Packages

```
#Main plotting function load in packages to get this
library (ggplot2)
library (tidyverse)
library (dplyr)
library(readxl)
library(readr)
library(magrittr)
library ("RColorBrewer")
library (pheatmap)
library (grid)
library(future)
library(furrr)
library(ppcor)
```

Import Anticancer Drugs Signatures

```
# Importing the Chemical Perturbation File (This may take a few minutes to run) Drugs from LINCS
Chem_Data <- read.delim("data/Inputs/I1000_cp.txt", header = FALSE)
# Extract the  Data, place keyword or metabolite of interest name in the REPLACENAMEHERE to pull out th
Female_CD <- Chem_Data[Chem_Data$V1 %like%"paclitaxel|ado-trastuzumabemtansine|everolimus|alpelisib|anas
write.csv(Female_CD, "data/Outputs/Female_CD_All.csv", row.names = FALSE, col.names = FALSE)
```

Microbe-RNA Data Preparation

```
###MEPH1
###Microbiota to RNA
Meph1_Mbiome_RNA <- read_excel("data/Inputs/Meph1_Mbiome_RNA.xlsx")
Meph1_Mbiome_R <- Meph1_Mbiome_RNA[,-c(1,3)]
Meph1_Mbiome_R_cols <- names(Meph1_Mbiome_R)[sapply(Meph1_Mbiome_R, is.numeric)]

Meph1_RNA_1_Mbiome <- read_excel("data/Inputs/Meph1_RNA_1_Mbiome.xlsx")
New <- read_excel("data/Inputs/Meph1_RNA_2_Mbiome.xlsx")
Merged_RNA_Mbiome <- merge(Meph1_RNA_1_Mbiome,New, by = "xptid_visit_code", all = TRUE)
colnames(Merged_RNA_Mbiome) <- gsub('.x','',names(Merged_RNA_Mbiome))
Meph1_RNA_M <- Merged_RNA_Mbiome[order(Merged_RNA_Mbiome$Lactobacillus_Dominance),]
Meph1_RNA <- Meph1_RNA_M [,-c(1,3)]
Meph1_RNA_cols <-  names(Meph1_RNA)[sapply(Meph1_RNA, is.numeric)]
```

Pairwise correlation Function

```
#Pairwise correlation function
pairwise_correlation <- function(.data, .var1, .var2) {
  corr <- cor.test(
    x = .data[[.var1]],
    y = .data[[.var2]],
    method = "spearman",
    exact = FALSE
  )
```

```
  data.frame(
    var1 = .var1,
    var2 = .var2,
    rho = corr$estimate,
    p = corr$p.value
  )
}
```

Microbe-RNA Spearman correlation

```
Microbe_RNA <- merge(Meph1_Mbiome_R, Meph1_RNA, by = "Lactobacillus_Dominance", all = TRUE)

vars <- tidyr::expand_grid(Meph1_Mbiome_R_cols, Meph1_RNA_cols)

future::plan(multisession, workers = 4)

microbe_rna_correlations <- furrr::future_map2(vars$Meph1_Mbiome_R_cols,
                                               vars$Meph1_RNA_cols,
                                               \(x, y) pairwise_correlation(Microbe_RNA, x, y)) |>
  purrr::list_rbind()

rownames(microbe_rna_correlations) <- NULL
colnames(microbe_rna_correlations) <- c("MICROBIOTA","RNA","Rho Value","P Value")
microbe_rna_correlations <- data.frame(na.omit(microbe_rna_correlations))
microbe_rna_correlations <- microbe_rna_correlations[!grepl('undistinguishable',microbe_rna_correlations
microbe_rna_correlations <- filter(microbe_rna_correlations,P.Value<=0.05)
microbe_rna_correlations$MICROBIOTA.num <- 0
microbe_rna_correlations <- transform(microbe_rna_correlations,MICROBIOTA.num = as.numeric(as.factor(MI
microbe_rna_correlations<- microbe_rna_correlations %>%
    group_by(MICROBIOTA.num) %>%
    mutate(pval.adj = p.adjust (P.Value, method='BH'))
write.csv(microbe_rna_correlations, file = "data/Outputs/Spearman test for Microbe_RNA_Meph1.csv")
```

