

# Review of decreasing cell adhesion protein during tumor progression

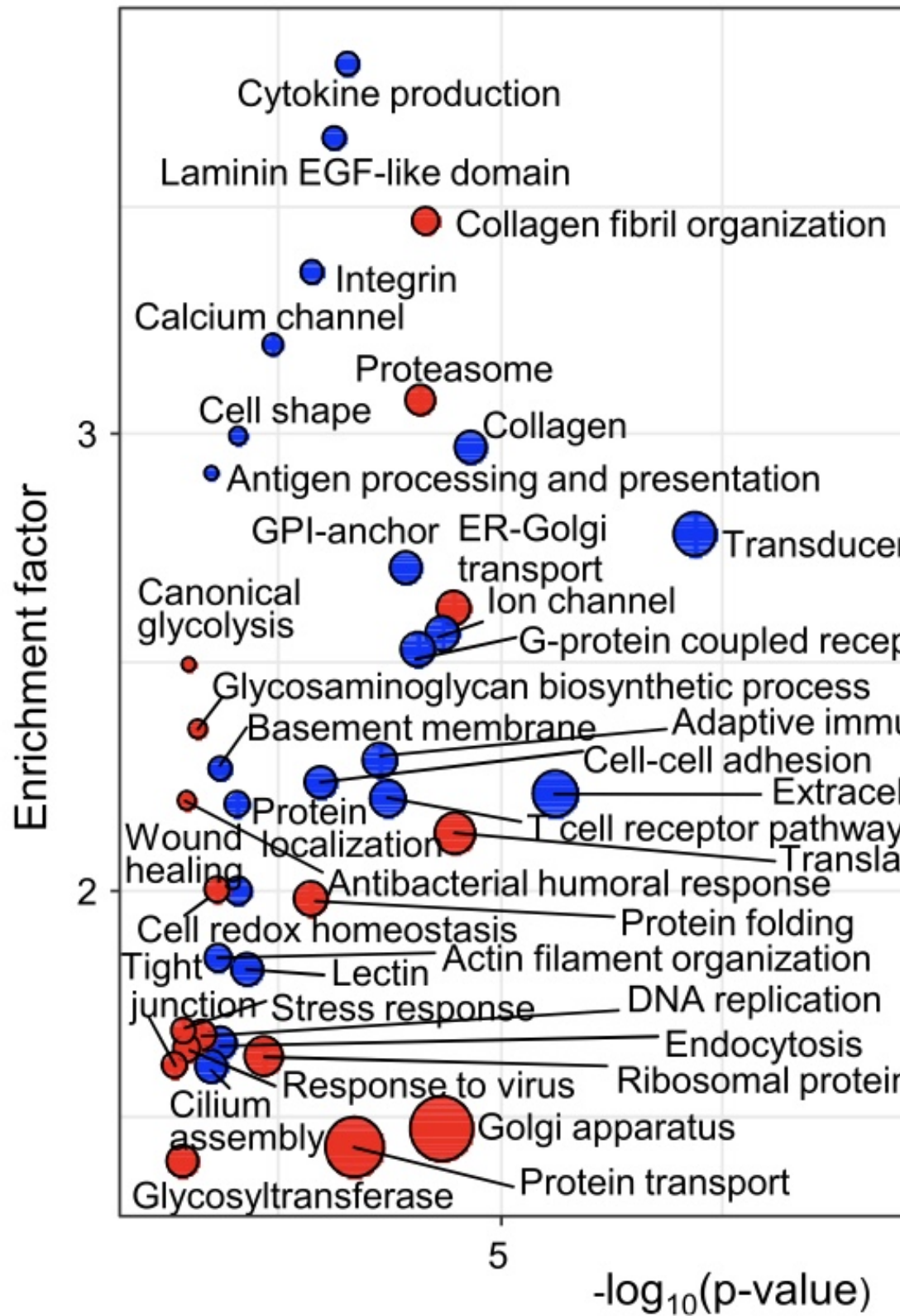
#Proteogenomics of Non-smoking Lung Cancer in East Asia Delinates Molecular Signatures of Pathogenesis and Progression

