A DeepFake framework for prediction of cell type specific transcriptional states induced by drug treatment

Ramzan Umarov[1], Erik Arner[1,2]

[1]Graduate School of Integrated Sciences for Life, Hiroshima University, Higashi-Hiroshima, 739-8528 Japan

[2]Laboratory for Applied Regulatory Genomics Network Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa, 230-0045 Japan

Correspondence: Erik Arner (arner@hiroshima-u.ac.jp)

Abstract

Drug treatment induces cell type specific transcriptional programs[1], and as the number of combinations of drugs and cell types grows, the cost for exhaustive screens measuring the transcriptional drug response becomes intractable. We developed DeepCellState, a neural network auto-encoder framework based on the DeepFake approach[2], for prediction of the induced transcriptional state in a cell type after drug treatment, based on the drug response in another cell type. Training the method on a large collection of transcriptional drug perturbation profiles[3], prediction accuracy improves significantly over baseline when applying the method to two cell types, with improved accuracy when generalizing the framework to additional cell types. Treatments with drugs or whole drug families not seen during training are predicted with similar accuracy, and the same framework can be used for predicting the results from other interventions such as gene knock-downs. Finally, analysis of the trained model shows that the internal representation is able to learn regulatory relationships between genes in a fully data driven manner.

Main

High throughput profiling of the transcriptional effects of drug treatment has proven to be useful for analysis of drug mode of action[4], drug repurposing[5] and prediction of off-target effects from drug treatment[6]. Although large repositories of gene expression profiles for a multitude of drugs in multiple cell types are available[3,7], it is not combinatorially feasible to profile all existing drugs in all relevant cell types, motivating a need for methods that can accurately predict cell type specific drug responses.

DeepFake technology is mainly used in image and video processing applications to generate synthetic media where the likeness of one person is replaced by that of another one[2] by training an autoencoder to compress the original input into a lower dimension latent space, with the same encoder part used to compress media depicting two or more people but with separate decoders for each person (**Supplementary Figure 1**). This enables similar facial expressions or actions to be encoded in a similar way in the latent space, while at the same time allowing for reconstruction of person specific facial details on the decoder side.

Noting that autoencoders have previously been used to create compressed representations of gene expression data[8–11], we hypothesized that DeepFake technology could be applied to predict cell type specific transcriptional response to drug treatment, using a network architecture featuring a common encoder and cell type specific decoders (**Figure 1a**). We initially developed the framework for two cell types (**Figure 1b**), retrieving 12737 and 12031 transcriptional profiles from the LINCS database[3] measuring the drug perturbation response in MCF-7 and PC-3 cell lines respectively, to use for initial training and evaluation (**Methods**). The transcriptional response profiles, measured at 978 “landmark genes” on the L1000 platform, were averaged across treatment time and doses in order to decrease noise[4]. Finally, we only kept profiles of drugs that were available for both MCF-7 and PC-3 which resulted in 1750 profiles for each cell type.

