



How to compete with COVID-19 with a computer?

Y-h. Taguchi, Professor at Chuo University in Japan, reveals how to compete with COVID-19 by using a computer

At the beginning of 2020, a rapidly rising pandemic, COVID-19, threatened the world. Since then, till the end of October 2021, over two and a half billion people have been infected, and over five million people have died. How to compete with this terrible pandemic was something that was urgently required.

Although many people have tried a wide range of attempts, one of the newly developed strategies, the computer-aided method, is promising. Especially, *in silico* drug repositioning is very attractive since it is cost-effective and safe. Although creating a new drug from scratch is not easy since drugs must be both adequate to the target diseases and safe for humans. Finding a new compound that satisfies these requirements simultaneously is not easy at all. On the other hand, if

we can use known drugs to treat a new disease since the governments have already approved the safety of drugs, then we can concentrate on finding which known drugs are effective for new infections.

Computer-aided drug repositioning strategies

There are various ways to find which known drugs are effective for a new disease by computers. The most popular form, ligand-based drug design (LBDD) that tries to find a new drug compound whose structure is similar to the known drugs, is useless since there are no-known drugs for the new diseases. The alternative strategy, structure-based drug design (SBDD), is too time-consuming and expensive to seek the correct one among more than a million known compounds. Thus, the third one, the phenotype

screening method, was raised. In the phenotype screening, candidate drugs were tested directly toward a phenotype, for example, disappearance or symptom improvement. The selected ones are more approved to be effective than those chosen by other strategies, LBDD and SBDD.

With the rapid development of recent measurements concerning comprehensive technology, gene expression has become more easily measured. Thus, the most cutting edge method of computer-aided phenotype screening employs gene expression instead of the symptom. Although replacement of symptoms with gene expression might look more indirect, it is not the case; in contrast to the impression, gene expression is more directly connected to the cause of diseases than symptoms. For example, since

numerous conditions can cause headaches, the fact that the patient has a headache cannot help the doctors find an effective drug that can resolve the fundamental cause of headaches. On the other hand, since gene expression profiles are more specific to the diseases and symptoms, gene expression profiles are more practical to find the fundamental causes of a particular sign. Thus, finding drugs that exhibit similar gene expression profiles of known drugs might be a good strategy. Unfortunately, we cannot apply this to a new cure for new diseases since we do not have any known drugs effective for a new infection.

Because of the limitations mentioned above, we need an unsupervised strategy applicable to new diseases.

Identification of effective drugs towards COVID-19

Recently, we developed a new unsupervised method for drug repositioning based upon gene expression and applied it to COVID-19 (1). Although we have no space to describe the mathematical background of this methodology (2), briefly speaking, it can use an old-fashioned linear algebra-based method, tensor decomposition (TD). Tensor, an extension of a matrix with only columns and rows, can deal with more than two "columns and rows." Gene expression profiles of five human lung cancer cell lines that SARS-CoV-2 causes the COVID-19 or mock virus to infect were measured three times over. We formatted gene expression profiles as a tensor to which we applied TD. Finally, we found 163 human genes that infection of SARS-CoV-2 primarily affects.

Compared with a total number of human genes, 20,000, although very limited, the 163 genes have many striking features. Since we published this paper in September 2020, we did not have much information about SARS-CoV-2. Thus, we have investigated these 163 genes based upon their connection to SARS-CoV, which is the most related virus to SARS-CoV-2. These 163 genes highly overlap with human genes known to interact with all SARS-CoV proteins when infection occurs. Since the virus cannot have all proteins for disease, it must make use of human genes. Thus, the detection of human genes that interact with the virus at the initial infection stage is essential. In addition to this, the 163 genes also largely overlapped with human genes, the infection drastically altered gene expression. These two indicated the success of our strategy. We could also find that interaction between these 163 human genes and SARS-CoV-2, not SARS-CoV, is highly significant and even more significant than between them and SARS-CoV proteins.

Since we successfully identified a minimal number of human genes, which is critical during the infection, we next tried to find known drugs that can target these identified 163 human genes. Using various databases that stored gene expression alteration because of drug treatment, we could locate promising drug candidate compounds with this strategy. Remarkably, they contain numerous anti-virus drugs; this fact also suggests the success of our approach. For example, the list of drugs includes ivermectin that some people believe is effective, although its efficacy was approved only *in vitro* but not *in vivo*.

The list of drugs also highly overlapped with the drugs reported as promising in the independent study.

Concluding remarks

Unfortunately, because of lack of funding, although we have not yet tested the effectiveness of these identified drugs experimentally, we are still seeking collaborators or companies willing to develop the drugs for COVID-19 with us. We would be glad if some of the readers of this manuscript could connect us with people eager to help us on this issue.

References

- (1) Taguchi Y-h, Turki T (2020) A new advanced in silico drug discovery method for novel coronavirus (SARS-CoV-2) with tensor decomposition-based unsupervised feature extraction. PLoS ONE 15(9): e0238907.
- (2) Taguchi Y-h, (2020) Unsupervised Feature Extraction Applied to Bioinformatics -A PCA Based and TD Based Approach-, Springer International.



Y-h. Taguchi
 Professor
 Chuo University
 Tel: +81 3 3817 1791
 tag@granular.com
https://researchmap.jp/Yh_Taguchi/
https://twitter.com/Yh_Taguchi