

Expression Analysis of Breast Cancer Cell-lines (E-GEOD-18494, GSE47533 and GSE41491)

Rede Alexandre (reunião 27-11-2020)

HIF1a, !VHL & !O2
 p53, !Mdm2
 Mdm2, p53 & !VHL
 VHL, HIF1a & !p53
 p300, ((p53 & HIF1a) & !VHL) | (!p53 & HIF1a) & VHL
 BIM, !MCL_1 & !BCLXL & !BCL2
 BAD, p53
 BID, (!HIF1a & (p53 & VHL)) | (!MCL_1 & !BCLXL & !BCL2)
 BIK, !MCL_1 & !BCLXL & !BCL2
 MCL_1, HIF1a
 BCLXL, HIF1a & !(p53 & VHL) & ((!Casp3 & !BAD) | (!Casp3 & BCL2))
 BCL2, HIF1a & !(p53 & VHL) & ((MCL_1 & !BIM & !BIK & !BAD) | (!BIM & !BIK & BCLXL & !BAD))
 IAPs, !DIABLO
 BAX, (BIM & !BCLXL) | (BIK & !BCLXL & !BCL2) | (BID & !BCLXL & !BCL2) | (BIM & BID) | (BIM & BIK) | (BIM & !BCL2) | (!MCL_1 & BIM)
 BAK, (!MCL_1 & BIM & !BCLXL) | (BID & !BCL2) | (BID & !BCLXL) | (!MCL_1 & BID) | (!MCL_1 & BIK & !BCLXL) | (BIM & BID) | (BIK & BID)
 DIABLO, BAX | BAK
 Cyto_C, BAX | BAK
 Casp9, Casp3 | (!IAPs & Cyto_C)
 Casp3, !IAPs & Casp9
 BCLXL ?
 No VHL in GSE41491

```
# Selected genes from HIF Axis
hif.symbols <- c("HIF1A", "TP53", "MDM2", "VHL", "EP300", "TMBIM1", "TMBIM4", "TMBIM6", "BAD", "BIK", "MCL1")

hif.probes <- anno.EGEOD18494$probes[anno.EGEOD18494$symbol %in% hif.symbols]

# Select the probes and genes
# EGEOD18494
expr.EGEOD18494.hif <- as.data.frame(expr.EGEOD18494) %>%
  rownames_to_column('probes') %>%
  filter(probes %in% hif.probes) %>%
  merge(anno.EGEOD18494[anno.EGEOD18494$symbol %in% hif.symbols, c("probes", "symbol")], by = "probes") %>%
  mutate(., symbol=ifelse(symbol %in% c("TMBIM1", "TMBIM4", "TMBIM6"), "BIM", symbol)) %>%
  mutate(., symbol=ifelse(symbol %in% c("BIRC2", "BIRC3", "BIRC5", "BIRC6", "BIRC7"), "IAPs", symbol)) %>%
  group_by(symbol) %>%
  summarise_at(vars(-probes), funs(mean(., na.rm=TRUE))) %>%
  column_to_rownames(var = "symbol") %>%
  dplyr::select(c(data.EGEOD18494$codes[data.EGEOD18494$cell_line == "MDA-MB231 breast cancer"])))
```

```
## Warning: `funs()` is deprecated as of dplyr 0.8.0.
## Please use a list of either functions or lambdas:
##
##   # Simple named list:
##   list(mean = mean, median = median)
##
##   # Auto named with `tibble::lst()`:
##   tibble::lst(mean, median)
##
##   # Using lambdas
##   list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_warnings()` to see where this warning was generated.
```

```
hif.probes <- anno.GSE47533$probes[anno.GSE47533$symbol %in% hif.symbols]
```

```
# GSE47533
```

```
expr.GSE47533.hif <- as.data.frame(expr.GSE47533) %>%
  rownames_to_column('probes') %>%
  filter(probes %in% hif.probes) %>%
  merge(anno.GSE47533[anno.GSE47533$symbol %in% hif.symbols, c("probes", "symbol")], by = "probes") %>%
  mutate(., symbol = ifelse(symbol %in% c("TMBIM1", "TMBIM4", "TMBIM6"), "BIM", symbol)) %>%
  mutate(., symbol = ifelse(symbol %in% c("BIRC2", "BIRC3", "BIRC5", "BIRC6", "BIRC7"), "IAPs", symbol)) %>%
  group_by(symbol) %>%
  summarise_at(vars(-probes), funs(mean(., na.rm = TRUE))) %>%
  column_to_rownames(var = "symbol")
```

```
hif.probes <- anno.GSE41491$probes[anno.GSE41491$symbol %in% hif.symbols]
```

```
# GSE41491
```

```
expr.GSE41491.hif <- as.data.frame(expr.GSE41491) %>%
  rownames_to_column('probes') %>%
  filter(probes %in% hif.probes) %>%
  merge(anno.GSE41491[anno.GSE41491$symbol %in% hif.symbols, c("probes", "symbol")], by = "probes") %>%
  mutate(., symbol = ifelse(symbol %in% c("TMBIM1", "TMBIM4", "TMBIM6"), "BIM", symbol)) %>%
  mutate(., symbol = ifelse(symbol %in% c("BIRC2", "BIRC3", "BIRC5", "BIRC6", "BIRC7"), "IAPs", symbol)) %>%
  group_by(symbol) %>%
  summarise_at(vars(-probes), funs(mean(., na.rm = TRUE))) %>%
  column_to_rownames(var = "symbol")
```

```
write.table(expr.GSE47533.hif, "expr.GSE47533.txt", sep = "\t")
write.table(expr.EGEOD18494.hif, "expr.EGEOD18494.txt", sep = "\t")
write.table(expr.GSE41491.hif, "expr.GSE41491.txt", sep = "\t")
```

```
expr.EGEOD18494.tdm <- tdm_transform(ref_file = "expr.GSE47533.txt", file = "expr.EGEOD18494.txt")
```

```
##
## Attaching package: 'data.table'
```

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

```

## The following object is masked from 'package:purrr':
##
##      transpose

##
## Attaching package: 'scales'

## The following object is masked from 'package:purrr':
##
##      discard

## The following object is masked from 'package:readr':
##
##      col_factor

expr.GSE41491.tdm <- tdm_transform(ref_file = "expr.GSE47533.txt", file = "expr.GSE41491.txt")

symbols <- expr.EGEOD18494.tdm$gene
expr.EGEOD18494.tdm$gene <- NULL

expr.EGEOD18494.tdm <- as.data.frame(matrix(as.numeric(unlist(expr.EGEOD18494.tdm)),
                                           nrow = dim(expr.EGEOD18494.tdm)[1],
                                           ncol = dim(expr.EGEOD18494.tdm)[2])),
colnames(expr.EGEOD18494.tdm) <- colnames(expr.EGEOD18494.hif)

rownames(expr.EGEOD18494.tdm) <- symbols

row_medians_assayData <-
  Biobase::rowMedians(as.matrix(expr.GSE47533))

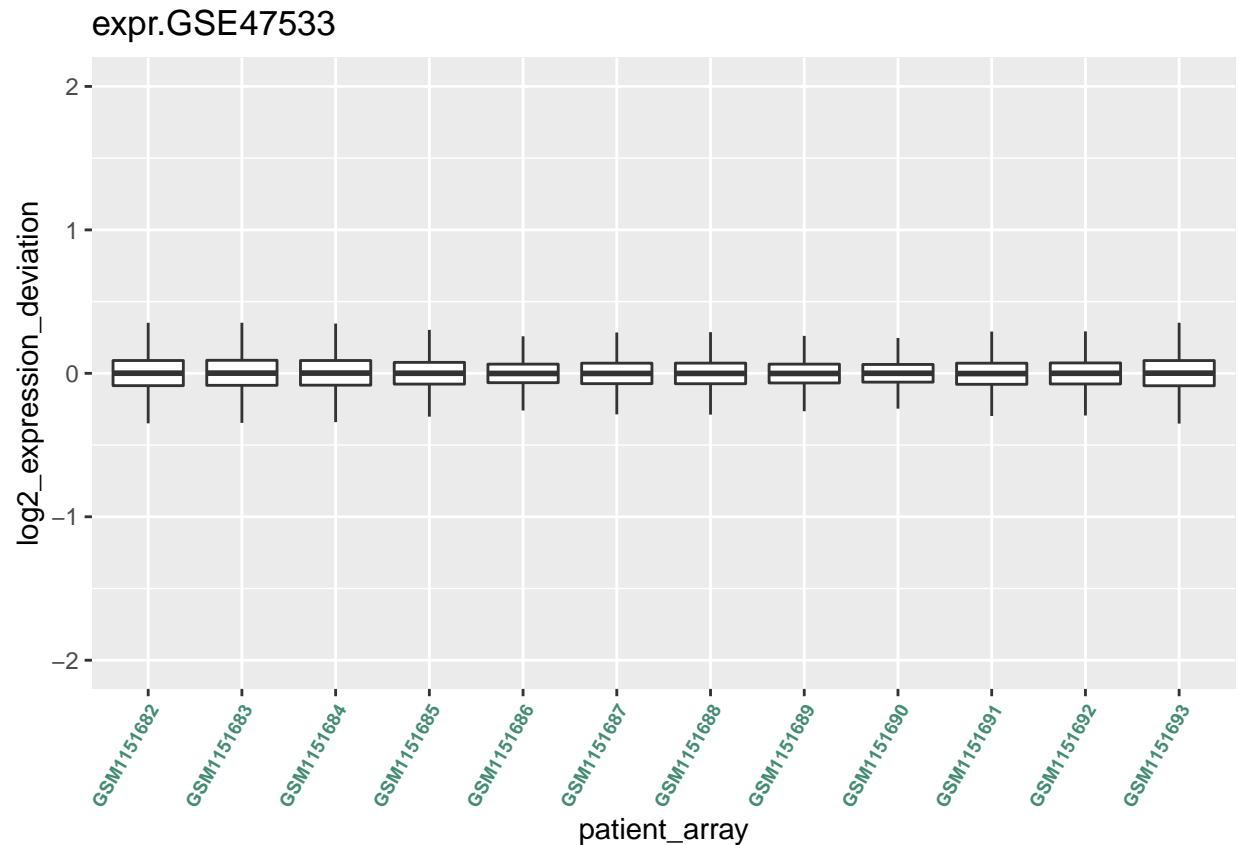
RLE_data <- sweep(expr.GSE47533, 1, row_medians_assayData)

RLE_data <- as.data.frame(RLE_data)
RLE_data_gathered <-
  tidyr::gather(RLE_data, patient_array, log2_expression_deviation)

ggplot2::ggplot(RLE_data_gathered, aes(patient_array,
                                       log2_expression_deviation)) +
  geom_boxplot(outlier.shape = NA) +
  ylim(c(-2, 2)) +
  ggtitle("expr.GSE47533") +
  theme(axis.text.x = element_text(colour = "aquamarine4",
                                    angle = 60, size = 6.5, hjust = 1 ,
                                    face = "bold"))

## Warning: Removed 142 rows containing non-finite values (stat_boxplot).

```

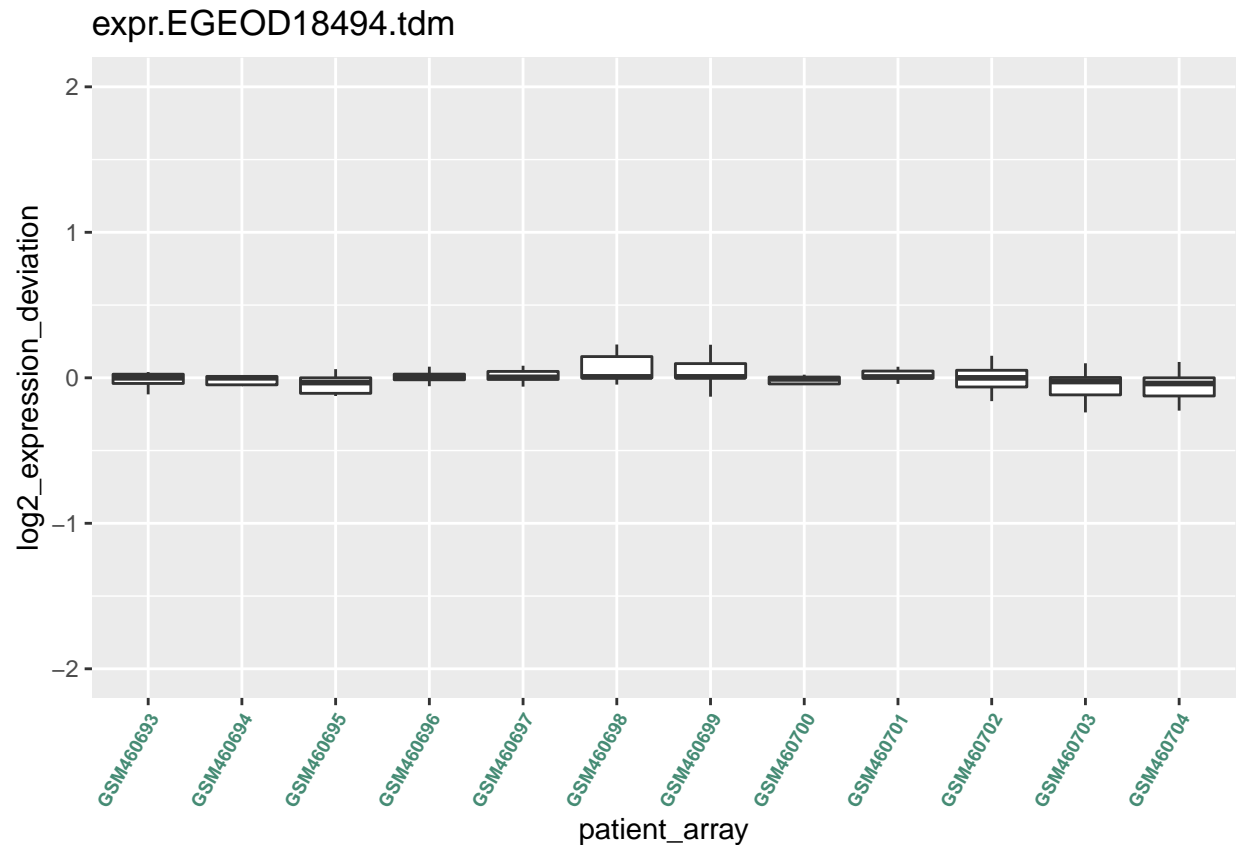


```
row_medians_assayData <-
  Biobase::rowMedians(as.matrix(expr.EGEOD18494.tdm))

RLE_data <- sweep(expr.EGEOD18494.tdm, 1, row_medians_assayData)

RLE_data <- as.data.frame(RLE_data)
RLE_data_gathered <-
  tidyr::gather(RLE_data, patient_array, log2_expression_deviation)

ggplot2::ggplot(RLE_data_gathered, aes(patient_array,
                                       log2_expression_deviation)) +
  geom_boxplot(outlier.shape = NA) +
  ylim(c(-2, 2)) +
  ggtitle("expr.EGEOD18494.tdm") +
  theme(axis.text.x = element_text(colour = "aquamarine4",
                                    angle = 60, size = 6.5, hjust = 1 ,
                                    face = "bold"))
```

