

Phenotypes, epistasis, and probability theory

David Angeles-Albores,
Ph.D.

Biology is ALL about phenotypes



Robber fly

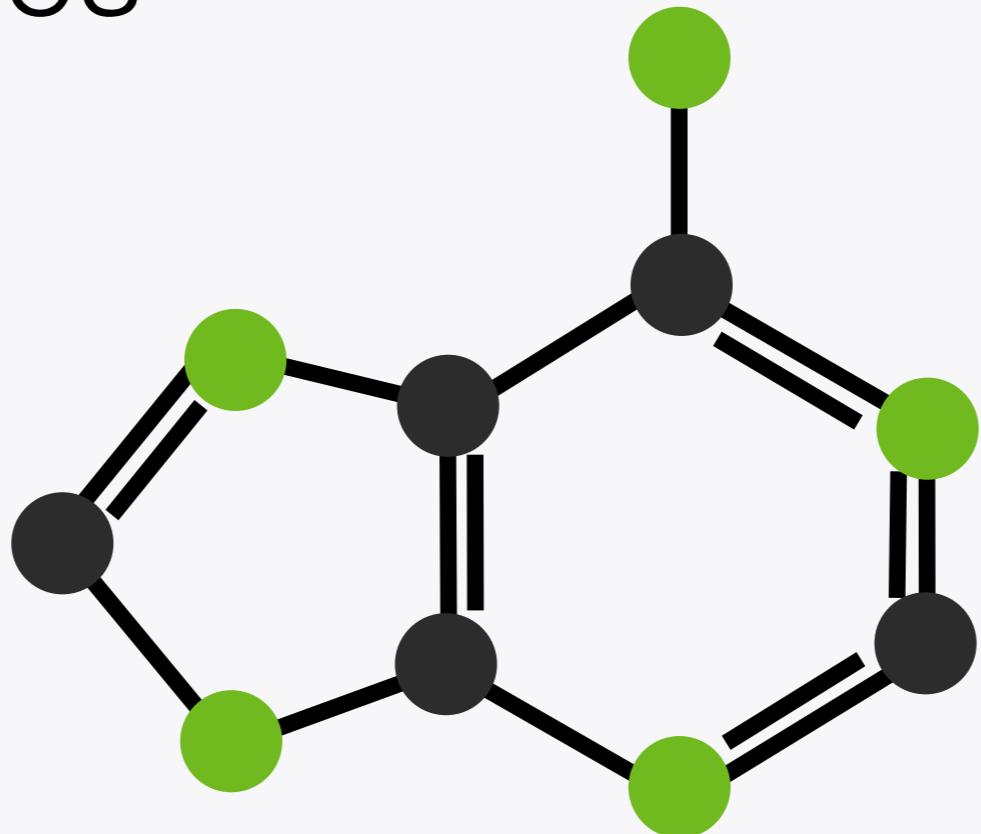


D. guanajuatensis

Classically, phenotypes are observable

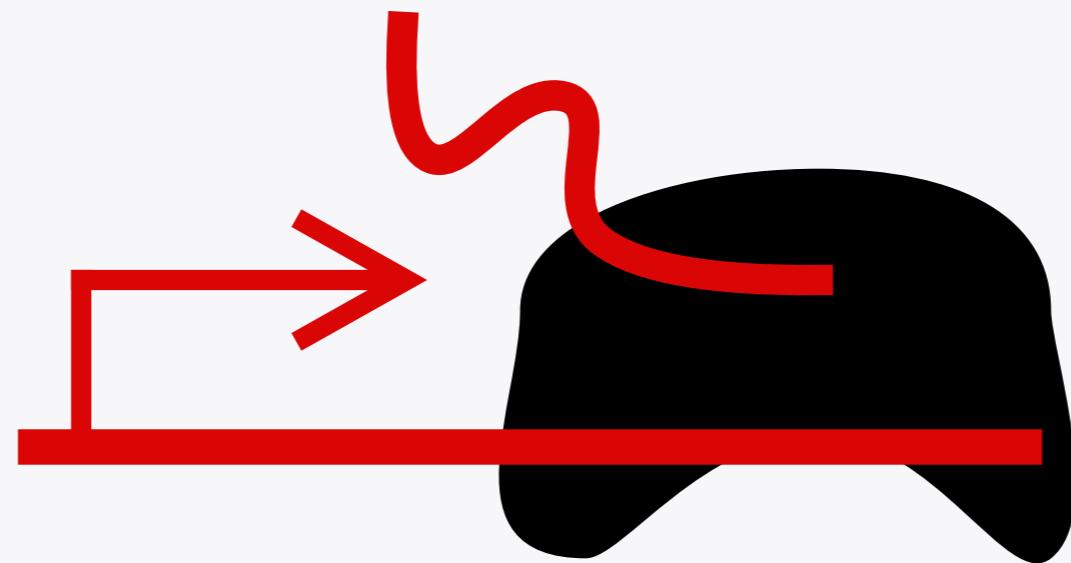
Photos my own

New phenotypes, new challenges, new discoveries



- **Phenotype:** Auxotrophy (metabolite biosynthesis)
- **Challenge:** Auxotroph isolation and maintenance
- **Discovery:** Proteins encoded by genes

New INVISIBLE phenotypes, new challenges, new discoveries



- **Phenotype:** Transcriptomes
- **Challenge:** Going beyond correlational experiments and analyses

Extending the geneticist's arsenal for **causal** analysis transcriptome-wide

Method

Transcriptome-wide

Epistasis (Null mutants)

Angeles-Albores *et al*, 2018a
Angeles-Albores *et al*, 2017

Dominance (Allelic series)

Angeles-Albores and
Sternberg, 2018b

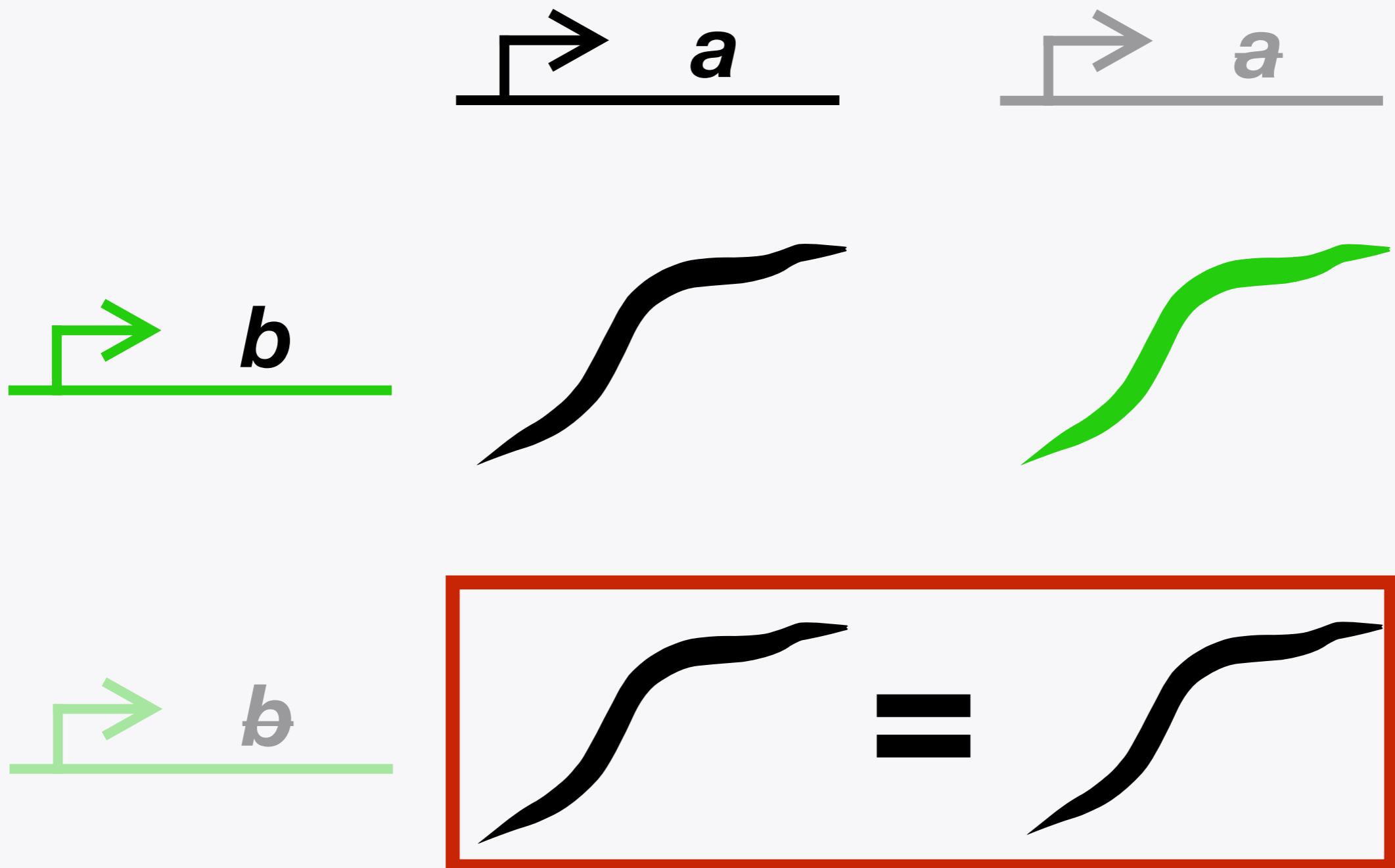
Maternal effects (Crosses)

Angeles-Albores *et al*,
manuscr. in prep.

