RBPs and ER, Hypoxia and Unfolded Protein Response

Manasa Ramakrishna, Robert Harvey

19 July, 2019

Contents

Startup

We start by installing and loading the libraries required for our analysis. Additionally, tell R where you are running your program by setting your working directory as shown below using the variable 'wd'. We will use this later on. Also make your input and output directories (indir/outdir) as shown below.

```r
suppressMessages(library(reshape2))
suppressMessages(library(ggplot2))
suppressMessages(library(ggsci))
suppressMessages(library(dplyr))
suppressMessages(library(MSnbase))
suppressMessages(library(ggbiplot))
suppressMessages(library(pRoloc))
suppressWarnings(library(mygene))
suppressWarnings(library(data.table))
suppressWarnings(library(patchwork))
suppressWarnings(library(venn))
suppressWarnings(library(gplots))
suppressWarnings(library(RColorBrewer))

# Setting working directories Note: Change the next
# line of card to point to your working directory
wd = "~/Documents/Work/TTT/15_UPR-Hypoxia-RHarvey/"
setwd(wd)
```

```r
# getwd()

# Declaring input and output directories
indir = paste(wd, "Input", sep = "/")
outdir = paste(wd, "Output", sep = "/")
plotdir = paste(wd, "Plots", sep = "/")

# If output and plots directory exist, clear them
# out and start afresh
if (exists(outdir)) {
    system(paste0("rm -r", outdir))
}
if (exists(plotdir)) {
    system(paste0("rm -r", plotdir))
}

dir.create(outdir)
dir.create(plotdir)
```

Introduction

The main aims of this analysis are to identify and classify known RBPs into those that are involved in Hypoxia, Endoplasmic Reticulim and also in eliciting an Unfolded Protein Respose (UPR). To do this, I will be using two sets of proteins as known RBPs - (1) From Trendel et al that have used RBP data from RNA Interactome Capture in HeLa, HEK293 and MCF7 cells (2) List of RBPs from Queiroz et al that have used Trizol -based OOPS to target RBPs.

01. Reading in data

1a. Reading in RNA Interactome Capture Data

I have obtained these data and stored them in text files in the "Input" folder and will bbe reading them and reformatting them for further analysis in this section.

```r
# 1. List of known RBPs across cell lines in the
# XRNAX paper (Table S2)
xrnax = read.delim(paste(indir, "xrnax-genelist.txt",
    sep = "/"), sep = "\t", header = T)

# Check how many are common to the cell lines in
# the XRNAX paper
xrnax %>% dplyr::select(MCF7.RBP:ihRBP) %>% apply(2,
    table, useNA = "always")
```

```
##              MCF7.RBP HEK293.RBP HeLa.RBP ihRBP
## non-poly(A)       617        698      565   775
## poly(A)           590        659      674   978
## <NA>             1276       1126     1244   730
```

```r
# Filter to only keep those that have been found by
# at least one cell line
xrnax.rbps = xrnax %>% dplyr::filter(!is.na(MCF7.RBP) |
    !is.na(HEK293.RBP) | !is.na(HeLa.RBP)) %>% dplyr::select(c(Uniprot.ID:Protein.name,
    MCF7.RBP:ihRBP))
# rownames(xrnax.rbps) = xrnax.rbps$Uniprot.ID
```

Note that in the protein lists, proteins have been classified as those that bind polyA RNA and those that bind non-polyA RNA. For now, i will leave all of them in but can change this later on.

1b. Reading in OOPS-based RBP data

```
# 2. List of RBPs from SILAC experiments using OOPS
# with 2 or more wash steps and CL/NC Ratio >1 i.e
# more than two-fold change in abundance (Table S1)
oops = read.delim(paste(indir, "oops-genelist.txt",
    sep = "/"), sep = "\t", header = T, row.names = NULL)
oops.rbps = oops %>% dplyr::filter(step > 1 & CL_NC_Ratio >=
    1) %>% dplyr::select(master_protein, step, RBP_glyco) %>%
    dplyr::filter(RBP_glyco != "Novel RBP, Glycoprotein") %>%
    dplyr::distinct(master_protein, .keep_all = TRUE)
rownames(oops.rbps) = oops.rbps$master_protein

# Merge both lists n = 2807
merged.rbp = union(xrnax$Uniprot.ID, oops$master_protein)
write.table(data.frame(merged.rbp), paste(outdir, "RBP-list-for-uniprot.txt",
    sep = "/"), sep = "\t", row.names = F, quote = F)
```

