Using AI to Derive Valuable Insights from Drug Discovery Data

EFMC-ISMC 2022 – 6th September 2022

Matthew Segall – matt@optibrium.com

Overview

- The challenges of drug discovery data
- Introduction to deep learning imputation
- Example applications
- Conclusions

Challenges of Using Data in Drug Discovery

- It’s impossible to measure all of my compounds in all of my assays, how do I make the most of the data I have?

- I know there is variability in my experiments, how do I avoid being led astray by artefacts and errors?

- What are the most valuable experiments to run? What data will give me the most information with which to make decisions?

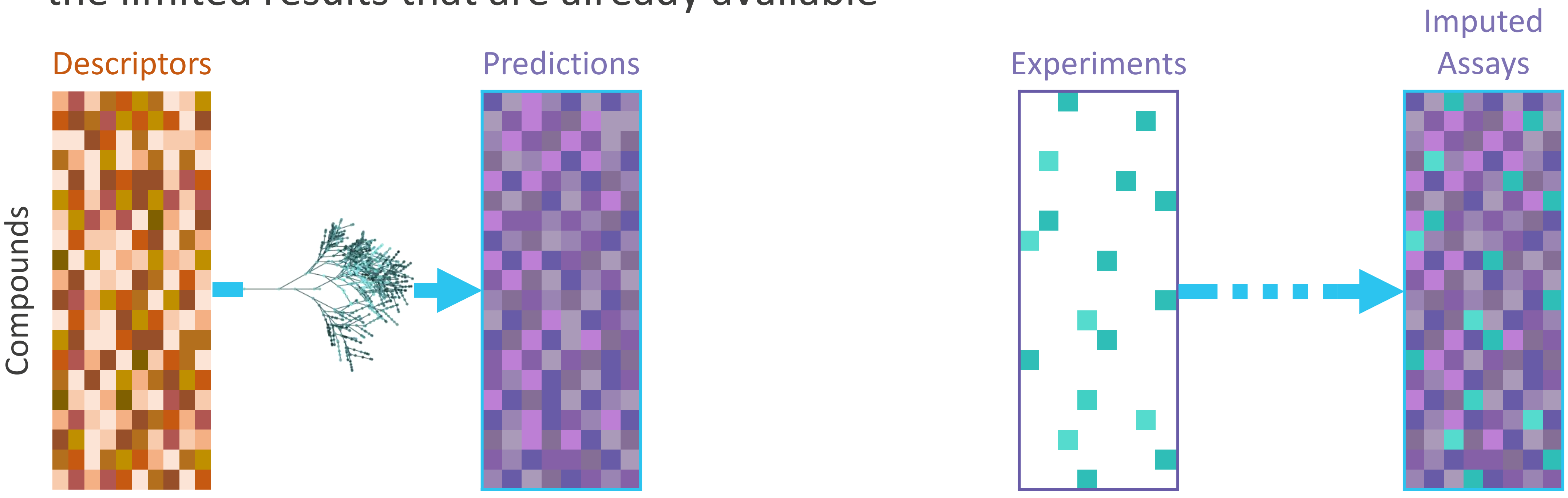
- How can I use the limited data I have to make better predictions for new compound designs, and choose the best ones for synthesis?

Introduction to Deep Learning Imputation

Prediction vs. Imputation

- Prediction uses input ‘features’ to predict one or more property values for a compound, e.g. QSAR models

- Imputation is the process of filling in the gaps in sparse experimental data using the limited results that are already available