How do we infer causal interactions



Phenotypes order genes along pathways using Batesonian epistasis

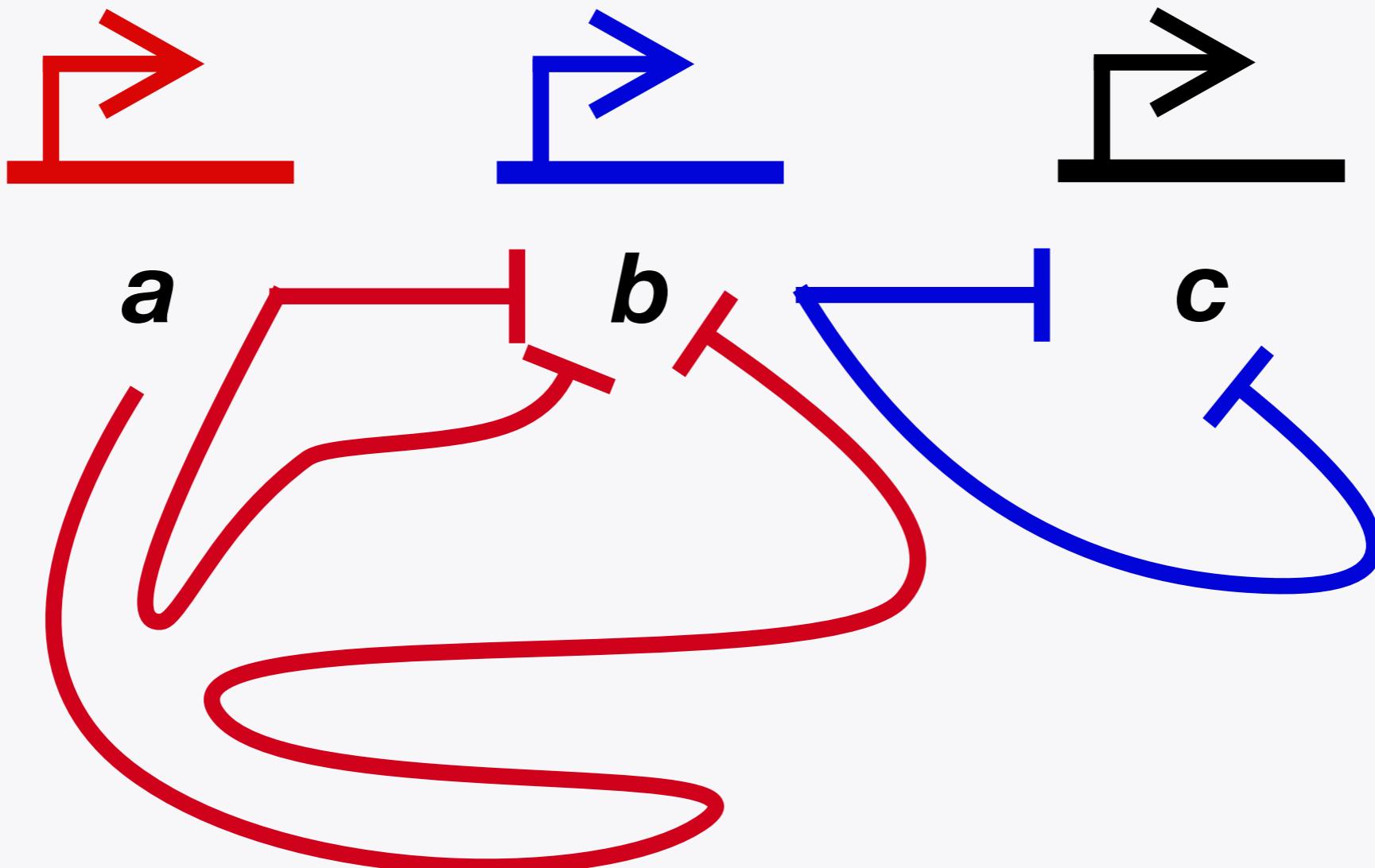


Batesonian Epistasis

A causal genetic pathway inferred through Batesonian epistasis

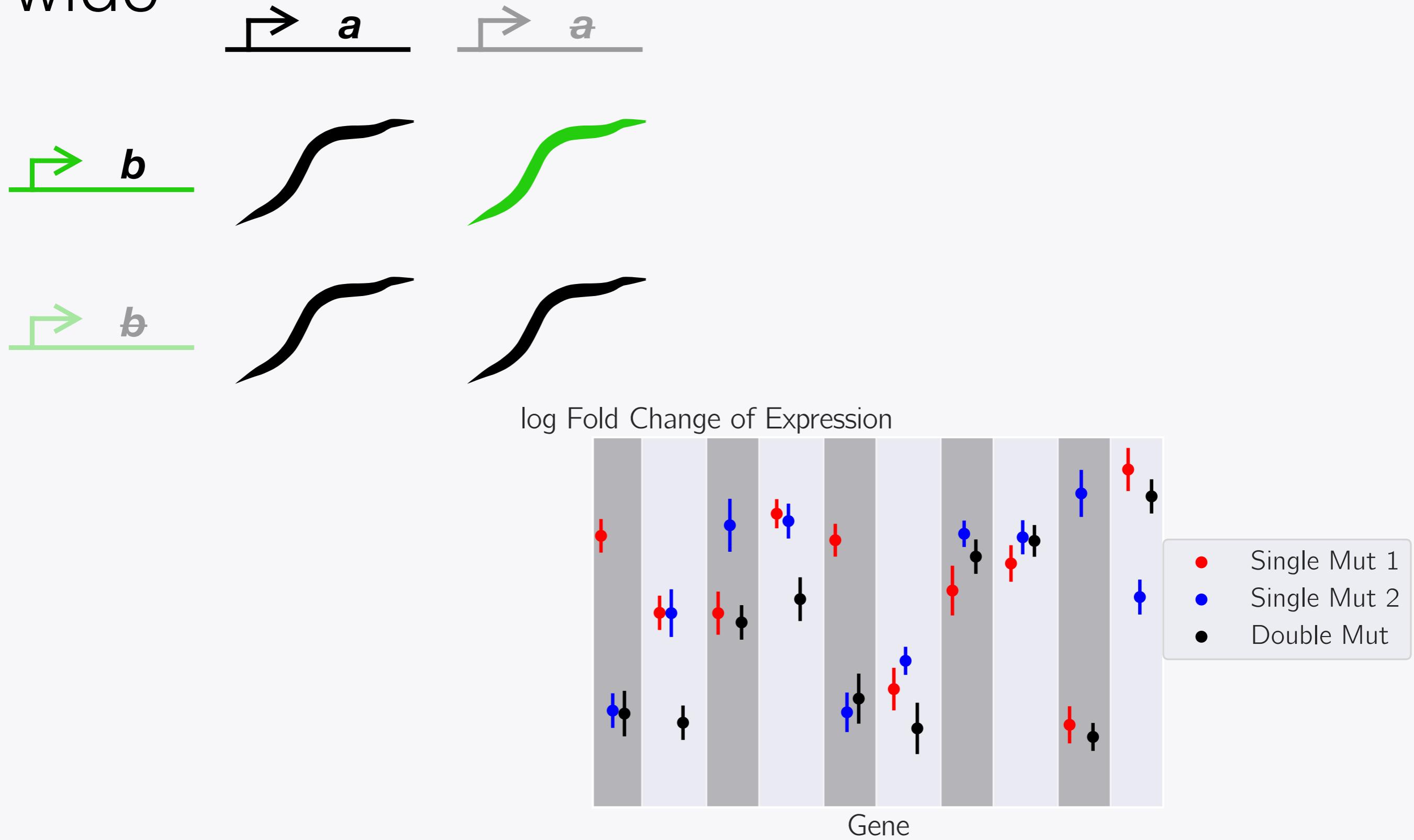


When does epistasis occur?



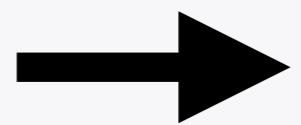
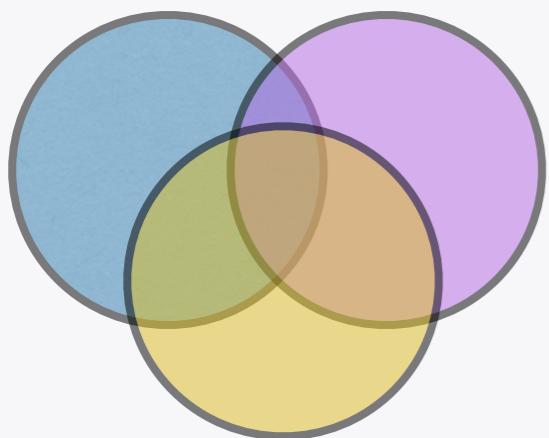
When all paths from **a** to **c** pass through **b**

Generalizing epistasis transcriptome-wide



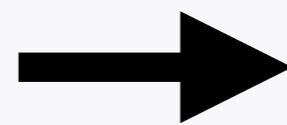
Transcriptome-wide epistasis analysis in a nutshell:

Multi-dimensional
phenotype

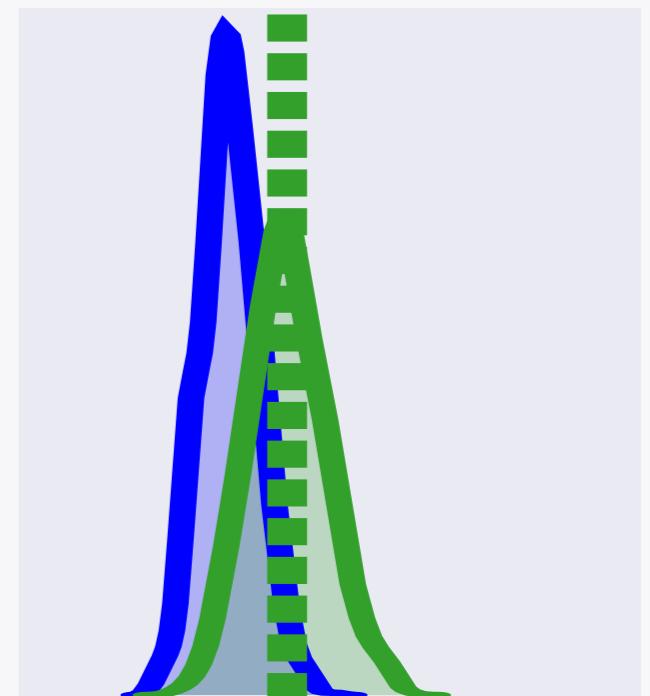


Turn phenotype
into a statistic

S



Check if statistic
is Batesonian

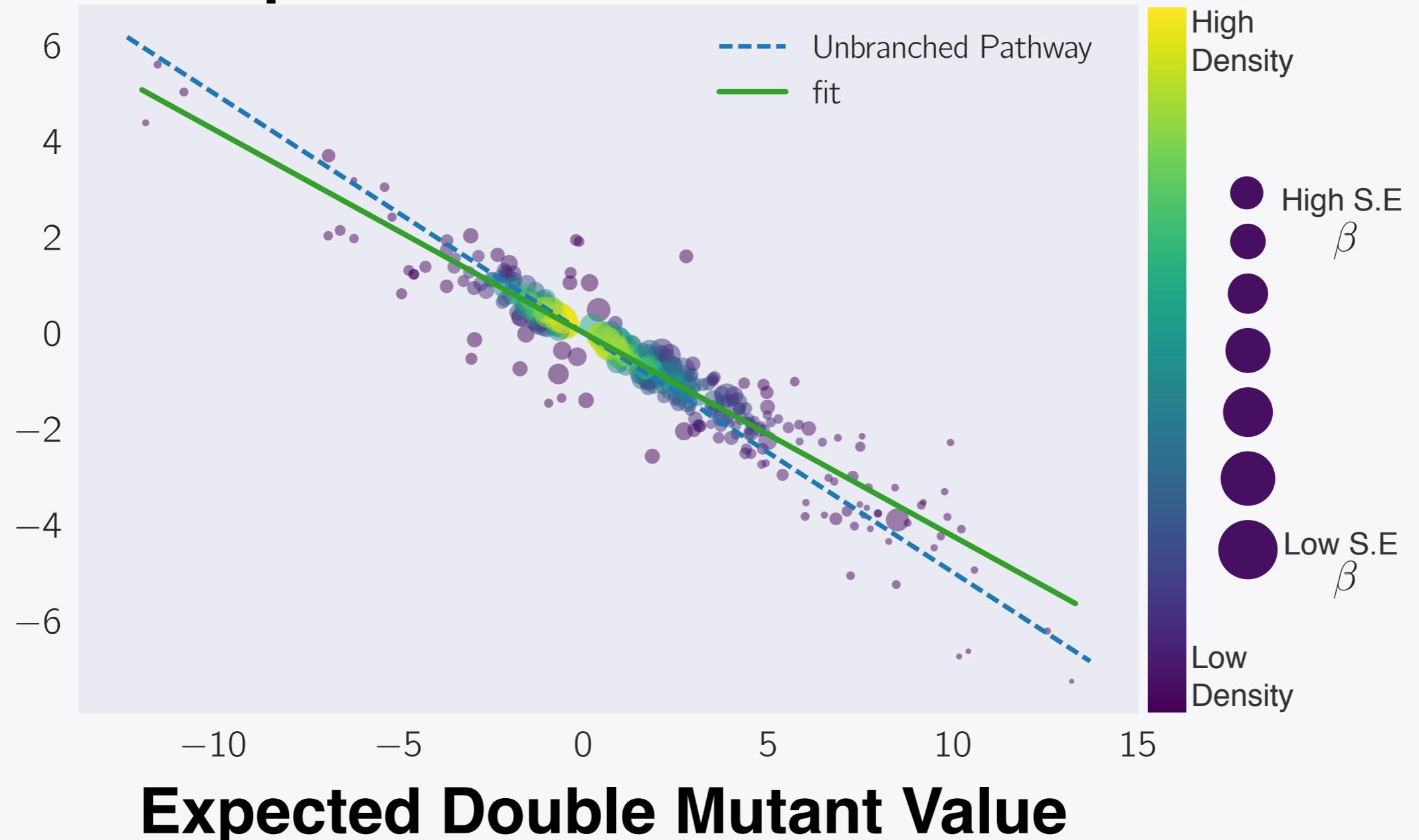


How do we compute this statistic?

