

# 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.symbols <- c("HIF1A", "TP53", "MDM2", "VHL", "EP300")
```

```
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::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::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))

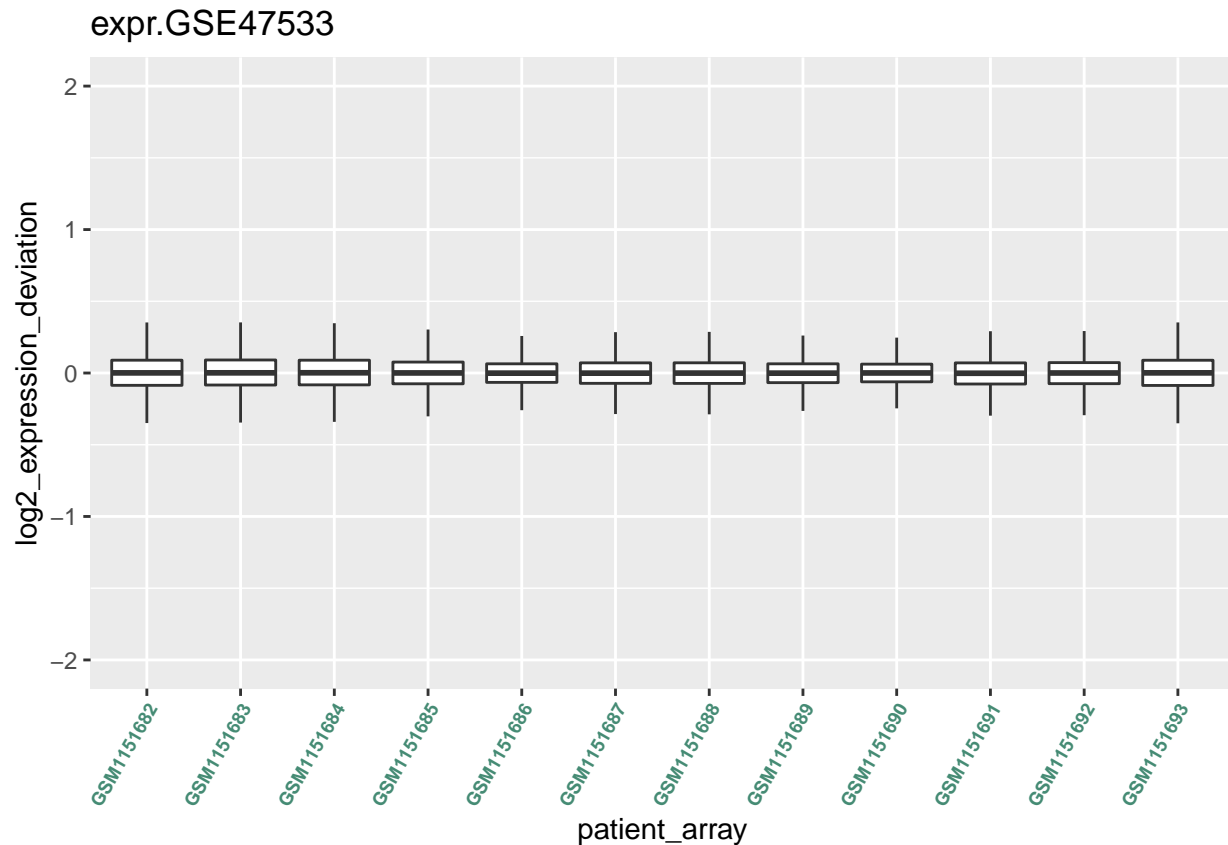
RLE_data <- sweep(expr.GSE47533, 1, row_medians_assayData)

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).

```

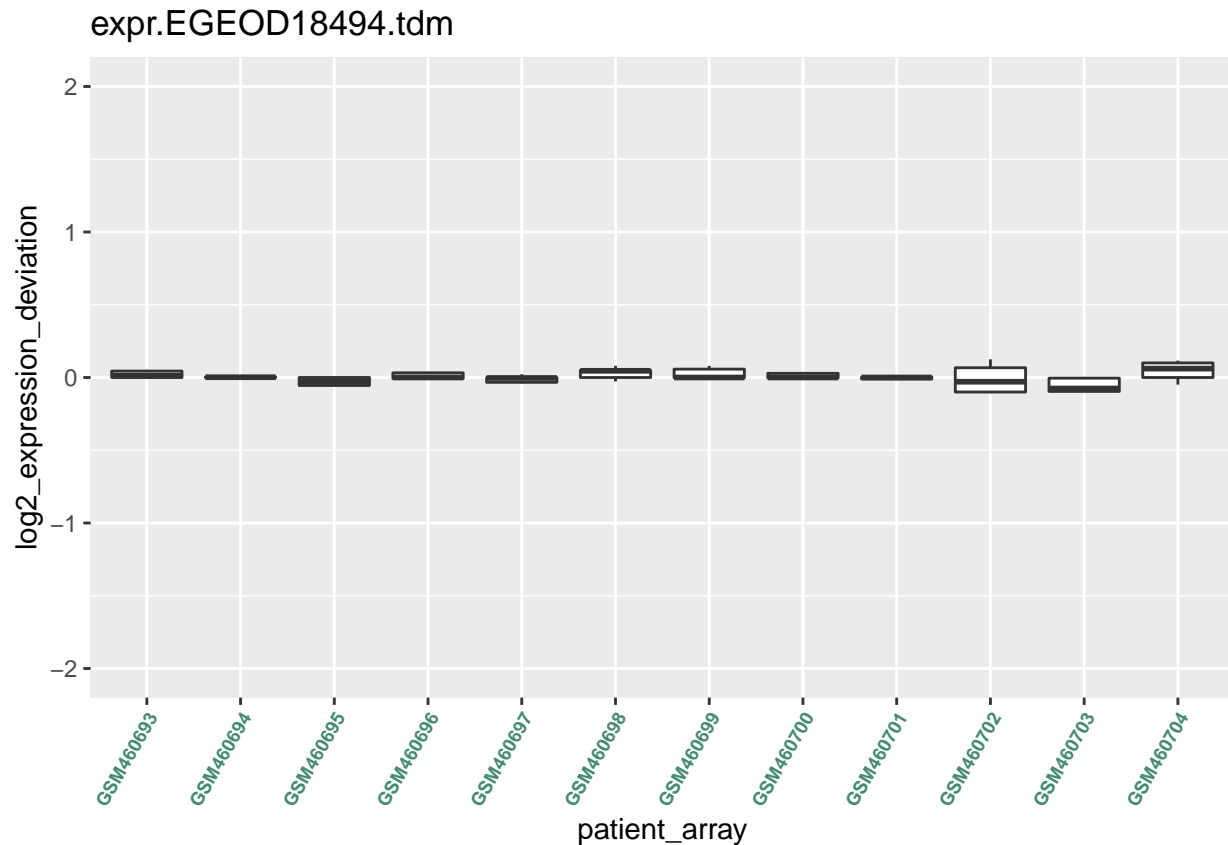


```
row_medians_assayData <-
  Biobase::rowMedians(as.matrix(expr.EGEOD18494.tdm))

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"))
```

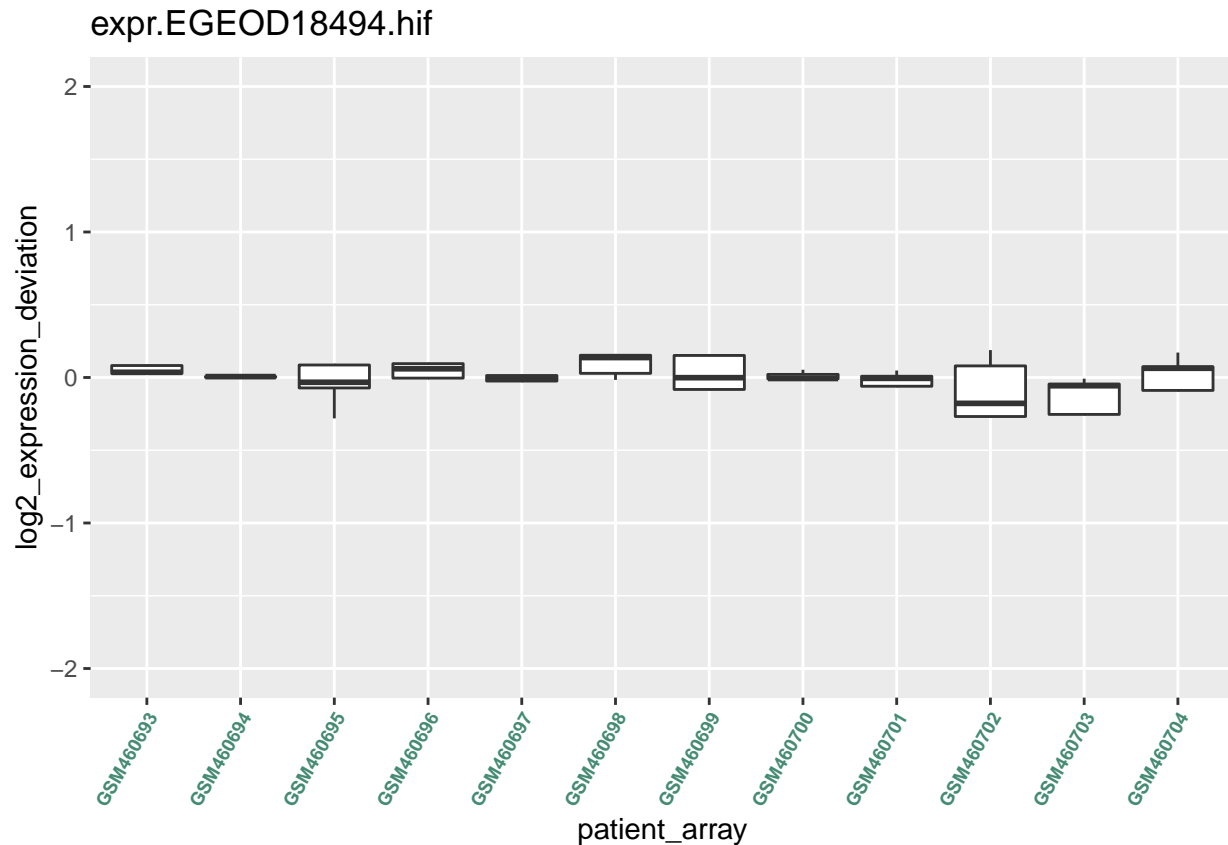


```
row_medians_assayData <-
  Biobase::rowMedians(as.matrix(expr.EGEOD18494.hif))

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]

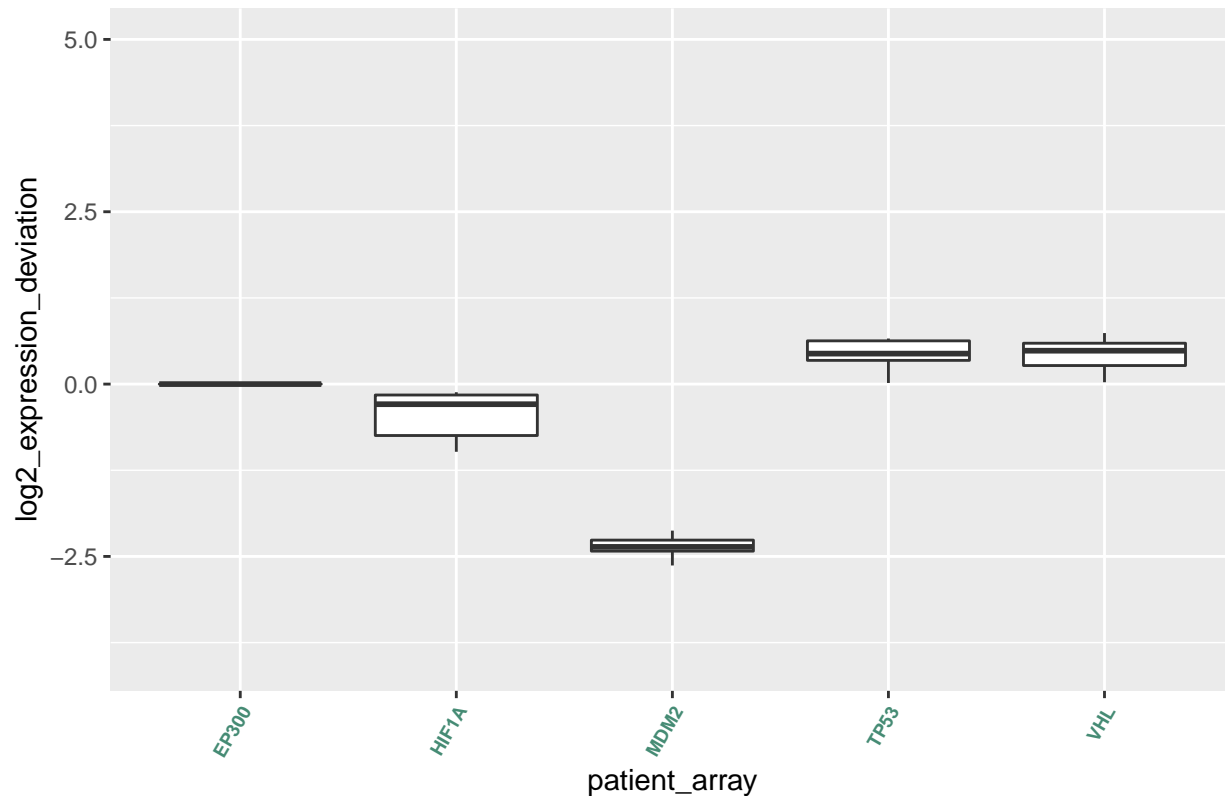
row_medians_assayData <-
  Biobase::rowMedians(as.matrix(t(expr.GSE47533.hif)))

