EGCG blood sample gene expression Analysis Comparing overweight females before and after EGCG and then with Males who are healthy and males who have myocardial infarctions

Janis Corona

12/25/2019

This Rmarkdown file uses R to analyze the gene expressions of two different National Center for Bioinformatics Gene Expression Omnibus (GEO) studies on blood samples. One study was done on overweight women treated with and without epigallocatechin (EGCG) which is a green tea flavanoid, and the other studied males who were healthy against those males with myocardial infarction or MI for short. These studies can be found and manually extracted for data samples pulling each sample data one by one in some cases. They are both array studies of blood gene expressions in their respective study. To verify this data, follow these GEO links:

MI males: - https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE141512

overweight females using EGCG before and after: - https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE74560

The results of this analysis provided insight into the top genes that showed the highest or the lowest fold change in comparison between five different groups between sample membership by using the means of each gene across each sample and the fold change in one group compared to another.

The gene HBZ is overexpressed in overweight females compared to either group of males by 30% and by 20% in overweight females using EGCG compared to overweight females not using EGCG. But shows a negligible change between males with MI compared to healthy males by less than 1%. HBZ is an iron ion binding and oxygen binding protein coding gene and this gene increases in overweight females when drinking EGCG green tea extract.

The gene THBS1 was found to be under expressed in overweight females who use EGCG and over expressed in males with MI.Given the gene function summary of THBS1 from genecards.org, this could imply that THBS1 relaxes the nervous system in the females and that it is involved in angiogenesis in males with the heart disease MI.Since healthy males have a 20% under expression when compared to overweight females who don't use EGCG, it could mean that the overweight females also have angiogenesis in their bodies related to heart disease, tumors forming, and blood clots from platelet aggregation. This gene could be a target to look for when determining the health of a person, the state of stress, or risk of other diseases just mentioned.

The gene OR7D2 was found in males with MI to have a 15% under expression compared to healthy males and females are 36-39% over expressing this gene compared to males with or without MI. Given the gene function for OR7D2, this could mean that males with MI are having smells blocked more than healthy males and that overweight females use their sense of smell more than males. MOre importantly, the drop in this gene expression in males with MI implies they have their smell blocked by other neuron responses like pain or side effects of having MI.

The gene CA1 is overexpressed in overweight females more than 375% the amount of healthy males and 367% more than males with MI. The changes in expression are almost negligible within each subcategory of gender comparisons. The gene summary for CA1 implies that overweight females might be expressing a lot of this gene than males because they have more gastric acid, or bone resorption of calcium, or have more saliva, create more cerebrospinal fluid, have more carbon dioxide in their blood that needs to be put through reversible hydration. Since the overweight females using EGCG show decreased gene expression levels of this gene by 3% compared to those overweight females not using EGCG, this could mean that EGCG lowers the need for this gene in those biological processes.

The gene MIR548AD is under expressed in overweight females compared to males with MI and only slightly in males with MI compared to healthy males and also more expressed in overweight females using EGCG compared to overweight females not using EGCG. This implies that males with MI and females using EGCG have more gene silencing or gene stabilization happening in their bodies. Since MIR548AD is a microRNA that is involved in stabilizing and translating messenger RNA (mRNA), it suggest that males have a faster metabolism than females and that overweight females who use EGCG increase their metabolism above overweight females who don't use EGCG.

Read in the EGCG files of pre and post treatment of overweight females given EGCG-green tea extract or EGCG AND flavanoids

```
#all pre EGCG samples totaling 14
PRE <- read.csv('pre_combined.csv', sep=',', header=TRUE,
                na.strings=c('',' '), row.names=1)

#EGCG treatment samples
Post_E <- read.csv('Post_EGCG_combine.csv', sep=',', header=TRUE,
                   na.strings=c('',' '), row.names=1)
#EGCG AND flavanoids in Quercentine with Vitamin C and fish oil
Post_Q <- read.csv('Post_EGCG_Quer_combine.csv', sep=',', header=TRUE,
                   na.strings=c('',' '), row.names=1)
```

Get the means of the pre-treatment samples with rowMeans() per gene and the post-treatment samples of EGCG and EGCG+Quercentin, then attach to each of those tables as the mean of those samples of gene expression values.

```
PRE$PRE_Means <- rowMeans(PRE)
Post_Q$Post_Q_Means <- rowMeans(Post_Q)
Post_E$Post_E_Means <- rowMeans(Post_E)
```

```
colnames(PRE)[1:14] <- paste('pre_',sep='', colnames(PRE)[1:14])

colnames(Post_E)[1:7] <- paste('post_EG_', sep='',colnames(Post_E)[1:7])

colnames(Post_Q)[1:7] <- paste('post_EQ_', sep='', colnames(Post_Q)[1:7])
```

Combine all tables into one with the means of each type of sample at the start of the columns

```
All <- cbind(PRE[,15],Post_E[,8],Post_Q[,8],PRE[,1:14], Post_E[1:7],
             Post_Q[1:7])
colnames(All)[1:3] <-
c('Pre_Means','Post_EGCG_Means','Post_EGCG_Quercentin_Means')
```

Now read in the gene expression data from the healthy and the myocardial infarction (MI) male samples with 6 each.

```
Healthy_MI <- read.csv('HealthyMales_6_bloodSamples_MI_study.csv',
                       sep=',', header=TRUE, na.strings=c('',' '))
MI <- read.csv('MI_nonHealthyMales_6_bloodSamples_MI_study.csv',
                       sep=',', header=TRUE, na.strings=c('',' '))
```

Attach the type of sample to the header names in the above data sets.

```
colnames(Healthy_MI)[2:7] <- paste('healthy_Males_',
                                   colnames(Healthy_MI)[2:7],sep='')
colnames(MI)[2:7] <- paste('MI_Males_',
                                   colnames(MI)[2:7],sep='')
```

