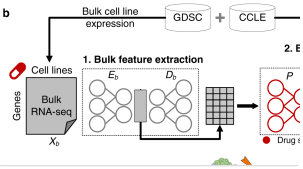
Deep transfer learning of cancer drug responses

Original paper:

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

Nature Communications - Single-cell RNA-seq data provide the opportunity to predict drug response in cancer while considering intratumour heterogeneity. Here, the authors develop a deep transfer...

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



Problem Context:

Cancer treatment often involves using drugs that can affect patients differently due to cancer heterogeneity among diverse states or cell fates.

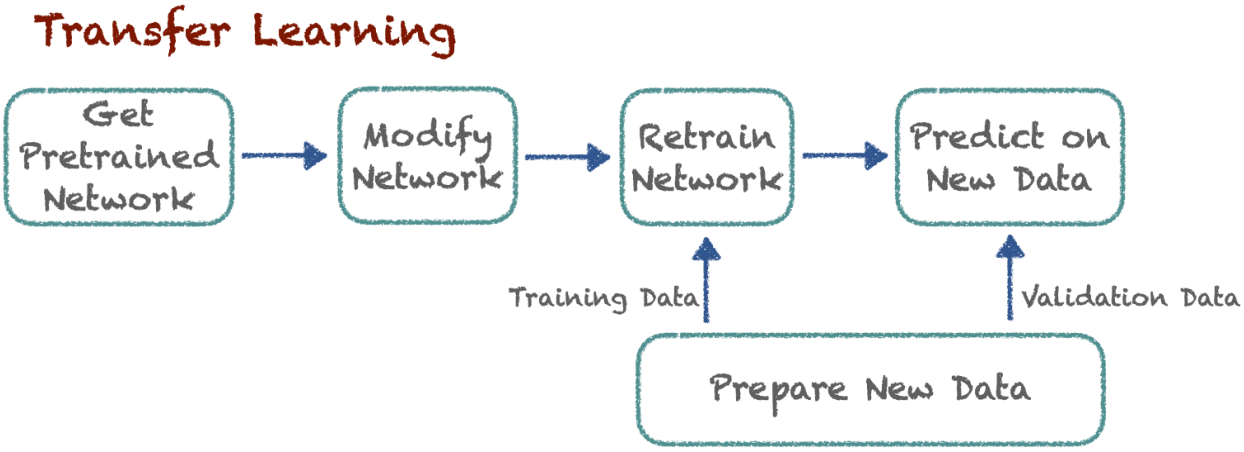
Such **heterogeneity of cells is responsible for differentiated responses of individual cells to a drug, leading to a minimal amount of cancerous residues remaining in the body, followed by, ultimately, cancer relapse**.

Traditional methods to predict how a patient will respond to a particular drug primarily use bulk RNA sequencing (bulk RNA-seq, the mix of all calls' RNA). Bulk RNA-seq provides an average expression profile of all the cells in a tumor sample, but it lacks detailed information about the diversity and heterogeneity within the tumor. Single-cell RNA sequencing (scRNA-seq), on the other hand, offers high-resolution data that reveals the heterogeneity of individual cells within a tumor. However, scRNA-seq data can be sparse and noisy due to technical limitations and high costs.

That's why authors combined those sources of information and created a precise neural network.

convolutional neural network (CNN) fine-tuning

In practice, training of entire CNNs is usually not done from scratch with arbitrary initialization. The reason is that it is usually not possible to find a dataset of sufficient size required for a network of the desired depth. Instead, it is most common to pre-train a DGNN on a very large data set and then use the weights of the trained DNN either as initialization or as feature extraction for a specific task.
The new dataset is relatively large and significantly different in content from the original dataset. Since the data set is very large, you can afford to train the entire GPSN from scratch. However, in practice, it often turns out to be more profitable to use a pre-trained model to initialize weights. In this case, we will have enough data to fine-tune the entire network.



To improve the prediction of drug responses in cancer, it is crucial to integrate the comprehensive, but less detailed bulk RNA-seq data with the detailed, but sparse scRNA-seq data. This integration aims to leverage the strengths of both data types to provide a more accurate and holistic understanding of tumor biology and drug response.

Authors' Proposal:

Deep Transfer Learning Framework:

The authors propose a novel deep transfer learning framework designed to integrate bulk RNA-seq and scRNA-seq data. They we develop scDEAL (single-cell Drug rEsponse AnaLysis) by adapting a Domain-adaptive Neural Network (DaNN) to predict drug responses from bulk and scRNA-seq data. Here's a breakdown of their approach:

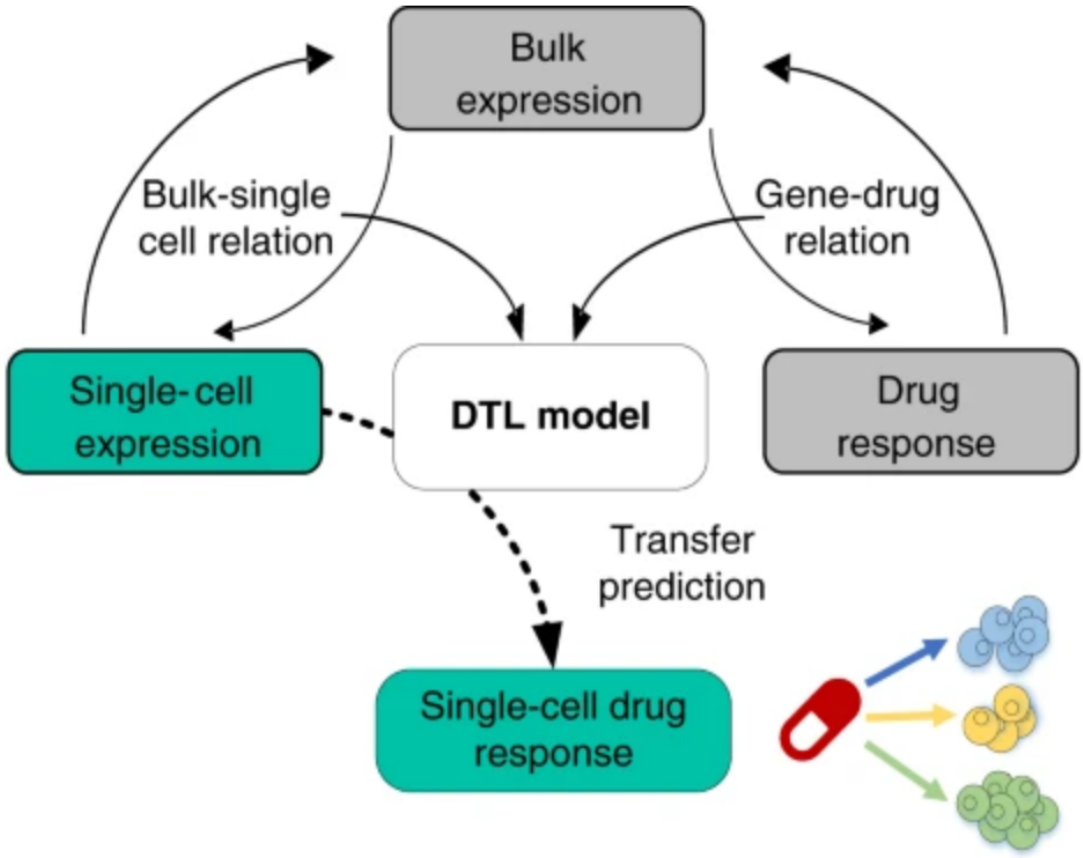
1. **Pre-training on Bulk RNA-seq Data:**
   - The model is initially trained on bulk RNA-seq data, where annotations for cell lines are available, which provides a broad and comprehensive view of gene expression across a large number of samples. This step helps the model learn general patterns and relationships within the data.
2. **Fine-tuning on scRNA-seq Data:**
   - After pre-training, the model is fine-tuned using scRNA-seq data. Fine-tuning allows the model to adjust and enhance its learned features with high-resolution, single-cell information. This step helps the model capture the intricate details and heterogeneity present in individual cells within the tumor.

Methodology and Metrics:

- **Input Data:**
  - Bulk RNA-seq data: Average gene expression profiles from tumor samples.
  - scRNA-seq data: Detailed gene expression profiles from individual cells within the tumor.
- **Processing:**

  The model uses a transfer learning approach where the knowledge gained from bulk RNA-seq data is transferred and refined using scRNA-seq data. This process enhances the model's ability to predict drug responses by combining the broad overview from bulk data with the detailed insights from single-cell data.
- **Evaluation Metrics:**

The performance of scDEAL in predicting drug responses was assessed using six publicly available datasets of single-cell RNA sequencing (scRNA-seq). These datasets involved the treatment of cells with five different drugs: Cisplatin, Gefitinib, I-BET-762, Docetaxel, and Erlotinib. The datasets included information regarding whether individual cells were sensitive or resistant to drugs, as indicated by ground-truth drug response annotations. The scDEAL predictions were assessed using seven metrics, which included F1-score, AUROC, AP score, precision, recall, AMI, and ARI. The findings indicated that scDEAL showed exceptional effectiveness in predicting drug response in individual cells. It achieved impressive average scores across all datasets, with values of 0.892 for F1-score, 0.898 for AUROC, 0.944 for AP score, 0.926 for precision, 0.899 for recall, 0.528 for AMI, and 0.608 for ARI.



Advantages and Limitations:

- **Advantages:**
  - By integrating bulk and single-cell RNA-seq data, the model achieves higher prediction accuracy and robustness compared to models using only one type of data.
  - This approach provides a more comprehensive understanding of tumor heterogeneity and its impact on drug response.
  - it can use a large amount of bulk-level drug response RNA-seq information from the Genomics of Drug Sensitivity in Cancer (GDSC) database and Cancer Cell Line Encyclopedia (CCLE) to train and optimize the model;

- 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;
- in order to avoid losing heterogeneity in scRNA-seq data, scDEAL includes cell cluster labels for loss function regularization in each training epoch;
- scDEAL's integrated gradient interpretation infers the signature genes of drug response predictions, which improves the interpretability of the model.
- scDEAL can find the CGs to drug response even though the sensitive and resistant tissues are derived from different patients and showcased the potential usage of scDEAL for combined data from different patients.

- **Limitations:**
  - The model's complexity may require significant computational resources.
  - Large datasets of both bulk and single-cell RNA-seq are necessary to train the model effectively.
  - remaining challenge in single-cell drug response prediction is the prediction across different species.

Conclusion:

The authors demonstrate that their deep transfer learning framework can significantly improve the prediction of cancer drug responses by effectively integrating bulk and single-cell RNA-seq data. This integrative approach offers a promising direction for advancing precision oncology and developing more personalized cancer treatments.