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"))
```

## GSE47533 MCF7 – Raw data



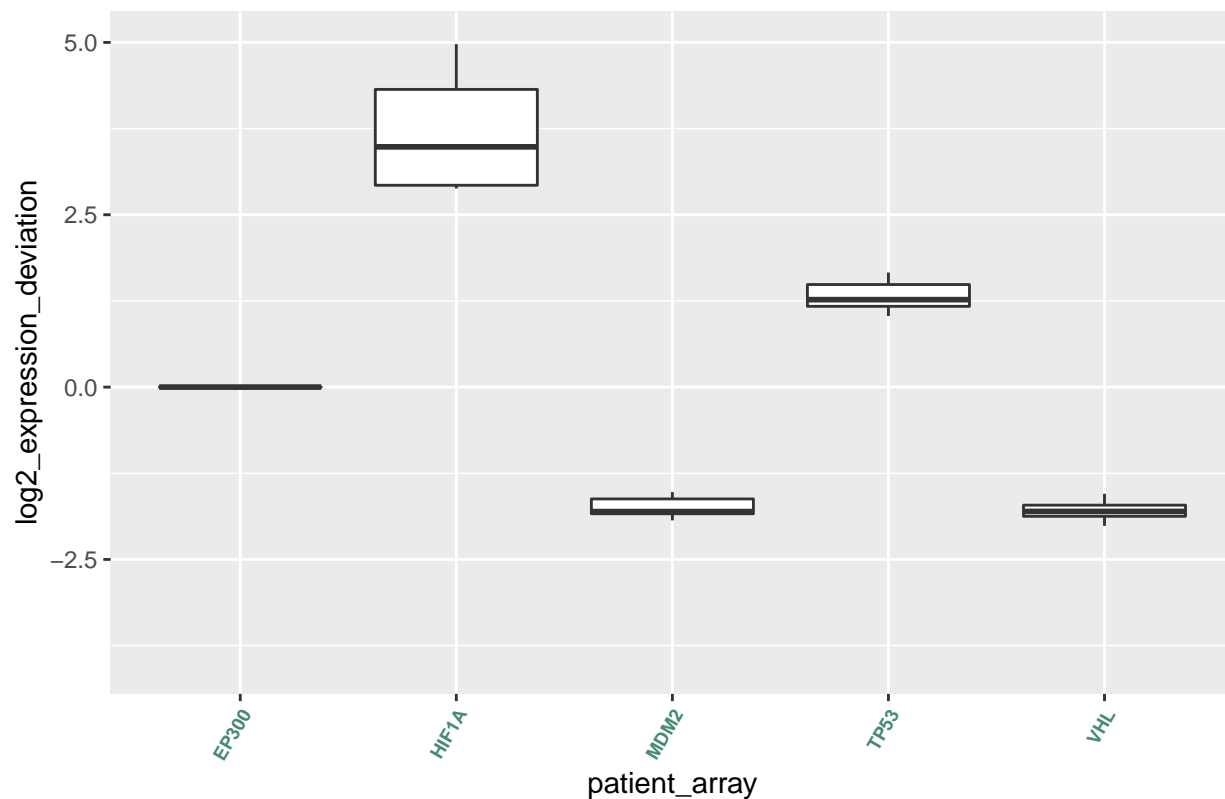
```
row_medians_assayData <-
  Biobase::rowMedians(as.matrix(t(expr.EGEOD18494.hif)))

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"))
```

## EGEOD18494 MDA-MB231 – Raw data



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

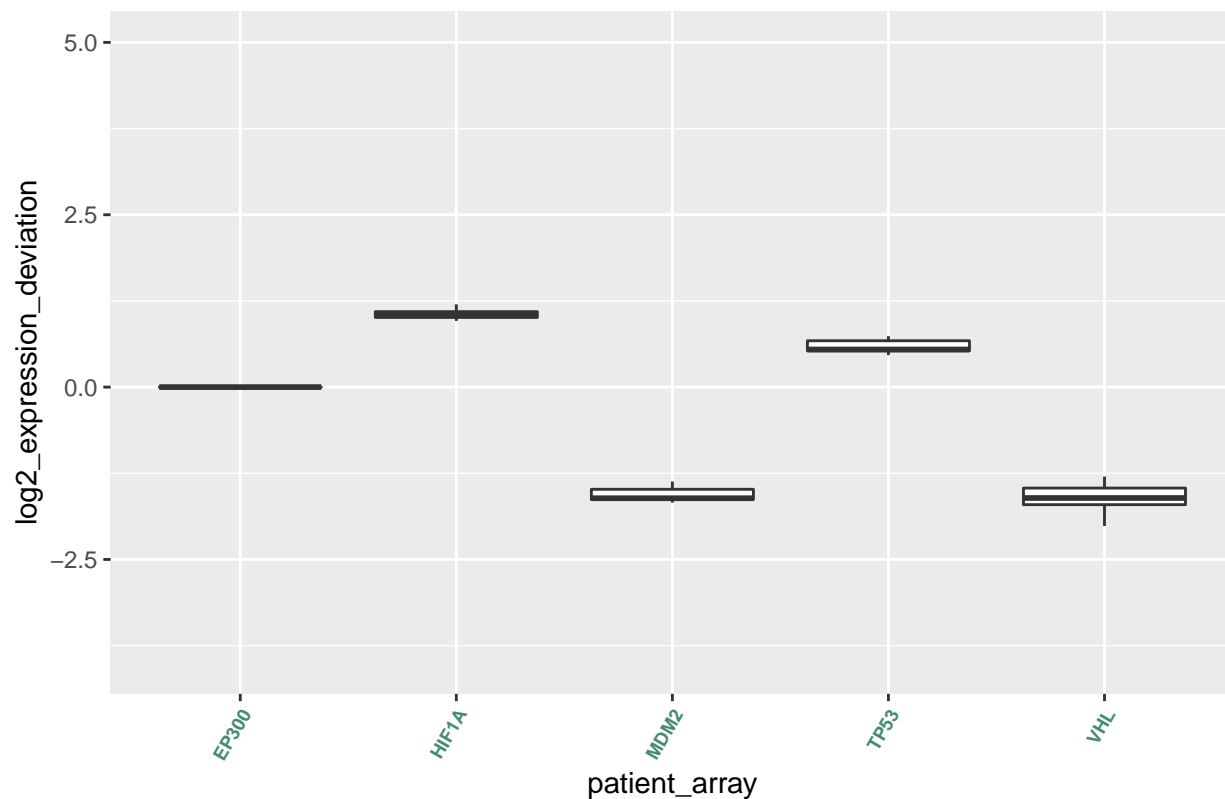
RLE_data <- sweep(t(expr.EGEOD18494.tdm), 1, row_medians_assayData)

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"))
```



## EGEOD18494 MDA-MB231 – TDM data



```
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, "p"),
  rep("p", nrow(expr.GSE47533.hif.bin)))

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

	Norm.O1	Norm.O2	Norm.O3	Hy.p.16	Hy.p.16	Hy.p.16	Hy.p.32	Hy.p.32	Hy.p.32	Hy.p.48	Hy.p.48	Hy.p.48	three shop	value	symbol
EP300	0	0	0	1	0	0	0	1	1	0	0	0	9.133766	6.386	EP300
HIF1A	1	1	1	0	0	0	0	0	0	1	1	1	8.699330	1.001	HIF1A

	Norm.O1	Norm.O2	Norm.OB	Hypo.1G	Hypo.1E	Hypo.1S	Hypo.3E	Hypo.3S	Hypo.3E	Hypo.4E	Hypo.4S	Hypo.4E	threshold	value	symbol
MDM2	1	1	1	1	1	1	1	0	1	1	1	1	6.619755	0.000	MDM2
TP53	0	1	1	1	1	1	0	0	1	0	0	0	9.531254	1.001	TP53
VHL	1	1	0	0	0	0	0	0	0	1	1	1	9.558766	0.806	VHL

```
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", "BAK", "PUMA", "NOXA", "FAS", "FADD", "TRAF3", "IKKα", "IKKβ", "NEMO", "cIAP1", "cIAP2",
                  "A1", "A2", "A3", "A4", "A5", "A6", "A7", "A8", "A9", "A10", "A11", "A12", "A13", "A14", "A15", "A16",
                  "A17", "A18", "A19", "A20", "A21", "A22", "A23", "A24", "A25", "A26", "A27", "A28", "A29", "A30", "A31",
                  "A32", "A33", "A34", "A35", "A36", "A37", "A38", "A39", "A40", "A41", "A42", "A43", "A44", "A45", "A46",
                  "A47", "A48", "A49", "A50", "A51", "A52", "A53", "A54", "A55", "A56", "A57", "A58", "A59", "A60", "A61",
                  "A62", "A63", "A64", "A65", "A66", "A67", "A68", "A69", "A70", "A71", "A72", "A73", "A74", "A75", "A76",
                  "A77", "A78", "A79", "A80", "A81", "A82", "A83", "A84", "A85", "A86", "A87", "A88", "A89", "A90", "A91",
                  "A92", "A93", "A94", "A95", "A96", "A97", "A98", "A99", "A100")

