PBPK/PD modeling and machine learning approaches to support the development of new drugs and regimens against tuberculosis

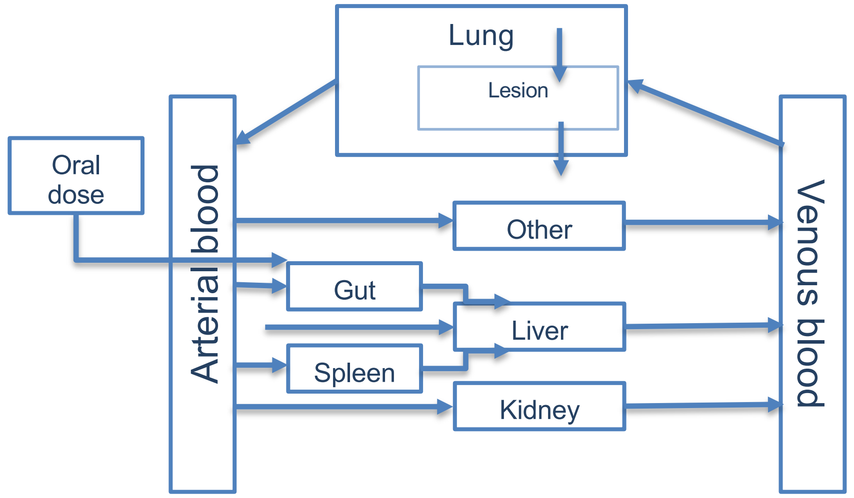
Fondazione The Microsoft Research - University of Trento
Centre for Computational and Systems Biology (COSBI)

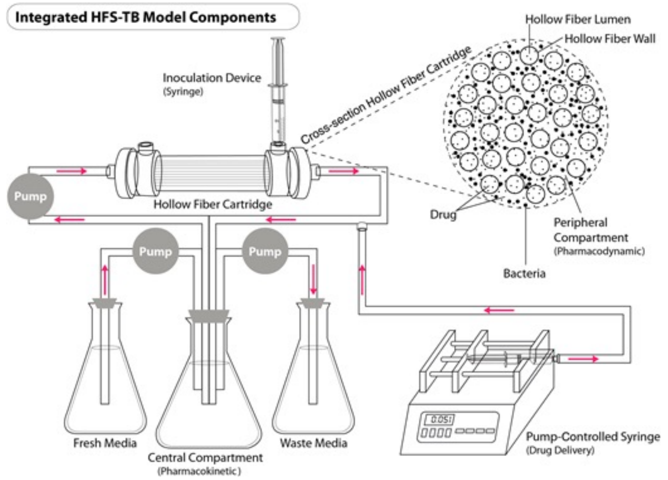
Federico Reali, PhD (reali@cosbi.eu)

PAGE 2022, Ljubljana
June 30th, 2022

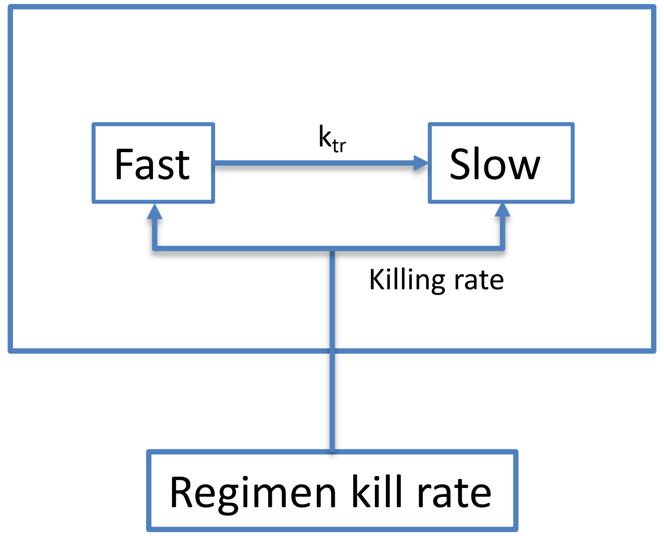
Collaborative projects on TB

Development of models and simulation tools for enabling in-silico translation





PD Model



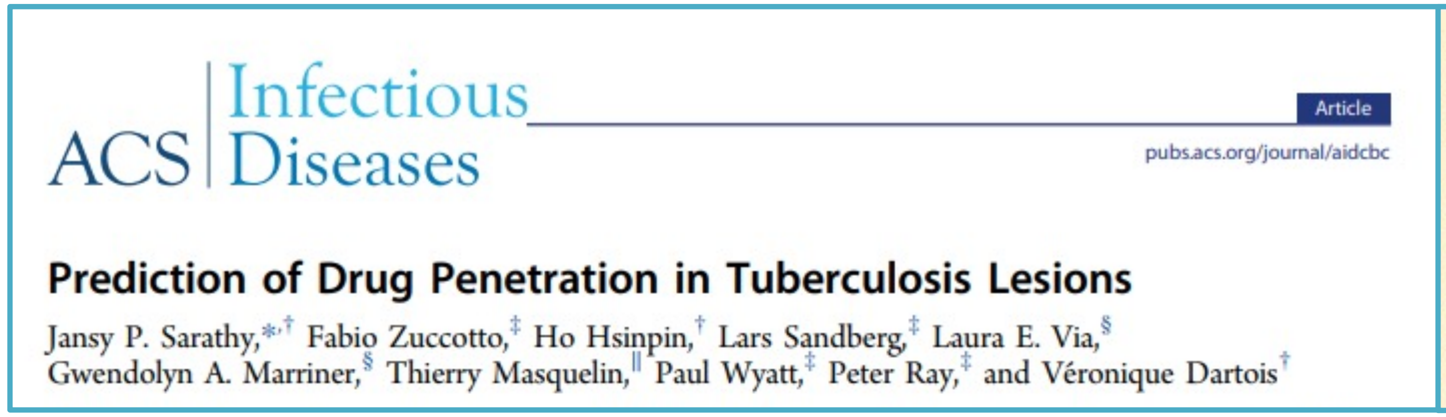
Computational approaches to predict drug lesion penetration in tuberculosis

COSBI Team:
Anna Fochesato
Roberto Visintainer
Luca Marchetti

Gates MRI:
Micha Levi
Shayne Watson
Karim Azer (former)
Chanchala Kaddi (former)

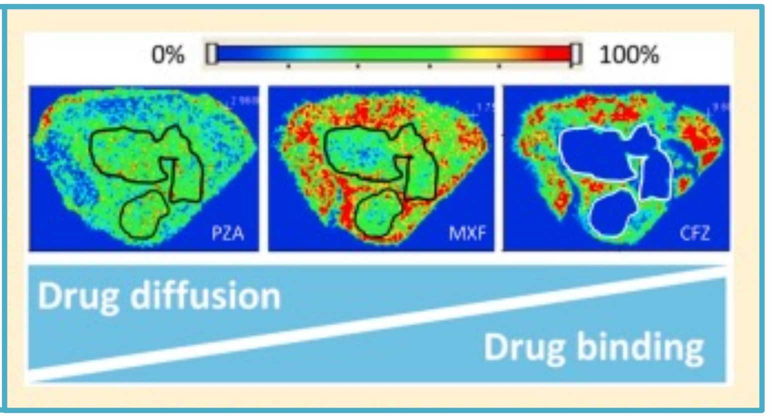
Hackensack Meridian Health:
Veronique Dartois

