ReMAP: Repurposing through Multi-Omics Analysis and Prediction

Abstract

The complexity and high costs of traditional drug development have driven increased interest in drug repurposing, a strategy that explores new therapeutic uses for existing drugs. This approach leverages established safety profiles and can significantly accelerate the drug development process. Recent advances in computational methods, particularly those that utilize multi-omics data, have further enhanced the potential for systemic drug repurposing. In this study, we present **ReMAP (Repurposing through Multi-Omics Analysis and Prediction), a model designed for predicting drug responses, with a focus on applications in cancer treatment**. By integrating somatic mutations, copy number aberrations, and gene expression data, ReMAP outperforms traditional single-omics and early integration approaches in predictive accuracy. Utilizing data from the PRISM database, we identify potential new indications for drugs, including Dacomitinib for Head & Neck Squamous cell cancer. This work demonstrates the potential of multi-omics integration and machine learning to revolutionize drug repurposing in oncology, bridging the gap between computational predictions and clinical applications.

Introduction

Pharmaceutical companies invest billions of dollars and undergo a rigorous process of multiple clinical trials to ensure a drug's dosage, efficacy, and safety before it reaches the market. High attrition rates, lengthy drug discovery timelines, and strict regulatory compliance are some of the major challenges companies face.

Drug repurposing, also known as drug repositioning, addresses these challenges by investigating existing drugs for new therapeutic uses. This approach has gained popularity due to its potential to accelerate drug development, reduce costs, and leverage the established safety profiles of existing medications. Over the past decade, interest in drug repurposing has surged, particularly for rare diseases and complex conditions like cancer.[1,2] In oncology, where the complexity and heterogeneity of cancers create significant hurdles, drug repurposing offers substantial promise. With only 5% of rare diseases having an approved treatment, innovative therapeutic strategies are critically needed.[3,4]

Developing new cancer therapies often encounters obstacles such as limited patient populations, disease complexity, and high costs. Drug repurposing provides a time- and cost-efficient alternative to traditional drug development, making it an attractive option for addressing unmet needs in cancer treatment.[5] Traditionally, successes in drug repurposing have often been the result of serendipitous discoveries.[6] For example, Thalidomide, initially developed as a sedative and later withdrawn due to safety concerns, was successfully repurposed for treating multiple myeloma and leprosy[7,8]. Similarly, cancer drugs like crizotinib, sorafenib, azacitidine, and decitabine, initially failed in their original indications but are now essential tools in cancer treatment.[9]

In recent years, computational approaches have surged in supporting systematic drug repurposing. These methods utilize diverse information sources, such as electronic health records, genome-wide association analyses, gene expression profiles, multi-omics data, pathway mappings, compound structures, target binding assays, and phenotypic profiling data[6,10,11,12]. By leveraging these data-driven techniques, researchers systematically analyze multiple components—such as chemical structures, adverse event profiles, compound-target interactions, and omics information—to suggest new therapeutic indications for existing drugs.[13]

Accurate drug response prediction is essential for successful repurposing, particularly in oncology, where the complexity and heterogeneity of cancers result in varied responses to treatment. Tailoring cancer therapy requires understanding how patients' molecular and clinical profiles influence their responses to anticancer drugs.[14] Extensive drug screening data from resources like the Cancer Cell Line Encyclopedia (CCLE)[15] and the Genomics of Drug Sensitivity in Cancer (GDSC)[16] project have been instrumental in developing computational models that predict drug-target interactions and pharmacological responses.

While gene expression profiles have proven effective for predicting drug responses, incorporating additional omics data can enhance predictive accuracy, especially in pan-cancer models.[17-20] We developed a new model, **ReMAP (Repurposing through Multi-Omics Analysis and Prediction), which integrates somatic mutations, copy number aberrations, and gene expression data to predict drug responses.** This approach is inspired by the MOLI (Multi-Omics Late Integration) method, enhancing its capabilities for drug repurposing, particularly in cancer treatment.[21]

In this case study, we evaluate ReMAP's suitability for modeling drug responses using data from the PRISM (Profiling Relative Inhibition Simultaneously in Mixtures) repurposing database. PRISM contains the growth inhibitory activity of 4,518 drugs tested across 578 human cancer cell lines and is designed to facilitate drug repurposing efforts, especially for cancer. The unique molecular barcoding method used in PRISM allows for the simultaneous screening of drugs against pooled cell lines, significantly increasing the efficiency of the process.

