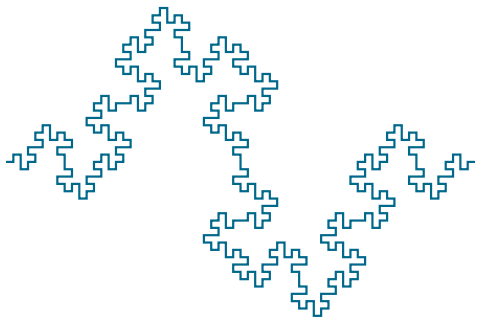
High-Throughput Sequencing Course

Statistical Interaction

Biostatistics and Bioinformatics

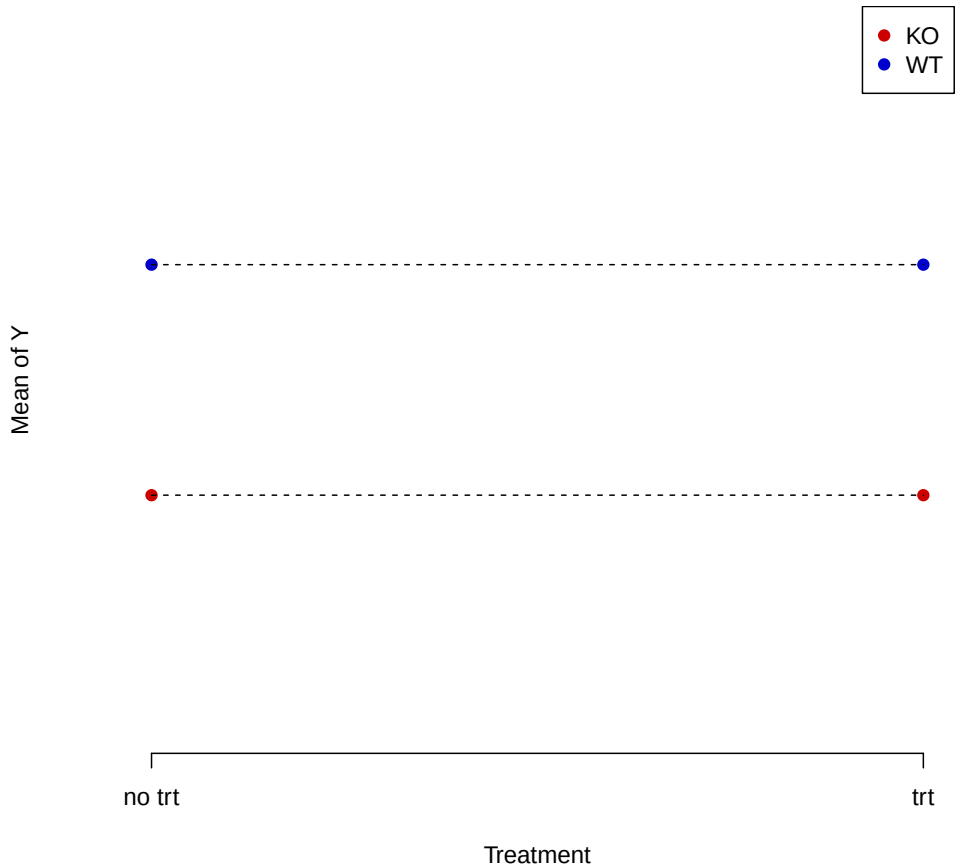


Summer 2018

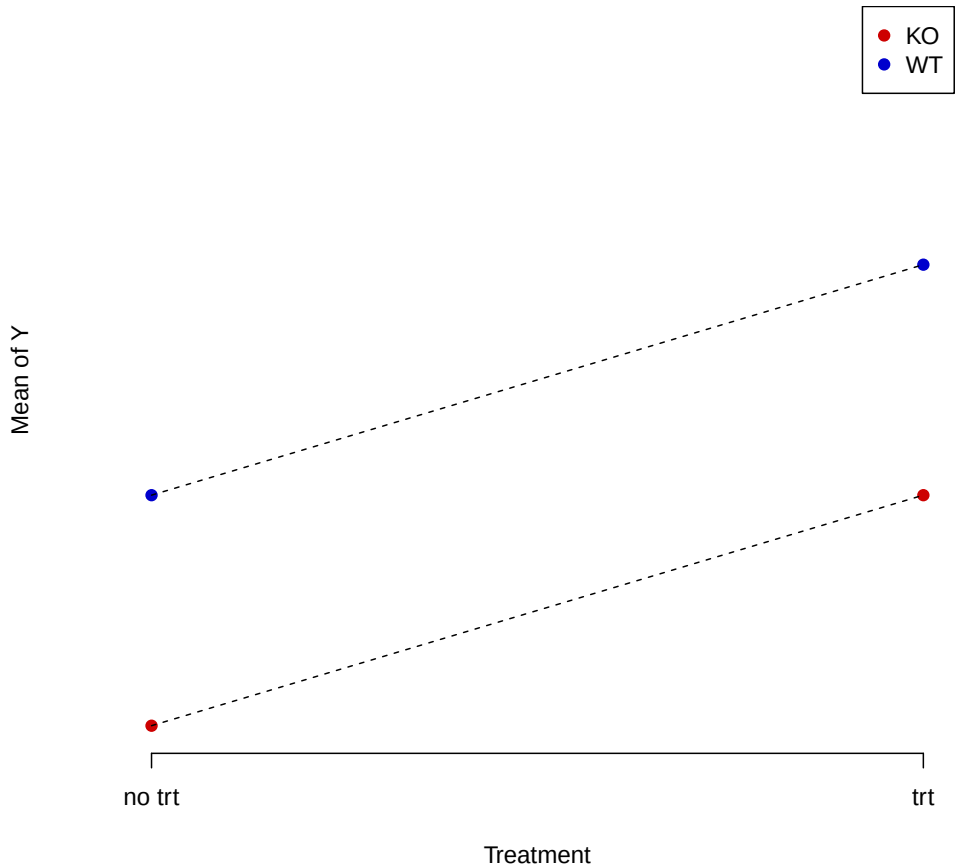
Duke University
School of Medicine

DUKE BIOSTATISTICS & BIOINFORMATICS

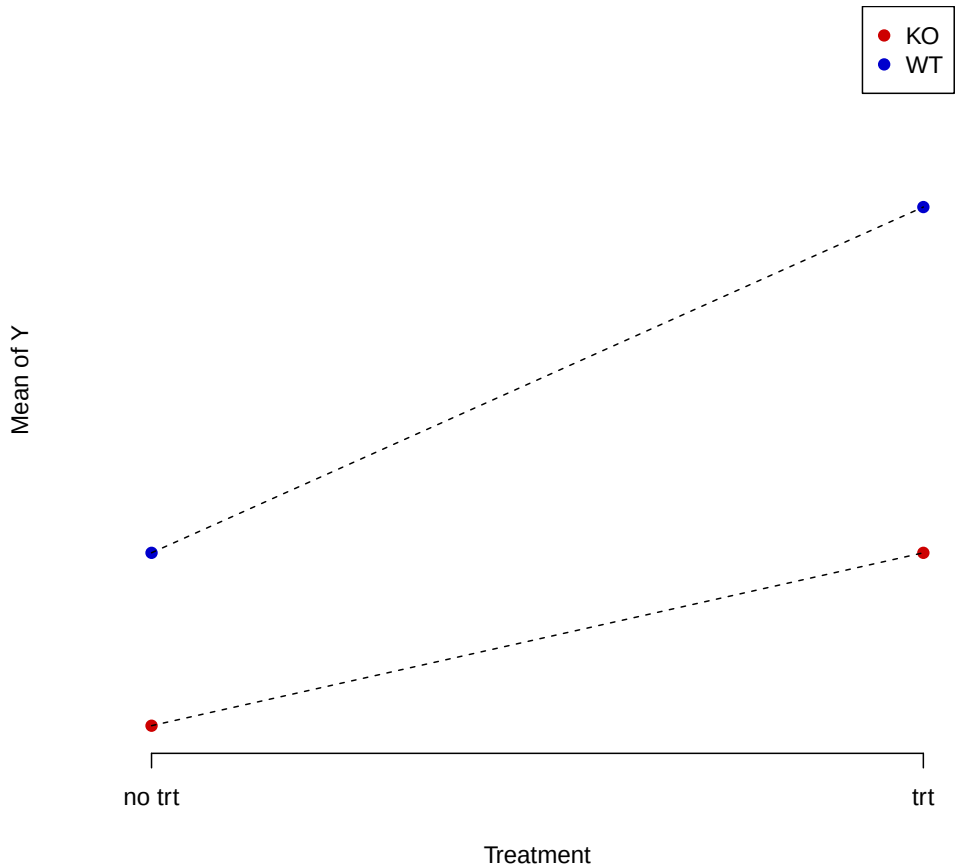
Example 1: No Interaction



Example 2: No Interaction



Example 3: Interaction



Model Interaction

- Y denotes the gene expression
- Let x denote the treatment indicator
  - $x = 0$ if not treated or 1 if treated
- Let z denote the knock-out indicator
  - $z = 0$ is WT or 1 otherwise
- The expected value of Y given treatment indicator x and knock out indicator z is denoted by

