Zonal wk1 EGCG/Ctrl

Using GEOquery to load in phenodata associated with count data file

```
library(GEOquery)
```

```
## Loading required package: Biobase
```

```
## Loading required package: BiocGenerics
```

```
##
## Attaching package: 'BiocGenerics'
```

```
## The following objects are masked from 'package:stats':
##
##     IQR, mad, sd, var, xtabs
```

```
## The following objects are masked from 'package:base':
##
##     anyDuplicated, aperm, append, as.data.frame, basename, cbind,
##     colnames, dirname, do.call, duplicated, eval, evalq, Filter, Find,
##     get, grep, grepl, intersect, is.unsorted, lapply, Map, mapply,
##     match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,
##     Position, rank, rbind, Reduce, rownames, sapply, setdiff, sort,
##     table, tapply, union, unique, unsplit, which.max, which.min
```

```
## Welcome to Bioconductor
##
##     Vignettes contain introductory material; view with
##     'browseVignettes()'. To cite Bioconductor, see
##     'citation("Biobase")', and for packages 'citation("pkgname")'.
```

```
## Setting options('download.file.method.GEOquery'='auto')
```

```
## Setting options('GEOquery.inmemory.gpl'=FALSE)
```

```
gse = getGEO("GSE124161") #unfortunately the data is not included in this file, so we need to load that in separately.
```

```
## Found 1 file(s)
```

```
## GSE124161_series_matrix.txt.gz
```

Loading in dataset previously downloaded from NCBI-GEO directly from stored computer location.

```
GSE124161_readcount <- read.delim("~/NYU/BIGY-7633 Transcriptomics/project/GSE124161_readcoun
t.txt", row.names=1)
```

readcount file needs to have the metadata associated with the sample ID names

since they re-ordered the count data, it is different than their metadata file, we need to re-create the metadata to fit the revised order that they utilized in the count data file.

Retreiving metadata from series_matrix file, int the order of the count data file

```
pheno_data <-gse[["GSE124161_series_matrix.txt.gz"]]@phenoData@data[["title"]]
D0 <-pheno_data[1:3]
W1T<-pheno_data[c(4,6,8)]
W1C<-pheno_data[c(5,7,9)]
W2T<-pheno_data[c(10,12,14)]
W2C<-pheno_data[c(11,13,15)]
W3T<-pheno_data[c(16,18,20)]
W3C<-pheno_data[c(17,19,21)]
W4T<-pheno_data[c(22,24,26)]
W4C<-pheno_data[c(23,25,27)]
W5T<-pheno_data[c(28,30,32)]
W5C<-pheno_data[c(29,31,33)]
W6T<-pheno_data[c(34,36,38)]
W6C<-pheno_data[c(35,37,39)]
W8T<-pheno_data[c(40,42,44)]
W8C<-pheno_data[c(41,43,45)]
count_pheno <-c(D0,W1T, W1C, W2T, W2C, W3T, W3C, W4T, W4C, W5T, W5C, W6T, W6C, W8T, W8C)
count_pheno
```

```
##  [1] "GT01 D0"                          "GT09 D0"
##  [3] "GT19 D0"                          "GT57 Week 1 Treated [GT57_1_T]"
##  [5] "GT58 Week 1 Treated [GT58_1_T]"   "GT59 Week 1 Treated [GT59_1_T ]"
##  [7] "GT57 Week 1 Control  [GT57_1_C ]" "GT58 Week 1 Control [GT58_1_C ]"
##  [9] "GT59 Week 1 Control [GT59_1_C]"   "GT51 Week 2 Treated [GT51_2_T]"
## [11] "GT52 Week 2 Treated [GT52_2_T]"   "GT53 Week 2 Treated [GT53_2_T]"
## [13] "GT51 Week 2 Control [GT51_2_C]"   "GT52 Week 2 Control [GT52_2_C]"
## [15] "GT53 Week 2 Control [GT53_2_C ]"  "GT45 Week 3 Treated [GT45_3_T]"
## [17] "GT46 Week 3 Treated [GT46_3_T]"   "GT47 Week 3 Treated [GT47_3_T]"
## [19] "GT45 Week 3 Control [GT45_3_C]"   "GT46 Week 3 Control [GT46_3_C]"
## [21] "GT47 Week 3 Control [GT47_3_C]"   "GT17 Week 4 Treated [GT17_4_T]"
## [23] "GT18 Week 4 Treated [GT18_4_T]"   "GT19 Week 4 Treated [GT19_4_T]"
## [25] "GT17 Week 4 Control [GT17_4_C]"   "GT18 Week 4 Control [GT18_4_C]"
## [27] "GT19 Week 4 Control [GT19_4_C]"   "GT39 Week 5 Treated [GT39_5_T]"
## [29] "GT40 Week 5 Treated [GT40_5_T]"   "GT41 Week 5 Treated [GT41_5_T]"
## [31] "GT39 Week 5 Control [GT39_5_C]"   "GT40 Week 5 Control [GT40_5_C]"
## [33] "GT41 Week 5 Control [GT41_5_C]"   "GT34 Week 6 Treated [GT34_6_T]"
## [35] "GT35 Week 6 Treated [GT35_6_T]"   "GT37 Week 6 Treated [GT37_6_T]"
## [37] "GT34 Week 6 Control [GT34_6_C]"   "GT35 Week 6 Control [GT35_6_C]"
## [39] "GT37 Week 6 Control [GT37_6_C]"   "GT27 Week 8 Treated [GT27_8_T]"
## [41] "GT28 Week 8 Treated [GT28_8_T]"   "GT29 Week 8 Treated [GT29_8_T]"
## [43] "GT27 Week 8 Control [GT27_8_C]"   "GT28 Week 8 Control [GT28_8_C]"
## [45] "GT29 Week 8 Control [GT29_8_C]"
```

Capturing count data file column names to match the metadata against sample names and treatment levels to be created in a dataframe below

```
count_cols <- names(GSE124161_readcount)#get the column names from the read count data
count_cols
```

```
##  [1] "GT01_D0"  "GT09_D0"  "GT19_D0"  "GT57_1_T" "GT58_1_T" "GT59_1_T"
##  [7] "GT57_1_C" "GT58_1_C" "GT59_1_C" "GT51_2_T" "GT52_2_T" "GT53_2_T"
## [13] "GT51_2_C" "GT52_2_C" "GT53_2_C" "GT45_3_T" "GT46_3_T" "GT47_3_T"
## [19] "GT45_3_C" "GT46_3_C" "GT47_3_C" "GT17_4_T" "GT18_4_T" "GT19_4_T"
## [25] "GT17_4_C" "GT18_4_C" "GT19_4_C" "GT39_5_T" "GT40_5_T" "GT41_5_T"
## [31] "GT39_5_C" "GT40_5_C" "GT41_5_C" "GT34_6_T" "GT35_6_T" "GT37_6_T"
## [37] "GT34_6_C" "GT35_6_C" "GT37_6_C" "GT27_8_T" "GT28_8_T" "GT29_8_T"
## [43] "GT27_8_C" "GT28_8_C" "GT29_8_C"
```

Constructing a data frame for later use with sample metadata

```
#this is matching the original count column names to the phenotype names I ordered in prior R code
pheno_df<-cbind(count_pheno,count_cols)
pheno_df<-as.data.frame(pheno_df)
```

Continue to build the metadata dataframe object

```
#I may need a factor separating the count data by week, so I am creating the information in an ordered fashion, to integrate into the data frame as a column.
day0 <-rep("D0", each = 3)
wknames <- c("W1", "W2", "W3", "W4", "W5", "W6", "W8")
weeks1_8 <-rep(wknames, each = 6)


all_weeks <-c(day0,weeks1_8)

pheno_df$weeks <-all_weeks
```

Keep building the dataframe, adding columns of metadata as necessary.

```
#adding another column to designate the treatment and control groups associated with the colum ns in the count file
group0 <-c("Day0","Day0","Day0")
expnames <- c("EGCG", "Placebo")
Group1_8 <-rep(expnames, each =3, times =7)

groups<-append(group0, Group1_8)

pheno_df$groups <-groups
```

Keep building the dataframe, adding columns of metadata as necessary.

```
#adding uninjured-injured to separate out the groups, in case we want to use this comparison.
uninjured <-c("uninjured","uninjured", "uninjured")
injured <-rep("injured", each =1, times =42)

treatment <-append(uninjured, injured)
pheno_df$treatment <-treatment #adding column "status" to pheno_df
```

Keep building the dataframe, adding columns of metadata as

necessary.

```
#adding "none", "zonal" and "direct" to pheno table to designate application type
none<- rep("none", each = 1, times =3)
zonal <-rep("zonal", each =1, times = 12)
direct <-rep("direct", each = 1, times = 30)

appl <- append(none, zonal)
application <-append(appl,direct)


pheno_df$application <-application #adding column "application" to pheno_df
```

Keep building the dataframe, adding columns of metadata as necessary

```
#adding "D0", "Zw1", "Zw2" and "Dw1", "Dw2", "Dw3", "Dw4", "Dw6" to pheno table to designate a
pplication type by week
non<- rep("W0", each = 1, times =3)
zw1 <-rep("Zw1", each =1, times = 6)
zw2 <-rep("Zw2", each =1, times = 6)
dw1 <-rep("Dw1", each = 1, times = 6)
dw2 <-rep("Dw2", each = 1, times = 6)
dw3 <-rep("Dw3", each = 1, times = 6)
dw4 <-rep("Dw4", each = 1, times = 6)
dw6 <-rep("Dw6", each = 1, times = 6)

zon1 <- append(non, zw1)
zon2 <- append(zon1, zw2)
dir1 <- append(zon2, dw1)
dir2 <- append(dir1, dw2)
dir3 <- append(dir2, dw3)
dir4 <- append(dir3, dw4)
dir6 <- append(dir4, dw6)

pheno_df$appl_by_wk <-dir6 #adding column "application" to pheno_df
```

Rename columns in pheno_df for clarity

```
#renaming column names in data frame for clarity
colnames(pheno_df) <-c("samples", "count_colnames", "week", "treatments", "status", "applicati
on", "appl_by_wk")
pheno_df
```

samples <chr>	count_colnames <chr>	w…	treatments <chr>	status <chr>	application <chr>	app <ch
GT01 D0	GT01_D0	D0	Day0	uninjured	none	W0
GT09 D0	GT09_D0	D0	Day0	uninjured	none	W0
GT19 D0	GT19_D0	D0	Day0	uninjured	none	W0

samples <chr>	count_colnames <chr>	w…	treatments <chr>	status <chr>	application <chr>	app <ch
GT57 Week 1 Treated [GT57_1_T]	GT57_1_T	W1	EGCG	injured	zonal	Zw1
GT58 Week 1 Treated [GT58_1_T]	GT58_1_T	W1	EGCG	injured	zonal	Zw1
GT59 Week 1 Treated [GT59_1_T]	GT59_1_T	W1	EGCG	injured	zonal	Zw1
GT57 Week 1 Control [GT57_1_C]	GT57_1_C	W1	Placebo	injured	zonal	Zw1
GT58 Week 1 Control [GT58_1_C]	GT58_1_C	W1	Placebo	injured	zonal	Zw1
GT59 Week 1 Control [GT59_1_C]	GT59_1_C	W1	Placebo	injured	zonal	Zw1
GT51 Week 2 Treated [GT51_2_T]	GT51_2_T	W2	EGCG	injured	zonal	Zw2

1-10 of 45 rows Previous **1** 2 3 4 5 Next

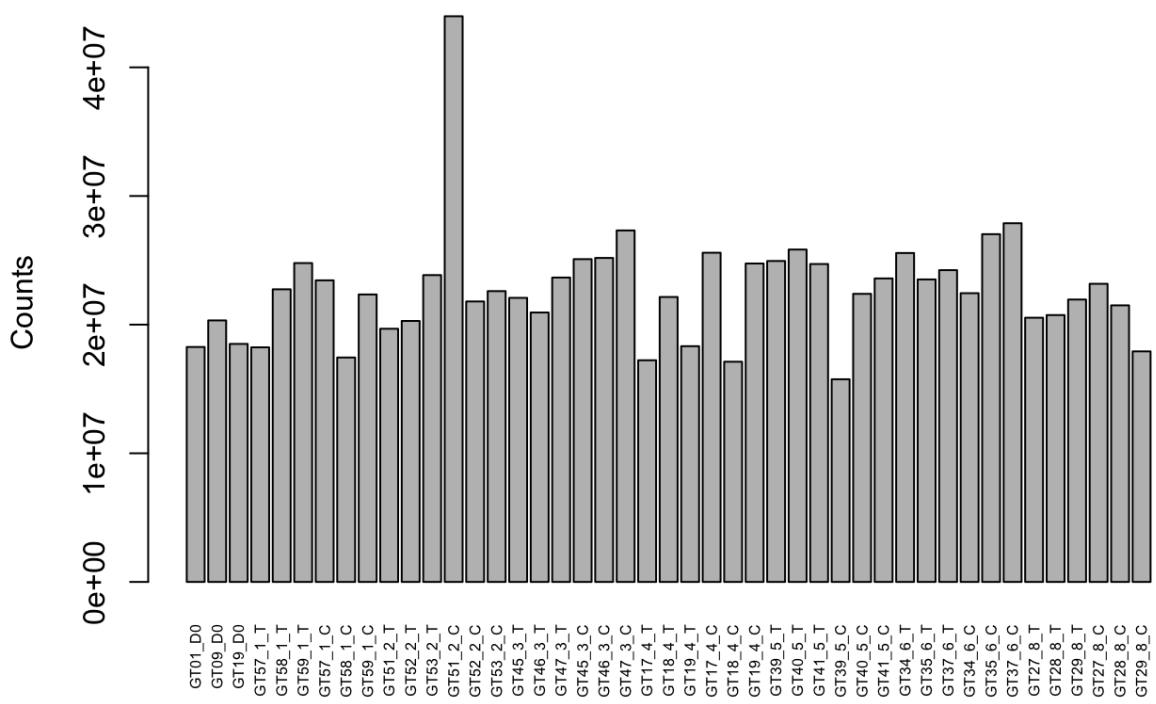
Quality Control

```
sums <- colSums(GSE124161_readcount)

#GT51_2_C, which is one of the week 2 control samples, is double the depth of the entire exper
iment 43,971,422 (divided in half it's = 21,985,711), so yes it is double. I wonder if they fo
und this, and removed it. One way is to check the week 2 heat maps to see if this sample has a
n extreme up regulation. It's there, so they did not exclude it, but nothing shows as extremel
y up-regulated???? What is going on here??
sums
```

```
##  GT01_D0  GT09_D0  GT19_D0 GT57_1_T GT58_1_T GT59_1_T GT57_1_C GT58_1_C
## 18261863 20328995 18501060 18235034 22746777 24785167 23443164 17444604
## GT59_1_C GT51_2_T GT52_2_T GT53_2_T GT51_2_C GT52_2_C GT53_2_C GT45_3_T
## 22345084 19680608 20286078 23851421 43971422 21802025 22604825 22081693
## GT46_3_T GT47_3_T GT45_3_C GT46_3_C GT47_3_C GT17_4_T GT18_4_T GT19_4_T
## 20941586 23660164 25090699 25184243 27321326 17230960 22148122 18323913
## GT17_4_C GT18_4_C GT19_4_C GT39_5_T GT40_5_T GT41_5_T GT39_5_C GT40_5_C
## 25588820 17116435 24752227 24947885 25837140 24712048 15753422 22393291
## GT41_5_C GT34_6_T GT35_6_T GT37_6_T GT34_6_C GT35_6_C GT37_6_C GT27_8_T
## 23591332 25568002 23513634 24235461 22444354 27031627 27882501 20537596
## GT28_8_T GT29_8_T GT27_8_C GT28_8_C GT29_8_C
## 20743466 21951239 23173809 21498675 17921576
```

```
barplot(sums,
        main = "Counts Across Samples",
        ylab = "Counts",
        cex.names = 0.5,
        las = 3)
```



Some Quick Analysis: Violon Plot

```
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✔ dplyr     1.1.2     ✔ readr     2.1.4
## ✔ forcats   1.0.0     ✔ stringr   1.5.0
## ✔ ggplot2   3.4.2     ✔ tibble    3.2.1
## ✔ lubridate 1.9.2     ✔ tidyr     1.3.0
## ✔ purrr     1.0.1
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::combine()    masks Biobase::combine(), BiocGenerics::combine()
## ✖ dplyr::filter()     masks stats::filter()
## ✖ dplyr::lag()        masks stats::lag()
## ✖ ggplot2::Position() masks BiocGenerics::Position(), base::Position()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to bec
ome errors
```

```
library(reshape2)
```

