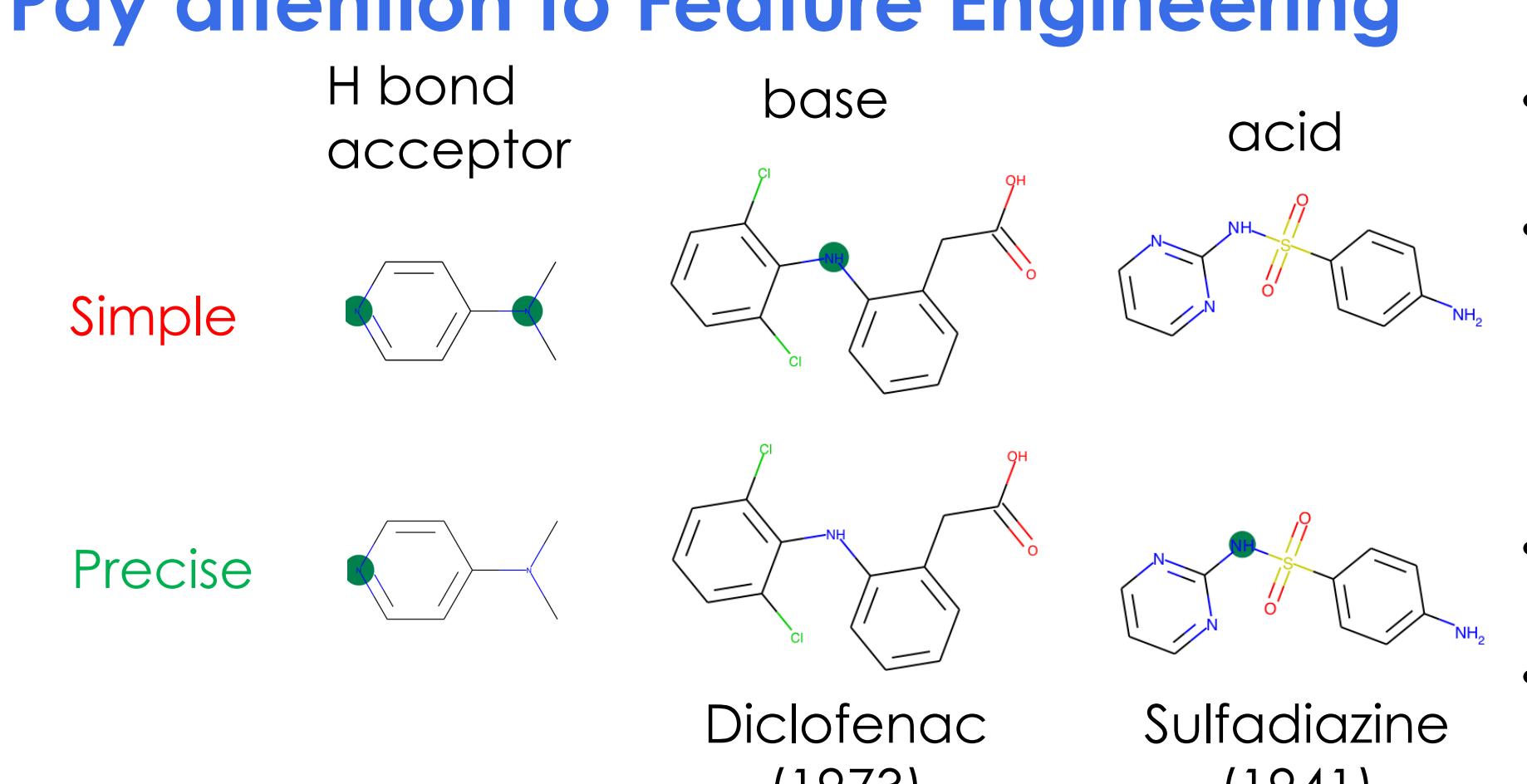
Application

Cardiac toxicity and Seizure are key toxicological risks

Cardiac	Seizure	Seizure
hERG ion channel inhibitor	Dopamine D1 receptor ant/ag	5HT 1A receptor antagonists
NaV 1.5 channel inhibitor	Dopamine D2 receptor ant/ag	5HT 1B receptor antagonists
Ca L type channel inhibitor	Cannabinoid CB1 receptor ant/ag	5HT receptor antagonists 2A
Ca T-type channel inhibitor	Acetylcholine $\alpha 1\beta 2$ receptor agonist / antagonists	GABA a1 antagonist
PDE 3A inhibitor	μ opioid agonist / antagonists	NMDA-NR1 agonist
	κ opioid agonist	5HT Transporter inhibitor
	δ opioid agonist/ antagonists	Dopamine Transporter inhib
	Muscarinic M1 receptor ant/ag	Noradrenaline Transporter inh
	Muscarinic M2 receptor ant/ag	Acetylcholine esterase inhibitor
		Monoamine oxidase inhibitor
		PDE 4D inhibitor