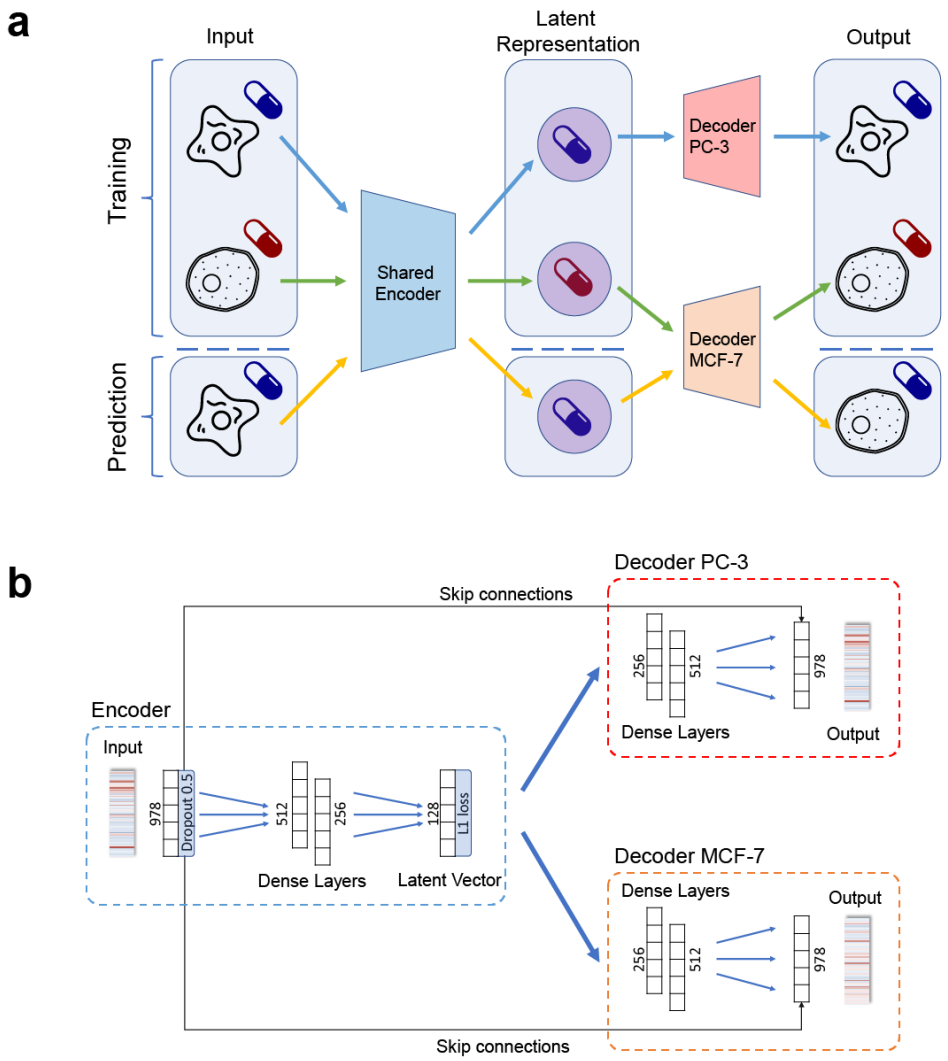

Fig. 1 The DeepFake framework of DeepCellState. **a.** Training and prediction procedures. Transcriptional profiles are encoded by a shared encoder that captures the drug response in a cell neutral way. The latent representations of the drug responses are decoded by cell specific decoders, which reconstruct the original input in a cell type specific way. After the shared encoder and decoders are trained, the response to a drug in a cell type can be predicted by using the drug's response in another cell type. **b.** Architecture of the encoder and the decoders used in DeepCellState. Dropout is applied to the input layer forcing the model to denoise the input. The input is encoded as a vector of size 128 with a sparsity constraint enforced by L1 regularization. The model also employs skip connections from the dropout layer directly to the output layer, allowing some of the input to be directly copied to the output during training.

A naive baseline method for predicting the transcriptional response in a cell type after treatment with a specific drug, is to predict that the response for that drug will be the same as in another cell type. After training our network and using 10-fold cross validation holding out 10% of drugs, we obtained an average Pearson correlation coefficient (PCC) of 0.60 between predicted and actual treatment response for drugs not seen by the network, a significant improvement (p<1e-300, t-test) compared to the baseline average correlation of 0.28 between the responses in input and output cell types (**Figure 2a,b,c,d, Supplementary Figure 2, Supplementary Figure 3**), corresponding to an average fold-increase of 2.14 over baseline (**Figure 2b**). In order to make sure that the training was not biased by being exposed to drugs with similar targets, we next revised the testing strategy to hold out entire drug families with the same modes of action[12], yielding similar results (**Supplementary Figure 2a, Supplementary Table 1, Methods**).

