R Notebook

Contents

BSMS_222 Assignment_1 **1**

Visualization of RNA and Protein expression of RMB10 with MAPK and PI3K pathways and EGFR 1

grouped by sex and existence of RMB10 SNVs. 1

BSMS_222 Assignment_1

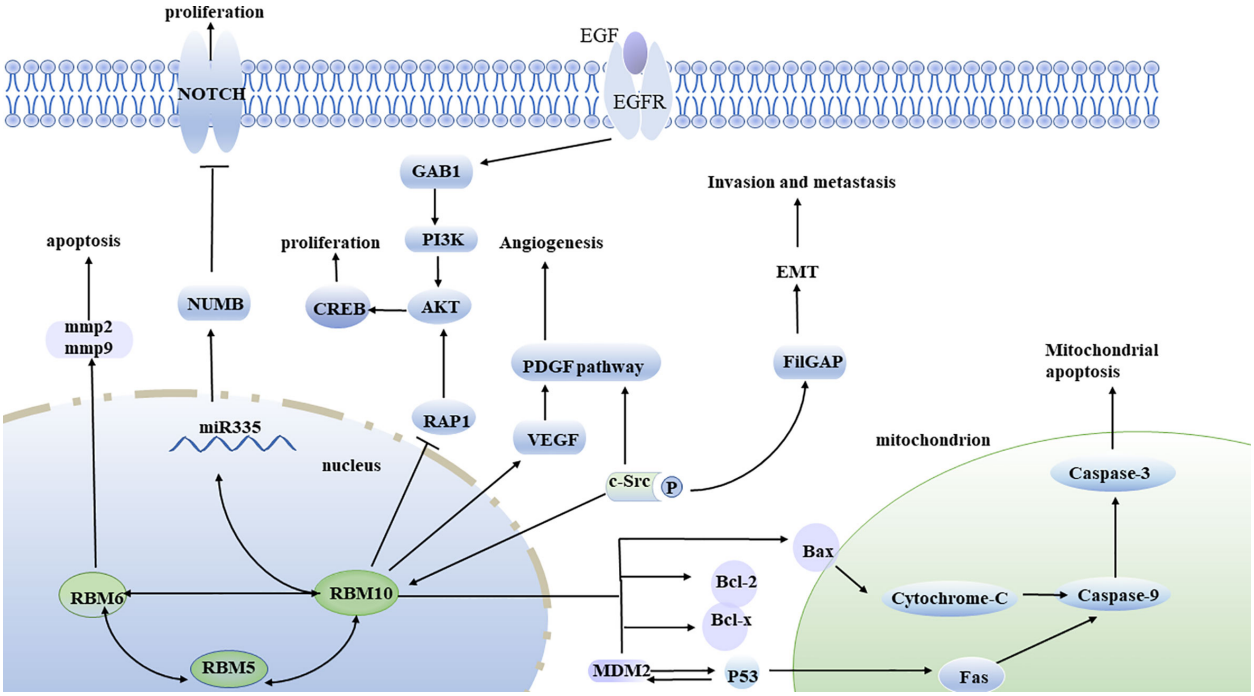
JiHye Lee [2020250155]

Visualization of RNA and Protein expression of RMB10 with MAPK and PI3K pathways and EGFR

grouped by sex and existence of RMB10 SNVs.

Introduction

The given paper was about LUAD in East Asia with the cohort of Taiwanese population (TW), with the intend to study non-smoking lung adenocarcinoma (Chen et al., 2020). Discussing mainly about female patients in the cohort, classified in age, stage, and the smoking status, the study identified certain genes and proteins that were distinct in certain classified groups. Among them RBM10 appeared to be in higher mutation frequency then other studies (TCGA, 2014; Imielinski et al., 2012; Campbell et al., 2016) compared in the paper. RBM10 also was distinct that it showed mutation exclusively with TP53, KRAS, XIP2, and ZNF804B. However, it showed decreased expression in RNA and Protein unlike the other genes such as the KRAS, LABM1, and PIK3CA in the according cohort. Curious about the role of RBM10, I learned that it plays an anti-cancer role by causing apoptosis and inhibiting proliferation as the image below. Other than the image, it also regulates the selective cleavage of *FAS*and block the MDM2-p53 loop.



