

# The Genetic, Transcriptomic and Protoemic Mutational Profile of Visceral Pleural Invasion in LUAD

based on Chen et al., 2020.

Yan Ting 2019250163

## Introduction

Visceral Pleural Invasion(VPI) is defined as penetration of tumor across the elastic layer of visceral pleura and is classified into 4 stages: PL0, PL1, PL2, and PL3. PL0 indicates tumor without pleural involvement beyond its elastic layer. PL1 indicates tumor that is not exposed on the pleural surface but penetrate beyond its elastic layer. While PL2 is defined as tumor that invades to the pleural surface, PL3 indicates those invading to the parietal pleura. (Travis, 2008). According to TNM classification 8th edition, the visceral pleural invasion status may increase the T staging of lung tumor below 3cm from T1 stage to T2 stage. Visceral pleural invasion is also generally considered as a poor prognosis factor in non-small cell lung cancer (Seok, 2017). Considering the importance of visceral pleural invasion as prognosis factor of NSCLC, the genetic mutation that may associate with the ability and pathway permitting visceral pleural invasion might be useful in finding new therapeutic targets while increasing the progression-free survival of LUAD patients.

## Visualising Data

### 1. Implication of Genetic Mutation (especially EGFR) on Visceral Pleural Invasion Ability

Here patients are divided into 2 categories: **patients with VPI (PL1, PL2, PL3)** and patients **without VPI (PL0)**. Analysis of gene mutation is conducted to determine the genes that may significantly associate with VPI. The data set recording genes implicated in cancer according to the Cancer Gene Census (COSMIC) is imported as reference to find the possible candidate genes associated with lung cancer. Two data sets are compared to obtain the genes that recur as high-ranking implicated genes.

```
##loading packages
library(dplyr)
library(tidyverse)
library(readxl)
library(data.table)
library(ggthemes)
library(ggsignif)
library(cowplot)

##plot 1a
#loading data
df <- read_excel("portfolio_6.xlsx", sheet = 2) %>% filter(`Gene name` == "na")
```

```

#setting header and tidying data
basic_info <- as.data.frame(t(df)) %>%
  set_names(. [1,]) %>%
  . [-(1:5),] %>% setDT(.,keep.rownames = "ID")

#determining the important candidate gene
TCGA_data <- read_excel("TCGA_data.xlsx", sheet = 4, skip = 2)
the_mutation <- unique(intersect(TCGA_data$Gene[1:100],gsub("_mut", "", df$id)))

#summarizing the percentage of mutation-implicated VPI vs. non-VPI patients
graph_table<- basic_info %>%
  filter(VisceralPleuralStatus != "NA") %>%
  mutate(VisceralPleuralStatus =
    case_when(VisceralPleuralStatus %in% c("P1","P2","P3") ~ "VPI",
              TRUE ~ "non-VPI")) %>%
  group_by(VisceralPleuralStatus, .drop = FALSE) %>%
  summarise(total = n(),
            sum_EGFR_del = sum(exon19del == "yes"),
            sum_EGFR_subs = sum(L858R == "yes"),
            sum_TP53 = sum(TP53_mut == "1"),
            sum_KRAS = sum(KRAS_mut == "1"),
            sum_RBM10 = sum(RBM10_mut == "1"),
            sum_NF1 = sum(NF1_mut == "1"),
            sum_RB1 = sum(RB1_mut == "1"),
            sum_PIK3CA = sum(PIK3CA_mut == "1"),
            sum_ATM = sum(ATM_mut == "1"),
            sum_CDC27 = sum(CDC27_mut == "1"),
            sum(APC_mut == "1")) %>%
  mutate(EGFR_exon19del = sum_EGFR_del/total*100,
        EGFR_L858R = sum_EGFR_subs/total*100,
        TP53 = sum_TP53/total*100,
        KRAS = sum_KRAS/total*100,
        RBM10 = sum_RBM10/total*100,
        NF1 = sum_NF1/total*100,
        RB1 = sum_RB1/total*100,
        PIK3CA = sum_PIK3CA/total*100,
        ATM = sum_ATM/total*100,
        CDC27 = sum_CDC27/total*100,
        APC = sum(APC/total*100)

#plotting graph
P1 <- graph_table %>%
  pivot_longer(cols = EGFR_exon19del:APC,
               names_to = "percent_name", values_to = "percent_value" ) %>%
  mutate(VisceralPleuralStatus2 =
    paste(VisceralPleuralStatus, "( N =", total, ")", sep = " ")) %>%
  mutate(VisceralPleuralStatus2 = factor(VisceralPleuralStatus2,
                                         levels = c("non-VPI ( N = 63 )","VPI ( N = 18 )")) ) %>%
  ggplot(aes(x = percent_name, y = percent_value, fill = VisceralPleuralStatus2)) +