Proceeding to generalize the method to multiple cell types, we incrementally added transcription profiles from A375, HA1E, HT29, HELA and YAPC cells (1796, 1796, 1750, 1570 and 1570 number of profiles respectively after averaging across treatment times and dosages), obtaining increased (PCC=0.66, p=1.88e-59) performance when prediction was averaged over multiple input cell types (**Supplementary Figure 2b**) compared to the two cell type case. In comparison, a previously developed tensor based method treating cell type specific response prediction as an imputation problem[13] was reported to achieve average correlation of 0.54 using multiple cell types as input. We also tested the prediction capability using completely unseen cell types as input obtaining lower performance, yet significantly higher than baseline (**Supplementary Figure 2c**, PCC=0.49 compared to baseline PCC=0.28, p<1e-300).

To further validate the performance of DeepCellState, we evaluated its predictive capabilities on independently generated data sets on expression profiling platforms other than the L1000 platform used for the training data. In the first dataset we used, the expression profiles of four statins were measured in HepG2, MCF-7, and THP-1 cells[14] using CAGE technology. Using HepG2 profiles as input and predicting MCF-7 response, the DeepCellState predictions showed a correlation of 0.43 with the true response in MCF-7 cells (1.64-fold over baseline, p=0.037, **Supplementary Table 2, Figure 2e**). Next, we tested our method using data from a study measuring the response for 15 anticancer agents[15] on Affymetrix Human Genome U133A 2.0 Array platform, achieving 0.41 PCC (**Figure 2f)** compared to 0.28 PCC for baseline (1.46-fold improvement, p=0.026, **Supplemental Table 3**). Performance on this external dataset could be further improved through transfer learning, using additional training of our model on a subset of the external data set. To this end, we used the model trained on the LINCS data and additionally trained it on 14 drugs from the anticancer set to predict the effects of treatment with the 15th drug. This was repeated 15 times, each time leaving out one drug, resulting in an average performance of 0.56 PCC (p=4.04e-06 compared to baseline). This result suggests that the transfer learning approach is useful

to complete a small drug response data set which is not large enough to train the deep learning model by itself.

