Journal of the American Medical Informatics Association, 28(1), 2021, 42–51
doi: 10.1093/jamia/ocaa212
Advance Access Publication Date: 11 October 2020
Research and Applications

Research and Applications

Anticancer drug synergy prediction in understudied tissues using transfer learning

Yejin Kim,[1] Shuyu Zheng,[2] Jing Tang,[2] Wenjin Jim Zheng,[1] Zhao Li,[1] and Xiaoqian Jiang[1]

[1]Center for Safe Artificial Intelligence for Healthcare, School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, Texas, USA and [2]Research Program in Systems Oncology, Faculty of Medicine, University of Helsinki, Helsinki, Finland

Corresponding Author: Yejin Kim, School of Biomedical Informatics, University of Texas Health Science Center at Houston, 7000 Fannin Street, Houston, TX 77030, USA (yejin.kim@uth.tmc.edu)

Received 1 April 2020; Editorial Decision 10 August 2020; Revised 21 July 2020; Accepted 14 August 2020

ABSTRACT

Objective: Drug combination screening has advantages in identifying cancer treatment options with higher efficacy without degradation in terms of safety. A key challenge is that the accumulated number of observations in in-vitro drug responses varies greatly among different cancer types, where some tissues are more understudied than the others. Thus, we aim to develop a drug synergy prediction model for understudied tissues as a way of overcoming data scarcity problems.

Materials and Methods: We collected a comprehensive set of genetic, molecular, phenotypic features for cancer cell lines. We developed a drug synergy prediction model based on multitask deep neural networks to integrate multimodal input and multiple output. We also utilized transfer learning from data-rich tissues to data-poor tissues.

Results: We showed improved accuracy in predicting synergy in both data-rich tissues and understudied tissues. In data-rich tissue, the prediction model accuracy was 0.9577 AUROC for binarized classification task and 174.3 mean squared error for regression task. We observed that an adequate transfer learning strategy significantly increases accuracy in the understudied tissues.

Conclusions: Our synergy prediction model can be used to rank synergistic drug combinations in understudied tissues and thus help to prioritize future in-vitro experiments. Code is available at https://github.com/yejinjkim/synergy-transfer.

Key words: ranking, data paucity, data scarcity, multitask learning, external validation, drug sensitivity

INTRODUCTION

Discovering synergistic drug combinations in cancer treatment has been widely studied as it identifies candidates with higher efficacy without degradation in terms of safety.[1–4] Many previous studies show great promise in finding potent drug combinations that cannot be identified by 1 drug–1 target approaches.[5–10] A considerable amount of drug synergy prediction methods has been investigated including pharmacokinetic and pharmacodynamic methods,[10,11] posttreatment transcriptome,[12,13] feature-based machine learning,[5,7,8,14,15] neural networks with multimodal data sources,[4,16] and large community efforts.[17,18] These studies utilize various features, such as molecular profiles, chemical structure, pre- or posttreatment transcriptome. Their prediction target usually remains in a specific pathway, cell line, or tissue. These studies are usually based on single databases, and the prediction was also made within the data-

base.[3–5,18] A systematic integration of multiple drug synergy databases has been recently proposed,[19,20] which sheds light on developing unbiased drug synergy prediction models.

On the developing unbiased and generalizable drug synergy prediction model, a key challenge is an imbalance of synergy experimental data, in which the accumulated number of observations in in-vitro drug responses varies greatly for different tissues. Traditional methods target studying commonly observed tissues such as breast, kidney, skin, and lungs.[4,7,17,21–25] These methods take known drug response data at certain cell lines and attempt to find other drug responses from other cell lines within the same data-rich tissues. The understudied tissues include bone, prostate, and pancreas (Figure 1a and b). A number of obstacles impede the drug response study in these tissues. For example, bone cancer drug response has been poorly understood due to physical difficulty of culturing bone tissue as cell lines, the rarity of tumors in sarcoma type, the difficulties of obtaining tumor tissue fragments from human patients for bone metastasis, and thus the limited number of cell models.[26] The lack of cell line models makes the high-throughput screening difficult, which in turn makes these tissues even more understudied. There is, therefore, a critical need to develop a computational drug combination prediction tool for the understudied tissues.

Our goal is to develop a generalizable in-silico drug synergy prediction model for understudied tissues. The most critical challenge is data scarcity. Understudied tissues inevitably lack enough training data; they have a limited number of experimental observations and also lack important features, such as post-treatment transcriptome. A potential way to mitigate this data scarcity problem is to utilize information from the data-rich tissues to the data-poor tissues as these different tissues share biological commonality, partly in terms of gene expression, and therefore respond to drugs in similar ways.[27] Several previous studies support that anticancer drug sensitivity in cell lines is not tissue-specific and that tissue-specific drugs can bring additional benefits to other tissues.[28,29] Accordingly we proposed to utilize interaction between drugs and cell lines learned from the data-rich tissues to help increase the performance of the data-poor tissue using *transfer learning*. Transfer learning is to transfer knowledge (in the form of parameters in machine learning models) from a previously learned model (teacher model) with more data to a new model (student model) with fewer data. Several methods have been proposed to utilize transfer learning in drug sensitivity prediction to link complementary data sources, different modality of auxiliary features, and different diseases.[30,31] To learn generalizable and thus transferable knowledge on drug target and cell line gene expressions, we integrated all types of available tissues into 1 model with a large dataset (ie, 4150 drugs; 112 cell lines; 710 242 monotherapy sensitivity; 1 175 220 combinations synergy from 15 different synergy studies). We maximally utilized various types of multimodal inputs (molecular, genetic, phenotypic features) and multiple outputs (drug sensitivity and synergy) using flexible neural network models, and we validated our ranking models and drug combinations with independent external databases.