```

```

geom_bar(stat = "identity", position = "dodge")+
scale_fill_brewer(palette = "Accent")+
theme_few()+
labs(x = "",
y = "Number of Cases / %",
title = "Number of Cases for each Gene Mutation")+
guides(fill = guide_legend(title = "Visceral Pleural Status")) +
scale_x_discrete(labels = c("EGFR_exon19del" = "EGFR del", "EGFR_L858R" = "EGFR L858R" ))+
theme(axis.text.x = element_text(size = 8, angle = 90, hjust = 1),
axis.text.y = element_text(size = 8))+
theme(plot.title = element_text(size = 9, face = "bold"),
legend.position = c(0.7, 0.8),
legend.title = element_text(size = 7),
legend.text = element_text(size = 7))

##plot1b

#loading data
transcriptome <- read_excel("portfolio_1.xlsx", sheet = 5, col_names = T, na = "")

proteome <- read_excel("portfolio_1.xlsx", sheet = 6, col_names = T, na = "NA")

basic_info2 <- read_excel("portfolio_1.xlsx", sheet = 2)

#merging and tidying up data
graph_table2 <- merge(
  as.data.frame(t(transcriptome %>% filter(gene %in% the_mutation | gene == "EGFR"))) %>%
    set_names(.[1,]) %>%
    .[-c(1:2),] %>%
    setDT(., keep.rownames = "ID"),
  as.data.frame(t(proteome %>% filter(Gene %in% the_mutation | Gene == "EGFR"))) %>%
    set_names(.[2,]) %>%
    .[-c(1:3),] %>%
    setDT(., keep.rownames = "ID"),
  by = "ID") %>%
  merge(., basic_info, by = "ID") %>%
  pivot_longer(cols = CDC27.x:NF1.x, names_to = "gene.x", values_to = "rna_expression") %>%
  pivot_longer(cols = EGFR.y:TP53.y, names_to = "gene.y", values_to = "protein_expression") %>%
  mutate(gene.x = gsub(".x", "", gene.x)) %>%
  mutate(gene.y = gsub(".y", "", gene.y)) %>%
  filter(gene.x == gene.y) %>%
  select(-gene.y) %%%
  rename(gene = gene.x) %>%
  mutate(rna_expression = as.numeric(rna_expression),
protein_expression = as.numeric(protein_expression))

#plotting graph
P2 <-
  merge(graph_table2, basic_info2 %>% select(ID, EGFR_Status), by = "ID", all.x = T) %>%
  mutate(VisceralPleuralStatus = ifelse(VisceralPleuralStatus %in% c("P1", "P2", "P3"),
                                         "VPI", "non-VPI")) %>%
  filter(gene == "EGFR" & !EGFR_Status %in% c("L858R.exon19del", "others")) %>%

```

```

ggplot(aes(x = rna_expression, y = protein_expression, col = VisceralPleuralStatus)) +
  geom_point()+
  scale_x_continuous(limits = c(-2, 2))+ 
  geom_xsideboxplot(aes(y = VisceralPleuralStatus), orientation = "y")+
  geom_ysideboxplot(aes(x = VisceralPleuralStatus), orientation = "x")+
  scale_xsidey_discrete(theme(axis.text = element_blank()))+
  scale_ysidex_discrete()
  facet_wrap(~EGFR_Status, ncol = 1, strip.position = "l")+
  theme(axis.text = element_text(size = 8))+
  theme(ggsave.panel.scale.x = .3,
        ggsave.panel.scale.y = .2)+
  scale_ysidex_discrete(guide = guide_axis(position = "none")) +
  scale_xsidey_discrete(guide = guide_axis(position = "none")) +
  labs(x = "RNA Expression",
       y = "Protein Expression",
       title = "Difference of Transcriptomic and Proteomic
       Expression among EGFR subtypes") +
  guides(fill = guide_legend(title = "Visceral Pleural Status"))+
  theme(plot.title = element_text(size = 8.5, face = "bold", hjust = 0),
        axis.text.x = element_text(size = 8),
        legend.position = "none")+
  scale_color_brewer(palette = "Accent")

#showing grpah
plot_grid(P1, P2, nrow = 1, rel_widths = c(0.9, 1))