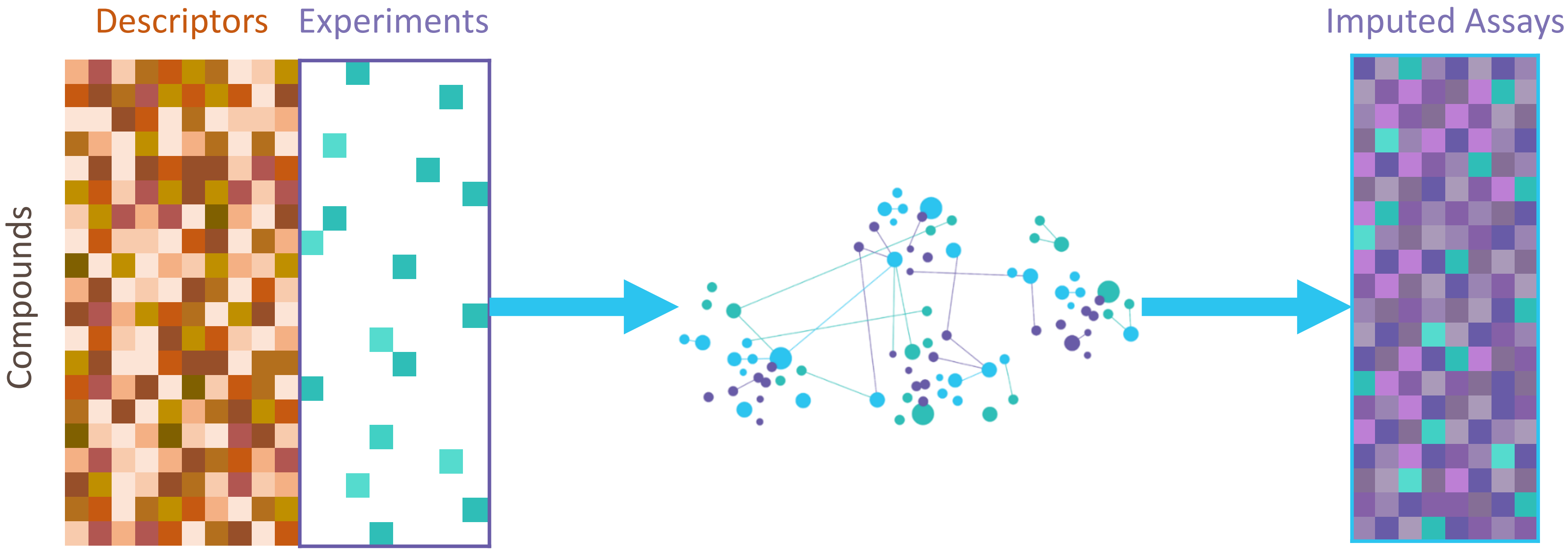


Prediction

Imputation

Alchemite™ Deep Learning Imputation
Optibrium's exclusive partnership with Intellegens

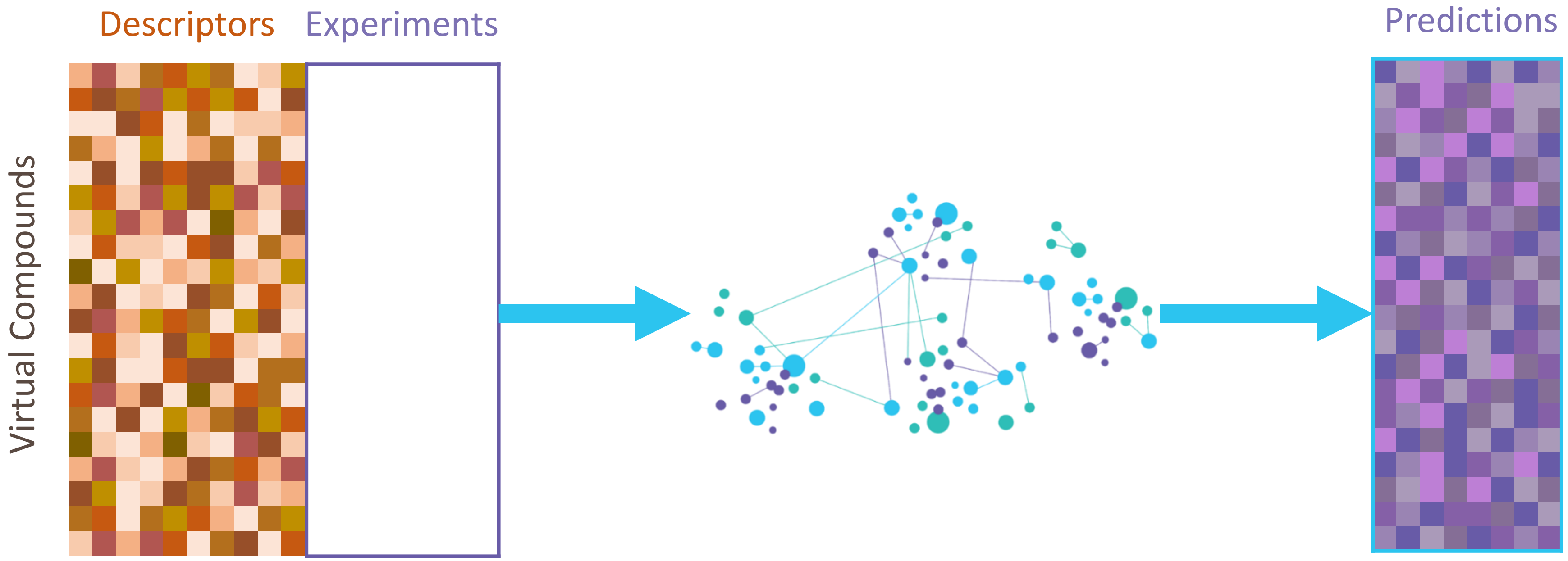
- Learns directly from relationships between experimental endpoints as well as SAR
  - Makes better use of sparse and noisy experimental data than conventional QSAR models
- 'Fills in' the gaps in your data and makes predictions for 'virtual' compounds
  - Generates more accurate predictions to target high-quality compounds



Whitehead *et al.* J. Chem Inf. Model. (2019) **59**(3) pp. 1197-1204, Irwin *et al.* J. Chem. Inf Model. (2020) **60**(6), pp. 2848–2857

Alchemite™ Deep Learning Imputation
Optibrium's exclusive partnership with Intellegens