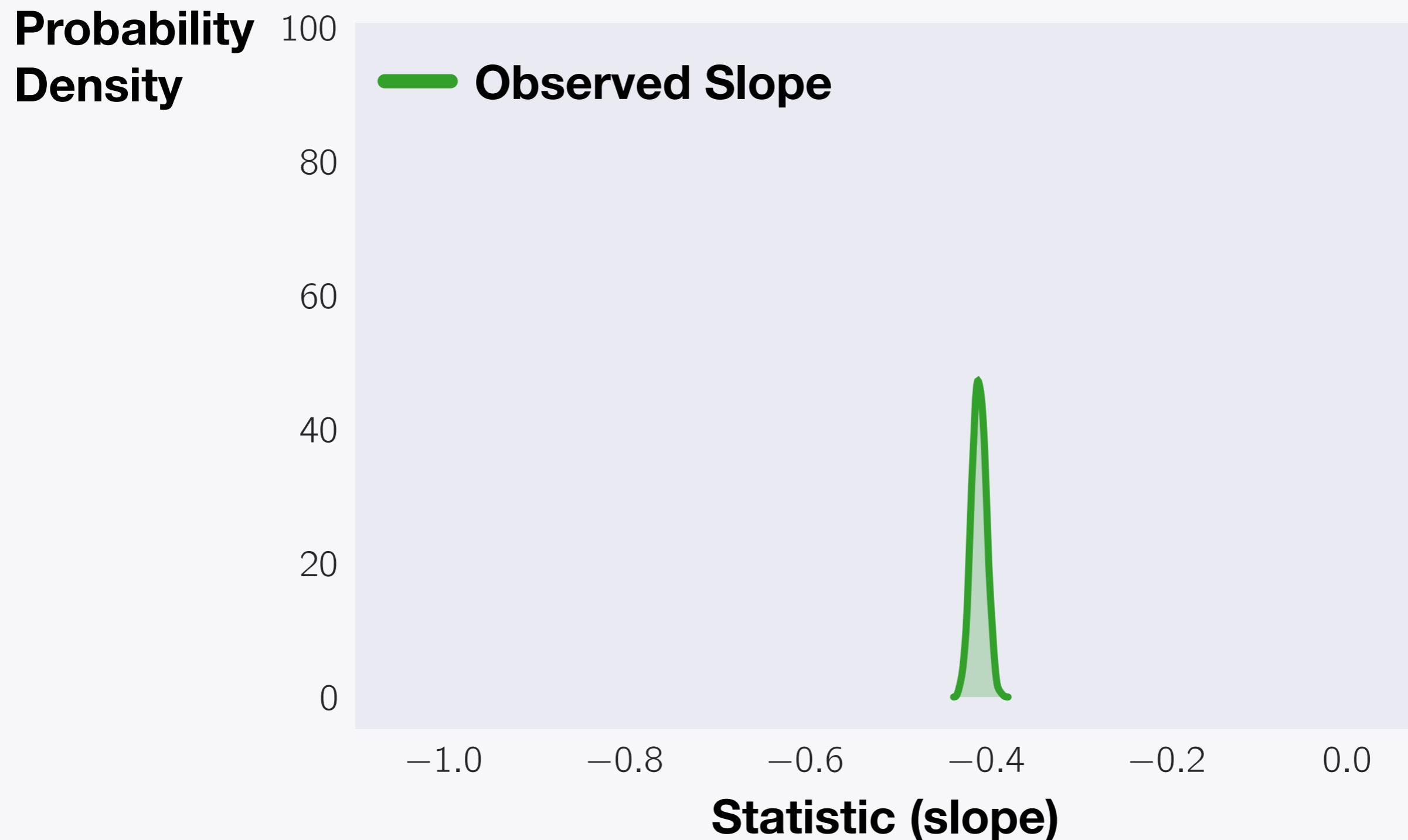
- (1) Calculate: **expected** double mutant value
(Add single mutants log Fold Changes)
- (2) Compute: **difference = observed - expected**
- (3) **Plot** difference vs. expected for all transcripts
and **find line of best fit**

An example: Calculating the statistic

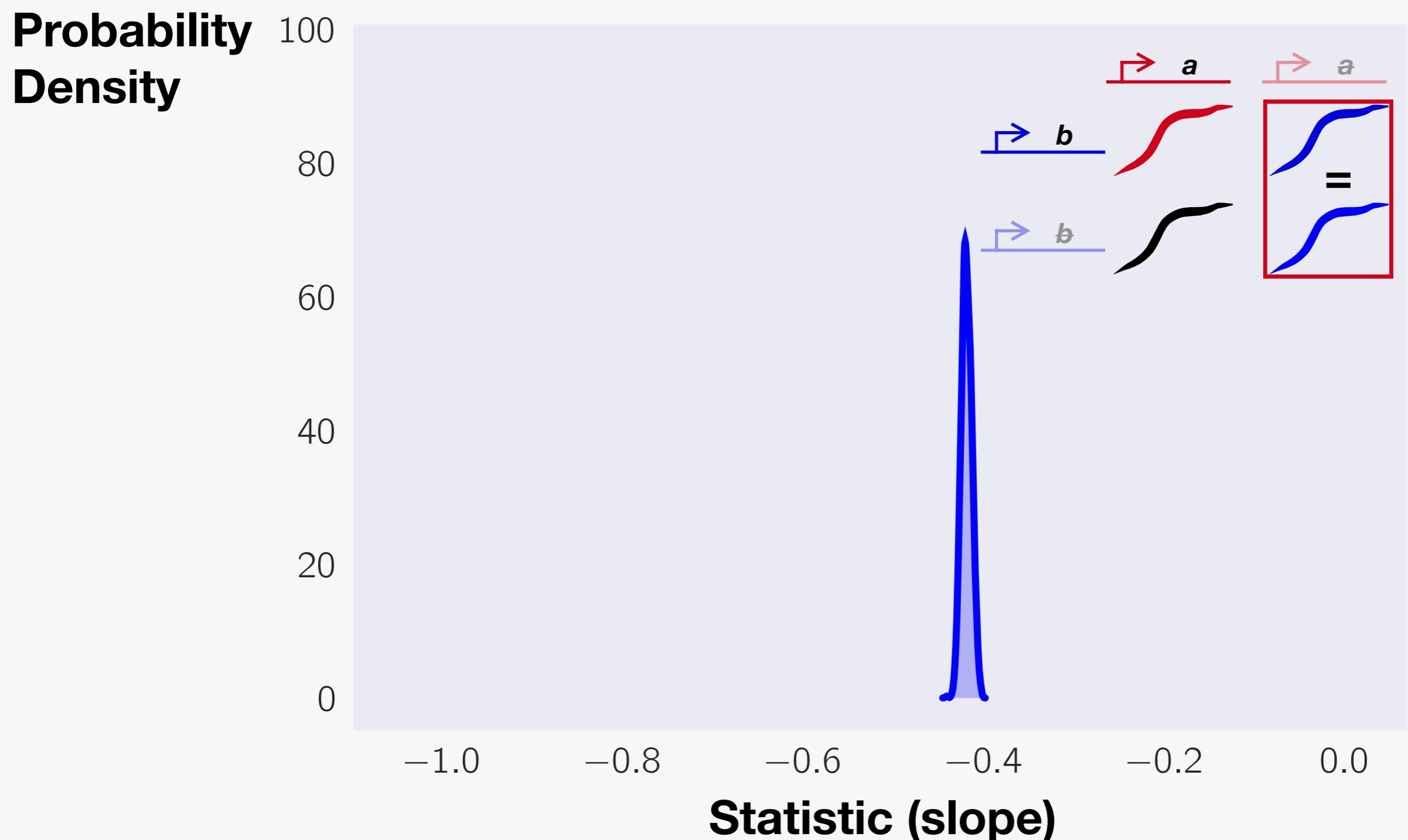
Observed - Expected



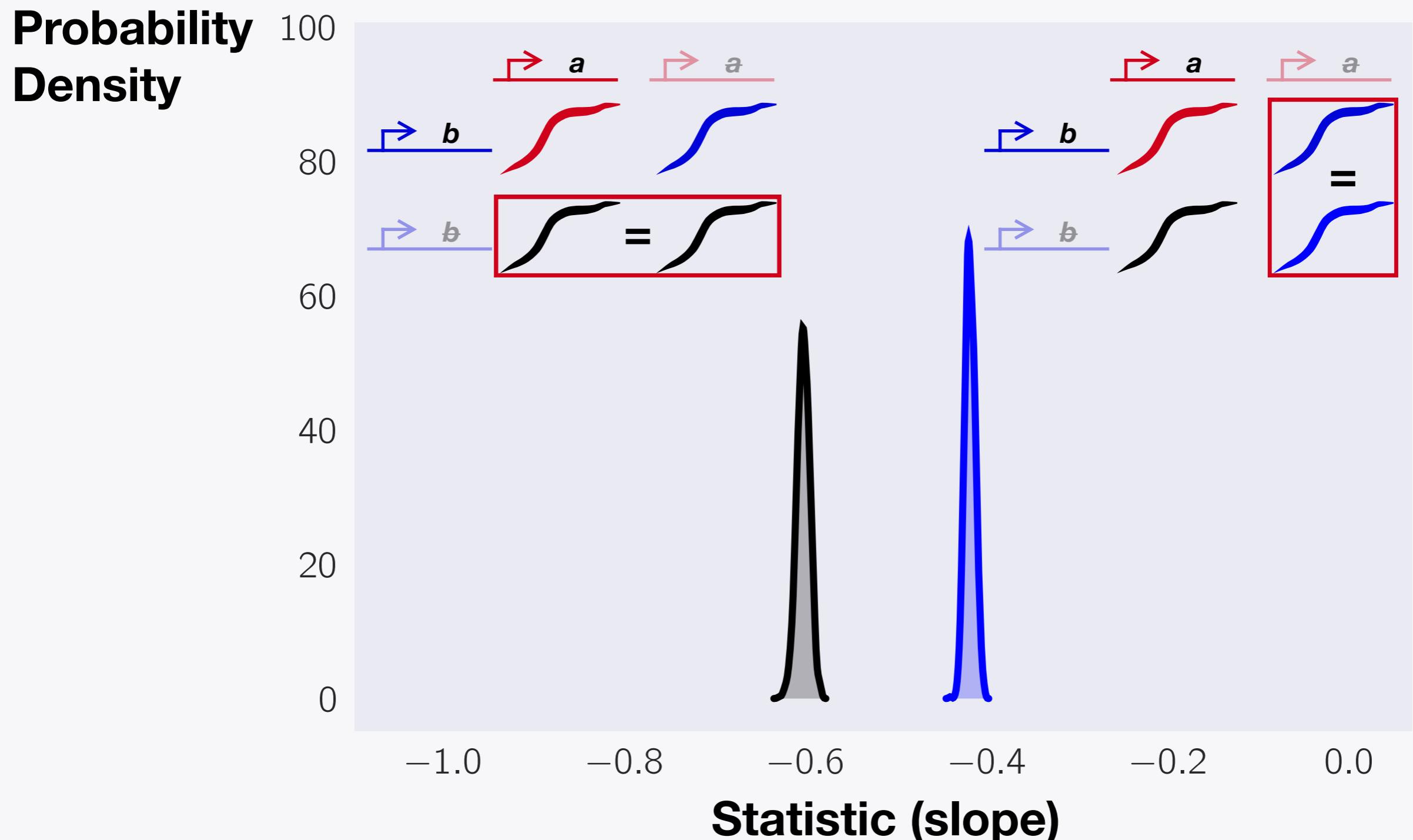
Density estimates show the confidence interval for the statistic (slope)



