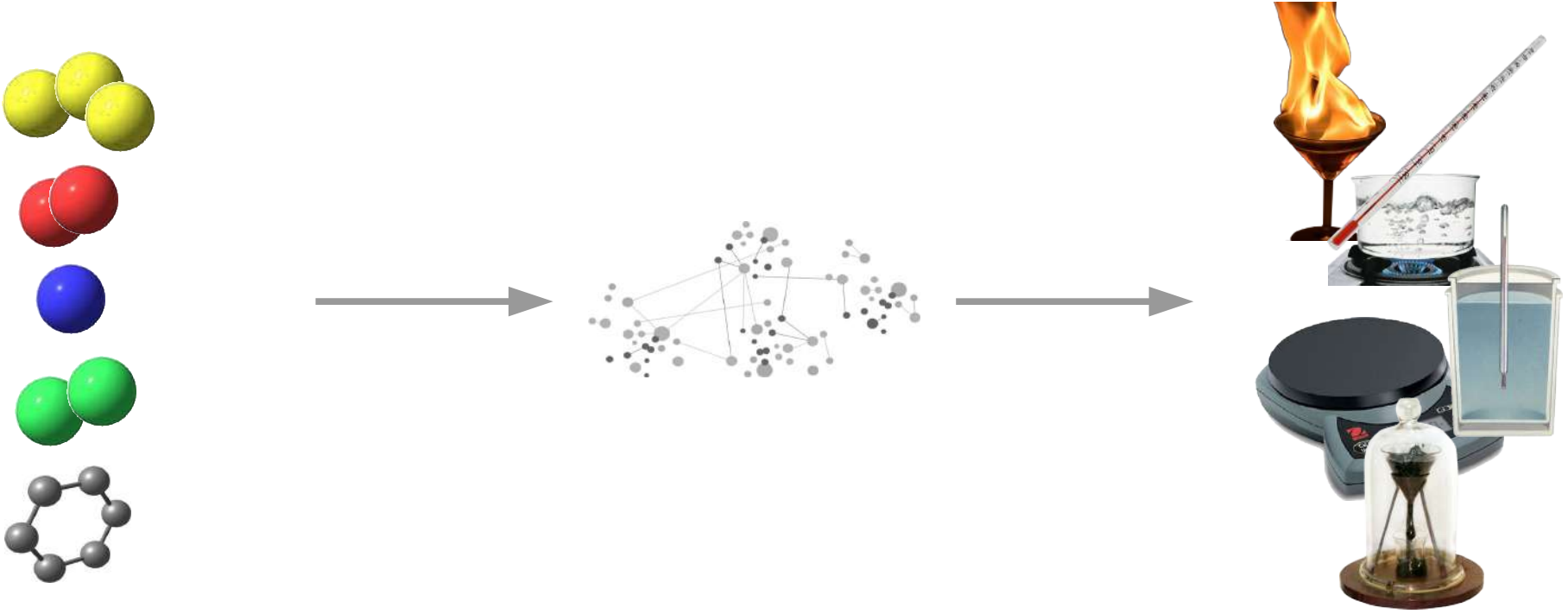
intellegens

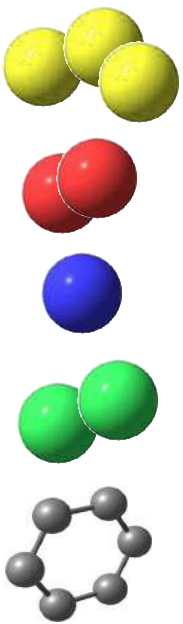
Imputation of assay activity data using deep learning

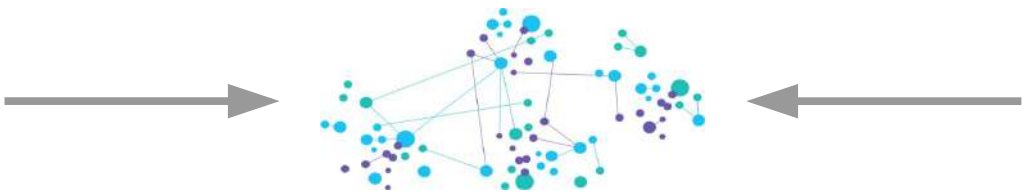
Gareth Conduit

Machine learning as a black box

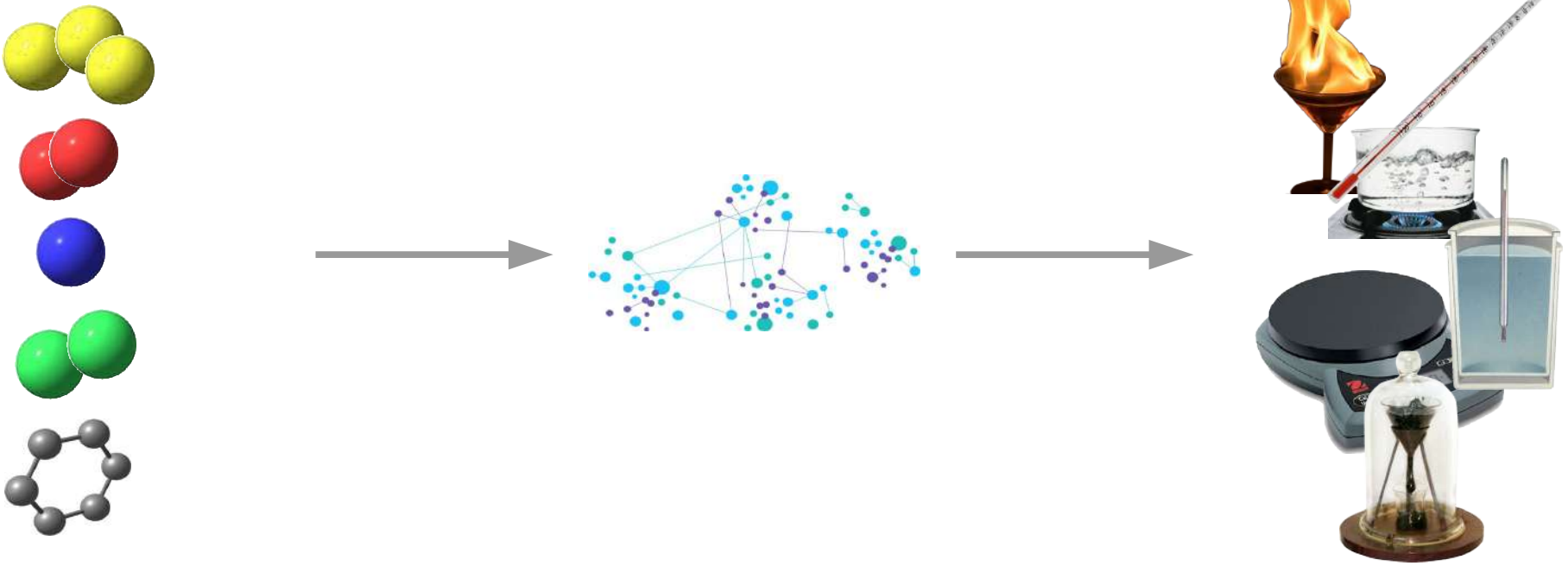


Train from historical data





Predict new chemicals



Alchemite™ machine learning tool to

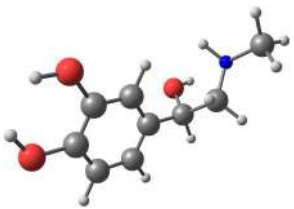
Reduce the need for experiments and accelerate discovery

Utilise all available information: computer simulations and real-life measurements

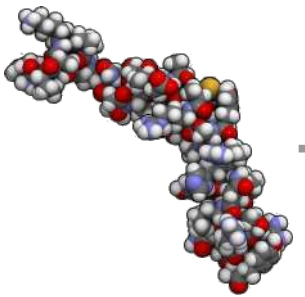
Impute values from sparse data

Broadly applicable with proven applications in drug design, industrial chemicals, and materials