MATERIALS

Drug sensitivity and synergy

We used a publicly available large-scale drug synergy database from DrugComb Portal,[19] which combines 15 drug synergy studies into 1 integrated database including O'Neil, Cloud, ALMANAC, NCATS Matrix, Forcina, Mathews, Phelan, Wilson, and Yohe (Supplementary Table S1).[3,32–38] The number of unique drugs and cell lines were 4150 and 112, respectively (Figure. 1). There were 2843 experimental drugs and 1307 FDA-approved drugs. There was a total of 710 242 monotherapy sensitivity (ie, a pair of drug and cell line) and 466 259 drug combinations synergy (ie, a triplet of drug1, drug2, and cell line). For monotherapy sensitivity we calculated relative inhibition (RI) from a dosage combination matrix (Supplementary Material S1). For drug combinations synergy we used Loewe synergy score (Figure 1), which ranged from −100 (antagonistic effect) to 75 (strong synergistic effect).[39] Loewe synergy score is to quantify the excess over the expected response if the 2 drugs are the same compound.[40,41] We selected Loewe score for comparison with the baseline study.[4]

Drug features

We extracted the drug's molecular and genomic features. For molecular features, we used both Molecular ACCess System (MACCS) fingerprints[42] and native chemical compounds. MACCS fingerprints contain 166 chemical structures, such as the number of oxygen, S-S bonds, and rings. We used RDKit (http://www.rdkit.org) to extract MACCS fingerprints. In addition, we represented drugs as a native chemical structure using SMILES. SMILES is a linear notation to represent chemical compounds uniquely; in the SMILES representation, atoms are represented as their atomic symbols (eg, c for carbon); special characters are also used to represent relationship (eg, "=": double bonds; "#":triple bonds; ".":ionic bond; ":": aromatic bond).[43] SMILES can provide richer features space that strictly represent functional substructures and express structural differences, such as a compound's chirality.[44] We used the state-of-the-art natural language processing model, Transformer, to encode the SMILES sequence.[45] For genomic features, we integrated 3 different drug databases—DrugBank,[46] Therapeutic Targets Database (TTD),[47] and NIH-LINC[48]—for a complete view of a drug's target genes (Supplementary Table S2). We filtered nonhuman target genes in TTD.

Cell line features

We used the cell line's cancer type and genomic features. There were 14 tissues including lung, ovary, and skin and 14 cancer types, including carcinoma, adenocarcinoma, and melanoma (Supplementary Table S3). We also extracted gene expression profiles by Fragments Per Kilobase of transcript per Million reads mapped (FPKM) from CCLE and COSMIC.[49,50] We found 75 cell lines with 37 260 genes from Broad Institute and 25 cell lines with 35 004 genes from Sangar (Supplementary Table S2). In total, we found 88 cell lines with 22 586 genes after excluding zero-variance genes. We normalized the FPKM values into z-score in a gene-wise manner, as FPKM varies greatly depending on gene. We only used baseline (before-treatment) gene expression profiles without after-treatment gene profiles because our objective is to test our models in understudied tissues, which rarely has after-treatment gene profiles.

METHODS

Method overview

Our transfer learning approach to enhance synergy prediction in data-poor tissue is to transfer parameters from a pretrained general prediction model to the specific tissue's prediction models. We first learned the general prediction model with data-abundant tissues (in-

