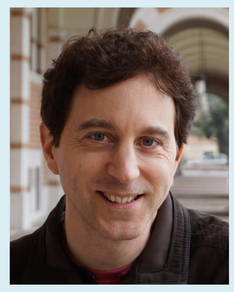


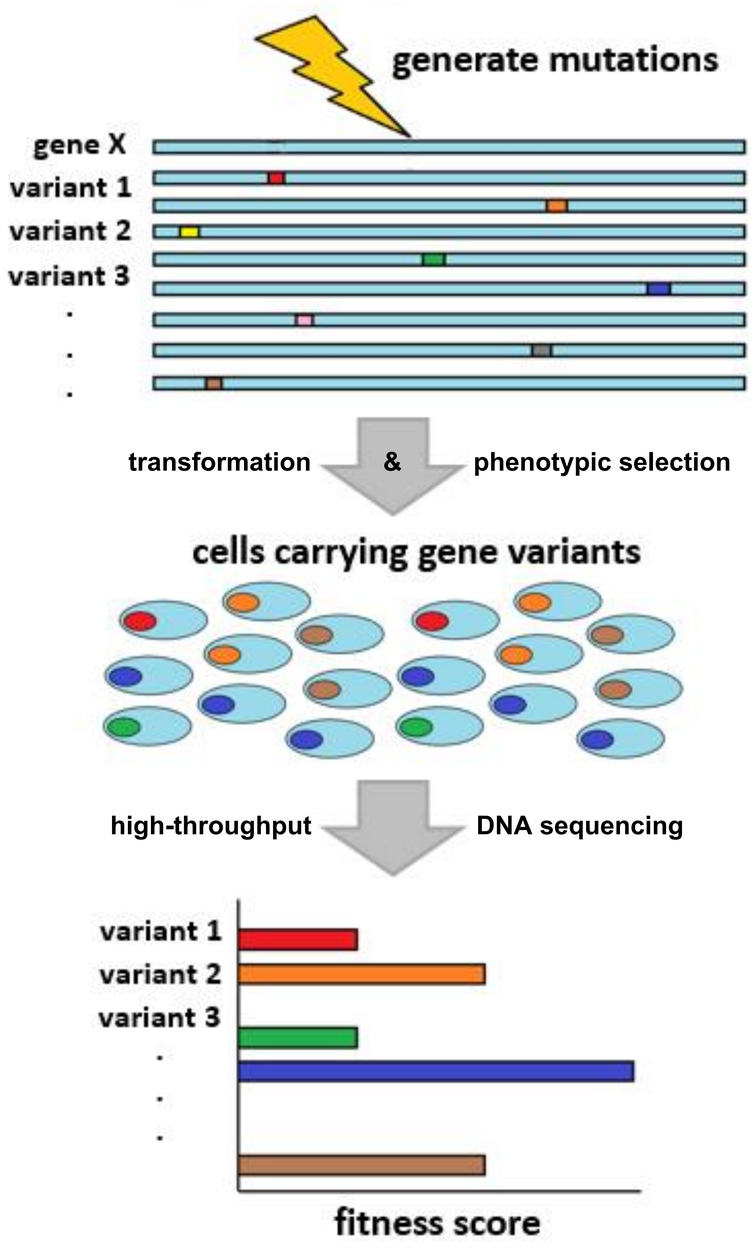
Didier Devaurs
is a cross-disciplinary post-doctoral fellow supported by funding from the University of Edinburgh and Medical Research Council (MC_UU_00009/2)

Interpreting the health impact of genetic variation via deep mutational scanning and machine learning

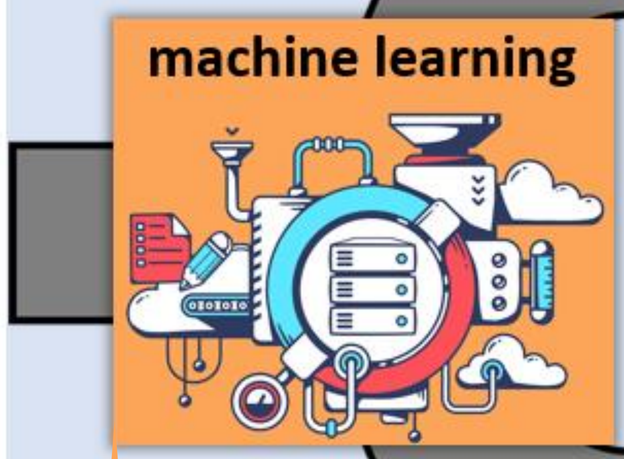
Didier Devaurs, Joe Marsh, Diego Oyarzún, Greg Kudla

