### Question 1

I downloaded the mutation files of lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) from cBioportal.

I used “data\_mutations\_extended.txt” as my data file. I could use either “data\_mutations\_extended.txt” or “data\_mutations\_mskcc.txt”. However, because the difference between them does not matter in answering the questions of this exercise and it is related to their isoform annotation, I selected “data\_mutations\_extended.txt” arbitrarily.

### Question 2

Table 1 illustrates the most frequently mutated genes in lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC)

|  |  |  |  |
| --- | --- | --- | --- |
| **LUAD** | **Freq** | **LUSC** | **Freq** |
| *TTN* | 46% | *TP53* | 79% |
| *TP53* | 45% | *TTN* | 70% |
| *MUC16* | 40% | *CSMD3\** | 46% |
| *CSMD3\** | 35% | *MUC16* | 43% |
| *RYR2* | 35% | *RYR2* | 43% |
| *KRAS* | 33% | *LRP1B* | 39% |
| *LRP1B* | 29% | *USH2A* | 38% |
| *USH2A* | 29% | *ZFHX4* | 37% |
| *ZFHX4* | 27% | *ADAM6* | 30% |
| *FLG* | 27% | *SYNE1* | 29% |

Table1. Most frequently mutated genes in lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC).

\*The results of my analysis differ slightly from the built-in analysis on the cbioportal website. For instance, CSMD3 is not listed among the 10 most frequently mutated genes in LUAD according to cbioportal. I am currently trying to determine the source of this discrepancy in the results.

There are various ways to display the most frequently mutated genes in cancer, such as bar plots and oncoplots. These methods can be used to visualize the data for different aims.