```



| The bar graph shows the number of cases of VPI and non-VPI for the top ranked gene mutation implicated in COSMIC and our data set. In accordance with several studies that discover the correlation between EGFR-L858R mutant and lung cancer invasion ability, for example in Tsai et al.,(2015), data analysis on Taiwan LUAD cohort also shows an overrepresentation of patients showing VPI in comparison to those who do not. In contrast, despite several studies also propose the role of EGFR-exon19del mutant in cancer cell invasion ability, the conclusion is not reproduced in Taiwan cohort, indicating the heterogeneity implicated by different types and position of EGFR mutation. Besides, NF1, RB1, and TP53 show positive correlation with VPI while KRAS gene mutation, which is generally recognized as cancer-driver mutation, is not enriched in Taiwan LUAD cohort, representing its potential as a negative VPI regulator.

In line with the suspected heterogeneity in EGFR mutant type, the point graph on the right shows the RNA expression(x-axis) and protein expression(y-axis) of EGFR corresponding to their mutant type. The data for patients who express both L858R and exon19del mutation other EGFR mutation type are not included because of the inadequacy of sample size( $n < 3$ ). Based on the comparison, it is shown that most of the patients, regardless of VPI status and EGFR mutant type, have their EGFR proteins negatively regulated while displaying increased expression at RNA level. However, there is no significant difference of EGFR expression between VPI and non-VPI patient at transcriptomic and proteomic level, despite the difference at genomic level. It is believed that larger sample size is required to validate the results considering the number of VPI-affected patients is too few.