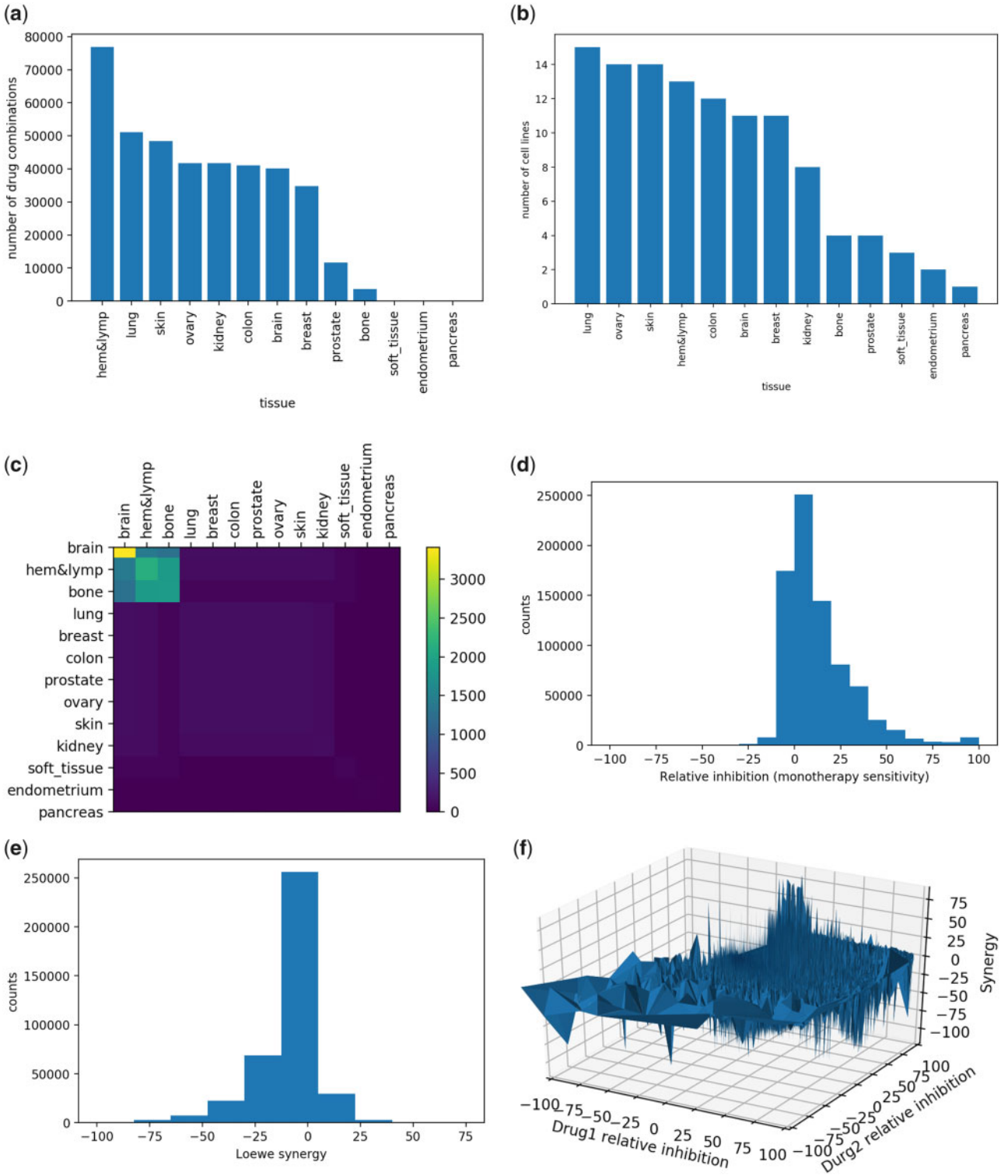

Figure 1. Summary statistics of the DrugComb database. (a) Number of drug combinations and (b) cell lines in each tissue. "Hem&lymp" = hematopoietic and lymphoid tissue. Data-rich tissues include hematopoietic and lymphoid, lung, skin, ovary, kidney, brain, breast, and colon, which have more than 20 000 drug combination tests with more than 5 different cell lines per each tissue. Data-poor tissues include prostate and bone, which have less than 15 000 drug combination tests within 3 or 4 cell lines per tissue. Soft tissue, endometrium, and pancreas have less than 1000 drug combination tests within less than 3 cell lines per tissue. (c) Number of tested drugs that overlap between tissues. Drugs for each tissue are rarely overlapped with drugs for other tissues except the brain, hematopoietic and lymphoid tissue, bone, and lung. (d) Distribution of monotherapy sensitivity (ie, relative inhibition) (e) Distribution of combination synergy score (f) Distribution of sensitivity and synergy. Concentrated and dense spikes lie in relative inhibition between 10 and 50, implying that many experiments have been performed within the relative inhibitions between 10 and 50. Loose spikes lie in relative inhibition below 0. Here we set the relative inhibition's binarization threshold at 50, as synergy scores show different distribution after 50.

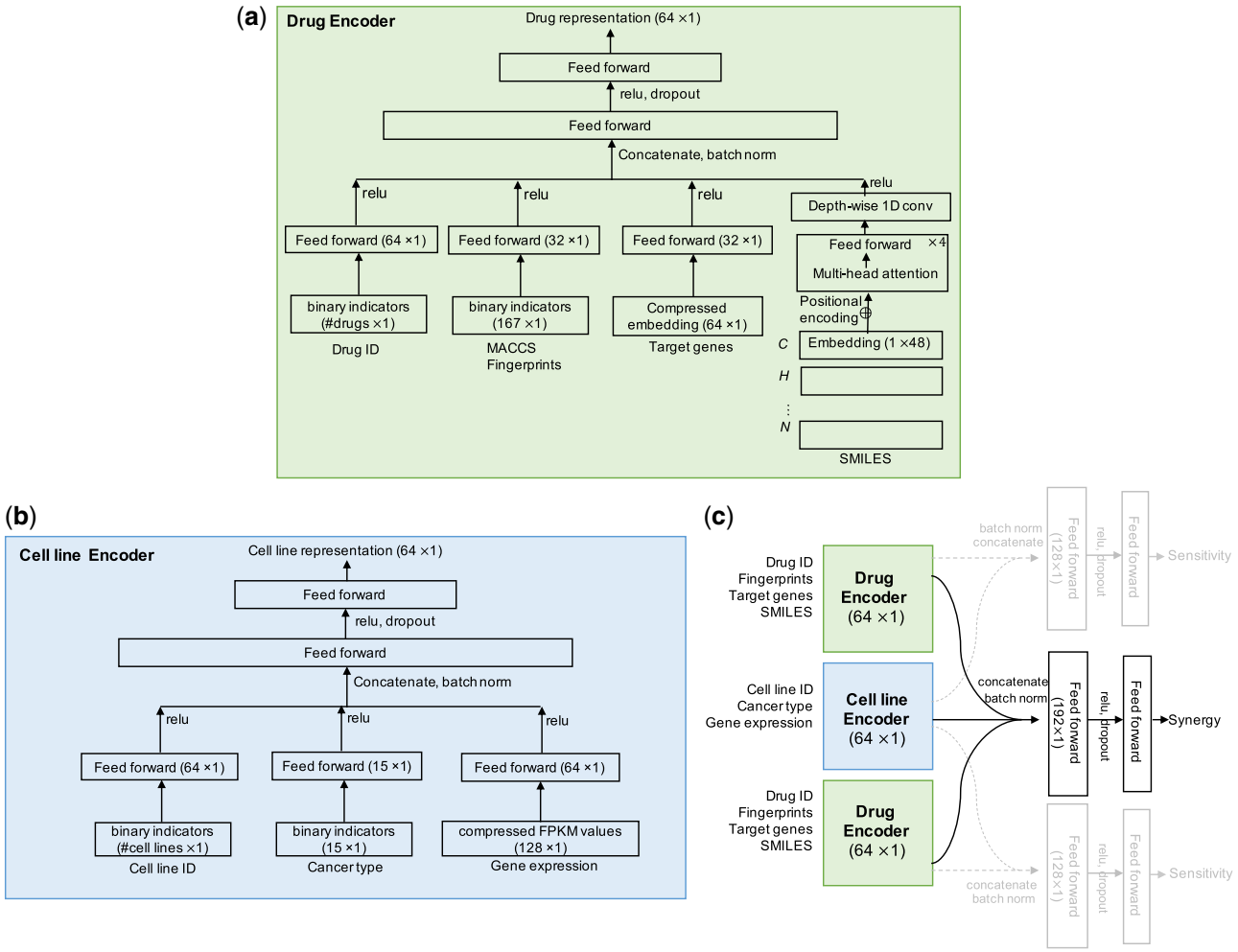

Figure 2. Drug synergy prediction model. (a) Drug encoder. It learns an embedding representation of a drug. Inputs are MACCS fingerprints, canonical SMILES, and target genes. (b) Cell line encoder. It learns an embedding representation of a cell line. Inputs are cancer type and gene expressions. (c) Merging drug encoder and cell line encoder. Sensitivity was an auxiliary output to boost synergy prediction.

cluding brain, breast, colon, endometrium, hematopoietic and lymphoid, kidney, lung, ovary, and skin) from the pooled drug synergy databases containing the 15 different databases. We used neural networks as the prediction model due to its flexibility in parameter sharing. The prediction model for the data-poor tissue was initialized with the pretrained general model's parameters and retrained with its own data accordingly.

General prediction model

Our objective is to predict whether unobserved drug combinations have synergistic effects in a given cell line and provide a list of combinations that researchers can prioritize for experiments. We formulated it as a prediction problem. Given an experimental block $x_n := \{d_i,\ d_j, c_k\}$ of drug combination $(d_i,\ d_j)$ and cell line c_k, the prediction model is a function f such that

$$y_{ijk} \approx f(d_i,\ d_j, c_k),$$

where y_{ijk} is the synergy score of the drug combination. In addition to the synergistic effect, we used the drug's monotherapy sensitivity as another prediction objective to boost the synergy prediction. For monotherapy sensitivity in the cell line, the prediction model is g:

$$z_{ik} \approx g(d_i, c_k) \text{ and } z_{jk} \approx g(d_j, c_k),$$

where z_{ik} is the sensitivity score of drug d_i in cell line c_k. We jointly optimize sensitivity and synergy as multitask learning. To incorporate multimodal sources of inputs and formulate nonlinear relationships between chemical and genomic features, we used deep neural networks as the prediction function. Our prediction model consists of drug encoders (Figure 2a), cell line encoder (Figure 2b), and merging layers (Figure 2c) for final prediction in an end-to-end manner. Using the estimated drug response, we can rank drug combinations at a cell line that are expected to have synergistic effect.