Attach the platform genes (a large 177 mb file) to those data files separately so that the genes can be identified by their gene symbol instead of Machine ID.

```
platform_MI <- read.delim('GPL17586-45144.txt', sep='\t', header=TRUE,
                          na.strings=c('',' ','---'), comment.char = '#')
```

keep only the ID and the gene_assignment fields

```
plf_MI <- platform_MI[,c(1,8)]
```

keep only the first value in the list of gene_assignments with multiple values separated by '//', but after keeping only the complete.cases(gene_assignment)

```
plf_MI_cc <- plf_MI[complete.cases(plf_MI$gene_assignment),]

Sym <- strsplit(as.character(plf_MI_cc$gene_assignment), '//')
Ref_Seq <- lapply(Sym,'[',1)
Symbol <- lapply(Sym,'[',2) #this is the gene_assignment interested in
HGNC <- lapply(Sym,'[',3)

plf_MI_cc$Symbol <- Symbol
plf_sym <- plf_MI_cc[,c(1,3)]
plf_sym <- plf_sym[complete.cases(plf_sym$Symbol),]
```

```
dash <- grep('---',plf_sym$Symbol)
platform <- plf_sym[-dash,]
```

Now merge the platform gene symbol with the IDs of the two MI data sets

```
HealthyMales <- merge(platform, Healthy_MI, by.x='ID', by.y='ID_REF')
MI_Males <- merge(platform, MI, by.x='ID', by.y='ID_REF')

#drop the machine ID now
HealthyMales <- HealthyMales[,-1]
MI_Males <- MI_Males[,-1]
```

Here we will get the means for the MI_Males and the HealthyMales data sets

```
library(dplyr)

Symbol1 <- t(as.data.frame(MI_Males$Symbol))
colnames(Symbol1) <- 'Symbol'

Symbol2 <- t(as.data.frame(HealthyMales$Symbol))
colnames(Symbol2) <- 'Symbol'

MI_males <- cbind(Symbol1,MI_Males[2:7])
Healthymales <- cbind(Symbol2, HealthyMales[2:7])

samples <- colnames(MI_Males)[2:7]
Samples <- as.vector(samples)
MI <- MI_males %>% group_by(Symbol) %>%
  summarise_at(vars(Samples), mean, na.rm=TRUE)

samples1 <- colnames(HealthyMales)[2:7]
Samples1 <- as.vector(samples1)
Healthy <- Healthymales %>% group_by(Symbol) %>%
  summarise_at(vars(Samples1), mean, na.rm=TRUE)
```

Add a row means of each gene in the healthy and MI data sets

```
Healthy <- as.data.frame(Healthy)
MI <- as.data.frame(MI)

rownames(Healthy) <- Healthy$Symbol
rownames(MI) <- MI$Symbol

Healthy <- Healthy[,-1]
MI <- MI[,-1]

Healthy$HealthyMale_Means <- rowMeans(Healthy)
MI$MI_Male_Means <- rowMeans(MI)

Healthy$Symbol <- row.names(Healthy)
MI$Symbol <- row.names(MI)
```

```
Healthy <- Healthy[,c(8,7,1:6)]
MI <- MI[,c(8,7,1:6)]

MI_means <- cbind(Healthy,MI)
MI_means1 <- MI_means[,c(1,2,10,3:8,11:16)]
FC_MI <- MI_means1 %>% mutate( FC_MI_males = MI_Male_Means/HealthyMale_Means)
FC_MI1 <- FC_MI[,c(1,16,2:15)]
row.names(FC_MI1) <- FC_MI$Symbol
write.csv(FC_MI1,'foldChange_MI_males.csv', row.names=TRUE)
```

The above last two sets of code produced means per gene and there were less observations in the new data sets afterwards with our two new data sets of original genes called 'Healthy' and 'MI' to work with.

We should merge the two data sets to see what genes they have in common between the female EGCG blood samples and the male myocardial infarction blood samples.

```
All$Symbol <- as.factor(row.names(All))
MI$Symbol <- gsub(' ','',MI$Symbol)
Healthy$Symbol <- gsub(' ','', Healthy$Symbol)

BothStudiesMI <- merge(All, MI, by.x='Symbol', by.y='Symbol')

BothStudies1 <- merge(BothStudiesMI, Healthy, by.x='Symbol', by.y='Symbol')
mn <- grep('Mean',colnames(BothStudies1))
BothStudies0 <- BothStudies1[,mn]
BothStudies2 <- BothStudies1[,-mn]
BothStudies3 <- cbind(BothStudies0, BothStudies2)
row.names(BothStudies3) <- BothStudies3$Symbol
BothStudies3 <- BothStudies3[,-6]
write.csv(BothStudies3, 'EGCG-MI-means.csv', row.names=TRUE)
```

There are 18278 genes in common between the two completely different studies on EGCG effects on overweight women and myocardial infarction in males that both used blood tissue for gene expression analysis.

```
row.names(BothStudies1) <- BothStudies1$Symbol
colnames(BothStudies1)[2] <- 'preEGCG_Means'

means <- grep('Means', colnames(BothStudies1))
Both <- BothStudies1[,c(means)]
#the above has only the gene means per group of samples
```

Write these means to a csv file to compare the differential expression and fold change between groups later.

```
write.csv(Both, 'BothStudies.csv', row.names=TRUE)
```

If you don't have BothStudies.csv in your environment, then read it in now

```
BothStudies <- read.csv("BothStudies.csv", sep=',', header=TRUE,
                        na.strings=c('',' '), row.names=1)
```

I want to compare five groups to each other: - post EGCG females to overweight females - MI males to healthy males - MI males to overweight females - post EGCG females to healthy males - MI males to post EGCG females

I will do this by using the fold change of the ratio of: - post EGCG females to overweight females - MI males to healthy males - MI males to overweight females - post EGCG females to healthy males - MI males to post EGCG females

```
library(dplyr)
```

