A flowering-time gene network model for association analysis in Arabidopsis thaliana

Karin Klotzbücher¹, Yasushi Kobayashi², Nino Shervashidze²,³, Karsten M. Borgwardt³ and Detlef Weigel²

¹Center for Bioinformatics Tübingen (ZBIT), Germany
²Max Planck Institute for Developmental Biology, Tübingen, Germany
³Max Planck Institute for Biological Cybernetics, Tübingen, Germany

In our project we want to determine a set of single nucleotide polymorphisms (SNPs), which have a major effect on the flowering time of Arabidopsis thaliana. Instead of performing a genome-wide association study on all SNPs in the genome of Arabidopsis thaliana, we examine the subset of SNPs from the flowering-time gene network model. We are interested in how the results of the association study vary when using only the ascertained subset of SNPs from the flowering network model, and when additionally using the information encoded by the structure of the network model. The network model is compiled from the literature by manual analysis and contains genes which have been found to affect the flowering time of Arabidopsis thaliana [Far+08; KW07]. The genes in this model are annotated with the SNPs that are located in these genes, or in near proximity to them. In a baseline comparison between the subset of SNPs from the graph and the set of all SNPs, we omit the structural information and calculate the correlation between the individual SNPs and the flowering time phenotype by use of statistical methods. Through this we can determine the subset of SNPs with the highest correlation to the flowering time. In order to further refine this subset, we include the additional information provided by the network structure by conducting a graph-based feature pre-selection. In the further course of this project we want to validate and examine the resulting set of SNPs and their corresponding genes with experimental methods.

References