p.MCF7 <- ggplot(aes(x = factor(time), y = value, group = symbol, color="red"),
               #data = expr.GSE47533.hif.pivot[expr.GSE47533.hif.pivot$symbol %in% c("HIF1A", "TP53", "MDM2", "VHL", "EP300", "TMBIM1", "TMBIM4", "TMBIM6", "BAD", "BIK", "MCL1", "BAX", "BAK", "PUMA", "NOXA", "FAS", "FADD", "TRAF3", "IKKα", "IKKβ", "NEMO", "cIAP1", "cIAP2", "A1", "A2", "A3", "A4", "A5", "A6", "A7", "A8", "A9", "A10", "A11", "A12", "A13", "A14", "A15", "A16", "A17", "A18", "A19", "A20", "A21", "A22", "A23", "A24", "A25", "A26", "A27", "A28", "A29", "A30", "A31", "A32", "A33", "A34", "A35", "A36", "A37", "A38", "A39", "A40", "A41", "A42", "A43", "A44", "A45", "A46", "A47", "A48", "A49", "A50", "A51", "A52", "A53", "A54", "A55", "A56", "A57", "A58", "A59", "A60", "A61", "A62", "A63", "A64", "A65", "A66", "A67", "A68", "A69", "A70", "A71", "A72", "A73", "A74", "A75", "A76", "A77", "A78", "A79", "A80", "A81", "A82", "A83", "A84", "A85", "A86", "A87", "A88", "A89", "A90", "A91", "A92", "A93", "A94", "A95", "A96", "A97", "A98", "A99", "A100"),
               data = expr.GSE47533.hif.pivot[expr.GSE47533.hif.pivot$symbol %in% c("HIF1A", "TP53", "MDM2", "VHL", "EP300", "TMBIM1", "TMBIM4", "TMBIM6", "BAD", "BIK", "MCL1", "BAX", "BAK", "PUMA", "NOXA", "FAS", "FADD", "TRAF3", "IKKα", "IKKβ", "NEMO", "cIAP1", "cIAP2", "A1", "A2", "A3", "A4", "A5", "A6", "A7", "A8", "A9", "A10", "A11", "A12", "A13", "A14", "A15", "A16", "A17", "A18", "A19", "A20", "A21", "A22", "A23", "A24", "A25", "A26", "A27", "A28", "A29", "A30", "A31", "A32", "A33", "A34", "A35", "A36", "A37", "A38", "A39", "A40", "A41", "A42", "A43", "A44", "A45", "A46", "A47", "A48", "A49", "A50", "A51", "A52", "A53", "A54", "A55", "A56", "A57", "A58", "A59", "A60", "A61", "A62", "A63", "A64", "A65", "A66", "A67", "A68", "A69", "A70", "A71", "A72", "A73", "A74", "A75", "A76", "A77", "A78", "A79", "A80", "A81", "A82", "A83", "A84", "A85", "A86", "A87", "A88", "A89", "A90", "A91", "A92", "A93", "A94", "A95", "A96", "A97", "A98", "A99", "A100")],
               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)
```

Figure 1 consists of five line graphs showing the relative gene expression of HIF1A, TP53, VHL, EP300, and MDM2 in A549 cells under four conditions: Normoxia, Hypoxia: 16h, Hypoxia: 32h, and Hypoxia: 48h. The y-axis represents Gene Expression, ranging from 0.00 to 1.00. The x-axis represents the conditions. The graphs are arranged in a 2x3 grid, with the last cell empty.

- HIF1A:** Expression is 1.00 at Normoxia, drops to 0.00 at 16h and 32h hypoxia, and returns to 1.00 at 48h hypoxia.
- TP53:** Expression is approximately 0.65 at Normoxia, peaks at 1.00 at 16h hypoxia, and decreases to 0.00 at 48h hypoxia.
- VHL:** Expression is approximately 0.65 at Normoxia, drops to 0.00 at 16h and 32h hypoxia, and returns to 1.00 at 48h hypoxia.
- EP300:** Expression is 0.00 at Normoxia and 48h hypoxia, increases to approximately 0.35 at 16h hypoxia, and peaks at approximately 0.65 at 32h hypoxia.
- MDM2:** Expression is 1.00 at Normoxia and 16h hypoxia, drops to approximately 0.65 at 32h hypoxia, and returns to 1.00 at 48h hypoxia.

Gene	Normoxia	Hypoxia: 16h	Hypoxia: 32h	Hypoxia: 48h
HIF1A	1.00	0.00	0.00	1.00
TP53	0.65	1.00	0.35	0.00
VHL	0.65	0.00	0.00	1.00
EP300	0.00	0.35	0.65	0.00
MDM2	1.00	1.00	0.65	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)

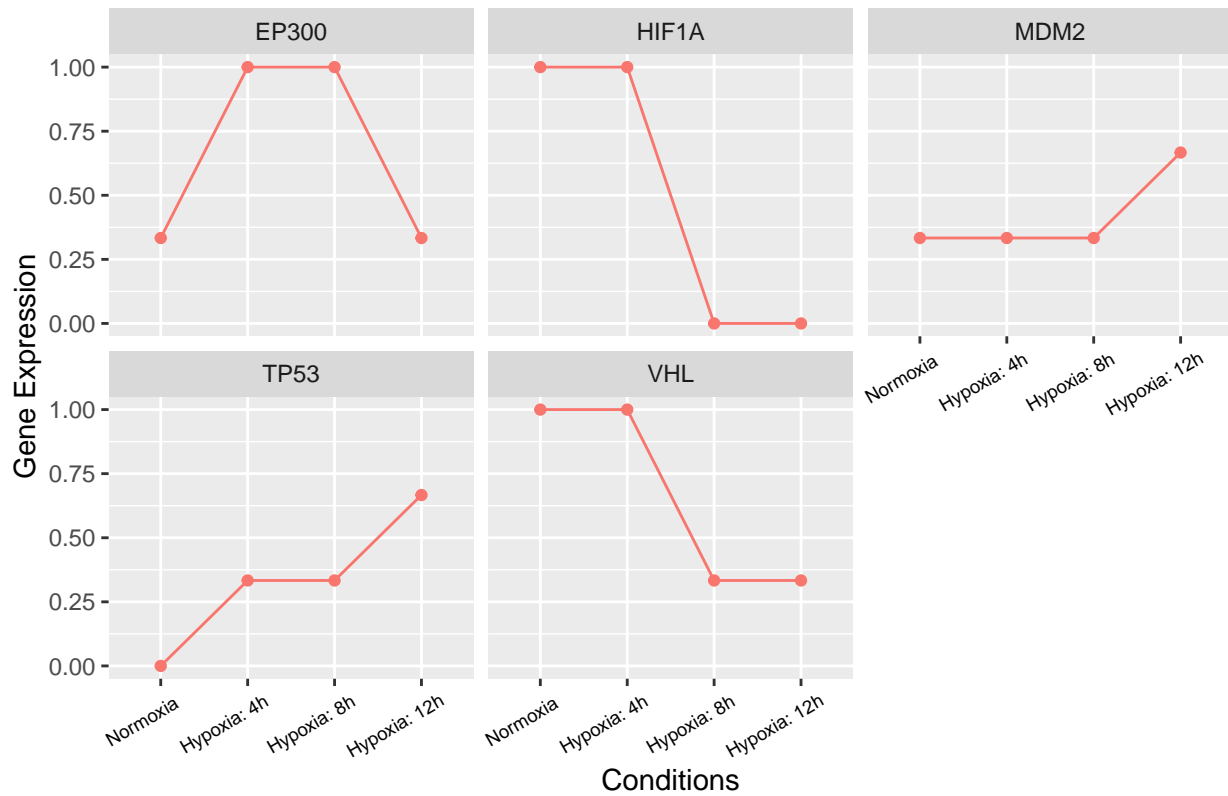
#expr.EGEOD18494.hif.bin <- expr.EGEOD18494.tdm

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", "gla", "luc", "mki", "mki2", "mki3", "mki4", "mki5", "mki6", "mki7", "mki8", "mki9", "mki10", "mki11", "mki12", "mki13", "mki14", "mki15", "mki16", "mki17", "mki18", "mki19", "mki20", "mki21", "mki22", "mki23", "mki24", "mki25", "mki26", "mki27", "mki28", "mki29", "mki30", "mki31", "mki32", "mki33", "mki34", "mki35", "mki36", "mki37", "mki38", "mki39", "mki40", "mki41", "mki42", "mki43", "mki44", "mki45", "mki46", "mki47", "mki48", "mki49", "mki50", "mki51", "mki52", "mki53", "mki54", "mki55", "mki56", "mki57", "mki58", "mki59", "mki60", "mki61", "mki62", "mki63", "mki64", "mki65", "mki66", "mki67", "mki68", "mki69", "mki70", "mki71", "mki72", "mki73", "mki74", "mki75", "mki76", "mki77", "mki78", "mki79", "mki80", "mki81", "mki82", "mki83", "mki84", "mki85", "mki86", "mki87", "mki88", "mki89", "mki90", "mki91", "mki92", "mki93", "mki94", "mki95", "mki96", "mki97", "mki98", "mki99", "mki100", "mki101", "mki102", "mki103", "mki104", "mki105", "mki106", "mki107", "mki108", "mki109", "mki110", "mki111", "mki112", "mki113", "mki114", "mki115", "mki116", "mki117", "mki118", "mki119", "mki120", "mki121", "mki122", "mki123", "mki124", "mki125", "mki126", "mki127", "mki128", "mki129", "mki130", "mki131", "mki132", "mki133", "mki134", "mki135", "mki136", "mki137", "mki138", "mki139", "mki140", "mki141", "mki142", "mki143", "mki144", "mki145", "mki146", "mki147", "mki148", "mki149", "mki150", "mki151", "mki152", "mki153", "mki154", "mki155", "mki156", "mki157", "mki158", "mki159", "mki160", "mki161", "mki162", "mki163", "mki164", "mki165", "mki166", "mki167", "mki168", "mki169", "mki170", "mki171", "mki172", "mki173", "mki174", "mki175", "mki176", "mki177", "mki178", "mki179", "mki180", "mki181", "mki182", "mki183", "mki184", "mki185", "mki186", "mki187", "mki188", "mki189", "mki190", "mki191", "mki192", "mki193", "mki194", "mki195", "mki196", "mki197", "mki198", "mki199", "mki200", "mki201", "mki202", "mki203", "mki204", "mki205", "mki206", "mki207", "mki208", "mki209", "mki210", "mki211", "mki212", "mki213", "mki214", "mki215", "mki216", "mki217", "mki218", "mki219", "mki220", "mki221", "mki222", "mki223", "mki224", "mki225", "mki226", "mki227", "mki228", "mki229", "mki230", "mki231", "mki232", "mki233", "mki234", "mki235", "mki236", "mki237", "mki238", "mki239", "mki240", "mki241", "mki242", "mki243", "mki244", "mki245", "mki246", "mki247", "mki248", "mki249", "mki250", "mki251", "mki252", "mki253", "mki254", "mki255", "mki256", "mki257", "mki258", "mki259", "mki260", "mki261", "mki262", "mki263", "mki264", "mki265", "mki266", "mki267", "mki268", "mki269", "mki270", "mki271", "mki272", "mki273", "mki274", "mki275", "mki276", "mki277", "mki278", "mki279", "mki280", "mki281", "mki282", "mki283", "mki284", "mki285", "mki286", "mki287", "mki288", "mki289", "mki290", "mki291", "mki292", "mki293", "mki294", "mki295", "mki296", "mki297", "mki298", "mki299", "mki300", "mki301", "mki302", "mki303", "mki304", "mki305", "mki306", "mki307", "mki308", "mki309", "mki310", "mki311", "mki312", "mki313", "mki314", "mki315", "mki316", "mki317", "mki318", "mki319", "mki320", "mki321", "mki322", "mki323", "mki324", "mki325", "mki326", "mki327", "mki328", "mki329", "mki330", "mki331", "mki332", "mki333", "mki334", "mki335", "mki336", "mki337", "mki338", "mki339", "mki340", "mki341", "mki342", "mki343", "mki344", "mki345", "mki346", "mki347", "mki348", "mki349", "mki350", "mki351", "mki352", "mki353", "mki354", "mki355", "mki356", "mki357", "mki358", "mki359", "mki360", "mki361", "mki362", "mki363", "mki364", "mki365", "mki366", "mki367", "mki368", "mki369", "mki370", "mki371", "mki372", "mki373", "mki374", "mki375", "mki376", "mki377", "mki378", "mki379", "mki380", "mki381", "mki382", "mki383", "mki384", "mki385", "mki386", "mki387", "mki388", "mki389", "mki390", "mki391", "mki392", "mki393", "mki394", "mki395", "mki396", "mki397", "mki398", "mki399", "mki400", "mki401", "mki402", "mki403", "mki404", "mki405", "mki406", "mki407", "mki408", "mki409", "mki410", "mki411", "mki412", "mki413", "mki414", "mki415", "mki416", "mki417", "mki418", "mki419", "mki420", "mki421", "mki422", "mki423", "mki424", "mki425", "mki426", "mki427", "mki428", "mki429", "mki430", "mki431", "mki432", "mki433", "mki434", "mki435", "mki436", "mki437", "mki438", "mki439", "mki440", "mki441", "mki442", "mki443", "mki444", "mki445", "mki446", "mki447", "mki448", "mki449", "mki450", "mki451", "mki452", "mki453", "mki454", "mki455", "mki456", "mki457", "mki458", "mki459", "mki460", "mki461", "mki462", "mki463", "mki464", "mki465", "mki466", "mki467", "mki468", "mki469", "mki470", "mki471", "mki472", "mki473", "mki474", "mki475", "mki476", "mki477", "mki478", "mki479", "mki480", "mki481", "mki482", "mki483", "mki484", "mki485", "mki486", "mki487", "mki488", "mki489", "mki490", "mki491", "mki492", "mki493", "mki494", "mki495", "mki496", "mki497", "mki498", "mki499", "mki500", "mki501", "mki502", "mki503", "mki504", "mki505", "mki506", "mki507", "mki508", "mki509", "mki510", "mki511", "mki512", "mki513", "mki514", "mki515", "mki516", "mki517", "mki518", "mki519", "mki520", "mki521", "mki522", "mki523", "mki524", "mki525", "mki526", "mki527", "mki528", "mki529", "mki530", "mki531", "mki532", "mki533", "mki534", "mki535", "mki536", "mki537", "mki538", "mki539", "mki540", "mki541", "mki542", "mki543", "mki544", "mki545", "mki546", "mki547", "mki548", "mki549", "mki550", "mki551", "mki552", "mki553", "mki554", "mki555", "mki556", "mki557", "mki558", "mki559", "mki560", "mki561", "mki562", "mki563", "mki564", "m
```