## 1.Introduction

This study are collected cohort from in Taiwan, representing early stage, predominantly female, non-smoking lung adenocarcinoma and provided clinical feature for Taiwan cohort. Each gene had distinct regulation patterns during tumors progression. Especially, I focused on cell adhesion gene and DNA replication gene. In Figure 2I, we can see that cell adhesion protein is decreasing as the tumor stage increases(blue circle).

```
library(knitr)
knitr::include_graphics("/Users/ddw20/Pictures/Saved Pictures/figure.jpg")
```

I



So I'll process supplementary table to figure out gene related with cell adhesion pathway and plot adhesion protein level in LUAD compared with DNA replication protein level. As a result, I'll check whether the actual TW cohort data shows down-regulation of adhesion pathway protein and consider the relationship between tumor and cell adhesion protein.

## 2. Data import

```
library(readxl)
library(tidyverse)
```

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

## v ggplot2 3.3.5      v purrr  0.3.4
## v tibble  3.1.4      v dplyr  1.0.7
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   2.0.1      v forcats 0.5.1

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

I need information about clinical features of patients, pathway of gene, and mRNA,protein log2Tumor/NAT value.

```
my_table1_S5 = read_excel('C:/Users/ddw20/Documents/bsms222_160_lee/TableS1.xlsx', sheet=5, na="NA")
my_table1_S6 = read_excel('C:/Users/ddw20/Documents/bsms222_160_lee/TableS1.xlsx', sheet=6, na="NA")
my_table3_S7 = read_excel('C:/Users/ddw20/Documents/bsms222_160_lee/TableS3.xlsx', sheet=7, na="NA")
my_table6_S2 = read_excel('C:/Users/ddw20/Documents/bsms222_160_lee/TableS6.xlsx', sheet=2, na="NA")
```

## 3. Data wrangling

Gene and related KEGG pathway is in my\_table3\_S7(Table S3F). I want to know some genes involved in the adhesion pathway and also filter genes related to DNA replication pathway. so i used filter and grepl.

```
adhesion<-my_table3_S7 %>% filter(grepl('adhesion', `C: KEGG pathway name`))%>%count(`T: Gene name`)%>%
replication<-my_table3_S7 %>% filter(grepl('replication', `C: KEGG pathway name`))%>%count(`T: Gene name`)%>%
```

my\_table6\_S2(Table S6A) has clinical features per patients. I converted row and column in my\_table6\_S2 to use tidy data.

```
my_table6_S22<-as.data.frame(t(my_table6_S2))
colnames(my_table6_S22)<-my_table6_S22[1,]
my_table6_S22<-my_table6_S22[-c(1,2,3,4,5),]
my_table6_S22<-tibble::rownames_to_column(my_table6_S22, "ID")
```

In my\_table1\_S6(Table S1E), there are so many genes and patients. I need expression level of only adhesion, replication gene.

```
my_table1_S66<-my_table1_S6%>%filter(Gene %in% c(adhesion,replication))%>%select(Gene, my_table6_S22$ID)
my_table1_S66<-as.data.frame(t(my_table1_S66))
colnames(my_table1_S66)<-my_table1_S66[1,]
my_table1_S66<-tibble::rownames_to_column(my_table1_S66, "ID")
my_table1_S66<-my_table1_S66[-1,]
```

Using `pivot_longer`, i made tidy data and added new column 'pathway' that distinguish DNA replication and Adhesion.

