

Differences in somatic mutations between smokers and Reformed smokers exposed to second-hand smoke among LSCC patients

Lee Sang Yoon 2016270905

1. Introduction

1.1. Background

Smoking is known to cause various diseases, including lung squamous cell carcinoma (LSCC). In the past, smoking was considered only an individual's preference, but now it is defined as a major risk factor that harms not only one's health but also those around him through second-hand smoke.

As a result of this change in perception, the proportion of daily smokers aged 15 or older decreased from 30% in the UK, 25.5% in the United States, and 28.5% in Japan in 1990 to 14.5%, 9.4% and 16.7% in 2020, respectively, and on the OECD average, decreased significantly to 16.1% in 2020.

if you want to download OECD Tobacco consumption dataset, click the link below

https://stats.oecd.org/Index.aspx?DataSetCode=HEALTH_STAT

```
setwd("C:/Users/user/Documents/R_project/bsms222_905_lee")
library(readr)

read_csv("HEALTH_LVNG_19112022063405406.csv")
```

```
## Rows: 1973 Columns: 11
## —— Column specification ——
—
## Delimiter: ","
## chr (8): VAR, Variable, UNIT, Measure, COU, Country, Flag Codes, Flags
## dbl (3): YEA, Year, Value
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

VAR										
<chr>										
TOBATBCT										
TOBATBCT										
TOBATBCT										
TOBATBCT										
TOBATBCT										
TOBATBCT										
TOBATBCT										
TOBATBCT										
TOBATBCT										

1.2. Aims

Reading [A proteogenomic portrait of lung squamous cell carcinoma](#) paper, I was able to identify a group of reformed-smokers with lung squamous cell carcinoma (LSCC) through [supplementary table S1](#). The group of reformed-smokers quit smoking for 15 years, but some of them were exposed to second-hand smoke.

In this portfolio, I will compare the occurrence of mutations in smokers who are not exposed to secondhand smoke and reformed smokers who are exposed to secondhand smoke through visualization.

2. Exploring data

2.1. Unboxing dataset

```
# Loading packages
```

```
library(tidyverse)
```

```
## — Attaching packages — tidyverse 1.3.2 —
## ✓ ggplot2 3.3.6   ✓ dplyr  1.0.10
## ✓ tibble  3.1.8   ✓ stringr 1.4.1
## ✓ tidyr   1.2.1   ✓ forcats 0.5.2
## ✓ purrr   0.3.4
## — Conflicts — tidyverse_conflicts() —
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

```
library(readxl)
library(ggplot2)
library(readr)
```

```
# Loading data
```

```
mmc1<- read_xlsx("mmc1.xlsx", sheet= 2)
```

2.2. Data wrangling

```
# Data wrangling
```

```
class(mmc1)
```

```
## [1] "tbl_df"      "tbl"        "data.frame"
```

```
Smoker<- mmc1%>%
  select(Sample.ID, Participant, Type, Smoking.History_modified, Secondhand.Smoke, Total.Mutation.Count.wes, Age, Smoking.score.wes)%>%
  filter(Smoking.History_modified=="Smoker" & Secondhand.Smoke != "Yes" & Type=="Tumor")
```

```
Secondary_smoker<- mmc1%>%
  select(Sample.ID, Participant, Type, Smoking.History_modified, Secondhand.Smoke, Total.Mutation.Count.wes, Age, Smoking.score.wes)%>%
  filter(Smoking.History_modified=="Reformed Smoker" & Secondhand.Smoke=="Yes" & Type=="Tumor")
```

mmc1's Total.Mutation.Count.wes column and Smoking.score.wes column contain charater.
Let's convert this into numeric

```
class(mmc1$ Total.Mutation.Count.wes)
```

```
## [1] "character"
```

```
class(mmc1$ Smoking.score.wes)
```

```
## [1] "character"
```

```
Smoker$ Total.Mutation.Count.wes1<- as.numeric(Smoker$ Total.Mutation.Count.wes)  
Smoker$ Smoking.score.wes1<- as.numeric(Smoker$ Smoking.score.wes)
```

