

Analysis

Jieqi Tu

12/30/2020

```
CNA_HNSC = read.delim("./Downloaded Data/CNA HNSC.txt")
mRNA_HNSC = read.delim("./Downloaded Data/mRNA Expression HNSC.txt")
HNSC = cbind.data.frame(CNA_HNSC, mRNA_HNSC)
HNSC = HNSC[-c(5,6)]
CNA_LSCC = read.delim("./Downloaded Data/CNA LSCC.txt")
mRNA_LSCC = read.delim("./Downloaded Data/mRNA Expression LSCC.txt")
LSCC = cbind.data.frame(CNA_LSCC, mRNA_LSCC)
LSCC = LSCC[-c(5,6)]
CNA_CSCC = read.delim("./Downloaded Data/CNA CSCC.txt")
mRNA_CSCC = read.delim("./Downloaded Data/mRNA expression CSCC.txt")
CSCC = cbind.data.frame(CNA_CSCC, mRNA_CSCC)
CSCC = CSCC[-c(5,6)]

FAT1_HNSC = HNSC[c(1, 2, 3, 5)]
colnames(FAT1_HNSC) = c("Study ID", "Sample ID", "Copy Number Alterations", "mRNA Expression")
FAT1_HNSC$`log2 (mRNA Expression)` = log(FAT1_HNSC$mRNA Expression`, base = 2)
# Pearson Correlation
cor.test(FAT1_HNSC$`Copy Number Alterations`, FAT1_HNSC$`log2 (mRNA Expression)`, method="pearson")

##
## Pearson's product-moment correlation
##
## data: FAT1_HNSC$`Copy Number Alterations` and FAT1_HNSC$`log2 (mRNA Expression)`
## t = 7.1847, df = 486, p-value = 2.549e-12
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.2273536 0.3879562
## sample estimates:
## cor
## 0.3098636

# Spearman Correlation
cor.test(FAT1_HNSC$`Copy Number Alterations`, FAT1_HNSC$`log2 (mRNA Expression)`, method="spearman")

## Warning in cor.test.default(FAT1_HNSC$`Copy Number Alterations`, FAT1_HNSC$`log2
## (mRNA Expression)`, : Cannot compute exact p-value with ties

##
## Spearman's rank correlation rho
##
```

```
## data: FAT1_HNSC$'Copy Number Alterations' and FAT1_HNSC$'log2 (mRNA Expression)'
## S = 14441643, p-value = 1.203e-08
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.2543926

# Kendall Correlation
cor.test(FAT1_HNSC$'Copy Number Alterations', FAT1_HNSC$'log2 (mRNA Expression)', method="kendall")

##
## Kendall's rank correlation tau
##
## data: FAT1_HNSC$'Copy Number Alterations' and FAT1_HNSC$'log2 (mRNA Expression)'
## z = 5.7466, p-value = 9.106e-09
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## 0.2017775

FAT1_HNSC$'Copy Number Alterations' = as.factor(FAT1_HNSC$'Copy Number Alterations')
summary(FAT1_HNSC[4:5]) %>% knitr::kable()
```

