

Sequence-function
Data

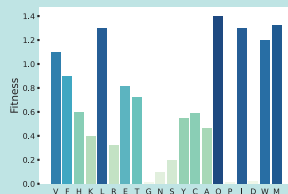
Amino Acid
Representation

Search Space Prescreening

Bayesian Optimization

Variant	Score
H76L, K78R	0.18
L102V, S116D	0.29
N127F	1.93
Q123M, L130P	6.89
E176V, M180A	4.21
Y189W, D190I	0.56
G192S, A196F	0.19
N44E, F45L	2.45

Average fitness of 20 AA
on the i-th site



AA	Value
V	1.12
F	0.91
H	0.62
K	0.43
L	1.31

Variant 1: P1 P2 A1 A2

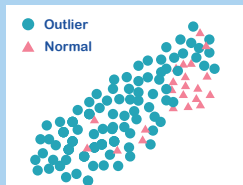
Variant N: P1 P2 A1 A2

Train

XGBOD

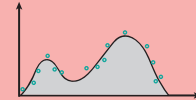


Remaining variants

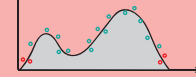


Screen out the low fitness
candidates and shrink the
total search space.

Update model



Predict remaining
variants



Select top-k variants



Recommendations for
next experiments

Next
trial