Drug repositioning using multiple gene expression profiles

openaccessgovernment.org/article/drug-repositioning-using-multiple-gene-expression-profiles/171213

12 December 2023

Chuo University's Professor Y-h. Taguchi places focus on drug repositioning using multiple gene expression profiles

In my previous manuscripts [1-3], I introduced our studies of in silico drug repositioning using gene expression profiles. Nevertheless, in these studies, we could use the single gene expression profile to perform in silico drug repositioning. In this manuscript, the last one in this series, I introduce our recent study [4] in which we could use multiple gene expression profiles. Please see my previous article for more details about the in silico drug repositioning. [5]

In the studies [1-3] introduced previously, I used tensor decomposition (TD) based unsupervised feature extraction (FE) [6,7], which was a mathematical/computational method for data processing and was recently implemented in Bioconductor packages. [8]

Bioconductor is a major R package repository specific to bioinformatics, a joint computer-oriented research field between biology and informatics. R is one of the popular computer programming languages used for data sciences. The packages [8] allow anyone unfamiliar with TD to perform TD without difficulties.

TD-based unsupervised FE

In the study [4] that I introduce in this article, we make use of TD-based unsupervised FE [6,7] to integrate multiple gene expression profiles related to Alzheimer's disease (AD), but without sample matching. To integrate multiple gene expression profiles, we usually need the shared labelling between samples, e.g., healthy control vs patients. If multiple profiles share the class label, profiles are classified into classes with each label.

For example, there are three gene expression profiles, each classified as two classes, A and B; each class can be composed of profiles retrieved from each of the three profiles. Suppose that A1, A2 and A3 are the sets of profiles that belong to class A in the 1st, 2nd and 3rd data sets and B1, B2, and B3 are the sets of profiles that belong to class B in the 1st, 2nd and 3rd data sets. The classes A and B can be composed of the joint sets (A1, A2, A3) and (B1, B2, B3), respectively.

Even so, if this is not the case, we are unsure how to integrate multiple profiles without shared class labels. Suppose the 1st data set is composed of two sets of profiles, A1 and B1, that belong to the classes A and B, respectively; the 2nd data set is composed of two

sets of profiles, C2 and D2, that belong to the classes C and D, respectively, and the 3rd data sets are composed of E3 and F3 that belong to the classes E and F, respectively. Can we imagine how we can integrate three profiles into one? The answer is, obviously, no.

In the study [4], I introduce in this article, we prepared three gene expression profiles without shared labels. The first one comprises three classes: cell lines with mutations, those with gene mutation corrected, and normal cell lines. The second one is composed of cell lines with two kinds of gene mutation and is classified into four classes depending on whether individual genes are mutated or not. The third one is composed of AD and normal cell lines. There are obviously no ways to integrate these three profiles directly.

Despite the difficulties mentioned above in integrating multiple profiles that lack common class labels, since TD-based unsupervised FE allows us to integrate these three profiles and provides us common gene expression profiles, the projections of each profile onto which commonly exhibits coincidence with individual class labels, we can make use of this gene expression profile as disease profiles as in the case of the study introduced previously. [2] We could integrate it with gene expression profiles of cell lines treated with various drugs.

The reason why TD-based unsupervised FE could integrate gene expression profiles without shared labels is that individual gene expression profiles were projected to the subspace that shares the number of dimensions with other profiles before the application of TD-based unsupervised FE by the mathematical technique called singular value decomposition (SVD).

Since SVD allows individual profiles to have the same number of dimensions attributed to samples, three profiles were successfully integrated by TD. Although the name of SVD was not frequently mentioned, the mathematics behind SVD does not go beyond the fundamental linear algebra taught in the first grade of the university.

The gene expression profiles integrated with gene expression profiles that the integration of three gene expression profiles with TD-based unsupervised FE could provide are those generated by drug treatments; we have already shown that TD-based unsupervised FE could correctly identify the relationship between gene expression and drug treatment. [3,9]

Drug treatments are considered

As many as 80 drug treatments are considered, among which TD-based unsupervised FE identify several promising candidates, including ruxolitinib, baricitinib, and vorinostat. The role of the JAK/STAT signaling pathway in the pathogenesis of Alzheimer's disease was identified (10), and ruxolitinib targets the JAK/STAT signaling pathway. Alzheimer's Drug Discovery Foundation approved baricitinib [11] and vorinostat [12] as promising candidates for AD. Thus, TD-based unsupervised FE could use multiple gene expression profiles for drug repositioning.

In the series of manuscripts [1-3], including this one, I introduced the studies in which TD-based unsupervised FE was successfully used repeatedly. I hope that TD-based unsupervised FE can be used for the broader range of drug repositioning in the future.

References

1. Y-h. Taguchi, Can we do drug repositioning without disease gene expression? 40(1) 282-283 Oct (2023) https://doi.org/10.56367/OAG-040-10651
2. 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
3. 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
4. Taguchi, Yh., Turki, T. A tensor decomposition-based integrated analysis applicable to multiple gene expression profiles without sample matching. Sci Rep 12, 21242 (2022). https://doi.org/10.1038/s41598-022-25524-4
5. Taguchi, Yh., In Silico Drug Discovery for COVID-19 Using an Unsupervised Feature Extraction Method, Scientia, (2021), https://doi.org/10.33548/SCIENTIA727
6. Y-h. Taguchi, Unsupervised feature extraction applied to bioinformatics, Research Outreach, No. 115, pp. 154-157, (2020), https://doi.org/10.32907/ro-115-154157
7. Y-h. Taguchi, Unsupervised Feature Extraction Applied to Bioinformatics: A PCA Based and TD Based Approach, Springer International, (2020).
8. Y-h. Taguchi and 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
9. Taguchi Yh and Turki T (2020) Universal Nature of Drug Treatment Responses in Drug-Tissue-Wide Model-Animal Experiments Using Tensor Decomposition-Based Unsupervised Feature Extraction. Front. Genet. 11:695. https://doi.org/10.3389/fgene.2020.00695
10. Rusek, M.; Smith, J.; El-Khatib, K.; Aikins, K.; Czuczwar, S.J.; Pluta, R. The Role of the JAK/STAT Signaling Pathway in the Pathogenesis of Alzheimer's Disease: New Potential Treatment Target. Int. J. Mol. Sci. 2023, 24, 864. https://doi.org/10.3390/ijms24010864
11. https://www.alzdiscovery.org/uploads/cognitive_vitality_media/Baricitinib.pdf
12. https://www.alzdiscovery.org/uploads/cognitive_vitality_media/Vorinostat-Cognitive-Vitality-For-Researchers.pdf

Please Note: This is a Commercial Profile

More About Stakeholder

In scilico drug repositioning targeting SARS-CoV-2

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