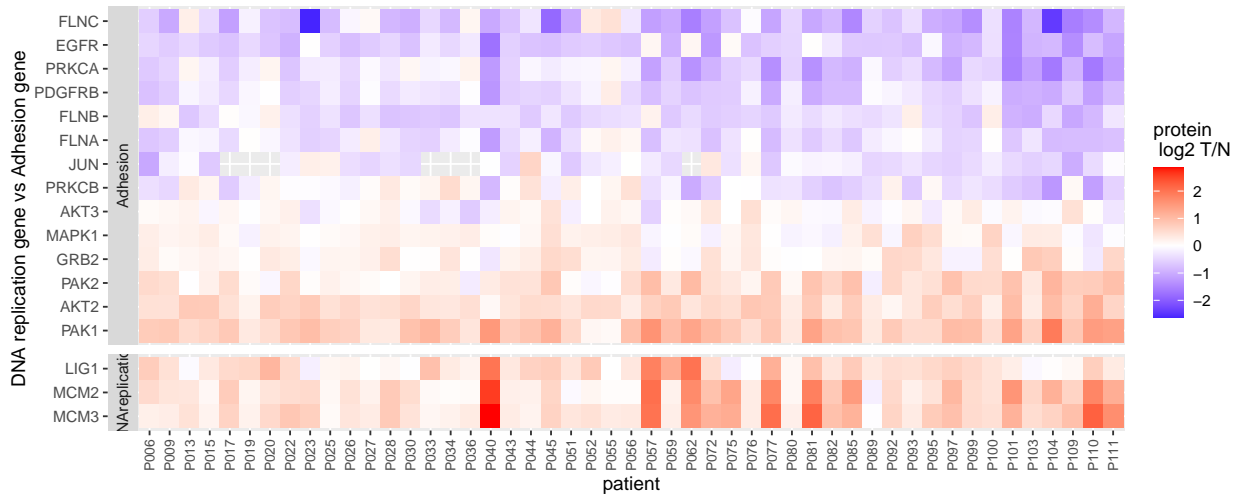
```
my_table1_S666<-my_table1_S66%>%pivot_longer(!ID,names_to="Gene",values_to="log2TN")%>%mutate(pathway=i
```

## 4. Visualizing

First, I merged `my_table1_S666` (Table S1E) and `my_table6_S22` (Table S6A) to make one data frame including patient ID, gene, pathway, protein log2 Tumor/NAT value. As we know, LUAD in TW cohort is characterized by predominant female patients. so I used only female patient information to make a plot. Through LUAD female cohort we can see DNA replication gene (MCM3, MCM2, LIG1) is up-regulation in tumor. It is natural because tumors proliferate abnormally. Then, how about cell adhesion gene? most adhesion genes were down-regulated in LUAD patients except AKT, GRB2, MAPK1, PAK gene. Despite all being involved in the adhesion pathway, Why does this difference occur? AKT, GRB2, MAPK1, PAK participate in other pathway. AKT is in not only focal adhesion but also EGFR signaling, autophagy, AMPK pathway etc. It can be seen that other proteins also participate in many different pathways. we can figure out through `my_table3_S7` and paper. especially, in Figure 2F, AKT, GRB2, MAPK is downstream protein of NLSCL pathway. and they cause antiapoptosis, proliferation..

```
merge(my_table1_S666, my_table6_S22, by='ID')%>%
  filter(!is.na(log2TN))%>%
  filter(Gender=="female")%>%
  select(ID, Gene, log2TN, pathway, AngiolymphaticInvasion)%>%
  mutate(log2TN=as.numeric(log2TN))%>%
  mutate(Gene=reorder(Gene, -log2TN))%>%
  filter(!is.na(AngiolymphaticInvasion))%>%
  ggplot(aes(ID, Gene, fill=log2TN))+
  geom_tile()+
  scale_fill_gradient2(high="red", mid="white", low="blue")+
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1, size=8))+
  facet_grid(pathway~., switch="both", scales = "free_y", space = "free_y")+
  labs(x="patient",
       y="DNA replication gene vs Adhesion gene",
       title="Figure1. Adhesion protein expression \n compared with DNA replication protein expression",
       fill="protein \n log2 T/N")
```