We train an AutoEncoder-based model on PRISM's multi-omics profiles to assess how ReMAP's performance varies with different data inputs. Additionally, we demonstrate how ReMAP can identify new drug indications, emphasizing the direct link between accurate drug response prediction and successful repurposing. We also address the challenge of generalizability by highlighting ReMAP's performance on PRISM data, underscoring the need for robust models capable of generalizing across different datasets. To illustrate the practical potential of this approach in drug repurposing, we trained ReMAP for Dacomitinib, initially used for non-small cell lung cancer, and predicted its response in CCLE cell lines lacking recorded drug response data. **Our findings suggest a potential new indication for Dacomitinib in Head & Neck Squamous cell cancer, consistent with Ather et al.'s findings on its efficacy in Head and Neck Cancer models.**[22]

Data

Researchers have mostly used the Genomics of Drug Sensitivity in Cancer (GDSC) database for building drug response prediction models for various cancer types16. The GDSC database, a collaboration between the Wellcome Sanger Institute and Massachusetts General Hospital's Center for Molecular Therapeutics, contains drug sensitivity data from over 1,000 genetically characterized human cancer cell lines. It integrates genomic and expression data to correlate genetic alterations with drug sensitivity profiles. With nearly 75,000 experiments across 138 drugs, GDSC is the largest public resource of its kind, supporting the discovery of therapeutic biomarkers and personalized cancer therapies. The data is freely accessible, fostering collaboration and innovation in cancer research. Despite the advent of state-of-the-art models on the GDSC dataset, the generalizability of such models on other datasets is still questionable and creates a barrier in selecting the perfect model for its application on real-life patient data.

For this case study, we wanted to experiment with ReMAP's architecture on a different dataset which is relatively less studied for drug response prediction models compared to the GDSC dataset in public research. Since ReMAP's architecture relies on multi-omics input (gene expression, copy number, and mutation), we chose the PRISM repurposing dataset. The unique molecular barcoding method used in PRISM allows for the simultaneous screening of drugs against pooled cell lines, significantly increasing the efficiency of the process. Notably, a substantial number of non-oncology drugs were found to selectively inhibit specific subsets of cancer cell lines, with these effects being predictable based on the molecular features of the cell lines.

To get the multi-omics data for the case study we leveraged data from the Cancer Dependency Map (DepMap) project, which provides a comprehensive molecular and pharmacological characterization of cancer cell lines. DepMap is widely recognized for its extensive collection of multi-omics data, including gene expression, copy number variations, and somatic mutations, across a diverse set of cancer cell lines. This data is crucial for understanding the molecular dependencies

of cancer cells, making it an excellent resource for drug repurposing studies. The integration of DepMap data allows for a more in-depth analysis of drug response prediction by correlating molecular features of cell lines with their sensitivities to various therapeutic agents. Table 1 provides the details of the data types we used in this study.

Table 1: Details of the data types used for Training

Data	File Name	Version	Description
Gene Expression Data	OmicsExpression ProteinCodingGenesTPMLogp1BatchCorrected	DepMap Public 24Q2	This file contains log2(TPM+1) values of protein-coding genes for DepMap cell lines, processed with RSEM and batch-corrected using ComBat.
Mutations Data	OmicsSomaticMutations.csv	DepMap Public 24Q2	This MAF-like file includes data on all somatic point mutations and indels identified in DepMap cell lines, generated using Mutect2.
Copy Number Data	OmicsAbsoluteCNGene.csv	DepMap Public 24Q2	This file contains gene-level absolute copy number data generated with PureCN. Genes overlapping segmental duplications or flagged by repeatMasker are masked. Values are based on the segment with the greatest overlap, using WGS or WES data as available.
Drug-Response Data	secondary-screen-dose-response-curve-parameters	DepMap Public 19Q4	This file contains parameters of dose-response curves fitted to replicate-level viability data using a four-parameter log-logistic function with the drc R package.

For the validation of ReMAP, we utilized the Patient-Derived Xenograft (PDX) data previously employed in the MOLI study. This comprehensive dataset, known as the PDX Encyclopedia, was originally published by Gao et al. in 2015. It encompasses over 300 PDX models representing various cancer types, each of which has been screened against a panel of 34 drugs, including both targeted therapies and traditional chemotherapeutic agents. PDX models are valuable for validation as they closely mimic human tumor biology, preserving the heterogeneity and complexity of the original cancer. By using this established dataset, we ensure consistency with previous research while leveraging a robust and diverse set of cancer models to assess ReMAP's predictive capabilities across different cancer types and treatment modalities.

