UNIVERSITY OF CAMBRIDGE

Modern day blacksmith: application of machine learning to materials and drug design

Gareth Conduit

Theory of Condensed Matter group

Machine learning algorithm to

Train from sparse datasets

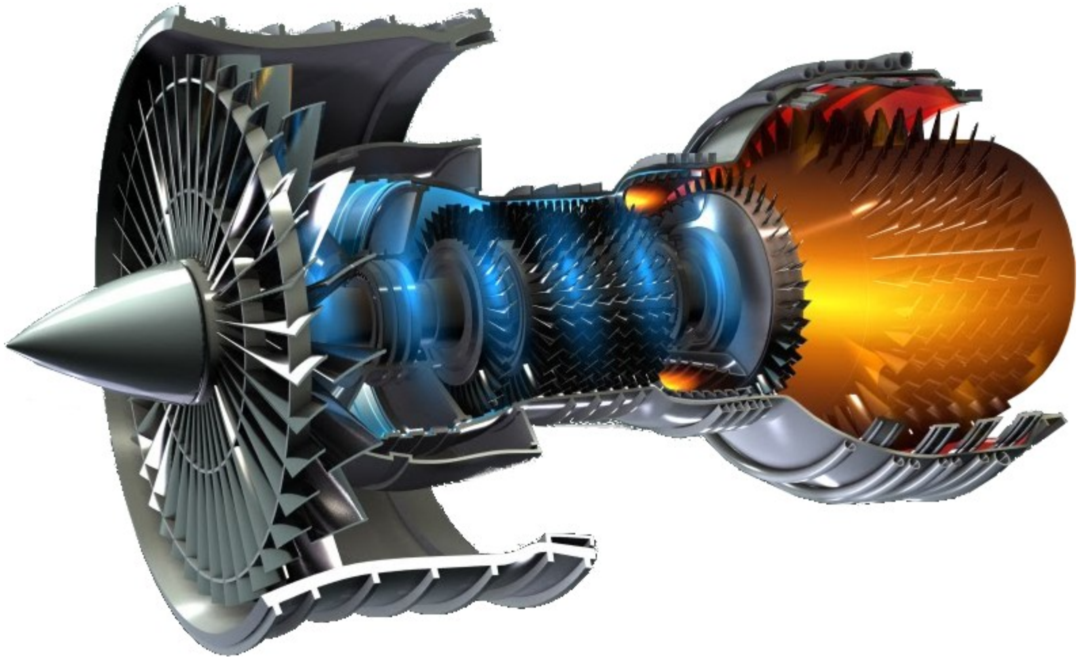
Merge simulations, physical laws, and experimental data

Reduce the need for expensive experimental development

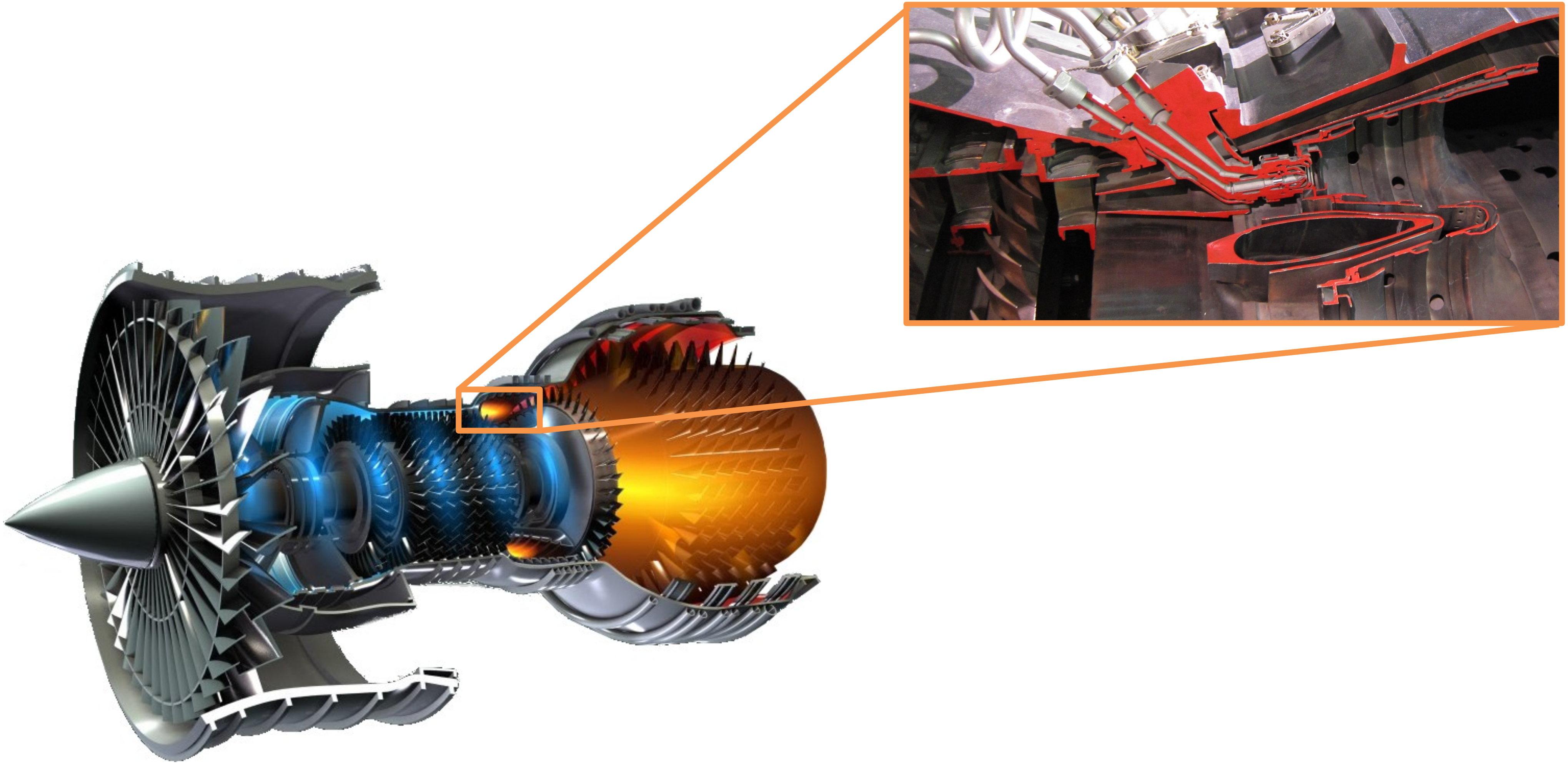
Accelerate materials and drugs discovery