## 2. Pathways related to visceral pleural invasion

Below attempts to find a correlation between Visceral Pleural Invasion status with EGFR signaling pathway-associated genes, cell cycle genes, and epithelial-mesothelial transition (EMT) associated genes. EGFR signaling pathway associated genes are obtained from Wikopathways, while gene set associated with cell cycle is obtained from KEGG. The genes associated to EMT was obtained by referring papers where their findings discovered genes where their mRNA expression levels are significantly associated EMT marker(Jung et al., 2020), and genes preliminarily identified as the key genes in LUAD. (Wang et al., 2020)

```
proteome <- read_excel("portfolio_1.xlsx", sheet = 6, col_names = T, na = "NA")

#pulling EGFR pathway related genes
EGFR_gene <- read.table("EGFR_signaling_associated_gene_set.txt") %>% pull(V1)

EGFR_gene2 <- as.data.frame(t(proteome %>% filter(Gene %in% EGFR_gene))) %>%
  set_names(.[2,]) %>%
  .[-c(1:3),] %>%
  .[, apply(., 2, function(x) !any(is.na(x)))] %>%
  setDT(., keep.rownames = "ID") %>%
  merge(., basic_info %>% select(ID, VisceralPleuralStatus), by= "ID") %>%
  mutate(VisceralPleuralStatus2 = ifelse(VisceralPleuralStatus %in% c("P1","P2","P3"),
                                         "VPI", "non-VPI")) %>%
  mutate_at(c(2:116), as.numeric) %>%
  group_by(VisceralPleuralStatus2) %>%
  summarise(across(2:116, median))

the_EGFR_gene <-
  as.data.frame(t(EGFR_gene2)) %>% set_names(.[1,]) %>%
  .[-1,] %>% setDT(., keep.rownames = "gene") %>%
  mutate_at(c(2:3),as.numeric) %>%
  mutate(foldchange = (VPI - `non-VPI`)) %>%
  mutate(ranking = ifelse(foldchange<0, foldchange*(-1), foldchange)) %>%
```

```

arrange(., desc(ranking)) %>%
head(10) %>%
pull(gene)

#pulling Cell cycle related genes
cell_cycle <- read_excel("portfolio_3.xlsx", sheet = 6) %>%
  set_names(.[1,]) %>% .[-1,] %>%
  filter(str_detect(.\$KEGG, "Cell cycle") == T ) %>% pull(Gene)

cell_cycle2 <- as.data.frame(t(proteome %>% filter(Gene %in% cell_cycle))) %>%
  set_names(.[2,]) %>%
  .[-c(1:3),] %>%
  .[, apply(., 2, function(x) !any(is.na(x)))] %>%
  setDT(., keep.rownames = "ID") %>%
  merge(., basic_info %>% select(ID, VisceralPleuralStatus), by= "ID") %>%
  mutate(VisceralPleuralStatus2 = ifelse(VisceralPleuralStatus %in% c("P1","P2","P3"),
                                         "VPI", "non-VPI")) %>%
  mutate_at(c(2:36), as.numeric) %>%
  group_by(VisceralPleuralStatus2) %>%
  summarise(across(2:36, median))

the_cell_cycle <-
  as.data.frame(t(cell_cycle2)) %>% set_names(.[1,]) %>%
  .[-1,] %>% setDT(., keep.rownames = "gene") %>%
  mutate_at(c(2:3),as.numeric) %>%
  mutate(foldchange = (VPI - `non-VPI`)) %>%
  mutate(ranking = ifelse(foldchange<0, foldchange*(-1), foldchange)) %>%
  arrange(., desc(ranking)) %>%
  head(10) %>%
  pull(gene)

#pulling EMT_marker
EMT_marker <- c("CDH1", "VIM", "CDH2", "FN1", "CEACAM5", "NQO1", "LCN2", "CXCL12" )

the_genes <- c(the_EGFR_gene, the_cell_cycle, EMT_marker)

#filtering proteomic data
plot_table1 <- as.data.frame(t(proteome %>% filter(Gene %in% the_EGFR_gene))) %>%
  set_names(.[2,]) %>%
  .[-c(1:3),] %>%
  .[, apply(., 2, function(x) !any(is.na(x)))] %>%
  setDT(., keep.rownames = "ID") %>%
  merge(., basic_info %>% select(ID, VisceralPleuralStatus), by= "ID") %>%
  pivot_longer(cols = 2:11, names_to = "gene", values_to = "protein_expression" ) %>%
  mutate(protein_expression = as.numeric(protein_expression)) %>%
  mutate(VisceralPleuralStatus2 = ifelse(VisceralPleuralStatus %in% c("P1","P2","P3"),
                                         "VPI", "non-VPI"))

plot_table2 <- as.data.frame(t(proteome %>% filter(Gene %in% the_cell_cycle))) %>%
  set_names(.[2,]) %>%
  .[-c(1:3),] %>%