However, there are studies that report the tumor-promoting role of RBM10. RBM10 was upregulated in LUAD cells and tissues (Sun X et al., 2019). Overexpression of RBM10 can lead to decrease in Bax and caspase-8, proteins that are pro-apoptotic, and increase of Bcl-2 expression which is anti-apoptotic. However, in the cohort TW, RBM10 showed decreased expression in RNA and Protein unlike the other genes such as the KRAS, LABM1, and PIK3CA. In some studies, RBM10 are hypothesized to be positively correlated with the expression of EGFR and the MAPK and PI3K signaling pathways. (Cao et al., 2021) I tried visualizing the RBM10 and EGFR, and MAPK, PI3K pathway(Yang et al., 2019) genes' RNA and protein expression, grouping the cohort according to sex and the existence of RBM10 SNV. What I hoped was to find a distinct characteristic in the TW cohort, that RBM10 plays a role in the cause of LUAD when SNV mutation did not occur.

Code

```
# library used
library(readr)
library(readxl)
library(dplyr)
```

```
##
##          : 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.4     v stringr 1.4.0
## v tidyr   1.1.4     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
library(ggridges)
library(GGally)
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
library(cowplot)

# basic data
## loading data
dir <- "C:/Users/USER/Desktop/bsms222_155_lee/portfolio_tables"
list.files(path = dir)
```

```
##  [1] "~$table_1.xlsx" "~$table_2.xlsx" "~$table_5.xlsx" "~$table_7.xlsx"
##  [5] "table_1.xlsx"   "table_2.xlsx"   "table_3.xlsx"   "table_4.xlsx"
##  [9] "table_5.xlsx"   "table_6.xlsx"   "table_7.xlsx"
```

```
fullpath_1 <- file.path(dir, "table_1.xlsx")



## had to earn these since the list of patients in two data were different
## patients with adc
adc_patients <- as.data.frame(read_excel(fullpath_1, sheet = 2)) %>%
  filter((`Histology Type` == 'ADC')) %>%
  pull(ID)

##snv column name
snv_columns <- read_excel(fullpath_1, sheet = 4, skip = 1, n_max = 0) %>%
    names()

data_1_female <- as.data.frame(read_excel(fullpath_1, sheet = 2)) %>%
  filter(`Histology Type` == 'ADC') %>%
  filter(ID %in% snv_columns & Gender == 'Female') %>%
  pull(ID)
```

```r
#snv data of adc patients
data_1_snv <- as.data.frame(read_excel(fullpath_1, sheet = 4, skip = 1)) %>%
  select(Gene, which(snv_columns %in% adc_patients))

rbm10_snv_patient <- as.data.frame(read_excel(fullpath_1, sheet = 4, skip = 1)) %>%
  select(Gene, which(snv_columns %in% adc_patients)) %>%
  filter(Gene == "RBM10") %>%
  pivot_longer(cols = -Gene, names_to = "patient", values_to = "RBM10 snv") %>%
  mutate(have_rbm10_snv = ifelse(`RBM10 snv` %in% NA, "N", "Y")) %>%
  filter(`have_rbm10_snv` == "Y") %>%  pull(patient)

## transcriptsome data (log2T/N) of ADC patients
transcriptsome_columns <- read_excel(fullpath_1, sheet = 5, n_max = 0) %>%
    names()
data_1_transcriptome <- as.data.frame(read_excel(fullpath_1, sheet = 5)) %>%
  select(gene, which(transcriptsome_columns %in% adc_patients & transcriptsome_columns %in% snv_columns)



## proteome data (log2T/N) of ADC patients
proteome_columns <- read_excel(fullpath_1, sheet = 6, na="NA", n_max = 0) %>%
    names()
data_1_proteome <- as.data.frame(read_excel(fullpath_1, sheet = 6, na = "NA")) %>%
  select(Gene,  which(proteome_columns %in% adc_patients & proteome_columns %in% snv_columns))
```

```r
## egfr
### egfr gene
rbm10_cancer_egfr <- c("RBM10", "EGFR")

### transcriptome
egfr_transcriptome <- data_1_transcriptome  %>% filter(gene %in% rbm10_cancer_egfr) %>%
  pivot_longer(cols = -gene, names_to = "patient", values_to = "transcriptome")%>%