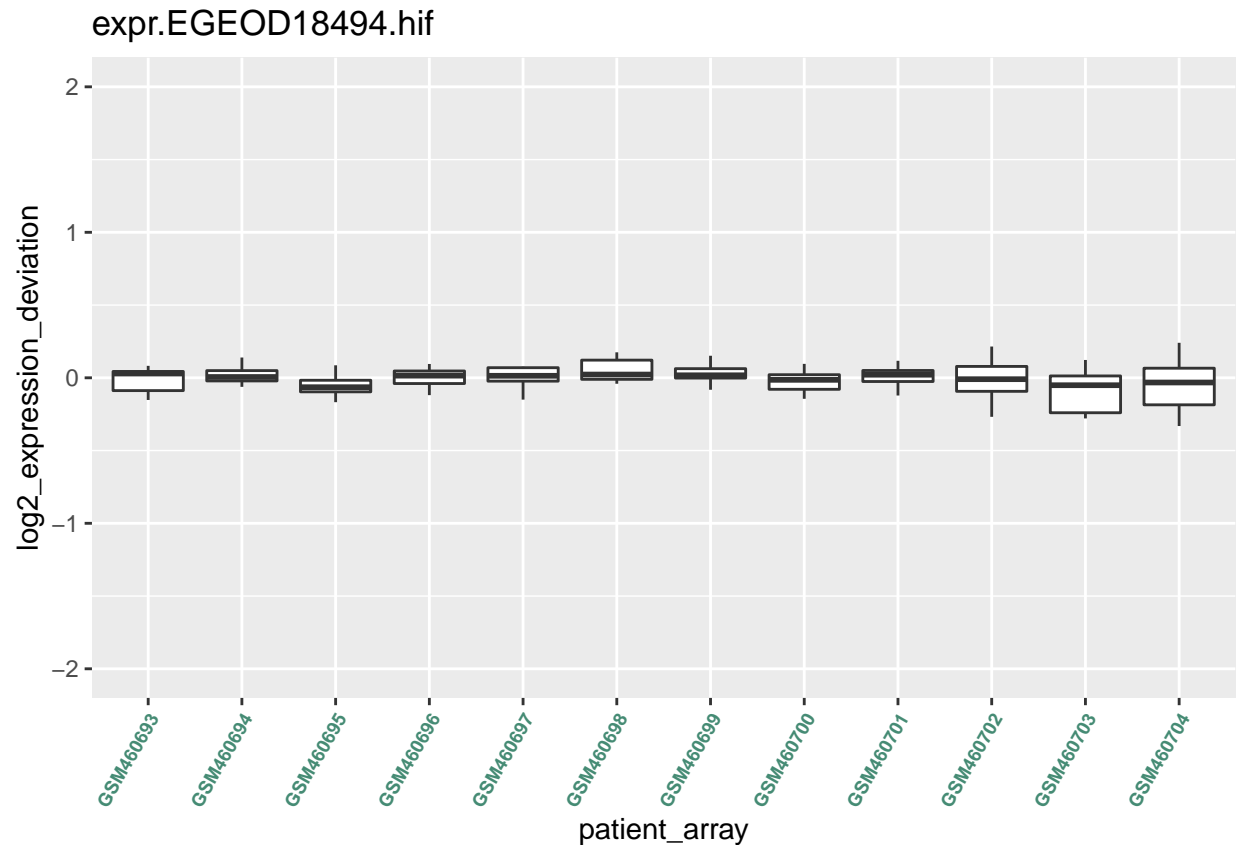


```
row_medians_assayData <-
  Biobase::rowMedians(as.matrix(expr.EGEOD18494.hif))

RLE_data <- sweep(expr.EGEOD18494.hif, 1, row_medians_assayData)

RLE_data <- as.data.frame(RLE_data)
RLE_data_gathered <-
  tidyr::gather(RLE_data, patient_array, log2_expression_deviation)

ggplot2::ggplot(RLE_data_gathered, aes(patient_array,
                                       log2_expression_deviation)) +
  geom_boxplot(outlier.shape = NA) +
  ylim(c(-2, 2)) +
  ggtitle("expr.EGEOD18494.hif") +
  theme(axis.text.x = element_text(colour = "aquamarine4",
                                    angle = 60, size = 6.5, hjust = 1 ,
                                    face = "bold"))
```



```
rm(RLE_data, RLE_data_gathered, row_medians_assayData)
```

```
hif.probes <- anno.GSE47533$probes[anno.GSE47533$symbol %in% hif.symbols]
```

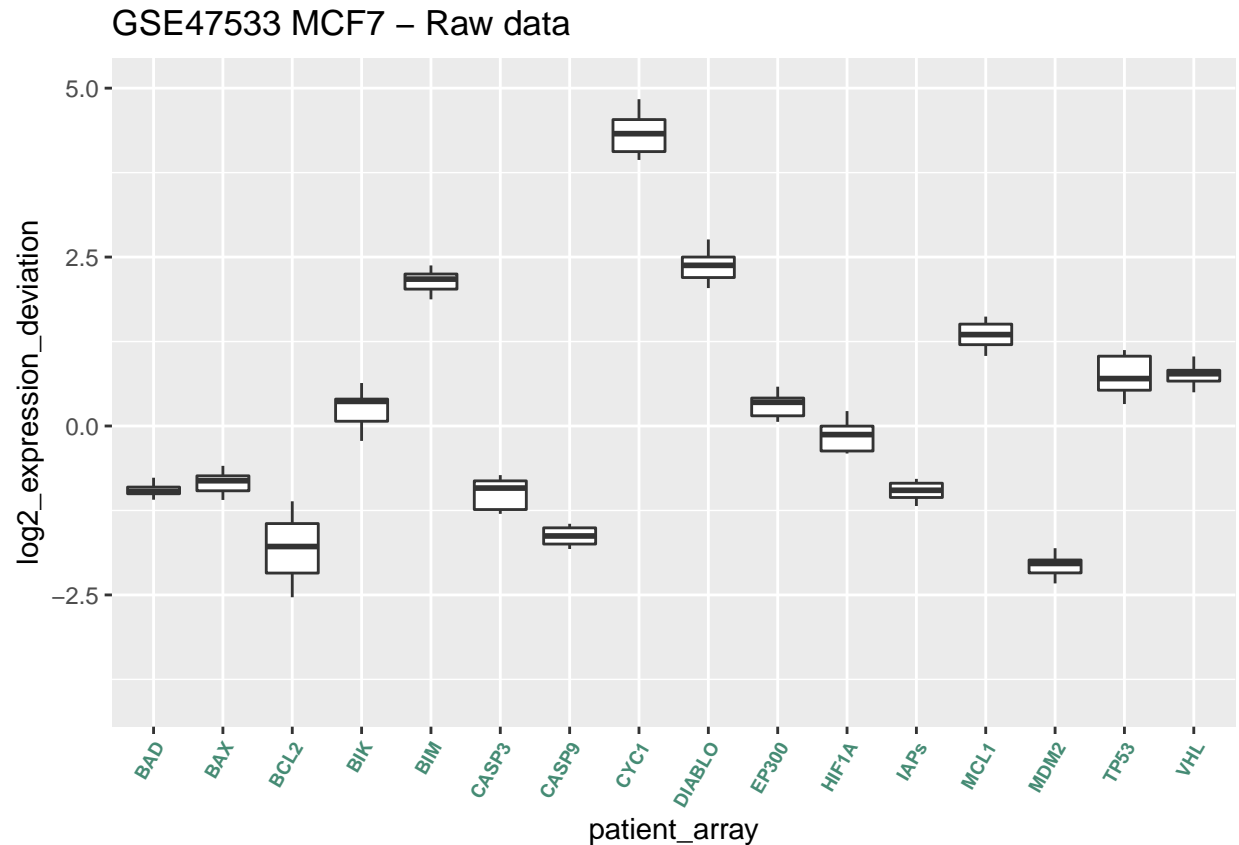
```
row_medians_assayData <-  
  Biobase::rowMedians(as.matrix(t(expr.GSE47533.hif)))
```

```
RLE_data <- sweep(t(expr.GSE47533.hif), 1, row_medians_assayData)
```

```
RLE_data <- as.data.frame(RLE_data)
```

```
RLE_data_gathered <-  
  tidyr::gather(RLE_data, patient_array, log2_expression_deviation)
```

```
ggplot2::ggplot(RLE_data_gathered, aes(patient_array,  
                                       log2_expression_deviation)) +  
  geom_boxplot(outlier.shape = NA) +  
  ylim(c(-4, 5)) +  
  ggtitle("GSE47533 MCF7 - Raw data") +  
  theme(axis.text.x = element_text(colour = "aquamarine4",  
                                    angle = 60, size = 6.5, hjust = 1 ,  
                                    face = "bold"))
```

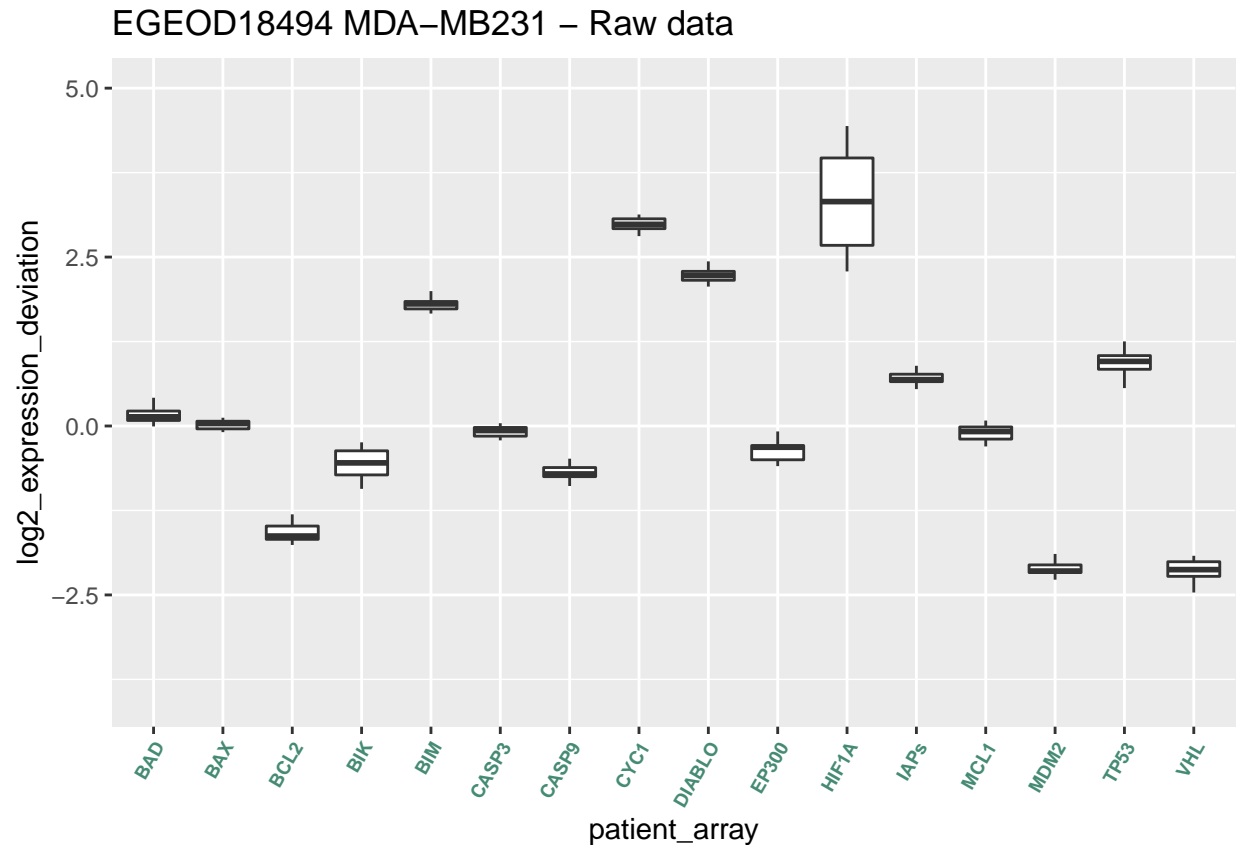


```
row_medians_assayData <-
  Biobase::rowMedians(as.matrix(t(expr.EGEOD18494.hif)))

RLE_data <- sweep(t(expr.EGEOD18494.hif), 1, row_medians_assayData)

RLE_data <- as.data.frame(RLE_data)
RLE_data_gathered <-
  tidyr::gather(RLE_data, patient_array, log2_expression_deviation)

ggplot2::ggplot(RLE_data_gathered, aes(patient_array,
                                       log2_expression_deviation)) +
  geom_boxplot(outlier.shape = NA) +
  ggtitle("EGEOD18494 MDA-MB231 - Raw data") +
  ylim(c(-4, 5)) +
  theme(axis.text.x = element_text(colour = "aquamarine4",
                                    angle = 60, size = 6.5, hjust = 1 ,
                                    face = "bold"))
```

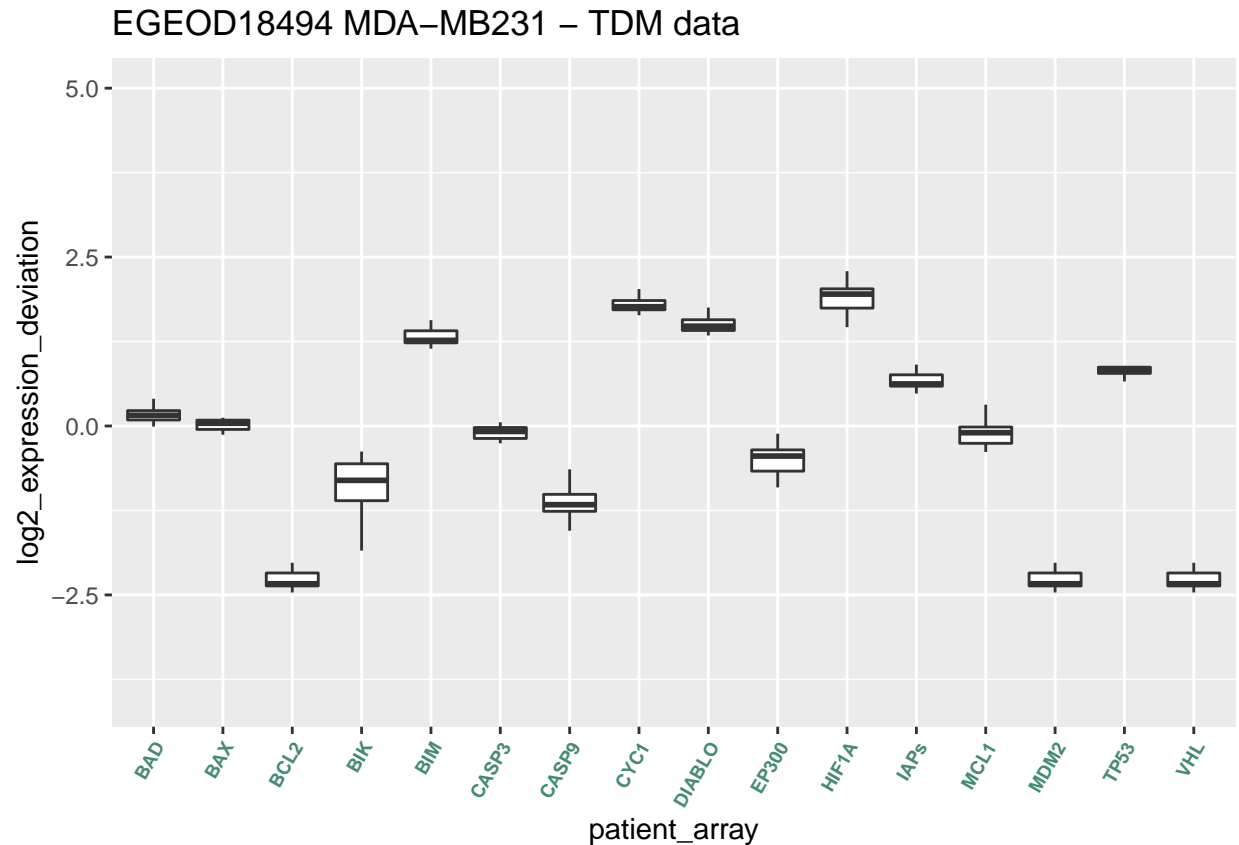


```
row_medians_assayData <-
  Biobase::rowMedians(as.matrix(t(expr.EGEOD18494.tdm)))

RLE_data <- sweep(t(expr.EGEOD18494.tdm), 1, row_medians_assayData)

RLE_data <- as.data.frame(RLE_data)
RLE_data_gathered <-
  tidyr::gather(RLE_data, patient_array, log2_expression_deviation)

ggplot2::ggplot(RLE_data_gathered, aes(patient_array,
                                       log2_expression_deviation)) +
  geom_boxplot(outlier.shape = NA) +
  ylim(c(-4, 5)) +
  ggtitle("EGEOD18494 MDA-MB231 - TDM data") +
  theme(axis.text.x = element_text(colour = "aquamarine4",
                                    angle = 60, size = 6.5, hjust = 1 ,
                                    face = "bold"))
```

```
rm(RLE_data, RLE_data_gathered, row_medians_assayData)
```

```
require(BiTrinA)
```

```
## Loading required package: BiTrinA
```

```
## Loading required package: diptest
```

```
expr.GSE47533.hif.bin <- binarizeMatrix(expr.GSE47533.hif,
  method = c("BASCA"),
  adjustment = "none")

expr.GSE47533.hif.bin$symbol <- row.names(expr.GSE47533.hif.bin)

expr.GSE47533.hif.bin <- expr.GSE47533.hif.bin[, c(as.character(data.GSE47533$codes), c("threshold", "p

names(expr.GSE47533.hif.bin) <- c(paste0(substr(data.GSE47533$condition,1,4), ".", data.GSE47533$time, "

head(expr.GSE47533.hif.bin) %>%
  knitr::kable(.)
```

