Pedro\_Salvador\_ADO\_PEC1

Pedro Salvador Escribano

7/4/2020

Table of Contents

<https://github.com/pesales/PEC1_Pedro_Salvador/>

# Abstract.

Se ha observado una función inmunitaria alterada en los astronautas que regresan de misiones espaciales. En este estudio se ha encontrado como posibles causas, la alteración de rutas relacionadas con la señalización del interferón gamma, la degranulación de neutrófilos o la regulación transcripcional de p53.

# Objetivos: Que se pretende con este estudio.

En este estudio se analiza la causa de la alteración de la función inmunitaria sufrida por los astronautas tras su regreso a la tierra. Para ello, se utiliza un modelo de gravedad alterada simulada con el que se simula la diferenciación de células inumitarias bajo estas condiciones atípicas. Se analiza las diferencias en la expresión génica de células T, tras su diferenciación en condiciones de gravedad normal (1g) o gravedad variante (vg).

# Materiales y Métodos.

## Aislamiento de células T, cultivo y activación:

Se aislaron leucocitos de sangre periférica de 3 donantes humanos mediante centrifugación de gradiente de densidad con Ficoll (Red Cross, Zurich, Suiza) y las células T fueron purifecadas usando columnas de alta afinidad para células T CD3 (R&D Systems, Minneapolis, MN). Las células fueron resuspendidas en RPMI-1640 con un 10% de suero bovino fetal a 3-8 millones de células/ml. Una porción de estas células fue activada con 5g/ml de concavalina A (Sigma, St. Louis, MO) y 4g/ml de antucuerpo anti-CD28 (PharMingen, San Diego, CA) e incubada a 37ºC durante 4 horas a 1g o en una Random positioning machine (RPM) (Dutch Space, Leiden, Países Bajos) rotando a 60º/s con controles inactivados.

## Extracción de RNA y lavado:

Se aisló el RNA usando isotiocianato de guanidinio y se lavó con columnas RNeasy® MinElute (Qiagen, Valencia, CA). La concentración y pureza de ARN se determinaron midiendo la absorbancia a 260 y 280 nm con un GeneQuant (Pfizer, Nueva York, NY); Se corrieron 0,3 g de ARN total en un gel desnaturalizante al 1%, o se cargaron 100 ng de ARN total en el bioanalizador 2100 (Agilent, Palo Alto, CA) para verificar la integridad del ARN.

## Análisis por microarray:

Se utilizó el Human Genome Focus Array (Affymetrix, Santa Clara, CA), compuesto por 8,796 conjuntos de sondas correspondientes a genes bien anotados de la base de datos RefSeq de NCBI; 1,8 g de ARN total fueron alicuotados para lasíntesis de para la síntesis de ADNc usando cebador oligo-(dT) purificado por HPLC (Affymetrix) y transcriptasa inversa 200 U SuperScript II (Invitrogen, Carlsbad, CA). Diez microlitros del ADNc se dividieron en alícuotas para la transcripción in vitro y el marcado de biotina usando BioArrayTM high yieldTM RNA transcription labeling kit (Enzo, New York, NY).. La reacción se incubó durante 10h a 37°C. Después de la fragmentación de cRNA, las muestras fueron se llevaron al Gladstone Institute Genomics Core. Diez microgramos del cRNA marcado se hibridaron a una concentración final de 0,05 g/μl en arrays Focus a 45°C durante 16 h. Los arrays se tiñeron usando Fluidics Station 400 (Affymetrix) y se escanearon usando un escáner GeneArray® (Affymetrix).

## Procedimiento de análisis de los datos:

Se siguió el procedimiento general de análisis de datos porcedentes de experimentos por microarray:

### Identificar que grupos hay y a qué grupo pertenece cada muestra.

Los archivos CEL de las muestras con número de acceso GSM24817, GSM29564, GSM29565, GSM29566, GSM29567, GSM29568, GSM29569, GSM29570, GSM29571, GSM29572, GSM29573, y GSM29574, con número de acceso de serie GSE170 se obtuvieron de <http://www.ncbi.nlm.nih.gov/geo/browse/?view=series>. Se identificó el grupo al que pertenecía cada una de las muestras y se creó el archivo targets.csv. Los archivos CEl y CSV fueron cargados dentro de la carpeta data.

### Control de calidad de los datos crudos

Los datos fueron cargados en la variable rawData y clasificados según el archivo targets en los diferentes grupos. El control de calidad de estos datos se almacenó en la carpeta “QCDir.Raw”, dentro de la carpeta de resultados. Dentro de esta carpeta se generó el archivo “index.html” que contiene la información del control de calidad. Se observaron discrepancias entre los datos que justificaron la posterior normalización de los datos. Se visualizó mediente un boxplot para la variabilidad de la intensidad de la señal entre muestras normalizadas. También se buscaron relaciones entre las muestras con un análisis de componentes principales en el que, a priori, no se detectaban relaciones fuertes entre las mismas.

### Normalización

Se procedió a la normalización de los datos en la variable eset\_rma.

### Control de calidad de los datos normalizados

El control de calidad de estos datos se almacenó en la carpeta “QCDir.Norm”, dentro de la carpeta de resultados. Dentro de esta carpeta se generó el archivo “index.html” que contiene la información del control de calidad. Tras la normalización de los datos, las diferencias entre arrays habían sido mitigadas. La variabilidad de la intensidad de la señal entre muestras normalizadas se visualizó mediente un boxplot. También se buscaron relaciones entre las muestras con un análisis de componentes principales en el que, a priori, no se detectaban relaciones fuertes entre las mismas.

### Identificación de genes diferencialmente expresados

Partiendo de los datos normalizados, se filtraron los genes menos variables y se guardaron en la variable eset\_filtered. Se definió la matriz de diseño y la matriz de contrastes. Se obtuvo la lista de genes que se expresan de manera diferente cuando se activan en función de si la gravedad es constante o variable (topTab\_VGVSGTREATED) y la lista de genes que se expresan de manera diferente en función de si la gravedad es constante o variable antes de activarse (topTab\_VGVSGNONTREATED).

### Anotación de los resultados

La anotación de los genes se hizo utilizando el paquete de anotación hgfocus.db, correspondiente al micorarray utilizado. Se obtuvo a lista anotada de genes que se expresan de manera diferente cuando se activan en función de si la gravedad es constante o variable (topAnnotated\_VGVSGTREATED) y la lista anotada de genes que se expresan de manera diferente en función de si la gravedad es constante o variable antes de activarse (topAnnotated\_VGVSGNONTREATED).

### Comparación entre distintas comparaciones (si hay más de una comparación, ver que genes han sido seleccionados en más de una comparación)

Se representó mediante un diagrama de Venn y mediante volcano Plots, los genes que estaban representados en cada una de las comparaciones. También se realizaron Heatmaps para visualizar el agrupamiento de las muestras según estos resultados.

### Análisis de significación biológica (“Gene Enrichment Analysis”)

Se agrupó los genes diferencialmente expresados dentro de funciones y rutas moleculares conocidas y se representó las principales de estas funciones que se encontraban alteradas en cada una de las comparaciones(ReactomePA.Results.VGVSGTREATED y ReactomePA.Results.VGVSGNONTREATED).

# Resultados.

Como resultado de este análisis se ha obtenido listas de genes y funciones diferencialmente expresadas en cada una de las comparaciones. Una lista detallada de estos resultados puede ser consultada en el último punto del apéndice, resumen de resultados.

# Discusión.

Los resultados aportados por este estudio deben ser interpretados con cautela, pues únicamente se ha utilizado un tipo celular para este experimento. Sería necesario completar este estudio con los diferentes tipos celulares pertenecientes al sistema inmunitario para obtener una visión global del estado del mismo en el individuo en condiciones de gravedad alterada.

