

Merging 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))

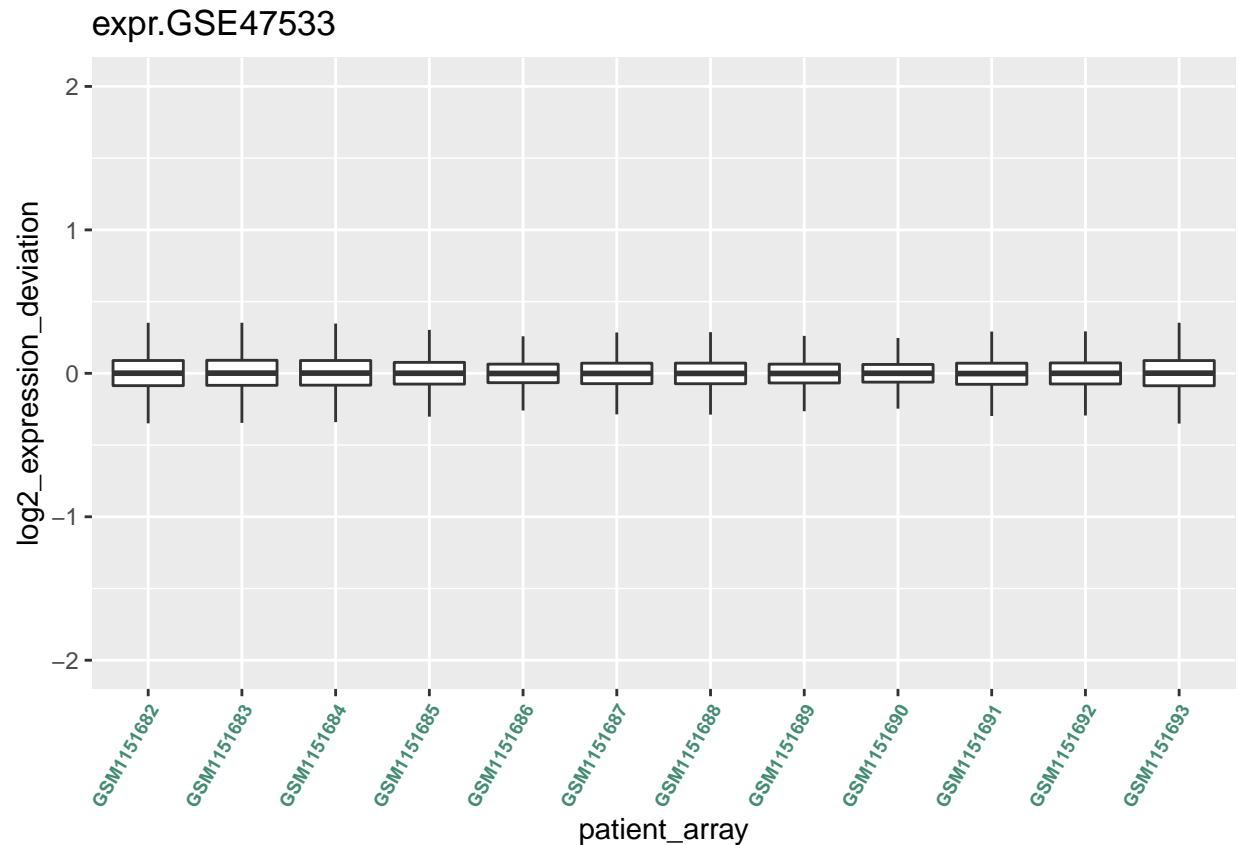
RLE_data <- sweep(expr.GSE47533, 1, row_medians_assayData)

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).

```

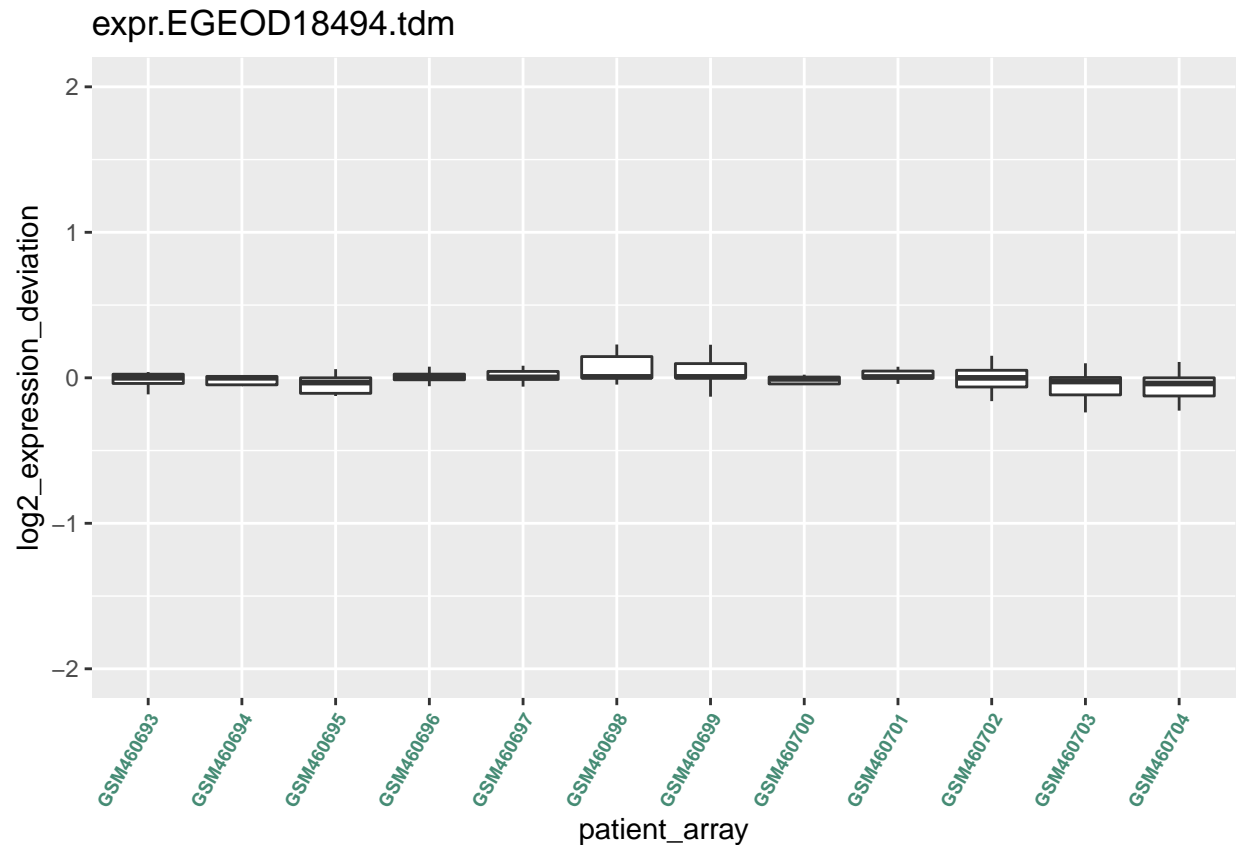