	Norm.0	Norm.1	Norm.2	Hypo.16h	Hypo.16h	Hypo.16h	Hypo.32h	Hypo.32h	Hypo.32h	Hypo.32h	Hypo.48h	Hypo.48h	Hypo.48h	log2value	symbol
BAD	0	0	0	0	0	0	0	0	0	1	1	1	1	7.925009	BAD
BAX	0	1	1	0	0	0	1	0	0	1	1	1	1	7.978194	BAX
BCL2	1	1	1	0	0	0	0	0	0	0	0	0	0	7.272830	BCL2
BIK	0	0	0	1	1	1	1	1	1	1	1	1	1	8.755742	BIK
BIM	1	1	0	1	0	0	1	0	1	1	0	1	1	10.921169	BIM
CASP3	1	1	1	0	0	1	1	0	0	0	0	0	0	7.799389	CASP3

```

expr.GSE47533.hif.mean <- expr.GSE47533.hif.bin %>%
  mutate(Norm = rowMeans(dplyr::select(., starts_with("Norm"))),
         Hypo.16h = rowMeans(dplyr::select(., starts_with("Hypo.16h"))),
         Hypo.32h = rowMeans(dplyr::select(., starts_with("Hypo.32h"))),
         Hypo.48h = rowMeans(dplyr::select(., starts_with("Hypo.48h")))) %>%
  dplyr::select(., -ends_with(c(".1", ".2", ".3")))

expr.GSE47533.hif.pivot <- expr.GSE47533.hif.mean %>%
  group_by(symbol) %>%
  pivot_longer(cols = starts_with(c("Norm", "Hypo")), names_to = "codes", values_to = "value")

expr.GSE47533.hif.pivot$codes <- factor(expr.GSE47533.hif.pivot$codes, levels = c("Norm", "Hypo.16h", "Hypo.32h", "Hypo.48h"))

expr.GSE47533.hif.pivot$time <- as.numeric(expr.GSE47533.hif.pivot$codes)

# hif.symbols <- c("HIF1A", "TP53", "MDM2", "VHL", "EP300", "TMBIM1", "TMBIM4", "TMBIM6", "BAD", "BIK", "MCL1", "BAX", "BCL2", "CASP3")

p.MCF7 <- ggplot(aes(x = factor(time), y = value, group = symbol, color="red"),
  data = expr.GSE47533.hif.pivot[expr.GSE47533.hif.pivot$symbol %in% c("HIF1A", "TP53", "MDM2")])
  geom_point() +
  geom_line() +
  scale_x_discrete(breaks = c(1, 2, 3, 4),
    labels = c("Normoxia", "Hypoxia: 16h", "Hypoxia: 32h", "Hypoxia: 48h")) +
  xlab("Conditions") + ylab("Gene Expression") +
  ggtitle("GSE47533 MCF7 - Raw data") +
  theme(legend.position = "none", axis.text.x=element_text(color = "black", size=7, angle=30, vjust=.8),
    #geom_line(aes(linetype=Symbol, color=Symbol)) +
    facet_wrap(~ symbol))
p.MCF7

```

Figure 2 displays 16 line graphs showing gene expression levels for various genes across four conditions: Normoxia, Hypoxia: 16h, Hypoxia: 32h, and Hypoxia: 48h. The y-axis represents Gene Expression (0.00 to 1.00). The genes are arranged in a 4x4 grid. The genes and their expression trends are summarized in the table below:

Gene	Normoxia	Hypoxia: 16h	Hypoxia: 32h	Hypoxia: 48h
BAD	0.00	0.00	0.00	1.00
BAX	0.65	0.00	0.35	1.00
BCL2	1.00	0.00	0.00	0.00
BIK	0.00	1.00	1.00	1.00
BIM	0.65	0.35	0.65	0.65
CASP3	1.00	0.35	0.35	0.00
CASP9	0.65	0.35	0.65	1.00
CYC1	1.00	0.00	0.00	0.00
DIABLO	1.00	0.00	0.00	0.00
EP300	0.00	0.35	0.65	0.00
HIF1A	1.00	0.00	0.00	1.00
IAPs	1.00	0.65	1.00	1.00
MCL1	1.00	0.35	0.65	0.00
MDM2	1.00	1.00	0.65	1.00
TP53	0.65	1.00	0.35	0.00
VHL	0.65	0.00	0.00	1.00

```
# expr.EGEOD18494.hif.bin <- binarizeMatrix(expr.EGEOD18494.tdm,
#                                     method = c("BASCA"),
#                                     tau = 0.15,
#                                     #sigma = 0.9,
#                                     adjustment = "none")

expr.EGEOD18494.hif.bin <- binarizeMatrix(expr.EGEOD18494.hif)

#expr.EGEOD18494.hif.bin <- expr.EGEOD18494.tdm

expr.EGEOD18494.hif.bin$symbol <- row.names(expr.EGEOD18494.hif.bin)

row <- data.EGEOD18494$cell_line == "MDA-MB231 breast cancer"
expr.EGEOD18494.hif.bin <- expr.EGEOD18494.hif.bin[, c(as.character(data.EGEOD18494$codes[row]), c("thr", "pro", "hif", "p53", "p21", "p27", "p29", "p31", "p35", "p37", "p39", "p42", "p44", "p51", "p53", "p57", "p63", "p70", "p73", "p75", "p79", "p80", "p85", "p86", "p90", "p91", "p93", "p97", "p101", "p105", "p107", "p110", "p115", "p116", "p119", "p121", "p125", "p127", "p130", "p135", "p137", "p140", "p143", "p145", "p147", "p150", "p153", "p155", "p157", "p160", "p161", "p164", "p167", "p170", "p171", "p173", "p175", "p177", "p180", "p183", "p185", "p187", "p188", "p191", "p193", "p195", "p197", "p199", "p201", "p203", "p205", "p207", "p209", "p211", "p213", "p215", "p217", "p219", "p221", "p223", "p225", "p227", "p229", "p231", "p233", "p235", "p237", "p239", "p241", "p243", "p245", "p247", "p249", "p251", "p253", "p255", "p257", "p259", "p261", "p263", "p265", "p267", "p269", "p271", "p273", "p275", "p277", "p279", "p281", "p283", "p285", "p287", "p289", "p291", "p293", "p295", "p297", "p299", "p301", "p303", "p305", "p307", "p309", "p311", "p313", "p315", "p317", "p319", "p321", "p323", "p325", "p327", "p329", "p331", "p333", "p335", "p337", "p339", "p341", "p343", "p345", "p347", "p349", "p351", "p353", "p355", "p357", "p359", "p361", "p363", "p365", "p367", "p369", "p371", "p373", "p375", "p377", "p379", "p381", "p383", "p385", "p387", "p389", "p391", "p393", "p395", "p397", "p399", "p401", "p403", "p405", "p407", "p409", "p411", "p413", "p415", "p417", "p419", "p421", "p423", "p425", "p427", "p429", "p431", "p433", "p435", "p437", "p439", "p441", "p443", "p445", "p447", "p449", "p451", "p453", "p455", "p457", "p459", "p461", "p463", "p465", "p467", "p469", "p471", "p473", "p475", "p477", "p479", "p481", "p483", "p485", "p487", "p489", "p491", "p493", "p495", "p497", "p499", "p501", "p503", "p505", "p507", "p509", "p511", "p513", "p515", "p517", "p519", "p521", "p523", "p525", "p527", "p529", "p531", "p533", "p535", "p537", "p539", "p541", "p543", "p545", "p547", "p549", "p551", "p553", "p555", "p557", "p559", "p561", "p563", "p565", "p567", "p569", "p571", "p573", "p575", "p577", "p579", "p581", "p583", "p585", "p587", "p589", "p591", "p593", "p595", "p597", "p599", "p601", "p603", "p605", "p607", "p609", "p611", "p613", "p615", "p617", "p619", "p621", "p623", "p625", "p627", "p629", "p631", "p633", "p635", "p637", "p639", "p641", "p643", "p645", "p647", "p649", "p651", "p653", "p655", "p657", "p659", "p661", "p663", "p665", "p667", "p669", "p671", "p673", "p675", "p677", "p679", "p681", "p683", "p685", "p687", "p689", "p691", "p693", "p695", "p697", "p699", "p701", "p703", "p705", "p707", "p709", "p711", "p713", "p715", "p717", "p719", "p721", "p723", "p725", "p727", "p729", "p731", "p733", "p735", "p737", "p739", "p741", "p743", "p745", "p747", "p749", "p751", "p753", "p755", "p757", "p759", "p761", "p763", "p765", "p767", "p769", "p771", "p773", "p775", "p777", "p779", "p781", "p783", "p785", "p787", "p789", "p791", "p793", "p795", "p797", "p799", "p801", "p803", "p805", "p807", "p809", "p811", "p813", "p815", "p817", "p819", "p821", "p823", "p825", "p827", "p829", "p831", "p833", "p835", "p837", "p839", "p841", "p843", "p845", "p847", "p849", "p851", "p853", "p855", "p857", "p859", "p861", "p863", "p865", "p867", "p869", "p871", "p873", "p875", "p877", "p879", "p881", "p883", "p885", "p887", "p889", "p891", "p893", "p895", "p897", "p899", "p901", "p903", "p905", "p907", "p909", "p911", "p913", "p915", "p917", "p919", "p921", "p923", "p925", "p927", "p929", "p931", "p933", "p935", "p937", "p939", "p941", "p943", "p945", "p947", "p949", "p951", "p953", "p955", "p957", "p959", "p961", "p963", "p965", "p967", "p969", "p971", "p973", "p975", "p977", "p979", "p981", "p983", "p985", "p987", "p989", "p991", "p993", "p995", "p997", "p999", "p1001", "p1003", "p1005", "p1007", "p1009", "p1011", "p1013", "p1015", "p1017", "p1019", "p1021", "p1023", "p1025", "p1027", "p1029", "p1031", "p1033", "p1035", "p1037", "p1039", "p1041", "p1043", "p1045", "p1047", "p1049", "p1051", "p1053", "p1055", "p1057", "p1059", "p1061", "p1063", "p1065", "p1067", "p1069", "p1071", "p1073", "p1075", "p1077", "p1079", "p1081", "p1083", "p1085", "p1087", "p1089", "p1091", "p1093", "p1095", "p1097", "p1099", "p1101", "p1103", "p1105", "p1107", "p1109", "p1111", "p1113", "p1115", "p1117", "p1119", "p1121", "p1123", "p1125", "p1127", "p1129", "p1131", "p1133", "p1135", "p1137", "p1139", "p1141", "p1143", "p1145", "p1147", "p1149", "p1151", "p1153", "p1155", "p1157", "p1159", "p1161", "p1163", "p1165", "p1167", "p1169", "p1171", "p1173", "p1175", "p1177", "p1179", "p1181", "p1183", "p1185", "p1187", "p1189", "p1191", "p1193", "p1195", "p1197", "p1199", "p1201", "p1203", "p1205", "p1207", "p1209", "p1211", "p1213", "p1215", "p1217", "p1219", "p1221", "p1223", "p1225", "p1227", "p1229", "p1231", "p1233", "p1235", "p1237", "p1239", "p1241", "p1243", "p1245", "p1247", "p1249", "p1251", "p1253", "p1255", "p1257", "p1259", "p1261", "p1263", "p1265", "p1267", "p1269", "p1271", "p1273", "p1275", "p1277", "p1279", "p1281", "p1283", "p1285", "p1287", "p1289", "p1291", "p1293", "p1295", "p1297", "p1299", "p1301", "p1303", "p1305", "p1307", "p
```

```

expr.EGEOD18494.hif.mean <- expr.EGEOD18494.hif.bin %>%
  mutate(norm = rowMeans(dplyr::select(., starts_with("norm"))),
         hypo.4h = rowMeans(dplyr::select(., starts_with("hypo.4h"))),
         hypo.8h = rowMeans(dplyr::select(., starts_with("hypo.8h"))),
         hypo.12h = rowMeans(dplyr::select(., starts_with("hypo.12h")))) %>%
  dplyr::select(., -ends_with(c(".1", ".2", ".3")))

expr.EGEOD18494.hif.pivot <- expr.EGEOD18494.hif.mean %>%
  group_by(symbol) %>%
  pivot_longer(cols = starts_with(c("Norm", "Hypo")), names_to = "codes", values_to = "value")

expr.EGEOD18494.hif.pivot$codes <- factor(expr.EGEOD18494.hif.pivot$codes, levels = c("norm", "hypo.4h", "hypo.8h", "hypo.12h"))

expr.EGEOD18494.hif.pivot$time <- as.numeric(expr.EGEOD18494.hif.pivot$codes)

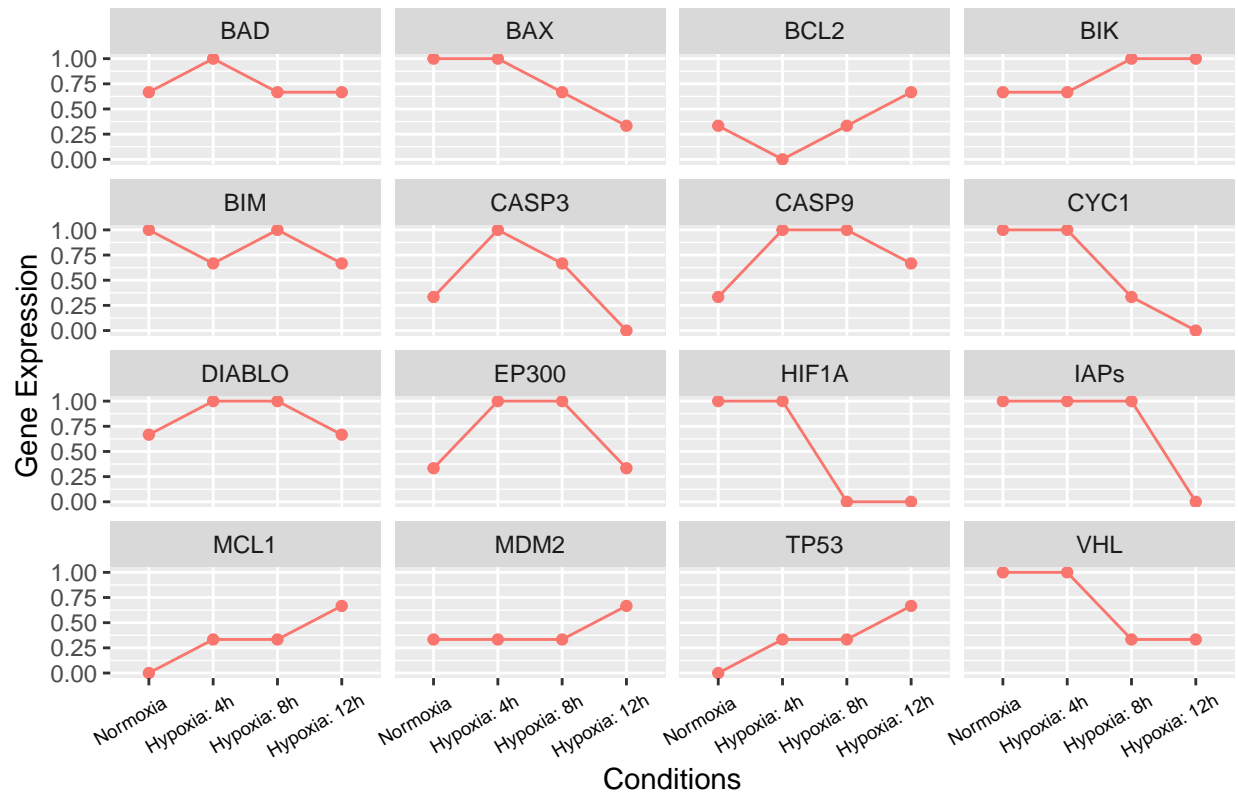
# hif.symbols <- c("HIF1A", "TP53", "MDM2", "VHL", "EP300", "TMBIM1", "TMBIM4", "TMBIM6", "BAD", "BIK", "MCL1", "FOXO3", "FOXO1", "FOXO2", "FOXO4", "FOXO6", "FOXO7", "FOXO8", "FOXO9", "FOXO10", "FOXO11", "FOXO12", "FOXO13", "FOXO14", "FOXO15", "FOXO16", "FOXO17", "FOXO18", "FOXO19", "FOXO20", "FOXO21", "FOXO22", "FOXO23", "FOXO24", "FOXO25", "FOXO26", "FOXO27", "FOXO28", "FOXO29", "FOXO30", "FOXO31", "FOXO32", "FOXO33", "FOXO34", "FOXO35", "FOXO36", "FOXO37", "FOXO38", "FOXO39", "FOXO40", "FOXO41", "FOXO42", "FOXO43", "FOXO44", "FOXO45", "FOXO46", "FOXO47", "FOXO48", "FOXO49", "FOXO50", "FOXO51", "FOXO52", "FOXO53", "FOXO54", "FOXO55", "FOXO56", "FOXO57", "FOXO58", "FOXO59", "FOXO60", "FOXO61", "FOXO62", "FOXO63", "FOXO64", "FOXO65", "FOXO66", "FOXO67", "FOXO68", "FOXO69", "FOXO70", "FOXO71", "FOXO72", "FOXO73", "FOXO74", "FOXO75", "FOXO76", "FOXO77", "FOXO78", "FOXO79", "FOXO80", "FOXO81", "FOXO82", "FOXO83", "FOXO84", "FOXO85", "FOXO86", "FOXO87", "FOXO88", "FOXO89", "FOXO90", "FOXO91", "FOXO92", "FOXO93", "FOXO94", "FOXO95", "FOXO96", "FOXO97", "FOXO98", "FOXO99", "FOXO100")

p.MDA <- ggplot(aes(x = factor(time), y = value, group = symbol, color="red"),
               data = expr.EGEOD18494.hif.pivot[expr.EGEOD18494.hif.pivot$symbol %in% c("HIF1A", "TP53", "MDM2", "VHL", "EP300", "TMBIM1", "TMBIM4", "TMBIM6", "BAD", "BIK", "MCL1", "FOXO3", "FOXO1", "FOXO2", "FOXO4", "FOXO6", "FOXO7", "FOXO8", "FOXO9", "FOXO10", "FOXO11", "FOXO12", "FOXO13", "FOXO14", "FOXO15", "FOXO16", "FOXO17", "FOXO18", "FOXO19", "FOXO20", "FOXO21", "FOXO22", "FOXO23", "FOXO24", "FOXO25", "FOXO26", "FOXO27", "FOXO28", "FOXO29", "FOXO30", "FOXO31", "FOXO32", "FOXO33", "FOXO34", "FOXO35", "FOXO36", "FOXO37", "FOXO38", "FOXO39", "FOXO40", "FOXO41", "FOXO42", "FOXO43", "FOXO44", "FOXO45", "FOXO46", "FOXO47", "FOXO48", "FOXO49", "FOXO50", "FOXO51", "FOXO52", "FOXO53", "FOXO54", "FOXO55", "FOXO56", "FOXO57", "FOXO58", "FOXO59", "FOXO60", "FOXO61", "FOXO62", "FOXO63", "FOXO64", "FOXO65", "FOXO66", "FOXO67", "FOXO68", "FOXO69", "FOXO70", "FOXO71", "FOXO72", "FOXO73", "FOXO74", "FOXO75", "FOXO76", "FOXO77", "FOXO78", "FOXO79", "FOXO80", "FOXO81", "FOXO82", "FOXO83", "FOXO84", "FOXO85", "FOXO86", "FOXO87", "FOXO88", "FOXO89", "FOXO90", "FOXO91", "FOXO92", "FOXO93", "FOXO94", "FOXO95", "FOXO96", "FOXO97", "FOXO98", "FOXO99", "FOXO100")],
               geom_point() +
               geom_line() +
               scale_x_discrete(breaks = c(1, 2, 3, 4),
                                labels = c("Normoxia", "Hypoxia: 4h", "Hypoxia: 8h", "Hypoxia: 12h")) +
               xlab("Conditions") + ylab("Gene Expression") +
               ggtitle("EGEOD18494 MDA-MB231 - Raw data") +
               theme(legend.position = "none", axis.text.x=element_text(color = "black", size=7, angle=30, vjust=.8),
                     #geom_line(aes(linetype=Symbol, color=Symbol)) +
                     facet_wrap(~ symbol))

p.MDA