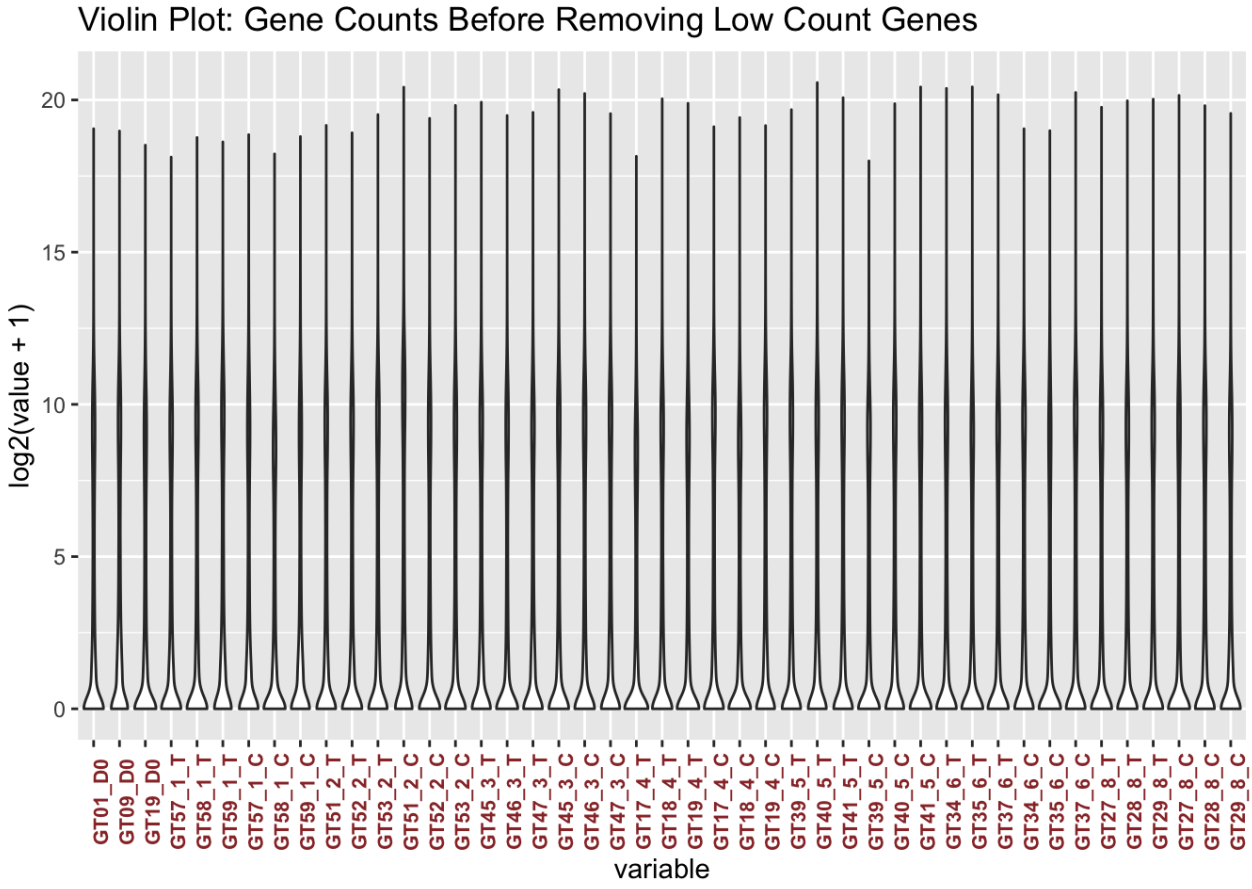
```
##
## Attaching package: 'reshape2'
##
## The following object is masked from 'package:tidyr':
##
##     smiths
```

```
datamat = apply(GSE124161_readcount, 2, as.integer)
data= as.data.frame(datamat)
rownames(data) = rownames(GSE124161_readcount)

data_wnames = data #data with gene names
data_wnames$gene = rownames(data) #creating a column $gene in the data stored in variable data
_wnames, using the rownames from the data variable
data_melt = melt(data_wnames) #melting the data into long form (see in environment) This is hu
ge, for every gene in the dataset 48,162 X 45 samples = 2,167,290 entries
```

```
## Using gene as id variables
```

```
ggplot(data_melt) +
  geom_violin(mapping = aes(x=variable, y=log2(value +1))) + theme(axis.text.x = element_text
(face="bold", color="#993333",
                           size=8, angle=90)) + labs(title="Violin Plot: Gene Counts Before Re
moving Low Count Genes")
```



```
# this violin plot gives you an idea on how the data is distributed. You would expected all of them to look the same and the samples should not vastly deviate from each other over the entire data set. So we are looking to see that all the samples and replicates do not have big differences

# in this example we want the variable as the x axis (the variable is the names in the melted graph, that will group/condense according to the name, and will be the samples in the graph, we should have 45 samples plotted)

# the (value+1) is added because if you have a value of 0, and you take the log of 0, you have a problem - it's undefined, so if you add 1 to all of the values, then a log of 1 will = 0 and everything will be scaled identically with 1 extra count added across the board, so we can get the log(1) = 0

# and we can see this in the violin plot as the thick base in the violin plot is the genes that actually have 0 counts.
```

Trim the dataset from low count genes

Removing low expressed genes. When you do the differential expression you do not want to run statistics on the background noise.Look at the data in the following histogram to determine how many genes are low expressed. 6,000 of the genes (look at the histogram below) are not going to give you results, they are too low, so you should trim the dataset.

```
genesum = rowSums(data) # we are asking R to calculate the sums of the counts in each row(gene) across all 45 samples, and we will graph the result as a histogram below.
hist(log2(genesum+1), ylim = c(0,10000), xlim = c(0,30), breaks = 25) #this generates the histogram below, and from looking at the histogram, we can see that 6K genes have a near-zero expression level, really low values, across the entire dataset.
```


```
sum(genesum ==0) #here we are asking exactly how many genes are equal to 0, which we get 4467 out of 48,162 total genes, this is not bad, it is expected, not all genes are going to be expressed in an RNAseq experiment.
```

```
## [1] 4467
```

```
sum(genesum <2) # and 6518 genes have a count across the 45 samples when totaled up are less than 2,
```

```
## [1] 6518
```

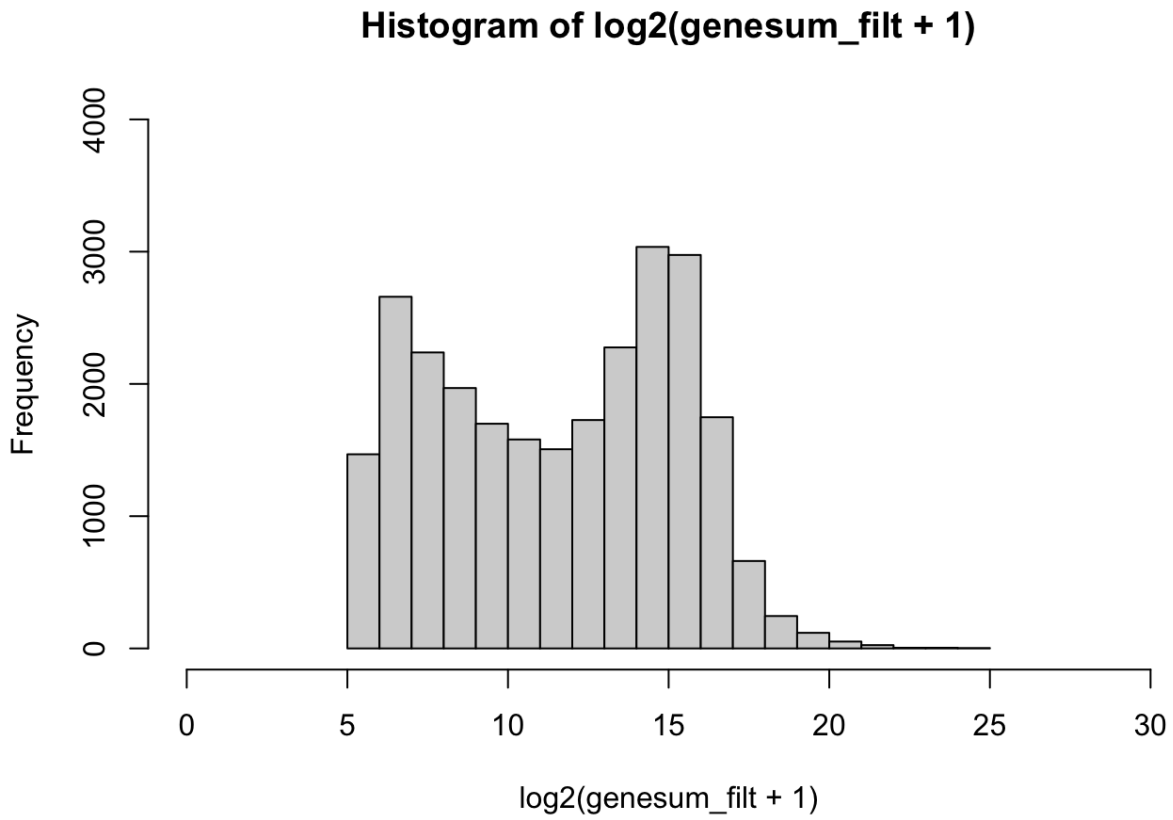
```
#For this filtering we wanted to keep the subset of data where the genesum was over 30
#genesum = 45 + 1 = 46 log2(46) = 5.52 = ~5.5, everything below 5 on the above graph, so we want to keep everything with a genesum count of 45 and above

data_filt = subset(data, genesum >45)
genesum = rowSums(data)
```

Take a look at the histogram again.

Is this the presence of 2 means forming in the data?

```
genesum_filt = rowSums(data_filt)
hist(log2(genesum_filt+1), ylim = c(0,4000), xlim = c(0,30), breaks = 25)
```




```
dim(data_filt)
```

```
## [1] 25995    45
```

Load limma and edgeR

```
library(limma)
```

```
##
## Attaching package: 'limma'
```

```
## The following object is masked from 'package:BiocGenerics':
##
##     plotMA
```

```
library(edgeR)
```

Create a design matrix for lm.

First we create the levels that we are potentially interested in. Some of these we may not use, but we have them in case we want to make a different comparison.

```
week_mm <- factor(pheno_df$week, levels = c("D0", "W1", "W2", "W3", "W4", "W5", "W6", "W8"))
treatments_mm <- factor(pheno_df$treatments, levels = c("Day0", "Placebo", "EGCG"))
status_mm <-factor(pheno_df$status, levels = c("uninjured", "injured"))
application_mm <-factor(pheno_df$application, levels = c("none", "zonal", "direct"))
appl_by_wk_mm <-factor(pheno_df$appl_by_wk, levels = c("W0", "Zw1", "Zw2", "Dw1", "Dw2", "Dw
3", "Dw4", "Dw6"))
```

This model matrix defines all of the things that you are interested in comparing It creates a matrix that defines the various experimental categories of the samples in your experiment that you want to compare. We will use the week_mm and treatments_mm as we are interested in comparing treatments across the weeks.

```
weektreat = factor(paste(week_mm,treatments_mm, sep=""))
weektreat
```

```
##  [1] D0Day0    D0Day0    D0Day0    W1EGCG    W1EGCG    W1EGCG    W1Placebo
##  [8] W1Placebo W1Placebo W2EGCG    W2EGCG    W2EGCG    W2Placebo W2Placebo
## [15] W2Placebo W3EGCG    W3EGCG    W3EGCG    W3Placebo W3Placebo W3Placebo
## [22] W4EGCG    W4EGCG    W4EGCG    W4Placebo W4Placebo W4Placebo W5EGCG
## [29] W5EGCG    W5EGCG    W5Placebo W5Placebo W5Placebo W6EGCG    W6EGCG
## [36] W6EGCG    W6Placebo W6Placebo W6Placebo W8EGCG    W8EGCG    W8EGCG
## [43] W8Placebo W8Placebo W8Placebo
## 15 Levels: D0Day0 W1EGCG W1Placebo W2EGCG W2Placebo W3EGCG W3Placebo ... W8Placebo
```

```
design = model.matrix(~0+weektreat)
design
```

```
##    weektreatD0Day0 weektreatW1EGCG weektreatW1Placebo weektreatW2EGCG
## 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
## 13               0                0                   0                0
## 14               0                0                   0                0
## 15               0                0                   0                0
## 16               0                0                   0                0
## 17               0                0                   0                0
## 18               0                0                   0                0
## 19               0                0                   0                0
## 20               0                0                   0                0
## 21               0                0                   0                0
## 22               0                0                   0                0
## 23               0                0                   0                0
## 24               0                0                   0                0
## 25               0                0                   0                0
## 26               0                0                   0                0
## 27               0                0                   0                0
## 28               0                0                   0                0
## 29               0                0                   0                0
## 30               0                0                   0                0
## 31               0                0                   0                0
## 32               0                0                   0                0
## 33               0                0                   0                0
## 34               0                0                   0                0
## 35               0                0                   0                0
## 36               0                0                   0                0
## 37               0                0                   0                0
## 38               0                0                   0                0
## 39               0                0                   0                0
## 40               0                0                   0                0
## 41               0                0                   0                0
## 42               0                0                   0                0
## 43               0                0                   0                0
## 44               0                0                   0                0
## 45               0                0                   0                0
##    weektreatW2Placebo weektreatW3EGCG weektreatW3Placebo weektreatW4EGCG
## 1                   0                0                   0                0
## 2                   0                0                   0                0
## 3                   0                0                   0                0
## 4                   0                0                   0                0
## 5                   0                0                   0                0
## 6                   0                0                   0                0
## 7                   0                0                   0                0
## 8                   0                0                   0                0
```

```
## 9                    0                0                    0                0
## 10                   0                0                    0                0
## 11                   0                0                    0                0
## 12                   0                0                    0                0
## 13                   1                0                    0                0
## 14                   1                0                    0                0
## 15                   1                0                    0                0
## 16                   0                1                    0                0
## 17                   0                1                    0                0
## 18                   0                1                    0                0
## 19                   0                0                    1                0
## 20                   0                0                    1                0
## 21                   0                0                    1                0
## 22                   0                0                    0                1
## 23                   0                0                    0                1
## 24                   0                0                    0                1
## 25                   0                0                    0                0
## 26                   0                0                    0                0
## 27                   0                0                    0                0
## 28                   0                0                    0                0
## 29                   0                0                    0                0
## 30                   0                0                    0                0
## 31                   0                0                    0                0
## 32                   0                0                    0                0
## 33                   0                0                    0                0
## 34                   0                0                    0                0
## 35                   0                0                    0                0
## 36                   0                0                    0                0
## 37                   0                0                    0                0
## 38                   0                0                    0                0
## 39                   0                0                    0                0
## 40                   0                0                    0                0
## 41                   0                0                    0                0
## 42                   0                0                    0                0
## 43                   0                0                    0                0
## 44                   0                0                    0                0
## 45                   0                0                    0                0
##    weektreatW4Placebo weektreatW5EGCG weektreatW5Placebo weektreatW6EGCG
## 1                    0                0                    0                0
## 2                    0                0                    0                0
## 3                    0                0                    0                0
## 4                    0                0                    0                0
## 5                    0                0                    0                0
## 6                    0                0                    0                0
## 7                    0                0                    0                0
## 8                    0                0                    0                0
## 9                    0                0                    0                0
## 10                   0                0                    0                0
## 11                   0                0                    0                0
## 12                   0                0                    0                0
## 13                   0                0                    0                0
## 14                   0                0                    0                0
## 15                   0                0                    0                0
## 16                   0                0                    0                0
## 17                   0                0                    0                0
## 18                   0                0                    0                0
```

```
## 19                  0                0                  0                0
## 20                  0                0                  0                0
## 21                  0                0                  0                0
## 22                  0                0                  0                0
## 23                  0                0                  0                0
## 24                  0                0                  0                0
## 25                  1                0                  0                0
## 26                  1                0                  0                0
## 27                  1                0                  0                0
## 28                  0                1                  0                0
## 29                  0                1                  0                0
## 30                  0                1                  0                0
## 31                  0                0                  1                0
## 32                  0                0                  1                0
## 33                  0                0                  1                0
## 34                  0                0                  0                1
## 35                  0                0                  0                1
## 36                  0                0                  0                1
## 37                  0                0                  0                0
## 38                  0                0                  0                0
## 39                  0                0                  0                0
## 40                  0                0                  0                0
## 41                  0                0                  0                0
## 42                  0                0                  0                0
## 43                  0                0                  0                0
## 44                  0                0                  0                0
## 45                  0                0                  0                0
##    weektreatW6Placebo weektreatW8EGCG weektreatW8Placebo
## 1                   0                0                  0
## 2                   0                0                  0
## 3                   0                0                  0
## 4                   0                0                  0
## 5                   0                0                  0
## 6                   0                0                  0
## 7                   0                0                  0
## 8                   0                0                  0
## 9                   0                0                  0
## 10                  0                0                  0
## 11                  0                0                  0
## 12                  0                0                  0
## 13                  0                0                  0
## 14                  0                0                  0
## 15                  0                0                  0
## 16                  0                0                  0
## 17                  0                0                  0
## 18                  0                0                  0
## 19                  0                0                  0
## 20                  0                0                  0
## 21                  0                0                  0
## 22                  0                0                  0
## 23                  0                0                  0
## 24                  0                0                  0
## 25                  0                0                  0
## 26                  0                0                  0
## 27                  0                0                  0
## 28                  0                0                  0
```

```
## 29                  0                 0                 0
## 30                  0                 0                 0
## 31                  0                 0                 0
## 32                  0                 0                 0
## 33                  0                 0                 0
## 34                  0                 0                 0
## 35                  0                 0                 0
## 36                  0                 0                 0
## 37                  1                 0                 0
## 38                  1                 0                 0
## 39                  1                 0                 0
## 40                  0                 1                 0
## 41                  0                 1                 0
## 42                  0                 1                 0
## 43                  0                 0                 1
## 44                  0                 0                 1
## 45                  0                 0                 1
## attr(,"assign")
##  [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## attr(,"contrasts")
## attr(,"contrasts")$weektreat
## [1] "contr.treatment"
```

```
colnames(design) = levels(weektreat)
design
```

```
##    D0Day0 W1EGCG W1Placebo W2EGCG W2Placebo W3EGCG W3Placebo W4EGCG W4Placebo
## 1       1       0          0       0          0       0          0       0          0
## 2       1       0          0       0          0       0          0       0          0
## 3       1       0          0       0          0       0          0       0          0
## 4       0       1          0       0          0       0          0       0          0
## 5       0       1          0       0          0       0          0       0          0
## 6       0       1          0       0          0       0          0       0          0
## 7       0       0          1       0          0       0          0       0          0
## 8       0       0          1       0          0       0          0       0          0
## 9       0       0          1       0          0       0          0       0          0
## 10      0       0          0       1          0       0          0       0          0
## 11      0       0          0       1          0       0          0       0          0
## 12      0       0          0       1          0       0          0       0          0
## 13      0       0          0       0          1       0          0       0          0
## 14      0       0          0       0          1       0          0       0          0
## 15      0       0          0       0          1       0          0       0          0
## 16      0       0          0       0          0       1          0       0          0
## 17      0       0          0       0          0       1          0       0          0
## 18      0       0          0       0          0       1          0       0          0
## 19      0       0          0       0          0       0          1       0          0
## 20      0       0          0       0          0       0          1       0          0
## 21      0       0          0       0          0       0          1       0          0
## 22      0       0          0       0          0       0          0       1          0
## 23      0       0          0       0          0       0          0       1          0
## 24      0       0          0       0          0       0          0       1          0
## 25      0       0          0       0          0       0          0       0          1
## 26      0       0          0       0          0       0          0       0          1
## 27      0       0          0       0          0       0          0       0          1
## 28      0       0          0       0          0       0          0       0          0
## 29      0       0          0       0          0       0          0       0          0
## 30      0       0          0       0          0       0          0       0          0
## 31      0       0          0       0          0       0          0       0          0
## 32      0       0          0       0          0       0          0       0          0
## 33      0       0          0       0          0       0          0       0          0
## 34      0       0          0       0          0       0          0       0          0
## 35      0       0          0       0          0       0          0       0          0
## 36      0       0          0       0          0       0          0       0          0
## 37      0       0          0       0          0       0          0       0          0
## 38      0       0          0       0          0       0          0       0          0
## 39      0       0          0       0          0       0          0       0          0
## 40      0       0          0       0          0       0          0       0          0
## 41      0       0          0       0          0       0          0       0          0
## 42      0       0          0       0          0       0          0       0          0
## 43      0       0          0       0          0       0          0       0          0
## 44      0       0          0       0          0       0          0       0          0
## 45      0       0          0       0          0       0          0       0          0
##    W5EGCG W5Placebo W6EGCG W6Placebo W8EGCG W8Placebo
## 1       0          0       0          0       0          0
## 2       0          0       0          0       0          0
## 3       0          0       0          0       0          0
## 4       0          0       0          0       0          0
## 5       0          0       0          0       0          0
## 6       0          0       0          0       0          0
## 7       0          0       0          0       0          0
## 8       0          0       0          0       0          0
```

```
## 9       0          0       0          0       0          0
## 10      0          0       0          0       0          0
## 11      0          0       0          0       0          0
## 12      0          0       0          0       0          0
## 13      0          0       0          0       0          0
## 14      0          0       0          0       0          0
## 15      0          0       0          0       0          0
## 16      0          0       0          0       0          0
## 17      0          0       0          0       0          0
## 18      0          0       0          0       0          0
## 19      0          0       0          0       0          0
## 20      0          0       0          0       0          0
## 21      0          0       0          0       0          0
## 22      0          0       0          0       0          0
## 23      0          0       0          0       0          0
## 24      0          0       0          0       0          0
## 25      0          0       0          0       0          0
## 26      0          0       0          0       0          0
## 27      0          0       0          0       0          0
## 28      1          0       0          0       0          0
## 29      1          0       0          0       0          0
## 30      1          0       0          0       0          0
## 31      0          1       0          0       0          0
## 32      0          1       0          0       0          0
## 33      0          1       0          0       0          0
## 34      0          0       1          0       0          0
## 35      0          0       1          0       0          0
## 36      0          0       1          0       0          0
## 37      0          0       0          1       0          0
## 38      0          0       0          1       0          0
## 39      0          0       0          1       0          0
## 40      0          0       0          0       1          0
## 41      0          0       0          0       1          0
## 42      0          0       0          0       1          0
## 43      0          0       0          0       0          1
## 44      0          0       0          0       0          1
## 45      0          0       0          0       0          1
## attr(,"assign")
##  [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## attr(,"contrasts")
## attr(,"contrasts")$weektreat
## [1] "contr.treatment"
```