The fold change for each group.

```
eg_2_ow <- BothStudies %>%
mutate(Fold_Change_Post_EGCG_to_Overweight_Females=Post_EGCG_Means/preEGCG_Me
ans)

MI_2_healthy <- eg_2_ow %>%
mutate(Fold_Change_MI_Males_to_Healthy_Males=MI_Male_Means/HealthyMale_Means)

ow_2_MI <- MI_2_healthy %>%
mutate(Fold_Change_Overweight_Females_to_MI_Males=preEGCG_Means/MI_Male_Means
)

ow_2_healthy <- ow_2_MI %>%
mutate(Fold_Change_Overweight_Females_to_Healthy_Males=preEGCG_Means/HealthyM
ale_Means)

eg_2_MI <- ow_2_healthy %>%
mutate(Fold_Change_Post_EGCG_Females_to_MI_Males=Post_EGCG_Means/MI_Male_Mean
s)

row.names(eg_2_MI) <- row.names(BothStudies)
```

The fold change shows how much the sample changed or is different in one group to another. If one group is double the change of another, then the fold change will be 2 or more, and if it is half the gene expression as the other it will be 0.5 or close to 0.5.

Lets write this file out as the fold change of four groups

```
write.csv(eg_2_MI, 'fold_change_5_groups-OW.csv', row.names=TRUE)
```

Read in the above file if you haven't already or if you are coming back to this file.

```
eg_2_MI <- read.csv('fold_change_5_groups-OW.csv', sep=',', row.names=1)
```

What genes were the 5 most expressed when comparing overweight females to overweight females treated with EGCG green tea extract?

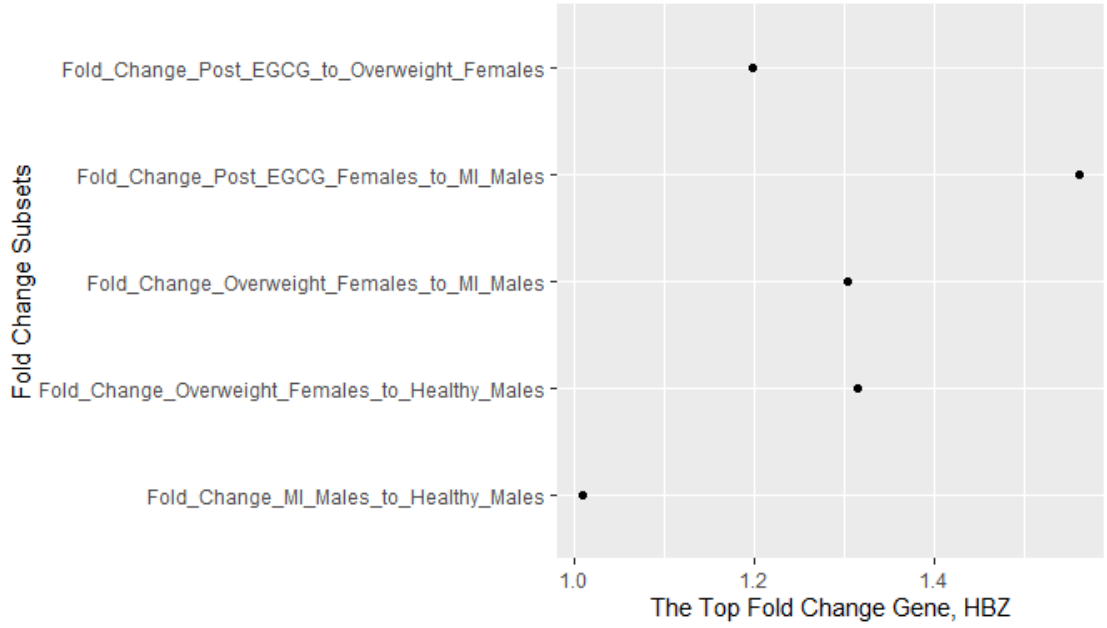
```
Top5_EGCG <-
eg_2_MI[order(eg_2_MI$Fold_Change_Post_EGCG_to_Overweight_Females,decreasing=
TRUE)[1:5],6:10]
row.names(Top5_EGCG) <-
row.names(eg_2_MI[order(eg_2_MI$Fold_Change_Post_EGCG_to_Overweight_Females,d
ecreasing=TRUE),])[1:5]
row.names(Top5_EGCG)

## [1] "HBZ"       "FAU"       "RNA5SP449" "TMEM176A"  "TMEM176B"
```

Plot the top gene most expressed, HBZ, in fold change of overweight females treated with EGCG compared to overweight females before treatment, and compare this gene expression in the other four groups.

```
library(ggplot2)
Top5_EGCG_t <- as.data.frame(t(Top5_EGCG))

ggplot(Top5_EGCG_t, aes(x = Top5_EGCG_t$HBZ, y = row.names(Top5_EGCG_t)))+
      geom_point() +
  ylab(label="Fold Change Subsets") +
  xlab("The Top Fold Change Gene, HBZ")
```



From the above plot, it is clear that HBZ is under expressed in males with myocardial infarction (MI) compared to overweight females who were not treated with EGCG by 20%

and those who were treated with EGCG by 30%. This is the same over expression of HBZ in overweight females not treated with EGCG compared to healthy males of approximately 30%.When comparing this expression of HBZ in MI males to healthy males it is only slightly increased with less than 1% by looking at the table.

```
Top5_EGCG_t[,1:2]
```

```
##                                                       HBZ       FAU
## Fold_Change_Post_EGCG_to_Overweight_Females     1.197956 1.1951983
## Fold_Change_MI_Males_to_Healthy_Males           1.008955 0.9841772
## Fold_Change_Overweight_Females_to_MI_Males      1.304346 0.4367278
## Fold_Change_Overweight_Females_to_Healthy_Males 1.316027 0.4298175
## Fold_Change_Post_EGCG_Females_to_MI_Males       1.562549 0.5219763
```

This could mean that HBZ doesn't really change much in males as it does in females compared to males. But using the table above, it is significantly overexpressed in overweight females by 30-56% when compared to males with MI who were not treated with EGCG and who were treated with EGCG respectively. Overweight females without EGCG treatment have 31.6% more HBZ gene expression than healthy males.

