Predicting drug response via transfer learning

Michela Meister
Cornell University, Ithaca, NY

Christopher Tosh
Memorial Sloan Kettering Cancer Center, New York, NY

Wesley Tansey
Memorial Sloan Kettering Cancer Center, New York, NY

Abstract

Functional precision oncology identifies promising treatments by conducting drug screens on patient-derived tumor models, such as cell lines, organoids, and mouse models. As the fidelity, cost, and duration of these experiments depends on the chosen model type, it is of great interest to effectively utilize cheaper, faster, and lower-quality experiments to inform potential outcomes on costlier, slower, and higher-quality experiments. In this work, we propose Bayesian Modality Transfer (BMT), a Bayesian transfer learning approach that leverages data from multiple modalities to make predictions on unseen experiments. In simulation studies, we demonstrate that BMT effectively transfers knowledge between three cell line datasets conducted under different experimental conditions and learns to map between in vitro and in vivo modalities. In a regime where data is scarce, we find that the traditional approach of conducting experiments independently would require an average of 1.55- to 4.27-times as many experiments as BMT.

Keywords: precision cancer medicine, transfer learning

1. Introduction

Drug screening has become a key component of precision cancer medicine (Corsello et al., 2020; Basu et al., 2013; Garnett et al., 2012; Lee et al., 2018; Gao et al., 2015). Identifying promising drugs requires running a progression of experiments which increase in both fidelity and duration. For example, candidate drugs may be first tested against 2D cell lines, then 3D patient-derived organoids, and finally in patient-derived xenografts or genetically engineered mouse models. A crucial question is how to select experiments to run in later stages, as inefficiency can be cost and time prohibitive. Our goal is to build a machine learning model that enables integration of previous experiments in the same or earlier stages and predicts drug response in future stages. A challenge is that drug response is not always consistent across stages: a drug which succeeds in an earlier stage may fail in a later stage, and vice versa. Even within the same stage, results can vary substantially between labs due to change in experimental conditions and natural variation (Errington et al., 2021).

In this paper, we apply a transfer learning approach: we use a *source dataset* S from an earlier study or stage and a *target dataset* T from the current study. The goal is to combine these two datasets to make predictions on unseen experiments in the later stage. Transfer learning has been successfully applied in a variety of contexts, including RNA feature-based neural networks to transfer predictions from cell lines to PDX models (Ma et al., 2021). In this work, we extend the reach of transfer learning to drug screening settings that are lacking in genomic information.

We propose *Bayesian Modality Transfer* (BMT), a lightweight, multi-view transfer model, which is agnostic to any biological, chemical, or genomic properties of the data. We benchmark BMT on a panel of cell line studies conducted in three different experimental settings. We find that BMT significantly improves prediction on all three cell line datasets which we study. In a regime where data is scarce, we find that the traditional approach of conducting experiments independently would require an average of 1.55- to 4.27-times as many experiments as BMT.

We then apply BMT to a study on 12 matched patient-derived organoids (PDOs) and xenografts (PDXs) where the raw data show no correlation on average between the two modalities, suggesting a

need to conduct expensive *in vivo* PDX experiments for every organoid. Even though the study involves only 108 experiments for each modality, BMT learns an effective transfer mapping, achieving a substantial improvement in Pearson correlation from the baseline average of 0.05 to an average of 0.22 and as high as 0.66.

Taken together with the efficiency gains BMT achieves on cell line drug response studies, these results suggest transfer learning has the potential to substantially reduce the cost and time burden of the precision medicine pipeline.

Further related work. Tang et al. (2022) use a transfer learning model incorporating molecular and genomic information to predict PDX response from cell line data. Mourragui et al. (2021) build a transfer learning system using cell line and PDX drug response data to predict patient response. Many models use machine learning to predict cell line drug response, (Liu et al., 2020; Chiu et al., 2019; Yuan et al., 2016; Adam et al., 2020). Pan and Yang (2009) and Weiss et al. (2016) survey transfer learning.

2. Model

Suppose there are patient samples $i = 1, \ldots, n$ and drugs $j = 1, \ldots, m$. Let S be an $n \times m$ matrix where S_{ij} is the response of sample i to drug j in the source dataset. Let T be an $n \times m$ matrix where T_{ij} is the response of sample i to drug j the target dataset. A sample i may be tested against only a subset of drugs, so S and T may be sparse.