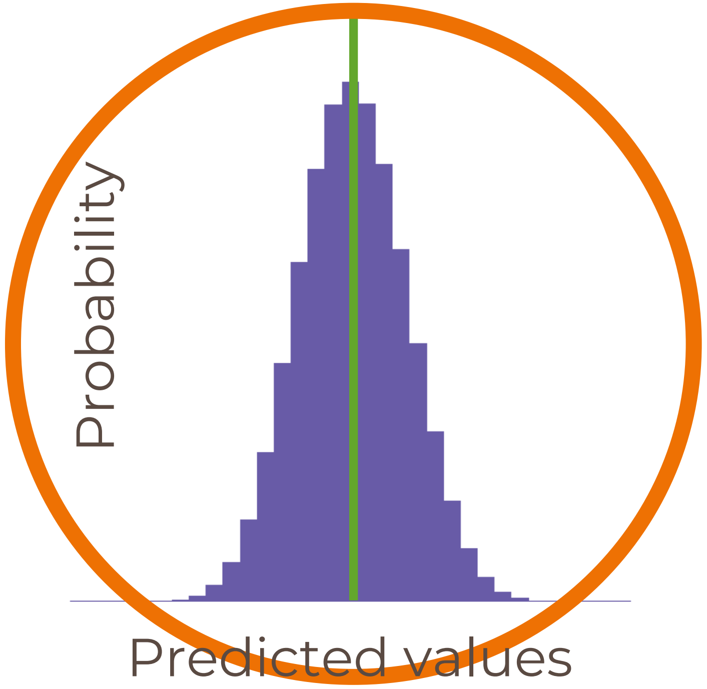
- Learns directly from relationships between experimental endpoints as well as SAR
  - Makes better use of sparse and noisy experimental data than conventional QSAR models
- 'Fills in' the gaps in your data and makes predictions for 'virtual' compounds
  - Generates more accurate predictions to target high-quality compounds

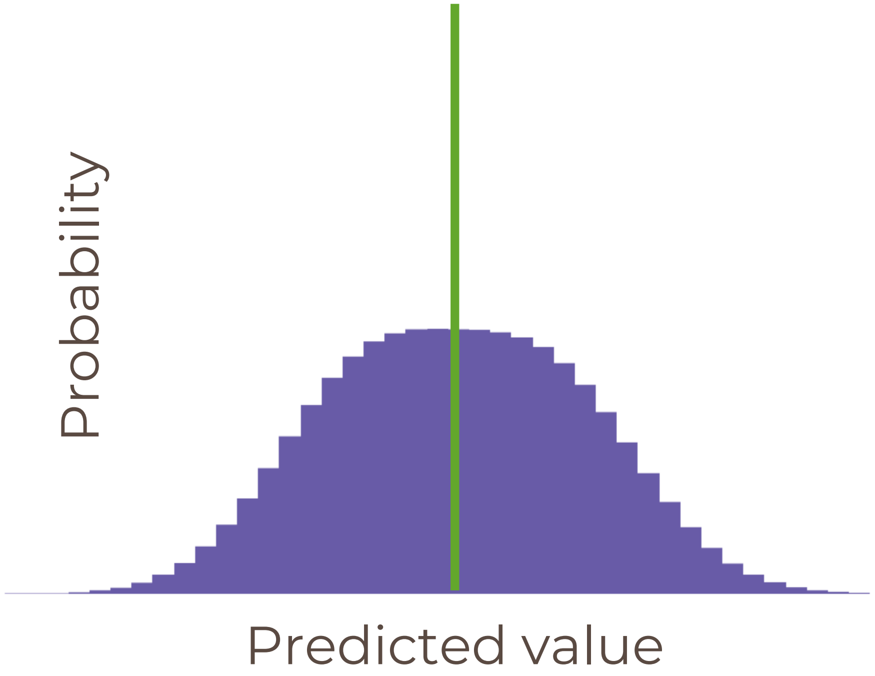


Whitehead *et al.* J. Chem Inf. Model. (2019) **59**(3) pp. 1197-1204, Irwin *et al.* J. Chem. Inf Model. (2020) **60**(6), pp. 2848–2857

Alchemite™ Deep Learning Imputation
Optibrium's exclusive partnership with Intellegens

- Estimates uncertainty in each individual prediction
  - Strong correlation between uncertainty estimates and observed accuracy on independent test sets
  - Highlights the most accurate predictions on which to base decisions
- Confidently targets high-quality compounds and prioritise experimental resources





Whitehead *et al.* J. Chem Inf. Model. (2019) **59**(3) pp. 1197-1204, Irwin *et al.* J. Chem. Inf Model. (2020) **60**(6), pp. 2848–2857

