a. Current drug therapies for various types of cancer have low efficacy and high relapse rates. This is due to the genetic heterogeneity of cancer cells: individual cells respond differently to a drug, leading to increased resistance. Single-cell RNA sequencing (scRNA-seq) provides the ability to detect heterogeneous gene expression in cancer subpopulations in response to drugs, so using these data to predict and interpret cancer drug response would be useful. There is therefore a need to develop computational approaches, particularly those based on deep learning techniques, for large-scale analysis of single-cell data.

b. Existing approaches developed for bulk data (e.g. DESeq2, edgeR and others[1,2]) could not be directly applied to single-cell data. At the same time, prediction methods for single-cell data (e.g. scGNN[3], scDrug[4], scDR[5]) suffered from insufficient learning performance due to the limited amount of benchmark data.

c. The authors propose **scDEAL**, an approach that integrates RNA-seq data with scRNA-seq data. It is based on deep transfer learning (DTL), which is able to transfer knowledge and relationship patterns from bulk data to single-cell data. **scDEAL involves data pre-processing and five main steps**:

0. *The pre-processing of bulk RNA-seq and scRNA-seq data*. For data from the Genomics of Drug Sensitivity in Cancer (GDSC) and Cancer Cell Line Encyclopedia (CCLE) databases, min-max scaling of bulk expression profile data and waterfall binarisation of drug viabilities were performed. The data were then split into training, validation and test datasets. For data from GEO single cell datasets, annotation pre-processing, gene and cell filtering, normalisation and search for highly variable genes were performed.
1. *Extraction of bulk gene features*. A denoising autoencoder (DAE) is used to introduce noise into the bulk data. It uses an encoder (Eb) and decoder (Db) to extract low-dimensional features. This pre-training is used to generate the initial weights of the neurons in the DTL model.
2. *Drug response prediction in each bulk cell line*. A fully connected predictor (P) is attached to the trained bulk feature extractor of the cell lines.
3. *Single cell gene feature extraction*. A similar strategy is used to select single cell features using separated DAEs (Es and Ds).
4. *Joint training and updating of all models from the previous steps*. The overall scheme is trained to maximise the mean discrepancy between low-dimensional single-cell feature spaces and bulk data, the cross-entropy loss between predicted cell line drug responses and validation-derived labels, and the regularisation of cell clusters predicted from scRNA-seq data. Eb, Es and P are simultaneously updated and optimised to achieve minimal overall loss.
5. *Transferring and applying the trained model to scRNA-seq data*. scDEAL transfers trained Es and P to predict single cell drug responses based on scRNA-seq data.

d. To evaluate the performance of scDEAL in predicting drug response, the authors used 6 public scRNA-seq datasets treated with five different drugs. These data contain binary sensitivity scores for individual cells. Compared to true tags, scDEAL prediction was evaluated using seven metrics: F1 score, area under receiver operating characteristic (AUROC), AP score, precision, recall, adjusted mutual information (AMI) and adjusted random index (ARI). The average scores of the six datasets are 0.892 (F1 score), 0.898 (AUROC), 0.944 (AP score), 0.926 (precision), 0.899 (recall), 0.528 (AMI) and 0.608 (ARI). Given that the AMI and ARI scores are sensitive to mislabelling, especially with only binary data, it can be said that the drug response labels predicted in scDEAL for most cells are well aligned with the ground truth.

To clarify the design rationale of the scDEAL structure, the authors also replaced or removed certain components in scDEAL and compared the results with those of the final structure. Comparisons were made:

- Training the model on bulk data only without transfer VS with transfer;
- Using bulk data only from the GDSC database VS only from the CCLE database VS a combination of GDSC and CCLE;
- Using conventional autoencoders VS DAE without cell type regularisation VS DAE with regularisation.

The strategies implemented in scDEAL showed a significant increase in F1 scores compared to the others. It was also shown that:

- Different grid parameter settings had no significant effect oon scDEAL performance (480 combinations of six hyperparameters were tested).
- Calculation of integral gradient (IG) scores between sensitive and resistant cells confirmed the success of the study and the transfer of the relationship between gene expression and drug response from the bulk level to the single cell level.

Furthermore, **comparisons with the original studies** (drug response of leukemic cells, oral squamous cell carcinoma) showed that scDEAL is able to predict the drug response of leukemic cells more stably than in the original study, to correctly identify critical genes and to predict the drug response strongly correlated with pseudotime analysis.

Limitations:

- Difficulties in predicting drug response for different species. scDEAL model was only trained on human cancer cell lines due to limited drug-treated mouse scRNA-seq reference data.
- The need to use additional bulk gene expression collections of cell lines. To improve the accuracy of the prediction results, the authors plan to update the training data with new databases.

e. Conclusion

scDEAL is a deep transfer learning system for cancer drug response prediction at the single cell level. By matching drug-related bulk RNA-seq data with scRNA-seq data and integrated gradient-based feature interpretation, scDEAL efficiently handles drug response label prediction, gene signature identification, and prediction matching with pseudotime analysis. scDEAL can aid in the study of cell reprogramming, drug selection, and drug repurposing to enhance therapeutic efficacy. Thus, scDEAL has significant potential to improve drug development at the single cell level.

References:

1. Madhukar, Neel S., and Olivier Elemento. "Bioinformatics approaches to predict drug responses from genomic sequencing." Cancer Systems Biology: Methods and Protocols (2018): 277-296.
2. Li, Yuanyuan, et al. "Predicting tumor response to drugs based on gene-expression biomarkers of sensitivity learned from cancer cell lines." BMC genomics 22 (2021): 1-18.
3. Wang, Juexin, et al. "scGNN is a novel graph neural network framework for single-cell RNA-Seq analyses." *Nature communications* 12.1 (2021): 1882.
4. Hsieh, Chiao-Yu, et al. "scDrug: from single-cell RNA-seq to drug response prediction." Computational and Structural Biotechnology Journal 21 (2023): 150-157.
5. Lei, Wanyue, et al. "scDR: predicting drug response at single-cell resolution." *Genes* 14.2 (2023): 268.