interact5 = W1EGCG + W1Placebo - W2EGCG - W2Placebo, # Global Differntial Expression from week 1 vs week 2

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

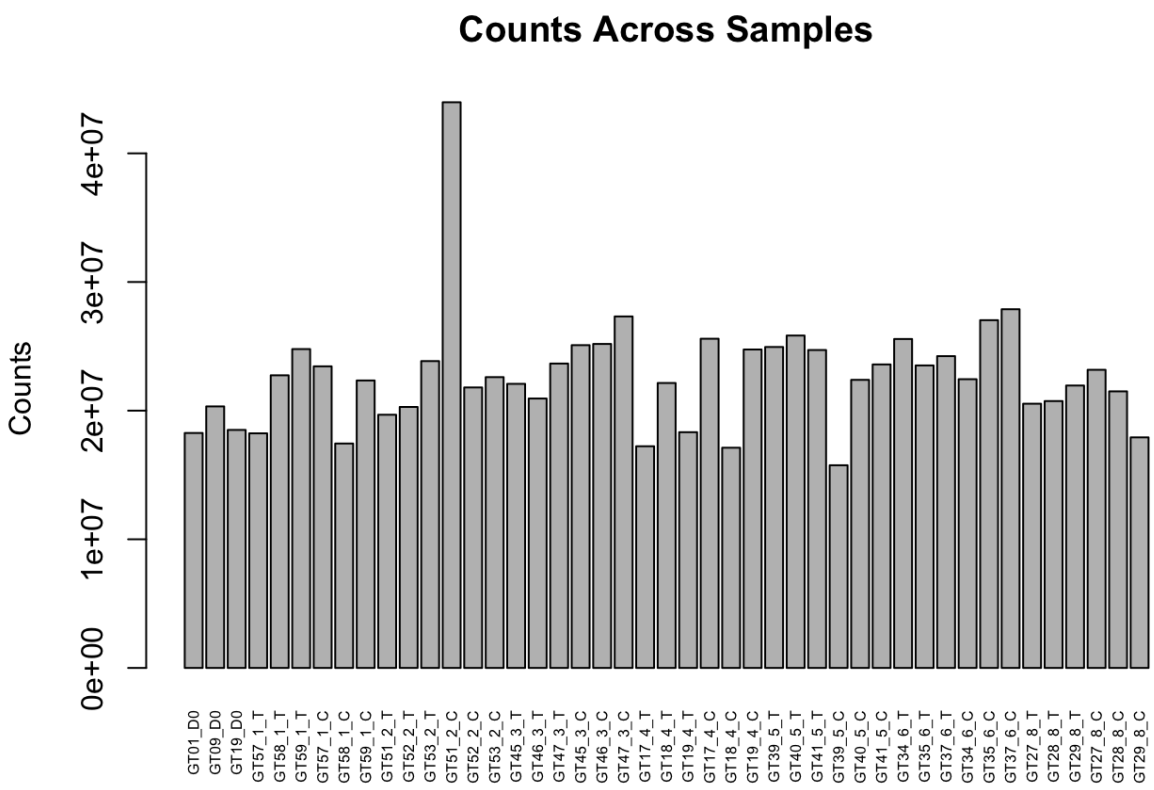
Quality Control

```
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: Violin 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)
```

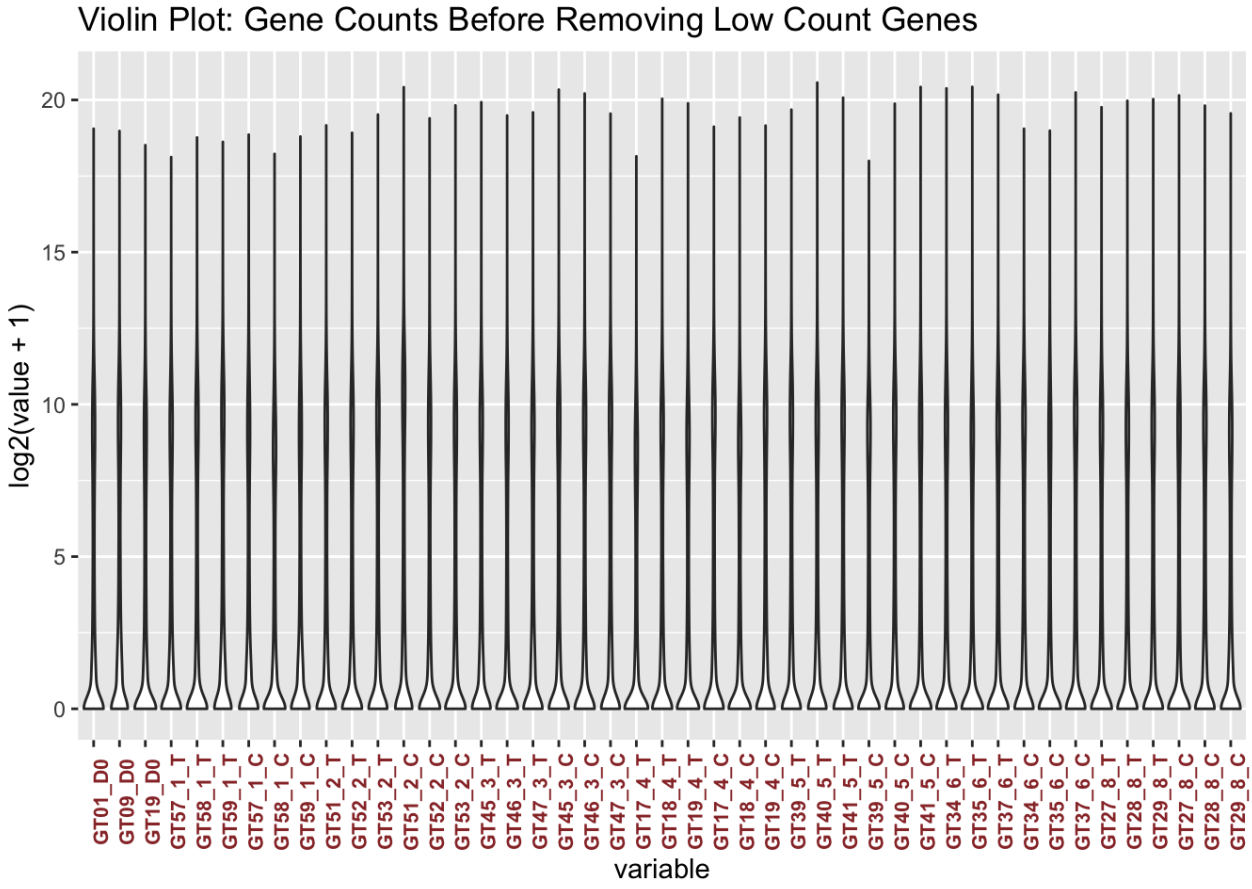
```
##
## Attaching package: 'reshape2'
##
## The following object is masked from 'package:tidyr':
##
##     smiths
```

```
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
```

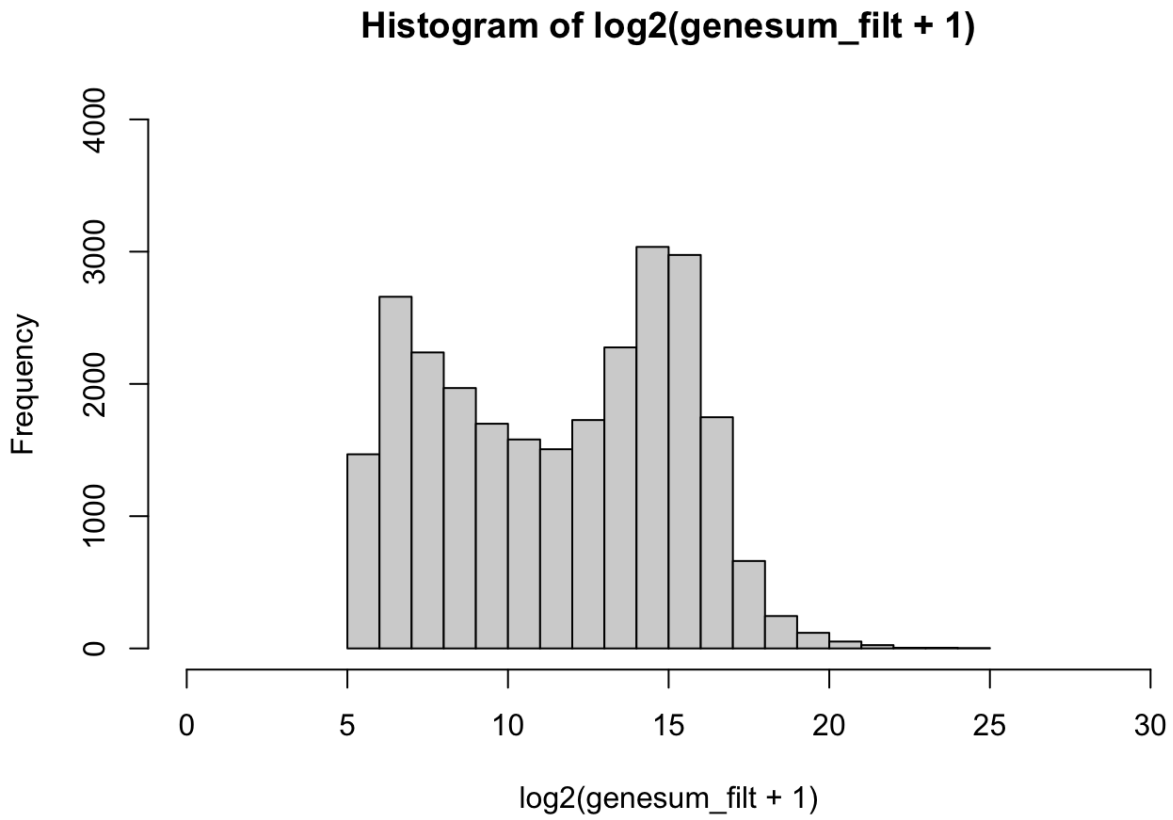
```
#For this filtering we wanted to keep the subset of data where the genesum was over 30
#genesum = 45 + 1 = 46 log2(46) = 5.52 = ~5.5, everything below 5 on the above graph, so we want to keep everything with a genesum count of 45 and above

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
```

