Full Metabolomics Analysis Protocol

Abrar Al-Shaer

October 21, 2019

This file shows an example metabolomics full analysis protocol (given validated metabolomics data). The example is using our lab's adipose tissue metabolomic data (from Nichole Reisdorph Lab). This was applied to the data for the manuscript by Pal et al. 2019 in the Shaikh Lab.

R Markdown

Loading in pakcages

```
rm(list=ls())
library("FactoMineR")
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.5.3
```

```
library("factoextra")
```

```
## Welcome! Related Books: `Practical Guide To Cluster Analysis in R` at https://goo.gl/13EFCZ
```

```
library("corrplot")
```

```
## corrplot 0.84 loaded
```

```
library(ggfortify)
```

```
## Warning: package 'ggfortify' was built under R version 3.5.3
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.5.3
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(gplots)
```

```
## Warning: package 'gplots' was built under R version 3.5.2
```

```
##
## Attaching package: 'gplots'
```

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

```
library(ComplexHeatmap)
```

```
## Loading required package: grid
```

```
## ========================================
## ComplexHeatmap version 1.20.0
## Bioconductor page: http://bioconductor.org/packages/ComplexHeatmap/
## Github page: https://github.com/jokergoo/ComplexHeatmap
## Documentation: http://bioconductor.org/packages/ComplexHeatmap/
##
## If you use it in published research, please cite:
## Gu, Z. Complex heatmaps reveal patterns and correlations in multidimensional
##   genomic data. Bioinformatics 2016.
## ========================================
```

```
library(reshape2)
library(ggpubr)
```

```
## Loading required package: magrittr
```

Creating PCA Plots

```
#for more detailed PCA code reference R script metabolites_PCA.R

#read in the file
#validated adipose data 10/15/19 - Con, HF, EPA
metData <- t(read.csv("Adipose Re-extraction Annotated_subset_rawData_Con_HF_EPA.csv", header =
TRUE, fill = TRUE, row.names = 2))[c(2:12,14:16),] #selecting from this file the samples (rows)
 I wish to include in the analysis

#example of the file read in
metData[1:14,1:3]
```

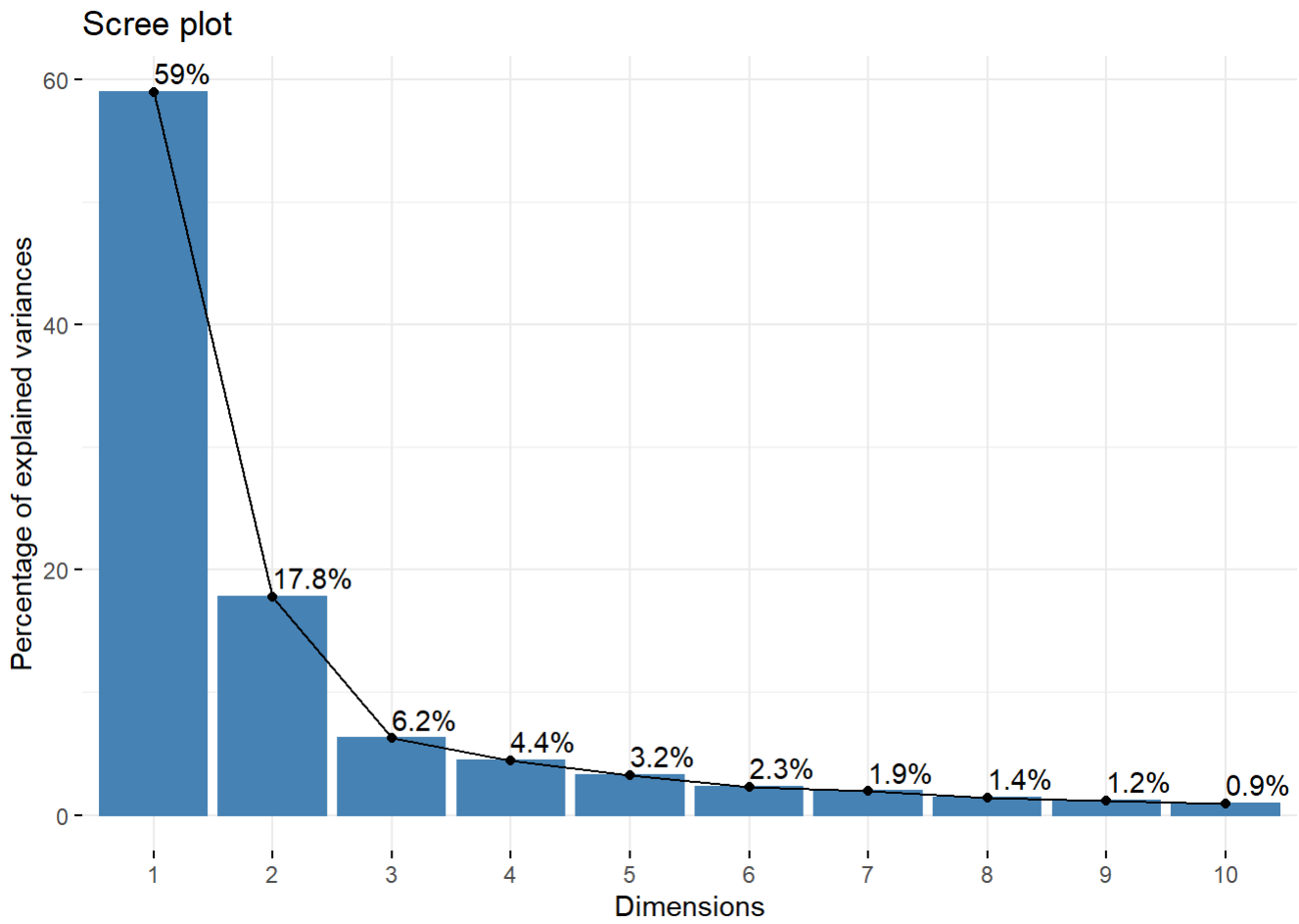
```
##          Diet     TG(58:12)   TG(62:16)
## Con.160  "Con "   "31674.855" "105.45054"
## Con.94   "Con "   "47160.934" "405.03076"
## Con.97   "Con "   "44973.47"  "418.18408"
## Con.98   "Con "   "34007.188" "130.7227"
## Con      "Con "   "40437.23"  "269.34964"
## EPA.122  "HF+EPA" "281643.4"  "20334.195"
## EPA.123  "HF+EPA" "335141.94" "20592.805"
## EPA.125  "HF+EPA" "193047.7"  "12608.929"
## EPA.129  "HF+EPA" "286710.22" "21367.342"
## EPA.130  "HF+EPA" "360116.8"  "18528.979"
## HF60.101 "HF"     "31650.309" "139.74446"
## HF60.106 "HF"     "37235.17"  "283.8175"
## HF60.108 "HF"     "34500.74"  "272.99994"
## HF60.110 "HF"     "1.00E-05"  "155.44041"
```

```
#convert the data frame to a numeric type
#2-77 are the columns that contain metabolite data
df2 <- data.frame(apply(metData[,2:77], 2, function(x) as.numeric(as.character(x))))
#assign the rownames to the dataframe
rownames(df2) <- rownames(metData)
#assign the column names to the dataframe
colnames(df2) <- colnames(metData)[2:77]
```

```
#Defining active variables for PCA
metData.active <- df2[,c(1:length(df2))]
#running PCA
res.PCA <- PCA(metData.active, ncp = 5, graph=FALSE) #ncp = # of principal components to store f
rom the PCA algorithm
```