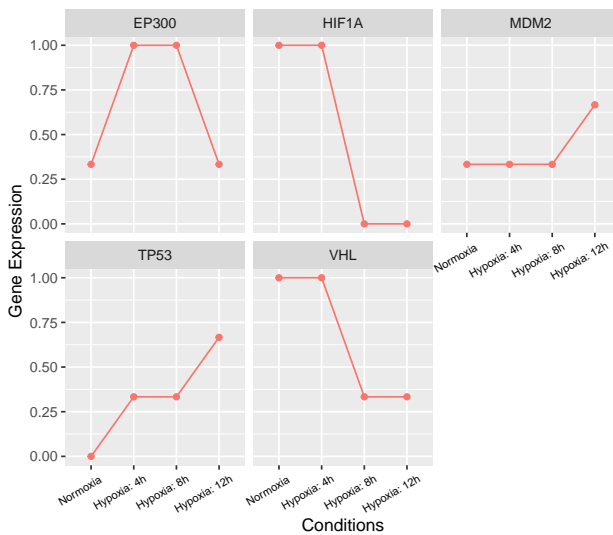
## EGEOD18494 MDA-MB231 – Raw data



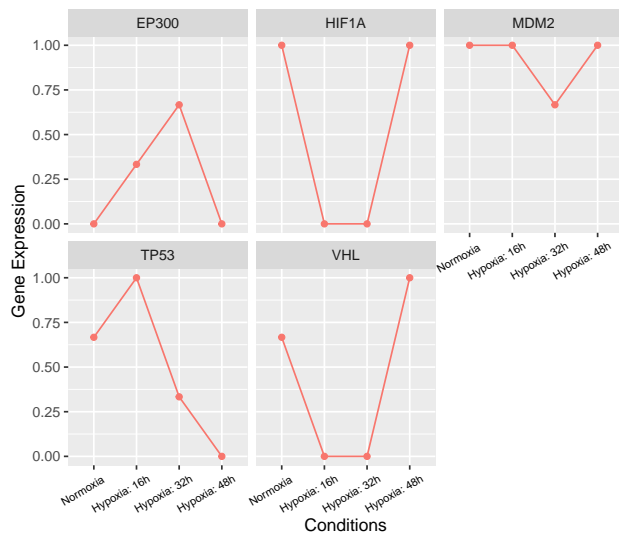
```
library(cowplot)
```

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

### A EGEOD18494 MDA-MB231 – Raw data



### B GSE47533 MCF7 – Raw data



# Heatmaps - EGEOD18494

## Multivariate Shapiro-Wilk normality test

From the output, the p-value > 0.05 implying that the distribution of the data are not significantly different from normal distribution. In other words, we can assume the normality.

```
#library(rstatix)

#rstatix::mshapiro_test(expr.EGEOD18494.hif)

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

## =====
## ComplexHeatmap version 2.4.3
## 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.
##
## This message can be suppressed by:
## suppressPackageStartupMessages(library(ComplexHeatmap))
## =====

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")
)

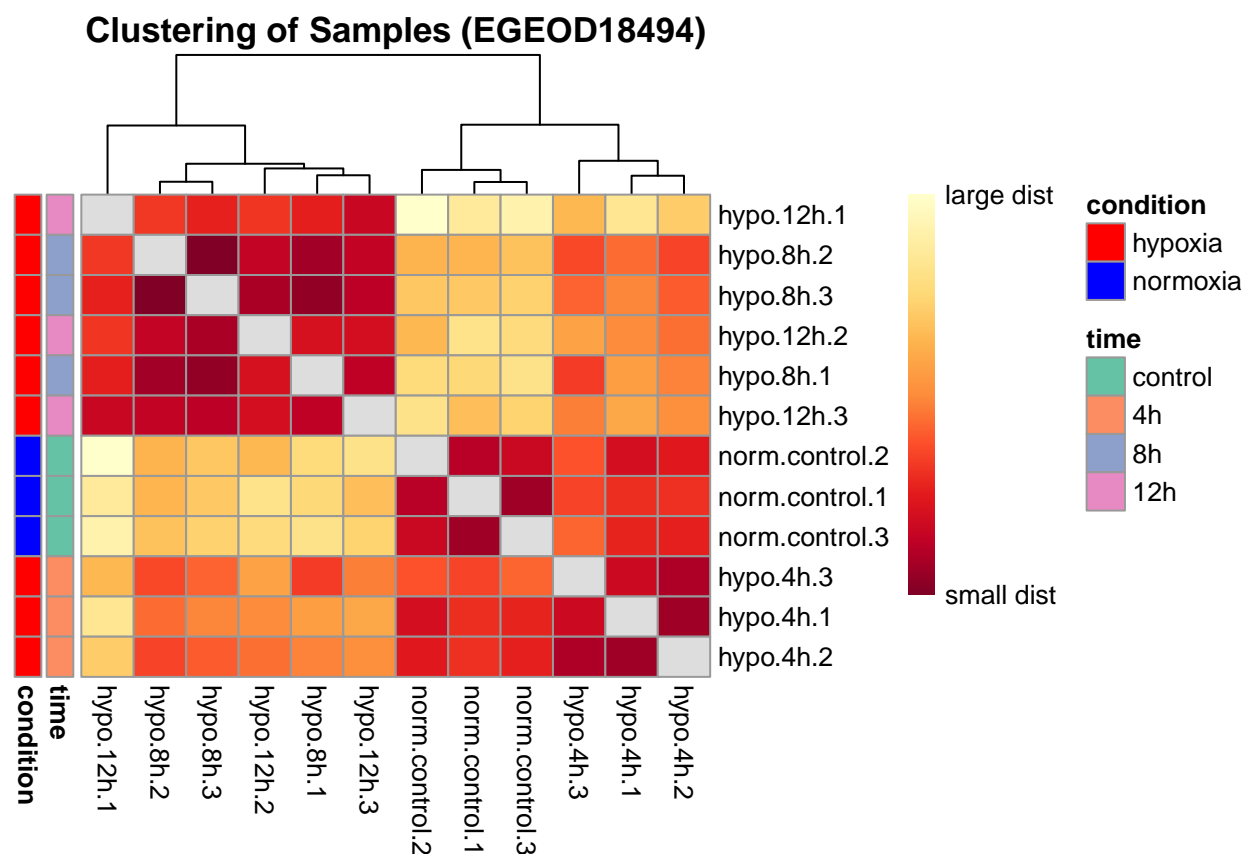
ann_colors

## $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 = (hmc col),
  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)")
```



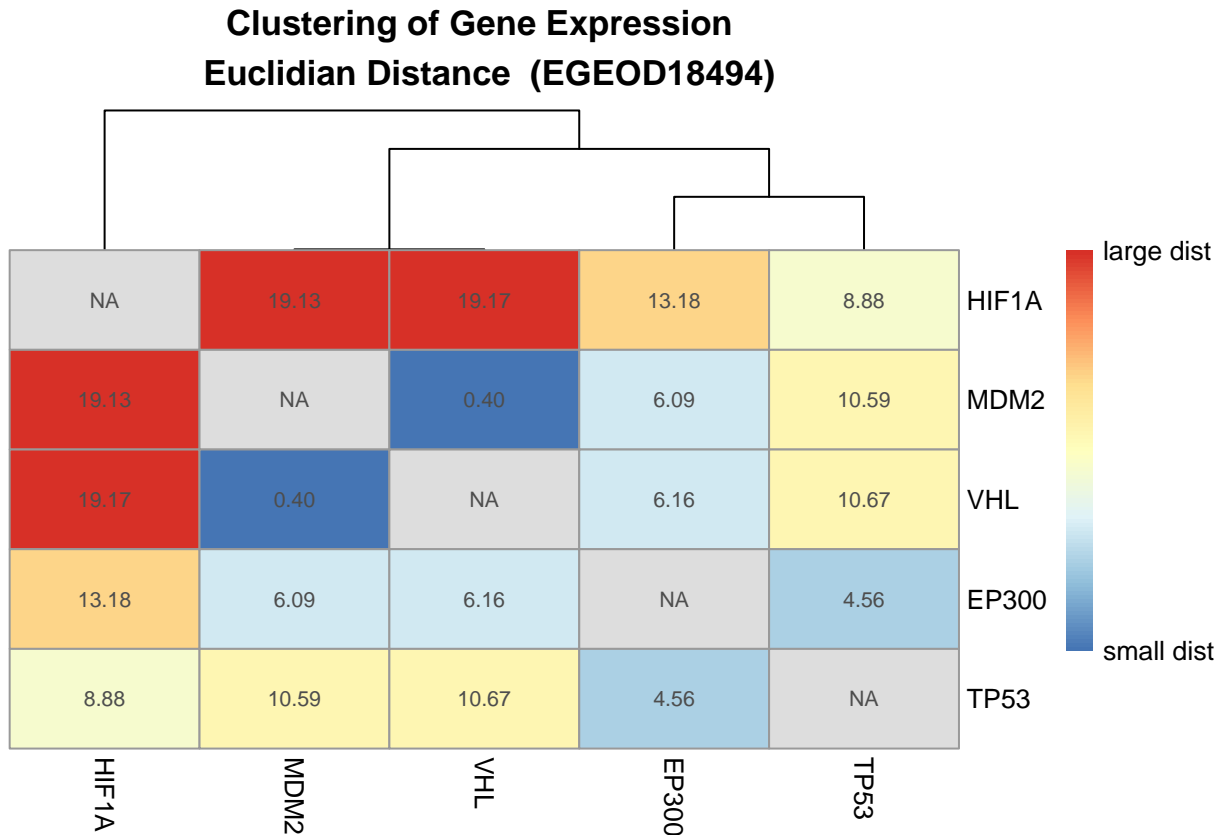
```
dists <- as.matrix(dist(expr.EGEOD18494.hif, method = "euclidean"))
rownames(dists) <- rownames(expr.EGEOD18494.hif)
colnames(dists) <- rownames(expr.EGEOD18494.hif)
diag(dists) <- NA

pheatmap(dists, #row = (hmc col),
  #annotation_col = annotation_for_heatmap,
  #annotation_colors = ann_colors,
  legend = TRUE,
  display_numbers = T,
  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)"