Drug encoder

Drug encoder learns an embedding representation of each drug d_i or d_j. Each drug's features are {MACCS fingerprints, canonical SMILES, target genes} (Table 1). Binary indicators of MACCS fingerprints were used as raw input. Binary vectors of target genes were represented as compressed embedding with a single layer of denoising autoencoder.[51] As SMILES has variable length, we needed more specific consideration. We used a Transformer encoder, a natural language processing model that converts the sequence into a representation.[45] Each symbol in SMILES was first represented as

Table 1. Multimodal input and multitask output to predict drug response

	Type	Size	Format
Input			
Drug features	MACCS fingerprint	166	Binary indicator vector
	Canonical SMILES	Variable length (max 288)	Sequences of characters (molecules and relation)
	Target gene	24 342 (Total number of unique genes)	Binary indicator vector
Cell line features	Cancer type	14	One-hot indicator vector
	Baseline gene expression	24 342 (Total number of unique genes)	Z-score of FPKM
Output			
Monotherapy sensitivity	Relative inhibition		Continuous value or binarized value at 50
Combination synergy	Loewe synergy score		Continuous value or binarized value at 30

Abbreviations: FPKM: Fragments Per Kilobase of transcript per Million reads mapped; SMILES: Simplified molecular-input line-entry system.

one-hot vectors with size of (#SMILES length * #unique symbols), where #unique symbols were 48 and the maximum SMILES length was 288. The symbol's one-hot vector was then represented as embedding. This embedding sequence was fed into a separate Transformer encoder, which consists of a multihead attention layer and a feed-forward layer with repeated residual connections. Once we derived all the embedding representations, they were concatenated into 1 and fed into 2 feed-forward layers with Relu activation and dropout (Figure 2a).

Cell line encoder

Cell line encoder learns an embedding representation of each cell line c_k. Each cell line's features are {cancer type, gene expressions} (Table 1). To combine the multimodal inputs, we derived embedding from each source and merged them into 1 embedding. The cancer types were represented as the same sized embedding. The gene expression of each cell line was represented as compressed embedding from the normalized FPKM values with 1 layer of denoising autoencoder. We concatenated all 3 embeddings into 1 and fed them into 2 feed-forward layers with Relu activation and dropout (Figure 2b).

Note that we did not include drug ID or cell line ID as an input. The IDs are useful in collaborative filtering techniques, which assumes that 2 drugs will react to a cell line similarly if these drugs have responded similarly to certain cell lines. In contrast, our focus was to derive transferable general knowledge in prediction models from a data-abundant domain to a data-scarce one. The cell lines between the 2 domains are disjointed from each other in most cases (cold start for all cases), and the data-scarce tissues usually do not have a past drug-response history.

Merging layers

For synergy prediction, we merged embedding representations of d_i, d_j, c_k into 1 and fed them into 2 feed-forward layers with Relu activation. For sensitivity prediction, we merged embedding representation from (d_i, c_k) and (d_j, c_k), respectively, and fed them into 2 feed-forward layers (Figure 2c).

Training loss

We trained the model as multitask learning that predicts synergy and sensitivity simultaneously. For combination synergy prediction, the training loss was mean squared error (MSE):

$$loss_{syn} = || \; y_{ijk} \; - f(d_i, d_j, c_k) \; ||^2$$

where y_{ijk} is the synergy score. The training loss for monotherapy (d_i, c_k) and (d_j, c_k) was similarly defined:

$$loss_{sen} = || \; z_{ik} \; - g(d_i, c_k) \; ||^2 \; + || \; z_{jk} \; - g(d_j, c_k) \; ||^2$$

for all drugs and cell lines in the training set. In addition to these regressions, we also developed classification models. We first binarized the drug responses by setting a threshold. That is, $_{y_{ijk}} = 1$if $y_{ijk} > threshold$; 0 otherwise, and $_{z_{ik}} = 1$if $z_{ik} > threshold$; 0 otherwise. The synergy and sensitivity thresholds were set at 30 and 50 following the previous study.[4,52] Then the classification model's training loss was binary cross entropy:

$$loss_{syn} = -\underline{y}_{ijk} \log \sigma(f(d_i, \; d_j, c_k)) - (1 - \underline{y}_{ijk}) \; log[1 - \sigma(f(d_i, \; d_j, c_k))]$$

$$loss_{sen} = -\underline{z}_{ik} \log \sigma(f(d_i, c_k)) - \; (1 - \underline{z}_{ik}) \; log[1 - \sigma(f(d_i, c_k))]$$

$$- \underline{z}_{jk} \log \sigma(f(d_j, c_k)) - \; (1 - \underline{z}_{jk}) \; log[1 - \sigma(f(d_j, c_k))]$$

where σ is a sigmoid function. We alternated the 2 optimization tasks with respect to synergy or sensitivity in every epoch. That is, we first minimized $loss_{syn}$ with all training batches and then switched to $loss_{sen}$ for the next training round. Optimizer was adaptive gradient descent with Autograd in PyTorch 1.3.0.