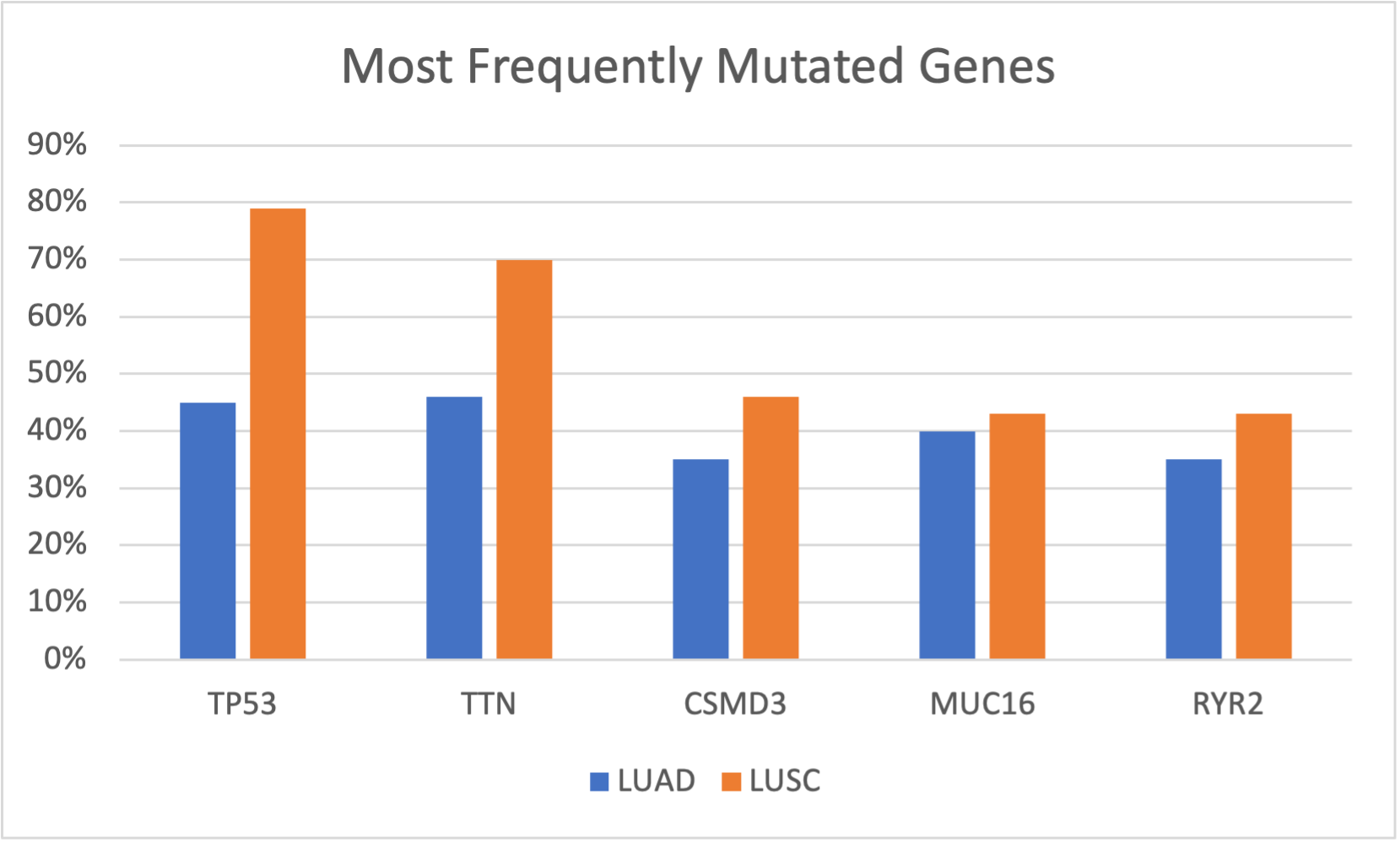
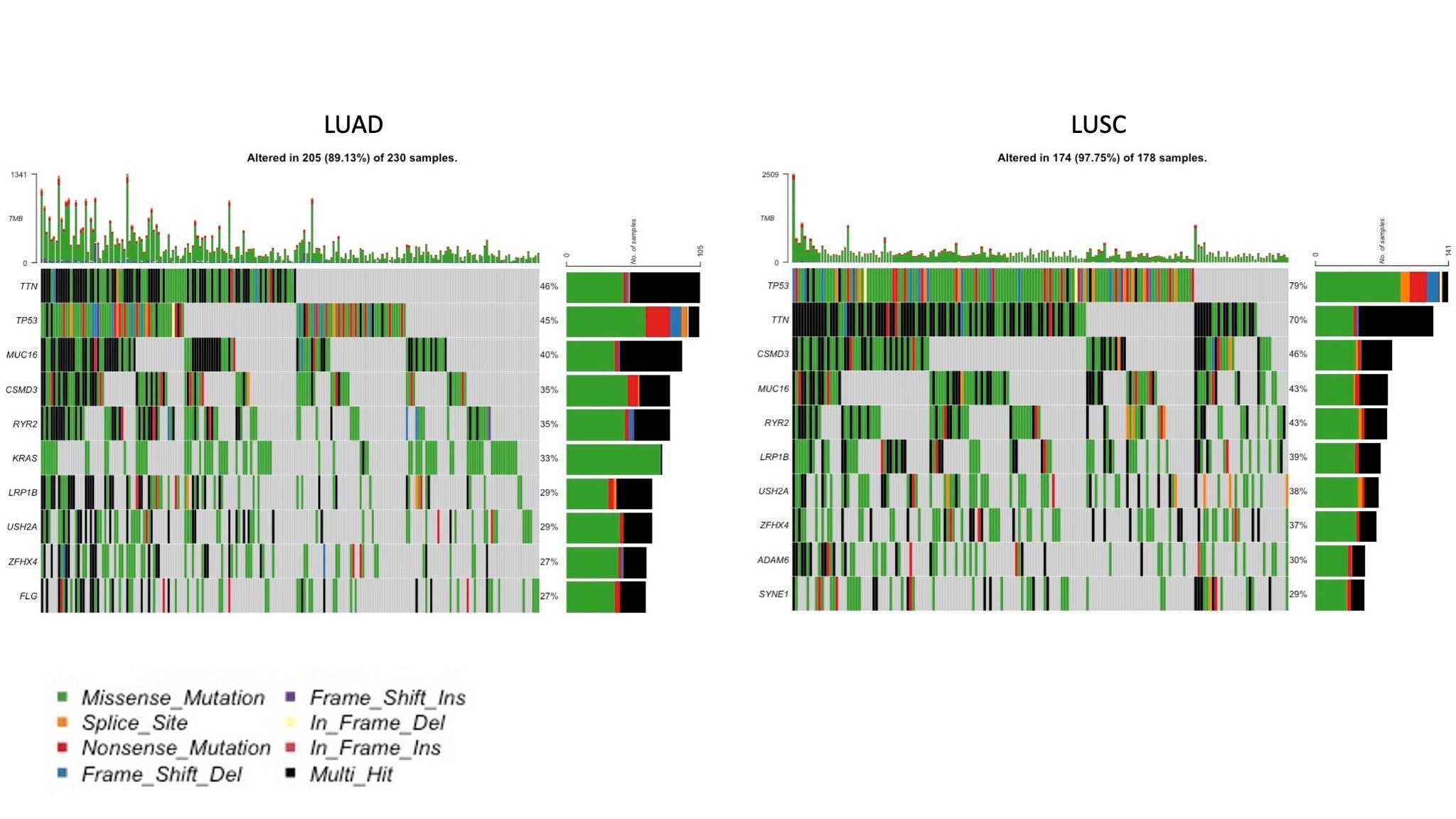


Fig 1. Most frequently mutated genes in lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) barplot.

Fig 2. Most frequently mutated genes in lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) oncoplot.

#### Do I think that the most frequently mutated genes are the most important genes for tumor development?

When I use the word 'important' in this context, I am referring to findings about causal relationships (it can be referred to other concepts such as survival rate). At this stage of my analysis, I cannot definitively say that these are the most important genes. It is important to note that correlation does not necessarily imply causation. While I have observed a relationship between certain genes and certain types of cancer, I am not yet confident in my ability to identify the causal relationship between the two.

