

A brief overview of Machine Learning and its application to Metabolic Engineering

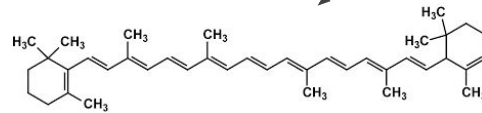
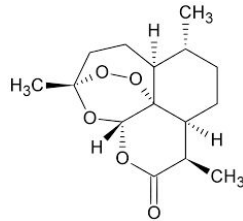
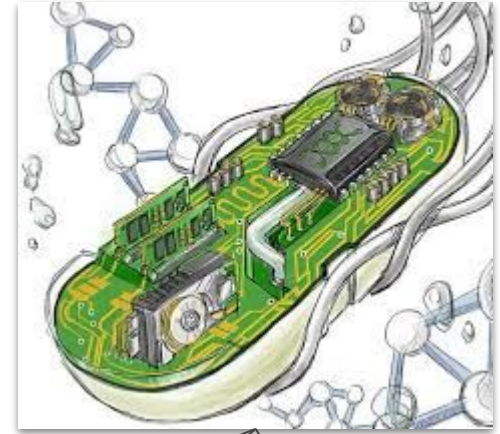
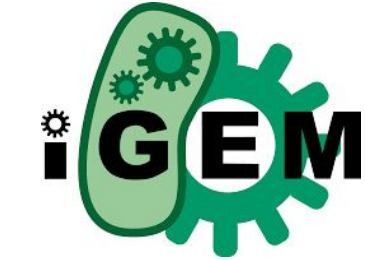
Kenny Workman, naive undergrad





Revisiting Synthetic Biology

To reduce biology to reusable and modular components.
Put the pieces back together to make something new.



Modern Bottlenecks in Metabolic Engineering

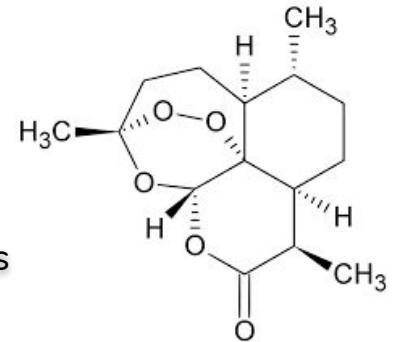
Synthetic biology
“remains a
collection of elegant
demonstrations
rather than a
systematic
discipline”.

amyris



150 Person-Years

Roosth, Sophia; Synthetic

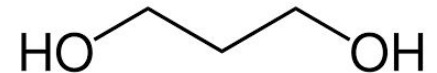


DUPONT

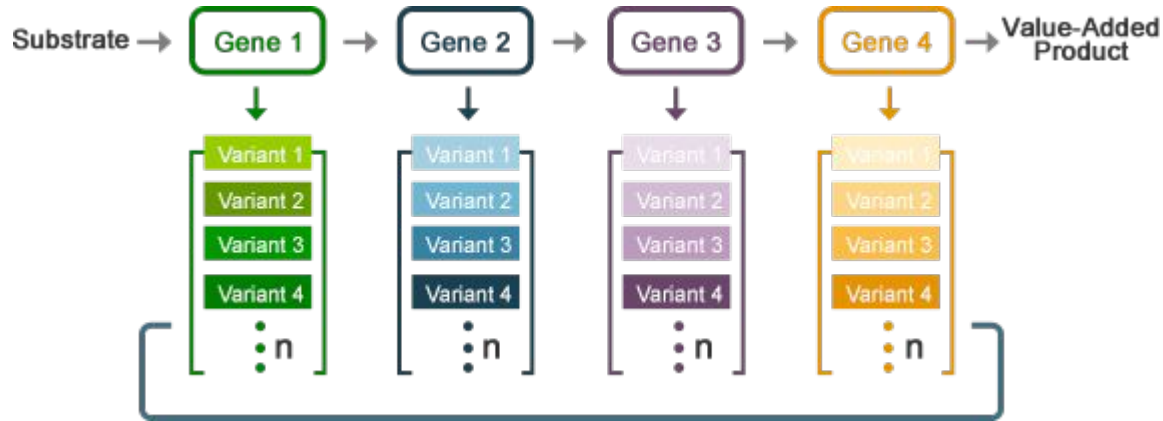


575 Person-Years

Roosth, Sophia; Synthetic



An Approach: Combinatorial Libraries and High-Throughput Assay



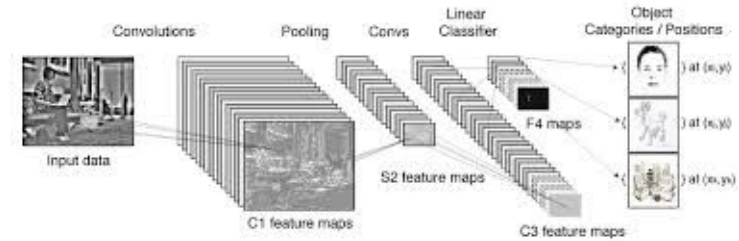
Alternative Approach: Machine Learning and multiple DBTL cycles



Use a fraction of your library constructs.

