Accelerate AI

BUSINESS & INNOVATION

The Future of AI Gathers Here.

Sept 19 - 22 | London

Imputation of assay activity data using deep learning

Tom Whitehead, Peter Hunt, Matt Segall, Gareth Conduit
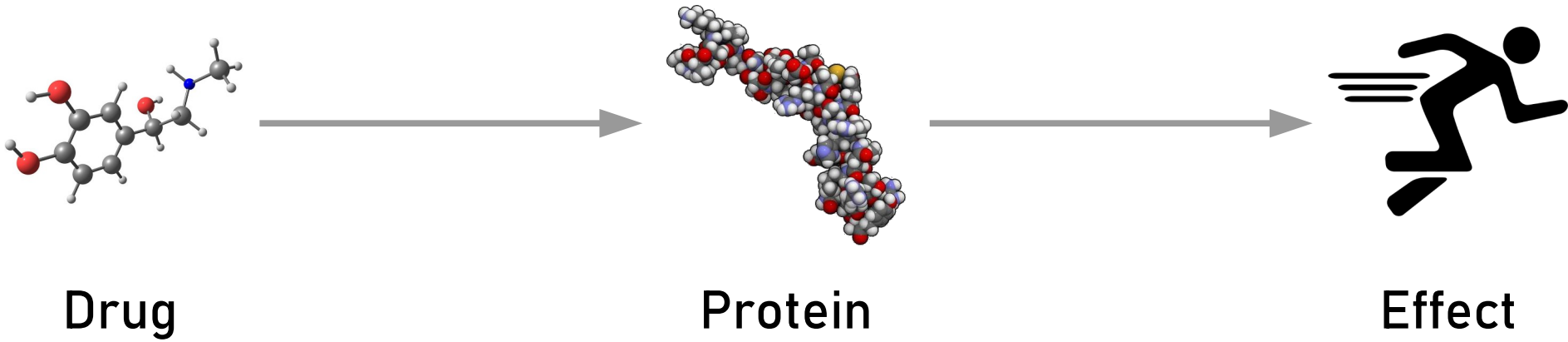
Neural network algorithm to

Reduce the need for experiments and **accelerate** drug discovery

Utilise **all available** information: computer simulations and real-life measurements
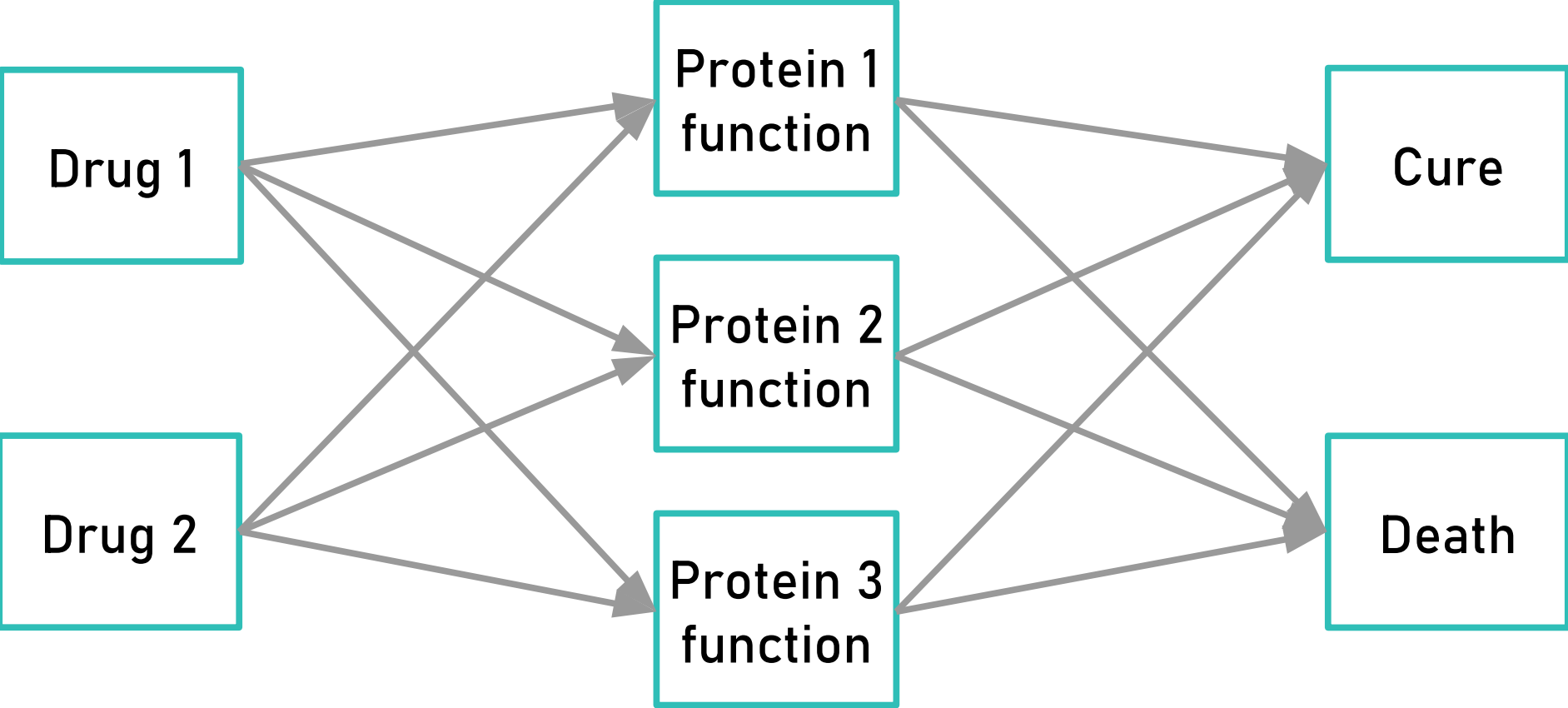
Impute values from sparse data

Generic with **proven** applications in drug design and materials discovery
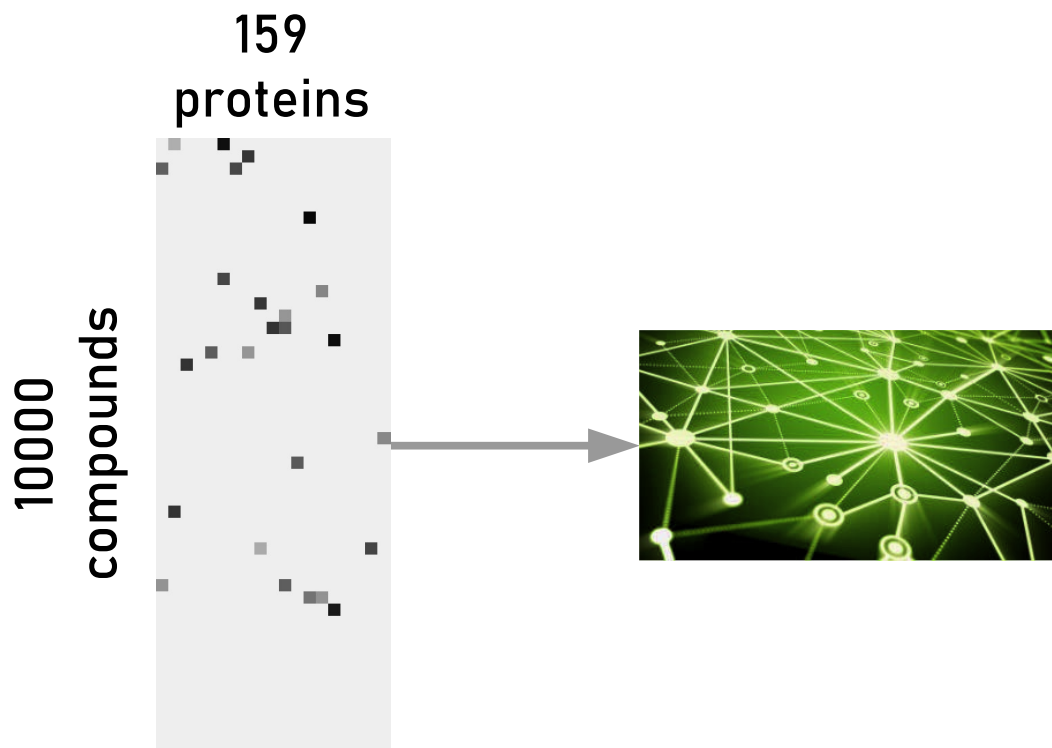
Action of a drug