# Apéndice (Código utilizado para el análisis).

## Cargado de datos y consideraciones previas:

Crear un directorio donde se almacenará todos los datos del análisis, incluidos datos brutos, intermedios y resultados:

> setwd("C:/Users/peri8/Google Drive/Master UOC/Análisis de datos ómicos/PEC1/Estudio\_gravedad")  
> dir.create("data")  
> dir.create("results")

Cargar el archivo que describe los datos:

> # Tras compiar los datos .CEL y target (que contienen la información de grupos y covariables de los diferentes archivos) se carga el archivo targets:  
> targets <- read.csv2("./data/targets.csv", header = TRUE, sep = ";")   
> knitr::kable(  
+ targets, booktabs = TRUE,  
+ caption = 'Content of the targets file used for the current analysis')

Content of the targets file used for the current analysis

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| FileName | Group | Gravity | Activation | ShortName |
| GSM24817.CEL | 1G.0H | 1G | 0H | 1G.0H.2 |
| GSM29567.CEL | 1G.0H | 1G | 0H | 1G.0H.4 |
| GSM29571.CEL | 1G.0H | 1G | 0H | 1G.0H.5 |
| GSM29564.CEL | 1G.4H | 1G | 4H | 1G.4H.2 |
| GSM29568.CEL | 1G.4H | 1G | 4H | 1G.4H.4 |
| GSM29572.CEL | 1G.4H | 1G | 4H | 1G.4H.5 |
| GSM29565.CEL | VG.0H | VG | 0H | VG.0H.2 |
| GSM29569.CEL | VG.0H | VG | 0H | VG.0H.4 |
| GSM29573.CEL | VG.0H | VG | 0H | VG.0H.5 |
| GSM29566.CEL | VG.4H | VG | 4H | VG.4H.2 |
| GSM29570.CEL | VG.4H | VG | 4H | VG.4H.4 |
| GSM29574.CEL | VG.4H | VG | 4H | VG.4H.5 |

## Instalación previa de paquetes requeridos:

> if (!requireNamespace("BiocManager", quietly = TRUE))  
+ install.packages("BiocManager")  
> BiocManager::install()

> install.packages("knitr")  
> install.packages("colorspace")  
> install.packages("gplots")  
> install.packages("ggplot2")  
> install.packages("ggrepel")  
> install.packages("htmlTable")  
> install.packages("prettydoc")  
> install.packages("devtools")  
> install.packages("BiocManager")  
> BiocManager::install("oligo")  
> BiocManager::install("pd.mogene.2.1.st")  
> BiocManager::install("arrayQualityMetrics")  
> BiocManager::install("pvca")  
> # NOT NEEDED UNTIL ANALYSES ARE PERFORMED  
> BiocManager::install("limma")  
> BiocManager::install("genefilter")  
> BiocManager::install("hgfocus.db")  
> BiocManager::install("annotate")  
> BiocManager::install("org.Mm.eg.db")  
> BiocManager::install("ReactomePA")  
> BiocManager::install("reactome.db")  
> BiocManager::install("org.Mm.eg.db")  
> BiocManager::install("org.Mm.egPATH")

## Lectura de archivos CEL:

> library(oligo)  
> celFiles <- list.celfiles("./data", full.names = TRUE)  
> library(Biobase)  
> my.targets <-read.AnnotatedDataFrame(file.path("./data","targets.csv"),   
+ header = TRUE, row.names = 1,   
+ sep=";")   
> rawData <- read.celfiles(celFiles, phenoData = my.targets)

> my.targets@data$ShortName->rownames(pData(rawData))  
> colnames(rawData) <-rownames(pData(rawData))   
>   
> head(rawData)

ExpressionFeatureSet (storageMode: lockedEnvironment)  
assayData: 1 features, 12 samples   
 element names: exprs   
protocolData  
 rowNames: 1G.0H.2 1G.0H.4 ... VG.4H.5 (12 total)  
 varLabels: exprs dates  
 varMetadata: labelDescription channel  
phenoData  
 rowNames: 1G.0H.2 1G.0H.4 ... VG.4H.5 (12 total)  
 varLabels: Group Gravity Activation ShortName  
 varMetadata: labelDescription channel  
featureData: none  
experimentData: use 'experimentData(object)'  
Annotation: pd.hg.focus

## Control de calidad de datos brutos:

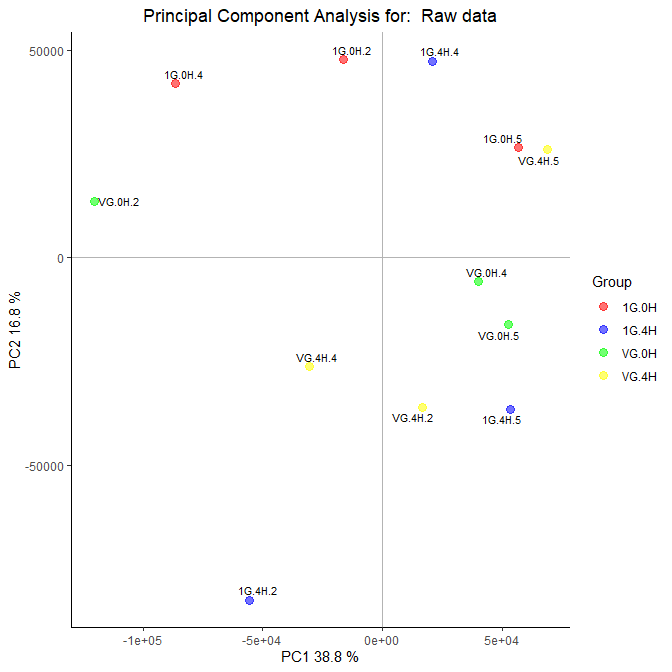
> library(arrayQualityMetrics)  
> arrayQualityMetrics(rawData, outdir = file.path("./results", "QCDir.Raw"), force=TRUE)

Tras esto se creará una carpeta “QCDir.Raw”, dentro de la carpeta de resultados. Dentro de esta carpeta hay que buscar un archivo“index.html” que contiene la información del control de calidad. Si las muestras no tienen mas de una marca, son aptas para continuar con el análisis y los problemas que presentan son pequeños. En caso contrario hay que proceder a la normalización de los datos.

Mediante un PCA se visualizan las relaciones enntre muestras:

> library(ggplot2)  
> library(ggrepel)  
> plotPCA3 <- function (datos, labels, factor, title, scale,colores, size = 1.5, glineas = 0.25) {  
+ data <- prcomp(t(datos),scale=scale)  
+ # plot adjustments  
+ dataDf <- data.frame(data$x)  
+ Group <- factor  
+ loads <- round(data$sdev^2/sum(data$sdev^2)\*100,1)  
+ # main plot  
+ p1 <- ggplot(dataDf,aes(x=PC1, y=PC2)) +  
+ theme\_classic() +  
+ geom\_hline(yintercept = 0, color = "gray70") +  
+ geom\_vline(xintercept = 0, color = "gray70") +  
+ geom\_point(aes(color = Group), alpha = 0.55, size = 3) +  
+ coord\_cartesian(xlim = c(min(data$x[,1])-5,max(data$x[,1])+5)) +  
+ scale\_fill\_discrete(name = "Group")  
+ # avoiding labels superposition  
+ p1 + geom\_text\_repel(aes(y = PC2 + 0.25, label = labels),segment.size = 0.25, size = size) +   
+ labs(x = c(paste("PC1",loads[1],"%")),y=c(paste("PC2",loads[2],"%"))) +   
+ ggtitle(paste("Principal Component Analysis for: ",title,sep=" "))+   
+ theme(plot.title = element\_text(hjust = 0.5)) +  
+ scale\_color\_manual(values=colores)  
+ }