```



```

#-----

expr.row <- (colnames(expr.EGEOD18494.hif) %in% data.EGEOD18494$codes[data.EGEOD18494$cell_line == "MDA-MB-231"])
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,
  display_numbers = T,
  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"])

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"

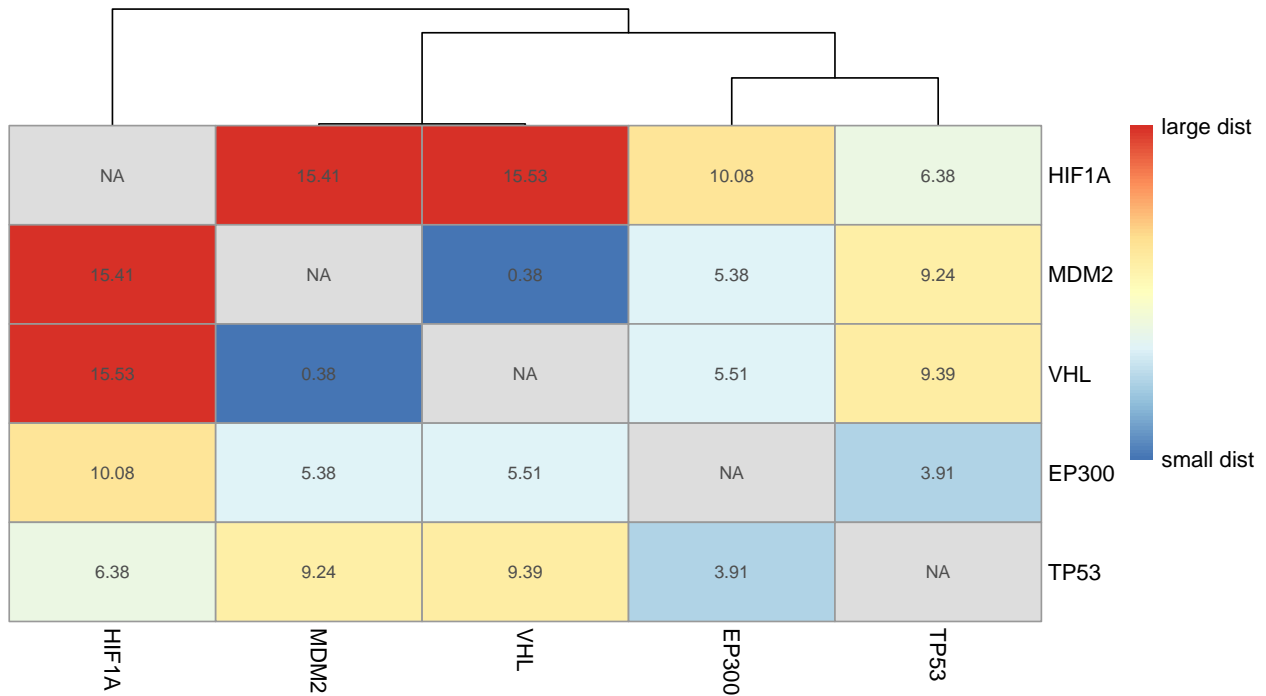
names(ann_colors$time) <- levels(data.EGEOD18494$time)
names(ann_colors$condition) <- levels(data.EGEOD18494$condition)

p2 <- pheatmap(dists, #row = (hmcol),
  #annotation_col = annotation_for_heatmap,
  #annotation_colors = ann_colors,
  legend = TRUE,
  display_numbers = T,
  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)

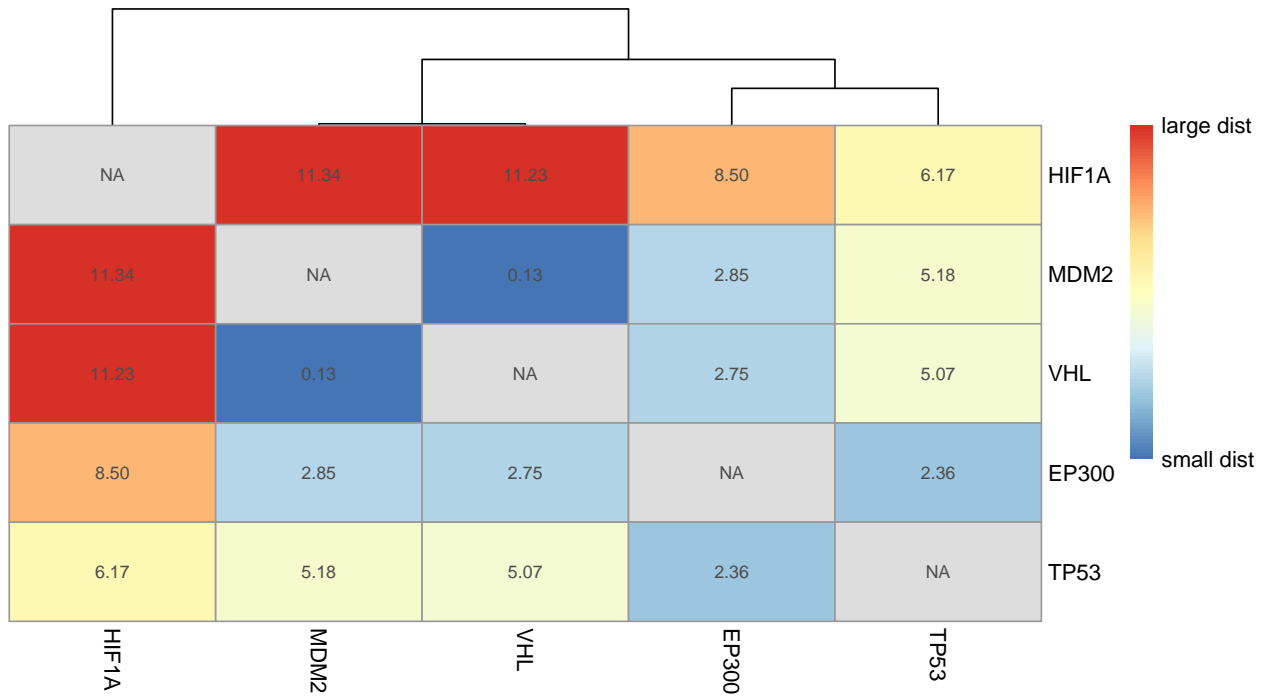
gridExtra::grid.arrange(grobs=list(p1$gtable, p2$gtable),
  nrow = 2 , labels=c('A', 'B'))

```

### 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"))

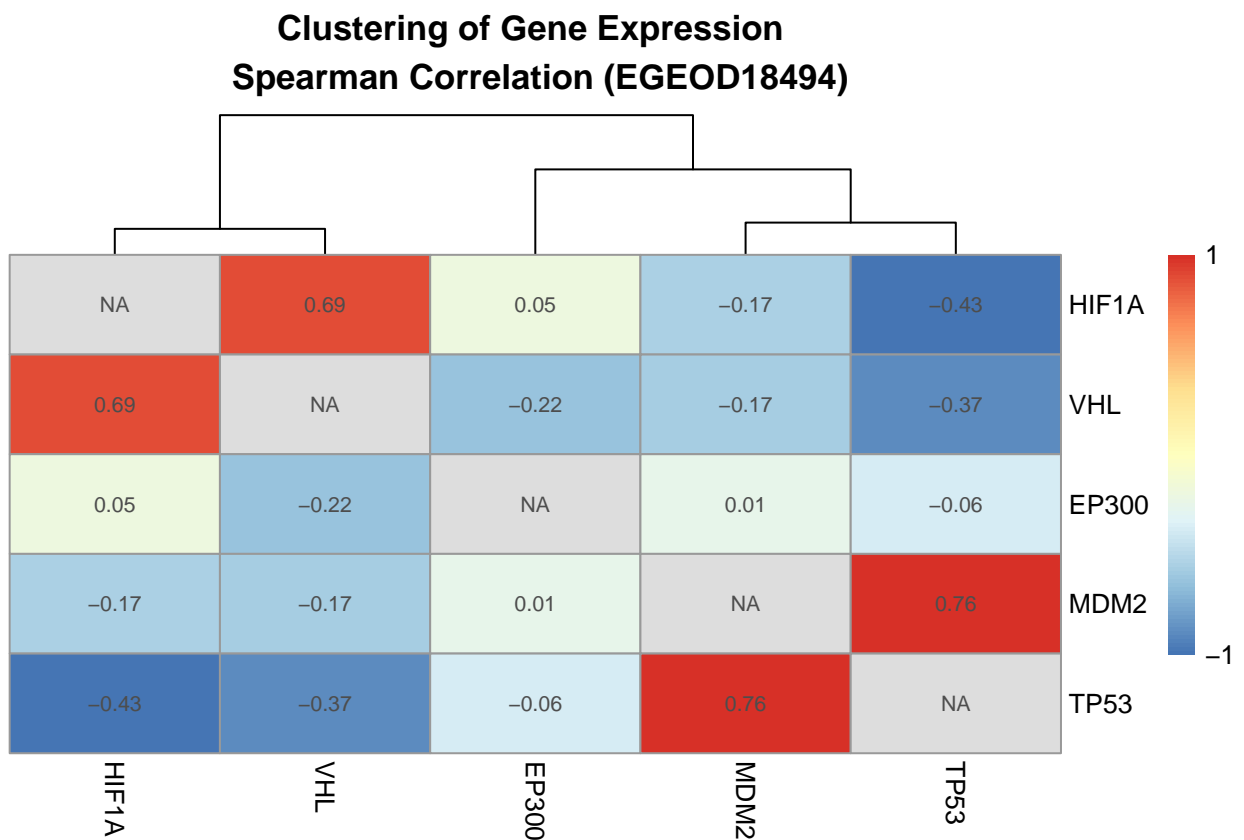
dists <- cor(t(expr.EGEOD18494.hif), use = "pairwise.complete.obs", method = "spearman")
rownames(dists) <- rownames(expr.EGEOD18494.hif)
```

```

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,
          display_numbers = T,
          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)")

```



```

dists <- cor(t(expr.EGEOD18494.hif), use = "pairwise.complete.obs", method = "pearson")
rownames(dists) <- rownames(expr.EGEOD18494.hif)
hmccl <- rev(colorRampPalette(RColorBrewer::brewer.pal(9, "YlOrRd"))(255))
colnames(dists) <- rownames(expr.EGEOD18494.hif)
diag(dists) <- NA

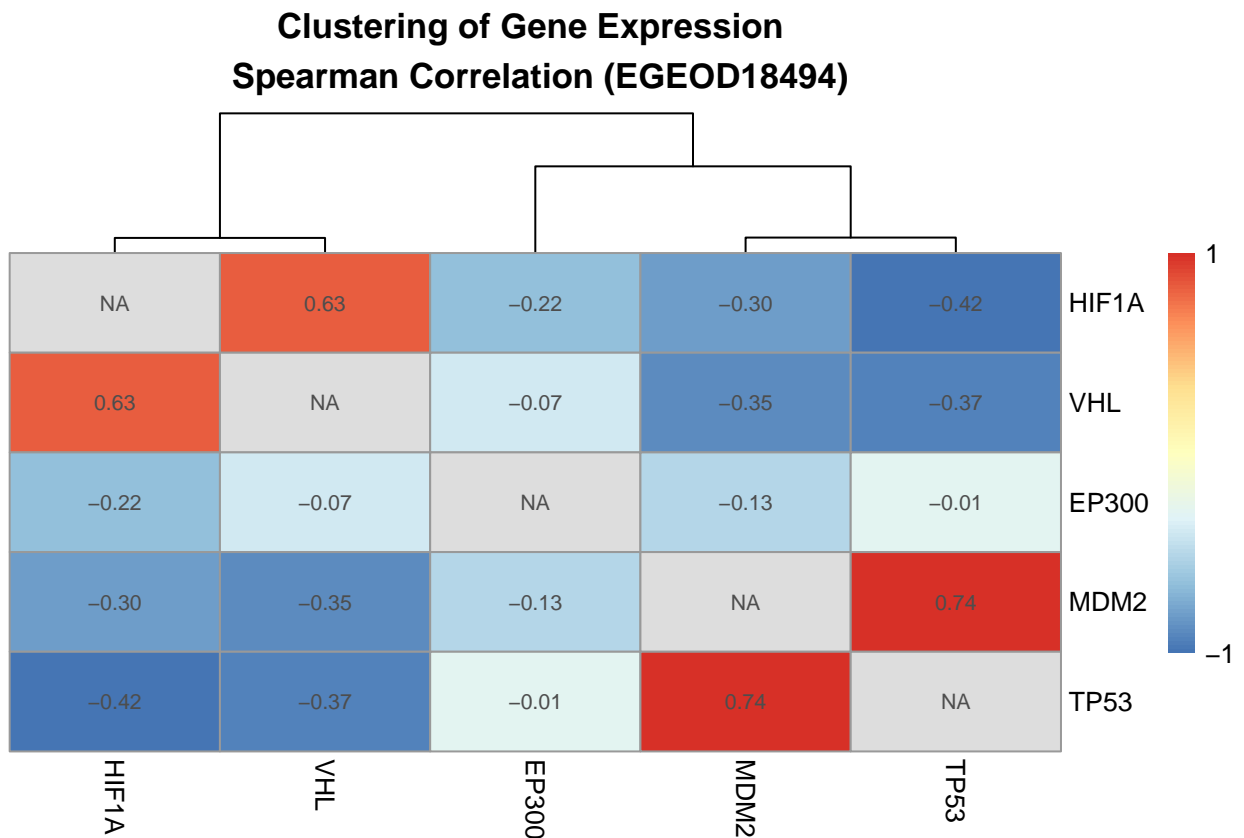
pheatmap(dists, #row = (hmccl),
          #annotation_col = annotation_for_heatmap,
          #annotation_colors = ann_colors,
          legend = TRUE,
          display_numbers = T,
          treeheight_row = 0,

```

```

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

### Multivariate Shapiro-Wilk normality test

From the output, the p-value > 0.05 implying that the distribution of the data are not significantly different from normal distribution. In other words, we can assume the normality.

```

# library(rstatix)
#
# rstatix::mshapiro_test(expr.GSE47533.hif) %>%
# knitr::kable(.)