Drug   Protein   Effect

Action of a drug

Novartis dataset to benchmark machine learning

159 kinase proteins, 10000 compounds, data 5% complete

159 proteins

10000 compounds

Want to impute missing entries
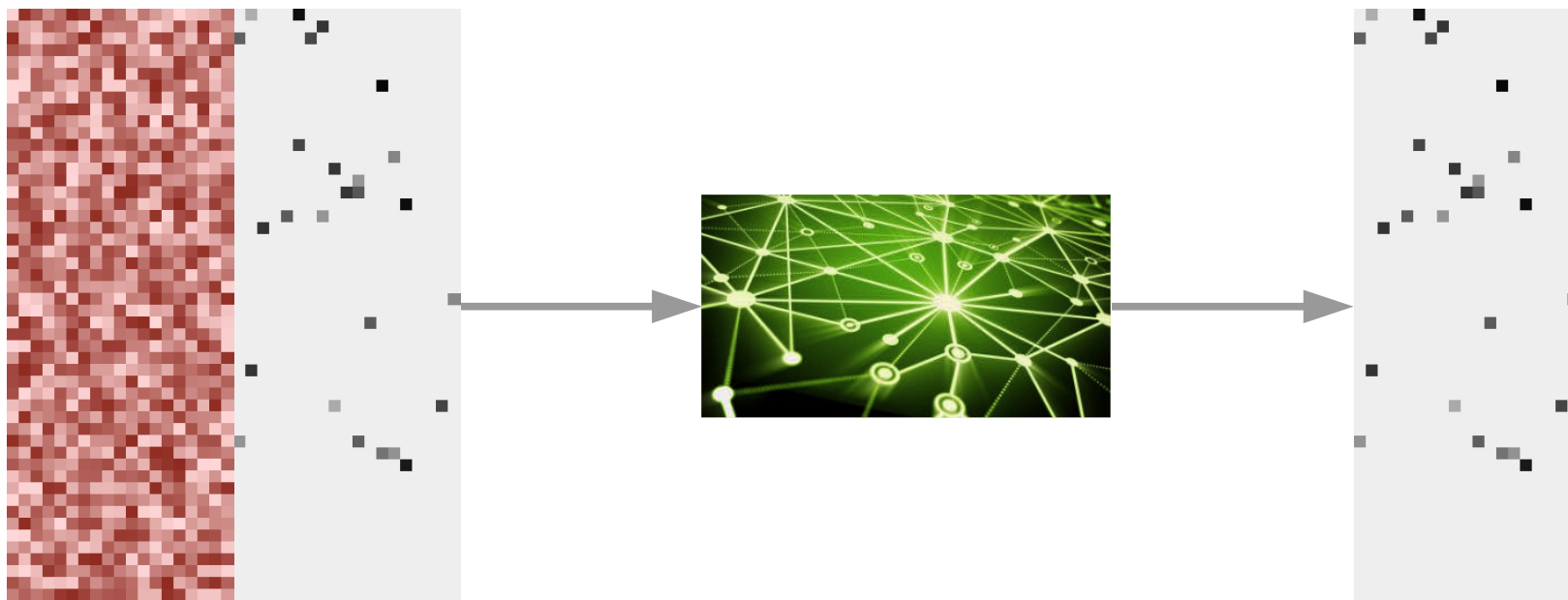

Random

Realistic

Want to impute missing entries

Validate using a realistically split holdout data set, extrapolate to new chemical space
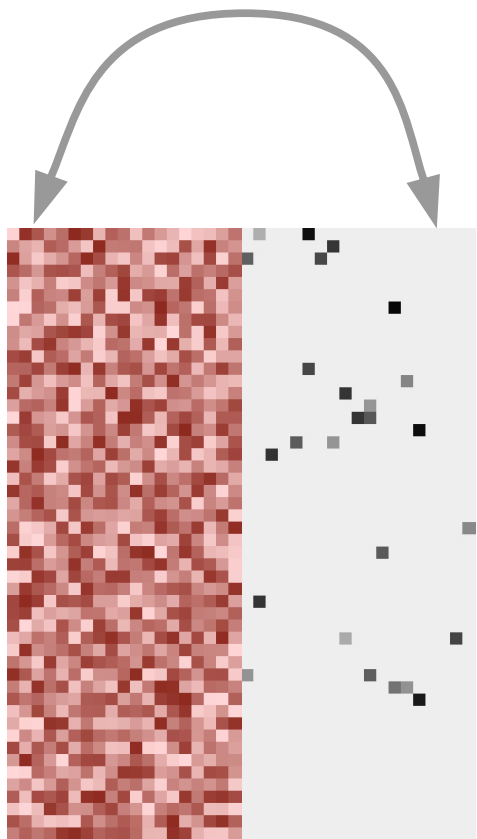
QSAR: quantitative structure–activity relationships
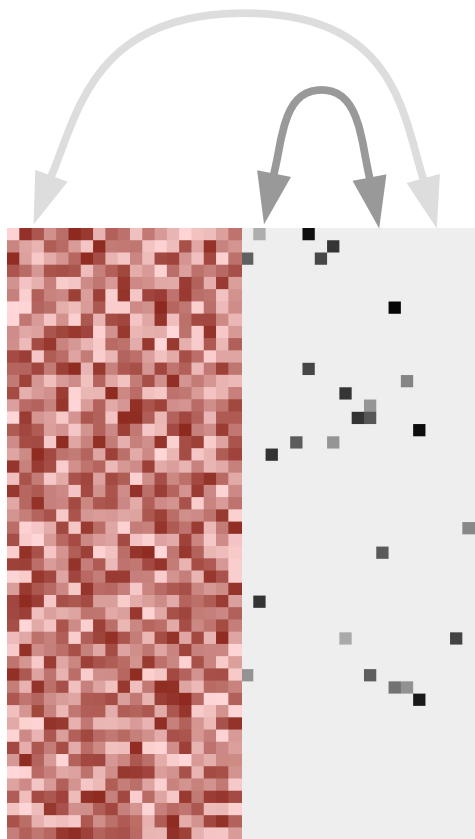

