# Precision Engineering of Biomolecular Function with Massively Multiplexed Genotype-Phenotype Measurements and Machine Learning

David Ross[1], Drew S. Tack[1], Peter D. Tonner[1], Abe Pressman[1], Nathan D. Olson[1], Sasha F. Levy[2], Eugenia F. Romantseva[1], Nina Alperovich[1], Olga Vasilyeva[1]
[1]National Institute of Standards and Technology (NIST), [2]SLAC National Accelerator Laboratory

## Goals

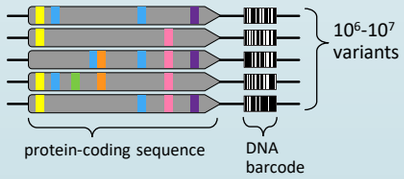
- Help turn engineering biology into a mature, quantitative engineering discipline
- Create an engineering workflow for precision, made-to-specification biomolecules for use in larger engineering projects
- Replace and/or supplement directed evolution with something better
- Create a synthetic biology design-build-test-learn framework that can scale to millions of designs simultaneously

### What do we want to engineer?

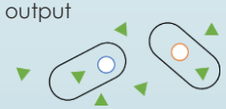
- Living Measurement Systems: cells engineered to sense their environment, make decisions, and respond

## Approach

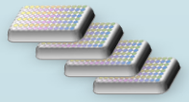
1. Create barcoded libraries of mutated variants of sensor proteins




2. Transform sensor libraries into bacteria and link cell growth rate to sensor output



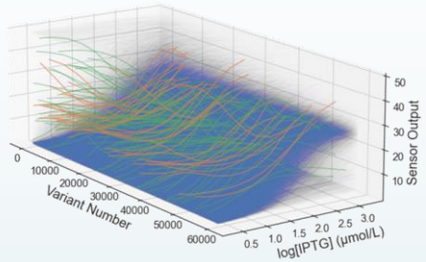
3. Test every variant over a range of ligand concentrations



4. Short-read (barcode) sequencing to measure dose-response (phenotype); long-read sequencing to measure matching genotype

## Results: Two Routes to Precision Engineering

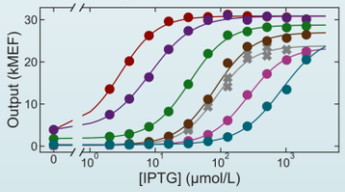
Quantitatively measure phenotype (dose-response) and matching genotype for every variant in large library



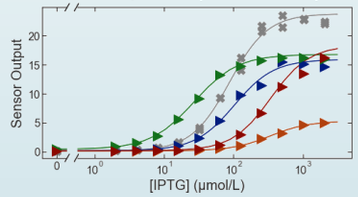

Results published in Molecular Systems Biology, DOI: 10.15252/msb.202010179

**1. In silico selection:** Identify DNA sequences from measured library with desired dose-response

Precision $EC_{50}$: 1.25-fold accuracy over 3 orders of magnitude

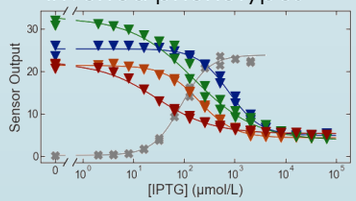



Multi-objective engineering: Select $EC_{50}$ and $G_\infty$ independently

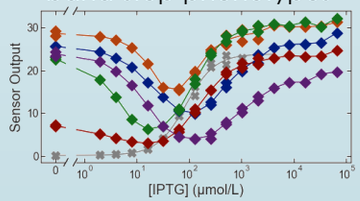



Find rare dose-response phenotypes

Genetically diverse inverted phenotypes



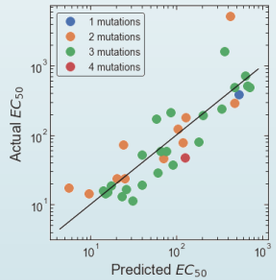

Completely novel band-stop phenotypes



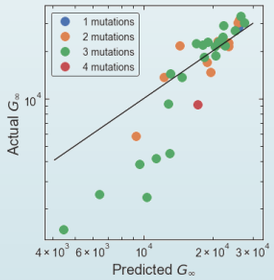

**2. Forward Engineering:** Learn design rules from large-scale data (ML, biophysical models); predict new DNA sequences

High-accuracy, interpretable ML model of genotype-phenotype landscape [Poster #112]

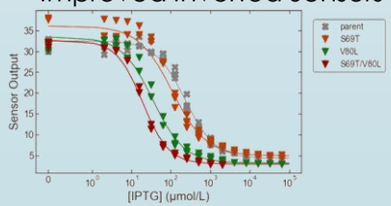
2.1-fold RMSE




1.6-fold RMSE




Extrapolative Engineering: Improved Inverted Sensors




**We have postdoctoral positions available and are also seeking collaborators interested in applying this measurement approach to other protein engineering projects; contact: david.ross@nist.gov**