Fine-tuning to data-poor tissues

We focused on transfer learning from the pooled data-rich tissues to bone and prostate (Figure 3a), respectively. For the transfer learning experiments, we chose the bone and prostate tissue (number of observations > 1000) over soft tissue, endometrium, and pancreas (number of observations < 1000) (Figure 1a), because they have too few observations to build their own prediction model for comparative experiments. We first pretrained the general drug response prediction model with the data-rich tissues, to learn the underlying mechanism between drugs and cell lines, and built a separate prediction model for the specific tissue of interest (ie, bone or prostate), transferred the pretrained model parameters from the general model, and fine-tuned the specific tissue models' parameters. Possible fine-tuning strategies include i) retraining all the parameters, ii) fixing drug encoders, retraining cell line encoders, and merging layers; iii) fixing drugs and cell line encoders and retraining merging layers; and iv) no retraining at all. We compared

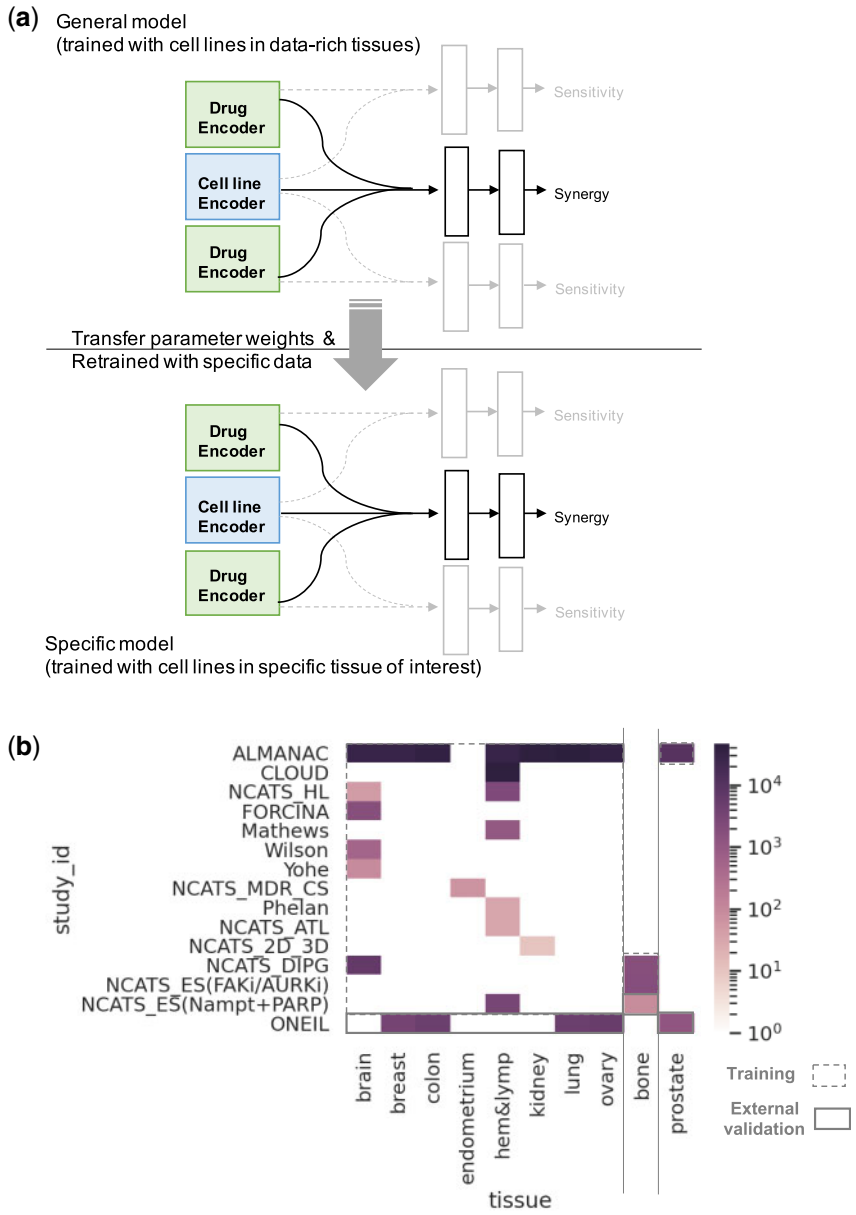

Figure 3. Transfer learning for understudied tissue. (a) Transfer model parameters from data-rich tissues to data-poor tissue, such as bone and prostate (b) Train/test split for data-rich tissues and data-poor tissue. Cell lines from different tissues are disjointed. Some cell lines from different tissues sometimes share drugs.

the accuracy of the 4 different strategies in each tissue and reported the best 1.

Evaluation

Training and test set

We evaluated the general and specific prediction models with cross and external validation. For cross validation, we randomly separated the pooled databases into training (80%) and testing sets (20%) (Figure 3b). For external validation, we set aside 1 independent database as a test set and used the remaining pooled databases as a training set (Figure 3b); the independent test set had only a few overlapping cell lines or drugs compared with that of the training set (eg, 9 cell lines and 36 drugs were common in training and test sets in data-rich tissues). Each data-poor tissue (bone and prostate) had 4 distinct cell lines, respectively (Figure 1a and b); different tissues sometimes shared drugs, such as the bone tissue sharing 1437 drugs and the prostate tissue sharing 127 drugs with the data-rich tissues.

Accuracy measures

We measured accuracy of the regression and classification task. The regression accuracy measure was MSE; the classification accuracy measures were area under the receiver operating curve (AUROC) and area under precision-recall curve (AUPRC).

Table 2. Drug synergy prediction accuracy for data-rich tissues

Train	Test	Methods	AUROC	AUPRC	MSE
15 pooled databases	15 pooled databases	collaborative filtering	0.9336	0.7484	151.9
14 pooled databases	O'Neil		0.6030	0.5081	706.0
O'Neil	O'Neil	Proposed method	0.9477	0.9388	181.7
ALMANAC	ALMANAC		0.9688	0.8376	75.4
15 pooled databases	15 pooled databases		0.9577	0.8335	132.4
14 pooled databases	O'Neil		0.6376	0.5564	742.1

Note: Cross validation with randomly split train/test from the 15 pooled databases. External validation with the O'Neil for test and the remaining 14 pooled databases for training. For classification tasks, AUROC and AUPRC were computed after binarizing synergy at threshold 30.

Abbreviations: AUPRC: area under the precision-recall curve; AUROC: area under the receiver operating curve; MSE: mean squared error.

Baselines

We compared our accuracy with two baseline methods: i) DeepSynergy that utilizes deep neural networks using drug's descriptors and cell line's gene expression on O'Neil database[4] and ii) ALMANAC + XGBoost that utilizes gradient boosting trees.[5] We also implemented the collaborative filtering method that uses drug and cell line ID as input to compare our contents-based approach with collaborative approach. Benchmark accuracy values of various machine learning methods (including support vector machines, gradient boosting machines, and random forest) are listed in the referred paper.[4] Note that each training or testing set was a triplet $\{d_i,\ d_j, c_k\}$ of the drug pair and cell line in corresponding tissue. We did not report sensitivity prediction accuracy because the train/test split was disjointed in terms of synergy triplets but sometimes overlapped in terms of sensitivity pairs. For example, let us say we split the data as $\{d_1,\ d_2, c_1\}$ in the training set and $\{d_1,\ d_3, c_1\}$ in the testing set for synergy prediction. This split implies $\{d_1, c_1\}$, $\{d_2, c_1\}$ in training and $\{d_1, c_1\}$, $\{d_3, c_1\}$ in the testing set for sensitivity prediction. Due to the overlapped pairs $\{d_1, c_1\}$, the sensitivity prediction accuracy can be too optimistic.

