S1 ANNOTATED TARGET GENES FOR THE STUDIED DRUGS

Drug	Annotated target genes from (Smirnov et al., 2017)
Bortezomib	PSMB2, PSMA1, PSMA2, PSMA3, PSMA4, PSMA5, PSMA6, PSMA7, PSMA8, PSMB1, PSMB10, PSMB11, PSMB3, PSMB4, PSMB5, PSMB6, PSMB7, PSMB8, PSMB9, PSMD1, PSMD2, RELA
Cisplatin	XIAP
Docetaxel	BCL2, MAP2, MAP4, MAPT, NR1I2, TUBB, TUBB1
Paclitaxel	BCL2, MAP2, MAP4, MAPT, NR1I2, TLR4, TUBB, TUBB1

S2 SELECTED HYPER-PARAMETERS

Selected hyper-parameters for MOLI (Sharifi-Noghabi et al., 2019):

Drug	Selected hyper-parameters
Bortezomib	64 (number of nodes in the feature extractor), 1.5 (margin for the triplet loss), 0.0001 (encoder subnetwork learning rate), 40 (epochs), 0.7 and 0.3 (encoder and classifier dropout rates), 0.01 (weight decay), 0.5 (classifier learning rate), 0.2 (regularization coefficient), 36 (batch size).
Cisplatin	64 (number of nodes in the feature extractor), 0.5 (margin for the triplet loss), 0.005 (encoder subnetwork learning rate), 40 (epochs), 0.5 and 0.5 (encoder and classifier dropout rates), 0.001 (weight decay), 0.001 (classifier learning rate), 0.2 (regularization coefficient), 64 (batch size).
Docetaxel	128 (number of nodes in the feature extractor), 1 (margin for the triplet loss), 0.05 (encoder subnetwork learning rate), 25 (epochs), 0.6 and 0.5 (encoder and classifier dropout rates), 0.001 (weight decay), 0.001 (classifier learning rate), 0.1 (regularization coefficient), 36 (batch size).
Paclitaxel	64 (number of nodes in the feature extractor), 1 (margin for the triplet loss), 0.0001 (encoder subnetwork learning rate), 15 (epochs), 0.5 and 0.5 (encoder and classifier dropout rates), 0.0001 (weight decay), 0.001 (classifier learning rate), 0.3 (regularization coefficient), 14 (batch size).

Selected hyper-parameters for ADDA (Tzeng et al., 2017):

Drug	Selected hyper-parameters
Bortezomib	256 (number of nodes in the feature extractor trained on the source samples, feature extractor of the target samples, and also the input layer of the classifier trained on the source samples), 64 (number of nodes in the hidden layer of the discriminator), 0.01 (learning rate), 20 (epochs), 0.3 and 0.7 (dropout rates for target samples feature extractor and the discriminator, respectively), no weight decay, 16 and 16 (batch size for source and target domains, respectively).
Cisplatin	256 (number of nodes in the feature extractor trained on the source samples, feature extractor of the target samples, and also the input layer of the classifier trained on the source samples), 64 (number of nodes in the hidden layer of the discriminator), 0.005 (learning rate), 20 (epochs), 0.3 and 0.6 (dropout rates for target samples feature extractor and the discriminator, respectively), no weight decay, 8 and 16 (batch size for source and target domains, respectively).
Docetaxel	1024 (number of nodes in the feature extractor trained on the source samples, feature extractor of the target samples, and also the input layer of the classifier trained on the source samples), 512 (number of nodes in the hidden layer of the discriminator), 5e-5 (learning rate), 15 (epochs), 0.3 and 0.5 (dropout rates for target samples feature extractor and the discriminator, respectively), 0.005 (weight decay), 16 and 32 (batch size for source and target domains, respectively).
Paclitaxel	NA.

Selected hyper-parameters for ProtoNet (Snell et al., 2017):

Drug	Selected hyper-parameters
Bortezomib	16 (number of nodes in the feature extractor), 5e-5 and 0.5 (learning rates for training on source and target domains), 15 (number of epochs for the source and target domains), 0.7 (dropout rate for the source and target domains), 2 and 8 (number of support and query), 100 (number of episodes).
Cisplatin	256 (number of nodes in the feature extractor), 0.0005 and 0.5 (learning rates for training on source and target domains), 15 and 10 (number of epochs for the source and target domains), 0.3 and 0.4 (dropout rate for the source and target domains), 2 and 4 (number of support and query), 100 (number of episodes).
Docetaxel	16 (number of nodes in the feature extractor), 0.0005 and 0.1 (learning rates for training on source and target domains), 10 and 30 (number of epochs for the source and target domains), 0.3 and 0.6 (dropout rate for the source and target domains), 4 and 8 (number of support and query), 100 (number of episodes).
Paclitaxel	NA.