So, what is HBZ? The site genecards.org calls it Hemoglobin Subunit Zeta with 5 functional genes and 2 pseudogenes. HBZ is an iron ion binding and oxygen binding protein coding gene. This gene increases in overweight females when drinking EGCG green tea extract.

What genes were the 5 least expressed when comparing overweight females to overweight females treated with EGCG green tea extract?

```
Least5_EGCG <-
eg_2_MI[order(eg_2_MI$Fold_Change_Post_EGCG_to_Overweight_Females,decreasing=
FALSE)[1:5],6:10]
row.names(Least5_EGCG) <-
row.names(eg_2_MI[order(eg_2_MI$Fold_Change_Post_EGCG_to_Overweight_Females,d
ecreasing=FALSE),])[1:5]
row.names(Least5_EGCG)
```

```
## [1] "VTRNA1-1"  "TMTC1"     "RNU4-2"    "LINC00189" "KIF4A"
```

Plot the least expressed genes in fold change of overweight females treated with EGCG to overweight females not treated with EGCG.

```
Least5_EGCG_t <- as.data.frame(t(Least5_EGCG))
Least5_EGCG_t[,1:2]
```

```
##                                                  VTRNA1-1     TMTC1
## Fold_Change_Post_EGCG_to_Overweight_Females     0.8705886 0.8968263
## Fold_Change_MI_Males_to_Healthy_Males           0.9802894 0.9919259
## Fold_Change_Overweight_Females_to_MI_Males      0.8676074 1.2593883
## Fold_Change_Overweight_Females_to_Healthy_Males 0.8505064 1.2492198
## Fold_Change_Post_EGCG_Females_to_MI_Males       0.7553291 1.1294525
```

```
ggplot(Least5_EGCG_t, aes(x = Least5_EGCG_t$`VTRNA1-1`, y =
row.names(Least5_EGCG_t)))+
      geom_point() +
  ylab(label="Fold Change Subsets") +
  xlab("The Least Fold Change Gene, VTRNA1-1")
```

The most under expressed gene, VTRNA1-1, in fold change of overweight women treated with EGCG compared to overweight women not treated with EGCG is 13% under expressed in overweight females who aren't taking EGCG. When looking at this gene in the other group comparisons, VTRNA1-1 is 15% under expressed in overweight females not treated with EGCG compared to healthy males. This gene is only 2% under expressed in males with MI compared to healthy males. It is 32% under expressed in overweight females treated with EGCG compared to males with MI, and 15% under expressed in overweight females not treated with EGCG compared to males with MI. This gene is clearly more under expressed in overweight females who take EGCG compared to males than overweight females not taking EGCG.

Genecards.org says this gene, VTRNA1-1, is called vault RNA 1-1, and is a large vault cytoplasmic ribonucleicprotein. It has one major vault protein and two minor vault proteins. It is of the class Vault_RNA. This isn't a whole lot to go on. Maybe looking up the other genes least and most expressed will provide results that prove or disprove there being benefits to females who use EGCG. But let us see what genes we can compare EGCG use on when comparing males with MI and healthy males.

What genes were the 5 most expressed when comparing males with Myocardial infarction to healthy males?

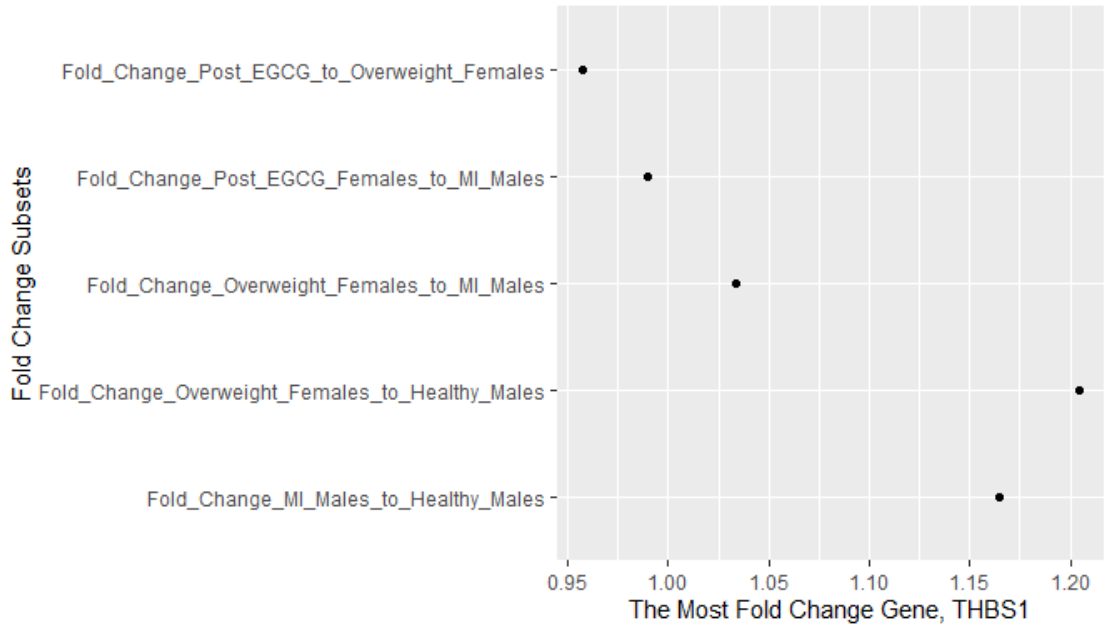
```
Top5_MI <-
eg_2_MI[order(eg_2_MI$Fold_Change_MI_Males_to_Healthy_Males,decreasing=TRUE)[
1:5],6:10]
row.names(Top5_MI) <-
row.names(eg_2_MI[order(eg_2_MI$Fold_Change_MI_Males_to_Healthy_Males,
decreasing=TRUE),])[1:5]
row.names(Top5_MI)

## [1] "THBS1" "VNN3"  "MERTK" "GPR15" "IFIT1"

Top5_MI_t <- as.data.frame(t(Top5_MI))
Top5_MI_t[,1:2]

##                                                    THBS1      VNN3
## Fold_Change_Post_EGCG_to_Overweight_Females     0.9573453 0.9827766
## Fold_Change_MI_Males_to_Healthy_Males           1.1649933 1.1616206
## Fold_Change_Overweight_Females_to_MI_Males      1.0340645 1.6756213
## Fold_Change_Overweight_Females_to_Healthy_Males 1.2046782 1.9464363
## Fold_Change_Post_EGCG_Females_to_MI_Males       0.9899568 1.6467614

ggplot(Top5_MI_t, aes(x = Top5_MI_t$THBS1, y = row.names(Top5_MI_t)))+
      geom_point() +
  ylab(label="Fold Change Subsets") +
  xlab("The Most Fold Change Gene, THBS1")
```