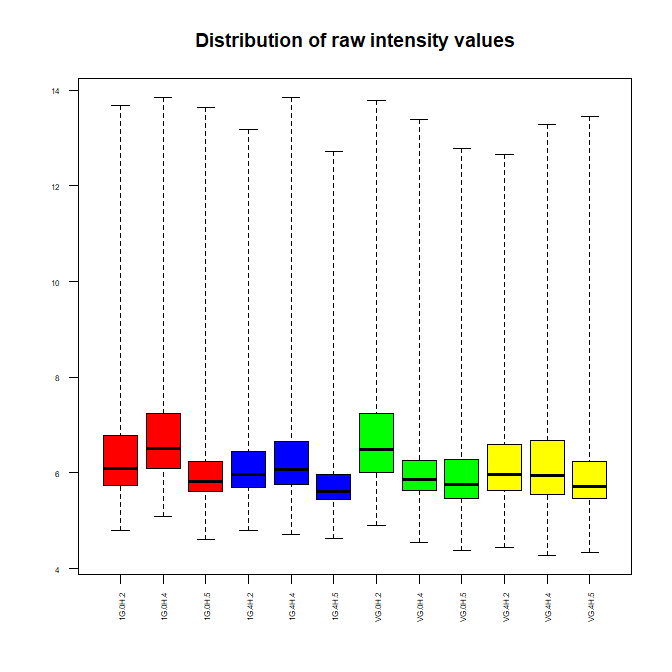
> plotPCA3(exprs(rawData), labels = targets$ShortName, factor = targets$Group,   
+ title="Raw data", scale = FALSE, size = 3,   
+ colores = c("red", "blue", "green", "yellow"))



Visualization of the two first Principal Components for raw data

Se realiza un boxplot para ver la variabilidad de la intensidad de la señal entre muestras:

> boxplot(rawData, cex.axis=0.5, las=2, which="all",   
+ col = c(rep("red", 3), rep("blue", 3), rep("green", 3), rep("yellow", 3)),  
+ main="Distribution of raw intensity values")



Boxplot for arrays intensities (Raw Data)

## Normalización de los datos:

> eset\_rma <- rma(rawData)

Background correcting  
Normalizing  
Calculating Expression

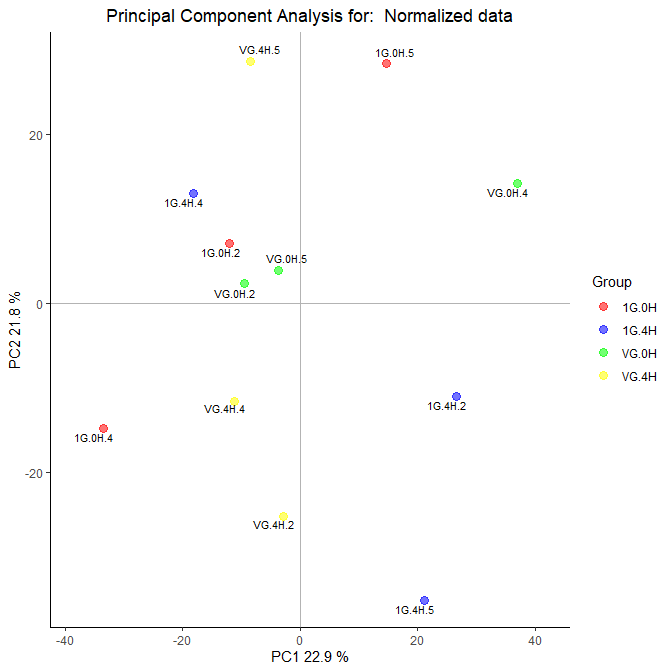
## Control de calidad de datos normalizados:

> arrayQualityMetrics(eset\_rma, outdir = file.path("./results", "QCDir.Norm"), force=TRUE)

Se crea una carpeta dentro de resultados que contiene el control de calidad de los datos normalizados. El archivo index.html contiene el report.

Visualizo el PCA con los datos normalizados

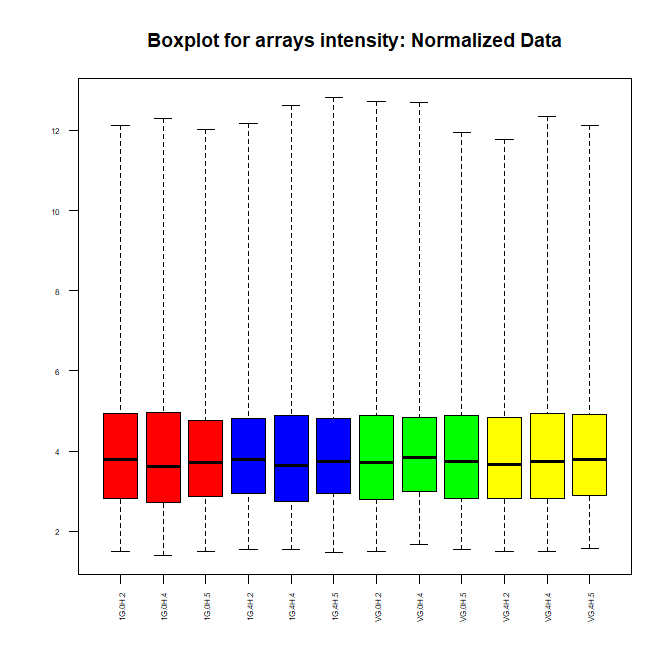
> plotPCA3(exprs(eset\_rma), labels = targets$ShortName, factor = targets$Group,   
+ title="Normalized data", scale = FALSE, size = 3,   
+ colores = c("red", "blue", "green", "yellow"))



Visualization of first two principal components for normalized data

Visualizo un boxplot para ver la variabilidad de la intensidad de la señal entre muestras normalizadas:

> boxplot(eset\_rma, cex.axis=0.5, las=2, which="all",   
+ col = c(rep("red", 3), rep("blue", 3), rep("green", 3), rep("yellow", 3)),  
+ main="Boxplot for arrays intensity: Normalized Data")

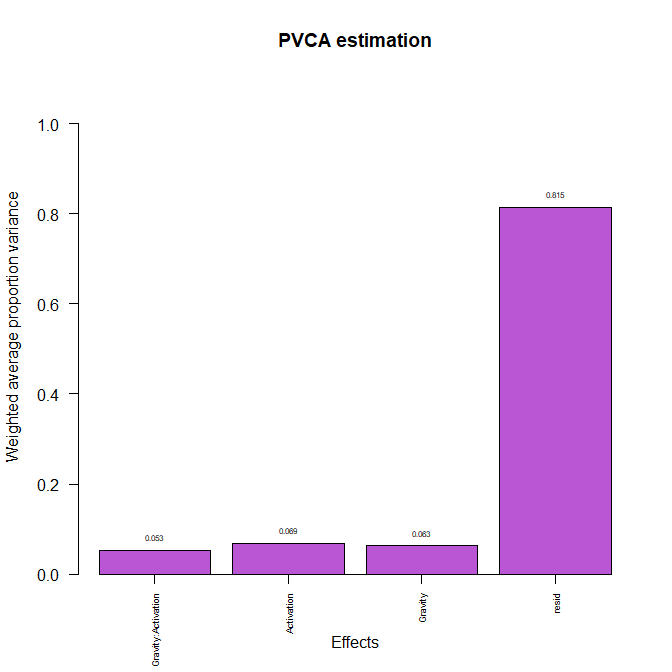


Distribution of intensities for normalized data

## Detección de lotes por Principal Variation Component Analysis:

> #load the library  
> library(pvca)  
> pData(eset\_rma) <- targets  
> #select the threshold  
> pct\_threshold <- 0.6  
> #select the factors to analyze  
> batch.factors <- c("Gravity", "Activation")  
> #run the analysis  
> pvcaObj <- pvcaBatchAssess (eset\_rma, batch.factors, pct\_threshold)

> #plot the results  
> bp <- barplot(pvcaObj$dat, xlab = "Effects",  
+ ylab = "Weighted average proportion variance",  
+ ylim= c(0,1.1),col = c("mediumorchid"), las=2,  
+ main="PVCA estimation")  
> axis(1, at = bp, labels = pvcaObj$label, cex.axis = 0.55, las=2)  
> values = pvcaObj$dat  
> new\_values = round(values , 3)  
> text(bp,pvcaObj$dat,labels = new\_values, pos=3, cex = 0.5)