```
row_medians_assayData <-
  Biobase::rowMedians(as.matrix(expr.EGEOD18494.tdm))

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"))
```

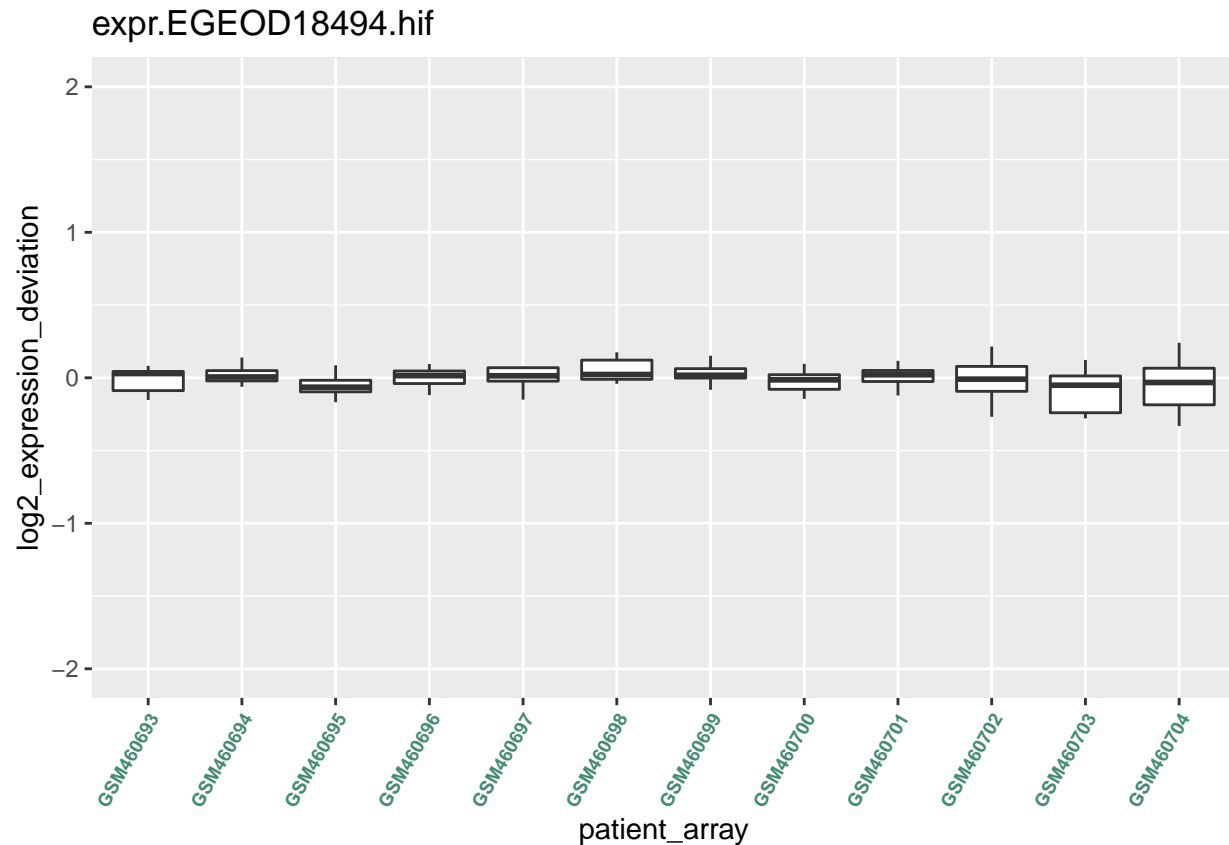


```
row_medians_assayData <-
  Biobase::rowMedians(as.matrix(expr.EGEOD18494.hif))

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]
```

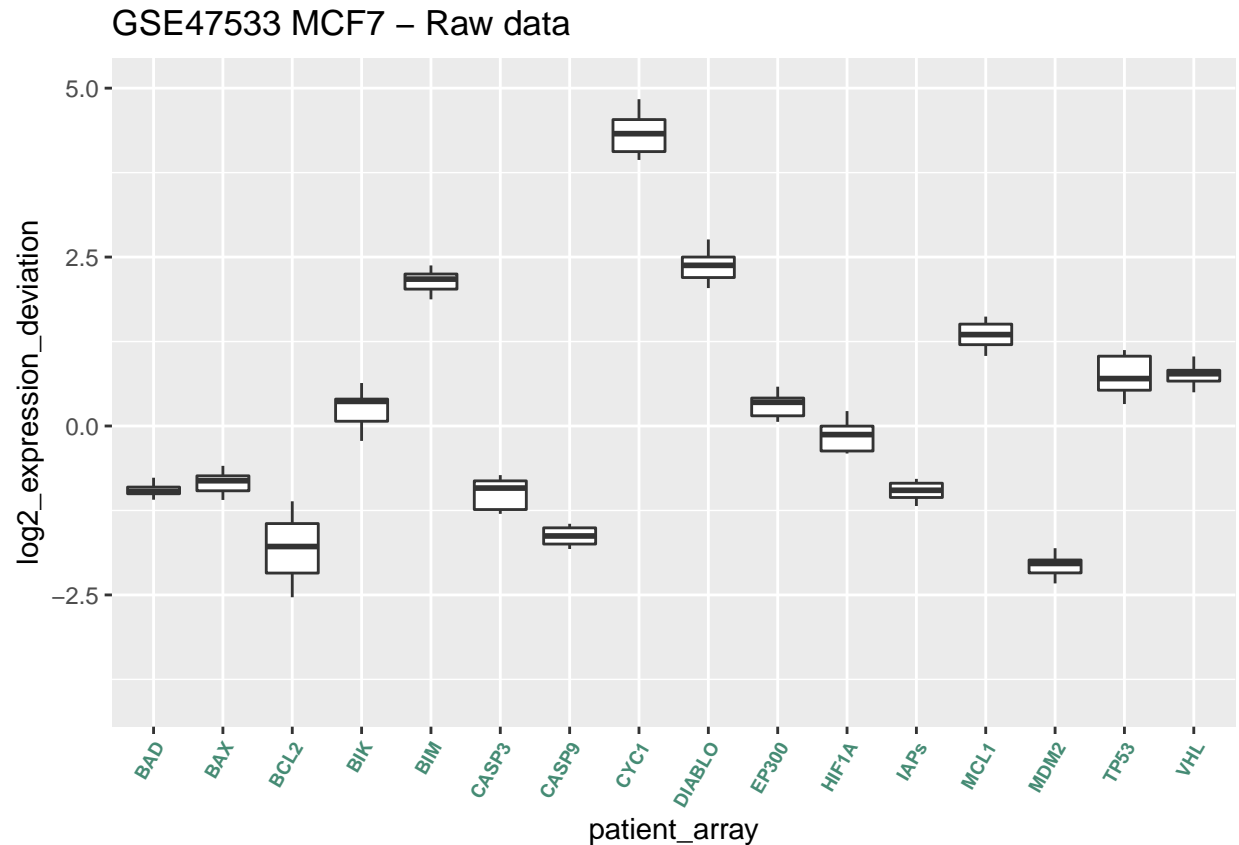
```
row_medians_assayData <-  
  Biobase::rowMedians(as.matrix(t(expr.GSE47533.hif)))
```