From the above table and plot of THBS1 gene expression as the most expressed gene in males with MI compared to healthy males by 16%. This gene is over expressed in

overweight females not treated with EGCG compared to healthy males by 20% and by 4% compared to males with MI. This gene is under expressed in overweight females taking EGCG by 1% compared to males with MI and by 5% when compared to overweight females not taking EGCG.

Genecards.org summarizes this gene as a protein coding gene named thrombospinden 1, and it codes a gene that mediates cell to cell and cell to matrix interactions. It plays a role in platelet aggregation, angiogenesis, and tumorigenesis. It also has a role in dentinogenesis or dental pulp and plays a role in ER stress response.

The fact that it is under expressed in females who use EGCG and over expressed in males with Myocardial Infarction, could mean that it relaxes the nervous system in the females and that it is involved in angiogenesis in males with the heart disease MI.Since healthy males have a 20% under expression when compared to overweight females who don't use EGCG, it could mean that the overweight females also have angiogenesis in their bodies related to heart disease, tumors forming, and blood clots from platelet aggregation. This gene could be a target to look for when determining the health of a person, the state of stress, or risk of other diseases just mentioned.

What genes were the 5 least expressed when comparing Myocardial infarction males to healthy males?

```
Least5_MI <-
eg_2_MI[order(eg_2_MI$Fold_Change_MI_Males_to_Healthy_Males,decreasing=FALSE)
[1:5],6:10]
row.names(Least5_MI) <-
row.names(eg_2_MI[order(eg_2_MI$Fold_Change_MI_Males_to_Healthy_Males,
decreasing=FALSE),])[1:5]
row.names(Least5_MI)

## [1] "OR7D2" "KLRC3" "KLRC1" "IFNG"  "KLRC2"

Least5_MI_t <- as.data.frame(t(Least5_MI))
Least5_MI_t[,1:2]

##                                                OR7D2     KLRC3
## Fold_Change_Post_EGCG_to_Overweight_Females     1.0260708 1.0088219
## Fold_Change_MI_Males_to_Healthy_Males           0.8554697 0.8674320
## Fold_Change_Overweight_Females_to_MI_Males      1.5910088 1.0366394
## Fold_Change_Overweight_Females_to_Healthy_Males 1.3610599 0.8992142
## Fold_Change_Post_EGCG_Females_to_MI_Males       1.6324877 1.0457845
```

Try plotting after formatting the data above into a column of classes of the row names for each group analyzed using tidy (gather()) then plot using ggplot with a bar chart by class but showing the genes.

```
#library(dplyr)
library(tidyr) #install.packages('tidyverse')
```

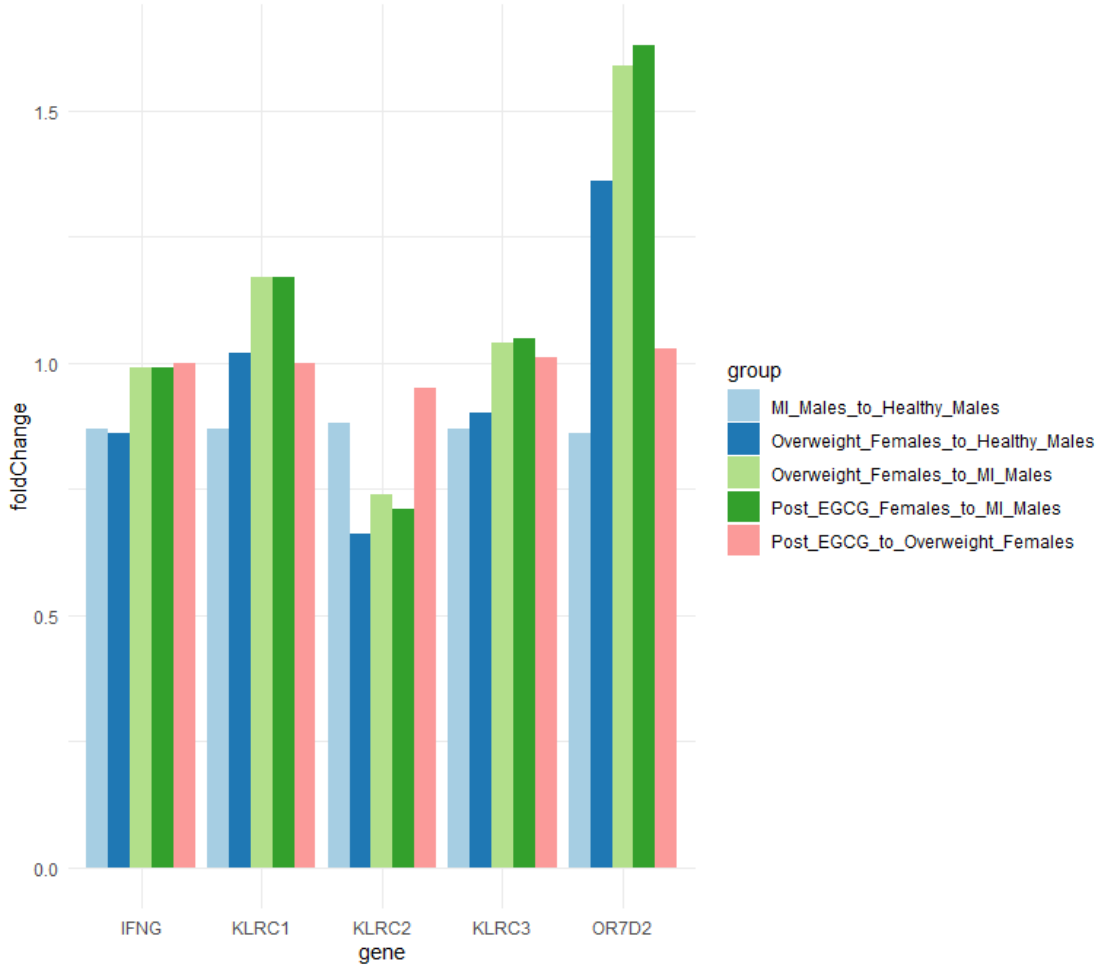
```
least_5_MI_tidy <- Least5_MI_t %>% mutate(group = row.names(Least5_MI_t))
least_5_MI_Tidy <- gather(least_5_MI_tidy, 'gene','foldChange',1:5)
least_5_MI_Tidy$foldChange <- round(least_5_MI_Tidy$foldChange,2)
head(least_5_MI_Tidy)
```

```
##                                             group  gene foldChange
## 1     Fold_Change_Post_EGCG_to_Overweight_Females OR7D2       1.03
## 2           Fold_Change_MI_Males_to_Healthy_Males OR7D2       0.86
## 3      Fold_Change_Overweight_Females_to_MI_Males OR7D2       1.59
## 4 Fold_Change_Overweight_Females_to_Healthy_Males OR7D2       1.36
## 5       Fold_Change_Post_EGCG_Females_to_MI_Males OR7D2       1.63
## 6     Fold_Change_Post_EGCG_to_Overweight_Females KLRC3       1.01
```