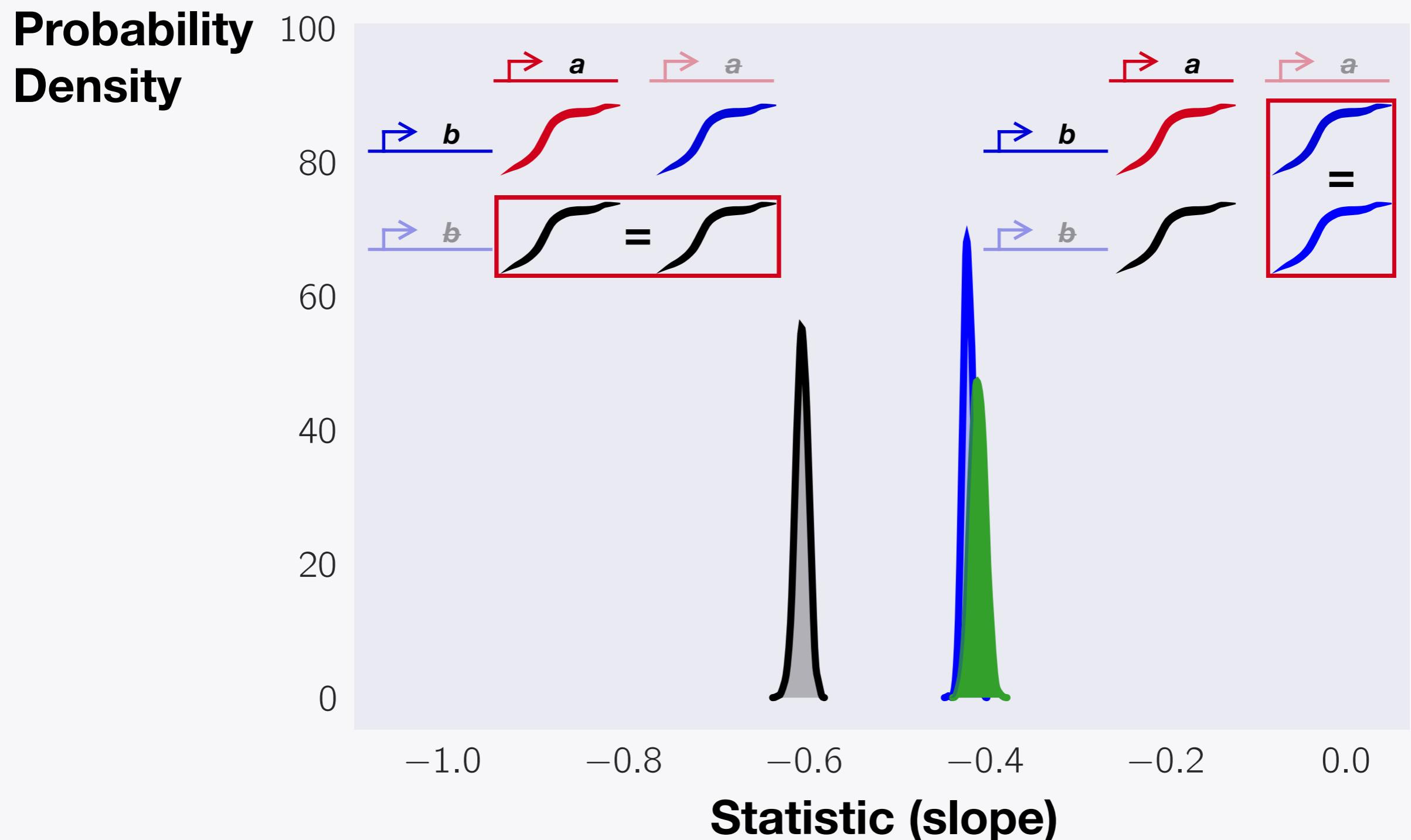
Simulations yield expected coefficient values for Batesonian epistasis



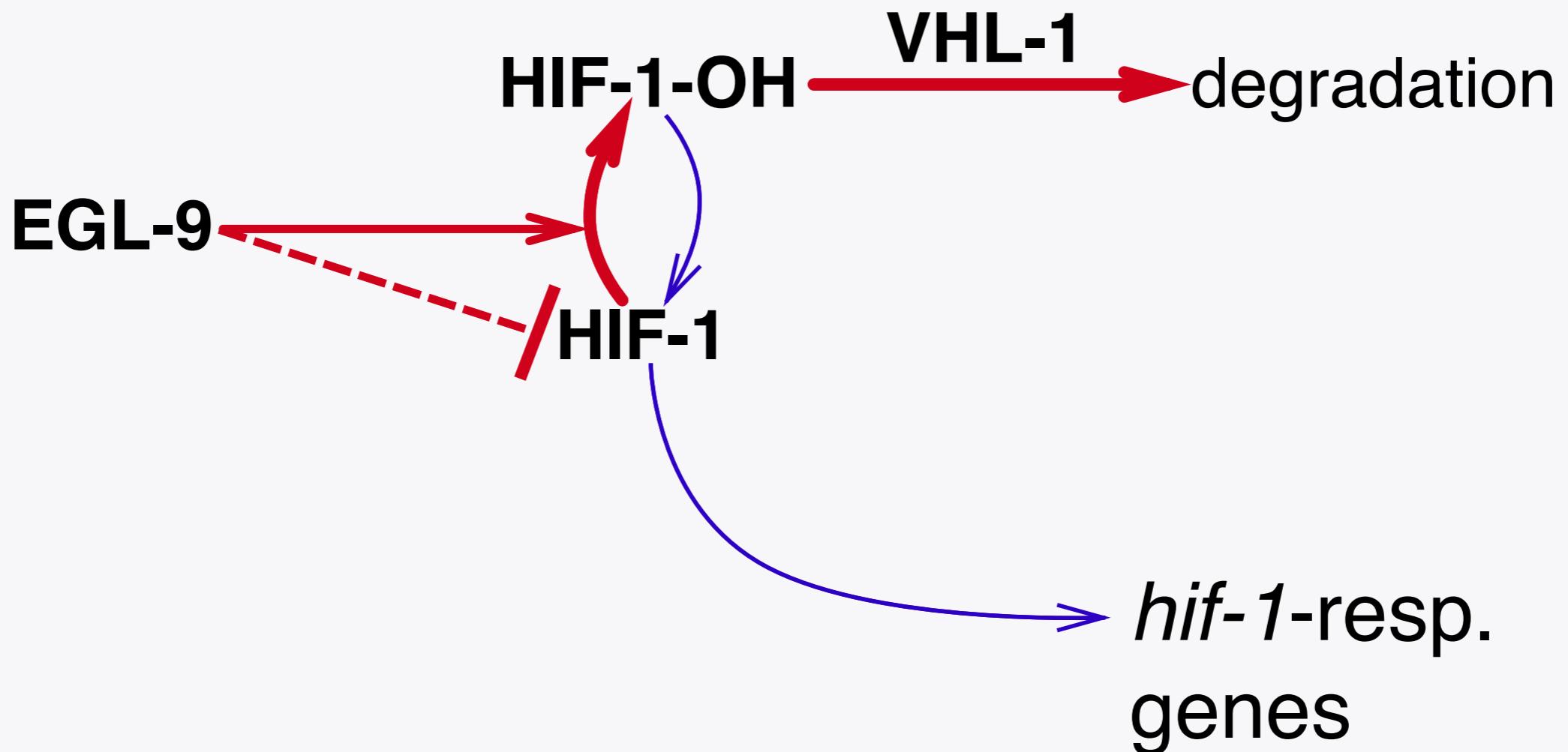
Simulations yield expected coefficient values for Batesonian epistasis



Checking for Batesonian epistasis

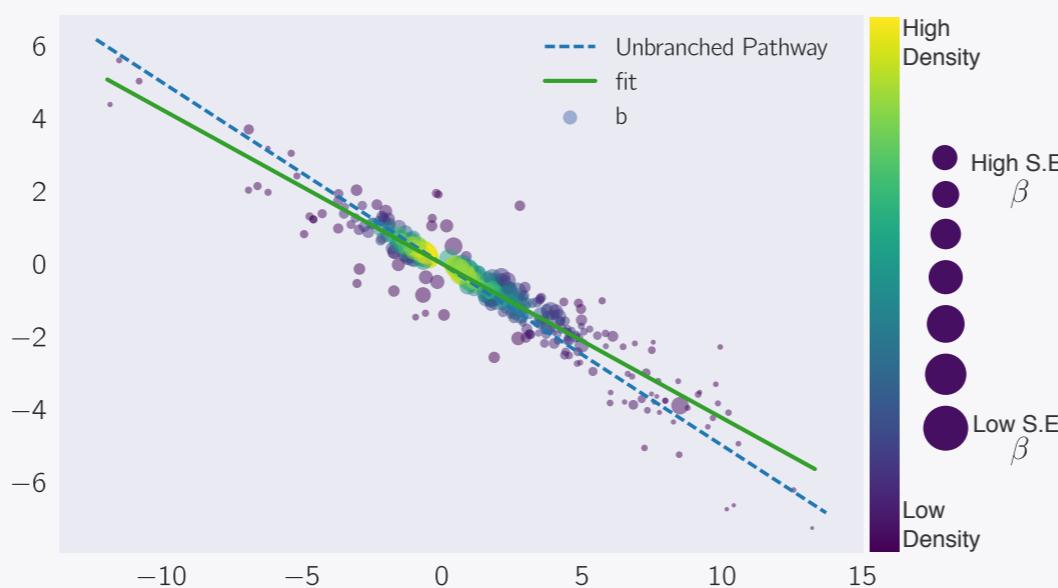


We reconstructed the *hif-1* pathway in *C. elegans*



Blind reconstruction using transcriptome-wide epistasis!

New phenotypes, new challenges, new discoveries



- **Phenotype:** Transcriptomes
- **Challenge:** Going beyond correlational experiments and analyses
- **Discovery:** The *C. elegans* female state

The geneticist's arsenal, automated

WormBase

Home

About

FAQ

Dropdown ▾

Server Status: Online

Alaska

A complete, automatic RNA-seq analysis pipeline.

Start a new project!