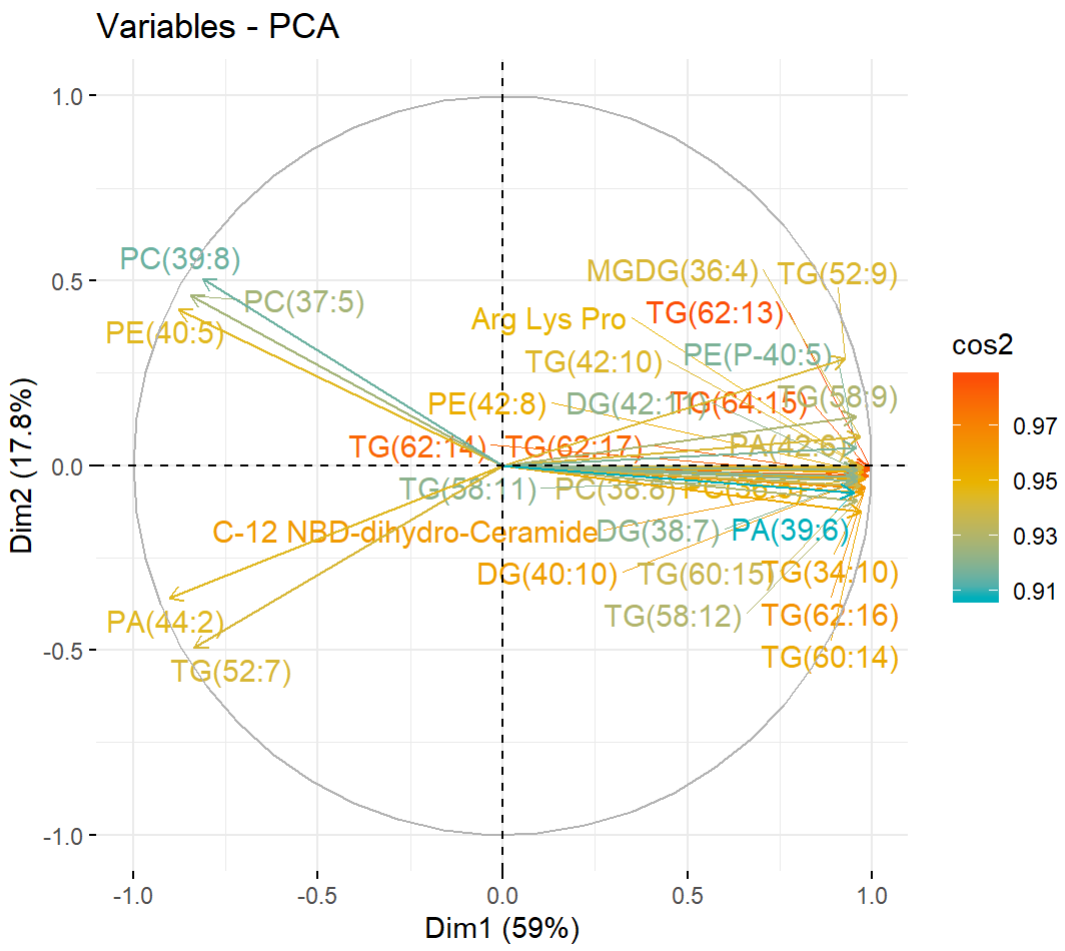
Visualize Scree Plot

```
#Scree plot
fviz_eig(res.PCA, addlabels = TRUE)
```