Also, although those genes are the most frequently mutated genes, it is important to note that a gene can have multiple different mutations. Therefore, just because a gene is frequently mutated does not necessarily mean that it has the most significant impact on the development and progression of cancer. It is important to carefully analyze the specific mutations present in a gene and how they affect the function of the gene in order to understand their potential impact on the cancer (See Fig 4).

To better understand the cuasual relationship between these mutations and cancer, it would be helpful to find correlations among these genes and between these genes and other factors. Additionally, examining the effects of these mutations on protein structure and function level can help us to further understand the causal relationship between these mutations and cancer.

For example, the exclusive/co-occurance analysis of the mutated genes can help us to get closer to understanding of the casual relationship.

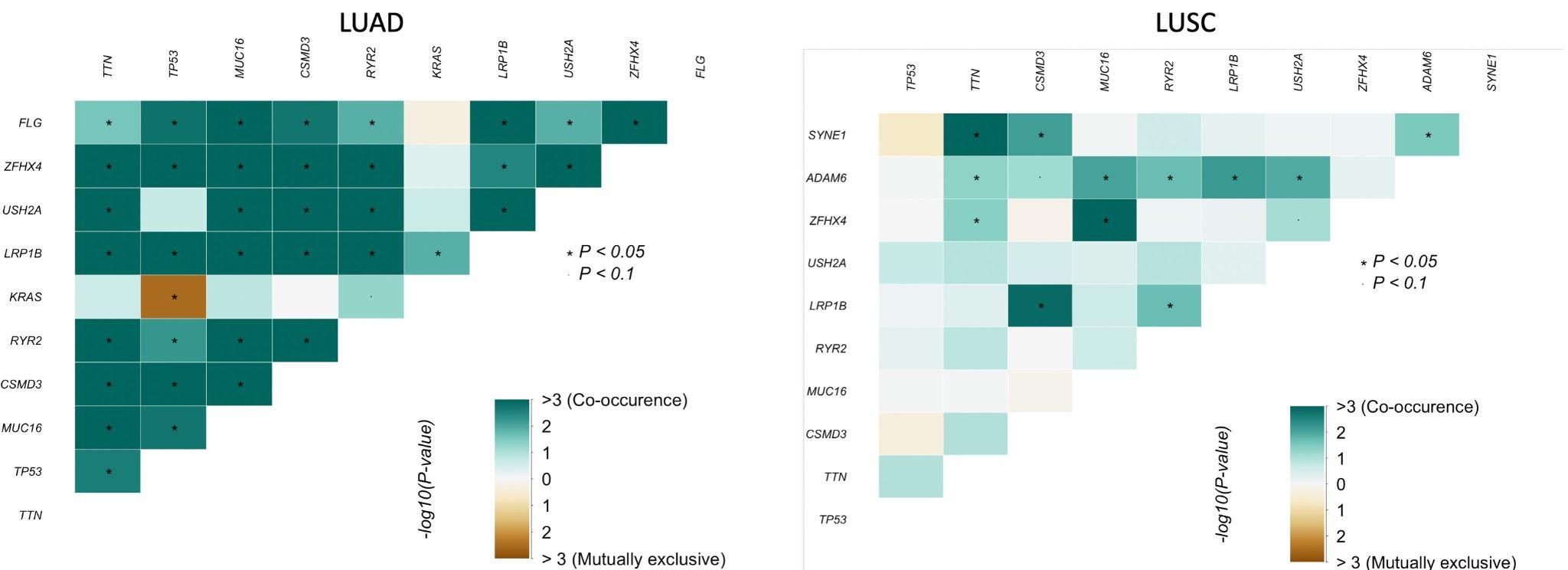


Fig 3. The Exclusive/co-occurance analysis of the mutated genes in lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC).

### 

### 

Fig 4. The Boxplot of Variant Allele Frequencies of the top most frequently mutated genes in lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC)

### Question 3

KRAS and TP53 are mutually exclusive (Fig 3. LUAD and Fig 5). KRAS and TP53 are both important regulators of cell proliferation and survival. KRAS mutations can cause uncontrolled cell growth and division, while TP53 mutations can disrupt the normal cell cycle and lead to cell death. When both KRAS and TP53 are mutated in the same cell, it can create a conflict that the cell is unable to resolve, leading to cell death.

Another reason might be that KRAS and TP53 mutations can have opposite effects on the cell's response to DNA damage. KRAS mutations can impair the cell's ability to repair DNA damage, while TP53 mutations can enhance the cell's ability to repair DNA damage. This can create a situation where cells with both KRAS and TP53 mutations are more resistant to DNA damage and more likely to survive and proliferate.

Also, KRAS and TP53 mutations can have different effects on the tumor microenvironment. KRAS mutations can promote inflammation and angiogenesis, while TP53 mutations can inhibit these processes. This can create a dynamic where cells with KRAS mutations are more likely to grow and spread, while cells with TP53 mutations are more likely to remain dormant or die.