Predict drug penetration in TB lesions

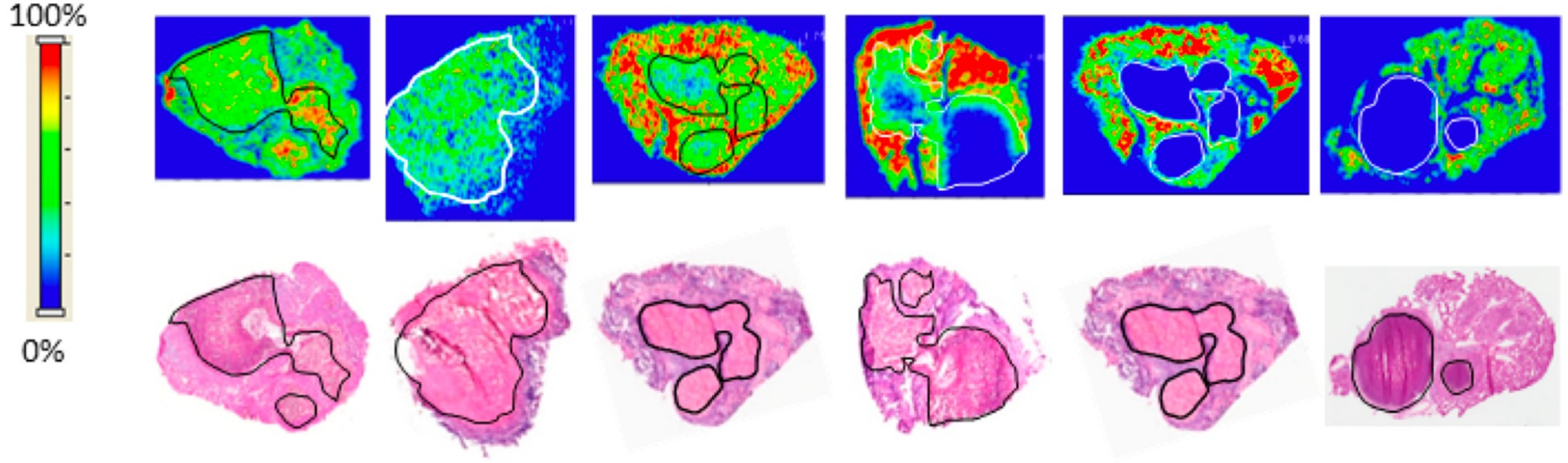

ACS Infectious Diseases
Article
pubs.acs.org/journal/aidcbc

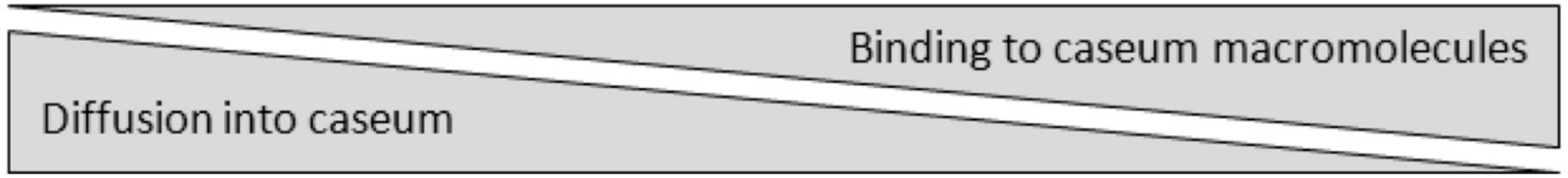
Prediction of Drug Penetration in Tuberculosis Lesions

Jansy P. Sarathy,[*,†] Fabio Zuccotto,[‡] Ho Hsinpin,[†] Lars Sandberg,[‡] Laura E. Via,[§]
Gwendolyn A. Marriner,[§] Thierry Masquelin,[∥] Paul Wyatt,[‡] Peter Ray,[‡] and Véronique Dartois[†]



Drug	Pyrazinamide	Acetyl-isoniazid	Moxifloxacin	Rifampicin	Clofazimine	Bedaquiline
***fu* (%)**	100	100	13.5	5.1	<0.01	<0.01




Binding to caseum macromolecules

Diffusion into caseum

Predict drug penetration in rabbit TB lesions

ACS Infectious Diseases

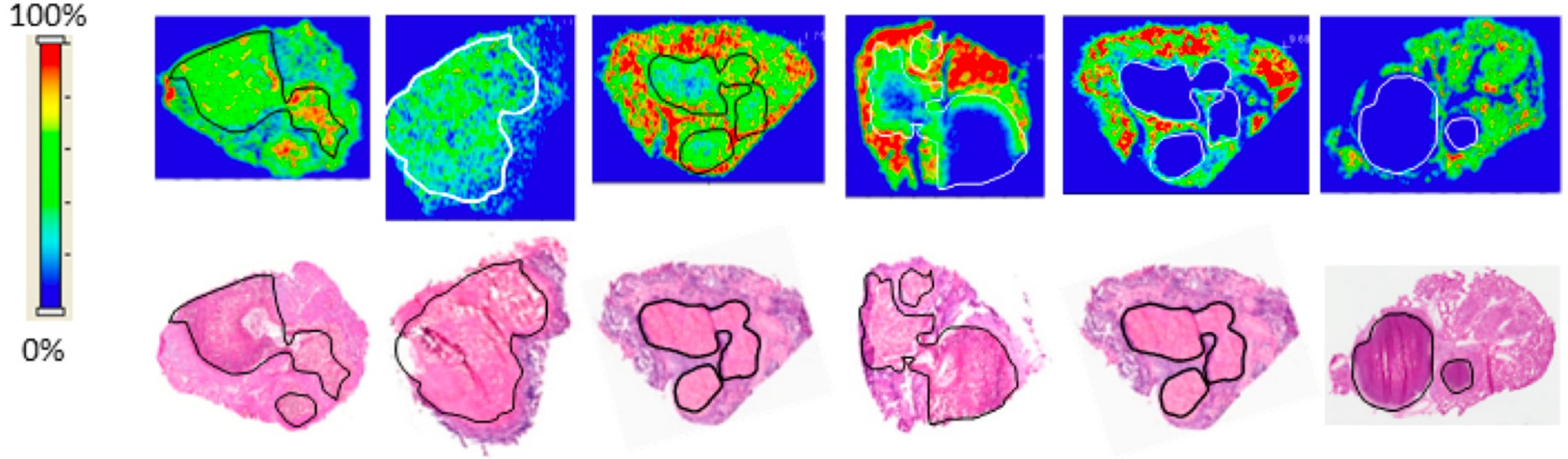
Article

pubs.acs.org/journal/aidcbc

Prediction of Drug Penetration in Tuberculosis Lesions

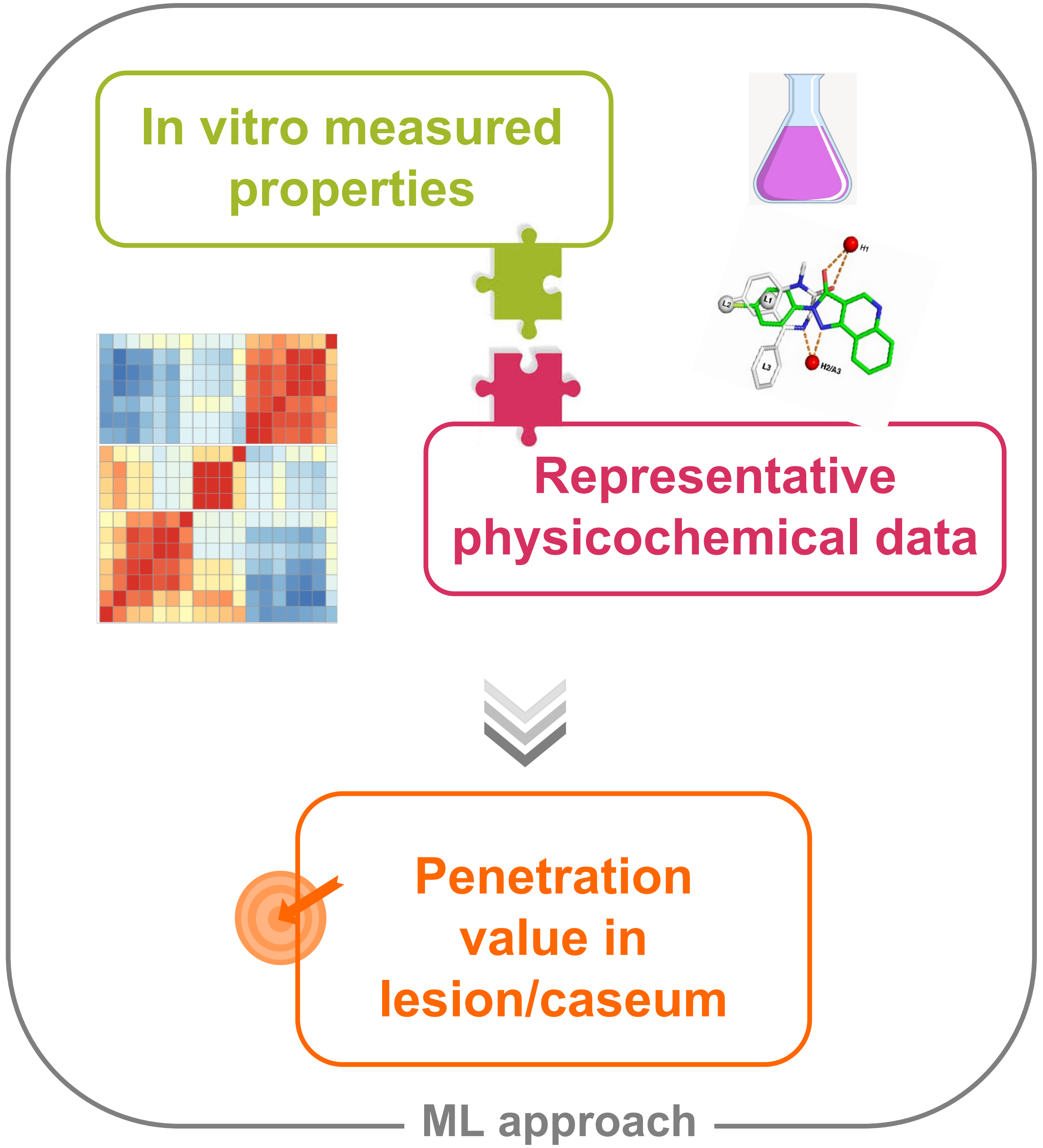
Jansy P. Sarathy,[*,†] Fabio Zuccotto,[‡] Ho Hsinpin,[†] Lars Sandberg,[‡] Laura E. Via,[§] Gwendolyn A. Marriner,[§] Thierry Masquelin,[∥] Paul Wyatt,[‡] Peter Ray,[‡] and Véronique Dartois[†]