Selected hyper-parameters for AITL:

Drug	Selected hyper-parameters
Bortezomib	1024 (number of nodes in the layer of the feature extractor), 1024 (number of nodes in the shared layer of the multi-task subnetwork), 1024 (number of nodes in the hidden layer of the regression tower), 0.0005 (learning rate), 0.2 and 0.4 (regularization for global and class-wise discriminators), 16 and 16 (mini-batch size for the source and target domains), 0.4 (dropout rate), 10 (epoch).
Cisplatin	512 (number of nodes in the hidden layer of the feature extractor), 16 (number of nodes in the shared layer of the multi-task subnetwork), 16 (number of nodes in the hidden layer of the regression tower), 0.05 (learning rate), 0.3 and 0.3 (regularization for global and class-wise discriminators), 32 and 8 (mini-batch size for the source and target domains), 0.15 (dropout rate), 25 (epoch).
Docetaxel	256 (number of nodes in the hidden layer of the feature extractor), 512 (number of nodes in the shared layer of the multi-task subnetwork), 512 (number of nodes in the hidden layer of the regression tower), 0.0001 (learning rate), 0.8 and 0.6 (regularization for global and class-wise discriminators), 32 and 32 (mini-batch size for the source and target domains), 0.5 (dropout rate), 35 (epoch).
Paclitaxel	1024 number of nodes in the layer of the feature extractor), 1024 (number of nodes in the shared layer of the multi-task subnetwork), 1024 (number of nodes in the hidden layer of the regression tower), 0.0001 (learning rate), 0.9 and 0.3 (regularization for global and class-wise discriminators), 32 and 32 (mini-batch size for the source and target domains), 0.5 (dropout rate), 20 (epoch).

Selected hyper-parameters for PRECISE (Mourragui et al., 2019):

Since this method used the same cell line dataset (source domain), we adopted the recommended default settings of the original paper and only used 3-fold cross validation to tune the predictor hyper-parameter.

Selected hyper-parameters for (Chen et al., 2017):

Drug	Selected hyper-parameters
Bortezomib	128 (number of nodes in the feature extractor), 32 (number of nodes in the hidden layer of discriminators), 0.0001 (learning rate), 20 (epochs), 0.0001 (weight decay), 0.8, 0.3, 0.2, 0.6, 0.2 (dropout rates in featuer extractor, global discriminator, responder class discriminator, non-responder class discriminator, and classifier, respectively), 0.9 and 0.6 (regularization coefficients for class-wise and global discriminators, respectively), 16 and 64 (batch size for source and target domains, respectively).
Cisplatin	512 (number of nodes in the feature extractor), 128 (number of nodes in the hidden layer of discriminators), 0.0001 (learning rate), 15 (epochs), 0.0001 (weight decay), 0.3, 0.3, 0.5, 0.8, 0.5 (dropout rates in featuer extractor, global discriminator, responder class discriminator, non-responder class discriminator, and classifier, respectively), 0.4 and 0.7 (regularization coefficients for class-wise and global discriminators, respectively), 8 and 32 (batch size for source and target domains, respectively).
Docetaxel	128 (number of nodes in the feature extractor), 64 (number of nodes in the hidden layer of discriminators), 0.0005 (learning rate), 5 (epochs), 0.0001 (weight decay), 0.6, 0.4, 0.3, 0.7, 0.4 (dropout rates in featuer extractor, global discriminator, responder class discriminator, non-responder class discriminator, and classifier, respectively), 1 and 0.4 (regularization coefficients for class-wise and global discriminators, respectively), 8 and 32 (batch size for source and target domains, respectively).
Paclitaxel	512 (number of nodes in the feature extractor), 16 (number of nodes in the hidden layer of discriminators), 0.0005 (learning rate), 10 (epochs), 0.1 (weight decay), 0.6, 0.8, 0.8, 0.7, 0.3 (dropout rates in featuer extractor, global discriminator, responder class discriminator, non-responder class discriminator, and classifier, respectively), 1 and 0.8 (regularization coefficients for class-wise and global discriminators, respectively), 64 and 16 (batch size for source and target domains, respectively).