To model this data, we introduce a Bayesian transfer-learning normal factorization model. Fix $k \in \mathbb{N}$. Each sample i is governed by a k-dimensional latent vector c_i, and each drug j is governed by a k-dimensional latent vector d_j. We model the response of sample i to drug j in the source dataset as

$$S_{ij} \sim \mathcal{N}(c_i^T d_j + \alpha, \sigma_s^2),$$

where α and σ_s are constant across all samples i and drugs j.

Recall that our goal is to model a mapping from the source dataset to the target dataset. We encode this mapping as a $k \times k$ latent matrix W. We model the response of sample i to drug j in the target dataset as

$$T_{ij} \sim \mathcal{N}((Wc_i)^T d_j + \alpha', \sigma_t^2),$$

where α' and σ_t are constant across all samples i and drugs j. We define W as the product of two $k \times k$ matrices U and V, indexed by r and l. We index embedding vectors c_i and d_j by l and place hierarchical Gaussian priors on the embeddings and fixed effects.

$$\alpha \sim \mathcal{N}(0, \sigma_\alpha^2), \; \alpha' \sim \mathcal{N}(0, \sigma_{\alpha'}^2), \; d_{jl} \sim \mathcal{N}(0, \sigma_d^2),$$

$$c_{il} \sim \mathcal{N}(0, \sigma_c^2), \; U_{rl} \sim \mathcal{N}(0, \sigma_w^2), \; V_{rl} \sim \mathcal{N}(0, \sigma_w^2),$$

We specify Gamma$(0.1, 0.1)$ priors for all σ variance parameters. The model is fit with black box variational inference in the `Pyro` python package. The embedding dimension k is chosen via cross-validation.

3. Simulation studies

For each experiment, we choose a source dataset S and a target dataset T. We hold out $H \subset T$. The training set comprises S and $T \setminus H$. We measure performance by computing the sample Pearson correlation coefficient of the maximum a posteriori (MAP) estimate and the held out set H.

3.1. Experiment 1: Predicting in-vitro response

We benchmark our model against three pan-cancer, cell line datasets, GDSC (Garnett et al., 2012), PRISM (Corsello et al., 2020), and CTD2 (Basu et al., 2013), which share 318 cell lines, 84 drugs, and 16,588 experiments. Each study tests cell lines against drugs at a range of doses and provides the percent-viablity curve (AUC). Dose ranges vary across the three datasets, so Corsello et al. (2020) compute the area under the percent-viability curve restricted to a single overlapping dose range ($\text{AUC}_\cap$). We define S_{ij} to be the log(AUC) for cell line i against drug j in the source dataset, and T_{ij} to be the log(AUC) of cell line i against drug j in the target dataset. We averged AUC values for experiments with multiple entries.

We begin by choosing a source dataset S and a target dataset T from PRISM, GDSC, and CTD2. We run the following experiment on each of the $\binom{3}{2} = 6$ possible choices.

Experimental Setup. We partition the cell lines into 10 sets $L = 1, \ldots, 10$. Using these sets, we partition T into 10 cross-validation folds $F_1, \ldots, F_{10}$, such that for each set L, we define fold $F_L = \{(i, j, T_{ij}) : i \in L\}$, which consists of all data involving cell lines in L. We choose k via 5-fold cross-validation. The final model is chosen by optimizing over 5 random restarts with the target training loss as the objective.

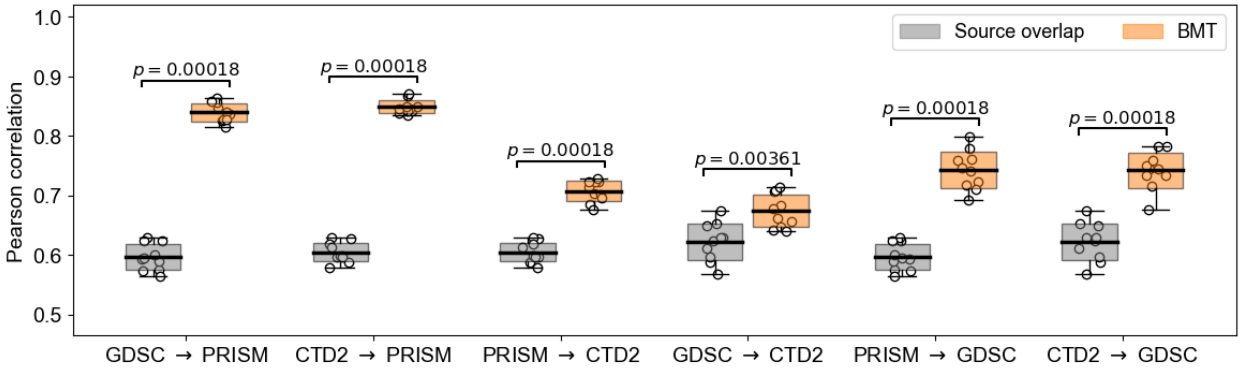

Figure 1: BMT incorporates data from a source dataset and a target dataset to significantly increase the accuracy of predictions in the target setting. $S \to T$ indicates S is the source dataset and T is the target dataset.