```

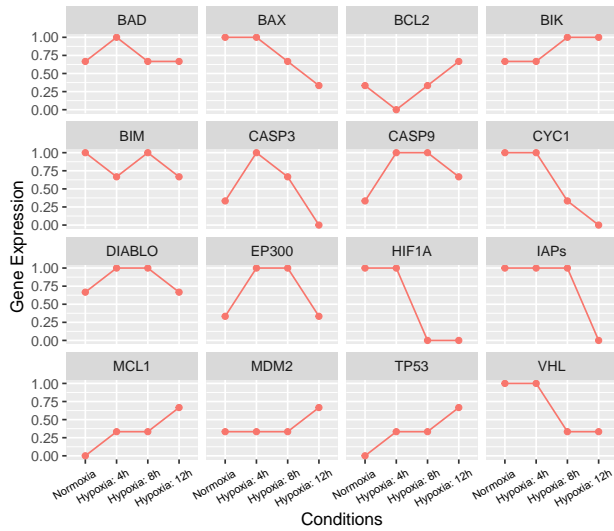
EGEOD18494 MDA-MB231 – Raw data



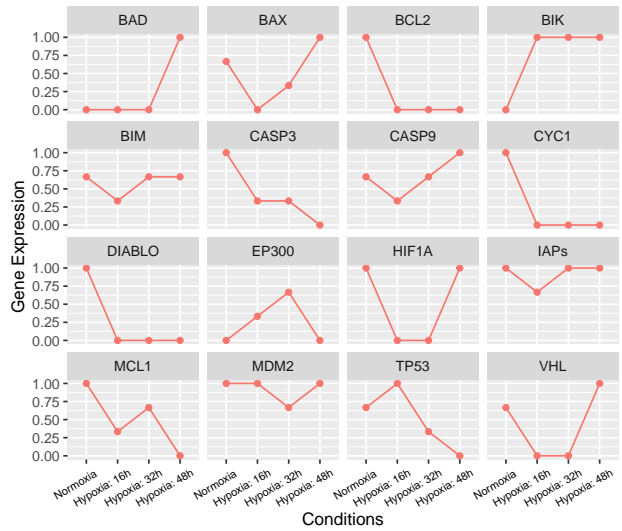
```
library(cowplot)
```

```
plot_grid(p.MDA, p.MCF7, labels = c('A', 'B'))
```

A EGEOD18494 MDA-MB231 – Raw data



B GSE47533 MCF7 – Raw data



Heatmaps - EGEOD18494

```
library("pheatmap")
library("ComplexHeatmap")

## =====
## ComplexHeatmap version 2.2.0
## Bioconductor page: http://bioconductor.org/packages/ComplexHeatmap/
## Github page: https://github.com/jokergoo/ComplexHeatmap
## Documentation: http://jokergoo.github.io/ComplexHeatmap-reference
##
## If you use it in published research, please cite:
## Gu, Z. Complex heatmaps reveal patterns and correlations in multidimensional
## genomic data. Bioinformatics 2016.
## =====

data.EGEOD18494$time <- factor(data.EGEOD18494$time, levels = c("control", "4h", "8h", "12h"))

row <- data.EGEOD18494$cell_line == "MDA-MB231 breast cancer"

annotation_for_heatmap <- droplevels(data.frame(time = data.EGEOD18494$time[row], condition = data.EGEOD18494$condition[row]))

row.names(annotation_for_heatmap) <- paste0(substr(data.EGEOD18494$condition[row], 1, 4), ".", data.EGEOD18494$time[row])

dists <- as.matrix(dist(t(expr.EGEOD18494.hif), method = "manhattan"))

rownames(dists) <- c(paste0(substr(data.EGEOD18494$condition[row], 1, 4), ".", data.EGEOD18494$time[row]),
  colnames(dists) <- c(paste0(substr(data.EGEOD18494$condition[row], 1, 4), ".", data.EGEOD18494$time[row]),

hmcol <- rev(colorRampPalette(RColorBrewer::brewer.pal(9, "YlOrRd"))(255))

diag(dists) <- NA

ann_colors <- list(
  time = RColorBrewer::brewer.pal(length(levels(data.EGEOD18494$time)), "Set2"),
  condition = c("red", "blue")
)

ann_colors

## $time
## [1] "#66C2A5" "#FC8D62" "#8DA0CB" "#E78AC3"
##
## $condition
## [1] "red" "blue"

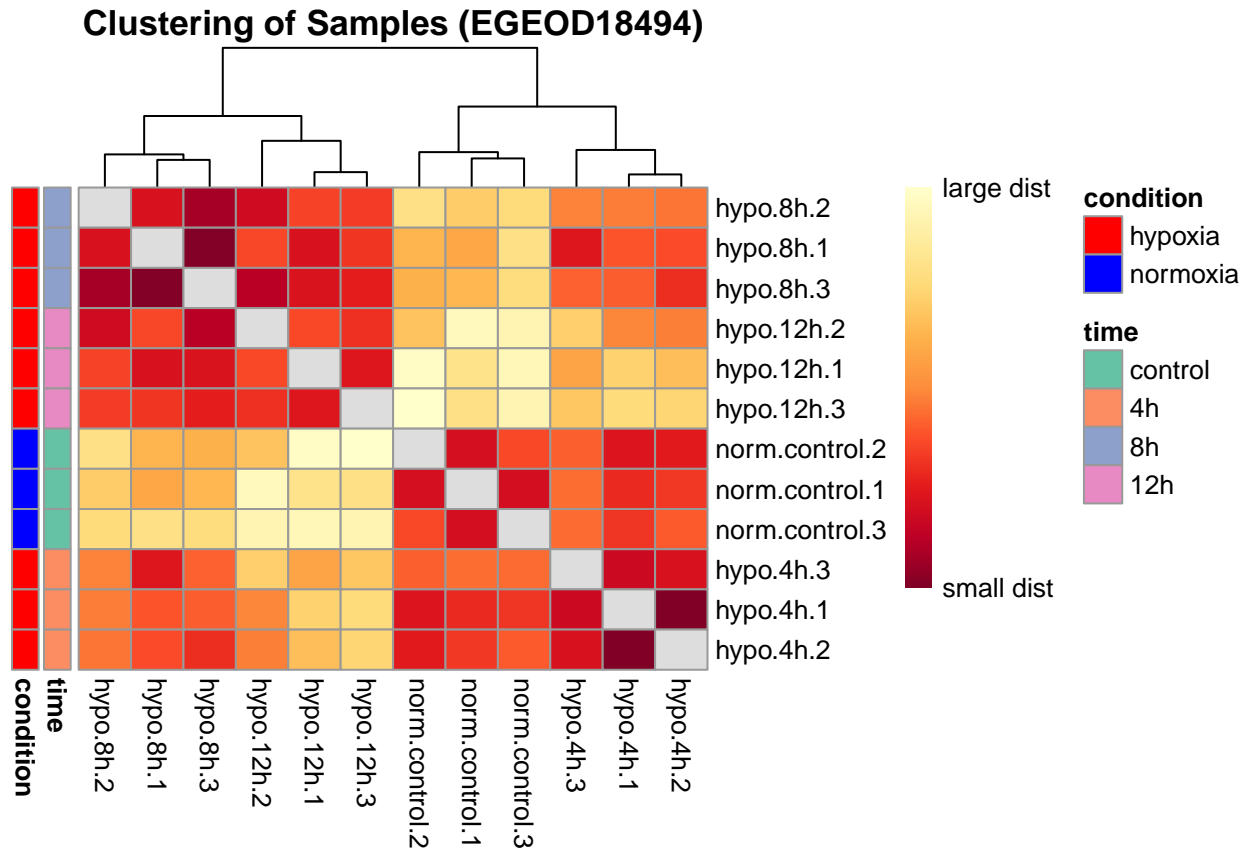
names(ann_colors$time) <- levels(data.EGEOD18494$time)
names(ann_colors$condition) <- levels(data.EGEOD18494$condition)

pheatmap(dists, col = (hmcol),
  annotation_row = annotation_for_heatmap,
```

```

annotation_colors = ann_colors,
legend = TRUE,
treeheight_row = 0,
legend_breaks = c(min(dists, na.rm = TRUE),
                    max(dists, na.rm = TRUE)),
legend_labels = (c("small dist", "large dist")),
main = "Clustering of Samples (EGEOD18494)"