RESULTS

Prediction accuracy of general model

We evaluated the accuracy of our synergy prediction models. The first experiment was to evaluate the accuracy of general models trained and tested with data-rich tissues (Table 2). We achieved {0.9577 AUROC, 0.8335 AUPRC, 132.4 MSE} for cross-validation (randomly split from 15 pooled databases) and {0.6376 AUROC, 0.5564 AUPRC, 742.1 MSE} for external validation (ie, trained with 14 pooled databases, tested with O'Neil). The collaborative filtering method based on drugs and cell IDs showed lower accuracy compared to our multimodal contents-based model. The lower accuracy in external validation compared to cross-validation might be partly due to different experimental settings on measuring the drug response scores and gene expression values across various synergy databases.

We trained and tested our model with O'Neil database in particular to compare our model's accuracy with DeepSynergy.[4] Our model achieved significantly higher AUROC/AUPRC and lower MSE. DeepSynergy showed {0.90 AUROC, 0.59 AUPRC, 255.5 MSE}, whereas our proposed model showed {0.95 AUROC, 0.94 AUPRC, 181.7 MSE}. In addition, we also trained our model with the ALMANAC database to compare our accuracy with that of NCI-ALMANAC + XGBoost model.[5] Our model achieved {0.97 AUROC, 0.84 AUPRC, 75.4 MSE}, whereas the baseline XGBoost model achieved 77.8 MSE. The increased accuracy was mainly due to more diverse multimodal features (such as target genes and native SMILES) and multitask learning with the monotherapy sensitivity.

Prediction accuracy of specific model with transfer learning

In the subsequent experiments, we shifted our focus to understudied tissues. These experiments compared the 4 different strategies on transferring model parameters from data-rich tissues to bone and prostate, respectively. We also compared accuracy with and without the transferred model parameters. We found that an adequate transfer learning increases accuracy in both bone and prostate (Table 3). In external validation of bone cell lines, we achieved {0.6647 AUROC, 0.4150 AUPRC, 297.4 MSE} without transfer learning, and {0.8015 AUROC, 0.4676 AUPRC, 195.9 MSE} with transfer learning (transferred all parameters and no fine-tuned). In the external validation of prostate cell lines, we achieved {0.6505 AUROC, 0.5220 APPRC, 528.5 MSE} without transfer learning, and {0.8542 AUROC, 0.7804 AUPRC, 327.9 MSE} with transfer learning (fixed drug and cell line encoders and retrained last merging layers).

Ranking

We listed predicted drug combinations for bone and prostate using the externally validated model. We selected the top 20 combinations based on estimated probability of being synergistic (Table 4, S3). For the bone cell lines, among the top 20, 5 combinations were actually synergistic ($5/20 = 25\%$ hit) in the independent validation set (ES Naampt_PARP). Considering the fact that only 1.56% of combinations are synergistic in bone, our model effectively ranked synergistic drug combinations. For the prostate cell lines, all top 20 combinations were actually synergistic in the independent O'Neil database (only 2.15% of combinations are synergistic in the prostate).

DISCUSSION

The objective of this study was to develop the drug combination synergy prediction model, which can be used even in understudied tissue with less observation. To meet this end, we i) collected a comprehensive set of multimodal features from multiple databases, ii) integrated the multimodal features and multiple tasks using deep neural networks, iii) transferred the prediction model from data-rich tissues to data-poor ones. As a result, the proposed model predicted drug synergy accurately with 0.8015 AUROC for bone and 0.8542 AUROC for prostate in external validation with an adequate transfer learning strategy.

Table 3. Drug synergy prediction accuracy for bone and prostate cancer (data-poor tissue) with different transfer learning strategies in cross or external validation

Tissue	Train	Test	Transfer	AUROC	AUPRC	MSE
Bone	DIPG, ES (FAKi/AURKi), ES (Naampt_PARP)	DIPG, ES (FAKi/AURKi), ES (Naampt_PARP)	No transfer	0.8283	0.5922	61.3
			Transfer and retrain all parameters	0.8357	0.5904	48.4
	DIPG, ES (FAKi/AURKi)	ES (Naampt_PARP)	No transfer	0.6647	0.4150	297.4
			Transfer and no retrain	0.8015	0.4676	195.9
Prostate	ALMANAC, ONEIL	ALMANAC, ONEIL	No transfer	0.9775	0.8575	78.5
			Transfer and retrain all parameters	0.9928	0.9628	47.4
	ALMANAC	ONEIL	No transfer	0.6505	0.5220	528.5
			Fix encoders and retrain last merging layers	0.8542	0.7804	327.9

Table 4. Top 5 drug combinations in bone and prostate

Drug1	Drug2	Cell line	Hit?
Bone			
AZD1775	AZACITIDINE	TC-71	
ZINC34894448	ZINC34894448	TC-71	✓
ZINC34894448	ZINC34894448	EW-8	
AZD1775	AZACITIDINE	EW-8	
thapsigargin	thapsigargin	TC-71	✓
Prostate			
TOPOTECAN	BEZ-235	LNCAP	✓
GELDANAMYCIN	BEZ-235	LNCAP	✓
DOXORUBICIN	BEZ-235	LNCAP	✓
BORTEZOMIB	BEZ-235	LNCAP	✓
5-FU	BEZ-235	LNCAP	✓

The bone model was trained with DIPG and ES (FAKi/AURKi), and tested with ES (Naampt_PARP). The prostate model was trained with ALMANAC and tested with O'Neil. "Hit" was marked if the predicted drug combinations show Loewe synergy score > 30. Full list of top 20 drug combinations in Supplementary Table S4.

The main contribution of our study is that we tackled understudied but critical tissues for drug synergy prediction. The difficulty of obtaining cell lines in these understudied tissues has limited the in-vitro experiments, which consequently become an obstacle to drug development in these tissues. As we focused on understudied tissues, we avoided using post-treatment gene expression profiles as an input feature. The posttreatment gene expression is the most powerful feature to estimate drug response, but it is only accessible after drug compounds are tested in cell line.[7] Our drug response prediction model was able to successfully bypass the posttreatment features while achieving competitive accuracy.