S3 STATISTICALLY SIGNIFICANT TARGET GENES IN DRUG RESPONSE PREDICTION FOR TCGA PATIENTS

S3.1 BREAST CANCER

Drug	Target gene (P-value)
Docetaxel	BCL2 ($P = 2.0 \times 10^{-}8$), MAP4 ($P < 1 \times 10^{-}10$), MAPT ($P < 1 \times 10^{-}10$).
Paclitaxel	BLC2 ($P = 1.7 \times 10^{-}4$), MAP2 ($P = 3.2 \times 10^{-}5$), MAP4 ($P < 1 \times 10^{-}10$), MAPT ($P = 3.7 \times 10^{-}9$), TLR4 ($P < 1 \times 10^{-}10$), TUBB ($P < 1 \times 10^{-}10$).
Bortezomib	PSMA1 ($P = 0.04$), PSMA4 ($P = 4.7 \times 10^{-}6$), PSMB1 ($P = 3.8 \times 10^{-}5$), PSMB4 ($P < 1 \times 10^{-}10$), PSMB6 ($P = 0.04$), PSMB8 ($P = 1.5 \times 10^{-}5$), PSMB9 ($P = 4.8 \times 10^{-}8$), PSMB10 ($P = 7.8 \times 10^{-}4$), PSMD1 ($P = 1.1 \times 10^{-}6$), RELA ($P < 1 \times 10^{-}10$).

S3.2 PROSTATE CANCER

Drug	Target gene (P-value)
Docetaxel	MAP2 ($P = 4.2 \times 10^{-}3$), MAP4 ($P < 1 \times 10^{-}10$).
Paclitaxel	BLC2 ($P < 1 \times 10^{-}10$), MAP2 ($P = 0.001$), MAP4 ($P < 1 \times 10^{-}10$), TLR4 ($P < 1 \times 10^{-}10$), TUBB ($P < 1 \times 10^{-}10$).
Bortezomib	PSMA1 ($P = 1.9 \times 10^{-}5$), PSMA3 ($P < 1 \times 10^{-}10$), PSMB2 ($P = 0.002$), PSMB4 ($P = 9.3 \times 10^{-}5$), PSMB7 ($P = 0.04$), PSMB8 ($P = 0.02$), PSMB10 ($P = 7.4 \times 10^{-}7$), PSMD1 ($P = 8.8 \times 10^{-}6$), RELA ($P = 2.2 \times 10^{-}4$).

S3.3 BLADDER CANCER

Drug	Target gene (P-value)
Docetaxel	MAP4 ($P < 1 \times 10^{-}10$).
Paclitaxel	BLC2 ($P = 2.4 \times 10^{-}5$), MAP4 ($P < 1 \times 10^{-}10$), TLR4 ($P = 7.3 \times 10^{-}4$), TUBB ($P < 1 \times 10^{-}10$).
Bortezomib	PSMA4 ($P = 0.001$), PSMB1 ($P = 0.04$), PSMB4 ($P = 4.7 \times 10^{-}6$), PSMB9 ($P = 2.2 \times 10^{-}6$), PSMB10 ($P = 1.3 \times 10^{-}9$), PSMD1 ($P = 0.006$), RELA ($P < 1 \times 10^{-}10$).

S3.4 KIDNEY CANCER

Drug	Target gene (P-value)
Docetaxel	BLC2 ($P = 1.4 \times 10^{-}5$), MAP4 ($P < 1 \times 10^{-}10$), MAPT ($P = 2.0 \times 10^{-}6$).
Paclitaxel	BLC2 ($P = 1.1 \times 10^{-}6$), MAP4 ($P < 1 \times 10^{-}10$), MAPT ($P < 1 \times 10^{-}10$),TLR4 ($P < 1 \times 10^{-}10$), TUBB ($P < 1 \times 10^{-}10$).
Bortezomib	PSMA2 ($P = 0.03$), PSMB4 ($P = 0.001$), PSMB9 ($P = 1 \times 10^{-}4$), PSMB10 ($P = 0.006$), PSMD2 ($P = 1 \times 10^{-}5$), RELA ($P = 5 \times 10^{-}5$).