1c. Subset proteins of interest from RBP list

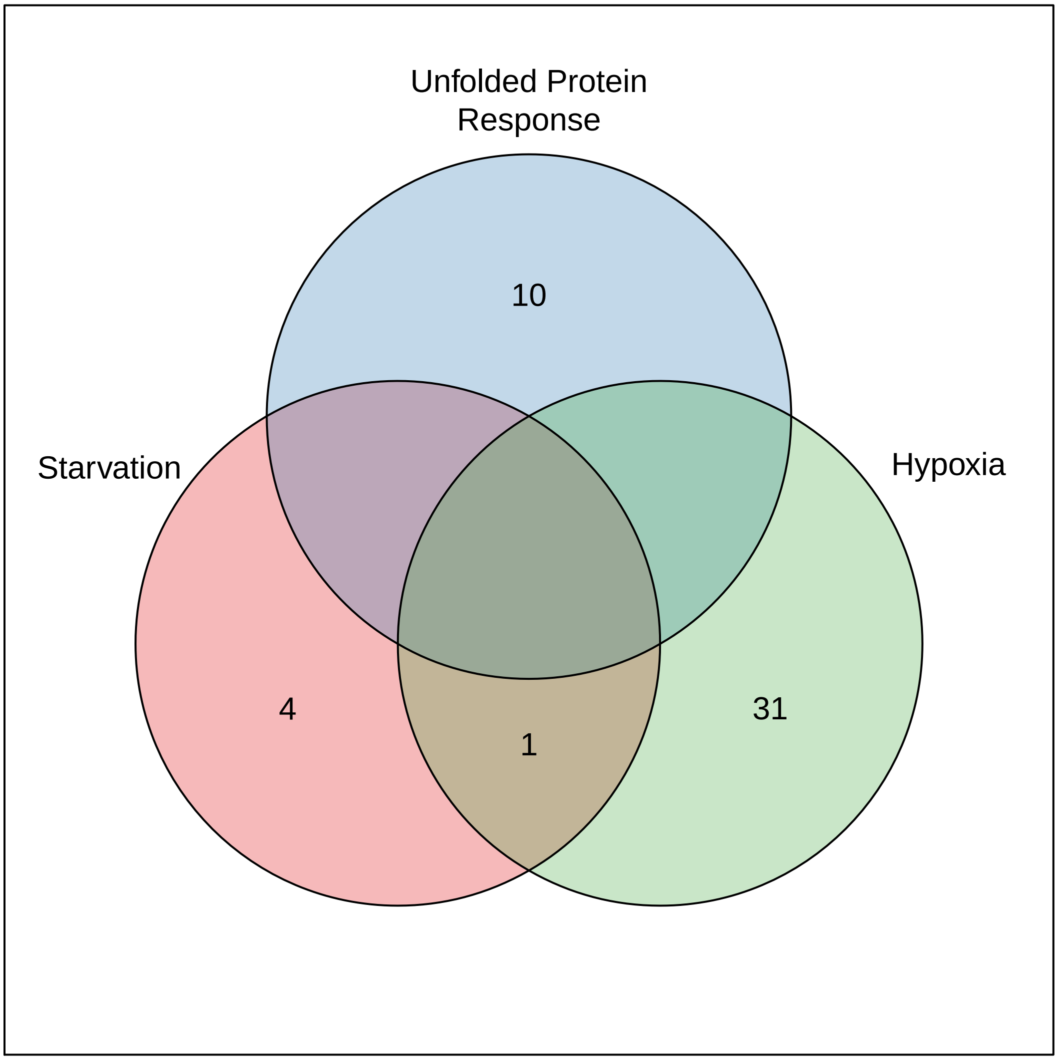
Here we are focussing on those RBPs that have a GO term relating them to UPR, Hypoxia or Endoplasmic Reticulim.

```
# Annotate OOPS/RIC RBPs with GO terms from Uniprot
go.rbp = read.delim(paste(indir, "human-rbp-uniprot.txt",
    sep = "/"), sep = "\t", stringsAsFactors = F, header = T)
colnames(go.rbp)[6] = "Gene.name"

# Hypoxia response
hypox = go.rbp[with(go.rbp, grepl("GO:0001666", paste(Gene.ontology..biological.process.,
    Gene.ontology..cellular.component., Gene.ontology..molecular.function.))),
    ]

starv = go.rbp[with(go.rbp, grepl("GO:0042594", paste(Gene.ontology..biological.process.,
    Gene.ontology..cellular.component., Gene.ontology..molecular.function.))),
    ]

upr = go.rbp[with(go.rbp, grepl("GO:0030968", paste(Gene.ontology..biological.process.,
    Gene.ontology..cellular.component., Gene.ontology..molecular.function.))),
    ]

# Concatenate genes into a list for output
all.rbps.out = rbind(cbind(hypox, Category = "Hypoxia-RBP"),
    cbind(upr, Category = "UPR-RBP"), cbind(starv,
        Category = "Starvation-RBP"))
```

```
write.table(all.rbps.out, paste(outdir, "UPR-Hypoxia-Starvation-RBPs.txt",
    sep = "/"), sep = "\t", row.names = F, quote = F)

# Plot intersect
input.v = list(Starvation = starv$Gene.name, `Unfolded Protein\nResponse` = upr$Gene.name,
    Hypoxia = hypox$Gene.name)
pdf(paste(plotdir, "UPR-Hypoxia-ER-RBPs-Venn.pdf",
    sep = "/"))
v = venn::venn(input.v, zcolor = brewer.pal(n = 3,
    name = "Set1"), cexil = 1, cexsn = 1)
dev.off()
```

```
## pdf
##   2
```

```
venn::venn(input.v, zcolor = brewer.pal(n = 3, name = "Set1"),
    cexil = 1, cexsn = 1)
```



This brings us to the end of the first part of the analysis which was all about known RBPs and how many of those are involved in the Unfolded Protein Response.

02. Cancer Cell Fitness genes

The analysis now focuses on the paper that uses CRISPR screens to identify genes that are required for cancer cell fitness. Based on Supplementary table S2 in this paper, we want to find
a. The RBPs that are required for cancer cell fitness
b. Find out how many of these RBPs are involved in stress response pathways – UPR and hypoxia
c. And how many are RNA modifying enzymes?

2a. Cancer fitness genes

The data are presented per cell-line (columns) and per gene (rows) in supplementary table S2 of the paper. There are 7470 fitness genes across 326 cell lines. The table also comes with metadata classifying each cell line into its tissue type. We need to extract this into a usable format before being able to comment on the occurence of fitness RBPs by cancer type.

We extract metadata information into "metadat" and fitness information into the object "fitness" in the code below. These two dataframes will be used for downstream analysis.

```
# HT-29 was not classified in the Tissue column so
# did this manually Fitness dataframe
fitness <- read.delim(paste(indir, "Cancer-Fitness-S2.txt",
    sep = "/"), sep = "\t", header = F, stringsAsFactors = F)
dim(fitness)
```

```
## [1] 7474  326
```

```
colnames(fitness) = fitness[4, ]
rownames(fitness) = fitness$Gene.CellLine

# Metadata
metadat = data.frame(t(fitness[1:4, 2:ncol(fitness)]))
colnames(metadat) = c("Tissue", "Cancer.Type", "CMP.id",
    "Cell.line")

# Resize 'fitness' dataframe and only keep RBP rows
fitness = fitness[5:nrow(fitness), ]
dim(fitness)
```

```
## [1] 7470  326
```

```
# Reshape
all.fit = reshape2::melt(fitness, id = "Gene.CellLine",
    na.rm = T)
colnames(all.fit) = c("Gene", "Cell.line", "Fitness")
all.fit = left_join(all.fit, metadat)
head(all.fit)
```

```
##      Gene Cell.line Fitness   Tissue        Cancer.Type    CMP.id
## 1    A1BG     22RV1       0 Prostate Prostate Carcinoma SIDM00499
## 2    A1CF     22RV1       0 Prostate Prostate Carcinoma SIDM00499
## 3    AAAS     22RV1       0 Prostate Prostate Carcinoma SIDM00499
## 4    AACS     22RV1       0 Prostate Prostate Carcinoma SIDM00499
## 5 AADACL2     22RV1       0 Prostate Prostate Carcinoma SIDM00499
## 6   AADAT     22RV1       0 Prostate Prostate Carcinoma SIDM00499
```

```
# Write fitness genes to a text file
write.table(as.data.frame(rownames(fitness)), sep = "\t",
    paste(outdir, "Cancer-fitness-list-for-uniprot.txt",
        sep = "/"), quote = F, row.names = F)
```

