

# Probabilistic Graphical Modeling of Gene Expression Modulation by CRISPR Perturbation

Jason Mohabir & Edward Moseley

26th November 2024  
BIOSTAT914

# EEGGO:

## Estimation of Enhancer-Gene Guide Outcomes



Jason Mohabir & Edward Moseley

26th November 2024  
BIOSTAT914

# Model Intuition & Formulation

We rely on empirically-derived statistics from **Control Samples**, to predict gene expression and beta coefficients proportional to the gene's level of inhibition by CRISPR.

**Goal:** Estimate gene expression values, and infer beta coefficients for each gene-enhancer pair given the following model:

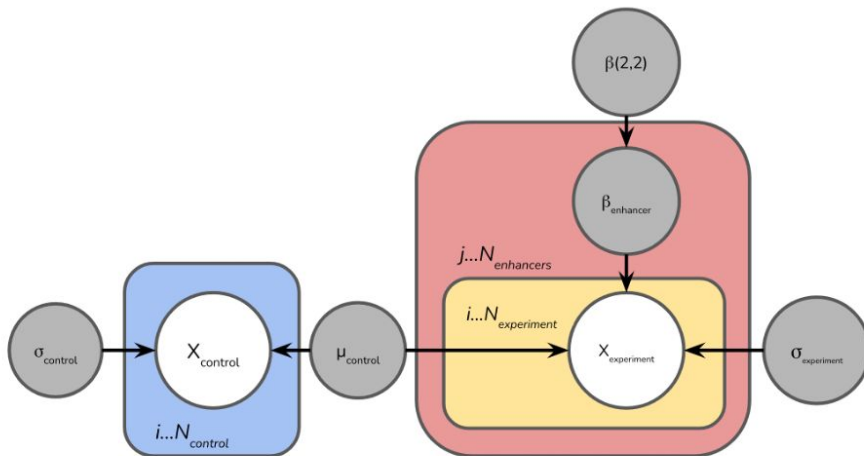
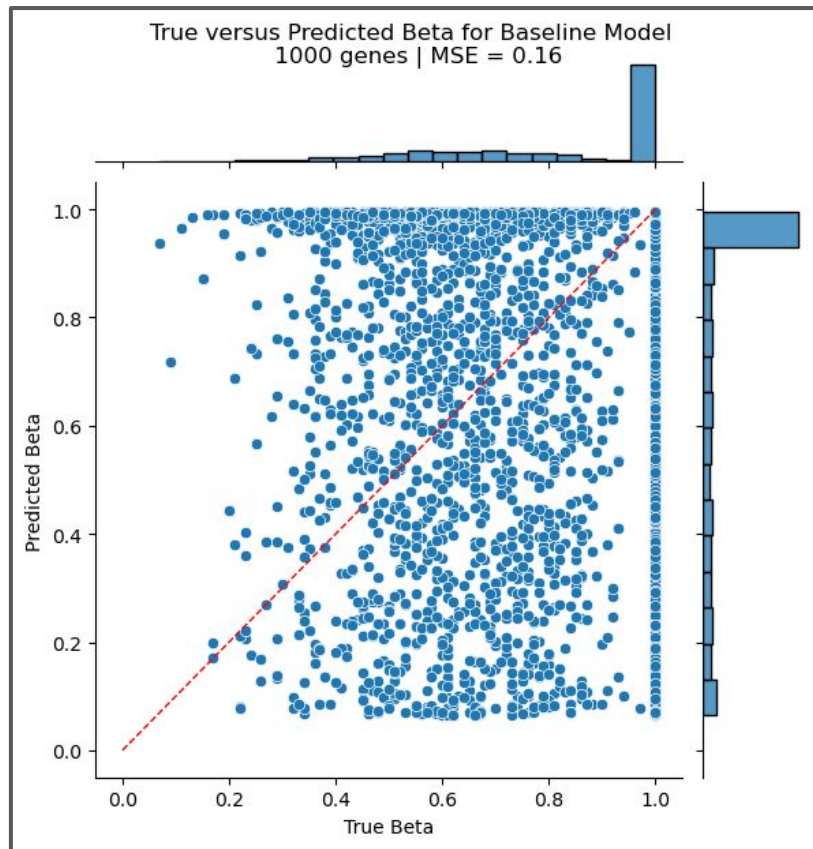
$$X_{\text{experiment}} \sim \beta_{\text{gene-enhancer}} \times \mu_{\text{control}}$$

$$\beta_{\text{gene-enhancer}} = \frac{\mu_{\text{experiment}}}{\mu_{\text{control}}}$$

With this formulation, expression is only modulated with a beta coefficient if it is **an experimental sample**.

We did not scale or otherwise transform the input data.