Baseline. We compute the Pearson correlation coefficient of the $\text{AUC}_\cap$ values in H and their corresponding $\text{AUC}_\cap$ values in the source dataset, a comparison used by Corsello et al. (2020).

Evaluation and results. As shown in Figure 1, BMT significantly increases the Pearson correlation coefficient for each combination of datasets (Mann-Whitney U-test; $p = 0.00361$ for (GDSC, CTD2) and $p = 0.00018$ for all other pairs). Relative to the baseline, BMT enjoys an average improvement in accuracy ranging from 8% to 41%.

3.2. Experiment 2: Comparing BMT to a model given access to only the target

We compare BMT to a model given access to only the target dataset, which we call the *target-only* model.

Defining the target-only model. Fix $k \in \mathbb{N}$. Each sample i is governed by a k-dimensional vector $\tilde{c}_i$ and each drug is governed by a k-dimensional vector $\tilde{d}_j$. We model the response of sample i to drug j in the target dataset as

$$T_{ij} \sim \mathcal{N}(\tilde{c}_i{}^T \tilde{d}_j + \tilde{\alpha}, \tilde{\sigma}^2),$$

where $\tilde{\alpha}$ and $\tilde{\sigma}$ are constant over all samples i and drugs j. The target-only model is fit via the same process as BMT. The full generative model is found in Appendix A.

Experimental set-up. We begin by choosing a source dataset S and a target dataset T from PRISM, GDSC, and CTD2. We select a γ-fraction of experiments uniformly at random to comprise the target training set (i.e. $T \setminus H$). Following the same procedure as in Experiment 1, BMT is fit on S and $T \setminus H$ and the target-only model is fit on $T \setminus H$. We run this process 10 times for each $\gamma \in \{5\%, 10\%, 15\%, 20\%\}$.

Evaluation and results. For a holdout set H, for a given γ, we interpolate the fraction of target data γ' required by target-only to match the accuracy of BMT. We call γ'/γ the *efficiency gain*.

Figure 2 reports efficiency gains in the small-data regime. For $\gamma = 5\%$, relative to the target-only model, BMT achieves mean efficiency gains ranging from 1.89 to 4.27. Given that there are 16,588 experiments total, this implies that the target-only model requires approximately 738 to 2,712 *additional* experiments to achieve similar respective accuracy. Efficiency gains decrease as γ increases, and for $\gamma = 20\%$ BMT achieves mean efficiency gains ranging from 1.55 to 2.02. Therefore for $\gamma = 20\%$ the target-only model requires an additional 1,825 to 3,384 experiments to achieve similar respective accuracy.

3.3. Experiment 3: Predicting in-vivo response

Schütte et al. (2017) study colorectal cancer PDOs and PDXs. They tested PDOs against drugs at concentrations ranging from 3.05nM to 60μM and tested PDXs in mouse models for 4 weeks. We study a subset of this dataset that includes 12 patient samples

