Correlation between Primary Tumor Location and Age in nonsmokers

Code

BSMS222 Jihyun Seo(2015250126)

1. Introduction

In resource paper ‘Proteogenomics of Non-smoking Lung Cancer in East Asia Delineates Molecular Signature of Pathogenesis and Progression’, focusing on clinical data from Taiwan(TW) cohort, what I was wondering is about primary tumor location in treatment-naive patients from TW, especially whose histology type is ADC(adenocacinoma) and who are nonsmokers. Which factor among paients’ characters such as age, gender, ADC stage or EGFR_status would have correlation to tumor location?

Description of tumor location in lung. RUL = Right Upper Lobe, RML = Right Middle Lobe, RLL = Right Lower Lobe LUL = Left Upper Lobe, LLL = Left Lower Lobe

2. Dataset $ Visualizing

Hide

```
library(dplyr)
```

```


다음의 패키지를 부착합니다: ‘dplyr’

The following objects are masked from ‘package:stats’:

    filter, lag

The following objects are masked from ‘package:base’:

    intersect, setdiff, setequal, union
```

Hide

```
library(ggplot2)
library(readxl)
library(tidyverse)
```

```
Registered S3 methods overwritten by 'dbplyr':
  method         from
  print.tbl_lazy
  print.tbl_sql
-- Attaching packages --------------------------------------------------------------- tidyver
se 1.3.1 --
√ tibble  3.1.4     √ purrr   0.3.4
√ tidyr   1.1.3     √ stringr 1.4.0
√ readr   2.0.1     √ forcats 0.5.1
-- Conflicts ------------------------------------------------------------------ tidyverse_con
flicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
```

Hide

```
d1 <- read_excel("mmc1.xlsx", sheet = "Table S1A_clinical_103patient")
head(d1)
```

ID	Proteome_Batch	Gen…	Age	Smoking Status	Histology Type	St…	EGFR_Stat…
<chr>	<chr>	<chr>	<dbl>	<chr>	<chr>	<chr>	<chr>
P002	B01-2	Male	73.77687	Nonsmoke	ADC	IB	others
P004	B01-4	Female	52.97741	Nonsmoke	SCC	IA	exon19del
P005	B02-1	Male	72.75017	Current_Smoker	SCC	IA	WT
P006	B02-2	Female	46.86105	Nonsmoke	ADC	IB	WT
P007	B02-3	Male	67.40589	Nonsmoke	ADC	IIA	WT
P009	B03-1	Female	53.80424	Nonsmoke	ADC	IIA	L858R

6 rows | 1-8 of 9 columns

First of all, select data of only nonsmokers and ADC patients from the clinical data source given from paper. And select 4 factors(Gender, Age, Stge, EGFR_Status) to focus on them.

Hide

```
d2 <- d1 %>%
  filter(`Smoking Status` == 'Nonsmoke', `Histology Type` == "ADC") %>%
  select('Gender', 'Age', 'Stage', 'EGFR_Status', 'Primary Tumor Location')
d2
```

Gender	Age	Stage	EGFR_Status	Primary Tumor Location
<chr>	<dbl>	<chr>	<chr>	<chr>
Male	73.77687	IB	others	LUL
Female	46.86105	IB	WT	RLL
Male	67.40589	IIA	WT	RLL
Female	53.80424	IIA	L858R	LLL

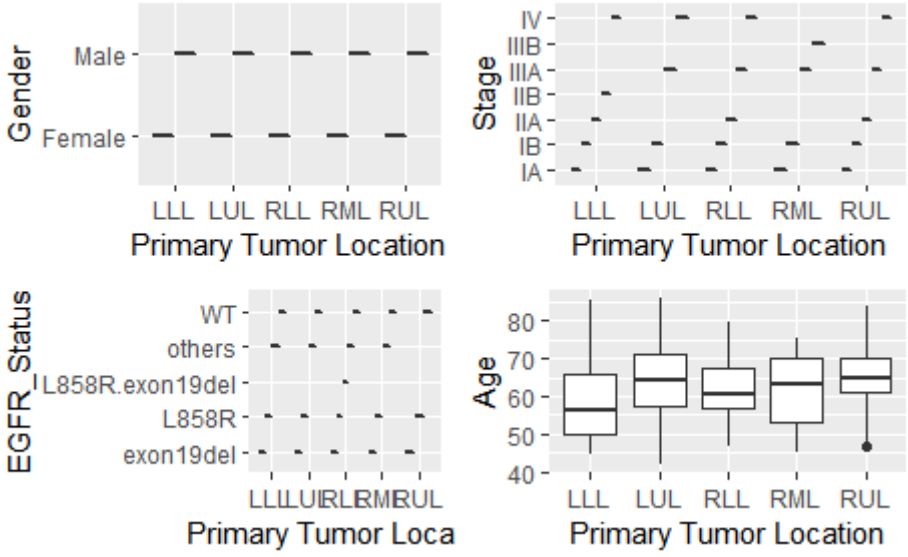
Gender <chr>	Age <dbl>	Stage <chr>	EGFR_Status <chr>	Primary Tumor Location <chr>
Female	56.47912	IB	exon19del	LUL
Male	59.02259	IA	exon19del	RLL
Male	61.85900	IB	exon19del	LLL
Female	44.91444	IA	WT	RML
Female	54.19576	IA	exon19del	RLL
Male	59.82752	IA	exon19del	RUL