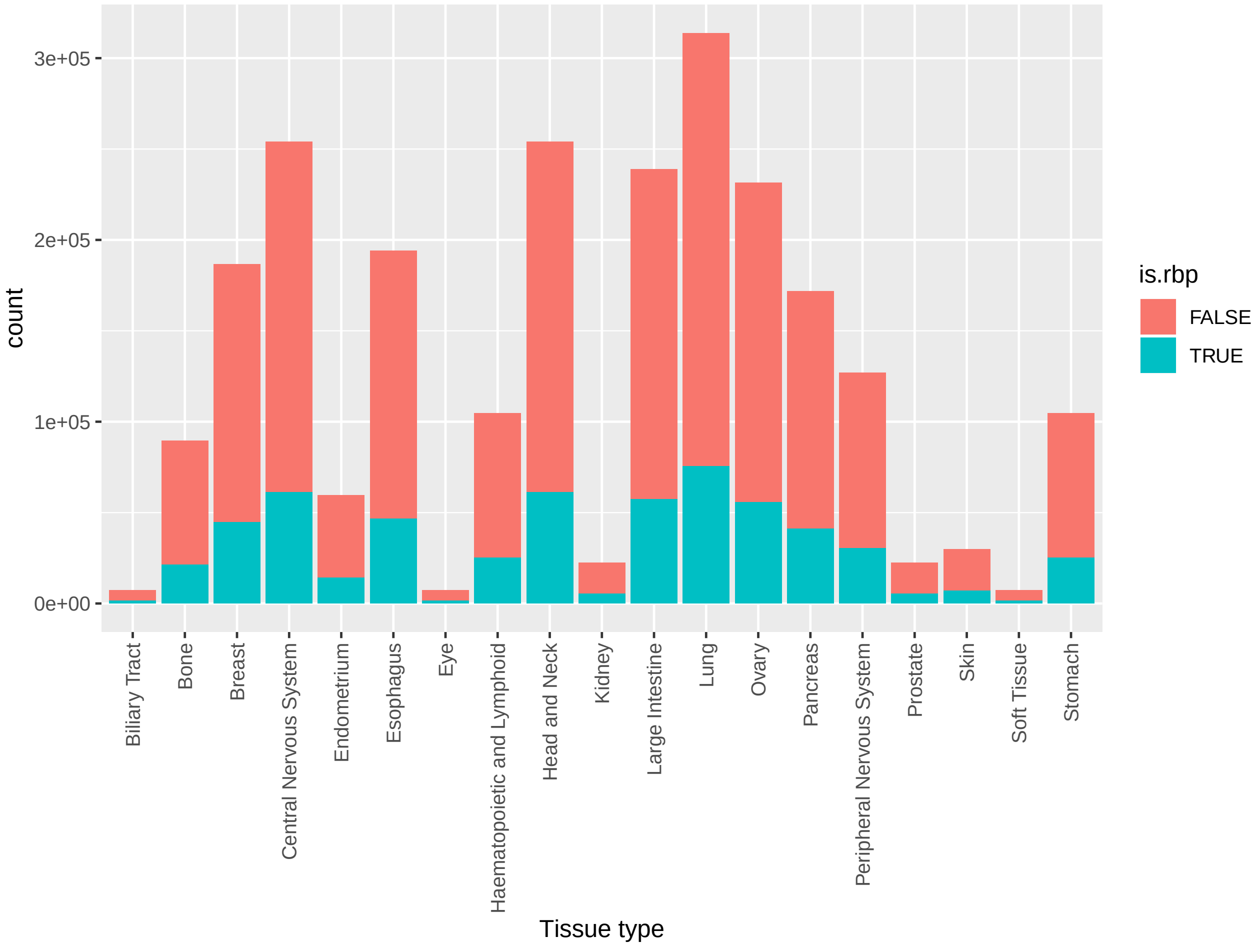
2b. Cancer fitness and RBP

The most straightforward way of workng out how many fitness genes are RBPs is to merge our list from Section 01 above to the fitness list which is what we do below.

```
# Add RBP information
all.fit = left_join(all.fit, go.rbp, by = c(Gene = "Gene.name"))
all.fit$is.rbp = TRUE
all.fit$is.rbp[which(is.na(all.fit$Entry))] = FALSE
head(all.fit)
```

```
##      Gene Cell.line Fitness   Tissue        Cancer.Type    CMP.id  Entry
## 1    A1BG     22RV1       0 Prostate Prostate Carcinoma SIDM00499   <NA>
## 2    A1CF     22RV1       0 Prostate Prostate Carcinoma SIDM00499   <NA>
## 3    AAAS     22RV1       0 Prostate Prostate Carcinoma SIDM00499 Q9NRG9
## 4    AACS     22RV1       0 Prostate Prostate Carcinoma SIDM00499   <NA>
## 5 AADACL2     22RV1       0 Prostate Prostate Carcinoma SIDM00499   <NA>
## 6   AADAT     22RV1       0 Prostate Prostate Carcinoma SIDM00499   <NA>
##   Entry.name      Protein.names             Organism Length
## 1       <NA>               <NA>                 <NA>     NA
## 2       <NA>               <NA>                 <NA>     NA
## 3 AAAS_HUMAN Aladin (Adracalin) Homo sapiens (Human)    546
## 4       <NA>               <NA>                 <NA>     NA
## 5       <NA>               <NA>                 <NA>     NA
## 6       <NA>               <NA>                 <NA>     NA
##
## 1
## 2
## 3 centrosome [GO:0005813]; cytosol [GO:0005829]; membrane [GO:0016020]; nuclear envelope [GO:0005635]
## 4
## 5
## 6
##
## 1
## 2
## 3 fertilization [GO:0009566]; learning [GO:0007612]; mRNA export from nucleus [GO:0006406]; nucleocy
## 4
## 5
## 6
##   Gene.ontology..molecular.function. is.rbp
## 1                               <NA>  FALSE
## 2                               <NA>  FALSE
## 3                                      TRUE
## 4                               <NA>  FALSE
## 5                               <NA>  FALSE
## 6                               <NA>  FALSE
```

```
# Plot distribution of RBPs by Tissue
ggplot(all.fit, aes(x = reorder(Tissue, -is.rbp), fill = is.rbp)) +
    geom_bar() + theme(axis.text.x = element_text(angle = 90,
    vjust = 0.5, hjust = 1)) + xlab("Tissue type")
```



Just noting here that given it is the same set of genes that have been queried across all cell-lines and only some of them are RBPs,the proportion of RBPs across all tissues and cell-lines will be identical (24%). The more important question is how many of these are also fitness genes and whether the proportion of "Fitness-RBPs" across cell-lines, tissues and cancer types varies.

03. Cancer fitness and RBPs

Our first task is to look at cancer fitness genes that are also known RBPs. We re-use the 'go.rbp' dataframe from the previous sections which consist of 2807 currently published RBPs (from interactome capture and OOPS) annotated with GO terms. Furthermore, we want to see if the number of RBPs involved in cancer fitness vary by cancer type or if there is a core set that is crucial to the survival of many cancer types.