```

```

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

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")
)

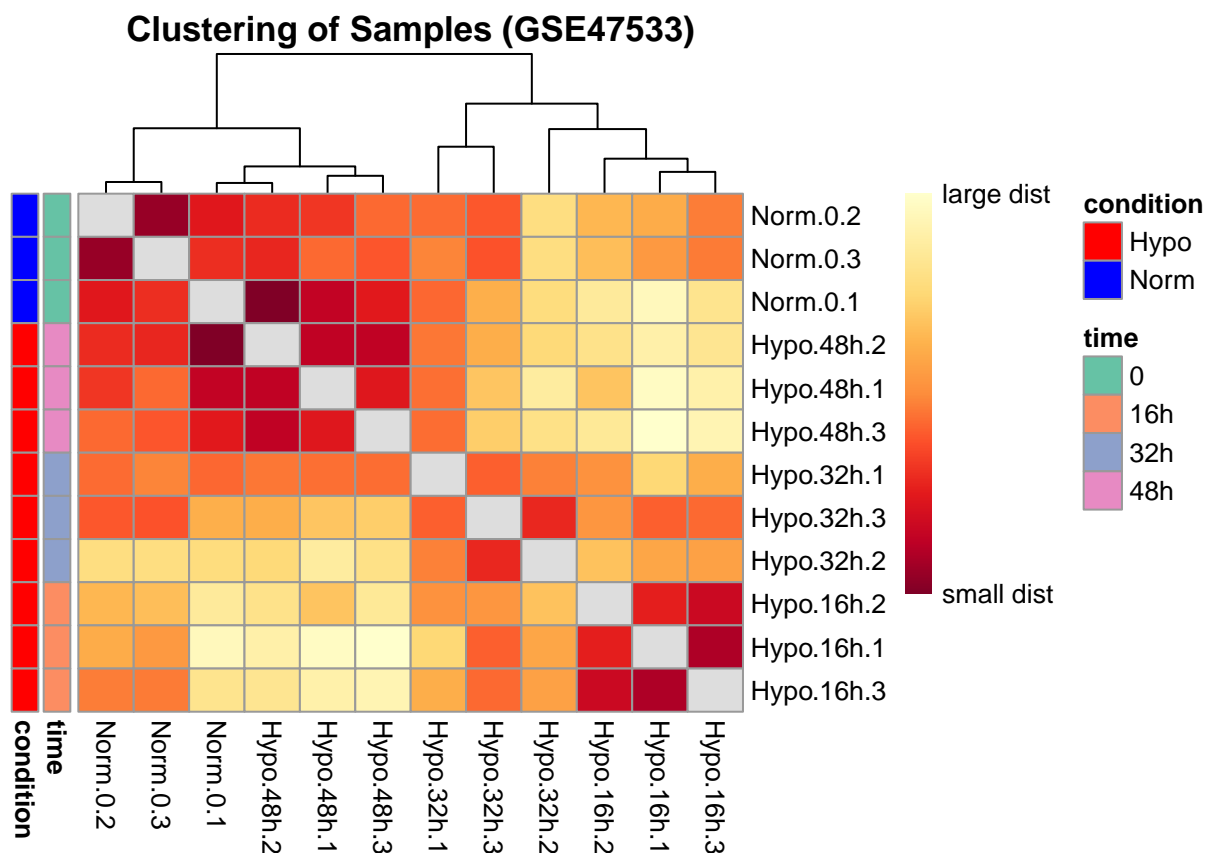
ann_colors

## $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)")

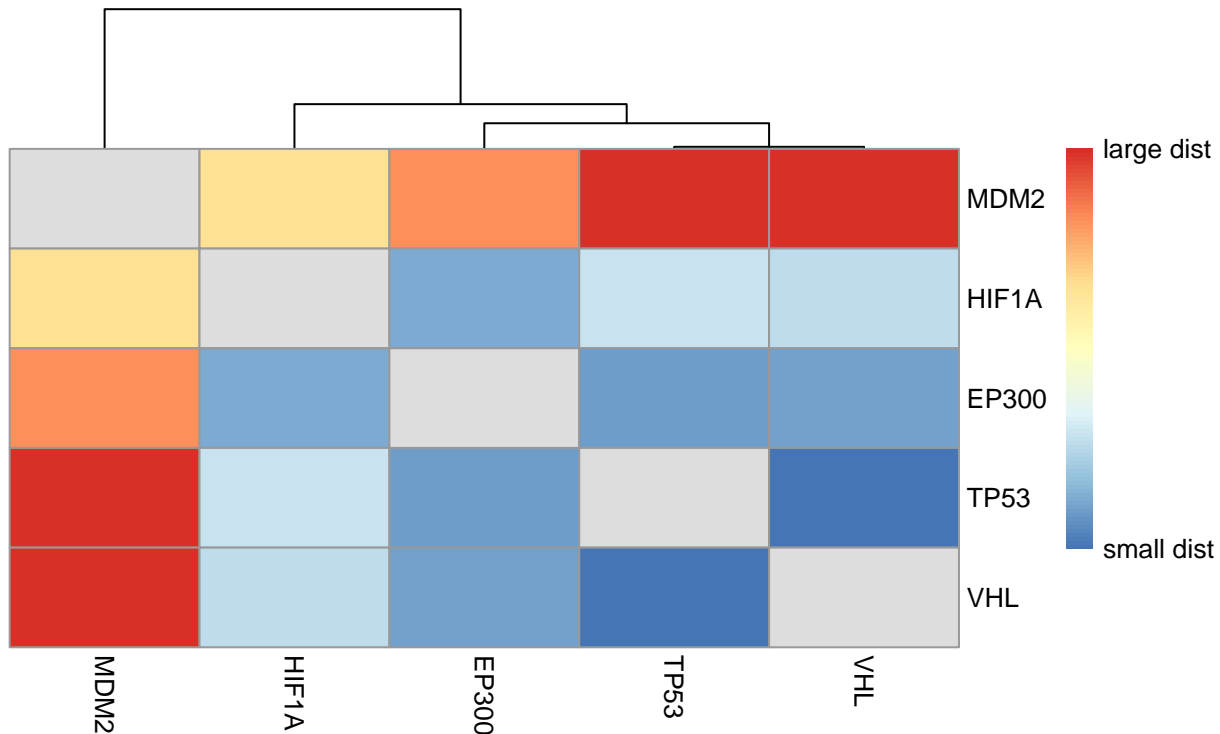
```



```
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
```

```

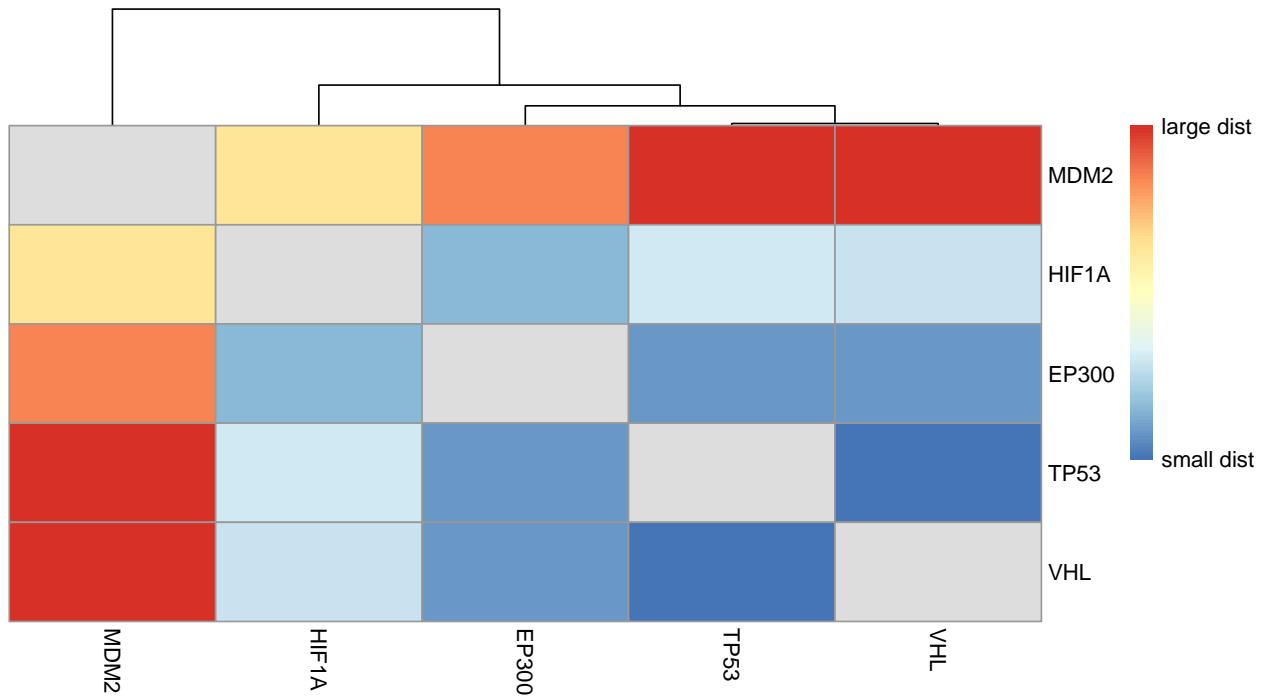
p2 <- pheatmap(dists,
  legend = TRUE,
  treeheight_row = 0,
  legend_breaks = c(min(dists, na.rm = TRUE),
    max(dists, na.rm = TRUE)),
  legend_labels = (c("small dist", "large dist")),
  main = "Clustering of Gene Expression on Normoxia \n Euclidian Distance (GSE47533)",
  silent=T)

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

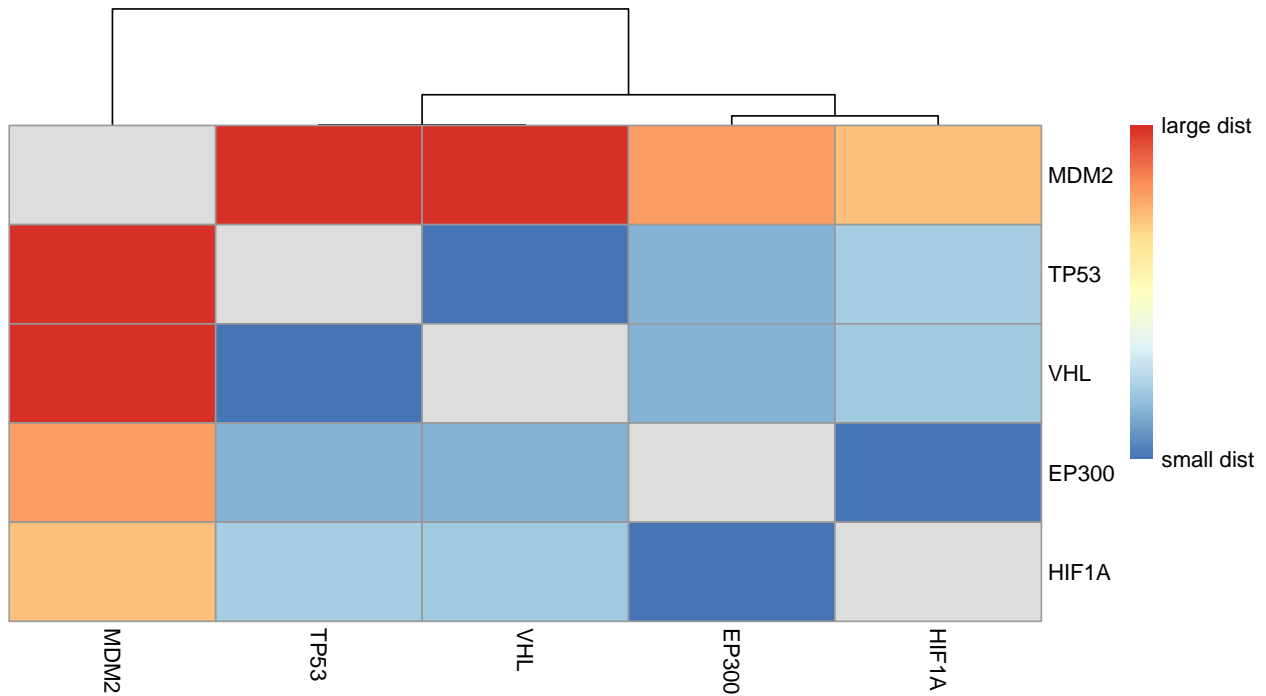
```