Generic with proven applications in materials discovery and drug design

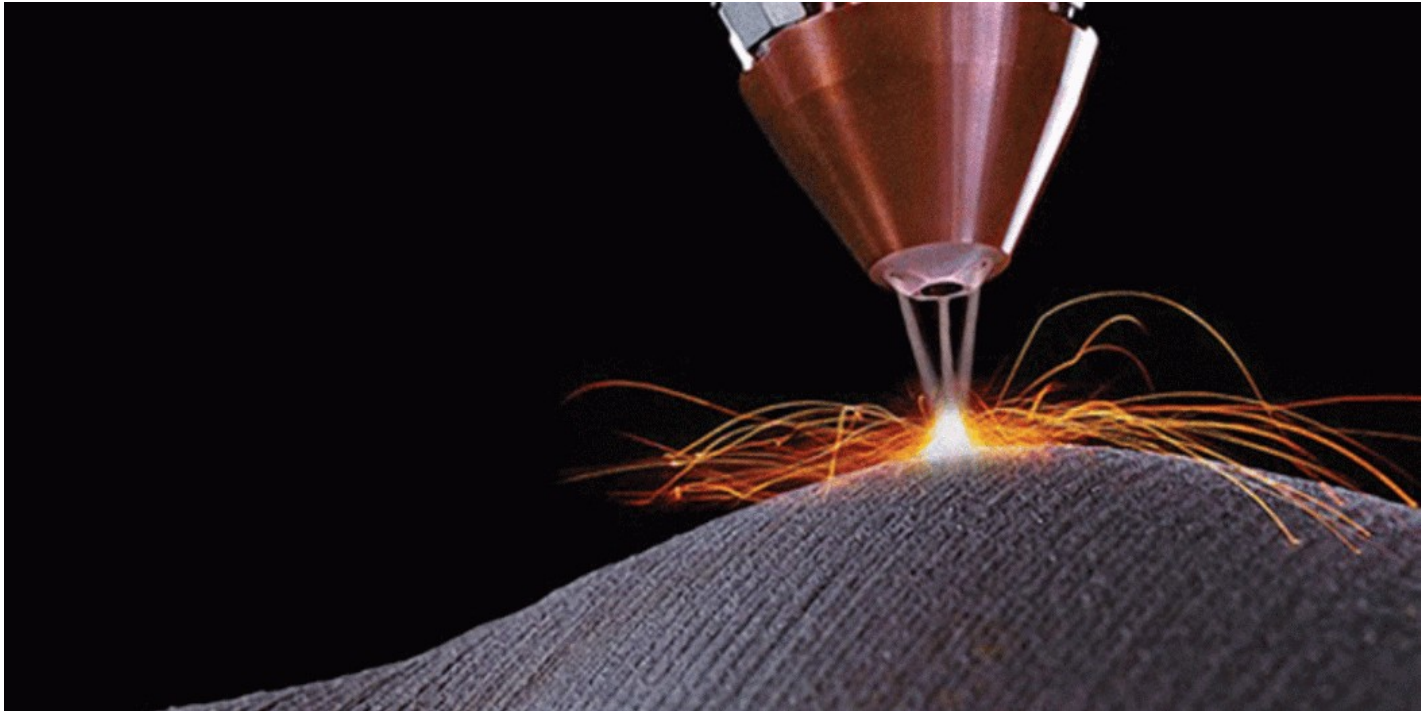
Schematic of a jet engine



Combustor in a jet engine



Direct laser deposition requires new alloys

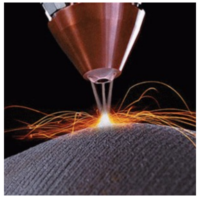


Black box machine learning for materials design

Composition

Properties

Defects



Fatigue

Welding

Train the neural network

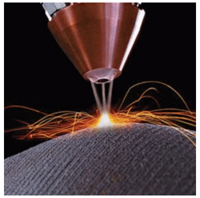
Composition

Neural network for materials design

Composition

Properties

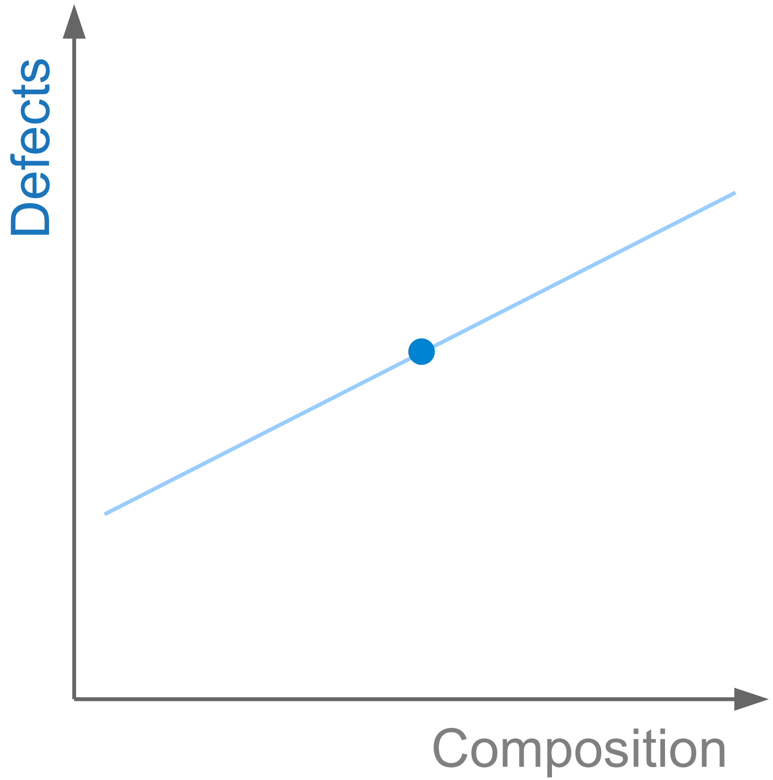
Defects



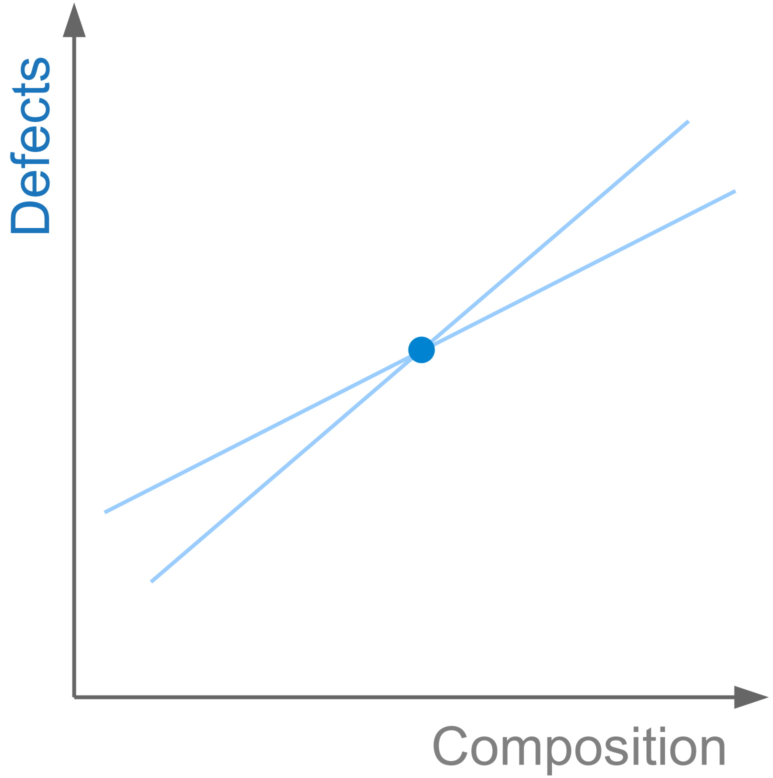
Fatigue

Welding

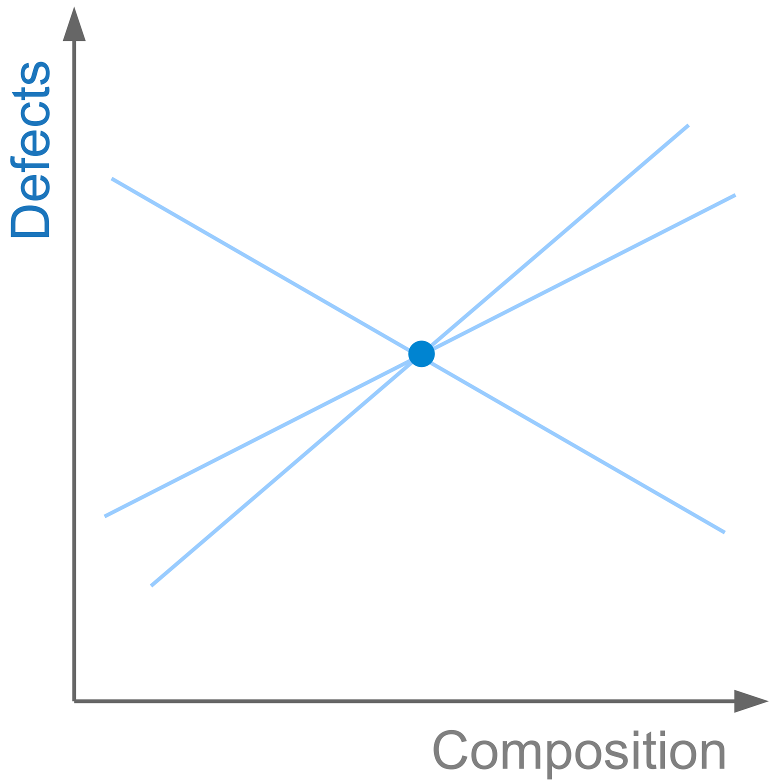
One point cannot define a straight line



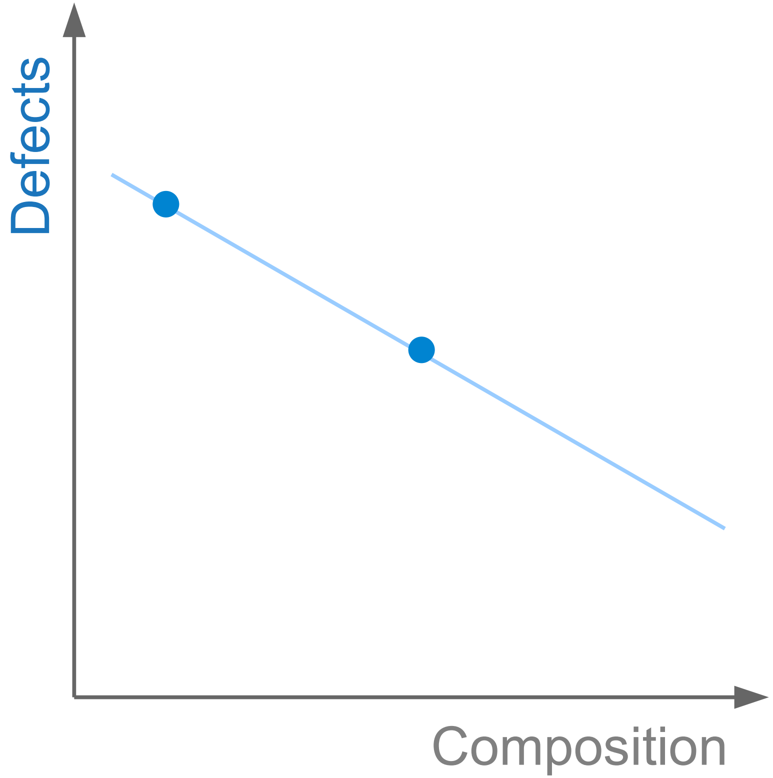
One point cannot define a straight line



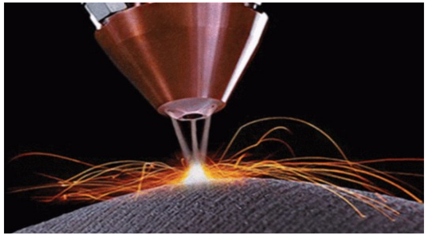
One point cannot define a straight line



Need at least two points to define a straight line



Data required for a defects model

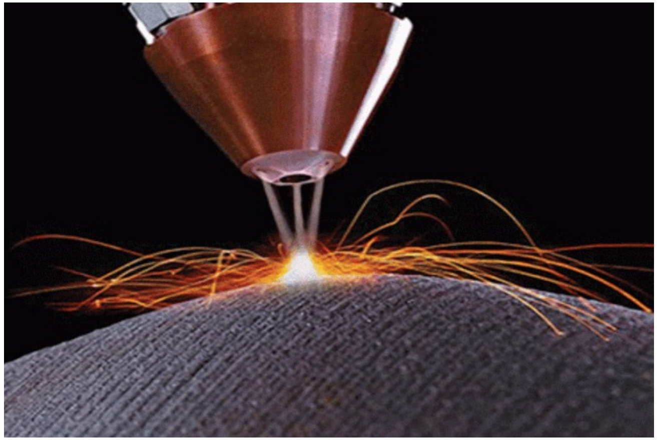


Composition and heat treatment space 30 dimensions