```
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"))
```



```

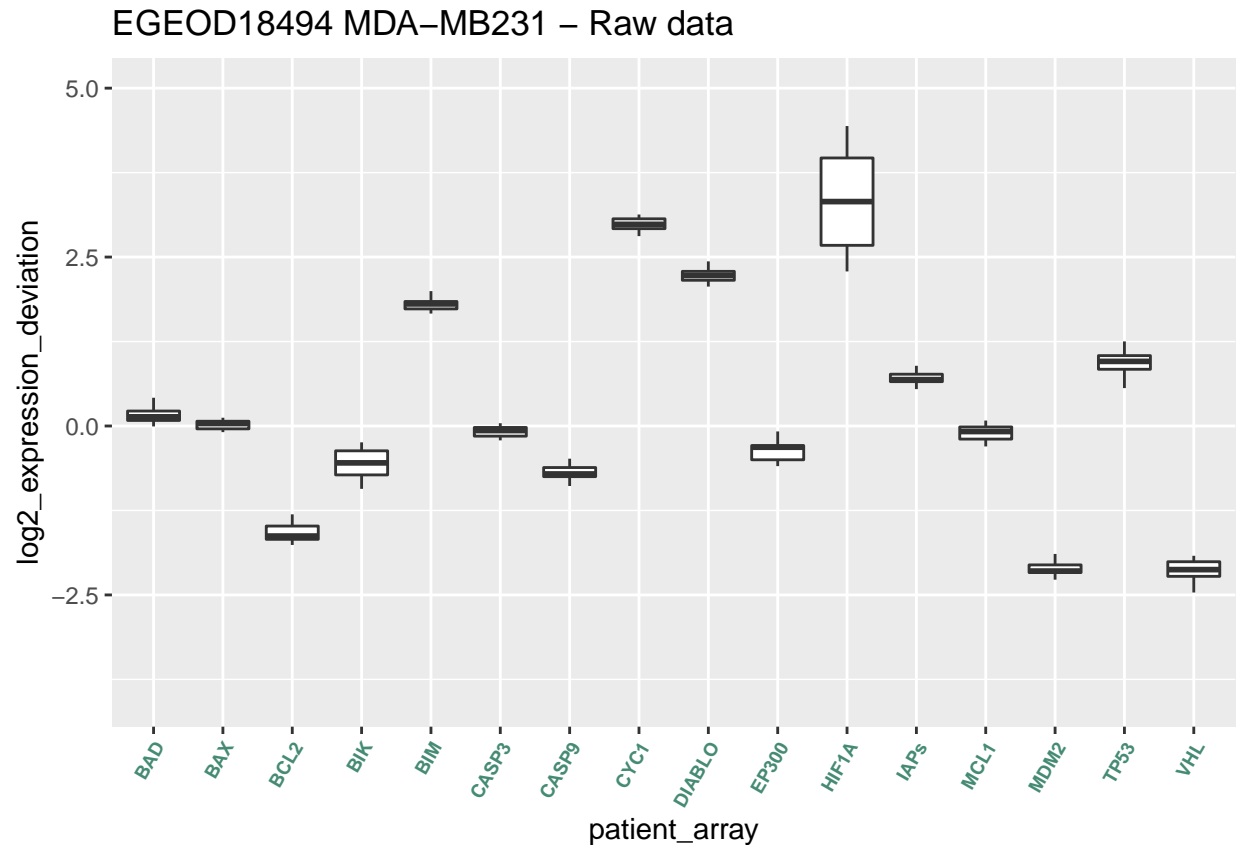
row_medians_assayData <-
  Biobase::rowMedians(as.matrix(t(expr.EGEOD18494.hif)))

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"))

```



```

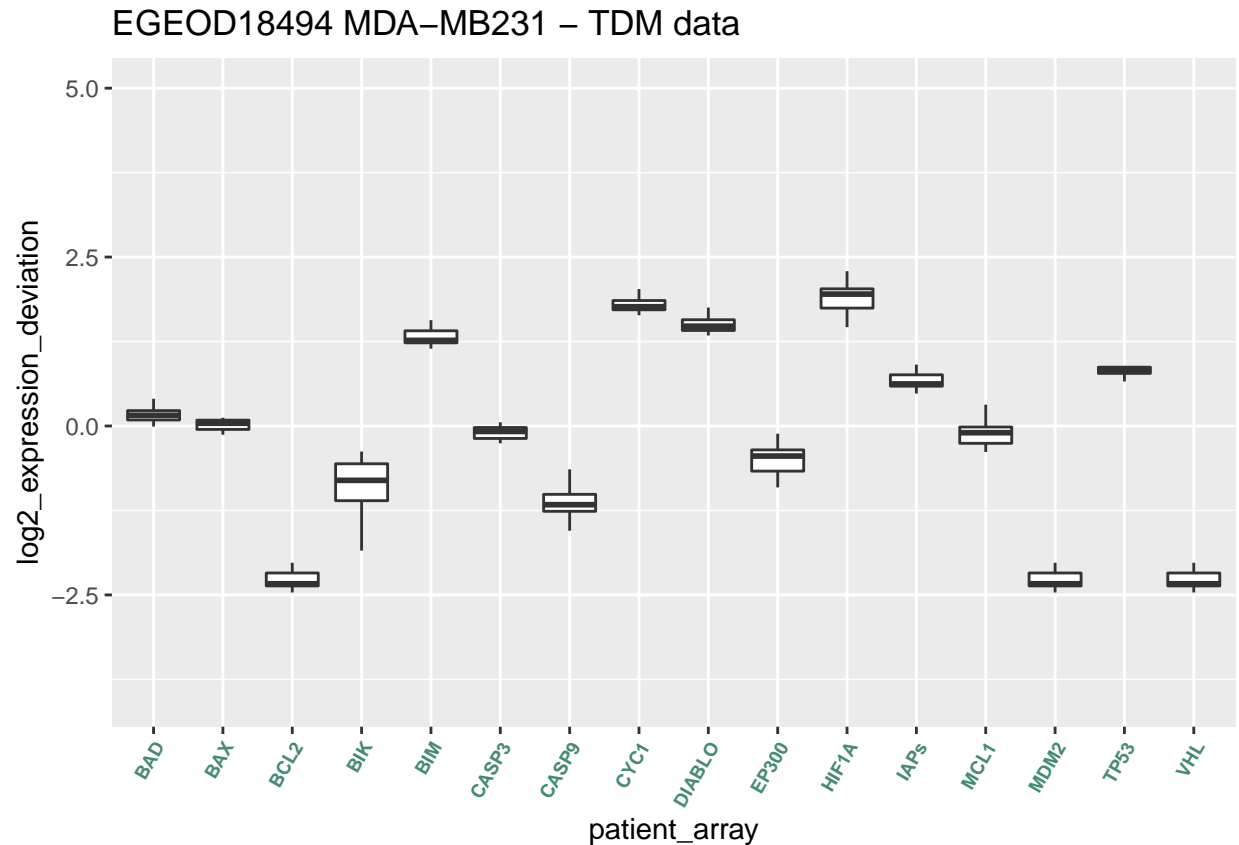
row_medians_assayData <-
  Biobase::rowMedians(as.matrix(t(expr.EGEOD18494.tdm)))

