
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	512 (number of nodes in the hidden layer of the feature extractor), 256 (number of nodes in the shared layer of the multi-task subnetwork), 256 (number of nodes in the hidden layer of the regression tower), 5e-5 (learning rate), 0.1 and 0.8 (regularization for global and class-wise discriminators), 16 and 32 (mini-batch size for the source and target domains), 0.4 (dropout rate), 20 (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 feature 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 feature 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 feature 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 feature 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	MAP4 ($P < 1 \times 10^{-10}$).
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 ($P < 1 \times 10^{-10}$), 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 ($P < 1 \times 10^{-10}$), NR1H2 ($P = 0.04$).
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 ($P = 5.5 \times 10^{-8}$), MAP4 ($P < 1 \times 10^{-10}$).
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	MAP2 ($P = 8.5 \times 10^{-4}$), 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 ($P < 1 \times 10^{-10}$).

REFERENCES

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