Action of a drug



Drug

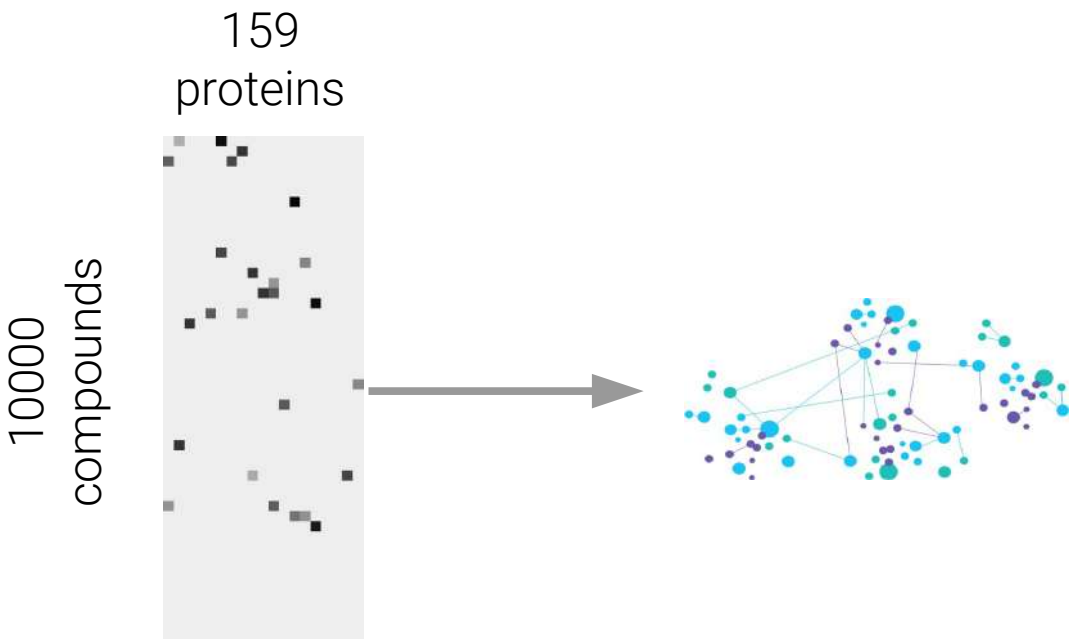


Protein

Effect

Novartis dataset to benchmark machine learning

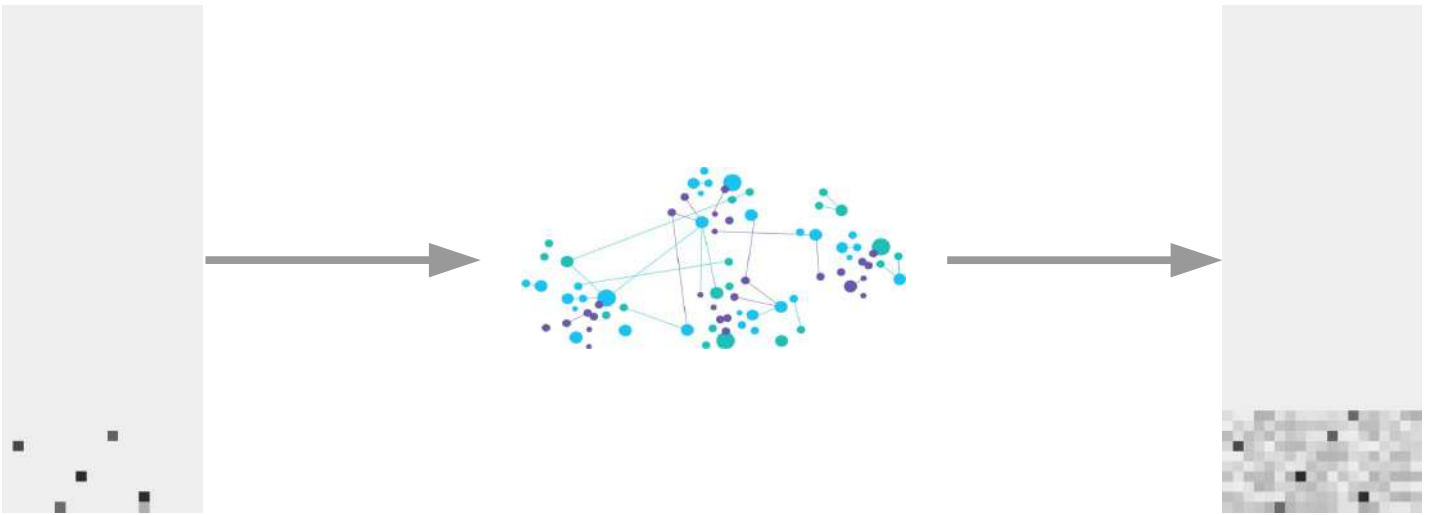
159 kinase proteins, 10000 compounds, data 5% complete



Data from ChEMBL
Martin, Polyakov, Tian, and Perez,
J. Chem. Inf. Model. 57, 2077 (2017)

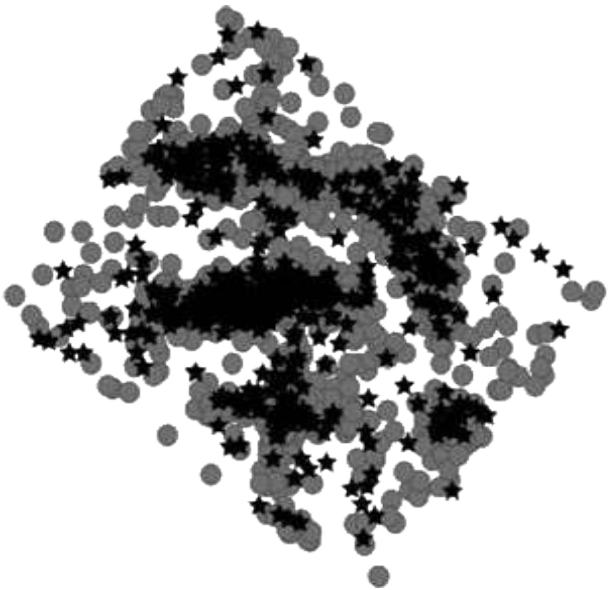
Validate imputation of missing entries

Realistically split holdout data set, extrapolate to new chemical space



Impute missing entries in new chemical space

Random



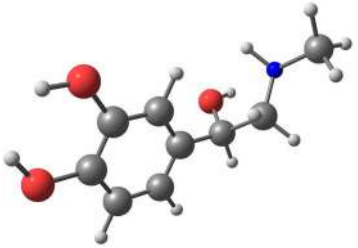
Realistic

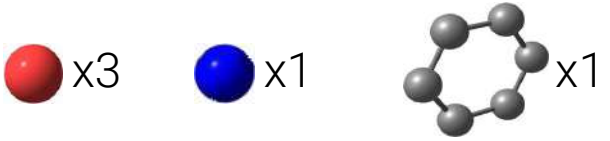
● Training

★ Validation

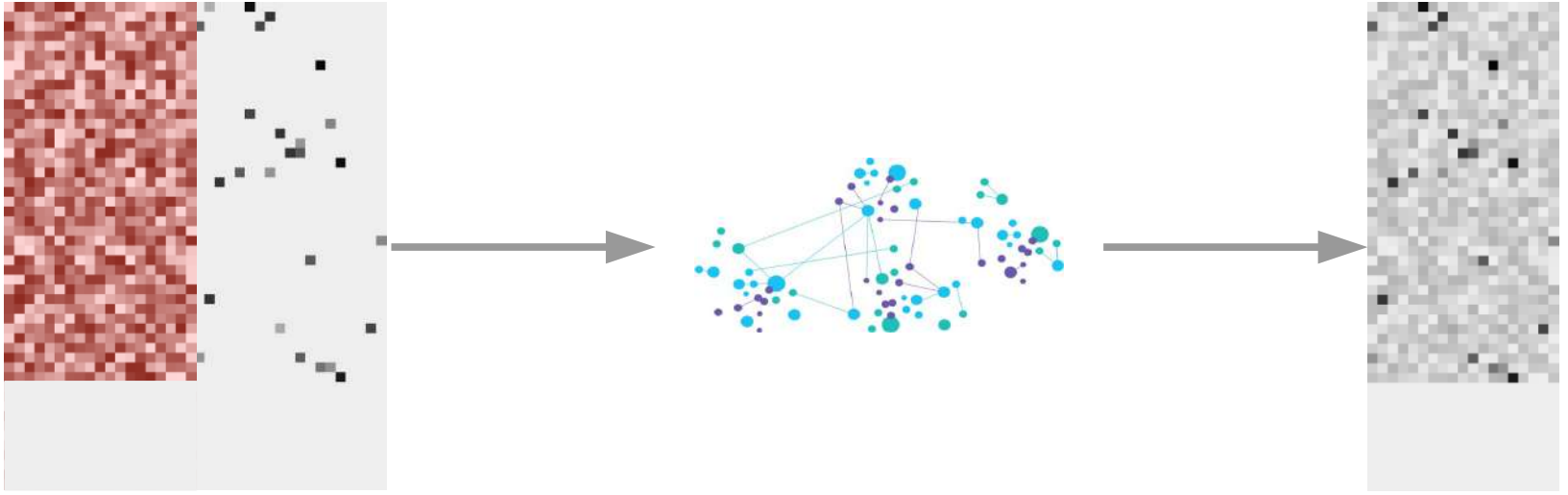
Data from ChEMBL
Martin, Polyakov, Tian, and Perez,
J. Chem. Inf. Model. 57, 2077 (2017)