3a. RBPs by cancer

We want to see how many fitness-RBPs there are per cancer type and also perhaps per tissue type. To do this, I take the full list of fitness genes and then filter to keep only those that are known RBPs from our 2807 list of RBPs above. Of 7470 fitness genes, we capture all

```
# Just keep those fitness genes that are also RBPs.
rbp.fitness = fitness[intersect(go.rbp$Gene.name, rownames(fitness)),
    ]

# Reshape data frame
```

```r
rbp.fit = reshape2::melt(rbp.fitness, id = "Gene.CellLine")
colnames(rbp.fit) = c("Gene", "Cell.line", "Fitness")
rbp.fit = left_join(rbp.fit, metadat)
head(rbp.fit)
```

```
##      Gene Cell.line Fitness    Tissue        Cancer.Type    CMP.id
## 1   AHNAK     22RV1       0 Prostate Prostate Carcinoma SIDM00499
## 2   MKI67     22RV1       0 Prostate Prostate Carcinoma SIDM00499
## 3 LRPPRC     22RV1       0 Prostate Prostate Carcinoma SIDM00499
## 4    PLEC     22RV1       0 Prostate Prostate Carcinoma SIDM00499
## 5 PDCD11     22RV1       0 Prostate Prostate Carcinoma SIDM00499
## 6   PRPF8     22RV1       1 Prostate Prostate Carcinoma SIDM00499
```

```r
rbp.fit.t = table(rbp.fit$Cancer.Type, rbp.fit$Fitness)/rowSums(table(rbp.fit$Cancer.Type,
    rbp.fit$Fitness))
melt.t = melt(rbp.fit.t)
colnames(melt.t) = c("Cancer.Type", "Fitness", "Percentage")
melt.t$Fitness = factor(melt.t$Fitness, levels = c(1,
    0))
melt.t$Percentage = round(melt.t$Percentage * 100,
    0)

pdf(paste(plotdir, "RBPs-by-cancer-type-and-fitness.pdf",
    sep = "/"))
p <- ggplot() + geom_bar(aes(y = Percentage, x = Cancer.Type,
    fill = Fitness), data = melt.t, stat = "identity")
p = p + theme_bw() + theme(plot.title = element_text(hjust = 0.5),
    axis.text.x = element_text(angle = 90, hjust = 0.95,
        vjust = 0.2))
p = p + scale_fill_manual(values = c("maroon", "lightblue")) +
    ggtitle("RNA Binding proteins by Cancer Type and Fitness Score")
print(p)
dev.off()
```

```
## pdf
##   2
```

```r
# Numbers of cell lines in each cancer type - need
# to normalise for that.
cancer.num = as.data.frame(table(metadat$Cancer.Type))
colnames(cancer.num)[1] = "Cancer.Type"
rbp.fit.num.by.cancer = as.data.frame.matrix(table(rbp.fit$Cancer.Type,
    rbp.fit$Fitness))
av.fit.rbp.per.cancer = round(rbp.fit.num.by.cancer/cancer.num$Freq,
    0)

# Cancer type distribution
pdf(paste(plotdir, "Cell-line-number-by-cancer-type.pdf",
    sep = "/"))
cancer.num = cancer.num[order(cancer.num$Freq, decreasing = T),
    ]
cancer.num$Cancer.Type = factor(cancer.num$Cancer.Type,
    levels = cancer.num$Cancer.Type)
ggplot(data = cancer.num) + geom_bar(aes(x = Cancer.Type,
    y = Freq), stat = "identity", fill = "thistle") +
    theme_bw() + theme(plot.title = element_text(hjust = 0.5),
    axis.text.x = element_text(angle = 90, hjust = 0.95,
        vjust = 0.2)) + ylab("Number of cell lines")
dev.off()
```

```
## pdf
##   2
```

While looking at only those genes that are in the known.rbp list and in the Cancer Fitness study, we have an overlap of 1801. On a cancer type level, the range of these RBPs that are involved in core-fitness is between 25 and 45% with the median at 37%. It is useful to note that solid tumours are more highly represented in the dataset than others with colorectal, breast, ovarian, glioblastoma and pancreatic making the top 5.