x3 x1 x1
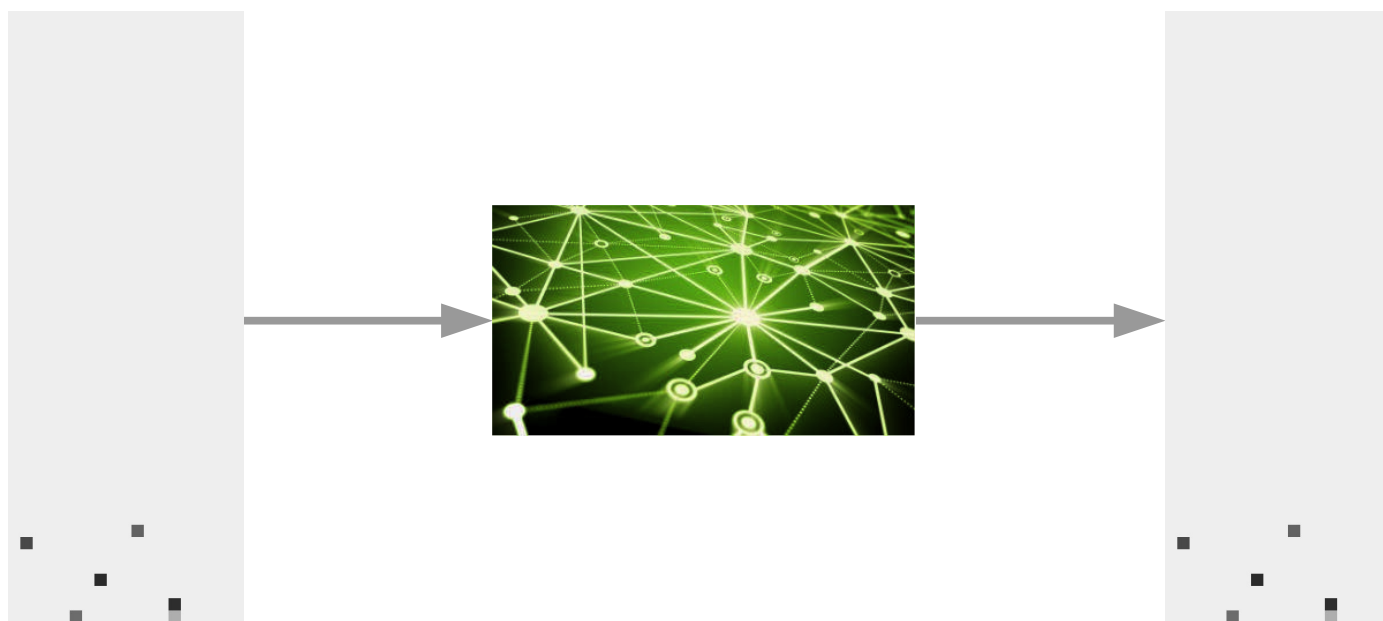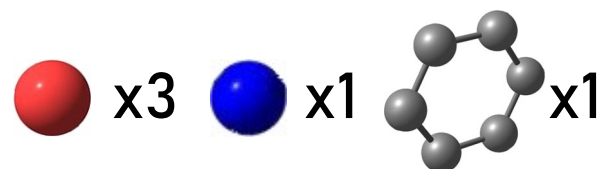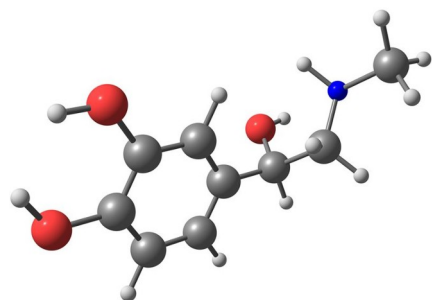
Molecular weight=183 Da

QSAR: quantitative structure–activity relationships

Standard methods learn chemical
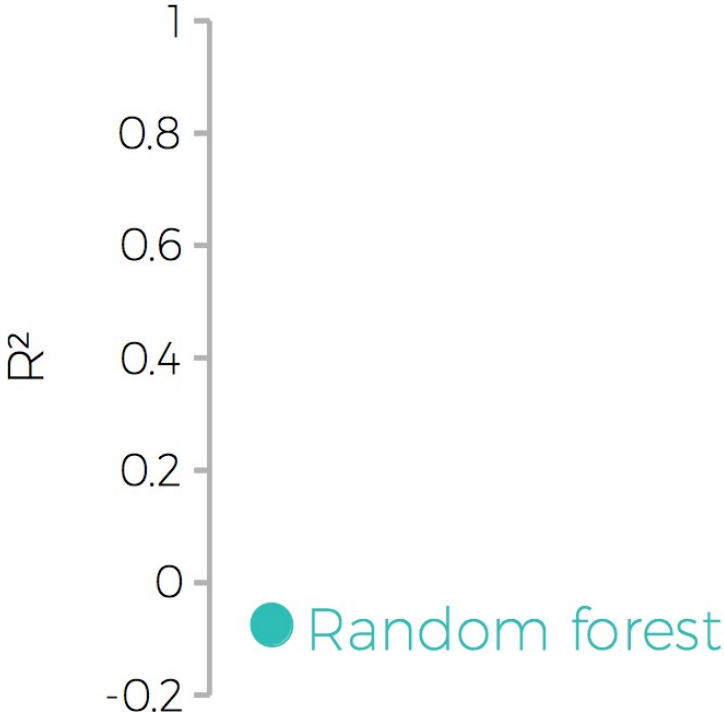descriptor–protein correlations

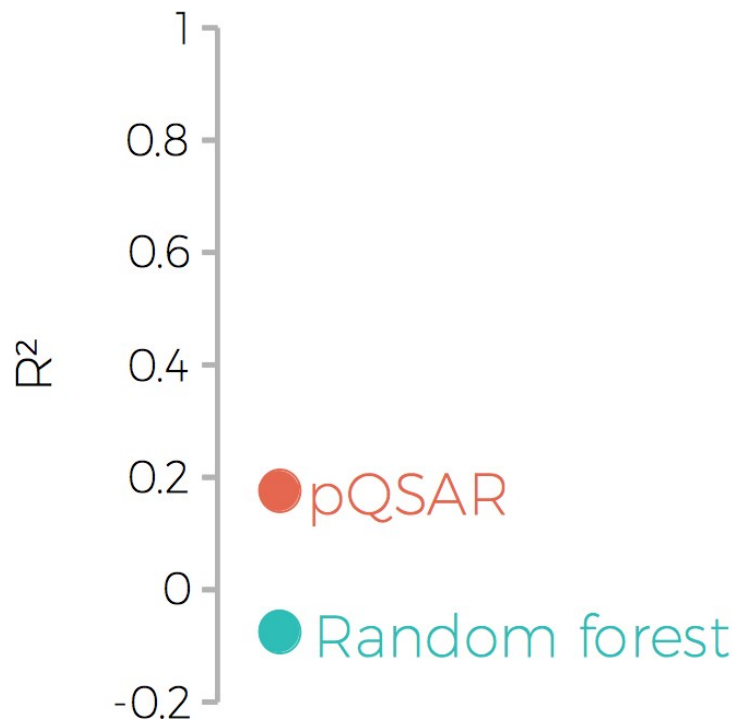# QSAR: quantitative structure–activity relationships