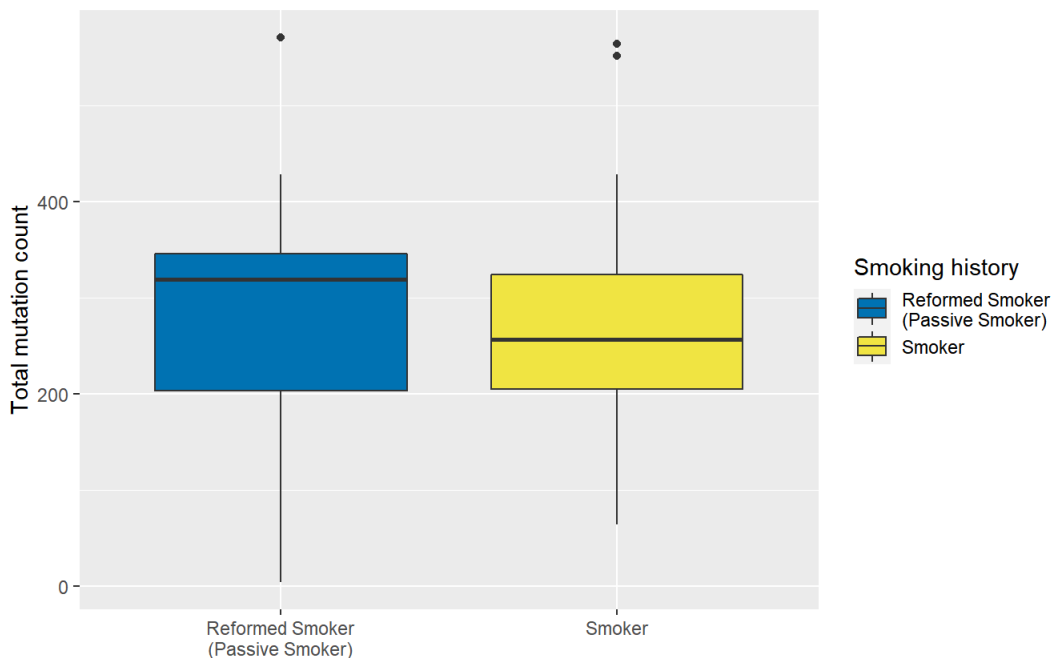
```
Secondary_smoker$ Total.Mutation.Count.wes1<- as.numeric(Secondary_smoker$ Total.Mutation.Count.wes)  
Secondary_smoker$ Smoking.score.wes1<- as.numeric(Secondary_smoker$ Smoking.score.wes)
```

3. Data Visualization

3.1. Boxplot

```
for_visualization<- rbind(Smoker, Secondary_smoker)  
  
for_color_blind_audience<- c("#0072b2", "#f0e442")  
  
for_visualization%>%  
  mutate(Smoking.History_modified= case_when(Smoking.History_modified== "Smoker"~ "Smoker", Smoking.History_modified== "Reformed Smoker"~ "  
Reformed Smoker  
(Passive Smoker)"))%>%  
  ggplot(aes(Smoking.History_modified, Total.Mutation.Count.wes1, fill= Smoking.History_modified))+  
  geom_boxplot()+  
  xlab("")+  
  ylab("Total mutation count")+  
  labs(fill="Smoking history")+  
  ggtitle("Differences in total number of mutations between  
smokers and reformed smokers exposed to secondhand smoke")+  
  scale_fill_manual(values=for_color_blind_audience)
```

Differences in total number of mutations between smokers and reformed smokers exposed to secondhand smoke