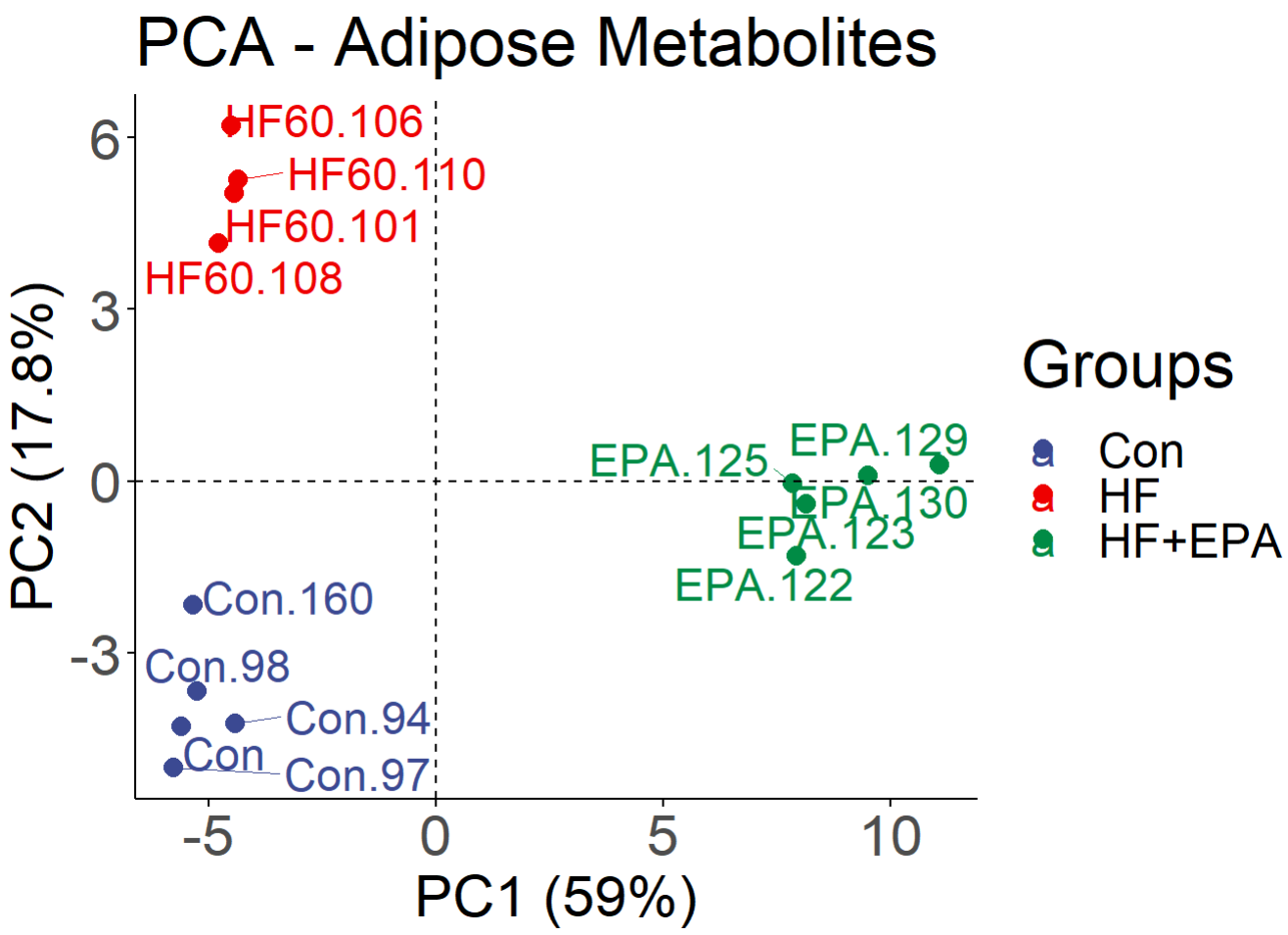
Visualize metabolites that contribute to the dimensions in the PCA

```
#provides a list of matrices containing all the results for the active variables
#(coordinates, correlation between variables and axes, squared cosine and contributions)
var <- get_pca_var(res.PCA)
#Visualize correlation between the variables (columns) and principal componenet (PC)
#color by cos2 values: quality on the factor map
#select.var = select variables with top 30 contribution scores
fviz_pca_var(res.PCA, col.var = "cos2", gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"), repe
l = TRUE, select.var = list(name = NULL, cos2 = NULL, contrib = 30)) #repel avoids overlapping l
abels
```



Visualize the Final PCA Graph

```
#assign the PCA graphical output to ind.p
ind.p <- fviz_pca_ind(res.PCA, geom.ind = c("point", "text"), axes = c(1,2), labelsize = 6, poin
tshape = 20, pointsize = 5, repel = TRUE, select.ind = list(contrib = 50), col.ind = metData[,1
], mean.point = FALSE, addEllipses = FALSE, legend.title = "Groups")
#for visualization purposes you can select individual samples (select.ind) that contribute to th
e PCA the most (top 50, contrib = 50). This does not change was the calculations of the PCA
#axes = c(1,2) defines the principal components you want to visualize (i.e.PC1 & PC2)
#col.ind = column you would like to color the groups by (I color them by column 1 of my metData
 dataframe)

ggpubr::ggpar(ind.p, title = "PCA - Adipose Metabolites", xlab = "PC1 (59%)", ylab = "PC2 (17.
8%)", legend.title = "Groups", legend.position = "top", ggtheme = theme_classic(), palette = "aa
as")+
theme(plot.title=element_text(size=26))+theme(plot.subtitle=element_text(size=20))+
  theme(axis.title.y = element_text(size = 22))+theme(axis.title.x = element_text(size = 22))+
  theme(legend.title=element_text(size=25))+theme(legend.text=element_text(size=18))+font("xy.te
xt", size = 22)
```