I would illustrate the distribution of KRAS and TP53 alteration across the genome with Lollipop plot. It seems that the allele frequency is substantially more in TP53 in comparison to KRAS (Fig.6, Fig.7, Fig.8). The VAFs of TP53 and KRAS may differ depending on the specific cancer type and population being studied.

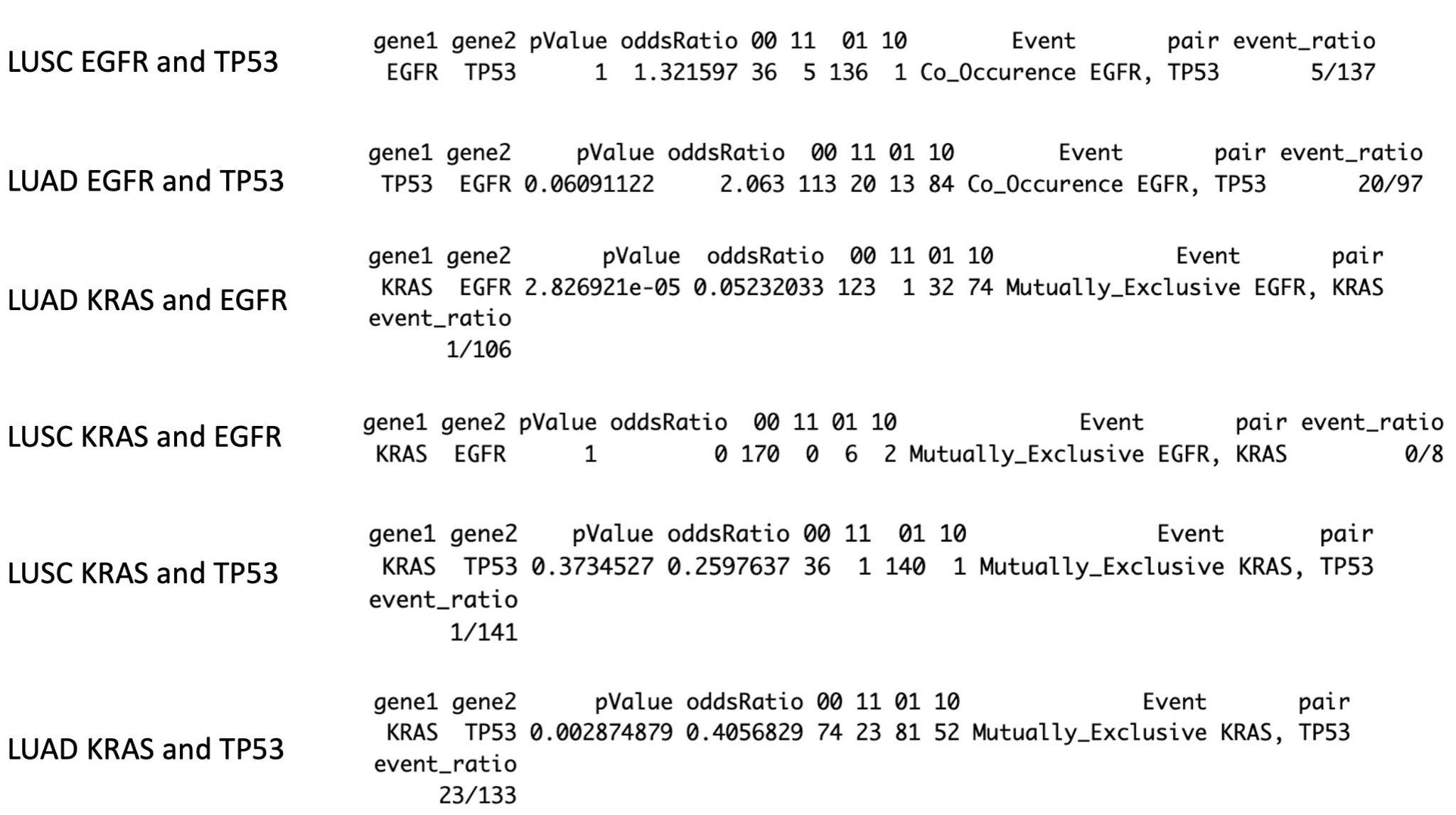


Fig 5. Exclusive/co-occurance relationships of TP53, KRAS and EGFR in LUSC and LUAD.