QSAR: quantitative structure-activity relationships



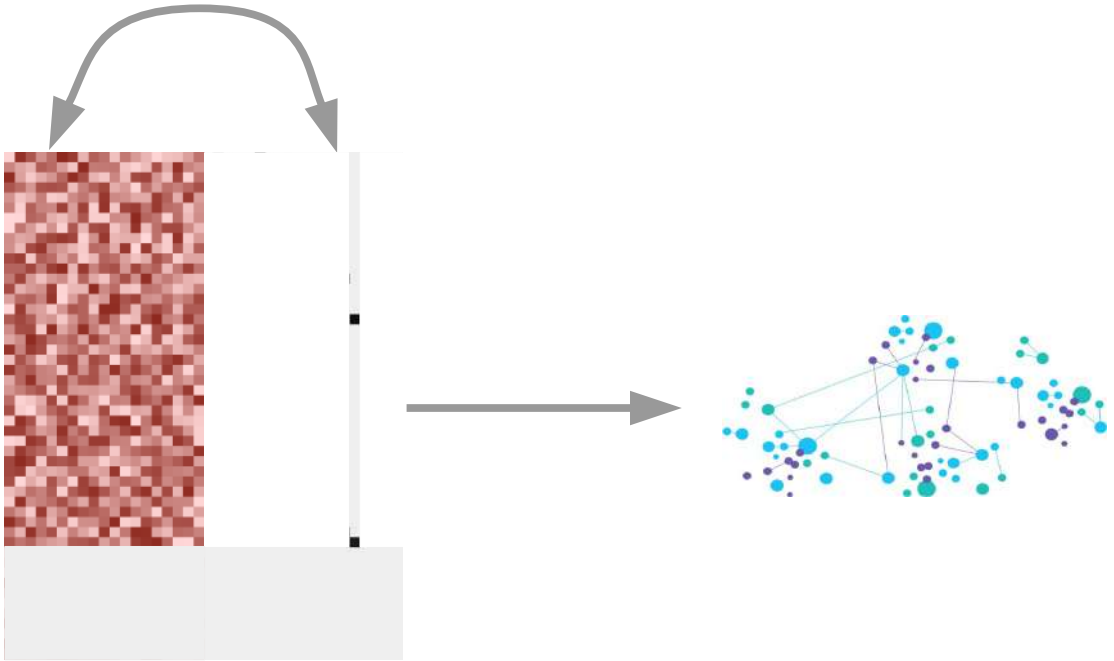


Molecular weight=183 Da

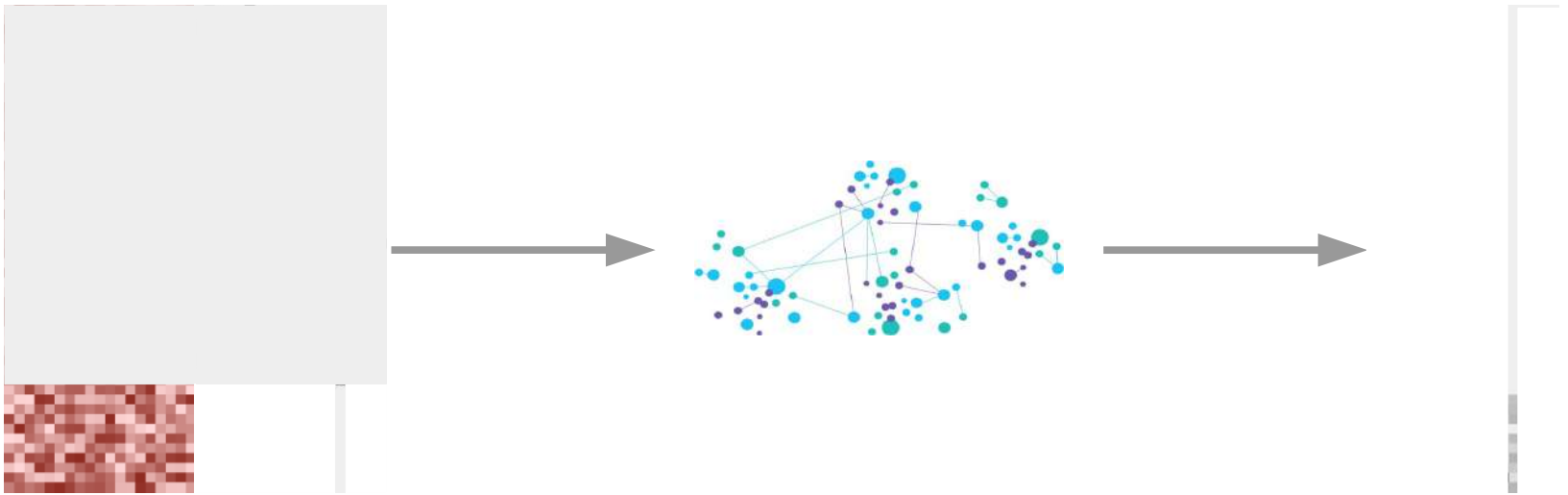


Train off one column at a time

Standard methods learn descriptor-protein correlations

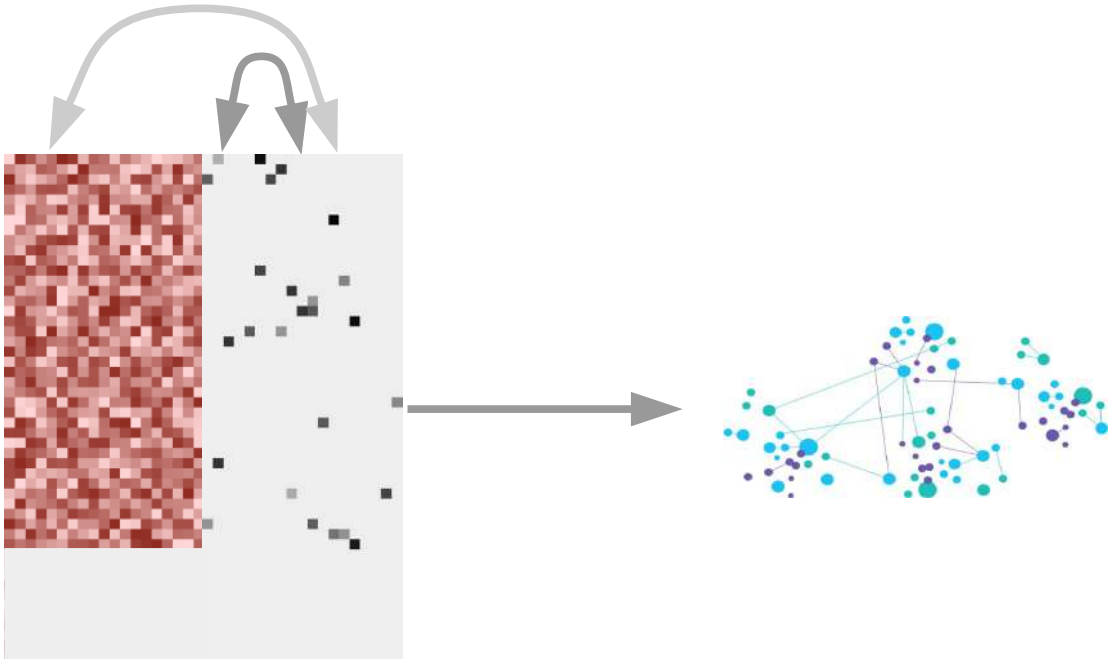


Train and predict one column at a time



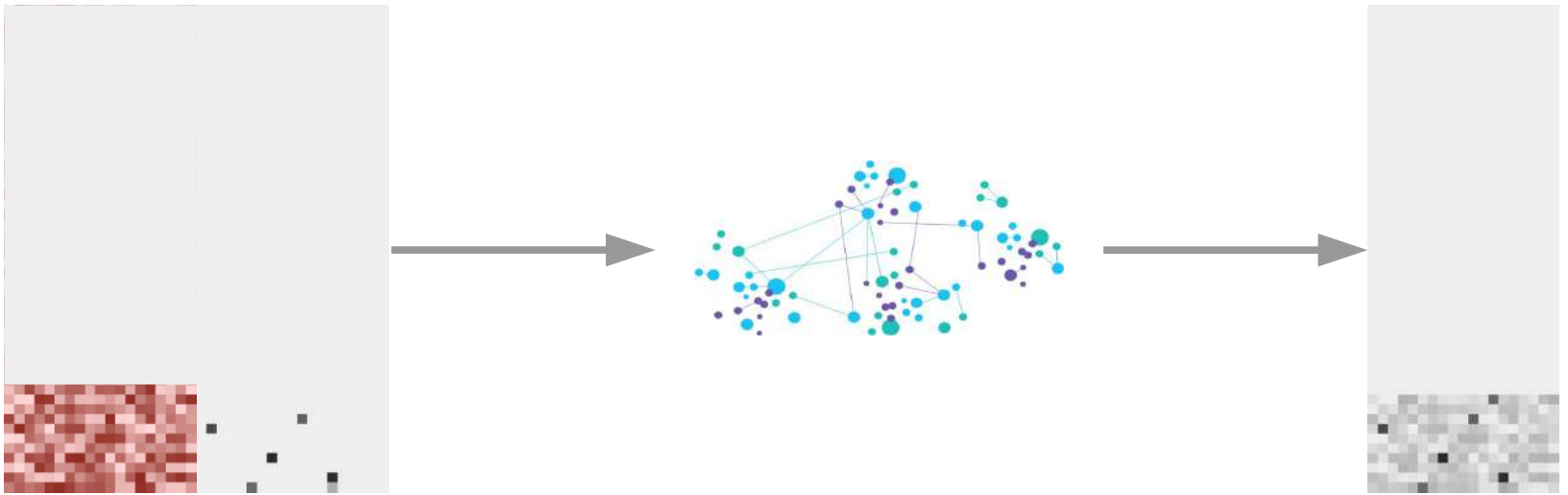
Alchemite™ uses all available data

Include protein-protein correlations

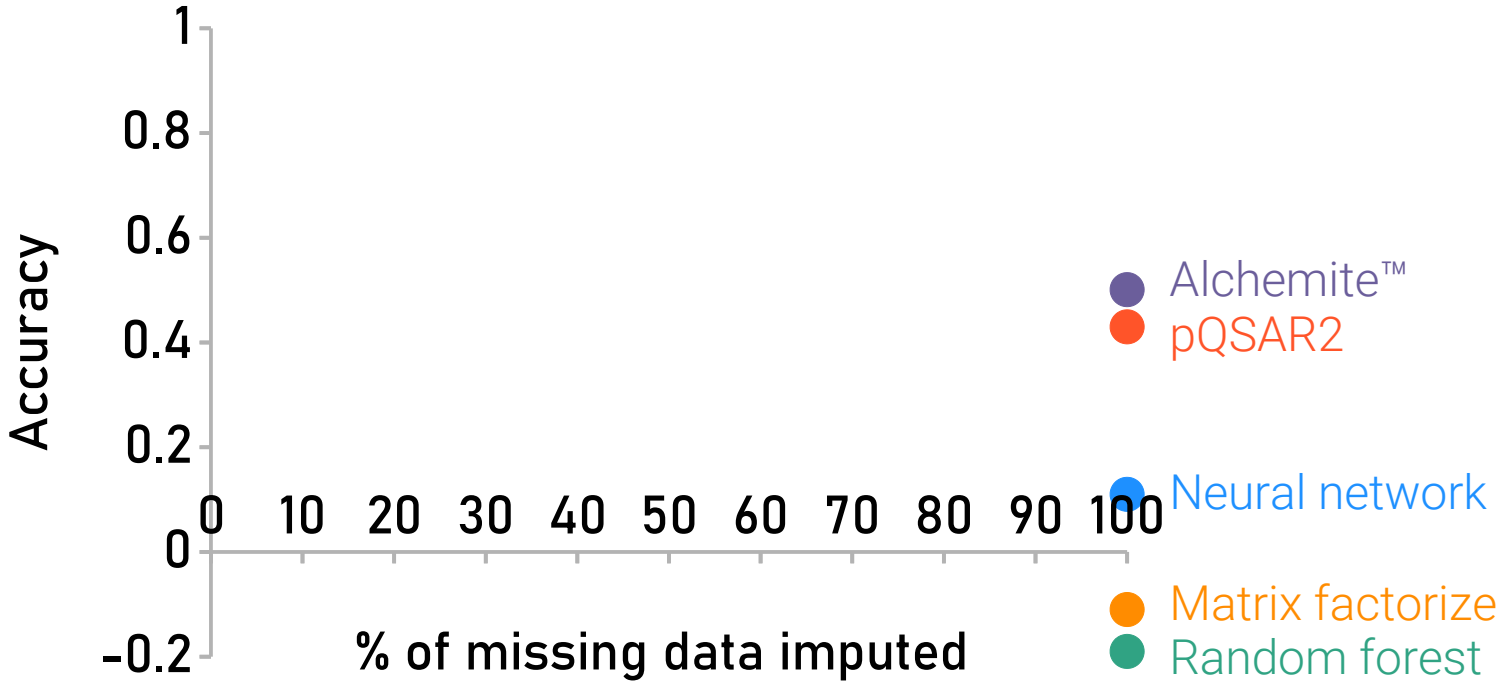


Validate imputation of missing entries

Realistically split holdout data set, extrapolate to new chemical space, and calculate the accuracy