$$\mu_{x,z} = E[Y|X = x, Z = z]$$

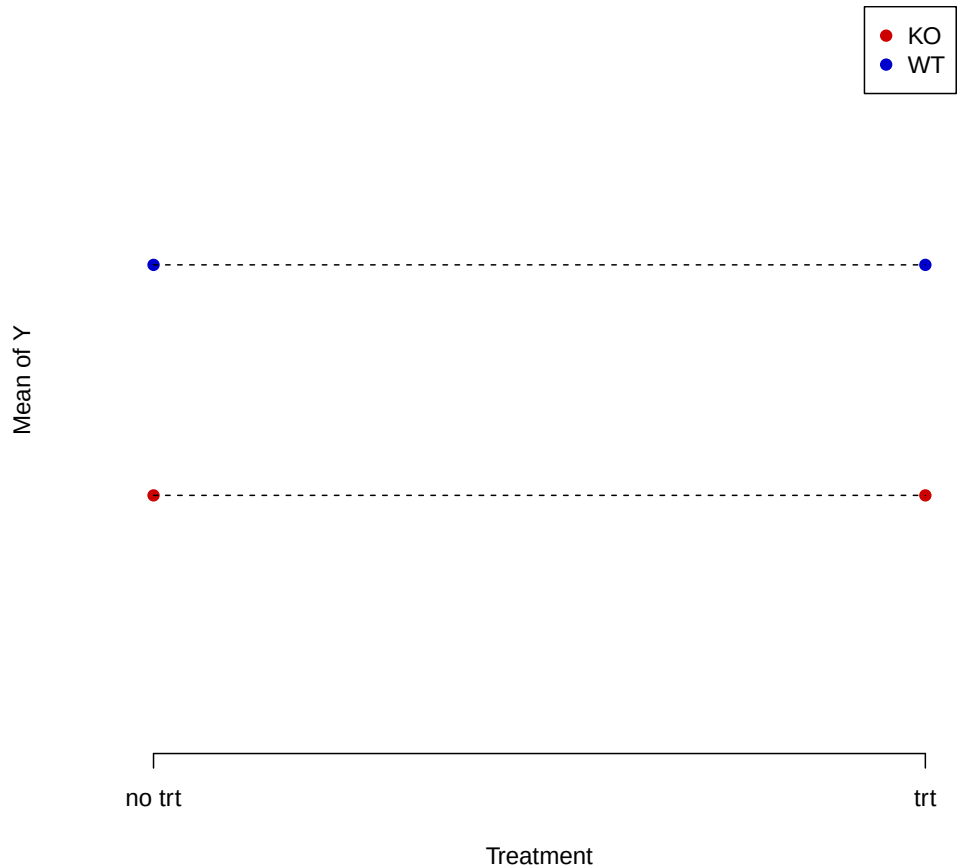
- The model will be

$$Y = \mu_{x,z} + \epsilon$$

where ϵ is a the measurement error

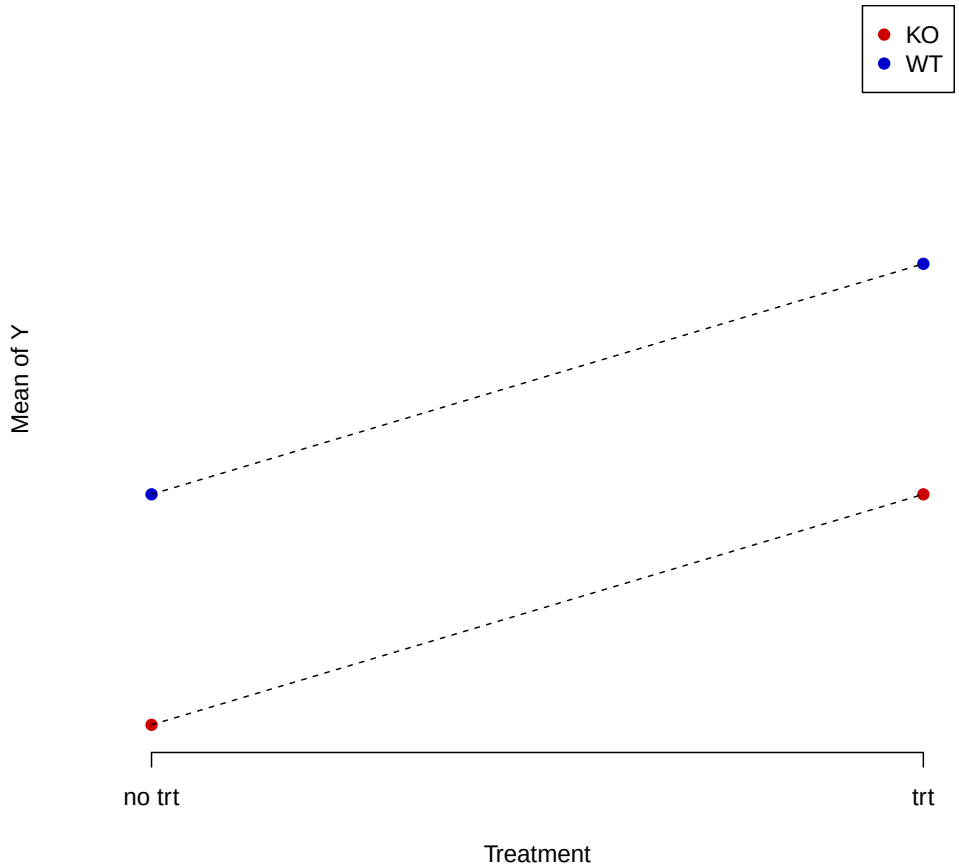
Example 1: Linear Model for No Interaction