Drug	Pyrazinamide	Acetyl-isoniazid	Moxifloxacin	Rifampicin	Clofazimine	Bedaquiline
fu (%)	100	100	13.5	5.1	<0.01	<0.01



Binding to caseum macromolecules

Diffusion into caseum



In vitro measured properties

Representative physicochemical data

Penetration value in lesion/caseum

ML approach

ML model definition

Prediction model:

Structure

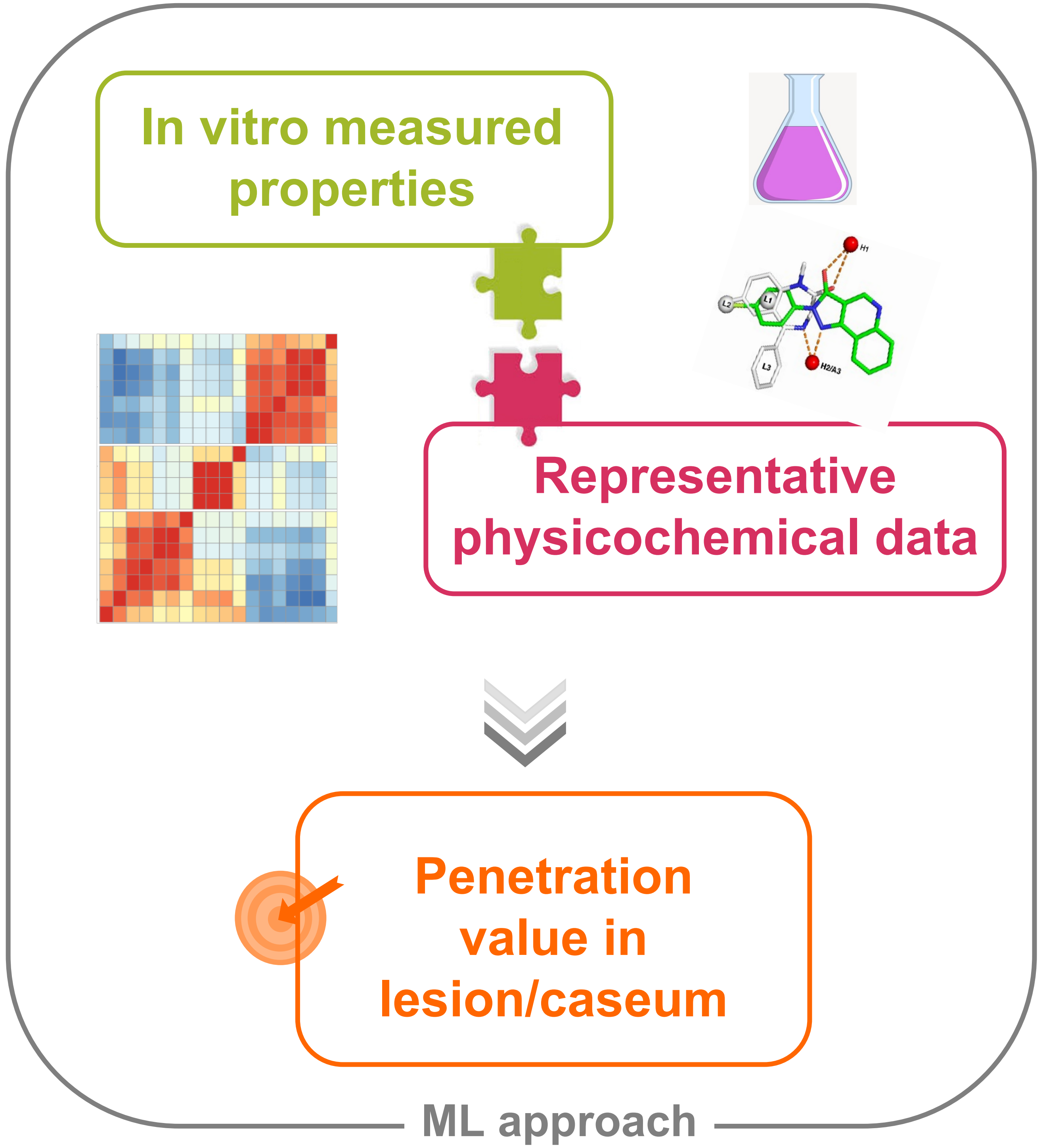
- Variables
- Target
- Algorithm

Estimation

- Cross-validation

Testing

- Accuracy & Stability



Physicochemical Variables

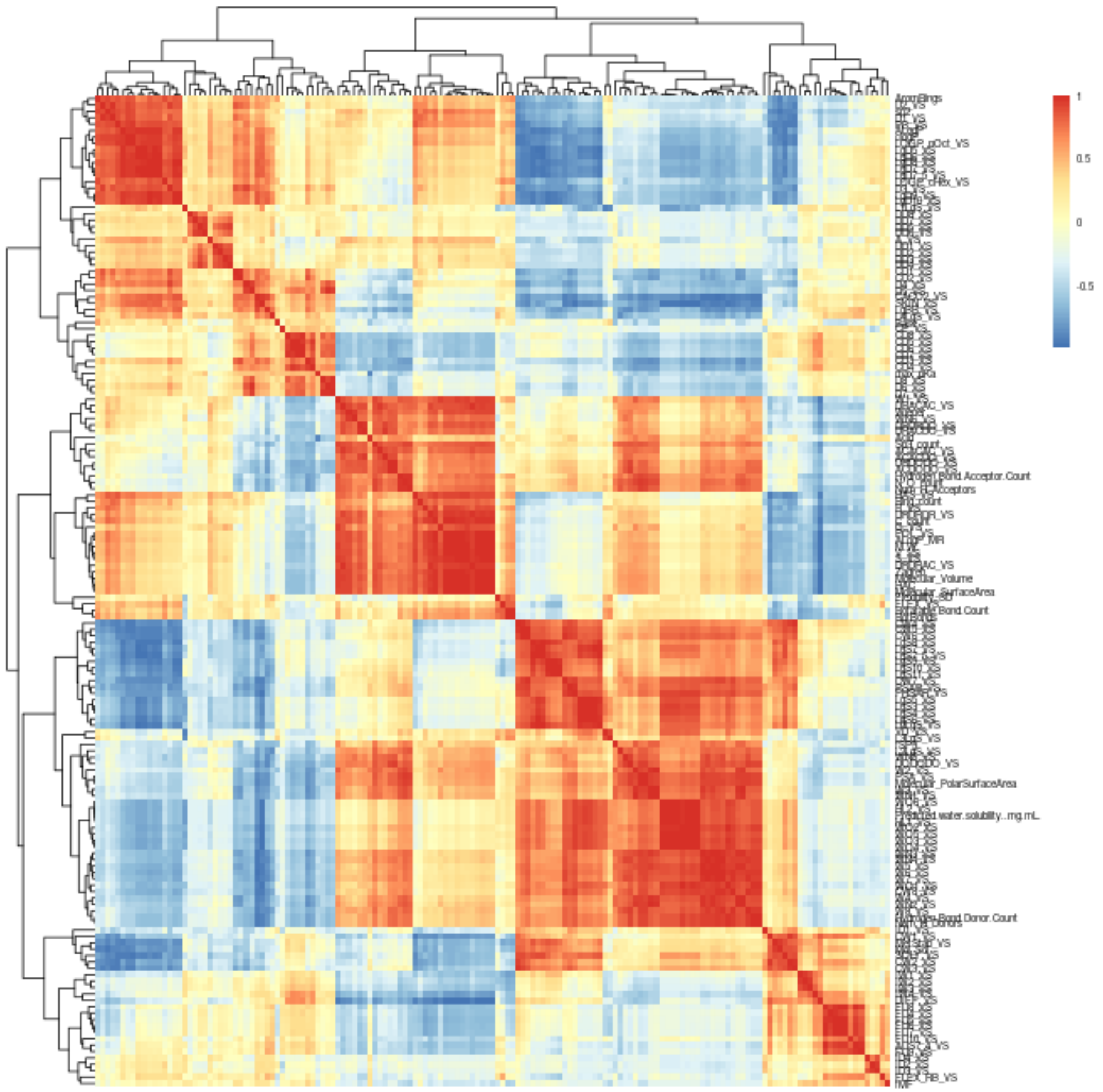
Collect variables from

- Literature
- Internal data
- Computational chemistry software
- Chis Cooper (TB Alliance)
- Fabio Zuccotto (Dundee University)

