Brief Introduction to R package gact

Palle Duun Rohde, Izel Fourie Sørensen, & Peter Sørensen

2022-10-31

Contents

The practical is based on the R package `gact` (Rohde et al.2022)). This package provides an infrastructure for working with large-scale genomic association data linked to different types of genomic features.

The most recent version of `gact` can be obtained from github:

```
library(devtools)
devtools::install_github("psoerensen/gact")
```

Load packages used

```
library(gact)
library(qgg)
library(corrplot)
```

Download and install GDT database

The function `gact()` dowload and install the GDT database:

```
# Set working for database
setwd("C://Users//au223366//Dropbox//GitHub//gdtdb")

# Download data bases from repository
GAlist <- gact(version = "t2dm-gact-0.0.1", task = "download")
```

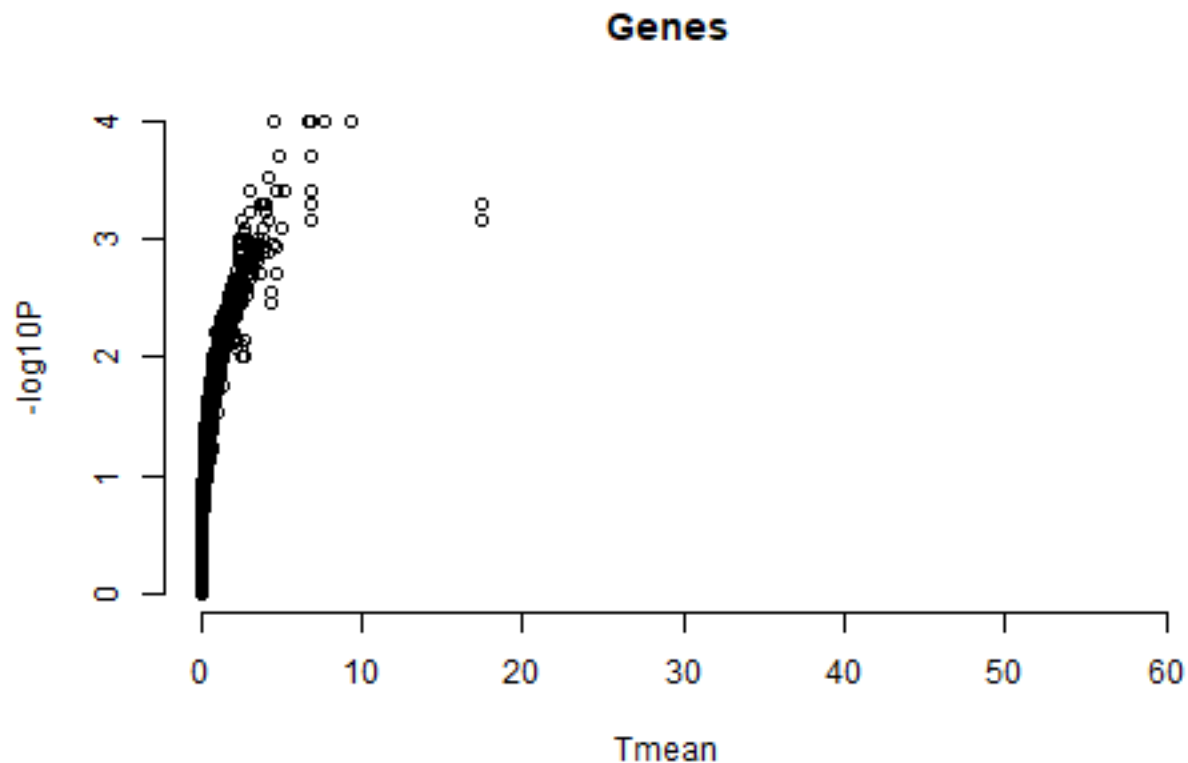## Extract data from GDT database

The function `getStat()` extract data from the database:

```
# Extract data from T2D for genomic feature Markers stat <-
# getStat(GAlist=GAlist, trait='t2d', feature='Markers') head(stat)

# Extract data from T2D for genomic feature Genes
stat <- getStat(GAlist = GAlist, trait = "t2d", feature = "Genes")
head(stat)
```

```
##                   Ensembl Gene ID   Symbol    m     stat       p
## ENSG00000121410 ENSG00000121410     A1BG   58  0.00000 1.0000
## ENSG00000175899 ENSG00000175899      A2M  153 20.78100 0.1477
## ENSG00000256069 ENSG00000256069     A2MP1  64  0.00000 1.0000
## ENSG00000171428 ENSG00000171428     NAT1  135  0.00000 1.0000
## ENSG00000156006 ENSG00000156006     NAT2  102 20.90052 0.0951
## ENSG00000196136 ENSG00000196136 SERPINA3  64  0.00000 1.0000
```