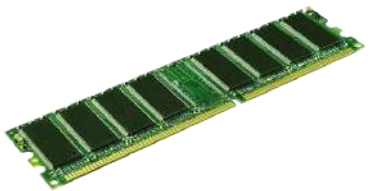


Alchemite™ outperforms other methods



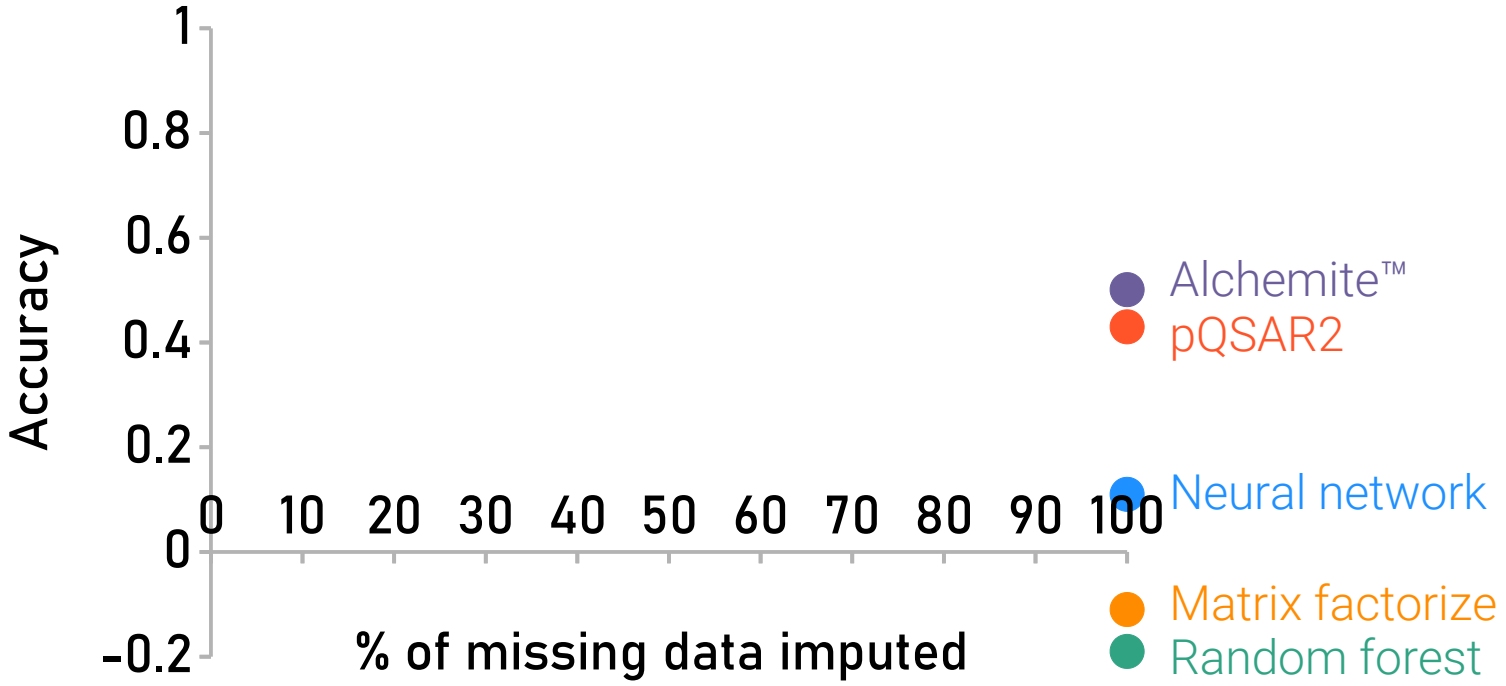
Computational cost

Seek R^2 = 0.465

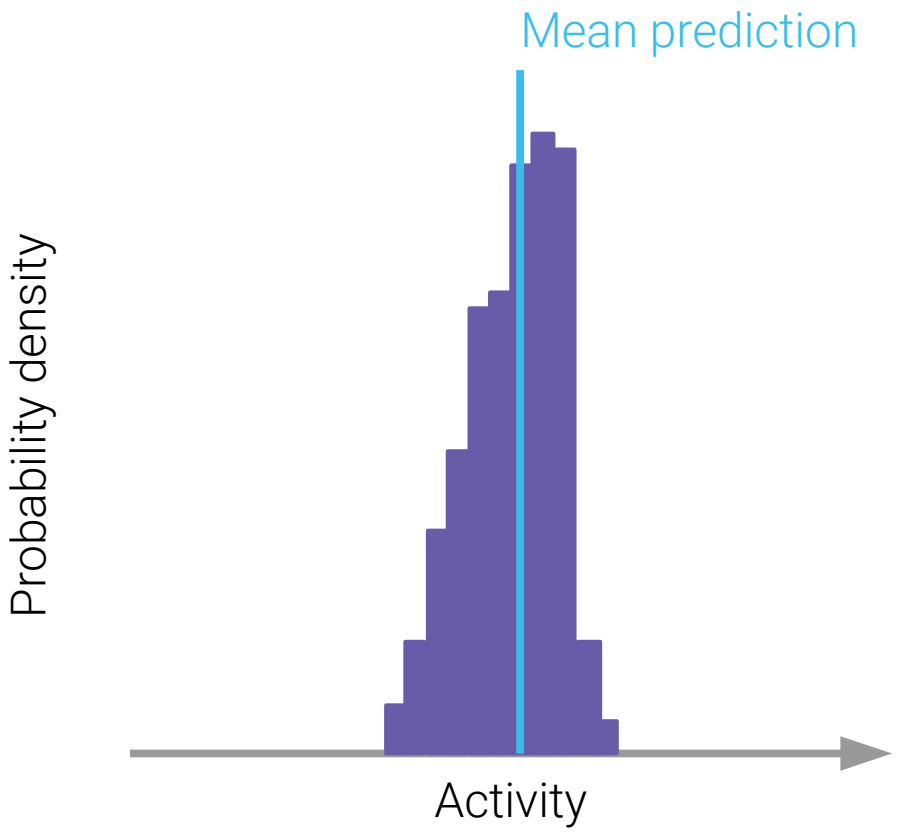


0.53 GB	13 s	7 GBs	Alchemite™
6.53 GB	154 s	1006 GBs	pQSAR2

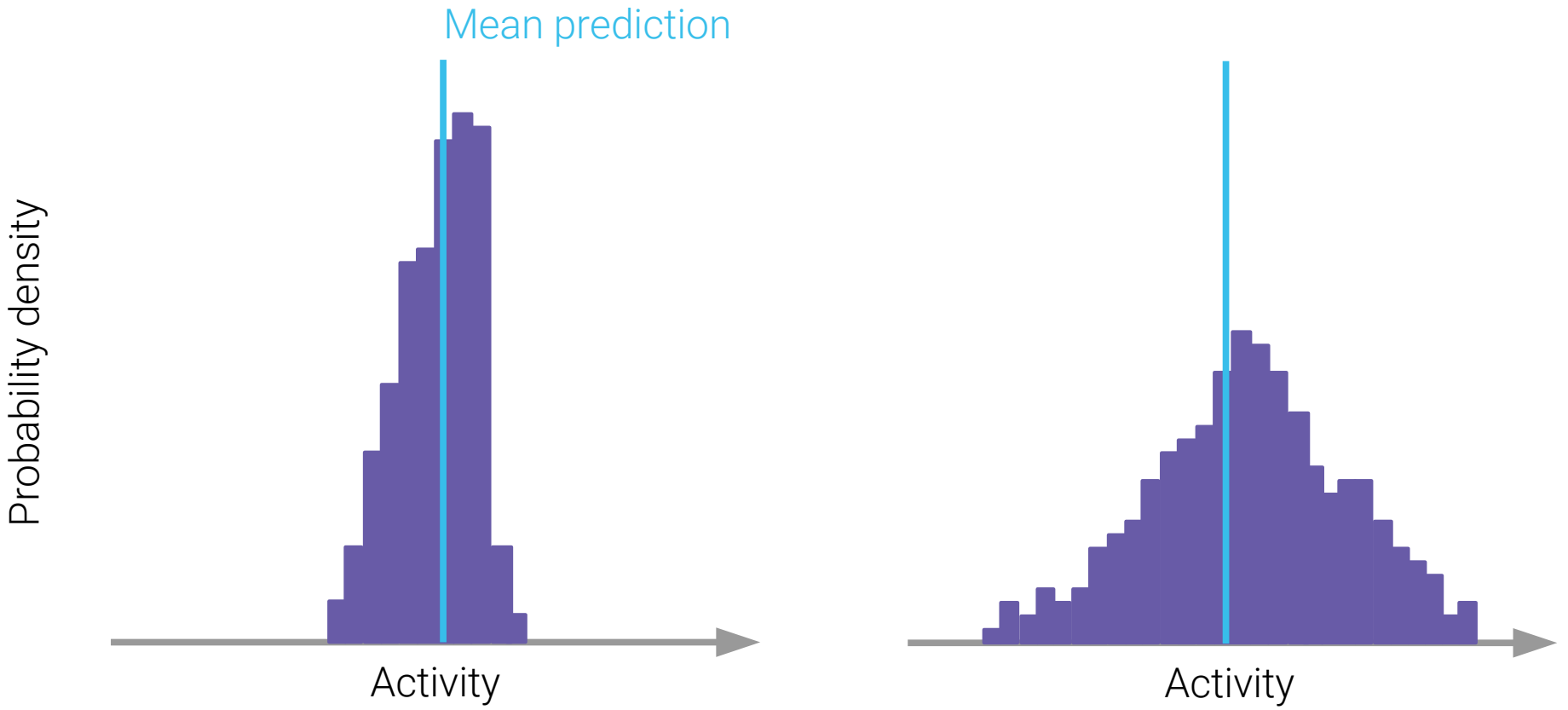
Alchemite™ outperforms other methods