Methods

Data Preparation

The PRISM data was processed and transformed for ReMAP. The IC50 value of each drug was modified to a binary value denoting 'resistant' or 'sensitive'[22]. The copy number data and Somatic mutation data were converted to binary values using the method described in MOLI.[21]

Model Training

In this study, we developed a model to use multi-omics for drug repurposing by using public datasets like PRISM. Instead of developing a general model, we drew inspiration from the MOLI paper and utilized a transfer learning approach. We trained ReMAP on EGFR-targeted drugs and validated it using PDX data. The MOLI authors trained their model on EGFR-based drugs and validated it on PDX data tested for Cetuximab and Erlotinib, arguing that models trained on EGFR drugs are more effective than a single, general model. However, since the PRISM dataset lacked drug response data for Cetuximab, we focused on training ReMAP solely on EGFR-based drugs to test the hypothesis.

Further, we trained a single drug model for Dacomitinib, which has the highest number of data samples as compared to the rest of the EGFR drugs. This trained Dacomitinib model was then used to predict drug responses on CCLE cell lines that did not have available DepMap PRISM drug response data. The goal of this was to identify potential new indications for Dacomitinib beyond its current use.

We employed hyperparameter tuning to optimize the model parameters for each experiment. Inspired by MOLI, we applied 5-fold cross-validation to tune hyperparameters based on the best AUC (Area Under the Curve). Given the class imbalance between sensitive and resistant samples, AUC was chosen for its ability to balance the true positive rate and false positive rate, offering a more accurate measure of model performance than accuracy. The tuned hyperparameters included the number of hidden layer nodes, learning rates, mini-batch size, weight decay, dropout rate, number of epochs, and margin and regularization terms (for triplet loss). All models were trained using the PyTorch framework.

For each of the experiments we report the training AUC and validation AUC using the best parameters from the hyper-parameter tuning in each case. The reported scores represent the average of 10 independent model runs to ensure result stability and reliability. The details of the experiments are presented in Table 2.

Table 2: Experiment Details and Results

Experiment	Drugs	Number of Cell-lines	Number of Samples[a]	Number of Common Genes	Validation	Training AUC[b]	Validation AUC[b]
Training on all EGFR targeted drugs	39 EGFR-targeted	386	4394 (R: 3292; S:1102)	12547	PDX-Cetuximab	0.66	0.54
Training on top 5 EGFR targeted drugs	Dacomitinib; Pelitinib; WZ8040; Afatinib; Lapatinib	374	1024 (R: 763; S: 261)	12547	PDX-Cetuximab	0.85	0.48
Training on single-drug Dacomitinib	Dacomitinib	234	234 (R: 173; S: 61)	12547	NA	NA	NA

[a] R: Resistant cell-lines or Non-Responder; S: Sensitive cell-lines or Responder
[b] The training AUC and Validation AUC reported here is the average of 10 iterations of the model run.

Results

Training ReMAP architecture on EGFR-targeted drugs from PRISM

We further selected the top 5 EGFR-targeted drugs with the most abundant resistant/sensitive data: dacomitinib, pelitinib, WZ8040, afatinib, and lapatinib. These drugs were primarily tested for non-small cell lung cancer and breast cancer, focusing on their efficacy in targeting EGFR.

