ClassifyGxT: Probabilistic classification of gene-by-treatment effects in molecular count phenotypes

Authors: William Valdar[1,2,3*], Yuriko Harigaya[1*], Nana Matoba[1,4], Brandon Le[1,4], Jordan Valone[1,4], Jason Stein[1,4,5], Michael Love[1,3*]
[1] Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [2] Lineberger Comprehensive Cancer Center, Chapel Hill, NC, USA. [3] Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.
[4] UNC Neuroscience Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [5] Carolina Institute for Developmental Disabilities, Carrboro, NC, USA.

* These authors contributed equally to this work.

Genetic effects on a phenotype of interest can differ, sometimes markedly, in response to an applied treatment. Such gene-by-treatment interactions (GxT) can be highly consequential in biomedicine and agriculture. A practical approach to identifying GxT signals and gaining insight into molecular mechanisms is mapping quantitative trait loci (QTL) of molecular count phenotypes, such as gene expression and chromatin accessibility, under multiple treatment conditions. Current practice, however, exhibits at least two limitations. First, a typical mapping analysis returns a list of feature-SNP pairs with significant GxT interactions but does not provide a principled way to prioritize GxT interactions of specific types. For example, treatment may have an impact on individuals of a certain genotype only. In other cases, with similarly significant GxT interactions, treatment may affect all individuals but to genotype-specific extents. Formally assigning probabilities to these cases can help prioritize response molecular QTLs for further investigation. A second potential limitation is the frequent assumption of linearity between the phenotype and the genotype after a variance-stabilizing transformation, such as the logarithm, which is routinely applied to molecular count phenotypes. This can lead to nontrivial model misspecification and inaccurate inference. Previous studies have shown that, consistent with the biologically reasonable assumption of common allelic additivity on molecular traits, such as expression, the linear relationship holds in the original count scale, but not in the transformed scale (PMID: 29021289, 29073327).

We describe a downstream method, ClassifyGxT, for categorizing response molecular QTLs. Our method uses Bayesian model selection and assigns posterior probabilities to different types of GxT interactions for a given feature-SNP pair. It also uses a non-linear model to capture the inherent relationship between genotype and phenotype in molecular phenotypes. Our method provides an intuitive way to report the evidence for different types of GxT interaction across a set of feature-SNP pairs.