[Midterm] Peer review

이름 : 오정민

학번 : 2018250106

과목 : BSMS222 (00) Biostatistics

제출일 : 2021.11.22.

윤세현

The topic of this report is “Visualizing APOBEC mutational signature at the phospho-proteome levels using kinase enrichment” and according to Introduction, it was found that the APOBEC mutational signature is associated with certain types of lung cancer patients, which can be investigated to develop targeted therapeutics at the phospho-proteome level.

In Processing the data, the data was filtered or combined based on kinase enrichment and then drew a histogram using the facet_grid function.

Through the plot, it was possible to compare the enrichment of kinase of APOBEC-high groups with APOBEC-low group. It was found that the concentration of kinase is related to the APOBEC mutation which can cause lung cancer. The report ended with the possibility of error due to distortion of information on the scale of the graph.

In the Data visualization, there is a phenomenon that the letters on the x-axis of the plot overlap. If the text size of the plot's x-axis part became smaller so that the letters do not overlap, it would be easier to identify the chromosome names.

The value of the y-axis is from -1 to 3 in APOBEC-high and -1 to 2 in APOBEC-low. Because they are different, it seems that an intuitive comparison is a little more difficult. It is necessary to set the range of the y-axis equally even if the y-axis space in the lower part (negative part) of the APOBEC-high and the upper part (positive part) of the APOBEC-low are left.

In the Discussion, there is a sentence 'kinase is highly enriched in patients with APOBEC-high compared to patients with APOBEC-low'. Comparing APOBEC-high group and APOBEC-low group only by looking at the given plot, it is somewhat difficult to distinguish which group is higher. It can be seen that the Enrichment of kinase has more enrichment in both groups, but the difference between two groups is rather subtle, so you might think that there is no difference. However, by looking only at patients with APOBEC-high, kinase enrichment can be compared. People can know what explanation talking about from the context, but it feels like the narrative is a bit ambiguous. If explain the results of the plot with a little more detail, the explanation will be more accurate and desirable. It can be a solution to add one more plot of comparing the high or low enrichment of kinase within each group.

이승엽

In instruction, the report subject is 'Correlation between Ber Protein and APOBEC3 families' and the APOBEC3 background shows that the APOBec3 enzyme induces gene mutations that cause cancer in several organs, including the lung.

After that, in the data processing, this report changed the columns and rows of the dataframe, selected data specific to APOBEC, and changed this dataframe to numeric to draw a point plot. Additional lines(abline function) with a slope of 1 was added to the plots so that data patterns could be compared.

As a result of plot generation, it was found that there was a significant correlation between the Ber protein and ABOBEC3 families.

In the introduction, it would be better to have a better to describe explain why the report's topic was decided like that.

In plotting, the code in the form of filter(!is.na(xx) & yy!='NA') is repeated several times. If possible, it is better to make the codes simpler by adding some codes at previous data-processing step. It can be a little more tidy. In addition, the geom_point() and geom_smooth() codes are often repeated. The codes would be more tidy if use the facet_grid function.

The plot title was cut off because the plot was not displayed through functions such as facet_grid, but a simple combination of several plots. The title should be modified so that it comes out completely.

Since the same legends are repeated 6 times, the size of each graph becomes smaller and the concentration towards the plot is also dispersed. If you remove all the legends and leave only one, the plot will be appeared clear.

The plots added lines with a slope of 1 to each plot, but since the x-axis and y-axis ratios are different for each plot, the slopes of the lines all seem to be different. For each plot, it would be better to adjust the ratio of the x-axis and y-axis uniformly so that the slope of the additional lines look same.

It is necessary to indicate the source for the instruction content. Also the interpretation of plot, result, conclusion, or discussion and suggestion should be included at the end.

이윤지

It is known that cell adhesion protein tends to decrease as the tumor stage increases. This report tried to confirm the relationship between the actual TW cohort data and downregulation of adhesion pathway proteins.

In Data wrangling, after extracting information related to the KEGG pathway, only data on the adhesion pathway and DNA replication pathway were extracted using the filter and grepl function. The final data was created by adding a new column in the dataframe and merging several tables.

By looking at the Figure1, it was revealed that most adhesion genes -except AKT, GRB2, MAPK1, and PAK- were LUAD patients-downregulated. Figure2 showed that adhesion protein expression level for angiolymphatic invasion-"yes" was decreased when compared to "no".

In Discussion, the conclusion that patients with Angiolymphatic Invasion downregulate adhesion gene protein levels and further suggestion that dysregulation of adhesion genes is associated with many pathological conditions, including vascular lymph invasion and cancer metastasis was showed.

The topic is appropriate and the data is well made in the desired form using only the necessary parts. Through the plot using geom_tile function in figure1, it was possible to compare the protein expression patterns of the DNA replication gene and the Adhesion gene at a glance. In figure2, by using boxplot appropriately, the plot pattern according to the difference in angiolymphatic invasion could be confirmed and "no" and "yes" in down-regulation could be clearly compared. As this report mentioned in the Introduction, the desired contents of the topic were clearly expressed by using effective functions and plots. I think there is nothing to add or subtract in the content part.

In Visualizing, it would be better to adjust the position of the plot explanation so that the explanation come after plot. If an explanation is given before the plot, it can be confusing whether this explanation describes the plot or is a separate description. Also, it seems that some of the contents of Visualizing should be moved to Discussion.

In Discussion, based on this report, I think contents like 'In what direction should the research be conducted in later' and 'what are the limitations of this report' would be additionally described.