Classical geneticists are mathematicians

Genetics

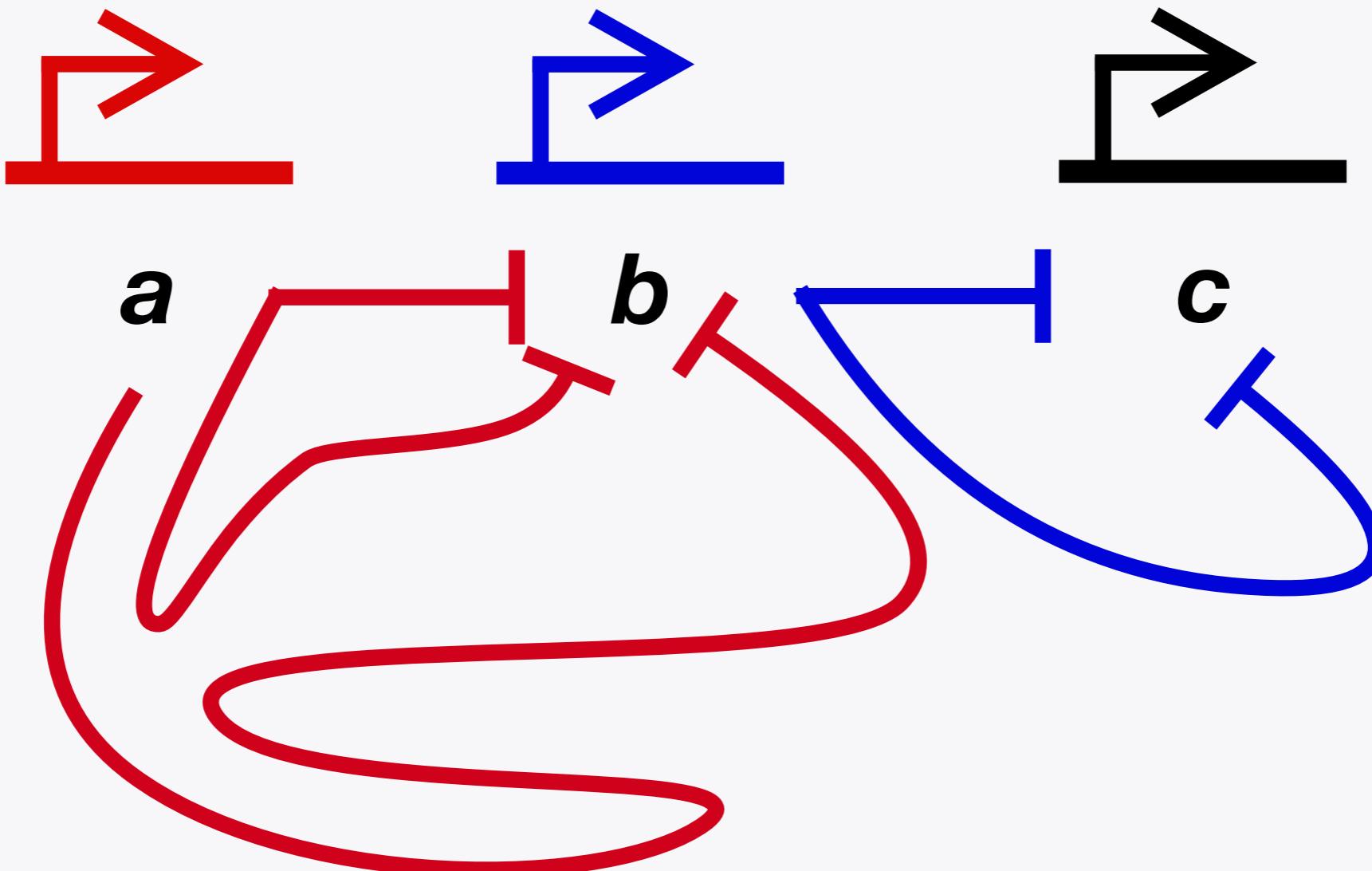
Epistasis

=

**Probability
Theory**

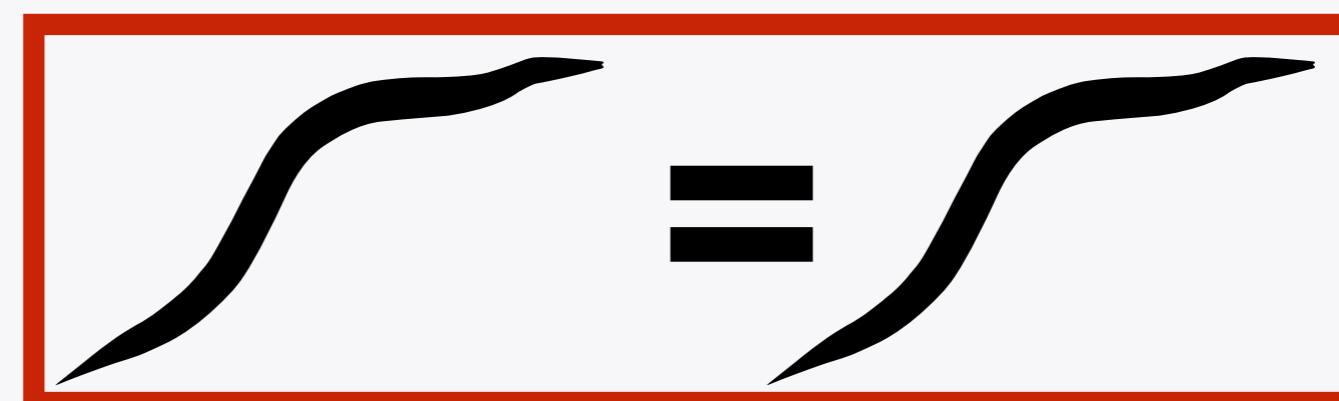
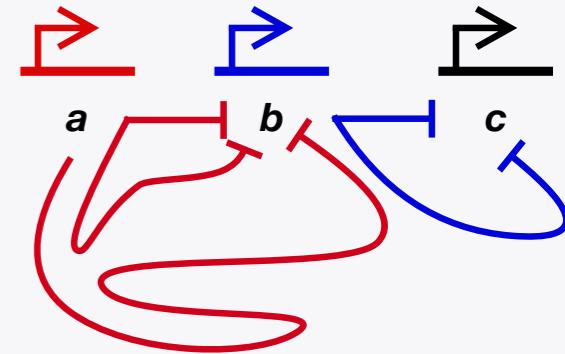
**Conditional
independence**

What is conditional independence?



All paths from **a** to **c** pass through **b**

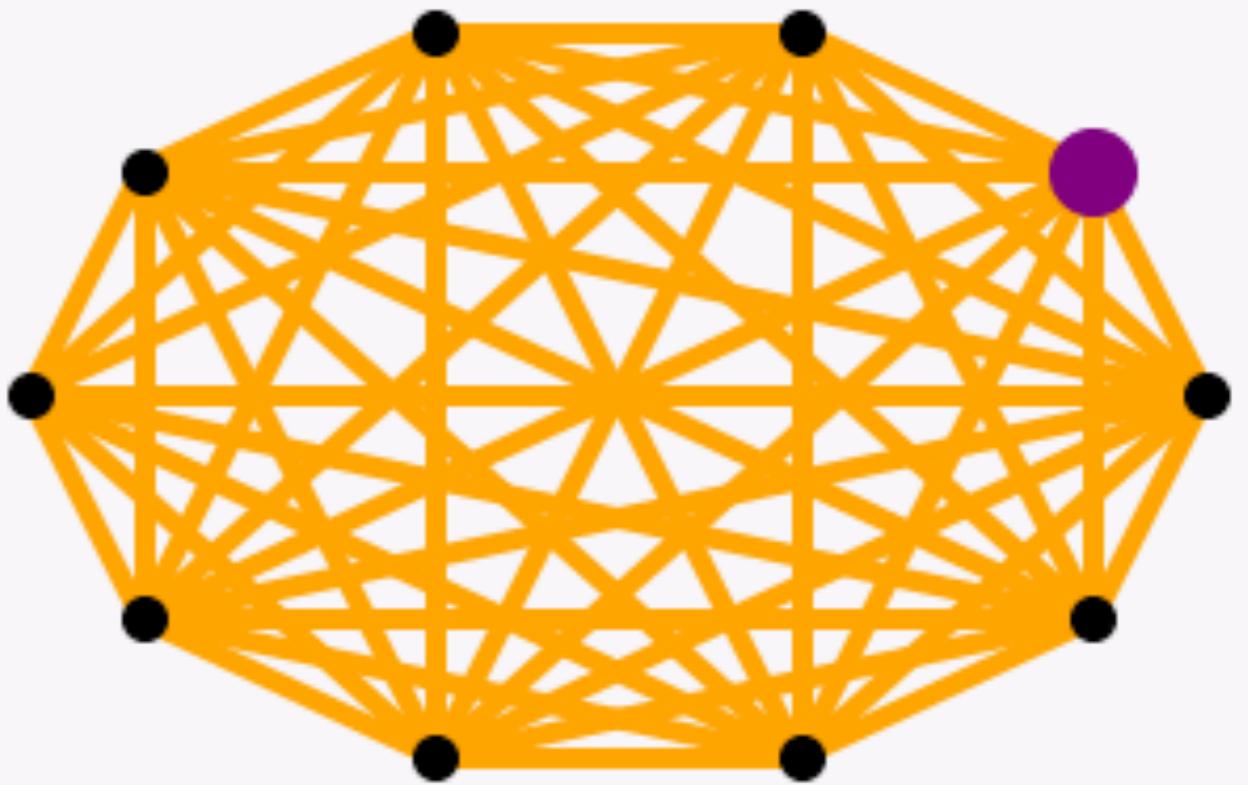
What is conditional independence?



Phenotype of b = Phenotype of ab

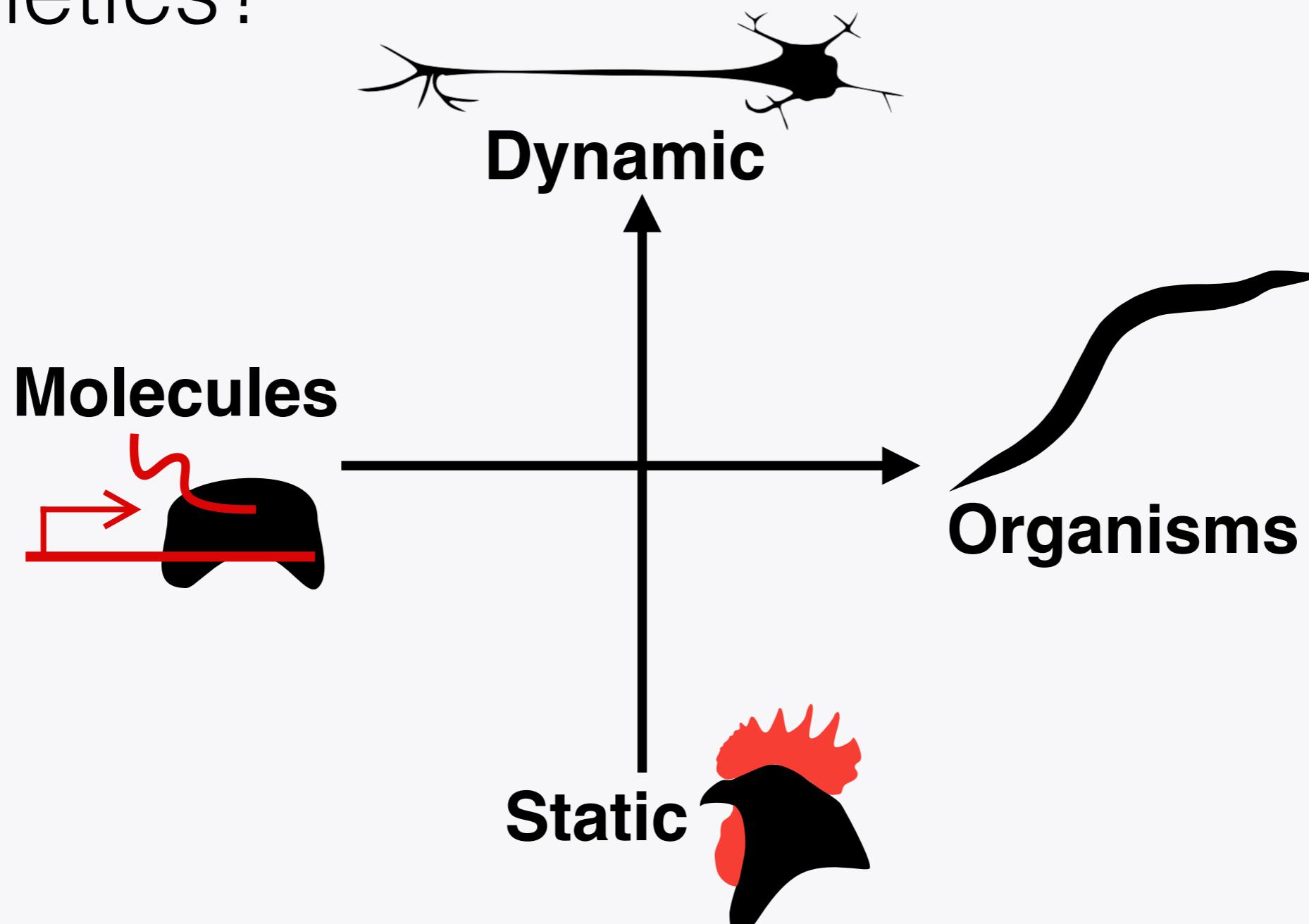
When faced with a complex network,
epistasis analyses are a great way to find
interactions

● **Mutated Gene (Null)**



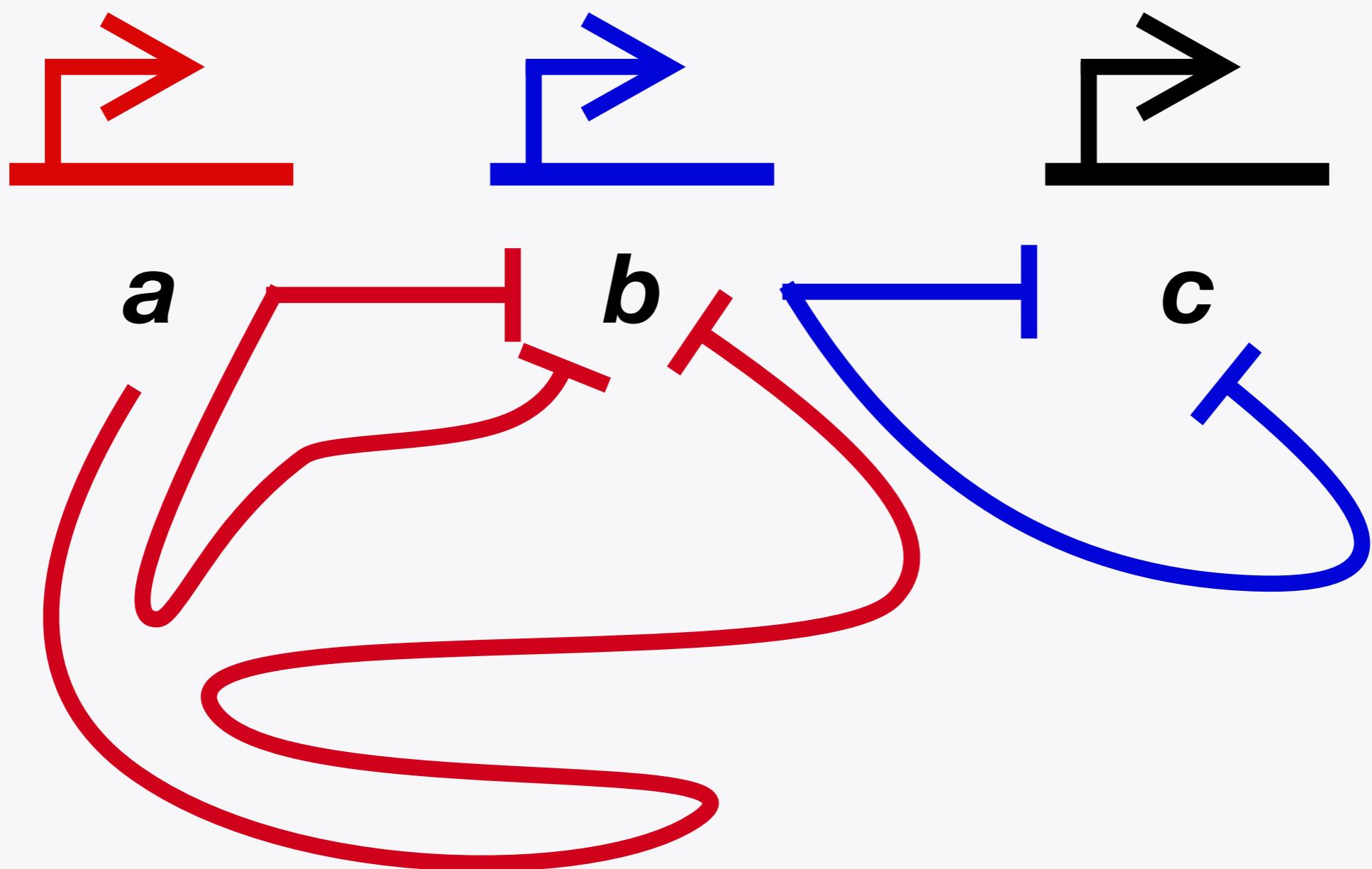
**Translating genetics into mathematics gives us tools
from computer science**

Phenotypes at scale: Why are they all good for genetics?