```

```

. [ , apply(., 2, function(x) !any(is.na(x)))] %>%
  setDT(., keep.rownames = "ID") %>%
merge(., basic_info %>% select(ID, VisceralPleuralStatus), by= "ID") %>%
pivot_longer(cols = 2:11, names_to = "gene", values_to = "protein_expression" ) %>%
mutate(protein_expression = as.numeric(protein_expression)) %>%
mutate(VisceralPleuralStatus2 = ifelse(VisceralPleuralStatus %in% c("P1","P2","P3"),
                                         "VPI", "non-VPI"))

plot_table3 <- as.data.frame(t(proteome %>% filter(Gene %in% EMT_marker))) %>%
  set_names(.[2,]) %>%
  .[-c(1:3),] %>%
  .[ , apply(., 2, function(x) !any(is.na(x)))] %>%
  setDT(., keep.rownames = "ID") %>%
merge(., basic_info %>% select(ID, VisceralPleuralStatus), by= "ID") %>%
pivot_longer(cols = 2:8, names_to = "gene", values_to = "protein_expression" ) %>%
mutate(protein_expression = as.numeric(protein_expression)) %>%
mutate(VisceralPleuralStatus2 = ifelse(VisceralPleuralStatus %in% c("P1","P2","P3"),
                                         "VPI", "non-VPI"))

#plotting graph
p1 <-
  plot_table1 %>%
  mutate(gene = factor(gene)) %>%
  mutate(gene = reorder(gene, protein_expression, median)) %>%
  ggplot(aes(x = ID, y = gene, fill = protein_expression))+ 
  geom_tile(color = "white") + 
  scale_fill_gradient2(low = "blue", high = "red", mid = "white",
                       name = "Protein \n Expression Level")+
  scale_x_discrete(limits=unique((plot_table1$ID)[order(plot_table1$VisceralPleuralStatus2)]))+ 
  theme(axis.text.x = element_blank(),
        legend.text=element_text(size = 9),
        legend.title = element_text(size = 9))+ 
  labs(y = "EGFR_pathway")

p2 <-
  plot_table2 %>%
  mutate(gene = factor(gene)) %>%
  mutate(gene = reorder(gene, protein_expression, median)) %>%
  ggplot(aes(x = ID, y = gene, fill = protein_expression))+ 
  geom_tile(color = "white") + 
  scale_fill_gradient2(low = "blue", high = "red", mid = "white",
                       name = "Expression Level")+
  scale_x_discrete(limits=unique((plot_table1$ID)[order(plot_table1$VisceralPleuralStatus2)]))+ 
  theme(axis.text.x = element_blank())+
  theme(legend.position = "none")+
  labs(x = "",
       y = "Cell cycle")

p3 <-
  plot_table3 %>%
  mutate(gene = factor(gene)) %>%
  mutate(gene = reorder(gene, protein_expression, median)) %>%

```

```

ggplot(aes(x = ID, y = gene, fill = protein_expression))+  

  geom_tile(color = "white") +  

  scale_fill_gradient2(low = "blue", high = "red", mid = "white",  

    name = "Expression Level") +  

  scale_x_discrete(limits=unique((plot_table1$ID)[order(plot_table1$VisceralPleuralStatus2)])) +  

  theme(axis.text.x = element_blank()) +  

  theme(legend.position = "none") +  

  labs(x = "",  

    y = "EMT-related")

p4 <-  

  plot_table1 %>%  

  select(ID, VisceralPleuralStatus2) %>%  

  mutate(VisceralPleuralStatus2 = factor(VisceralPleuralStatus2,  

    levels = c("VPI", "non-VPI")) %>%  

  ggplot(aes(x = ID, y = 1, fill = VisceralPleuralStatus2)) +  

  scale_x_discrete(limits=unique((plot_table1$ID)[order(plot_table1$VisceralPleuralStatus2)])) +  

  geom_tile() +  

  theme_void() +  

  scale_fill_grey(start=0, end = 0.9, name = "Visceral Pleural Status") +  

  theme(legend.text=element_text(size = 9),  

    legend.title = element_text(size = 9))

p5 <- basic_info2 %>%  

  select(ID, Gender, Age, `Smoking Status`) %>%  

  mutate(Gender = factor(Gender)) %>%  

  mutate(`Smoking Status` = factor(`Smoking Status`)) %>%  

  mutate(Status = ifelse(`Smoking Status` == "Nonsmoke", "Non-smoker", "Current Smoker\nor Ex-smoker"))  

  ggplot(aes(x = ID, y = 1, fill = Status)) +  

  scale_x_discrete(limits=unique((plot_table1$ID)[order(plot_table1$VisceralPleuralStatus2)])) +  

  geom_tile() +  

  theme_void() +  

  scale_fill_brewer(palette = "Set1") +  

  theme(legend.text=element_text(size = 9),  

    legend.title = element_text(size = 9))

legend <- plot_grid(get_legend(p1), get_legend(p4), get_legend(p5), ncol = 1)

p1 <- p1 +  

  theme(legend.position = "none") +  

  labs(x = "")  

p4 <- p4 +theme(legend.position = "none")  

p5 <- p5 +theme(legend.position = "none")

plot <- plot_grid(p5, p4, p1, p2, p3, align = "v", ncol = 1, axis = "tb", rel_heights = c(1,1,10,10,10))

plot_grid(plot, legend, nrow = 1, rel_widths = c(4, 1), hjust= (c(0, -1)))