Standard methods learn chemical descriptor–protein correlations

Deep learning also learns the strong protein–protein correlations
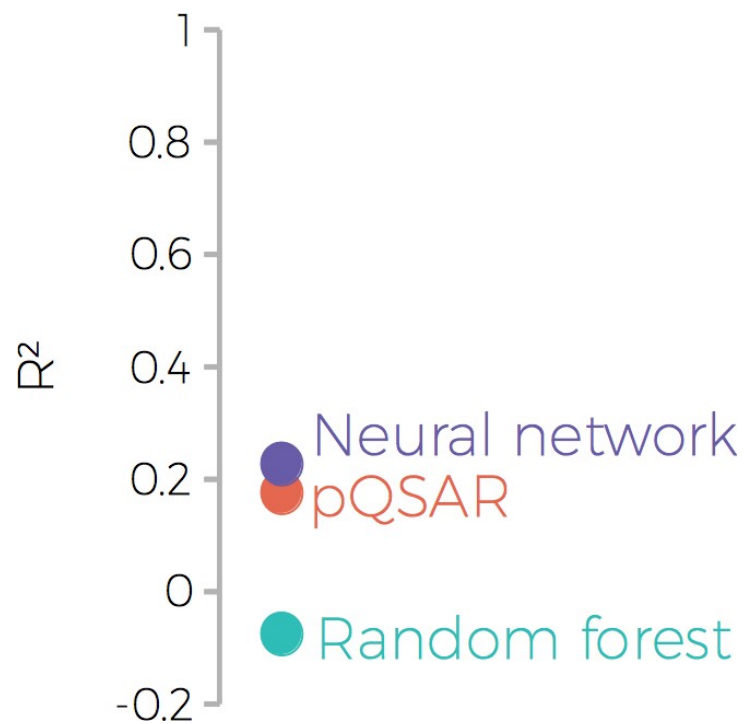
Random forest
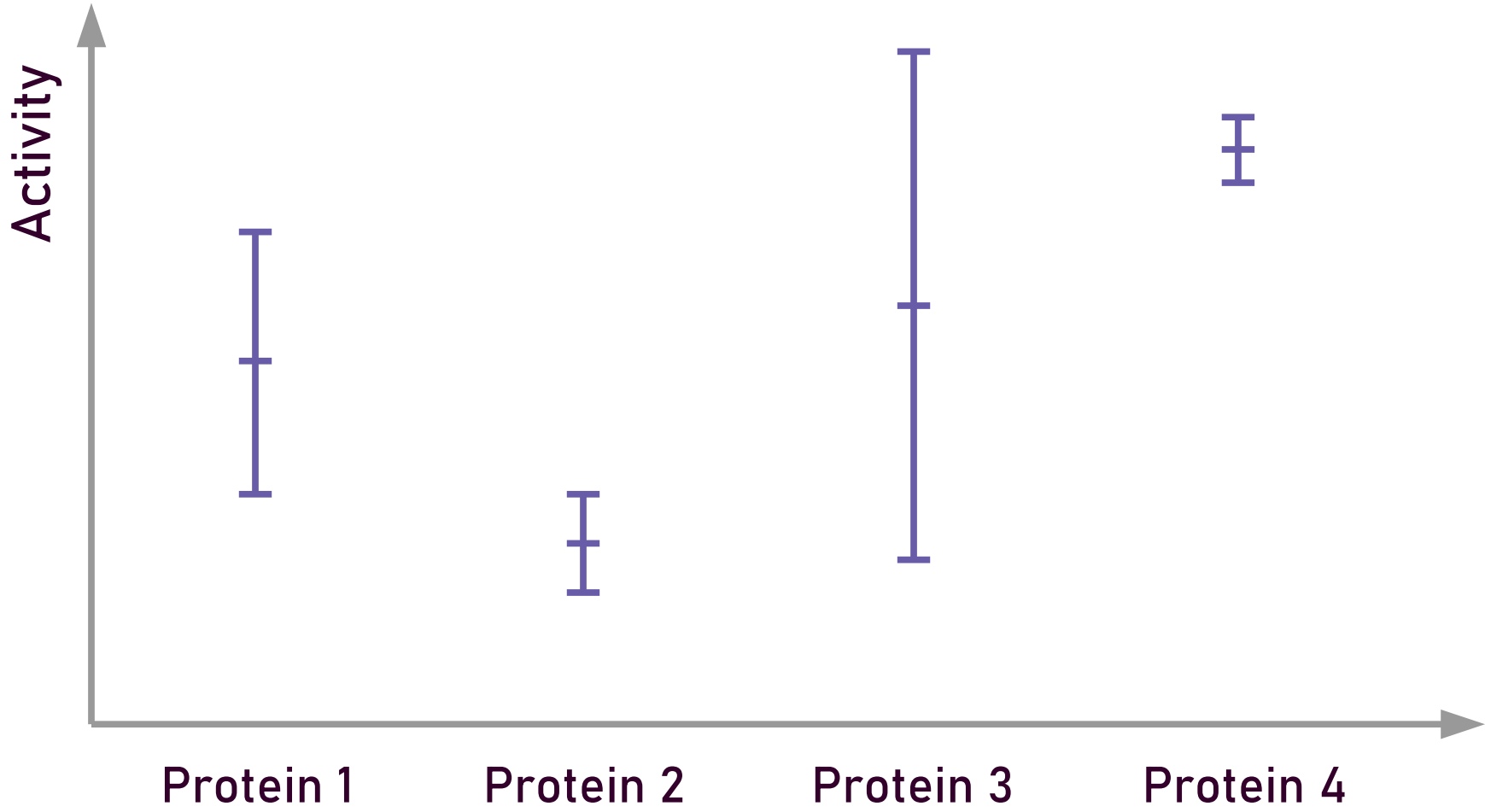
Predictions from pQSAR