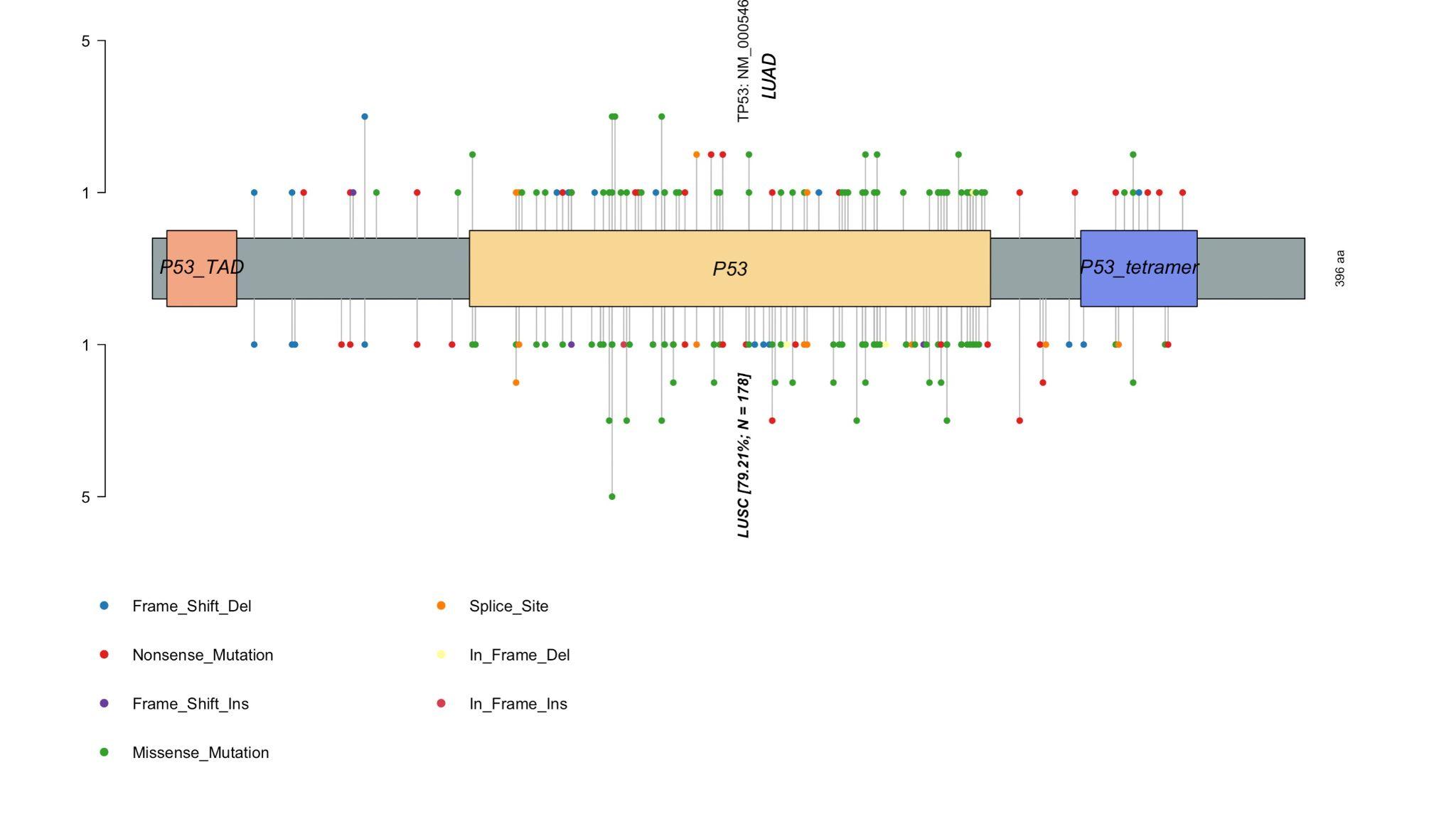


Fig 6. The plot illustrate the the location of the alteration on the TP53 in LUAD (top) and LUSC (bottom).