Although our study's focus is on predicting synergy in understudied tissues, our model achieved better accuracy even on general data-rich tissues than that of the baseline model. This increased accuracy is possibly due to the large-scale and multimodal data with multitask optimization. We used the 15 complementary drug synergy databases including the NCI-ALMANAC, Cloud, O'Neil, and NCATS Matrix. We incorporated diverse and comprehensive multimodal sources of features. This large-scale data collection allowed us to maximize the power of data-driven computational models based on deep neural networks.

Our model was designed to predict combination synergy together with monotherapy sensitivity simultaneously. Here monotherapy sensitivity was an auxiliary output to boost accuracy of synergy prediction. A previous study uses monotherapy sensitivity as an input feature to predict synergy,[7] but understudied tissues do not have enough experimental observation including monotherapy response features. To utilize this partial input feature in data-rich tissues (not in data-poor tissues), we used the monotherapy sensitivity as an auxiliary output in data-rich tissue models so that we can still transfer the information to data-poor tissues via the model parameters when minor tissues do not have monotherapy sensitivity.

The major limitation of this study is that the general prediction model was not able to achieve high accuracy in external validation. A possible explanation for this might be that the various synergy studies and various tissues may have different experimental settings on measuring the drug responses and gene expressions values, thus validation on the mixed tissues in the independent synergy experiment could cause the discrepancy in the predicted and observed values.

CONCLUSION

In conclusion, our model is an end-to-end drug synergy prediction model that learns interaction between drugs and cell lines. Based on the fact that different tissues share common gene expression and therefore respond to drugs in similar ways, we used transfer learning from data-rich tissues to data-poor tissues to make the synergy prediction model work in data-poor tissues. For future work, our drug prediction model for understudied tissues can potentially shed light onto other diseases that share drug targets and underlying mechanisms and offer a novel way of efficient and low-cost drug discovery.

FUNDING

This work was supported by CPRIT RP180012, NIH R01GM124111, U01TR002062, R01AG066749, Christopher Sarofim Family Professorship, UT Stars award, UTHealth startup award, NIH R01GM114612, CPRIT RP170668, the European Research Council (ERC) DrugComb (No. 716063), and Academy of Finland No. 317680.

AUTHOR CONTRIBUTIONS

XJ, WZ, and JT conceptualized the study; SZ, ZL prepared the necessary databases; JT, SZ, and WZ provided important insight for model development; XJ and YK designed the prediction model; YK, SZ implemented models and prepared the manuscript; SZ, JT, WZ, and XJ provided proof reading.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

CONFLICT OF INTEREST STATEMENT

None declared.