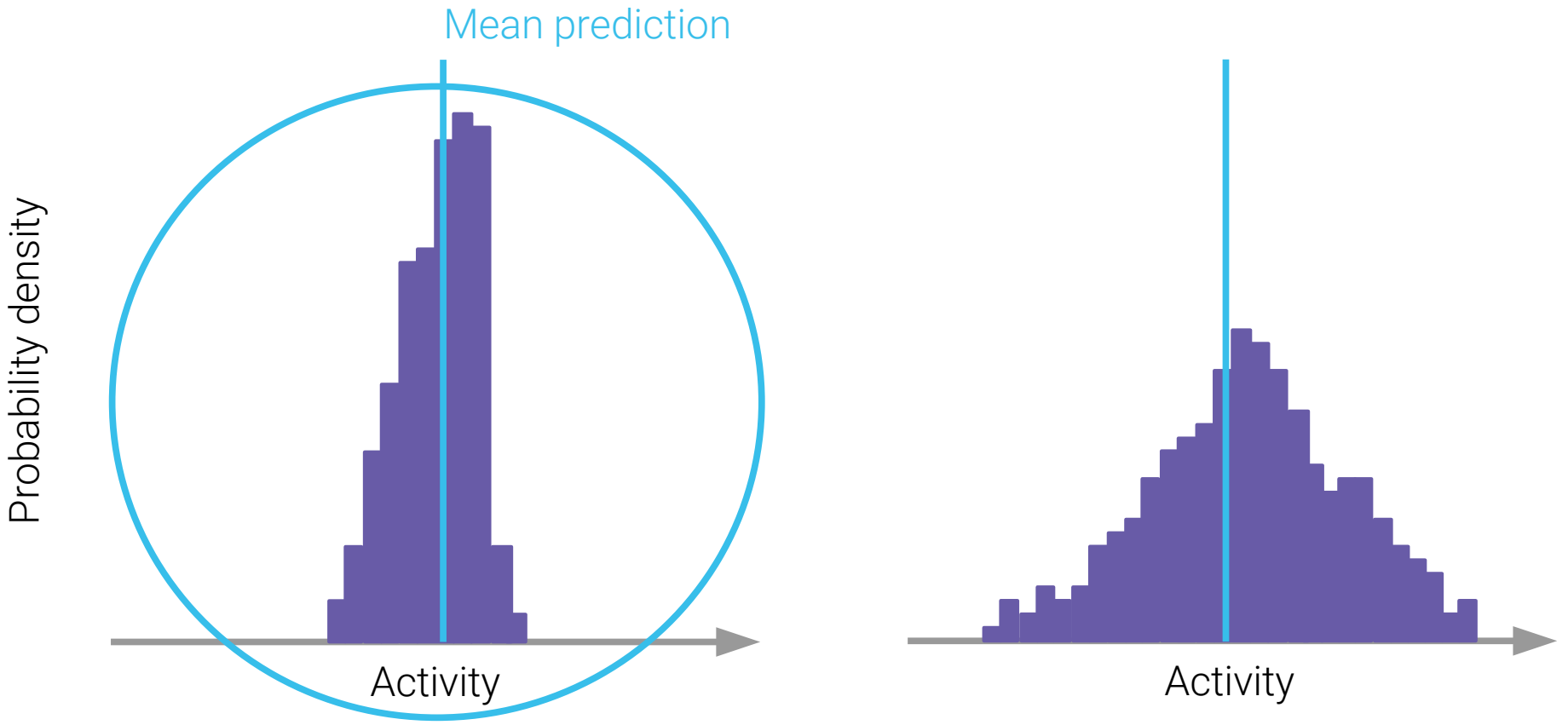
Calculate probability distribution



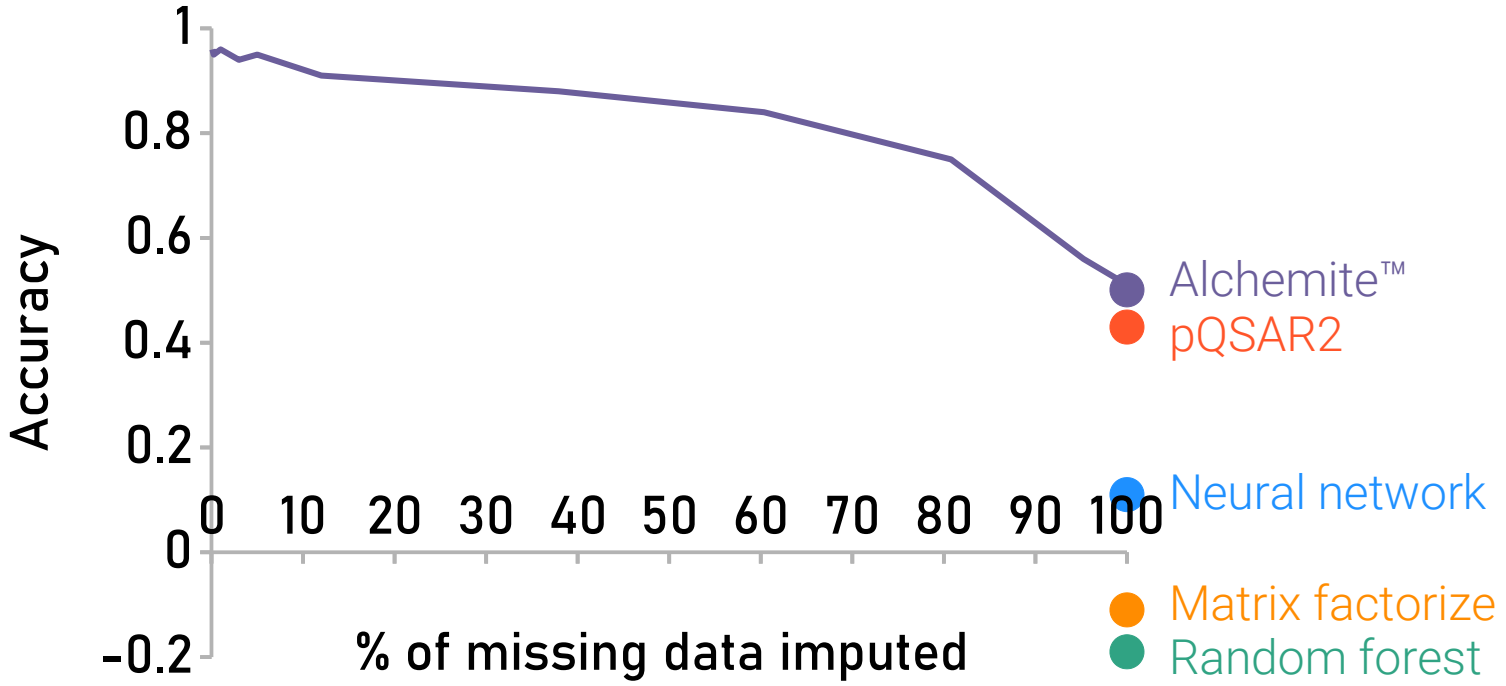
Less confident prediction



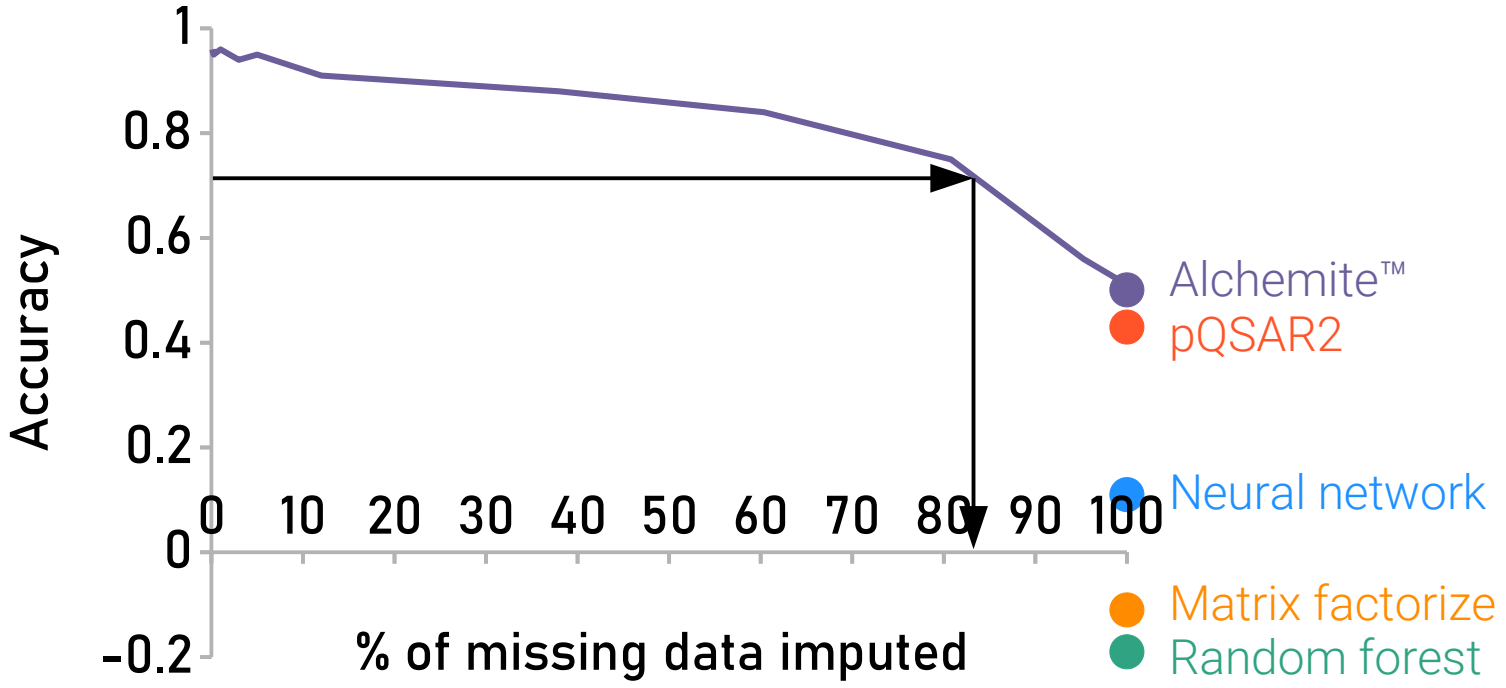
Focus on most confident predictions



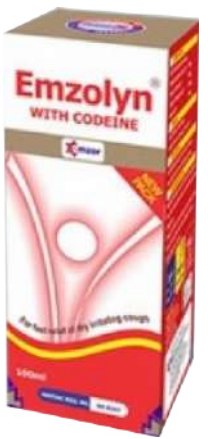
Reporting on only most confident predictions