04. Cancer Fitness and the Unfolded Protein Response

Here we want to see how many of the ~7500 cancer fitness genes are related to the unfolded protein response, hypoxia and endoplasmic reticulum as well as the integrated stress reponse. I'm going to do as I did with the RBPs, upload the list to Uniprot and download GO terms followed by filtering.

4a. Cancer fitness gene annotation

We have 7470 genes and we are going to try and get the GO terms for all of these. Follow up questions could be - how many UPR genes per cancer ? Is there a particular cancer with a high number of UPR/Hypoxia/Stress

```
# Annotate Cancer fitness genes with GO terms from
# Uniprot
go.fit = read.delim(paste(indir, "cancer-fitness-uniprot.txt",
    sep = "/"), sep = "\t", stringsAsFactors = F, header = T)
colnames(go.fit)[6] = "Gene.name"

# Hypoxia response
hypox.fit = go.fit[with(go.fit, grepl("GO:0001666",
    paste(Gene.ontology..biological.process., Gene.ontology..cellular.component.,
        Gene.ontology..molecular.function.))), ]

upr.fit = go.fit[with(go.fit, grepl("GO:0030968", paste(Gene.ontology..biological.process.,
    Gene.ontology..cellular.component., Gene.ontology..molecular.function.))),
    ]

starv.fit = go.fit[with(go.fit, grepl("GO:0042594",
    paste(Gene.ontology..biological.process., Gene.ontology..cellular.component.,
        Gene.ontology..molecular.function.))), ]

print(length(upr.fit$Gene.name))
```

```
## [1] 21
```

```
print(length(hypox.fit$Gene.name))
```

```
## [1] 59
```

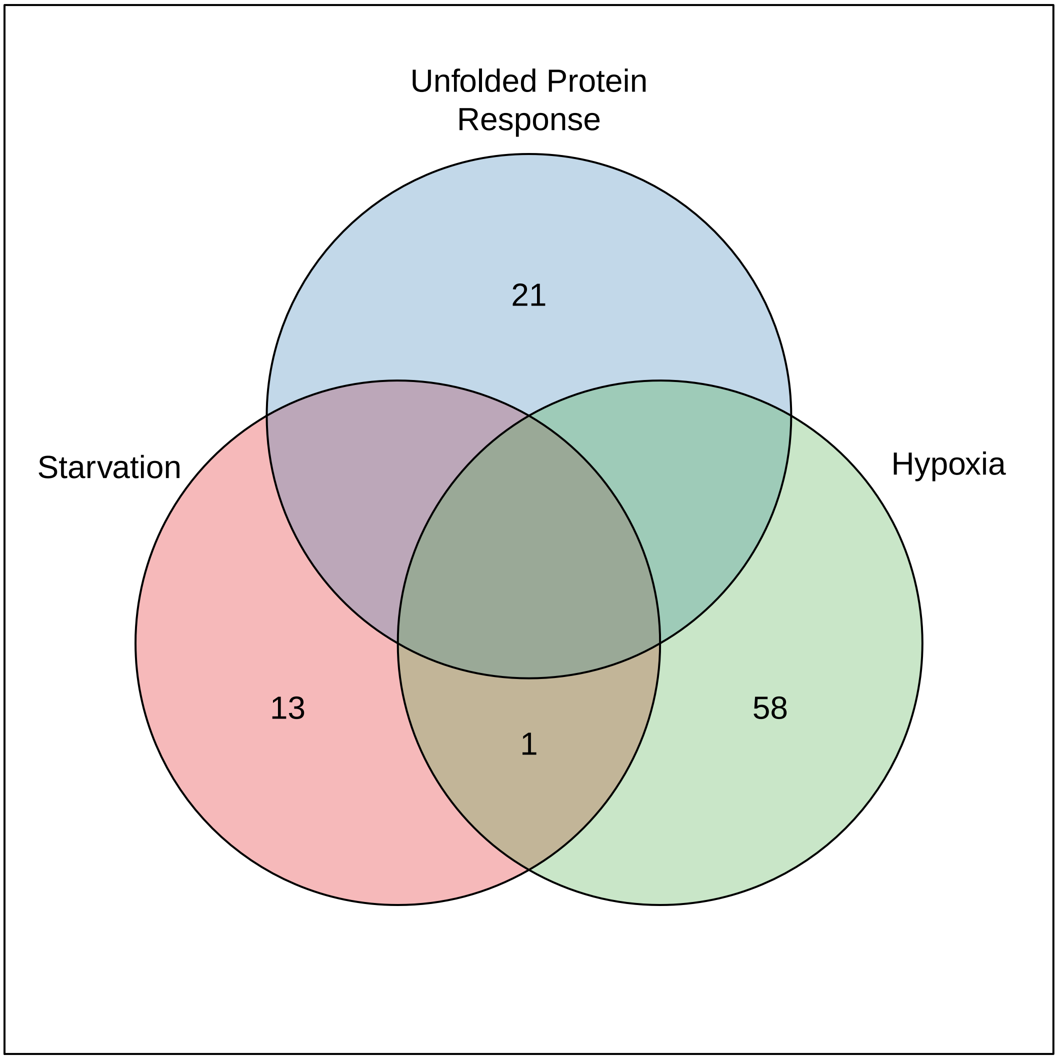
```
print(length(starv.fit$Gene.name))
```

```
## [1] 14
```

```
# Concatenate genes into a list for output
all.fit.out = rbind(cbind(hypox.fit, Category = "Hypoxia-Cancer-Fitness"),
    cbind(upr.fit, Category = "UPR-Cancer-Fitness"),
    cbind(starv.fit, Category = "Starvation-Cancer-Fitness"))

write.table(all.fit.out, paste(outdir, "UPR-Hypoxia-Starvation-Cancer-Fitness-Genes.txt",
    sep = "/"), sep = "\t", row.names = F, quote = F)

# Write lists of proteins to file
write.table(hypox.fit, paste(outdir, "Hypoxia-Fitness-genes.txt",
    sep = "/"), sep = "\t", row.names = F, quote = F)
write.table(starv.fit, paste(outdir, "Starvation-Fitness-genes.txt",
    sep = "/"), sep = "\t", row.names = F, quote = F)
write.table(upr.fit, paste(outdir, "UPR-Fitness-genes.txt",
    sep = "/"), sep = "\t", row.names = F, quote = F)

# Plot intersect
input.v.fit = list(Starvation = starv.fit$Gene.name,
    `Unfolded Protein\nResponse` = upr.fit$Gene.name,
    Hypoxia = hypox.fit$Gene.name)
pdf(paste(plotdir, "UPR-Hypoxia-ER-Cancer-fitness-genes-Venn.pdf",
    sep = "/"))
v = venn::venn(input.v.fit, zcolor = brewer.pal(n = 4,
    name = "Set1"), cexil = 1, cexsn = 1, main = "Cancer Fitness Genes & the Integrated Stress Response
dev.off()
```

```
## pdf
##   2
```

```
venn::venn(input.v.fit, zcolor = brewer.pal(n = 3,
    name = "Set1"), cexil = 1, cexsn = 1, main = "Cancer Fitness Genes & the Integrated Stress Response
```