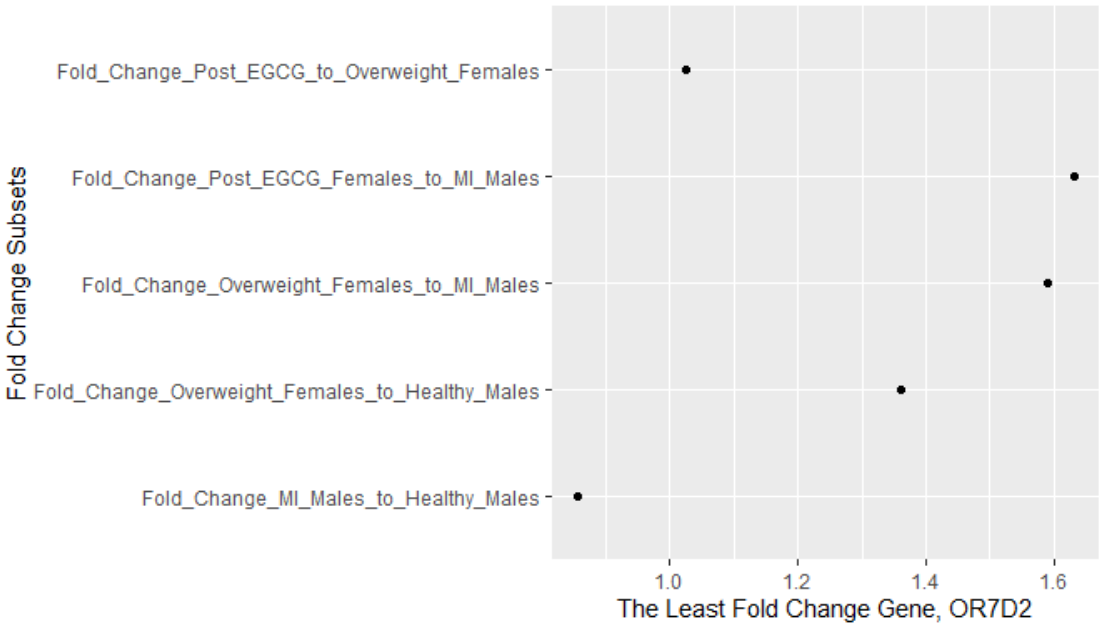
Now plot a bar chart of the fold change values of each gene in each of the five groups being analyzed using ggplot2

```
#remove the 'Fold_Change_' from the group names to plot better
least_5_MI_Tidy$group <- gsub('Fold_Change_','', least_5_MI_Tidy$group)

ggplot(data = least_5_MI_Tidy, aes(x=gene, y=foldChange, fill=group)) +
  geom_bar(stat='identity', position=position_dodge())+
  #geom_text(aes(label=foldChange), vjust = 1.0, color = 'black', position =
#position_dodge(0.5), size=2.9) +
  scale_fill_brewer(palette='Paired') + theme_minimal()
```



```
ggplot(Least5_MI_t, aes(x = Least5_MI_t$OR7D2, y = row.names(Least5_MI_t)))+
      geom_point() +
  ylab(label="Fold Change Subsets") +
  xlab("The Least Fold Change Gene, OR7D2")
```



The gene least expressed in comparing males with MI to healthy males is 15% under expressed in males with MI compared to healthy males. This gene is also over expressed 36% in overweight females compared to healthy males but also 59-63% over expressed in females who do and who don't take EGCG compared to males with MI. When comparing overweight females taking EGCG to those overweight females not taking EGCG this gene is overexpressed 2-3%.

Genecards.org summarizes this gene, OR7D2 or olfactory receptor family 7 subfamily D member 2, as a protein coding gene that is an olfactory receptor interacting with oderous molecules in the nose that stimulate a neuronal response and return the perception of a smell. The olfactory receptors share a 7-transmembrane domain structure with many neurotransmitter and hormone receptors and are responsible for the recognition and G protein-mediated transduction of odorant signals.