What if we let LIMMA and edgeR select our low expressed genes for us?

how would that be different than the genesum cutoff we chose of 45?

```
dge = DGEList(counts = GSE124161_readcount)
dim(dge$counts) #before filtering
```

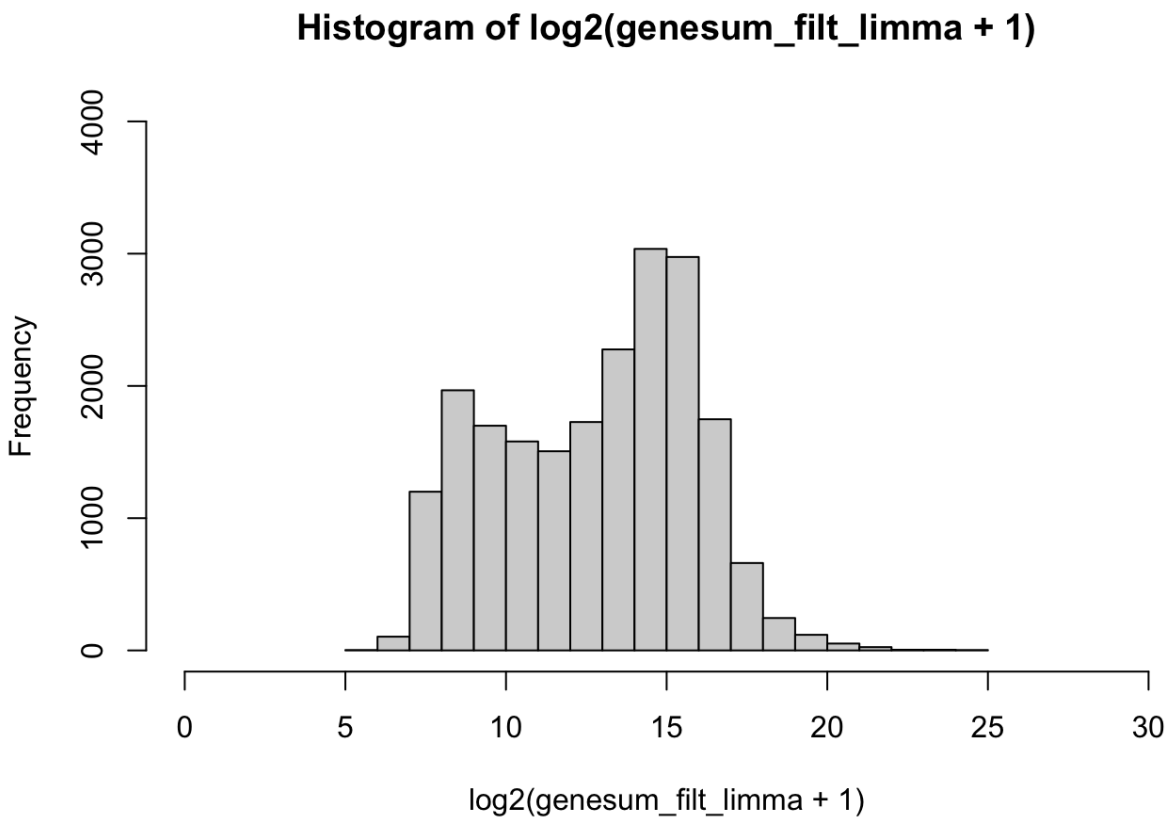
```
## [1] 48162    45
```

```
keep = filterByExpr(dge, design)
dge = dge[keep,,keep.lib.sizes=FALSE]

dim(dge$counts) #after filtering there are 20,935 genes, this is a lot stricter than the 25,99
5 genes we kept by filtering using genesum which was arbitrary and selected by judgement.
```

```
## [1] 20935    45
```

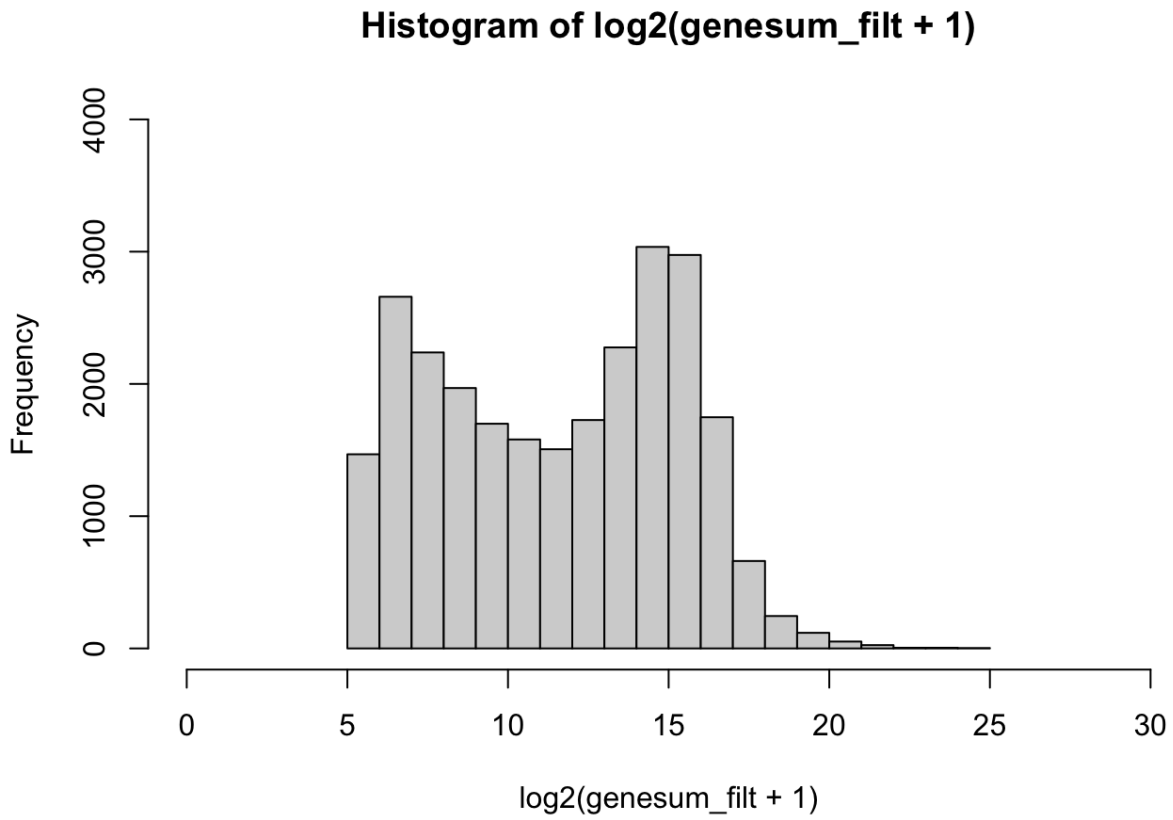
```
genesum_filt_limma = rowSums(dge$counts)
hist(log2(genesum_filt_limma+1), ylim = c(0,4000), xlim = c(0,30), breaks = 25)
```



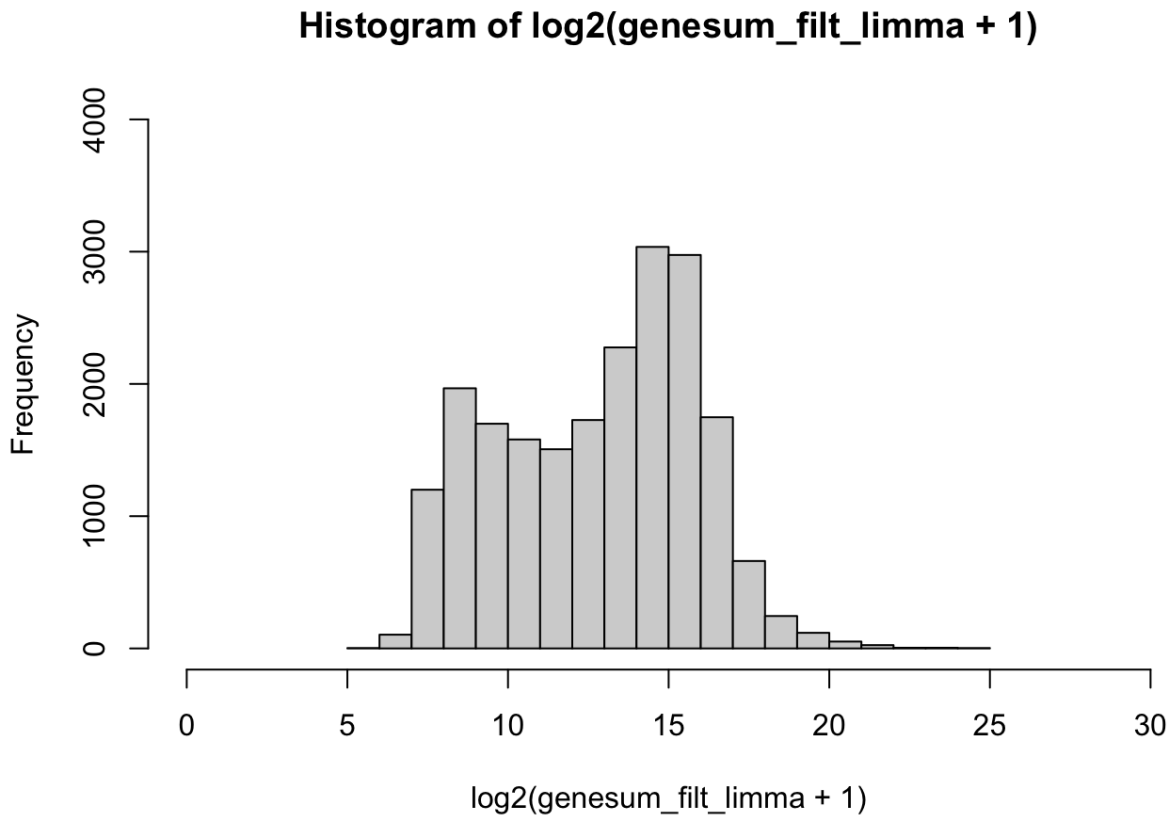
Merge of

plots “Low Counts Trimmed Data: Mannual vs Limma”

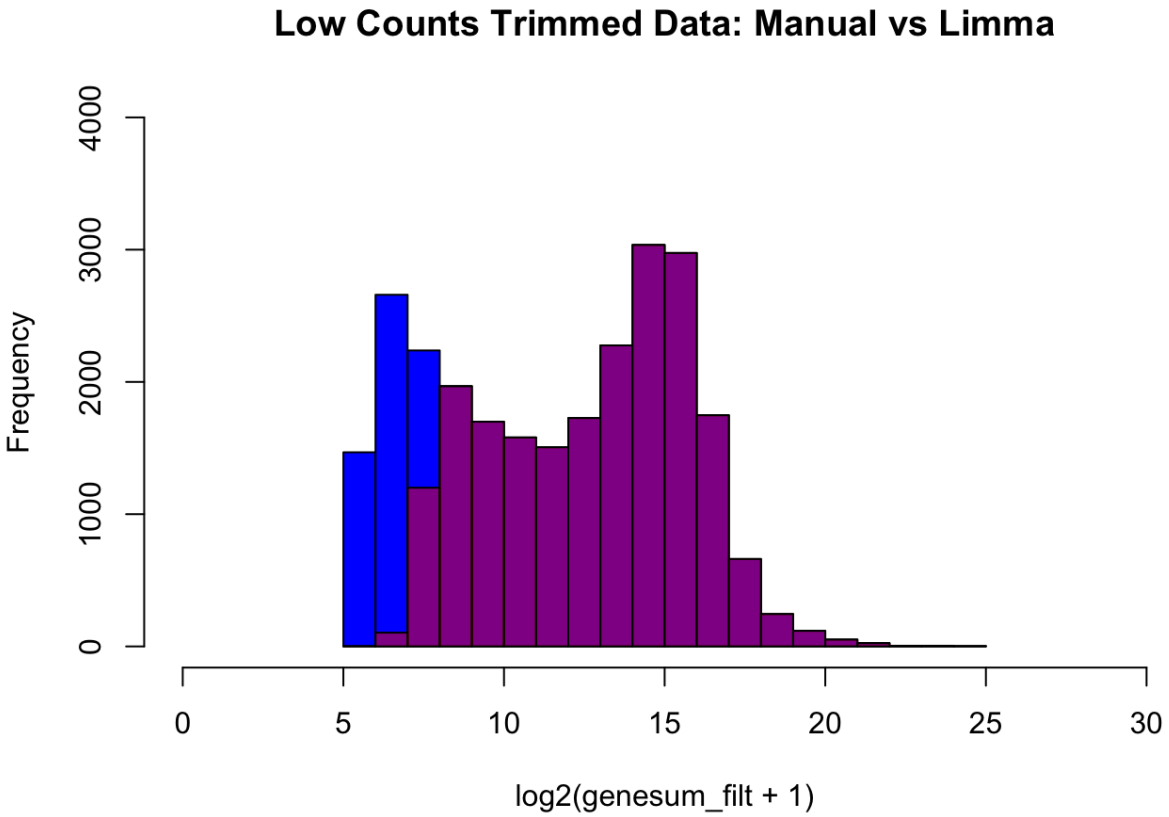
```
histfilt <- hist(log2(genesum_filt+1), ylim = c(0,4000), xlim = c(0,30), breaks = 25)
```




```
histlimma <-hist(log2(genesum_filt_limma+1), ylim = c(0,4000), xlim = c(0,30), breaks = 25)
```



```
plot(histfilt, ylim = c(0,4000), xlim = c(0,30), col= rgb(blue = 1, green=0, red=0, alpha =
1), main="Low Counts Trimmed Data: Manual vs Limma")
plot(histlimma, ylim = c(0,4000), xlim = c(0,30), col=rgb(red = 1, blue=0, green=0, alpha = 0.
5), add = TRUE) #note red is transparent and data is common to all "blue" so limma dataset pre
sents as purple.
```



Create a PCA plot, after low expressed genes are filtered out

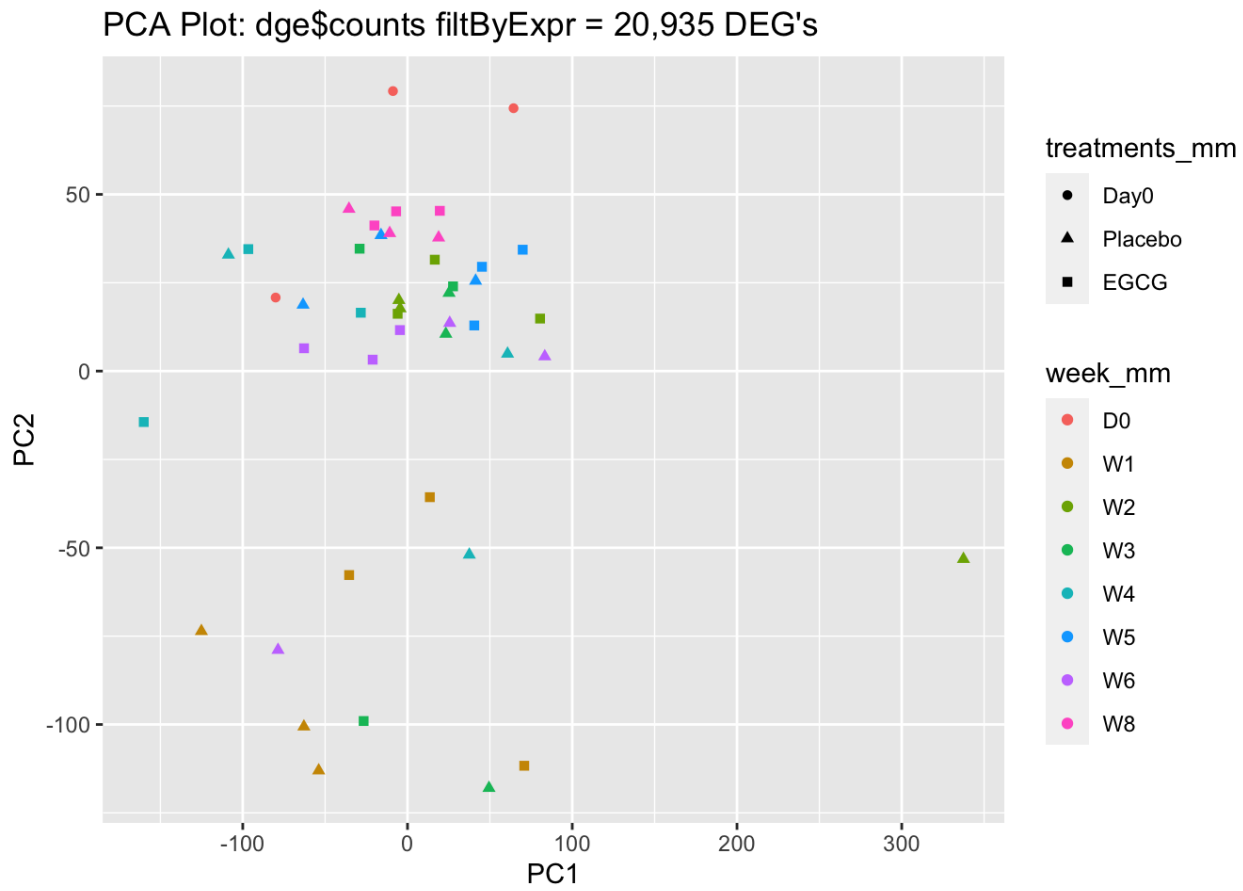
```
library(dplyr)

data_prcomp = prcomp(t(dge$counts), scale=TRUE, center=TRUE)

library(ggplot2)

coords2draw = as.data.frame(data_prcomp$x)

ggplot(coords2draw) +
  geom_point(mapping=aes(x = PC1, y= PC2,
                         col = week_mm, shape = treatments_mm)) +
  labs(title = "PCA Plot: dge$counts filtByExpr = 20,935 DEG's")
```