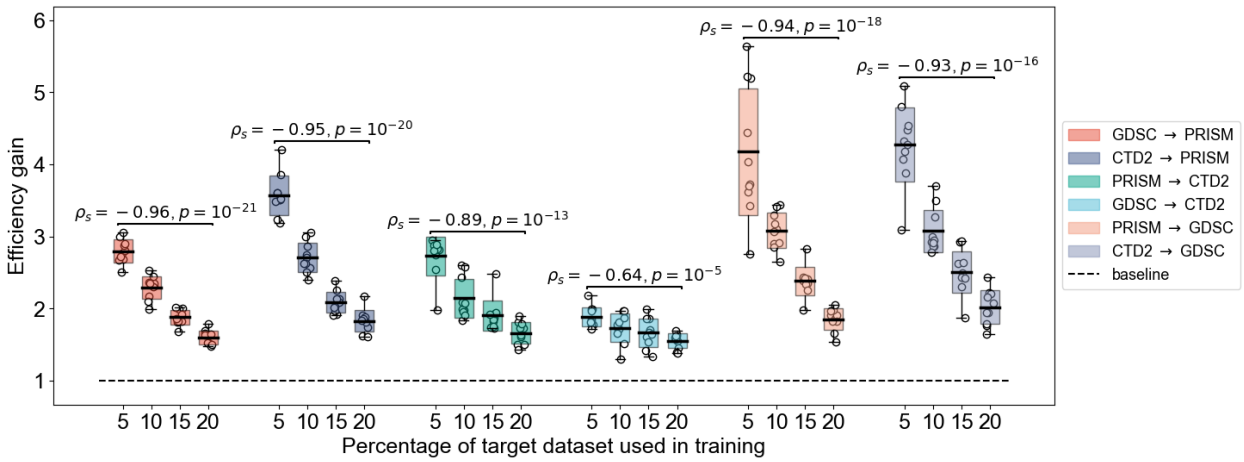

Figure 2: In the small-data regime, the traditional approach of conducting experiments independently would require 1.55- to 4.27-times as many experiments as BMT on average. $S \to T$ indicates S is the source dataset and T is the target dataset. ρ_s is Spearman's correlation coefficient.

each tested against the same 9 drugs, resulting in 108 observations in each of the PDO and PDX settings. The study provided IC50 measurements for PDO samples, so we computed S_{ij} as the $\log_{10}(IC50)$. The study provided tumor volumes for both control and treated PDX samples, so we calculated T_{ij} as the ratio of the treatment volume and the control volume at 4 weeks.

Experimental set-up. We split the data into 12 cross-validation folds such that each fold contains all data points corresponding to a single sample. The remainder of the experiment follows the same procedure as Experiment 1, except that k is chosen from $\{1, 2, \ldots, 9\}$. We compare BMT predictions to the Pearson correlation coefficient between H and the raw source data.

Evaluation and results. Figure 3 reports our results. A Mann-Whitney U-test shows that BMT increases the Pearson correlation coefficient ($p = .026$) from 0.05 to 0.22. Across the entire dataset, 83% (10 / 12) of PDXs saw an improvement from using the BMT transfer predictions instead of the raw PDO results.

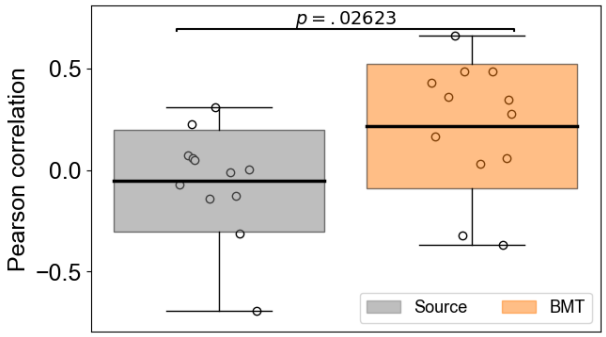

Figure 3: BMT learns to map between PDO and PDX modalities on a study of 12 matched PDOs and PDXs with 108 experiments in total for each modality. Relative to the baseline mean of 0.05, BMT increases the mean Pearson correlation coefficient to 0.22.

References

George Adam, Ladislav Rampášek, Zhaleh Safikhani, Petr Smirnov, Benjamin Haibe-Kains, and Anna Goldenberg. Machine learning approaches to drug response prediction: challenges and recent progress. *NPJ precision oncology*, 4(1):19, 2020.

Amrita Basu, Nicole E Bodycombe, Jaime H Cheah, Edmund V Price, Ke Liu, Giannina I Schaefer, Richard Y Ebright, Michelle L Stewart, Daisuke Ito, Stephanie Wang, et al. An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell*, 154(5):1151–1161, 2013.

Yu-Chiao Chiu, Hung-I Harry Chen, Tinghe Zhang, Songyao Zhang, Aparna Gorthi, Li-Ju Wang, Yufei Huang, and Yidong Chen. Predicting drug response of tumors from integrated genomic profiles by deep neural networks. *BMC medical genomics*, 12(1):143–155, 2019.

Steven M Corsello, Rohith T Nagari, Ryan D Spangler, Jordan Rossen, Mustafa Kocak, Jordan G Bryan, Ranad Humeidi, David Peck, Xiaoyun Wu, Andrew A Tang, et al. Discovering the anticancer potential of non-oncology drugs by systematic viability profiling. *Nature cancer*, 1(2):235–248, 2020.

Timothy M Errington, Maya Mathur, Courtney K Soderberg, Alexandria Denis, Nicole Perfito, Elizabeth Iorns, and Brian A Nosek. Investigating the replicability of preclinical cancer biology. *Elife*, 10: e71601, 2021.