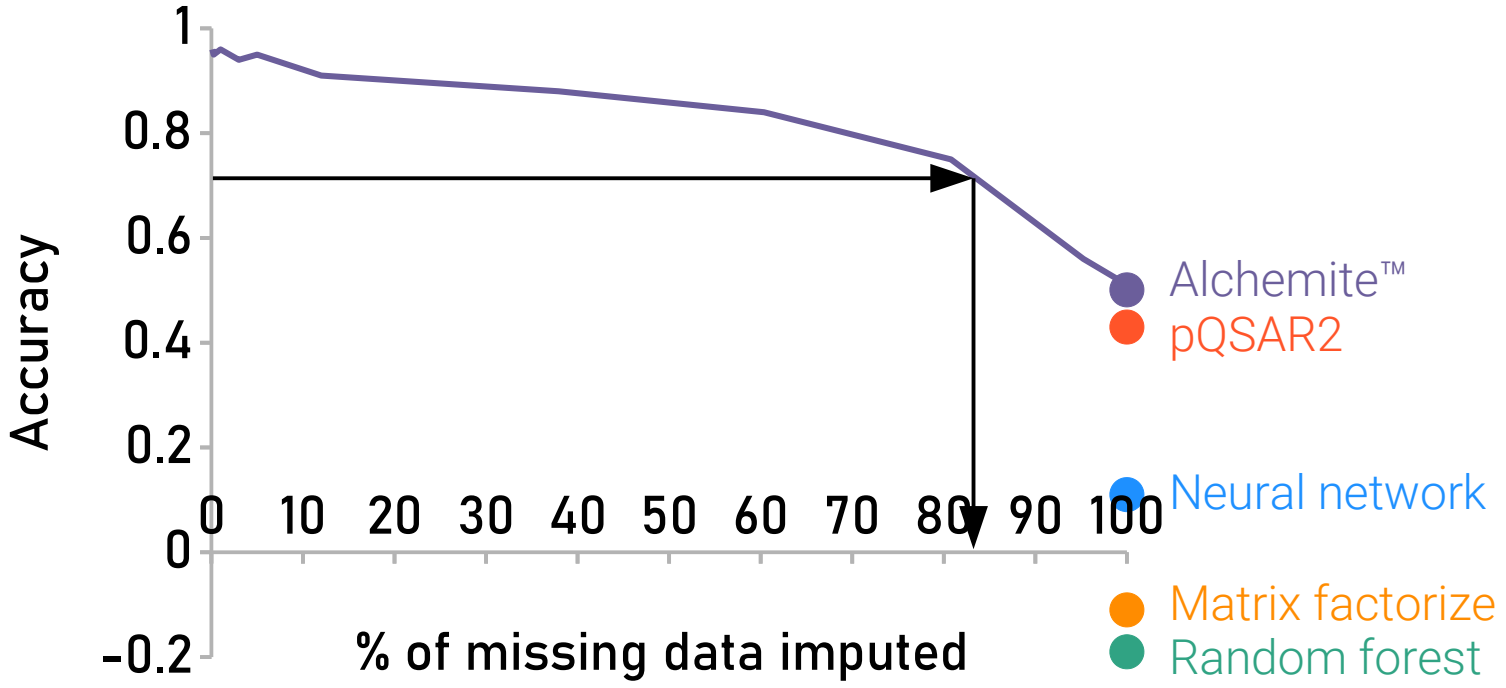
Select performance level



Different drugs can treat the same ailment



Focus on most promising hits



Open Source Malaria competition

OPEN SOURCE MALARIA

Looking for New Medicines

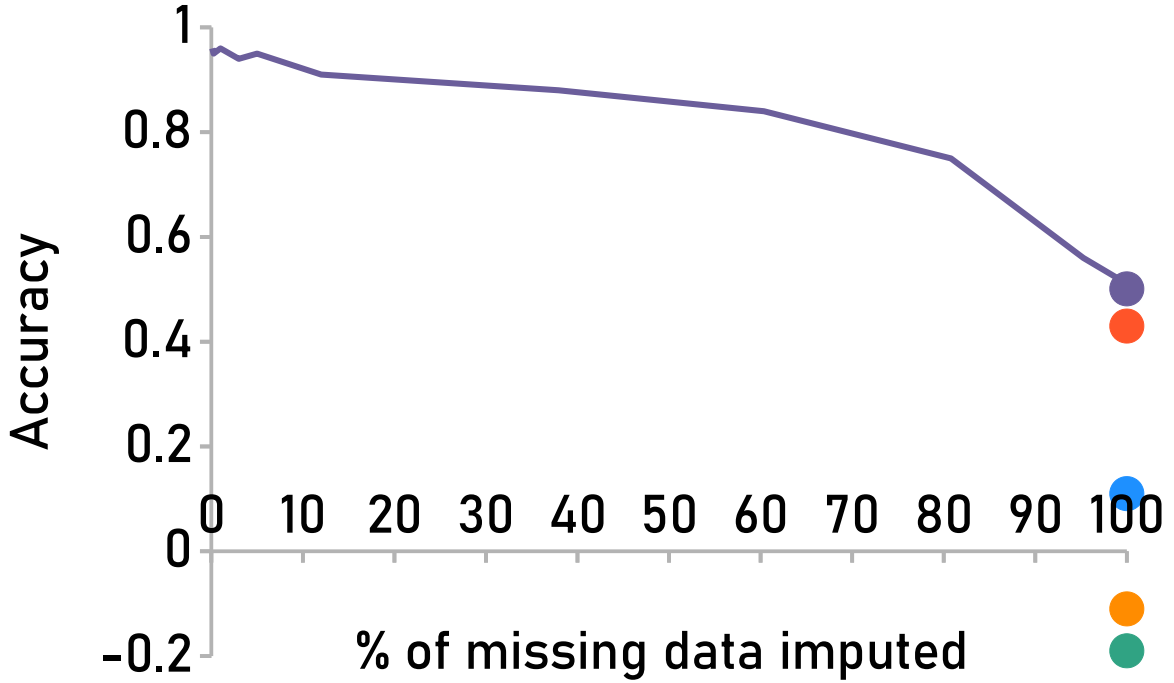
Open Source Malaria entrants

Entrant	Precision
Molomics	82%
Davy Guan	82%
Optibrium & Intellegens	81%
Exscientia	81%
Slade Matthews	64%
Auromind	58%
Raymond Lui	58%
KCL	36%
Interlinked TX	36%

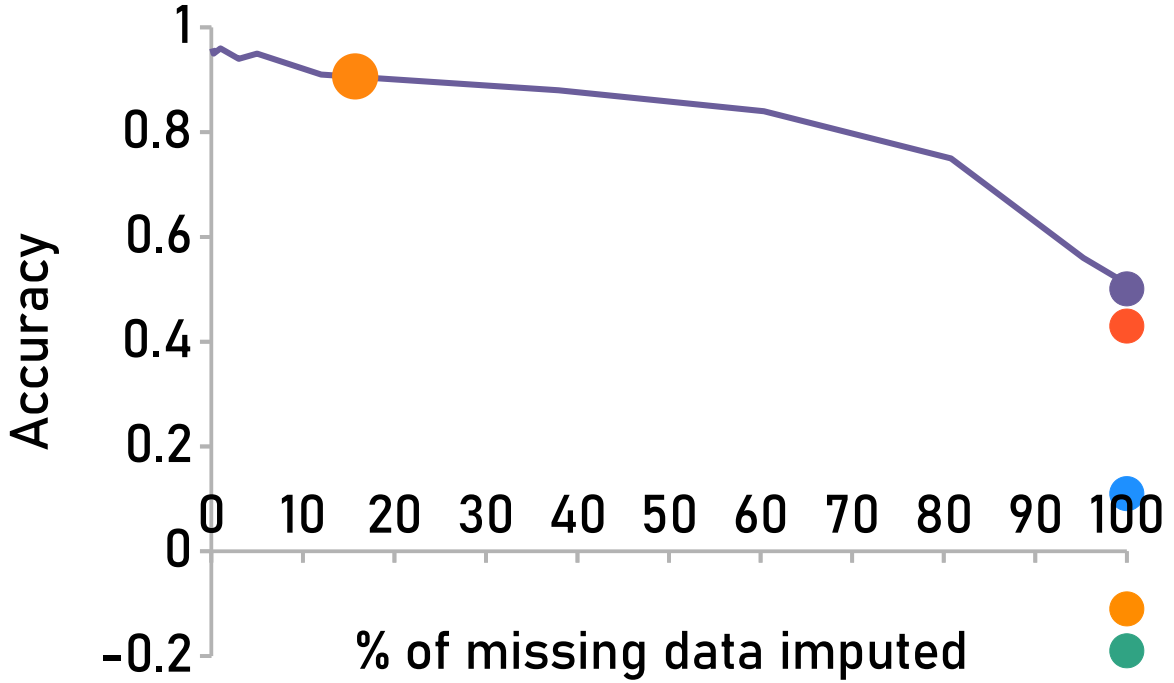
Open Source Malaria entrants

Entrant	Precision
Molomics	82%
Davy Guan	82%
Optibrium & Intellegens	81%
Exscientia	81%
Slade Matthews	64%
Auromind	58%
Raymond Lui	58%
KCL	36%
Interlinked TX	36%

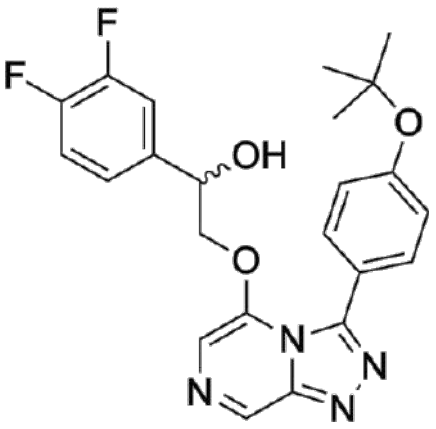
Focus on compounds with low uncertainty



Focus on compounds with low uncertainty



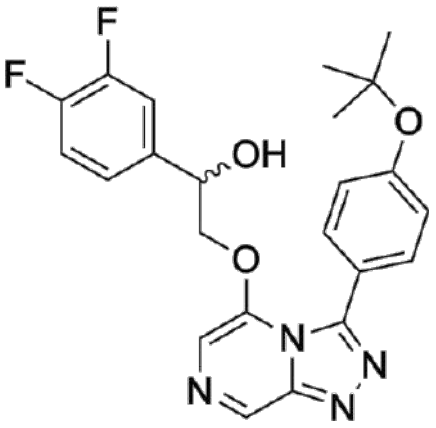
Open Source Malaria experimental validation