4b. Cancer fitness genes that relate to Hypoxia

Here, we are looking at cancer fitness genes that are also known to play a role in hypoxia and if that list differs across cancer types.

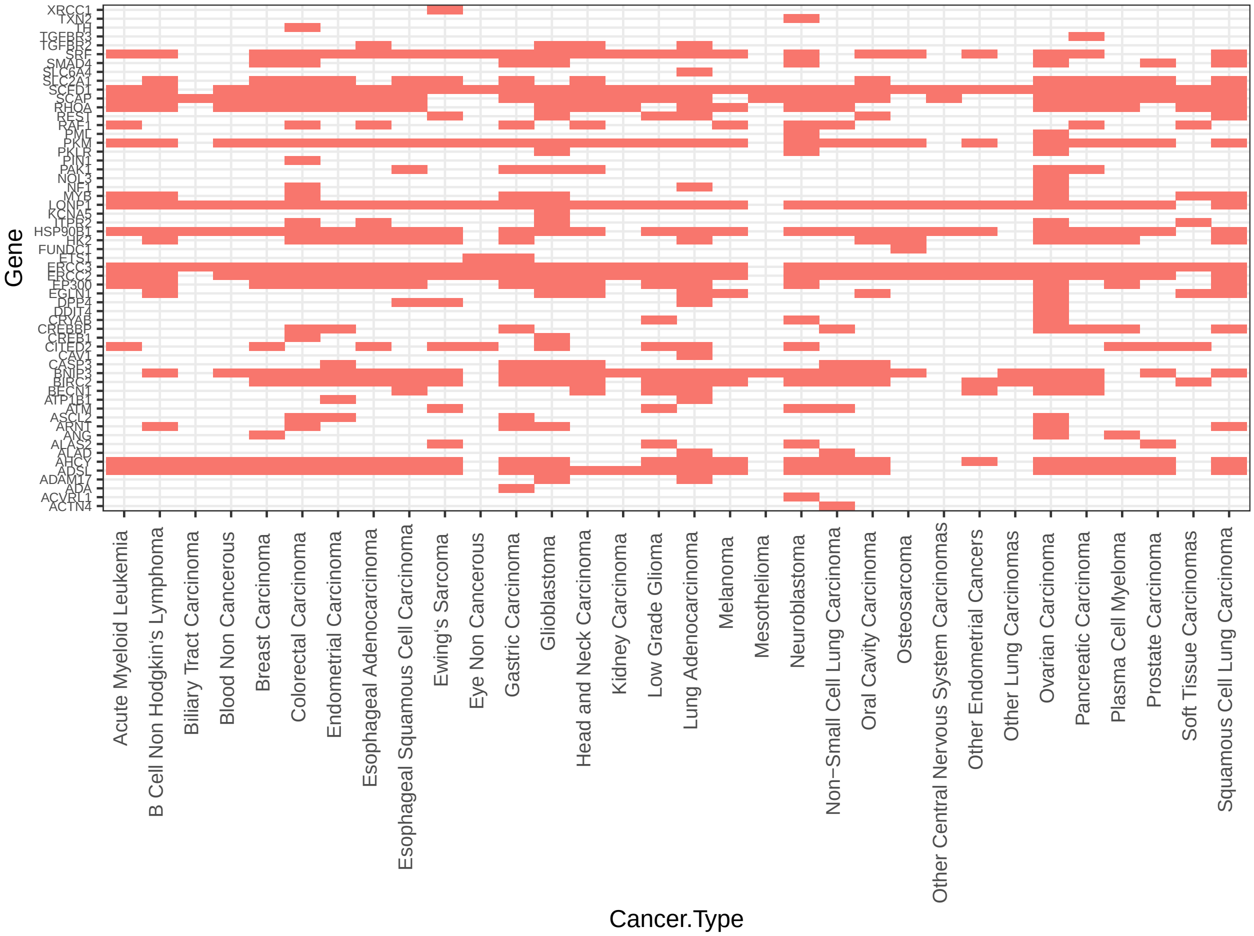
```
# Isolate hypoxia related cancer fitness genes and
# see if they occur in a particular cancer type
# hypox.fit$Gene.name
hypox.fitness = fitness[hypox.fit$Gene.name, ]

# Reshape data frame
hypox.fitness.melt = reshape2::melt(hypox.fitness,
    id = "Gene.CellLine", na.rm = T)
colnames(hypox.fitness.melt) = c("Gene", "Cell.line",
    "Fitness")
hypox.fitness.melt = left_join(hypox.fitness.melt,
    metadat)
# head(hypox.fitness.melt)
# Fitness only hypoxia genes
fit.hypox.1 = hypox.fitness.melt[which(hypox.fitness.melt$Fitness ==
    1), ]
p <- ggplot(fit.hypox.1, aes(y = Gene, x = Cancer.Type)) +
    geom_tile(aes(fill = Fitness)) + theme_bw() + theme(plot.title = element_text(hjust = 0.5),
    axis.text.x = element_text(angle = 90, hjust = 0.95,
        vjust = 0.2), axis.text.y = element_text(size = 6),
    legend.title = element_blank(), legend.position = "none")

# Plot the heatmap of genes
p
pdf(paste(plotdir, "Hypoxia-heatmap.pdf", sep = "/"))
print(p)
dev.off()
```

```
## pdf
##   2
```

```
# Create a data table showing number of hypoxia
# fitness genes by cancer
dt = data.frame(table(fit.hypox.1$Cancer.Type))
colnames(dt) = c("Cancer.Type", "Hypoxia.Genes")
dt = left_join(dt, cancer.num)
colnames(dt)[3] = "Cell.Lines"
dt$Av.Hypoxia = ceiling(dt$Hypoxia.Genes/dt$Cell.Lines)
dt = dt[order(dt$Av.Hypoxia, decreasing = T), ]
dt
```

```
##                                  Cancer.Type Hypoxia.Genes Cell.Lines
## 1                     Acute Myeloid Leukemia            25          2
## 2              B Cell Non Hodgkin`s Lymphoma            51          5
## 29                       Plasma Cell Myeloma            53          5
## 25                 Other Endometrial Cancers            10          1
## 31                    Soft Tissue Carcinomas            10          1
## 4                        Blood Non Cancerous            17          2
## 12                         Gastric Carcinoma           114         14
## 18                                  Melanoma            36          4
## 32              Squamous Cell Lung Carcinoma            89         11
## 6                       Colorectal Carcinoma           235         32
## 7                      Endometrial Carcinoma            51          7
## 8                  Esophageal Adenocarcinoma            52          7
## 9         Esophageal Squamous Cell Carcinoma           143         19
## 10                           Ewing`s Sarcoma            79         10
## 11                         Eye Non Cancerous             8          1
## 13                              Glioblastoma           174         24
## 14                   Head and Neck Carcinoma           127         16
## 17                       Lung Adenocarcinoma           147         20
## 20                             Neuroblastoma           121         17
## 21             Non-Small Cell Lung Carcinoma            68          9
## 27                         Ovarian Carcinoma           241         31
## 28                      Pancreatic Carcinoma           169         23
## 30                        Prostate Carcinoma            24          3
## 16                          Low Grade Glioma            63          9
## 22                     Oral Cavity Carcinoma           125         18
## 3                    Biliary Tract Carcinoma             6          1
## 5                          Breast Carcinoma           147         25
## 15                         Kidney Carcinoma            18          3
## 23                             Osteosarcoma            12          2
## 24 Other Central Nervous System Carcinomas             6          1
## 26                    Other Lung Carcinomas             6          1
## 19                             Mesothelioma             3          1
##    Av.Hypoxia
## 1          13
## 2          11
## 29         11
## 25         10
## 31         10
## 4           9
## 12          9
## 18          9
## 32          9
## 6           8
## 7           8
## 8           8
## 9           8
## 10          8
## 11          8
## 13          8
## 14          8
## 17          8
## 20          8
## 21          8
## 27          8
## 28          8
## 30          8
## 16          7
## 22          7
## 3           6
## 5           6
## 15          6
## 23          6
## 24          6
## 26          6
## 19          3
```

```
# Of all hypoxia genes, what percentage are fitness
# related by cancer-type (i.e Fitness = 1)
hypox.tab = round(100 * table(hypox.fitness.melt$Cancer.Type,
    hypox.fitness.melt$Fitness)/rowSums(table(hypox.fitness.melt$Cancer.Type,
    hypox.fitness.melt$Fitness)), 0)
hypox.tab = hypox.tab[order(hypox.tab[, 2], decreasing = T),
    ]
# hypox.tab
```