1-10 of 79 rows Previous **1** 2 3 4 5 6 … 8 Next

Let's make plot for four points of view 1) gender and tumor location 2) stage and tumor location 3) EGFR_Status and tumor location 4) age and tumor location.

Hide

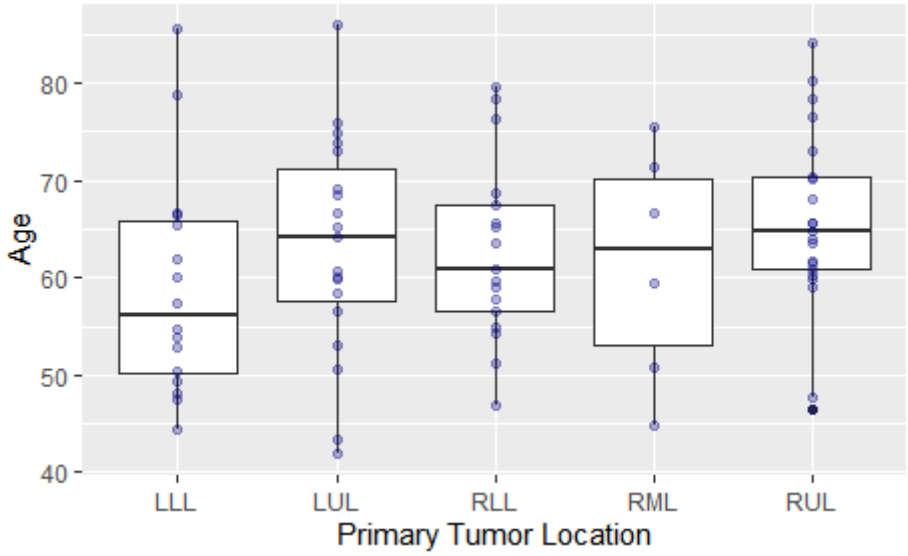
```
p1 <- d2 %>% ggplot(aes(`Primary Tumor Location`, Gender)) +
  geom_boxplot()
p2 <- d2 %>% ggplot(aes(`Primary Tumor Location`, Stage)) +
  geom_boxplot()
p3 <- d2 %>% ggplot(aes(`Primary Tumor Location`, EGFR_Status)) +
  geom_boxplot()
p4 <- d2 %>% ggplot(aes(`Primary Tumor Location`, Age)) +
  geom_boxplot()

plot_grid(p1, p2, p3, p4)
```



Among these, 4)age and tumor location had meaningful correlation.

Hide

```
d2 %>% ggplot(aes(`Primary Tumor Location`, Age)) +
  geom_boxplot() +
  geom_point(alpha=0.3, size = 1.4, color='navy')
```



3. Discussion

According to the figure, in “left” lung lobes, median of LUL was higher in age than one of LLL. Likewise in “right” lung lobes, median of RUL was the highest, followed by RML and then RUL. In terms of median, left and right lungs both look right-top direction. And it leads to positive correlation between age and location of ADC(upper-middle-lower).

4. Reference

Yi-JuChen, Theodoros I.Roumeliotis, Ya-Hsuan Chang, Ching-Tai Chen, Chia-Li Han, Miao-Hsia Lin, Huei-Wen Chen , Gee-Chen Chang, Yih-Leong Chang, Chen-Tu Wu, Mong-Wei Lin, Min-Shu Hsieh, Yu-Tai Wang, Yet-Ran Chen, Inge Jonassen, Fatemeh Zamanzad Ghavidel, Ze-Shiang Lin, Kuen-Tyng Lin1 …Yu-Ju Chen, Proteogenomics of Non-smoking Lung Cancer in East Asia Delineates Molecular Signatures of Pathogenesis and Progression, Cell, Volume 182, Issue 1, 9 July 2020, Pages 226-244.e17