Requires 31 points to fit a hyperplane

Just 8 data points available

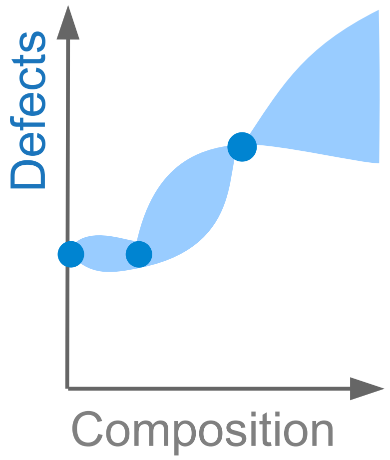
Neural networks for materials design



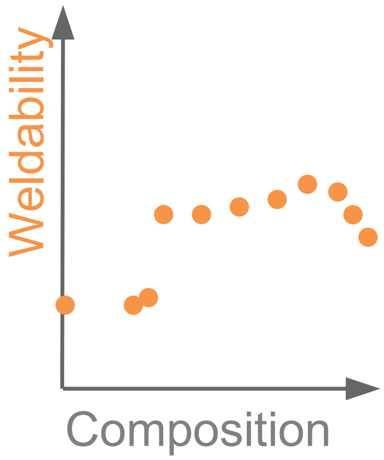
Laser

Electricity

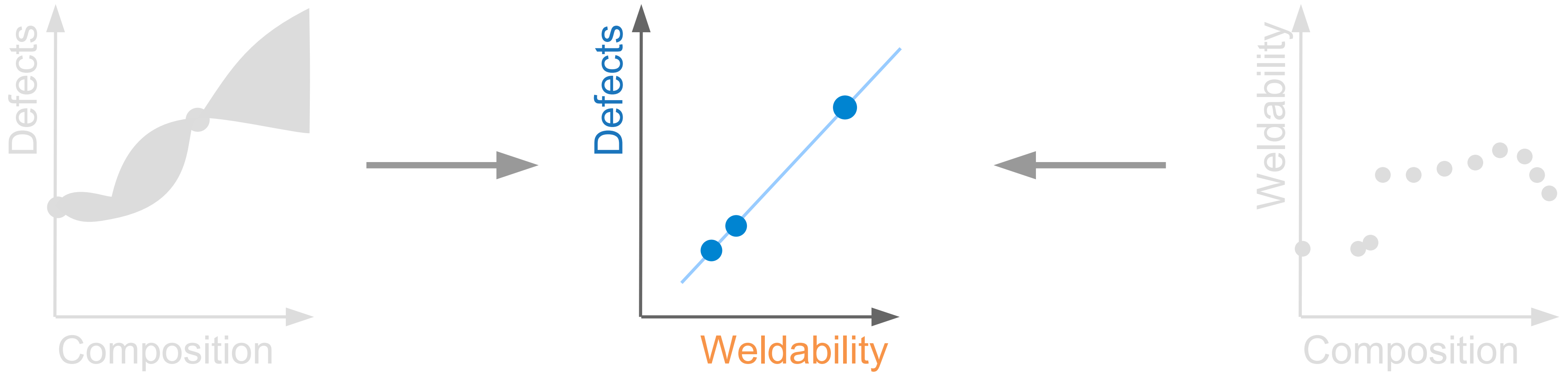
Insufficient data for processability



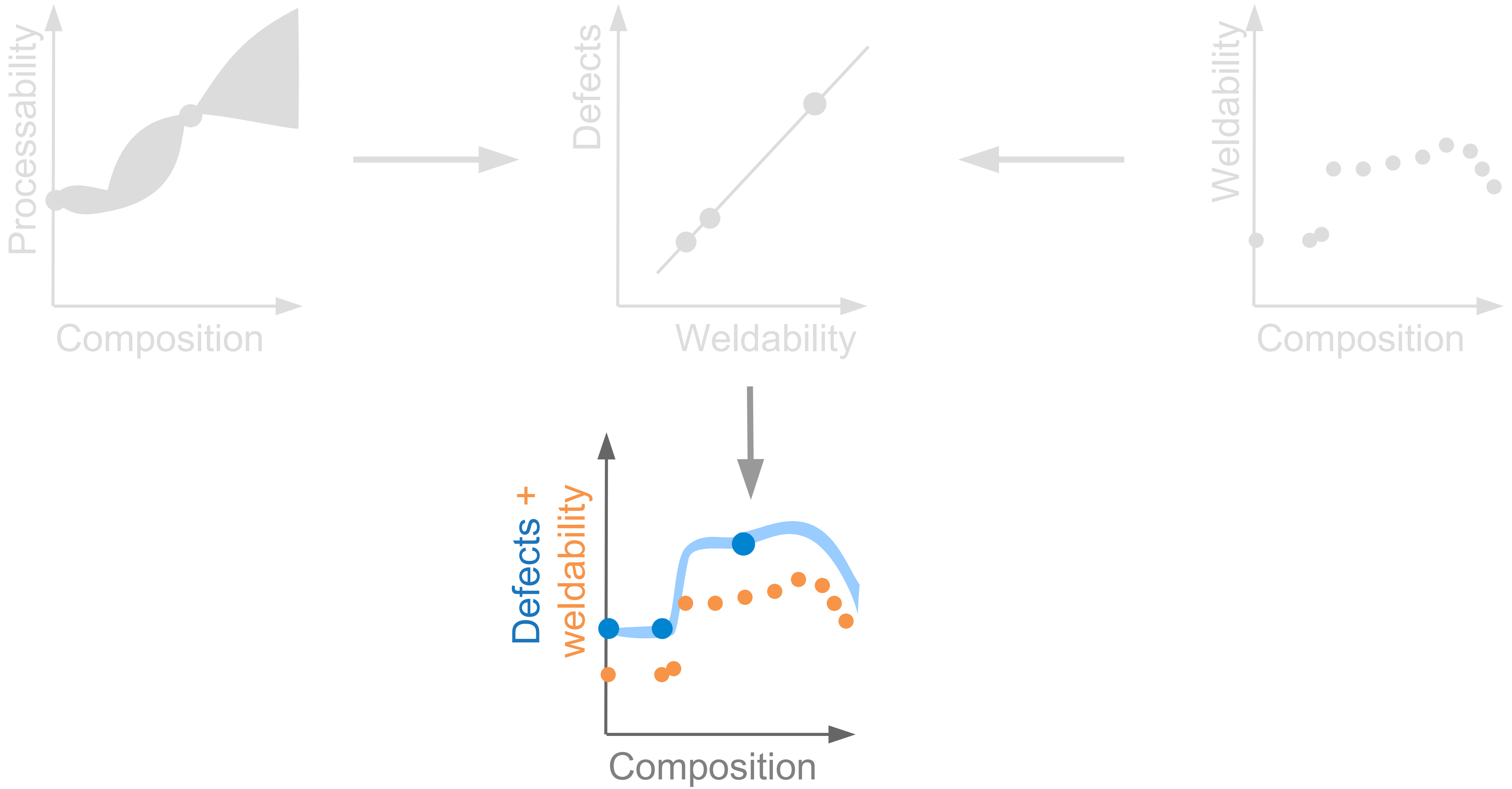
Welding is analogous to direct laser deposition



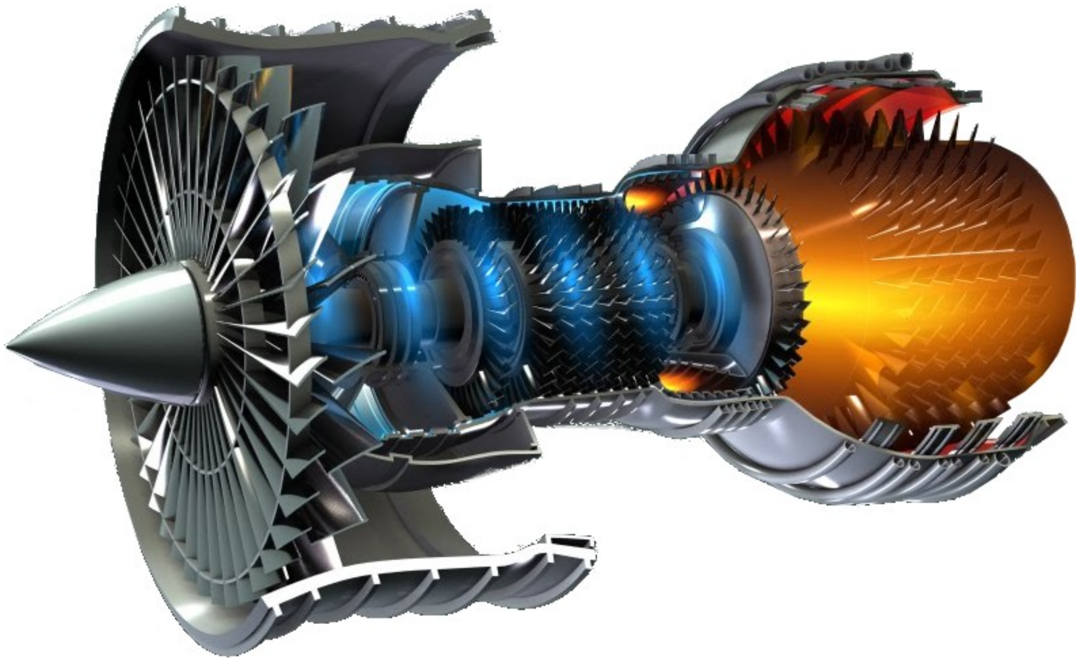
Simple processability-welding relationship



Merging properties with the neural network



Schematic of a jet engine



Target properties

Elemental cost	$< 25\ \$kg^{-1}$
Density	$< 8500\ kgm^{-3}$
γ’ content	< 25 wt%
Oxidation resistance	$< 0.3\ mgcm^{-2}$
Defects	$< 0.15\%$ defects
Phase stability	> 99.0 wt%
γ’ solvus	$> 1000^{\circ}C$
Thermal resistance	$> 0.04\ K\Omega^{-1}m^{-3}$
Yield stress at 900˚C	> 200 MPa
Tensile strength at 900˚C	> 300 MPa
Tensile elongation at 700˚C	$> 8\%$
1000hr stress rupture at 800˚C	> 100 MPa
Fatigue life at 500 MPa, 700˚C	$> 10^5$ cycles