  mutate(with_rbm10_snv = ifelse(patient %in% rbm10_snv_patient, "Y", "N")) %>%
  mutate(gender = ifelse(patient %in% data_1_female, "F", "M")) %>%
  mutate(groups = case_when(gender == "F"& with_rbm10_snv == "Y" ~ "Female with RBM10 SNV",
         gender == "F"& with_rbm10_snv == "N" ~ "Female without RBM10 SNV",
         gender == "M"& with_rbm10_snv == "Y" ~ "Male with RBM10 SNV",
         gender == "M"& with_rbm10_snv == "N" ~ "Male without RBM10 SNV")) %>%
    select(-c(with_rbm10_snv, gender)) %>%
  pivot_wider( names_from = "gene", values_from = "transcriptome")


mylabels <- c(`1` = "Female with RBM10 SNV",
              `2` = "Female without RBM10 SNV",
         `3` =  "Male with RBM10 SNV",
         `4` = "Male without RBM10 SNV")

plot_egfr_trans <- egfr_transcriptome %>% ggparcoord(
    columns = 3:4,
    groupColumn = NULL,
    showPoints = TRUE,
    alphaLines = 1,
    boxplot = TRUE) +
```

```
    scale_color_brewer(palette = "Set1") +
  scale_x_discrete(limits = rbm10_cancer_egfr) +
  theme_bw() +
      labs(y = 'Log2T/N',
           x = NULL,
       title = "RBM10 and EGFR RNA Expression")+
  facet_wrap(. ~ groups, ncol = 4, labeller = labeller(groups = mylabels))


### proteome
egfr_proteome <- data_1_proteome  %>% filter(Gene %in% rbm10_cancer_egfr) %>%
  pivot_longer(cols = -Gene, names_to = "patient", values_to = "proteome") %>%
  mutate(with_rbm10_snv = ifelse(patient %in% rbm10_snv_patient, "Y", "N")) %>%
  mutate(gender = ifelse(patient %in% data_1_female, "F", "M")) %>%
  mutate(groups = case_when(gender == "F"& with_rbm10_snv == "Y" ~ "Female with RBM10 SNV",
         gender == "F"& with_rbm10_snv == "N" ~ "Female without RBM10 SNV",
         gender == "M"& with_rbm10_snv == "Y" ~ "Male with RBM10 SNV",
         gender == "M"& with_rbm10_snv == "N" ~ "Male without RBM10 SNV")) %>%
    select(-c(with_rbm10_snv, gender)) %>%
  mutate_if(is.numeric, round, 3) %>%
  pivot_wider( names_from = "Gene", values_from = "proteome")


plot_egfr_prot <- egfr_proteome %>% ggparcoord(
    columns = 3:4,
    groupColumn = NULL,
    showPoints = TRUE,
    alphaLines = 1,
    boxplot = TRUE) +
    scale_color_brewer(palette = "Set1") +
  scale_x_discrete(limits = rbm10_cancer_egfr) +
  theme_bw()+
      labs(y = 'Log2T/N',
           x = NULL,
       title = "RBM10 and EGFR Protein Expression")+
    facet_wrap(. ~ groups, ncol = 4,labeller = labeller(groups = mylabels))

##mapk
### mapk pathway genes
rbm10_cancer_mapk <- c("RBM10", "AKT2", "AKT3", "BAD", "FOXO3", "MAP2K2", "MAPK1", "MAPK3", "PLCG1", "PI

### filtering basic data and changing into tidy form
mapk_transcriptome <- data_1_transcriptome  %>% filter(gene %in% rbm10_cancer_mapk) %>%
  pivot_longer(cols = -gene, names_to = "patient", values_to = "transcriptome")


mapk_proteome <- data_1_proteome  %>% filter(Gene %in% rbm10_cancer_mapk) %>%
  pivot_longer(cols = -Gene, names_to = "patient", values_to = "proteome")

#### unifying column name
names(mapk_proteome)[names(mapk_proteome) == 'Gene'] <- 'gene'
```

```r
### merging data and mutating into four groups
mapk <- merge(mapk_transcriptome, mapk_proteome, by= c("gene","patient"), all = TRUE) %>%
  pivot_longer(cols = -c(gene, patient), names_to = "data_type", values_to = "value") %>%
  mutate(with_rbm10_snv = ifelse(patient %in% rbm10_snv_patient, "Y", "N")) %>%
  mutate(gender = ifelse(patient %in% data_1_female, "F", "M")) %>%
  mutate(groups = case_when(gender == "F"& with_rbm10_snv == "Y" ~ "Female with RBM10 SNV",
         gender == "F"& with_rbm10_snv == "N" ~ "Female without RBM10 SNV",