S3.5 LUNG CANCER

Drug	Target gene (P-value)
Docetaxel	MAP4 ($P < 1 \times 10^{-}10$).
Paclitaxel	BLC2 ($P = 2.9 \times 10^{-}5$), MAP4 ($P < 1 \times 10^{-}10$), TLR4 ($P = 4.8 \times 10^{-}9$), TUBB ($P < 1 \times 10^{-}10$).
Bortezomib	PSMA1 ($P = 0.005$), PSMA4 (0.007), PSMB1 (0.006), PSMB8 ($P = 1.9 \times 10^{-}4$), PSMB9 ($P < 1 \times 10^{-}10$), PSMB10 ($P = 3.9 \times 10^{-}9$), RELA ($P < 1 \times 10^{-}10$).

S4 SIGNIFICANT GENES IN TCGA ANALYSIS

The obtained results are in concordance with previous studies. For example, we observed that Microtubule-Associated Proteins (MAPs) were significant for Docetaxel and Paclitaxel in the studied cancers which aligns with previous research on this family of proteins (Yang et al., 2017; Smoter et al., 2011; Bhat & Setaluri, 2007). For Bortezomib, we observed significant associations for different proteasome subunits such as subunit alpha (PSMA) and beta (PSMB). These subunits have been shown to be key players across different cancers (Rouette et al., 2016; Tsvetkov et al., 2017; Li et al., 2017). We also observed significant associations for RELA (also known as Transcription Factor p65) in all of the studied cancers which aligns with its oncogenic role across different cancers (Collignon et al., 2018), and moreover, with its reported associations with Bortezomib in breast cancer (Hideshima et al., 2014), prostate cancer (Manna et al., 2013), and lung cancer (Zhao et al., 2015).

S5 DATA PREPROCESSING STEPS

Microarray CEL files for clinical trial cohorts were downloaded from Gene Expression Omnibus (see identifiers listed in Table 1) and for GDSC cell lines from the ArrayExpress database (E-MTAB-3610). All raw CEL files were subjected to robust multi-array average normalization (Irizarry et al., 2003) using the justRMA() function from the affy (v1.54.0) R package and CDF library files and probe set annotations provided by BrainArray v22.0.0 (http://brainarray.mbni.med.umich.edu) (Dai et al., 2005). Probe sets ambiguously mapped to Entrez genes were excluded. Expressions of the remaining probe sets were aggregated to gene level using collapseRows() function (Miller et al., 2011) from WGCNA (v 1.64.1) R package with method="Average".
TCGA expression data were obtained from Firehose (http://gdac.broadinstitute.org/), the version published on 28.01.2016. PDX expressions were taken from supplementary data of(Gao et al., 2015). TCGA and PDX expression values were converted to TPM and log2-transformed. All gene names were mapped to Entrez gene ids. All expression values were standardized according to the normalization parameters of the training data.

S6 SENSITIVITY TO HYPER-PARAMETERS

AITL can be quite sensitive to the selection of hyper-parameters, especially to the learning rate, number of training epochs, and the dropout rate. We observe that lower learning rates tend to yield better performance for the AITL models. In addition, a smaller number of training epochs also tends to produce better results, which makes sense because we have limited amounts of training data, and training with higher epochs would overfit the model. Lastly, we observe that dropout rates of around 0.4 - 0.5 result in the highest performing AITL models.

REFERENCES

Kumar MR Bhat and Vijayasaradhi Setaluri. Microtubule-associated proteins as targets in cancer chemotherapy. *Clinical Cancer Research*, 13(10):2849–2854, 2007.

Yi-Hsin Chen, Wei-Yu Chen, Yu-Ting Chen, Bo-Cheng Tsai, Yu-Chiang Frank Wang, and Min Sun. No more discrimination: Cross city adaptation of road scene segmenters. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1992–2001, 2017.

Evelyne Collignon, Annalisa Canale, Clémence Al Wardi, Martin Bizet, Emilie Calonne, Sarah Dedeurwaerder, Soizic Garaud, Céline Naveaux, Whitney Barham, Andrew Wilson, et al. Immunity drives tet1 regulation in cancer through nf-κb. *Science advances*, 4(6):eaap7309, 2018.

Manhong Dai, Pinglang Wang, Andrew D Boyd, Georgi Kostov, Brian Athey, Edward G Jones, William E Bunney, Richard M Myers, Terry P Speed, Huda Akil, et al. Evolving gene/transcript definitions significantly alter the interpretation of genechip data. *Nucleic acids research*, 33(20): e175–e175, 2005.

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, 2015.

Hiromasa Hideshima, Yasuhiro Yoshida, Hiroshi Ikeda, Maya Hide, Akinori Iwasaki, Kenneth C Anderson, and Teru Hideshima. Ikkβ inhibitor in combination with bortezomib induces cytotoxicity in breast cancer cells. *International journal of oncology*, 44(4):1171–1176, 2014.