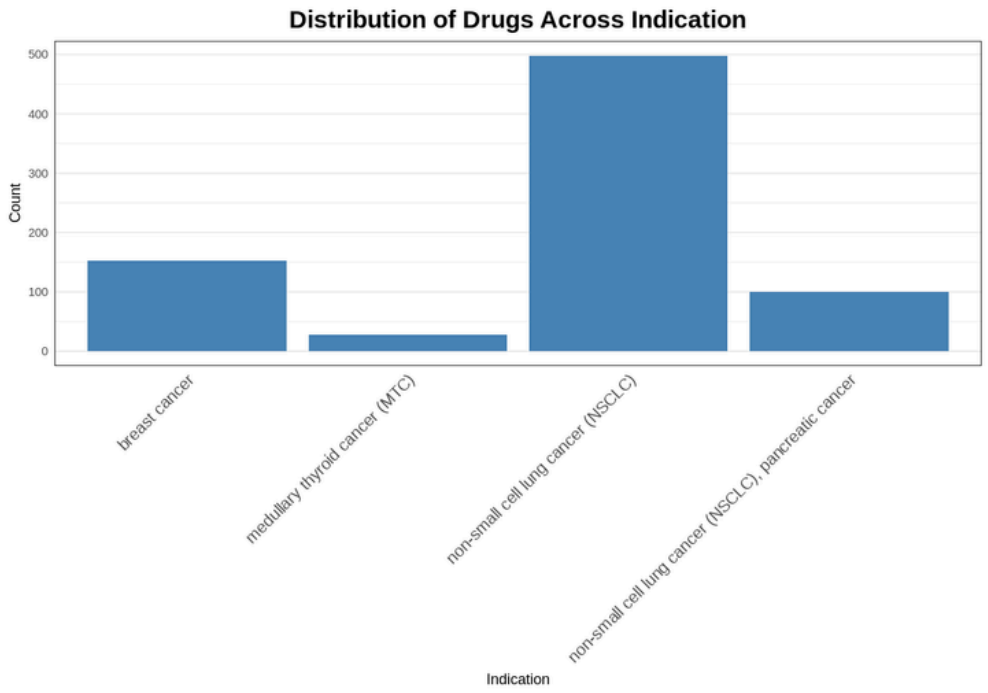

Figure 1: Distribution of EGFR-target Drugs Across Indication

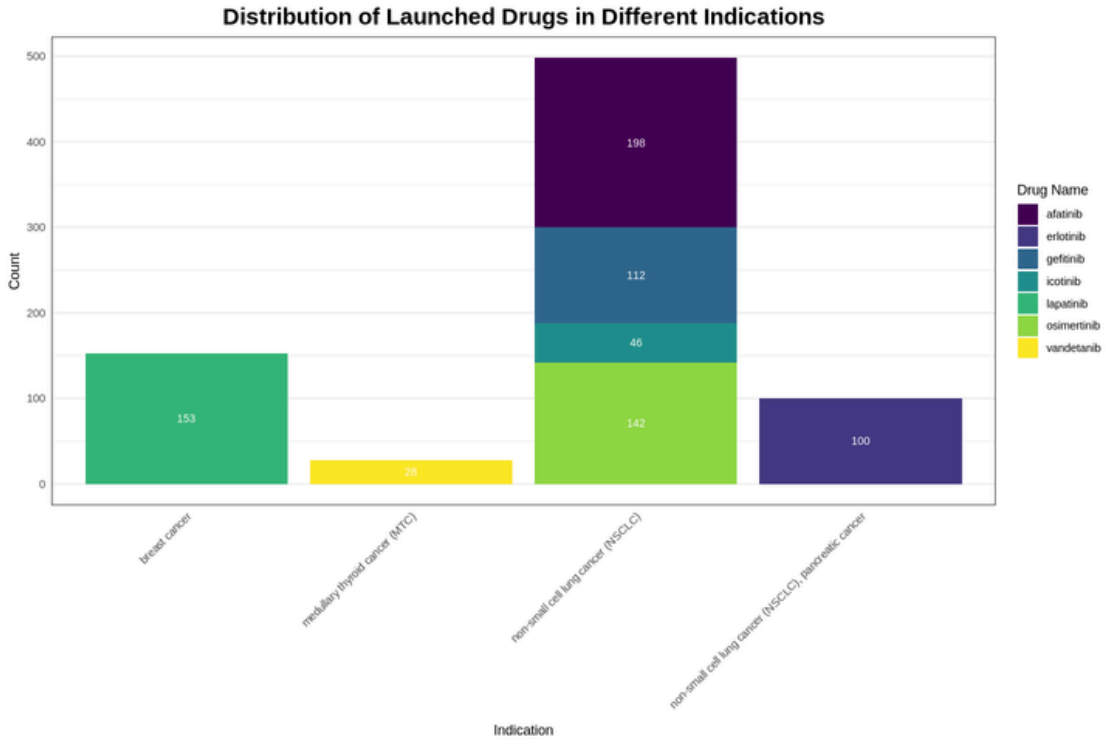

Figure 2: Distribution of Launched EGFR-target Drugs in Different Indications

ReMAP Trained on top 5 EGFR-Targeted Drugs from DepMap Data

We further selected the top 5 EGFR-targeted drugs with the most abundant resistant/sensitive data: dacomitinib, pelitinib, WZ8040, afatinib, and lapatinib. These drugs were primarily tested for non-small cell lung cancer and breast cancer, focusing on their efficacy in targeting EGFR.

