Fully Bayesian Analysis of High-Throughput Targeted Metabolomics Assays

James L. Crooks, Denise K. MacMillan and Jane E. Gallagher
National Health and Environmental Effects Research Laboratory, U.S EPA, Research Triangle Park, NC

High-throughput metabolomic assays that allow simultaneous targeted screening of hundreds of metabolites have recently become available in kit form. Such assays provide a window into understanding changes to biochemical pathways due to chemical exposure or disease, and are useful tools for toxicological studies and biomarker discovery. The manufacturer has recently begun including a statistical analysis package with these kits, which the analytical chemistry community has found very helpful. This package supports preliminary data exploration and basic analyses, but it is less suited to complex experimental designs, and it lacks multiple-testing adjustment and other tools for promoting reproducibility. Fortunately, the smaller number of features measured by such assays relative to that for genomic assays makes a fully Bayesian approach practical on standard computing hardware. We present a novel model for analyzing the output of high-throughput metabolomic assays in a statistically rigorous manner. The model includes flexible spline-based normalization across samples, as well as Bayesian variable selection and multiplicity adjustment. It also allows a choice of likelihood specification and link function. The model can be used either in differential expression-style studies where the metabolite concentrations are treated as dependent variables or in susceptibility-style studies where the metabolite concentrations are treated as independent variables. Examples for both study types are given using data from the Mechanistic Indicators of Childhood Asthma (MICA) study targeting 186 metabolites associated with a variety of metabolic disorders.