The modern broiler (meat) chicken, selected for production characteristics over the past 70 years, demonstrates marked improvements in feed efficiency and muscle growth. The liver plays a primary role in this, especially over the first three weeks’ post-hatch while the birds transition from reliance on stored yolk nutrients to a carbohydrate-rich diet. A preliminary comparison of hepatic transcriptome and metabolome from Day 4 and Day 20 post-hatch has identified multiple differences, especially in core metabolism and stages of liver growth. This work exploits a longitudinal high-throughput dataset containing thousands of features. High-dimensional data like this is often unwieldy, and prioritizing groups of features to generate biological meaning is a persistent challenge. Additionally, groups of biological replicates often display heterogeneity, yet this can contain meaningful information. Here, we develop an automated workflow to rank large sets of features with respect to a measurable trait, and reduce them to the simplest model describing underlying biological variation that is present.

Finite mixture models are used to detect latent “classes” of variation within a group of individuals. These classes lend themselves to regression methods using prioritized groups of features from high-throughput data. First, Spearman correlation is used to efficiently screen a set of 30,000 genes and metabolites by relationship to a trait of interest: in this case Normalized Liver Mass. The top candidates are then used in lasso regression to further reduce the pool by eliminating less meaningful features. Finally, we refine a model to find the simplest set of candidate features that can explain unknown variation in the trait of interest, uncovering biological meaning contained in the relationships, leading to inference quantifying the specific nature of those relationships, and generating testable hypotheses for future experiments.