**BBNet: A Bayesian Bi-Clustering Approach that Identifies Associations Between Metabolites and Gene Expression in Populus Trichocarpa**

Piet Jones, Tim Tschaplinski, Gerald Tuskan, Jay Chen, Wellington Muchero, and
Daniel Jacobson

Oak Ridge National Laboratory

jonespc@ornl.gov

**Abstract**

An important feedstock of interest for bioenergy production is Populus trichocarpa. Using metabolomics and gene expression we can better understand and characterize the phenotypic variances associated with various genotypes of this species. Here we present a bi-clustering method that combines a Dirichlet Process Mixture with XGBoost, to determine associations between variables and features, respectively. The Dirichlet Process Mixture is used to propose a clustering of variables, while XGBoost is used to determine which features are important in discriminating between the respective clusters. Together with Markov Chain Monte Carlo sampling, this approach provides a weighted matrix that can be visualized as a bipartite network. We apply this approach to metabolomics data, sampled from members of a Populus trichocarpa GWAS population arrayed in a common garden, which are treated as our genotypes. The resultant weighted bipartite network shows putative associations between metabolites and genotypes. Similarly, we apply our approach to transcriptome data from the same population, to determine the association between the expression level of genes and the respective genotypes. These two weighted matrices are then combined. Our method provides an interactive visual framework, which can be used to uncover the associations between metabolites and gene expression by means of their putative genotype association. The structure of the network can then be used to explore alternative hypotheses that may improve our understanding of the molecular function and assist in biofuel production.