Optibrium & Intellegens

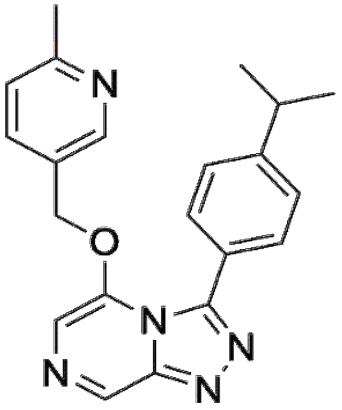
0.647 µM

Open Source Malaria other compounds



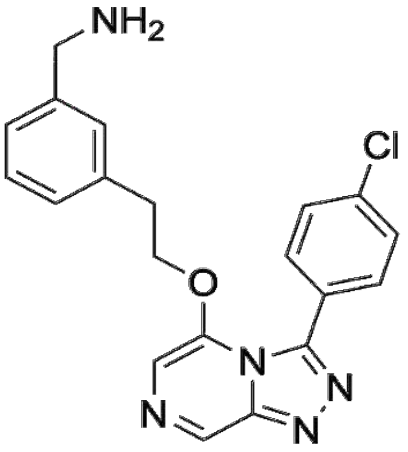
Optibrium & Intellegens

0.647 µM



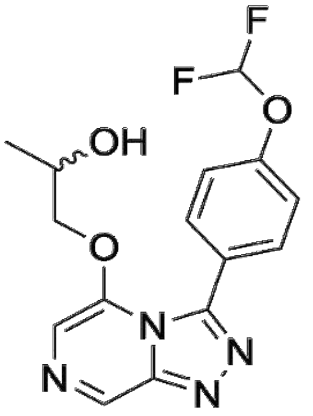
Davy Guan

>25 µM



Exscientia

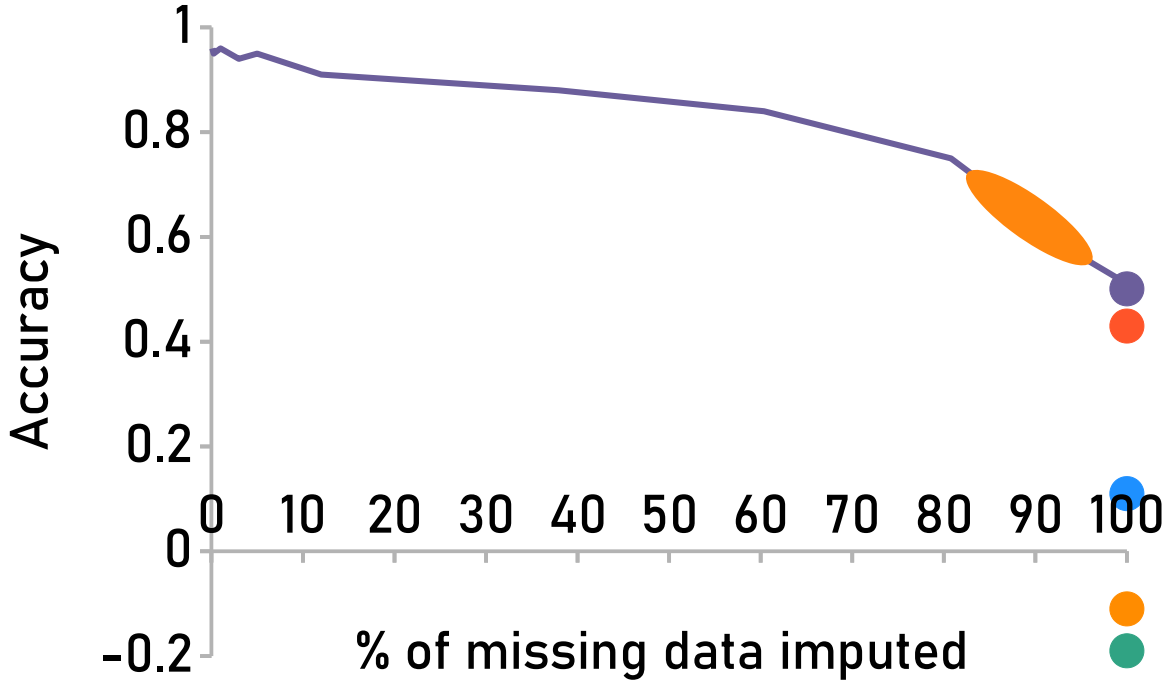
10.9 µM



Molomics

>25 µM

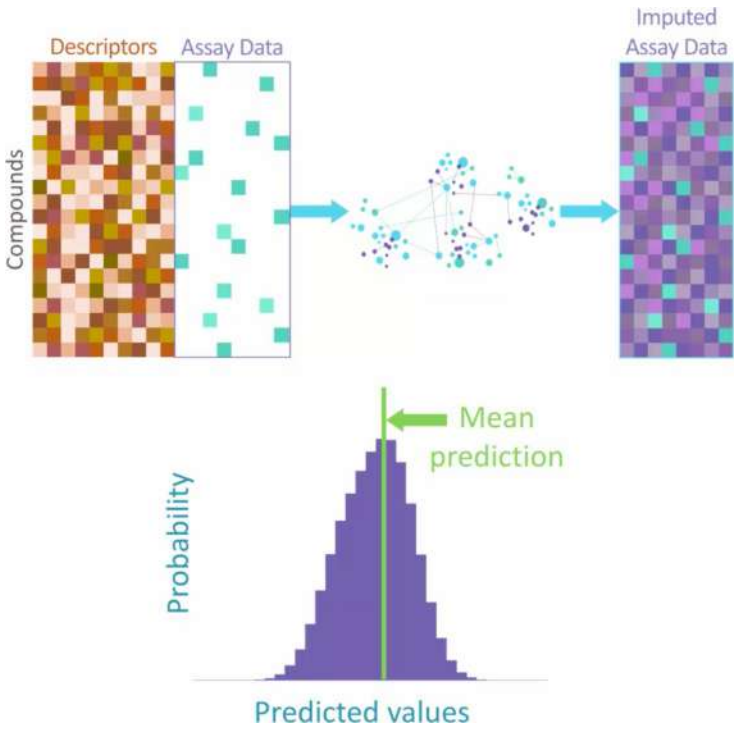
Open Source Malaria other compounds



Imputation Versus Prediction and Applications in Drug Discovery

Matthew Segall, Benedict Irwin, Thomas Whitehead, Gareth Conduit

P36



Practical applications:

- Project optimisation
  - 2,453 compounds and 18 heterogeneous endpoints
- Global pharma data
  - 678,994 compounds and 1,166 endpoints
- Combined with generative methods
  - Design of active anti-malarial compounds

matt@optibrium.com Skype: matthew.d.segall

Whitehead *et. al*. J. Chem. Inf. Model (2019) **59**(3) p. 1197

Irwin *et al*. Future Drug Discovery (2020) **2**(2) DOI: 10.4155/fdd-2020-0008

Irwin *et al*. J. Chem. Inf Model. (2020) **60**(6), p. 2848

Demonstrated Benefits of Alchemite in Drug Discovery

Exclusive Partnership with Optibrium

P36

- ‘Fill in’ missing data to proactively highlight high-quality compounds
  - Identify new opportunities with confidence
- Identify experimental outliers
  - Highlight unlikely results – experimental errors or unexpected SAR
  - Highlight potential false negatives
- Suggest the most valuable measurements to improve predictions for target assays and chemistry
  - Prioritise experimental resources
  - Confidently progress the best compounds to expensive, downstream experiments
- Virtual screening across endpoints for multi-parameter optimisation

optibrium

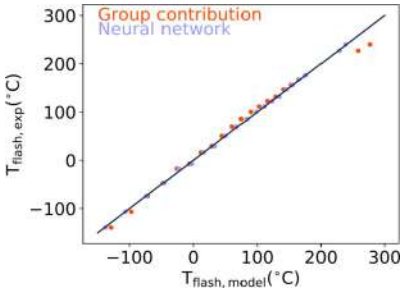
Lubricants

Reduce friction between surfaces, protect surfaces from wear, transfer heat, remove dirt, prevent surface corrosion

Molecules are heavy hydrocarbons with branches and functional groups, and a lubricant blends many molecules and inorganic additives

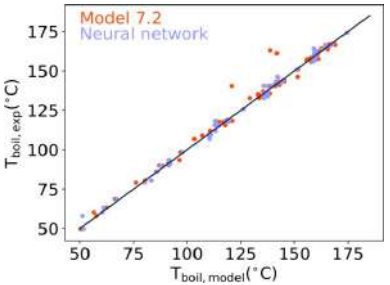
Machine learning to juxtapose experimental data and computational methods including molecular dynamics

