

Random Forests for eQTL Analysis: A Performance Comparison

Jacob J. Michaelson and Andreas Beyer

Biotechnology Center, TU Dresden, Germany

In recent years quantitative trait locus (QTL) methods have been combined with microarrays, using gene expression as a quantitative trait for genetic linkage analysis. Finding genetic loci significantly linked to the expression of a gene can help to identify regulators of the expressed gene. Traditional QTL methods used to find expression quantitative trait loci (eQTL) typically apply a univariate model to each genotyped locus in order to assess linkage to the quantitative trait. This univariate approach makes it difficult to uncover the interacting genes in the upstream regulatory pathway of the target. As has been previously suggested[1], in this work we view the eQTL problem as one of multivariate model selection: finding the genotyped loci which together best explain the variability of target gene expression in a population. We performed regression with Random Forests using the genotyped loci as predictor variables and the gene expression as the response. Measures of variable importance returned by Random Forests were used in locating eQTL. To assess whether this was a valid approach to eQTL, we determined eQTL for transcriptional targets of several canonical regulatory pathways using both Random Forests and several conventional QTL methods provided by the R `qt1` package. Gene expressions derived from several tissues of recombinant inbred mouse strains were used, and each eQTL method was evaluated for its ability to recapitulate known members of the canonical regulatory pathways. The results of our work demonstrate the biological validity and performance advantages of using Random Forests as a tool for finding eQTL.

References

- [1] K. W. Broman and T. P. Speed, “A model selection approach for the identification of quantitative trait loci in experimental crosses,” *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, vol. 64, no. 4, pp. 641–656, 2002.