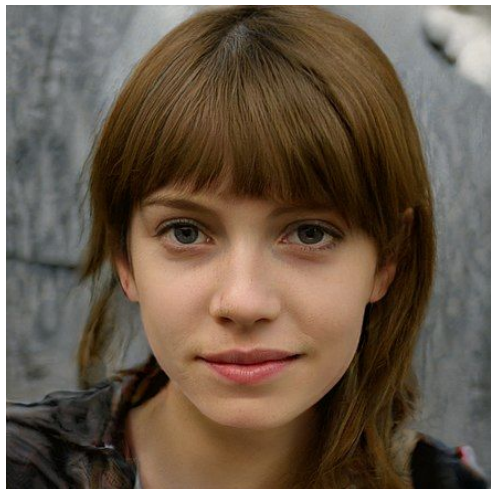
Emphasize the **learning** to extrapolate from the data.

Traditional Machine Learning Problem Domains





Novel Solution Generation w/ GANs and NLP



Look closely, but you won't catch it. This woman *does not exist*.



Legolas and Gimli advanced on the orcs, raising their weapons and shouting in unison. The orcs swung their weapons at them, but the two heroes were ready. They leapt from their horses and charged the orcs, slashing and slashing until their blades were dull and their weapons were broken. They were the first to fall, but the orcs were not done yet. They rushed the two heroes, and they fell in a hail of arrows.

The orcs were not finished yet, however. They charged the two heroes again, and they fell in a hail of arrows. The orcs were not finished yet. They charged the two heroes again, and they fell in a hail of arrows.

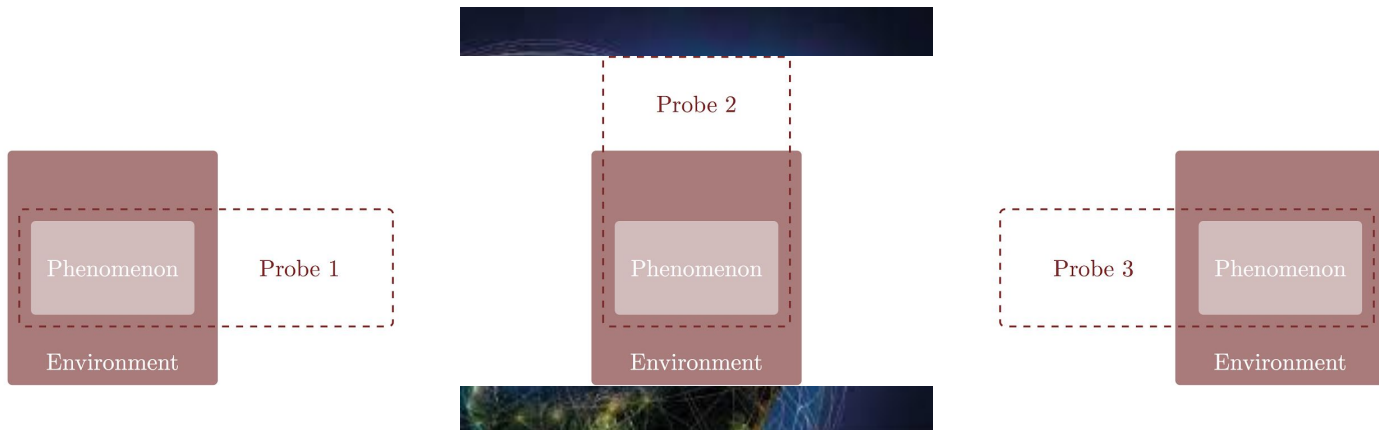
The orcs were not finished yet, however. They charged the two heroes again, and they fell in a hail of arrows. The orcs were not finished yet.. The orcs were not far off, and the two of them were not far from the edge of the cliff.