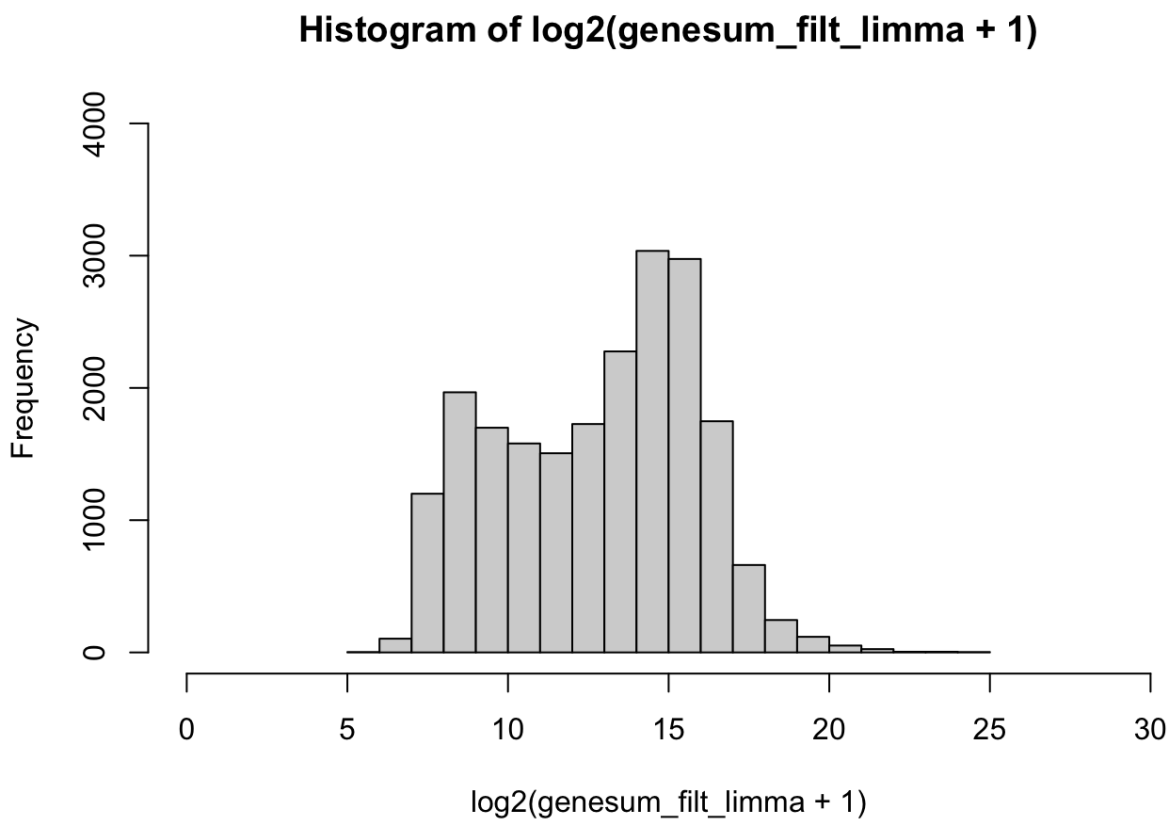
```
## [1] 48162    45
```

```
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.
```

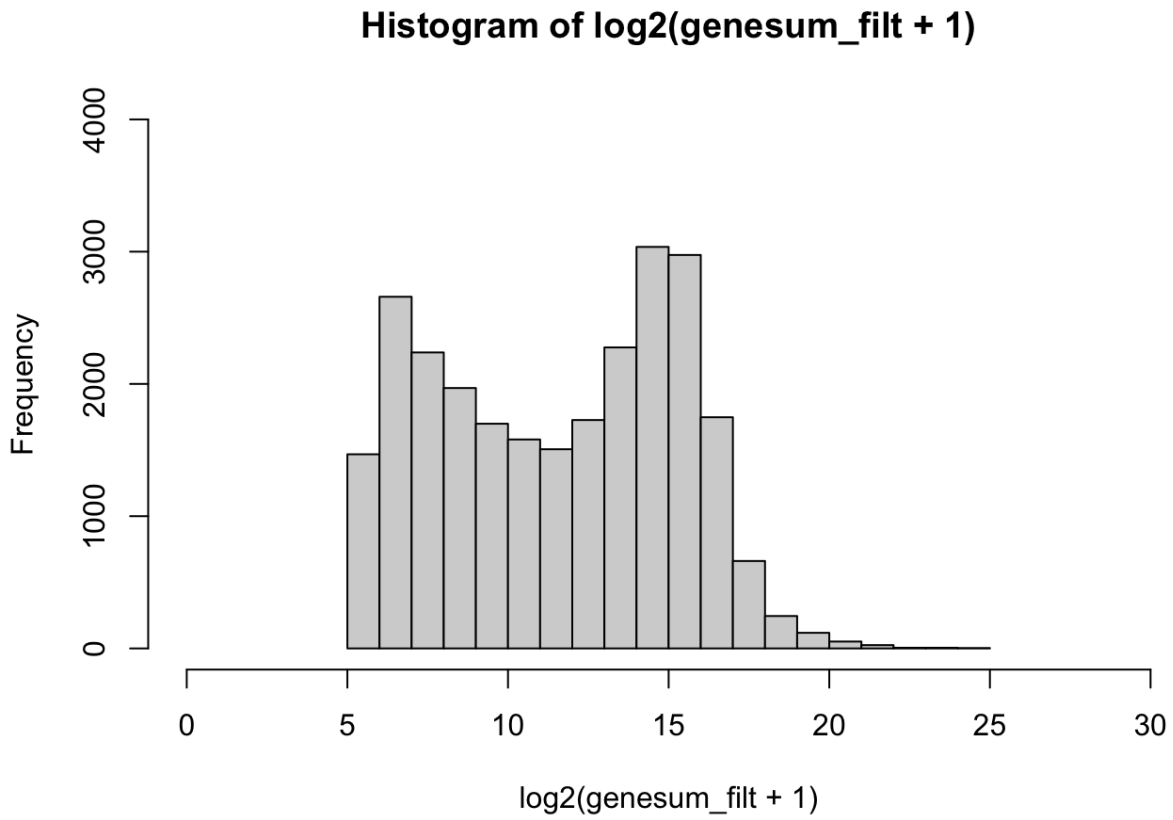
```
## [1] 20935    45
```

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

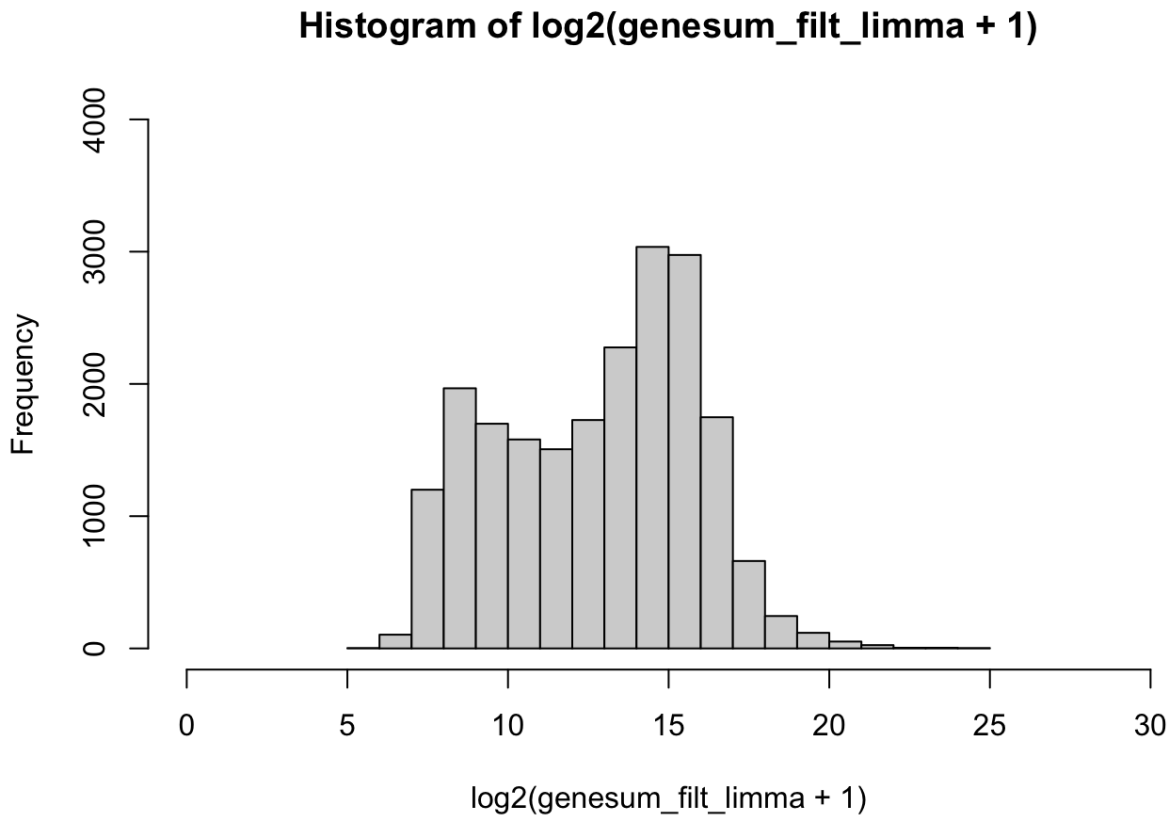


Merge of plots “Low Counts Trimmed Data: Mannual vs Limma”

