BSMS222 Biostatistics

<PEER REVIEW>

2017250105

최승환

<Portfolio>

1. 최정후
2. 한필우
3. 경우진

Portfolio by 최정후

■ Summary

The genesis of this portfolio is based on the positive correlation between MCM and TP53 and dysregulated MCM in lung cancer. At first, MCM gene family was separated by their different functions: licensing factors in DNA replication (MCM2-7) and DNA recombination repairs (MCM 8,9). As consequences, DNA replication helicase and licensing factor is MCMs' core function in lung cancer. Also, RB1 was found to have higher correlation with MCM 2-7, even more than TP53. As both MCM 2-7 and RB1 are involved in cell cycle, especially during G1 phase to S phase, they should be considered in cancer therapy and research.

■ Comment

As this portfolio is well organized and written, it is easy to read and understand without much background knowledge. Some statistical analyzes, however, need more caution. Mean is the most frequently used measure for understanding central tendency, however, median is more preferred measure when there are some extreme values in data distribution. It would be better if you examine median expression level of MCM 2-7 in Figure 2A.

At Figure 1, blue lines representing 0 point are quite helpful. A difference between two MCM group is well delivered. Some boxplots, however, were not easy to understand at first, especially at figure of protein expression levels of DNA recombination repair. Short explanations about some unique boxplot here would make this figure more comprehendible. In addition, as a side note, it would be easier to compare the two groups if there are some lines or something that represent the IQR or median for each MCM group. Furthermore, according to the portfolio, MCMs are expressed higher in younger age group, and higher in IB late-like stage than other stages. It would be nice to show how significant these differences are, and I would like to know your opinions on these results.

The discussion part, especially assumption parts that RB1 mutation would cause continuous cell cycle and increase MCM expression level was very interesting and clear. Thank you for sharing your thoughts and findings from your intensive analysis of MCM!

Portfolio by 한필우

■ Summary

Tumors can avoid immune responses by means of numerous mechanisms. One of them is to eliminate or loss MHC class 1 molecules in cell surface. In this portfolio, expression levels of MHC1-related genes in lung cancer cohorts were analyzed. The MHC1-related proteomes are less expressed in tumor tissues as expected while RNA expression levels were increased as effects of negative feedback. Also, the MHC1-related proteomes were reported to be generally less expressed in younger age. Because avoiding immune responses is more important mechanism in younger age than elderly group who have already gone through a reduced immunity, it might be more efficient to use NK cells in cancer immunotherapy.

■ Comment

As this portfolio is easily written and organized, it was easy to read and follow. Also, main concept that tumor would have decreased immunogenicity is interesting approach as a method of analyzing lung cancer patients. Regarding decreased immunogenicity, however, other well-known proteins need to be included in this portfolio. The other proteins would include PD-L1, known as repressing T-cell and directly inhibiting immune responses, or MDSCs which inhibit T-cell activation within tumor. It would make this portfolio more informative and powerful.

MHC gene, HLA in human, are highly polymorphic. In other words, there are many different alleles in the different individuals inside a population. Here in this portfolio, many different MHC alpha chain proteins are listed as expected. As all these protein isoforms have identical function to present a peptide, it would be okay not to differentiate them according to their specific protein name. The only difference is the peptide that can bind to each MHC isoforms, and it does not need to be considered here. At figure 1B, just simply classify them into HLA-A, HLA-B, and HLA-C like in Figure1A. It would be fine.

Some of hypothesis and discussion part in this portfolio is quite persuasive and well written. At discussion part about Figure1, however, I disagree with the parts that since MHC1-related protein expression levels decreased, RNA translation levels increased as negative feedback. I would like to say that MHC1-related genes are regulated at protein levels not RNA. Lastly, I would like to identify exact numerical difference between young and old group in Figure2. Thank you for sharing your interesting idea!

Portfolio by 경우진

■ Summary

In this portfolio, RNA expression levels of each gene were analyzed by whether angiolymphatic invasion is occurred. A p-value of each RNA expression was calculated by t-test (Figure1). Additionally, only with RNA which have p-value lower than 0.01, rescaled log2TN was used to show correlation between these RNA expression levels and angiolymphatic invasion (Figure2).

■ Comment

I can see you really interested in angiolymphatic invasion process. However, it was quite hard to follow up you trying to show data you found. It would be much better if you additionally explain some background information about your main interest.
Regarding Figure1, two sample *t*-test is a test of the hypothesis that two population means are equal. Null hypothesis is not that RNA correlates with angiolymphatic invasion. Also, t-test that you used here assumes that data are normally distributed. Before adjusting t-test, some preliminary tests should be performed to make sure that this data follows a normal distribution. If not, you should use other statistical tests: non-parametric tests are recommended.

Also, it would be better to title the plot and legend appropriately. As there are already many RNA with p-value lower than 0.05, it would be better to label RNA only with p-value lower than 0.005 or some specific value that you think as efficient value. Most of all, your own interpretation about this figure will make this visualization more meaningful.

Let's look at Figure2. Only with RNA which have p-value lower than 0.01, rescaled log2TN was used. Original log2T/N data, however, should have not been rescaled. log2T/N tells really important information here: whether RNA is expressed higher than NAT or not. This information must not be changed by inadequate normalization. Even without rescaling, differences between two groups can be more clearly delivered. Also, as I said, it would be much better if your own opinion or discussion are included after each figure. Thank you for sharing your portfolio!