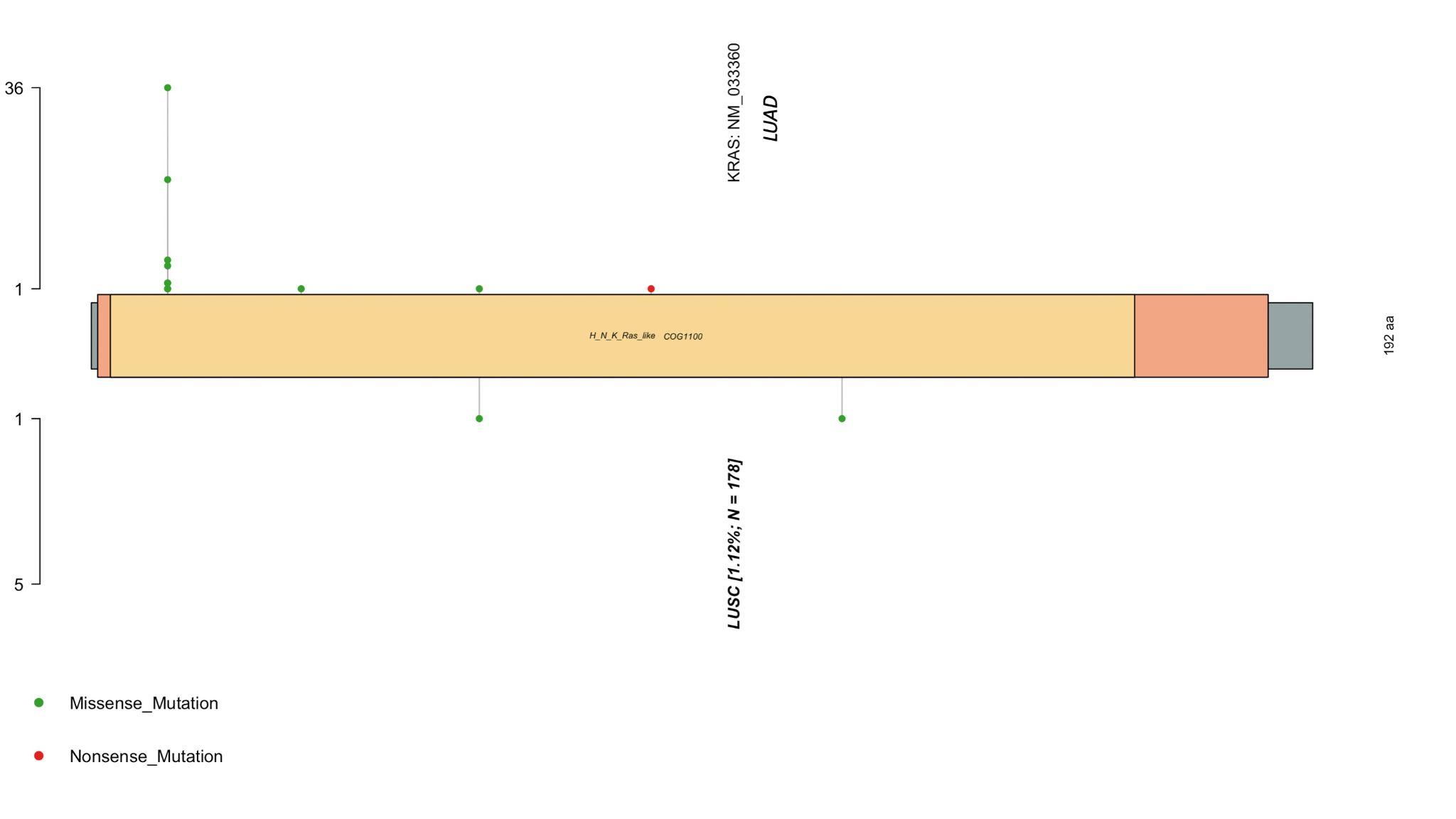


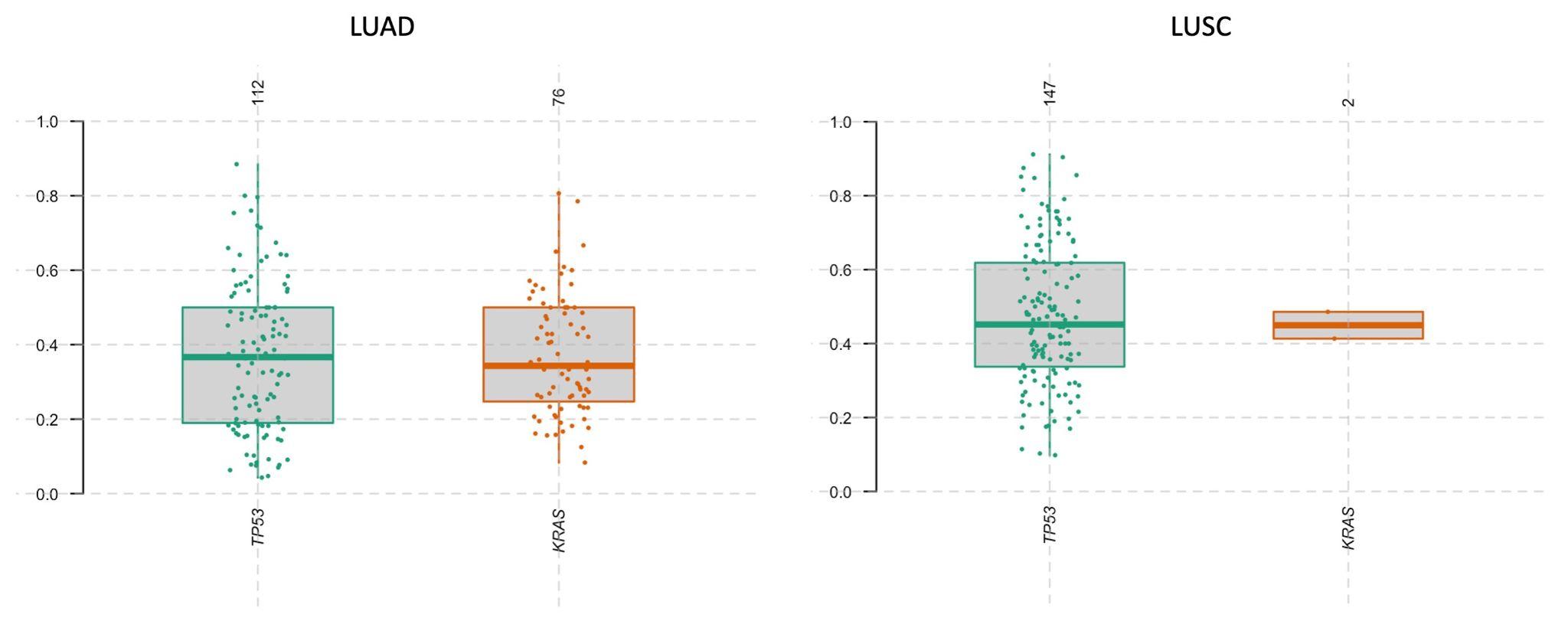
Fig 7. The plot illustrate the the location of the alteration on the KRAS in LUAD (top) and LUSC (bottom).   
  