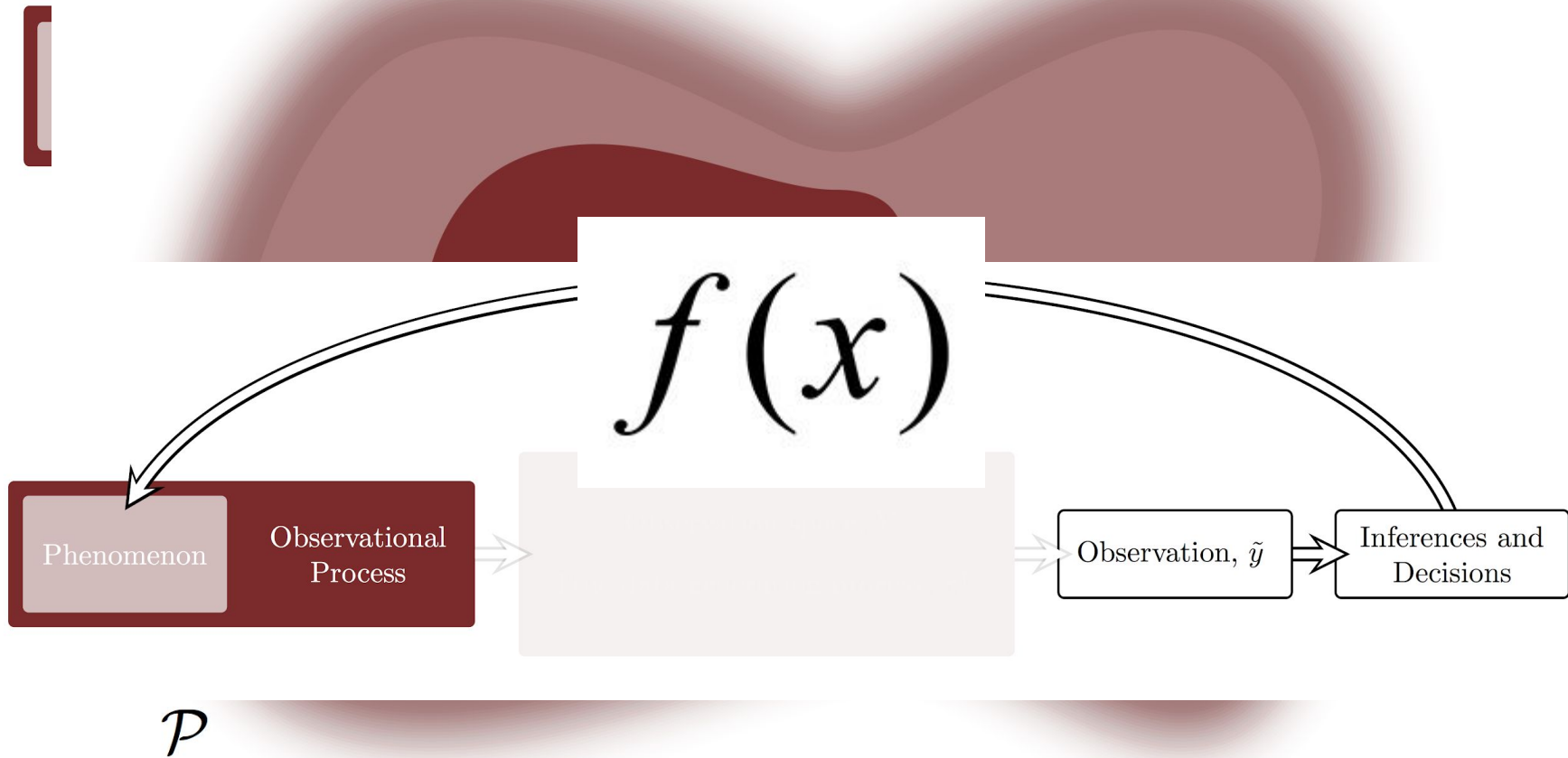
Gimli turned to face the orcs, who were now close enough to see him. He raised his hands and shouted, "I'm Gimli, son of Thráin, and I'm here to help you!"

The orcs were confused. They had seen Gimli before, but he was not the same man who had saved them from the orcs in the cave. Gimli was a tall and powerful man, and he had a beard and a moustache. He was also a dwarf, and he had a strong build, and he was covered in tattoos. He was not a man who looked like a hobbit.