Example Applications

Applications of Deep Learning Imputation

University of Dundee

Mapping accessible property space

Constellation Pharmaceuticals

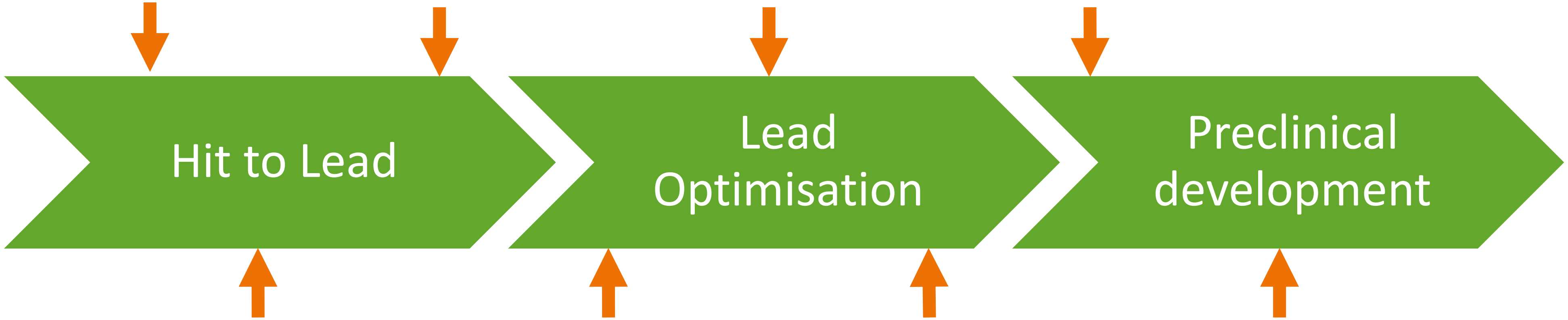
Quickly improve model quality

Genentech
A Member of the Roche Group

Efficient selectivity profiling

AstraZeneca

Predicting *in vivo* PK



Exploring novel chemical space

OSM

Prospective activity prediction

Takeda | ONCOLOGY

Predicting secondary activity assays

Constellation Pharmaceuticals

Predicting clinical safety from preclinical data

Non-pharma applications:

Imputation of *in vivo* sensory properties

Prediction of agrochemical bioactivity profiles

iff

FMC

Watch our webinar at https://bit.ly//AI-solutions-webinar

optibrium™

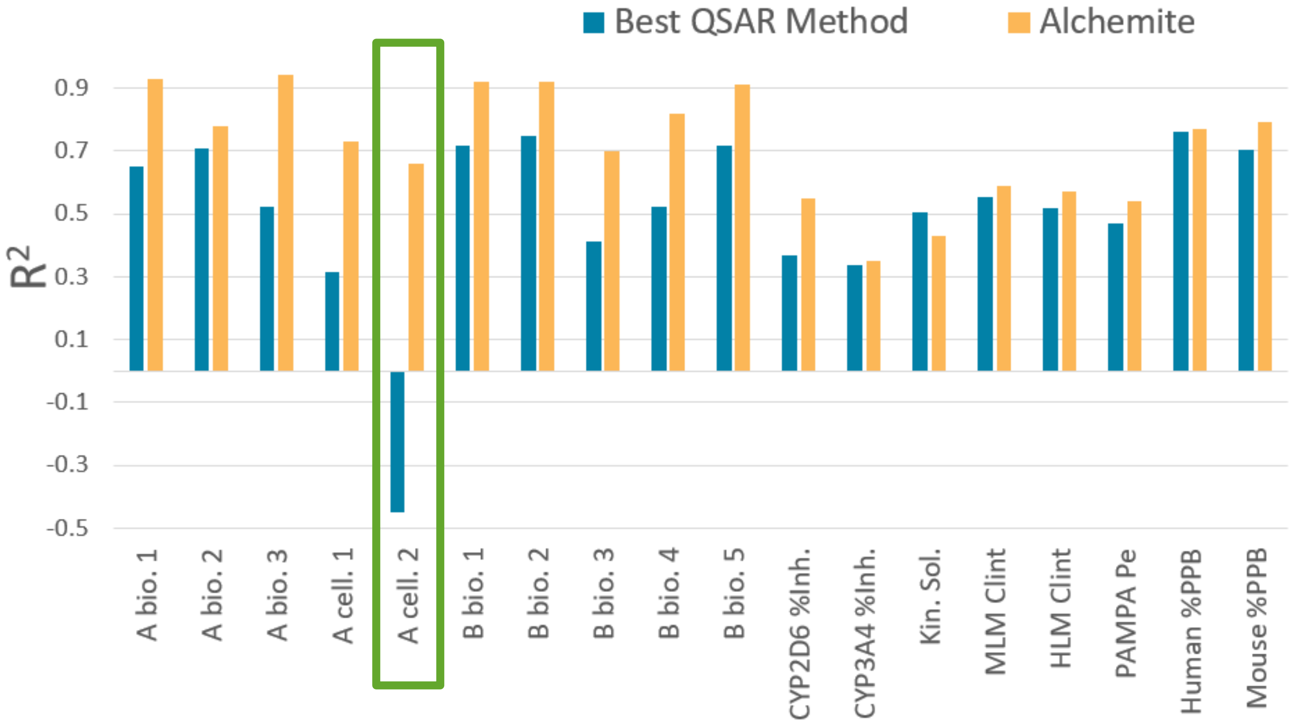
Alchemite Application to Project Data

- Application to **heterogeneous** data across two projects
  - Target and phenotypic activities and ADME endpoints
  - 2453 compounds across 18 endpoints
- Significant improvement in accuracy

	Average R^2
Best QSAR	0.50
Alchemite™	**0.72**

- Example of value delivered:
  - Few false negatives among confidently-predicted inactives – could have saved 24 FTE-months in unnecessary synthesis

Constellation PHARMACEUTICALS



Irwin *et al.* J. Chem. Inf Model. (2020) **60**(6), pp. 2848–2857
Watch our webinar: http://bit.ly/practical_deeplearning

Collaboration with Open Source Malaria (OSM)