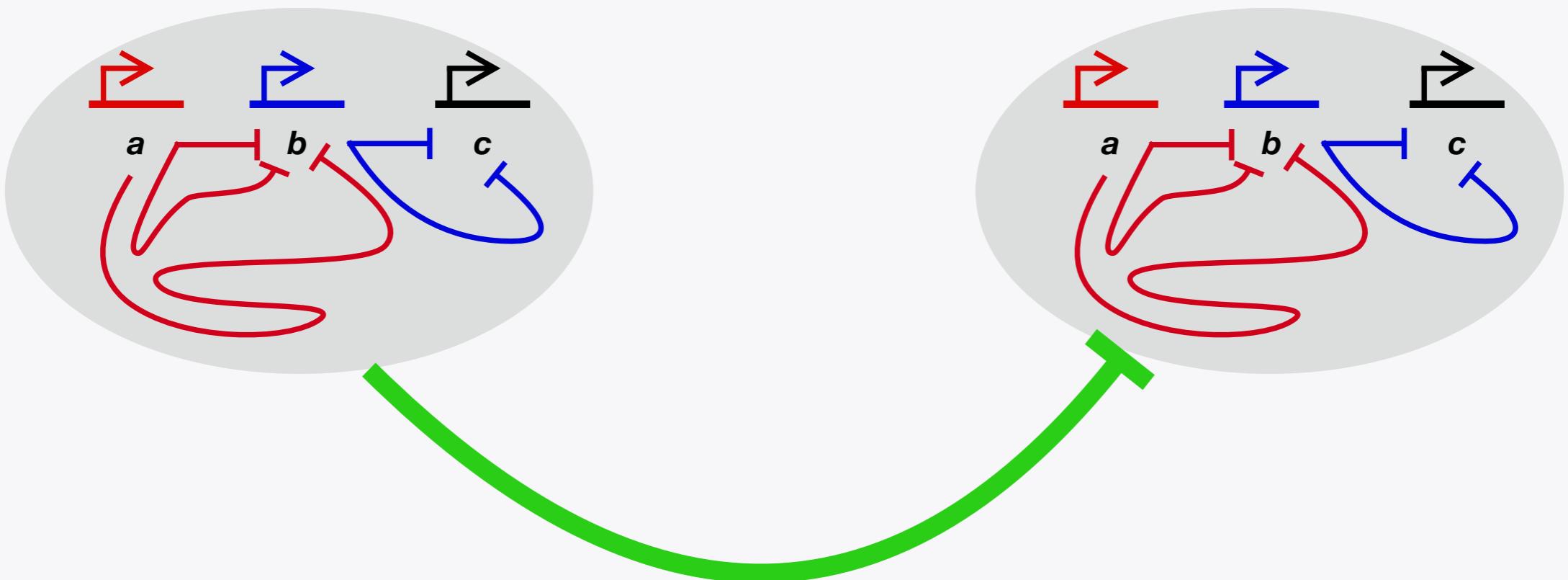


Conditional independence is scale-free!

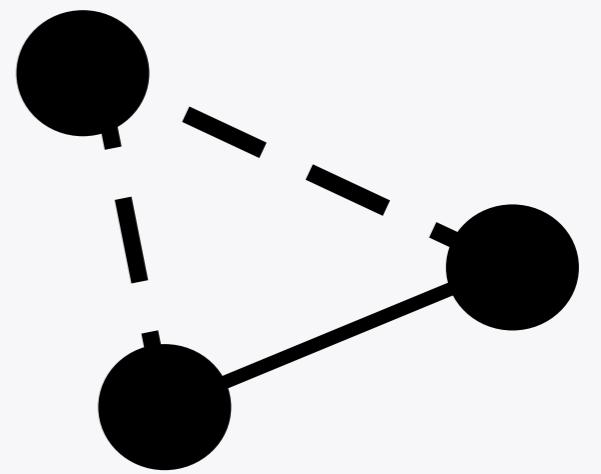
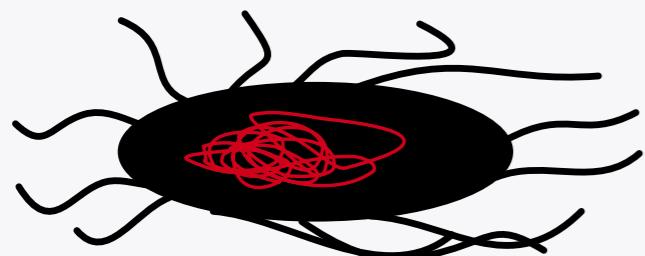
Epistasis can reconstruct pathways among individual genes



...but epistasis analyses can be extended to groups of genes!



Genetic principles will elucidate the **causal** interactions in the microbiome



Bacterium



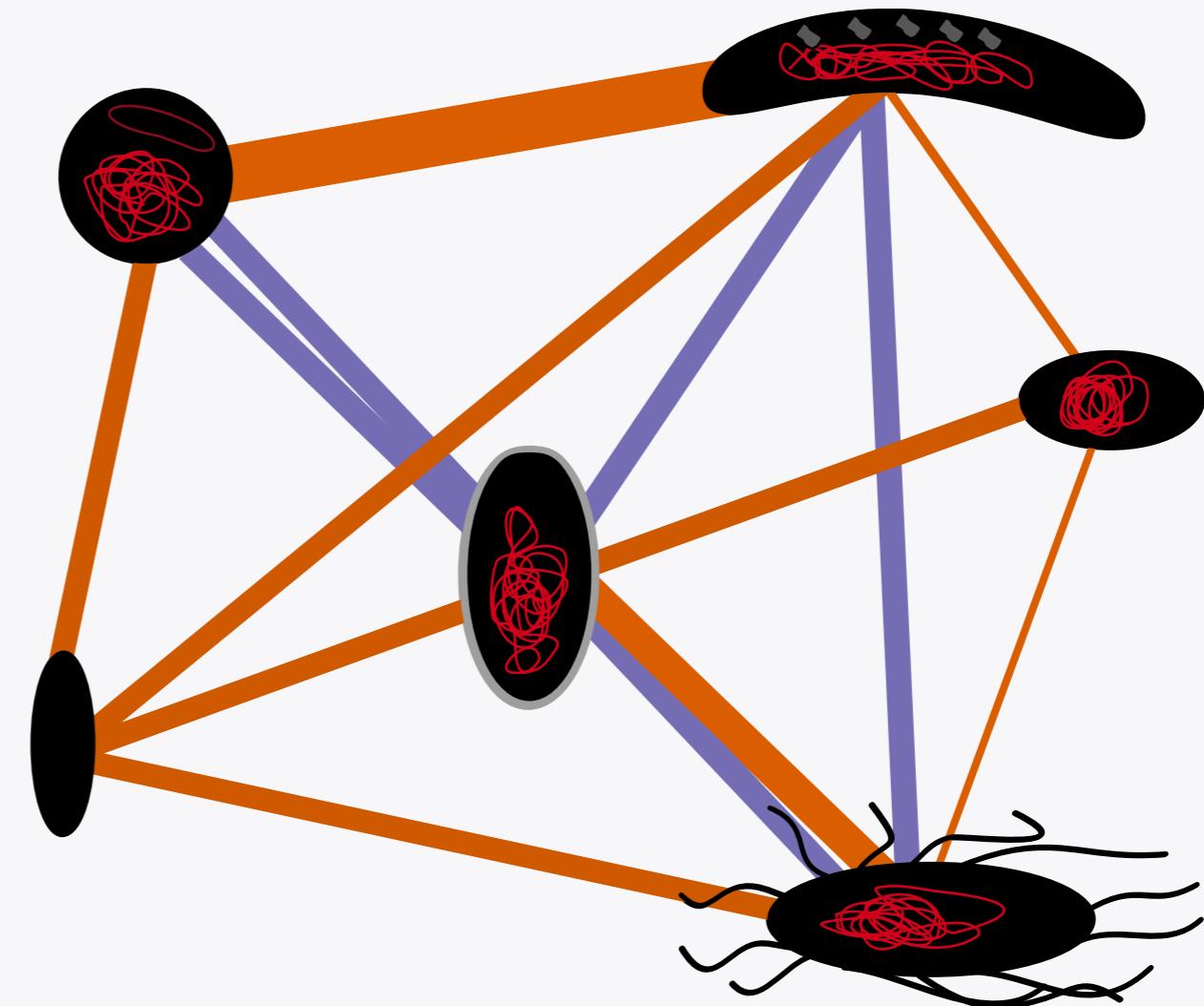
Bacterium

b

Community

How sparse or dense are microbial networks?

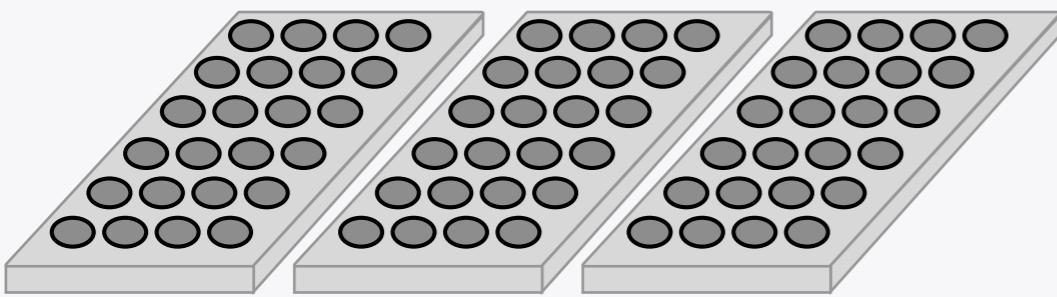
 Mutualism
 Antagonism



How many community members does one species interact with?

