김연재

2018250103

BSMS222 Biostatistics

Submission date : (re) 2021.12.01

김예주

<Finding the patterns of phosphorylation of NSCLC pathway related genes by mutational profiles>

[Summary]

In this paper, the patterns of phosphorylation of NSCLC pathway related genes are analyzed. The writer select the genes by mutational profiles, which are founded by comparing Table S2D “Differentially expressed proteins in NSCLC pathway” and Table S5D. “Proteins enriched in the five carcinogen groups”. Common genes in these two tables are selected and each genes’ enriched pathways were explored. Therefore, Gene “BAD”, “CDK4”, “CDK6”, ”EGFR”, “ERBB2”, “MAP2K2”, “PLCG1”, “PRKCA” ,“RB1” , “SOS1” were chosen, and grouped mutational profiles such as “Taiwan Mixed” , “Alkylating ,”Radiation”, “Nitro-PAHs", "PAHs", "Nitrosamine" are expressed. The figure shows that the EGFR, ERBB2, and PRKCA genes are remarkably downregulated, and the MAP2K2 gene is upregulated. The result seems reasonable because MAP2K2 is involved in the metabolic and signaling pathways of cancer cells. In addition, as the expression of EGFR and ERBB2, which are involved in adherens junctions, is reduced, it can be considered in association with EMT.

[Peer Opinion]

As this paper contains only one Figure, I looked at the data processing process step by step. In this paper 5 Table was opened and processed. The writer select the common gene in Table 5D and Table 2D and then connected them to the pathway by intersecting Table 5D’s “enriched pathway” and Table 5C’s “carcinogen enriched pathway”. Following this data, “d_5d_rank” Data frame was made, however, I cannot find the column “variable” and “Rank Median G#” contents. It would have been easy to understand if a kind explanation had been attached according to the flow of progress, but it was difficult to understand because only a brief explanation was written.

Additionally, this paper’s topic is finding the patterns of phosphorylation of NSCLC pathway-related genes. If what I understand is correct, it seems that only the amount of gene expression is expressed now. I expect the writer to create a figure that fits your theme.

김용구

<Are age, gender, and tobacco exposure confounding variables in the correlation analysis of lung cancer?>

[Summary]

The Chen's paper showed that the proportion of non-smokers among female lung cancer patients in East Asia was high. This time, we would like to proceed with the analysis by adding age to the variable. Therefore, a boxplot of patients' age and their tobacco exposure are drawn. By this data, we can say that lung cancer in East Asia occurs more in females than males even if they are younger. Also, we can say that lung cancer in East Asia occurs more to non-smokers than smokers even if they are younger.

[Peer Opinion]

First, I don't know what the conclusion was obtained through this analysis. The writer seems to have tried to analyze the age cohort of Taiwanese patients, but was this box plot the right choice? This box plot shows nothing other than the distribution of samples. There is no tendency, and it does not come up with what more to explore through these three findings(1. age of the male are older than female 2. age of smokers are older than non-smokers 3. females are much less exposed to tobacco than males). I think it would have been more effective to visualize this content into a bar plot by dividing sections by age group, but I don't think this has much meaning either. Why don't we divide it by age group and see the amount of gene expression of interest? I've thought a lot about it, but I can't find it attractive to divide it by age group. This data has vast amounts of data on gene expression and protein expression in lung cancer patients, and it is not well understood that none of them were used and simply looked at the distribution of the sample. In order to present a more advanced direction in this data for the final exam, I think I need to change the subject or have more specific questions. For example, as a result of this analysis, there is a difference in exposure to cigarettes only in men. Using this, how about extracting the difference in gene expression between patients exposed to cigarettes and those not exposed to cigarettes and investigating whether there is a connection between them?

김정욱

< Sexual differences in EGFR mutation patterns by smoking status>

[Summary]

By comparing EGFR mutation patterns by smoking status in sexual difference, the writer wants to find mutations difference and suitable treatment methods for gender group.

[Peer Opinion]

In this paper, "EGFR status by Gender and smoking status" is expressed by alluvium plot. When I first encountered the plot, I wondered what it was expressing, but soon realized that it was a plot that connected data quite simply. Obviously, the writer said that he was exploring whether there was a treatment suitable for each gender through the pattern of EGFR mutation, but it seems that the inquiry into the content was not revealed through this figure. In addition, if divided by mutation patterns of EGFR, it is questionable that the differences, commonalities, and expressions between them should be compared. This plot is simply a Figure of the table in Table S1A. I wonder why the S1H table opened by the writer was not used. In addition, it is understandable that Figure 2 intends to show the number of samples through lines, but it seems that there was no tendency or interpretable part. According to the writer, EGFR mutations are associated with cell division, cell growth, and apoptosis of cells, and if so, it seems that investigations of pathways related to them or related genes should have been conducted. According to the writer, EGFR mutations are associated with cell division, cell growth, and apoptosis of cells, and if so, it seems that investigations of pathways related to them or related genes should have been conducted. And if the writer wanted to deal with the treatment method, I think it is necessary to investigate whether there is currently a drug targeting the gene of interest, EGFR, what kind, and whether there is a marker to confirm the effectiveness of the drug. It would be nice to find this and check the difference in the effectiveness of the drug according to the pattern of mutation if there are several types of drugs, and proceed with the analysis according to gender. Overall, this paper does not seem to have approached the topic that was initially set up, but I hope that through the process of upgrading this portfolio, more specific questions will be created and solved.