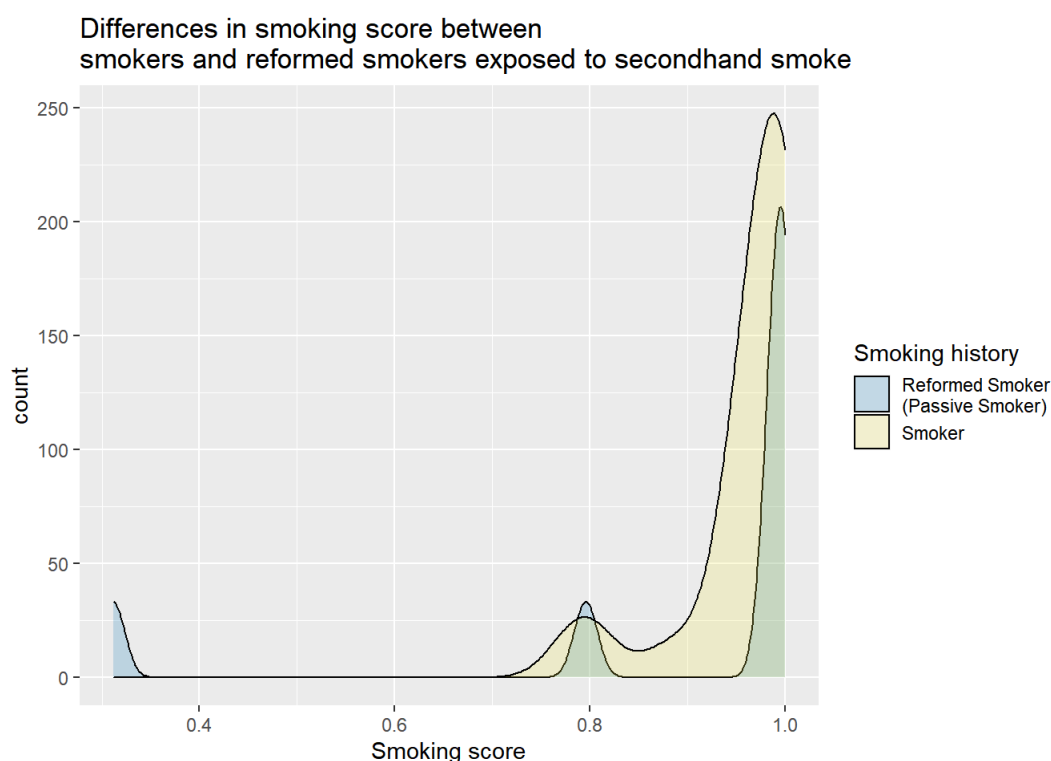


Although smoking was stopped, it was confirmed that more mutation occurred in the group exposed to second-hand smoke than in the group that continued smoking.

3.2. Density plot

For more information, I will draw a density plot and compare the smoking score between the two groups.

```
for_visualization%>%  
  mutate(Smoking.History_modified= case_when(Smoking.History_modified== "Smoker"~ "Smoker", Smoking.History_modified== "Reformed Smoker"~ "  
Reformed Smoker  
(Passive Smoker)"))%>%  
  ggplot(aes(Smoking.score.wes1, y= ..count.., fill= Smoking.History_modified))+  
  geom_density(alpha= 0.2, adjust= 2)+  
  xlab("Smoking score")+  
  labs(fill="Smoking history")+  
  ggtitle("Differences in smoking score between  
smokers and reformed smokers exposed to secondhand smoke")+  
  scale_fill_manual(values=for_color_blind_audience)
```



we can confirm that there is no significant difference in smoking scores of the two groups.

4. Discussions

We can confirm that there is little difference in total mutation count and smoking score between the smoker group and the reformed-smoker group exposed to secondhand smoke, and even the median value of total mutation count is higher in the reformed-smoker group exposed to secondhand smoke. In other words, we could see that second-hand smoke had an effect on mutation as much as smoking.

Today, the smoking rate has decreased significantly due to increased interest in health. However, contact with smokers at work, at home, and in society is overshadowing this, exposing reformed-smokers and non-smokers to the risk of disease.

The end of the year is coming soon, and appointments with acquaintances are increasing. I would like to suggest that we should start quitting smoking for the health of you and the people around you.

5. Reference

Shankha Satpathy, Karsten Krug, Pierre M.Jean Beltran, Sara R.Savage, ...& Ziad Hanhan. (2021). [A proteogenomic portrait of lung squamous cell carcinoma](#)