A platform for customizable microbiomes

Bacterial Library
(Global Microbiome
Conservancy Project)

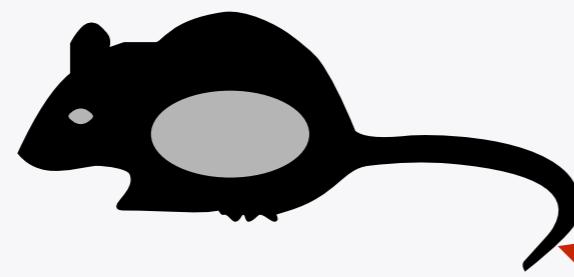


Poyet*, Groussin* *et al, in review*

Customizable Communities
in vitro
(> 20 species)

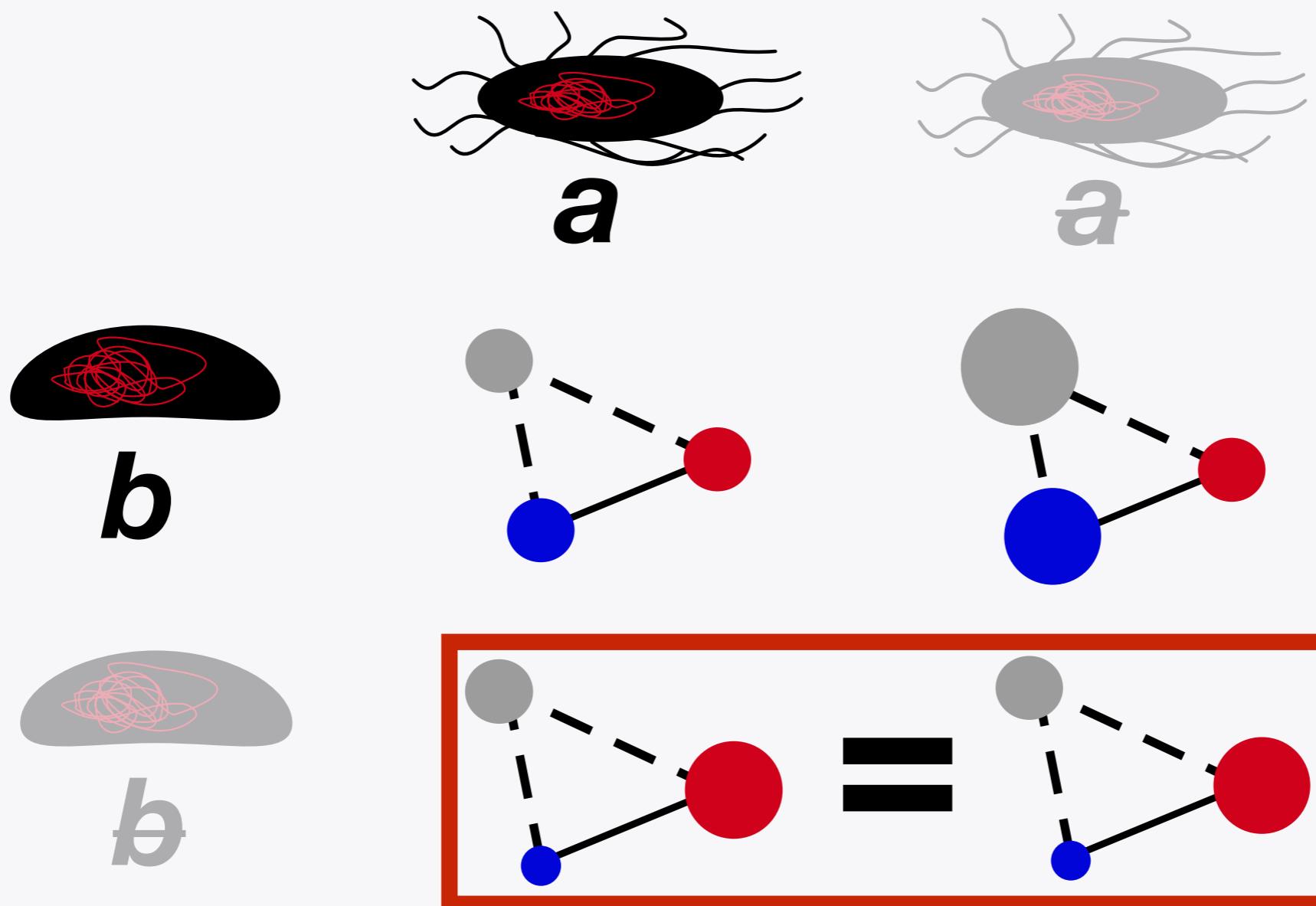


Customizable Communities
in vivo



future

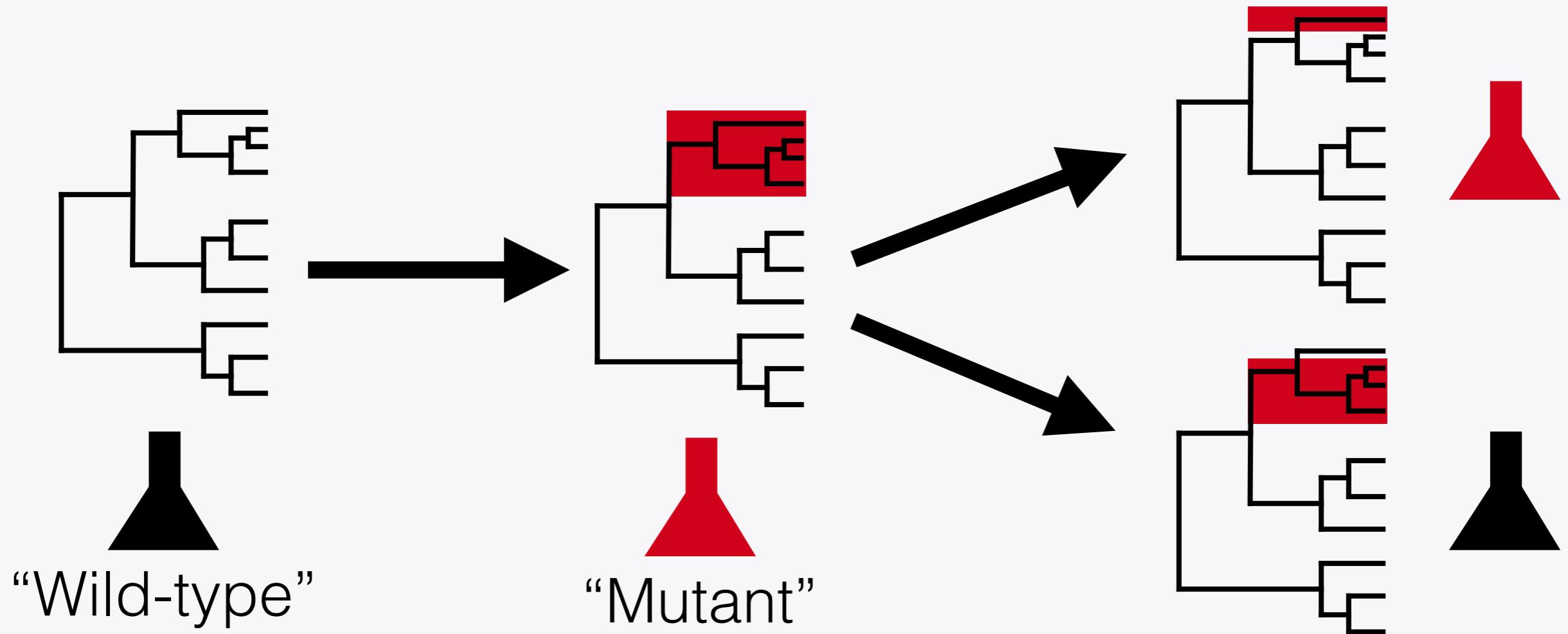
Determining bacterial interactions through epistasis using 16S community abundances



No genetically encoded tools required!

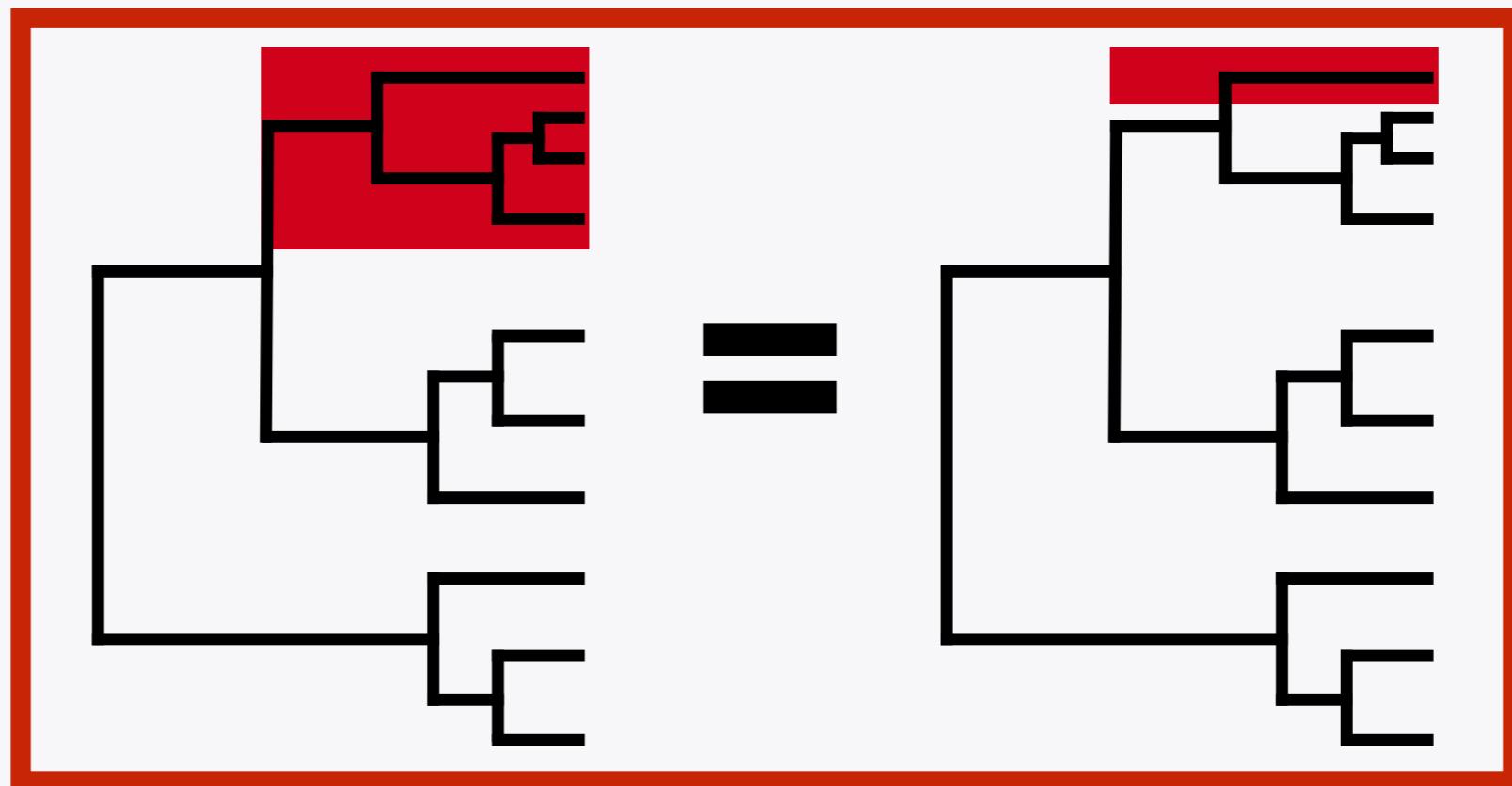
“Binary sorting” genetic analyses

Sort bacteria phylogenetically
Remove subset ('Double' Knockout)
Split subset ('Single' Knockout)