Metabolite-RNA Data Preparation

```
#Metabolite to RNA

Meph1_Metabolomics_RNA_1 <- read_excel("data/Inputs/Meph1_Metabolomics_RNA_1.xlsx")
Meph1_Met_R <- Meph1_Metabolomics_RNA_1[,-c(1,2,3)]
Meph1_Met_R_Condition <- Meph1_Metabolomics_RNA_1[,-c(1)]
Meph1_Met_R <- na.omit(Meph1_Met_R)
#set dataset to strictly numeric
colnames(Meph1_Met_R) <- NULL
rownames(Meph1_Met_R) <- NULL

Meph1_RNA_Met_Mbiome2 <- read_excel("data/Inputs/Meph1_RNA_Met_Mbiome2.xlsx")
Meph1_Mbiome_R_M <- Meph1_RNA_Met_Mbiome2[,-c(1,2)]
Meph1_Mbiome_R_M_Condition <- Meph1_RNA_Met_Mbiome2[,-c(1)]
Meph1_Mbiome_R_M <- na.omit(Meph1_Mbiome_R_M)
#set dataset to strictly numeric
```

```r
colnames(Meph1_Mbiome_R_M) <- NULL
rownames(Meph1_Mbiome_R_M) <- NULL

Meph1_RNA_1_Met <- read_excel("data/Inputs/Meph1_RNA_1_Met.xlsx")
New_2 <- read_excel("data/Inputs/Meph1_RNA_2_Met.xlsx")
Merged_RNA_Met <- merge(Meph1_RNA_1_Met,New_2, by = "xptid_visit_code", all = TRUE)
colnames(Merged_RNA_Met) <- gsub('.x','',names(Merged_RNA_Met))
Meph1_RNA_Met <- Merged_RNA_Met[order(Merged_RNA_Met$Lactobacillus_Dominance),]
Meph1_RNA_Met <- Meph1_RNA_Met [,4:23485]
Meph1_RNA_Met_num <- Meph1_RNA_Met
colnames(Meph1_RNA_Met_num) <- NULL
rownames(Meph1_RNA_Met_num) <- NULL
```

Metabolite-RNA Spearman Preparation

```r
#Runs for long hours
#vectors for rho values, p values, and names of both rna and metabolite
spear_test_rho <- vector()
spear_test_p <- vector()
names_RNA <- vector()
names_met_R <- vector()
mbiome_R_vect_num <- (as.vector(sapply(Meph1_RNA_Met_Mbiome2$Lactobacillus, as.numeric)))

#Loop
for (i in 1:ncol(Meph1_RNA_Met_num)){ #outside loop through all rna
   RNA_vect_num <- (as.vector(sapply(Meph1_RNA_Met_num[,i], as.numeric)))

for (j in 1:ncol(Meph1_Met_R)){ #inside loops for all metabolite - each RNA gets compared to all metabo
    met_R_vect_num <- (as.vector(sapply(Meph1_Met_R[,j], as.numeric)))


     #spearman test
   rhoval <- pcor.test(RNA_vect_num, met_R_vect_num, mbiome_R_vect_num,
                           method = "spearman")$estimate

    names(rhoval) <- NULL

    pval <- pcor.test(RNA_vect_num, met_R_vect_num, mbiome_R_vect_num,
                           method = "spearman")$p.value


#filter on p-value, save all 4 necessary information pieces into arrays
      if (pval <= 0.05){
      names_RNA <- append(names_RNA, colnames(Meph1_RNA_Met[i]))
      names_met_R <- append(names_met_R, colnames(Meph1_Met_R_Condition[j+1]))
      spear_test_p <- append(spear_test_p, pval)
      spear_test_rho <- append(spear_test_rho, rhoval)
       }
     }
   }

spear_prog_met <- cbind(names_RNA,names_met_R,spear_test_p, spear_test_rho)
```

```r
colnames(spear_prog_met) <- c("RNA","METABOLITE", "P Value", "Rho Value")
spear_prog_met <- data.frame(na.omit(spear_prog_met))
spear_prog_met <- spear_prog_met[!grepl('Numeric',spear_prog_met$METABOLITE),]
spear_prog_met <- spear_prog_met[!grepl('undistinguishable',spear_prog_met$METABOLITE),]
spear_prog_met <- transform(spear_prog_met,METABOLITE.num = as.numeric(as.factor(METABOLITE)))
spear_prog_met<- spear_prog_met %>%
    group_by(METABOLITE.num) %>%
    mutate(pval.adj = p.adjust (P.Value, method='BH'))
write.csv(spear_prog_met, file = "data/Outputs/Spearman test for Metabolite_RNA_Meph1.csv")
```