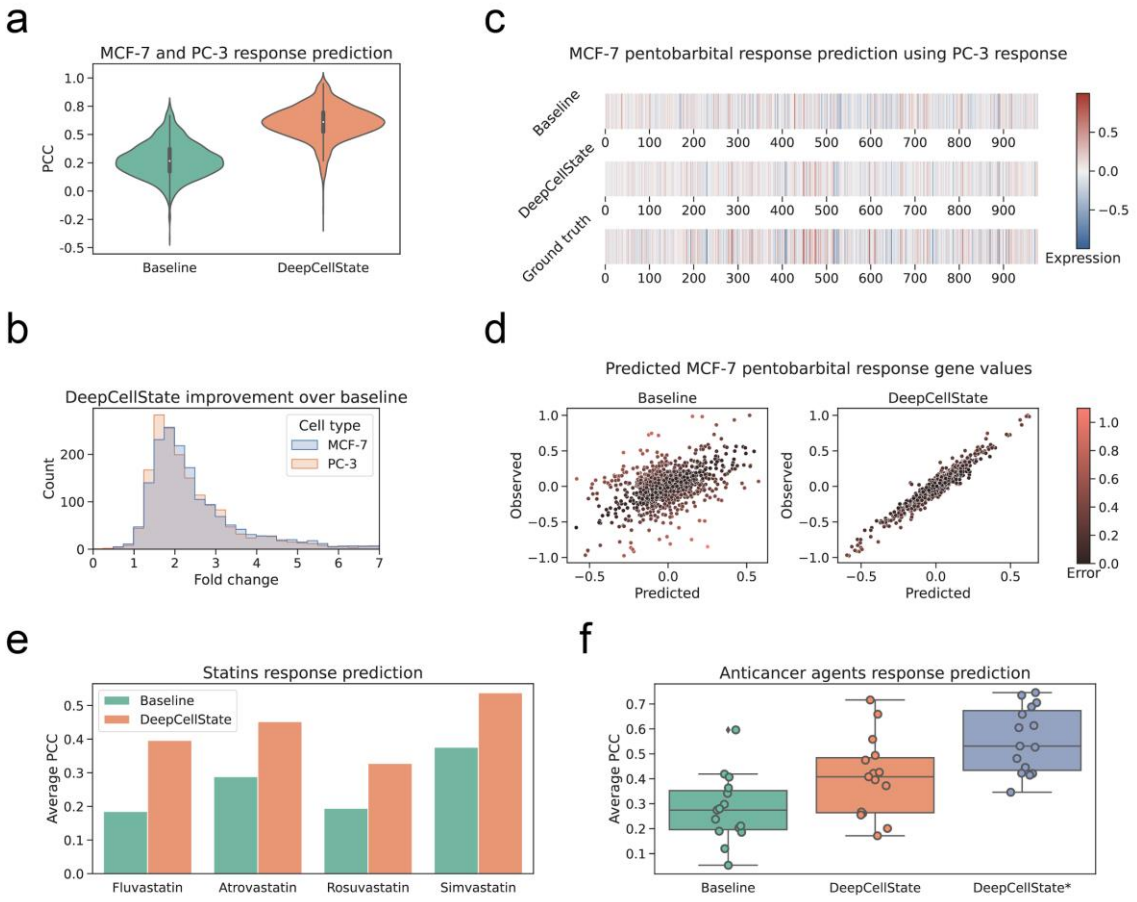

Fig. 2 Evaluation of the DeepCellState performance. **a.** 10-fold cross-validation results. DeepCellState performance compared to the baseline. **b.** Distribution of fold-change in DeepCellState performance over baseline for predicting MCF-7 (blue) and PC-3 (red) response. **c.** From top to bottom pentobarbital response in PC-3 (baseline), MCF-7 pentobarbital response predicted by DeepCellState based on the PC-3 response, and the observed MCF-7 response (ground truth). **d.** Scatter plot of individual gene values predicted for pentobarbital response compared to the observed response in MCF-7 cells. **e**. Comparison of DeepCellState and baseline performance for prediction of statin response from an external data set. **f.** Comparison of predicted anticancer drugs responses. DeepCellState* represents results obtained by applying transfer learning: performing additional training on 14 drugs and predicting one which is repeated 15 times each time leaving out one drug.

We further hypothesized that the DeepCellState approach may be useful for predicting the response to cellular perturbations other than drug treatment. To test this, we added transcriptional profiles from loss of function (LoF) experiments from the same database to the training set. Using the updated set, we obtained an average correlation of 0.56 (corresponding to an average fold-increase over baseline of 1.6, $p<1e-300$, **Supplementary Figure 2d**) on LoF experiments, the lower performance compared to drug perturbations likely due to that the number of available profiles for LoF was substantially lower than the number of drug treatment profiles (10230 LoF profiles in total compared to 24768 drug perturbation profiles for MCF-7 and PC-3 before

averaging). This result prompted us to investigate more deeply the relationship between the number of input profiles used in training and the prediction performance. Results showed that significant improvement over baseline was possible already for a low number of training examples, with higher performance as the number of training examples increased (**Figure 3a**).To study the impact of the latent layer size on the performance of DeepCellState we tested five different layer sizes (8, 16, 32, 64 and 128 neurons) and trained our model 10 times for each size. Average results for the test set (shown in **Figure 3b**) showed that performance started to drop significantly at a latent layer size below 32.

Finally, we set out to analyze general and cell type specific regulatory properties captured by the trained model, expecting that the latent vector itself should capture a general response to the input perturbations whereas the decoder parts should represent behavior related to functionality specific to each cell type. To this end, we computed the PCC between the output of the nodes in the latent layer for the drugs in the test set between MCF-7 and PC-3 cells. Indeed, although considerably different, the inputs from the same drug in the two cell types consistently yielded similar output in the latent layer as illustrated in a t-SNE plot (**Figure 3c)** in input and latent space, where drugs with profiles divergent for different cell types in the input space (**Figure 3c**, left panel) could be observed to be encoded using very similar latent representations (**Figure 3c**, right panel). Similarly, a matrix containing correlations between latent responses for MCF-7 and PC-3 cells, sorted by the magnitude of the correlation of the input profiles, showed a consistently strong signal on the diagonal **(Figure 3d)**, independent of the correlation between the responses in input space.