         gender == "M"& with_rbm10_snv == "Y" ~ "Male with RBM10 SNV",
         gender == "M"& with_rbm10_snv == "N" ~ "Male without RBM10 SNV"))

### plot

plot_mapk <- mapk%>% ggplot(aes(value, gene, fill = groups)) +
  geom_density_ridges(alpha = 0.5) +
  scale_x_continuous(limits = c(-3, 3)) +
  theme_ridges(grid = FALSE) +
  scale_y_discrete(limits = rbm10_cancer_mapk) +
  scale_color_brewer(palette = "Set1") +
  labs(x = 'Log2T/N',
       y = 'MAPK pathway genes',
       title = "RBM10 and MAPK Pathway Gene Expression") +
  theme_bw() +
  facet_grid(data_type ~ .) +
  theme(plot.title = element_text(size=11),
        legend.position = "none")



## PI3K
rbm10_cancer_pi3k <- c("RBM10", "AKT", "MTOR", "PTEN", "FRAP", "FRAP1", "FRAP2", "RAFT1", "RAPT1")

### filtering basic data and changing into tidy form
pi3k_transcriptome <- data_1_transcriptome  %>% filter(gene %in% rbm10_cancer_pi3k) %>%
  pivot_longer(cols = -gene, names_to = "patient", values_to = "transcriptome")


pi3k_proteome <- data_1_proteome  %>% filter(Gene %in% rbm10_cancer_pi3k) %>%
  pivot_longer(cols = -Gene, names_to = "patient", values_to = "proteome")

#### unifying column name
names(pi3k_proteome)[names(pi3k_proteome) == 'Gene'] <- 'gene'


### merging data and mutating into four groups
pi3k <- merge(pi3k_transcriptome, pi3k_proteome, by= c("gene","patient"), all = TRUE) %>%
  pivot_longer(cols = -c(gene, patient), names_to = "data_type", values_to = "value") %>%
  mutate(with_rbm10_snv = ifelse(patient %in% rbm10_snv_patient, "Y", "N")) %>%
  mutate(gender = ifelse(patient %in% data_1_female, "F", "M")) %>%
  mutate(groups = case_when(gender == "F"& with_rbm10_snv == "Y" ~ "Female with RBM10 SNV",
         gender == "F"& with_rbm10_snv == "N" ~ "Female without RBM10 SNV",
         gender == "M"& with_rbm10_snv == "Y" ~ "Male with RBM10 SNV",
         gender == "M"& with_rbm10_snv == "N" ~ "Male without RBM10 SNV"))
```

```
### plot
plot_pi3k <- pi3k%>% ggplot(aes(value, gene, fill = groups)) +
  geom_density_ridges(alpha = 0.5) +
  scale_x_continuous(limits = c(-3, 3)) +
  theme_ridges(grid = FALSE) +
  scale_y_discrete(limits = rbm10_cancer_pi3k) +
  scale_color_brewer(palette = "Set1") +
  labs(x = 'Log2T/N',
       y = 'PI3K pathway genes',
       title = "RBM10 and PI3K Pathway Gene Expression") +
  theme_bw() +
  facet_grid(data_type ~ .) +
   theme(plot.title = element_text(size=11),
         legend.key.size = unit(0.3, 'cm'),
         legend.key.height = unit(0.3, 'cm'),
         legend.key.width = unit(0.3, 'cm'),
         legend.title = element_text(size=9),
         legend.text = element_text(size=9),
         legend.position = c(.5,.9),)



# all of four plots
plots_top <- plot_grid(
  plot_mapk, plot_pi3k,
  labels = "AUTO", ncol = 2)
```

```
## Picking joint bandwidth of 0.112
```

```
## Picking joint bandwidth of 0.24
```

```
## Warning: Removed 26 rows containing non-finite values (stat_density_ridges).
```

```
## Picking joint bandwidth of 0.0968
```

```
## Picking joint bandwidth of 0.264
```

```
## Warning: Removed 3 rows containing non-finite values (stat_density_ridges).
```

```
plots_bottom <- plot_grid(
  plot_egfr_trans, plot_egfr_prot, labels = c("C", "D"), ncol = 1)

title <- ggdraw() +
  draw_label(
    "RBM10 and Expression of Possible Cancer Produing Pathways",
    fontface = 'bold',
    x = 0,
    hjust = 0,
    size = 15
  ) +
  theme(
```

```
    plot.margin = margin(5, 5, 5, 7)
  )
figure_1 <- plot_grid(
  title, plots_top, plots_bottom,
  ncol = 1,
  rel_heights = c(0.1, 6, 3 )
)

ggsave("figure_1.png", figure_1, width = 8, height = 14)
```