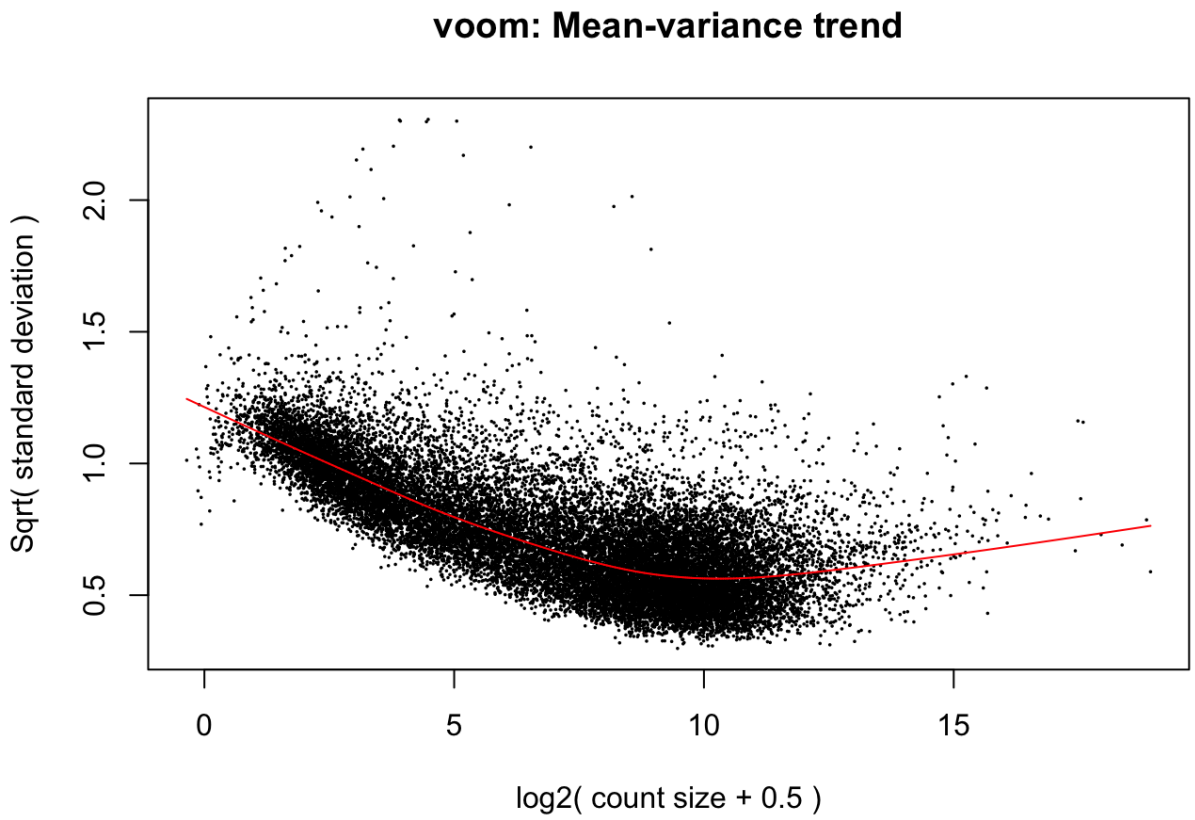
Now that we created a matrix that defines the various experimental categories of the samples, now we want to normalize the data. We need to normalize the data first before making the comparisons, as the normalized data is needed to proceed in next steps.

Voom normalization

Voom provides data in a format that can be used for standard limma methods. In the Limma manual, another normalization process is called “eset”, which is normalization through the AFFY package, however we are normalizing this data with “voom”, so “v” is the object we are storing the normalized data in.

Voom is the normalization method that allows us to use the data in downstream analyses. For this expression, voom is acting on the 20,935 genes captured in “dge”, using the design outlined

```
v = voom(dge, design, plot=TRUE, normalize="quantile")
```



Then create the lmfit (this calculates the “within” variance). This fits a linear model to the data.

```
nfit = lmFit(v,design)
```

Now specifically compare the different coefficients for the comparison

This gives us a lot more control to make the specific comparisons we want. This gives us a lot of control over a complex data set, one with a lot of levels, time-series data.

```
newcontrasts = makeContrasts(Zw1EGCG_vs_Zw1_placebo = W1EGCG - W1Placebo,#these are comparing the Treatment to the control at a single time point
                             Zw2EGCG_vs_Zw2_placebo = W2EGCG - W2Placebo,
                             Dw1_EGCG_vs_Dw1_placebo = W3EGCG - W3Placebo,
                             Dw2_EGCG_vs_Dw2_placebo  = W4EGCG - W4Placebo,
                             Dw3_EGCG_vs_Dw3_placebo  = W5EGCG - W5Placebo,
                             Dw4_EGCG_vs_Dw4_placebo  = W6EGCG - W6Placebo,
                             Dw6_EGCG_vs_Dw6_placebo  = W8EGCG - W8Placebo,
                             interact = (W1EGCG - W1Placebo) - (W2EGCG - W2Placebo), #Change in expression levels from Zonal week1 to wk2 differs between the ECGC-treated group and the placebo-treated group. If statistically significant, it would suggest that the change in expression levels over time (from week1 --> week 2) differs between the EGCG-treated group and the placebo-treated group.
                             interact2 = (W1EGCG - W1Placebo) - (W3EGCG - W3Placebo), #Change in expression levels for Zonal wk 1 to Direct wk 1. If statistically significant, it would suggest that the change in expression levels over time (from week1 --> week 2) differs between the EGCG-treated group and the placebo-treated group.
                             interact3 = (W2EGCG - W2Placebo) - (W4EGCG - W4Placebo), #Change in Expression levels for Zonal wk 2 and Direct wk 2
                             interact4 = W2EGCG - W1EGCG - W2Placebo + W1Placebo, #EGCG vs Placebo, Significant means that EGCG is having a statistically differential response between the two time points W1-W2
                             interact5 = W1EGCG + W1Placebo - W2EGCG - W2Placebo, #Global gene expression of week 1 vs week 2
                             levels = weektreat)
```

```
newcontrasts
```

```
##            Contrasts
## Levels      Zw1EGCG_vs_Zw1_placebo Zw2EGCG_vs_Zw2_placebo
##   D0Day0                        0                      0
##   W1EGCG                        1                      0
##   W1Placebo                    -1                      0
##   W2EGCG                        0                      1
##   W2Placebo                     0                     -1
##   W3EGCG                        0                      0
##   W3Placebo                     0                      0
##   W4EGCG                        0                      0
##   W4Placebo                     0                      0
##   W5EGCG                        0                      0
##   W5Placebo                     0                      0
##   W6EGCG                        0                      0
##   W6Placebo                     0                      0
##   W8EGCG                        0                      0
##   W8Placebo                     0                      0
##            Contrasts
## Levels      Dw1_EGCG_vs_Dw1_placebo Dw2_EGCG_vs_Dw2_placebo
##   D0Day0                         0                       0
##   W1EGCG                         0                       0
##   W1Placebo                      0                       0
##   W2EGCG                         0                       0
##   W2Placebo                      0                       0
##   W3EGCG                         1                       0
##   W3Placebo                     -1                       0
##   W4EGCG                         0                       1
##   W4Placebo                      0                      -1
##   W5EGCG                         0                       0
##   W5Placebo                      0                       0
##   W6EGCG                         0                       0
##   W6Placebo                      0                       0
##   W8EGCG                         0                       0
##   W8Placebo                      0                       0
##            Contrasts
## Levels      Dw3_EGCG_vs_Dw3_placebo Dw4_EGCG_vs_Dw4_placebo
##   D0Day0                         0                       0
##   W1EGCG                         0                       0
##   W1Placebo                      0                       0
##   W2EGCG                         0                       0
##   W2Placebo                      0                       0
##   W3EGCG                         0                       0
##   W3Placebo                      0                       0
##   W4EGCG                         0                       0
##   W4Placebo                      0                       0
##   W5EGCG                         1                       0
##   W5Placebo                     -1                       0
##   W6EGCG                         0                       1
##   W6Placebo                      0                      -1
##   W8EGCG                         0                       0
##   W8Placebo                      0                       0
##            Contrasts
## Levels      Dw6_EGCG_vs_Dw6_placebo interact interact2 interact3 interact4
##   D0Day0                         0        0         0         0         0
##   W1EGCG                         0        1         1         0        -1
```

```
##   W1Placebo                      0        -1         -1          0          1
##   W2EGCG                         0        -1          0          1          1
##   W2Placebo                      0         1          0         -1         -1
##   W3EGCG                         0         0         -1          0          0
##   W3Placebo                      0         0          1          0          0
##   W4EGCG                         0         0          0         -1          0
##   W4Placebo                      0         0          0          1          0
##   W5EGCG                         0         0          0          0          0
##   W5Placebo                      0         0          0          0          0
##   W6EGCG                         0         0          0          0          0
##   W6Placebo                      0         0          0          0          0
##   W8EGCG                         1         0          0          0          0
##   W8Placebo                     -1         0          0          0          0
##            Contrasts
## Levels      interact5
##   D0Day0             0
##   W1EGCG             1
##   W1Placebo          1
##   W2EGCG            -1
##   W2Placebo         -1
##   W3EGCG             0
##   W3Placebo          0
##   W4EGCG             0
##   W4Placebo          0
##   W5EGCG             0
##   W5Placebo          0
##   W6EGCG             0
##   W6Placebo          0
##   W8EGCG             0
##   W8Placebo          0
```

Fit the data to new contrasts and then calculate the p-value for each gene.

```
nfit2= contrasts.fit(nfit, newcontrasts)
nfit2 = eBayes(nfit2)
topTable(nfit2,  adjust="BH") #BH = one of the multiple hypothesis testing methods we talked a
bout the FDR correction.
```

	Zw1EGCG_vs_Zw1_placebo <dbl>	Zw2EGCG_vs_Zw2_placebo <dbl>	Dw1_EGCG_vs_Dw1_placebo <dbl>
ENSG00000140519	-1.6248092	0.51406004	1.0167165352
ENSG00000021355	-0.2395233	-0.01146786	-0.0999755754
ENSG00000171848	0.4620292	0.08929963	0.0575356387
ENSG00000183696	-0.5946972	0.94157598	-0.0009506325
ENSG00000163209	-0.9096414	0.75378615	-0.9568299708
ENSG00000189410	-0.2724934	0.93794466	0.3123970609

	Zw1EGCG_vs_Zw1_placebo <dbl>	Zw2EGCG_vs_Zw2_placebo <dbl>	Dw1_EGCG_vs_Dw1_placebo <dbl>
ENSG00000128965	-1.8278921	0.72243711	-0.0759112571
ENSG00000115602	-0.6573179	0.18092313	-0.0095954582
ENSG00000074317	-0.9754783	1.15879278	0.2942885039
ENSG00000106819	0.3898805	-0.46008831	-0.0828503291

1-10 of 10 rows | 1-4 of 17 columns

Coeff = Zw1EGCG_vs_Zw1_placebo

get details of specific coeff defined in the contrast. Selected contrast “Zw1EGCG_vs_Zw1_placebo”

topTable() is a function in limma which summarizes the results of the linear model, perform hypothesis tests, and adjust the p-values for multiple testing.Results include (log2) fold changes, standard errors, t-statistics and p-values. A number of summary statistics are presented by topTable() for the top genes and the selected contrast:
Zw1EGCG_vs_Zw1_placebo = W1EGCG - W1Placebo

```
topTable(nfit2, coef = "Zw1EGCG_vs_Zw1_placebo",  adjust="BH") #we want to specify a specific coefficient, we can look at the interaction of #Zonal wk 1 and Direct wk 1
```

	logFC <dbl>	AveExpr <dbl>	t <dbl>	P.Value <dbl>	adj.P.Val <dbl>	B <dbl>
ENSG00000206560	1.0411315	5.4347981	6.839136	6.856949e-08	0.001435502	8.110385
ENSG00000128965	-1.8278921	0.3528284	-6.469926	2.047805e-07	0.002143540	6.981398
ENSG00000138071	0.9487888	8.4230263	5.694576	2.089023e-06	0.009355054	4.882916
ENSG00000185338	-1.8444434	2.7058248	-5.627166	2.558631e-06	0.009355054	4.699070
ENSG00000136694	-2.2964641	-4.1871643	-5.991831	8.551580e-07	0.005967578	4.681823
ENSG00000158352	1.2260299	3.8522816	5.611616	2.681171e-06	0.009355054	4.652817
ENSG00000173762	-1.9961761	2.2048275	-5.464843	4.169653e-06	0.009883011	4.240381
ENSG00000138778	1.2587972	2.7599442	5.404928	4.993132e-06	0.009883011	4.079783
ENSG00000137710	0.7716775	5.7848368	5.391886	5.192888e-06	0.009883011	4.043962
ENSG00000244094	-2.6820388	-2.9183119	-5.418382	4.795107e-06	0.009883011	3.992881

1-10 of 10 rows

```
Top_10_genes <-as.data.frame(rownames(topTable(nfit2, coef = "Zw1EGCG_vs_Zw1_placebo",  adjust ="BH")))
Top_10_genes
```

rownames(topTable(nfit2, coef = "Zw1EGCG_vs_Zw1_placebo", adjust = "BH"))
<chr>
ENSG00000206560
ENSG00000128965
ENSG00000138071
ENSG00000185338
ENSG00000136694
ENSG00000158352
ENSG00000173762
ENSG00000138778
ENSG00000137710
ENSG00000244094

1-10 of 10 rows

```
colnames(Top_10_genes) <-c("GeneIDs")

#adding Gene SYMBOL
library("AnnotationDbi")
```

```
## Loading required package: stats4
```

```
## Loading required package: IRanges
```

```
## Loading required package: S4Vectors
```

```
##
## Attaching package: 'S4Vectors'
```

```
## The following objects are masked from 'package:lubridate':
##
##     second, second<-
```

```
## The following objects are masked from 'package:dplyr':
##
##     first, rename
```

```
## The following object is masked from 'package:tidyr':
##
##     expand
```

```
## The following objects are masked from 'package:base':
##
##     expand.grid, I, unname
```

```
##
## Attaching package: 'IRanges'
```

```
## The following object is masked from 'package:lubridate':
##
##     %within%
```

```
## The following objects are masked from 'package:dplyr':
##
##     collapse, desc, slice
```

```
## The following object is masked from 'package:purrr':
##
##     reduce
```

```
##
## Attaching package: 'AnnotationDbi'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
library(org.Hs.eg.db)
```

```
##
```

```
Top_10_genes$GeneSymbol = mapIds(org.Hs.eg.db,
                     keys=rownames(topTable(nfit2, coef = "Zw1EGCG_vs_Zw1_placebo",  adjust="B
H")), #Column containing Ensembl gene ids
                     column="SYMBOL",
                     keytype="ENSEMBL",
                     multiVals="first") #This selects the first gene alias, if there are multip
le gene names under the single EntrezID
```

```
## 'select()' returned 1:1 mapping between keys and columns
```

```
Top_10_genes
```

GeneIDs <chr>	GeneSymbol <chr>
ENSG00000206560	ANKRD28
ENSG00000128965	CHAC1

GeneIDs <chr>	GeneSymbol <chr>
ENSG00000138071	ACTR2
ENSG00000185338	SOCS1
ENSG00000136694	IL36A
ENSG00000158352	SHROOM4
ENSG00000173762	CD7
ENSG00000138778	CENPE
ENSG00000137710	RDX
ENSG00000244094	SPRR2F

1-10 of 10 rows

```
top_table <-as.data.frame(topTable(nfit2, coef = "Zw1EGCG_vs_Zw1_placebo",  adjust="BH"))
top_table
```

	logFC <dbl>	AveExpr <dbl>	t <dbl>	P.Value <dbl>	adj.P.Val <dbl>	B <dbl>
ENSG00000206560	1.0411315	5.4347981	6.839136	6.856949e-08	0.001435502	8.110385
ENSG00000128965	-1.8278921	0.3528284	-6.469926	2.047805e-07	0.002143540	6.981398
ENSG00000138071	0.9487888	8.4230263	5.694576	2.089023e-06	0.009355054	4.882916
ENSG00000185338	-1.8444434	2.7058248	-5.627166	2.558631e-06	0.009355054	4.699070
ENSG00000136694	-2.2964641	-4.1871643	-5.991831	8.551580e-07	0.005967578	4.681823
ENSG00000158352	1.2260299	3.8522816	5.611616	2.681171e-06	0.009355054	4.652817
ENSG00000173762	-1.9961761	2.2048275	-5.464843	4.169653e-06	0.009883011	4.240381
ENSG00000138778	1.2587972	2.7599442	5.404928	4.993132e-06	0.009883011	4.079783
ENSG00000137710	0.7716775	5.7848368	5.391886	5.192888e-06	0.009883011	4.043962
ENSG00000244094	-2.6820388	-2.9183119	-5.418382	4.795107e-06	0.009883011	3.992881

1-10 of 10 rows

```
Top_10_genes$logFC<-top_table$logFC
Top_10_genes$adj.p.val <-top_table$adj.P.Val
Top_10_genes
```

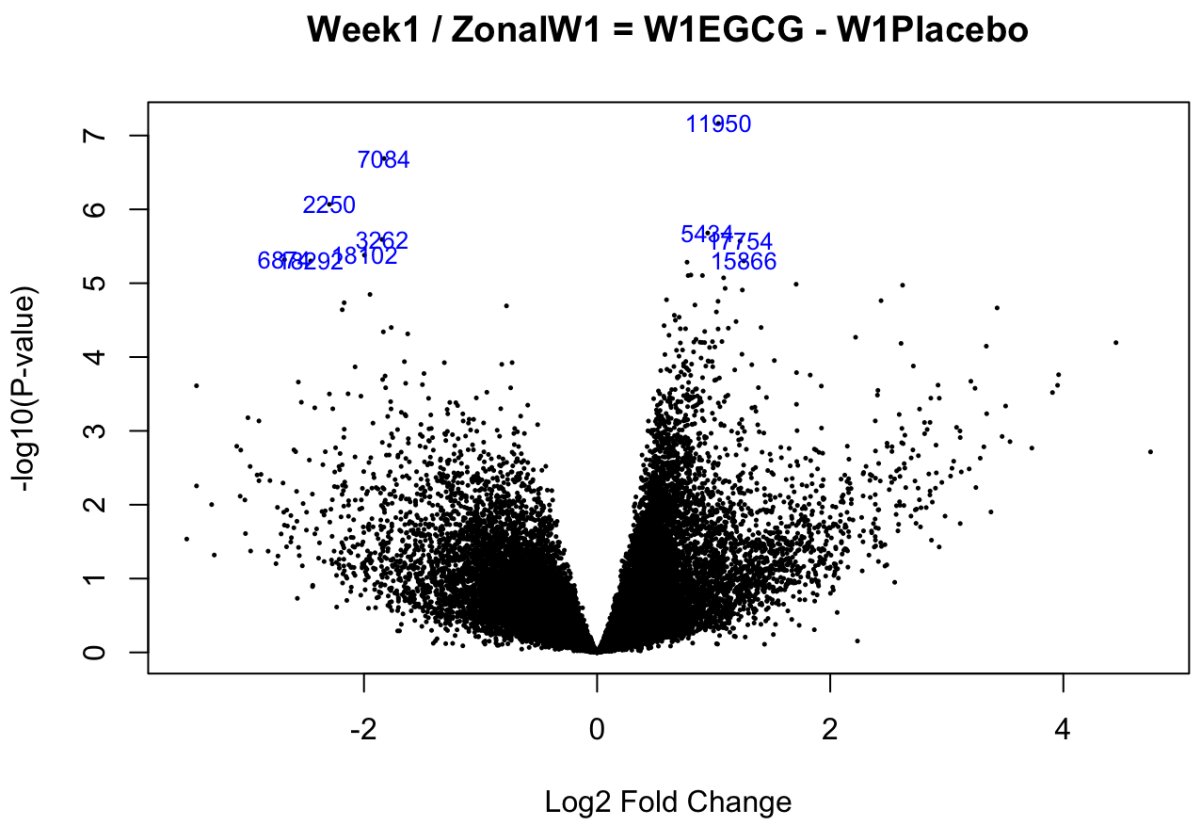
GeneIDs <chr>	GeneSymbol <chr>	logFC <dbl>	adj.p.val <dbl>
ENSG00000206560	ANKRD28	1.0411315	0.001435502
ENSG00000128965	CHAC1	-1.8278921	0.002143540