From the above summary of this gene, it has been shown that males with MI have a 15% under expression compared to healthy males and females are 36-39% over expressing this gene compared to male with or without MI. Judging on an outlook perspective, this could mean that males with MI are having smells blocked more than healthy males and that females use their smell to guide them more than males. MOre importantly, the drop in this gene expression in males with MI could be having their smell blocked by other neuron responses like pain or side effects of having MI.

The following group comparisons display many of the same genes in each group that is being compared.

What genes were the most expressed when comparing males with MI to overweight females?

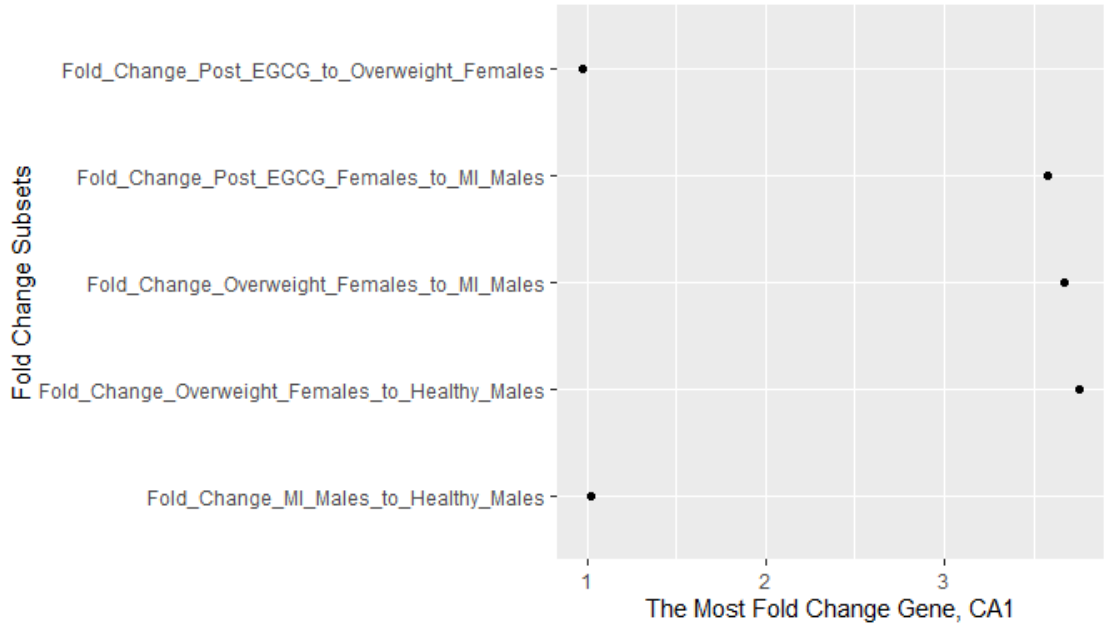
```
Top5_MI_fem <-
eg_2_MI[order(eg_2_MI$Fold_Change_Overweight_Females_to_MI_Males,decreasing=T
RUE)[1:5],6:10]
row.names(Top5_MI_fem) <-
row.names(eg_2_MI[order(eg_2_MI$Fold_Change_Overweight_Females_to_MI_Males,
decreasing=TRUE),])[1:5]
row.names(Top5_MI_fem)

## [1] "CA1"    "GYPA"   "AHSP"   "IFIT1B" "GYPB"

Top5_MI_fem_t <- as.data.frame(t(Top5_MI_fem))
Top5_MI_fem_t[,1:2]

##                                                    CA1      GYPA
## Fold_Change_Post_EGCG_to_Overweight_Females     0.9743263 0.9561337
## Fold_Change_MI_Males_to_Healthy_Males           1.0233859 0.9923072
## Fold_Change_Overweight_Females_to_MI_Males      3.6722875 3.6276161
## Fold_Change_Overweight_Females_to_Healthy_Males 3.7581674 3.5997096
## Fold_Change_Post_EGCG_Females_to_MI_Males       3.5780064 3.4684861
```

```
ggplot(Top5_MI_fem_t, aes(x = Top5_MI_fem_t$CA1, y =
row.names(Top5_MI_fem_t)))+
      geom_point() +
  ylab(label="Fold Change Subsets") +
  xlab("The Most Fold Change Gene, CA1")
```



The gene CA1 is most expressed in overweight females compared to males with MI by 367%. This gene is only slightly expressed more in males with MI compared to healthy males by 2% and only slightly under expressed in overweight females using EGCG compared to overweight females not using EGCG by 3%. There is also a huge over expression of this gene in overweight females not using EGCG compared to healthy males by more than 3 3/4 folds or 375% of an increase in gene expression.

The genecards.org summary for CA1 or Carbonic Anhydrase 1 is as a protein coding gene involved in many biological processes that includes erythrocytes that take up carbon dioxide and release oxygen. Some of the biological processes involved in are respiration, calcification, bone resorption, acid-base balance, and the formation of the following: aqueous humor, saliva, cerebrospinal fluid, and gastric acid.

This given summary above for this gene could mean many things when observing that it is overexpressed in overweight females more than 375% the amount of healthy males and 367% more than males with MI. The changes in expression are almost negligible within each subcategory of gender comparisons. Overweight females might be expressing a lot of this gene because they have more gastric acid, or bone resorption of calcium, or have more saliva, create more cerebrospinal fluid, have more carbon dioxide in their blood that needs to be put through reversible hydration. Since the overweight females using EGCG show decreased gene expression levels of this gene by 3% compared to those overweight females not using EGCG, this could mean that EGCG lowers the need for this gene in those biological processes.

What genes were the least expressed when comparing males with MI to overweight females?