The Figure is at the end of the assignment

Discussion

The A and B plot, which is showing the expression of RNA and protein expression of MAPK and PI3K Pathway genes, dividing the group into Female with RBM10 SNV, and Female without RBM10 SNV, and Male without RBM10 SNV. The plots did not show a distinct characteristic of females with RBM10 SNV. The gene with notably low log2 T/N data is only RBM10. Although the log2 T/N data of protein expression of BRAF, MAPK3, and the RNA expression of MAPK3 showed bimodality that is different with the other groups, it is highly that the cause would be the small number of members that are in the group.

In the C and D plot, females with RBM10 SNV, has overall less log2 T/N data in expression in both RNA and Protein than females that has RBM10 SNV. Although there is worry that this might also have been affected by the small sample in all four groups, the log2 T/N data of EGFR are similar, but the range is narrow in scope in females with RBM10 SNV.

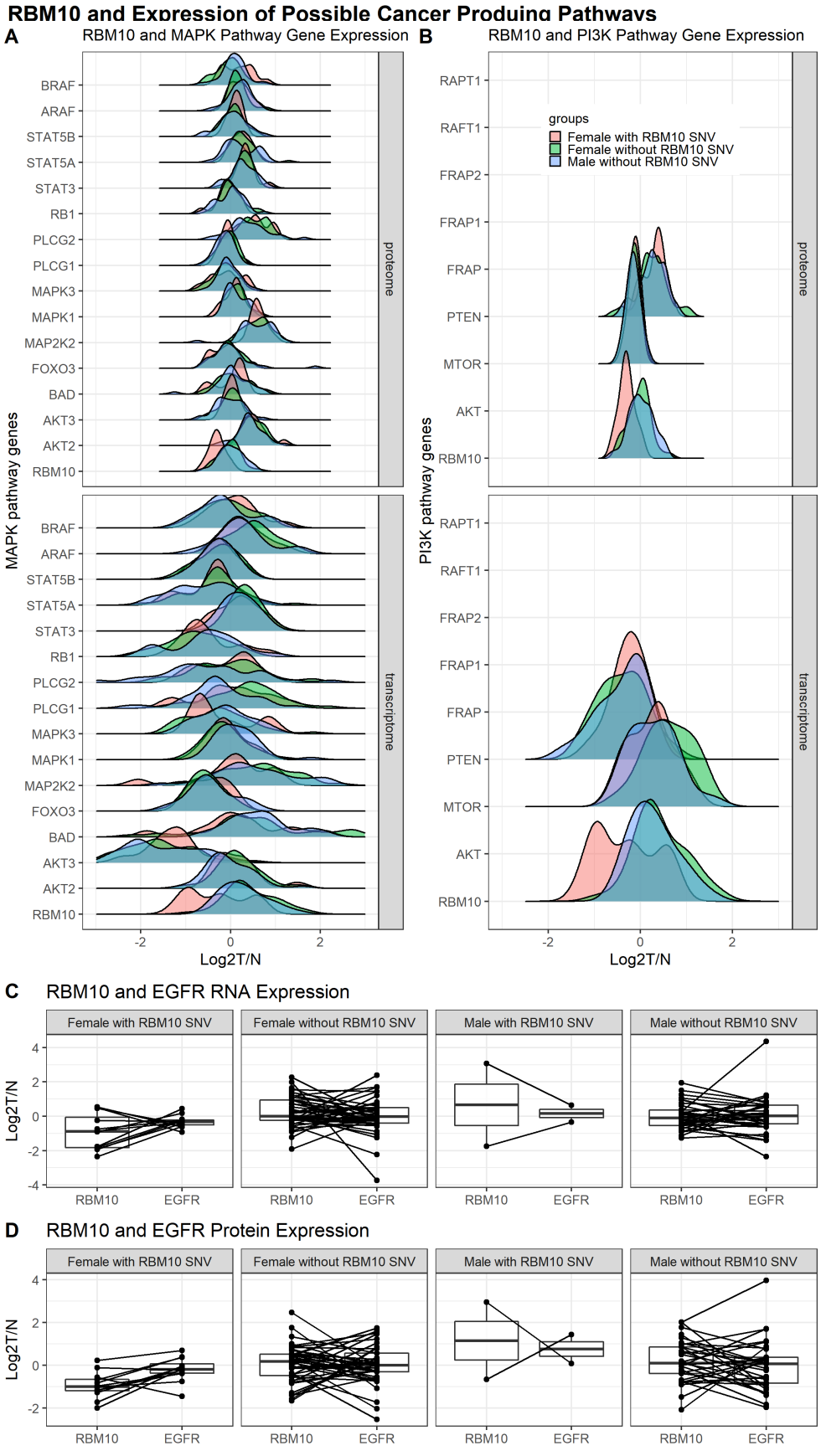
0.0.1 References {-}s