```

By comparing the protein expression level of three pathways between patients who show visceral pleural invasion and patients who do not, modest correlation is observed, except EMT-related genes like NQO1

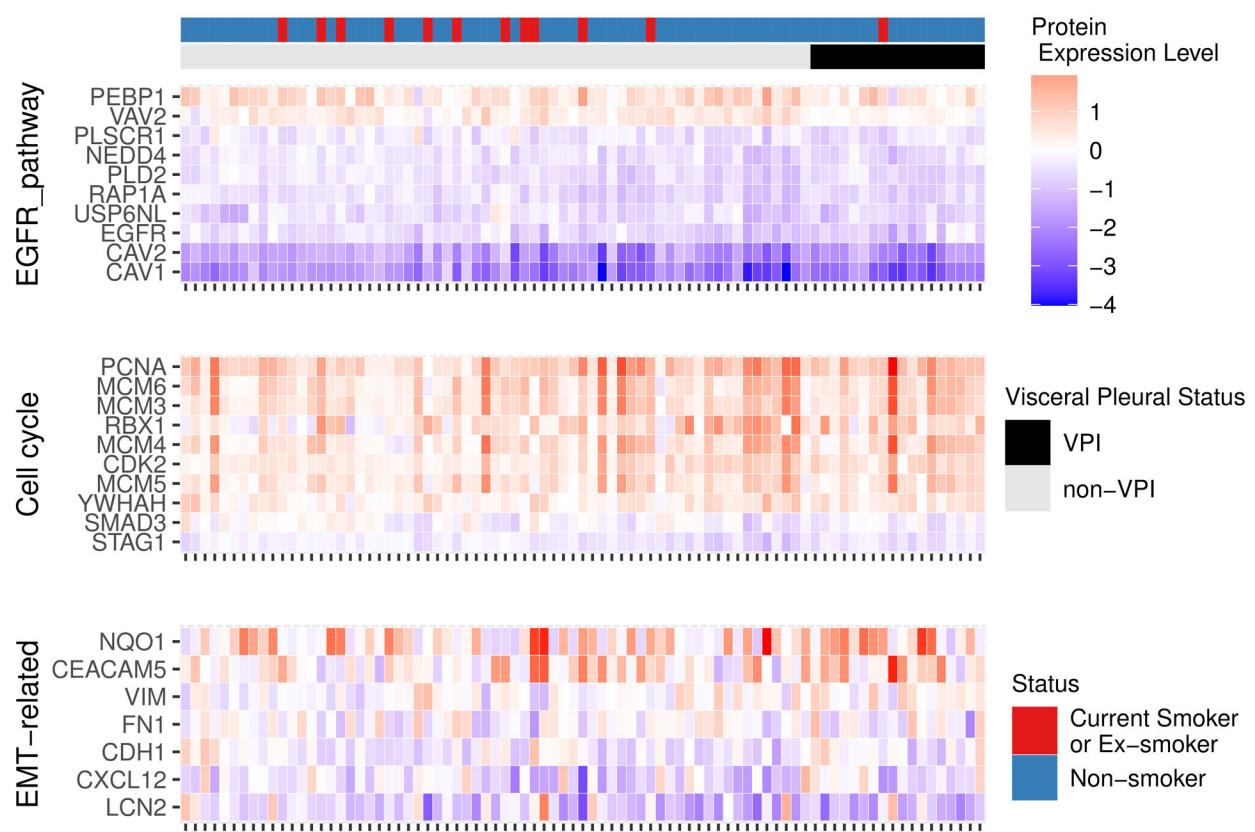


Figure 1: Protein expression of EGFR signaling pathway, cell cycle, and EMT associated proteins

and CEACAM5 showing slightly greater upregulation, and LCN2 showing slightly greater downregulation in VPI-positive group, highlighting the role of EMT-pathway in activating tumor migration and invasion ability. Besides, it is observed that most of the VPI-positive patients are never-smoker, indicating the effect exerted on tumor invasion ability from environmental or other exterior factors rather than smoking.