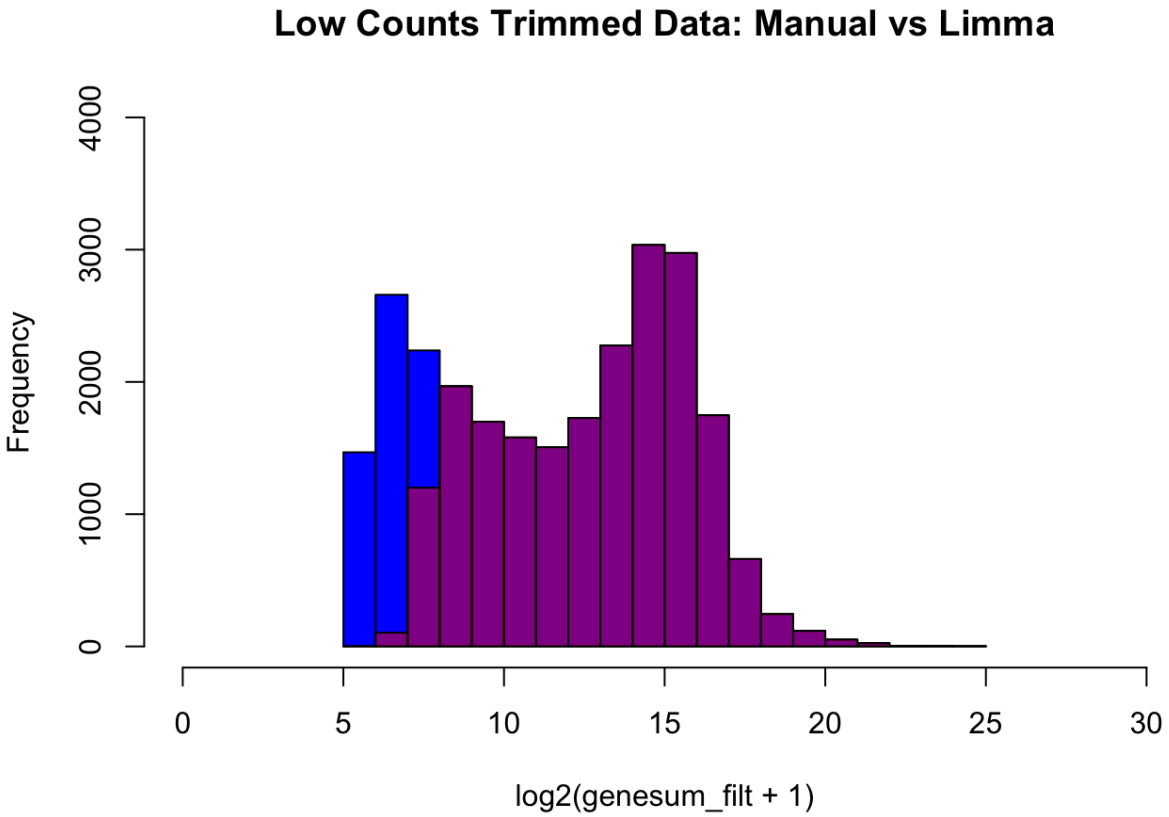
```
histfilt <- hist(log2(genesum_filt+1), ylim = c(0,4000), xlim = c(0,30), breaks = 25)
```




```
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

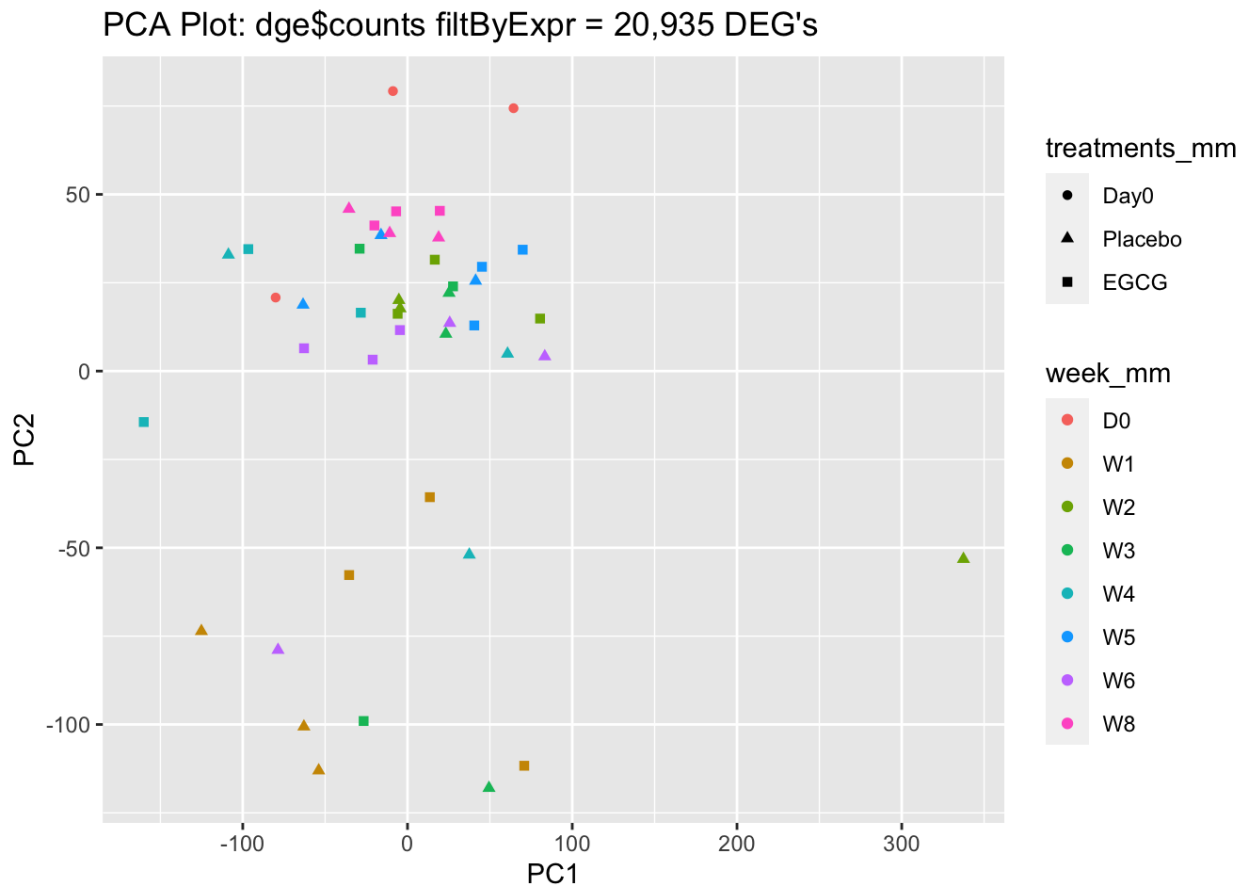
```
library(dplyr)

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")
```



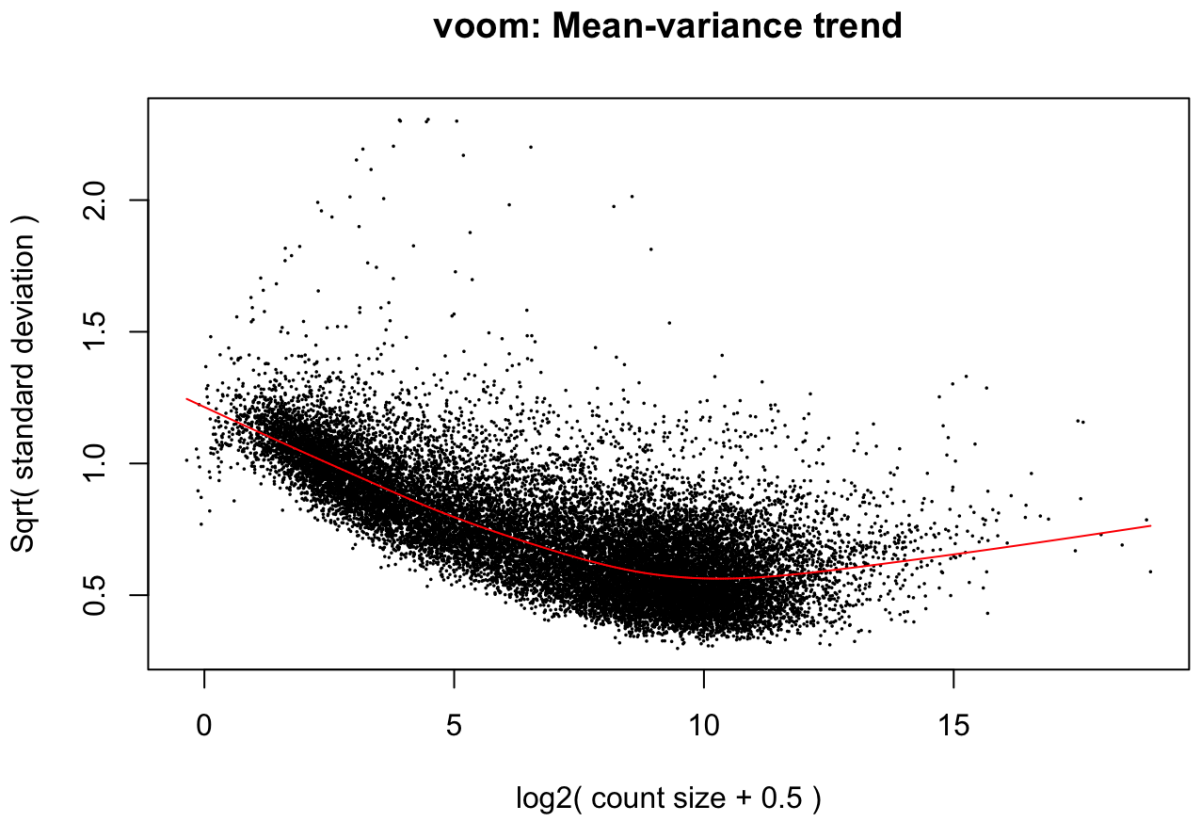
Now that we created a matrix that defines the various experimental categories of the samples, now we want to normalize the data. We need to normalize the data first before making the comparisons, as the normalized data is needed to proceed in next steps.

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 = INTERACT5

get details of specific coeff defined in the contrast. Selected contrast “interact4”

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.

interact5 = W1EGCG + W1Placebo - W2EGCG - W2Placebo, # Significant means that EGCG is having a statistically differential response between the two time points, relative to the control

```
topTable(nfit2, coef = "interact5",  adjust="BH") #Global gene expression of week 1 vs week 2
```

	logFC <dbl>	AveExpr <dbl>	t <dbl>	P.Value <dbl>	adj.P.Val <dbl>	B <dbl>
ENSG00000021355	5.398868	5.6992582	13.73003	1.704989e-15	2.759329e-11	25.16351
ENSG00000140519	8.296636	2.0845790	13.47454	2.943551e-15	2.759329e-11	24.41837
ENSG00000171848	4.043708	4.0747034	13.33783	3.954138e-15	2.759329e-11	24.35641
ENSG00000183696	6.103838	4.4857629	12.76551	1.391846e-14	4.856383e-11	23.13621
ENSG00000189410	11.148766	-0.8341036	12.98246	8.600142e-15	4.501099e-11	22.18823
ENSG00000115602	6.170956	1.7099245	12.29540	4.025744e-14	1.203985e-10	21.78421
ENSG00000163209	10.491986	-2.8936690	12.77534	1.361650e-14	4.856383e-11	21.59765
ENSG00000106819	-9.712416	8.7259723	-11.55994	2.236541e-13	5.852749e-10	20.40319
ENSG00000175592	5.437045	2.1858640	11.48703	2.660563e-13	6.188766e-10	20.14505
ENSG00000138772	7.846384	2.5481564	11.18186	5.542740e-13	1.082157e-09	19.41625

1-10 of 10 rows