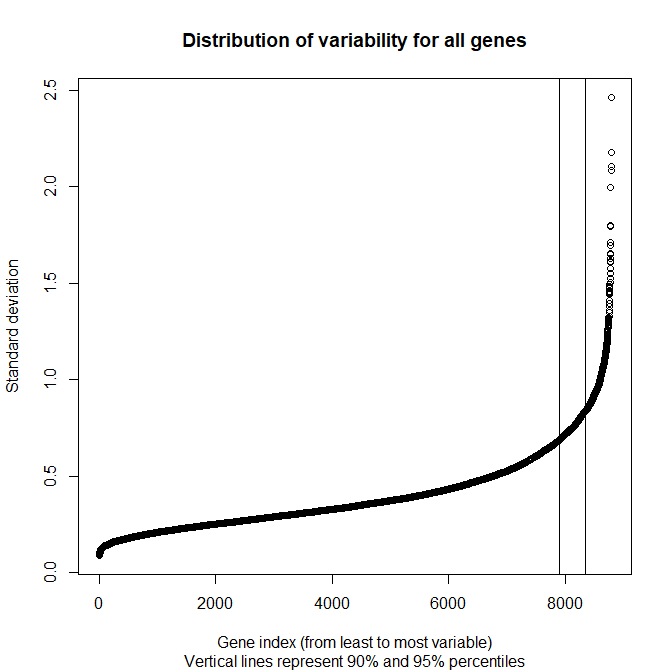


Relative importance of the different factors -genotype, temperature and interaction- affecting gene expression

## Detección de genes mas variables:

Aquí se mostrarán aquellos genes cuya desviación estandard esté por encima del 90 y 95% de todas las desviaciones estandard

> sds <- apply (exprs(eset\_rma), 1, sd)  
> sdsO<- sort(sds)  
> plot(1:length(sdsO), sdsO, main="Distribution of variability for all genes",  
+ sub="Vertical lines represent 90% and 95% percentiles",  
+ xlab="Gene index (from least to most variable)", ylab="Standard deviation")  
> abline(v=length(sds)\*c(0.9,0.95))



Values of standard deviations allong all samples for all genes ordered from smallest to biggest

## Filtrado de genes menos variables:

> library(genefilter)  
> library(hgfocus.db)  
> annotation(eset\_rma) <- "hgfocus.db"  
> filtered <- nsFilter(eset\_rma,   
+ require.entrez = TRUE, remove.dupEntrez = TRUE,  
+ var.filter=TRUE, var.func=IQR, var.cutoff=0.75,   
+ filterByQuantile=TRUE, feature.exclude = "^AFFX")

## Guardado de datos filtrados y normalizados:

> print(filtered$filter.log)

$numDupsRemoved  
[1] 120  
  
$numLowVar  
[1] 6078  
  
$numRemoved.ENTREZID  
[1] 559  
  
$feature.exclude  
[1] 10

> eset\_filtered <-filtered$eset

> write.csv(exprs(eset\_rma), file="./results/normalized.Data.csv")  
> write.csv(exprs(eset\_filtered), file="./results/normalized.Filtered.Data.csv")  
> save(eset\_rma, eset\_filtered, file="./results/normalized.Data.Rda")

## Definición de la matriz de diseño:

> if (!exists("eset\_filtered")) load (file="./results/normalized.Data.Rda")

> library(limma)  
> designMat<- model.matrix(~0+Group, pData(eset\_filtered))  
> colnames(designMat) <- c("G.NONTREATED", "G.TREATED", "VG.NONTREATED", "VG.TREATED")  
> print(designMat)

G.NONTREATED G.TREATED VG.NONTREATED VG.TREATED  
1 1 0 0 0  
2 1 0 0 0  
3 1 0 0 0  
4 0 1 0 0  
5 0 1 0 0  
6 0 1 0 0  
7 0 0 1 0  
8 0 0 1 0  
9 0 0 1 0  
10 0 0 0 1  
11 0 0 0 1  
12 0 0 0 1  
attr(,"assign")  
[1] 1 1 1 1  
attr(,"contrasts")  
attr(,"contrasts")$Group  
[1] "contr.treatment"

## Definición de comparaciones con matriz de contrastes:

> cont.matrix <- makeContrasts (VGVSGTREATED = VG.TREATED-G.TREATED,  
+ VGVSGNONTREATED = VG.NONTREATED-G.NONTREATED,  
+ INT = (VG.TREATED-G.TREATED) - (VG.NONTREATED-G.NONTREATED),  
+ levels=designMat)  
> print(cont.matrix)

Contrasts  
Levels VGVSGTREATED VGVSGNONTREATED INT  
 G.NONTREATED 0 -1 1  
 G.TREATED -1 0 -1  
 VG.NONTREATED 0 1 -1  
 VG.TREATED 1 0 1

## Estimación del modelo y selección de genes:

> library(limma)  
> fit<-lmFit(eset\_filtered, designMat)  
> fit.main<-contrasts.fit(fit, cont.matrix)  
> fit.main<-eBayes(fit.main)  
> class(fit.main)

[1] "MArrayLM"  
attr(,"package")  
[1] "limma"

## Obtención de genes diferencialmente expresados:

Para la primera comparación: Genes que se expresan de manera diferente cuando se activan en función de si la gravedad es constante o variable.

> topTab\_VGVSGTREATED <- topTable (fit.main, number=nrow(fit.main), coef="VGVSGTREATED", adjust="fdr")   
> head(topTab\_VGVSGTREATED)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | logFC | AveExpr | t | P.Value | adj.P.Val | B |
| 201909\_at | 4.450880 | 5.702507 | 5.733260 | 0.0000170 | 0.0344119 | -1.470859 |
| 217232\_x\_at | -3.293567 | 10.488364 | -4.762406 | 0.0001416 | 0.1434877 | -2.029984 |
| 205000\_at | 2.598167 | 2.970576 | 3.805787 | 0.0012253 | 0.8274579 | -2.687190 |
| 208894\_at | 1.783379 | 4.856287 | 3.353896 | 0.0033971 | 0.9978433 | -3.028036 |
| 206700\_s\_at | 1.722553 | 4.245418 | 3.305463 | 0.0037865 | 0.9978433 | -3.065376 |
| 208610\_s\_at | -1.673263 | 7.356095 | -3.103175 | 0.0059425 | 0.9978433 | -3.222484 |

Para la segunda comparación: Genes que se expresan de manera diferente en función de si la gravedad es constante o variable antes de activarse

> topTab\_VGVSGNONTREATED <- topTable (fit.main, number=nrow(fit.main), coef="VGVSGNONTREATED", adjust="fdr")   
> head(topTab\_VGVSGNONTREATED)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | logFC | AveExpr | t | P.Value | adj.P.Val | B |
| 200873\_s\_at | -2.578050 | 5.792393 | -4.774253 | 0.0001380 | 0.2794921 | -3.112913 |
| 203665\_at | -3.059741 | 6.076885 | -3.955459 | 0.0008729 | 0.6797011 | -3.414116 |
| 201068\_s\_at | -1.880674 | 6.041268 | -3.808148 | 0.0012187 | 0.6797011 | -3.473264 |
| 202768\_at | -2.240810 | 5.230130 | -3.765617 | 0.0013420 | 0.6797011 | -3.490599 |
| 217911\_s\_at | -1.951585 | 6.470501 | -3.433439 | 0.0028412 | 0.9851903 | -3.629645 |
| 211506\_s\_at | -2.005053 | 4.216521 | -3.283863 | 0.0039740 | 0.9851903 | -3.694131 |

Para la tercera comparación (INT): Genes que se comportan diferente entre la comparación 1 y 2:

> topTab\_INT <- topTable (fit.main, number=nrow(fit.main), coef="INT", adjust="fdr")   
> head(topTab\_INT)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | logFC | AveExpr | t | P.Value | adj.P.Val | B |
| 208894\_at | 3.316861 | 4.856287 | 4.410808 | 0.0003117 | 0.6314486 | -4.511859 |
| 200910\_at | 3.247301 | 5.702287 | 3.857743 | 0.0010892 | 0.8427838 | -4.524011 |
| 200873\_s\_at | 2.900146 | 5.792393 | 3.797685 | 0.0012480 | 0.8427838 | -4.525417 |
| 201068\_s\_at | 2.542602 | 6.041268 | 3.640521 | 0.0017810 | 0.9020803 | -4.529175 |
| 211506\_s\_at | 2.967051 | 4.216521 | 3.436126 | 0.0028241 | 0.9599724 | -4.534221 |
| 210125\_s\_at | 2.460431 | 6.111311 | 3.338135 | 0.0035193 | 0.9599724 | -4.536700 |

## Anotación génica:

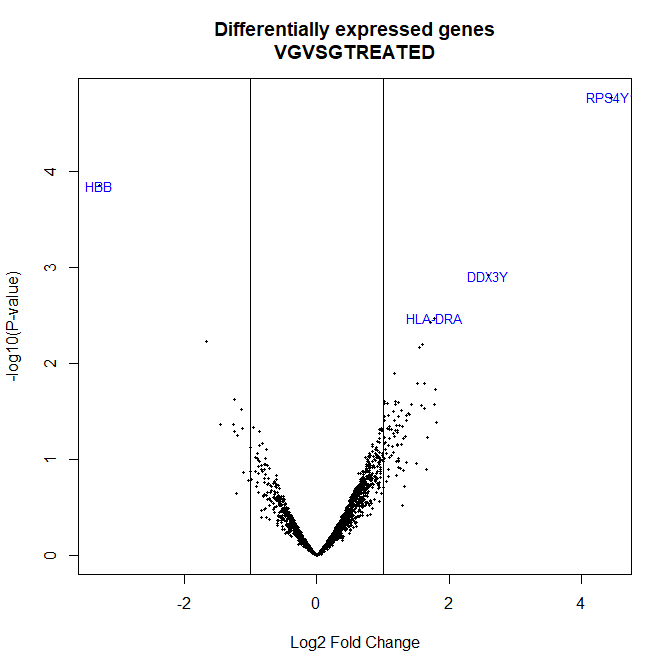
> annotatedTopTable <- function(topTab, anotPackage)  
+ {  
+ topTab <- cbind(PROBEID=rownames(topTab), topTab)  
+ myProbes <- rownames(topTab)  
+ thePackage <- eval(parse(text = anotPackage))  
+ geneAnots <- select(thePackage, myProbes, c("SYMBOL", "ENTREZID", "GENENAME"))  
+ annotatedTopTab<- merge(x=geneAnots, y=topTab, by.x="PROBEID", by.y="PROBEID")  
+ return(annotatedTopTab)  
+ }

> topAnnotated\_VGVSGTREATED <- annotatedTopTable(topTab\_VGVSGTREATED,  
+ anotPackage="hgfocus.db")  
> topAnnotated\_VGVSGNONTREATED <- annotatedTopTable(topTab\_VGVSGNONTREATED,  
+ anotPackage="hgfocus.db")  
> topAnnotated\_INT <- annotatedTopTable(topTab\_INT,  
+ anotPackage="hgfocus.db")  
> write.csv(topAnnotated\_VGVSGTREATED, file="./results/topAnnotated\_VGVSGTREATED.csv")  
> write.csv(topAnnotated\_VGVSGNONTREATED, file="./results/topAnnotated\_VGVSGNONTREATED.csv")  
> write.csv(topAnnotated\_INT, file="./results/topAnnotated\_INT.csv")

PROBEID SYMBOL ENTREZID GENENAME  
1 117\_at HSPA6 3310 heat shock protein family A (Hsp70) member 6  
2 200001\_at CAPNS1 826 calpain small subunit 1  
3 200002\_at RPL35 11224 ribosomal protein L35  
4 200004\_at EIF4G2 1982 eukaryotic translation initiation factor 4 gamma 2  
5 200006\_at PARK7 11315 Parkinsonism associated deglycase

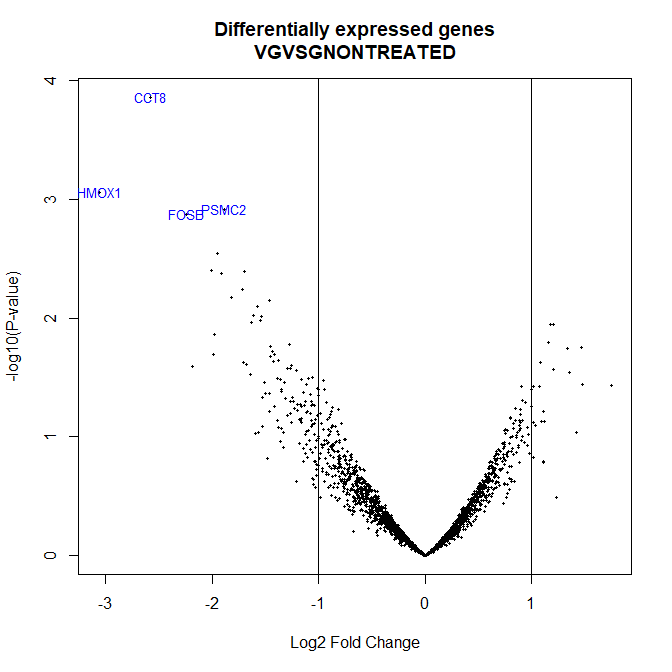
## Visualización de expresión diferencial:

> library(hgfocus.db)  
> geneSymbols <- select(hgfocus.db, rownames(fit.main), c("SYMBOL"))  
> SYMBOLS<- geneSymbols$SYMBOL  
> volcanoplot(fit.main, coef=1, highlight=4, names=SYMBOLS,   
+ main=paste("Differentially expressed genes", colnames(cont.matrix)[1], sep="\n"))  
> abline(v=c(-1,1))



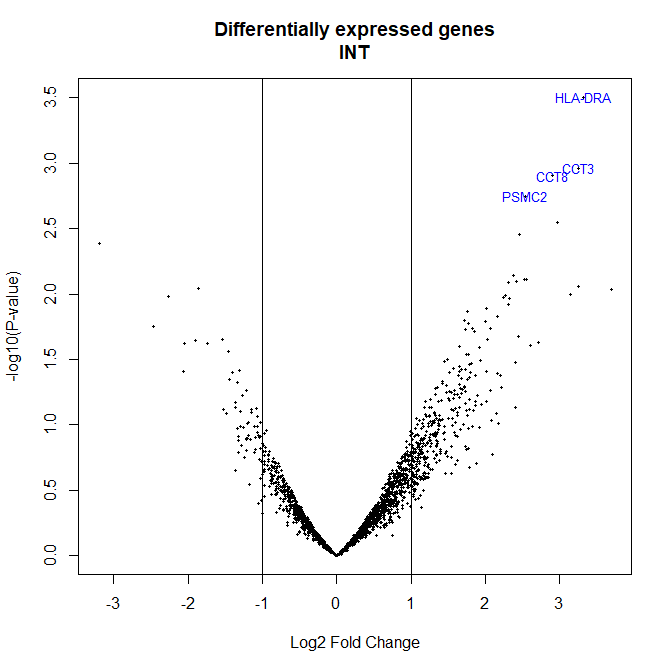
Volcano plot. Genes que se expresan de manera diferente cuando se activan en función de si la gravedad es constante o variable. Se muestran los nombres de los 4 más variables (i.e. cuatro primeros genes en topTable)

> library(hgfocus.db)  
> geneSymbols <- select(hgfocus.db, rownames(fit.main), c("SYMBOL"))  
> SYMBOLS<- geneSymbols$SYMBOL  
> volcanoplot(fit.main, coef=2, highlight=4, names=SYMBOLS,   
+ main=paste("Differentially expressed genes", colnames(cont.matrix)[2], sep="\n"))  
> abline(v=c(-1,1))



