**Supplementary file 3: differential gene expression analysis of rVSVΔG-ZEBOV-GP vaccinated volunteers**

Import of data tables containing counts data and samples’ metadata 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 vaccinated volunteers, excluding the placebo volunteers

V<- desc\_tab**$**Treatment **%in%** "V"  
desc\_V<- desc\_tab[V,]  
counts\_V<- counts\_tab[, V]

Creation of a DGElist object (“x”) 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

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

## [1] 14163 303

Normalization of raw library sizes by calculating TMM normalization

x.1<- **calcNormFactors**(x.1)

Creation of a design matrix to define the experimental design and integrate the sequencing library prepation batch as a variable in addition to the day of sampling (day 0, day 1, day 2, day 3, day 7, day 14, day 21, day 28 and day 35 after vaccination). The design matrix is used during common dispersion estimation. Including “batch” variable in design matrix allows to take into account this unintentional aspect of the experimental design. In this way different time points are compared only within each batch. In case of multiple factors experiment edgeR uses the Cox-Reid profile-adjusted likelihood (CR) method.

batch<- **factor**(desc\_V**$**Library.Batch)  
desc\_V**$**Group.Day<- **relevel**(desc\_V**$**Group.Day, ref = "0")  
design<- **model.matrix**(**~**batch**+**desc\_V**$**Group.Day)  
x.2<- **estimateDisp**(x.1, design)

Fitting general linear models to each feature

qlfit\_new<- **glmQLFit**(x.2, design)

Application of a quasi-likelihood negative binomial generalized log-linear model to count data. The contrast is indicated with coefficient and refers to the design matrix. During the differential expression analysis, a threshold of 1.2 on log2 fold-change is introduced to reduce the number of disregulated genes. Only genes that exceed the threshold value will be declared statistically significant. To do that, we use glmTreat function which tests differential expression relative to a threshold: instead of testing for genes which have log-fold-changes different from zero, it tests whether the log2-fold-change is greater than lfc in absolute value. glmTreat detects whether the argument is produced by glmFit or glmQLFit and it can be used in both pipelines. The decide test is used to identify the significant differentially expressed genes, and the “toptags” command orders the list by FDR.

*#Day1 vs Day 0*  
DEA\_Day1<- **glmTreat**(qlfit\_new, coef=6, lfc=**log2**(1.2))   
**summary**(**decideTestsDGE**(DEA\_Day1))

## desc\_V$Group.Day1  
## Down 2996  
## NotSig 8694  
## Up 2473

DEA\_Day1<- **topTags**(DEA\_Day1, n=Inf)  
**write.csv2**(DEA\_Day1, "DEA\_Day1.csv")  
  
*#Day2 vs Day 0*  
DEA\_Day2<- **glmTreat**(qlfit\_new, coef=9, lfc=**log2**(1.2))   
**summary**(**decideTestsDGE**(DEA\_Day2))

## desc\_V$Group.Day2  
## Down 630  
## NotSig 12806  
## Up 727

DEA\_Day2<- **topTags**(DEA\_Day2, n=Inf)  
**write.csv2**(DEA\_Day2, "DEA\_Day2.csv")  
  
*#Day3 vs Day 0*  
DEA\_Day3<- **glmTreat**(qlfit\_new, coef=12, lfc=**log2**(1.2))   
**summary**(**decideTestsDGE**(DEA\_Day3))

## desc\_V$Group.Day3  
## Down 272  
## NotSig 13596  
## Up 295

DEA\_Day3<- **topTags**(DEA\_Day3, n=Inf)  
**write.csv2**(DEA\_Day3, "DEA\_Day3.csv")  
  
*#Day7 vs Day 0*  
DEA\_Day7<- **glmTreat**(qlfit\_new, coef=15, lfc=**log2**(1.2))   
**summary**(**decideTestsDGE**(DEA\_Day7))

## desc\_V$Group.Day7  
## Down 183  
## NotSig 13691  
## Up 289

DEA\_Day7<- **topTags**(DEA\_Day7, n=Inf)  
**write.csv2**(DEA\_Day7, "DEA\_Day7.csv")  
  
*#Day14 vs Day 0*  
DEA\_Day14<- **glmTreat**(qlfit\_new, coef=7, lfc=**log2**(1.2))   
**summary**(**decideTestsDGE**(DEA\_Day14))

## desc\_V$Group.Day14  
## Down 3  
## NotSig 14136  
## Up 24

DEA\_Day14<- **topTags**(DEA\_Day14, n=Inf)  
**write.csv2**(DEA\_Day14, "DEA\_Day14.csv")  
  
*#Day21 vs Day 0*  
DEA\_Day21<- **glmTreat**(qlfit\_new, coef=10, lfc=**log2**(1.2))   
**summary**(**decideTestsDGE**(DEA\_Day21))

## desc\_V$Group.Day21  
## Down 0  
## NotSig 14163  
## Up 0

DEA\_Day21<- **topTags**(DEA\_Day21, n=Inf)  
**write.csv2**(DEA\_Day21, "DEA\_Day21.csv")  
  
*#Day28 vs Day 0*  
DEA\_Day28<- **glmTreat**(qlfit\_new, coef=11, lfc=**log2**(1.2))   
**summary**(**decideTestsDGE**(DEA\_Day28))

## desc\_V$Group.Day28  
## Down 0  
## NotSig 14163  
## Up 0

DEA\_Day28<- **topTags**(DEA\_Day28, n=Inf)  
**write.csv2**(DEA\_Day28, "DEA\_Day28.csv")  
  
*#Day35 vs Day 0*  
DEA\_Day35<- **glmTreat**(qlfit\_new, coef=14, lfc=**log2**(1.2))   
**summary**(de<- **decideTestsDGE**(DEA\_Day35))

## desc\_V$Group.Day35  
## Down 0  
## NotSig 14159  
## Up 4

DEA\_Day35<- **topTags**(DEA\_Day35, n=Inf)  
**write.csv2**(DEA\_Day35, "DEA\_Day35.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