## Conclusion and Discussion

Overall, the data exploration shows that EGFR L858R, TP53 and few other genes are enriched in visceral pleural invasion while KRAS and ATM are underrepresented in LUAD VPI group. Particularly, among EGFR subtypes, only EGFR L858R shows positive correlation to VPI, and it is noted the correlation is not represented at transcriptomic neither proteomic level. Besides, modest correlation is discovered between VPI and protein expression of EGFR pathway related as well as cell cycle genes, despite their robust positive correlation with TP53 (Chen et al., 2020). However, a slight hint of correlation is shown between protein expression of EMT-related genes and VPI. It is believed that a larger sample size with adequate amount of VPI-positive cases is required to produce more reliable analysis.

Visceral Pleural Invasion is one of the important hallmarks in LUAD tumor progression before metastasis. The postoperative survival probability in patients who have N0 or N1 metastasis could decrease significantly with the presence of visceral pleural invasion (Adachi et al., 2015). Therefore if we successfully figure out the potential associated driver gene mutations or aberrations at transcriptomic and proteomic level that distinguish a tumor cell which invades visceral pleura and one which does not, new therapeutic target could be established.

## References

- Chen YJ, Roumeliotis TI, Chang YH, Chen CT, Han CL, Lin MH, Chen HW, Chang GC, Chang YL, Wu CT, Lin MW, Hsieh MS, Wang YT, Chen YR, Jonassen I, Ghavidel FZ, Lin ZS, Lin KT, Chen CW, Sheu PY, Hung CT, Huang KC, Yang HC, Lin PY, Yen TC, Lin YW, Wang JH, Raghav L, Lin CY, Chen YS, Wu PS, Lai CT, Weng SH, Su KY, Chang WH, Tsai PY, Robles AI, Rodriguez H, Hsiao YJ, Chang WH, Sung TY, Chen JS, Yu SL, Choudhary JS, Chen HY, Yang PC, Chen YJ. Proteogenomics of Non-smoking Lung Cancer in East Asia Delineates Molecular Signatures of Pathogenesis and Progression. *Cell.* 2020 Jul 9;182(1):226-244.e17. doi: 10.1016/j.cell.2020.06.012. PMID: 32649875.
- Travis WD, Brambilla E, Rami-Porta R, Vallières E, Tsuboi M, Rusch V, Goldstraw P; International Staging Committee. Visceral pleural invasion: pathologic criteria and use of elastic stains: proposal for the 7th edition of the TNM classification for lung cancer. *J Thorac Oncol.* 2008 Dec;3(12):1384-90. doi: 10.1097/JTO.0b013e31818e0d9f. PMID: 19057261.
- Seok Y, Jeong JY, Lee E. Extent of visceral pleural invasion and the prognosis of surgically resected node-negative non-small cell lung cancer. *Thorac Cancer.* 2017 May;8(3):197-202. doi: 10.1111/1759-7714.12424. Epub 2017 Feb 21. PMID: 28220643; PMCID: PMC5415476.
- Wang W, He J, Lu H, Kong Q, Lin S. KRT8 and KRT19, associated with EMT, are hypomethylated and over-expressed in lung adenocarcinoma and link to unfavorable prognosis. *Biosci Rep.* 2020;40(7):BSR20193468. doi:10.1042/BSR20193468
- Tsai, MF., Chang, TH., Wu, SG. et al. EGFR-L858R mutant enhances lung adenocarcinoma cell invasive ability and promotes malignant pleural effusion formation through activation of the CXCL12-CXCR4 pathway. *Sci Rep* 5, 13574 (2015). <https://doi.org/10.1038/srep13574>
- Adachi H, Tsuboi M, Nishii T, Yamamoto T, Nagashima T, Ando K, Ishikawa Y, Woo T, Watanabe K, Kumakiri Y, Maehara T, Morohoshi T, Nakayama H, Masuda M. Influence of visceral pleural invasion on survival in completely resected non-small-cell lung cancer. *Eur J Cardiothorac Surg.* 2015 Nov;48(5):691-7; discussion 697. doi: 10.1093/ejcts/ezu515. Epub 2015 Jan 5. PMID: 25564209.