Volcano plot. Genes que se expresan de manera diferente en función de si la gravedad es constante o variable antes de activarse. Se muestran los nombres de los 4 más variables (i.e. cuatro primeros genes en topTable)

> library(hgfocus.db)  
> geneSymbols <- select(hgfocus.db, rownames(fit.main), c("SYMBOL"))  
> SYMBOLS<- geneSymbols$SYMBOL  
> volcanoplot(fit.main, coef=3, highlight=4, names=SYMBOLS,   
+ main=paste("Differentially expressed genes", colnames(cont.matrix)[3], sep="\n"))  
> abline(v=c(-1,1))



Volcano plot. Genes que se comportan diferente entre la comparación 1 y 2. Se muestran los nombres de los 4 más variables (i.e. cuatro primeros genes en topTable)

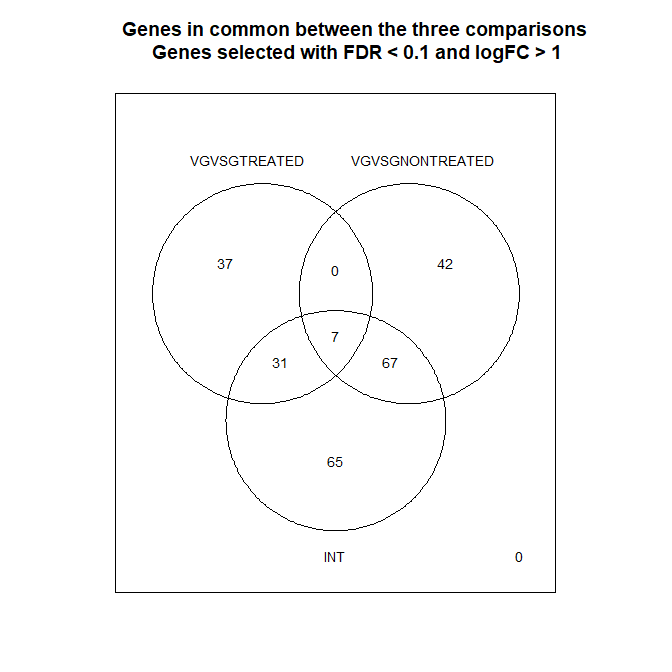
## Comparaciones múltiples:

> library(limma)  
> res<-decideTests(fit.main, method="separate", adjust.method="none", p.value=0.1, lfc=1)

> sum.res.rows<-apply(abs(res),1,sum)  
> res.selected<-res[sum.res.rows!=0,]   
> print(summary(res))

VGVSGTREATED VGVSGNONTREATED INT  
Down 10 98 29  
NotSig 1951 1910 1856  
Up 65 18 141

> vennDiagram (res.selected[,1:3], cex=0.9)  
> title("Genes in common between the three comparisons\n Genes selected with FDR < 0.1 and logFC > 1")

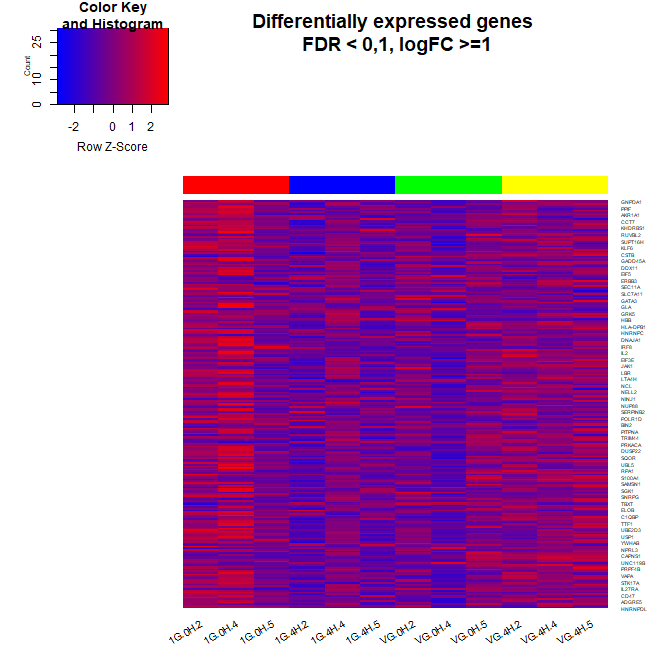


Venn diagram showing the genes in common between the three comparisons performed

## Heatmaps

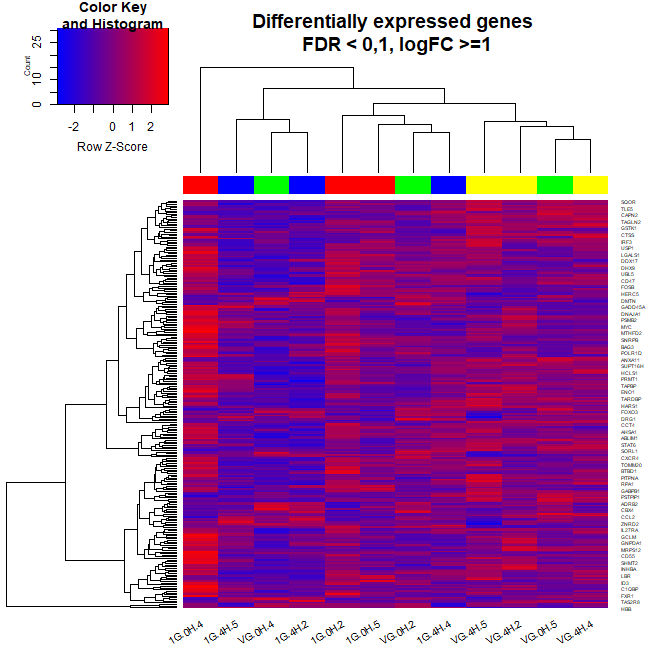
> probesInHeatmap <- rownames(res.selected)  
> HMdata <- exprs(eset\_filtered)[rownames(exprs(eset\_filtered)) %in% probesInHeatmap,]  
>   
> geneSymbols <- select(hgfocus.db, rownames(HMdata), c("SYMBOL"))  
> SYMBOLS<- geneSymbols$SYMBOL  
> rownames(HMdata) <- SYMBOLS  
> write.csv(HMdata, file = file.path("./results/data4Heatmap.csv"))

> my\_palette <- colorRampPalette(c("blue", "red"))(n = 299)  
> library(gplots)  
>   
> heatmap.2(HMdata,  
+ Rowv = FALSE,  
+ Colv = FALSE,  
+ main = "Differentially expressed genes \n FDR < 0,1, logFC >=1",  
+ scale = "row",  
+ col = my\_palette,  
+ sepcolor = "white",  
+ sepwidth = c(0.05,0.05),  
+ cexRow = 0.5,  
+ cexCol = 0.9,  
+ key = TRUE,  
+ keysize = 1.5,  
+ density.info = "histogram",  
+ ColSideColors = c(rep("red",3),rep("blue",3), rep("green",3), rep("yellow",3)),  
+ tracecol = NULL,  
+ dendrogram = "none",  
+ srtCol = 30)



Heatmap for expression data without any grouping

> heatmap.2(HMdata,  
+ Rowv = TRUE,  
+ Colv = TRUE,  
+ dendrogram = "both",  
+ main = "Differentially expressed genes \n FDR < 0,1, logFC >=1",  
+ scale = "row",  
+ col = my\_palette,  
+ sepcolor = "white",  
+ sepwidth = c(0.05,0.05),  
+ cexRow = 0.5,  
+ cexCol = 0.9,  
+ key = TRUE,  
+ keysize = 1.5,  
+ density.info = "histogram",  
+ ColSideColors = c(rep("red",3),rep("blue",3), rep("green",3), rep("yellow",3)),  
+ tracecol = NULL,  
+ srtCol = 30)



Heatmap for expression data grouping genes (rows) and samples (columns) by their similarity