```r
plot(x = stat$stat/stat$m, y = -log10(stat$p), ylab = "-log10P", xlab = "Tmean",
     frame.plot = FALSE, main = "Genes")
```



```r
# Extract data from T2D for genomic feature Gene Ontology (GO) where
# output format is a data frame
stat <- getStat(GAlist = GAlist, trait = "t2d", feature = "GO")
head(stat)
```

```
##                 GO ID    m      stat       p
## GO:0000002 GO:0000002 2165 143.36901 0.3208
## GO:0000012 GO:0000012 1885 107.38110 0.3577
## GO:0000027 GO:0000027 1867  72.77531 0.5031
## GO:0000028 GO:0000028  861  71.43454 0.2242
## GO:0000038 GO:0000038 2397  63.58818 0.6599
## GO:0000045 GO:0000045 8223 749.36205 0.2368
```

```
plot(x = stat$stat/stat$m, y = -log10(stat$p), ylab = "-log10P", xlab = "Tmean",
    frame.plot = FALSE, main = "GO")
```
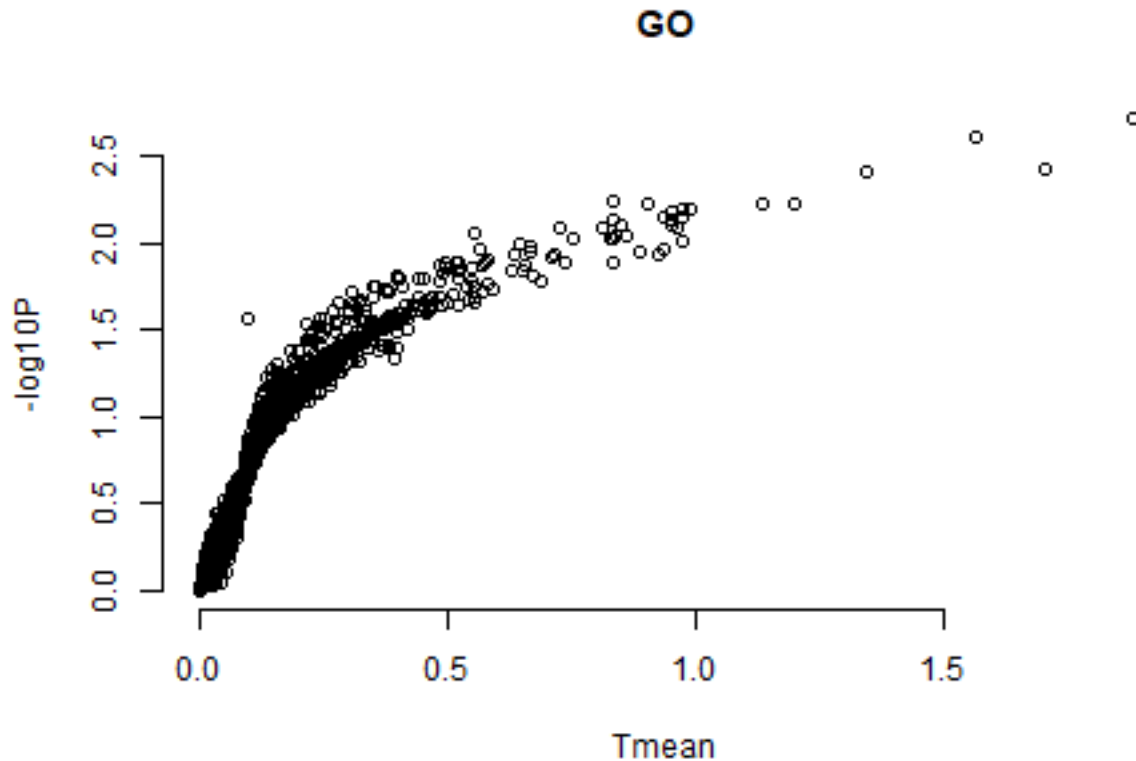


```
# Extract data from T2D for genomic feature Gene Ontology (GO) where
# output format is list
stat <- getStat(GAlist = GAlist, trait = "t2d", feature = "GO", format = "list",
    cls = c("p=1e.04", "p=0.001", "p=0.05", "p=0.1", "p=0.95"))
str(stat)
```

```