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	Norm.3	Norm.4	Norm.5	Norm.6	Norm.7	Norm.8	Norm.9	Norm.10	Norm.11	Norm.12	Norm.13	Norm.14	Norm.15	Norm.16	Norm.17	Norm.18	Norm.19	Norm.20	Norm.21	Norm.22	Norm.23	Norm.24	Norm.25	Norm.26	Norm.27	Norm.28	Norm.29	Norm.30	Norm.31	Norm.32	Norm.33	Norm.34	Norm.35	Norm.36	Norm.37	Norm.38	Norm.39	Norm.40	Norm.41	Norm.42	Norm.43	Norm.44	Norm.45	Norm.46	Norm.47	Norm.48	Norm.49	Norm.50	Norm.51	Norm.52	Norm.53	Norm.54	Norm.55	Norm.56	Norm.57	Norm.58	Norm.59	Norm.60	Norm.61	Norm.62	Norm.63	Norm.64	Norm.65	Norm.66	Norm.67	Norm.68	Norm.69	Norm.70	Norm.71	Norm.72	Norm.73	Norm.74	Norm.75	Norm.76	Norm.77	Norm.78	Norm.79	Norm.80	Norm.81	Norm.82	Norm.83	Norm.84	Norm.85	Norm.86	Norm.87	Norm.88	Norm.89	Norm.90	Norm.91	Norm.92	Norm.93	Norm.94	Norm.95	Norm.96	Norm.97	Norm.98	Norm.99	Norm.100	Norm.101	Norm.102	Norm.103	Norm.104	Norm.105	Norm.106	Norm.107	Norm.108	Norm.109	Norm.110	Norm.111	Norm.112	Norm.113	Norm.114	Norm.115	Norm.116	Norm.117	Norm.118	Norm.119	Norm.120	Norm.121	Norm.122	Norm.123	Norm.124	Norm.125	Norm.126	Norm.127	Norm.128	Norm.129	Norm.130	Norm.131	Norm.132	Norm.133	Norm.134	Norm.135	Norm.136	Norm.137	Norm.138	Norm.139	Norm.140	Norm.141	Norm.142	Norm.143	Norm.144	Norm.145	Norm.146	Norm.147	Norm.148	Norm.149	Norm.150	Norm.151	Norm.152	Norm.153	Norm.154	Norm.155	Norm.156	Norm.157	Norm.158	Norm.159	Norm.160	Norm.161	Norm.162	Norm.163	Norm.164	Norm.165	Norm.166	Norm.167	Norm.168	Norm.169	Norm.170	Norm.171	Norm.172	Norm.173	Norm.174	Norm.175	Norm.176	Norm.177	Norm.178	Norm.179	Norm.180	Norm.181	Norm.182	Norm.183	Norm.184	Norm.185	Norm.186	Norm.187	Norm.188	Norm.189	Norm.190	Norm.191	Norm.192	Norm.193	Norm.194	Norm.195	Norm.196	Norm.197	Norm.198	Norm.199	Norm.200	Norm.201	Norm.202	Norm.203	Norm.204	Norm.205	Norm.206	Norm.207	Norm.208	Norm.209	Norm.210	Norm.211	Norm.212	Norm.213	Norm.214	Norm.215	Norm.216	Norm.217	Norm.218	Norm.219	Norm.220	Norm.221	Norm.222	Norm.223	Norm.224	Norm.225	Norm.226	Norm.227	Norm.228	Norm.229	Norm.230	Norm.231	Norm.232	Norm.233	Norm.234	Norm.235	Norm.236	Norm.237	Norm.238	Norm.239	Norm.240	Norm.241	Norm.242	Norm.243	Norm.244	Norm.245	Norm.246	Norm.247	Norm.248	Norm.249	Norm.250	Norm.251	Norm.252	Norm.253	Norm.254	Norm.255	Norm.256	Norm.257	Norm.258	Norm.259	Norm.260	Norm.261	Norm.262	Norm.263	Norm.264	Norm.265	Norm.266	Norm.267	Norm.268	Norm.269	Norm.270	Norm.271	Norm.272	Norm.273	Norm.274	Norm.275	Norm.276	Norm.277	Norm.278	Norm.279	Norm.280	Norm.281	Norm.282	Norm.283	Norm.284	Norm.285	Norm.286	Norm.287	Norm.288	Norm.289	Norm.290	Norm.291	Norm.292	Norm.293	Norm.294	Norm.295	Norm.296	Norm.297	Norm.298	Norm.299	Norm.300	Norm.301	Norm.302	Norm.303	Norm.304	Norm.305	Norm.306	Norm.307	Norm.308	Norm.309	Norm.310	Norm.311	Norm.312	Norm.313	Norm.314	Norm.315	Norm.316	Norm.317	Norm.318	Norm.319	Norm.320	Norm.321	Norm.322	Norm.323	Norm.324	Norm.325	Norm.326	Norm.327	Norm.328	Norm.329	Norm.330	Norm.331	Norm.332	Norm.333	Norm.334	Norm.335	Norm.336	Norm.337	Norm.338	Norm.339	Norm.340	Norm.341	Norm.342	Norm.343	Norm.344	Norm.345	Norm.346	Norm.347	Norm.348	Norm.349	Norm.350	Norm.351	Norm.352	Norm.353	Norm.354	Norm.355	Norm.356	Norm.357	Norm.358	Norm.359	Norm.360	Norm.361	Norm.362	Norm.363	Norm.364	Norm.365	Norm.366	Norm.367	Norm.368	Norm.369	Norm.370	Norm.371	Norm.372	Norm.373	Norm.374	Norm.375	Norm.376	Norm.377	Norm.378	Norm.379	Norm.380