Predicting properties to lubricants



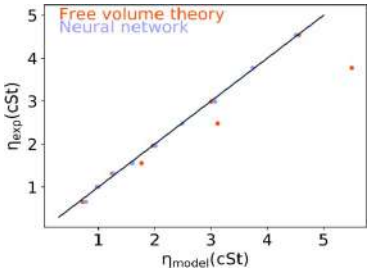
Flash point

Alchemite™ R^2=0.997
Group contribution R^2=0.971



Boiling point

Alchemite™ R^2=0.992
Fitting model R^2=0.976



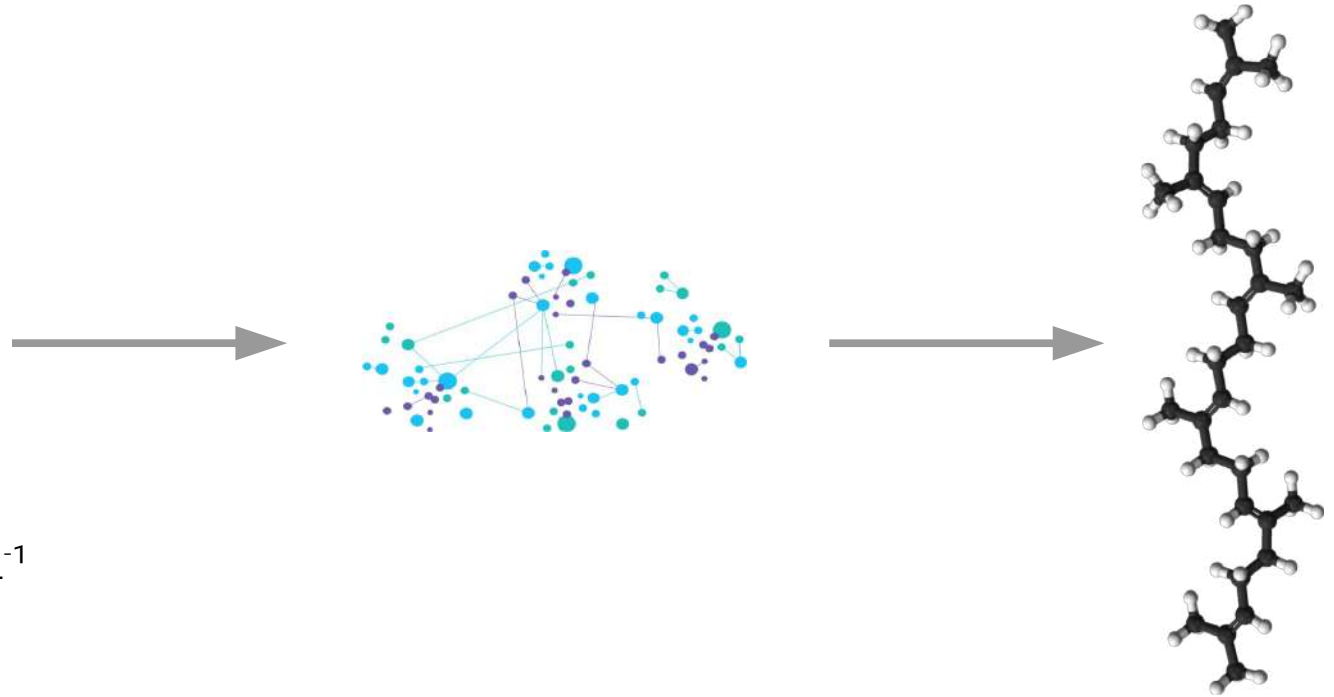
Kinematic viscosity

Alchemite™ R^2=0.998
Free volume theory R^2=0.899

Fluid Phase Equilibria 501, 112259 (2019); Journal of Chemical Physics 153, 014102 (2020)

Designing a lubricant

Flashpoint	> 122.55°C
Viscosity	< 3.78cSt
Boiling point	> 270°C
Density	< 769mgmL^{-1}



Fluid Phase Equilibria 501, 112259 (2019); Journal of Chemical Physics 153, 014102 (2020)

Improving inks

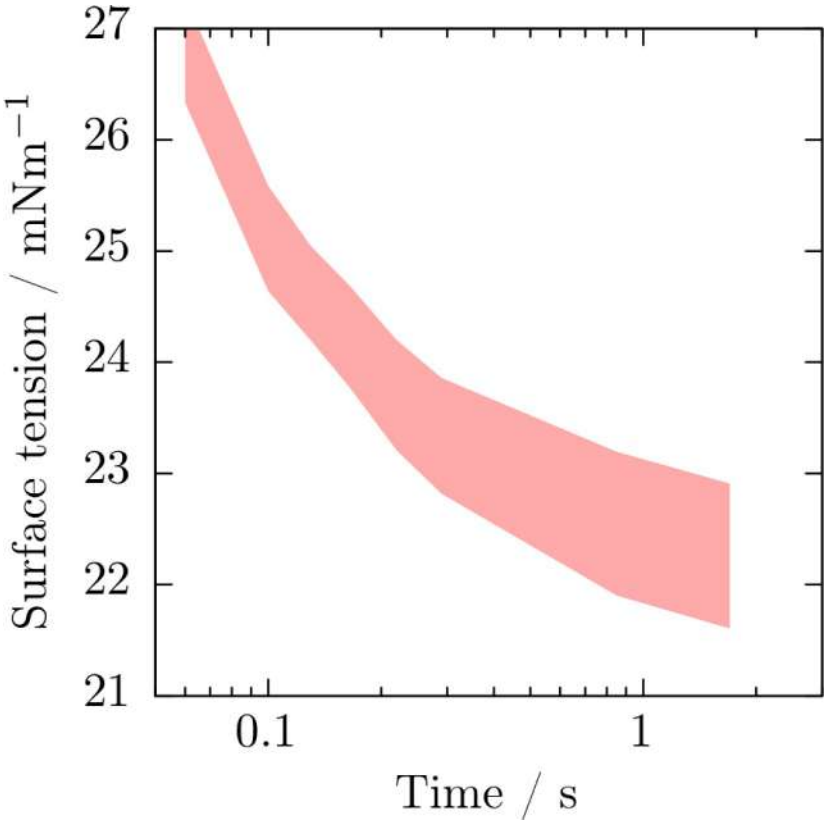
Inks can comprises over 30 individual chemicals

Domino Ink wish to remove two chemicals to improve sustainability

Limited access to laboratory during COVID-19 so turn to machine learning

Dr Josie Harries

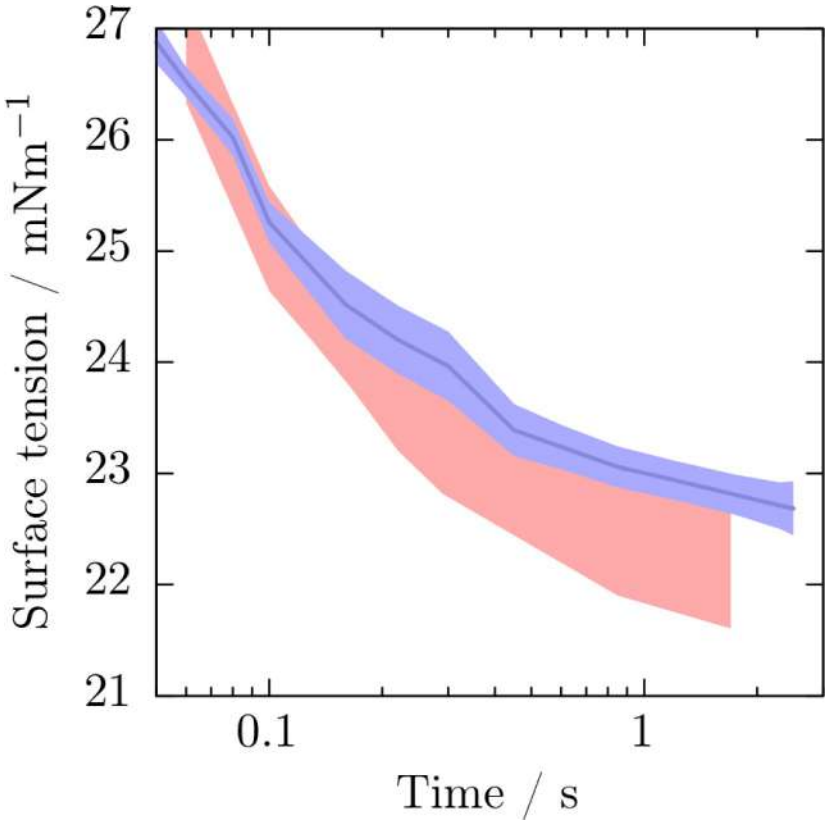
Target for a new ink



Dr Josie Harries

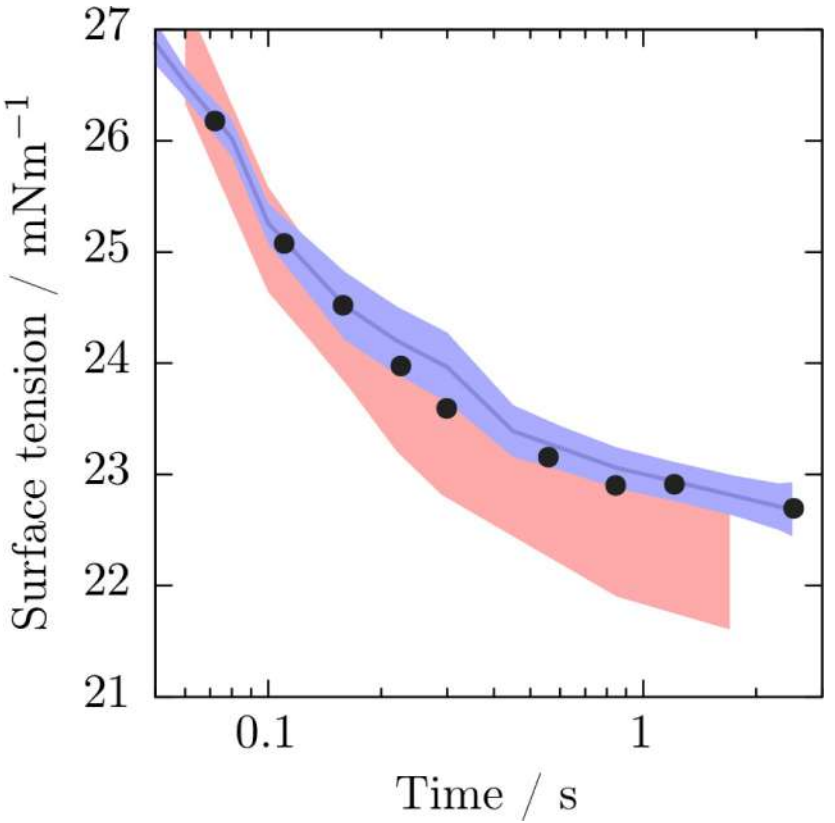
DOMINO

Alchemite™ proposes a new ink



Dr Josie Harries

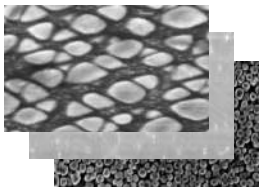
Experimental validation of the proposed ink



Dr Josie Harries

Materials

High temperature alloys



Batteries

Direct laser deposition

Summary

Alchemite™ trains across all endpoints to capture property-property correlations

Understand and exploit probability distribution to focus on most confident results

Impute results of missing properties to high accuracy, enabling computational screening of compounds to identify new hits

Partnership with Optibrium for small molecule drug design

gareth@intellegens.ai