Rafael A Irizarry, Benjamin M Bolstad, Francois Collin, Leslie M Cope, Bridget Hobbs, and Terence P Speed. Summaries of affymetrix genechip probe level data. *Nucleic acids research*, 31(4): e15–e15, 2003.

Yunhai Li, Jing Huang, Jiazheng Sun, Shili Xiang, Dejuan Yang, Xuedong Ying, Mengqi Lu, Hongzhong Li, and Guosheng Ren. The transcription levels and prognostic values of seven proteasome alpha subunits in human cancers. *Oncotarget*, 8(3):4501, 2017.

Subrata Manna, Bipradeb Singha, Sai Aung Phyo, Himavanth Reddy Gatla, Tzu-Pei Chang, Shannon Sanacora, Sitharam Ramaswami, and Ivana Vancurova. Proteasome inhibition by bortezomib increases il-8 expression in androgen-independent prostate cancer cells: the role of ikkα. *The Journal of Immunology*, 191(5):2837–2846, 2013.

Jeremy A Miller, Chaochao Cai, Peter Langfelder, Daniel H Geschwind, Sunil M Kurian, Daniel R Salomon, and Steve Horvath. Strategies for aggregating gene expression data: the collapserows r function. *BMC bioinformatics*, 12(1):322, 2011.

Soufiane Mourragui, Marco Loog, Mark A van de Wiel, Marcel JT Reinders, and Lodewyk FA Wessels. Precise: a domain adaptation approach to transfer predictors of drug response from pre-clinical models to tumors. *Bioinformatics*, 35(14):i510–i519, 2019.

Alexandre Rouette, Assya Trofimov, David Haberl, Geneviève Boucher, Vincent-Philippe Lavallée, Giovanni D'Angelo, Josée Hébert, Guy Sauvageau, Sébastien Lemieux, and Claude Perreault. Expression of immunoproteasome genes is regulated by cell-intrinsic and–extrinsic factors in human cancers. *Scientific reports*, 6:34019, 2016.

Hossein Sharifi-Noghabi, Olga Zolotareva, Colin C Collins, and Martin Ester. MOLI: multi-omics late integration with deep neural networks for drug response prediction. *Bioinformatics*, 35(14): i501–i509, 2019. ISSN 1367-4803. doi: 10.1093/bioinformatics/btz318.

Petr Smirnov, Victor Kofia, Alexander Maru, Mark Freeman, Chantal Ho, Nehme El-Hachem, George-Alexandru Adam, Wail Ba-alawi, Zhaleh Safikhani, and Benjamin Haibe-Kains. Pharmacodb: an integrative database for mining in vitro anticancer drug screening studies. *Nucleic acids research*, 46(D1):D994–D1002, 2017.

Marta Smoter, Lubomir Bodnar, Renata Duchnowska, Rafał Stec, Bartłomiej Grala, and Cezary Szczylik. The role of tau protein in resistance to paclitaxel. *Cancer chemotherapy and pharmacology*, 68(3):553–557, 2011.

Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems*, pp. 4077–4087, 2017.

Peter Tsvetkov, Ethan Sokol, Dexter Jin, Zarina Brune, Prathapan Thiru, Mahmoud Ghandi, Levi A Garraway, Piyush B Gupta, Sandro Santagata, Luke Whitesell, et al. Suppression of 19s proteasome subunits marks emergence of an altered cell state in diverse cancers. *Proceedings of the National Academy of Sciences*, 114(2):382–387, 2017.

Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7167–7176, 2017.

Jian Yang, Yang Yu, Wei Liu, Zhi Li, Zhongqing Wei, and Rongjiang Jiang. Microtubule-associated protein tau is associated with the resistance to docetaxel in prostate cancer cell lines. *Research and reports in urology*, 9:71, 2017.

Yujie Zhao, Nathan R Foster, Jeffrey P Meyers, Sachdev P Thomas, Donald W Northfelt, Kendrith M Rowland Jr, Bassam I Mattar, David B Johnson, Julian R Molina, Sumithra J Mandrekar, et al. A phase i/ii study of bortezomib in combination with paclitaxel, carboplatin, and concurrent thoracic radiation therapy for non–small-cell lung cancer: North central cancer treatment group (ncctg)-n0321. *Journal of Thoracic Oncology*, 10(1):172–180, 2015.