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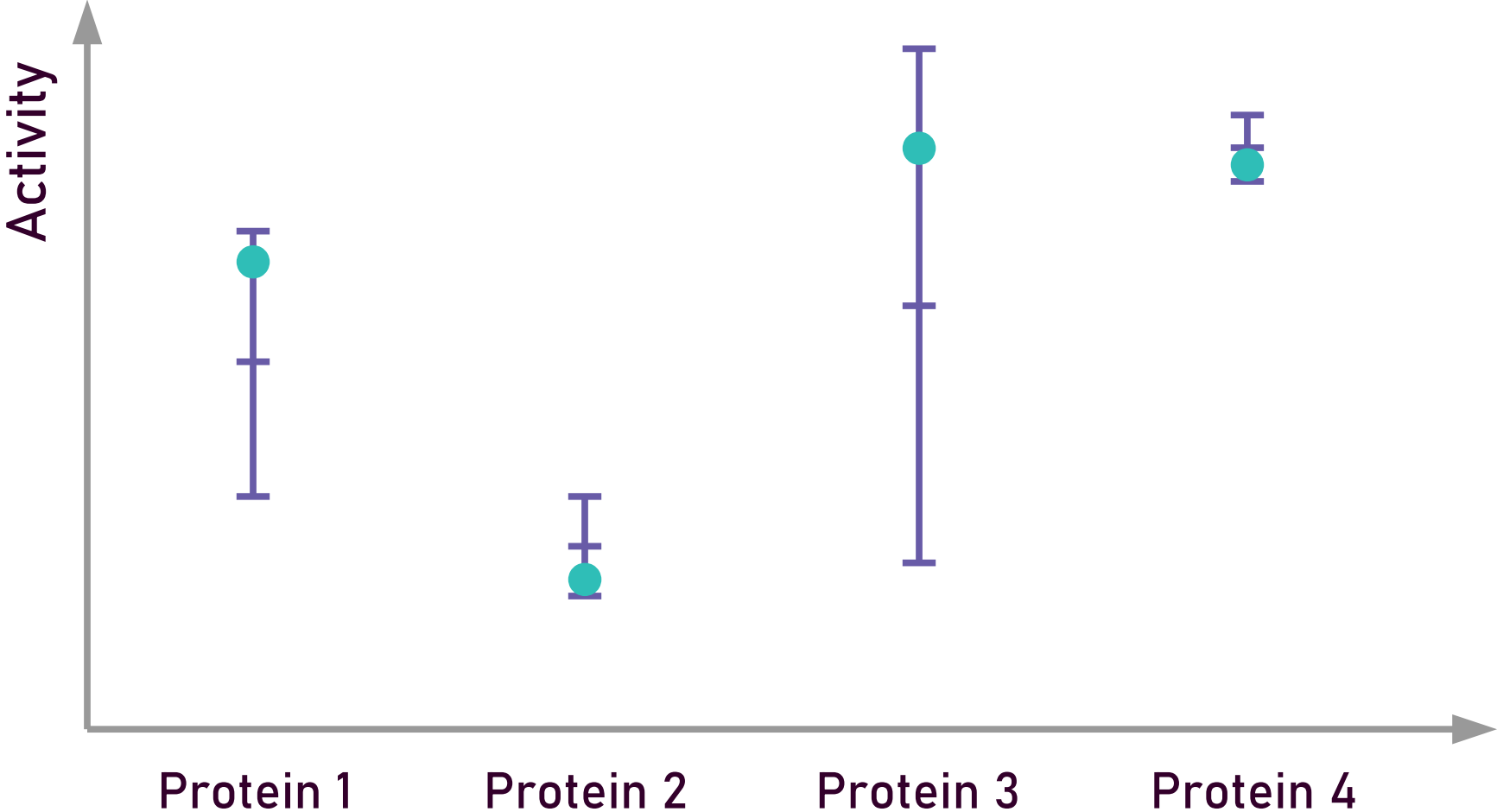J. Chem. Inf. Model. 57, 2077 (2017)

Predictions by the neural network

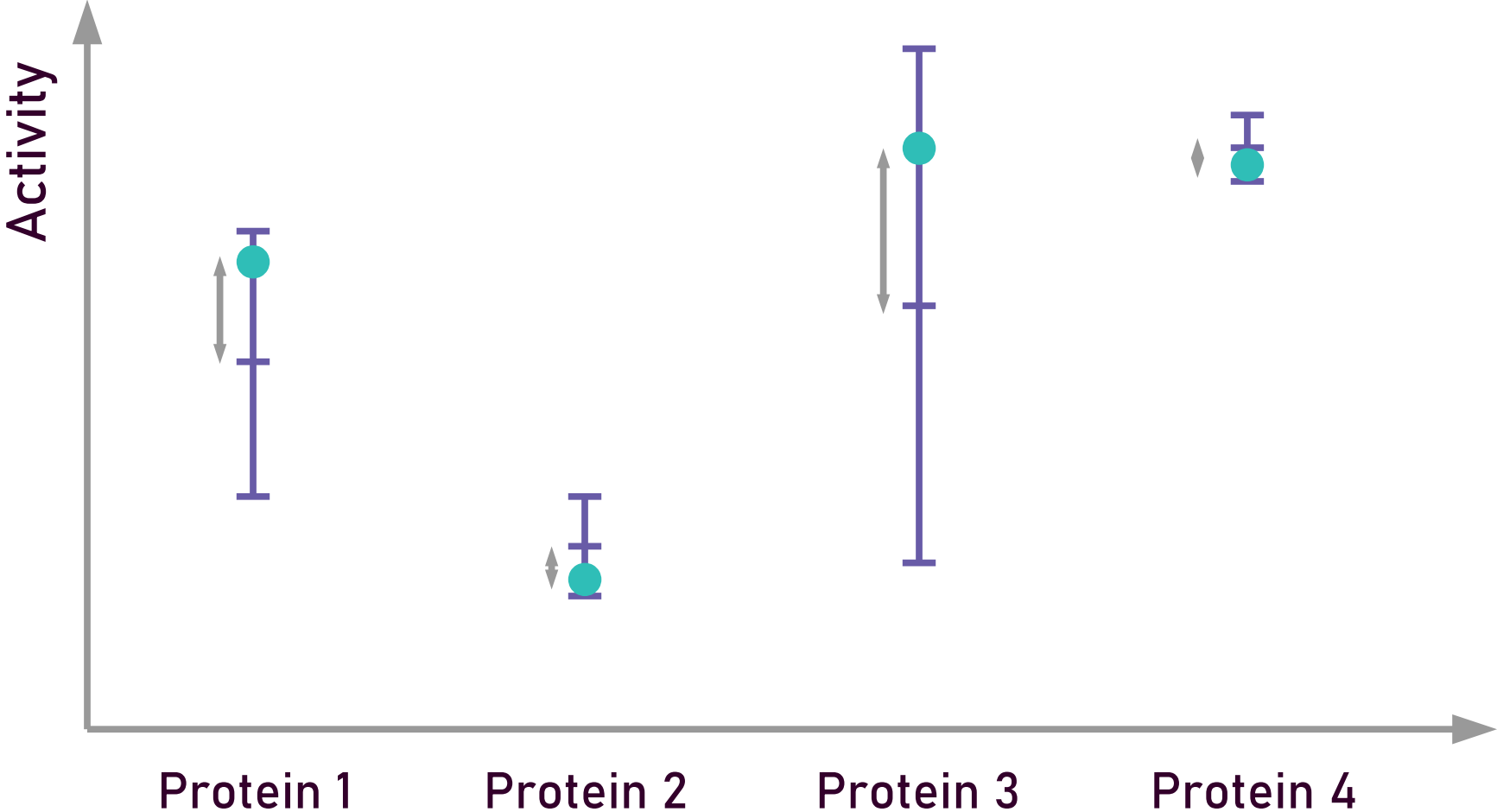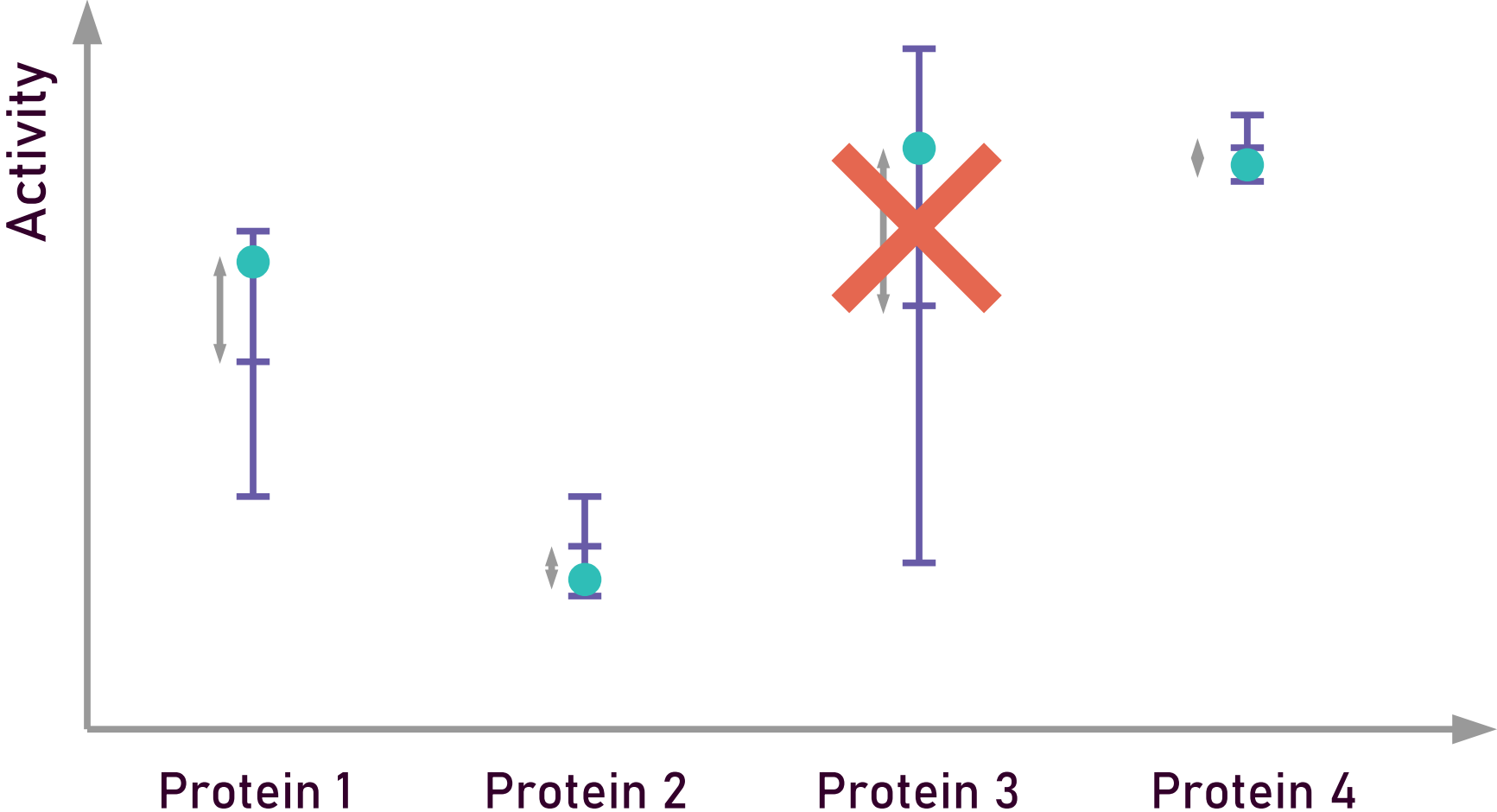# Predicted activities have an uncertainty