It has been shown that autoencoders are able to capture TF-gene regulatory interactions[8,16]. To test if this was the case also for the DeepCellState framework, we used TF target gene lists from ChIP-Atlas[17] which include MCF-7 specific targets. After filtering we computed the overlap between the targets implied by our model (see **Methods**) and targets based on ChIP-seq binding data, obtaining average fold enrichment of 2.5 (**Figure 3e**), with higher enrichment for MCF-7 targets in four out of six transcription factors, compared to the general set of targets according to ChIP-Atlas. Identification of the most important genes in the cell type specific decoder response (see **Methods**) revealed enrichment of biological processes[18] that matched known properties of each cell type (**Figure 3d**). For the top genes common to the decoders, the high ranking hits were enriched for terms including “response to abiotic stimulus” (p=1.44e-02) and “cellular response to chemical stimulus” (p=1.18e-02), expected since the model was trained using drug profiles (**Supplemental Table 4**). Additionally, several significantly enriched terms (p=1.12e-04, p=2.32e-04) were related to positive regulation of cell death and apoptotic process, likely due to some doses being toxic as previously observed[19]. Conversely, top genes of the MCF-7 decoder included several genes involved in mammary gland development (SRC, STAT5B, EGF). Similarly, BMP4, GLI2, and FGFR2 which are known to be involved in prostate morphogenesis were among the

top important genes of the PC-3 decoder. Taken together, these results suggest that the trained network is able to capture known biological processes and regulatory relationships, and that additional, deeper analysis may reveal relationships previously unnoticed.