Testing for Statistical Significance & Calculating Fold Change

```
#reading in transposed validated adipose datwith only HF & EPA groups
metData <- read.csv("Adipose Re-extraction Annotated_subset_rawData_HF_EPA_transpose.csv", heade
r = TRUE, fill = TRUE, row.names = 1)[c(1:6,8:10),] #subset the samples (rows) you wish to inclu
de in the analysis (I'm excluding HF_105 because it's a clear outlier)

#assing a variables to the column names of the dataframe that contain METABOLITE names only (not
other identifiers)
columns <- colnames(metData[,2:77])

#example of the file read in
metData[1:9,1:3] #only HF & EPA samples exist (i.e. 2 groups only) because we are also calculati
ng fold change. So the dataframe must only contain 2 groups.
```

```
##            Diet   TG.62.14.  TG.42.10.
## EPA 122  HF+EPA 3231.585200 14058.0070
## EPA 123  HF+EPA 2644.176800 13985.6950
## EPA 125  HF+EPA 2442.924300  8620.6620
## EPA 129  HF+EPA 3620.994900 18574.1110
## EPA 130  HF+EPA 2922.708500 16452.7230
## HF60 101     HF    2.485317  1855.6494
## HF60 106     HF  364.999880  2424.4192
## HF60 108     HF  135.371410  1919.1084
## HF60 110     HF  358.767460   924.3308
```

```
#running the for loop that is iterating through all the columns & running the proper test & calc
ulating FC
#make sure toreplace metData[1:5] with the rows for group 1 and metData[6:9] with the rows for g
roup 2
datalist = list()
fold_change <- list()
i = 1
for(col in columns) #only loop through the column names specified above
{
  #print(col) #printing column name
  formula = as.formula(paste(col, "~ Diet")) #creating the comparison formula with the column

  if(shapiro.test(metData[,col])$p.value > 0.05){
    result = compare_means(formula,  data = metData, method = "t.test") #if data is normal
    if (result[[4]] < 0.05) {
      datalist[[i]] = print(compare_means(formula,  data = metData, method = "t.test"))

      if(mean(metData[1:5,col]) > mean(metData[6:9,col])){ #if the mean of EPA is > than HF, the
n it's a (+)FC
        fold_change[[i]] = mean(metData[1:5,col])/mean(metData[6:9,col]) #mean of EPA / mean of
 HF
      } else { fold_change[[i]] = (mean(metData[1:5,col])/mean(metData[6:9,col]))*-1 } #(-)FC ->
EPA < than HF
    }
  } else {
    result = compare_means(formula,  data = metData, method = "wilcox.test") #if the data is not
normal
    if (result[[4]] < 0.05) {
      datalist[[i]] = print(compare_means(formula,  data = metData, method = "wilcox.test"))

      if(mean(metData[1:5,col]) > mean(metData[6:9,col])){ #if the mean of EPA is > than HF, the
n it's a (+)FC
        fold_change[[i]] = mean(metData[1:5,col])/mean(metData[6:9,col]) #mean of EPA / mean of
 HF
      } else { fold_change[[i]] = (mean(metData[1:5,col])/mean(metData[6:9,col]))*-1 } #(-)FC ->
EPA < than HF

    }
  }
  i = i+1
}
```

```
## # A tibble: 1 x 8
##   .y.       group1 group2      p p.adj p.format p.signif method
##   <chr>     <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 TG.62.14. HF     HF+EPA 0.0200  0.02 0.02     *        Wilcoxon
## # A tibble: 1 x 8
##   .y.       group1 group2       p  p.adj p.format p.signif method
##   <chr>     <chr>  <chr>    <dbl>  <dbl> <chr>    <chr>    <chr>
## 1 TG.42.10. HF     HF+EPA 0.00134 0.0013 0.0013   **       T-test
## # A tibble: 1 x 8
##   .y.       group1 group2      p p.adj p.format p.signif method
##   <chr>     <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 TG.62.17. HF     HF+EPA 0.0200  0.02 0.02     *        Wilcoxon
## # A tibble: 1 x 8
##   .y.       group1 group2      p p.adj p.format p.signif method
##   <chr>     <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 TG.64.15. HF     HF+EPA 0.0200  0.02 0.02     *        Wilcoxon
## # A tibble: 1 x 8
##   .y.       group1 group2      p p.adj p.format p.signif method
##   <chr>     <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 TG.60.14. HF     HF+EPA 0.0200  0.02 0.02     *        Wilcoxon
## # A tibble: 1 x 8
##   .y.       group1 group2      p p.adj p.format p.signif method
##   <chr>     <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 TG.62.16. HF     HF+EPA 0.0200  0.02 0.02     *        Wilcoxon
## # A tibble: 1 x 8
##   .y.       group1 group2        p   p.adj p.format p.signif method
##   <chr>     <chr>  <chr>     <dbl>   <dbl> <chr>    <chr>    <chr>
## 1 TG.58.12. HF     HF+EPA 0.000408 0.00041 0.00041  ***      T-test
## # A tibble: 1 x 8
##   .y.                  group1 group2      p p.adj p.format p.signif method
##   <chr>                <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 X3.Isopropylcatechol HF     HF+EPA 0.0200  0.02 0.02     *        Wilcox~
## # A tibble: 1 x 8
##   .y.         group1 group2        p   p.adj p.format p.signif method
##   <chr>       <chr>  <chr>     <dbl>   <dbl> <chr>    <chr>    <chr>
## 1 Arg.Lys.Pro HF     HF+EPA 0.000472 0.00047 0.00047  ***      T-test
## # A tibble: 1 x 8
##   .y.             group1 group2       p   p.adj p.format p.signif method
##   <chr>           <chr>  <chr>    <dbl>   <dbl> <chr>    <chr>    <chr>
## 1 Bouillonamide.A HF     HF+EPA 0.00237 0.00240 0.0024   **       T-test
## # A tibble: 1 x 8
##   .y.                  group1 group2      p p.adj p.format p.signif method
##   <chr>                <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 C.12.NBD.dihydro.Ce~ HF     HF+EPA 0.0200  0.02 0.02     *        Wilcox~
## # A tibble: 1 x 8
##   .y.            group1 group2       p  p.adj p.format p.signif method
##   <chr>          <chr>  <chr>    <dbl>  <dbl> <chr>    <chr>    <chr>
## 1 Concanamycin.A HF     HF+EPA 0.00730 0.0073 0.0073   **       T-test
## # A tibble: 1 x 8
##   .y.         group1 group2      p p.adj p.format p.signif method
##   <chr>       <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 deslanoside HF     HF+EPA 0.0211 0.021 0.021    *        T-test
## # A tibble: 1 x 8
```

```
##   .y.      group1 group2      p p.adj p.format p.signif method
##   <chr>    <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 DG.38.7. HF     HF+EPA 0.0200  0.02 0.02     *        Wilcoxon
## # A tibble: 1 x 8
##   .y.       group1 group2      p p.adj p.format p.signif method
##   <chr>     <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 DG.40.10. HF     HF+EPA 0.0200  0.02 0.02     *        Wilcoxon
## # A tibble: 1 x 8
##   .y.       group1 group2       p p.adj p.format p.signif method
##   <chr>     <chr>  <chr>    <dbl> <dbl> <chr>    <chr>    <chr>
## 1 DG.42.10. HF     HF+EPA 0.00102 0.001 0.001    **       T-test
## # A tibble: 1 x 8
##   .y.       group1 group2        p   p.adj p.format p.signif method
##   <chr>     <chr>  <chr>     <dbl>   <dbl> <chr>    <chr>    <chr>
## 1 DG.42.11. HF     HF+EPA 0.000621 0.00062 0.00062  ***      T-test
## # A tibble: 1 x 8
##   .y.           group1 group2      p p.adj p.format p.signif method
##   <chr>         <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 GalCer.d42.1. HF     HF+EPA 0.0236 0.024 0.024    *        T-test
## # A tibble: 1 x 8
##   .y.         group1 group2      p p.adj p.format p.signif method
##   <chr>       <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 Gln.His.Thr HF     HF+EPA 0.0200  0.02 0.02     *        Wilcoxon
## # A tibble: 1 x 8
##   .y.         group1 group2       p p.adj p.format p.signif method
##   <chr>       <chr>  <chr>    <dbl> <dbl> <chr>    <chr>    <chr>
## 1 Gln.Pro.Pro HF     HF+EPA 0.00398 0.004 0.004    **       T-test
## # A tibble: 1 x 8
##   .y.           group1 group2      p p.adj p.format p.signif method
##   <chr>         <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 LacCer.d32.0. HF     HF+EPA 0.0172 0.017 0.017    *        T-test
## # A tibble: 1 x 8
##   .y.          group1 group2       p  p.adj p.format p.signif method
##   <chr>        <chr>  <chr>    <dbl>  <dbl> <chr>    <chr>    <chr>
## 1 LysoPE.22.5. HF     HF+EPA 0.00177 0.0018 0.0018   **       T-test
## # A tibble: 1 x 8
##   .y.        group1 group2        p   p.adj p.format p.signif method
##   <chr>      <chr>  <chr>     <dbl>   <dbl> <chr>    <chr>    <chr>
## 1 MGDG.36.4. HF     HF+EPA 0.000620 0.00062 0.00062  ***      T-test
## # A tibble: 1 x 8
##   .y.         group1 group2      p p.adj p.format p.signif method
##   <chr>       <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 Momordin.Ie HF     HF+EPA 0.0138 0.014 0.014    *        T-test
## # A tibble: 1 x 8
##   .y.      group1 group2       p  p.adj p.format p.signif method
##   <chr>    <chr>  <chr>    <dbl>  <dbl> <chr>    <chr>    <chr>
## 1 PA.29.0. HF     HF+EPA 0.00249 0.0025 0.0025   **       T-test
## # A tibble: 1 x 8
##   .y.      group1 group2      p p.adj p.format p.signif method
##   <chr>    <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 PA.35.3. HF     HF+EPA 0.0200  0.02 0.02     *        Wilcoxon
## # A tibble: 1 x 8
##   .y.      group1 group2      p p.adj p.format p.signif method
##   <chr>    <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
```

