Learning and applying general principles of cancer cell drug sensitivity

ShT-04.6-4

F. Carli[I], P. Di Chiaro[II], M. Morelli[III], C.M. Mazzanti[III], G. Natoli[II], F. Raimondi[I]

[I]Bio@SNS, Scuola Normale Superiore, Pisa, Italy, [II]Department of Experimental Oncology, European Institute of Oncology - IEO, Milano, Italy, [III]Fondazione Pisana Scienza, Pisa, Italy

High-throughput screening platforms for the profiling of sensitivity of hundreds of cancer cell lines (CCLs) to drugs have generated large datasets that hold the potential to unlock new targeted, anti-tumor therapies.

In this study, we leveraged these datasets to train explainable machine learning algorithms by employing cell line transcriptomics to predict the growth inhibitory potential (e.g. IC50) of drugs. We used large language models (LLMs) to expand descriptions of the mechanisms of action (MOA) for each drug starting from available annotations, which were matched to the semantically closest pathways from reference knowledge bases. Through this AI-curated resource, and thanks to the inherent interpretability of our model, we demonstrated that pathways enriched for genes most important for predictiveness often match known drug-MOAs and essential genes, suggesting that most of the models learned the molecular determinants of drug response. We then showed that by using the LLM-curated MOA-genes, we were able to further improve the predictive performances of drug models.

To enhance translatability to clinical samples, we employed a pipeline to align bulk RNAseq from CCLs, used for training the models, to the ones from patient samples, used for inference. We showcased the utility of our approach on transcriptomics data from TCGA samples, for which patients samples that are best scored for each drug are consistent with therapies prescribed for the corresponding cancer type. We further demonstrated the applicability of our method by inferring effective drugs on samples from pancreatic cancer and glioblastoma patients. Predictions were experimentally validated on commercial as well as primary cell lines, respectively, confirming the reliability of our model.

In summary, we demonstrated that our method facilitates the inference and interpretation of cancer cell line drug sensitivity, and holds potential to effectively translate them into new cancer therapeutics.