

The identification of cis-regulatory sequence motifs in gene promoters based on SNP information. Additional information and software utilities.

[Korkuc, P.; Walther, D.: The Identification of Cis-Regulatory Sequence Motifs in Gene Promoters Based on SNP Information.

PLANT SYNTHETIC PROMOTERS: METHODS AND PROTOCOLS, S. 31 - 47 (2016)]

Content:

1. R-script for expression difference computation
2. GO-term enrichment analysis
  - 2.1 Fisher Exact Test - C-code and Instructions

Files:

- expressionTest.R
- GO\_term\_general.pl
- fisher\_oneSided.c
- fisher\_routine.c

1. R-script to compute pairwise correlation coefficients of expression data for a set of genes and compare two gene sets for difference of pairwise co-expression. See comments within the R-code for further instructions.

2. GO-term enrichment Perl program

Generic Perl program to compute the enrichment of GO-terms by comparing two gene sets. Over- or underrepresented GO terms will be listed along with their p-values (corrected for multiple testing according to Benjamini & Hochberg)

Invoke with `GO_term_general.pl <file1> <file2>`

where file1 and file2 are lists of gene ids of two gene sets (one serving as the target set, the other as the reference set). Format: One gene\_id per line.

Make sure that program fisher.out (generated by compiling the provided C-code fisher\_oneSided.c") is in the same directory.

Reads GO-terms from the file "GO-termFile.txt" - with multiple GO-terms per genes allowed.

Format:

Gene\_ID[tab]GO-term

Example:

AT1G01010 multicellular organismal development

AT1G01020 sphingolipid metabolic process

AT1G01020 sterol metabolic process

AT1G01030 flower development

AT1G01030 leaf development

AT1G01040 ATP catabolic process

AT1G01040 cytokinesis

AT1G01040 embryonic pattern specification

AT1G01040 flower development

AT1G01040 mRNA cleavage involved in gene silencing by miRNA

AT1G01040 primary miRNA processing

## 2.1 Fisher Exact Test - C-code and Instructions

C-code for performing the Fisher exact test.

Compile with: `cc fisher_oneSided.c -lm -o fisher.out` (make sure `fisher_routine.c` is in the same directory)

Returns the smaller of the two one-sided(!) Fisher exact p-values. If the left-sided p-value is smaller, the p-values are designated by a minus sign (THEY ARE NOT NEGATIVE THOUGH!!)

Example:

```
fisher.out 10 1 2 12
```

```
returns: 1.952e-04
```

For contingency table:  $x_{11}=10$   $x_{12}=1$   $x_{21}=2$   $x_{22}=12$

Alternatively: `fisher.out 1 10 12 2`

returns:  $-1.952e-04$  <<< negative sign to indicate that the left-sided p-value was smaller than the right-handed p-value...the p-value itself is:  $\text{abs}(\text{p-value})!$

For contingency table:  $x_{11}=1$   $x_{12}=10$   $x_{21}=12$   $x_{22}=2$