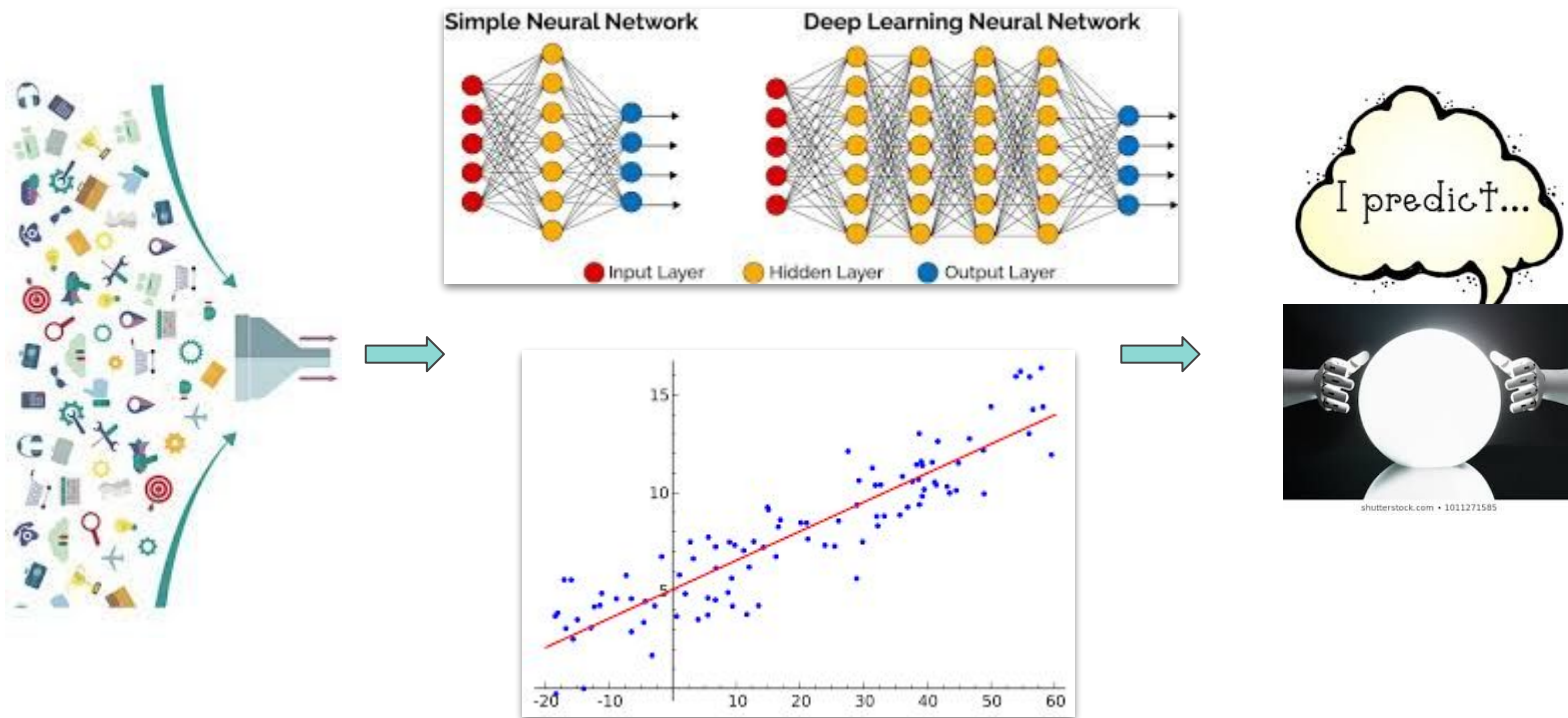


Fundamentals of Probability Theory

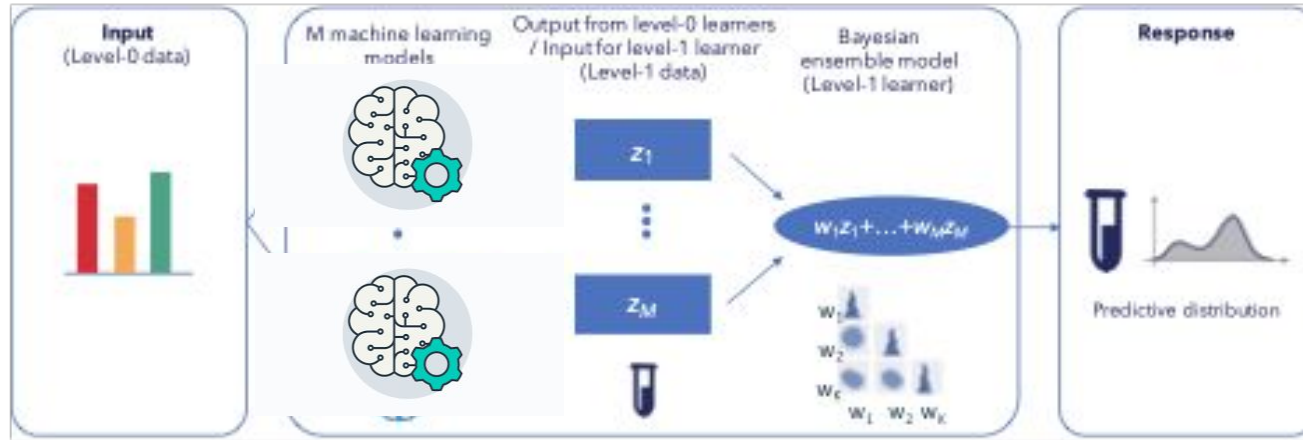




How do we model latent phenomena?




$$F: y = \mathbf{w}^T \mathbf{f}(\mathbf{x}) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2),$$



ART: an Automated Recommendation Tool





ART as both a predictive tool and a recommender



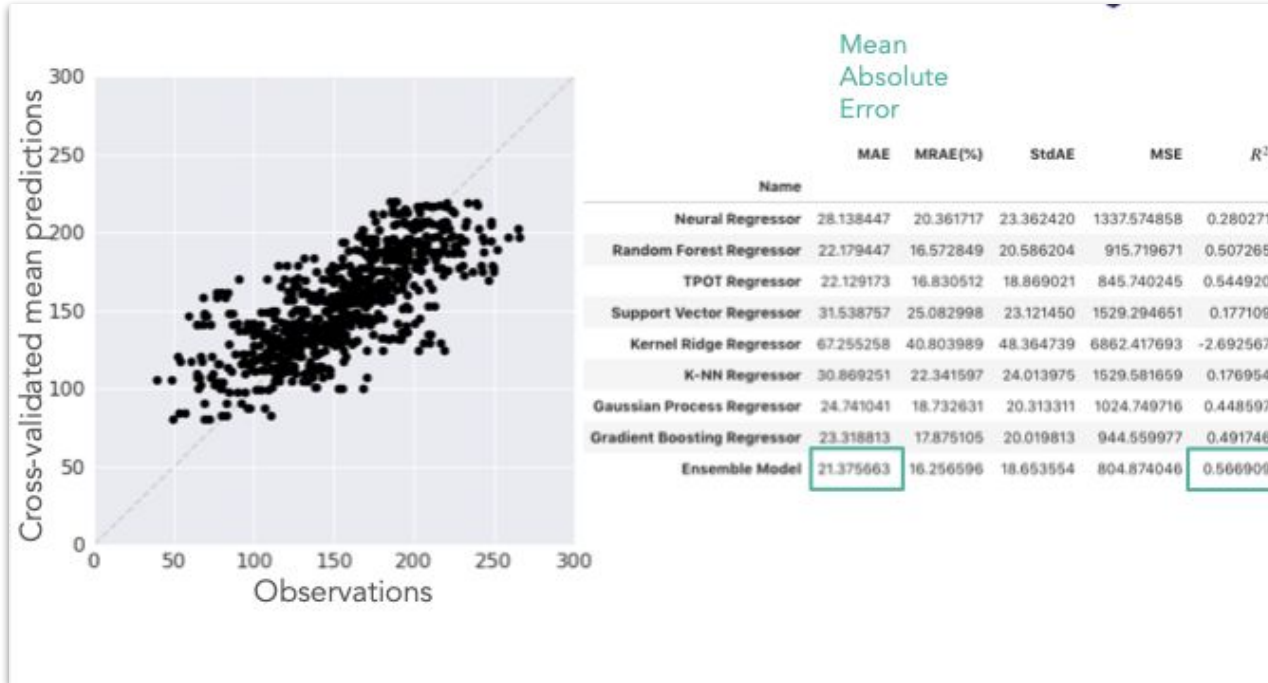
Generates recommendations to fit some objective:

- Maximization
- Minimization
- Specification

Quantifies Likelihood of the prediction being true:

There is a 15% chance that protein D will be expressed at x concentration.