Composition

Cr 19%

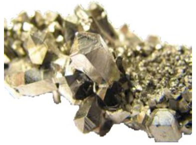
Co 4%

Mo 4.9%

W 1.2%

Zr 0.05%

Nb 3%



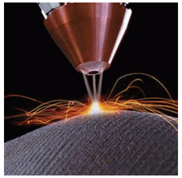
Al 2.9%

C 0.04%

B 0.01%

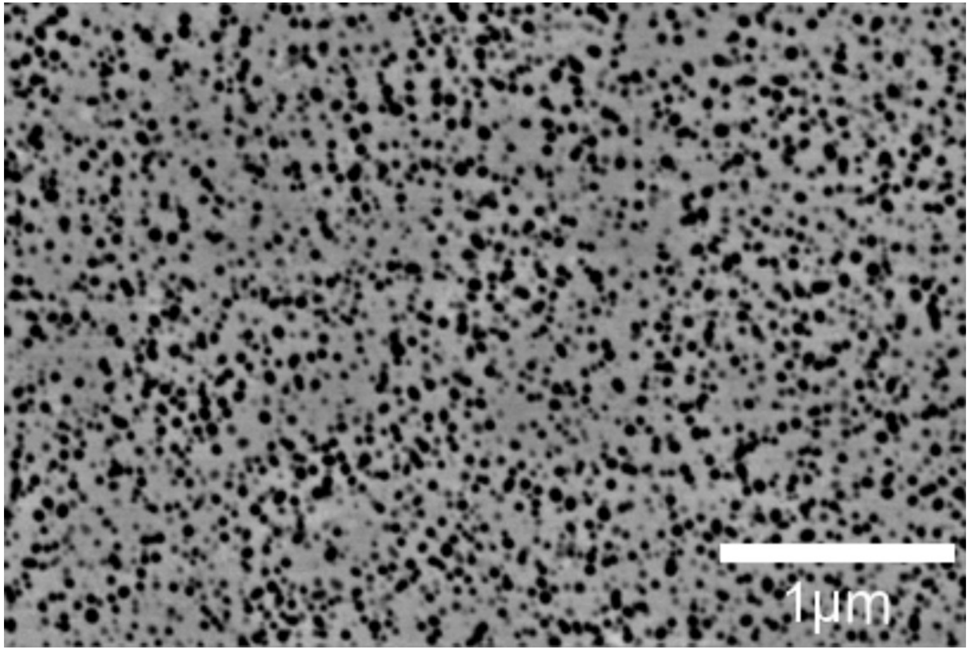
Ni

Expose 0.8



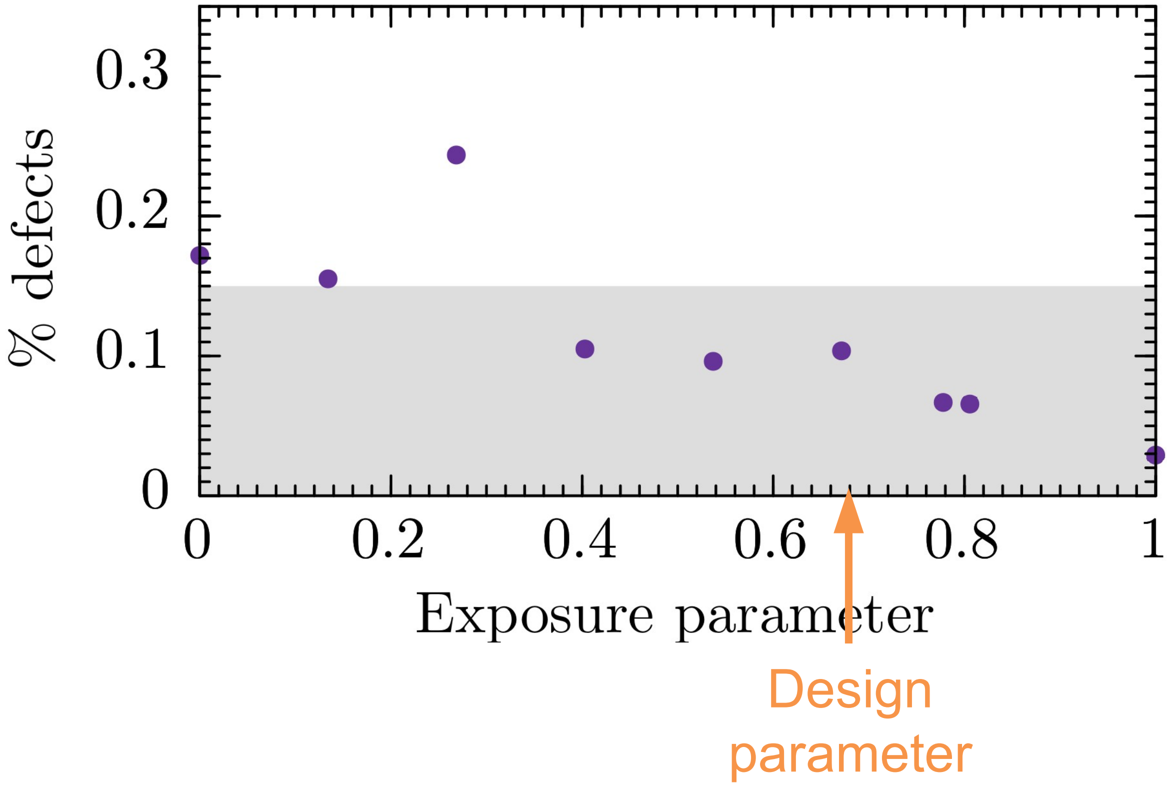
T_{HT} 1300ºC

Microstructure



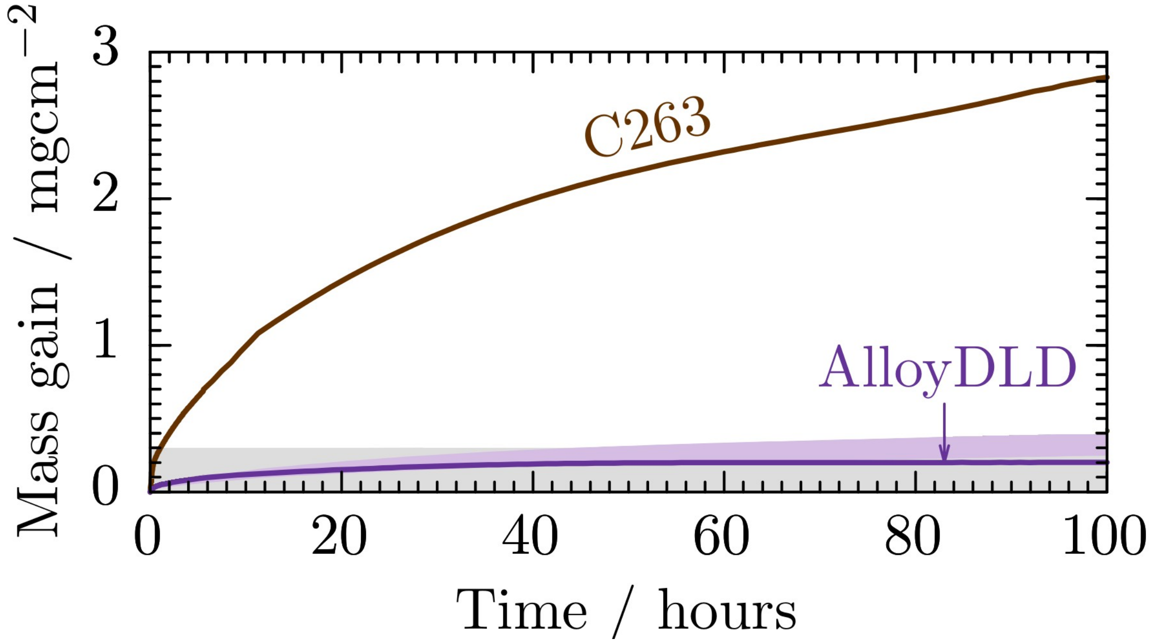
Probabilistic neural network identification of an alloy for direct laser deposition
Materials & Design 168, 107644 (2019)

Testing the defect density



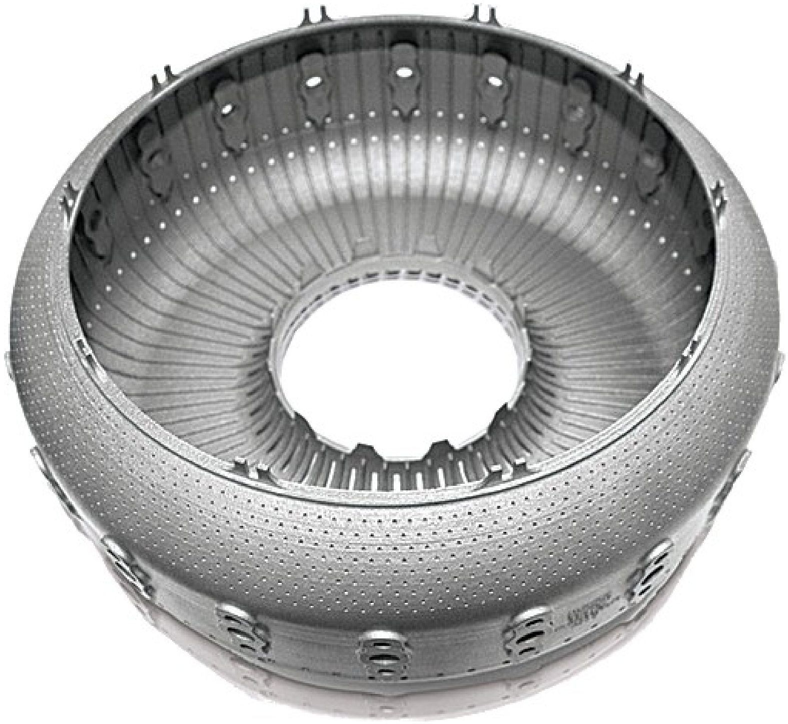
Probabilistic neural network identification of an alloy for direct laser deposition
Materials & Design 168, 107644 (2019)

Testing the oxidation resistance



Probabilistic neural network identification of an alloy for direct laser deposition
Materials & Design 168, 107644 (2019)

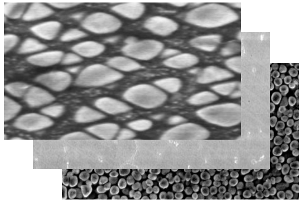
Printing components for an engine



Probabilistic neural network identification of an alloy for direct laser deposition
Materials & Design 168, 107644 (2019)

More materials designed

Nickel and molybdenum



ROLLS ROYCE

Steel for welding

Experiment and DFT for batteries

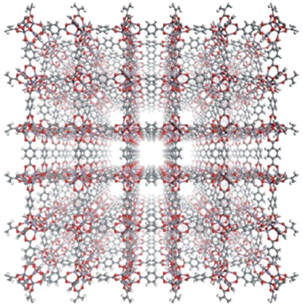
SAMSUNG

Application to industrial chemicals

Ink formulations

DOMINO

Metal organic framework

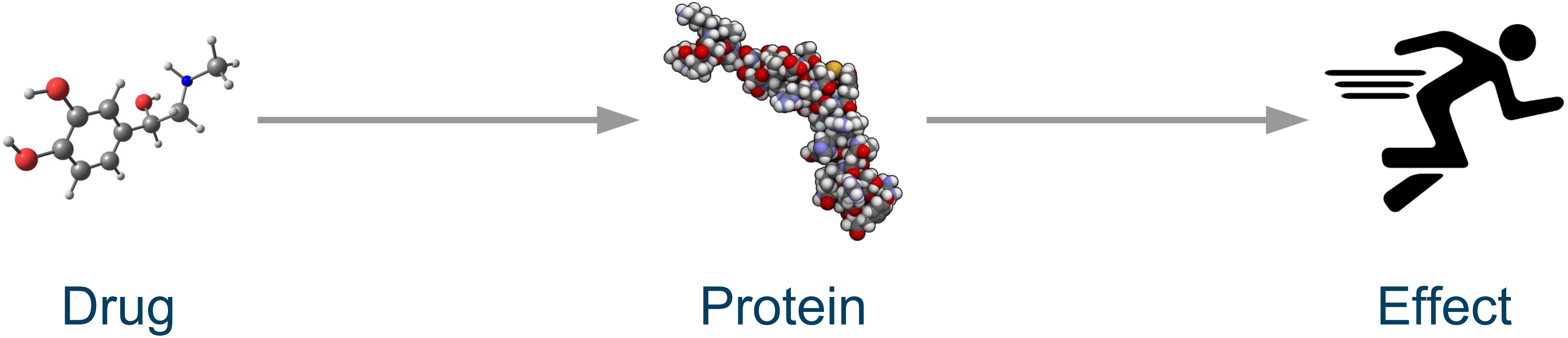


UNIVERSITY OF CAMBRIDGE

Lubricants with molecular dynamics and experiments

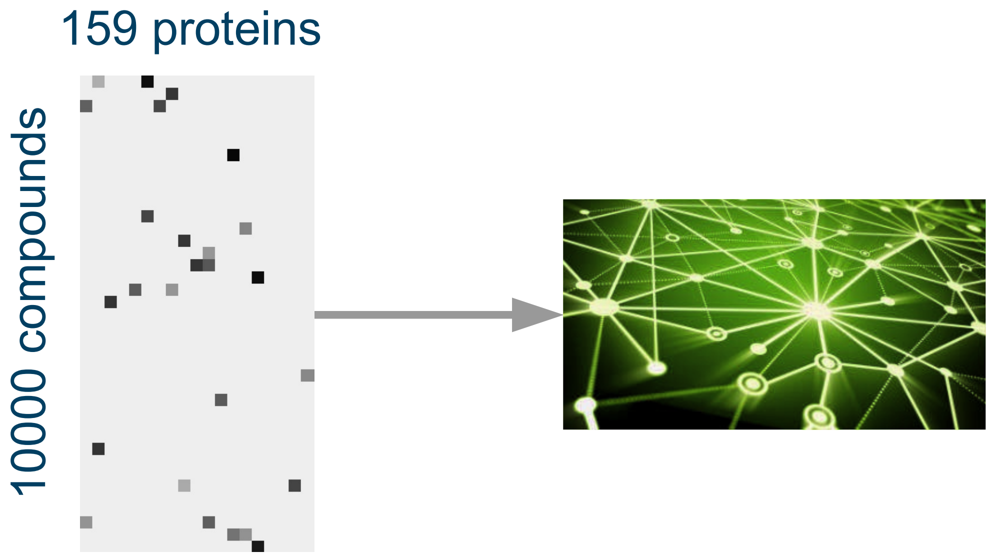
bp

Action of a drug



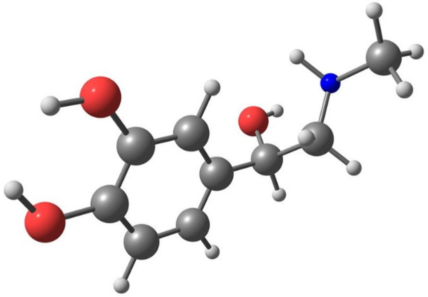
Novartis dataset to benchmark machine learning

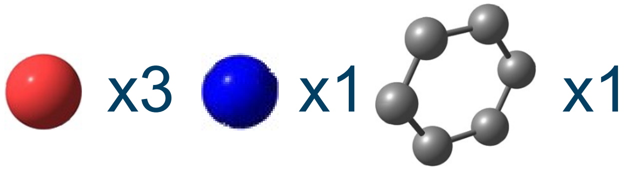
159 kinase proteins, 10000 compounds, data 5% complete