Institute of Genetics and Cancer, The University of Edinburgh

deep mutational scanning (DMS) experiment



problem:
genetic variants of uncertain significance

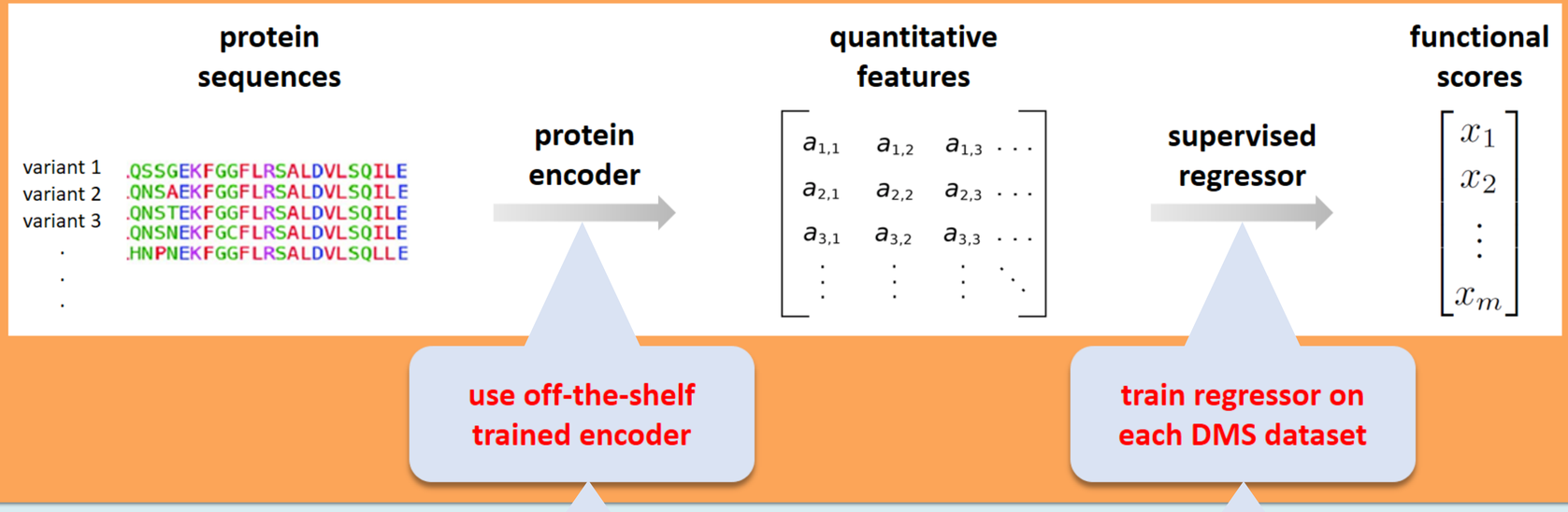


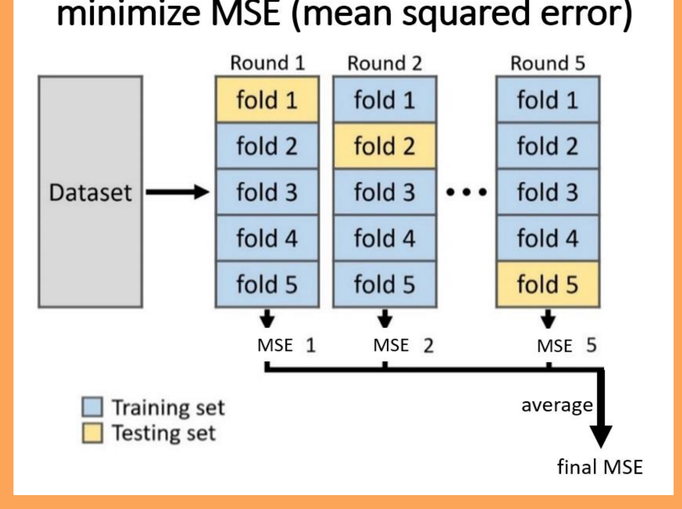
solution:
decipher the relationship between fitness scores and the health impact of genetic variants

good mutations

bad mutations

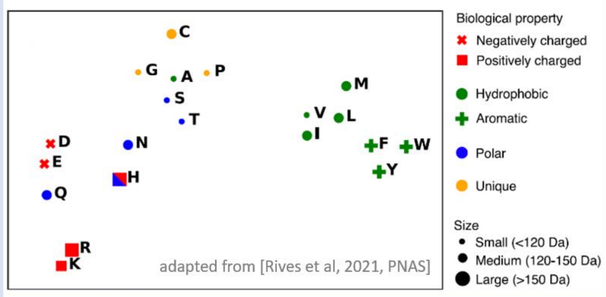
bilayer machine learning architecture



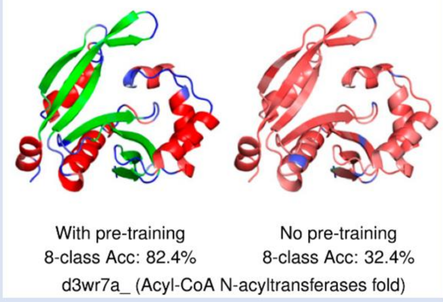


protein encoder: Evolutionary Scale Modelling (ESM) developed by Facebook AI research

recapitulate amino acid properties



predict protein secondary structure



17 encoders: deep neural networks (Transformer) trained on UniProt protein sequences

20 regressors:
linear regression
lasso
partial least squares regression
Gaussian process regression
elastic-net
kernel ridge regression
support vector machines
k-nearest neighbours
stochastic gradient descent
multi-layer Perceptron
decision tree
random forest
extremely randomized trees
AdaBoost
gradient boosted decision trees
XGBoost
4 voting regressors

short-term impact: addressing quality issues in DMS datasets, e.g., by reducing noise, imputing missing scores, and improving overall interpretability

long-term impact: interpreting genetic variant effects in people directly from experimental DMS data (i.e., DMS fitness scores)