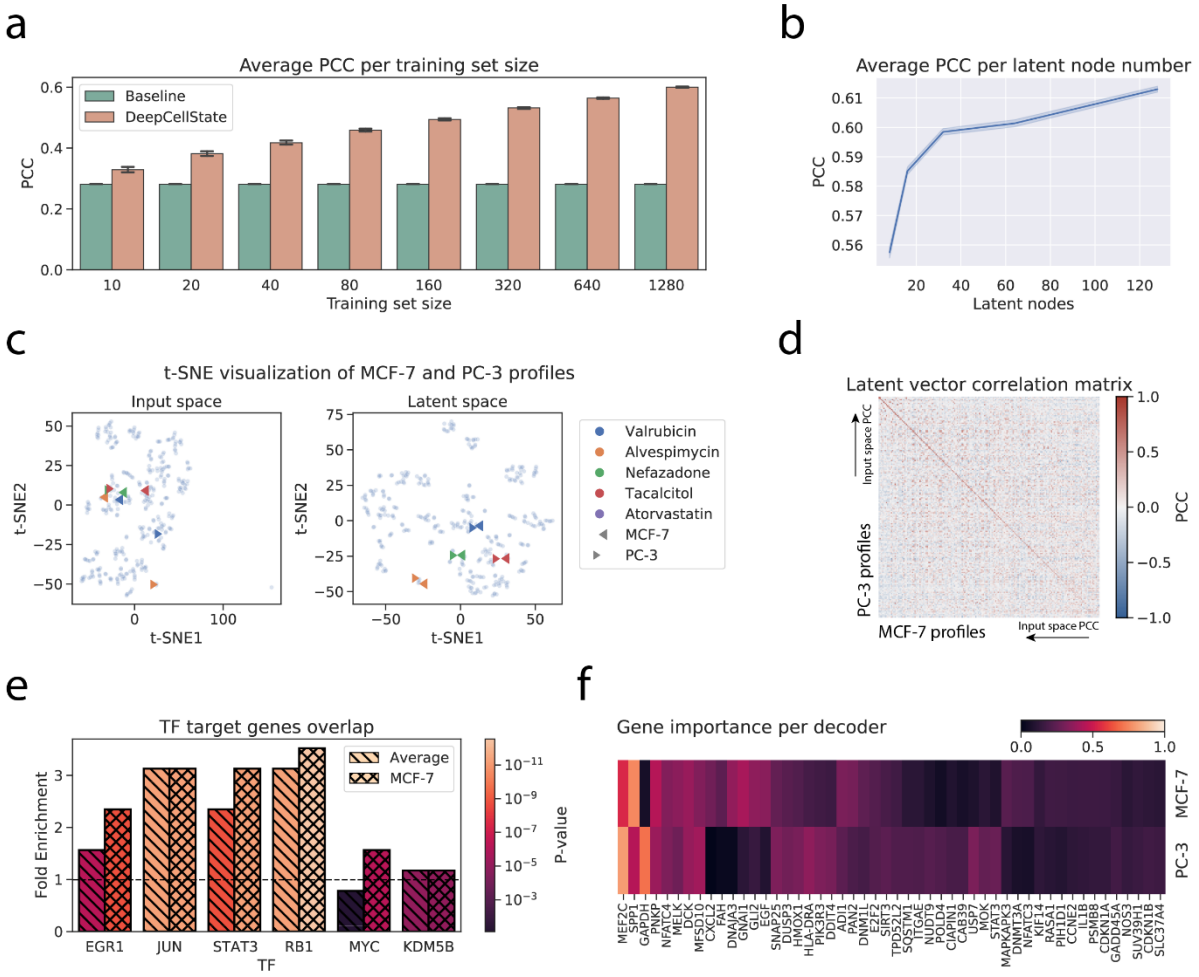

Fig. 3 Analysis of the developed method. **a.** Average PCC obtained for different size subsets of the training set. **b.** Dependency between the number of latent nodes and average PCC. **c.** t-SNE clustering of profiles in input and latent space. **d.** Matrix displaying latent vector response correlation. Drugs on each axis are sorted from high to low correlation in input space. **e.** Overlaps of top target genes of TFs based on chromatin immunoprecipitation data compared to target gene prediction based on DeepCellState trained model. **f.** Heatmap of most important genes for MCF-7 and PC-3 decoders.

To conclude, we have demonstrated that the DeepFake neural network architecture can be successfully implemented to predict the transcriptional drug response in a cell type based on the response in another cell type. Moreover, our results training on even a relatively small amount of LoF data suggest that the approach may be able to capture changes in cellular state in a more general way. The results here serve as proof of principle and may be improved by various modifications to the architecture and training strategy. In addition to tuning parameters such as the number of nodes in the different

layers of the network, the use of generative adversarial networks (GANs) has previously been shown to improve performance in the traditional DeepFake setting[20]. In this case, a separate model (discriminator) is trained to distinguish between generated and real media. The decoders and discriminators compete with each other, which results in decoders producing more realistic media. The addition of GANs to the architecture of DeepCellState may be one way of increasing performance. Additionally, the performance may be increased by measuring a larger set of genes than the 978 "landmark genes" measured on the L1000 platform, or by adding measurements of additional transcriptional features such as non-coding RNAs and transcribed enhancers. Performance is also expected to increase as more data capturing additional cellular states becomes available, especially if produced under controlled and standardized experimental conditions. Finally, our results suggest that DeepFake technology may have a wider range of applications than previously appreciated, not exclusive to image analysis but also biological, and perhaps other, domains.

Methods

LINCS data processing

Drug response expression profiles from LINCS phase 2 dataset (GSE70138) were used for training and evaluation. The latest level 5 data (signatures from aggregating replicates and converted to z-scores) was used (downloaded May 27, 2020). There are usually several treatment times and doses for each drug treatment; we averaged the profiles for the same drug across different timepoints and doses to reduce noise. Finally, the profiles were normalized to the range from -1 to +1 based on the whole data matrix.

