Quantifying epistatic conservation across genetic and environmental backgrounds

Andrew D. Mathis[1], Judith Boldt[1,3], and Kimberly A. Reynolds[1,2]

***Short Abstract* — Epistasis is a measure of how the effect of one gene is influenced by other genes. Quantifying epistasis will improve genotype-to-phenotype predictions, but our ability to generalize these results beyond model systems is dependent on (1) how much epistasis is conserved and (2) the sufficiency of conserved epistasis to improve phenotype prediction. To determine this, we are measuring how epistasis varies across *E. coli* strains and nutrient conditions. Additionally, we infer conserved epistasis statistically by analyzing gene co-evolution across bacterial species. Combining these data may point to a general strategy for predicting conserved epistasis and estimating phenotype in non-model organisms.**

***Keywords* — Epistasis, *E. coli*, evolutionary statistics, genetic background, environmental background, genotype-to-phenotype.**

I. Environment, genetic background, and epistasis

Epistatic interactions between genes are the basis for the complex traits, and a careful mapping of these genetic interactions is an important step towards making better genotype-to-phenotype predictions. Prior work has focused on mapping epistasis computationally via flux-balance analysis or experimentally using high throughput cell growth assays [1-3]. However, it is unclear how epistasis measurements made *in silico* or within a given model organism will generalize across different genetic and environmental backgrounds. Additionally, how well can conserved epistatic interactions be predicted? And is knowledge of only conserved epistatic interactions sufficient to improve genotype-to-phenotype predictions? We are developing complementary experimental and computational strategies to address these questions.

II. Experimental measurements of variation in epistasis

We selected 22 enzymes from bacterial folate and purine metabolism as a model system for measuring epistatic conservation. Folate and purine metabolism are conserved pathways, which allows for more powerful statistical co-evolution analyses, and perturbations in these pathways result in experimentally measureable growth phenotypes. To measure epistasis, we use CRISPR interference (CRISPRi) to knockdown gene expression [4]. CRISPRi specificity is directed using a guide RNA (gRNA) library that targets our 22 selected enzymes individually and in all possible combinations. We then grow this library in a continuous culture device (turbidostat), regularly sample the bacterial population, and count allele frequencies using next generation sequencing. Our method can resolve a 2% difference in relative doubling time per hour, is applicable in a variety of *E. coli* strains (and other organisms), simplifies targeting essential genes, and easily scales for studies in different environments. Additionally, we have developed a strategy for post hoc removal of "escapers", which occur when an adaptive mutation renders CRISPRi ineffective. Using this method, we are currently measuring epistasis in seven strains of non-pathogenic *E. coli* and in multiple environmental conditions (variations in media, culture density) [5].

III. A statistical model of conserved epistasis

An alternative approach is to estimate epistatic interactions using quantitative statistical models of co-evolution [6]. The basic premise is that epistatic constraints between genes will drive their co-evolution, leading to detectable statistical correlations like conserved proximity on the chromosome (synteny), joint presence and absence, or amino acid sequence covariation. Because we search for these correlations in large and diverse genomic databases, the interactions we find likely represent conserved epistasis, rather than background specific idiosyncrasies. We have conducted statistical co-evolution analyses across folate and purine metabolism using a database of ~2000 bacterial species, and hypothesize that these data will be more like the conserved epistatic interactions we find experimentally than those sampled from any single organism.

Acknowledgements: This work is funded by Gordon and Betty Moore Foundation Data-Driven Discovery grant: GBMF4557.

[1]Green Center for Systems Biology, [2]Department of Biophysics, University of Texas Southwestern Medical Center, Dallas. E-mail: Andrew.Mathis@UTSouthwestern.edu, Kimberly.Reynolds@UTSouthwestern.edu

[3]Center for Interdisciplinary Research in Biology, Collège de France, Paris

References

1. Butland, G. et al., eSGA: E. coli synthetic genetic array analysis. *Nature methods* 2008**,** *5* (9), 789-795.
2. Costanzo, M. et al., A global genetic interaction network maps a wiring diagram of cellular function. *Science* 2016**,** *353* (6306), aaf1420.
3. Segre, D. et al., Modular epistasis in yeast metabolism. *Nature genetics* 2005**,** *37* (1), 77-83.
4. Qi, L. S. et al., Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 2013**,** *152* (5), 1173-1183.
5. Monk, J. M. et al., Multi-omics quantification of species variation of Escherichia coli links molecular features with strain phenotypes. *Cell systems* 2016**,** *3* (3), 238-251. e12.
6. Schober, A. F. et al., An evolutionary module in central metabolism. *bioRxiv* 2017, 120006.