```



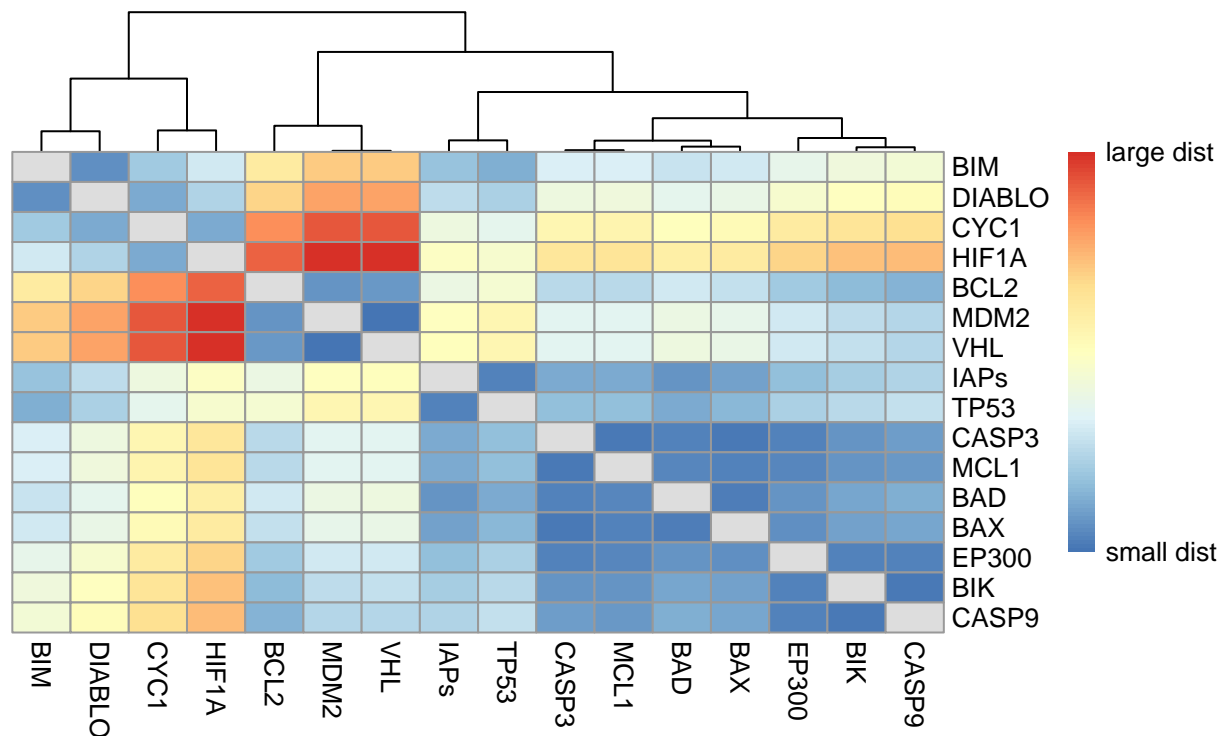
```

dists <- as.matrix(dist(expr.EGEOD18494.hif, method = "euclidean"))
rownames(dists) <- rownames(expr.EGEOD18494.hif)
colnames(dists) <- rownames(expr.EGEOD18494.hif)
diag(dists) <- NA

pheatmap(dists, #row = (hmc),
          #annotation_col = annotation_for_heatmap,
          #annotation_colors = ann_colors,
          legend = TRUE,
          treeheight_row = 0,
          legend_breaks = c(min(dists, na.rm = TRUE),
                              max(dists, na.rm = TRUE)),
          legend_labels = (c("small dist", "large dist")),
          main = "Clustering of Gene Expression \n Euclidian Distance (EGEOD18494)")

```

Clustering of Gene Expression Euclidian Distance (EGEOD18494)



```
#-----

expr.row <- (colnames(expr.EGEOD18494.hif) %in% data.EGEOD18494$codes[data.EGEOD18494$cell_line == "MDA-MB231 breast cancer"])
dists <- as.matrix(dist(expr.EGEOD18494.hif[expr.row], method = "euclidean"))
rownames(dists) <- rownames(expr.EGEOD18494.hif[expr.row])
colnames(dists) <- rownames(expr.EGEOD18494.hif[expr.row])
diag(dists) <- NA

p1 <- pheatmap(dists,
  legend = TRUE,
  treeheight_row = 0,
  legend_breaks = c(min(dists, na.rm = TRUE),
    max(dists, na.rm = TRUE)),
  legend_labels = (c("small dist", "large dist")),
  main = "Clustering of Gene Expression on Hypoxia \n Euclidian Distance (EGEOD18494)",
  silent=T)

#-----

row <- data.EGEOD18494$cell_line == "MDA-MB231 breast cancer" & data.EGEOD18494$condition == "normoxia"
annotation_for_heatmap <- droplevels(data.frame(time = data.EGEOD18494$time[row], condition = data.EGEOD18494$condition[row]))
expr.row <- (colnames(expr.EGEOD18494.hif) %in% data.EGEOD18494$codes[data.EGEOD18494$cell_line == "MDA-MB231 breast cancer"])
row.names(annotation_for_heatmap) <- colnames(expr.EGEOD18494.hif[expr.row])
```



```

dists <- as.matrix(dist(expr.EGEOD18494.hif[expr.row], method = "euclidean"))

rownames(dists) <- rownames(expr.EGEOD18494.hif[expr.row])

hmcol <- rev(colorRampPalette(RColorBrewer::brewer.pal(9, "YlOrRd"))(255))
colnames(dists) <- rownames(expr.EGEOD18494.hif[expr.row])
diag(dists) <- NA

ann_colors <- list(
  time = RColorBrewer::brewer.pal(length(levels(data.EGEOD18494$time)), "Set2"),
  condition = c("#EF8A62", "#67A9CF")
)

ann_colors

## $time
## [1] "#66C2A5" "#FC8D62" "#8DA0CB" "#E78AC3"
##
## $condition
## [1] "#EF8A62" "#67A9CF"

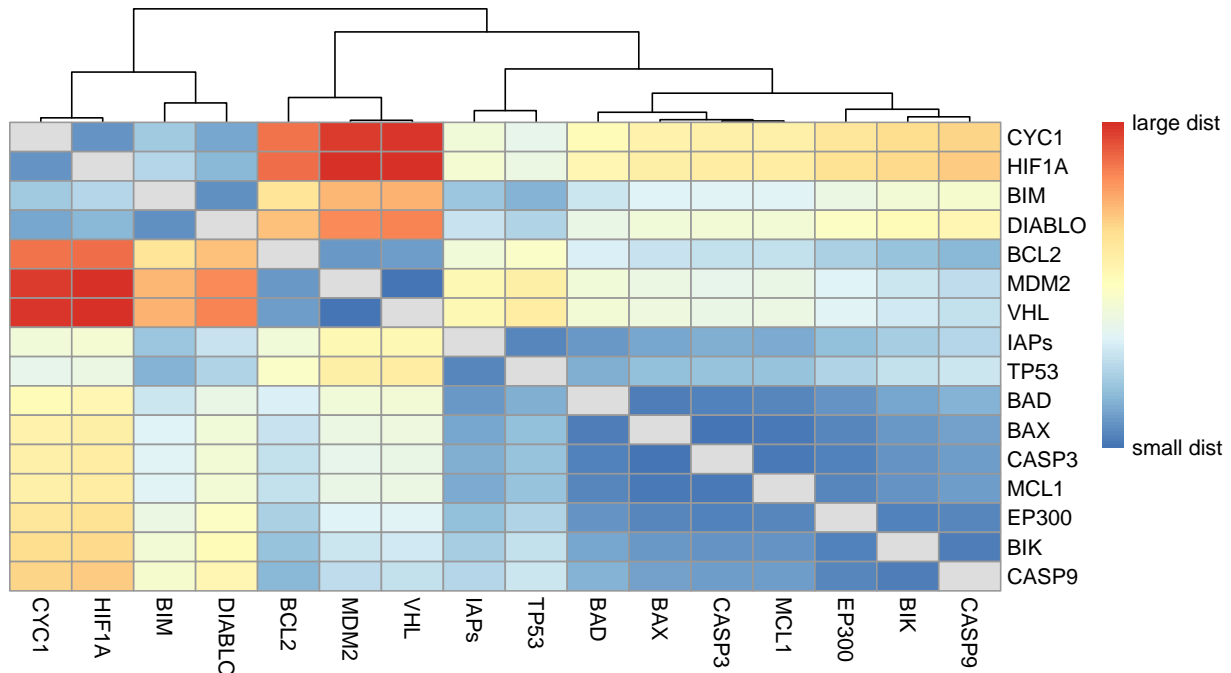
names(ann_colors$time) <- levels(data.EGEOD18494$time)
names(ann_colors$condition) <- levels(data.EGEOD18494$condition)

p2 <- pheatmap(dists, #row = (hmcol),
  #annotation_col = annotation_for_heatmap,
  #annotation_colors = ann_colors,
  legend = TRUE,
  treeheight_row = 0,
  legend_breaks = c(min(dists, na.rm = TRUE),
    max(dists, na.rm = TRUE)),
  legend_labels = (c("small dist", "large dist")),
  main = "Clustering of Gene Expression on Normoxia \n Euclidian Distance (EGEOD18494)",
  silent=T)

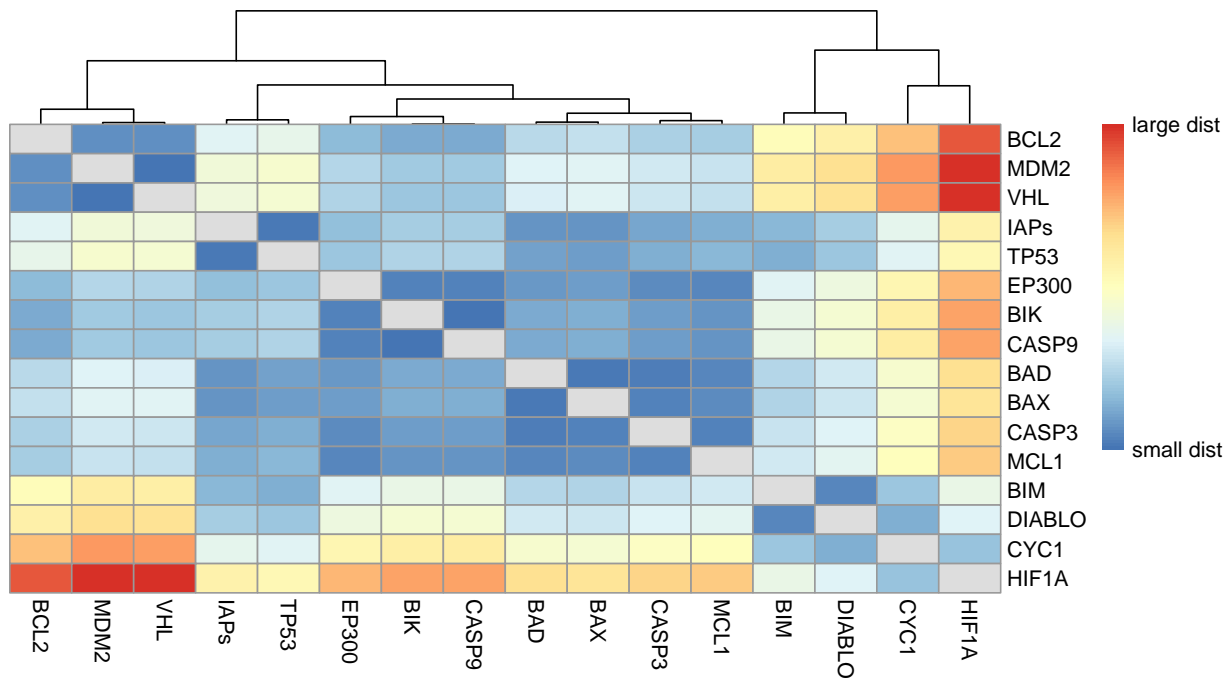
gridExtra::grid.arrange(grobs=list(p1$gtable, p2$gtable),
  nrow = 2 , labels=c('A', 'B'))

```

Clustering of Gene Expression on Hypoxia Euclidian Distance (EGEOD18494)



Clustering of Gene Expression on Normoxia Euclidian Distance (EGEOD18494)



```
data.EGEOD18494$time <- factor(data.EGEOD18494$time, levels = c("control", "4h", "8h", "12h 4h"))

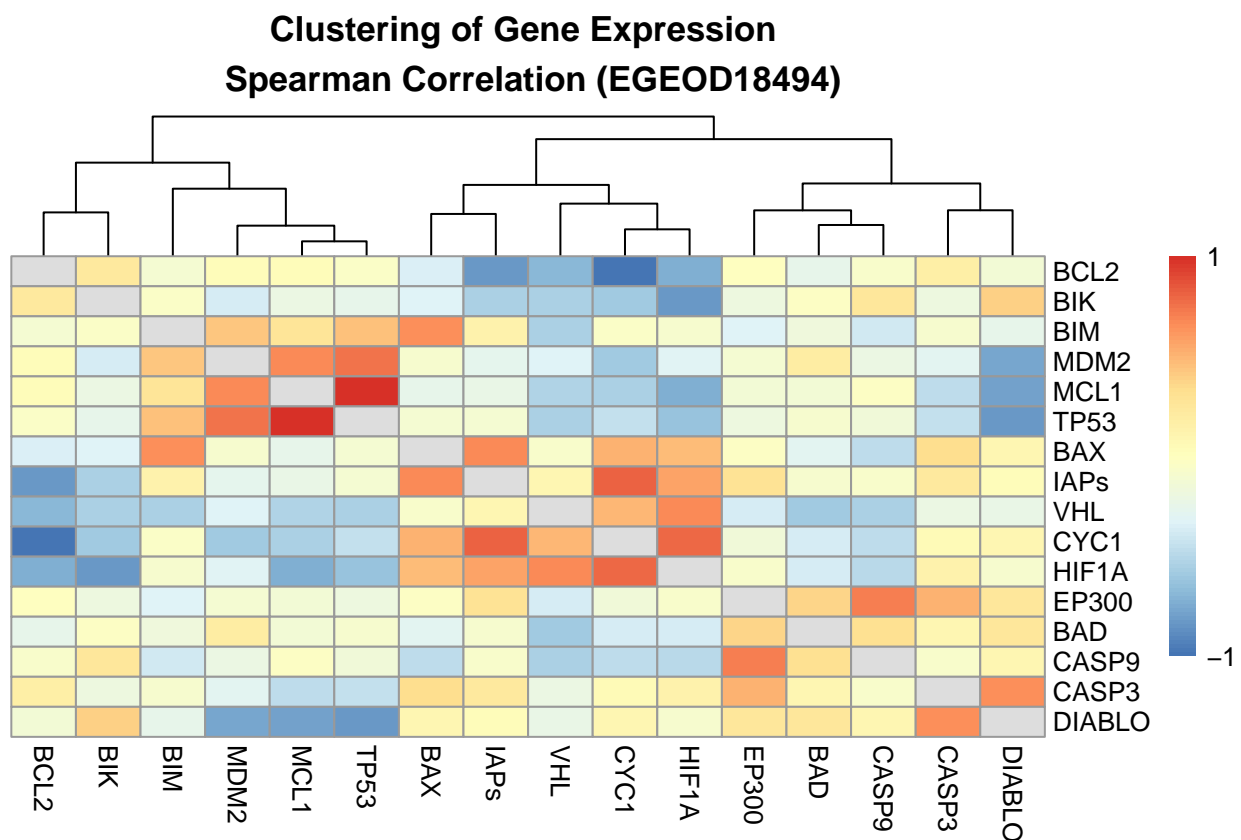
dists <- cor(t(expr.EGEOD18494.hif), use = "pairwise.complete.obs", method = "spearman")
rownames(dists) <- rownames(expr.EGEOD18494.hif)
```

```

hmccl <- rev(colorRampPalette(RColorBrewer::brewer.pal(9, "YlOrRd"))(255))
colnames(dists) <- rownames(expr.EGEOD18494.hif)
diag(dists) <- NA

pheatmap(dists, #row = (hmccl),
          #annotation_col = annotation_for_heatmap,
          #annotation_colors = ann_colors,
          legend = TRUE,
          treeheight_row = 0,
          legend_breaks = c(min(dists, na.rm = TRUE),
                             max(dists, na.rm = TRUE)),
          legend_labels = (c("-1", "1")),
          main = "Clustering of Gene Expression \n Spearman Correlation (EGEOD18494)")