Model architecture and training

The architecture of the autoencoder used in our method is shown in **Figure 1b**. The encoder consists of two dense layers with 512 and 256 neurons respectively. Dropout is applied to the input to increase the generalization capabilities of the model. The dropout rate is 0.5 which means half of the input genes will be set to 0 during training. This makes our model act as a denoising autoencoder since all values in the output layer need to be predicted. The last layer of the encoder is flattened and fed into a dense layer with 128 nodes, which represents our latent vector. The decoder takes the latent vector as an input and consists of two layers with 256 and 512 neurons respectively. The output layer has a direct connection to the dropout layer. The motivation is that some of the input can be "leaked" into the output[21].

We use L1 regularization for the latent layer, which acts as a sparsity constraint on the activity of the latent representations. Activation function is leaky relu[22] with alpha equal to 0.2 for all layers except the output layer, which uses tanh activation. The optimizer

used is ADAM[23] with a learning rate equal to 1e-4 and batch size is 128. Validation set is employed for early stopping with patience of 5 epochs.

To initialize our model, we first train a regular autoencoder using profiles of all cell types as both input and output. Then we create a separate decoder for each cell type by copying the weights from the autoencoder. These decoders are trained one by one together with the shared encoder, where the inputs are profiles from all cell types and the outputs are profiles with the decoder's cell type only. In the final stage, when there has been no improvement in the validation set for 5 epochs, the encoder weights are frozen and only decoders are trained until there is no improvement. When making a prediction for a certain drug and cell combination, any cell profile can be used as an input. The cell specific decoder is used to make the final prediction.

DeepCellState is implemented in Python 3 using TensorFlow[24] library and the models were trained using NVIDIA Quadro RTX 6000 GPU. Training two cell decoders takes about one hour to complete, however new predictions can be made in less than a second.

External data processing

We trained an encoder and three decoders (HepG2, MCF-7 and PC-3) for external validation. The common drugs from the two external data sets were removed from the training and validation sets. The statin data set was generated by converting CAGE data to gene expression data by mapping to GENCODE version 34. CAGE tags in a 1000 base pair window around a landmark gene start were summed. We added a small value c = 2 to each gene value to avoid problems when calculating log fold change. Each generated profile was multiplied by a million and divided by the total sum of the profile. The resulting profiles for controls and treatments were averaged across replicates followed by a computation of log fold change after treatment for each gene. For the anticancer data we directly calculated log fold change between the average control and treatment values. In both statin and anticancer tests, we only used landmark genes, setting missing gene values to 0. Only the 24 h time point was used as it is the most common time point in LINCS phase 2 compound data (96% of the data).

Methods related to analysis/interpretation of model

Since it is hard to directly learn the inner-workings of the deep learning model, we study it by feeding in different inputs and analyzing the output. To calculate the top genes for each decoder, we iterated through all the data, for each profile using 100 random genes subsets of size 100 to input in the model. The gene sets which performed the best to make the prediction were considered to be important. For each profile we picked the top 10 subsets that performed the best. After doing this for all the profiles, we ranked the genes based on their frequency in these subsets.

For the TF analysis, we downloaded target genes for 22 TFs profiled in MCF-7 cells from CHIP-Atlas[17]. We filtered them by retaining TFs that were upregulated in at least one drug treatment profile (i.e. has value > 0.5), and that had at least 50 targets in ChIP-Atlas (with targets defined as genes having a binding score of at least 100). This resulted in 6 remaining TFs and we picked the top 50 target genes for each. To make a prediction of target genes using DeepCellState, we started with a random profile and performed gradient descent optimization to find the input profile that maximizes the TF. As the obtained profile will also have other upregulated genes, the top 50 with the highest values are the DeepCellState prediction for the TF targets. For each TF we calculated fold enrichment and p-values using binomial statistic for the overlap between the DeepCellState set and the ChIP-Atlas set, using a similar method as described on the PANTHER website[18].

Data and code availability

Trained models, code to generate them, and code for analysis and figures described in this study are available at the following GitHub repository: https://github.com/umarov90/DeepFake. All data used in training and validation is publicly available through databases referenced above.

Author contributions

EA conceived the study and supervised the research. RU performed the research and drafted the manuscript together with EA.

References

1. Niepel, M. *et al.* Common and cell-type specific responses to anti-cancer drugs revealed by high throughput transcript profiling. *Nat. Commun.* **8**, 1186 (2017).