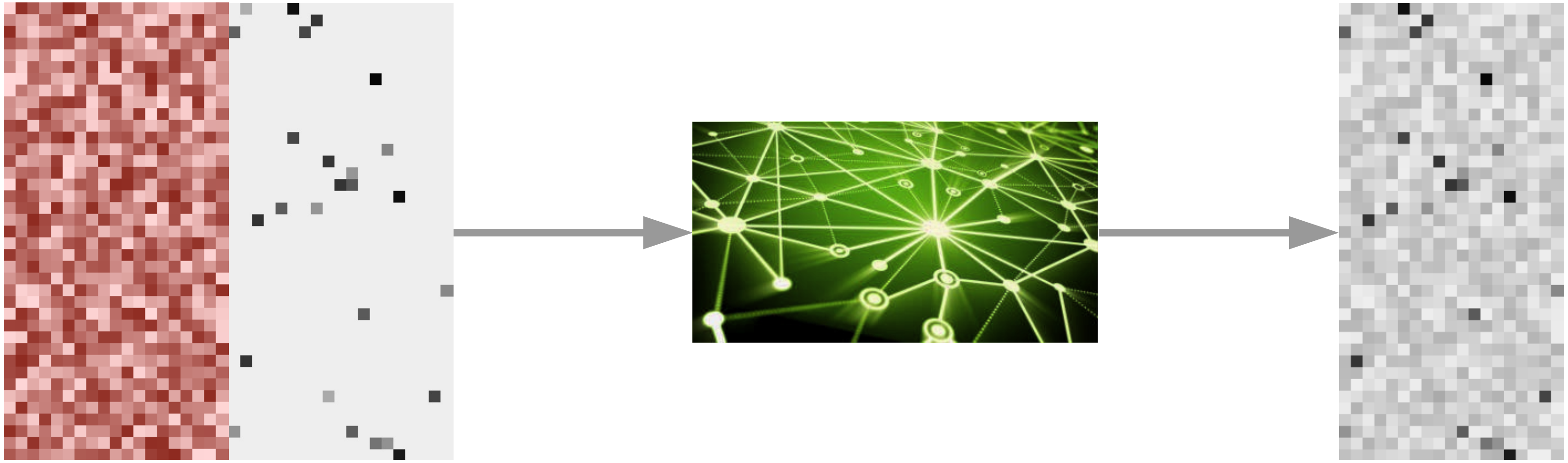
Imputation of Assay Bioactivity Data using Deep Learning
Journal of Chemical Information and Modeling, 59, 1197 (2019)

Quantitative structure-activity relationships



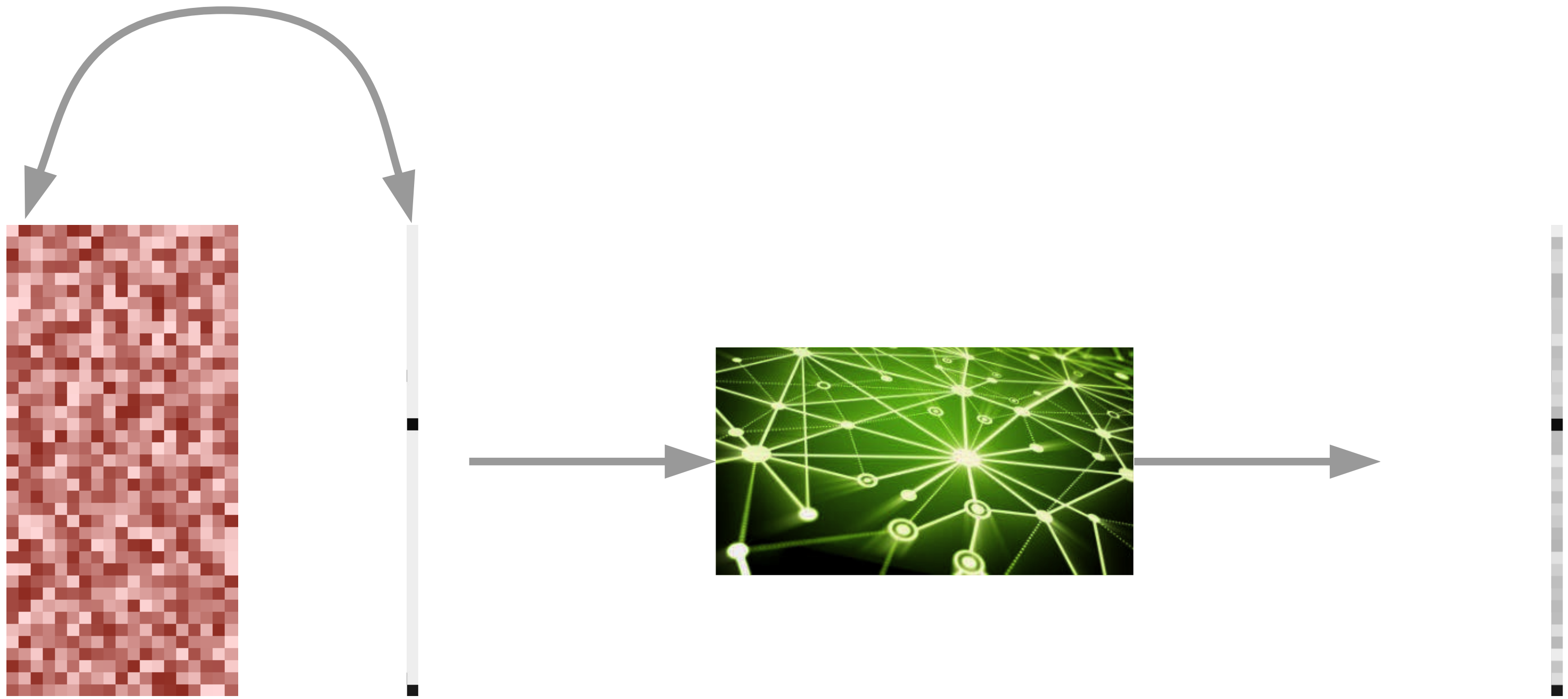


Molecular weight=183 Da



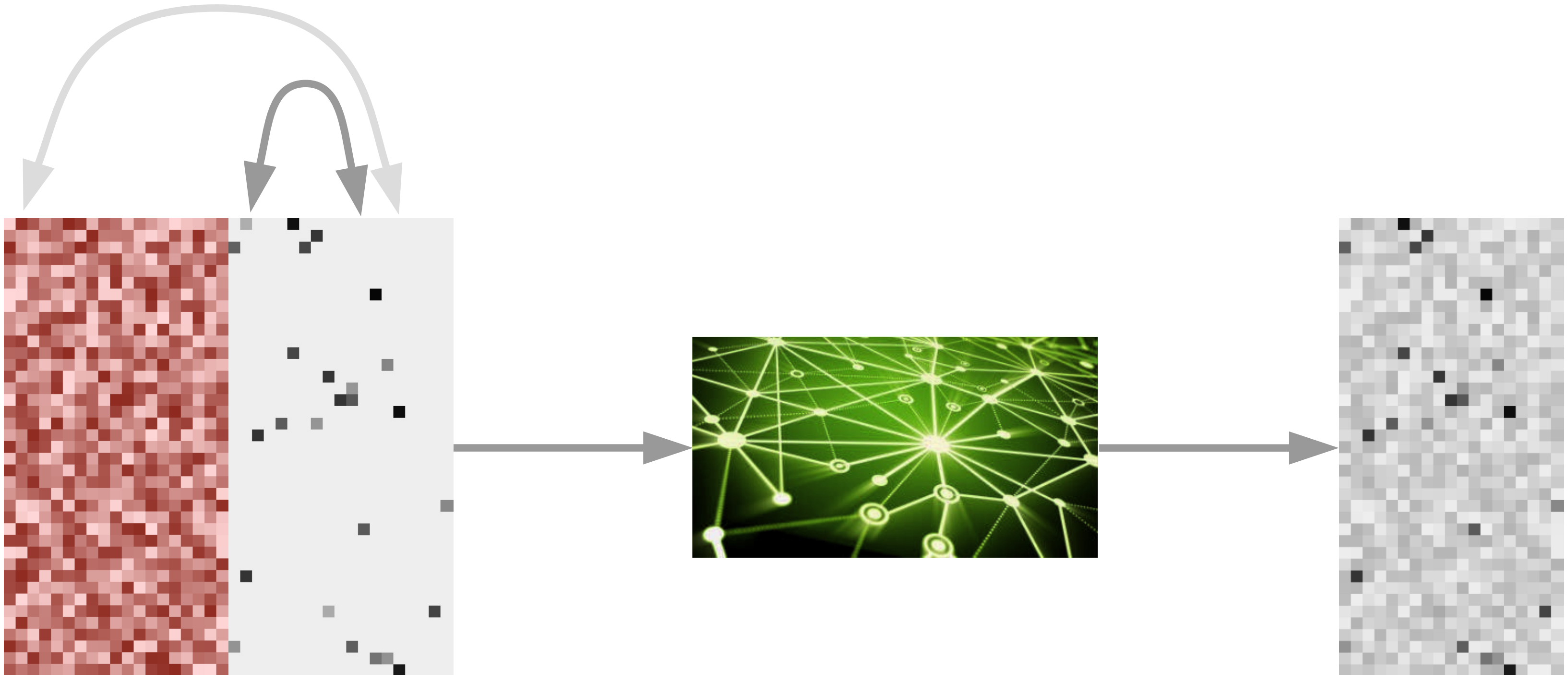
Imputation of Assay Bioactivity Data using Deep Learning
Journal of Chemical Information and Modeling, 59, 1197 (2019)

Quantitative structure-activity relationships



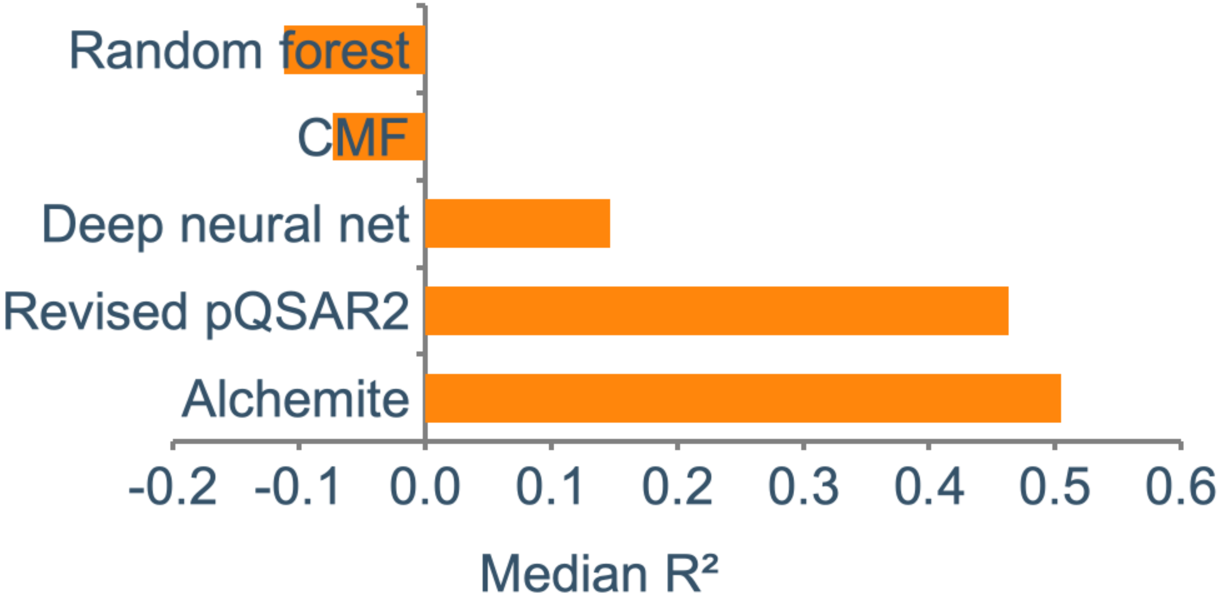
Imputation of Assay Bioactivity Data using Deep Learning
Journal of Chemical Information and Modeling, 59, 1197 (2019)