mRNA Expression	log2 (mRNA Expression)
Min. : 97.33	Min. : 6.605
1st Qu.: 3215.97	1st Qu.:11.651
Median : 6208.62	Median :12.600
Mean : 7342.56	Mean :12.318
3rd Qu.: 9771.61	3rd Qu.:13.254
Max. :29544.90	Max. :14.851
NA's :8	NA's :8

```
P1.1 =
FAT1_HNSC %>%
  ggplot(aes(x = 'Copy Number Alterations', y = 'mRNA Expression')) + geom_boxplot() + theme_bw() +
  labs(caption = "FAT1 in HNSCC")

P1.2 =
FAT1_HNSC %>%
  ggplot(aes(x = 'Copy Number Alterations', y = 'log2 (mRNA Expression)')) + geom_boxplot() + theme_bw() +
  labs(caption = "FAT1 in HNSCC")

# Anova analysis
analysis1 = aov(FAT1_HNSC$'log2 (mRNA Expression)'~FAT1_HNSC$'Copy Number Alterations')
summary.aov(analysis1)
```

```
##
## FAT1_HNSC$'Copy Number Alterations'      Df Sum Sq Mean Sq F value    Pr(>F)
## Residuals                                483  846.3      1.75
## ---
```

```

## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 8 observations deleted due to missingness

EGFR_HNSC = HNSC[c(1, 2, 4, 6)]
colnames(EGFR_HNSC) = c("Study ID", "Sample ID", "Copy Number Alterations", "mRNA Expression")
EGFR_HNSC$`log2 (mRNA Expression)` = log(EGFR_HNSC$mRNA Expression`, base = 2)
# Pearson Correlation
cor.test(EGFR_HNSC$`Copy Number Alterations`, EGFR_HNSC$`log2 (mRNA Expression)`, method="pearson")

##
## Pearson's product-moment correlation
##
## data:  EGFR_HNSC$`Copy Number Alterations` and EGFR_HNSC$`log2 (mRNA Expression)`
## t = 13.204, df = 486, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.4453884 0.5763150
## sample estimates:
##      cor
## 0.5138375

# Spearman Correlation
cor.test(EGFR_HNSC$`Copy Number Alterations`, EGFR_HNSC$`log2 (mRNA Expression)`, method="spearman")

## Warning in cor.test.default(EGFR_HNSC$`Copy Number Alterations`, EGFR_HNSC$`log2
## (mRNA Expression)`, : Cannot compute exact p-value with ties

##
## Spearman's rank correlation rho
##
## data:  EGFR_HNSC$`Copy Number Alterations` and EGFR_HNSC$`log2 (mRNA Expression)`
## S = 10285277, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.4689816

# Kendall Correlation
cor.test(EGFR_HNSC$`Copy Number Alterations`, EGFR_HNSC$`log2 (mRNA Expression)`, method="kendall")

##
## Kendall's rank correlation tau
##
## data:  EGFR_HNSC$`Copy Number Alterations` and EGFR_HNSC$`log2 (mRNA Expression)`
## z = 10.78, p-value < 2.2e-16
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## 0.3796477

```

```
EGFR_HNSC$'Copy Number Alterations' = as.factor(EGFR_HNSC$'Copy Number Alterations')
summary(EGFR_HNSC[4:5]) %>% knitr::kable()
```

mRNA Expression	log2 (mRNA Expression)
Min. : 26.67	Min. : 4.737
1st Qu.: 1539.41	1st Qu.:10.588
Median : 2849.16	Median :11.476
Mean : 4860.16	Mean :11.386
3rd Qu.: 4566.32	3rd Qu.:12.157
Max. :82763.50	Max. :16.337
NA's :8	NA's :8

```
P2.1 =
EGFR_HNSC %>%
  ggplot(aes(x = 'Copy Number Alterations', y = 'mRNA Expression')) + geom_boxplot() + theme_bw() +
  labs(caption = "EGFR in HNSCC")

P2.2 =
EGFR_HNSC %>%
  ggplot(aes(x = 'Copy Number Alterations', y = 'log2 (mRNA Expression)')) + geom_boxplot() + theme_bw() +
  labs(caption = "EGFR in HNSCC")

# Anova analysis
analysis2 = aov(EGFR_HNSC$'log2 (mRNA Expression)'~EGFR_HNSC$'Copy Number Alterations')
summary.aov(analysis2)
```

```
##                                Df Sum Sq Mean Sq F value Pr(>F)
## EGFR_HNSC$'Copy Number Alterations'  3   337.0   112.33    68.84 <2e-16 ***
## Residuals                        484   789.7     1.63
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 8 observations deleted due to missingness
```

```
FAT1_LSCC = LSCC[c(1, 2, 3, 5)]
colnames(FAT1_LSCC) = c("Study ID", "Sample ID", "Copy Number Alterations", "mRNA Expression")
FAT1_LSCC$'log2 (mRNA Expression)' = log(FAT1_LSCC$'mRNA Expression', base = 2)
# Pearson Correlation
cor.test(FAT1_LSCC$'Copy Number Alterations', FAT1_LSCC$'log2 (mRNA Expression)', method="pearson")
```

```
##
## Pearson's product-moment correlation
##
## data:  FAT1_LSCC$'Copy Number Alterations' and FAT1_LSCC$'log2 (mRNA Expression)'
## t = 5.5346, df = 464, p-value = 5.225e-08
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1616738 0.3321825
## sample estimates:
##          cor
## 0.2488554
```

```
# Spearman Correlation
cor.test(FAT1_LSCC$`Copy Number Alterations`, FAT1_LSCC$log2 (mRNA Expression), method="spearman")

## Warning in cor.test.default(FAT1_LSCC$`Copy Number Alterations`, FAT1_LSCC$log2
## (mRNA Expression)', : Cannot compute exact p-value with ties

##
## Spearman's rank correlation rho
##
## data: FAT1_LSCC$`Copy Number Alterations` and FAT1_LSCC$log2 (mRNA Expression)`
## S = 12519855, p-value = 1.671e-08
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.2576738
```

```
# Kendall Correlation
cor.test(FAT1_LSCC$`Copy Number Alterations`, FAT1_LSCC$log2 (mRNA Expression), method="kendall")

##
## Kendall's rank correlation tau
##
## data: FAT1_LSCC$`Copy Number Alterations` and FAT1_LSCC$log2 (mRNA Expression)`
## z = 5.6598, p-value = 1.516e-08
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
## tau
## 0.2054575
```

```
FAT1_LSCC$`Copy Number Alterations` = as.factor(FAT1_LSCC$`Copy Number Alterations`)
summary(FAT1_LSCC[4:5]) %>% knitr::kable()
```

