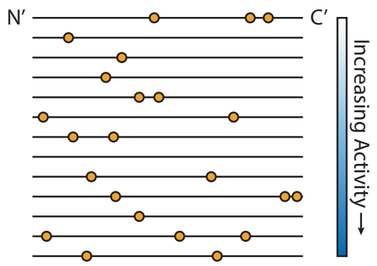
Hierarchical Bayesian Modelling for High Throughput Measurement

Arkadij Kummer and Eric J. Ma, Novartis Institutes for Biomedical Research

Protein Engineering

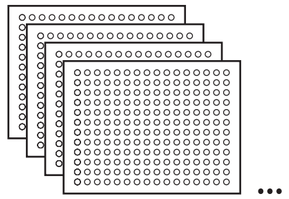


● Mutation

Goal: Estimate true activity of protein, given replicate measurements. (Sometimes singleton replicates.)

Application: Use downstream in constructing prediction models.

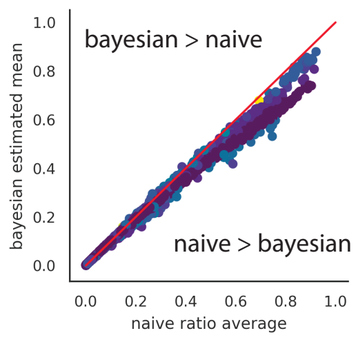
High throughput assays



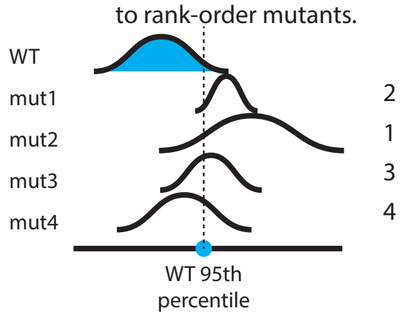
num measurements
~ 10^4-10^5

Problems:
1. Some samples are measured only once.
2. Plate shift effects.

Regularization and uncertainty in decision-making



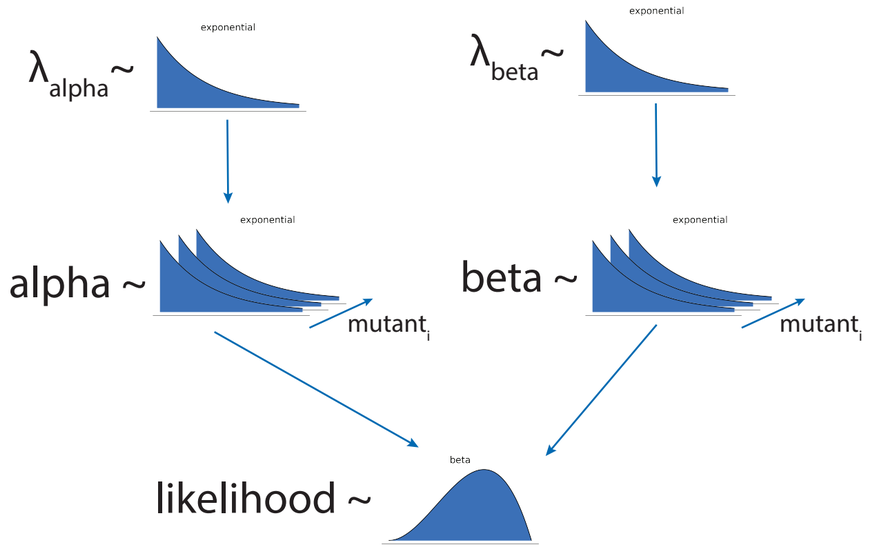
Use posterior likelihood distribution to rank-order mutants.



- Bayesian posteriors give us uncertainties, even for singleton measurements.
- Regularization guards against extreme values produced by random chance.
- Posterior likelihoods give us statistically principled path towards prioritization of samples.

Estimation Model

Hierarchical Beta distribution modelling



Fitting: ADVI because large number of measurements (10^4-10^5), and large number of measurements (~10^4)

Upgraded Estimation Model

Factoring in plate shift effects