GeneIDs <chr>	GeneSymbol <chr>	logFC <dbl>	adj.p.val <dbl>
ENSG00000138071	ACTR2	0.9487888	0.009355054
ENSG00000185338	SOCS1	-1.8444434	0.009355054
ENSG00000136694	IL36A	-2.2964641	0.005967578
ENSG00000158352	SHROOM4	1.2260299	0.009355054
ENSG00000173762	CD7	-1.9961761	0.009883011
ENSG00000138778	CENPE	1.2587972	0.009883011
ENSG00000137710	RDX	0.7716775	0.009883011
ENSG00000244094	SPRR2F	-2.6820388	0.009883011

1-10 of 10 rows

Make a volcano plot of this data

```
volcanoplot(nfit2, "Zw1EGCG_vs_Zw1_placebo", highlight = 10, main = "Week1 / ZonalW1 = W1EGCG - W1Placebo") #The highlight=10 highlights the top 10, but gives the rownames... Changing to Ensembl gene name truncates the name, so it is not useful. However, we can capture the row name and get the gene name, but why bother, as the names are already in the top10 gene list above from the toptable() function. So see above ^)
```



Where are the normalized values from the Zoom normalization for all of the comparisons made in the dataset. ### Normalized values are stored in v$E

```
normexpvalues = v$E
dim(normexpvalues)
```

```
## [1] 20935    45
```

```
head(normexpvalues)
```

```
##                   GT01_D0    GT09_D0   GT19_D0  GT57_1_T  GT58_1_T  GT59_1_T
## ENSG00000123159  7.271149  7.4122624 8.0319992  6.987033  7.314820  7.098878
## ENSG00000233005 -1.235546 -0.5587764 0.4922798 -2.100866 -2.745624 -2.038246
## ENSG00000131242  5.666664  5.4804867 5.3694893  3.815836  4.336988  3.616872
## ENSG00000139168  5.134622  5.1683002 5.9914752  4.952089  5.118815  4.801069
## ENSG00000115541  2.893089  2.6645463 3.4987717  3.383017  3.165651  3.065886
## ENSG00000105486  4.717573  5.0625303 4.3603528  4.447360  4.814297  4.264415
##                  GT57_1_C  GT58_1_C  GT59_1_C   GT51_2_T  GT52_2_T  GT53_2_T
## ENSG00000123159  7.518985  7.684010  7.597574  6.9607315  7.261658  7.128385
## ENSG00000233005 -1.485629 -1.488546 -2.827593 -0.8579596 -1.345410 -1.195722
## ENSG00000131242  4.264950  2.849192  4.520628  4.5330654  4.342466  4.672421
## ENSG00000139168  4.827302  6.253074  4.654369  4.9533066  5.269447  4.875076
## ENSG00000115541  4.253053  5.025696  3.750854  2.3880567  3.052152  2.288165
## ENSG00000105486  4.273606  3.904416  4.233130  4.3465493  4.168258  4.423441
##                  GT51_2_C   GT52_2_C  GT53_2_C  GT45_3_T  GT46_3_T  GT47_3_T
## ENSG00000123159 7.3061832  7.3284764  6.961145  6.967745  7.288389  6.772299
## ENSG00000233005 0.2076713 -0.2889084 -2.032891 -1.941631 -1.101122 -1.509705
## ENSG00000131242 4.7960280  4.8491830  4.562683  3.794021  4.511481  3.548694
## ENSG00000139168 5.1067209  5.2165167  5.089568  4.833474  4.960957  5.778976
## ENSG00000115541 2.4973033  2.5490538  3.021289  2.534274  2.621711  4.571800
## ENSG00000105486 4.0003567  4.1640820  3.937719  3.937066  3.949419  4.026190
##                  GT45_3_C  GT46_3_C  GT47_3_C  GT17_4_T  GT18_4_T  GT19_4_T
## ENSG00000123159  7.098878  7.014791  6.528873 7.7103974  7.340138  7.341442
## ENSG00000233005 -1.779650 -1.905098 -2.262058 0.6500853 -1.509705 -1.862430
## ENSG00000131242  4.375879  4.206759  3.789523 3.9281970  4.786679  4.504358
## ENSG00000139168  5.038846  4.960489  5.702488 6.5497827  5.216517  4.944889
## ENSG00000115541  2.524548  2.909242  4.306894 4.1797316  2.504025  2.556141
## ENSG00000105486  3.984542  4.044198  4.230057 2.7321736  4.247714  4.249908
##                   GT17_4_C   GT18_4_C  GT19_4_C  GT39_5_T  GT40_5_T  GT41_5_T
## ENSG00000123159  7.6458680  7.5539732  7.442093  6.964603  7.354397  7.247190
## ENSG00000233005 -0.5900602 -0.7800433 -0.853282 -2.838380 -3.840517 -1.674588
## ENSG00000131242  3.9592267  4.7818214  5.126004  4.459279  4.398804  4.546997
## ENSG00000139168  5.7875683  5.4415423  5.021878  4.945479  4.684047  4.905898
## ENSG00000115541  3.3237744  2.4725512  2.682647  2.477482  3.143244  2.368212
## ENSG00000105486  3.3436448  4.3932545  4.612627  4.277105  4.225955  4.354620
##                  GT39_5_C  GT40_5_C   GT41_5_C  GT34_6_T  GT35_6_T  GT37_6_T
## ENSG00000123159  7.005782  7.101623  7.1745103  7.381414  7.048813  7.162279
## ENSG00000233005 -0.853282 -2.341753 -0.6415115 -2.077288 -1.146907 -1.320596
## ENSG00000131242  3.348706  4.118488  4.5110221  4.214178  4.067442  4.869286
## ENSG00000139168  5.569744  4.852133  4.9444328  5.032897  5.375121  5.465900
## ENSG00000115541  3.212673  2.332835  2.3253250  2.640732  2.825147  2.737686
## ENSG00000105486  4.223734  4.181322  4.1387828  4.193323  4.328932  4.112569
##                  GT34_6_C GT35_6_C  GT37_6_C  GT27_8_T  GT28_8_T  GT29_8_T
## ENSG00000123159  7.375818 7.469538  6.956143  6.991880  6.861242  6.612324
## ENSG00000233005 -2.182059 1.275623 -2.032891 -5.481118 -5.481118 -4.304903
## ENSG00000131242  5.073132 4.490853  4.758655  4.710662  4.344061  4.303919
## ENSG00000139168  4.980795 6.445494  5.427946  4.946310  5.131683  5.223881
## ENSG00000115541  2.706947 4.458751  2.419008  2.608850  2.498930  2.192418
## ENSG00000105486  4.786679 2.885908  4.092467  4.161710  4.073175  4.160029
##                  GT27_8_C  GT28_8_C  GT29_8_C
## ENSG00000123159  6.854044  6.726125  6.568899
## ENSG00000233005 -1.228285 -2.077288 -1.281308
## ENSG00000131242  4.690596  4.141276  4.222744
## ENSG00000139168  5.104816  5.132168  5.217227
```

```
## ENSG00000115541  2.354893  2.398867  2.812218
## ENSG00000105486  4.086042  3.890042  4.019271
```

Get the genes that have adjpvalue < 0.2 and absolute log2fc > 1.5

```
interact_sig = topTable(nfit2,
                coef = "Zw1EGCG_vs_Zw1_placebo",
                adjust="BH", #method used to adjust the p-values for multiple testing. Option
s, in increasing conservatism, include                                   "none", "BH", "BY", "ho
lm"
                p.value=0.2, #cutoff value for adjusted p-values. Only genes with lower p-valu
es are listed
                number=10000, #max number of genes to list
                sort.by = "P", #sort by p-value
                lfc=log2(1.5)) #log fold change cutoff, the minimum absolute log2-fold-change
required
```

Get the voom values for these genes.

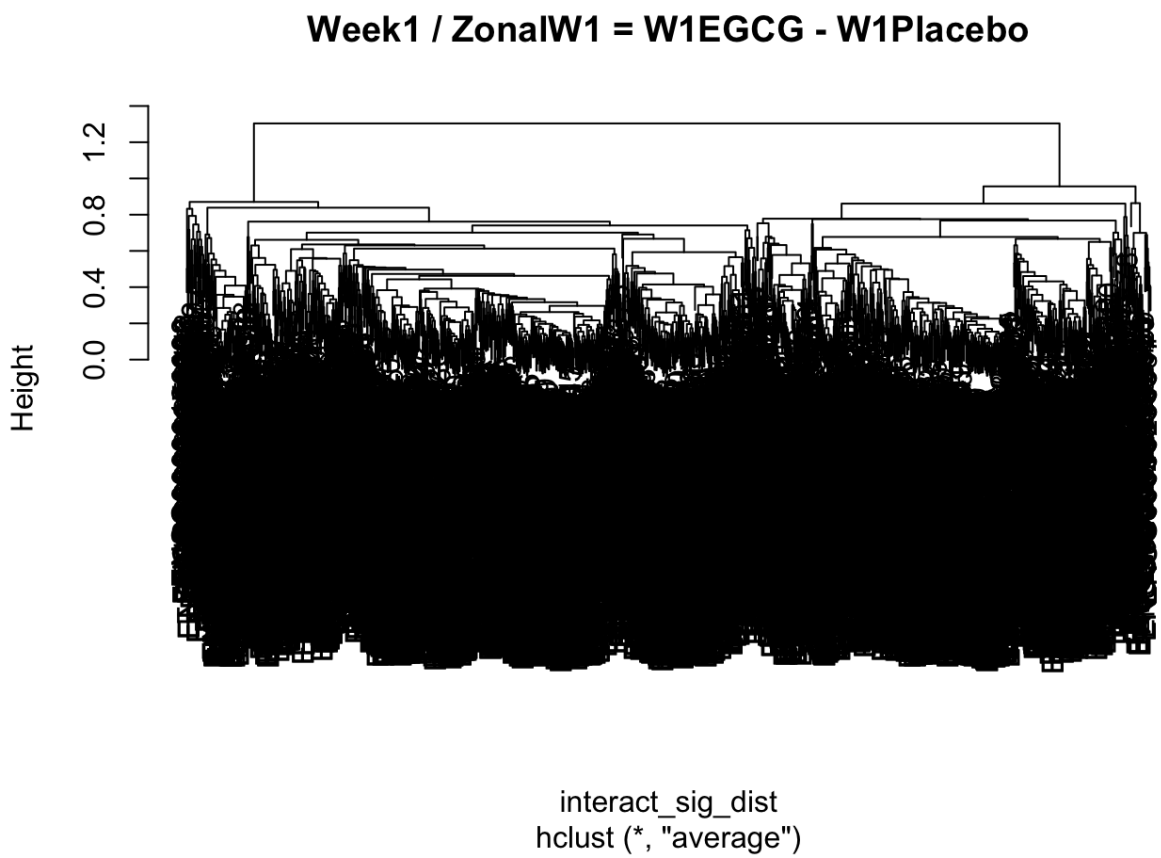
```
interact_sig_normvalues = normexpvalues[rownames(interact_sig),]
```

Calculate the distance using pairwise correlation of genes.

Use hclust to perform the clustering.

This is the interaction of Zw1EGCG_vs_Zw1_placebo = W1EGCG - W1Placebo

```
interact_sig_dist = as.dist(1 - cor(t(interact_sig_normvalues))) #this is correlation, not euc
lidean ()
interact_sig_hclust = hclust(interact_sig_dist, method="average")
plot(interact_sig_hclust, main = "Week1 / ZonalW1 = W1EGCG - W1Placebo")
```



Let's

determing the ideal number of cluster by calculating the avg silhouette width at each cutting.

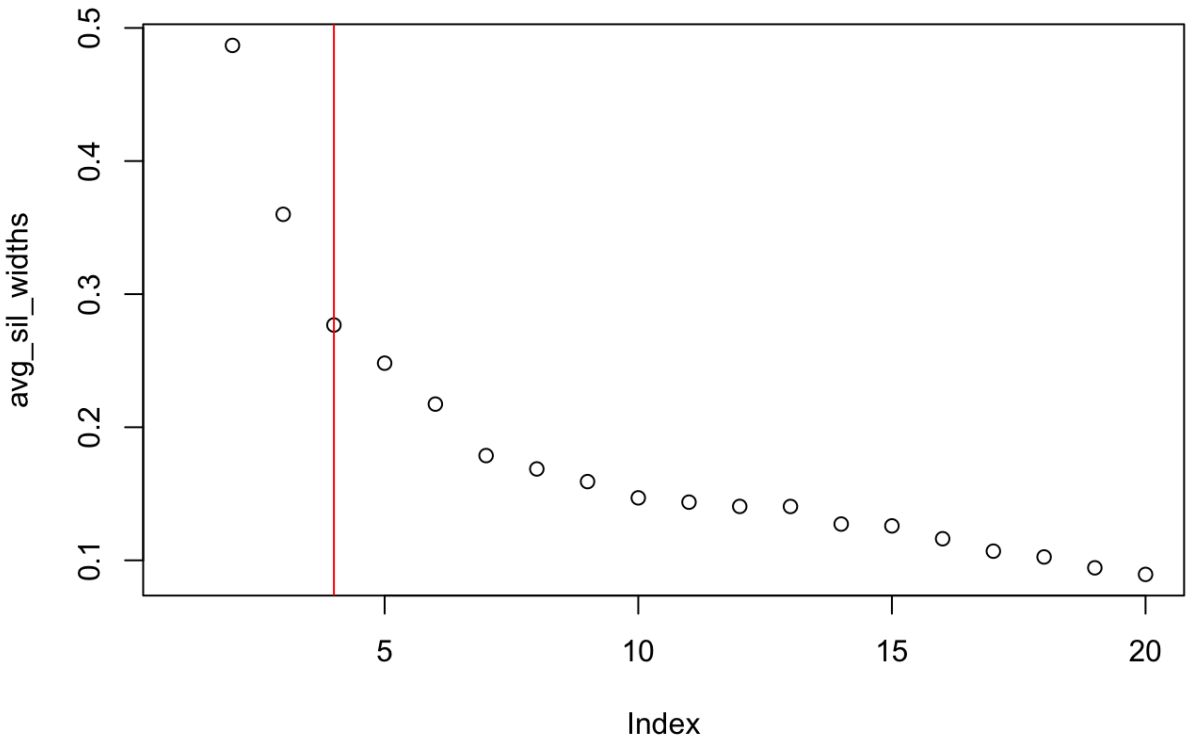
```
library(cluster)

avg_sil_widths = numeric()
for ( i in 2:20) {
  tempclust = cutree(interact_sig_hclust, k = i)
  avg_sil_widths[i] = mean(silhouette(tempclust, interact_sig_dist)[,"sil_width"])

}
```

2 & 4 looks promising. Let's go with 4 for now.

```
plot(avg_sil_widths)
abline(v=4, col="red")
```



Create

the groups. Notice the result is actually a vector of number and the gene names are the labels.

```
interact_sig_hclust_4 = cutree(interact_sig_hclust, k=4)
head(interact_sig_hclust_4)
```

```
## ENSG00000206560 ENSG00000128965 ENSG00000136694 ENSG00000138071 ENSG00000185338
##               1               2               2               1               2
## ENSG00000158352
##               1
```

To get the gene names that are in the different groups, use the which command to find out which genes are in the different groups, but then use the names function to get the actual names.

```
interact_sig_hclust_g1= normexpvalues[names(which(interact_sig_hclust_4==1)),]
interact_sig_hclust_g2= normexpvalues[names(which(interact_sig_hclust_4==2)),]
interact_sig_hclust_g3= normexpvalues[names(which(interact_sig_hclust_4==3)),]
interact_sig_hclust_g4= normexpvalues[names(which(interact_sig_hclust_4==4)),]
```