Bowes J., et al. Reducing safety-related drug attrition: the use of in vitro pharmacological profiling. Nature Reviews Drug Discovery 2012;11:909–22.

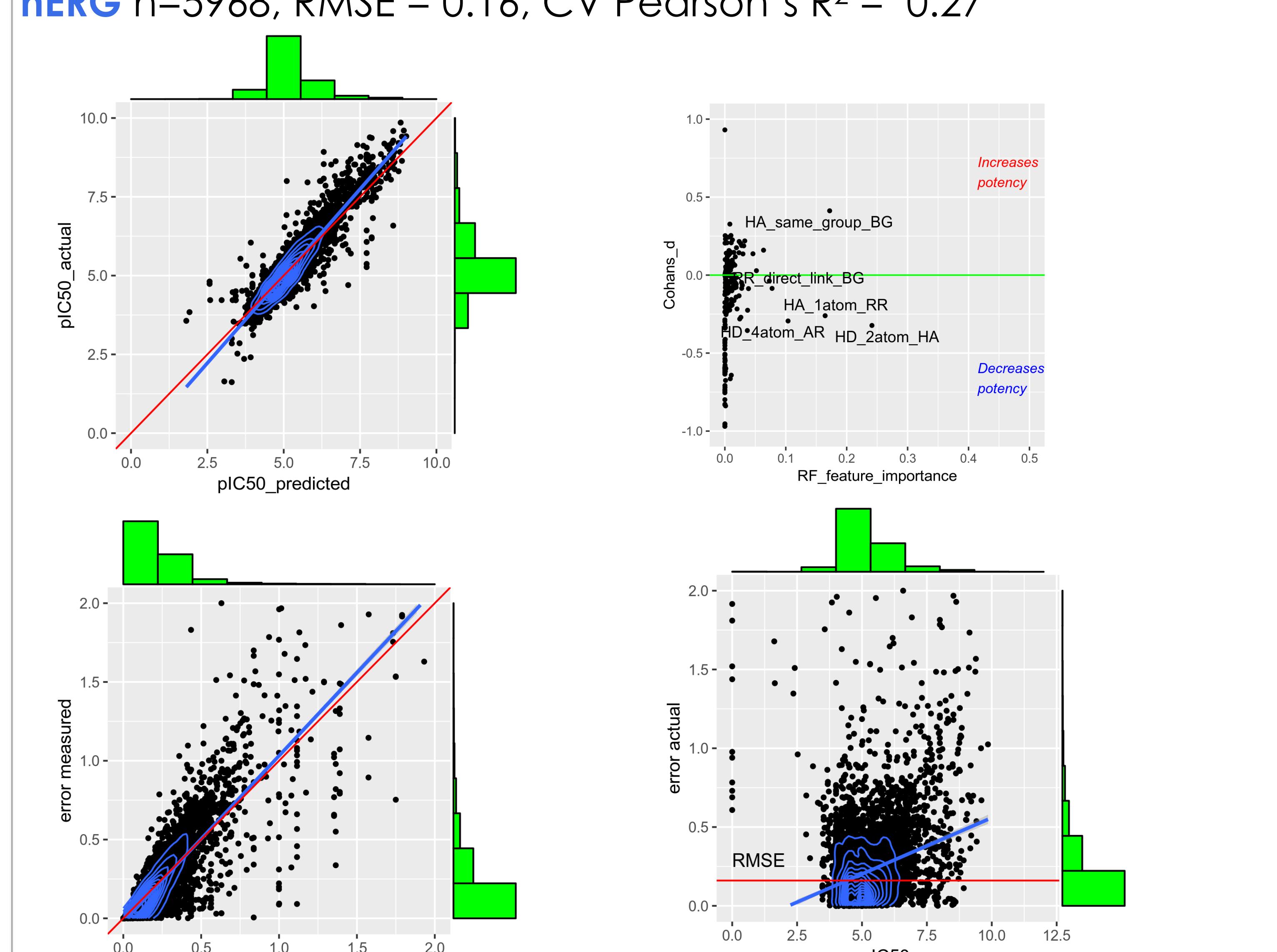
Pay attention to Feature Engineering



Pharmacophore Pairs

- Feature 1 – topological distance - Feature 2
- Engineered for chemical relevance – pairs can be