### 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
```

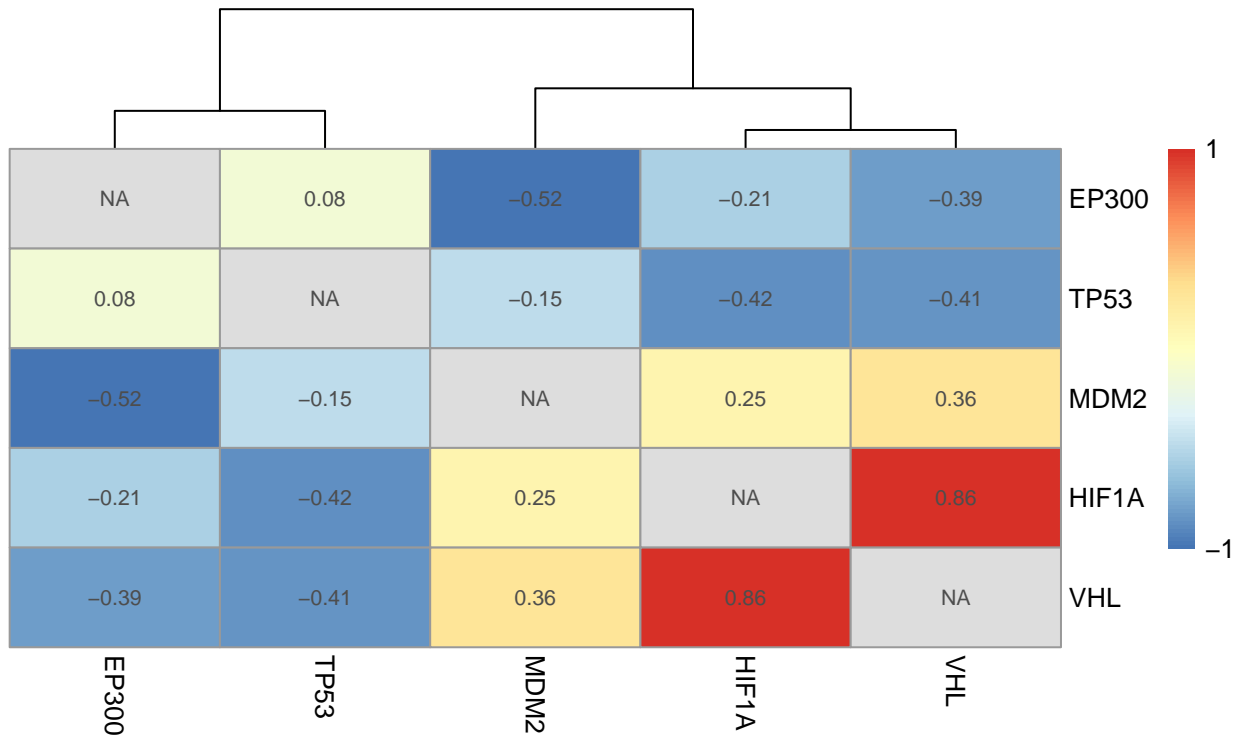
```
pheatmap(dists,
  legend = TRUE,
  display_numbers = T,
  treeheight_row = 0,
  legend_breaks = c(min(dists, na.rm = TRUE),
    max(dists, na.rm = TRUE)),
  legend_labels = (c("-1", "1")),
  main = "Clustering of Gene Expression \n Spearman Correlation (GSE47533)")
```



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

pheatmap(dists,
  legend = TRUE,
  display_numbers = T,
  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 Pearson Correlation (GSE47533)")
```

## Clustering of Gene Expression Pearson 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)

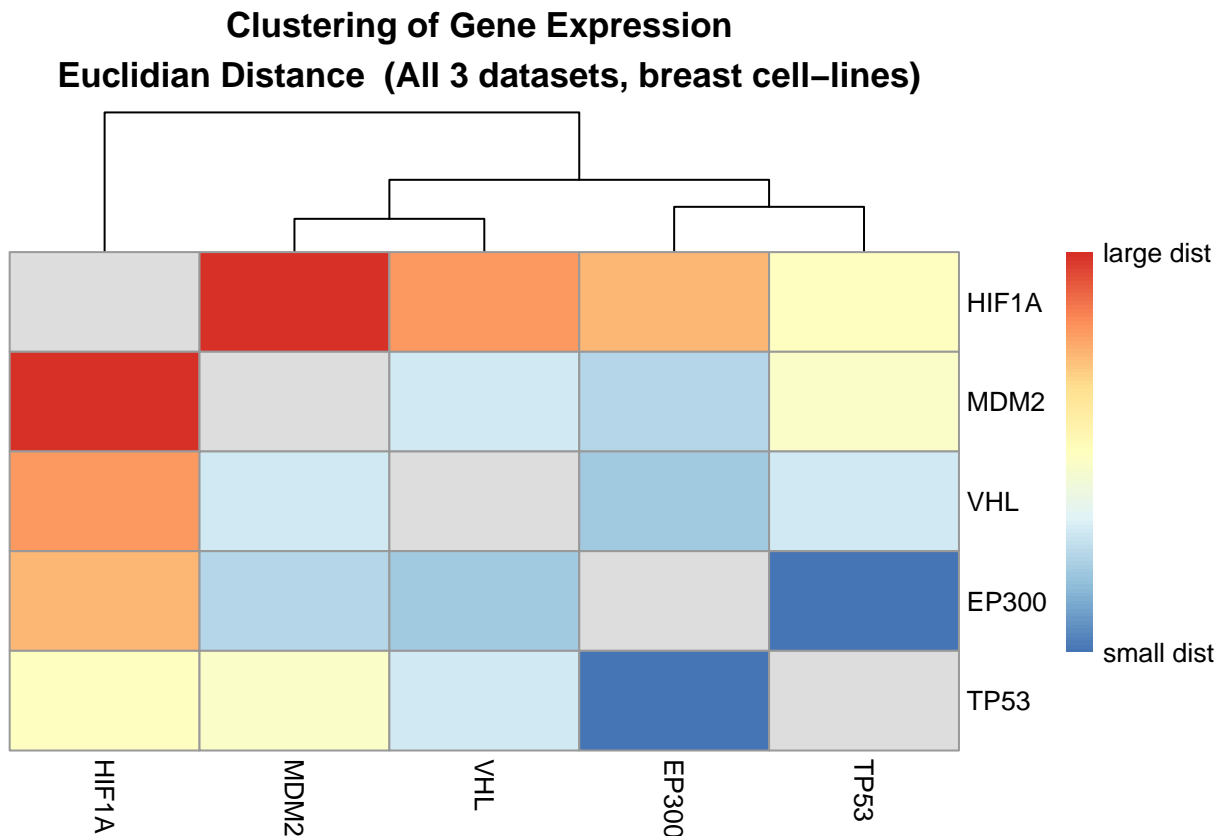
expr.all.hif <- cbind(expr.GSE47533.hif, expr.EGEOD18494.hif, expr.GSE41491.hif)

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)
```



```
#-----

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, breast cancer) \n
    silent=T)

#-----

col_norm <- union(data.GSE47533$codes[data.GSE47533$condition == "Norm"],
                  union(data.EGEO18494$codes[data.EGEO18494$condition == "normoxia"],
                        data.GSE41491$codes[data.GSE41491$condition == "no"]))

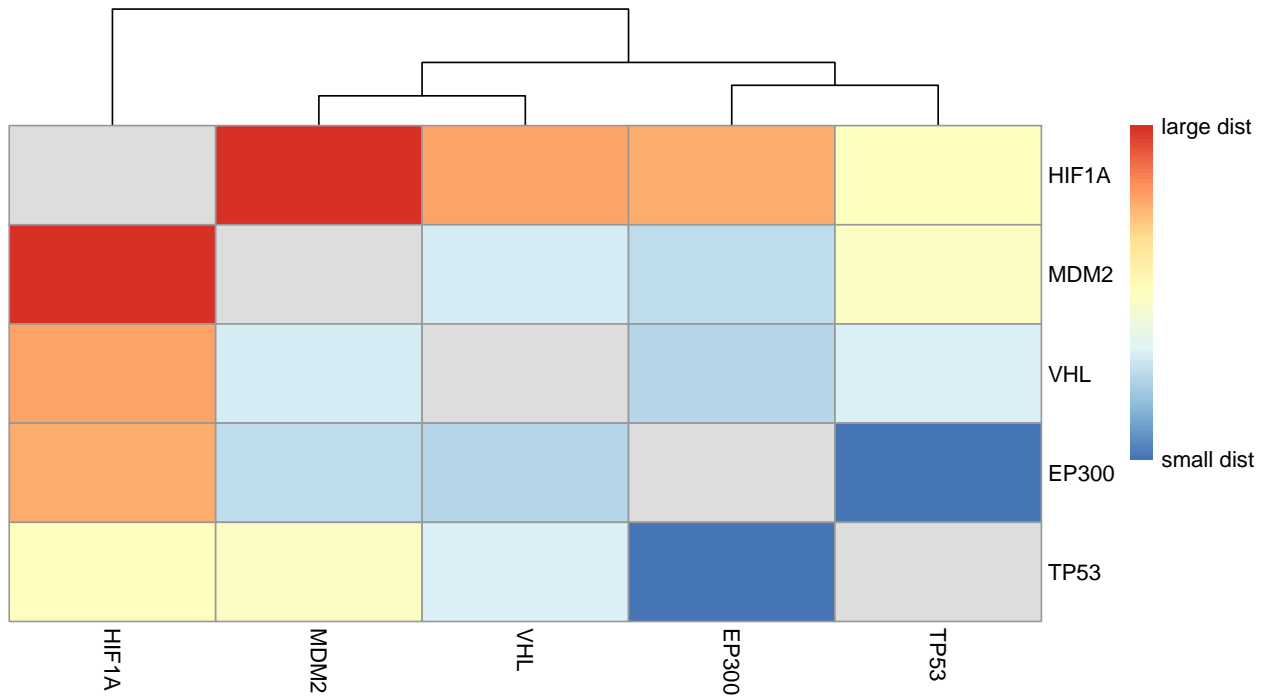
expr.row <- (colnames(expr.all.hif) %in% col_norm)
dists <- as.matrix(dist(expr.all.hif[expr.row], method = "euclidean"))
rownames(dists) <- rownames(expr.all.hif[expr.row])
colnames(dists) <- rownames(expr.all.hif[expr.row])
diag(dists) <- NA

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, breast cancer) \n
               silent=T)

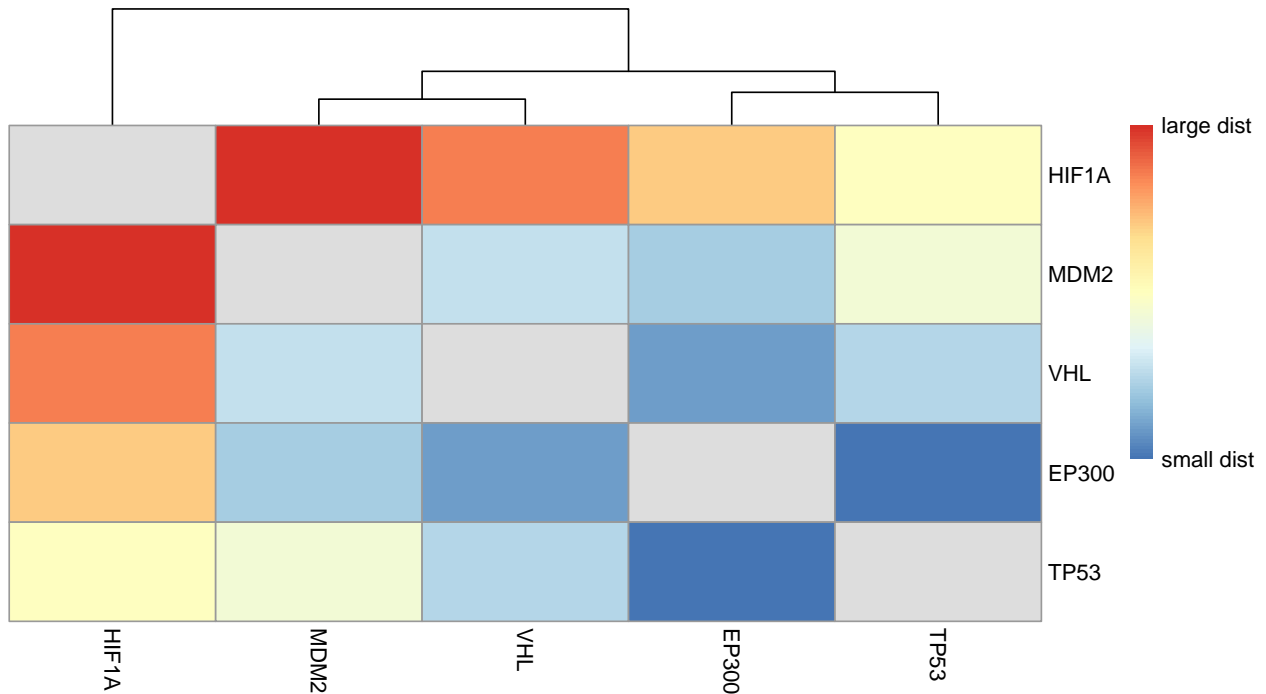
gridExtra::grid.arrange(grobs=list(p1$gtable, p2$gtable),
                        nrow = 2 , labels=c('A', 'B'))

```

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

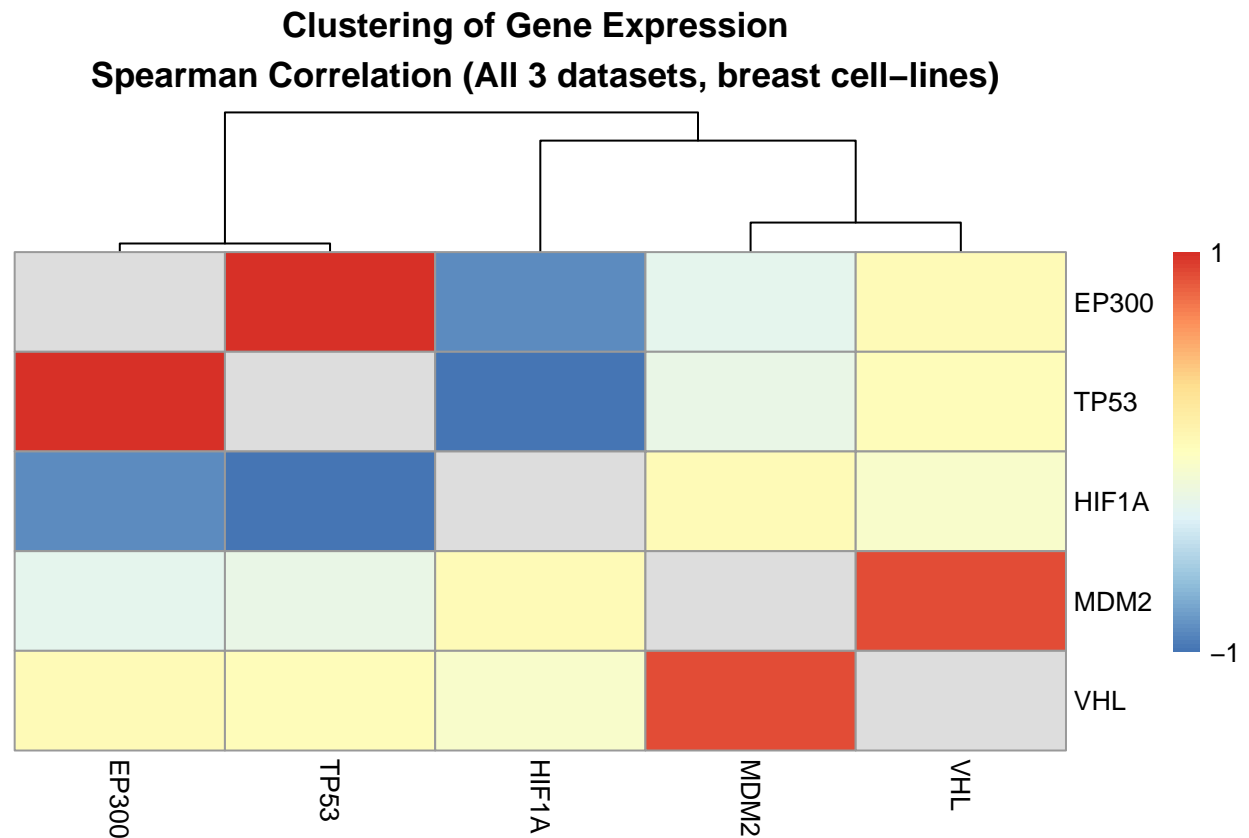


### Clustering of Gene Expression on Normoxia Euclidian Distance (All 3 datasets, breast cell-lines)