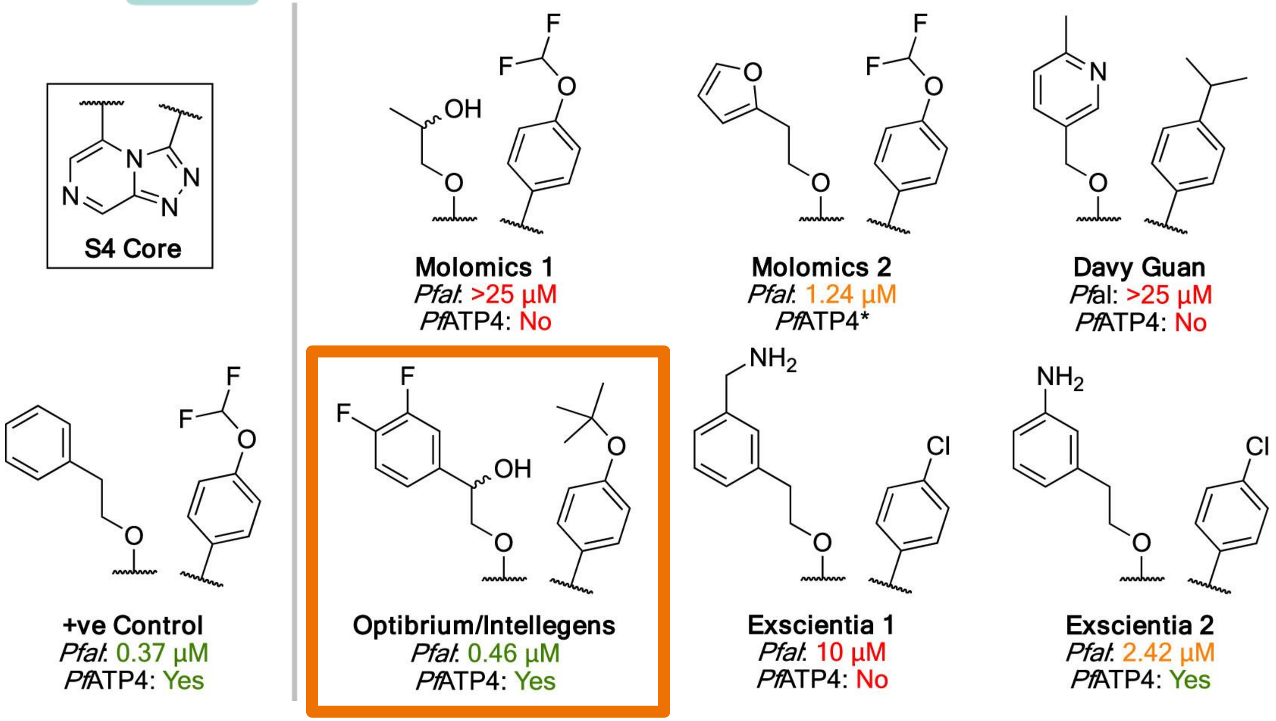
Combination with generative methods

- Application to **sparse antimalarial activity data**
  - Targeting novel MoA – *Pf*ATP4
  - Alchemite generated one of the top-ranked models
- New compound ideas were generated using the Nova™ module in StarDrop™
  - Prioritised with Alchemite model
  - Good activity profile and properties
- A **confidently** predicted compound was synthesised and tested by OSM
  - **Only confirmed active** of those proposed by four organisations
- "[this] suggestion… was thought by the human team to be a certain inactive… yet this compound displayed good potency and is a particularly useful outcome (i.e., **the "Machine Overlords" class**)"*

Watch our webinar http://bit.ly/ai_antimalarials

*Tse et al. J. Med. Chem. (2021) **64**(22) pp 1645-16463

FINAL SELECTIONS



S4 Core

Molomics 1
Pfal: >25 µM
*Pf*ATP4: No

Molomics 2
Pfal: 1.24 µM
*Pf*ATP4*

Davy Guan
Pfal: >25 µM
*Pf*ATP4: No

+ve Control
Pfal: 0.37 µM
*Pf*ATP4: Yes

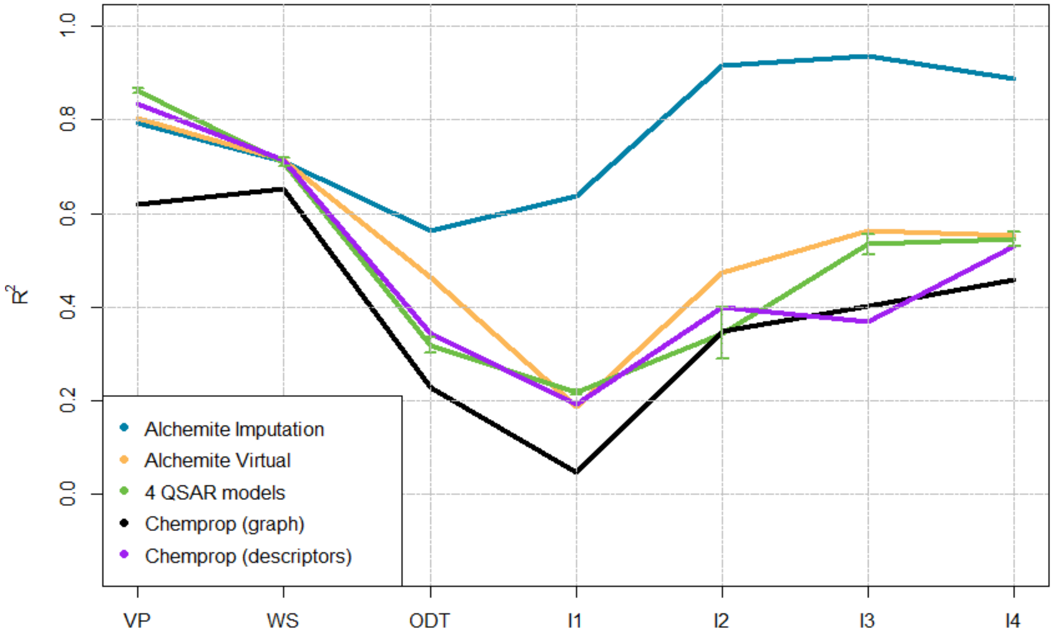
Optibrium/Intellegens
Pfal: 0.46 µM
*Pf*ATP4: Yes