Validation data within one standard deviation

Activity

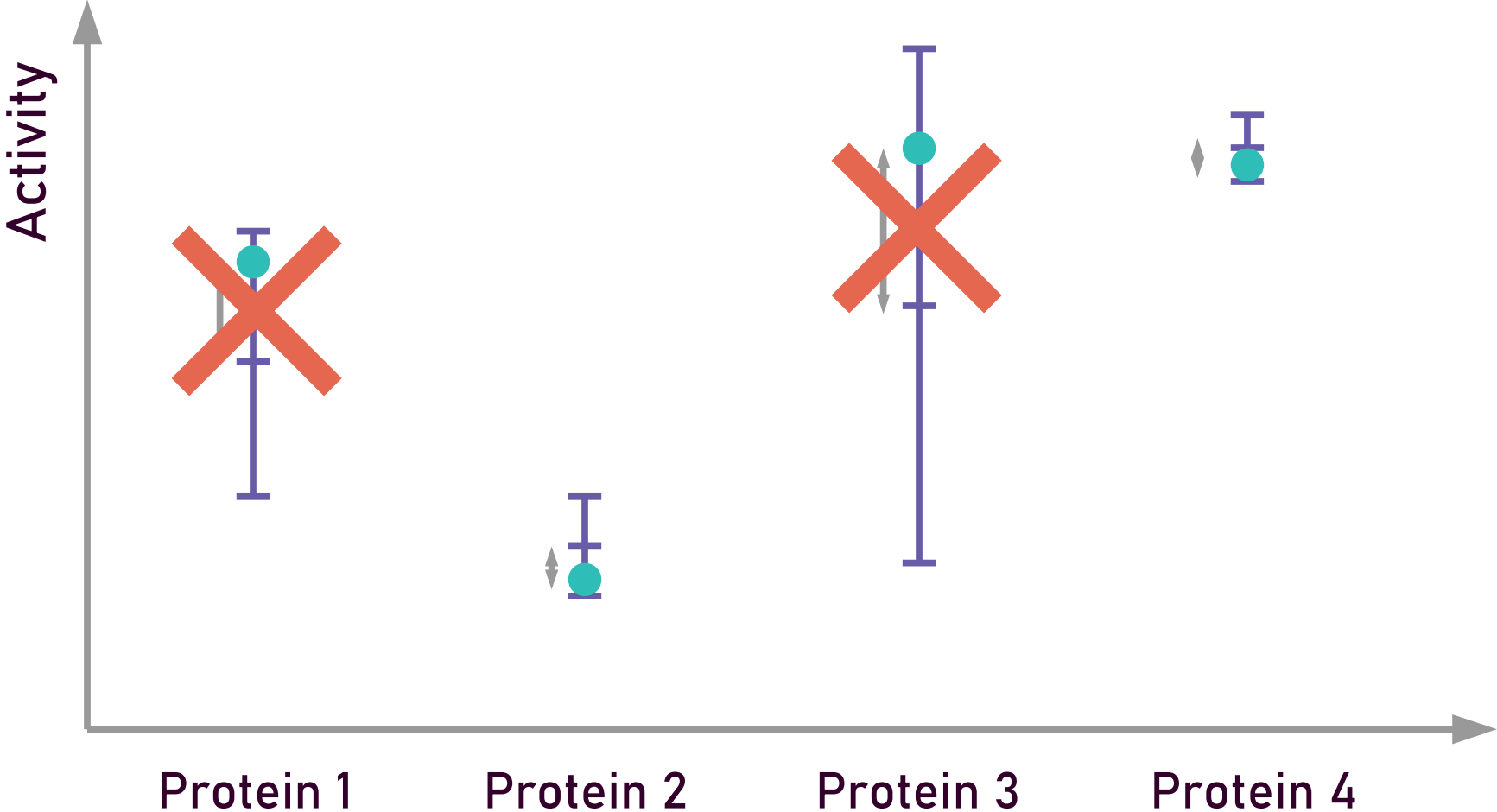Protein 1   Protein 2   Protein 3   Protein 4