```
Top_10_genes <-as.data.frame(rownames(topTable(nfit2, coef = "interact5",  adjust="BH")))
Top_10_genes
```

rownames(topTable(nfit2, coef = "interact5", adjust = "BH"))
<chr>
ENSG00000021355
ENSG00000140519
ENSG00000171848
ENSG00000183696
ENSG00000189410
ENSG00000115602
ENSG00000163209
ENSG00000106819
ENSG00000175592
ENSG00000138772

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 = "interact5",  adjust="BH")), #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>
ENSG00000021355	SERPINB1
ENSG00000140519	RHCG

GeneIDs <chr>	GeneSymbol <chr>
ENSG00000171848	RRM2
ENSG00000183696	UPP1
ENSG00000189410	SH2D5
ENSG00000115602	IL1RL1
ENSG00000163209	SPRR3
ENSG00000106819	ASPN
ENSG00000175592	FOSL1
ENSG00000138772	ANXA3

1-10 of 10 rows

```
top_table <-as.data.frame(topTable(nfit2, coef = "interact5",  adjust="BH"))
top_table
```

	logFC <dbl>	AveExpr <dbl>	t <dbl>	P.Value <dbl>	adj.P.Val <dbl>	B <dbl>
ENSG00000021355	5.398868	5.6992582	13.73003	1.704989e-15	2.759329e-11	25.16351
ENSG00000140519	8.296636	2.0845790	13.47454	2.943551e-15	2.759329e-11	24.41837
ENSG00000171848	4.043708	4.0747034	13.33783	3.954138e-15	2.759329e-11	24.35641
ENSG00000183696	6.103838	4.4857629	12.76551	1.391846e-14	4.856383e-11	23.13621
ENSG00000189410	11.148766	-0.8341036	12.98246	8.600142e-15	4.501099e-11	22.18823
ENSG00000115602	6.170956	1.7099245	12.29540	4.025744e-14	1.203985e-10	21.78421
ENSG00000163209	10.491986	-2.8936690	12.77534	1.361650e-14	4.856383e-11	21.59765
ENSG00000106819	-9.712416	8.7259723	-11.55994	2.236541e-13	5.852749e-10	20.40319
ENSG00000175592	5.437045	2.1858640	11.48703	2.660563e-13	6.188766e-10	20.14505
ENSG00000138772	7.846384	2.5481564	11.18186	5.542740e-13	1.082157e-09	19.41625

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
```

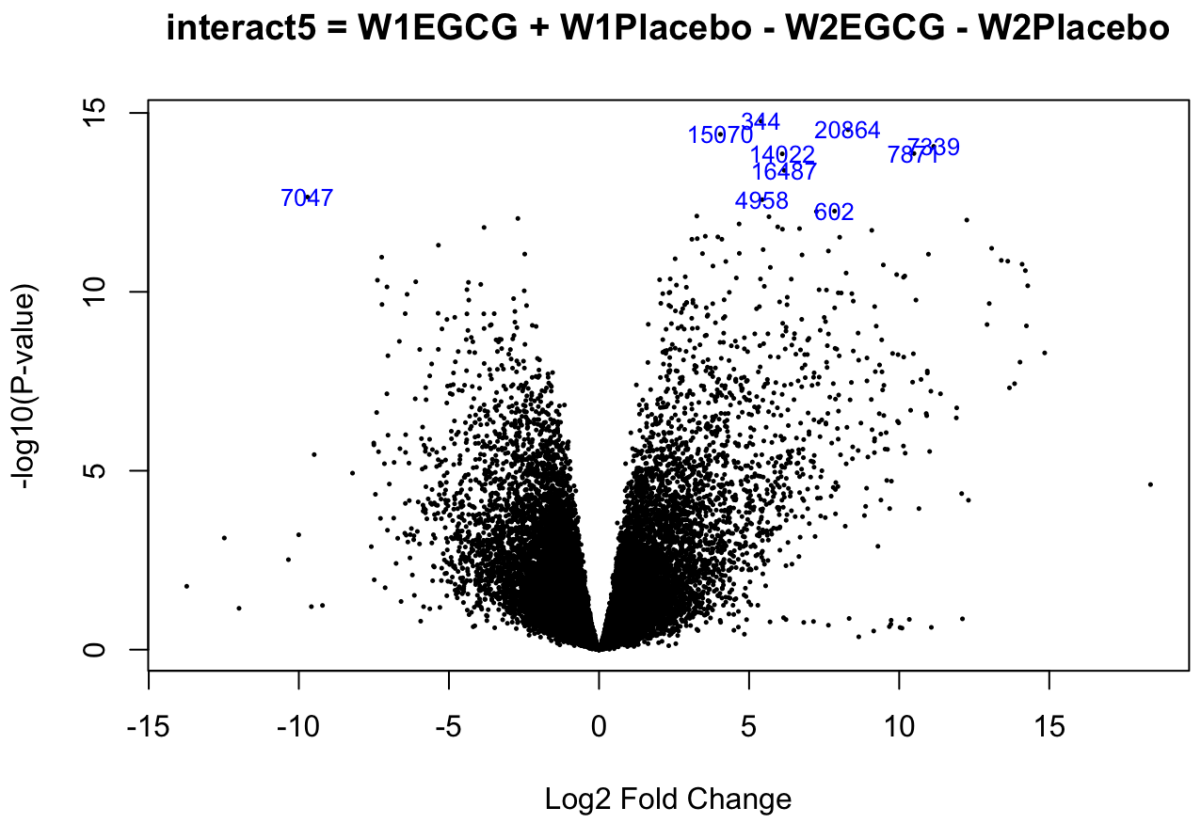
GeneIDs <chr>	GeneSymbol <chr>	logFC <dbl>	adj.p.val <dbl>
ENSG00000021355	SERPINB1	5.398868	2.759329e-11
ENSG00000140519	RHCG	8.296636	2.759329e-11

GeneIDs <chr>	GeneSymbol <chr>	logFC <dbl>	adj.p.val <dbl>
ENSG00000171848	RRM2	4.043708	2.759329e-11
ENSG00000183696	UPP1	6.103838	4.856383e-11
ENSG00000189410	SH2D5	11.148766	4.501099e-11
ENSG00000115602	IL1RL1	6.170956	1.203985e-10
ENSG00000163209	SPRR3	10.491986	4.856383e-11
ENSG00000106819	ASPN	-9.712416	5.852749e-10
ENSG00000175592	FOSL1	5.437045	6.188766e-10
ENSG00000138772	ANXA3	7.846384	1.082157e-09

1-10 of 10 rows

Make a volcano plot of this data

```
volcanoplot(nfit2, "interact5", highlight = 10, main="interact5 = W1EGCG + W1Placebo - W2EGCG - W2Placebo")#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

The p-value was selected for 0.2, and L2FC of 1.5 to keep analysis parameters for signifigance and L2FC equivalent across the analysis.

This uses the coefficient “interact5” of the following interaction of Global gene expression of week 1 vs week 2 interact5 = W1EGCG + W1Placebo - W2EGCG - W2Placebo

```
interact_sig = topTable(nfit2,
                coef = "interact5",
                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.

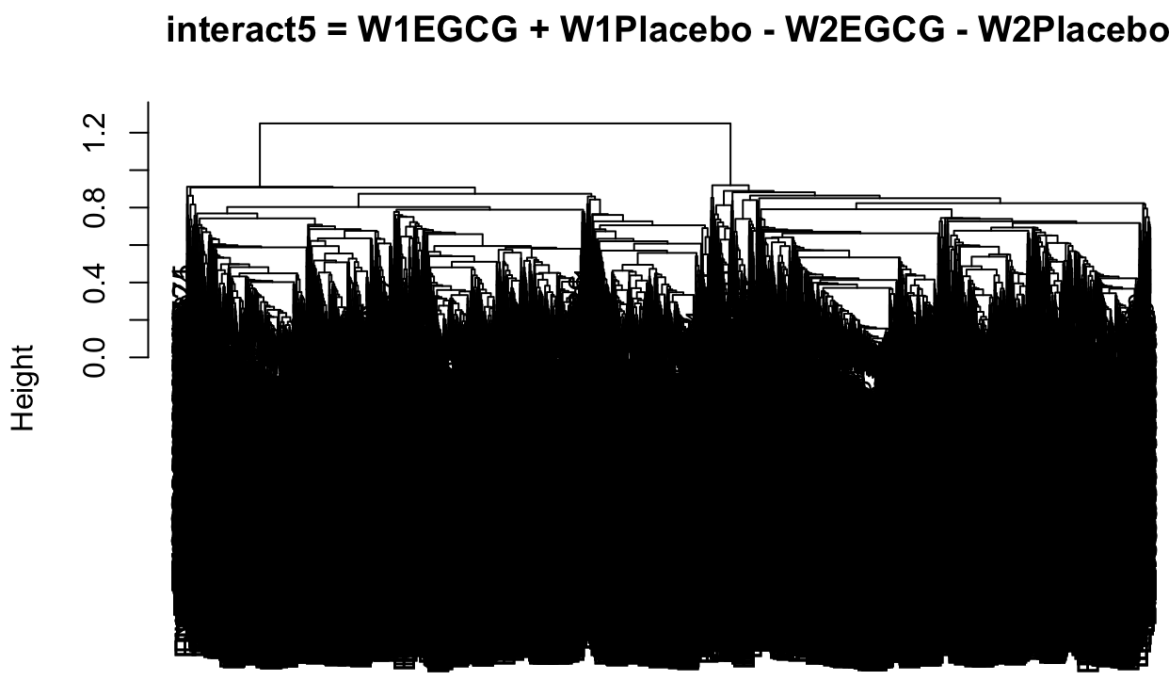
```
interact_sig_normvalues = normexpvalues[rownames(interact_sig),]
```

Calculate the distance using pairwise correlation of genes.