Vaginal Microbe/Metabolite vs Anticancer Drugs Gene Similarity Calculation

Upregulated

```r
#UPREGULATED CALCULATION
Table_1_Up <- read_excel("data/Inputs/Table 1_Up.xlsx")
MCF_Female_CD_up <- read_excel("data/Inputs/MCF_Female_CD_up.xlsx")
Tab_1_Up <- Table_1_Up
Tab_2_Up <- MCF_Female_CD_up
x_0 <- list()
x_1 <- list()
x2 <- list()
x3 <- list()

for (i in 1:ncol(Tab_2_Up)){
  Drug <- Tab_2_Up[,i]
  colnames(Drug) <- "temp"

  for (j in 1:ncol(Tab_1_Up)){
    Met <- Tab_1_Up[,j]
    colnames(Met) <- "temp"

    x2[j] <- as.data.frame(na.omit(union(unlist(as.vector(Drug)),unlist(as.vector(Met)))))
    x3[j] <- as.data.frame(na.omit(intersect(unlist(as.vector(Drug)),unlist(as.vector(Met)))))
  }
  x_0[i] <- list(x2)
  x_1[i] <- list(x3)
}

r1 <- rapply(x_1, length, how = "list")

Drug_Comp_Count_up <- as.matrix(do.call(rbind, r1))
Drug_Comp_List_up <- as.matrix(do.call(rbind, x_1))

Mb_Met_Count_up <- data.frame(Results = names(Tab_1_Up), Totals = sapply (Tab_1_Up, function(x) length(

MCF_Female_Drug_Count_up <- data.frame(Results = names(Tab_2_Up), Totals = sapply (Tab_2_Up, function(x

write.csv(Drug_Comp_Count_up,"data/Outputs/Drug_Comp_Count_up_Female_Meph1.csv", row.names = TRUE)
write.csv(Mb_Met_Count_up,"data/Outputs/Mb_Met_Count_up_Meph1.csv", row.names = FALSE)
write.csv(MCF_Female_Drug_Count_up,"data/Outputs/MCF_Female_Drug_Count_up_Meph1.csv", row.names = FALSE
```

```
#FISHERS TEST
UP <- read.table("data/Inputs/MCF_Up_Fishers(Final)_2.txt", header = FALSE)
Drug_pvalues_Up<-apply(UP,1,function(x) fisher.test(matrix(x,nr=2))$p.value)
write.csv(Drug_pvalues_Up,"data/Outputs/MCF_Up_Pvalues(Final)_2.csv", row.names = FALSE)
  #Group and put back in
MCF_Up_Pvalues_Final_2 <- read.csv("data/Inputs/MCF_Up_Pvalues(Final)_2.csv")
dat_up <- MCF_Up_Pvalues_Final_2
dat_up_adj <- dat_up %>%
    group_by(Group) %>%
    mutate(pval.adj = p.adjust (Pvalues, method='BH'))
write.csv(dat_up_adj,"data/Outputs/MCF_Up_Pvalues_adjusted(Final)_2.csv", row.names = FALSE)
```