4c. Cancer fitness genes that relate to Unfolded Protein Response

Here, we are looking at cancer fitness genes that are also known to play a role in Unfolded Protein response and if that list differs across cancer types.

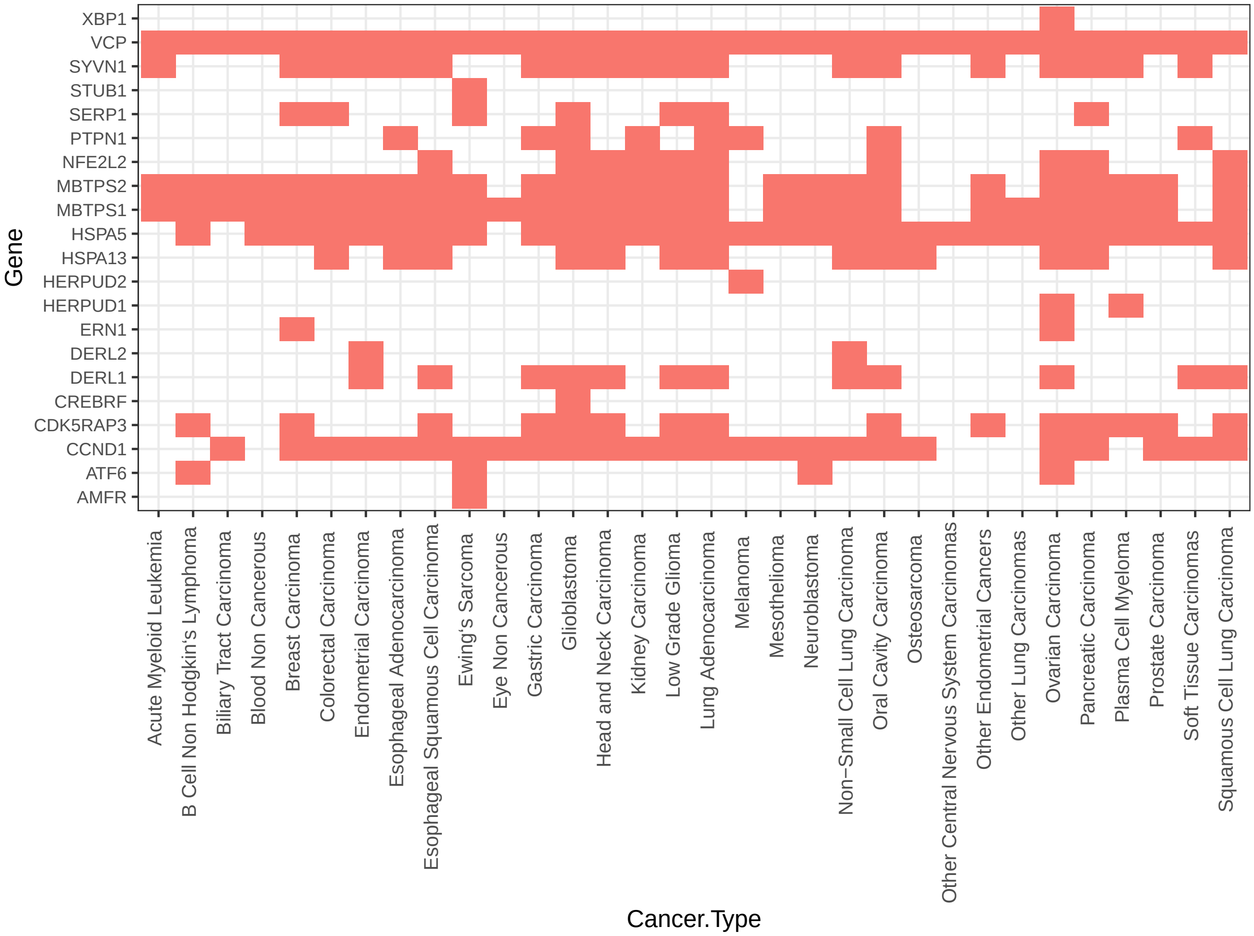
```
# Isolate UPR related cancer fitness genes and see
# if they occur in a particular cancer type
# upr.fit$Gene.name
upr.fitness = fitness[upr.fit$Gene.name, ]

# Reshape data frame
upr.fitness.melt = reshape2::melt(upr.fitness, id = "Gene.CellLine",
    na.rm = T)
colnames(upr.fitness.melt) = c("Gene", "Cell.line",
    "Fitness")
upr.fitness.melt = left_join(upr.fitness.melt, metadat)
# head(upr.fitness.melt)

# Fitness only UPR genes
fit.upr.1 = upr.fitness.melt[which(upr.fitness.melt$Fitness ==
    1), ]
q <- ggplot(fit.upr.1, aes(y = Gene, x = Cancer.Type)) +
    geom_tile(aes(fill = Fitness)) + theme_bw() + theme(plot.title = element_text(hjust = 0.5),
    axis.text.x = element_text(angle = 90, hjust = 0.95,
        vjust = 0.2), axis.text.y = element_text(size = 8),
    legend.title = element_blank(), legend.position = "none")

# Heatmap
q
pdf(paste(plotdir, "UPR-heatmap.pdf", sep = "/"))
print(q)
dev.off()
```

```
## pdf
##   2
```

```
# Create a data table showing number of UPR fitness
# genes by cancer
dt = data.frame(table(fit.upr.1$Cancer.Type))
colnames(dt) = c("Cancer.Type", "UPR.Genes")
dt = left_join(dt, cancer.num)
colnames(dt)[3] = "Cell.Lines"
dt$Av.UPR = ceiling(dt$UPR.Genes/dt$Cell.Lines)
dt = dt[order(dt$Av.UPR, decreasing = T), ]
dt
```

```
##                                 Cancer.Type UPR.Genes Cell.Lines Av.UPR
## 25              Other Endometrial Cancers         6          1      6
## 31                 Soft Tissue Carcinomas         6          1      6
## 9      Esophageal Squamous Cell Carcinoma        80         19      5
## 10                        Ewing`s Sarcoma        41         10      5
## 15                       Kidney Carcinoma        13          3      5
## 16                       Low Grade Glioma        41          9      5
## 17                    Lung Adenocarcinoma        85         20      5
## 19                           Mesothelioma         5          1      5
## 28                   Pancreatic Carcinoma        93         23      5
## 3                 Biliary Tract Carcinoma         4          1      4
## 4                     Blood Non Cancerous         7          2      4
## 7                   Endometrial Carcinoma        22          7      4
## 12                      Gastric Carcinoma        50         14      4
## 13                           Glioblastoma        93         24      4
## 14                Head and Neck Carcinoma        57         16      4
## 20                          Neuroblastoma        54         17      4
## 21          Non-Small Cell Lung Carcinoma        36          9      4
## 22                  Oral Cavity Carcinoma        62         18      4
## 27                      Ovarian Carcinoma       112         31      4
## 29                    Plasma Cell Myeloma        20          5      4
## 32           Squamous Cell Lung Carcinoma        43         11      4
## 1                  Acute Myeloid Leukemia         5          2      3
## 2           B Cell Non Hodgkin`s Lymphoma        14          5      3
## 5                        Breast Carcinoma        63         25      3
## 6                    Colorectal Carcinoma        85         32      3
## 8               Esophageal Adenocarcinoma        21          7      3
## 11                      Eye Non Cancerous         3          1      3
## 18                               Melanoma        11          4      3
## 23                           Osteosarcoma         6          2      3
## 26                  Other Lung Carcinomas         3          1      3
## 30                     Prostate Carcinoma         9          3      3
## 24 Other Central Nervous System Carcinomas          2           1       2
```

```
# Of all UPR genes, what percentage are fitness
# related by cancer-type (i.e Fitness = 1)
upr.tab = round(100 * table(upr.fitness.melt$Cancer.Type,
    upr.fitness.melt$Fitness)/rowSums(table(upr.fitness.melt$Cancer.Type,
    upr.fitness.melt$Fitness)), 0)
upr.tab = upr.tab[order(upr.tab[, 2], decreasing = T),
    ]
# upr.tab
```



4d. Cancer fitness genes that relate to Starvation

Here, we are looking at cancer fitness genes that are also known to play a role in Starvation response in a cell and if that list differs across cancer types.

```
# Isolate starvation response related cancer
# fitness genes and see if they occur in a
# particular cancer type starv.fit$Gene.name
starv.fitness = fitness[starv.fit$Gene.name, ]

# Reshape data frame
starv.fitness.melt = reshape2::melt(starv.fitness,
    id = "Gene.CellLine", na.rm = T)
```