```
## 1 PA.39.6. HF     HF+EPA 0.0200  0.02 0.02     *        Wilcoxon
## # A tibble: 1 x 8
##   .y.      group1 group2      p p.adj p.format p.signif method
##   <chr>    <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 PA.42.6. HF     HF+EPA 0.0200  0.02 0.02     *        Wilcoxon
## # A tibble: 1 x 8
##   .y.      group1 group2      p p.adj p.format p.signif method
##   <chr>    <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 PA.44.1. HF     HF+EPA 0.0201  0.02 0.02     *        T-test
## # A tibble: 1 x 8
##   .y.      group1 group2       p  p.adj p.format p.signif method
##   <chr>    <chr>  <chr>    <dbl>  <dbl> <chr>    <chr>    <chr>
## 1 PA.44.2. HF     HF+EPA 0.00195 0.0019 0.0019   **       T-test
## # A tibble: 1 x 8
##   .y.      group1 group2       p  p.adj p.format p.signif method
##   <chr>    <chr>  <chr>    <dbl>  <dbl> <chr>    <chr>    <chr>
## 1 PC.35.5. HF     HF+EPA 0.00586 0.0059 0.0059   **       T-test
## # A tibble: 1 x 8
##   .y.      group1 group2      p p.adj p.format p.signif method
##   <chr>    <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 PC.36.5. HF     HF+EPA 0.0200  0.02 0.02     *        Wilcoxon
## # A tibble: 1 x 8
##   .y.      group1 group2        p   p.adj p.format p.signif method
##   <chr>    <chr>  <chr>     <dbl>   <dbl> <chr>    <chr>    <chr>
## 1 PC.37.4. HF     HF+EPA 0.000365 0.00037 0.00037  ***      T-test
## # A tibble: 1 x 8
##   .y.      group1 group2      p p.adj p.format p.signif method
##   <chr>    <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 PC.37.5. HF     HF+EPA 0.0200  0.02 0.02     *        Wilcoxon
## # A tibble: 1 x 8
##   .y.      group1 group2         p    p.adj p.format p.signif method
##   <chr>    <chr>  <chr>      <dbl>    <dbl> <chr>    <chr>    <chr>
## 1 PC.38.8. HF     HF+EPA 0.0000745 0.000074 7.4e-05  ****     T-test
## # A tibble: 1 x 8
##   .y.      group1 group2      p p.adj p.format p.signif method
##   <chr>    <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 PC.39.8. HF     HF+EPA 0.0200  0.02 0.02     *        Wilcoxon
## # A tibble: 1 x 8
##   .y.      group1 group2      p p.adj p.format p.signif method
##   <chr>    <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 PC.40.5. HF     HF+EPA 0.0116 0.012 0.012    *        T-test
## # A tibble: 1 x 8
##   .y.      group1 group2       p  p.adj p.format p.signif method
##   <chr>    <chr>  <chr>    <dbl>  <dbl> <chr>    <chr>    <chr>
## 1 PE.36.5. HF     HF+EPA 0.00331 0.0033 0.0033   **       T-test
## # A tibble: 1 x 8
##   .y.      group1 group2      p p.adj p.format p.signif method
##   <chr>    <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 PE.40.4. HF     HF+EPA 0.0200  0.02 0.02     *        Wilcoxon
## # A tibble: 1 x 8
##   .y.        group1 group2       p  p.adj p.format p.signif method
##   <chr>      <chr>  <chr>    <dbl>  <dbl> <chr>    <chr>    <chr>
## 1 PE.P.40.5. HF     HF+EPA 0.00183 0.0018 0.0018   **       T-test
## # A tibble: 1 x 8
```

```
##   .y.      group1 group2      p p.adj p.format p.signif method
##   <chr>    <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 PE.40.5. HF     HF+EPA 0.0179 0.018 0.018    *        Wilcoxon
## # A tibble: 1 x 8
##   .y.       group1 group2      p p.adj p.format p.signif method
##   <chr>     <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 PE.42.11. HF     HF+EPA 0.0394 0.039 0.039    *        T-test
## # A tibble: 1 x 8
##   .y.      group1 group2      p p.adj p.format p.signif method
##   <chr>    <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 PE.42.8. HF     HF+EPA 0.0200  0.02 0.02     *        Wilcoxon
## # A tibble: 1 x 8
##   .y.        group1 group2      p p.adj p.format p.signif method
##   <chr>      <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 PE.P.40.6. HF     HF+EPA 0.0145 0.014 0.014    *        T-test
## # A tibble: 1 x 8
##   .y.       group1 group2      p p.adj p.format p.signif method
##   <chr>     <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 PI.42.11. HF     HF+EPA 0.0101  0.01 0.01     *        T-test
## # A tibble: 1 x 8
##   .y.      group1 group2       p   p.adj p.format p.signif method
##   <chr>    <chr>  <chr>    <dbl>   <dbl> <chr>    <chr>    <chr>
## 1 PS.39.8. HF     HF+EPA 0.00237 0.00240 0.0024   **       T-test
## # A tibble: 1 x 8
##   .y.      group1 group2      p p.adj p.format p.signif method
##   <chr>    <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 PS.42.5. HF     HF+EPA 0.0352 0.035 0.035    *        T-test
## # A tibble: 1 x 8
##   .y.       group1 group2      p p.adj p.format p.signif method
##   <chr>     <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 TG.34.10. HF     HF+EPA 0.0200  0.02 0.02     *        Wilcoxon
## # A tibble: 1 x 8
##   .y.      group1 group2      p p.adj p.format p.signif method
##   <chr>    <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 TG.52.7. HF     HF+EPA 0.0309 0.031 0.031    *        T-test
## # A tibble: 1 x 8
##   .y.      group1 group2      p p.adj p.format p.signif method
##   <chr>    <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 TG.52.9. HF     HF+EPA 0.0200  0.02 0.02     *        Wilcoxon
## # A tibble: 1 x 8
##   .y.       group1 group2      p p.adj p.format p.signif method
##   <chr>     <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 TG.56.11. HF     HF+EPA 0.0200  0.02 0.02     *        Wilcoxon
## # A tibble: 1 x 8
##   .y.      group1 group2      p p.adj p.format p.signif method
##   <chr>    <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 TG.56.7. HF     HF+EPA 0.0491 0.049 0.049    *        T-test
## # A tibble: 1 x 8
##   .y.       group1 group2        p    p.adj p.format p.signif method
##   <chr>     <chr>  <chr>     <dbl>    <dbl> <chr>    <chr>    <chr>
## 1 TG.58.11. HF     HF+EPA 0.000277 0.000280 0.00028  ***      T-test
## # A tibble: 1 x 8
##   .y.       group1 group2      p p.adj p.format p.signif method
##   <chr>     <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
```

```
## 1 TG.58.13. HF     HF+EPA 0.0151 0.015 0.015    *        Wilcoxon
## # A tibble: 1 x 8
##   .y.      group1 group2        p   p.adj p.format p.signif method
##   <chr>    <chr>  <chr>     <dbl>   <dbl> <chr>    <chr>    <chr>
## 1 TG.58.9. HF     HF+EPA 0.000231 0.00023 0.00023  ***      T-test
## # A tibble: 1 x 8
##   .y.       group1 group2      p p.adj p.format p.signif method
##   <chr>     <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 TG.60.15. HF     HF+EPA 0.0151 0.015 0.015    *        Wilcoxon
## # A tibble: 1 x 8
##   .y.       group1 group2      p p.adj p.format p.signif method
##   <chr>     <chr>  <chr>   <dbl> <dbl> <chr>    <chr>    <chr>
## 1 TG.62.13. HF     HF+EPA 0.0200  0.02 0.02     *        Wilcoxon
```

```
big_data1 = do.call(rbind, datalist) #combines all previous dataframes from for loop
big_data2 = data.frame(do.call(rbind, fold_change)) #contains fold change calculations
colnames(big_data2) = "Fold Change EPA:HF" #assign column name
big_data_final = dplyr::bind_cols(big_data1, big_data2) #combine all the dataframes together (st
ats & fold changes)
#write.csv(big_data_final, "Validated_Adipose_metabolites_FoldChange+signif_HF_EPA_unadjust.cs
v")

#adjust p-values
big_data_final$BH.adjust <- p.adjust(big_data_final$p, method = "BH")
#write.csv(big_data_final, "Validated_Adipose_metabolites_FoldChange+signif_HF_EPA_BH_adjust.cs
v")

#visualize big_data_final dataframe to show final output
big_data_final[1:5,1:10]
```

```
## # A tibble: 5 x 10
##   .y.   group1 group2       p  p.adj p.format p.signif method
##   <chr> <chr>  <chr>    <dbl>  <dbl> <chr>    <chr>    <chr>
## 1 TG.6~ HF     HF+EPA 0.0200  0.02   0.02     *        Wilco~
## 2 TG.4~ HF     HF+EPA 0.00134 0.0013 0.0013   **       T-test
## 3 TG.6~ HF     HF+EPA 0.0200  0.02   0.02     *        Wilco~
## 4 TG.6~ HF     HF+EPA 0.0200  0.02   0.02     *        Wilco~
## 5 TG.6~ HF     HF+EPA 0.0200  0.02   0.02     *        Wilco~
## # ... with 2 more variables: `Fold Change EPA:HF` <dbl>, BH.adjust <dbl>
```