superimposed or directly linked, e.g. enables a group to be both a hydrogen bond acceptor and a base
- Used as *unfolded* 280 bit fingerprints
- A bit identifies a pharmacophore pair e.g.: Aromatic - 3 bonds - Base

hERG n=5968, RMSE = 0.16, CV Pearson's R² = 0.27



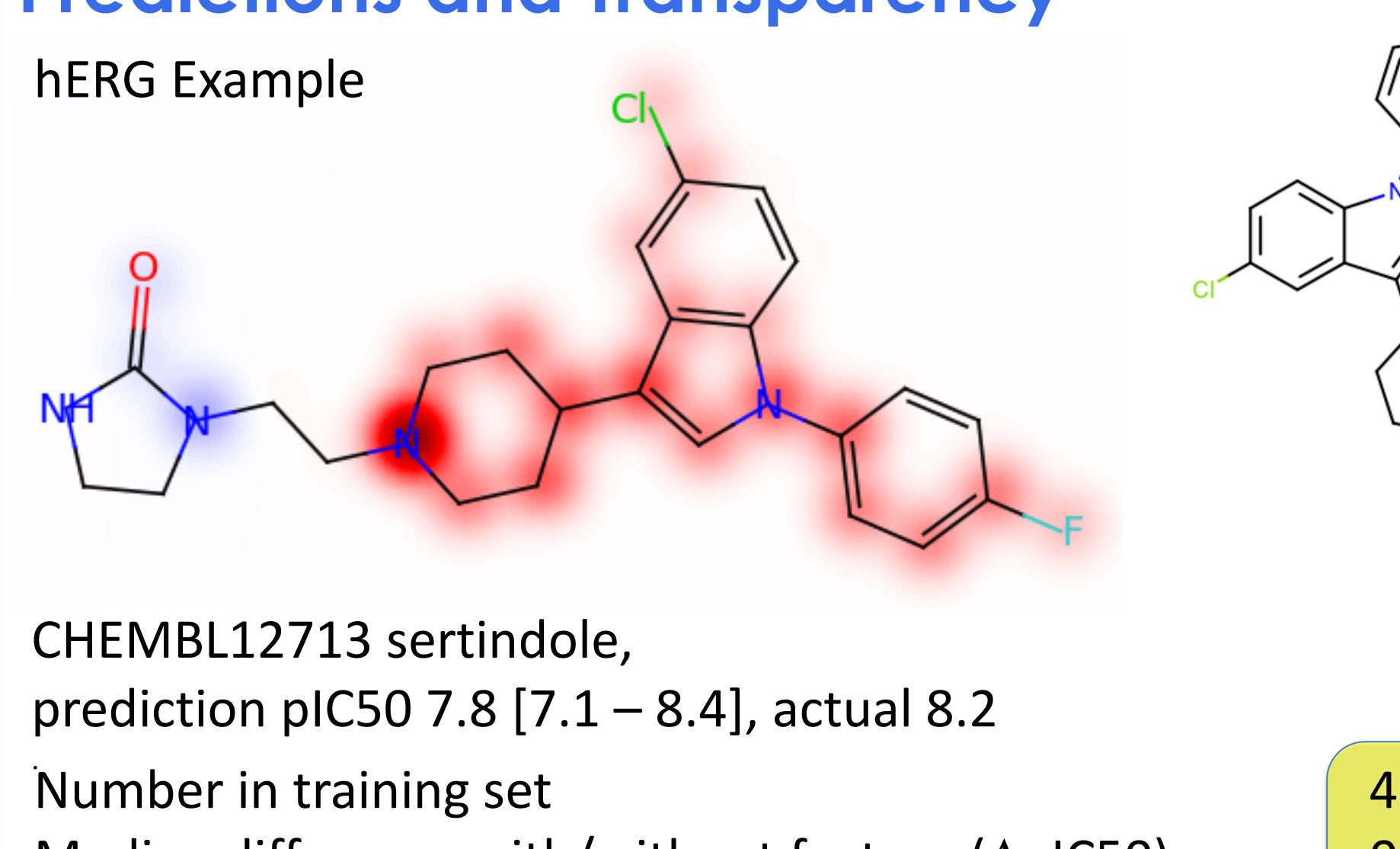
Clear definitions enables identifying key features