Hui Gao, Joshua M Korn, Stéphane Ferretti, John E Monahan, Youzhen Wang, Mallika Singh, Chao Zhang, Christian Schnell, Guizhi Yang, Yun Zhang, et al. High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nature medicine*, 21(11): 1318–1325, 2015.

Mathew J Garnett, Elena J Edelman, Sonja J Heidorn, Chris D Greenman, Anahita Dastur, King Wai Lau, Patricia Greninger, I Richard Thompson, Xi Luo, Jorge Soares, et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature*, 483(7391):570–575, 2012.

Jin-Ku Lee, Zhaoqi Liu, Jason K Sa, Sang Shin, Jiguang Wang, Mykola Bordyuh, Hee Jin Cho, Oliver Elliott, Timothy Chu, Seung Won Choi, et al. Pharmacogenomic landscape of patient-derived tumor cells informs precision oncology therapy. *Nature genetics*, 50(10):1399–1411, 2018.

Chuanying Liu, Dong Wei, Ju Xiang, Fuquan Ren, Li Huang, Jidong Lang, Geng Tian, Yushuang Li, and Jialiang Yang. An improved anticancer drug-response prediction based on an ensemble method integrating matrix completion and ridge regression. *Molecular Therapy-Nucleic Acids*, 21:676–686, 2020.

Jianzhu Ma, Samson H Fong, Yunan Luo, Christopher J Bakkenist, John Paul Shen, Soufiane Mourragui, Lodewyk FA Wessels, Marc Hafner, Roded Sharan, Jian Peng, et al. Few-shot learning creates predictive models of drug response that translate from high-throughput screens to individual patients. *Nature Cancer*, 2(2):233–244, 2021.

Soufiane MC Mourragui, Marco Loog, Daniel J Vis, Kat Moore, Anna G Manjon, Mark A van de Wiel, Marcel JT Reinders, and Lodewyk FA Wessels. Predicting patient response with models trained on cell lines and patient-derived xenografts by nonlinear transfer learning. *Proceedings of the National Academy of Sciences*, 118(49):e2106682118, 2021.

Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.

Moritz Schütte, Thomas Risch, Nilofar Abdavi-Azar, Karsten Boehnke, Dirk Schumacher, Marlen Keil, Reha Yildiriman, Christine Jandrasits, Tatiana Borodina, Vyacheslav Amstislavskiy, et al. Molecular dissection of colorectal cancer in pre-clinical models identifies biomarkers predicting sensitivity to egfr inhibitors. *Nature communications*, 8(1): 14262, 2017.

Yi-Ching Tang, Reid T Powell, and Assaf Gottlieb. Molecular pathways enhance drug response prediction using transfer learning from cell lines to tumors and patient-derived xenografts. *Scientific Reports*, 12(1):16109, 2022.

Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. A survey of transfer learning. *Journal of Big data*, 3(1):1–40, 2016.

Han Yuan, Ivan Paskov, Hristo Paskov, Alvaro J González, and Christina S Leslie. Multitask learn-

ing improves prediction of cancer drug sensitivity. *Scientific reports*, 6(1):31619, 2016.

Appendix A. Supplement

Data Availability. All datasets studied are publicly available.

Defining the target-only model in full. Fix $k \in \mathbb{N}$. Each sample i is governed by a k-dimensional vector $\tilde{c_i}$ and each drug is governed by a k-dimensional vector $\tilde{d_j}$. We model the response of sample i to drug j in the target dataset as

$$T_{ij} \sim \mathcal{N}(\tilde{c_i}^T \tilde{d_j} + \tilde{\alpha}, \tilde{\sigma}^2),$$

where $\tilde{\alpha}$ and $\tilde{\sigma}$ are constant over all samples i and drugs j.

We index embedding vectors c_i and d_j by l and place hierarchical Gaussian priors on the embeddings and fixed effects.

$$\tilde{\alpha} \sim \mathcal{N}(0, \tilde{\sigma}_\alpha^2)\,,\ \tilde{d}_{jl} \sim \mathcal{N}(0, \tilde{\sigma}_d^2)\,, \tilde{c}_{il} \sim \mathcal{N}(0, \tilde{\sigma}_c^2)\,,$$

We specify Gamma$(0.1, 0.1)$ priors for all σ variance parameters. The model is fit with black box variational inference in the `Pyro` python package. Parameter k is chosen via cross-validation.