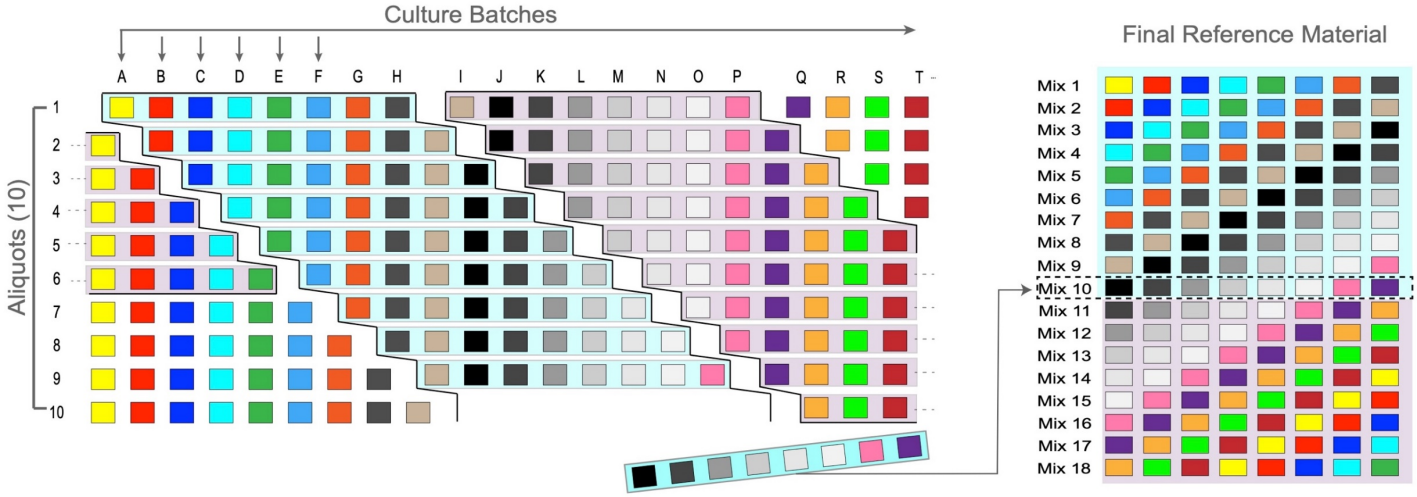
Development of a *Caenorhabditis elegans* Reference Material for Long-Term Metabolomics Quality Control and Unknown Compound Identification

Goncalo J. Gouveia,[1] David L. Blum,[1] Emerson Ferreira Queiroz,[2] Jean-Luc Wolfender,[2] Lauren M. McIntyre,[3] Arthur S. Edison[1]

1) Department of Biochemistry, University of Georgia, 2) Faculty of Sciences, University of Geneva, 3) Department of Molecular Genetics and Microbiology and Genetics Institute, University of Florida

Metabolomics is a well-established tool applied to numerous scientific problems across various disciplines. Despite this, two main challenges still remain: reproducibility and identification of unknown features. One critical component for reproducibility is biological reference material that is stable over time. Small sets of chemical standards are crucial but fail to provide an adequate reference for NMR or LC-MS spectra in untargeted studies. Ideally, a biological reference material should be measured as part of quality assurance/quality control in every batch of experimental samples in a large-scale study that might last for years and involve multiple labs. Use of a common constant biological reference material in combination with chemical standards should enable integration of studies of the same organism, tissues, or cells. The question we are addressing in this study: **How do we make and characterize a biological reference material that is stable over many years?**

To generate this reference material, we utilize a moving average. The concept is straightforward: individual batches are combined to generate a reference pool that is averaged over batches. For each iteration one batch is added and another removed, which results in most of the batches used in the mixture staying constant. Originally developed for delivering consistent products from distilling, batch averaging has modern theory in industrial manufacturing supporting this concept. We are utilizing small batch averaging to develop a robust biological reference material using *C. elegans* and its bacterial food. Multiple batches of *E. coli* (food) and *C. elegans* are grown in bio-reactors under strictly controlled conditions. Multiple aliquots are constructed from each batch. We minimize inter-batch variability, creating a stable biological reference material over long periods of time. Each aliquot is characterized by NMR and Biosorter measurements, which provide life stage distribution data of *C. elegans*. A set of quality control measures is being developed for acceptance or rejection of the batch. Batches will be included in the reference only if the individual metrics fall within the standards set for the batch.



This reference material is being used as an external pooled control in a large study funded by the NIH Metabolomics Common Fund to utilize *C. elegans* genetics pathways in unknown compound identification. We will be collecting LC-MS data on hundreds of *C. elegans* natural isolates and mutants, and the reference material will be used to quantify samples collected across machines and over time.

In addition, we are making a large-scale fraction library of the standard reference material. These fractions will be collected using the same chromatography as LC-MS data. NMR 1D and 2D data will be collected on all concentrated fractions, and quantum mechanical chemical shift calculations will be used to match LC-MS and NMR data using methods similar to SUMMIT[1] developed by the Brüschweiler lab.

[1] Bingol, K., Bruschweiler-Li, L., Yu, C., Somogyi, A., Zhang, F., and Bruschweiler, R. (2015) Metabolomics beyond spectroscopic databases: a combined MS/NMR strategy for the rapid identification of new metabolites in complex mixtures, *Anal Chem 87*, 3864-3870.

Supported by the NIH Common Fund 1U2CES030167-01 (ASE)