Downregulated

```
#DOWNREGULATED CALCULATION

Table_1_Down <- read_excel("data/Inputs/Table 1_Down.xlsx")
MCF_Female_CD_down <- read_excel("data/Inputs/MCF_Female_CD_down.xlsx")
Tab_1_Down <- Table_1_Down
Tab_2_Down <- MCF_Female_CD_down
x_0_D <- list()
x_1_D <- list()
x2 <- list()
x3 <- list ()

for (i in 1:ncol(Tab_2_Down)){
  Drug <- Tab_2_Down[,i]
  colnames(Drug) <- "temp"

  for (j in 1:ncol(Tab_1_Down)){
    Met <- Tab_1_Down[,j]
    colnames(Met) <- "temp"

    x2[j] <- as.data.frame(na.omit(union(unlist(as.vector(Drug)),unlist(as.vector(Met)))))
    x3[j] <- as.data.frame(na.omit(intersect(unlist(as.vector(Drug)),unlist(as.vector(Met)))))

  }
  x_0_D[i] <- list(x2)
  x_1_D[i] <- list(x3)
}

r1 <- rapply(x_1_D, length, how = "list")

Drug_Comp_Count_down <- as.matrix(do.call(rbind, r1))

Drug_Comp_List_down <- as.matrix(do.call(rbind, x_1_D))

Mb_Met_Count_down <- data.frame(Results = names(Tab_1_Down), Totals = sapply (Tab_1_Down, function(x) le

MCF_Female_Drug_Count_down <- data.frame(Results = names(Tab_2_Down), Totals = sapply (Tab_2_Down, funct

write.csv(Drug_Comp_Count_down,"data/Outputs/Drug_Comp_Count_down_Female(Final).csv", row.names = TRUE)
write.csv(Mb_Met_Count_down,"data/Outputs/Mb_Met_Count_down_F(Final).csv", row.names = FALSE)
```

```
write.csv(MCF_Female_Drug_Count_down,"data/Outputs/MCF_Female_Drug_Count_down(Final).csv", row.names = F

#FISHERS TEST
DOWN <- read.table("data/Inputs/MCF_Down_FISHERS(Final)_2.txt", header = FALSE)
Drug_pvalues_Down<-apply(DOWN,1,function(x) fisher.test(matrix(x,nr=2))$p.value)
write.csv(Drug_pvalues_Down,"data/Outputs/MCF_Down_Pvalues(Final)_2.csv", row.names = FALSE)
#Group and put back in
MCF_Down_Pvalues_Final_2 <- read.csv("data/Inputs/MCF_Down_Pvalues(Final)_2.csv")
dat_down <- MCF_Down_Pvalues_Final_2
dat_down_adj <- dat_down %>%
    group_by(Group) %>%
    mutate(pval.adj = p.adjust (Pvalues, method='BH'))
write.csv(dat_down_adj,"data/Outputs/MCF_Down_Pvalues_adjusted(Final)_2.csv", row.names = FALSE)
```

Visualizations

Heatmaps

```
#HEATMAPS
##Upregulated
Heatmap_up_adj_microbe_70_ <- read.csv("data/Inputs/Heatmap_up_adj_microbe_70%.csv", row.names = "Drugs
#Row annotation
Drug_Class <- read.csv("data/Inputs/Drug Class.csv", row.names = "Drugs")
Drugs <- data.matrix(Heatmap_up_adj_microbe_70_)
pdf("PP.pdf",width = 5, height = 6)
P1 <- as.ggplot(pheatmap(Drugs, annotation_row = Drug_Class, main =  "Upregulated", cutree_rows = 6, cu
dev.off()

##Downregulated
Heatmap_down_adj_microbe_70_ <- read.csv("data/Inputs/Heatmap_down_adj_microbe_70%.csv", row.names = "R
#Row annotation
Drug_Class_2 <- read.csv("data/Outputs/Drug Class_2.csv", row.names = "Results")
Drugs <- data.matrix(Heatmap_down_adj_microbe_70_)
pdf("PP2.pdf",width = 5, height = 6)
P2 <- as.ggplot(pheatmap(Drugs, annotation_row = Drug_Class, main =  "Downregulated", cutree_rows = 6,
dev.off()
```