Use hclust to perform the clustering.

This is the interaction of EGCG vs Treatment in the interaction of week 1 and 2

```
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 = "interact5 = W1EGCG + W1Placebo - W2EGCG - W2Placebo")
```



Let's

interact_sig_dist
hclust (*, "average")

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

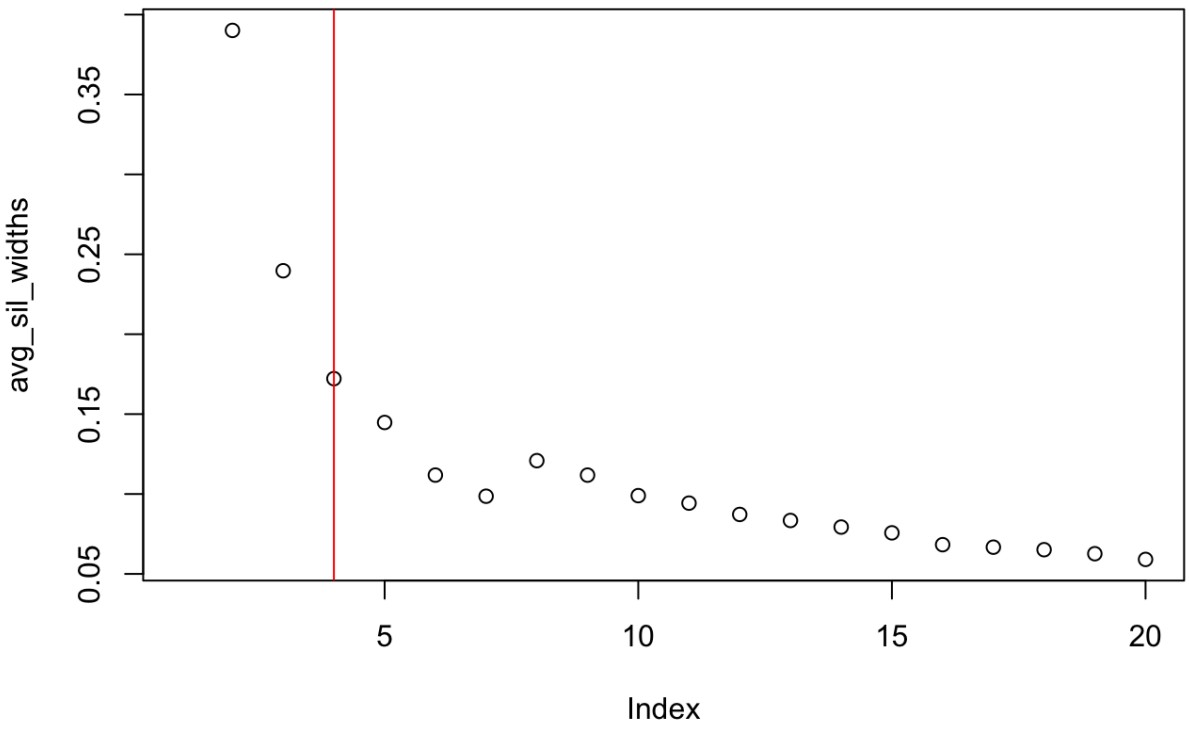
```
library(cluster)

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"])

}
```

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)
```

```
## ENSG00000021355 ENSG00000140519 ENSG00000171848 ENSG00000189410 ENSG00000163209
##               1               1               1               1               1
## ENSG00000183696
##               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

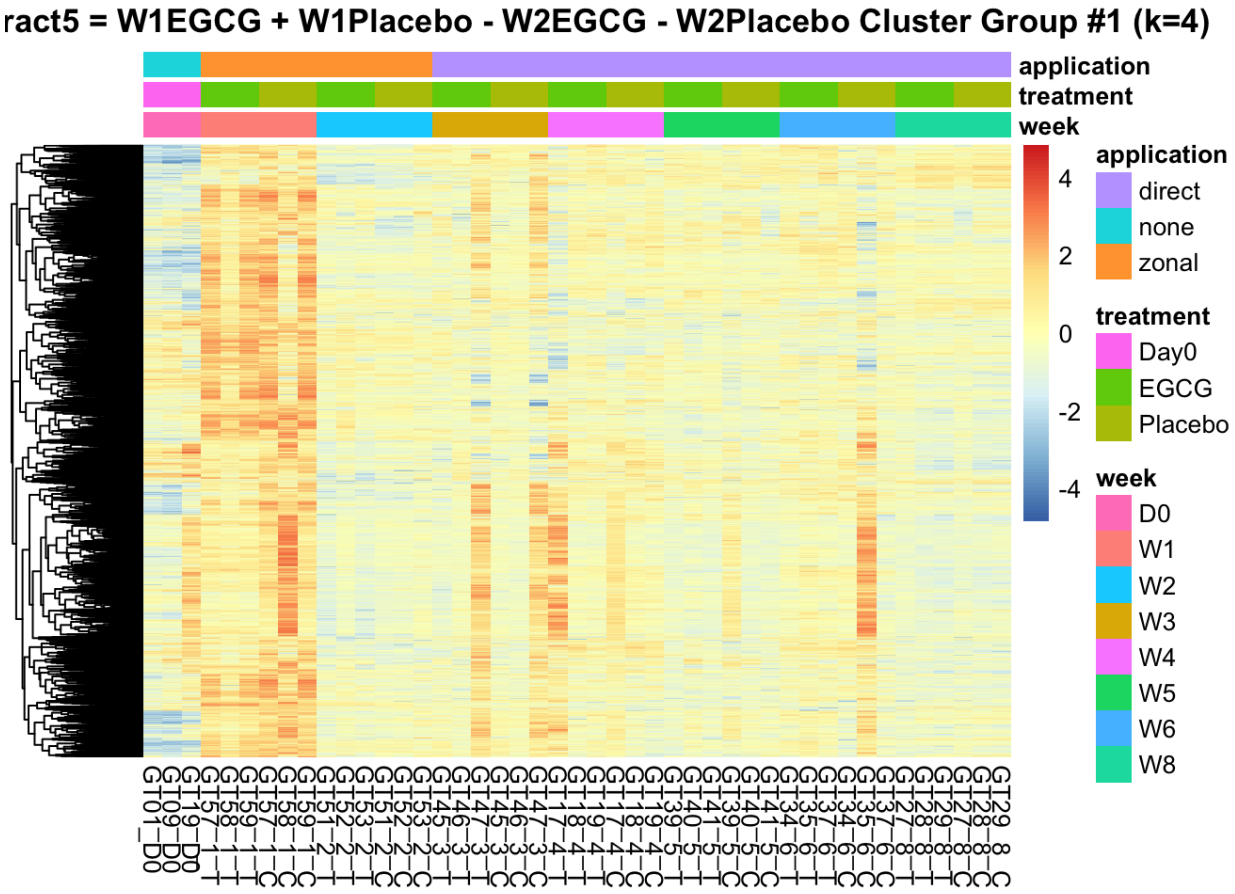
Cluster#1

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 = "interact5 = W1EGCG + W1Placebo - W2EGCG - W2Placebo 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
## 11068 GO BP ids tested (1462 have p < 0.01)
## Selected gene set size: 3469
##     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:0006955	6.875325e-56	2.384590	369.431717	641	1838
GO:0002376	2.333915e-53	2.087573	550.127644	859	2737
GO:0045321	8.874173e-46	2.770193	194.765687	381	969
GO:0002682	3.292776e-45	2.369642	289.234081	508	1439
GO:0001775	5.608168e-42	2.520896	222.905209	411	1109
GO:0006952	6.885088e-41	2.158489	342.096182	564	1702
GO:0009607	1.145718e-40	2.238687	301.896865	512	1502
GO:0006950	1.554274e-40	1.777583	775.846805	1076	3860
GO:0006954	2.764704e-39	2.828330	155.772351	311	775
GO:0043207	3.998907e-39	2.222746	294.459992	498	1465

1-10 of 1,332 rows | 1-6 of 7 columns Previous **1** 2 3 4 5 6 … 134 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>
GO:0006955	6.875325e-56
GO:0002376	2.333915e-53
GO:0045321	8.874173e-46
GO:0002682	3.292776e-45
GO:0001775	5.608168e-42
GO:0006952	6.885088e-41
GO:0009607	1.145718e-40
GO:0006950	1.554274e-40
GO:0006954	2.764704e-39
GO:0043207	3.998907e-39

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	GeneRa
	<chr>	<chr>	<chr>
hsa04061	hsa04061	Viral protein interaction with cytokine and cytokine receptor	58/1761
hsa04110	hsa04110	Cell cycle	72/1761
hsa04064	hsa04064	NF-kappa B signaling pathway	54/1761
hsa04060	hsa04060	Cytokine-cytokine receptor interaction	113/176
hsa04640	hsa04640	Hematopoietic cell lineage	51/1761
hsa03030	hsa03030	DNA replication	26/1761

