**Supplementary file 2: DEA between placebo and vaccinated samples at day 0**

Import the data tables containing counts data and features of each sample in the R environment

desc\_tab<-read.csv2("Descriptive\_Table.csv")  
counts\_tab<-read.csv2("Counts\_Table.csv", row.names=1)

Selection of data from day 0 to test whether there is difference between samples from placebo and vaccinated volunteers before vaccination

D0<- desc\_tab$Group.Day %in% "0"  
desc\_D0<- desc\_tab[D0,]  
counts\_D0<- counts\_tab[, D0]

Creation of a DGElist object (“y”) for storing read counts, gene names and information from samples or libraries. Rows correspond to counts and columns to samples. “counts” is a numeric matrix of counts, each row corresponds to a gene and each column to a sample. Within the DGElist, “Samples” is a data frame which stores in each row a sample and in columns group, library size and the normalization factors. The experimental design is defined with the “group” command. Data were filtered to exclude the unexpressed and lowly expressed genes applying a filter of 1 count per million (cpm), which corresponds to 6-7 counts, in at least 10 libraries.

library(edgeR)

## Loading required package: limma

y<- DGEList(counts = counts\_D0, genes = rownames(counts\_D0), group = desc\_D0$Group.Day)  
keep<-rowSums(cpm(y)>1)>=10  
y.1<-y[keep,]  
dim(y.1)

## [1] 12762 64

Normalization of raw library sizes by calculating TMM normalization

y.1<- calcNormFactors(y.1)

Creation of a design matrix to define the experimental design and integrate the sequencing library preparation batch as a variable in addition to the treatment (placebo or vaccinated). The design matrix is used during common dispersion estimation. In case of multiple factors experiment edgeR uses the Cox-Reid profile-adjusted likelihood (CR) method.

batch<- factor(desc\_D0$Library.Batch)  
design<- model.matrix(~batch+desc\_D0$Treatment)  
y.2<- estimateDisp(y.1, design)

Fitting general linear models to each feature

qlfit\_treatment<- glmQLFit(y.2, design)

Application of a quasi-likelihood negative binomial generalized log-linear model to count data. The contrast is indicated with the coefficient and refers to the design matrix. The decide test is used to identify the significant differentially expressed genes, and the “toptags” command orders the list by FDR.

DEA\_Placebo\_Vaccinated<- glmQLFTest(qlfit\_treatment, coef=6)  
summary(decideTestsDGE(DEA\_Placebo\_Vaccinated))

## desc\_D0$TreatmentV  
## Down 0  
## NotSig 12762  
## Up 0

DEA\_Placebo\_Vaccinated<- topTags(DEA\_Placebo\_Vaccinated, n=Inf)  
write.csv2(DEA\_Placebo\_Vaccinated, "DEA\_Placebo\_Vaccinated.csv")

This analysis was conducted on:

sessionInfo()

## R version 3.5.1 (2018-07-02)  
## Platform: x86\_64-w64-mingw32/x64 (64-bit)  
## Running under: Windows 7 x64 (build 7601) Service Pack 1  
##   
## Matrix products: default  
##   
## locale:  
## [1] LC\_COLLATE=Italian\_Italy.1252 LC\_CTYPE=Italian\_Italy.1252   
## [3] LC\_MONETARY=Italian\_Italy.1252 LC\_NUMERIC=C   
## [5] LC\_TIME=Italian\_Italy.1252   
##   
## attached base packages:  
## [1] stats graphics grDevices utils datasets methods base   
##   
## other attached packages:  
## [1] edgeR\_3.24.1 limma\_3.38.3  
##   
## loaded via a namespace (and not attached):  
## [1] locfit\_1.5-9.1 Rcpp\_1.0.0 lattice\_0.20-38 digest\_0.6.18   
## [5] grid\_3.5.1 magrittr\_1.5 evaluate\_0.12 stringi\_1.2.4   
## [9] rmarkdown\_1.11 splines\_3.5.1 tools\_3.5.1 stringr\_1.3.1   
## [13] xfun\_0.4 yaml\_2.2.0 compiler\_3.5.1 htmltools\_0.3.6  
## [17] knitr\_1.21