BAD	0	0	0	0	0	0	0	0	0	1	1	1	7.925009	791	BAD																																																																																																																																																																																																																																																																																																																																																																														
BAX	0	1	1	0	0	0	1	0	0	1	1	1	7.978194	000	BAX																																																																																																																																																																																																																																																																																																																																																																														
BCL2	1	1	1	0	0	0	0	0	0	0	0	0	7.272830	930	BCL2																																																																																																																																																																																																																																																																																																																																																																														
BIK	0	0	0	1	1	1	1	1	1	1	1	1	8.755742	001	BIK																																																																																																																																																																																																																																																																																																																																																																														
BIM	1	1	0	1	0	0	1	0	1	1	0	1	10.921166	962	BIM																																																																																																																																																																																																																																																																																																																																																																														
CASP3	1	1	1	0	0	1	1	0	0	0	0	0	7.799389	001	CASP3																																																																																																																																																																																																																																																																																																																																																																														

```
# 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

names(expr.EGEOD18494.hif.bin) <- c(paste0(substr(data.EGEOD18494$condition[row],1,4),".", data.EGEOD18494$

#names(expr.EGEOD18494.hif.bin) <- c(paste0(substr(data.EGEOD18494$condition[row],1,4),".", data.EGEOD18494$

# head(expr.EGEOD18494.hif.bin) %>%
#   knitr::kable(.)
```

```

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)

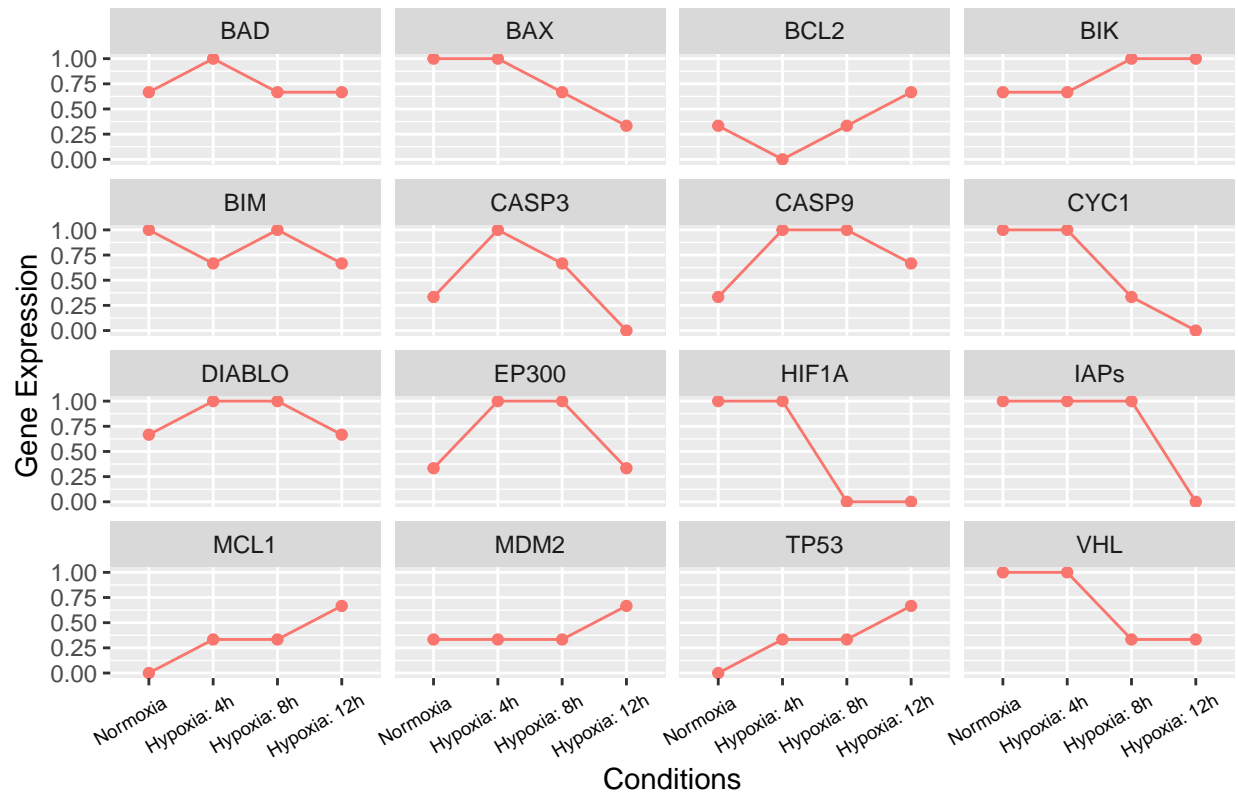
# hif.symbols <- c("HIF1A", "TP53", "MDM2", "VHL", "EP300", "TMBIM1", "TMBIM4", "TMBIM6", "BAD", "BIK", "MCL1", "FOXO3", "FOXO1", "FOXO2", "FOXO4", "FOXO6", "FOXO7", "FOXO8", "FOXO9", "FOXO10", "FOXO11", "FOXO12", "FOXO13", "FOXO14", "FOXO15", "FOXO16", "FOXO17", "FOXO18", "FOXO19", "FOXO20", "FOXO21", "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