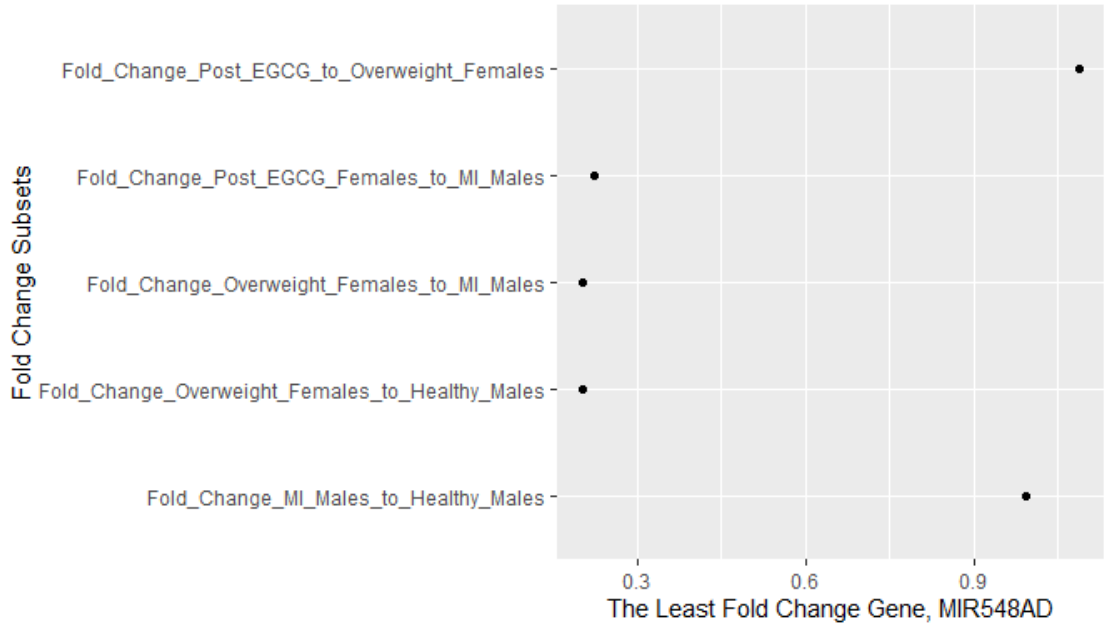
```
Least5_MI_fem <-
eg_2_MI[order(eg_2_MI$Fold_Change_Overweight_Females_to_MI_Males,decreasing=F
ALSE)[1:5],6:10]
row.names(Least5_MI_fem) <-
row.names(eg_2_MI[order(eg_2_MI$Fold_Change_Overweight_Females_to_MI_Males,
decreasing=FALSE),])[1:5]
row.names(Least5_MI_fem)

## [1] "MIR548AD"  "RNU5D-1"   "SNORD63"   "RNA5SP77"  "RNA5SP183"

Least5_MI_fem_t <- as.data.frame(t(Least5_MI_fem))
Least5_MI_fem_t[,1:2]

##                                                MIR548AD   RNU5D-1
## Fold_Change_Post_EGCG_to_Overweight_Females     1.0882736 1.0077962
## Fold_Change_MI_Males_to_Healthy_Males           0.9928044 0.9962153
## Fold_Change_Overweight_Females_to_MI_Males      0.2047564 0.3060531
## Fold_Change_Overweight_Females_to_Healthy_Males 0.2032830 0.3048948
## Fold_Change_Post_EGCG_Females_to_MI_Males       0.2228309 0.3084391
```

```
ggplot(Least5_MI_fem_t, aes(x = Least5_MI_fem_t$MIR548AD, y =
row.names(Least5_MI_fem_t)))+
      geom_point() +
  ylab(label="Fold Change Subsets") +
  xlab("The Least Fold Change Gene, MIR548AD")
```



Genecards.org summarizes this gene, MIR548AD or microRNA 548ad, as a micro RNA gene or miRNA that are short nucleotide non coding RNA involved in post-transcriptional regulation of gene expression in multicellular organisms by affecting the stability and translation of messenger RNA or mRNA. Micro RNAs can be inhibitors of translating mRNA or destabilizing it.

Given the summary above on this gene and the fact that it is under expressed in overweight females compared to males with MI and only slightly in males with MI compared to healthy males and also more expressed in overweight females using eGCG compared to overweight females not using EGCG, this could mean that males with MI and females using EGCG have more gene silencing or gene stabilization happening in their bodies. Earlier it was suggested that EGCG might relax overweight females by under expression of the gene THBS1 involved in the ER stress response.Given that this gene MIR548AD is a microRNA that is involved in stabilizing and translating mRNA, this gene could indicate or suggest that males have a faster metabolism and that females who use EGCG increase their metabolism as well.

What genes were the most expressed when comparing overweight females treated with EGCG to healthy males?

```
row.names(eg_2_MI[order(eg_2_MI$Fold_Change_Overweight_Females_to_Healthy_Mal
es,decreasing=TRUE),])[1:5]

## [1] "CA1"    "GYPA"   "AHSP"   "IFIT1B" "HBD"
```

What genes were the least expressed when comparing overweight females treated with EGCG to healthy males?

```
row.names(eg_2_MI[order(eg_2_MI$Fold_Change_Overweight_Females_to_Healthy_Mal
es,decreasing=FALSE),])[1:5]

## [1] "MIR548AD"  "SNORD63"   "RNU5D-1"   "RNA5SP77"  "RNA5SP183"
```

What genes were the most expressed when comparing males with MI to overweight females treated with EGCG?

```
row.names(eg_2_MI[order(eg_2_MI$Fold_Change_Post_EGCG_Females_to_MI_Males,dec
reasing=TRUE),])[1:5]

## [1] "CA1"    "GYPA"   "AHSP"   "IFIT1B" "H3F3C"
```

What genes were the least expressed when comparing males with MI to overweight females treated with EGCG?

```
row.names(eg_2_MI[order(eg_2_MI$Fold_Change_Post_EGCG_Females_to_MI_Males,dec
reasing=FALSE),])[1:5]

## [1] "MIR548AD"  "SNORD63"   "RNU5D-1"   "RNA5SP183" "RNU6-14P"
```