Generating Heatmaps

```
#take the output file from the above code & add in a MetaboliteGroup column where you manually a
nnotate the metabolite class of each metabolite (also add a Type column with an abbreviation of
 the class). Next take the fold change column & add in a Log2 FC column where you Log2 all the f
old changes (for negative fold changes multiply the FC by -1 then calculate the Log2).
metabolites <- read.csv("Validated_Adipose_metabolites_FoldChange+signif_HF_EPA_BH_adjust.csv",
 header = TRUE, fill = TRUE)
#read in the transposed raw metabolomics data with any undesired samples removed (I removed HF_1
05 because it was an outlier)
metData <- read.csv("Adipose Re-extraction Annotated_subset_rawData_HF_EPA_transpose_rmHF105.cs
v", header = TRUE, fill = TRUE, row.names = 1)
```

Selecting only significant metabolites

```
#extract metabolite names from the file that contains significant metabolites
compounds <- unique(as.character(metabolites$Metabolite))

#selects all occurances/matches for that list of significant compound
metData_select <- metData[,colnames(metData) %in% compounds]
#add the sampleID label from the rownames of the previous dataframe (metData)
row.names(metData_select) <- rownames(metData)

#transpose the metData_select dataframe
metData_select_t <- t(metData_select)

#write to a CSV the list of significant metabolites & their abundance values (raw data)
#write.csv(metData_select_t,"Validated_Adipose_metabolites_FoldChange+signif_HF_EPA_BH_adjust_me
taboliteValues.csv")

#view output of dataframe
metData_select_t[1:6,1:9]
```