```
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)

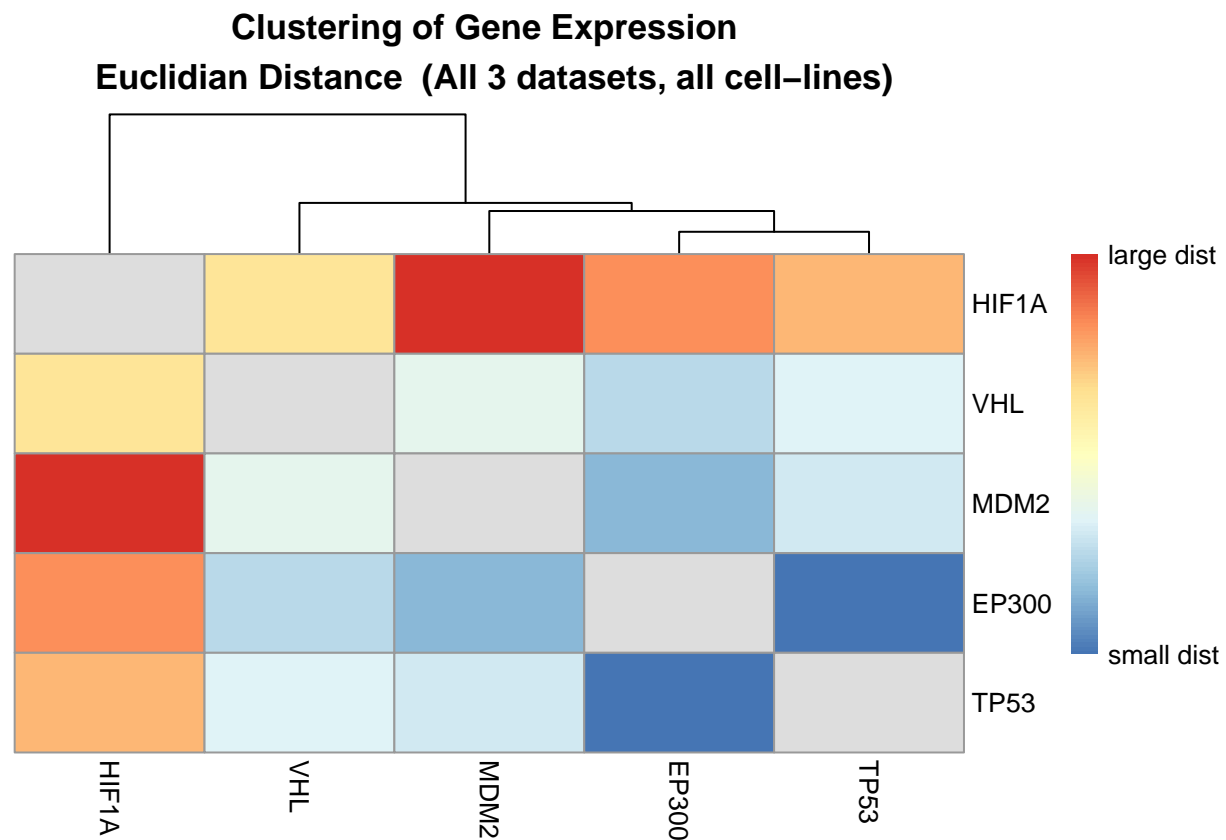
- E-GEOD-18494 2012 / HepG2, U87, MDA-MB231 / hepatoma, glioma, breast / 4h, 8h, 12h / microarray
- GSE41491 2012 / DU145, HT29, MCF7 / prostate, colon, breast / 1h, 2h, 4h, 8h, 12h, 16h, 24h / microarray
- GSE47534 2014 / MCF7 / breast / normoxia, 16h, 32h, 48h / mRNA

```
# 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.EGEOD18494.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)")
```



```
#-----
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, all c
silent=T)

#-----

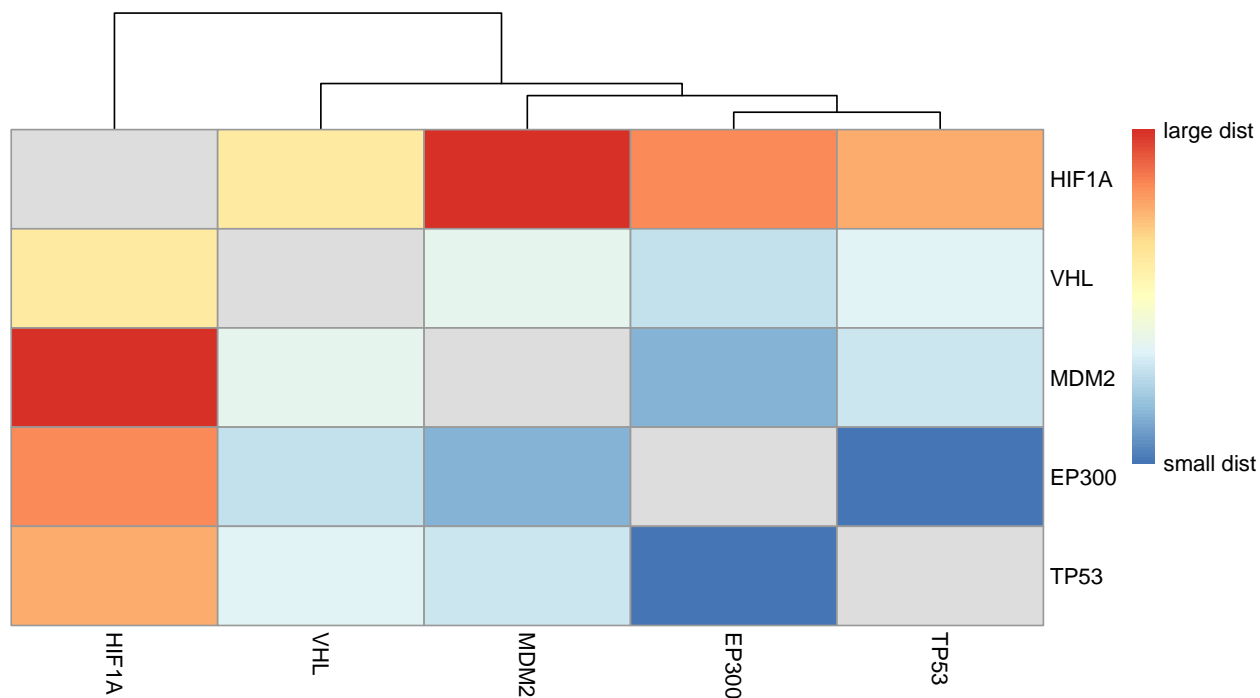
expr.row <- (colnames(expr.all.hif) %in% data.GSE47533$codes[data.GSE47533$condition == "Norm"])
dists <- as.matrix(dist(expr.all.hif[expr.row], method = "euclidean"))
rownames(dists) <- rownames(expr.all.hif[expr.row])
colnames(dists) <- rownames(expr.all.hif[expr.row])
diag(dists) <- NA

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)

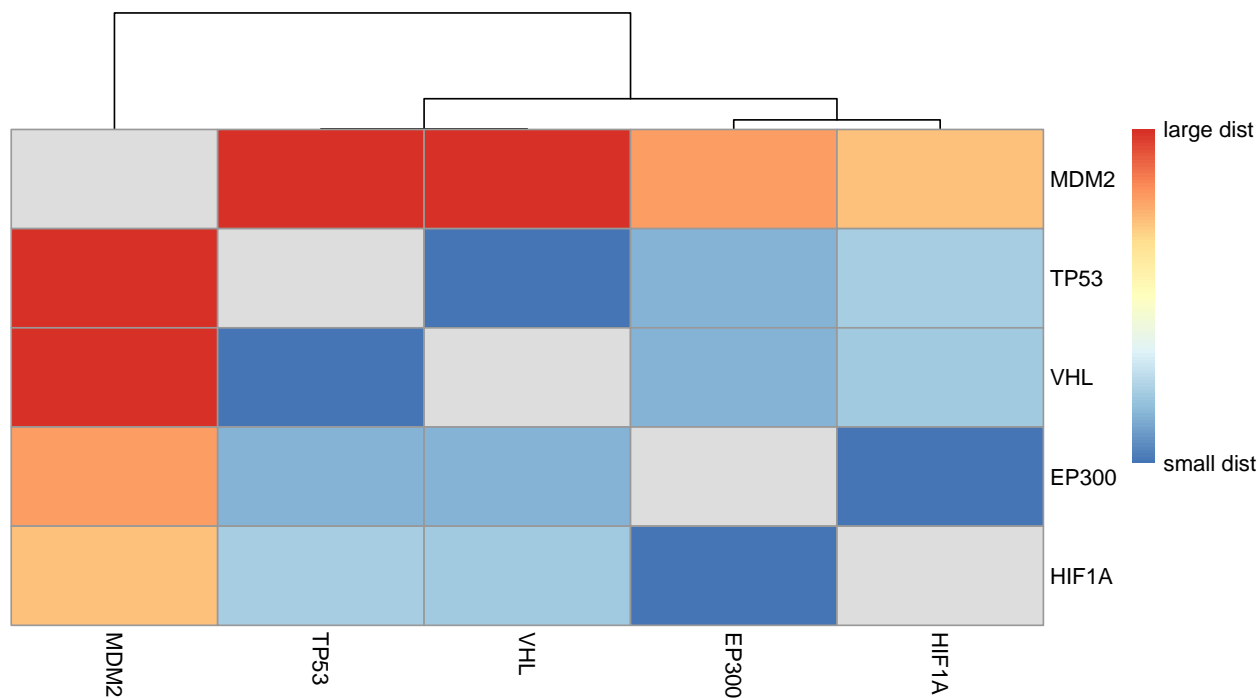
gridExtra::grid.arrange(grobs=list(p1$gtable, p2$gtable),
                        nrow = 2 , labels=c('A', 'B'))

```

### 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,
  display_numbers = T,
  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)
```