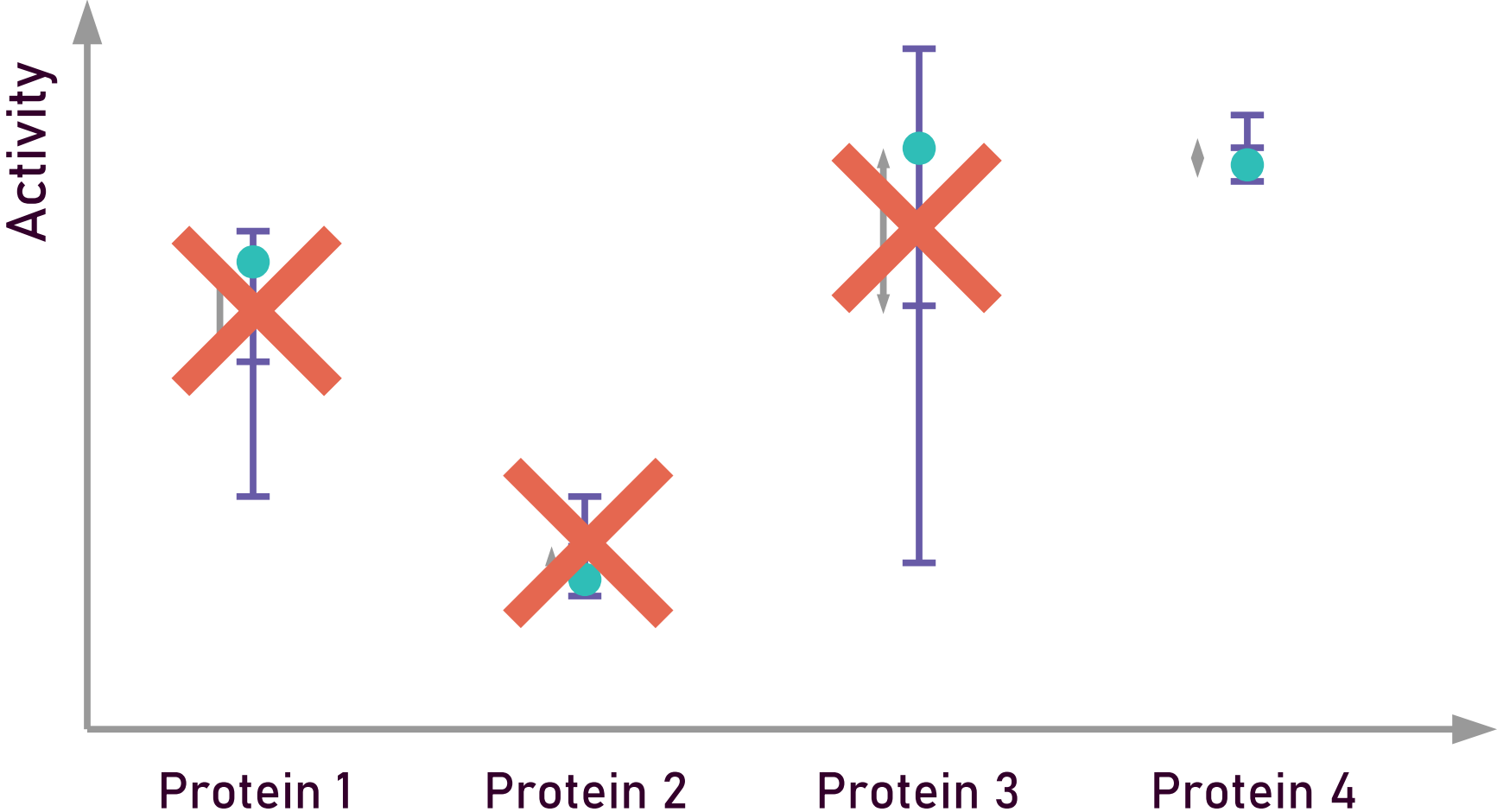R^2 metric calculated with difference from mean

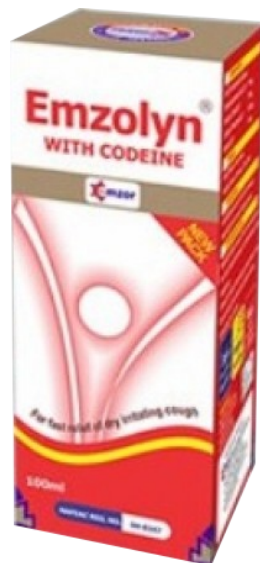# Impute 75% of data with smallest uncertainty

Impute 50% of data with smallest uncertainty

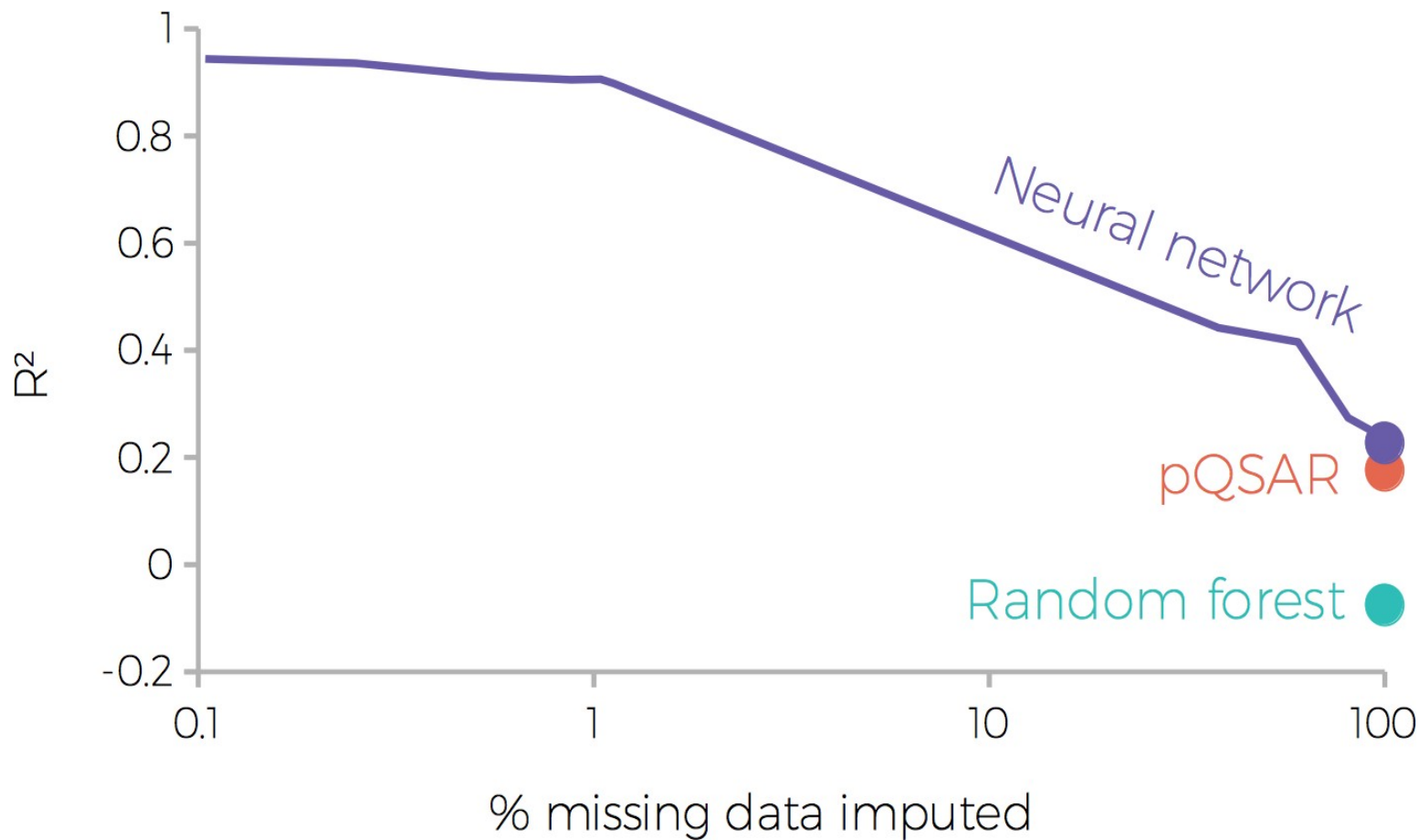# Impute 25% of data with smallest uncertainty