```



Heatmaps - GSE47533

```

library("pheatmap")
library("ComplexHeatmap")

data.GSE47533$time <- factor(data.GSE47533$time, levels = c("0", "16h", "32h", "48h"))

annotation_for_heatmap <- droplevels(data.frame(time = data.GSE47533$time, condition = data.GSE47533$condition))

```

```

row.names(annotation_for_heatmap) <- paste0(substr(data.GSE47533$condition,1,4),".", data.GSE47533$time

dists <- as.matrix(dist(t(expr.GSE47533.hif), method = "manhattan"))

rownames(dists) <- c(paste0(substr(data.GSE47533$condition,1,4),".", data.GSE47533$time, ".", data.GSE4
colnames(dists) <- c(paste0(substr(data.GSE47533$condition,1,4),".", data.GSE47533$time, ".", data.GSE4

hmcol <- rev(colorRampPalette(RColorBrewer::brewer.pal(9, "YlOrRd"))(255))

diag(dists) <- NA

ann_colors <- list(
  time = RColorBrewer::brewer.pal(length(levels(data.GSE47533$time)), "Set2"),
  condition = c("red", "blue")
)

ann_colors

```

```

## $time
## [1] "#66C2A5" "#FC8D62" "#8DA0CB" "#E78AC3"
##
## $condition
## [1] "red" "blue"

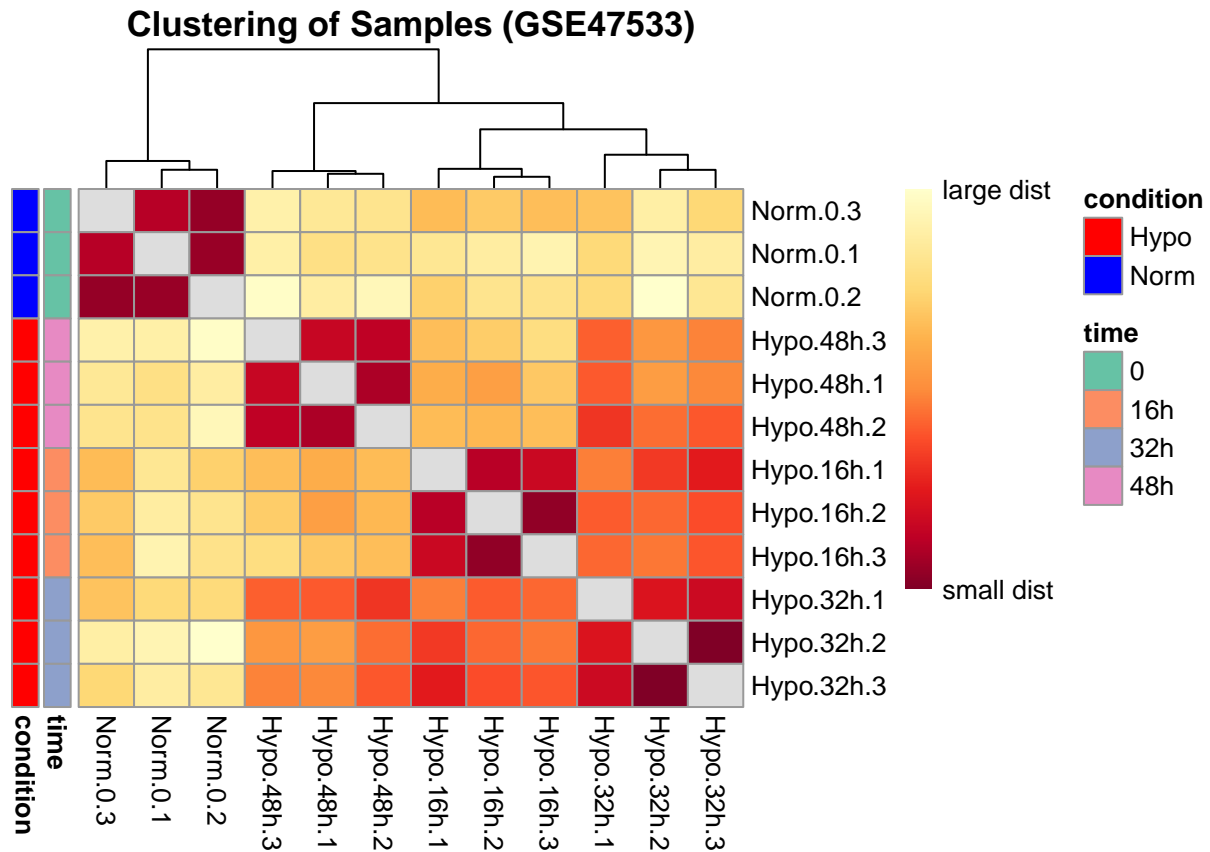
```

```

names(ann_colors$time) <- levels(data.GSE47533$time)
names(ann_colors$condition) <- levels(data.GSE47533$condition)

pheatmap(dists, col = (hmcol),
  annotation_row = annotation_for_heatmap,
  annotation_colors = ann_colors,
  legend = TRUE,
  treeheight_row = 0,
  legend_breaks = c(min(dists, na.rm = TRUE),
                     max(dists, na.rm = TRUE)),
  legend_labels = (c("small dist", "large dist")),
  main = "Clustering of Samples (GSE47533)")

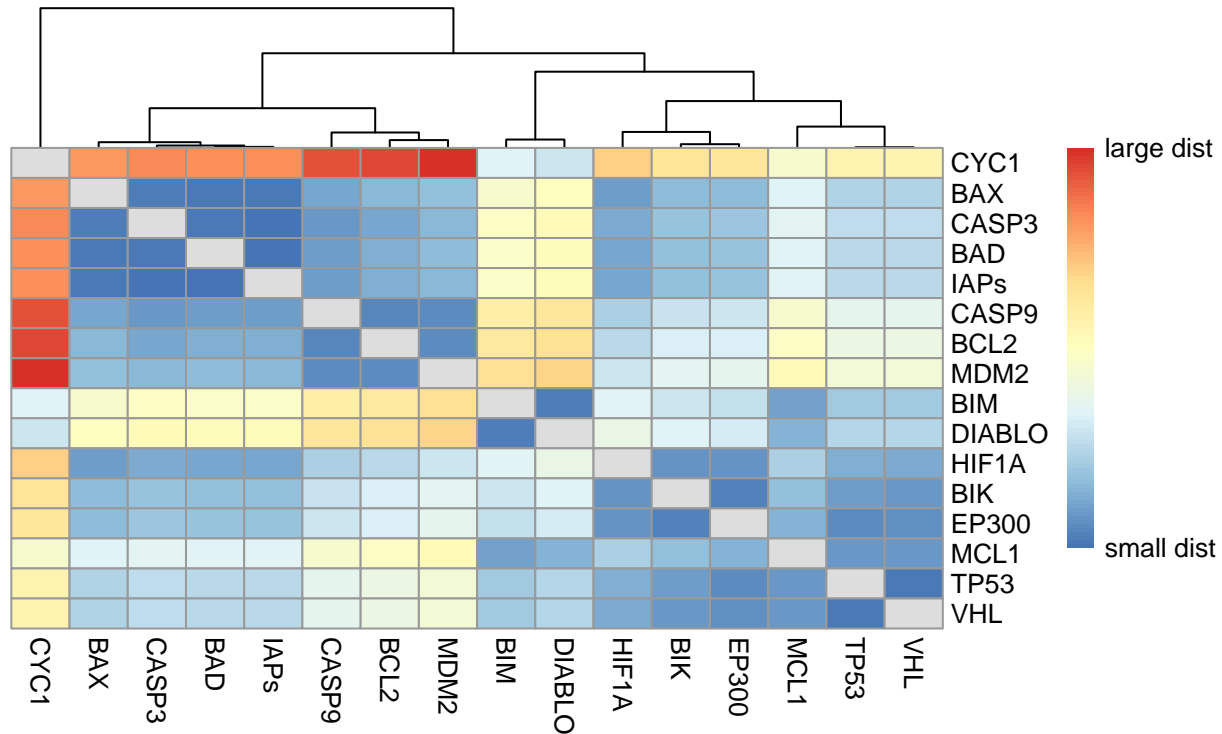
```



```
dists <- as.matrix(dist(expr.GSE47533.hif, method = "euclidean"))
rownames(dists) <- rownames(expr.GSE47533.hif)
colnames(dists) <- rownames(expr.GSE47533.hif)
diag(dists) <- NA

pheatmap(dists,
  legend = TRUE,
  treeheight_row = 0,
  legend_breaks = c(min(dists, na.rm = TRUE),
                     max(dists, na.rm = TRUE)),
  legend_labels = (c("small dist", "large dist")),
  main = "Clustering of Gene Expression \n Euclidian Distance (GSE47533)")
```

Clustering of Gene Expression Euclidian Distance (GSE47533)



```
#-----
expr.row <- (colnames(expr.GSE47533.hif) %in% data.GSE47533$codes[data.GSE47533$condition == "Hypo"])
dists <- as.matrix(dist(expr.GSE47533.hif[expr.row], method = "euclidean"))
rownames(dists) <- rownames(expr.GSE47533.hif[expr.row])
colnames(dists) <- rownames(expr.GSE47533.hif[expr.row])
diag(dists) <- NA

p1 <- pheatmap(dists,
  legend = TRUE,
  treeheight_row = 0,
  legend_breaks = c(min(dists, na.rm = TRUE),
    max(dists, na.rm = TRUE)),
  legend_labels = (c("small dist", "large dist")),
  main = "Clustering of Gene Expression on Hypoxia \n Euclidian Distance (GSE47533)",
  silent=T)

#-----

expr.row <- (colnames(expr.GSE47533.hif) %in% data.GSE47533$codes[data.GSE47533$condition == "Norm"])
dists <- as.matrix(dist(expr.GSE47533.hif[expr.row], method = "euclidean"))
rownames(dists) <- rownames(expr.GSE47533.hif[expr.row])
colnames(dists) <- rownames(expr.GSE47533.hif[expr.row])
diag(dists) <- NA
```

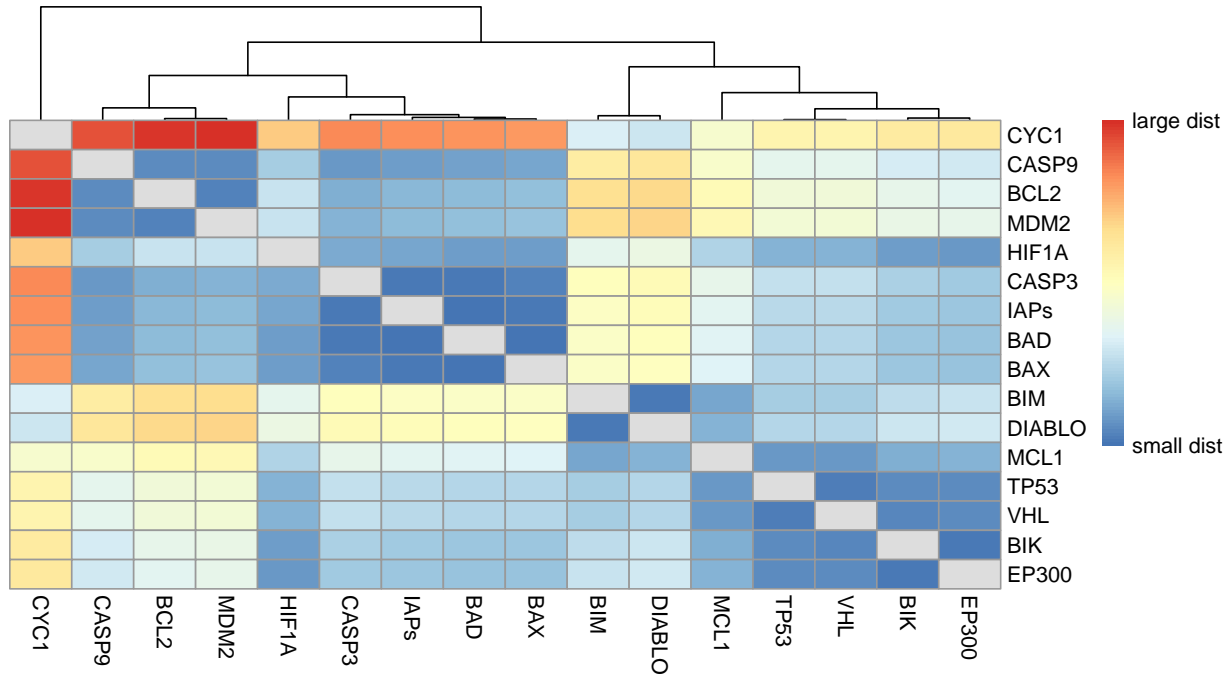
```

p2 <- pheatmap(dists,
  legend = TRUE,
  treeheight_row = 0,
  legend_breaks = c(min(dists, na.rm = TRUE),
    max(dists, na.rm = TRUE)),
  legend_labels = (c("small dist", "large dist")),
  main = "Clustering of Gene Expression on Normoxia \n Euclidian Distance (GSE47533)",
  silent=T)

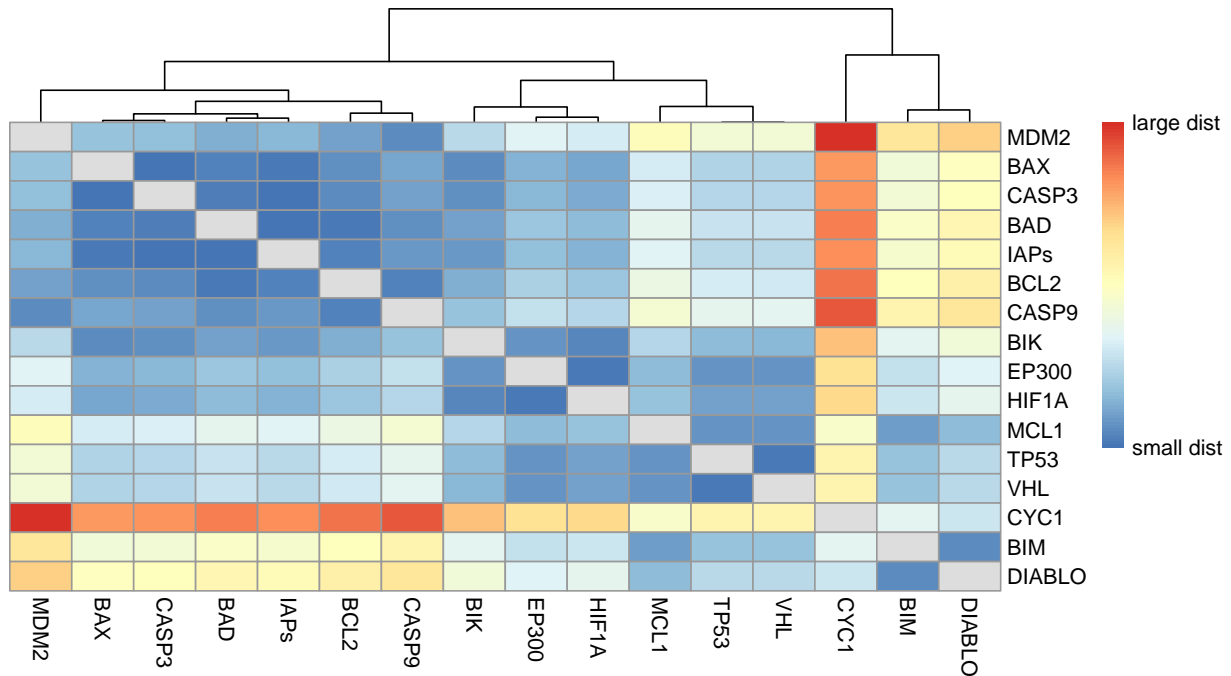
gridExtra::grid.arrange(grobs=list(p1$gtable, p2$gtable),
  nrow = 2 , labels=c('A', 'B'))