## Significación biológica de los resultados:

> listOfTables <- list(VGVSGTREATED = topTab\_VGVSGTREATED,   
+ VGVSGNONTREATED = topTab\_VGVSGNONTREATED,   
+ INT = topTab\_INT)  
> listOfSelected <- list()  
> for (i in 1:length(listOfTables)){  
+ # select the toptable  
+ topTab <- listOfTables[[i]]  
+ # select the genes to be included in the analysis  
+ whichGenes<-topTab["P.Value"]<0.15  
+ selectedIDs <- rownames(topTab)[whichGenes]  
+ # convert the ID to Entrez  
+ EntrezIDs<- select(hgfocus.db, selectedIDs, c("ENTREZID"))  
+ EntrezIDs <- EntrezIDs$ENTREZID  
+ listOfSelected[[i]] <- EntrezIDs  
+ names(listOfSelected)[i] <- names(listOfTables)[i]  
+ }  
> sapply(listOfSelected, length)

VGVSGTREATED VGVSGNONTREATED INT   
 207 259 269

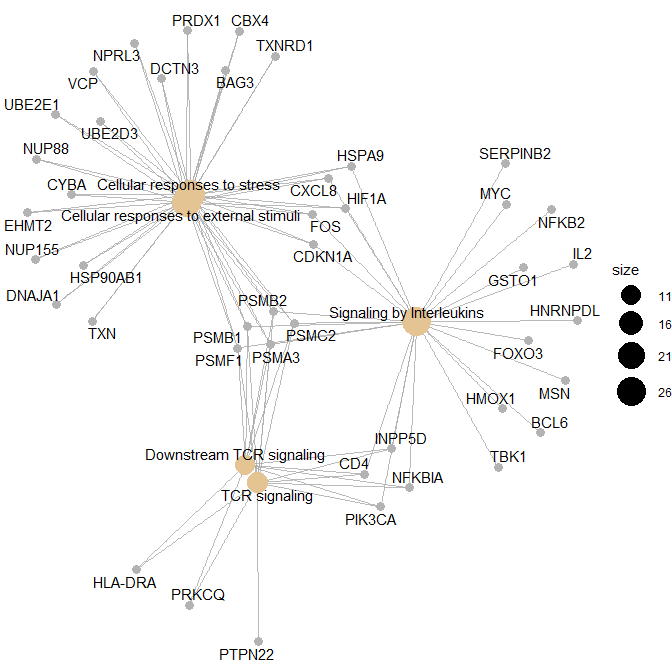
> library(ReactomePA)  
>   
> listOfData1 <- listOfSelected[1]  
> comparisonsNames1 <- names(listOfData1)  
>   
> for (i in 1:length(listOfData1)){  
+ genesIn1 <- listOfData1[[i]]  
+ comparison1 <- comparisonsNames1[i]  
+ enrich.result1 <- enrichPathway(gene = genesIn1,  
+ pvalueCutoff = 0.05,  
+ readable = T,  
+ pAdjustMethod = "BH",  
+ organism = "human")  
+   
+ cat("##################################")  
+ cat("\nComparison1: ", comparison1,"\n")  
+ print(head(enrich.result1))  
+   
+ if (length(rownames(enrich.result1@result)) != 0) {  
+ write.csv(as.data.frame(enrich.result1),   
+ file =paste0("./results/","ReactomePA.Results.",comparison1,".csv"),   
+ row.names = FALSE)  
+   
+ pdf(file=paste0("./results/","ReactomePABarplot.",comparison1,".pdf"))  
+ print(barplot(enrich.result1, showCategory = 15, font.size = 4,   
+ title = paste0("Reactome Pathway Analysis for ", comparison1,". Barplot")))  
+ dev.off()  
+   
+ pdf(file = paste0("./results/","ReactomePAcnetplot.",comparison1,".pdf"))  
+ print(cnetplot(enrich.result1, categorySize = "geneNum", schowCategory = 15,   
+ vertex.label.cex = 0.75))  
+ dev.off()  
+ }  
+ }

##################################  
Comparison1: VGVSGTREATED   
 ID  
R-HSA-6798695 R-HSA-6798695  
R-HSA-877300 R-HSA-877300  
R-HSA-8862803 R-HSA-8862803  
R-HSA-8863678 R-HSA-8863678  
R-HSA-5628897 R-HSA-5628897  
R-HSA-3700989 R-HSA-3700989  
 Description  
R-HSA-6798695 Neutrophil degranulation  
R-HSA-877300 Interferon gamma signaling  
R-HSA-8862803 Deregulated CDK5 triggers multiple neurodegenerative pathways in Alzheimer's disease models  
R-HSA-8863678 Neurodegenerative Diseases  
R-HSA-5628897 TP53 Regulates Metabolic Genes  
R-HSA-3700989 Transcriptional Regulation by TP53  
 GeneRatio BgRatio pvalue p.adjust qvalue  
R-HSA-6798695 26/168 480/10654 3.037643e-08 2.311646e-05 1.717067e-05  
R-HSA-877300 9/168 92/10654 1.377995e-05 3.703329e-03 2.750796e-03  
R-HSA-8862803 5/168 22/10654 1.946559e-05 3.703329e-03 2.750796e-03  
R-HSA-8863678 5/168 22/10654 1.946559e-05 3.703329e-03 2.750796e-03  
R-HSA-5628897 8/168 86/10654 6.038203e-05 7.899761e-03 5.867863e-03  
R-HSA-3700989 17/168 365/10654 6.228458e-05 7.899761e-03 5.867863e-03  
 geneID  
R-HSA-6798695 HBB/CTSS/TYROBP/BIN2/CD47/CSTB/LGALS3/LTA4H/ITGB2/IMPDH2/DDOST/PSAP/FCER1G/PNP/NIT2/XRCC5/PTPN6/CPNE1/SERPINB6/CCT2/PKM/CTSB/GDI2/CYBA/PSMD11/IDH1  
R-HSA-877300 HLA-DRA/HLA-DPB1/JAK1/IRF3/MT2A/GBP2/OAS3/PTPN6/STAT1  
R-HSA-8862803 CAST/CAPNS1/PRDX2/CDC25B/LMNA  
R-HSA-8863678 CAST/CAPNS1/PRDX2/CDC25B/LMNA  
R-HSA-5628897 YWHAB/PRDX2/TXN/TXNRD1/TSC2/SCO2/RRAGA/YWHAZ  
R-HSA-3700989 YWHAB/RPA1/ELOB/PLK3/SUPT16H/PRDX2/CCNH/BRD7/TXN/CNOT8/BCL2L14/BNIP3L/TXNRD1/TSC2/SCO2/RRAGA/YWHAZ  
 Count  
R-HSA-6798695 26  
R-HSA-877300 9  
R-HSA-8862803 5  
R-HSA-8863678 5  
R-HSA-5628897 8  
R-HSA-3700989 17

> library(ReactomePA)  
>   
> listOfData2 <- listOfSelected[2]  
> comparisonsNames2 <- names(listOfData2)  
>   
> for (i in 1:length(listOfData2)){  
+ genesIn2 <- listOfData2[[i]]  
+ comparison2 <- comparisonsNames2[i]  
+ enrich.result2 <- enrichPathway(gene = genesIn2,  
+ pvalueCutoff = 0.05,  
+ readable = T,  
+ pAdjustMethod = "BH",  
+ organism = "human")  
+   
+ cat("##################################")  
+ cat("\nComparison2: ", comparison2,"\n")  
+ print(head(enrich.result2))  
+   
+ if (length(rownames(enrich.result2@result)) != 0) {  
+ write.csv(as.data.frame(enrich.result2),   
+ file =paste0("./results/","ReactomePA.Results.",comparison2,".csv"),   
+ row.names = FALSE)  
+   
+ pdf(file=paste0("./results/","ReactomePABarplot.",comparison2,".pdf"))  
+ print(barplot(enrich.result2, showCategory = 15, font.size = 4,   
+ title = paste0("Reactome Pathway Analysis for ", comparison2,". Barplot")))  
+ dev.off()  
+   
+ pdf(file = paste0("./results/","ReactomePAcnetplot.",comparison2,".pdf"))  
+ print(cnetplot(enrich.result2, categorySize = "geneNum", schowCategory = 15,   
+ vertex.label.cex = 0.75))  
+ dev.off()  
+ }  
+ }

