5. Supplementary Information

5.1. *Implementation*

We implemented the DeepGRMF models in Python 3.7.4 with PyTorch.[28] For training graph regularized matrix factorization, we set the learning rate to 0.001 and mini-batch size as 5096. The latent dimension of cell line and drug factor matrix should be the same, and we set them to be 64. Both λ_c and λ_d were 0.2. We only considered top 10 and 5 nearest neighbors for cell line and drug, respectively. An early stop was applied after 30 consecutive epochs without loss reduction. For training two neural networks, we used different sets of hyperparameters and model structures due to the difference in feature dimensions of cell line and drug. For the neural network I (f_θ), the batch size was 25 with a learning rate 0.00004, and the hidden layers were [2048, 512, 128]. For the neural network II (f_ϕ), the batch size was 10, the learning rate was 0.0002 and hidden layers were [192, 128]. We set the dropout rate at 0.2 for both neural networks for each layer except the output layer to prevent overfitting. We also applied early stop for training these two neural networks. For both matrix factorization and neural networks, we used Adaptive Moment Estimation[29] as our gradient descent optimization algorithm. We also reduced the learning rate 10 times as loss stopped decreasing for 10 consecutive training epochs.

5.2. *Comparison between using gene expression and cell line factor as the features of cell line*

Learning more informative features of cell lines is extremely significant in drug sensitivity prediction, since it could improve the accuracy of our predictions. We can provide a better representation of cell lines by using neural network I (f_θ) trained in our model. DeepGRMF is a separate model, at step II, we trained a mapping function for cell lines that can map the gene expression data to the cell line factor via neural network I (f_θ). To predict drug sensitivity of a new cell line, most methods directly utilize the gene expression profiling of the cell lines as input. We can alternatively provide a cell line factor by applying the trained neural network I (f_θ) to the gene expression profiling. We used the lasso model to compare these two representations as the features of cell lines. The result is shown in Table A1. The performance of using the cell line factor as the feature is significantly improved for both AUROC and AUPR.

Table A1. Performance of using original expression data and cell line factor as features of cell lines to predict drug sensitivity of new cell lines.

		Per Cell Line		Per Drug		Micro	
Dataset	Features	AUROC	AUPR	AUROC	AUPR	AUROC	AUPR
GDSC	Gene Expression	79.1	53.8	67.1	38.2	79.3	55.4
	Cell Line Factor	**82.6**	**59.5**	**70.8**	**41.5**	**82.5**	**61.4**
CCLE	Gene Expression	79.2	67.5	66.2	38.2	74.1	50.5
	Cell Line Factor	**82.0**	**70.4**	**67.5**	**40.7**	**75.6**	**52.7**

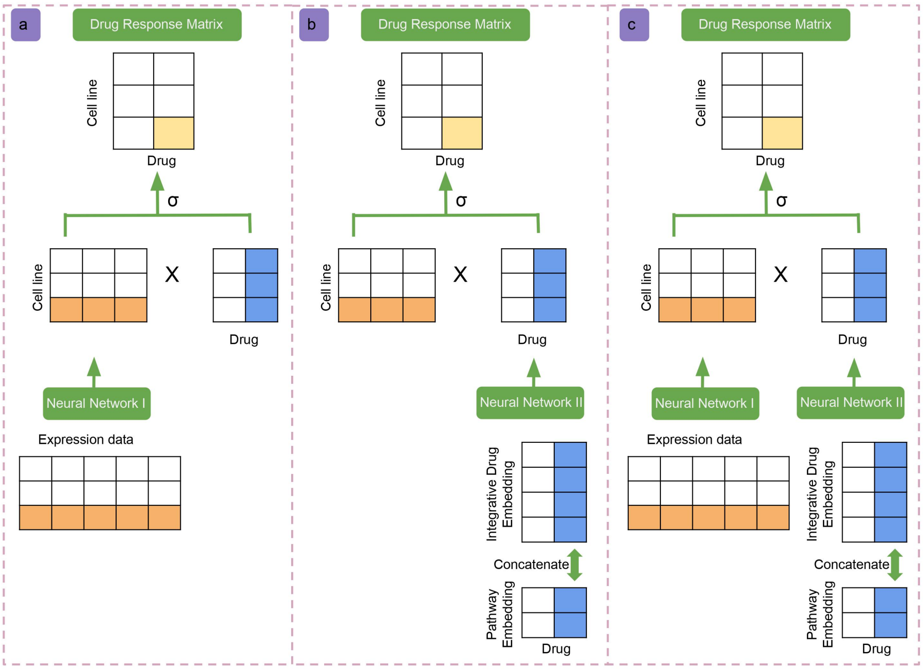

Fig. A1. Diagram of DeepGRMF model in three different tasks at the time of testing. Task a is to predict drug response of new cell lines to existing drugs; Task b is to predict drug response of existing cell lines to new drugs; Task c is to predict drug response of new cell lines to new drugs.

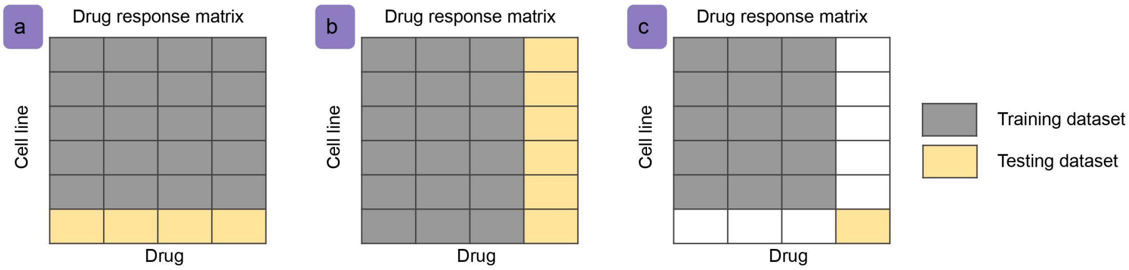

Fig. A2. Train-test split strategies in three different tasks. Task a is to predict drug sensitivity of new cell lines to existing drugs; Task b is to predict drug sensitivity of existing cell lines to new drugs; Task c is to predict drug sensitivity of new cell lines to existing drugs.

Table A2. Performance of comparison of using different similarities to regulate the matrix factorization. We calculated the mean AUROC and AUPR in the graph regularized matrix factorization step.

Type of similarity	AUROC	AUPR
No similarity	79.4	56.8
Cell line similarity	83.7	65.6
Drug similarity	82.5	64.4
Both similarities	**86.5**	**71.0**

Table A3. Performance comparison of using different models to map gene expression data to MF embedding to predict drug sensitivity of new cell lines to existing drugs.

		Per Cell Line		Per Drug		Micro	
Dataset	Model	AUROC	AUPR	AUROC	AUPR	AUROC	AUPR
GDSC	Random Forest	82.5	58.7	68.5	39.5	82.2	60.1
	Elastic Net	83.0	60.0	70.1	40.7	82.8	61.4
	Neural Network	**83.2**	**60.1**	**70.9**	**41.8**	**83.1**	**62.0**