

GARFIELD

Sandro Morganello, Valentina Iotchkova

October 30, 2017

1 Introduction

GARFIELD is a non-parametric functional enrichment analysis approach described in the paper GARFIELD: GWAS analysis of regulatory or functional information enrichment with LD correction. Briefly, it is a method that leverages GWAS findings with regulatory or functional annotations (primarily from ENCODE and Roadmap epigenomics data) to find features relevant to a phenotype of interest. It performs greedy pruning of GWAS SNPs ($LD\ r^2 > 0.1$) and then annotates them based on functional information overlap. Next, it quantifies Fold Enrichment (FE) at various GWAS significance cutoffs and assesses them by permutation testing, while matching for minor allele frequency, distance to nearest transcription start site and number of LD proxies ($r^2 > 0.8$). Within this framework, GARFIELD accounts for major sources of confounding that current methods do not offer.

We have implemented GARFIELD into a standalone tool using C++ for data pre-processing, enrichment estimation and significance testing and R for visualisation. It provides a way for assessing the enrichment of association analysis signals in 1005 features extracted from ENCODE, GENCODE and Roadmap Epigenomics projects, including genic annotations, chromatin states, histone modifications, DNaseI hypersensitive sites and transcription factor binding sites, among others, in a number of publicly available cell lines.

2 Installation

The package can be installed directly from Bioconductor with

```
source("https://bioconductor.org/biocLite.R")
biocLite("garfield")
```

and loaded as follows

```
library(garfield)
```

For more help, and to see all of the functions in the package use the following command:

```
help(package = garfield)
```