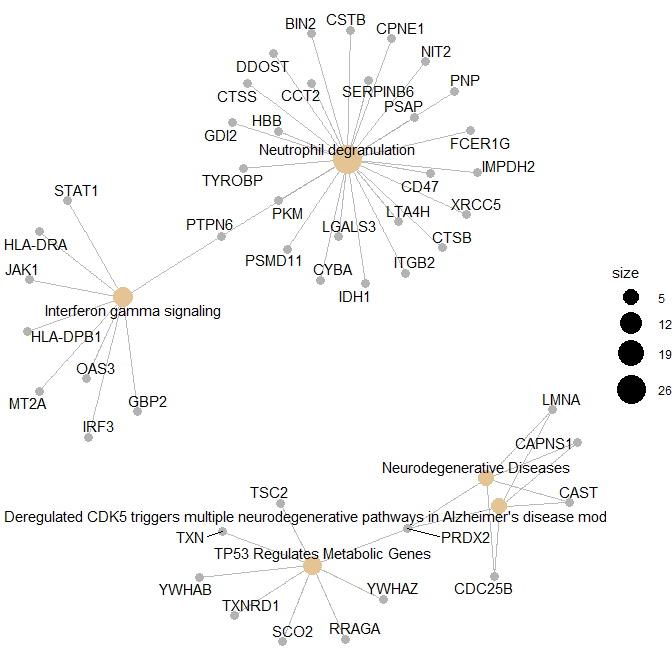
##################################  
Comparison2: VGVSGNONTREATED   
 ID Description GeneRatio  
R-HSA-2262752 R-HSA-2262752 Cellular responses to stress 26/201  
R-HSA-449147 R-HSA-449147 Signaling by Interleukins 25/201  
R-HSA-8953897 R-HSA-8953897 Cellular responses to external stimuli 26/201  
R-HSA-202424 R-HSA-202424 Downstream TCR signaling 11/201  
R-HSA-202403 R-HSA-202403 TCR signaling 12/201  
R-HSA-5689880 R-HSA-5689880 Ub-specific processing proteases 15/201  
 BgRatio pvalue p.adjust qvalue  
R-HSA-2262752 479/10654 1.095326e-06 0.0003873925 0.0002993396  
R-HSA-449147 461/10654 1.827535e-06 0.0003873925 0.0002993396  
R-HSA-8953897 493/10654 1.877463e-06 0.0003873925 0.0002993396  
R-HSA-202424 98/10654 2.191850e-06 0.0003873925 0.0002993396  
R-HSA-202403 119/10654 2.373728e-06 0.0003873925 0.0002993396  
R-HSA-5689880 220/10654 1.744254e-05 0.0023721848 0.0018329962  
 geneID  
R-HSA-2262752 PSMC2/BAG3/CXCL8/UBE2D3/NUP88/TXN/PSMA3/NUP155/NPRL3/PSMF1/UBE2E1/EHMT2/VCP/HSPA9/DNAJA1/PSMB2/CBX4/TXNRD1/CDKN1A/PRDX1/HSP90AB1/FOS/CYBA/PSMB1/DCTN3/HIF1A  
R-HSA-449147 HMOX1/PSMC2/CXCL8/BCL6/HNRNPDL/FOXO3/GSTO1/NFKB2/PSMA3/PSMF1/PIK3CA/SERPINB2/HSPA9/PSMB2/NFKBIA/CDKN1A/IL2/MYC/MSN/FOS/CD4/PSMB1/INPP5D/HIF1A/TBK1  
R-HSA-8953897 PSMC2/BAG3/CXCL8/UBE2D3/NUP88/TXN/PSMA3/NUP155/NPRL3/PSMF1/UBE2E1/EHMT2/VCP/HSPA9/DNAJA1/PSMB2/CBX4/TXNRD1/CDKN1A/PRDX1/HSP90AB1/FOS/CYBA/PSMB1/DCTN3/HIF1A  
R-HSA-202424 PSMC2/HLA-DRA/PSMA3/PSMF1/PIK3CA/PSMB2/NFKBIA/CD4/PSMB1/PRKCQ/INPP5D  
R-HSA-202403 PSMC2/HLA-DRA/PSMA3/PSMF1/PIK3CA/PSMB2/NFKBIA/PTPN22/CD4/PSMB1/PRKCQ/INPP5D  
R-HSA-5689880 PSMC2/ADRB2/PSMA3/PSMF1/USP3/PSMB2/NFKBIA/MYC/TADA3/USP14/USP16/SMAD7/TOMM20/PSMB1/HIF1A  
 Count  
R-HSA-2262752 26  
R-HSA-449147 25  
R-HSA-8953897 26  
R-HSA-202424 11  
R-HSA-202403 12  
R-HSA-5689880 15

> cnetplot(enrich.result2, categorySize = "geneNum", schowCategory = 15,   
+ vertex.label.cex = 0.75)



Network obtained from the Reactome enrichment analysis on the list obtained from the comparison between VG and 1G non activated

> cnetplot(enrich.result1, categorySize = "geneNum", schowCategory = 15,   
+ vertex.label.cex = 0.75)



Network obtained from the Reactome enrichment analysis on the list obtained from the comparison between VG and 1G non activated

First rows and columns for Reactome results on VGVSGTREATED.csv comparison

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Description | GeneRatio | BgRatio | pvalue | p.adjust |
| R-HSA-6798695 | Neutrophil degranulation | 26/168 | 480/10654 | 3.03764253168675e-08 | 2.31164596661362e-05 |
| R-HSA-877300 | Interferon gamma signaling | 9/168 | 92/10654 | 1.37799538771663e-05 | 0.00370332892163295 |
| R-HSA-8862803 | Deregulated CDK5 triggers multiple neurodegenerative pathways in Alzheimer’s disease models | 5/168 | 22/10654 | 1.94655922293453e-05 | 0.00370332892163295 |
| R-HSA-8863678 | Neurodegenerative Diseases | 5/168 | 22/10654 | 1.94655922293453e-05 | 0.00370332892163295 |

First rows and columns for Reactome results on VGVSGNONTREATED.csv comparison

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Description | GeneRatio | BgRatio | pvalue | p.adjust |
| R-HSA-2262752 | Cellular responses to stress | 26/201 | 479/10654 | 1.0953257232416e-06 | 0.000387392478430706 |
| R-HSA-449147 | Signaling by Interleukins | 25/201 | 461/10654 | 1.82753465586944e-06 | 0.000387392478430706 |
| R-HSA-8953897 | Cellular responses to external stimuli | 26/201 | 493/10654 | 1.87746316820232e-06 | 0.000387392478430706 |
| R-HSA-202424 | Downstream TCR signaling | 11/201 | 98/10654 | 2.19185022281151e-06 | 0.000387392478430706 |

## Resumen de resultados:

List of files generated in the analysis

|  |
| --- |
| List\_of\_Files |
| data4Heatmap.csv |
| normalized.Data.csv |
| normalized.Data.Rda |
| normalized.Filtered.Data.csv |
| QCDir.Norm |
| QCDir.Raw |
| ReactomePA.Results.VGVSGNONTREATED.csv |
| ReactomePA.Results.VGVSGTREATED.csv |
| ReactomePABarplot.VGVSGNONTREATED.pdf |
| ReactomePABarplot.VGVSGTREATED.pdf |
| ReactomePAcnetplot.VGVSGNONTREATED.pdf |
| ReactomePAcnetplot.VGVSGTREATED.pdf |
| topAnnotated\_INT.csv |
| topAnnotated\_VGVSGNONTREATED.csv |
| topAnnotated\_VGVSGTREATED.csv |