Exscientia 1
Pfal: 10 µM
*Pf*ATP4: No

Exscientia 2
Pfal: 2.42 µM
*Pf*ATP4: Yes

Imputation of Sensory Properties

- Sensory properties are measured in panels of human subjects
  - Expensive and subjective
  - Noisy data
- Deep learning imputation is more accurate than QSAR methods
  - Including multi-target deep neural networks
- Accurate prediction of activity cliffs that are missed by QSAR methods
  - Small changes in structure that drive a large change in property

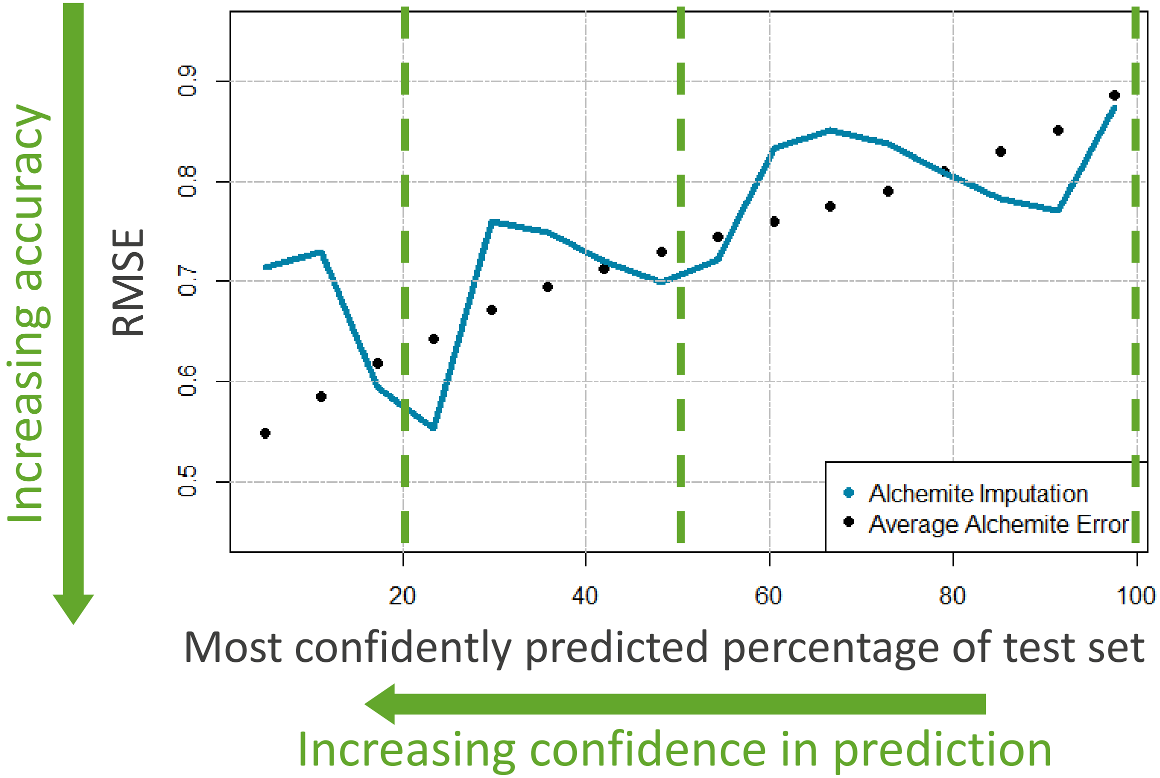


Mahmoud *et al.* J. Comput. Aided Mol. Des. (2021) **35**(11) pp. 1125-1140

Watch our webinar https://bit.ly/SensoryWebinar

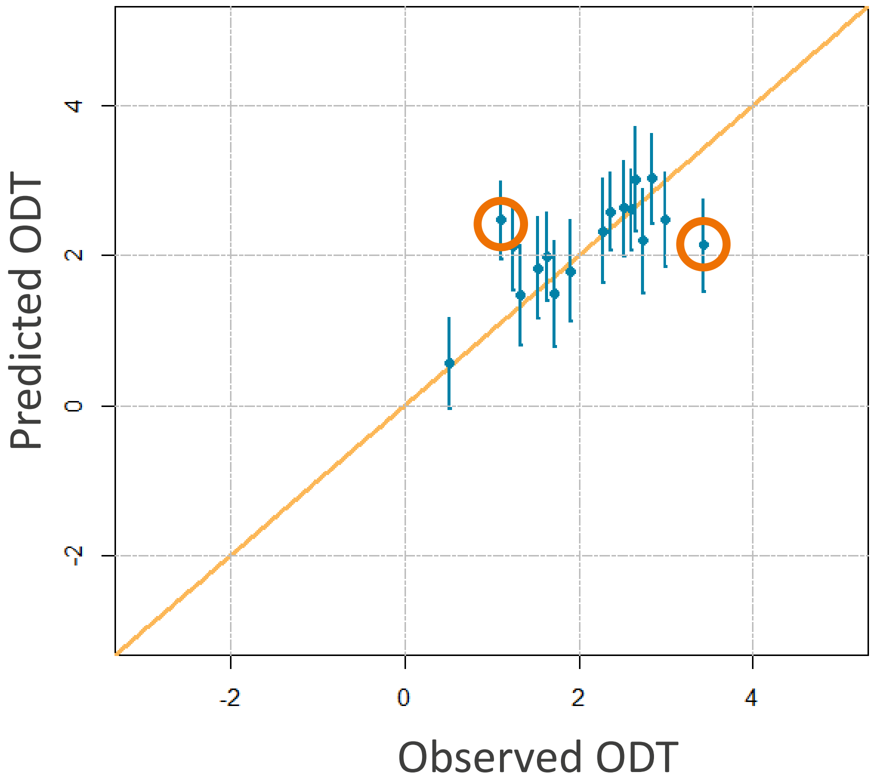
Focusing on the Most Confident Results

ODT Endpoint



Increasing accuracy

Increasing confidence in prediction



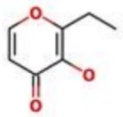
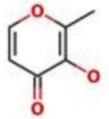
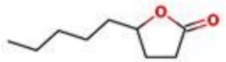
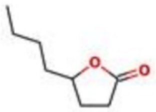
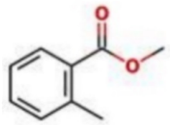