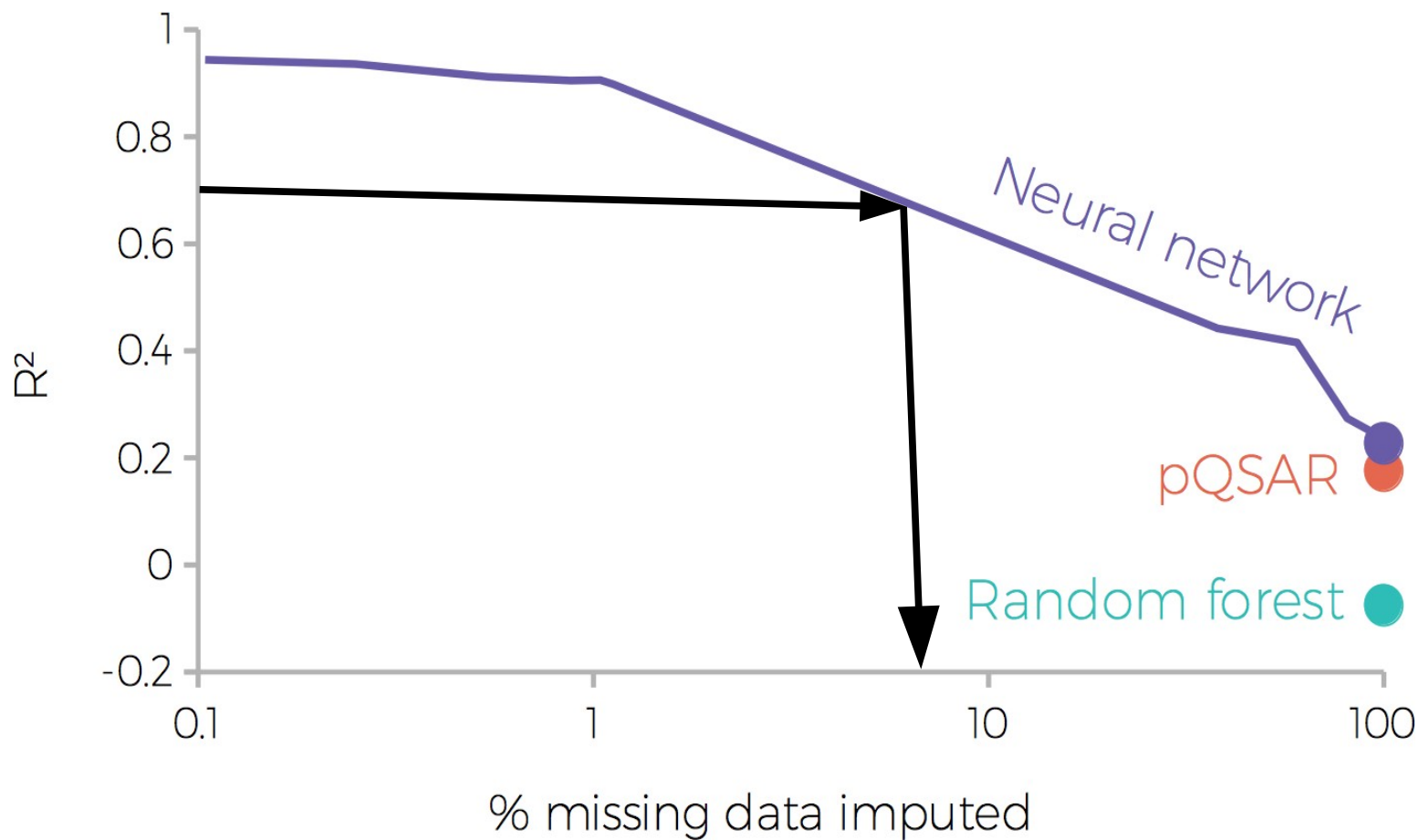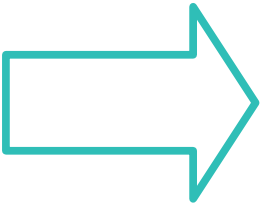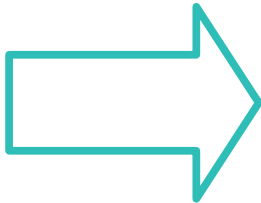# Different drugs can treat the same ailment
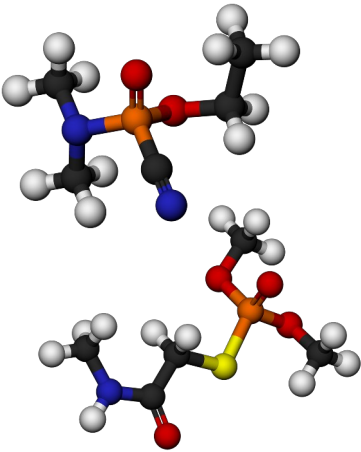
Improved performance by exploiting uncertainties
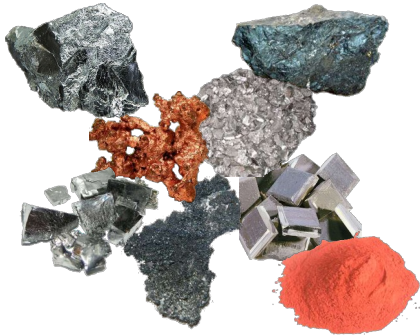
Improved performance by exploiting uncertainties

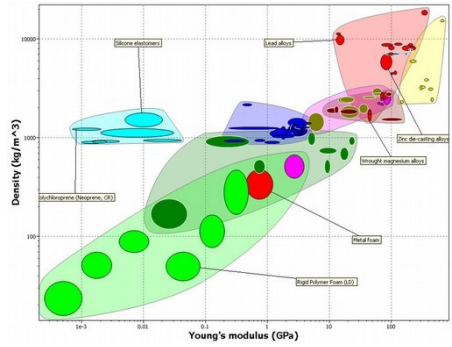# Improved performance by exploiting uncertainties

Collaboration with Optibrium

Neural networks for materials design

Materials designed

Summary

Impute values in sparse matrix to high accuracy, enables identification of **new hits** and activity profiling of compounds

Understand and exploit **uncertainties** to dial-in on most confident results

Reduce the need for experiments and **accelerate** discovery

Intellegens
gareth@intellegens.ai

optibrium
info@optibrium.com