## List of 3
##  $ m   : Named int [1:4547] 2165 1885 1867 861 2397 8223 1050 3519 1077 5192 ...
##   ..- attr(*, "names")= chr [1:4547] "GO:0000002" "GO:0000012" "GO:0000027" "GO:0000028" ...
##  $ stat: num [1:4547, 1:5] 0 0 0 17 0 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:4547] "GO:0000002" "GO:0000012" "GO:0000027" "GO:0000028" ...
##   .. ..$ : chr [1:5] "p=1e.04" "p=0.001" "p=0.05" "p=0.1" ...
##  $ p   : num [1:4547, 1:5] 1 1 1 0.0781 1 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:4547] "GO:0000002" "GO:0000012" "GO:0000027" "GO:0000028" ...
##   .. ..$ : chr [1:5] "p=1e.04" "p=0.001" "p=0.05" "p=0.1" ...
```

```
# Plot results for genomic feature Gene Ontology (GO)
colbar <- colorRampPalette(c("#FFFFFF", "#D1E5F0", "#92C5DE", "#4393C3",
    "#2166AC", "#053061"))
```

```
corrplot(-log10(stat$p[1:10, ]), is.corr = FALSE, tl.cex = 0.7, tl.srt = 45,
    col = colbar(6), cl.pos = "n", mar = c(1, 1, 1, 1))
```
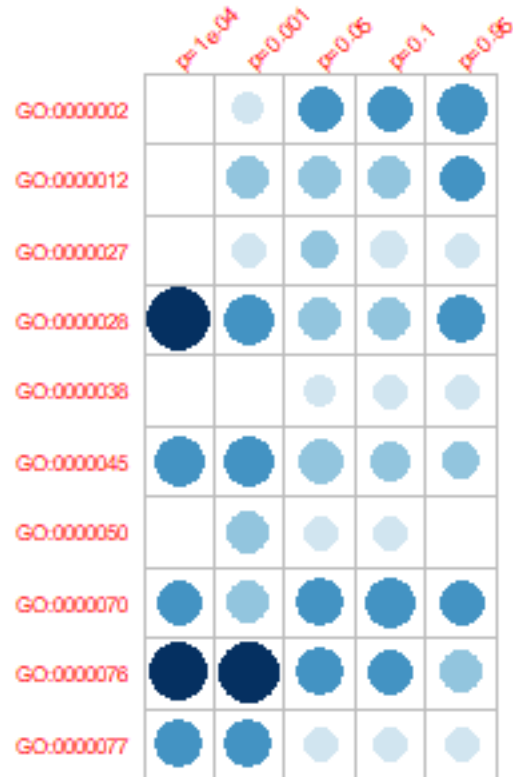

Extract and write data from GDT database

The function `writeStat()` extract and write data from the database:

```
writeStat(GAlist = GAlist, feature = "GO", trait = "t2d", file.csv = "go_t2dm_gcta.csv")
writeStat(GAlist = GAlist, feature = "Pathways", trait = "t2d", file.csv = "pathways_t2dm_gcta.csv")
writeStat(GAlist = GAlist, feature = "ProteinComplexes", trait = "t2d",
    file.csv = "proteincomplexes_t2dm_gcta.csv")
writeStat(GAlist = GAlist, feature = "ChemicalComplexes", trait = "t2d",
    file.csv = "chemicalcomplexes_t2dm_gcta.csv")
writeStat(GAlist = GAlist, feature = "Genes", trait = "t2d", file.csv = "genes_t2dm_gcta.csv")
```