Exploit protein-protein correlations

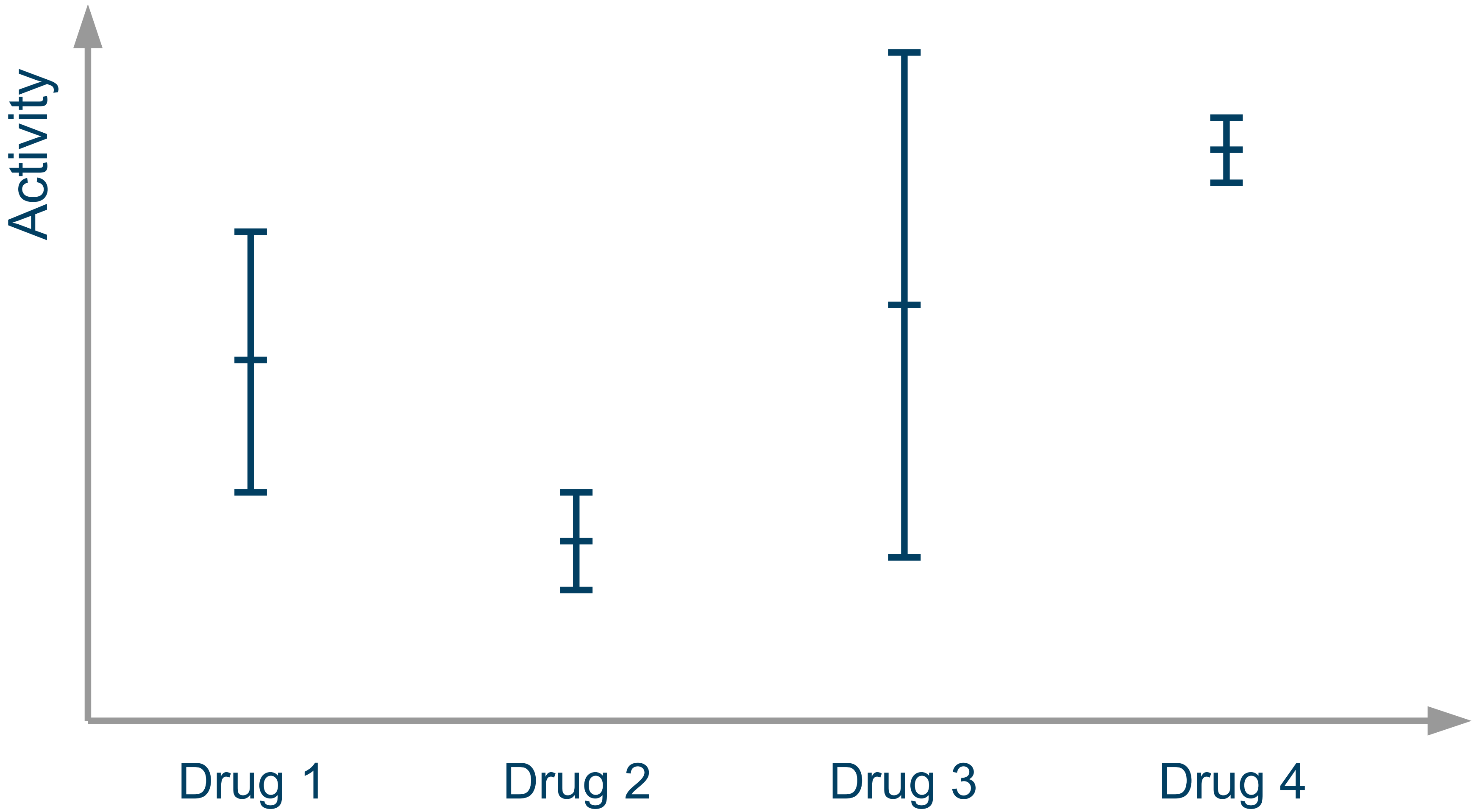


Imputation of Assay Bioactivity Data using Deep Learning
Journal of Chemical Information and Modeling, 59, 1197 (2019)