```

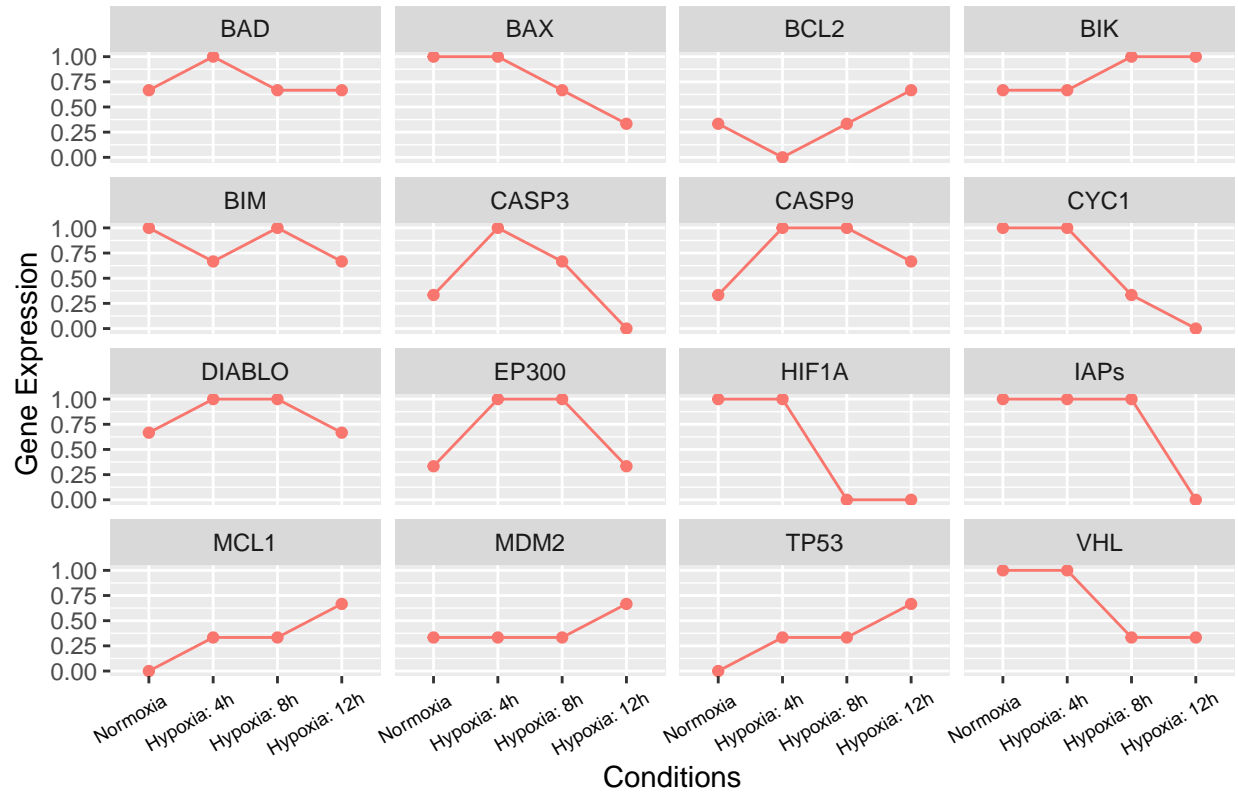
EGEOD18494 MDA-MB231 – Raw data



```
par(mfrow = c(1,2))
```

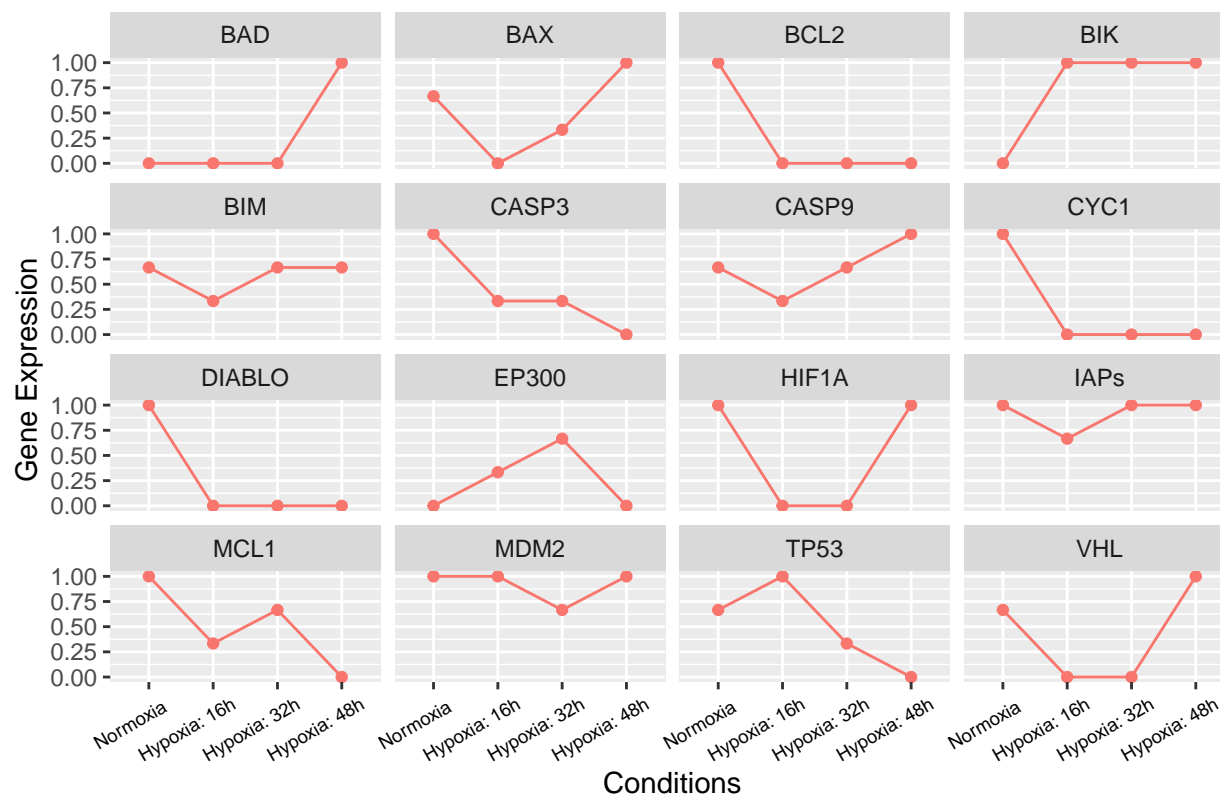
```
p.MDA
```

EGEOD18494 MDA-MB231 – Raw data



p.MCF7

GSE47533 MCF7 – Raw data



```
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) <- colnames(expr.EGEOD18494.hif)
row.names(annotation_for_heatmap) <- paste0(substr(data.EGEOD18494$condition[row], 1, 4), ".", data.EGEOD18494$cell_line[row])

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

```

#rownames(dists) <- colnames(expr.EGEOD18494.hif)