Yang, J., Nie, J., Ma, X. et al. Targeting PI3K in cancer: mechanisms and advances in clinical trials. *Mol Cancer* 18, 26 (2019).

Cell, ISSN: 0092-8674, Vol: 182, Issue: 1, Page: 226-244.e17

Sun X, Jia M, Sun W, Feng L, Gu C, Wu T. Functional role of RBM10 in lung adenocarcinoma proliferation. *Int J Oncol* (2019) 54(2):467–78. doi: 10.3892/ijo.2018.4643

Cao Y, Di X, Zhang Q, Li R and Wang K (2021) RBM10 Regulates Tumor Apoptosis, Proliferation, and Metastasis. *Front. Oncol.* 11:603932. doi: 10.3389/fonc.2021.603932


RBM10 and Expression of Possible Cancer Producing Pathways
A
RBM10 and MAPK Pathway Gene Expression
B
RBM10 and PI3K Pathway Gene Expression
groups
Female with RBM10 SNV
Female without RBM10 SNV
Male without RBM10 SNV
MAPK pathway genes
PI3K pathway genes
proteome
transcriptome
BRAF
ARAF
STAT5B
STAT5A
STAT3
RB1
PLCG2
PLCG1
MAPK3
MAPK1
MAP2K2
FOXO3
BAD
AKT3
AKT2
RBM10
RAPT1
RAFT1
FRAP2
FRAP1
FRAP
PTEN
MTOR
AKT
-2
0
2
Log2T/N
C
RBM10 and EGFR RNA Expression
Female with RBM10 SNV
Female without RBM10 SNV
Male with RBM10 SNV
Male without RBM10 SNV
Log2T/N
4
2
0
-2
-4
RBM10
EGFR
D
RBM10 and EGFR Protein Expression