mRNA Expression	log2 (mRNA Expression)
Min. : 86.63	Min. : 6.437
1st Qu.: 2255.52	1st Qu.:11.139
Median : 3784.18	Median :11.886
Mean : 4704.11	Mean :11.772
3rd Qu.: 5734.70	3rd Qu.:12.485
Max. :47344.60	Max. :15.531
NA's :3	NA's :3

```
P3.1 =
FAT1_LSCC %>%
  ggplot(aes(x = `Copy Number Alterations`, y = `mRNA Expression`)) + geom_boxplot() + theme_bw() +
  labs(caption = "FAT1 in Lung SCC")

P3.2 =
FAT1_LSCC %>%
  ggplot(aes(x = `Copy Number Alterations`, y = `log2 (mRNA Expression)`)) + geom_boxplot() + theme_bw()
```

```

labs(caption = "FAT1 in Lung SCC")

# Anova analysis
analysis3 = aov(FAT1_LSCC$log2 (mRNA Expression)~FAT1_LSCC$Copy Number Alterations')
summary.aov(analysis3)

##
##              Df Sum Sq Mean Sq F value    Pr(>F)
## FAT1_LSCC$Copy Number Alterations'    4    42.7   10.678    8.127 2.44e-06 ***
## Residuals                        461   605.7    1.314
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 3 observations deleted due to missingness

EGFR_LSCC = LSCC[c(1, 2, 4, 6)]
colnames(EGFR_LSCC) = c("Study ID", "Sample ID", "Copy Number Alterations", "mRNA Expression")
EGFR_LSCC$log2 (mRNA Expression)' = log(EGFR_LSCC$mRNA Expression', base = 2)
# Pearson Correlation
cor.test(EGFR_LSCC$Copy Number Alterations', EGFR_LSCC$log2 (mRNA Expression)', method="pearson")

##
## Pearson's product-moment correlation
##
## data:  EGFR_LSCC$Copy Number Alterations' and EGFR_LSCC$log2 (mRNA Expression)'
## t = 10.481, df = 464, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3610248 0.5081542
## sample estimates:
##      cor
## 0.4375131

# Spearman Correlation
cor.test(EGFR_LSCC$Copy Number Alterations', EGFR_LSCC$log2 (mRNA Expression)', method="spearman")

## Warning in cor.test.default(EGFR_LSCC$Copy Number Alterations', EGFR_LSCC$log2
## (mRNA Expression)', : Cannot compute exact p-value with ties

##
## Spearman's rank correlation rho
##
## data:  EGFR_LSCC$Copy Number Alterations' and EGFR_LSCC$log2 (mRNA Expression)'
## S = 9921580, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.4117305

# Kendall Correlation
cor.test(EGFR_LSCC$Copy Number Alterations', EGFR_LSCC$log2 (mRNA Expression)', method="kendall")