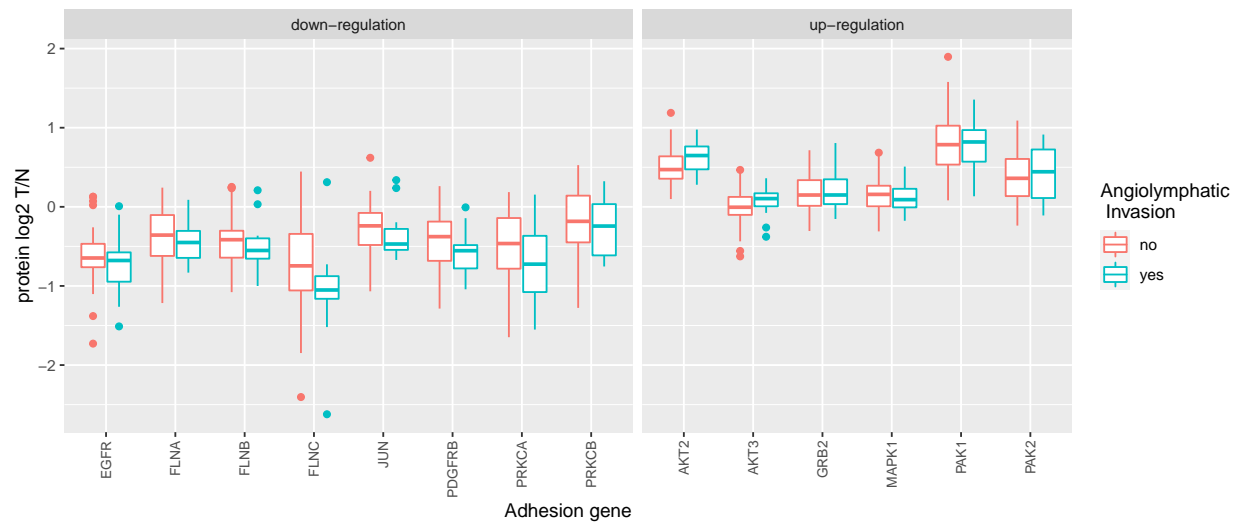
Figure1. Adhesion protein expression compared with DNA replication protein expression in female TW cohort



I associated angiolympathic invasion with cell adhesion. Instead of classifying into tumor stages, I used angiolympathic invasion information to compare only the presence or absence of metastasis with adhesion. We can see that protein log2 T/N of down-regulated adhesion genes in the first plot is decreasing when angiolympathic invasion is "yes". Median protein log2 T/N of boxplot is decreased. (AKT, GRB2, PAK still remain an exception; they also participate in other pathways.) Cells are connected to ECM, each other cell by cell-cell adhesion including adherens junctions and focal adhesions. As adhesion genes are down-regulated, tumor cells invade and migrate into blood and lymph node.

```
merge(my_table1_S666, my_table6_S22, by='ID') %>%
  filter(!is.na(log2TN)) %>%
  filter(Gender == "female") %>%
  select(ID, Gene, log2TN, pathway, AngiolympathicInvasion) %>%
  mutate(log2TN = as.numeric(log2TN)) %>%
  filter(!is.na(AngiolympathicInvasion)) %>%
  filter(pathway == "Adhesion") %>%
  mutate(regulation = ifelse(Gene %in% c("AKT2", "AKT3", "GRB2", "MAPK1", "PAK1", "PAK2"), "up-regulation", "down-regulation"))
ggplot(aes(Gene, log2TN, color = AngiolympathicInvasion)) +
  geom_boxplot() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1, size = 8)) +
  facet_grid(~regulation, scales = "free_x", space = "free_x") +
  labs(x = "Adhesion gene",
       y = "protein log2 T/N",
       title = "Figure2. Adhesion protein expression level based on Angiolympathic Invasion ",
       color = "Angiolympathic \n Invasion")
```

Figure2. Adhesion protein expression level based on Angiolymphatic Invasion



## 5. Discussion

Patients with Angiolymphatic Invasion have down-regulation adhesion gene protein levels. Because cell migrations are depending on the cell adhesion properties. Dysregulation of adhesion gene is associated with many pathological states including Angiolymphatic Invasion, cancer metastasis. In addition to the genes shown in the Figure, there are many genes participating in cell to cell adhesion or communication. Also, the genes in the Figure are not only involved in the adhesion pathway. There are genes that is related with other pathway(eg.NSCLC pathway) as well as adhesion pathway, so some genes are upregulated despite of their adhesion pathway. In general, as the tumor stage increases, the adhesion protein log2Tumor/NAT decreases.

```
tinytex::install_tinytex()
```