Select a short list of representative

- easy to retrieve
- reduce model complexity
- cope with collinearity in the data

Such as:
Molecular Weight
Predicted water solubility
Octanol/Water Partition Coefficient
Protein Bindings
THP-1 uptake



The correlation matrix of the whole physicochemical dataset shows high collinearity between many variables

AUC initial data

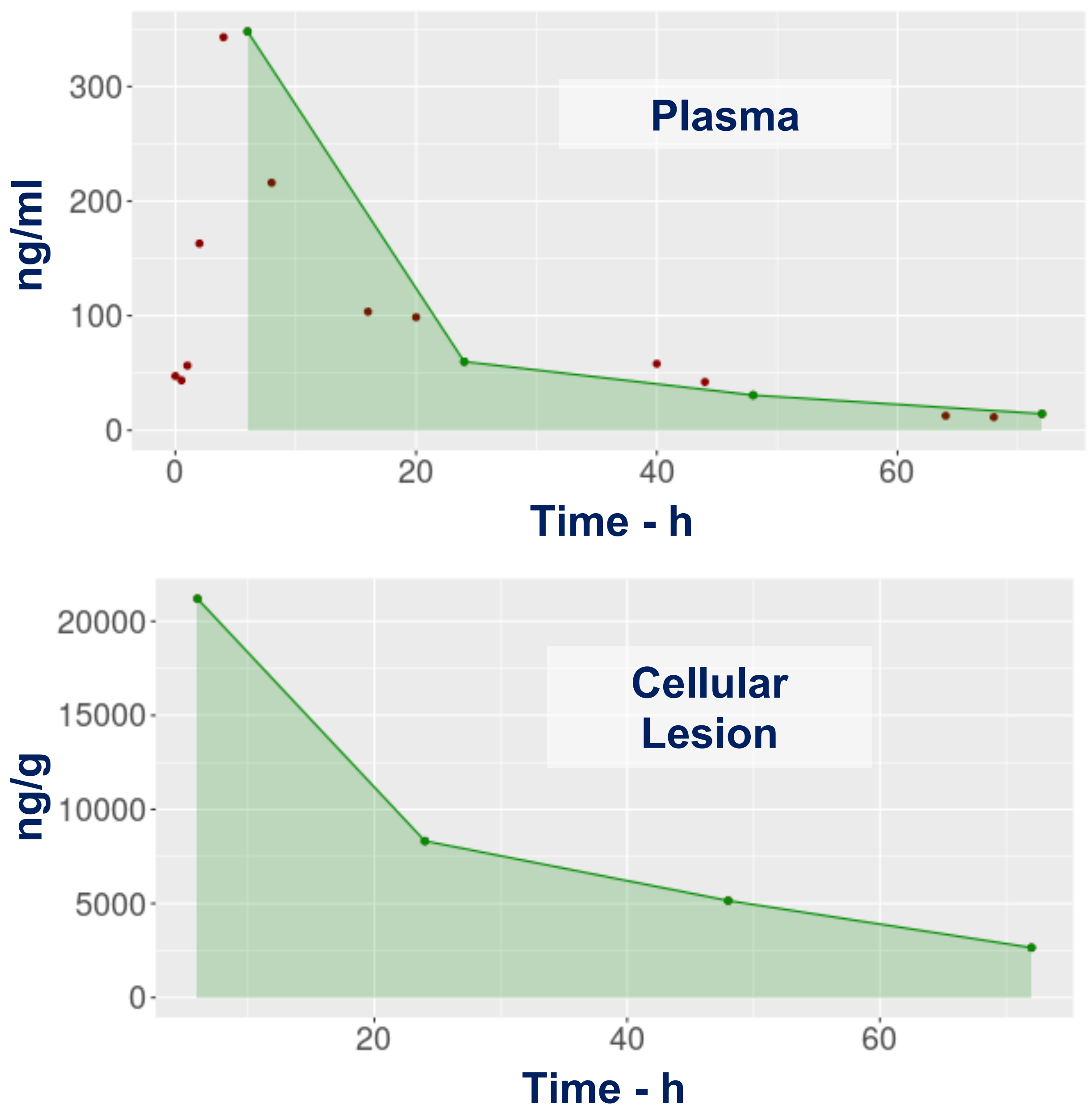
$AUC_L: \log_{10}(AUC_{cellularLesion} / AUC_{Plasma})$

$AUC_C: \log_{10}(AUC_{Caseum} / AUC_{Plasma})$

We select the measurements at common times between Cellular Lesion and Plasma to compute the respective AUC 24 hours. Analogously, we selected data for Caseum.