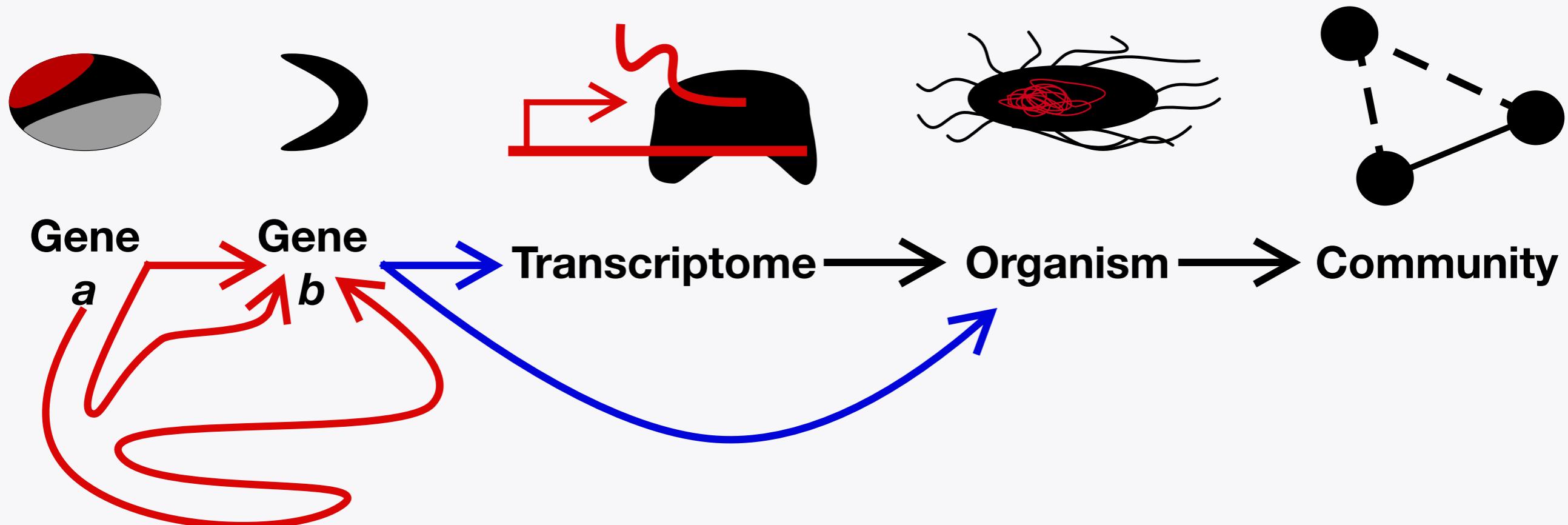
Algorithms can minimize the number of experiments and maximize information gain

“Binary sorting” genetic analyses



Instead of moving UP (single mutants, then double, then triple) moving DOWN allows us to optimize experimental designs

Biology is all about phenotypes



Paul Sternberg, Matt Thomson, Eric Alm



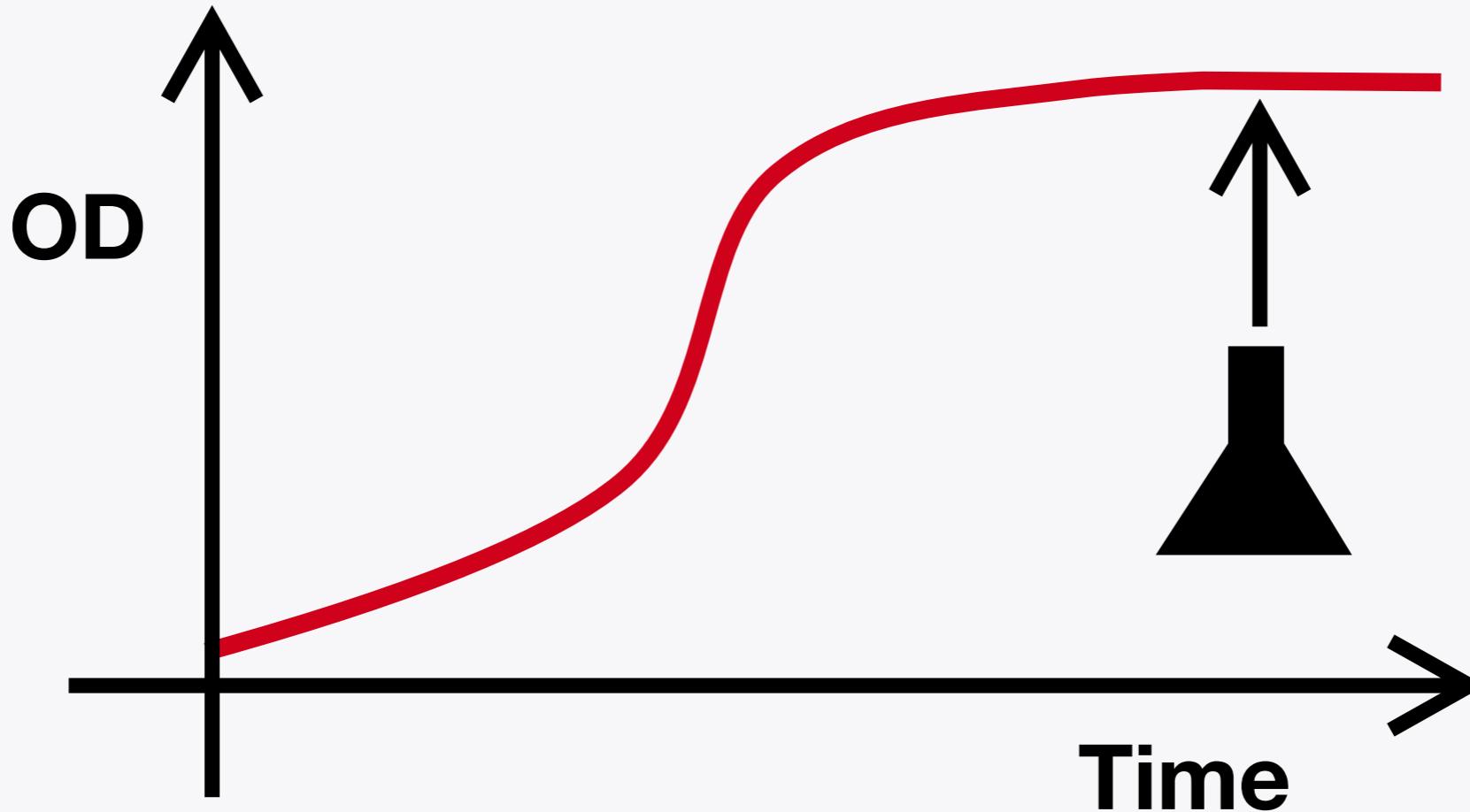
Hillel Schwartz
Carmie Puckett
Daniel Leighton
Kyung Hoi Min

Heather Curtis
Igor Antoshechkin
Vijaya Kumar
Erich Schwarz

Tiffany Tsou
Barbara Wold
Brian Williams
Dianne Newman



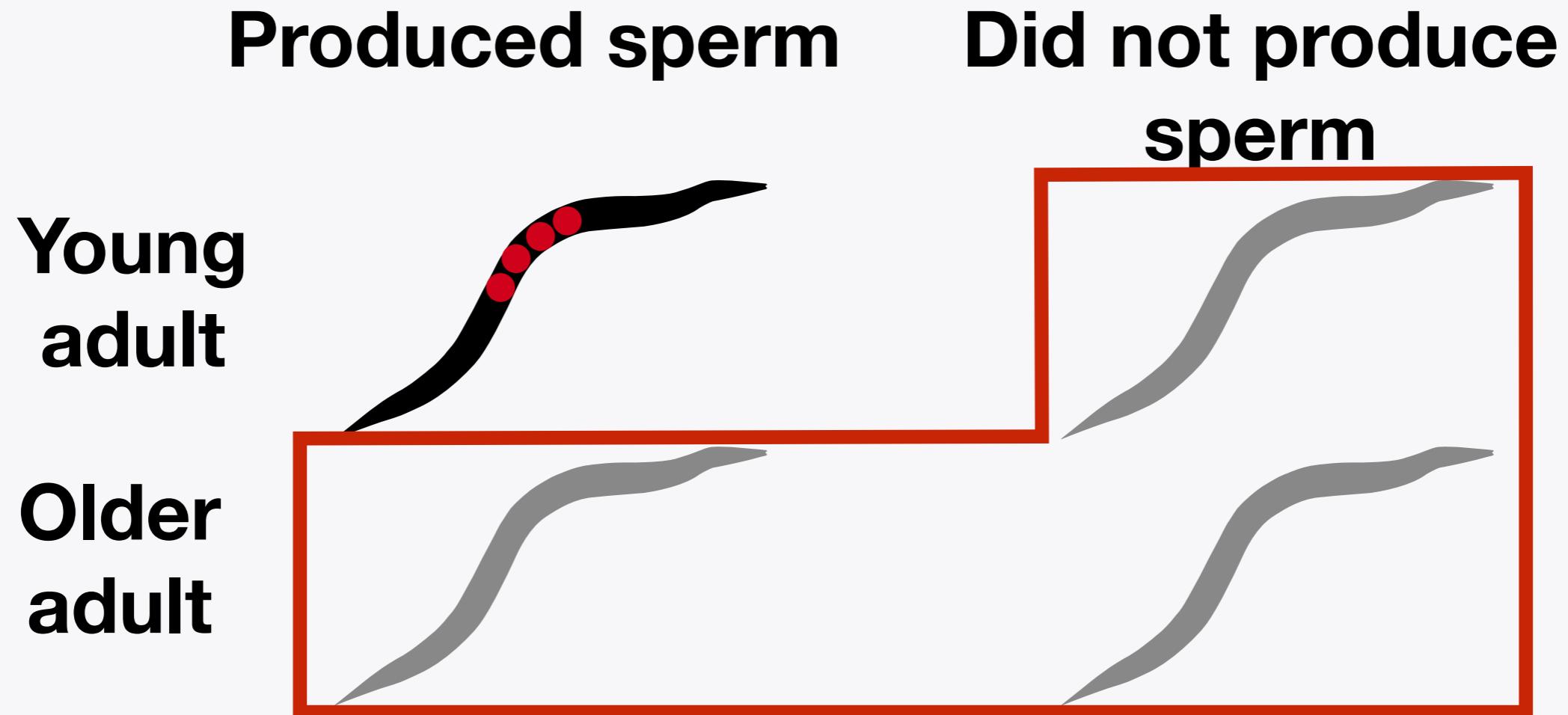
16S abundances as phenotypes



Community 16S measured when OD has saturated

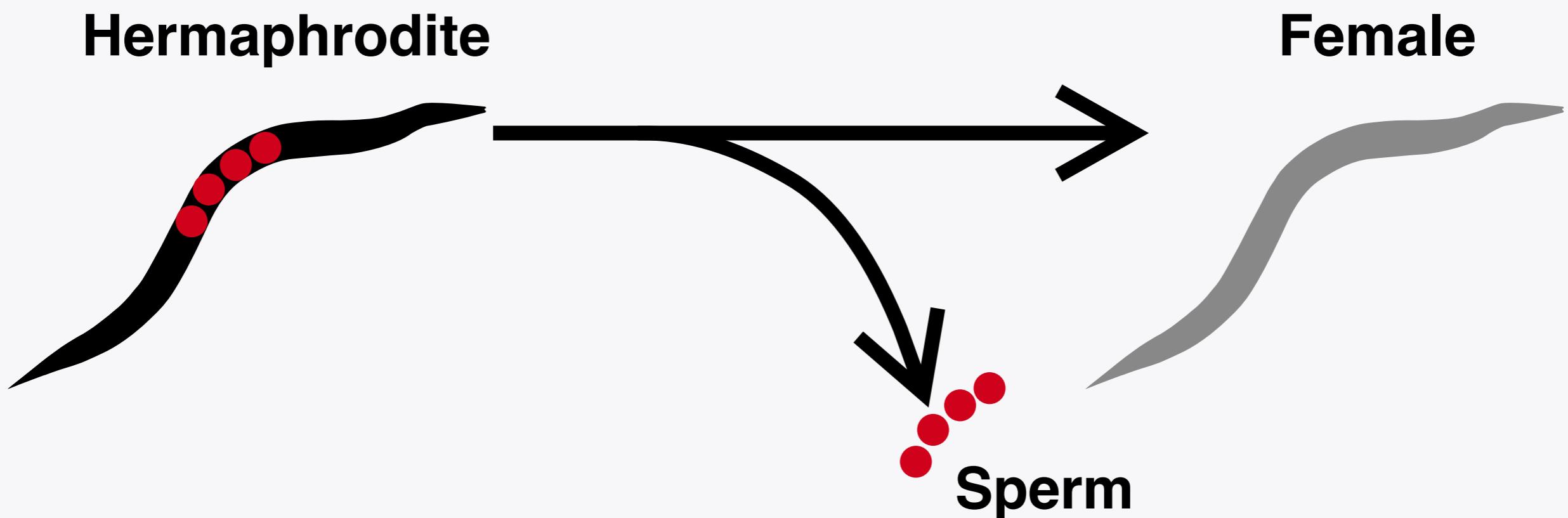
Hopefully, communities will be robust to initial conditions

A 2-factor experimental design decoupled sperm-status from aging effects



Transcriptome-wide epistasis analysis shows sperm loss, independent of mechanism, causes gene expression changes

C. elegans hermaphrodites become female upon sperm loss



**Transition into the female state causes
transcriptomic, metabolic and behavioral changes**

Geneticists use phenotypes to determine causality