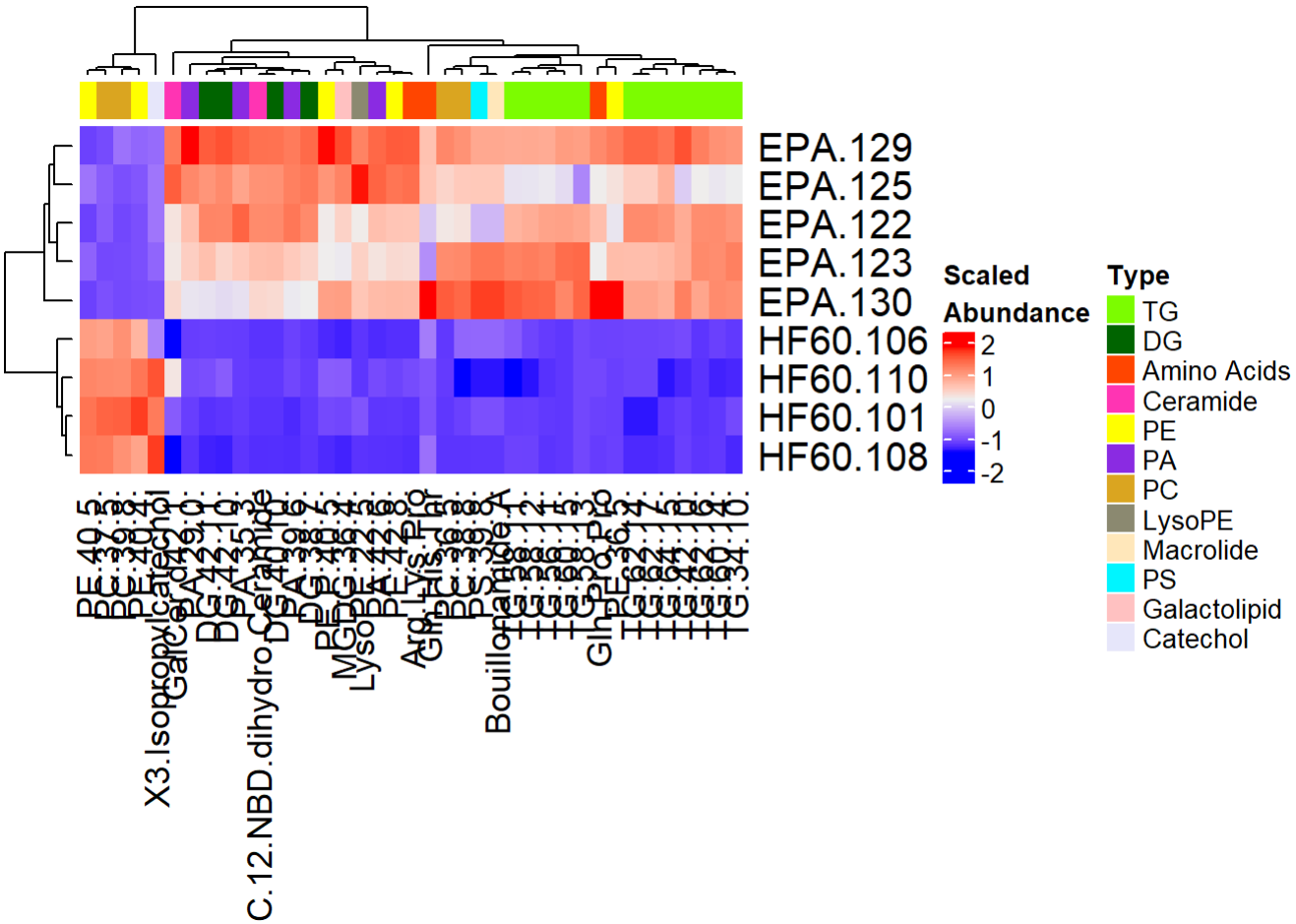
```
##             EPA 122   EPA 123   EPA 125   EPA 129   EPA 130    HF60 101
## TG.62.14.  3231.585  2644.177  2442.924  3620.995  2922.709    2.485317
## TG.42.10. 14058.007 13985.695  8620.662 18574.111 16452.723 1855.649400
## TG.62.17.  3231.585  2644.177  2442.924  3620.995  2922.709    2.485317
## TG.64.15.  3767.060  3298.833  3394.267  4148.719  3429.416  612.233950
## TG.60.14. 46148.582 46045.960 28216.729 45286.844 46270.086 3624.804400
## TG.62.16. 20334.195 20592.805 12608.929 21367.342 18528.979  139.744460
##            HF60 106  HF60 108  HF60 110
## TG.62.14.  364.9999  135.3714  358.7675
## TG.42.10. 2424.4192 1919.1084  924.3308
## TG.62.17.  364.9999  135.3714  358.7675
## TG.64.15.  771.9772  547.4090  358.1771
## TG.60.14. 4717.0195 3685.7737 1133.5958
## TG.62.16.  283.8175  272.9999  155.4404
```

```
#Take the above dataframe that you just wrote into a CSV and sort the dataframe & add the metabo
lite Group names (same metabolite group names as the metabolites dataframe). Also, filter the da
taframe by fold changes above 1.5 or below -1.5
metabolites2 <- na.omit(read.csv("Validated_Adipose_metabolites_FoldChange+signif_HF_EPA_BH_adju
st_metaboliteValues_1.5FC.csv", header = TRUE, fill = TRUE))
nams = metabolites2[,1] #names of the metabolites
#assign rownames to the dataframe based on metabolite names
rownames(metabolites2) = make.names(nams, unique=TRUE) #unique = TRUE to avoid duplicate names e
rror

#view output of dataframe
metabolites2[1:5,1:15]
```

```
##           Metabolite MetaboliteGroup n3PUFA_containing Type    Log2FC
## TG.58.13.  TG(58:13)   Triglycerides               YES   TG 32.368486
## TG.60.15.  TG(60:15)   Triglycerides               YES   TG 31.674860
## TG.62.16.  TG(62:16)   Triglycerides               YES   TG  6.454991
## TG.56.11.  TG(56:11)   Triglycerides               YES   TG  4.903492
## TG.34.10.  TG.34.10.   Triglycerides                NO   TG  4.058451
##           GreaterDiet   EPA.122   EPA.123   EPA.125   EPA.129   EPA.130
## TG.58.13.         EPA  57415.08  72613.10  14115.96  59542.65  73552.74
## TG.60.15.         EPA  34756.56  41995.62  20507.48  35580.82  38576.19
## TG.62.16.         EPA  20334.19  20592.81  12608.93  21367.34  18528.98
## TG.56.11.         EPA 259749.10 293047.78 169899.66 250470.02 318699.03
## TG.34.10.         EPA  24975.09  26867.81  16144.44  24856.15  25559.86
##              HF60.101    HF60.106   HF60.108   HF60.110
## TG.58.13.     0.00001     0.00001    0.00001    0.00001
## TG.60.15.     0.00001     0.00001    0.00001    0.00001
## TG.62.16.   139.74446   283.81750  272.99994  155.44041
## TG.56.11. 10611.68700 12845.62400 6676.69700 4396.98930
## TG.34.10.  2910.93070  1714.12500  548.07117  511.97780
```