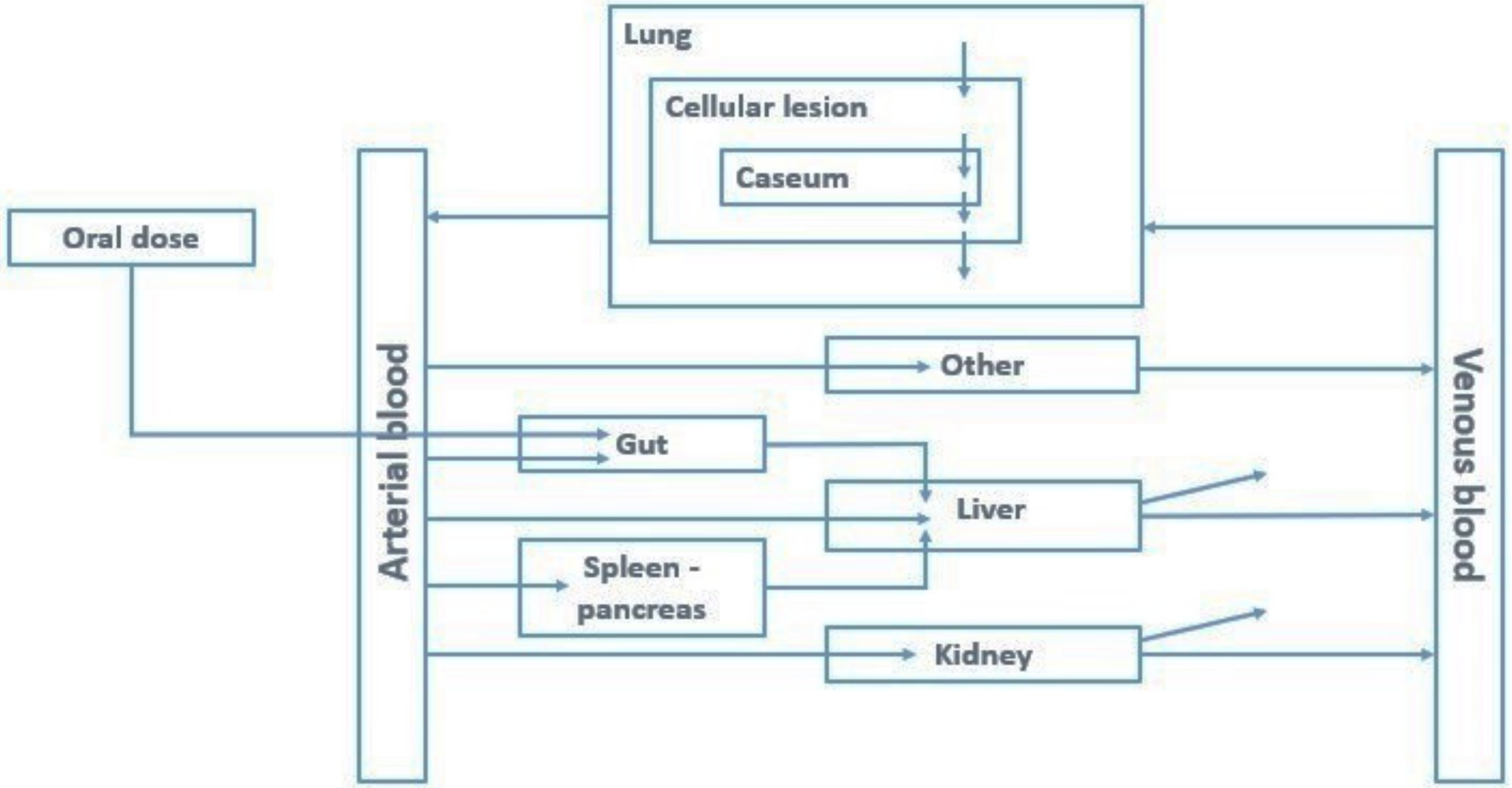
Drug		
LVX	BDQ / BDQ-M2	DprE1 inhibitor
LZD	SUT / SUT-M1	EMB
GTX	TBAJ587 / TBAJ587-M3	INH
RIF	TBAJ876 / TBAJ876-M3	TBI223
RBT	AMK	MXF
RPT	KAN	DEL / DM-6705
PZA	CFZ	RZL
PTM	CLA	

Drug: TBAJ-876 Dose: 125 mg Flat

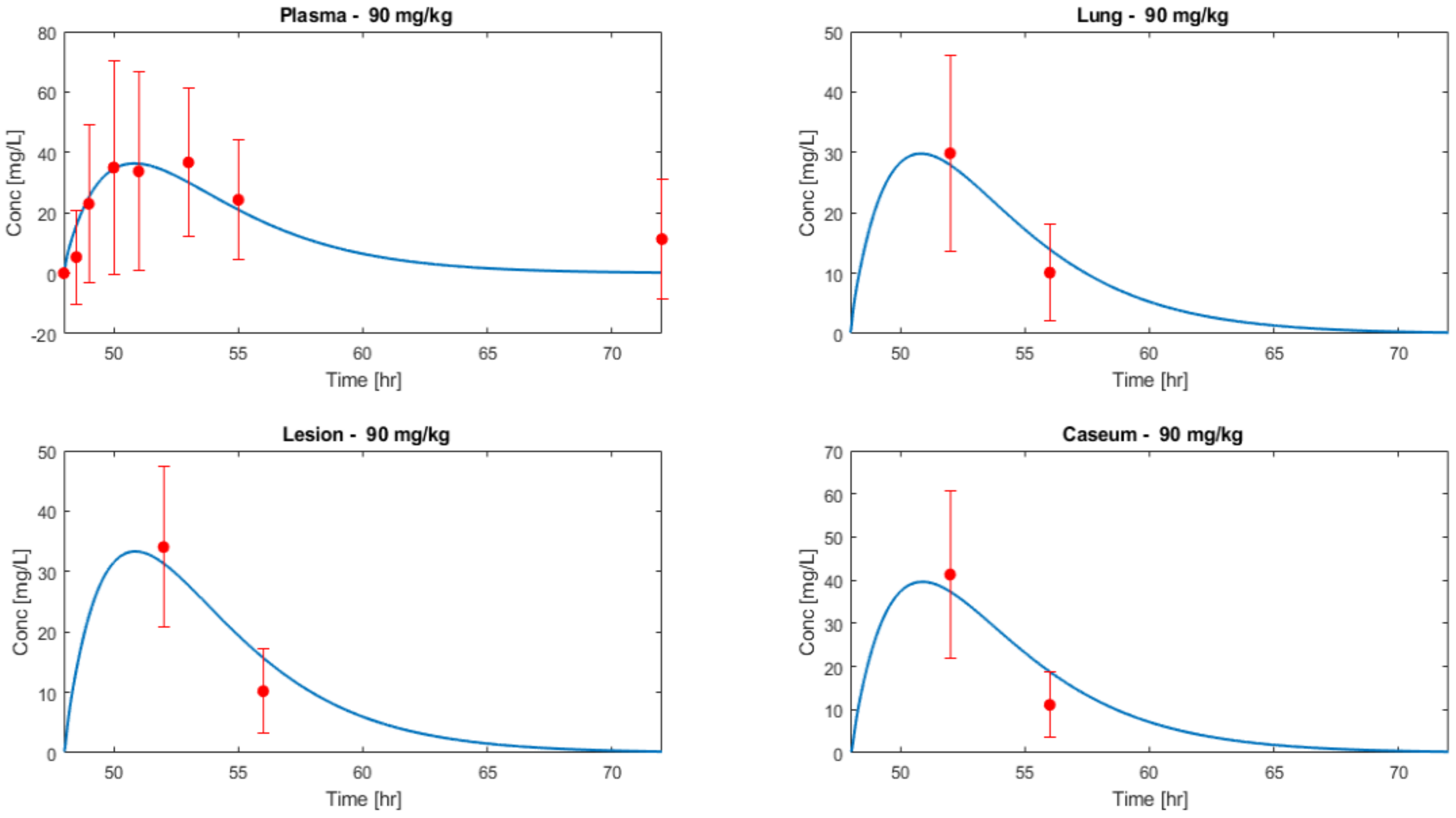


AUC improvement: mPBPK model

Minimal PBPK model to
- Track drug distribution
- Harmonize data (Time points, steady state, doses)
- Compute AUCs (24 hours)



Levofloxacin



ML model structure

Variables

Literature and Internal data + Molsoft
Literature and Internal data + Molsoft + Volsurf + QSAR module + Biovia

Target

AUC_L: $\log_{10}(AUC_{cellularLesion} / AUC_{Plasma})$
AUC_C: $\log_{10}(AUC_{Caseum} / AUC_{Plasma})$

Algorithm

- LM
- SVM
- Random forest
- Xgboost
- ANN (single and multi layer)

Model training: Cellular Lesion

Prediction model:

Structure

Variables

Target

Algorithm

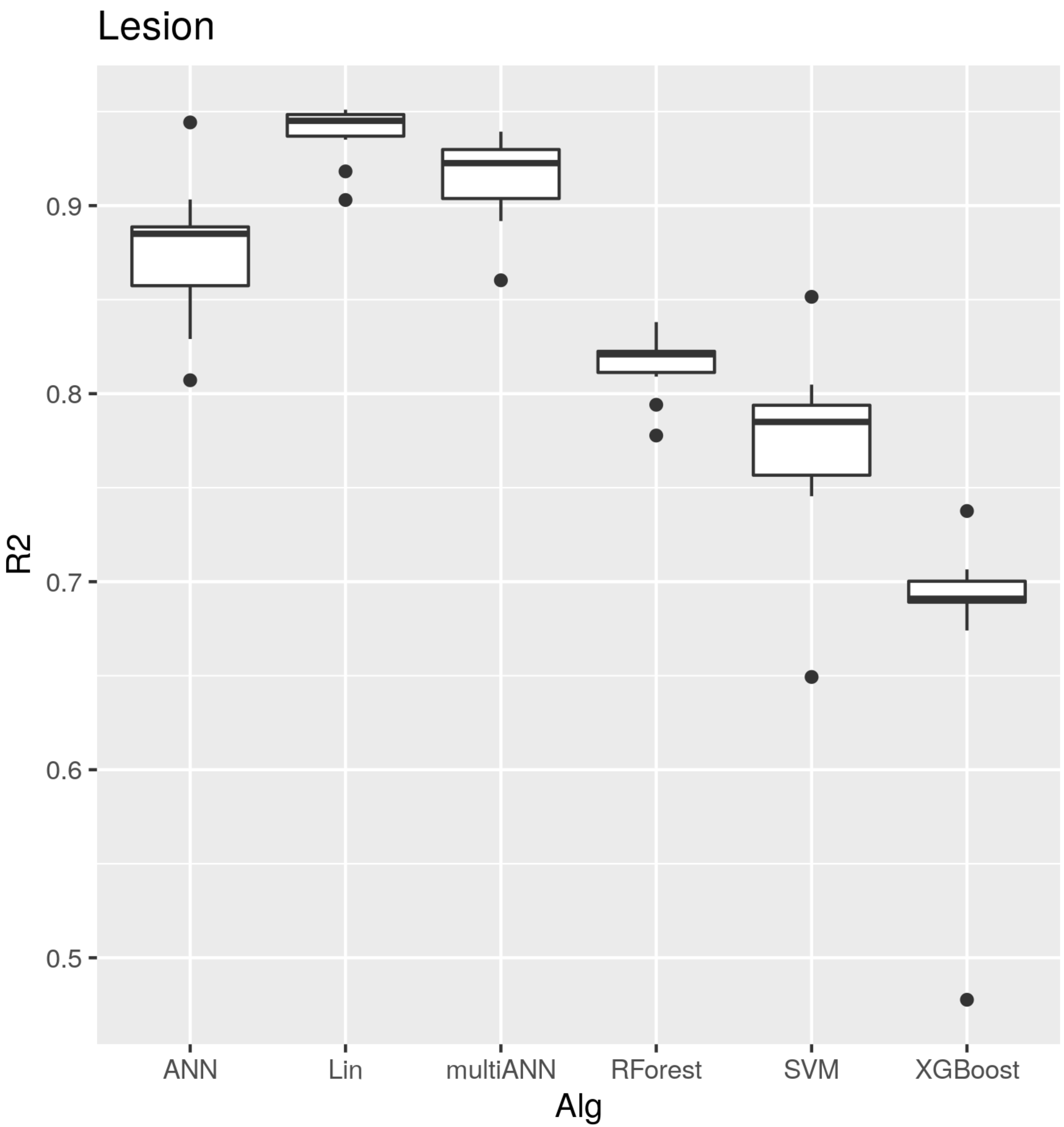
Estimation

Cross-validation

10 x 10-fold cross-validation

For each drug, we get
10 independent predictions

This provides a suggestion of the model stability



ML model definition

Prediction model:

Structure

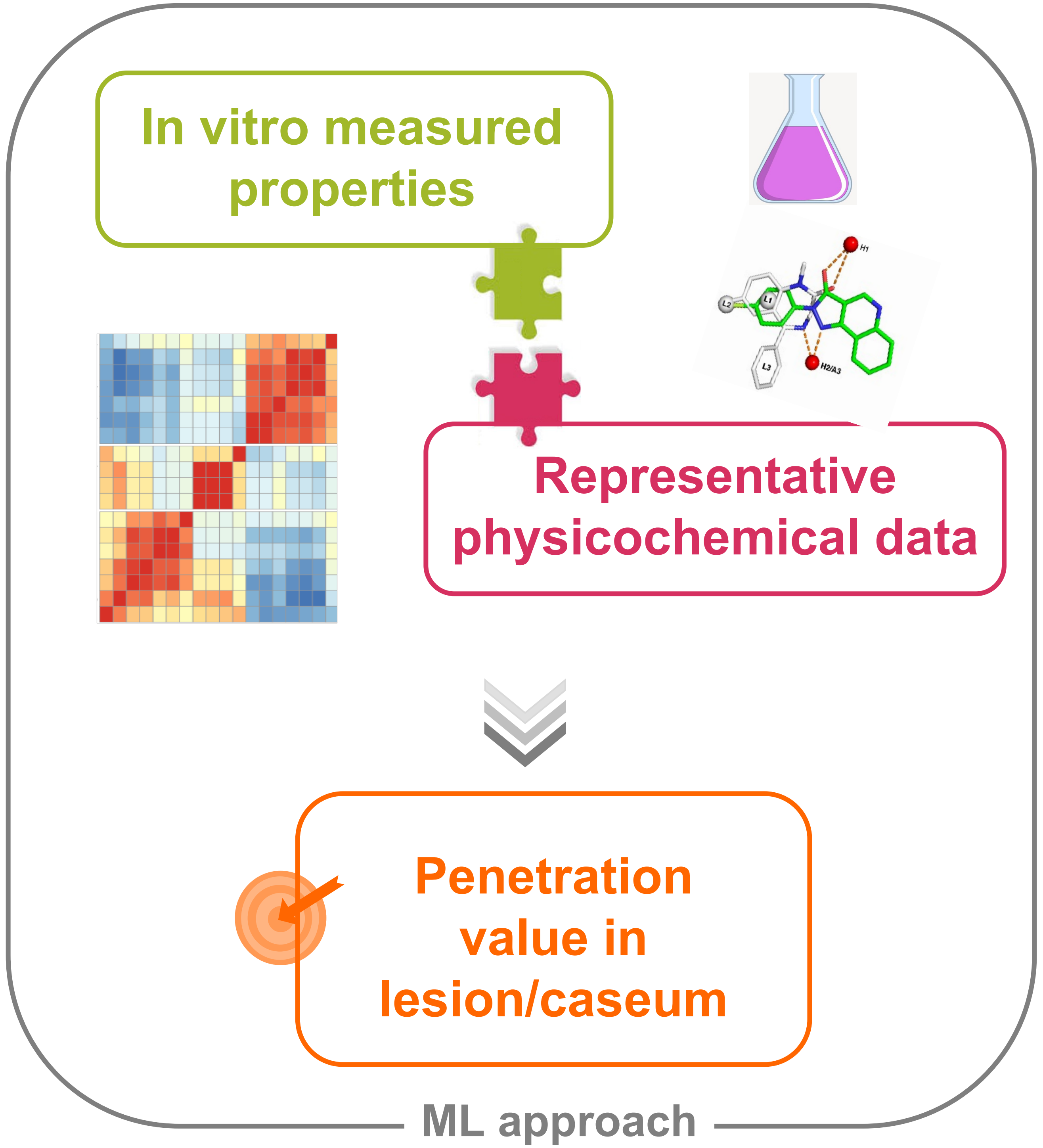
- Variables
- Target
- Algorithm

Estimation

- Cross-validation

Testing

- Accuracy & Stability



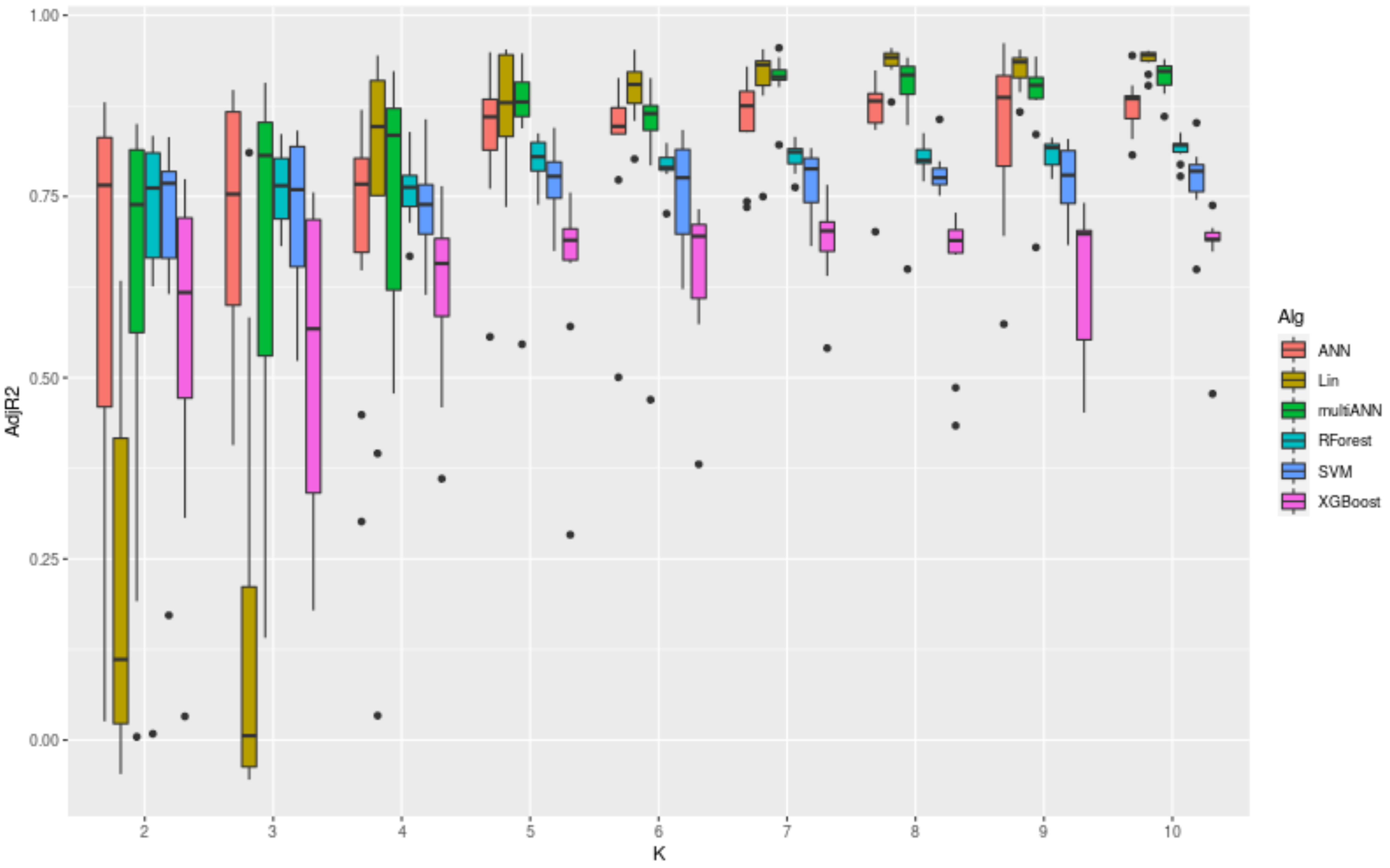
In vitro measured properties

Representative physicochemical data

Penetration value in lesion/caseum

ML approach

Model analysis: Cellular Lesion



At the varying of k, we repeated the k-fold cross-validation

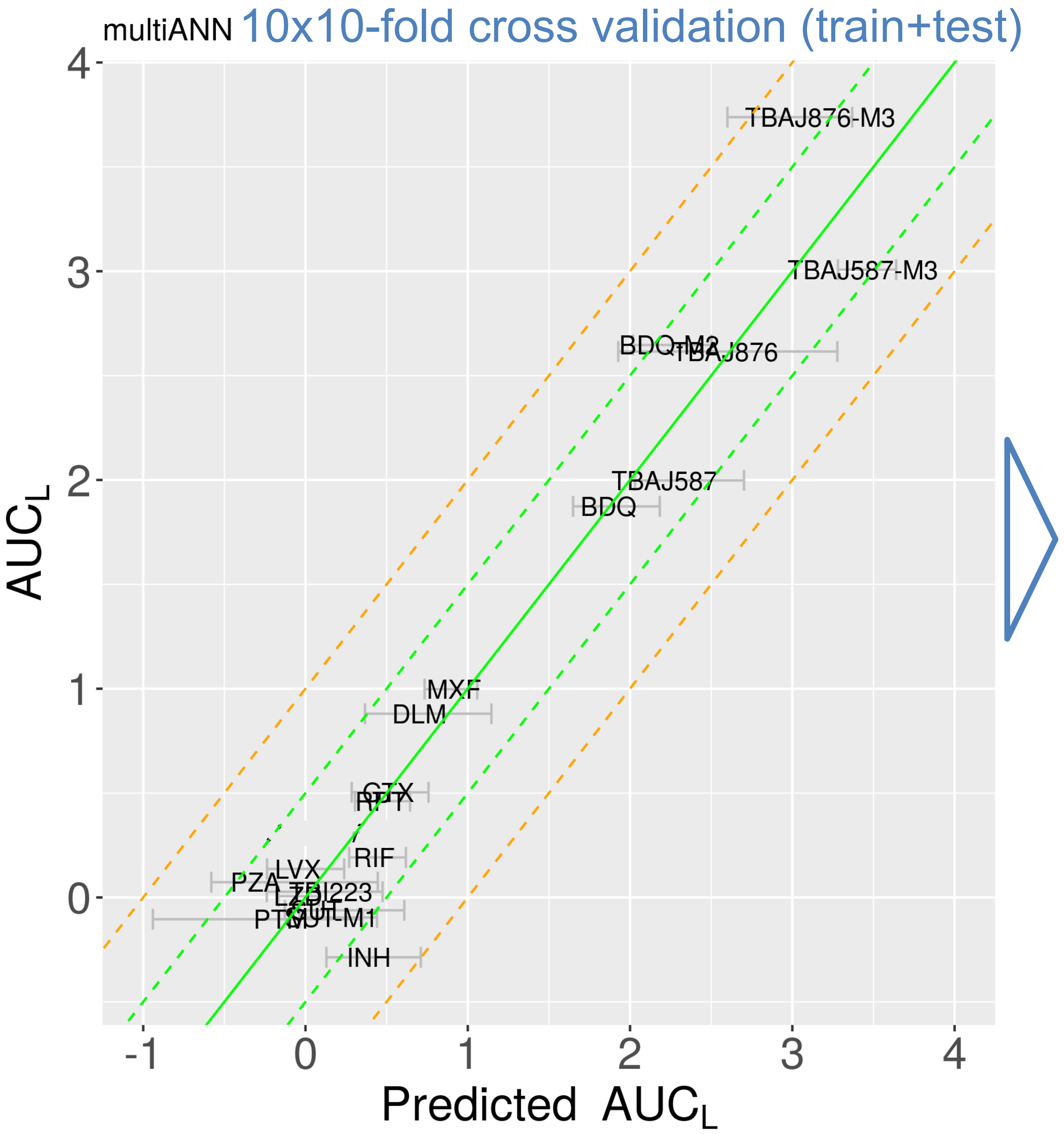
For each drug and each k, we get 10 independent predictions