```

Clustering of Gene Expression on Hypoxia Euclidian Distance (GSE47533)



Clustering of Gene Expression on Normoxia Euclidian Distance (GSE47533)



```
dists <- cor(t(expr.GSE47533.hif), use = "pairwise.complete.obs", method = "spearman")
rownames(dists) <- rownames(expr.GSE47533.hif)
colnames(dists) <- rownames(expr.GSE47533.hif)
diag(dists) <- NA
```



```
pheatmap(dists,
  legend = TRUE,
  treeheight_row = 0,
  legend_breaks = c(min(dists, na.rm = TRUE),
    max(dists, na.rm = TRUE)),
  legend_labels = (c("-1", "1")),
  main = "Clustering of Gene Expression \n Spearman Correlation (GSE47533)")
```



Heatmaps - All datasets Breast Cell-lines (E-GEOD-18494, GSE47533, and GSE41491)

- E-GEOD-18494 2012 / MDA-MB231 / breast / 4h, 8h, 12h / microarray
- GSE41491 2012 / MCF7 / breast / 1h, 2h, 4h, 8h, 12h, 16h, 24h / microarray
- GSE47534 2014 / MCF7 / breast / normoxia, 16h, 32h, 48h / mRNA

```
# Imput the mean of all VHL values
mean.vhl <- mean(unlist(expr.GSE47533.hif["VHL",], expr.EGEOD18494.hif["VHL",]))
expr.GSE41491.hif["VHL",] <- rep(mean.vhl, 24)

expr.all.hif <- cbind(expr.GSE47533.hif, expr.EGEOD18494.hif, expr.GSE41491.hif)
```

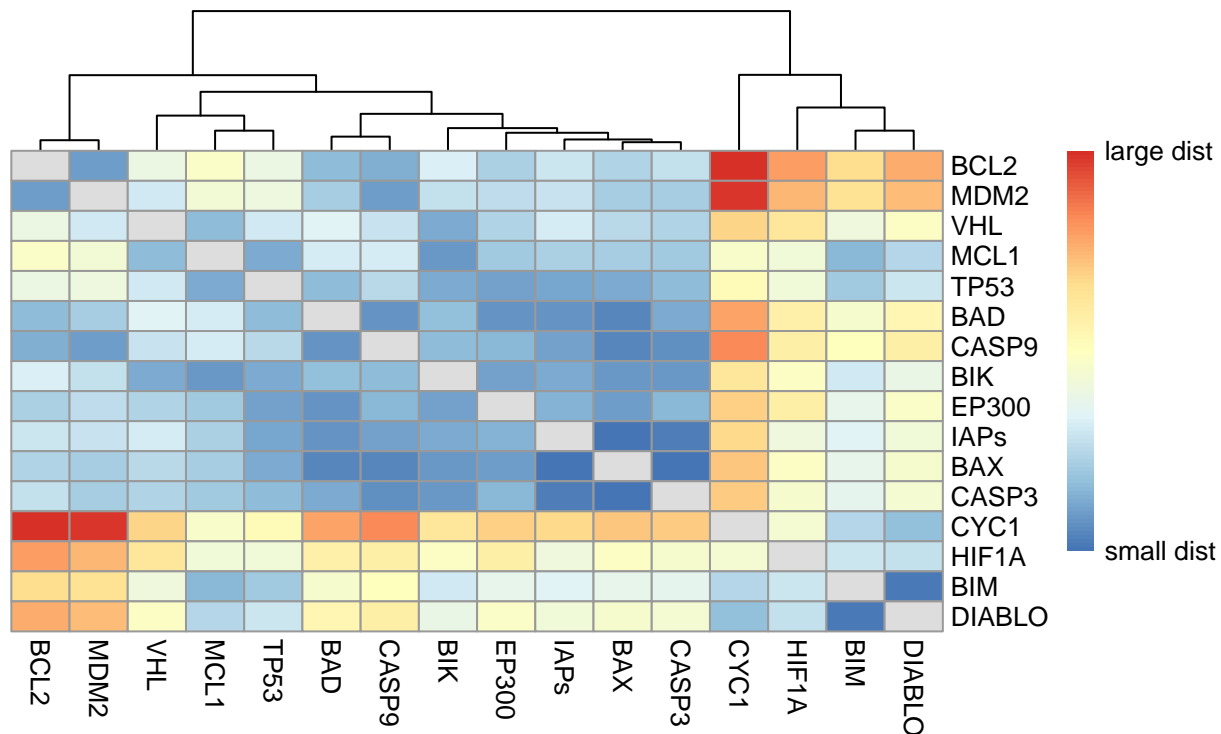
```
col_brca <- union(data.GSE47533$codes[data.GSE47533$cell_line == "MCF7"],
  union(data.EGEOD18494$codes[data.EGEOD18494$cell_line == "MDA-MB231 breast cancer"],
    data.GSE41491$codes[data.GSE41491$cell_line == "MCF7"]))

expr.all.hif <- expr.all.hif[, (colnames(expr.all.hif) %in% col_brca)]

dists <- as.matrix(dist(expr.all.hif, method = "euclidean"))
rownames(dists) <- rownames(expr.all.hif)
colnames(dists) <- rownames(expr.all.hif)
diag(dists) <- NA

pheatmap(dists,
  legend = TRUE,
  treeheight_row = 0,
  legend_breaks = c(min(dists, na.rm = TRUE),
    max(dists, na.rm = TRUE)),
  legend_labels = (c("small dist", "large dist")),
  main = "Clustering of Gene Expression \n Euclidian Distance (All 3 datasets, breast cell-lines)"))
```

Clustering of Gene Expression Euclidian Distance (All 3 datasets, breast cell-lines)



```
#-----

col_hypo <- union(data.GSE47533$codes[data.GSE47533$condition == "Hypo"],
  union(data.EGEOD18494$codes[data.EGEOD18494$condition == "hypoxia"],
    data.GSE41491$codes[data.GSE41491$condition == "hy"]))
```

```

expr.row <- (colnames(expr.all.hif) %in% col_hypo)
dists <- as.matrix(dist(expr.all.hif[expr.row], method = "euclidean"))
rownames(dists) <- rownames(expr.all.hif[expr.row])
colnames(dists) <- rownames(expr.all.hif[expr.row])
diag(dists) <- NA

p1 <- pheatmap(dists,
  legend = TRUE,
  treeheight_row = 0,
  legend_breaks = c(min(dists, na.rm = TRUE),
                     max(dists, na.rm = TRUE)),
  legend_labels = (c("small dist", "large dist")),
  main = "Clustering of Gene Expression on Hypoxia \n Euclidian Distance (All 3 datasets, break
  silent=T)

#-----

col_norm <- union(data.GSE47533$codes[data.GSE47533$condition == "Norm"],
  union(data.EGEOD18494$codes[data.EGEOD18494$condition == "normoxia"],
        data.GSE41491$codes[data.GSE41491$condition == "no"]))

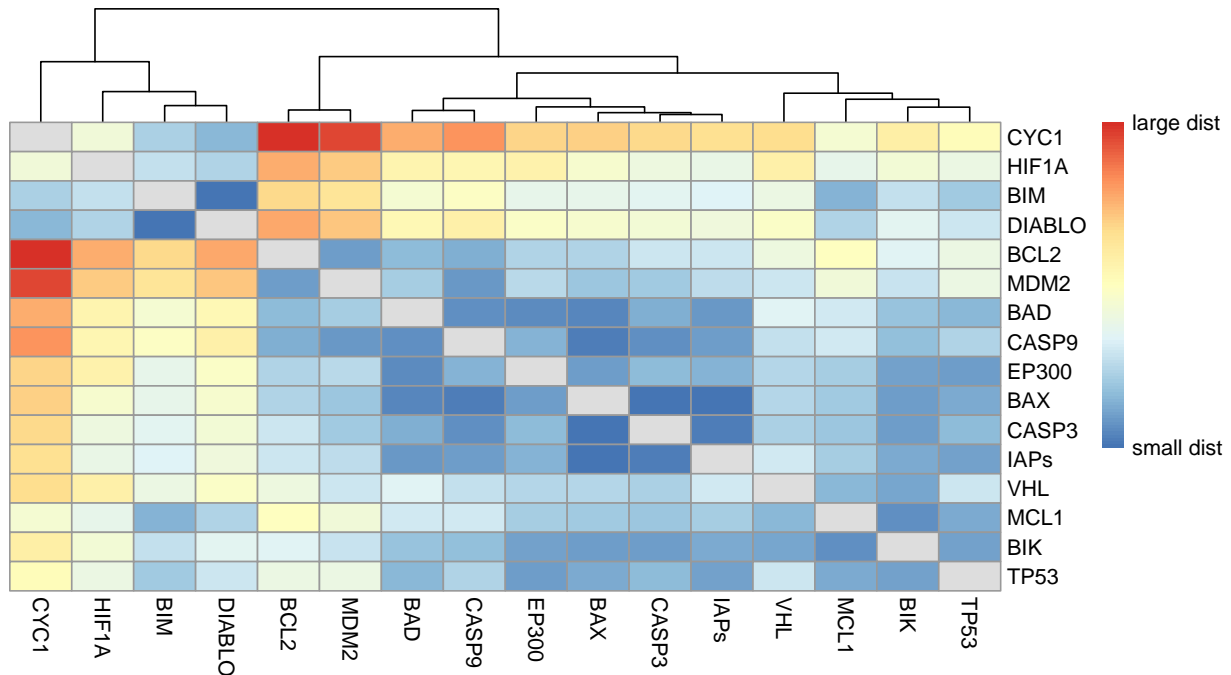
expr.row <- (colnames(expr.all.hif) %in% col_norm)
dists <- as.matrix(dist(expr.all.hif[expr.row], method = "euclidean"))
rownames(dists) <- rownames(expr.all.hif[expr.row])
colnames(dists) <- rownames(expr.all.hif[expr.row])
diag(dists) <- NA

p2 <- pheatmap(dists,
  legend = TRUE,
  treeheight_row = 0,
  legend_breaks = c(min(dists, na.rm = TRUE),
                     max(dists, na.rm = TRUE)),
  legend_labels = (c("small dist", "large dist")),
  main = "Clustering of Gene Expression on Normoxia \n Euclidian Distance (All 3 datasets, break
  silent=T)

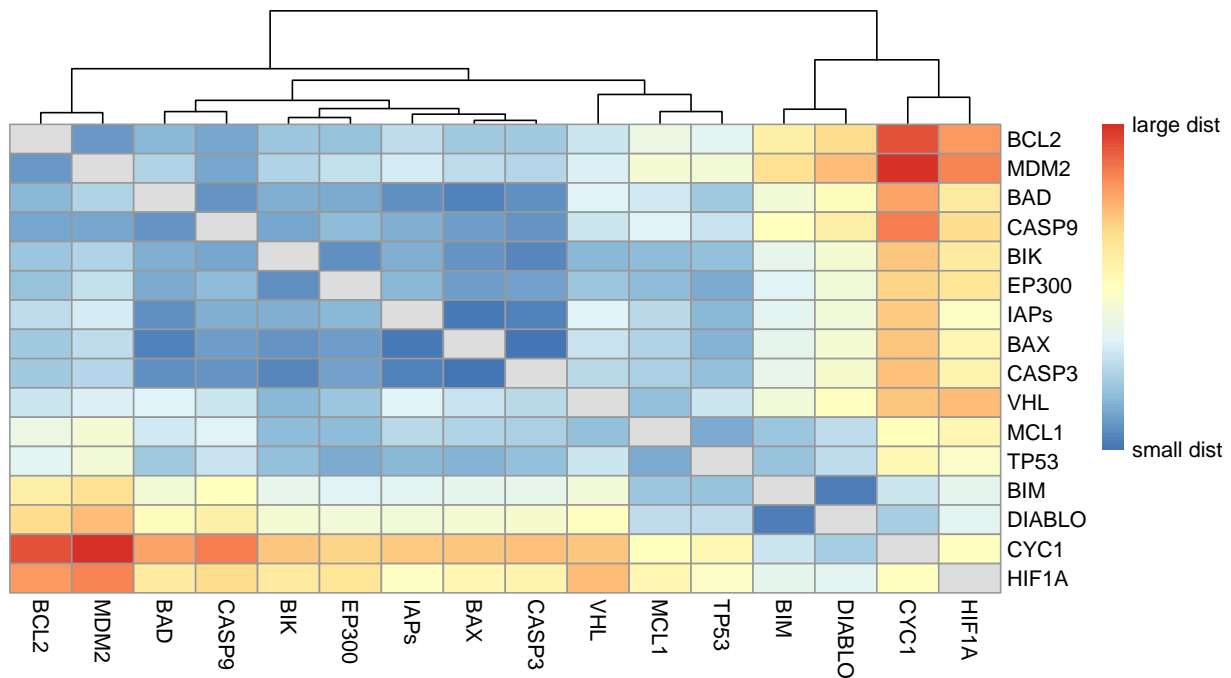
gridExtra::grid.arrange(grobs=list(p1$gtable, p2$gtable),
  nrow = 2 , labels=c('A', 'B'))

```

Clustering of Gene Expression on Hypoxia Euclidian Distance (All 3 datasets, breast cell-lines)