Bayesian Ensemble approach can extrapolate from limited data





The math behind ART

Training Model
Parameters

$$F : y = \mathbf{w}^T \mathbf{f}(\mathbf{x}) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2),$$

Producing a
distribution of
confidence in
prediction

$$p(y|\mathbf{x}^*, \mathcal{D}) = \int p(y|\mathbf{x}^*, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathcal{D}) d\boldsymbol{\theta} = \int \mathcal{N}(y; \mathbf{w}^T \mathbf{f}, \sigma^2) p(\boldsymbol{\theta}|\mathcal{D}) d\boldsymbol{\theta}.$$

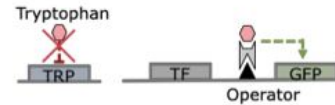
Optimizing this
parameter space to
some objective

$$G(\mathbf{x}) = \begin{cases} (1 - \alpha)\mathbf{E}(y) + \alpha\text{Var}(y)^{1/2} & \text{(maximization case)} \\ -(1 - \alpha)\mathbf{E}(y) + \alpha\text{Var}(y)^{1/2} & \text{(minimization case)} \\ -(1 - \alpha)\|\mathbf{E}(y) - \mathbf{y}^*\|_2^2 + \alpha\text{Var}(y)^{1/2} & \text{(specification case)} \end{cases}$$

A general strategy for applying Machine Learning to pathway design.

1. Remove feedback regulation and integrate biosensor

(use biosensor readout as a proxy for intracellular tryptophan accumulation)

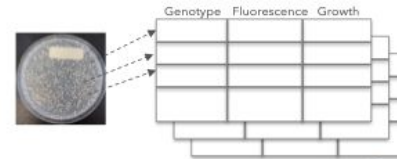


2. Create library by replacing promoters of 5 native genes

(6 promoters for each gene → 7776 possible combinations)



3. Characterize random sample

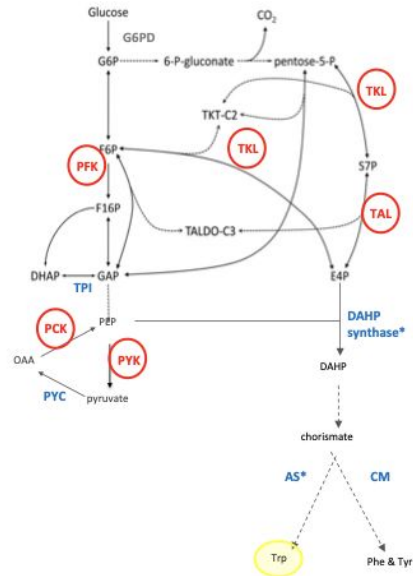


4. Find optimal combinations



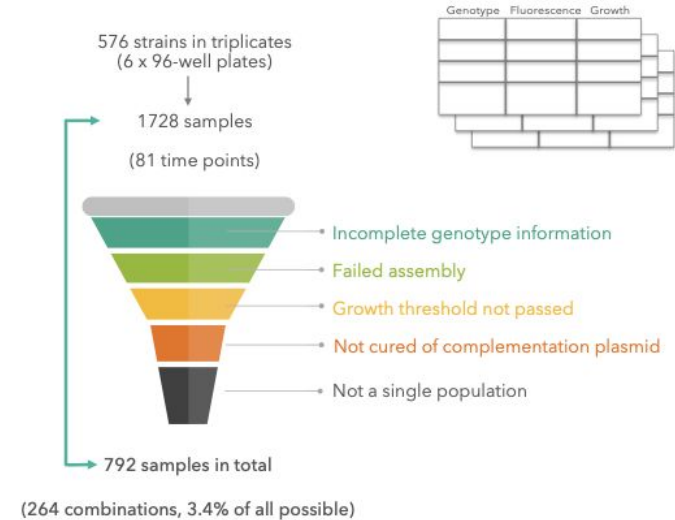
Petersen S. et al., in preparation

A Case Study: Optimizing Tryptophan production in *S. cerevisiae*



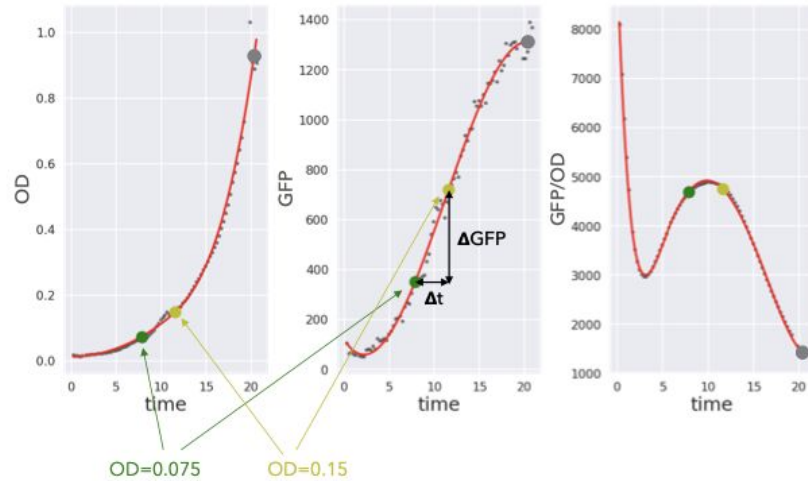
5 reactions chosen through Flux Variability Analysis (FVA) so as to most impact final yield.