- Random Forest feature importance and Cohan's d for effect size allow identification of critical features in models
- Highlight atoms by Σ Feature Importance coloured by direction of Cohan's d
- Show statistics on the effect and variance of each feature
- Drill back to precise features and original compounds with data supporting that feature – *complete transparency*

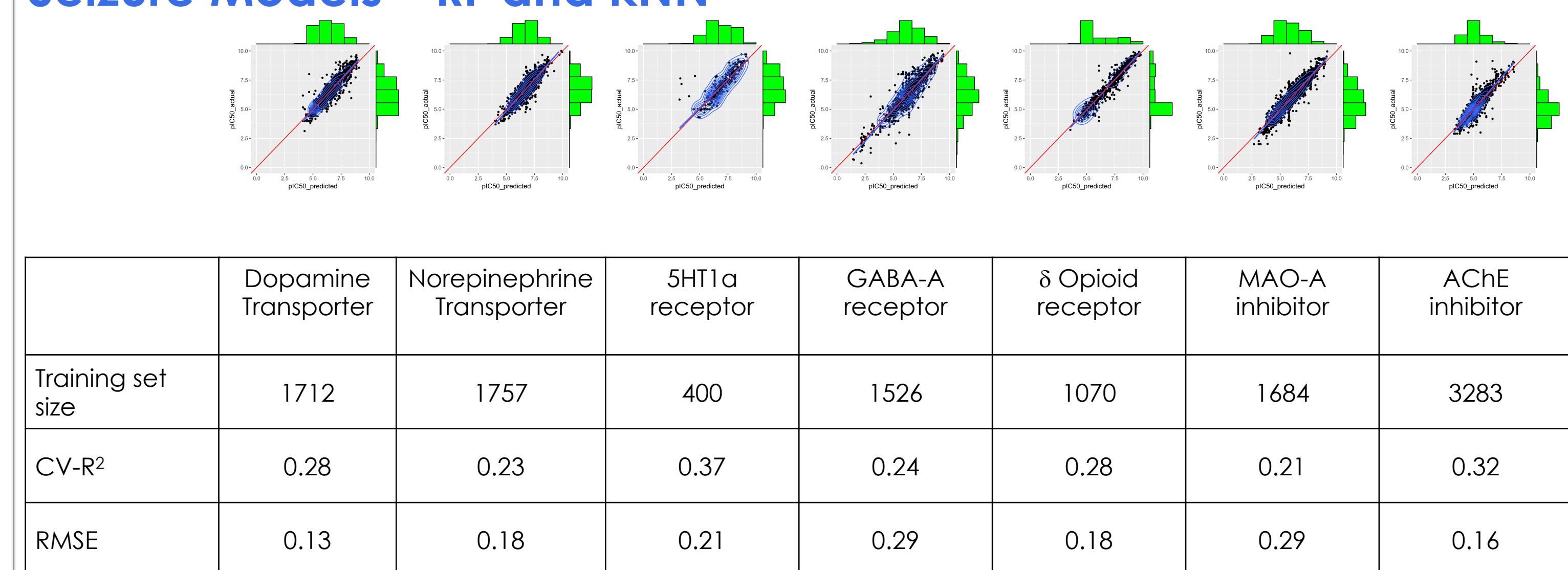
Model 'quality', Error models and Domain of applicability

- Build models with 10 fold CV – report CV-Pearson's R² and CV RMSE
- Build a Random Forest error model to generate predicted error for each compound
- Error model can be used to flag compounds out of Domain of Applicability

Predictions and Transparency



Seizure Models – RF and kNN



Learning

- Models must be transparent and show structures to influence chemists
- Random Forest models with the correct descriptors can be used to show important features as pharmacophores and the evidence supporting them
- Error models can give a measure of confidence to predictions beyond use of an RMSE.