Report

nature communications

Article

https://doi.org/10.1038/s41467-022-34277-7

Deep transfer learning of cancer drug responses by integrating bulk and single-cell RNA-seq data

Received: 6 August 2021

Accepted: 19 October 2022

Junyi Chen[1,6], Xiaoying Wang[2,6], Anjun Ma[1,3] ✉, Qi-En Wang[4], Bingqiang Liu[2], Lang Li[1], Dong Xu[5] & Qin Ma[1,3] ✉

In this article the authors introduce scDEAL, which stands for **s**ingle-**c**ell **D**rug **R**esponse **A**nal**y**sis. It is a deep transfer learning framework for prediction of cancer drug response. The main impact of the article is the integration of both bulk and single cell RNA-sequencing data analysis in order to achieve improved drug response prediction.

Currently there are several approaches for studying cancer drug response. A range of in vitro studies have already been conducted, and cell-line drug response sequencing data is accumulating [1,2]. Drug screening data from massive bulk gene expression databases can be analyzed to determine the optimal clinical application of cancer drugs. At the same time, cancer drug treatments suffer from low efficacies and high relapse rates caused by cancer heterogeneity. This leads to differentiated responses of individual cells to a drug, and complicates drug response prediction.

And that's where analysis of single-cell data comes to rescue. ScRNAseq data analysis and application of deep learning provide valuable insights into prediction of drug response as it facilitates the discovery of the heterogeneous gene expressions of cancer subpopulations in response to specific drugs. Deep learning models applied to scRNA-seq data have already achieved competitive performances in gene expression imputation, cell clustering, batch correction, etc. [3-5]. However, there are some obstacles regarding this approach, such as insufficient training power due to to the limited number of publicly available and high-quality data, which leads to lower prediction quality.

Intuitively, drug-related bulk RNA-seq data can be effective complementary resources to infer gene expression - drug response relations in support of the drug response predictions at the single-cell level [6,7]. And deep transfer learning (DTL) can transfer knowledge and relation patterns from bulk data to single-cell data, which can be a means to overcome the issue of limited training. The DTL model has been applied as an effective strategy in leveraging multiple bulk data sources for cancer drug response predictions [8]; however, so far, its capabilities in transferring valuable bulk-level knowledge to the single-cell level are declared to be under-investigated.

Therefore, the authors developed an approach for combining both bulk and scRNA data analysis to perform high quality model training and further analysis. Large-scale RNA sequencing data were accessed from GDSC (https://www.cancerrxgene.org/) and CCLE (https://sites.broadinstitute.org/ccle/) databases. The drug response prediction performances were evaluated on six public scRNA-seq datasets treated by five drugs, namely, Cisplatin, Gefitinib, I-BET-762, Docetaxel, and Erlotinib.

As for the framework, first, scDEAL models relations between the gene expression feature and drug response at the bulk level. Then, the shared low-dimensional feature space between single-cell and bulk data is identified in order to harmonize the relation between gene expression and drug response. The DTL model is trained to learn the optimized solution to these two relations. Finally, the single cell–drug response relations can be built through the meta-relation of gene expression at the single-cell level, gene expression at the bulk level, and drug response in the DTL model. The scDEAL framework involves such steps as extracting bulk gene features, predicting drug response in each bulk cell line using extracted features, extracting single-cell gene features, jointly training and updating all the models, and transferring and applying the trained model to scRNA-seq data to predict drug responses. The output of scDEAL is the predicted potential drug response of individual cells.

To evaluate the prediction of scDEAL, the authors applied several metrics. They are:

Precision - represents the ability of the model to correctly predict positive numbers among all positive predictions;

Recall - represents the model ability to correctly predict positivity from actual positive samples;

F1-score can be interpreted as a weighted average of precision and recall;

AUROC score - computes the area under the receiver operating characteristic (ROC) curve. The ROC curve's x-axis is the true positive rate and the y-axis is the false positive rate derived from prediction scores;

AP score - summarizes a precision-recall curve (PRC) as the weighted mean of precisions achieved at each threshold, with the increase in recall from the previous threshold used as the weight;

AMI score is an adjustment of the Mutual Information (MI) score to account for chance.

All in all, scDEAL is very powerful at predicting single-cell level drug sensitivity as it establishes bridges among drug sensitivity, gene features in single cells, and gene features in bulk samples.

scDEAL highlights the following aspects: (i) it can use a large amount of bulk-level drug response RNA-seq information from the GDSC and CCLE databases (ii) in order to account for data-structure differences between bulk and scRNA-seq data, scDEAL harmonizes single-cell and bulk embeddings to ensure that the drug response labels are transferable from bulk to single cells; (iii) in order to avoid losing heterogeneity in scRNA-seq data, scDEAL includes cell cluster labels for loss function regularization in each training epoch; (iv) scDEAL's integrated gradient interpretation infers the signature genes of drug response predictions, which improves the interpretability of the model.

However, there is one remaining challenge in single-cell drug response prediction declared in the article, which is the prediction across different species. Considering the genetic variation, drug response in one species cannot be directly transferred to predict the other, for example, regarding human and mouse data. Due to the limited number of open access drug-treated mouse scRNA-seq data, it was not possible for the scDEAL developers to properly evaluate and optimize the trans-species reliability.

All in all, scDEAL, a deep transfer learning framework for prediction of cancer drug response, was introduced in this article. scDEAL enables the deployment of the DTL model in single-cell drug response prediction by means of using abundant bulk RNA sequencing data for model training. Hopefully, scDEAL could help study cell reprogramming, drug development, repurposing, and cancer treatment for improving therapeutic efficacy.

References

1. Verjans, E. T., Doijen, J., Luyten, W., Landuyt, B. & Schoofs, L. Three-dimensional cell culture models for anticancer drug screening: Worth the effort? J. Cell. Physiol. 233, 2993–3003 (2018).

2. Schirle, M. & Jenkins, J. L. Identifying compound efficacy targets in phenotypic drug discovery. Drug Discovery Today 21, 82–89 (2016).

3. Wang, J. et al. scGNN is a novel graph neural network framework for single-cell RNA-Seq analyses. Nat. Commun. 12, 1882 (2021).

4. Gayoso, A. et al. Joint probabilistic modeling of single-cell multi-omic data with totalVI. Nat. Methods 18, 272–282 (2021).

5. Ma, Q. & Xu, D. Deep learning shapes single-cell data analysis. Nat Rev Mol Cell Biol 23, 303–304 (2022).

6. Wang, J. et al. Data denoising with transfer learning in single-cell transcriptomics. Nat. Methods 16, 875–878 (2019).

7. Wu, Z. et al. Single-Cell Techniques and Deep Learning in Predicting Drug Response. Trends Pharmacol. Sci. 41, 1050–1065 (2020).

8. Dhruba, S. R., Rahman, R., Matlock, K., Ghosh, S. & Pal, R. Application of transfer learning for cancer drug sensitivity prediction. BMC Bioinformatics 19, 497 (2018).