Scatter plots

Metabolites

```
#SCATTER PLOTS
##METABOLITES
###Fulvestrant
Metabolite_All_Scatter_Plot <- read_excel("data/Inputs/Metabolite_All Scatter Plot.xlsx")
Met_Scat <- Metabolite_All_Scatter_Plot
p<- ggplot(Met_Scat, aes(fulvestrant_up, fulvestrant_down, label = Drugs)) +    # ggplot2 plot with lab
 geom_hline(yintercept=5.3) +
 geom_vline(xintercept=5.3) +
 geom_point() +
 geom_jitter(width = 4) +
 ggtitle("Fulvestrant") +
```

```
 xlim(0,13) +
 ylim(0,12) +
 xlab(expression("-log10(Qvalue)Up")) +
 ylab(expression("-log10(Qvalue)Down")) +
 geom_text(data=subset(Met_Scat, fulvestrant_up >= 10 | fulvestrant_down >= 8),
           aes(fulvestrant_up, fulvestrant_down, label = Drugs), hjust = - 0.5, size = 10/5.5)
gridExtra::grid.arrange(egg::set_panel_size(p=p, width=unit(7, "cm"), height=unit(7, "cm")))
ggsave("Fulvestrant_Met Scatter Plot.pdf",height = 3.75, width = 3.75, unit = "in")

###Raloxifene
p<- ggplot(Met_Scat, aes(raloxifene_up, raloxifene_down, label = Drugs)) +    # ggplot2 plot with label
 geom_hline(yintercept=5.3) +
 geom_vline(xintercept=5.3) +
 geom_point() +
 geom_jitter(width = 4) +
 ggtitle("Raloxifene") +
 xlim(0,17) +
 ylim(0,15) +
 xlab(expression("-log10(Qvalue)Up")) +
 ylab(expression("-log10(Qvalue)Down")) +
 geom_text(data=subset(Met_Scat, raloxifene_up >= 12 | raloxifene_down >= 10),
           aes(raloxifene_up, raloxifene_down, label = Drugs), hjust = - 0.5, size = 10/5.5)
gridExtra::grid.arrange(egg::set_panel_size(p=p, width=unit(7, "cm"), height=unit(7, "cm")))
ggsave("Raloxifene_Met Scatter Plot.pdf",height = 3.75, width = 3.75, unit = "in")

##Etoposide
p<- ggplot(Met_Scat, aes(etoposide_up, etoposide_down, label = Drugs)) +    # ggplot2 plot with labels
 geom_hline(yintercept=5.3) +
 geom_vline(xintercept=5.3) +
 geom_point() +
 geom_jitter(width = 4) +
 ggtitle("Etoposide") +
 xlim(0,17) +
 ylim(0,15) +
 xlab(expression("-log10(Qvalue)Up")) +
 ylab(expression("-log10(Qvalue)Down")) +
 geom_text(data=subset(Met_Scat, etoposide_up >= 12 | etoposide_down >= 10),
           aes(etoposide_up, etoposide_down, label = Drugs), hjust = - 0.5, size = 10/5.5)
gridExtra::grid.arrange(egg::set_panel_size(p=p, width=unit(7, "cm"), height=unit(7, "cm")))
ggsave("Etoposide_Met Scatter Plot.pdf",height = 3.75, width = 3.75, unit = "in")
```