Fig 8. The number and variant allele frequencies of KRAS is substantially lower in LUSC in comparison to LUAD.

### Question 4

KRAS and EGFR are mutually exclusive. (Fig 5. And Fig 9)

## Appendix

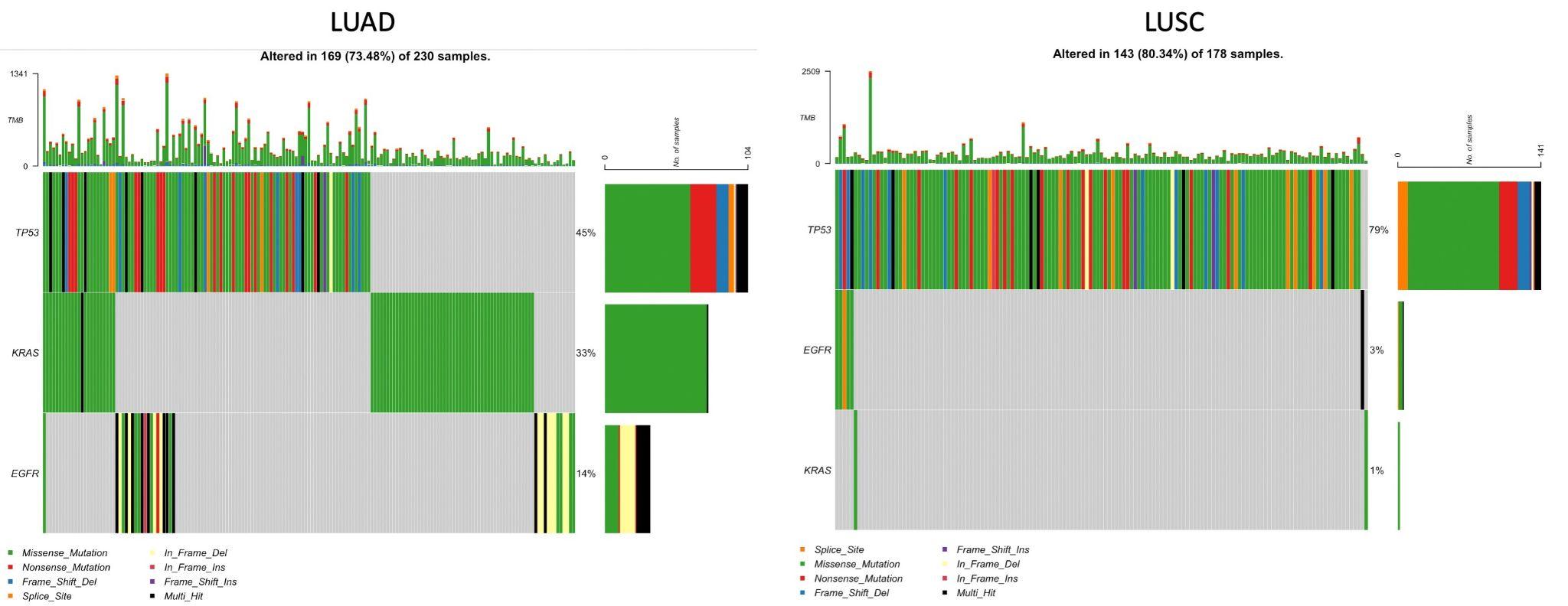


Fig 9. The oncoplot illustrate the distribution of 3 mutant genes (TP53, KRAS and EGFR) in the sample for LUAD and LUSC.