```

```
##
## Kendall's rank correlation tau
##
## data: EGFR_LSCC$'Copy Number Alterations' and EGFR_LSCC$'log2 (mRNA Expression)'
## z = 9.2136, p-value < 2.2e-16
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## 0.3296031
```

```
EGFR_LSCC$'Copy Number Alterations' = as.factor(EGFR_LSCC$'Copy Number Alterations')
summary(EGFR_LSCC[4:5]) %>% knitr::kable()
```

mRNA Expression	log2 (mRNA Expression)
Min. : 22.68	Min. : 4.504
1st Qu.: 1264.14	1st Qu.:10.304
Median : 2084.55	Median :11.026
Mean : 3558.23	Mean :10.917
3rd Qu.: 3289.62	3rd Qu.:11.684
Max. :80121.90	Max. :16.290
NA's :3	NA's :3

```
P4.1 =
EGFR_LSCC %>%
  ggplot(aes(x = 'Copy Number Alterations', y = 'mRNA Expression')) + geom_boxplot() + theme_bw() +
  labs(caption = "EGFR in Lung SCC")

P4.2 =
EGFR_LSCC %>%
  ggplot(aes(x = 'Copy Number Alterations', y = 'log2 (mRNA Expression)')) + geom_boxplot() + theme_bw() +
  labs(caption = "EGFR in Lung HNSCC")

# Anova analysis
analysis4 = aov(EGFR_LSCC$'log2 (mRNA Expression)'~EGFR_LSCC$'Copy Number Alterations')
summary.aov(analysis4)
```

```
##
##          Df Sum Sq Mean Sq F value Pr(>F)
## EGFR_LSCC$'Copy Number Alterations'    4   311.6    77.90   43.97 <2e-16 ***
## Residuals                            461   816.8     1.77
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 3 observations deleted due to missingness
```

```
FAT1_CSCC = CSCC[c(1, 2, 3, 5)]
colnames(FAT1_CSCC) = c("Study ID", "Sample ID", "Copy Number Alterations", "mRNA Expression")
FAT1_CSCC$'log2 (mRNA Expression)' = log(FAT1_CSCC$'mRNA Expression', base = 2)
# Pearson Correlation
cor.test(FAT1_CSCC$'Copy Number Alterations', FAT1_CSCC$'log2 (mRNA Expression)', method="pearson")
```

```
##
```

```
## Pearson's product-moment correlation
##
## data: FAT1_CSCC$'Copy Number Alterations' and FAT1_CSCC$'log2 (mRNA Expression)'
## t = 4.9332, df = 273, p-value = 1.408e-06
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.1736830 0.3911375
## sample estimates:
## cor
## 0.2860896

# Spearman Correlation
cor.test(FAT1_CSCC$'Copy Number Alterations', FAT1_CSCC$'log2 (mRNA Expression)', method="spearman")

## Warning in cor.test.default(FAT1_CSCC$'Copy Number Alterations', FAT1_CSCC$'log2
## (mRNA Expression)', : Cannot compute exact p-value with ties

##
## Spearman's rank correlation rho
##
## data: FAT1_CSCC$'Copy Number Alterations' and FAT1_CSCC$'log2 (mRNA Expression)'
## S = 2727472, p-value = 0.0003726
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.2131006

# Kendall Correlation
cor.test(FAT1_CSCC$'Copy Number Alterations', FAT1_CSCC$'log2 (mRNA Expression)', method="kendall")

##
## Kendall's rank correlation tau
##
## data: FAT1_CSCC$'Copy Number Alterations' and FAT1_CSCC$'log2 (mRNA Expression)'
## z = 3.5793, p-value = 0.0003446
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
## tau
## 0.1695754

FAT1_CSCC$'Copy Number Alterations' = as.factor(FAT1_CSCC$'Copy Number Alterations')
summary(FAT1_CSCC[4:5]) %>% knitr::kable()
```