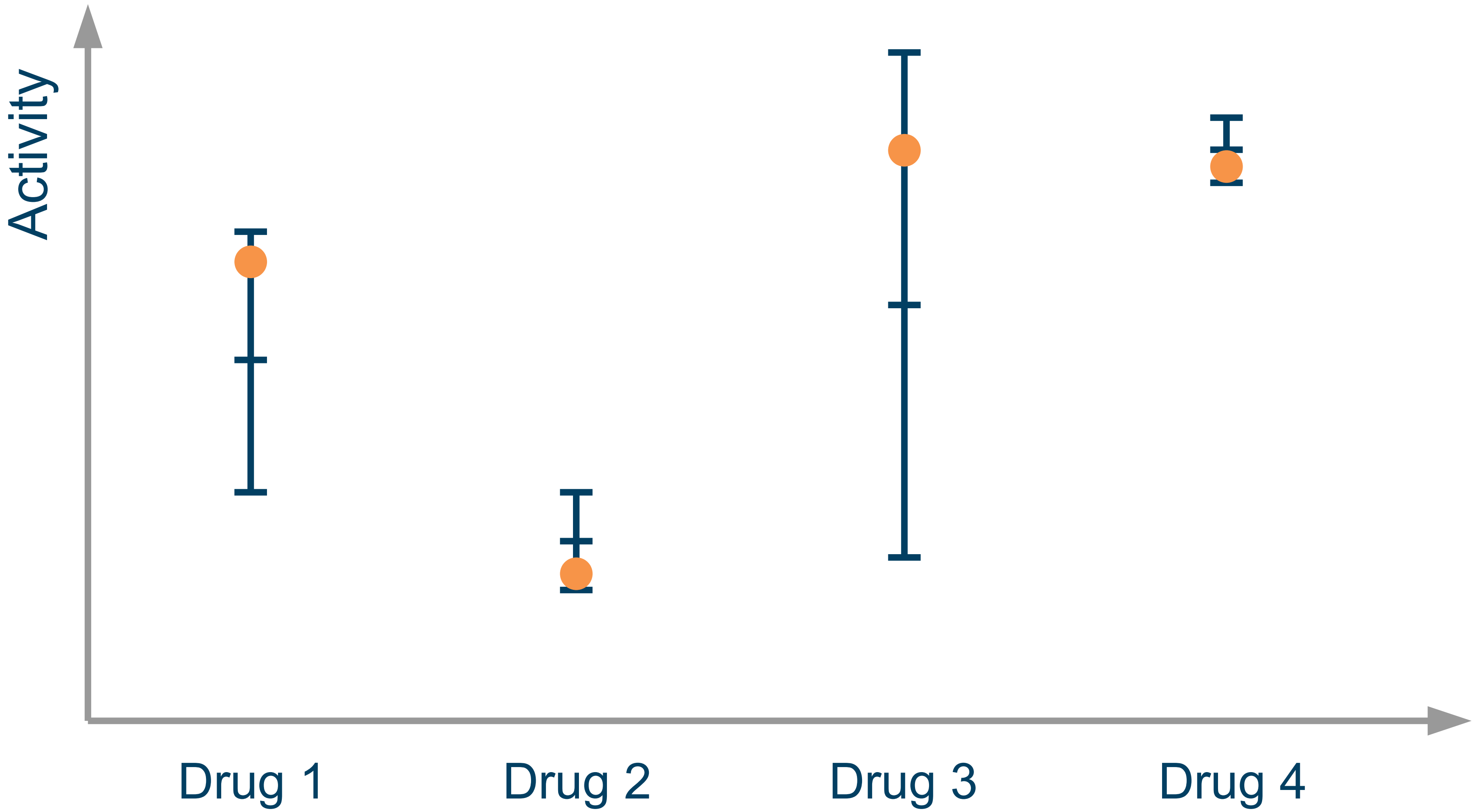
Comparison of techniques



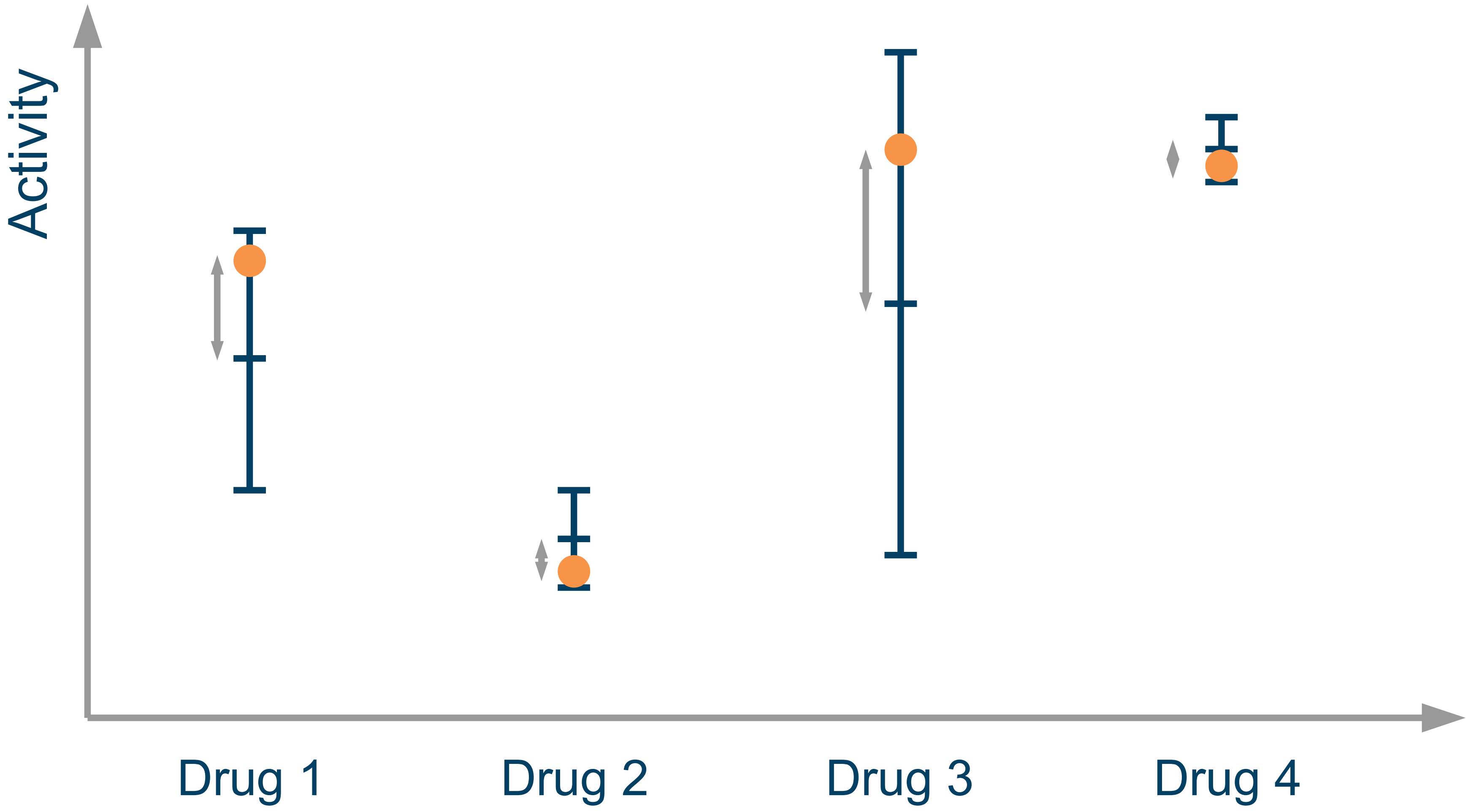
Predictions have an uncertainty



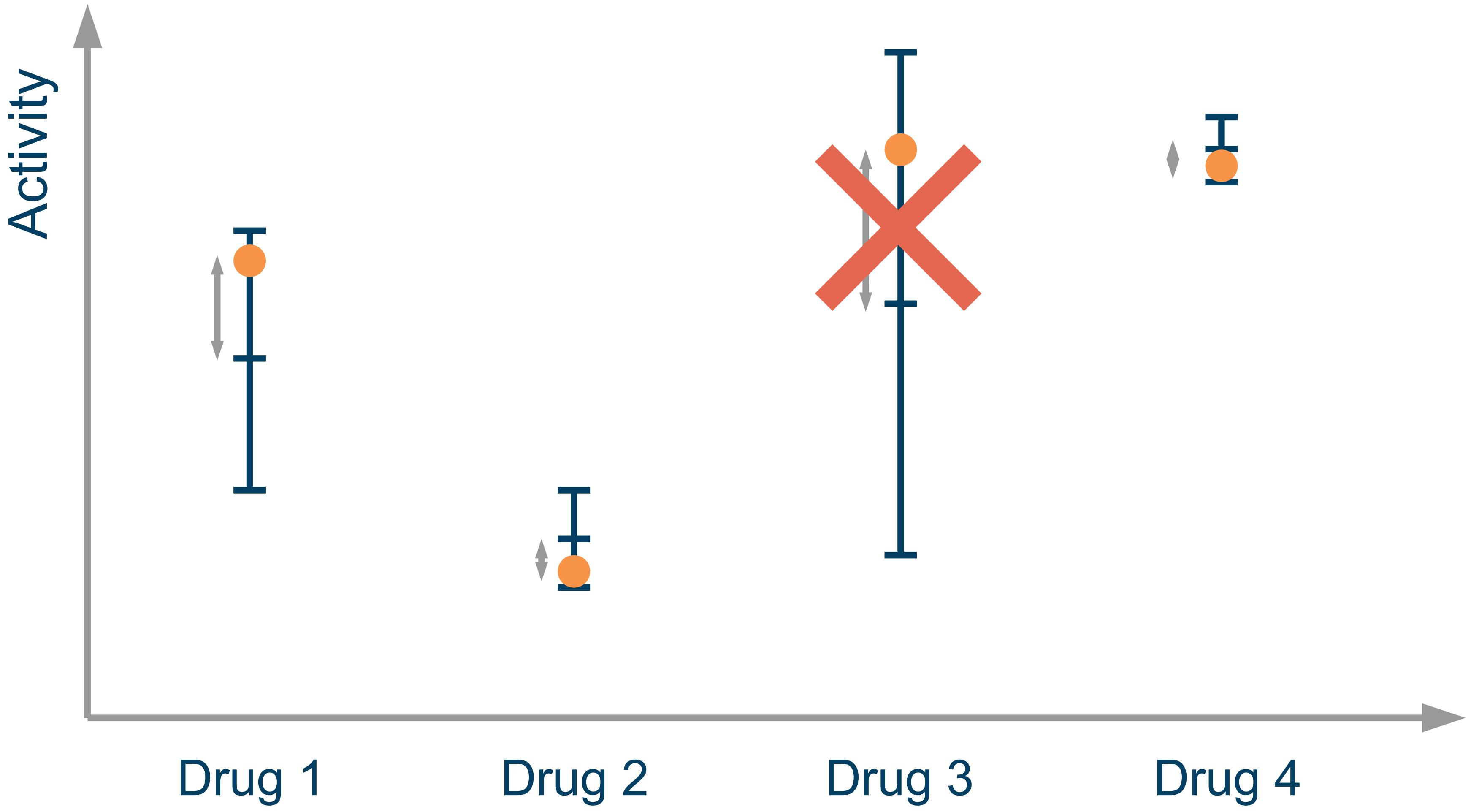
Validation data typically within one standard deviation



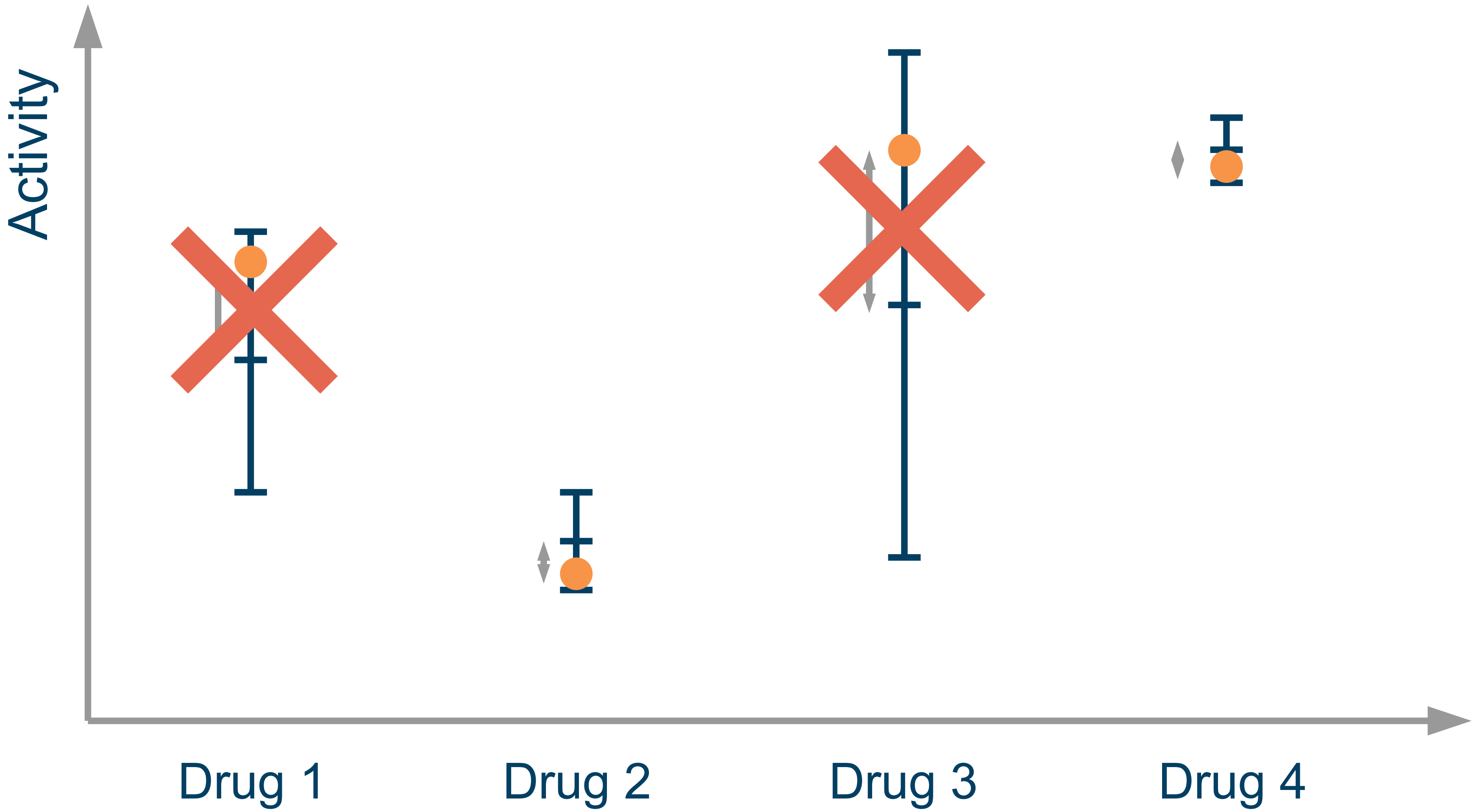
R^2 metric calculated with difference from mean



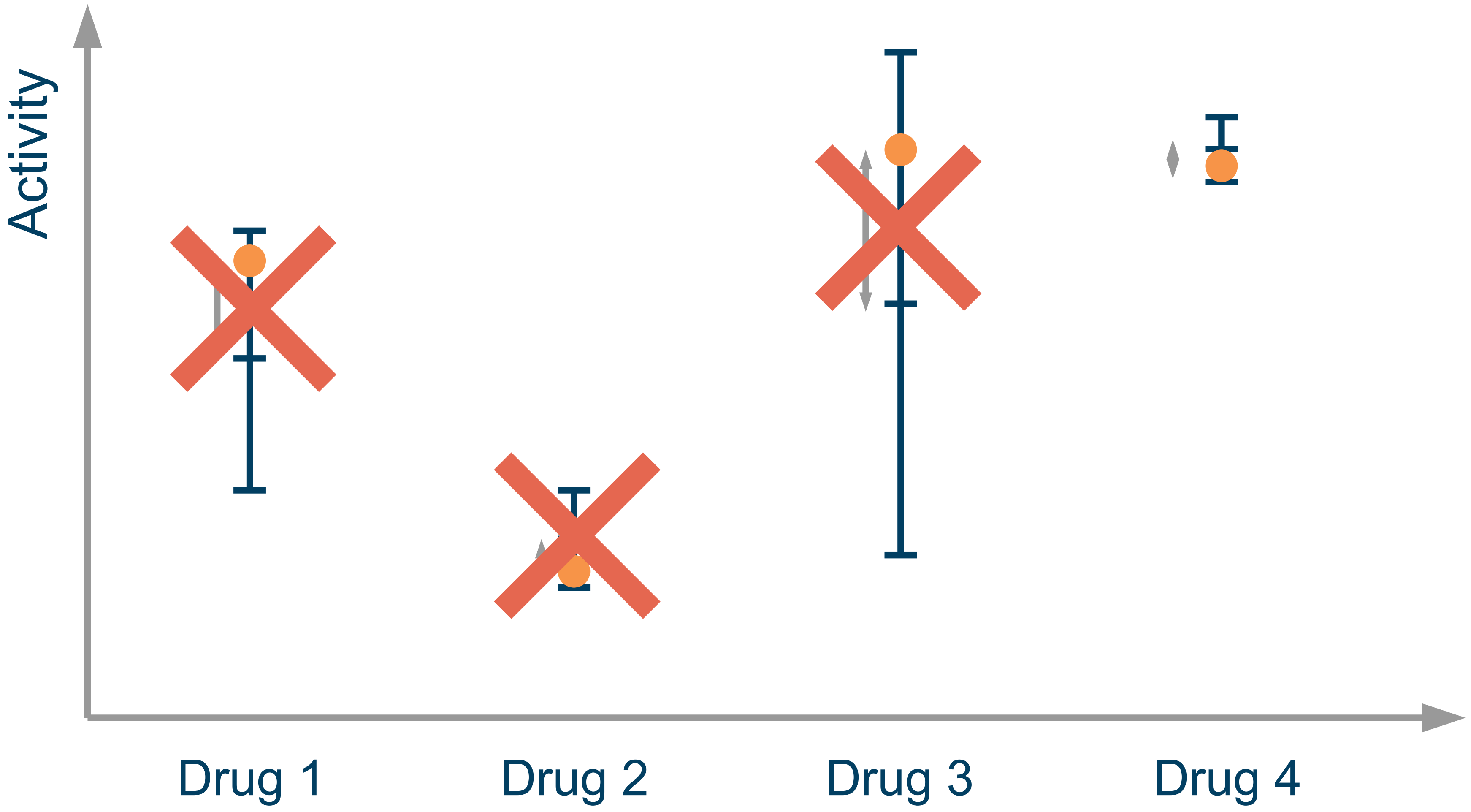
Impute 75% of data with smallest uncertainty