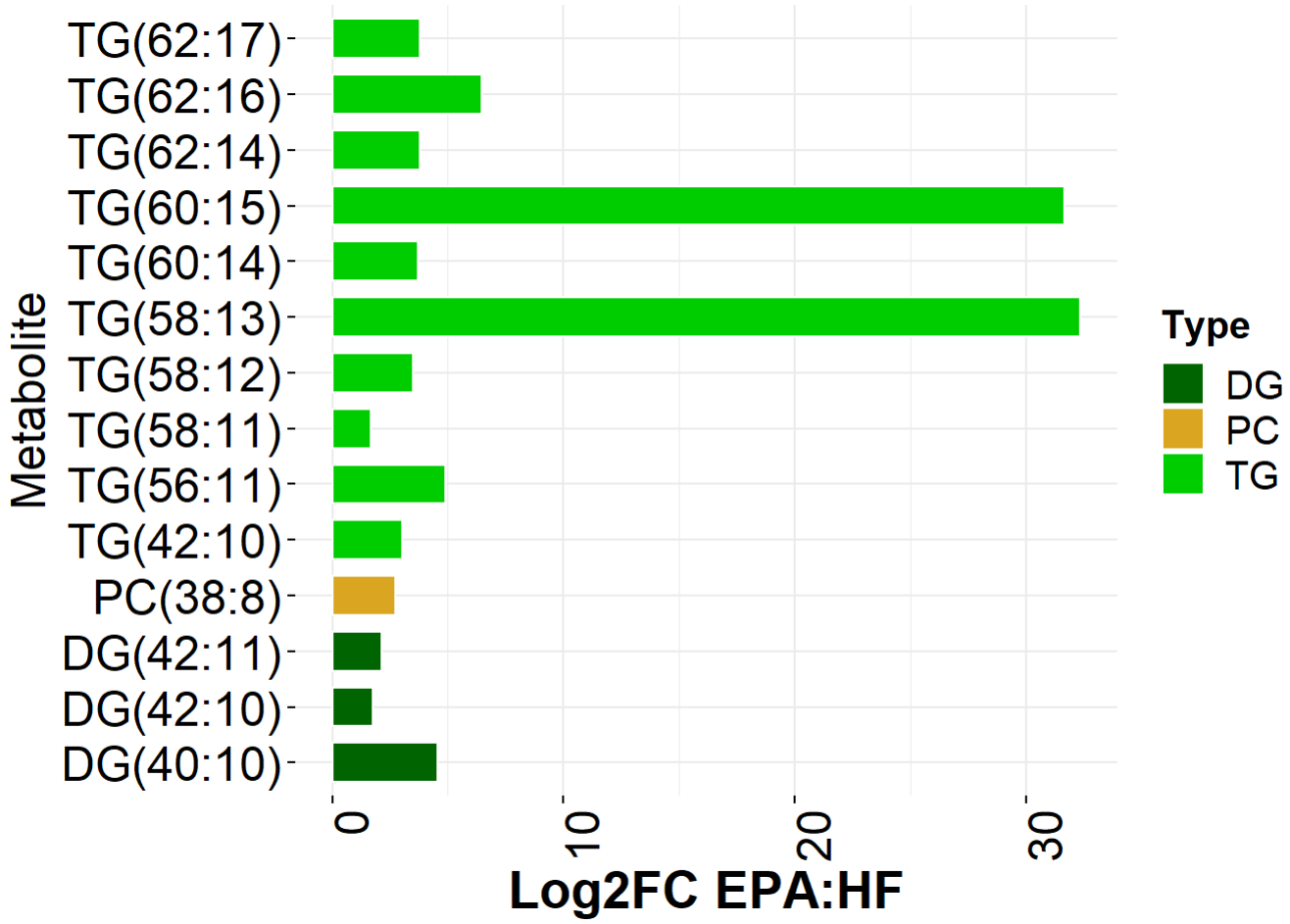
Creating the heatmap w/annotations

```
# Annotation data frame
annot_df <- data.frame(Type = metabolites2$Type) #Log2FC = metabolites2$Log2FC (you can also add
Log2FC annotation)
#Define colors for each levels of qualitative variables
#Define gradient color for continuous variables (FC)
col = list(Type = c("TG" = "lawngreen", "DG" = "darkgreen", "Amino Acids" = "orangered", "Cerami
de" = "maroon1",
                    "PE" = "yellow", "PA" = "blueviolet", "PC" = "goldenrod", "LysoPE" = "lemonc
hiffon4",
                    "Macrolide" = "wheat1", "PS" = "turquoise1", "Galactolipid" = "rosybrown1",
"Catechol" = "lavender")) #Log2FC = circlize::colorRamp2(c(-5,34), c("lightblue", "purple"))
# Create the heatmap annotation
ha <- HeatmapAnnotation(annot_df, col = col)

# Combine the heatmap and the annotation
Heatmap(scale(t(metabolites2[,7:15])), name = "Scaled \nAbundance", top_annotation = ha, row_nam
es_gp = gpar(fontsize = 15), column_names_gp = gpar(fontsize = 13), cluster_columns = TRUE, clus
ter_rows = TRUE)
```



Fold Change Plot

```
#subsetted (1.5 FC cutoff) validated adipose data
metData <- read.csv("Validated_Adipose_metabolites_FoldChange+signif_HF_EPA_BH_adjust_metabolite
Values_1.5FC.csv", header = TRUE, fill = TRUE)
#filter metData by only n3 PUFA containing metabolites
metData_sub <- metData[metData$n3PUFA_containing == "YES",]

#FOLD CHANGE BAR GRAPH
#color pallet
colors = c("TG" = "green3", "DG" = "darkgreen", "PC" = "goldenrod")

fc_plot <- ggbarplot(metData_sub, x = "Metabolite", y = "Log2FC",
          fill = "Type",           # change fill color by Type
          color = "white",            # Set bar border colors to white
          palette = colors,            # jco journal color palett is also an option. see ?ggpar
          sort.val = "none",          # If you want it sorted by FC then put (desc) to sort in d
escending order
          sort.by.groups = FALSE,     # Don't sort inside each group
          x.text.angle = 90,          # Rotate vertically x axis texts
          ylab = "Log2FC EPA:HF",
          legend.title = "Type",
          rotate = TRUE,              #if Rotate = FALSE it will be a horizontal graph
          ggtheme = theme_minimal()
)

fc_plot + theme(axis.text.y = element_text(color = "black", size = 18),
                axis.text.x = element_text(color = "black", size = 18),
      axis.title.x = element_text(color = "black", face = "bold", size = 20),
      axis.title.y = element_text(color = "black", size = 18)) +
  theme(legend.title = element_text(size=15, face="bold")) + theme(legend.text = element_text(si
ze=15))
```


TG(62:17)
TG(62:16)
TG(62:14)
TG(60:15)
TG(60:14)
TG(58:13)
TG(58:12)
TG(58:11)
TG(56:11)
TG(42:10)
PC(38:8)
DG(42:11)
DG(42:10)
DG(40:10)
Metabolite
0
10
20
30
Log2FC EPA:HF
Type
DG
PC
TG