

Can we do drug repositioning without disease gene expression?

openaccessgovernment.org/article/can-we-do-drug-repositioning-without-disease-gene-expression/167127

26 September 2023

Chuo University's Professor Y-h. Taguchi examines the application of cutting-edge single-cell-based measurements in drug repositioning

Single-cell analysis is a new trend in genomic science. Prior to its development, measurements were made on whole tissues. However, because tissues are composed of millions of cells, measuring them can miss important information. Imagine trying to understand the economic status of a big city with millions of people based only on the average income.

Even if two cities have the same average income, one might be a mixture of a small number of rich people and many poor people, while the other might be composed of people who have almost the same income. Without detailed information about the distribution of income, we would not be able to distinguish between the two cities from an economic point of view.

In the same way, single-cell measurements have opened the door to a true understanding of the genomic state of living material. By measuring the gene expression of individual cells, we can identify the different cell types present in a tissue and their relative abundance. This information can be used to understand how tissues function and how they respond to different stimuli.

Mice and Alzheimer's disease

In my previous studies on in silico drug repositioning,^(1, 2) I only analyzed tissue-based measurements. In this article, I discuss how to use single cell-based measurements, a cutting-edge technology, for drug repositioning.⁽³⁾

The target of this study is Alzheimer's disease (AD), which destroys human brain cognitive function and lacks effective treatment. The data set analyzed includes two mouse brain tissues, the cortex and hippocampus, which are thought to be related to AD. Additionally, the measurements were performed in two distinct genotypes, which means that there are slight differences between the two groups to avoid the effects of small changes in the genome. Two sexes were also considered to avoid the effects of sex.

Then, the gene expression of single cells in these two tissues was measured at four time points: 3, 6, 12, and 21 weeks after birth. This means that the purpose of this measurement is to monitor how gene expression in the brain changes with age. Since

aging is a primary factor in AD, it is expected that the dependence of brain gene expression on aging is deeply related to AD progression. However, no direct measurement of AD was performed.

Is it possible to find a drug candidate without direct information about the disease?

The method we used is tensor decomposition – TD. ^(4, 5) I will not discuss the method in detail here because of its mathematical nature. Those who are interested in the method itself can read my previous articles. It is not easy to integrate this complicated set of single-cell measurements successfully. Each measurement is associated with four plates, each of which contains 96 cells, for a total of 400 single cells. How can we integrate these data sets?

Ideally, the expression of the selected genes should be independent of sex, genotype, and brain location but should be dependent on aging, if possible, monotonically. This is not an easy task, as there is no practical way to select genes whose expression is independent of various conditions.

Using various statistical tests, we can evaluate whether the gene expression is different between two conditions by evaluating the probability under the null hypothesis that the two conditions have the same distribution. (If the probability of the difference is very small, we can conclude that the difference is statistically significant.) However, even if the probability is not small enough, we cannot say that the distributions under the two conditions are identical; we can only say “no conclusion.”

Tensor decomposition allows us to do this easily, as it can tell us which genes are independent of various conditions. Another advantage of TD is that it is an unsupervised learning method. Popular machine learning methods, such as ChatGPT or Stable Diffusion, require massive datasets from which the machine learning algorithm can learn. These methods also require a long computation time to learn. Unsupervised methods are much faster, as they do not have to learn anything.

Using TD, we were able to identify as many as 401 genes. Although this number may seem large, it is only a small percentage of the total number of mouse genes, which is 20,000. This means that TD was successful in reducing the number of genes to those that are most promising.

Using the data

The screening of known drugs using the list of 401 genes was simple. There are many databases that store gene expression changes caused by drug treatment. We can compare the 401 genes we selected with those whose expression is changed by individual drug treatments. We can then rank the drugs based on the number of matches between the 401 genes we selected and the genes whose expression is changed by a specific drug treatment.

Surprisingly, even though we did not deal with gene expression of AD, but only those associated with aging, we were able to identify several promising candidate compounds. Two of the top compounds based on the “LINCS L1000 Chem Pert up” database, alvocidib and AZD-8055, were once tested as promising candidates for AD drugs.

Even using another database, “DrugMatrix”, the top, fifth, and tenth-ranked compound, “cyclosporin-A”, was also tested as a promising candidate for AD. We also tested the “Drug Perturbations from GEO up” database, and the top ranked compound was imatinib, which was also previously tested for AD.

In contrast to previous studies, ^(1, 2) which required disease gene expression, we were able to successfully show that we do not even need disease gene expression to identify effective drugs for diseases. This is possible if we use my TD method, which can be easily used by anyone with access to the two freely available Bioconductor packages, TDbasedUFE and TDbasedUFEadv. ⁽⁶⁾

I encourage readers to try this method themselves, as they may be able to identify new candidate drugs that could help many patients. This can be done by simply using computers and public-domain datasets!

In conclusion, we have proposed a new method that can identify drugs without using disease gene expression. This method is very promising, and it has the potential to revolutionize the way we develop new drugs.

References

1. Y-h. Taguchi, Drug repositioning without the gene expression of disease cells treated with drugs 39(1) 28-29 July (2023) <https://doi.org/10.56367/OAG-039-10651>.
2. Y-h. Taguchi, The link between gene expression and machine learning, Open Access Government 38(1) 296-297 April (2023) <https://doi.org/10.56367/OAG-038-10651>.
3. Y-h. Taguchi and Turki Turki, Neurological Disorder Drug Discovery from Gene Expression with Tensor Decomposition, Current Pharmaceutical Design, 25(43) 2019, pp. 4589-4599 <https://doi.org/10.2174/1381612825666191210160906>.
4. Y-h. Taguchi, Unsupervised feature extraction applied to bioinformatics, Research Outreach, No.115, pp.154-157.
5. Y-h. Taguchi, Unsupervised Feature Extraction Applied to Bioinformatics: A PCA Based and TD Based Approach, Springer International, (2020).
6. Y-h. Taguchi and Turki Turki, Application note: TDbasedUFE and TDbasedUFEadv: bioconductor packages to perform tensor decomposition based unsupervised feature extraction, TECHNOLOGY AND CODE article, Sec. Medicine and Public Health, Front. Artif. Intell. 6 (2023) <https://doi.org/10.3389/frai.2023.1237542>.

Please Note: This is a Commercial Profile



This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

More About Stakeholder



In silico drug repositioning targeting SARS-CoV-2

Development of effective drugs toward COVID-19 is urgently required, and so research is being implemented with in silico drug repositioning.