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

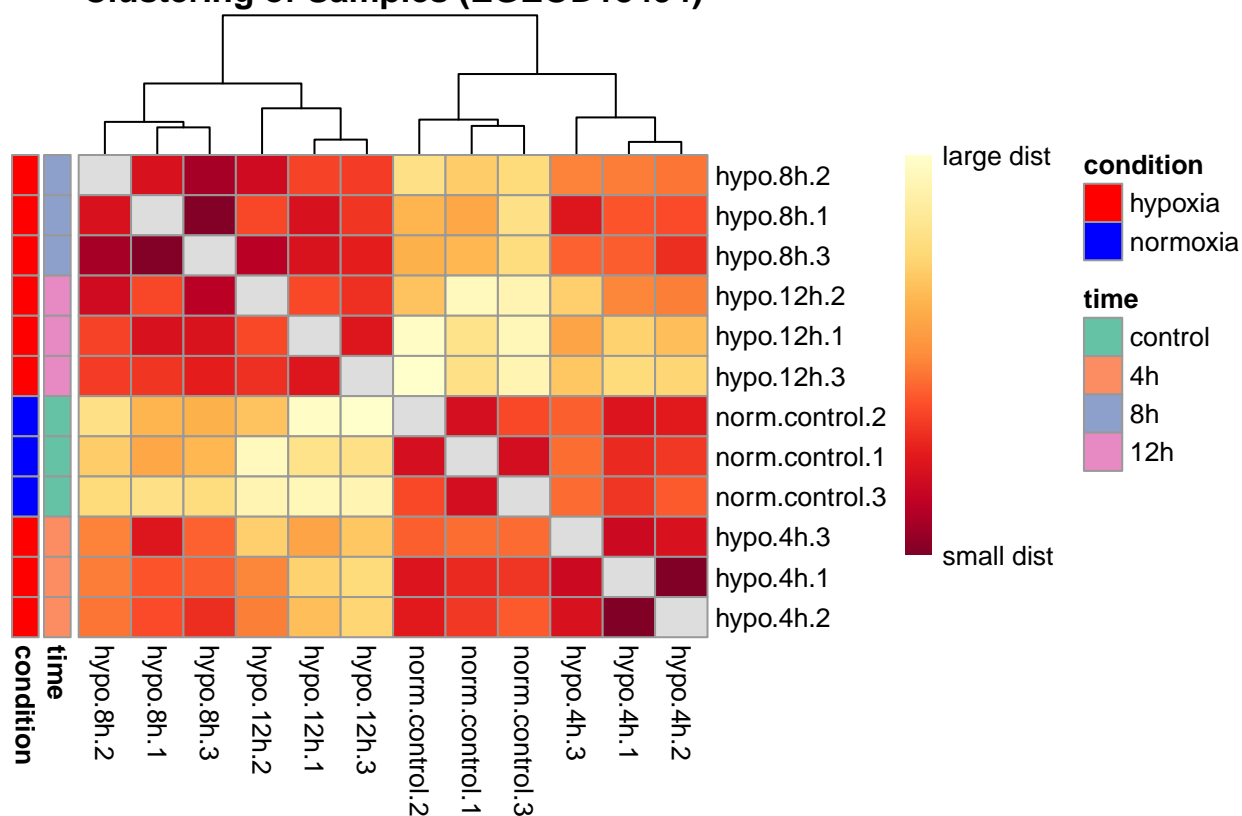
## $time
## [1] "#66C2A5" "#FC8D62" "#8DA0CB" "#E78AC3"
##
## $condition
## [1] "red" "blue"

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)")

```


Clustering of Samples (EGEOD18494)



```
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) <- colnames(expr.EGEOD18494.hif)

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

rownames(dists) <- rownames(expr.EGEOD18494.hif)

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

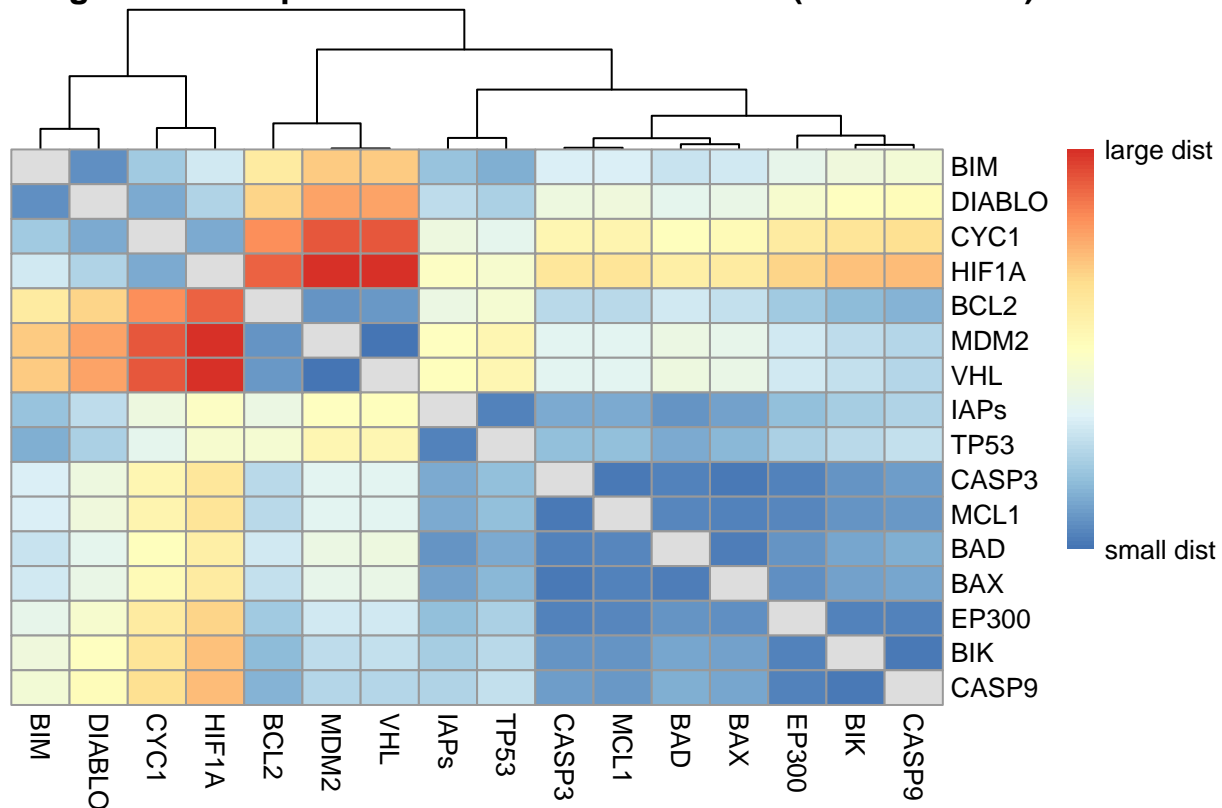
```

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

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 - Euclidian Distance (EGEOD18494)")

