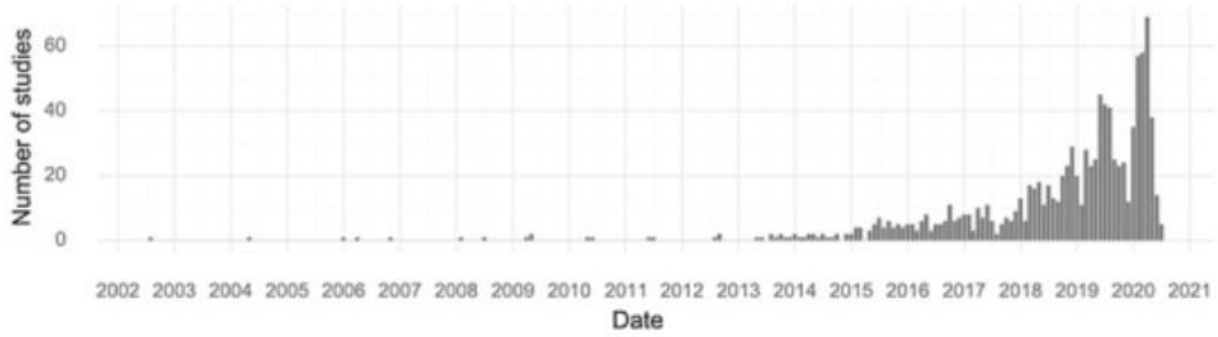
Deep transfer learning of cancer drug responses by integrating bulk and single-cell RNA-seq data.

The authors describe the problem of the heterogeneity of drug responses for cancer cell subpopulations. Therefore, the heterogeneity of a tumor is responsible for the differentiated response of individual cells to the drug, which can lead to a minimum amount of cancer residues remaining in the body and these cancer cells, eventually, may be the cause of recurrence.

This article was relevant because at that time (2022 year) the amount of data on single-cell RNA sequencing increased rapidly and continues to increase to this day. This fact allowed to use the accumulated data set for accurate prediction of cancer, whereas, at that time, existing drug-response prediction methods were developed for bulk data and couldn't be used for single-cell data.



Authors decided to use deep transfer learning (DTL) because this method can transfer knowledge and relation patterns from bulk data to single-cell data that can help solve a problem of the lack of scRNA data. Their approach is based on a Domain-adaptive Neural Network that is a specific type of adversarial-based DTL approach. Their tool (scDEAL) has following steps:

Firstly, necessary input data was collected (scRNA-seq from GEO, bulk RNA-seq database and bulk drug response from the Genomics of Drug Sensitivity in Cancer (GDSC) database and Cancer Cell Line Encyclopedia (CCLE)). Gene features were extracted from bulk and single cell data, this step included using Denoising Autoencoders (DAE) to remove noise from data, making it more suitable for analysis. Then, extracted bulk features were used to predict drug response in each bulk cell line. After this step, results from preprocessed single cell data and bulk drug response prediction trained jointly and updated the condition of previous steps. The most important thing that the training reduces decoding loss, making low-dimensional features informative, that we obtained after DAE, enough to represent the original gene expressions, and make better the predictor in a gene-drug relation (minimizing the difference between the prediction results and the database-provided drug responses). Finally, we can use scDEAL to get the predicted potential drug response of individual cells.

They had 7 metrics for evaluating this tool: F1-score, area under the receiver operating characteristic (AUROC), AP score, precision, recall, Adjusted Mutual Information (AMI), and Adjusted Rand Index (ARI). In this article authors often referred to F1 score, that can be interpreted as a weighted average of precision and recall. Additionally, authors calculated differential IG scores between sensitive and resistant cells to validate that the gene-drug relation learned and transferred to the single-cell level successfully.

For evaluating the prediction performance of scDEAL five drugs from six public scRNA-seq datasets were used.

Firstly, they checked the significance of the various stages and the data that was used for prediction like transfer learning, bulk data from CCLE or GDSC, DAE and cell-type regularization. Authors demonstrated their significance running the tool without some data or steps and compared resalts with a standard procedure.

One of the most important things, that they found, is the capability of scDEAL to identify critical genes responsible for drug response. Authors conducted the analysis for oral squamous cell carcinoma (OSCC) that was treated with Cisplatin and 85% of cells were correctly predicted as either sensitive or resistant to this drug. Then the list of genes was extracted using following parameters: adjusted p values <0.05, log-fold change <0.1 and cell percentage in either comparison group higher than 0.2. Thus, critical genes were identified. GO enrichment analysis of drug-resistant critical genes showed that they involve in processes that cause Cisplatin resistance, as DNA repair, cell division pathways

Whereas, authors found another important fact, that drug response prediction is highly correlated with pseudotime analysis. They noticed that the expression levels of some genes matched the trajectory of pseudotime analysis and predicted drug response probability scores. Using predicted CGs and DEGs and applying the Pearson's correlations they confirmed that the drug response results and CGs predicted in scDEAL have strong correlations to, at least, the I-BET treated cell pseudotime trajectory.

This approach has a problem with the prediction across different species, because we can't use the data from one species to predict some features in another one directly due to the genetic variation between them. It is more complex issue and for this purpose there was not enough scRNA-seq data in the public domain for mice at that time.

To summarize, we can say that scDEAL provided new results and an approach that allows predicting the response to a drug at the single-cell level. In addition to predicting the response to cancer drugs, this tool can allow to find the CGs that can be used as targets for experimental validations on drug–gene relations