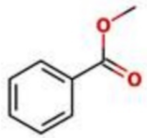
- Excellent correlation between model confidence (error bars) and observed accuracy
- The model can reliably identify the most accurate predictions
- Identify experimental outliers for retest

Imputation of Sensory Properties

- Sensory properties are measured in panels of human subjects
  - Expensive and subjective
  - Noisy data

- Deep learning imputation is more accurate than QSAR methods
  - Including multi-target deep neural networks

- Accurate prediction of activity cliffs that are missed by QSAR methods
  - Small changes in structure that drive a large change in property

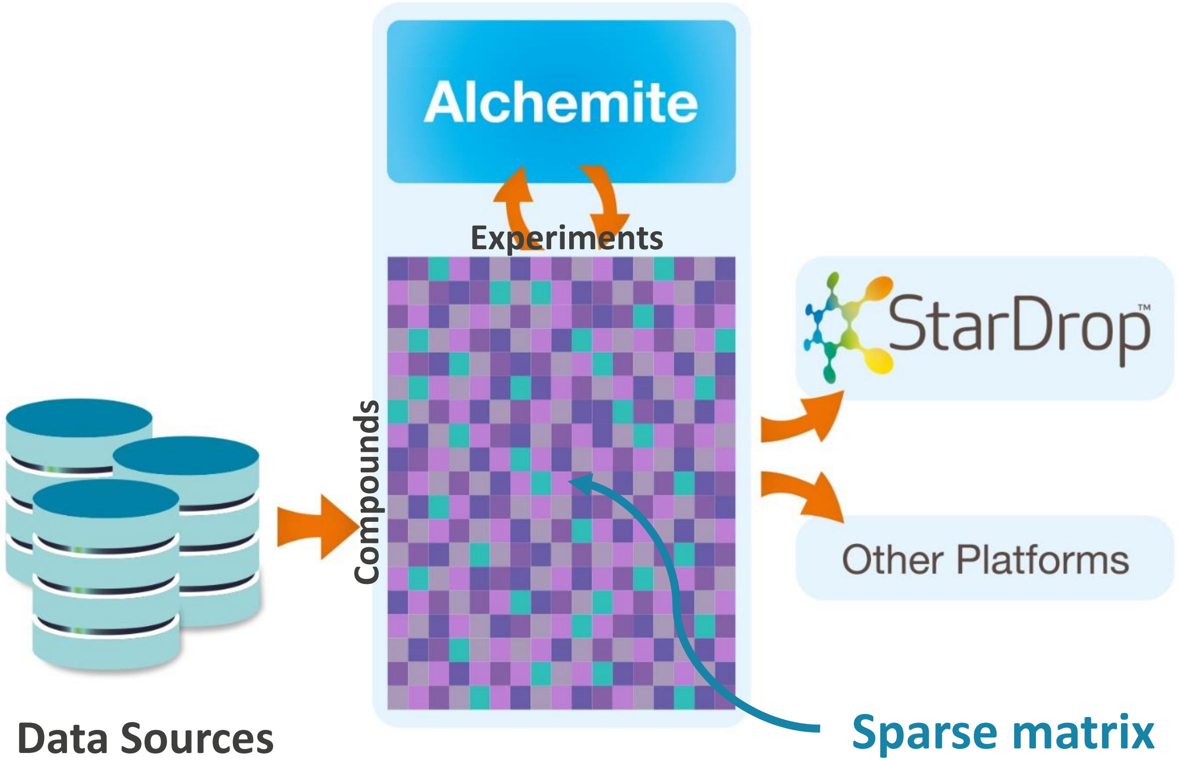
Structure of Nearest Neighbour in Training Set	Observed ODT of Nearest Neighbour in Training Set	Test Compound	Observed ODT of Test Compound	Predicted ODT by the Alchemite Imputation model	Predicted ODT by the best QSAR model
	0.93		-0.6	-0.7	1.43
	1.93		-0.12	-0.05	0.98
	0.52		1.99	1.89	0.89