Create heatmap of each cluster group

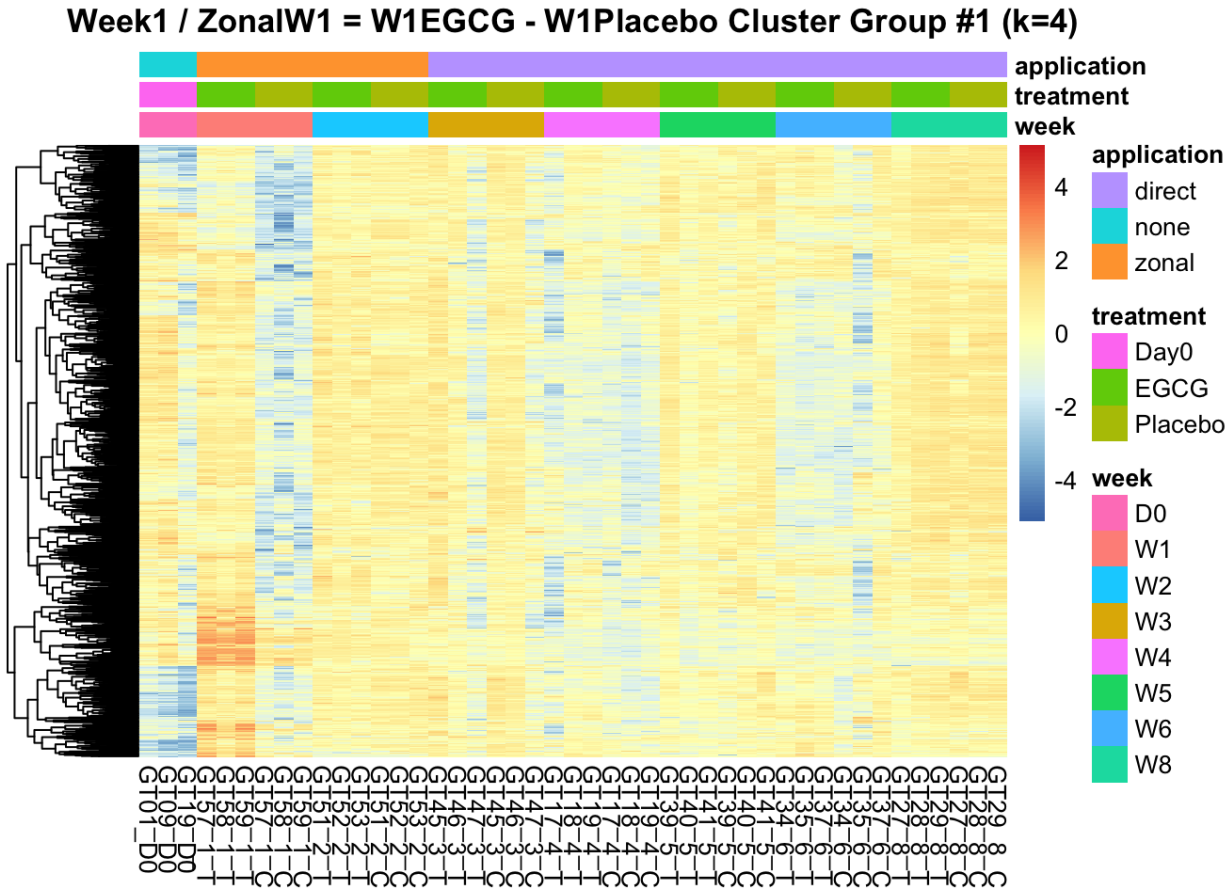
Cluster#1

Use “pheatmap” to draw cluster. “annot_col” defines how to create the legend. “scale” allows us to see the pattern for each gene. To make it easier to compare the different groups, I asked the columns not to be clustered “cluster_cols = F” , and to not show the gene names “show_rownames = F”.

```
library(pheatmap)

annotation <-as.data.frame(cbind(pheno_df$week, pheno_df$treatments, pheno_df$application))
colnames(annotation) <- c('week','treatment', 'application')
rownames(annotation) <-pheno_df$count_colnames

pheatmap(interact_sig_hclust_g1,annotation_col = annotation, scale="row", cluster_cols = F, sh
ow_rownames = F, main = "Week1 / ZonalW1 = W1EGCG - W1Placebo Cluster Group #1 (k=4)" )
```



Perform Go-Term Enrichment analysis

```
# Load the proper packages
library(GOstats)
```

```
## Loading required package: Category
```

```
## Loading required package: Matrix
```

```
##
## Attaching package: 'Matrix'
```

```
## The following object is masked from 'package:S4Vectors':
##
##     expand
```

```
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

```
## Loading required package: graph
```

```
##
## Attaching package: 'graph'
```

```
## The following object is masked from 'package:stringr':
##
##     boundary
```

```
##
```

```
##
## Attaching package: 'GOstats'
```

```
## The following object is masked from 'package:AnnotationDbi':
##
##     makeGOGraph
```

```
library(GO.db)
library(Category)
library(org.Hs.eg.db)
```

Go-Term Enrichment Part 1

Create HyperGoparpam

Converting the Ensemble to Entrez was achieved with this code: https://www.biostars.org/p/441386/ (https://www.biostars.org/p/441386/)

```
library("AnnotationDbi")

#adding ENTREZ ID's to global gene data file
GSE124161_readcount$entrez = mapIds(org.Hs.eg.db,
                     keys=rownames(GSE124161_readcount), #Column containing Ensembl gene ids
                     column="ENTREZID",
                     keytype="ENSEMBL",
                     multiVals="first") #This selects the first gene alias, if there are multip
le gene names under the single EntrezID
```

```
## 'select()' returned 1:many mapping between keys and columns
```

```
#Wrangling the ensemble gene ID's to Entrez in the interact_sig_hclust_g1
diffexpgenes_names_df <-rownames(as.data.frame(interact_sig_hclust_g1))
diffexpgenes_names_df <-as.data.frame(diffexpgenes_names_df)

diffexpgenes_names_df$entrez = mapIds(org.Hs.eg.db,
                    keys= diffexpgenes_names_df$diffexpgenes_names_df, #Column containing Ense
mbl gene ids
                    column="ENTREZID",
                    keytype="ENSEMBL",
                    multiVals="first") #This selects the first gene alias, if there are multip
le gene names under the single EntrezID
```

```
## 'select()' returned 1:many mapping between keys and columns
```

```
diffexpgenes_names <-diffexpgenes_names_df$entrez
readcount_names <-GSE124161_readcount$entrez

#Utilized following resource for below code format https://bioconductor.org/packages/release/b
ioc/vignettes/GOstats/inst/doc/GOstatsHyperG.pdf

params <- new("GOHyperGParams",
              geneIds = diffexpgenes_names, #don't use quotes here, it will not work, you will
get an error message. This is the variable name where you stored your differentially expressed
gene names
              universeGeneIds = readcount_names, #don't use quotes here, it will not work, you
will get an error message. This is the variable name where you stored all of the gene names fr
om the whole unfiltered data set. Its the whole list of the "universe" of gene IDs for your ar
ray or reference genome.
              annotation = "org.Hs.eg",
              ontology = "BP",
              pvalueCutoff=0.01, #don't use quotes here, it will not work, you will get an err
or message
              testDirection = "over")
```

```
## Warning in makeValidParams(.Object): removing duplicate IDs in geneIds
```

```
## Warning in makeValidParams(.Object): removing duplicate IDs in universeGeneIds
```

```
hypGO <- hyperGTest(params)
hypGO
```

```
## Gene to GO BP  test for over-representation
## 6340 GO BP ids tested (418 have p < 0.01)
## Selected gene set size: 693
##     Gene universe size: 17259
##     Annotation package: org.Hs.eg
```

The summary function returns a data.frame summarizing the result.

By default, only the results for terms with a p-value less than the cutoff specified in the parameter instance will be shown.You can also set a minimum number of genes for each term using the “categorySize” argument. I chose a grouping of 10.

```
sumGo <- summary(hypGO, categorySize =10)
sumGo
```

GOBPID <chr>	Pvalue <dbl>	OddsRatio <dbl>	ExpCount <dbl>	Count <int>	Size <int>
GO:0006996	3.996605e-13	1.877538	139.4110899	218	3472
GO:1903047	2.208067e-11	2.638362	30.4359465	71	758
GO:0007010	7.835833e-11	2.114681	58.9043977	110	1467
GO:0007049	8.975647e-11	2.017920	70.4282983	125	1754
GO:0022402	1.537748e-10	2.205668	48.5047801	95	1208
GO:0000278	3.879509e-10	2.360894	36.5391969	77	910
GO:0071840	1.096805e-09	1.601884	255.9349904	332	6374
GO:0016043	5.836788e-09	1.570268	247.8642447	320	6173
GO:0010564	8.069818e-09	2.421487	27.9464627	61	696
GO:0051301	8.303734e-09	2.501457	25.2963671	57	630

1-10 of 372 rows | 1-6 of 7 columns Previous **1** 2 3 4 5 6 … 38 Next

```
GoPlot <- data.frame(sumGo$GOBPID,sumGo$Pvalue,sumGo$Term)
colnames(GoPlot) <-c("GO_ID_BP", "P-value", "Term")
GoPlot
```

GO_ID_BP <chr>	P-value <dbl>	Term <chr>
GO:0006996	3.996605e-13	organelle organization
GO:1903047	2.208067e-11	mitotic cell cycle process
GO:0007010	7.835833e-11	cytoskeleton organization
GO:0007049	8.975647e-11	cell cycle
GO:0022402	1.537748e-10	cell cycle process
GO:0000278	3.879509e-10	mitotic cell cycle
GO:0071840	1.096805e-09	cellular component organization or biogenesis
GO:0016043	5.836788e-09	cellular component organization
GO:0010564	8.069818e-09	regulation of cell cycle process
GO:0051301	8.303734e-09	cell division

KEGG ENRIGHMENT Part1

```
#install Libraries needed for KEGG Enrichment Analysis
library(clusterProfiler)
```

```
## clusterProfiler v4.6.2  For help: https://yulab-smu.top/biomedical-knowledge-mining-book/
##
## If you use clusterProfiler in published research, please cite:
## T Wu, E Hu, S Xu, M Chen, P Guo, Z Dai, T Feng, L Zhou, W Tang, L Zhan, X Fu, S Liu, X Bo,
and G Yu. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The In
novation. 2021, 2(3):100141
```

```
##
## Attaching package: 'clusterProfiler'
```

```
## The following object is masked from 'package:AnnotationDbi':
##
##     select
```

```
## The following object is masked from 'package:IRanges':
##
##     slice
```

```
## The following object is masked from 'package:S4Vectors':
##
##     rename
```

```
## The following object is masked from 'package:purrr':
##
##     simplify
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
library(pathview)
```

```
## ##############################################################################
## Pathview is an open source software package distributed under GNU General
## Public License version 3 (GPLv3). Details of GPLv3 is available at
## http://www.gnu.org/licenses/gpl-3.0.html. Particullary, users are required to
## formally cite the original Pathview paper (not just mention it) in publications
## or products. For details, do citation("pathview") within R.
##
## The pathview downloads and uses KEGG data. Non-academic uses may require a KEGG
## license agreement (details at http://www.kegg.jp/kegg/legal.html).
## ##############################################################################
```

```
library(gage)
library(gageData)

#Now perform KEGG ENRICHMENT

keggEnrich <- enrichKEGG(
  diffexpgenes_names_df$entrez,
  organism = "hsa",
  keyType = "kegg",
  pvalueCutoff = 0.05, #adjust this if you are not seeing any results
  pAdjustMethod = "BH",
  )
```

```
## Reading KEGG annotation online: "https://rest.kegg.jp/link/hsa/pathway"...
```

```
## Reading KEGG annotation online: "https://rest.kegg.jp/list/pathway/hsa"...
```

```
#Show results from enrichKEGG
head(keggEnrich)
```

	ID	Description	GeneRatio	BgRatio	pvalue
	<chr>	<chr>	<chr>	<chr>	<dbl>
hsa04810	hsa04810	Regulation of actin cytoskeleton	26/313	229/8393	3.514036e-07
hsa05205	hsa05205	Proteoglycans in cancer	23/313	205/8393	2.100642e-06
hsa04520	hsa04520	Adherens junction	13/313	93/8393	3.709343e-05
hsa04218	hsa04218	Cellular senescence	17/313	156/8393	6.682878e-05
hsa05100	hsa05100	Bacterial invasion of epithelial cells	11/313	77/8393	1.204430e-04
hsa04510	hsa04510	Focal adhesion	19/313	203/8393	1.974970e-04

6 rows | 1-6 of 10 columns

```
keggEnrich
```

```
## #
## # over-representation test
## #
## #...@organism     hsa
## #...@ontology     KEGG
## #...@keytype      kegg
## #...@gene     chr [1:812] "23243" "10097" "57477" "1062" "5962" "57590" "79801" "950" ...
## #...pvalues adjusted by 'BH' with cutoff <0.05
## #...21 enriched terms found
## 'data.frame':    21 obs. of  9 variables:
##  $ ID         : chr  "hsa04810" "hsa05205" "hsa04520" "hsa04218" ...
##  $ Description: chr  "Regulation of actin cytoskeleton" "Proteoglycans in cancer" "Adherens
junction" "Cellular senescence" ...
##  $ GeneRatio  : chr  "26/313" "23/313" "13/313" "17/313" ...
##  $ BgRatio    : chr  "229/8393" "205/8393" "93/8393" "156/8393" ...
##  $ pvalue     : num  3.51e-07 2.10e-06 3.71e-05 6.68e-05 1.20e-04 ...
##  $ p.adjust   : num  9.59e-05 2.87e-04 3.38e-03 4.56e-03 6.58e-03 ...
##  $ qvalue     : num  8.03e-05 2.40e-04 2.82e-03 3.82e-03 5.50e-03 ...
##  $ geneID     : chr  "10097/5962/1398/4659/6093/3688/81624/9475/3685/23365/3680/10787/3845/
3690/10000/200576/5295/2247/128239/8503/36"| __truncated__ "5962/4659/5781/6093/3688/9475/368
5/23365/51196/1278/3845/3690/10000/3091/1499/5295/2247/1634/8503/3236/71/2335/867" "6093/9475/
9411/57493/7048/5787/2241/5797/1499/25945/889/71/7046" "824/2113/7048/3845/10000/5054/5295/89
0/10111/472/204851/7248/8503/5934/9134/7046/54822" ...
##  $ Count      : int  26 23 13 17 11 19 10 12 10 11 ...
## #...Citation
##  T Wu, E Hu, S Xu, M Chen, P Guo, Z Dai, T Feng, L Zhou, W Tang, L Zhan, X Fu, S Liu, X Bo,
and G Yu.
##  clusterProfiler 4.0: A universal enrichment tool for interpreting omics data.
##  The Innovation. 2021, 2(3):100141
```

```
#Generate a graph for the two KEGG results
#Edit the pathway id to that which is appropriate based on the ID column from the enrichKEGG o
utput

#These will generate images that will be saved to the working directory or the downloads folde
r
#Repeat for however many results you get from keggEnrich

pv.out_htmp1a <- pathview(gene.data = diffexpgenes_names_df$entrez, pathway.id = "hsa04810", s
pecies = "hsa")
```

```
## 'select()' returned 1:1 mapping between keys and columns
```

```
## Info: Working in directory /Users/dawndestefano/NYU/BIGY-7633 Transcriptomics/project
```

```
## Info: Writing image file hsa04810.pathview.png
```

```
#Repeat for the second result
pv.out_htmp1b <- pathview(gene.data = diffexpgenes_names_df$entrez, pathway.id = "hsa05205", s
pecies = "hsa")
```

```
## 'select()' returned 1:1 mapping between keys and columns
```

```
## Info: Working in directory /Users/dawndestefano/NYU/BIGY-7633 Transcriptomics/project
```

```
## Info: Writing image file hsa05205.pathview.png
```

```
#Also show the genes involved in the pathway
#These correspond to the elements included in the image of the KEGG pathway generated earlier
pv.out_htmp1a$plot.data.gene
```

	kegg.names <chr>	labels <chr>	all.mapped <chr>	ty… <chr>	x <dbl>	y <dbl>	width <dbl>	height <dbl>	mol.data <dbl>
35	2147	F2		gene	460	102	46	17	*NA*
40	7114	TMSB4X		gene	1128	464	46	17	*NA*
41	10097	ACTR2	10097	gene	1037	351	46	17	1
42	5216	PFN1		gene	976	505	46	17	*NA*
43	10163	WASF2		gene	894	526	46	17	*NA*
44	8936	WASF1		gene	894	474	46	17	*NA*
45	55845	BRK1		gene	894	457	46	17	*NA*
46	324	APC		gene	940	416	46	17	*NA*
47	8976	WASL		gene	894	374	46	17	*NA*
48	50649	ARHGEF4		gene	830	416	46	17	*NA*

1-10 of 79 rows | 1-10 of 11 columns Previous **1** 2 3 4 5 6 … 8 Next

```
pv.out_htmp1b$plot.data.gene
```

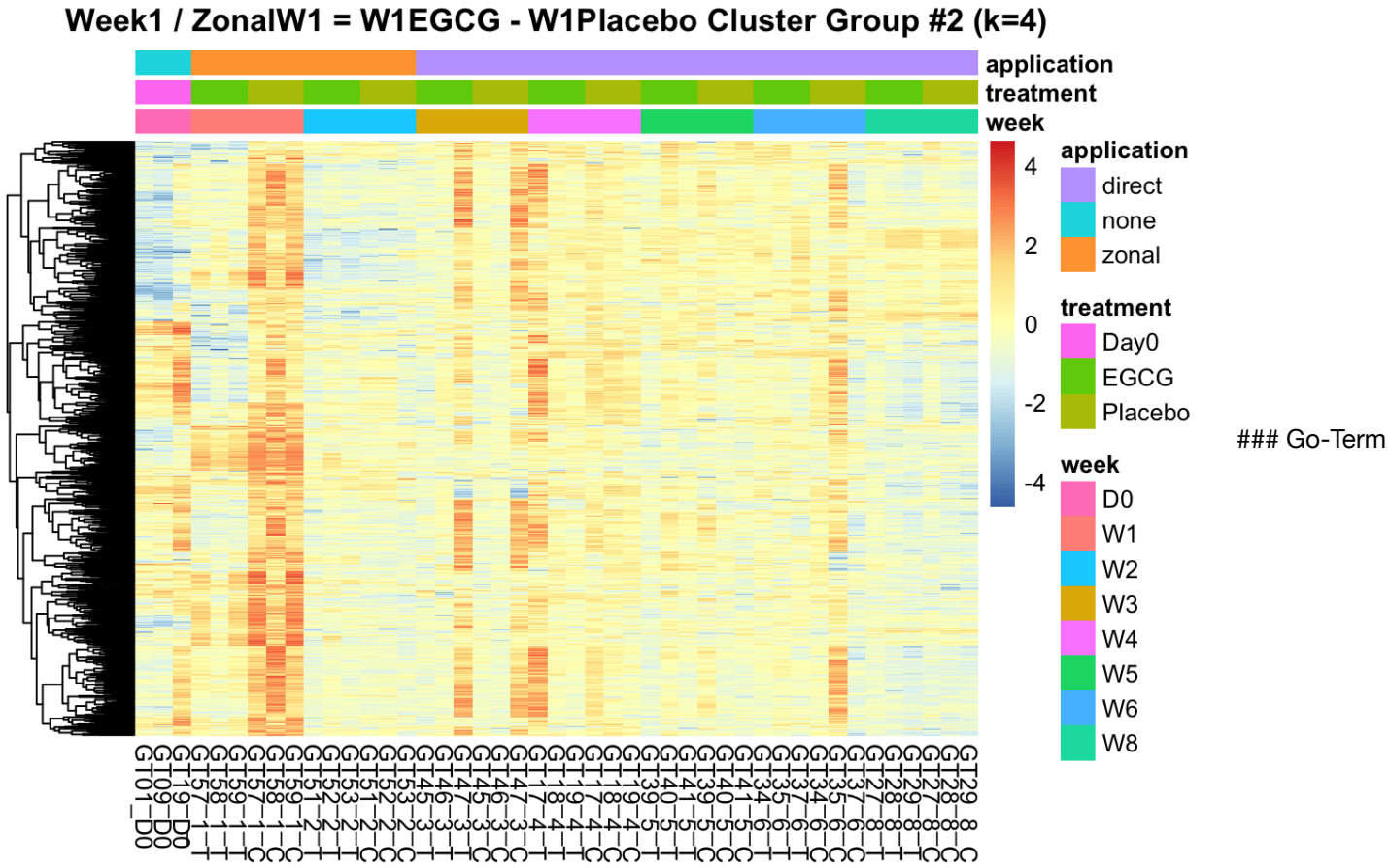
	kegg.names <chr>	labels <chr>	all.mapped <chr>	type <chr>	x <dbl>	y <dbl>	width <dbl>	height <dbl>	mol.data <dbl>
25	960	CD44		gene	198	488	46	17	*NA*
26	7074	TIAM1		gene	294	392	46	17	*NA*
28	2064	ERBB2		gene	198	209	46	17	*NA*
31	387	RHOA		gene	393	488	46	17	*NA*
33	6093	ROCK1	6093,9475	gene	491	487	46	17	2
36	286	ANK1		gene	294	607	46	17	*NA*
39	5295	PIK3R1	5295,8503	gene	684	445	46	17	2
41	10000	AKT3	10000	gene	783	444	46	17	1
47	4659	PPP1R12A	4659	gene	590	531	46	17	1

	kegg.names <chr>	labels <chr>	all.mapped <chr>	type <chr>	x <dbl>	y <dbl>	width <dbl>	height <dbl>	mol.data <dbl>
48	960	CD44		gene	198	607	46	17	*NA*

1-10 of 180 rows | 1-10 of 11 columns Previous **1** 2 3 4 5 6 … 18 Next

Cluster#2