## Code:

#Calling the library

library(maftools)

# lung adenocarcinoma(LUAD)

#loading the mutation file

luad\_maf = "/Users/amin/Desktop/Cancer\_Maftool/luad\_tcga/data\_mutations\_extended.txt"

luad = read.maf(maf = luad\_maf)

#lung squamous cell carcinoma (LUSC)

#loading the mutation file

lusc\_maf = "/Users/amin/Desktop/Cancer\_Maftool/lusc\_tcga/data\_mutations.txt"

lusc = read.maf(maf = lusc\_maf)

#gene summary

print("\_\_\_\_\_\_\_\_\_\_\_\_LUAD\_\_\_\_Summary\_\_\_\_\_\_\_\_\_\_\_\_\_\_")

getGeneSummary(luad)

print("\_\_\_\_\_\_\_\_\_\_\_\_LUSC\_\_\_\_Summary\_\_\_\_\_\_\_\_\_\_\_\_\_\_")

getGeneSummary(luad)

#A general visualization

#LUAD

plotmafSummary(

maf = luad, rmOutlier = TRUE,

addStat = 'mean', dashboard = TRUE,

titvRaw = FALSE)

#LUSC

plotmafSummary(

maf = lusc, rmOutlier = TRUE,

addStat = 'mean', dashboard = TRUE,

titvRaw = FALSE)

#oncoplot for top ten most frequent mutated genes.

#LUAD

oncoplot(maf = luad, top = 10, keepGeneOrder = FALSE, sortByMutation=FALSE)

#LUSC

oncoplot(maf = lusc, top = 10, keepGeneOrder = FALSE, sortByMutation=FALSE)

#LUAD exclusive/co-occurance event analysis on top 10 mutated genes.

somaticInteractions(maf = luad, top = 10, pvalue = c(0.05, 0.1),

fontSize = 0.9, sigSymbolsSize = 2, sigSymbolsFontSize = 2,

nShiftSymbols = 2,showSum = FALSE

)

#LUSC exclusive/co-occurance event analysis on top 10 mutated genes.

somaticInteractions(maf = lusc, top = 10, pvalue = c(0.05, 0.1),

fontSize = 0.9, sigSymbolsSize = 2, sigSymbolsFontSize = 2 ,

nShiftSymbols = 2, showSum = FALSE)

#LUad exclusive/co-occurance of TP53 and KRAS

somaticInteractions(maf = luad, pvalue = c(0.05, 0.1),

genes = c("TP53", "KRAS"))

#LUSC exclusive/co-occurance of TP53 and KRAS

somaticInteractions(maf = lusc, pvalue = c(0.05, 0.1),

genes = c("TP53", "KRAS"))

#LUad exclusive/co-occurance of EGFR and KRAS

somaticInteractions(maf = luad, pvalue = c(0.05, 0.1),

genes = c("EGFR", "KRAS"))

#LUSC exclusive/co-occurance of EGFR and KRAS

somaticInteractions(maf = lusc, pvalue = c(0.05, 0.1),

genes = c("EGFR", "KRAS"))

#LUad exclusive/co-occurance of TP53 and EGFR

somaticInteractions(maf = luad, pvalue = c(0.05, 0.1),

genes = c("EGFR", "TP53"))

#LUSC exclusive/co-occurance of TP53 and EGFR

somaticInteractions(maf = lusc, pvalue = c(0.05, 0.1),

genes = c("EGFR", "TP53"))

#visualize the distribution of mutant genes in sample

oncoplot(maf = luad,genes = c("EGFR", "TP53","KRAS" ))

oncoplot(maf = lusc,genes = c("EGFR", "TP53","KRAS" ))

#variant allele frequency

#LUAD

plotVaf(maf = luad)

plotVaf(maf = luad, gene\_fs = 1, axis\_fs = 1,

height = 20, width = 20)

plotVaf(maf = luad, genes = c("KRAS","TP53"), height = 15, width = 15 )

#LUSC

plotVaf(maf = lusc, gene\_fs = 1, axis\_fs = 1)

plotVaf(maf = lusc, genes = c("KRAS","TP53"), height = 15, width = 15 )

#comparing LUAD and LUSC

pt.vs.rt <- mafCompare(m1 = luad, m2 = lusc, m1Name = 'LUAD', m2Name = 'LUSC',

minMut = 5)

print(pt.vs.rt)

#TP53

lollipopPlot2(m1 = luad, m2 = lusc,

gene = "TP53",

m1\_name = "LUAD", m2\_name = "LUSC",

domainLabelSize = 1.6,

legendTxtSize = 1.3, verbose = FALSE,

colors = NULL, alpha = 1,

axisTextSize = c(1.3,1.3), pointSize = 1.2

)

#KRAS

lollipopPlot2(m1 = luad, m2 = lusc,

gene = "KRAS",

m1\_name = "LUAD", m2\_name = "LUSC",

domainLabelSize = 0.65,

legendTxtSize = 1.3, verbose = FALSE,

colors = NULL, alpha = 1,

axisTextSize = c(1.3,1.3), pointSize = 1.2

)

#Appendix

luad.titv = titv(maf = luad, plot = FALSE, useSyn = TRUE)

#plot titv summary

plotTiTv(res = luad.titv)

luad.sig = oncodrive(

maf = luad,AACol = 'Protein\_Change',

minMut = 5, pvalMethod = 'zscore')

plotOncodrive(res = luad.sig,

fdrCutOff = 0.1,

useFraction = TRUE, labelSize = 0.5)