# Baseline: Control-Aware Normally Distributed Expression



$$\forall_{ij} X_{ij} \sim \begin{cases} \mathcal{N}(\mu_{\text{control}}, \sigma_{\text{control}}) \\ \mathcal{N}(\beta_j \mu_{\text{control}}, \sigma_{\text{experimental}}) \end{cases}$$

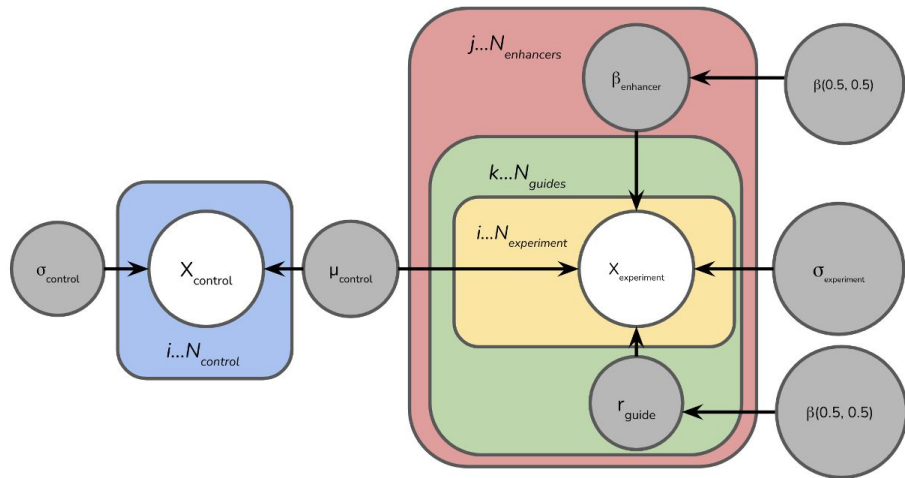
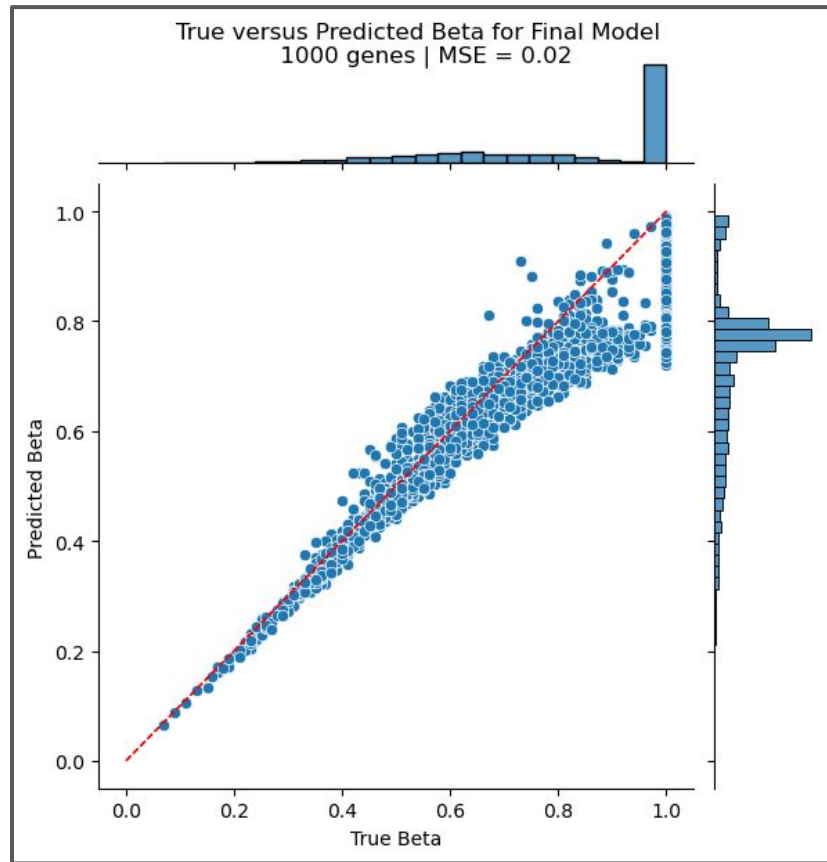
$$\begin{aligned} \text{isControl} &== 1 \\ \text{isControl} &== 0 \end{aligned}$$

- This control-aware model infers the beta value only in cases where the observation is not a control.
- Expression values,  $X_{ij}$ , and  $\mu_{\text{control}}$ ,  $\sigma_{\text{control}}$  are observed, and control statistics are defined **globally** across all 1,000 genes
- This model ignores guideIDs.

Performance (Beta Coefficients):

**Mean Squared Error (MSE): 0.16402**

# Final Model: Control-Aware Mixture Model for Guide Potency



$$X_{ijk} \sim (r_{jk}) N(\beta_j \mu_{j \text{ control}}, \sigma_{j \text{ experimental}}) + (1 - r_{jk}) N(\mu_{\text{control}}, \sigma_{j \text{ control}})$$

We introduce the latent mixture proportion,  $r_d$ , with a  $\beta(0.5, 0.5)$  prior to our formulation, which is meant to account for GuideID functionality

Assumption: Not all GuideIDs are functional

Performance (Beta Coefficients):

**Mean Squared Error (MSE): 0.02228**