```
pheatmap(interact_sig_hclust_g2,annotation_col = annotation, scale="row", cluster_cols = F, sh
ow_rownames = F, main = "Week1 / ZonalW1 = W1EGCG - W1Placebo Cluster Group #2 (k=4)" )
```



Go-Term

Enrichment Part 2

Create HyperGoparpam

Converting the Ensemble to Entrez was achieved with this code: https://www.biostars.org/p/441386/ (https://www.biostars.org/p/441386/)

```
library("AnnotationDbi")

#adding ENTREZ ID's to global gene data file
GSE124161_readcount$entrez = mapIds(org.Hs.eg.db,
                    keys=rownames(GSE124161_readcount), #Column containing Ensembl gene ids
                    column="ENTREZID",
                    keytype="ENSEMBL",
                    multiVals="first") #This selects the first gene alias, if there are multip
le gene names under the single EntrezID
```

```
## 'select()' returned 1:many mapping between keys and columns
```

```
#Wrangling the ensemble gene ID's to Entrez in the interact_sig_hclust_g2
diffexpgenes_names_df <-rownames(as.data.frame(interact_sig_hclust_g2))
diffexpgenes_names_df <-as.data.frame(diffexpgenes_names_df)

diffexpgenes_names_df$entrez = mapIds(org.Hs.eg.db,
                    keys= diffexpgenes_names_df$diffexpgenes_names_df, #Column containing Ense
mbl gene ids
                    column="ENTREZID",
                    keytype="ENSEMBL",
                    multiVals="first") #This selects the first gene alias, if there are multip
le gene names under the single EntrezID
```

```
## 'select()' returned 1:many mapping between keys and columns
```

```
diffexpgenes_names <-diffexpgenes_names_df$entrez
readcount_names <-GSE124161_readcount$entrez

#Utilized following resource for below code format https://bioconductor.org/packages/release/b
ioc/vignettes/GOstats/inst/doc/GOstatsHyperG.pdf

params <- new("GOHyperGParams",
              geneIds = diffexpgenes_names, #don't use quotes here, it will not work, you will
get an error message. This is the variable name where you stored your differentially expressed
gene names
              universeGeneIds = readcount_names, #don't use quotes here, it will not work, you
will get an error message. This is the variable name where you stored all of the gene names fr
om the whole unfiltered data set. Its the whole list of the "universe" of gene IDs for your ar
ray or reference genome.
              annotation = "org.Hs.eg",
              ontology = "BP",
              pvalueCutoff=0.01, #don't use quotes here, it will not work, you will get an err
or message
              testDirection = "over")
```

```
## Warning in makeValidParams(.Object): removing duplicate IDs in geneIds
```

```
## Warning in makeValidParams(.Object): removing duplicate IDs in universeGeneIds
```

```
hypGO <- hyperGTest(params)
hypGO
```

```
## Gene to GO BP  test for over-representation
## 5085 GO BP ids tested (365 have p < 0.01)
## Selected gene set size: 514
##     Gene universe size: 17259
##     Annotation package: org.Hs.eg
```

```
sumGo <- summary(hypGO, categorySize =10)
sumGo
```

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size
<chr>	<dbl>	<dbl>	<dbl>	<int>	<int>
GO:0002376	1.354829e-18	2.484911	81.5121386	160	2737
GO:0006955	2.600013e-16	2.602992	54.7385132	118	1838
GO:0042110	4.771148e-14	3.797650	15.9033548	52	534
GO:0002682	2.124595e-13	2.562577	42.8556695	94	1439
GO:0045321	1.545714e-12	2.828299	28.8583348	71	969
GO:0046649	2.813807e-12	2.983277	24.1826294	63	812
GO:0002250	1.340786e-10	2.965241	20.1918999	53	678
GO:0002684	2.153314e-10	2.633432	27.4586013	64	922
GO:0046631	2.440852e-10	5.568577	5.0628658	24	170
GO:0001775	2.867738e-10	2.467473	33.0277536	72	1109

1-10 of 330 rows | 1-6 of 7 columns Previous **1** 2 3 4 5 6 … 33 Next

```
GoPlot <- data.frame(sumGo$GOBPID,sumGo$Pvalue,sumGo$Term)
colnames(GoPlot) <-c("GO_ID_BP", "P-value", "Term")
GoPlot
```

GO_ID_BP	P-value	Term
<chr>	<dbl>	<chr>
GO:0002376	1.354829e-18	immune system process
GO:0006955	2.600013e-16	immune response
GO:0042110	4.771148e-14	T cell activation
GO:0002682	2.124595e-13	regulation of immune system process
GO:0045321	1.545714e-12	leukocyte activation
GO:0046649	2.813807e-12	lymphocyte activation

GO_ID_BP	P-value	Term
<chr>	<dbl>	<chr>
GO:0002250	1.340786e-10	adaptive immune response
GO:0002684	2.153314e-10	positive regulation of immune system process
GO:0046631	2.440852e-10	alpha-beta T cell activation
GO:0001775	2.867738e-10	cell activation

1-10 of 330 rows Previous **1** 2 3 4 5 6 … 33 Next

KEGG ENRIGHMENT Part2

```
#Now perform KEGG ENRICHMENT

keggEnrich <- enrichKEGG(
  diffexpgenes_names_df$entrez,
  organism = "hsa",
  keyType = "kegg",
  pvalueCutoff = 0.05, #adjust this if you are not seeing any results
  pAdjustMethod = "BH",
  )
```

```
#Show results from enrichKEGG
head(keggEnrich)
```

	ID	Description	GeneRa
	<chr>	<chr>	<chr>
hsa04060	hsa04060	Cytokine-cytokine receptor interaction	25/239
hsa04660	hsa04660	T cell receptor signaling pathway	14/239
hsa04061	hsa04061	Viral protein interaction with cytokine and cytokine receptor	13/239
hsa05340	hsa05340	Primary immunodeficiency	8/239
hsa04640	hsa04640	Hematopoietic cell lineage	11/239
hsa04064	hsa04064	NF-kappa B signaling pathway	11/239

6 rows | 1-4 of 10 columns

```
keggEnrich
```

```
## #
## # over-representation test
## #
## #...@organism     hsa
## #...@ontology     KEGG
## #...@keytype      kegg
## #...@gene     chr [1:610] "79094" "27179" "8651" "924" "6705" "1668" NA "5790" "84937" ...
## #...pvalues adjusted by 'BH' with cutoff <0.05
## #...14 enriched terms found
## 'data.frame':    14 obs. of  9 variables:
##  $ ID         : chr  "hsa04060" "hsa04660" "hsa04061" "hsa05340" ...
##  $ Description: chr  "Cytokine-cytokine receptor interaction" "T cell receptor signaling pa
thway" "Viral protein interaction with cytokine and cytokine receptor" "Primary immunodeficien
cy" ...
##  $ GeneRatio  : chr  "25/239" "14/239" "13/239" "8/239" ...
##  $ BgRatio    : chr  "295/8393" "104/8393" "100/8393" "38/8393" ...
##  $ pvalue     : num  9.24e-07 1.28e-06 4.60e-06 8.98e-06 1.10e-04 ...
##  $ p.adjust   : num  0.000162 0.000162 0.000387 0.000566 0.005527 ...
##  $ qvalue     : num  0.00014 0.00014 0.000336 0.000491 0.004802 ...
##  $ geneID     : chr  "27179/9235/7124/8784/10148/3601/7293/3552/6361/4283/4050/6367/2833/10
563/6363/1236/959/8742/4049/6364/939/6375/9173/6346/3559" "7124/915/916/917/925/3932/5605/2985
1/959/4792/919/1493/4794/3265" "7124/6361/4283/6367/2833/10563/6363/1236/4049/6364/6375/6346/3
559" "915/973/916/925/3932/29851/959/8625" ...
##  $ Count      : int  25 14 13 8 11 11 10 12 15 15 ...
## #...Citation
##  T Wu, E Hu, S Xu, M Chen, P Guo, Z Dai, T Feng, L Zhou, W Tang, L Zhan, X Fu, S Liu, X Bo,
and G Yu.
##  clusterProfiler 4.0: A universal enrichment tool for interpreting omics data.
##  The Innovation. 2021, 2(3):100141
```

```
#Generate a graph for the two KEGG results
#Edit the pathway id to that which is appropriate based on the ID column from the enrichKEGG o
utput

#These will generate images that will be saved to the working directory or the downloads folde
r
#Repeat for however many results you get from keggEnrich

pv.out_htmp2a <- pathview(gene.data = diffexpgenes_names_df$entrez, pathway.id = "hsa04060", s
pecies = "hsa")
```

```
## 'select()' returned 1:1 mapping between keys and columns
```

```
## Info: Working in directory /Users/dawndestefano/NYU/BIGY-7633 Transcriptomics/project
```

```
## Info: Writing image file hsa04060.pathview.png
```

```
#Repeat for the second result
pv.out_htmp2b <- pathview(gene.data = diffexpgenes_names_df$entrez, pathway.id = "hsa04660", s
pecies = "hsa")
```

```
## 'select()' returned 1:1 mapping between keys and columns
```

```
## Info: Working in directory /Users/dawndestefano/NYU/BIGY-7633 Transcriptomics/project
```

```
## Info: Writing image file hsa04660.pathview.png
```

```
## Info: some node width is different from others, and hence adjusted!
```

```
#Also show the genes involved in the pathway
#These correspond to the elements included in the image of the KEGG pathway generated earlier
pv.out_htmp2a$plot.data.gene
```

	kegg.names <chr>	labels <chr>	all.mapped <chr>	type <chr>	x <dbl>	y <dbl>	width <dbl>	height <dbl>	mol.data <dbl>
50	53833	IL20RB		gene	881	929	46	17	*NA*
51	53833	IL20RB		gene	881	886	46	17	*NA*
52	3565	IL4		gene	580	739	46	17	*NA*
53	659	BMPR2		gene	1748	399	46	17	*NA*
54	93	ACVR2B		gene	1531	779	46	17	*NA*
55	91	ACVR1B		gene	1531	675	46	17	*NA*
56	92	ACVR2A		gene	1531	735	46	17	*NA*
57	3588	IL10RB		gene	881	976	46	17	*NA*
58	58985	IL22RA1		gene	881	959	46	17	*NA*
59	3561	IL2RG		gene	666	455	46	17	*NA*

1-10 of 372 rows | 1-10 of 11 columns Previous **1** 2 3 4 5 6 … 38 Next

```
pv.out_htmp2b$plot.data.gene
```

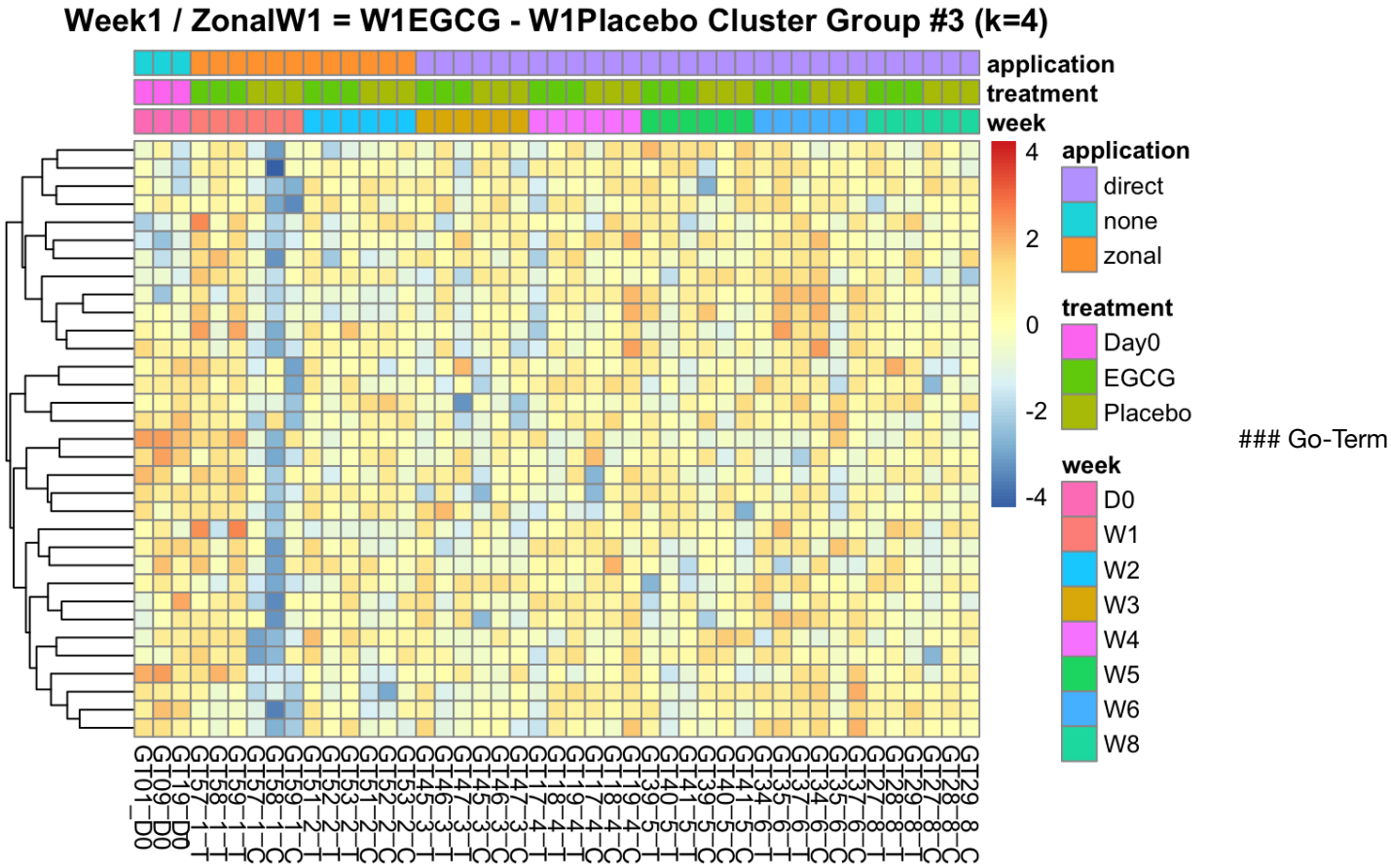
	kegg.names <chr>	labels <chr>	all.mapped <chr>	type <chr>	x <dbl>	y <dbl>	width <dbl>	height <dbl>	mol.data <dbl>
9	1019	CDK4		gene	1140	733	46	17	*NA*
10	7124	TNF	7124	gene	1140	699	46	17	1
11	1437	CSF2		gene	1140	678	46	17	*NA*
12	3458	IFNG		gene	1140	657	46	17	*NA*
13	3586	IL10		gene	1140	636	46	17	*NA*
14	3567	IL5		gene	1140	615	46	17	*NA*
15	3565	IL4		gene	1140	594	46	17	*NA*
16	3558	IL2		gene	1140	573	45	17	*NA*

	kegg.names <chr>	labels <chr>	all.mapped <chr>	type <chr>	x <dbl>	y <dbl>	width <dbl>	height <dbl>	mol.data <dbl>
17	4792	NFKBIA	4792,4794	gene	828	815	46	17	2
18	4790	NFKB1		gene	828	784	46	17	*NA*

1-10 of 66 rows | 1-10 of 11 columns Previous **1** 2 3 4 5 6 7 Next

Cluster#3

```
pheatmap(interact_sig_hclust_g3,annotation_col = annotation, scale="row", cluster_cols = F, sh
ow_rownames = F, main = "Week1 / ZonalW1 = W1EGCG - W1Placebo Cluster Group #3 (k=4)" )
```



Go-Term

Enrichment Part 3

Create HyperGoparpam

Converting the Ensemble to Entrez was achieved with this code: https://www.biostars.org/p/441386/ (https://www.biostars.org/p/441386/)

```
library("AnnotationDbi")

#adding ENTREZ ID's to global gene data file
GSE124161_readcount$entrez = mapIds(org.Hs.eg.db,
                    keys=rownames(GSE124161_readcount), #Column containing Ensembl gene ids
                    column="ENTREZID",
                    keytype="ENSEMBL",
                    multiVals="first") #This selects the first gene alias, if there are multip
le gene names under the single EntrezID
```

```
## 'select()' returned 1:many mapping between keys and columns
```

```
#Wrangling the ensemble gene ID's to Entrez in the interact_sig_hclust_g3
diffexpgenes_names_df <-rownames(as.data.frame(interact_sig_hclust_g3))
diffexpgenes_names_df <-as.data.frame(diffexpgenes_names_df)

diffexpgenes_names_df$entrez = mapIds(org.Hs.eg.db,
                    keys= diffexpgenes_names_df$diffexpgenes_names_df, #Column containing Ense
mbl gene ids
                    column="ENTREZID",
                    keytype="ENSEMBL",
                    multiVals="first") #This selects the first gene alias, if there are multip
le gene names under the single EntrezID
```

```
## 'select()' returned 1:1 mapping between keys and columns
```

```
diffexpgenes_names <-diffexpgenes_names_df$entrez
readcount_names <-GSE124161_readcount$entrez

#Utilized following resource for below code format https://bioconductor.org/packages/release/b
ioc/vignettes/GOstats/inst/doc/GOstatsHyperG.pdf

params <- new("GOHyperGParams",
              geneIds = diffexpgenes_names, #don't use quotes here, it will not work, you will
get an error message. This is the variable name where you stored your differentially expressed
gene names
              universeGeneIds = readcount_names, #don't use quotes here, it will not work, you
will get an error message. This is the variable name where you stored all of the gene names fr
om the whole unfiltered data set. Its the whole list of the "universe" of gene IDs for your ar
ray or reference genome.
              annotation = "org.Hs.eg",
              ontology = "BP",
              pvalueCutoff=0.01, #don't use quotes here, it will not work, you will get an err
or message
              testDirection = "over")
```

```
## Warning in makeValidParams(.Object): removing duplicate IDs in geneIds
```

```
## Warning in makeValidParams(.Object): removing duplicate IDs in universeGeneIds
```

```
hypGO <- hyperGTest(params)
hypGO
```

```
## Gene to GO BP  test for over-representation
## 866 GO BP ids tested (51 have p < 0.01)
## Selected gene set size: 16
##     Gene universe size: 17259
##     Annotation package: org.Hs.eg
```