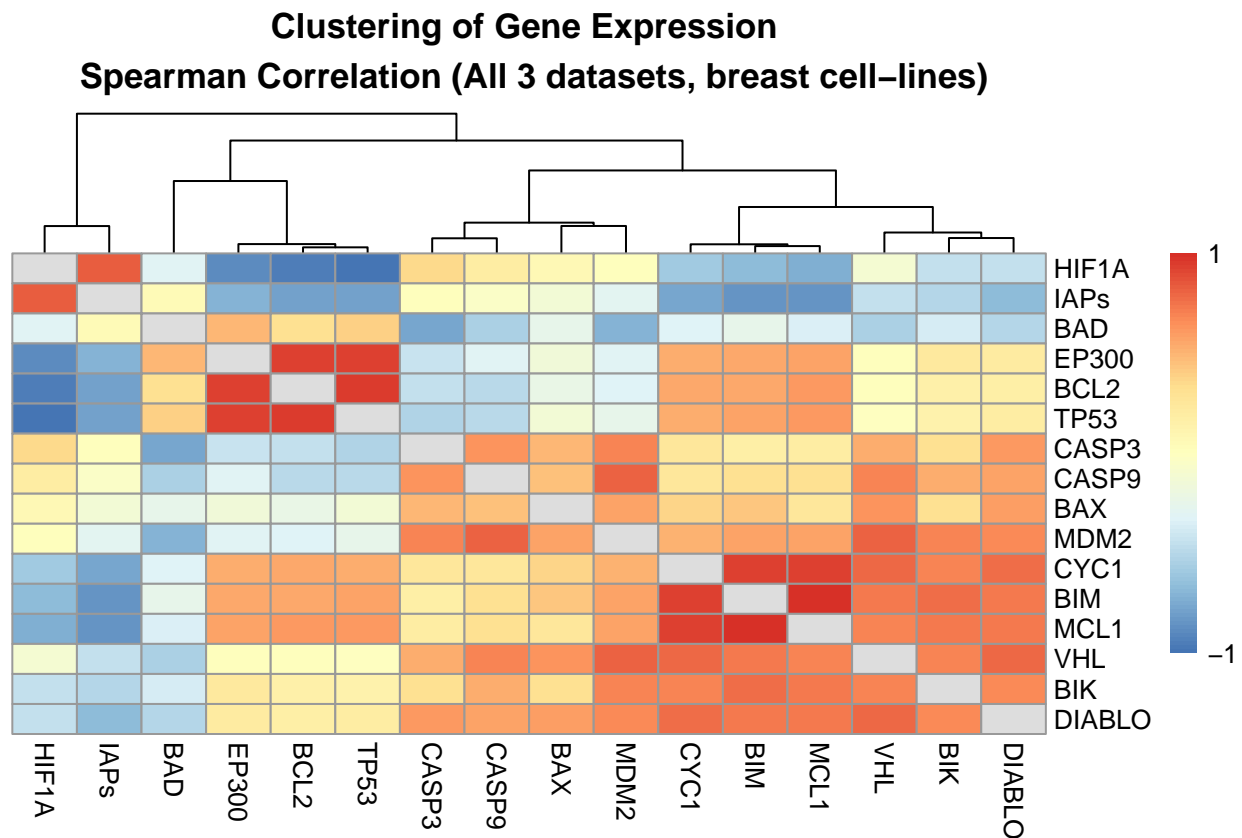


Clustering of Gene Expression on Normoxia Euclidian Distance (All 3 datasets, breast cell-lines)



```
dists <- cor(t(expr.all.hif), use = "pairwise.complete.obs", method = "spearman")
rownames(dists) <- rownames(expr.all.hif)
colnames(dists) <- rownames(expr.all.hif)
diag(dists) <- NA
```

```
pheatmap(dists,
  legend = TRUE,
  treeheight_row = 0,
  legend_breaks = c(min(dists, na.rm = TRUE),
    max(dists, na.rm = TRUE)),
  legend_labels = (c("-1", "1")),
  main = "Clustering of Gene Expression \n Spearman Correlation (All 3 datasets, breast cell-lines)
```



Heatmaps - All datasets All Cell-lines (E-GEOD-18494, GSE47533, and GSE41491)

- E-GEOD-18494 2012 / HepG2, U87, MDA-MB231 / hepatoma, glioma, breast / 4h, 8h, 12h / microarray
- GSE41491 2012 / DU145, HT29, MCF7 / prostate, colon, breast / 1h, 2h, 4h, 8h, 12h, 16h, 24h / microarray
- GSE47534 2014 / MCF7 / breast / normoxia, 16h, 32h, 48h / mRNA

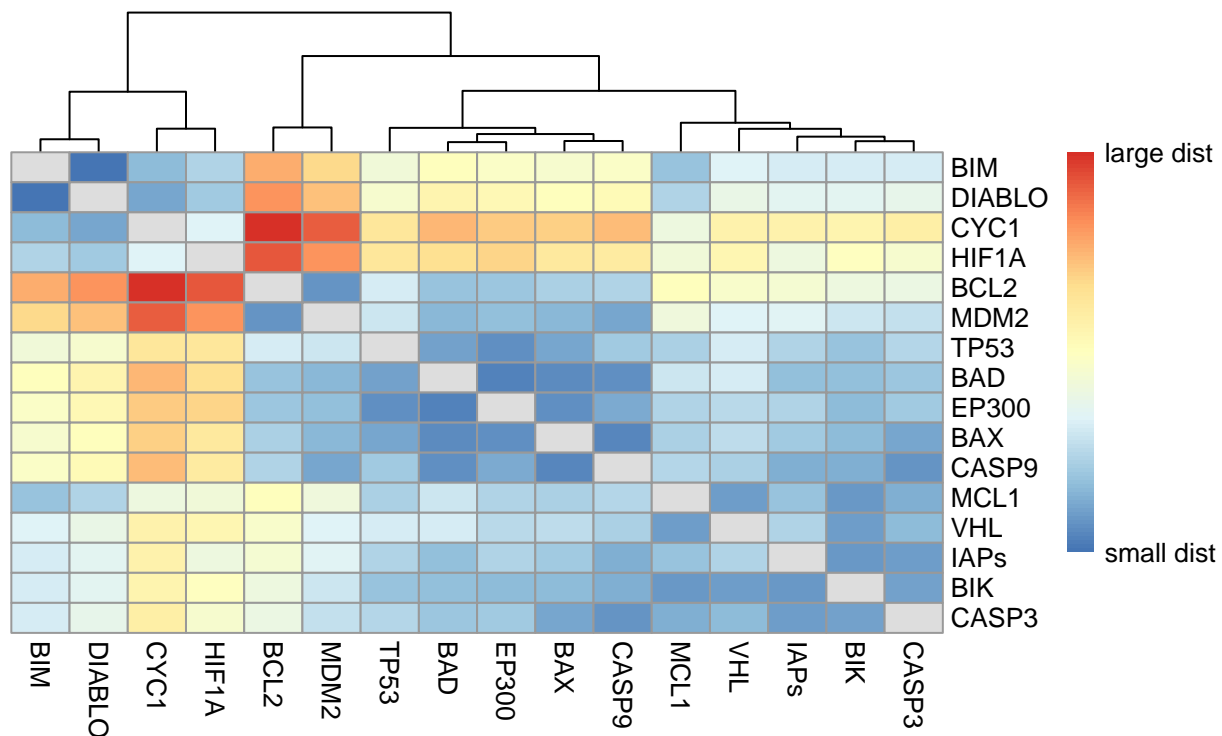
```
# Imput the mean of all VHL values
mean.vhl <- mean(unlist(expr.GSE47533.hif["VHL",], expr.EGEOD18494.hif["VHL",]))
expr.GSE41491.hif["VHL",] <- rep(mean.vhl, 24)
```

```
expr.all.hif <- cbind(expr.GSE47533.hif, expr.EGEO18494.hif, expr.GSE41491.hif)
```

```
dists <- as.matrix(dist(expr.all.hif, method = "euclidean"))
rownames(dists) <- rownames(expr.all.hif)
colnames(dists) <- rownames(expr.all.hif)
diag(dists) <- NA
```

```
pheatmap(dists,
  legend = TRUE,
  treeheight_row = 0,
  legend_breaks = c(min(dists, na.rm = TRUE),
                    max(dists, na.rm = TRUE)),
  legend_labels = (c("small dist", "large dist")),
  main = "Clustering of Gene Expression \n Euclidian Distance (All 3 datasets, all cell-lines)"))
```

Clustering of Gene Expression Euclidian Distance (All 3 datasets, all cell-lines)



```
#-----
col_hypo <- union(data.GSE47533$codes[data.GSE47533$condition == "Hypo"],
  union(data.EGEO18494$codes[data.EGEO18494$condition == "hypoxia"],
    data.GSE41491$codes[data.GSE41491$condition == "hy"]))

expr.row <- (colnames(expr.all.hif) %in% col_hypo)
dists <- as.matrix(dist(expr.all.hif[expr.row], method = "euclidean"))
rownames(dists) <- rownames(expr.all.hif[expr.row])
```

```

colnames(dists) <- rownames(expr.all.hif[expr.row])
diag(dists) <- NA

p1 <- pheatmap(dists,
  legend = TRUE,
  treeheight_row = 0,
  legend_breaks = c(min(dists, na.rm = TRUE),
                     max(dists, na.rm = TRUE)),
  legend_labels = (c("small dist", "large dist")),
  main = "Clustering of Gene Expression on Hypoxia \n Euclidian Distance (All 3 datasets, all c
  silent=T)

#-----

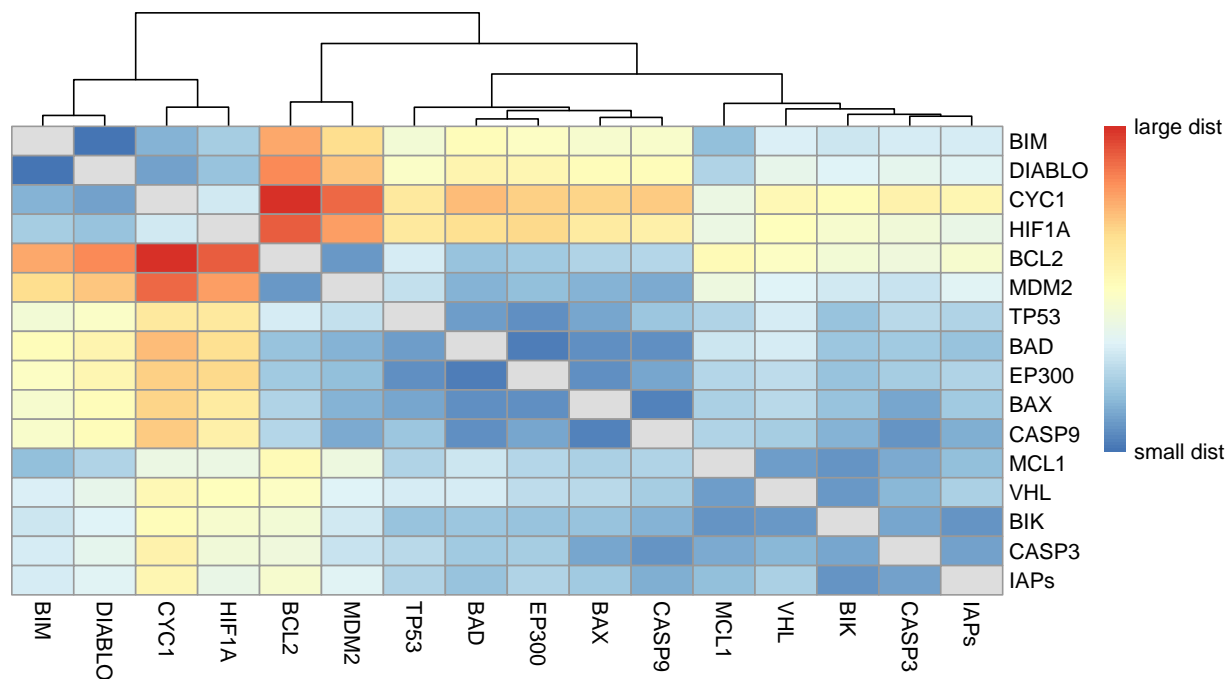
expr.row <- (colnames(expr.all.hif) %in% data.GSE47533$codes[data.GSE47533$condition == "Norm"])
dists <- as.matrix(dist(expr.all.hif[expr.row], method = "euclidean"))
rownames(dists) <- rownames(expr.all.hif[expr.row])
colnames(dists) <- rownames(expr.all.hif[expr.row])
diag(dists) <- NA

p2 <- pheatmap(dists,
  legend = TRUE,
  treeheight_row = 0,
  legend_breaks = c(min(dists, na.rm = TRUE),
                     max(dists, na.rm = TRUE)),
  legend_labels = (c("small dist", "large dist")),
  main = "Clustering of Gene Expression on Normoxia \n Euclidian Distance (All 3 datasets, all c
  silent=T)

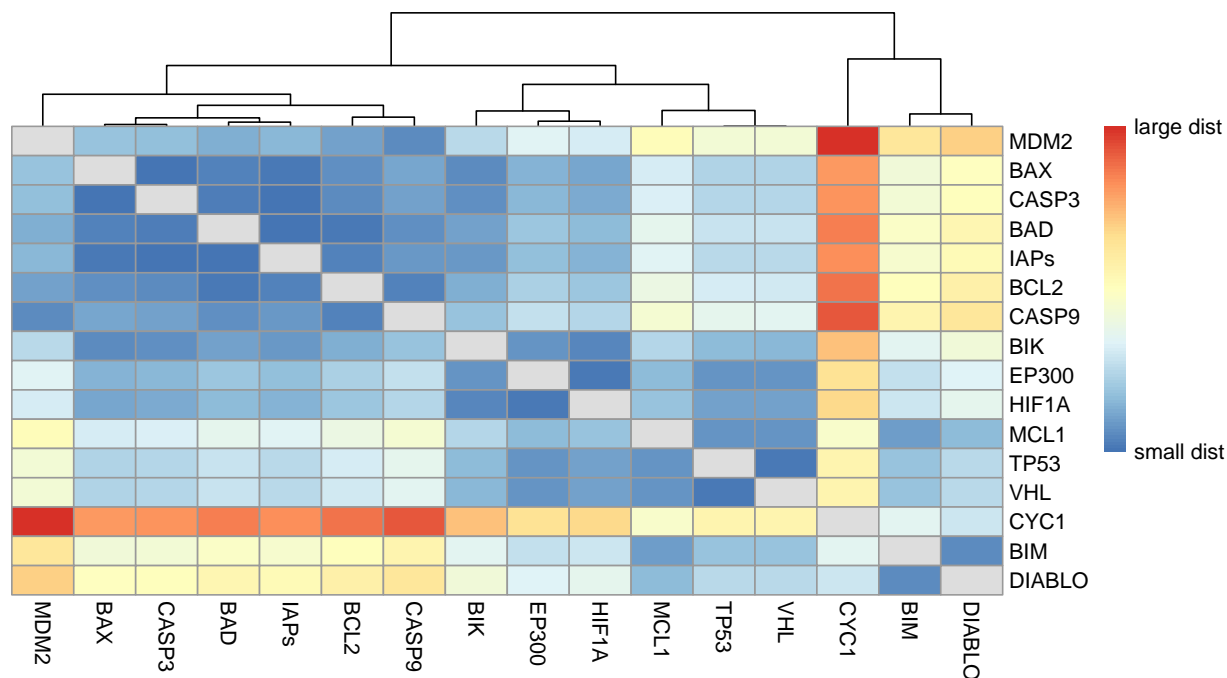
gridExtra::grid.arrange(grobs=list(p1$gtable, p2$gtable),
  nrow = 2 , labels=c('A', 'B'))

```

Clustering of Gene Expression on Hypoxia Euclidian Distance (All 3 datasets, all cell-lines)



Clustering of Gene Expression on Normoxia Euclidian Distance (All 3 datasets, all cell-lines)



```
dists <- cor(t(expr.all.hif), use = "pairwise.complete.obs", method = "spearman")
rownames(dists) <- rownames(expr.all.hif)
colnames(dists) <- rownames(expr.all.hif)
diag(dists) <- NA
```



```
pheatmap(dists,
  legend = TRUE,
  treeheight_row = 0,
  legend_breaks = c(min(dists, na.rm = TRUE),
    max(dists, na.rm = TRUE)),
  legend_labels = (c("-1", "1")),
  main = "Clustering of Gene Expression \n Spearman Correlation (All 3 datasets, all cell-lines)
```