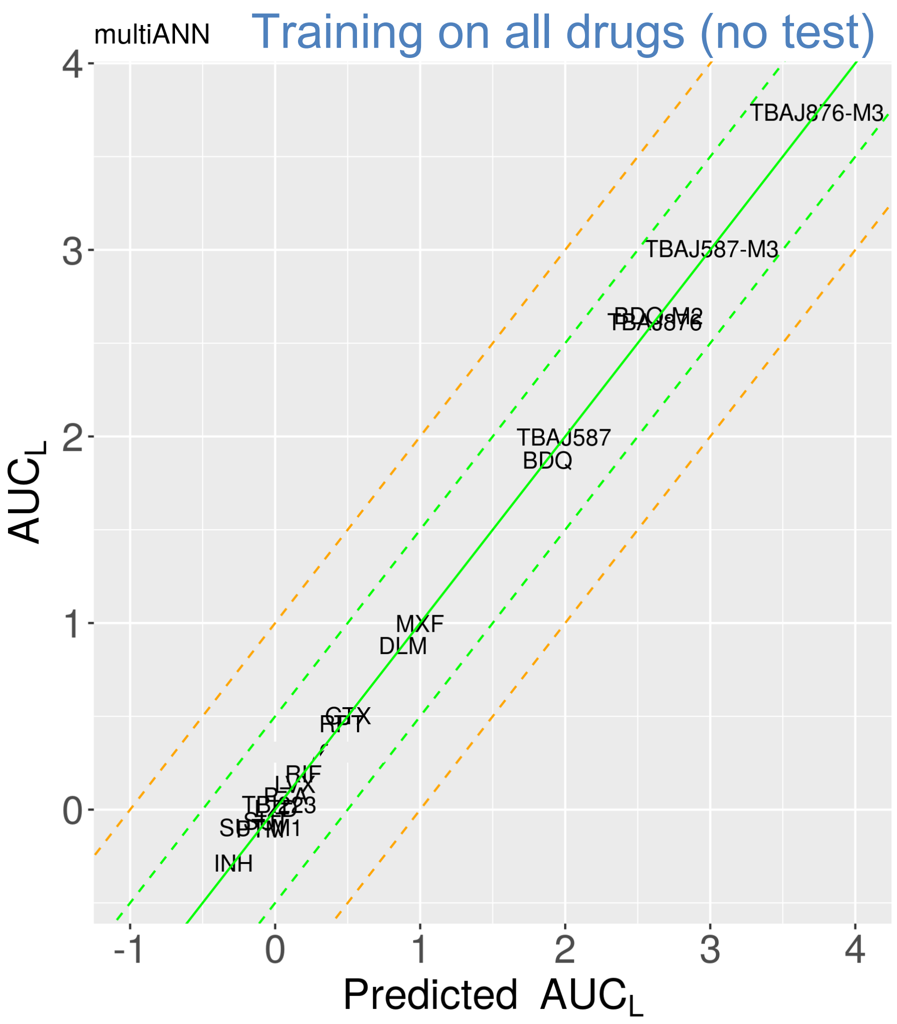
Lower values of k provide suggestions of the model response to “new” input

MultiANN is the best trade off between accuracy and stability

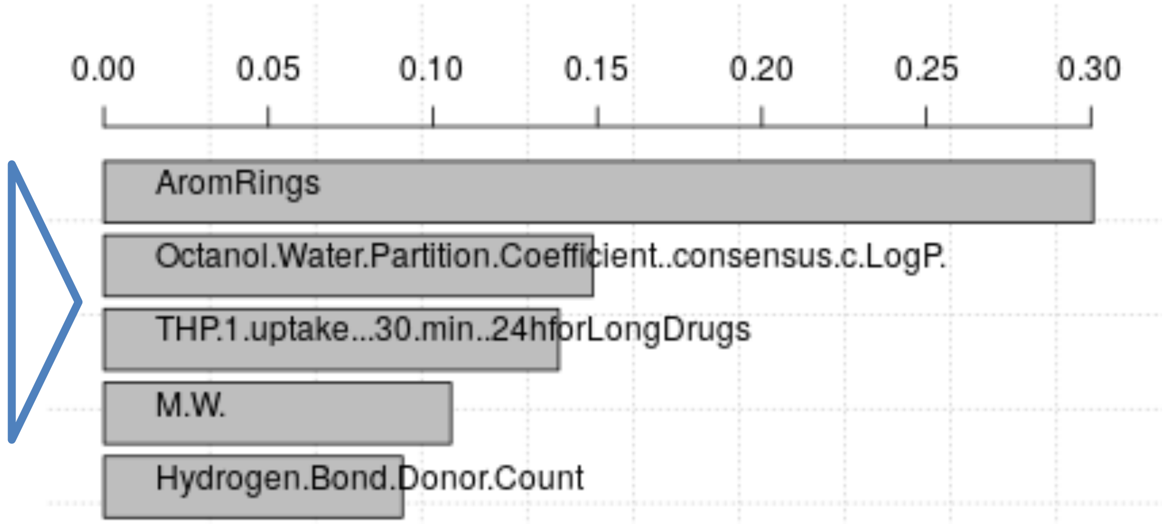
MultiANN Model analysis: Cellular Lesion

Accuracy & Stability


multiANN 10x10-fold cross validation (train+test)

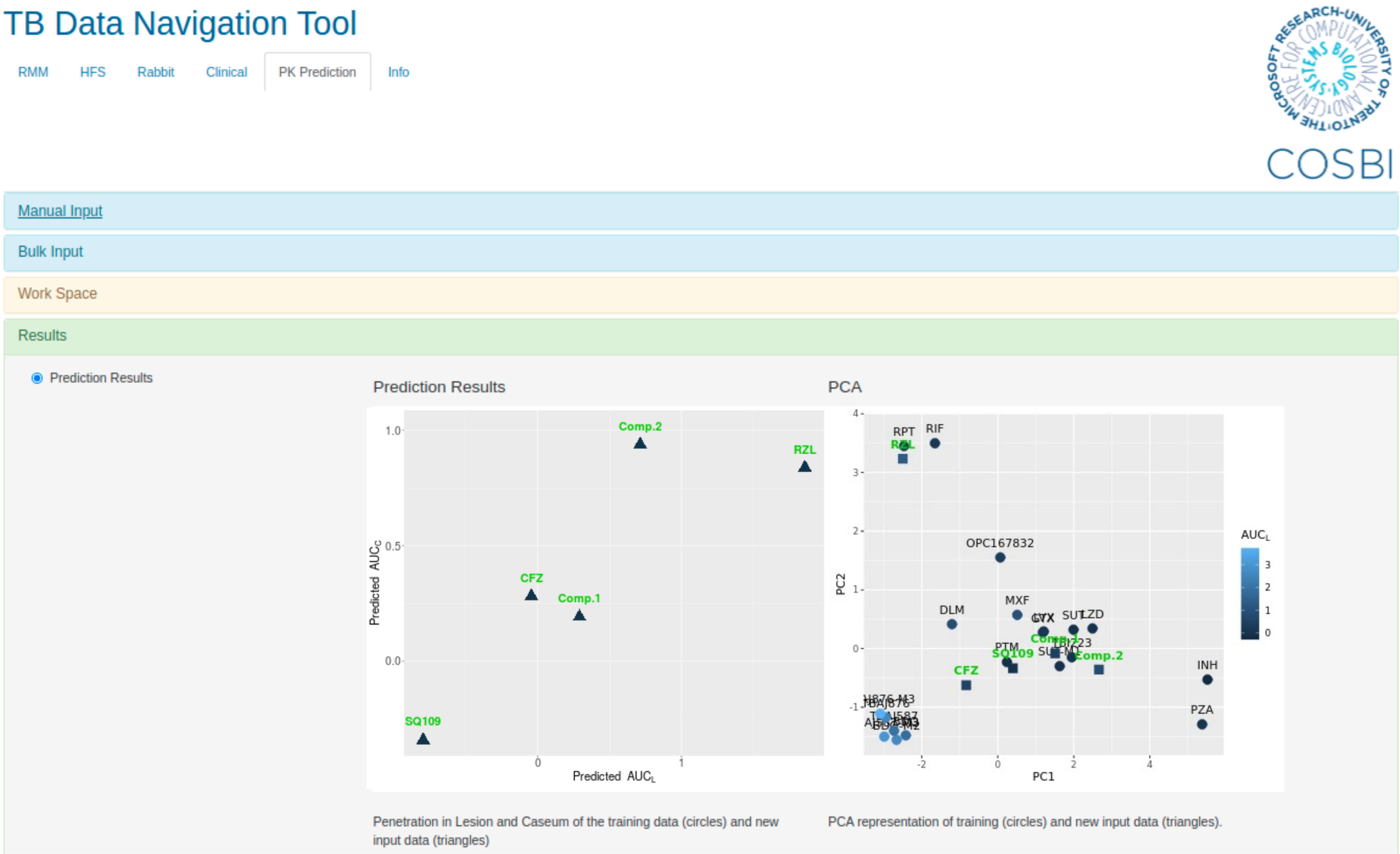

multiANN Training on all drugs (no test)

Variance based model sensitivity analysis



Same strategy for caseum

R interactive app



TB Data Navigation Tool

Penetration in Lesion and Caseum of the training data (circles) and new input data (triangles)

PCA representation of training (circles) and new input data (triangles).

Summary

- We developed a ML pipeline that takes advantage of minimal PBPK model predictions to increase the quality of the data.
- The tool requires minimal physiochemical properties and *in-vitro* inputs
- Can be used in early pre-clinical stages to prioritize new drug candidates
- We plan to:
    - Integrate it in other machine learning tools
    - Specialize the output to inform the mPBPK on the lesion/caseum dynamics

Thank you!