```

Clustering of Gene Expression - Euclidian Distance (EGEOD18494)



```

#-----
row <- data.EGEOD18494$cell_line == "MDA-MB231 breast cancer" & data.EGEOD18494$condition == "hypoxia"
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

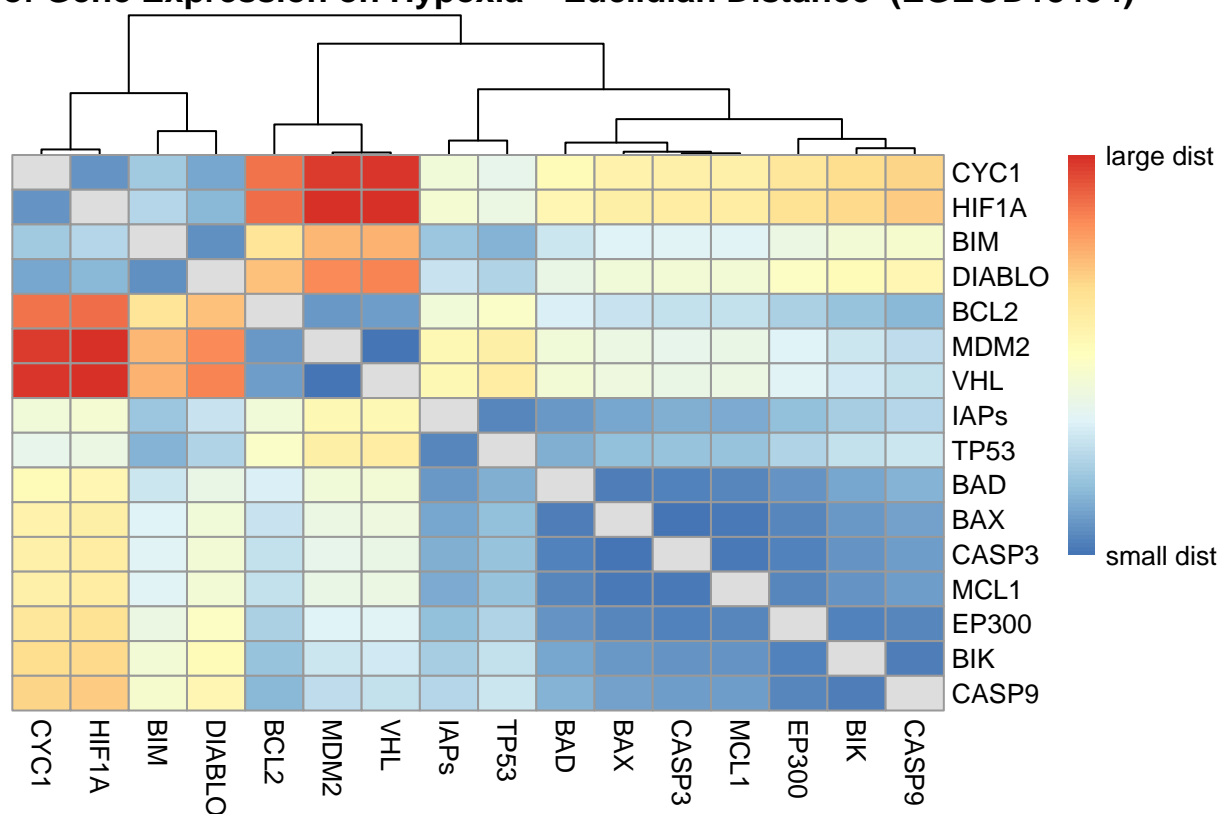
## $time
## [1] "#66C2A5" "#FC8D62" "#8DA0CB" "#E78AC3"
##
## $condition
## [1] "#EF8A62" "#67A9CF"

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

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 Hypoxia - Euclidian Distance (EGEOD18494)")

```

of Gene Expression on Hypoxia – Euclidian Distance (EGEOD18494)



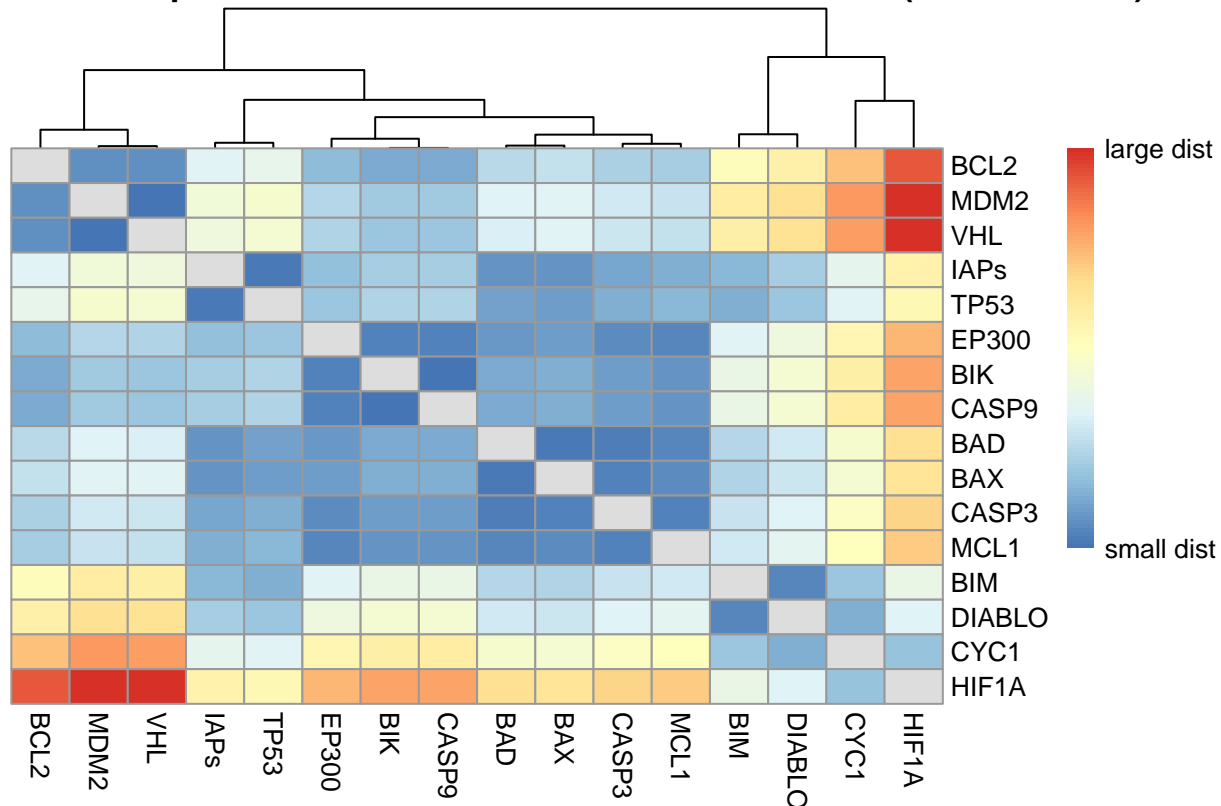
```
#-----
row <- data.EGEOD18494$cell_line == "MDA-MB231 breast cancer" & data.EGEOD18494$condition == "normoxia"
annotation_for_heatmap <- droplevels(data.frame(time = data.EGEOD18494$time[row], condition = data.EGEOD18494$condition[row]))
expr.row <- (colnames(expr.EGEOD18494.hif) %in% data.EGEOD18494$codes[data.EGEOD18494$cell_line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

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 on Normoxia - Euclidian Distance (EGEOD18494)")
```

of Gene Expression on Normoxia – Euclidian Distance (EGEOD18494)



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

```

#dists <- as.matrix(dist(expr.EGEOD18494.hif, method = "pearson"))
dists <- cor(t(expr.EGEOD18494.hif), use = "pairwise.complete.obs", method = "pearson")

rownames(dists) <- rownames(expr.EGEOD18494.hif)

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

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("-1", "1")),
  main = "Clustering of Gene Expression - Pearson Correlation (EGEOD18494)")

```

ustering of Gene Expression – Pearson Correlation (EGEOD18494)