6 rows | 1-4 of 10 columns

```
keggEnrich
```

```
## #
## # over-representation test
## #
## #...@organism     hsa
## #...@ontology     KEGG
## #...@keytype      kegg
## #...@gene     chr [1:4069] "1992" "51458" "6241" "400745" "6707" "7378" "9173" "8061" ...
## #...pvalues adjusted by 'BH' with cutoff <0.05
## #...104 enriched terms found
## 'data.frame':    104 obs. of  9 variables:
##  $ ID         : chr  "hsa04061" "hsa04110" "hsa04064" "hsa04060" ...
##  $ Description: chr  "Viral protein interaction with cytokine and cytokine receptor" "Cell
cycle" "NF-kappa B signaling pathway" "Cytokine-cytokine receptor interaction" ...
##  $ GeneRatio  : chr  "58/1761" "72/1761" "54/1761" "113/1761" ...
##  $ BgRatio    : chr  "100/8393" "157/8393" "104/8393" "295/8393" ...
##  $ pvalue     : num  4.99e-16 1.69e-12 2.45e-12 3.25e-12 1.47e-11 ...
##  $ p.adjust   : num  1.67e-13 2.71e-10 2.71e-10 2.71e-10 9.84e-10 ...
##  $ qvalue     : num  1.12e-13 1.83e-10 1.83e-10 1.83e-10 6.64e-10 ...
##  $ geneID     : chr  "5473/6374/6372/2919/3577/3559/3576/7852/6355/3579/6367/11009/1236/879
4/50604/2920/8807/8740/2921/3586/7132/6347"| __truncated__ "891/8318/991/9319/990/113130/5347/
993/9212/699/9088/10403/9133/890/151648/4085/57082/898/4998/81620/896/983/511"| __truncated__
"3929/5743/3553/2919/3576/5971/4067/597/10673/2920/7128/929/4615/8740/2921/8837/7132/4050/330/
3932/4792/4616/709"| __truncated__ "9173/3575/27179/5473/5008/6374/6372/3553/3552/1441/944/143
8/2919/3577/3559/3576/7852/6355/3579/6367/5617/11009/"| __truncated__ ...
##  $ Count      : int  58 72 54 113 51 26 59 42 45 78 ...
## #...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 = "hsa04061", 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 hsa04061.pathview.png
```

```
#Repeat for the second result
pv.out_htmp1b <- pathview(gene.data = diffexpgenes_names_df$entrez, pathway.id = "hsa04110", 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 hsa04110.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
#Commented out line, no pathway identified
pv.out_htmp1a$plot.data.gene
```

	kegg.na…	labels	all.mapped	t…	x	y	wi…	height	mol.data
	<chr>	<chr>	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
2	6354	CCL7	6354,6355,6357	gene	165	177	46	17	3
6	1230	CCR1	1230	gene	258	177	46	17	1
8	6348	CCL3		gene	165	207	46	17	*NA*
14	6348	CCL3		gene	165	231	46	17	*NA*
18	6354	CCL7	6354	gene	165	294	46	17	1
22	6354	CCL7	6354	gene	165	318	46	17	1
26	1230	CCR1	1230	gene	258	365	46	17	1
32	6347	CCL2	6347,6354,6355,6357	gene	165	456	46	17	4
34	729230	CCR2	729230	gene	258	456	46	17	1
36	6347	CCL2	6347	gene	165	489	46	17	1

1-10 of 147 rows | 1-10 of 11 columns Previous **1** 2 3 4 5 6 … 15 Next

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

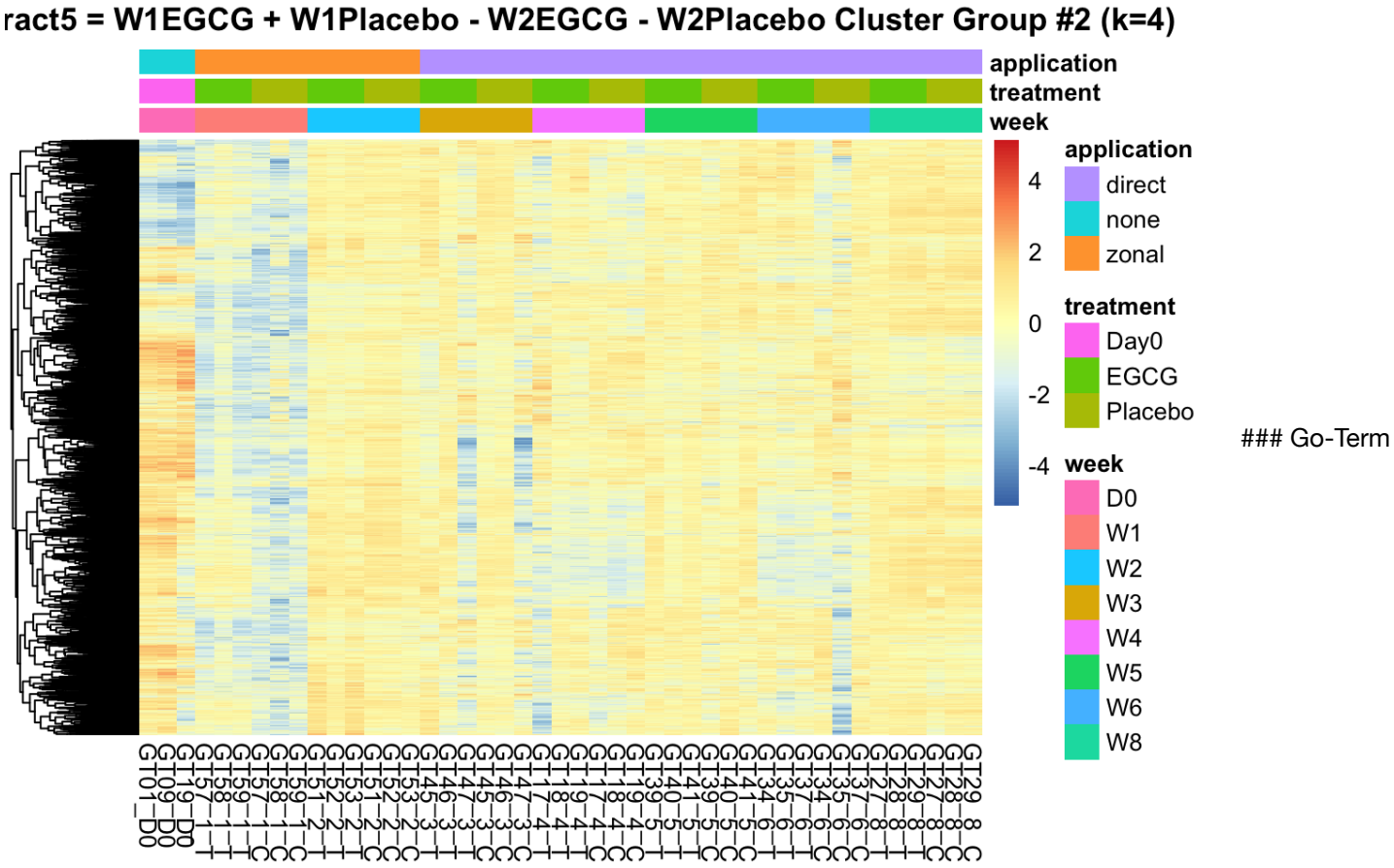
	kegg.nam…	labels	all.mapped	ty…	x	y	wi…	height	mol.data
	<chr>	<chr>	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
4	1029	CDKN2A		gene	532	218	46	17	*NA*
5	51343	FZR1		gene	981	630	46	17	*NA*
6	4171	MCM2	4171,4172,4173,4174,4175,4176	gene	553	681	46	17	6
7	4998	ORC1	4998,23594	gene	494	681	46	17	2
8	51529	ANAPC11	51529,246184	gene	981	392	46	17	2

	kegg.nam…	labels	all.mapped	ty…	x	y	wi…	height	mol.data
	<chr>	<chr>	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
9	51529	ANAPC11	51529,246184	gene	981	613	46	17	2
10	6500	SKP1		gene	188	613	46	17	*NA*
11	6500	SKP1		gene	432	285	46	17	*NA*
24	983	CDK1	983	gene	780	562	46	17	1
25	701	BUB1B	701	gene	873	392	46	17	1

1-10 of 112 rows | 1-10 of 11 columns Previous **1** 2 3 4 5 6 … 12 Next

Cluster#2

```
pheatmap(interact_sig_hclust_g2,annotation_col = annotation, scale="row", cluster_cols = F, sh
ow_rownames = F, main = "interact5 = W1EGCG + W1Placebo - W2EGCG - W2Placebo 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
## 11295 GO BP ids tested (1043 have p < 0.01)
## Selected gene set size: 3980
##     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:0007399	1.054629e-48	2.024726	563.827568	858	2445
GO:0007275	8.340263e-48	1.762116	1090.989049	1456	4731
GO:0048731	4.351900e-44	1.748143	985.602874	1325	4274
GO:0009653	2.686458e-43	1.905223	615.482936	901	2669
GO:0048856	3.575570e-42	1.664018	1319.748537	1678	5723
GO:0032502	2.114507e-37	1.601060	1452.346022	1795	6298
GO:0022008	8.667301e-37	2.047773	377.268671	592	1636
GO:0048699	1.839933e-35	2.103808	329.302972	528	1428
GO:0030182	1.973337e-32	2.067897	312.930065	498	1357
GO:0009887	1.222732e-30	2.204016	239.367287	399	1038

1-10 of 979 rows | 1-6 of 7 columns Previous **1** 2 3 4 5 6 … 98 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:0007399	1.054629e-48	nervous system development
GO:0007275	8.340263e-48	multicellular organism development
GO:0048731	4.351900e-44	system development
GO:0009653	2.686458e-43	anatomical structure morphogenesis
GO:0048856	3.575570e-42	anatomical structure development
GO:0032502	2.114507e-37	developmental process

GO_ID_BP	P-value	Term
<chr>	<dbl>	<chr>
GO:0022008	8.667301e-37	neurogenesis
GO:0048699	1.839933e-35	generation of neurons
GO:0030182	1.973337e-32	neuron differentiation
GO:0009887	1.222732e-30	animal organ morphogenesis

1-10 of 979 rows Previous **1** 2 3 4 5 6 … 98 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	GeneRatio	Bg
	<chr>	<chr>	<chr>	<c
hsa04360	hsa04360	Axon guidance	90/1736	18
hsa04550	hsa04550	Signaling pathways regulating pluripotency of stem cells	63/1736	14
hsa05217	hsa05217	Basal cell carcinoma	36/1736	63
hsa04390	hsa04390	Hippo signaling pathway	67/1736	15
hsa04512	hsa04512	ECM-receptor interaction	45/1736	89
hsa04310	hsa04310	Wnt signaling pathway	71/1736	17

6 rows | 1-5 of 10 columns