```
sumGo <- summary(hypGO, categorySize =10)
sumGo
```

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size
<chr>	<dbl>	<dbl>	<dbl>	<int>	<int>
GO:0007292	0.0003790957	25.607892	0.145547251	3	157
GO:2000241	0.0007905140	19.764979	0.187264616	3	202
GO:0051445	0.0015708682	39.587558	0.059331363	2	64
GO:1903510	0.0034213793	26.344086	0.088069992	2	95
GO:0048477	0.0038571348	24.738817	0.093632308	2	101
GO:0030203	0.0056630937	20.214876	0.114027464	2	123
GO:0022412	0.0064129302	9.266018	0.391216177	3	422
GO:0006022	0.0066860411	18.518398	0.124225042	2	134
GO:0048609	0.0070875938	6.326767	0.803754563	4	867
GO:0032504	0.0082688739	6.038803	0.839909612	4	906

1-10 of 13 rows | 1-6 of 7 columns Previous **1** 2 Next

```
GoPlot <- data.frame(sumGo$GOBPID,sumGo$Pvalue,sumGo$Term)
colnames(GoPlot) <-c("GO_ID_BP", "P-value", "Term")
GoPlot
```

GO_ID_BP	P-value	Term
<chr>	<dbl>	<chr>
GO:0007292	0.0003790957	female gamete generation
GO:2000241	0.0007905140	regulation of reproductive process
GO:0051445	0.0015708682	regulation of meiotic cell cycle
GO:1903510	0.0034213793	mucopolysaccharide metabolic process
GO:0048477	0.0038571348	oogenesis
GO:0030203	0.0056630937	glycosaminoglycan metabolic process

GO_ID_BP	P-value	Term
<chr>	<dbl>	<chr>
GO:0022412	0.0064129302	cellular process involved in reproduction in multicellular organism
GO:0006022	0.0066860411	aminoglycan metabolic process
GO:0048609	0.0070875938	multicellular organismal reproductive process
GO:0032504	0.0082688739	multicellular organism reproduction

1-10 of 13 rows Previous **1** 2 Next

KEGG ENRIGHMENT Part3

```
#Now perform KEGG ENRICHMENT

keggEnrich <- enrichKEGG(
  diffexpgenes_names_df$entrez,
  organism = "hsa",
  keyType = "kegg",
  pvalueCutoff = 0.2, #adjust this if you are not seeing any results
  pAdjustMethod = "BH",
  )
```

```
#Show results from enrichKEGG
head(keggEnrich)
```

	ID	Description
	<chr>	<chr>
hsa04060	hsa04060	Cytokine-cytokine receptor interaction
hsa00603	hsa00603	Glycosphingolipid biosynthesis - globo and isoglobo series
hsa00604	hsa00604	Glycosphingolipid biosynthesis - ganglio series
hsa00511	hsa00511	Other glycan degradation
hsa00531	hsa00531	Glycosaminoglycan degradation
hsa00532	hsa00532	Glycosaminoglycan biosynthesis - chondroitin sulfate / dermatan sulfate

6 rows | 1-3 of 10 columns

```
keggEnrich
```

```
## #
## # over-representation test
## #
## #...@organism     hsa
## #...@ontology     KEGG
## #...@keytype      kegg
## #...@gene     chr [1:21] NA "3178" "3074" "79695" "84986" "2661" "105370683" "284071" ...
## #...pvalues adjusted by 'BH' with cutoff <0.2
## #...18 enriched terms found
## 'data.frame':    18 obs. of  9 variables:
##  $ ID         : chr  "hsa04060" "hsa00603" "hsa00604" "hsa00511" ...
##  $ Description: chr  "Cytokine-cytokine receptor interaction" "Glycosphingolipid biosynthes
is - globo and isoglobo series" "Glycosphingolipid biosynthesis - ganglio series" "Other glyca
n degradation" ...
##  $ GeneRatio  : chr  "3/8" "1/8" "1/8" "1/8" ...
##  $ BgRatio    : chr  "295/8393" "15/8393" "15/8393" "18/8393" ...
##  $ pvalue     : num  0.00211 0.01421 0.01421 0.01704 0.01797 ...
##  $ p.adjust   : num  0.0401 0.0629 0.0629 0.0629 0.0629 ...
##  $ qvalue     : num  0.0178 0.0279 0.0279 0.0279 0.0279 ...
##  $ geneID     : chr  "2661/57007/3976" "3074" "3074" "3074" ...
##  $ Count      : int  3 1 1 1 1 1 1 1 1 1 ...
## #...Citation
##  T Wu, E Hu, S Xu, M Chen, P Guo, Z Dai, T Feng, L Zhou, W Tang, L Zhan, X Fu, S Liu, X Bo,
and G Yu.
##  clusterProfiler 4.0: A universal enrichment tool for interpreting omics data.
##  The Innovation. 2021, 2(3):100141
```

```
#Generate a graph for the two KEGG results
#Edit the pathway id to that which is appropriate based on the ID column from the enrichKEGG o
utput

#These will generate images that will be saved to the working directory or the downloads folde
r
#Repeat for however many results you get from keggEnrich

pv.out_htmp3a <- pathview(gene.data = diffexpgenes_names_df$entrez, pathway.id = "hsa04060", s
pecies = "hsa")
```

```
## 'select()' returned 1:1 mapping between keys and columns
```

```
## Info: Working in directory /Users/dawndestefano/NYU/BIGY-7633 Transcriptomics/project
```

```
## Info: Writing image file hsa04060.pathview.png
```

```
#pv.out_htmp3b <- pathview(gene.data = diffexpgenes_names_df$entrez, pathway.id = "hsa00603", species = "hsa")# this will not pull from KEGG correctly, I think it has something to do with the double zero in the kegg pathway.id name "hsa00..." as it ONLY happens to genes with a 00 in the ID name. It's a glitch in the pathview() code somehow...So we captured the pathway from KEGG website.
```

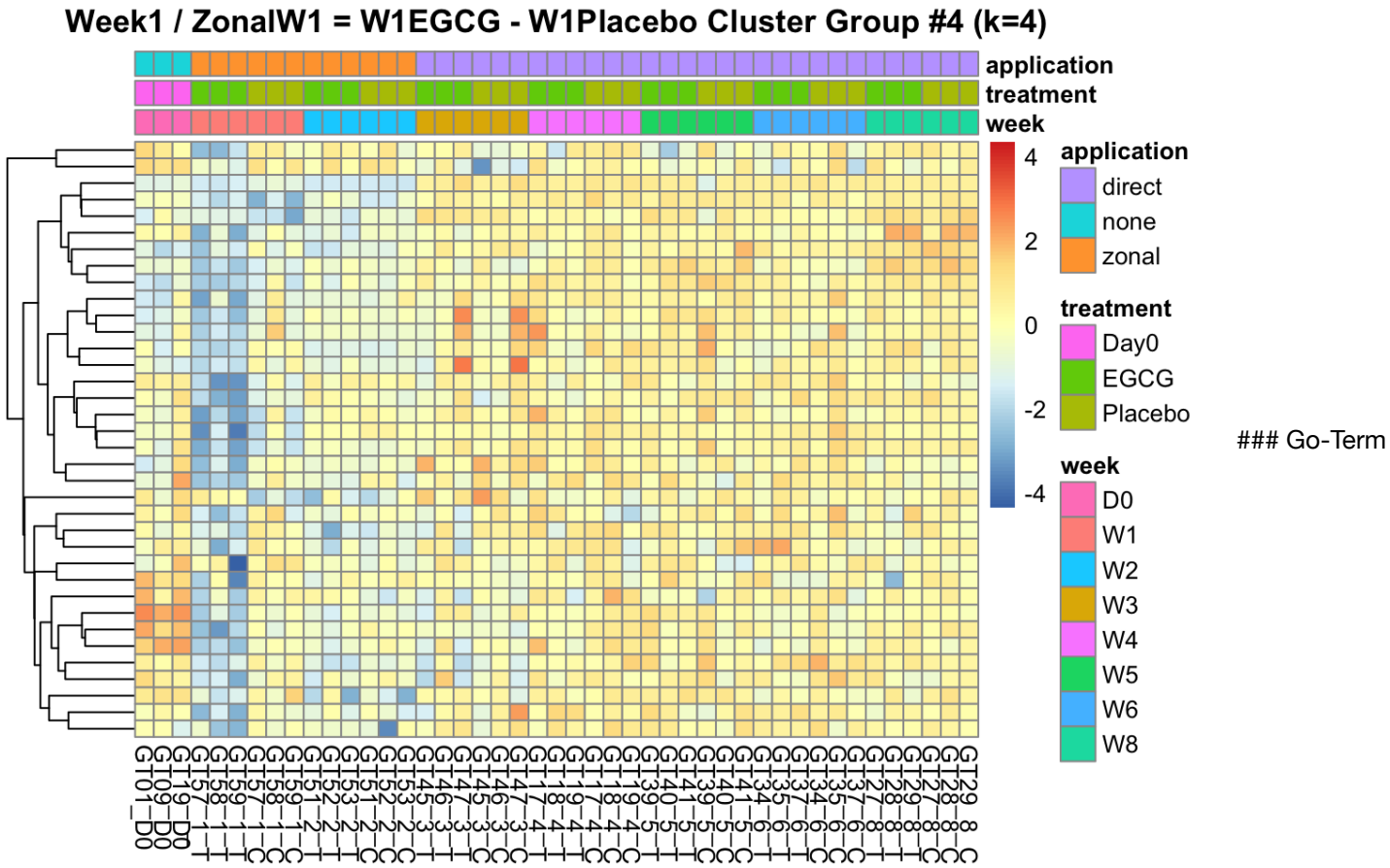
```
#Also show the genes involved in the pathway
#These correspond to the elements included in the image of the KEGG pathway generated earlier
pv.out_htmp3a$plot.data.gene
```

	kegg.names <chr>	labels <chr>	all.mapped <chr>	type <chr>	x <dbl>	y <dbl>	width <dbl>	height <dbl>	mol.data <dbl>
50	53833	IL20RB		gene	881	929	46	17	*NA*
51	53833	IL20RB		gene	881	886	46	17	*NA*
52	3565	IL4		gene	580	739	46	17	*NA*
53	659	BMPR2		gene	1748	399	46	17	*NA*
54	93	ACVR2B		gene	1531	779	46	17	*NA*
55	91	ACVR1B		gene	1531	675	46	17	*NA*
56	92	ACVR2A		gene	1531	735	46	17	*NA*
57	3588	IL10RB		gene	881	976	46	17	*NA*
58	58985	IL22RA1		gene	881	959	46	17	*NA*
59	3561	IL2RG		gene	666	455	46	17	*NA*

1-10 of 372 rows | 1-10 of 11 columns Previous **1** 2 3 4 5 6 … 38 Next

Cluster#4

```
pheatmap(interact_sig_hclust_g4,annotation_col = annotation, scale="row", cluster_cols = F, show_rownames = F, main = "Week1 / ZonalW1 = W1EGCG - W1Placebo Cluster Group #4 (k=4)" )
```



Go-Term

Enrichment Part 4

Create HyperGoparpam

Converting the Ensemble to Entrez was achieved with this code: https://www.biostars.org/p/441386/ (https://www.biostars.org/p/441386/)

```
library("AnnotationDbi")

#adding ENTREZ ID's to global gene data file
GSE124161_readcount$entrez = mapIds(org.Hs.eg.db,
                     keys=rownames(GSE124161_readcount), #Column containing Ensembl gene ids
                     column="ENTREZID",
                     keytype="ENSEMBL",
                     multiVals="first") #This selects the first gene alias, if there are multip
le gene names under the single EntrezID
```

```
## 'select()' returned 1:many mapping between keys and columns
```

```
#Wrangling the ensemble gene ID's to Entrez in the interact_sig_hclust_g4
diffexpgenes_names_df <-rownames(as.data.frame(interact_sig_hclust_g4))
diffexpgenes_names_df <-as.data.frame(diffexpgenes_names_df)

diffexpgenes_names_df$entrez = mapIds(org.Hs.eg.db,
                    keys= diffexpgenes_names_df$diffexpgenes_names_df, #Column containing Ense
mbl gene ids
                    column="ENTREZID",
                    keytype="ENSEMBL",
                    multiVals="first") #This selects the first gene alias, if there are multip
le gene names under the single EntrezID
```

```
## 'select()' returned 1:1 mapping between keys and columns
```

```
diffexpgenes_names <-diffexpgenes_names_df$entrez
readcount_names <-GSE124161_readcount$entrez

#Utilized following resource for below code format https://bioconductor.org/packages/release/b
ioc/vignettes/GOstats/inst/doc/GOstatsHyperG.pdf

params <- new("GOHyperGParams",
              geneIds = diffexpgenes_names, #don't use quotes here, it will not work, you will
get an error message. This is the variable name where you stored your differentially expressed
gene names
              universeGeneIds = readcount_names, #don't use quotes here, it will not work, you
will get an error message. This is the variable name where you stored all of the gene names fr
om the whole unfiltered data set. Its the whole list of the "universe" of gene IDs for your ar
ray or reference genome.
              annotation = "org.Hs.eg",
              ontology = "BP",
              pvalueCutoff=0.01, #don't use quotes here, it will not work, you will get an err
or message
              testDirection = "over")
```

```
## Warning in makeValidParams(.Object): removing duplicate IDs in geneIds
```

```
## Warning in makeValidParams(.Object): removing duplicate IDs in universeGeneIds
```

```
hypGO <- hyperGTest(params)
hypGO
```

```
## Gene to GO BP  test for over-representation
## 987 GO BP ids tested (47 have p < 0.01)
## Selected gene set size: 27
##     Gene universe size: 17259
##     Annotation package: org.Hs.eg
```

```
sumGo <- summary(hypGO, categorySize =10)
sumGo
```

GOBPID <chr>	Pvalue <dbl>	OddsRatio <dbl>	ExpCount <dbl>	Count <int>	Size <int>
GO:0001865	0.0001052416	172.240000	0.01564401	2	10
GO:0032823	0.0001818911	125.243636	0.02033722	2	13
GO:0001779	0.0006915264	59.857391	0.03911003	2	25
GO:0030574	0.0019524237	34.384000	0.06570485	2	42
GO:0032814	0.0019524237	34.384000	0.06570485	2	42
GO:1903018	0.0023379659	31.250909	0.07196245	2	46
GO:0022617	0.0038169276	24.105263	0.09229967	2	59
GO:0010951	0.0059319060	9.040957	0.37232748	3	238
GO:0010466	0.0065720644	8.702869	0.38640709	3	247
GO:0042110	0.0090367402	5.480558	0.83539023	4	534

1-10 of 12 rows | 1-6 of 7 columns Previous **1** 2 Next

```
GoPlot <- data.frame(sumGo$GOBPID,sumGo$Pvalue,sumGo$Term)
colnames(GoPlot) <-c("GO_ID_BP", "P-value", "Term")
GoPlot
```

GO_ID_BP <chr>	P-value <dbl>	Term <chr>
GO:0001865	0.0001052416	NK T cell differentiation
GO:0032823	0.0001818911	regulation of natural killer cell differentiation
GO:0001779	0.0006915264	natural killer cell differentiation
GO:0030574	0.0019524237	collagen catabolic process
GO:0032814	0.0019524237	regulation of natural killer cell activation
GO:1903018	0.0023379659	regulation of glycoprotein metabolic process
GO:0022617	0.0038169276	extracellular matrix disassembly
GO:0010951	0.0059319060	negative regulation of endopeptidase activity
GO:0010466	0.0065720644	negative regulation of peptidase activity
GO:0042110	0.0090367402	T cell activation

1-10 of 12 rows Previous **1** 2 Next

KEGG ENRIGHMENT Part4

```
#Now perform KEGG ENRICHMENT

keggEnrich <- enrichKEGG(
  diffexpgenes_names_df$entrez,
  organism = "hsa",
  keyType = "kegg",
  pvalueCutoff = 0.2, #adjust this if you are not seeing any results
  pAdjustMethod = "BH",
  )
```

```
#Show results from enrichKEGG
head(keggEnrich)
```

	ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue
	<chr>	<chr>	<chr>	<chr>	<dbl>	<dbl>	<dbl>
hsa04514	hsa04514	Cell adhesion molecules	3/16	157/8393	0.00300667	0.1353002	0.1297616

1 row | 1-8 of 10 columns

```
keggEnrich
```

```
## #
## # over-representation test
## #
## #...@organism     hsa
## #...@ontology     KEGG
## #...@keytype      kegg
## #...@gene     chr [1:33] "5104" "6366" "100505832" "1471" "81558" "9934" "10522" "4519" ...
## #...pvalues adjusted by 'BH' with cutoff <0.2
## #...1 enriched terms found
## 'data.frame':    1 obs. of  9 variables:
##  $ ID         : chr "hsa04514"
##  $ Description: chr "Cell adhesion molecules"
##  $ GeneRatio  : chr "3/16"
##  $ BgRatio    : chr "157/8393"
##  $ pvalue     : num 0.00301
##  $ p.adjust   : num 0.135
##  $ qvalue     : num 0.13
##  $ geneID     : chr "926/57689/7122"
##  $ Count      : int 3
## #...Citation
##  T Wu, E Hu, S Xu, M Chen, P Guo, Z Dai, T Feng, L Zhou, W Tang, L Zhan, X Fu, S Liu, X Bo,
and G Yu.
##  clusterProfiler 4.0: A universal enrichment tool for interpreting omics data.
##  The Innovation. 2021, 2(3):100141
```

```
#Generate a graph for the last KEGG result
#Edit the pathway id to that which is appropriate based on the ID column from the enrichKEGG o
utput

#These will generate images that will be saved to the working directory or the downloads folde
r
#Repeat for however many results you get from keggEnrich

pv.out_htmp4a <- pathview(gene.data = diffexpgenes_names_df$entrez, pathway.id = "hsa04514", s
pecies = "hsa")
```

```
## 'select()' returned 1:1 mapping between keys and columns
```

```
## Info: Working in directory /Users/dawndestefano/NYU/BIGY-7633 Transcriptomics/project
```

```
## Info: Writing image file hsa04514.pathview.png
```

```
#Also show the genes involved in the pathway
#These correspond to the elements included in the image of the KEGG pathway generated earlier
pv.out_htmp4a$plot.data.gene
```

	kegg.names <chr>	labels <chr>	all.mapped <chr>	type <chr>	x <dbl>	y <dbl>	width <dbl>	height <dbl>	mol.data <dbl>
77	214	ALCAM		gene	83	582	46	17	*NA*
78	923	CD6		gene	167	582	46	17	*NA*
79	6693	SPN		gene	168	819	46	17	*NA*
80	6614	SIGLEC1		gene	84	819	46	17	*NA*
81	5788	PTPRC		gene	168	780	46	17	*NA*
82	933	CD22		gene	84	780	46	17	*NA*
83	926	CD8B	926	gene	273	249	46	17	1
84	3688	ITGB1		gene	933	476	46	17	*NA*
85	8516	ITGA8		gene	933	459	46	17	*NA*
86	3689	ITGB2		gene	465	644	46	17	*NA*