$$Y = \beta_0 + \beta_1 z + \epsilon \ (\beta_2 = 0)$$



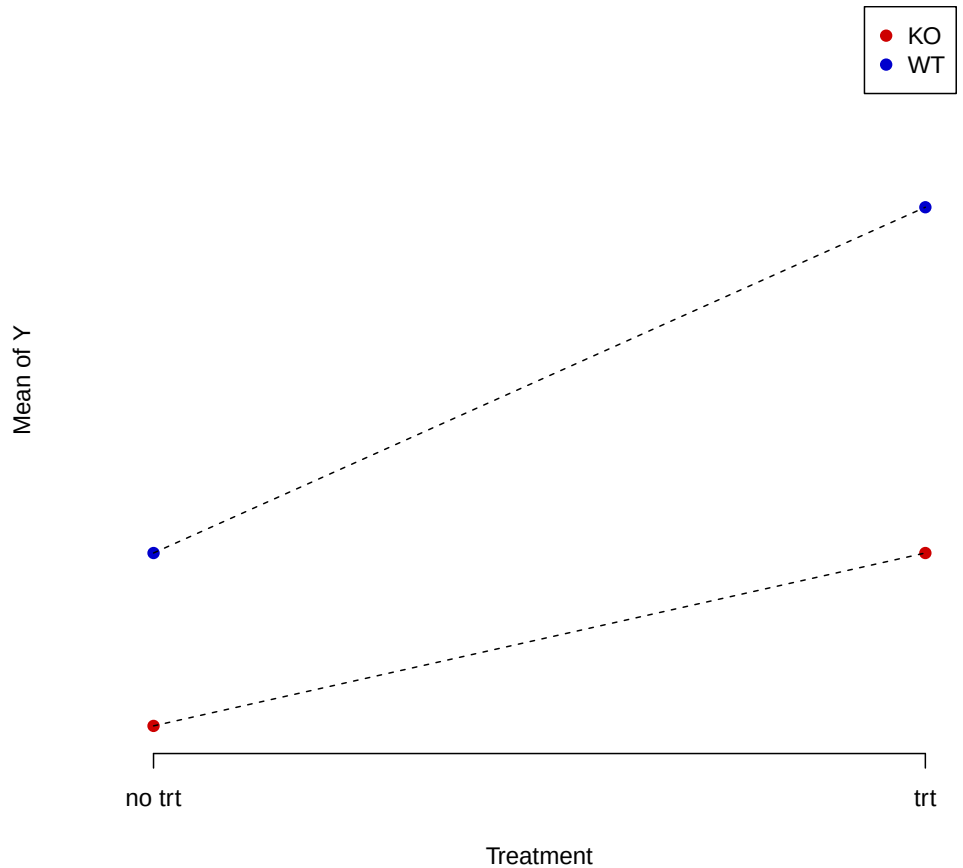
Example 2: No Interaction

$$Y = \beta_0 + \beta_1 z + \beta_2 x + \epsilon$$



Example 3: Interaction

$$Y = \beta_0 + \beta_1 z + \beta_2 x + \beta_3 xz + \epsilon$$



Interaction Examples

- Example 1: What are the signs for β_0 and β_1?
- Example 2: What are the signs for β_0, β_1 and β_2?
- Example 2: What are the signs for $\beta_0, \beta_1, \beta_2$ and β_3?

Incorporating Interactions into the NB Model

- Conditional on x_j and z_j, the observed number of reads mapped to gene i for sample j, K_{ij}, follows a negative binomial distribution with
  - Mean μ_{ij}
  - Dispersion parameter α_i (gene specific)
- Conditional on the treatment status of sample j ($x_j = 0$ or 1) and the temperature z_j, the expected value of K_{ij} is

$$\mu_{ij} = s_j \times q_{ij}$$

  where

$$\log q_{ij} = \beta_{i0} + \beta_{i1} x_j + \beta_{i2} z_j + \beta_{i3} x_j z_j$$

- The normalization parameters are assumed to be sample (not gene) specific ($s_{ij} = s_j$)