Slight changes can improve much for algorithms looking at gene expressions

🔓 **openaccessgovernment.org**

Y-h. Taguchi, a Professor at Chuo University, looks at the slight changes made to algorithms when looking at the COVID-19 virus and gene expressions

Improving something through slight changes is always difficult, especially nowadays where all things have already been optimised and connected with one another. For example, if a train company would like to speed up the train, the trains might have to reduce the stations where they stop. In this case, the number of passengers who can ride these trains unavoidably decreases.

Nevertheless, there are a few topics that can be improved easily without sacrificing anything. For <u>example, computer algorithms</u>. When one would like to add the numbers from one to 10, instead of simply summing up them, one can compute the sum by adding one and 10 and multiplying five to the sum of these two numbers. Although the latter is obviously easier than the former, both result in the same number, 55. It can happen even in real applications.

As we have described in the series of articles [1,2,3], we have developed a computational method [4] applicable to drug repositioning targeting COVID-19. Recently we have found that slight changes in algorithmic improvement can drastically improve achievement [5].

COVID-19 and gene expressions

Although it is impossible to fully describe the details since it is highly mathematical, the improvement is really small. In our method, we consider the expression of human genes whose number is as many as a few tens of thousands when SARS- COV-2, a virus which causes COVID-19, infects human cells. Then we generate new variables that discriminate infected cells from not infected ones, by summing up gene expressions.

Genes whose contribution to the new variables is large are regarded as playing critical roles during infection. In this analysis, we assume that the level of contributions from individual genes obey Gaussian distribution that appear when independent random variables are summed up. Thus, genes whose contribution is too large to obey Gaussian, we can regard as important ones. The slight change but very effective improvement we found were the estimation of width of Gaussian distribution, known as the standard deviation. Improved estimation of the standard deviation could improve achievement, too.

First of all, the number of genes selected have increased much and interact much, much more likely with virus proteins than the previous studies [1,2]. Since the virus does not have enough proteins and must borrow human proteins for its purpose, the estimation of human genes that likely interact with virus protein is a critical process for drug repositioning.

Next, we sought the known drugs reported to target these selected genes. There are multiple databases that list genes whose expression is altered by the treatment of known drugs; such drugs are expected to be COVID-19 drugs since they are expected to reverse the gene expression altered by infection.

Interestingly, the drugs selected by this criterion were previously reported to be effective drugs toward COVID-19. Using one of such databases, imatinib was listed at the top in the selected drugs.

Imatinib as a response to COVID

Imatinib is known to be an anti-cancer drug which is supposed to be effective against multiple cancers. There are several studies that test this drug toward COVID-19. Although the effectiveness has not yet been approved, the fact that this drug is still under consideration suggests that our methodology is not wrong.

As described in the previous article [1], binding affinity of drug compounds toward the target protein, in this case that of SARS-CoV-2, is important, since the compounds often suppress functionality of target proteins by binding to target proteins.

The second top compound, apratoxin A, turned out to have binding affinity to one of SARS-CoV-2 proteins, SARS-CoV-2 Mpro, based upon the computational estimation. Although we consider interaction between compound and human proteins based upon gene expression, it is interesting to know that binding affinity of a compound with a SARS-CoV-2 protein is detected. This drug is also a candi- date for anti-cancer treatment. It also supports the effectiveness of our methodology.

The seventh one, trovafloxacin, was also validated to have binding affinity to SARS-CoV-2 M-pro, but experimentally in this case. It is another example that binding affinity to one of SARS- CoV-2 is identified. Trovafloxacin is a known antibiotic. There are more additional compounds found.

In total, nine out of top 10 compounds retrieved from this database were previously reported to be candidate compounds towards COVID-19. This is a highly improved outcome if com- pared with the previous study described in the previous article [1]. Compounds identified are not antivirus, but antibiotic or anti-cancers. This suggests that the method is very suitable for drug repositioning toward SARS-CoV-2. More compounds were identified using other databases [5], and performance is as good as the above.

Although the gene expression profile of human cell lines used in this study is the same as those used in the previous study [1], the outcome is drastically improved by slight changes in algorithm. In a typical biological experimental study, in order to get progress, we need additional experiments. Nevertheless, in computational studies, as can be seen in the above, sometimes we do not need additional datasets at all, and just an improvement of the algorithm can result in progress. This is one of the advantages and fun in computational study.

References

1. Y-h. Taguchi, How to compete with COVID-19 with a computer? Open Access Government, issue 33, Jan. (2022) pp. 210-211.
2. Y-h. Taguchi, Can mice be an effective model animal for Covid-19? Open Access Government, issue 34, April (2022) pp.112-113.
3. Y-h. Taguchi, Is human blood better than cell lines as a COVID-19 infection model? Open Access Government, issue 35, July (2022) pp.182-183.
4. Y-h. Taguchi, Unsupervised feature extraction applied to bioinformatics, Research Outreach Vol. 115 (2020) pp.154-157. doi:10.32907/ro-115-154157
5. Y-h. Taguchi and Turki Turki, Tensor decomposition- and principal component analysis-based unsupervised feature extraction to select more reasonable differentially expressed genes: Opti- mization of standard deviation versus state-of-art methods, bioRxiv 2022.02.18.481115; doi:10.1101/2022.02.18.481115

Please Note: This is a Commercial Profile