2. Nguyen, T. T., Nguyen, C. M., Nguyen, D. T., Nguyen, D. T. & Nahavandi, S. Deep learning for deepfakes creation and detection. *ArXiv Prepr. ArXiv190911573* **1**, (2019).

3. Subramanian, A. *et al.* A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171**, 1437-1452.e17 (2017).

4. Iorio, F. *et al.* Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 14621–14626 (2010).

5. Iorio, F., Rittman, T., Ge, H., Menden, M. & Saez-Rodriguez, J. Transcriptional data: a new gateway to drug repositioning? *Drug Discov. Today* **18**, 350–357 (2013).

6. Ye, C. *et al.* DRUG-seq for miniaturized high-throughput transcriptome profiling in drug discovery. *Nat. Commun.* **9**, 4307 (2018).

7. Lamb, J. *et al.* The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**, 1929–1935 (2006).

8. Chen, L., Cai, C., Chen, V. & Lu, X. Learning a hierarchical representation of the yeast transcriptomic machinery using an autoencoder model. *BMC Bioinformatics* **17**, S9 (2016).

9. Danaee, P., Ghaeini, R. & Hendrix, D. A. A DEEP LEARNING APPROACH FOR CANCER DETECTION AND RELEVANT GENE IDENTIFICATION. *Pac. Symp. Biocomput. Pac. Symp. Biocomput.* **22**, 219–229 (2017).

10. Xie, R., Wen, J., Quitadamo, A., Cheng, J. & Shi, X. A deep auto-encoder model for gene expression prediction. *BMC Genomics* **18**, 845 (2017).

11. Rampášek, L., Hidru, D., Smirnov, P., Haibe-Kains, B. & Goldenberg, A. Dr.VAE: improving drug response prediction via modeling of drug perturbation effects. *Bioinforma. Oxf. Engl.* **35**, 3743–3751 (2019).

12. Huang, C.-T. *et al.* Perturbational Gene-Expression Signatures for Combinatorial Drug Discovery. *iScience* **15**, 291–306 (2019).

13. Hodos, R. *et al.* Cell-specific prediction and application of drug-induced gene expression profiles. *Pac. Symp. Biocomput. Pac. Symp. Biocomput.* **23**, 32–43 (2018).

14. Huang, Y. *et al.* A framework for identification of on- and off-target transcriptional responses to drug treatment. *Sci. Rep.* **9**, 17603 (2019).

15. Monks, A. *et al.* The NCI Transcriptional Pharmacodynamics Workbench: A Tool to Examine Dynamic Expression Profiling of Therapeutic Response in the NCI-60 Cell Line Panel. *Cancer Res.* **78**, 6807–6817 (2018).

16. Tan, J. *et al.* Unsupervised Extraction of Stable Expression Signatures from Public Compendia with an Ensemble of Neural Networks. *Cell Syst.* **5**, 63-71.e6 (2017).

17. Oki, S. *et al.* ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. *EMBO Rep.* **19**, (2018).

18. Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* **47**, D419–D426 (2019).

19. Sirci, F. *et al.* Comparing structural and transcriptional drug networks reveals signatures of drug activity and toxicity in transcriptional responses. *NPJ Syst. Biol. Appl.* **3**, 23 (2017).

20. Shen, T., Liu, R., Bai, J. & Li, Z. “Deep Fakes” using Generative Adversarial Networks (GAN). (2018).

21. Mirsky, Y. & Lee, W. The Creation and Detection of Deepfakes: A Survey. *ArXiv200411138 Cs Eess* (2020) doi:10.1145/3425780.

22. Maas, A. L., Hannun, A. Y. & Ng, A. Y. Rectifier nonlinearities improve neural network acoustic models. in *Proc. icml* vol. 30 3 (2013).

23. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *ArXiv Prepr. ArXiv14126980* (2014).

24. Abadi, M. *et al.* Tensorflow: A system for large-scale machine learning. in *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)* 265–283 (2016).