Microbes

```
#MICROBES
##Doxorubicin
Microbe_Scatter_Plot <- read_excel("data/Inputs/Microbe_Scatter _Plot.xlsx")
Mb_Scat <- Microbe_Scatter_Plot
p2<-ggplot(Mb_Scat, aes(doxorubicin_up, doxorubicin_down, label = Results)) +    # ggplot2 plot with la
 geom_hline(yintercept=1.3) +
 geom_vline(xintercept=10) +
 geom_point() +
 geom_jitter(width = 4) +
 ggtitle("Doxorubicin") +
```

```
 xlim(0,35) +
 ylim(0,7) +
 xlab(expression("-log10(Qvalue)Up")) +
 ylab(expression("-log10(Qvalue)Down")) +
 geom_text(data=subset(Mb_Scat, doxorubicin_up >= 2.3 | doxorubicin_down >= 2.3),
           aes(doxorubicin_up, doxorubicin_down, label = Results), hjust = - 0.5, size = 10/5.5)
gridExtra::grid.arrange(egg::set_panel_size(p=p2, width=unit(5, "cm"), height=unit(5, "cm")))
ggsave("Doxorubicin_Mb Scatter Plot.pdf",height = 3.75, width = 3.75, unit = "in")
p2

##Everolimus
p2<-ggplot(Mb_Scat, aes(everolimus_up, everolimus_down, label = Results)) +    # ggplot2 plot with labe
 geom_hline(yintercept=10) +
 geom_vline(xintercept=1.3) +
 geom_point() +
 geom_jitter(width = 4) +
 ggtitle("Everolimus") +
 xlim(0,6) +
 ylim(0,25) +
 xlab(expression("-log10(Qvalue)Up")) +
 ylab(expression("-log10(Qvalue)Down")) +
 geom_text(data=subset(Mb_Scat, everolimus_up >= 2.3 | everolimus_down >= 2.3),
           aes(everolimus_up, everolimus_down, label = Results), hjust = - 0.5, size = 10/5.5)
gridExtra::grid.arrange(egg::set_panel_size(p=p2, width=unit(5, "cm"), height=unit(5, "cm")))
ggsave("Everolimus_Mb Scatter Plot.pdf",height = 3.75, width = 3.75, unit = "in")

##Raloxifene
p2<-ggplot(Mb_Scat, aes(raloxifene_up, raloxifene_down, label = Results)) +    # ggplot2 plot with labe
 geom_hline(yintercept=5.3) +
 geom_vline(xintercept=5.3) +
 geom_point() +
 geom_jitter(width = 4) +
 ggtitle("Raloxifene") +
 xlim(5,25) +
 ylim(5,20) +
 xlab(expression("-log10(Qvalue)Up")) +
 ylab(expression("-log10(Qvalue)Down")) +
 geom_text(data=subset(Mb_Scat, raloxifene_up >= 2.3 | raloxifene_down >= 2.3),
           aes(raloxifene_up, raloxifene_down, label = Results), hjust = - 0.5, size = 10/5.5)
gridExtra::grid.arrange(egg::set_panel_size(p=p2, width=unit(5, "cm"), height=unit(5, "cm")))
ggsave("Raloxifene_Mb Scatter Plot.pdf",height = 3.75, width = 3.75, unit = "in")
```

Box plots

```
#BOX PLOTS
##Suspension
###Cytosine
Log2FC_Sig_Metabolite_Suspension <- read_excel("data/Inputs/Log2FC Sig Metabolite_Suspension.xlsx")
Log2FC_Sig_Metabolite_Suspension$Metabolite <- factor(Log2FC_Sig_Metabolite_Suspension$Metabolite, level
myColors <- ifelse(levels(Log2FC_Sig_Metabolite_Suspension$Metabolite)=="L_cris_S", "purple",
              ifelse(levels(Log2FC_Sig_Metabolite_Suspension$Metabolite)=="L_iners_S", "gray",
               ifelse(levels(Log2FC_Sig_Metabolite_Suspension$Metabolite)=="G_vag_S", "firebrick3",
                        "black" )))
```

```
pdf("Cytosine_S.pdf", width = 5, height = 5 )
boxplot(Log2FC_Sig_Metabolite_Suspension$cytosine ~ Log2FC_Sig_Metabolite_Suspension$Metabolite,
data=Log2FC_Sig_Metabolite_Suspension,
main="Cytosine",
xlab="Bacterial Taxa (Suspension S)",
ylab="log2Fold Change(Bacteria:Media)",
col=myColors
  )
dev.off()
###Taurine
Log2FC_Sig_Metabolite_Suspension <- read_excel("data/Inputs/Log2FC Sig Metabolite_Suspension.xlsx")
Log2FC_Sig_Metabolite_Suspension$Metabolite <- factor(Log2FC_Sig_Metabolite_Suspension$Metabolite, level
myColors <- ifelse(levels(Log2FC_Sig_Metabolite_Suspension$Metabolite)=="L_cris_S", "purple",
              ifelse(levels(Log2FC_Sig_Metabolite_Suspension$Metabolite)=="L_iners_S", "gray",
               ifelse(levels(Log2FC_Sig_Metabolite_Suspension$Metabolite)=="G_vag_S", "firebrick3",
                        "black" )))

pdf("Taurine_S.pdf", width = 5, height = 5 )
boxplot(Log2FC_Sig_Metabolite_Suspension$taurine ~ Log2FC_Sig_Metabolite_Suspension$Metabolite,
data=Log2FC_Sig_Metabolite_Suspension,
main="Taurine",
xlab="Bacterial Taxa (Suspension S)",
ylab="log2Fold Change(Bacteria:Media)",
col=myColors
  )
dev.off()
```