Figure from J. Zhang, 2017

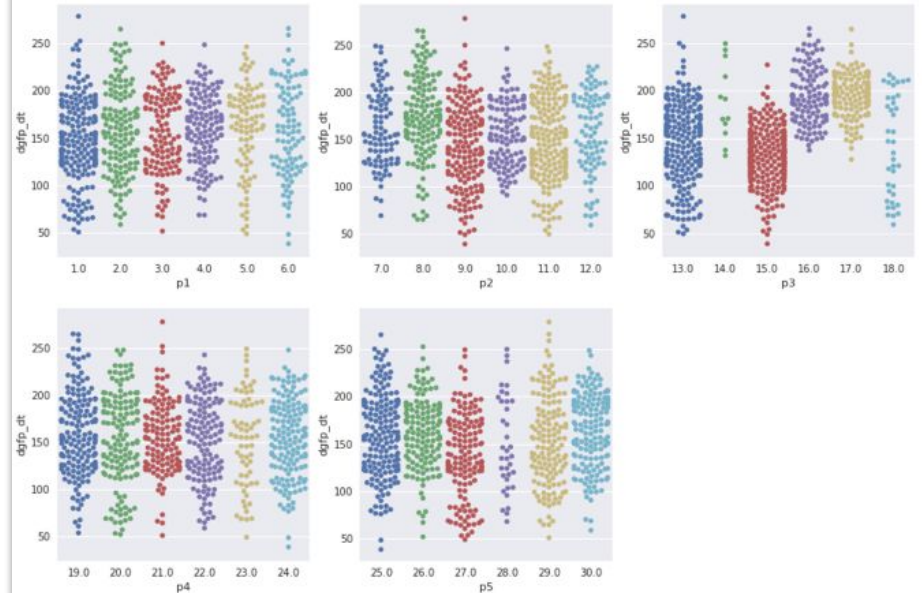


Petersen S. et al., in preparation

The relationship between production and metabolomics data is unclear.