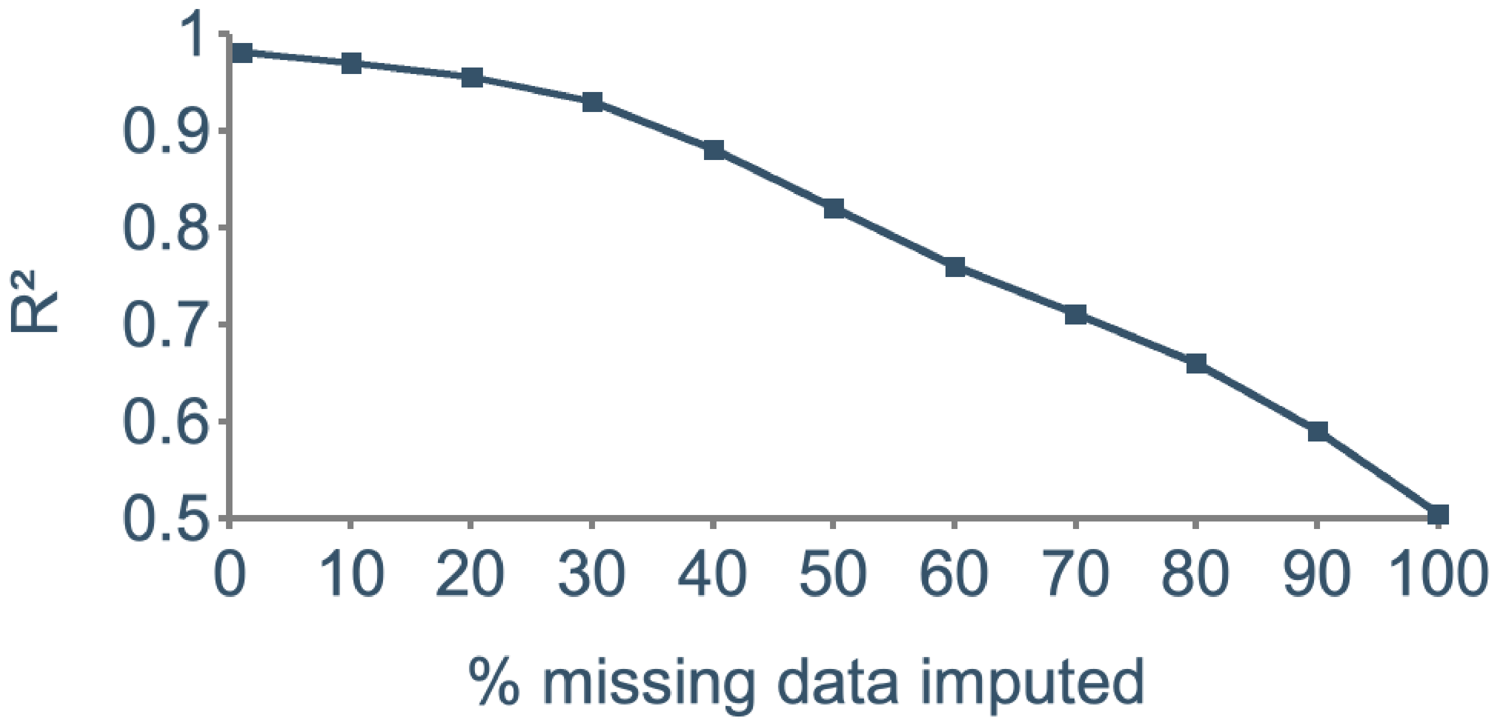
Impute 50% of data with smallest uncertainty



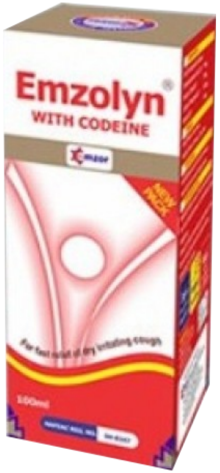
Impute 25% of data with smallest uncertainty



Improved performance by exploiting uncertainty



Different drugs can treat the same ailment

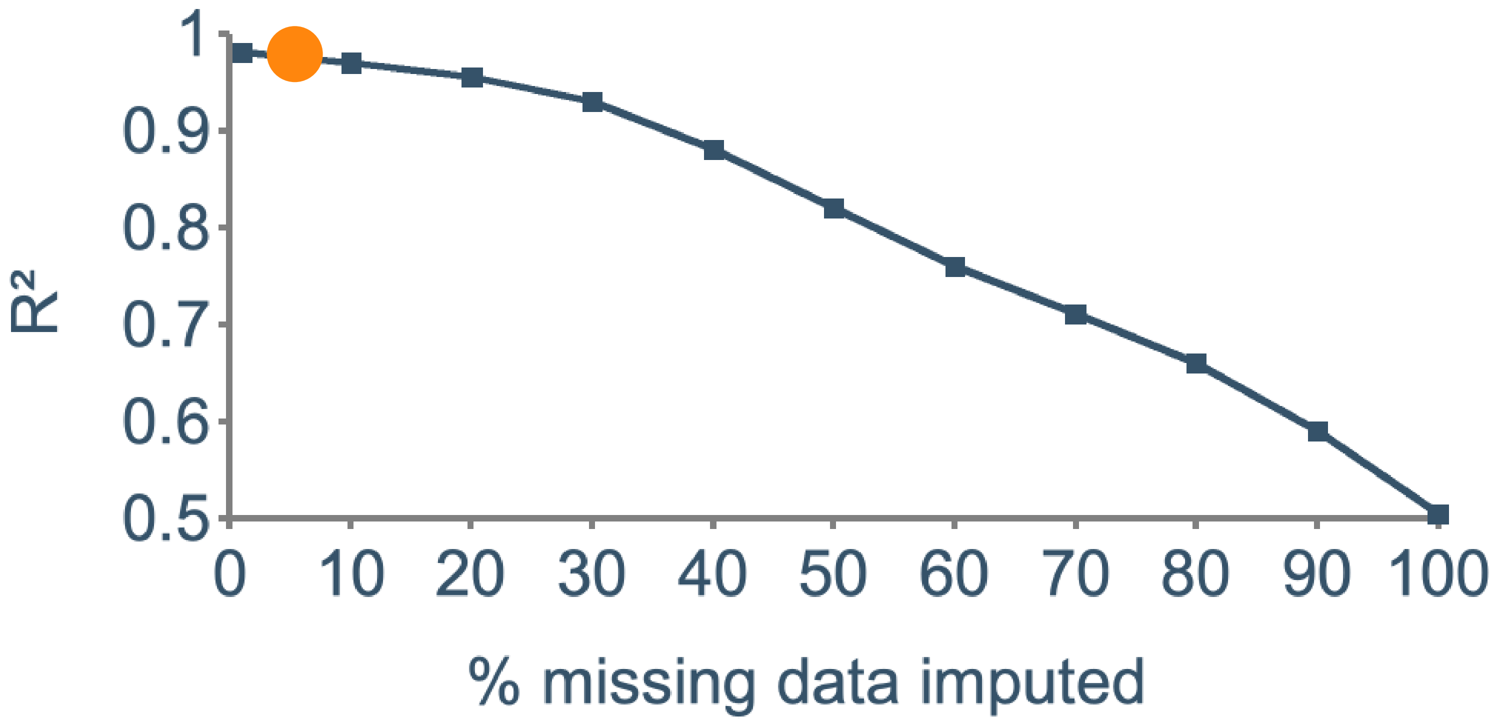


Open Source Malaria contest

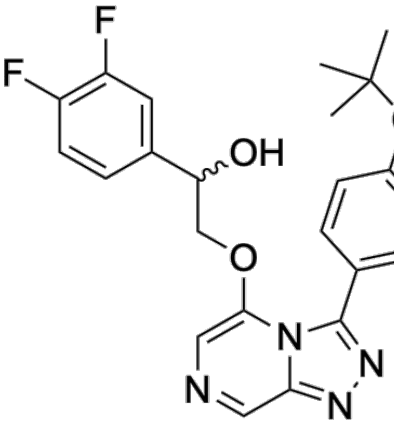
OPEN SOURCE MALARIA

Looking for New Medicines

Focus on compounds with low uncertainty



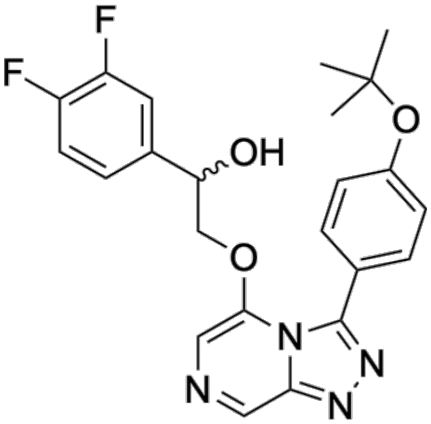
Open Source Malaria experimental validation



Optibrium & Intellegens

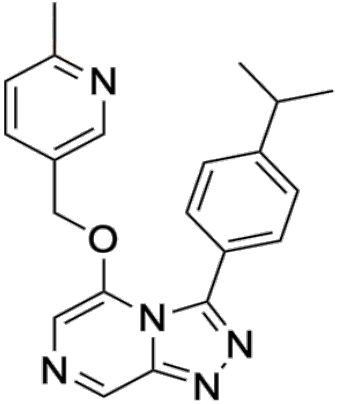
0.647 µM

Open Source Malaria other compounds



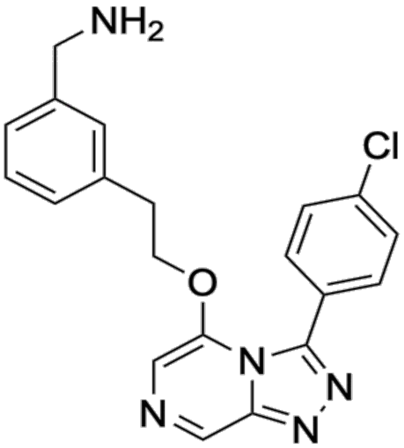
Optibrium & Intellegens

0.647 μM



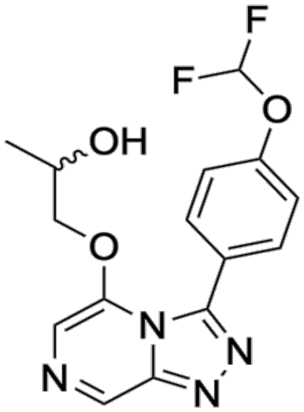
Davy Guan

>25 μM



Exscientia

10.9 μM



Molomics

>25 μM

Commercialization

Alchemite Analytics platform for materials and chemicals with Intellegens

Machine learning tool embedded into next generation of Optibrium software
Cerella™ released in October 2020

intellegens

optibrium

Summary

Merge different experimental quantities and computer simulations into a **holistic** design tool

Designed and experimentally verified alloy for **direct laser deposition**

Designed experimentally verified **drug** in Open Source Malaria competition

Taken to market through startup **Intellegens**