Report\_episcan\_package

Federico Melograna

9/12/2020

## R Markdown

In this markdown we want to show the 2 possible tests of the Episcan package, the 1step, quick and less computational intensive. If the phenotype is binary it is done with the Epiblaster function and the second, with quantitative phenotype, with the epiHSIC1geno function.

# Hilbert-Schmitt Independence Criterion (HSIC)  
  
# Reference: https://amber0309.github.io/2017/01/13/hsic/  
  
# CODE  
  
# Reference: https://cran.r-project.org/web/packages/episcan/index.html  
  
# install.packages("episcan")  
  
library(episcan)  
  
# help(episcan)  
# simulate some data C:\Users\federico\Desktop\Work\Epistasis\Episcan\_epiblaster\Esperimento\_episcan  
  
# the right explanaition is this  
# I have looked into the package and my code, I think that I have found some interesting answers about the https://cran.r-project.org/web/packages/episcan/episcan.pdf package. The episcan, the main function in the lot, will select the function epiblaster or epiHSIC1 depending if the phenotype is respectively binary ("case-control") or quantitative ("quantitative"), for the first case, the computation is basically a Pearson correlation between the SNPs and the second case is basically an HSIC https://amber0309.github.io/2017/01/13/hsic/ computation. As I have seen in the paper, this is only the first step of the computation, so I have to ask for the second one.   
  
  
setwd("C:/Users/federico/Desktop/Work/Epistasis/Episcan\_epiblaster/Esperimento\_episcan")  
data\_experimental <- read.csv("C:/Users/federico/Desktop/Work/Epistasis/Episcan\_epiblaster/Esperimento\_episcan/100\_100.csv", row.names=1)  
pheno = data\_experimental$Y  
pheno[50:100] = 0   
geni = as.matrix(data\_experimental[,-ncol(data\_experimental)])  
print(head(geni[,1:10]))

## rs9729550 rs1815606 rs7515488 rs11260562 rs6603785 rs11804831 rs6603788  
## 0 0 0 0 0 0 0 0  
## 1 2 2 0 0 0 0 0  
## 2 0 0 0 0 0 0 0  
## 3 1 1 1 0 0 1 0  
## 4 0 0 0 0 0 0 0  
## 5 2 2 2 2 1 2 1  
## rs1739855 rs12103 rs12142199  
## 0 0 0 0  
## 1 0 0 0  
## 2 0 0 0  
## 3 1 1 1  
## 4 0 0 0  
## 5 2 2 2

str(geni)

## int [1:100, 1:101] 0 2 0 1 0 2 0 0 1 1 ...  
## - attr(\*, "dimnames")=List of 2  
## ..$ : chr [1:100] "0" "1" "2" "3" ...  
## ..$ : chr [1:101] "rs9729550" "rs1815606" "rs7515488" "rs11260562" ...

set.seed(123)  
  
#1 step with case control phenotype  
  
episcan(geno1 = geni,   
 geno2 = NULL,   
 pheno = pheno,   
 phetype = "case-control",  
 outfile = "episcan\_1step\_cc",   
 suffix = ".txt",   
 zpthres = 0.9,   
 chunksize = 10,   
 scale = TRUE) #it scales the geno and the pheno also if it is quantitative

## p-value threshold of Z test for output: 0.9   
## set chunksize: 10   
## [1] "episcan starts:"  
## [1] "Fri Dec 11 16:38:47 2020"  
## [1] "1 chunk loop: Fri Dec 11 16:38:47 2020"  
## [1] "2 chunk loop: Fri Dec 11 16:38:48 2020"  
## [1] "3 chunk loop: Fri Dec 11 16:38:49 2020"  
## [1] "4 chunk loop: Fri Dec 11 16:38:50 2020"  
## [1] "5 chunk loop: Fri Dec 11 16:38:50 2020"  
## [1] "6 chunk loop: Fri Dec 11 16:38:51 2020"  
## [1] "7 chunk loop: Fri Dec 11 16:38:51 2020"  
## [1] "8 chunk loop: Fri Dec 11 16:38:51 2020"  
## [1] "9 chunk loop: Fri Dec 11 16:38:52 2020"  
## [1] "10 chunk loop: Fri Dec 11 16:38:52 2020"  
## [1] "11 chunk loop: Fri Dec 11 16:38:52 2020"  
## [1] "epiblaster calculation is over!"  
## [1] "Fri Dec 11 16:38:52 2020"

# take a look at the result  
  
res <- read.table("episcan\_1step\_cc.txt",   
 header = TRUE,   
 stringsAsFactors = FALSE)  
head(res)

## SNP1 SNP2 Zscore ZP  
## 1 rs9729550 rs1815606 -0.3495935 0.72664382  
## 2 rs9729550 rs7515488 -1.8109037 0.07015576  
## 3 rs1815606 rs7515488 -1.9949372 0.04604973  
## 4 rs9729550 rs11260562 -1.4324657 0.15201058  
## 5 rs1815606 rs11260562 -1.6054104 0.10840351  
## 6 rs7515488 rs11260562 0.5025510 0.61528000

#1 step with quantitative phenotype  
#it is equivalent to the episcan function up, with the quantitative flag and scale = True  
pheno = unlist(scale(pheno))[,1]  
#genes are scaled  
epiHSIC1geno(geno = scale(geni), pheno = pheno, chunk = 1000, zpthres = 0.9, #1e-05,  
 outfile = "episcan\_1stepQuant", suffix = ".txt")

## [1] "1 chunk loop: Fri Dec 11 16:38:52 2020"  
## [1] "epiHSIC calculation is over!"  
## [1] "Fri Dec 11 16:38:52 2020"

#pheno has to be normalized scale(pheno)  
  
  
res1 <- read.table("episcan\_1stepQuant.txt",  
 header = TRUE,  
 stringsAsFactors = FALSE)  
  
head(res1)

## SNP1 SNP2 Zscore ZP  
## 1 rs9729550 rs1815606 -1.317431 0.187694315  
## 2 rs9729550 rs7515488 -2.321099 0.020281475  
## 3 rs1815606 rs7515488 -2.479215 0.013167180  
## 4 rs9729550 rs11260562 -1.763147 0.077875659  
## 5 rs1815606 rs11260562 -1.898231 0.057665603  
## 6 rs9729550 rs6603785 -2.731613 0.006302507