REFERENCES

1. Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov* 2004; 3 (8): 711–5.
2. Chou T-C. Theoretical basis, experimental design, and computerized simulation of synergism and antagonism in drug combination studies. *Pharmacol Rev* 2006; 58 (3): 621–81.
3. O'Neil J, Benita Y, Feldman I, *et al.* An unbiased oncology compound screen to identify novel combination strategies. *Mol Cancer Ther* 2016; 15 (6): 1155–62.
4. Preuer K, Lewis RPI, Hochreiter S, Bender A, Bulusu KC, Klambauer G. DeepSynergy: predicting anti-cancer drug synergy with deep learning. *Bioinformatics* 2018; 34 (9): 1538–46.
5. Sidorov P, Naulaerts S, Ariey-Bonnet J, Pasquier E, Ballester PJ. Predicting synergism of cancer drug combinations using NCI-ALMANAC data. *Front Chem* 2019; 7: 509.
6. Zhang C, Yan G. Synergistic drug combinations prediction by integrating pharmacological data. *Synth Syst Biotechnol* 2019; 4 (1): 67–72.
7. Celebi R, Bear Don't Walk O, Movva R, Alpsoy S, Dumontier M. In-silico prediction of synergistic anti-cancer drug combinations using multi-omics data. *Sci Rep* 2019; 9 (1): 8949.
8. Cuvitoglu A, Zhou JX, Huang S, Isik Z. Predicting drug synergy for precision medicine using network biology and machine learning. *J Bioinform Comput Biol* 2019; 17 (02): 1950012.
9. Li X, Qin G, Yang Q, Chen L, Xie L. Biomolecular network-based synergistic drug combination discovery. *Biomed Res Int* 2016; 2016: 1–11.
10. Sun X, Bao J, You Z, Chen X, Cui J. Modeling of signaling crosstalk-mediated drug resistance and its implications on drug combination. *Oncotarget* 2016; 7 (39): 63995–4006.
11. Feala JD, Cortes J, Duxbury PM, Piermarocchi C, McCulloch AD, Paternostro G. Systems approaches and algorithms for discovery of combinatorial therapies. *Wires Syst Biol Med* 2010; 2 (2): 181–93.
12. Goswami CP, Pankaj Goswami C, Cheng L, Alexander PS, Singal A, Li L. A new drug combinatory effect prediction algorithm on the cancer cell based on gene expression and dose-response curve. *CPT Pharmacometrics Syst Pharmacol* 2015; 4 (2): 80–90.
13. Yang J, Tang H, Li Y, *et al.* DIGRE: drug-induced genomic residual effect model for successful prediction of multidrug effects. *CPT Pharmacometrics Syst Pharmacol* 2015; 4 (2): e1.
14. Li P, Huang C, Fu Y, *et al.* Large-scale exploration and analysis of drug combinations. *Bioinformatics* 2015; 31 (12): 2007–16.
15. Wildenhain J, Spitzer M, Dolma S, *et al.* Prediction of synergism from chemical-genetic interactions by machine learning. *Cell Syst* 2015; 1 (6): 383–95.
16. Chiu Y-C, Chen H-IH, Zhang T, *et al.* Predicting drug response of tumors from integrated genomic profiles by deep neural networks. *BMC Med Genomics* 2019; 12 (S1): 18.
17. Bansal M, Yang J, Karan C, *et al.* A community computational challenge to predict the activity of pairs of compounds. *Nat Biotechnol* 2014; 32 (12): 1213–22.
18. Menden MP, Wang D, Mason MJ, *et al.* Community assessment to advance computational prediction of cancer drug combinations in a pharmacogenomic screen. *Nat Commun* 2019; 10 (1): 1–17.
19. Zagidullin B, Aldahdooh J, Zheng S, *et al.* DrugComb: an integrative cancer drug combination data portal. *Nucleic Acids Res* 2019; 47 (W1): W43–51.
20. Liu H, Zhang W, Zou B, Wang J, Deng Y, Deng L. DrugCombDB: a comprehensive database of drug combinations toward the discovery of combinatorial therapy. *Nucleic Acids Res* 2020; 48 (D1): D871–81.
21. Sun X, Vilar S, Tatonetti NP. High-throughput methods for combinatorial drug discovery. *Sci Transl Med* 2013; 5 (205): 205rv1.
22. Huang L, Li F, Sheng J, *et al.* DrugComboRanker: drug combination discovery based on target network analysis. *Bioinformatics* 2014; 30 (12): i228–36.
23. Zhao X-M, Iskar M, Zeller G, Kuhn M, van Noort V, Bork P. Prediction of drug combinations by integrating molecular and pharmacological data. *PLoS Comput Biol* 2011; 7 (12): e1002323.
24. Chen G, Tsoi A, Xu H, Jim Zheng W. Predict effective drug combination by deep belief network and ontology fingerprints. *J Biomed Inform* 2018; 85: 149–54.
25. Tang J, Gautam P, Gupta A, *et al.* Network pharmacology modeling identifies synergistic Aurora B and ZAK interaction in triple-negative breast cancer. *NPJ Syst Biol Appl* 2019; 5 (1): 20.
26. Cortini M, Baldini N, Avnet S. New advances in the study of bone tumors: a lesson from the 3D environment. *Front Physiol* 2019; 10: 814.
27. Xu C, Ai D, Shi D, *et al.* Accurate drug repositioning through non-tissue-specific core signatures from cancer transcriptomes. *Cell Rep* 2019; 29 (4): 1055.
28. Jaeger S, Duran-Frigola M, Aloy P. Drug sensitivity in cancer cell lines is not tissue-specific. *Mol Cancer* 2015; 14 (1): 40–4.
29. Yao F, Madani Tonekaboni SA, Safikhani Z, *et al.* Tissue specificity of in vitro drug sensitivity. *J Am Med Inform Assoc* 2018; 25 (2): 158–66.
30. Dhruba SR, Rahman R, Matlock K, Ghosh S, Pal R. Application of transfer learning for cancer drug sensitivity prediction. *BMC Bioinformatics* 2018; 19 (S17): 497.
31. Turki T, Wei Z, Wang JTL. A transfer learning approach via procrustes analysis and mean shift for cancer drug sensitivity prediction. *J Bioinform Comput Biol* 2018; 16 (03): 1840014.
32. Licciardello MP, Ringler A, Markt P, *et al.* A combinatorial screen of the CLOUD uncovers a synergy targeting the androgen receptor. *Nat Chem Biol* 2017; 13 (7): 771–8.
33. Holbeck SL, Camalier R, Crowell JA, *et al.* The National Cancer Institute ALMANAC: a comprehensive screening resource for the detection of anticancer drug pairs with enhanced therapeutic activity. *Cancer Res* 2017; 77 (13): 3564–76.
34. Forcina GC, Conlon M, Wells A, Cao JY, Dixon SJ. Systematic quantification of population cell death kinetics in mammalian cells. *Cell Syst* 2017; 4 (6): 600–10.e6.
35. Mathews Griner LA, Guha R, Shinn P, *et al.* High-throughput combinatorial screening identifies drugs that cooperate with ibrutinib to kill activated B-cell-like diffuse large B-cell lymphoma cells. *Proc Natl Acad Sci USA* 2014; 111 (6): 2349–54.
36. Wilson KM, Mathews-Griner LA, Williamson T, *et al.* Mutation profiles in glioblastoma 3D oncospheres modulate drug efficacy. *SLAS Technol* 2019; 24 (1): 28–40.
37. Yohe ME, Gryder BE, Shern JF, *et al.* MEK inhibition induces MYOG and remodels super-enhancers in RAS-driven rhabdomyosarcoma. *Sci Transl Med* 2018; 10 (448): eaan4470.
38. Phelan JD, Young RM, Webster DE, *et al.* A multiprotein supercomplex controlling oncogenic signalling in lymphoma. *Nature* 2018; 560 (7718): 387–91.
39. Greco WR, Bravo G, Parsons JC. The search for synergy: a critical review from a response surface perspective. *Pharmacol Rev* 1995; 47 (2): 331–85.
40. Ianevski A, He L, Aittokallio T, Tang J. SynergyFinder: a web application for analyzing drug combination dose-response matrix data. *Bioinformatics* 2017; 33 (15): 2413–5.
41. Yadav B, Wennerberg K, Aittokallio T, Tang J. Searching for drug synergy in complex dose-response landscapes using an interaction potency model. *Comput Struct Biotechnol J* 2015; 13: 504–13.
42. Polton DJ. Installation and operational experiences with MACCS (Molecular Access System). *Online Review* 1982; 6 (3): 235–42.
43. Segler MHS, Kogej T, Tyrchan C, Waller MP. Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS Cent Sci* 2018; 4 (1): 120–31.
44. Hirohara M, Saito Y, Koda Y, Sato K, Sakakibara Y. Convolutional neural network based on SMILES representation of com-

pounds for detecting chemical motif. *BMC Bioinformatics* 2018; 19 (S19): 526.

45. Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need.In: Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017; December 4–9, 2017: 5998–6008; Long Beach, CA.
46. Wishart DS, Knox C, Guo AC, *et al.* DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res* 2008; 36 (suppl 1): D901–6.
47. Chen X, Ji ZL, Chen YZ. TTD: therapeutic target database. *Nucleic Acids Res* 2002; 30 (1): 412–5.
48. HMS LINCS (Harvard Medical School). LINCS MCF 10A Common Project: rolling-time-point sensitivity measures of the MCF 10A breast cell line to 8 small molecule perturbagens. Dataset 8 of 15: end-point dose-response metrics for biological replicate 2. http://lincsportal.ccs.miami.edu/datasets/#/view/LDS-1415 Accessed December 2019.
49. Barretina J, Caponigro G, Stransky N, *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 2012; 483 (7391): 603–7.
50. Tate JG, Bamford S, Jubb HC, *et al.* COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res* 2019; 47 (D1): D941–7.
51. Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol P-A. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. *J Mach Learn Res* 2010; 11 (110): 3371–408.
52. Malyutina A, Majumder MM, Wang W, Pessia A, Heckman CA, Tang J. Drug combination sensitivity scoring facilitates the discovery of synergistic and efficacious drug combinations in cancer. *PLoS Comput Biol* 2019; 15 (5): e1006752.