mRNA Expression	log2 (mRNA Expression)
Min. : 14.7	Min. : 3.878
1st Qu.: 2106.8	1st Qu.:11.041
Median : 3432.4	Median :11.745
Mean : 3997.2	Mean :11.503
3rd Qu.: 5038.0	3rd Qu.:12.299
Max. :16725.8	Max. :14.030
NA's :3	NA's :3


```

P5.1 =
FAT1_CSCC %>%
  ggplot(aes(x = 'Copy Number Alterations', y = 'mRNA Expression')) + geom_boxplot() + theme_bw() +
  labs(caption = "FAT1 in Cervical SCC")

P5.2 =
FAT1_CSCC %>%
  ggplot(aes(x = 'Copy Number Alterations', y = 'log2 (mRNA Expression)')) + geom_boxplot() + theme_bw() +
  labs(caption = "FAT1 in Cervical SCC")

# Anova analysis
analysis5 = aov(FAT1_CSCC$log2 (mRNA Expression) ~ FAT1_CSCC$Copy Number Alterations)
summary.aov(analysis5)

##
##              Df Sum Sq Mean Sq F value    Pr(>F)
## FAT1_CSCC$Copy Number Alterations'    3   80.1   26.697    16.83 4.71e-10 ***
## Residuals                        271  429.9    1.586
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 3 observations deleted due to missingness

EGFR_CSCC = CSCC[c(1, 2, 4, 6)]
colnames(EGFR_CSCC) = c("Study ID", "Sample ID", "Copy Number Alterations", "mRNA Expression")
EGFR_CSCC$log2 (mRNA Expression)' = log(EGFR_CSCC$mRNA Expression', base = 2)
# Pearson Correlation
cor.test(EGFR_CSCC$Copy Number Alterations', EGFR_CSCC$log2 (mRNA Expression)', method="pearson")

##
## Pearson's product-moment correlation
##
## data:  EGFR_CSCC$Copy Number Alterations' and EGFR_CSCC$log2 (mRNA Expression)'
## t = 5.8643, df = 273, p-value = 1.3e-08
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.2251040 0.4355345
## sample estimates:
##      cor
## 0.3344819

# Spearman Correlation
cor.test(EGFR_CSCC$Copy Number Alterations', EGFR_CSCC$log2 (mRNA Expression)', method="spearman")

## Warning in cor.test.default(EGFR_CSCC$Copy Number Alterations', EGFR_CSCC$log2
## (mRNA Expression)', : Cannot compute exact p-value with ties

##
## Spearman's rank correlation rho
##
## data:  EGFR_CSCC$Copy Number Alterations' and EGFR_CSCC$log2 (mRNA Expression)'
## S = 2496435, p-value = 2.446e-06
## alternative hypothesis: true rho is not equal to 0

```

```
## sample estimates:
##      rho
## 0.2797567

# Kendall Correlation
cor.test(EGFR_CSCC$`Copy Number Alterations`, EGFR_CSCC$`log2 (mRNA Expression)`, method="kendall")

##
## Kendall's rank correlation tau
##
## data: EGFR_CSCC$`Copy Number Alterations` and EGFR_CSCC$`log2 (mRNA Expression)`
## z = 4.7227, p-value = 2.328e-06
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## 0.225934

EGFR_CSCC$`Copy Number Alterations` = as.factor(EGFR_CSCC$`Copy Number Alterations`)
summary(EGFR_CSCC[4:5]) %>% knitr::kable()
```

mRNA Expression	log2 (mRNA Expression)
Min. : 2.0	Min. : 1.003
1st Qu.: 440.4	1st Qu.: 8.783
Median : 833.6	Median : 9.703
Mean : 1584.6	Mean : 9.585
3rd Qu.: 1621.1	3rd Qu.:10.663
Max. :53811.0	Max. :15.716
NA's :3	NA's :3

```
P6.1 =
EGFR_CSCC %>%
  ggplot(aes(x = `Copy Number Alterations`, y = `mRNA Expression`)) + geom_boxplot() + theme_bw() +
  labs(caption = "EGFR in Cervical SCC")

P6.2 =
EGFR_CSCC %>%
  ggplot(aes(x = `Copy Number Alterations`, y = `log2 (mRNA Expression)`) + geom_boxplot() + theme_bw() +
  labs(caption = "EGFR in Cervical HNSCC")

# Anova analysis
analysis6 = aov(EGFR_CSCC$`log2 (mRNA Expression)`~EGFR_CSCC$`Copy Number Alterations`)
summary.aov(analysis6)
```

```
##
##      Df Sum Sq Mean Sq F value    Pr(>F)
## EGFR_CSCC$`Copy Number Alterations`    3   129.3    43.08    15.91 1.47e-09 ***
## Residuals                          271   733.6     2.71
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 3 observations deleted due to missingness
```

```
ggarrange(P1.1, P2.1,
          P3.1, P4.1,
          P5.1, P6.1,
          ncol = 2, nrow = 3)
```

```
## Warning: Removed 8 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 8 rows containing non-finite values (stat_boxplot).
```

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

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

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

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