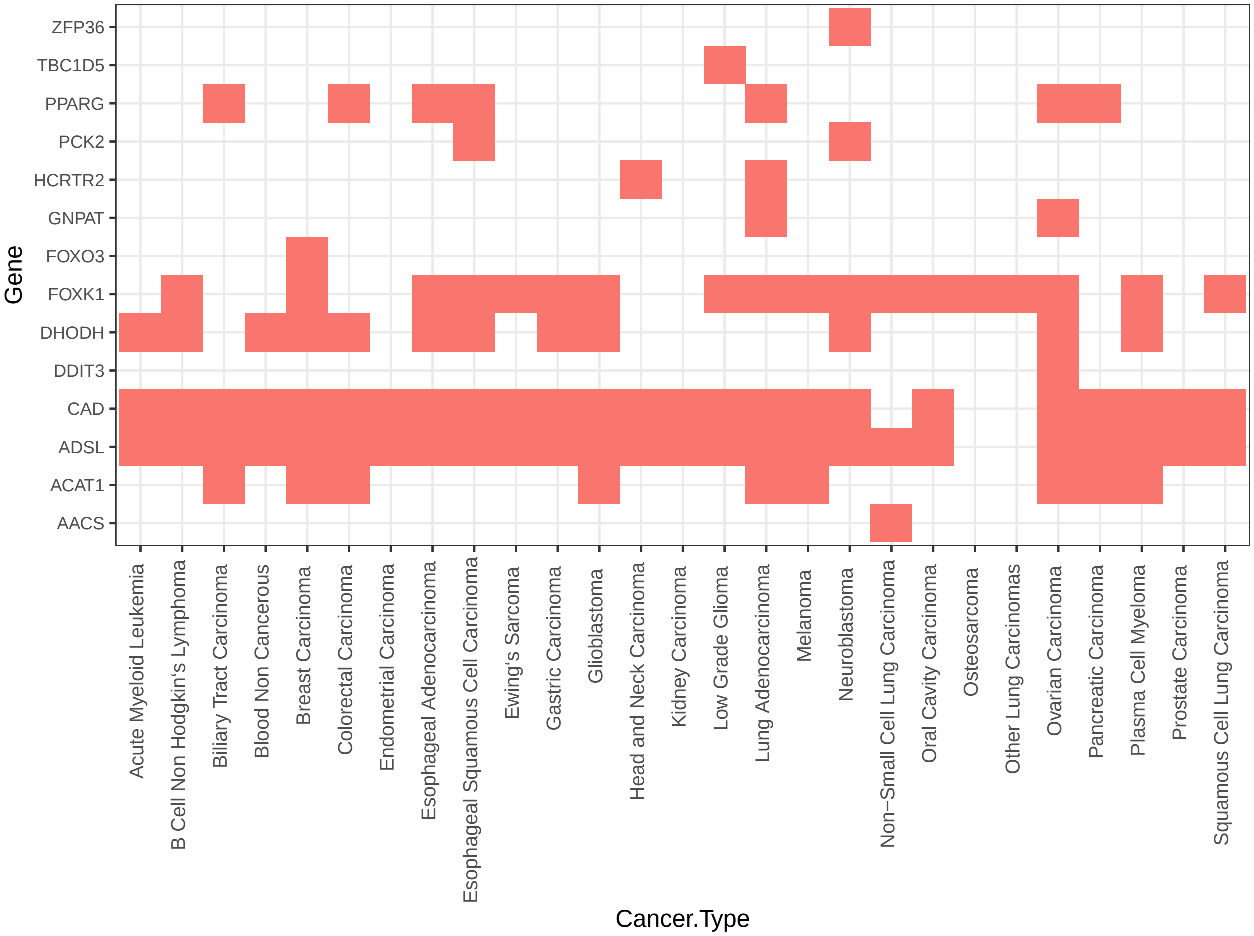
```
colnames(starv.fitness.melt) = c("Gene", "Cell.line",
    "Fitness")
starv.fitness.melt = left_join(starv.fitness.melt,
    metadat)
# head(starv.fitness.melt)

# Fitness only starvation genes
fit.starv.1 = starv.fitness.melt[which(starv.fitness.melt$Fitness ==
    1), ]
r <- ggplot(fit.starv.1, aes(y = Gene, x = Cancer.Type)) +
    geom_tile(aes(fill = Fitness)) + theme_bw() + theme(plot.title = element_text(hjust = 0.5),
    axis.text.x = element_text(angle = 90, hjust = 0.95,
        vjust = 0.2), axis.text.y = element_text(size = 8),
    legend.title = element_blank(), legend.position = "none")

# Heatmap
r
pdf(paste(plotdir, "Starv-heatmap.pdf", sep = "/"))
print(r)
dev.off()
```

```
## pdf
##   2
```

```
# Create a data table showing number of starvation
# fitness genes by cancer
dt = data.frame(table(fit.starv.1$Cancer.Type))
colnames(dt) = c("Cancer.Type", "Starv.Genes")
dt = left_join(dt, cancer.num)
colnames(dt)[3] = "Cell.Lines"
dt$Av.UPR = ceiling(dt$Starv.Genes/dt$Cell.Lines)
dt = dt[order(dt$Av.UPR, decreasing = T), ]
dt
```

```
##                                 Cancer.Type Starv.Genes Cell.Lines Av.UPR
## 3                   Biliary Tract Carcinoma           4          1      4
## 1                    Acute Myeloid Leukemia           6          2      3
## 2             B Cell Non Hodgkin`s Lymphoma          12          5      3
## 4                       Blood Non Cancerous           6          2      3
## 29                      Plasma Cell Myeloma          13          5      3
## 5                          Breast Carcinoma          34         25      2
## 6                      Colorectal Carcinoma          50         32      2
## 8                 Esophageal Adenocarcinoma          12          7      2
## 12                        Gastric Carcinoma          22         14      2
## 17                      Lung Adenocarcinoma          23         20      2
## 18                                 Melanoma           5          4      2
## 30                       Prostate Carcinoma           4          3      2
## 7                     Endometrial Carcinoma           4          7      1
## 9        Esophageal Squamous Cell Carcinoma          17         19      1
## 10                          Ewing`s Sarcoma           8         10      1
## 13                             Glioblastoma           8         24      1
## 14                  Head and Neck Carcinoma          11         16      1
## 15                         Kidney Carcinoma           2          3      1
## 16                         Low Grade Glioma           6          9      1
## 20                            Neuroblastoma          13         17      1
## 21          Non-Small Cell Lung Carcinoma        7          9       1
## 22                  Oral Cavity Carcinoma       11         18       1
## 23                           Osteosarcoma        1          2       1
## 26                  Other Lung Carcinomas        1          1       1
## 27                      Ovarian Carcinoma       31         31       1
## 28                    Pancreatic Carcinoma       22         23       1
## 32             Squamous Cell Lung Carcinoma       11         11       1
## 11                      Eye Non Cancerous        0          1       0
## 19                           Mesothelioma        0          1       0
## 24 Other Central Nervous System Carcinomas        0          1       0
## 25                Other Endometrial Cancers        0          1       0
## 31                   Soft Tissue Carcinomas        0          1       0
```

```
# Of all Starvation genes, what percentage are
# fitness related by cancer-type (i.e Fitness = 1)
starv.tab = round(100 * table(starv.fitness.melt$Cancer.Type,
    starv.fitness.melt$Fitness)/rowSums(table(starv.fitness.melt$Cancer.Type,
    starv.fitness.melt$Fitness)), 0)
starv.tab = starv.tab[order(starv.tab[, 2], decreasing = T),
    ]
# starv.tab
```


ZFP36
TBC1D5
PPARG
PCK2
HCRTR2
GNPAT
FOXO3
FOXK1
DHODH
DDIT3
CAD
ADSL
ACAT1
AACS
Gene
Acute Myeloid Leukemia
B Cell Non Hodgkin's Lymphoma
Biliary Tract Carcinoma
Blood Non Cancerous
Breast Carcinoma
Colorectal Carcinoma
Endometrial Carcinoma
Esophageal Adenocarcinoma
Esophageal Squamous Cell Carcinoma
Ewing's Sarcoma
Gastric Carcinoma
Glioblastoma
Head and Neck Carcinoma
Kidney Carcinoma
Low Grade Glioma
Lung Adenocarcinoma
Melanoma
Neuroblastoma
Non−Small Cell Lung Carcinoma
Oral Cavity Carcinoma
Osteosarcoma
Other Lung Carcinomas
Ovarian Carcinoma
Pancreatic Carcinoma
Plasma Cell Myeloma
Prostate Carcinoma
Squamous Cell Lung Carcinoma
Cancer.Type