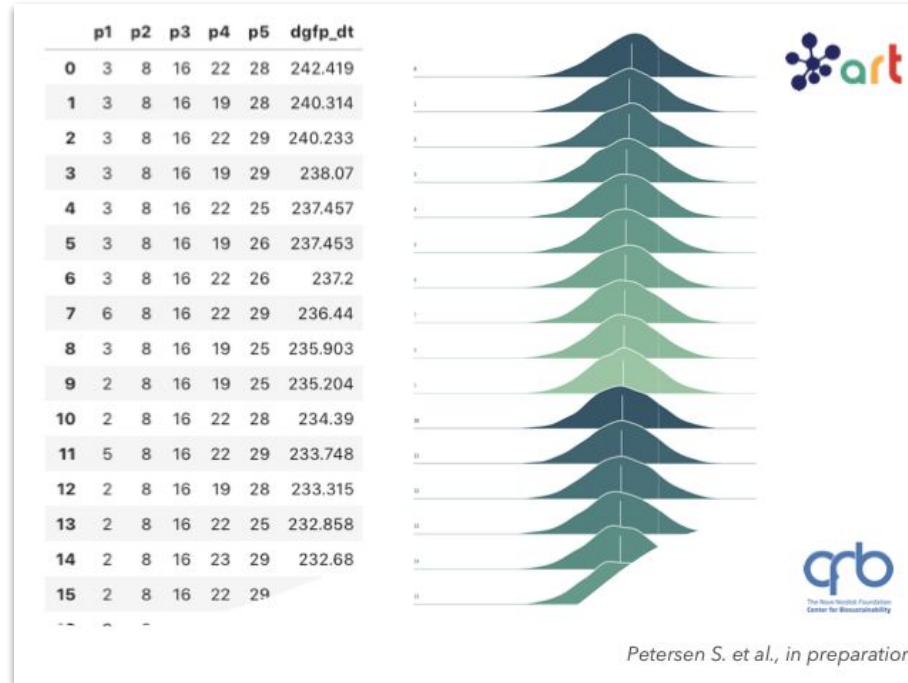


Petersen S. et al., in preparation



Petersen S. et al., in preparation

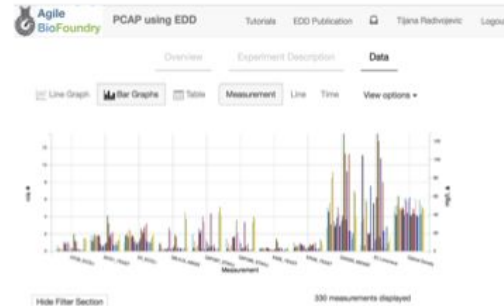
Constructs are currently being tested



A Contrived Case Study: Improving Limonene production in *E. coli*

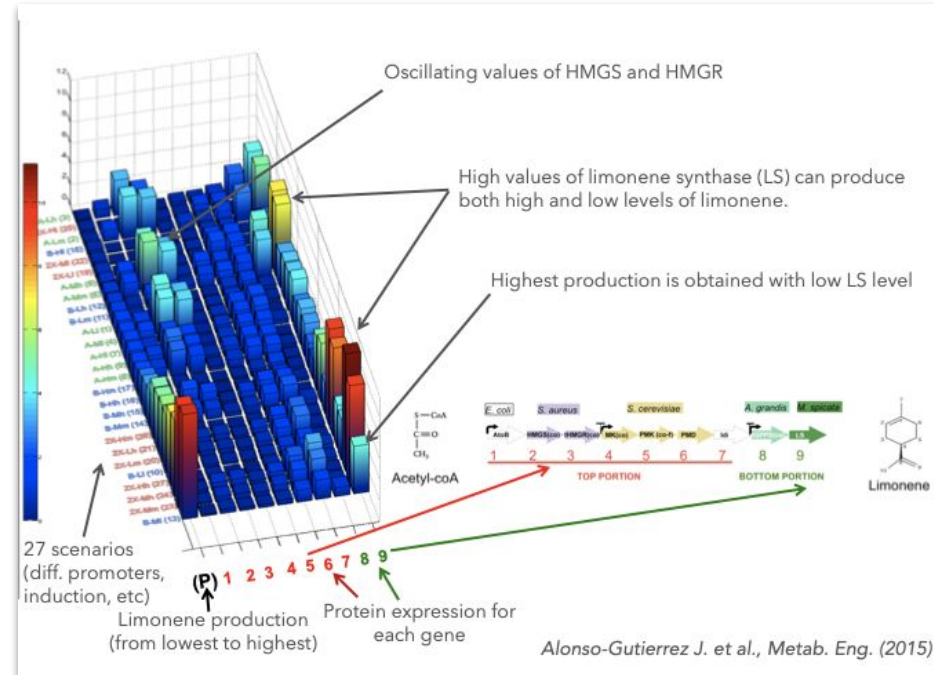
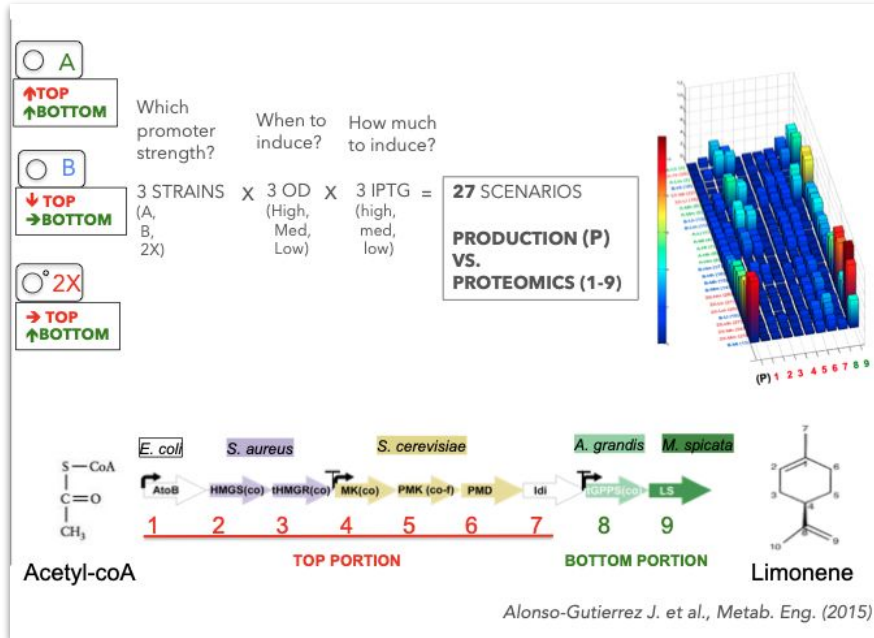


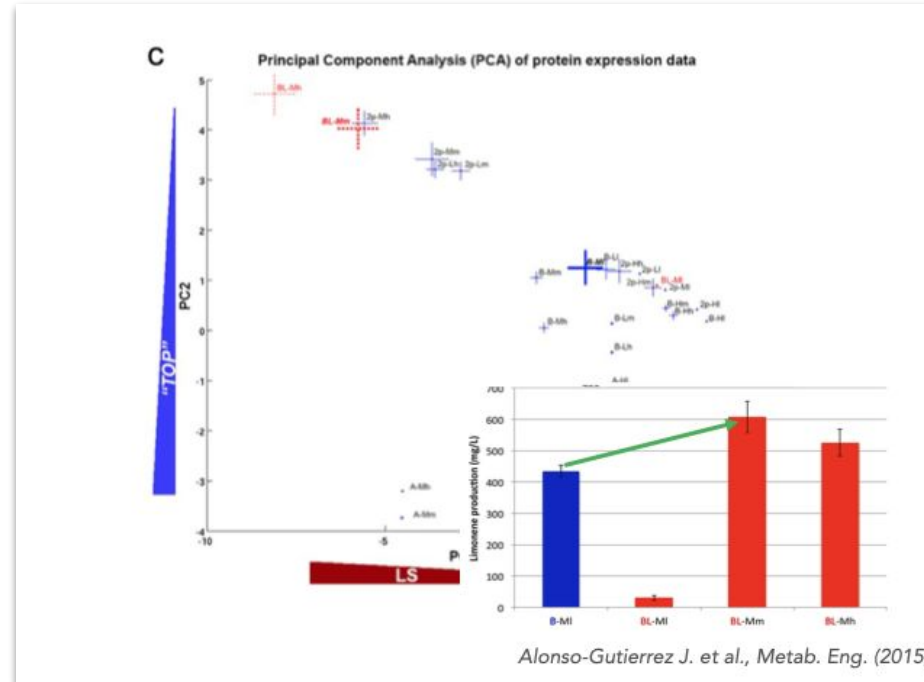
- **Data:** collection of proteomics and limonene production data



