This repository contains Jupyter notebooks, python files, and data files to reproduce data reported in the paper *Genotype sampling for deep-learning assisted experimental mapping of fitness landscapes* by Andreas Wagner,

**hypertune\_regr**: hypertunes several neural network architectures on fitness data by Papkou et al., Science 2023 for nonlinear regression prediction of fitness from genotypes

**hypertune\_binclass**: hypertunes several neural network architectures on fitness data by Papkou et al., Science 2023 for binary classification into viable and non-viable genotypes

**sampling,ipynb**: trains three hypertuned deep learning neural network architectures on training data sets of different sizes and sampled in multiple different ways. Determines generalization performance of the resulting neural networks on test data.

**deep\_funcs\_pub**: encodes multiple utilities and functions, such as for loading data, encoding data, and sampling genotypes

**fitness\_data\_science\_papkou2023.tsv**: contains the fitness data reported in Papkou et al., (Science 2023) and used in the above publication for neural network training, in a simple, tab-delimited five-column format (record number, genotype, amino acid sequence, fitness of genotype, standard error of fitness)