```
keggEnrich
```

```
## #
## # over-representation test
## #
## #...@organism     hsa
## #...@ontology     KEGG
## #...@keytype      kegg
## #...@gene     chr [1:4688] "54829" "94241" "1513" "148741" "152503" "4969" "114899" ...
## #...pvalues adjusted by 'BH' with cutoff <0.05
## #...53 enriched terms found
## 'data.frame':    53 obs. of  9 variables:
##  $ ID         : chr  "hsa04360" "hsa04550" "hsa05217" "hsa04390" ...
##  $ Description: chr  "Axon guidance" "Signaling pathways regulating pluripotency of stem ce
lls" "Basal cell carcinoma" "Hippo signaling pathway" ...
##  $ GeneRatio  : chr  "90/1736" "63/1736" "36/1736" "67/1736" ...
##  $ BgRatio    : chr  "182/8393" "143/8393" "63/8393" "157/8393" ...
##  $ pvalue     : num  2.47e-18 1.81e-10 2.21e-10 2.60e-10 3.18e-10 ...
##  $ p.adjust   : num  8.15e-16 1.69e-08 1.69e-08 1.69e-08 1.69e-08 ...
##  $ qvalue     : num  6.34e-16 1.31e-08 1.31e-08 1.31e-08 1.31e-08 ...
##  $ geneID     : chr  "6586/7869/4756/1949/23365/6092/59277/2043/57556/7976/80031/7225/9165
3/2051/10512/2042/54437/56288/818/6608/2231"| __truncated__ "3400/8324/11211/8323/8313/7482/22
60/7976/92/51384/4089/657/7473/2263/4086/1856/659/652/4090/7475/84333/3977/463"| __truncated__
"8324/11211/8323/8313/7482/7976/2737/51384/6608/7473/1856/652/1643/5727/7475/54361/8321/51684/
89780/80326/7481/6"| __truncated__ "7042/25937/8324/11211/8323/154796/8313/7482/7976/1741/174
0/7161/56288/51384/4089/657/7473/4086/166824/1856/659/"| __truncated__ ...
##  $ Count      : int  90 63 36 67 45 71 81 115 62 77 ...
## #...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 = "hsa04360", 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 hsa04360.pathview.png
```

```
#Repeat for the second result
pv.out_htmp2b <- pathview(gene.data = diffexpgenes_names_df$entrez, pathway.id = "hsa04550", 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 hsa04550.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_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>
5	84552	PARD6G	84552,84612	gene	778	1254	46	17	2
6	56288	PARD3	56288	gene	778	1220	46	17	1
16	2770	GNAI1	2770	gene	233	606	46	17	1
17	2770	GNAI1	2770	gene	603	606	46	17	1
18	7852	CXCR4		gene	509	606	46	17	*NA*
19	5594	MAPK1	5594	gene	690	993	46	17	1
20	5881	RAC3	5881	gene	603	870	46	17	1
21	5881	RAC3	5881	gene	690	529	46	17	1
22	5881	RAC3	5881	gene	777	240	46	17	1
23	1630	DCC		gene	509	259	46	17	*NA*

1-10 of 139 rows | 1-10 of 11 columns Previous **1** 2 3 4 5 6 … 14 Next

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

	kegg.na… <chr>	labels <chr>	all.mapped <chr>	t… <chr>	x <dbl>	<…
16	7473	WNT3	7473,7475,7480,7481,7482,7483,51384,54361,80326,89780	gene	148	4
17	10297	APC2	10297	gene	518	4
18	8313	AXIN2	8313	gene	472	4
19	1856	DVL2	1856,1857	gene	339	4
20	7976	FZD3	7976,8321,8323,8324,11211	gene	246	4
22	1499	CTNNB1	1499	gene	512	4

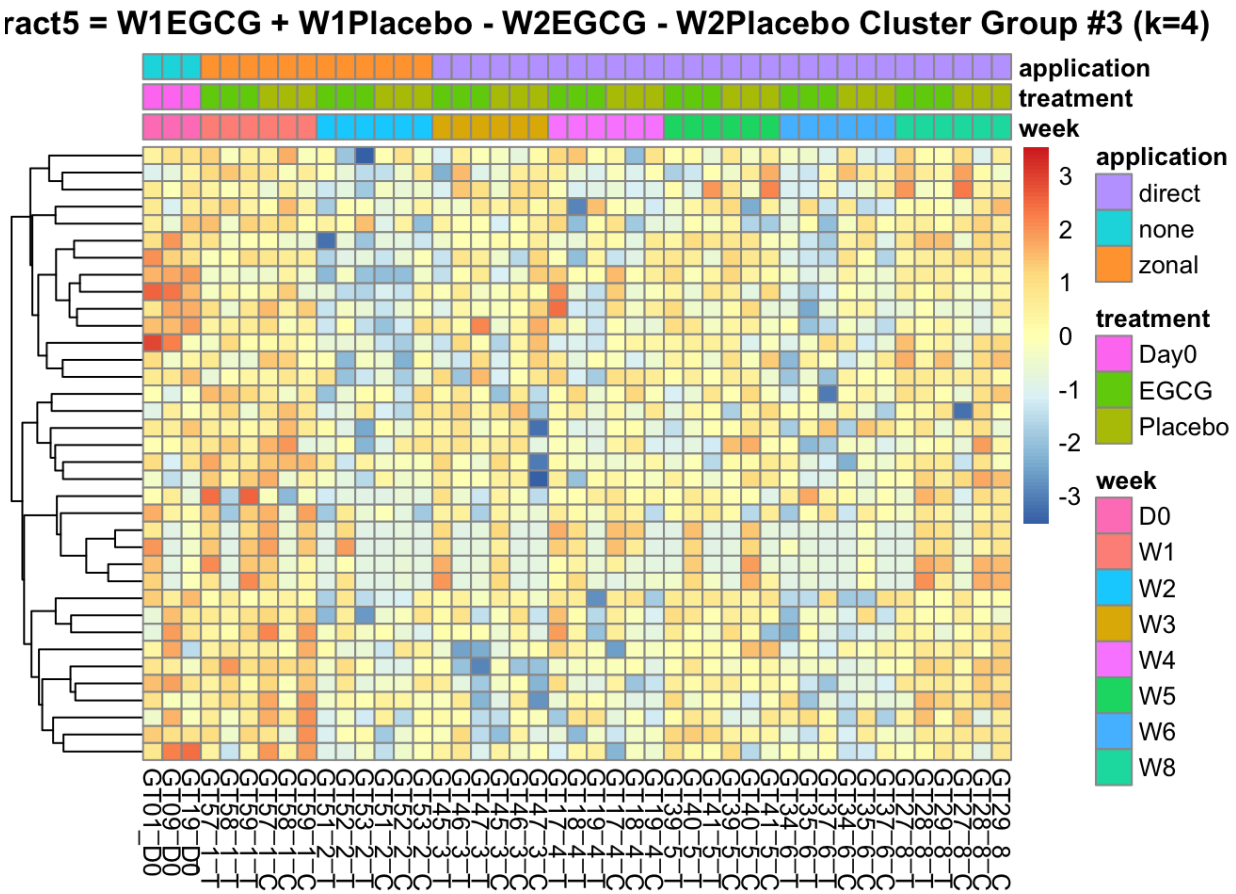
	kegg.na…	labels	all.mapped	t…	x
	<chr>	<chr>	<chr>	<chr>	<dbl>
23	2932	GSK3B	2932	gene	429
24	6932	TCF7		gene	611
38	3976	LIF		gene	148
39	3977	LIFR	3977	gene	244

1-10 of 98 rows | 1-7 of 11 columns

Previous 1 2 3 4 5 6 … 10 Next

Cluster#3

```
pheatmap(interact_sig_hclust_g3,annotation_col = annotation, scale="row", cluster_cols = F, sh
ow_rownames = F, main = "interact5 = W1EGCG + W1Placebo - W2EGCG - W2Placebo 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: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
## 778 GO BP ids tested (44 have p < 0.01)
## Selected gene set size: 18
##     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:0048711	7.958967e-05	195.795455	0.01355814	2	13
GO:0048710	4.101989e-04	79.694444	0.03024509	2	29
GO:0045687	8.223630e-04	55.134615	0.04276030	2	41
GO:0010001	1.485409e-03	15.545205	0.23153137	3	222
GO:0014015	1.993310e-03	34.635081	0.06674778	2	64
GO:0045685	2.514791e-03	30.662500	0.07509126	2	72
GO:0048708	3.018268e-03	27.863636	0.08239180	2	79
GO:0042063	3.463428e-03	11.449324	0.31183730	3	299
GO:0014013	4.694000e-03	22.092784	0.10325048	2	99
GO:1904029	5.660520e-03	20.016355	0.11367982	2	109

1-10 of 17 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:0048711	7.958967e-05	positive regulation of astrocyte differentiation
GO:0048710	4.101989e-04	regulation of astrocyte differentiation
GO:0045687	8.223630e-04	positive regulation of glial cell differentiation
GO:0010001	1.485409e-03	glial cell differentiation
GO:0014015	1.993310e-03	positive regulation of gliogenesis
GO:0045685	2.514791e-03	regulation of glial cell differentiation

GO_ID_BP <chr>	P-value <dbl>	Term <chr>
GO:0048708	3.018268e-03	astrocyte differentiation
GO:0042063	3.463428e-03	gliogenesis
GO:0014013	4.694000e-03	regulation of gliogenesis
GO:1904029	5.660520e-03	regulation of cyclin-dependent protein kinase activity

1-10 of 17 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 <chr>	Description <chr>	GeneRatio <chr>	BgRatio <chr>	pvalue <dbl>
hsa04630	hsa04630	JAK-STAT signaling pathway	2/6	166/8393	0.005535674
hsa04060	hsa04060	Cytokine-cytokine receptor interaction	2/6	295/8393	0.016816682
hsa05321	hsa05321	Inflammatory bowel disease	1/6	65/8393	0.045590179
hsa04512	hsa04512	ECM-receptor interaction	1/6	89/8393	0.061979387
hsa04658	hsa04658	Th1 and Th2 cell differentiation	1/6	92/8393	0.064011445
hsa05222	hsa05222	Small cell lung cancer	1/6	92/8393	0.064011445

6 rows | 1-7 of 10 columns

```
keggEnrich
```

```
## #
## # over-representation test
## #
## #...@organism     hsa
## #...@ontology     KEGG
## #...@keytype      kegg
## #...@gene     chr [1:28] "4884" "1285" "374286" "100506305" "79628" "196913" "221442" ...
## #...pvalues adjusted by 'BH' with cutoff <0.2
## #...20 enriched terms found
## 'data.frame':    20 obs. of  9 variables:
##  $ ID         : chr  "hsa04630" "hsa04060" "hsa05321" "hsa04512" ...
##  $ Description: chr  "JAK-STAT signaling pathway" "Cytokine-cytokine receptor interaction"
"Inflammatory bowel disease" "ECM-receptor interaction" ...
##  $ GeneRatio  : chr  "2/6" "2/6" "1/6" "1/6" ...
##  $ BgRatio    : chr  "166/8393" "295/8393" "65/8393" "89/8393" ...
##  $ pvalue     : num  0.00554 0.01682 0.04559 0.06198 0.06401 ...
##  $ p.adjust   : num  0.122 0.131 0.131 0.131 0.131 ...
##  $ qvalue     : num  0.111 0.119 0.119 0.119 0.119 ...
##  $ geneID     : chr  "3595/3976" "3595/3976" "3595" "1285" ...
##  $ Count      : int  2 2 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 = "hsa04630", 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 hsa04630.pathview.png
```