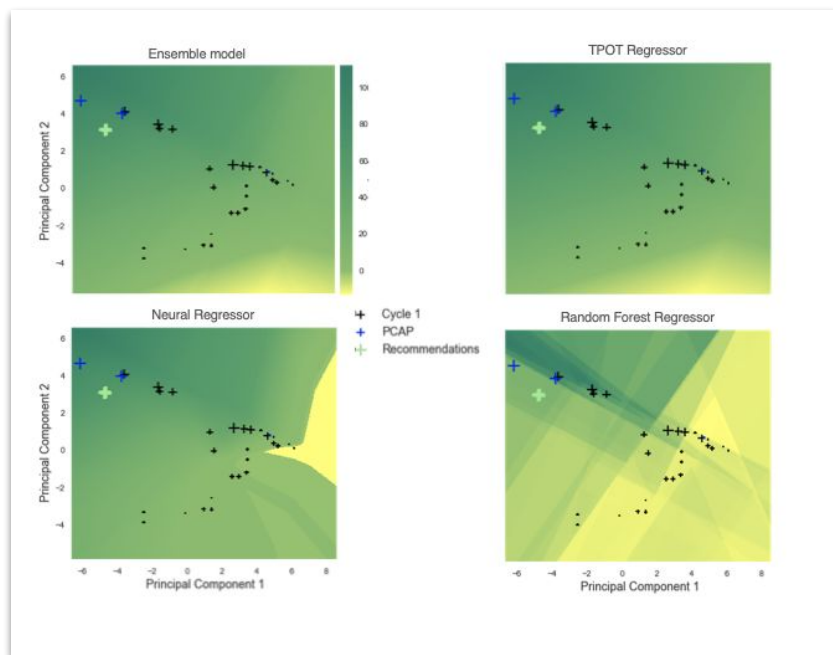
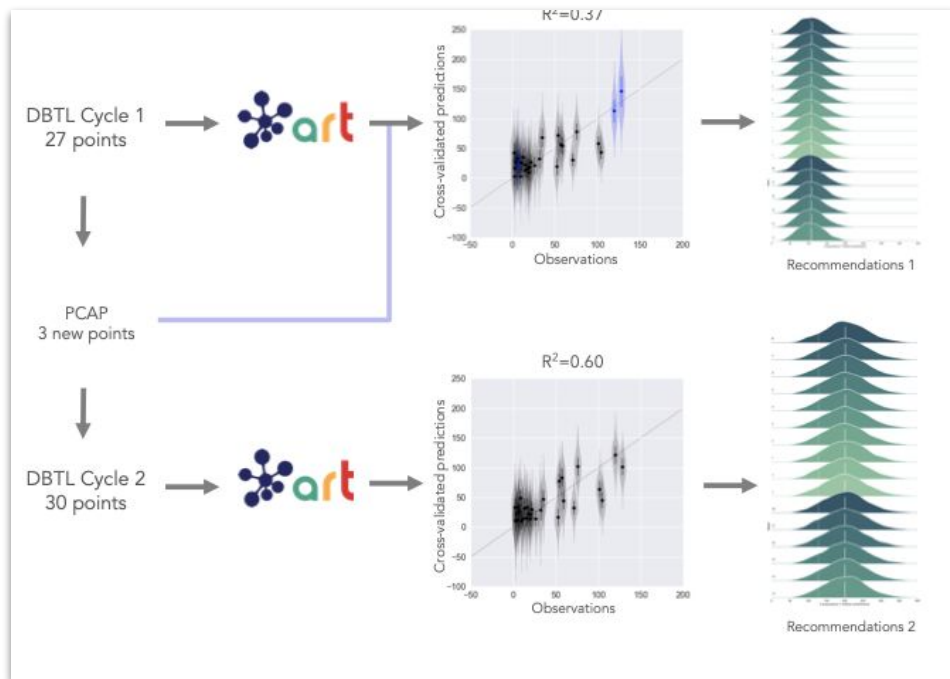
- **Objective:** identify proteins that need to have their expression level adjusted to maximize limonene production

Combinatorial libraries yield only 24 constructs from which to extrapolate.





ARTs predictions match and expand upon experimentally validated PCAP.



The importance of generalizable Machine Learning in Synthetic Biology

Opaque data; Pipeline not reproducible.



Easy integration with virtually any repeatable experiment



```
from py_art.art import *

edd_study_slug = 'limonene-data-for-art'
edd_server = 'edd.jbei.org'
df = utils.load_study(edd_study_slug=edd_study_slug, edd_server=edd_server)
# data_file = '../data/Limonene_data_for_ART.csv'
# df = utils.load_study(data_file=data_file)

Password for tradivojevic: *****

art_params = {}
art_params['input_var'] = ['ATOB_ECOLI', 'ERG8_YEAST', 'IDI_ECOLI', 'KIME_YEAST',
                          'MVD1_YEAST', 'Q40322_MENSP', 'Q8LKJ3_ABIGR',
                          'Q9FD86_STAAU', 'Q9FD87_STAAU']
art_params['response_var'] = ['4-isopropenyl-1-methyl-cyclohexene']
art_params['objective'] = 'maximize'
art_params['threshold'] = 0.2
art_params['seed'] = 5

art = RecommendationEngine(df, **art_params)
```