Mahmoud *et al.* J. Comput. Aided Mol. Des. (2021) **35**(11) pp. 1125-1140
Watch our webinar: https://bit.ly/SensoryWebinar

Cerella
Enabling active learning in drug discovery

Cerella™
an Augmented Chemistry® software product

- **Automatically** updates and prepares experimental data for model building
  - Connects seamlessly to data repositories
  - Applies cleaning, business rules and transformations to data for best model performance
- **Automatically** updates Alchemite models as new data become available
  - Always work with results based on the latest information
  - Remove the burden of manually building and updating models
- Manage 'massive matrix' of imputed results for easy access
  - May contain O(10^{10}) data points!
- Provide seamless access to results
  - Using StarDrop™ or any platform via a RESTful API



Data Sources

- Data mart
- ELN
- …

Watch our webinar at http://bit.ly/cerella_active

optibrium™

Conclusions
Reducing the time and cost of discovery cycles

Deep learning imputation gains more value than prediction from experimental data than conventional compounds

- Proactively **highlight high-quality compounds** by more accurately ‘filling in’ sparse data (imputation)
- **Increase confidence** in decision making, identify **hidden opportunities,** flag outliers and false negatives
- Translate AI insights into planning of experiments and **focus on the most valuable measurements**
- Gain more value from your compound data, **accurately predicting complex endpoints**, intractable with conventional QSAR modelling

For more information: www.optibrium.com, matt@optibrium.com or booth #35

Acknowledgements

- Optibrium
  - Samar Mahmoud
  - Mario Öeren
  - Benedict Irwin (now at MS Research)
  - Alexander Wade (University of Cambridge)
- Intellegens
  - Gareth Conduit
  - Tom Whitehead
- Prof. Matthew Todd, Dr Edwin Tse and the rest of the Open Source Malaria team
- Takeda
  - Scott Rowland

References

- Imputation of Assay Bioactivity Data Using Deep Learning
  - Whitehead *et al.* J. Chem Inf. Model. (2019) **59**(3) pp. 1197-1204
- Practical Applications of Deep Learning to Impute Heterogeneous Drug Discovery Data
  - Irwin *et al.* J. Chem. Inf. Model. (2020) **60**(6), pp. 2848–2857
- Guiding Drug Optimisation Using Deep Learning Imputation and Compound Generation
  - Irwin *et al.* Int. Pharm. Ind. (2020) **12**(2) pp. 28-31
- Deep Imputation on Large-Scale Drug Discovery Data
  - Irwin *et al.* App. AI Lett. (2021) **2**(3) p. e31 DOI: 10.1002/ail2.31
- Imputation of Sensory Properties Using Deep Learning
  - Mahmoud *et al.* J. Comput. Aided Mol. Des. (2021) **35**(11) pp. 1125-1140
- An Open Drug Discovery Competition: Experimental Validation of Predictive Models in a Series of Novel Antimalarials
  - Tse *et al.* J. Med. Chem. (2021) **64**(22) pp 1645-16463
- Prediction of In Vivo Pharmacokinetic Parameters and Time–Exposure Curves in Rats Using Machine Learning from the Chemical Structure
  - Obrezanova *et al.* Mol. Pharm. (2022) DOI: 10.1021/acs.molpharmaceut.2c00027

Webinars

- Practical Applications of Deep Learning to Imputation of Drug Discovery Data
  - http://bit.ly/practical_deeplearning
- Large Scale Imputation of Drug Discovery Data using Deep Learning
  - http://bit.ly/largescale_imputation
- A Global Deep Learning Model for Global Health Drug Discovery
  - http://bit.ly/deep_learning_global
- AI-guided Design of Antimalarials with In Vitro Validation
  - http://bit.ly/ai_antimalarials
- Predicting Pharmacokinetic Parameters and Curves
  - http://bit.ly/pk_prediction_az
- Optimising Kinase Profiling Programmes with Deep Learning
  - https://bit.ly/deep_learning_kinase_profiling
- Imputation of Sensory Properties Using Deep Learning
  - https://bit.ly/SensoryWebinar
- AI Solutions from Hit to Candidate
  - https://bit.ly//AI-solutions-webinar