```
pv.out_htmp3b <- 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
```

```
#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>
97	6772	STAT1		gene	369	270	46	17	*NA*
98	3716	JAK1		gene	245	270	46	17	*NA*
101	3559	IL2RA		gene	199	270	46	17	*NA*
102	6772	STAT1		gene	492	261	46	17	*NA*
103	3558	IL2		gene	105	270	46	17	*NA*
105	6774	STAT3		gene	369	524	46	17	*NA*
106	3717	JAK2		gene	245	524	46	17	*NA*
109	1441	CSF3R		gene	199	524	46	17	*NA*
110	6774	STAT3		gene	492	515	46	17	*NA*
111	1440	CSF3		gene	105	524	46	17	*NA*

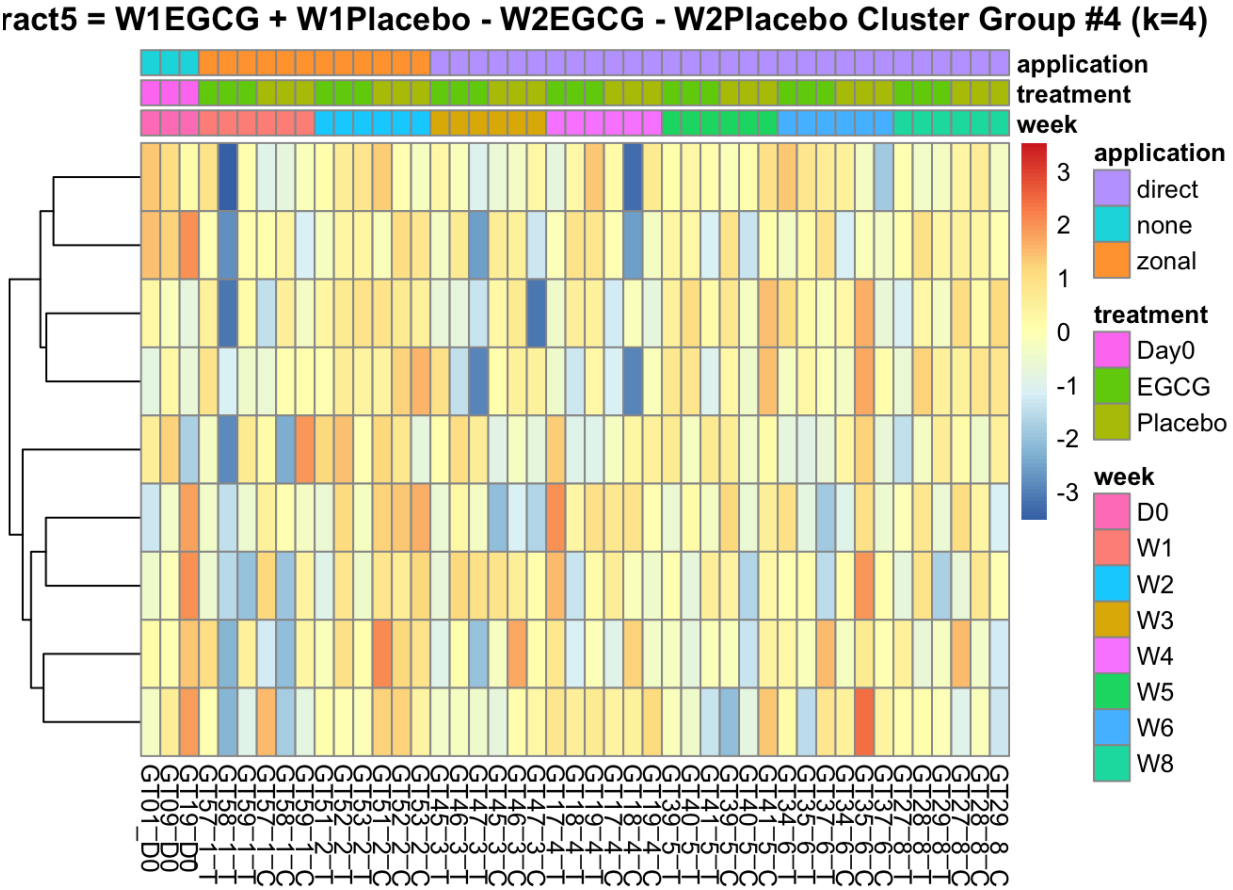
1-10 of 85 rows | 1-10 of 11 columns Previous **1** 2 3 4 5 6 … 9 Next

```
pv.out_htmp3b$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*

Cluster#4

```
pheatmap(interact_sig_hclust_g4,annotation_col = annotation, scale="row", cluster_cols = F, sh
ow_rownames = F, main = "interact5 = W1EGCG + W1Placebo - W2EGCG - W2Placebo 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
## 85 GO BP ids tested (4 have p < 0.01)
## Selected gene set size: 3
##     Gene universe size: 17259
##     Annotation package: org.Hs.eg
```

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

GOBPID	Pvalue	OddsRatio	ExpCount	Co…	S…	Term
<chr>	<dbl>	<dbl>	<dbl>	<int>	<int>	<chr>
GO:0006957	0.002778741	574.7000	0.002781158	1	16	complement activation, alternative pathway
GO:0019835	0.004166179	374.6304	0.004171736	1	24	cytolysis
GO:0030574	0.007283208	209.9390	0.007300539	1	42	collagen catabolic process
GO:0060976	0.007974997	191.2333	0.007995828	1	46	coronary vasculature development

4 rows

```
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:0006957	0.002778741	complement activation, alternative pathway
GO:0019835	0.004166179	cytolysis
GO:0030574	0.007283208	collagen catabolic process
GO:0060976	0.007974997	coronary vasculature development

4 rows

KEGG ENRIGHMENT Part4

```
#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	GeneRatio	BgRatio	pvalue	p.a
	<chr>	<chr>	<chr>	<chr>	<dbl>	
hsa04610	hsa04610	Complement and coagulation cascades	1/1	86/8393	0.01024663	0.032

	ID <chr>	Description <chr>	GeneRatio <chr>	BgRatio <chr>	pvalue <dbl>	p.a
hsa05146	hsa05146	Amoebiasis	1/1	102/8393	0.01215298	0.032
hsa05322	hsa05322	Systemic lupus erythematosus	1/1	136/8393	0.01620398	0.032
hsa04810	hsa04810	Regulation of actin cytoskeleton	1/1	229/8393	0.02728464	0.032
hsa05171	hsa05171	Coronavirus disease - COVID-19	1/1	232/8393	0.02764208	0.032
hsa05020	hsa05020	Prion disease	1/1	273/8393	0.03252711	0.032

6 rows | 1-7 of 10 columns

```
keggEnrich
```

```
## #
## # over-representation test
## #
## #...@organism     hsa
## #...@ontology     KEGG
## #...@keytype      kegg
## #...@gene     chr [1:6] "735" "92806" NA "116328" "118856" "51059"
## #...pvalues adjusted by 'BH' with cutoff <0.05
## #...6 enriched terms found
## 'data.frame':    6 obs. of  9 variables:
##  $ ID         : chr  "hsa04610" "hsa05146" "hsa05322" "hsa04810" ...
##  $ Description: chr  "Complement and coagulation cascades" "Amoebiasis" "Systemic lupus ery
thematosus" "Regulation of actin cytoskeleton" ...
##  $ GeneRatio  : chr  "1/1" "1/1" "1/1" "1/1" ...
##  $ BgRatio    : chr  "86/8393" "102/8393" "136/8393" "229/8393" ...
##  $ pvalue     : num  0.0102 0.0122 0.0162 0.0273 0.0276 ...
##  $ p.adjust   : num  0.0324 0.0324 0.0324 0.0325 0.0325 ...
##  $ qvalue     : logi  NA NA NA NA NA NA
##  $ geneID     : chr  "735" "735" "735" "735" ...
##  $ Count      : int  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_htmp4a <- pathview(gene.data = diffexpgenes_names_df$entrez, pathway.id = "hsa04610", 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 hsa04610.pathview.png
```

```
pv.out_htmp4b <- pathview(gene.data = diffexpgenes_names_df$entrez, pathway.id = "hsa05146", 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 hsa05146.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>
373	2155	F7		gene	90	287	46	17	*NA*
375	2152	F3		gene	114	483	46	17	*NA*
381	2161	F12		gene	412	218	46	17	*NA*
383	2160	F11		gene	340	252	46	17	*NA*
389	2158	F9		gene	268	287	46	17	*NA*
390	7450	VWF		gene	299	350	46	17	*NA*
391	2157	F8		gene	299	367	46	17	*NA*
396	2159	F10		gene	155	425	46	17	*NA*
399	2147	F2		gene	102	734	46	17	*NA*

	kegg.names <chr>	labels <chr>	all.mapped <chr>	type <chr>	x <dbl>	y <dbl>	width <dbl>	height <dbl>	mol.data <dbl>
403	2153	F5		gene	186	486	46	17	*NA*

1-10 of 74 rows | 1-10 of 11 columns Previous **1** 2 3 4 5 6 … 8 Next

```
pv.out_htmp4b$plot.data.gene
```

	kegg.names <chr>	labels <chr>	all.mapped <chr>	type <chr>	x <dbl>	y <dbl>	width <dbl>	height <dbl>	mol.data <dbl>
23	3553	IL1B		gene	382	185	46	17	*NA*
24	4583	MUC2		gene	397	234	46	17	*NA*
26	2335	FN1		gene	397	288	46	17	*NA*
27	3908	LAMA2		gene	397	307	46	17	*NA*
35	4790	NFKB1		gene	519	185	46	17	*NA*
36	3576	CXCL8		gene	1138	503	46	17	*NA*
37	3569	IL6		gene	1138	429	46	17	*NA*
38	836	CASP3		gene	489	466	46	17	*NA*
41	2919	CXCL1		gene	1138	448	46	17	*NA*
42	1437	CSF2		gene	1138	467	46	17	*NA*

1-10 of 46 rows | 1-10 of 11 columns Previous **1** 2 3 4 5 Next