Predicting Drug Responses in Unseen CCLE Cell Lines Using ReMAP Trained on EGFR Drug Dacomitinib with Extensive Response Data

We trained the architecture on a single drug, Dacomitinib, which had the most extensive drug-response data among the EGFR-targeted drugs. The trained model was then used to predict drug responses on an unseen CCLE cell line dataset that lacked recorded drug responses. The objective was to identify new indications for Dacomitinib using a multi-omics model.

After training on Dacomitinib, we predicted drug responses across 949 CCLE cell lines that had multi-omics data but no recorded drug responses. To ensure accuracy, predictions for each cell line were averaged over 10 iterations. Our study demonstrated that Dacomitinib exhibits high sensitivity in Head & Neck Squamous Cell Cancer and prostate adenocarcinoma (Figure 3-4). This finding aligns with a previous study which showed that Dacomitinib significantly inhibited growth in multiple cancer types.[23-27]

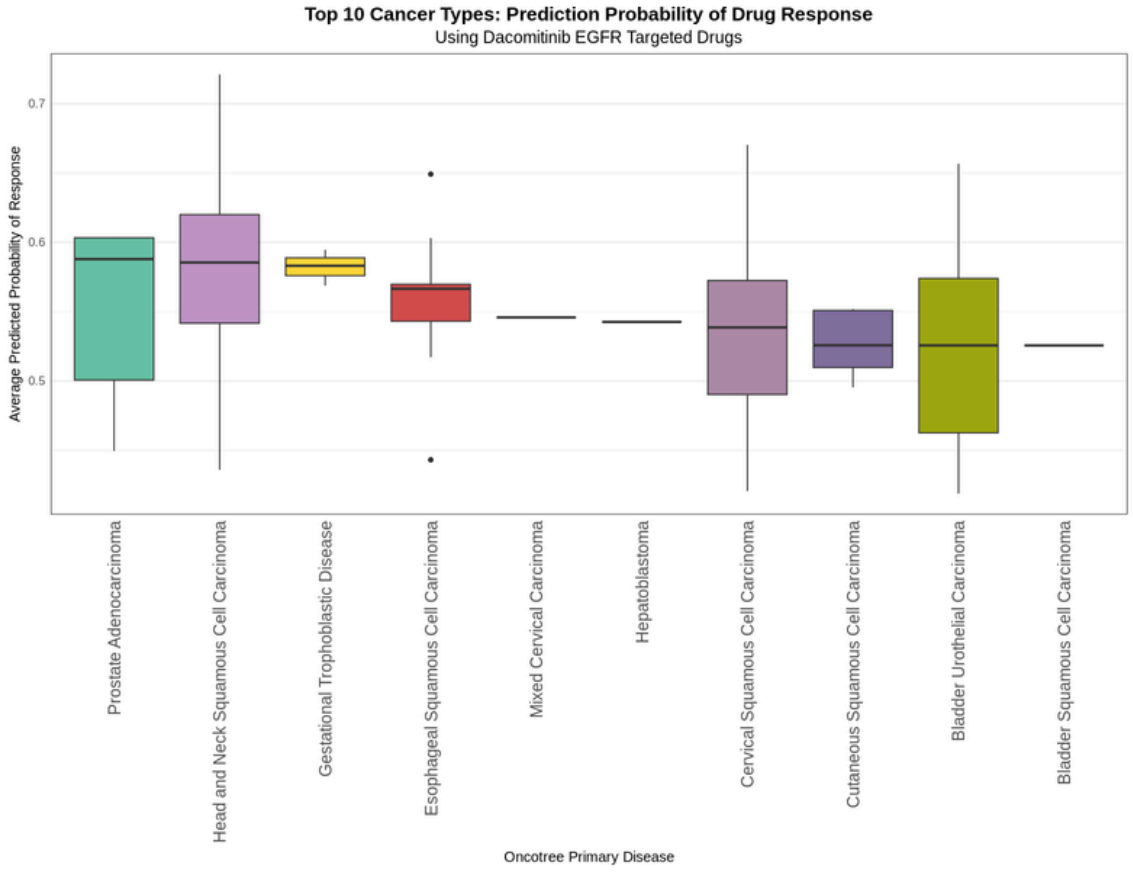

Figure 3: Prediction probability of Dacomitinib response to top 10 cancer types.

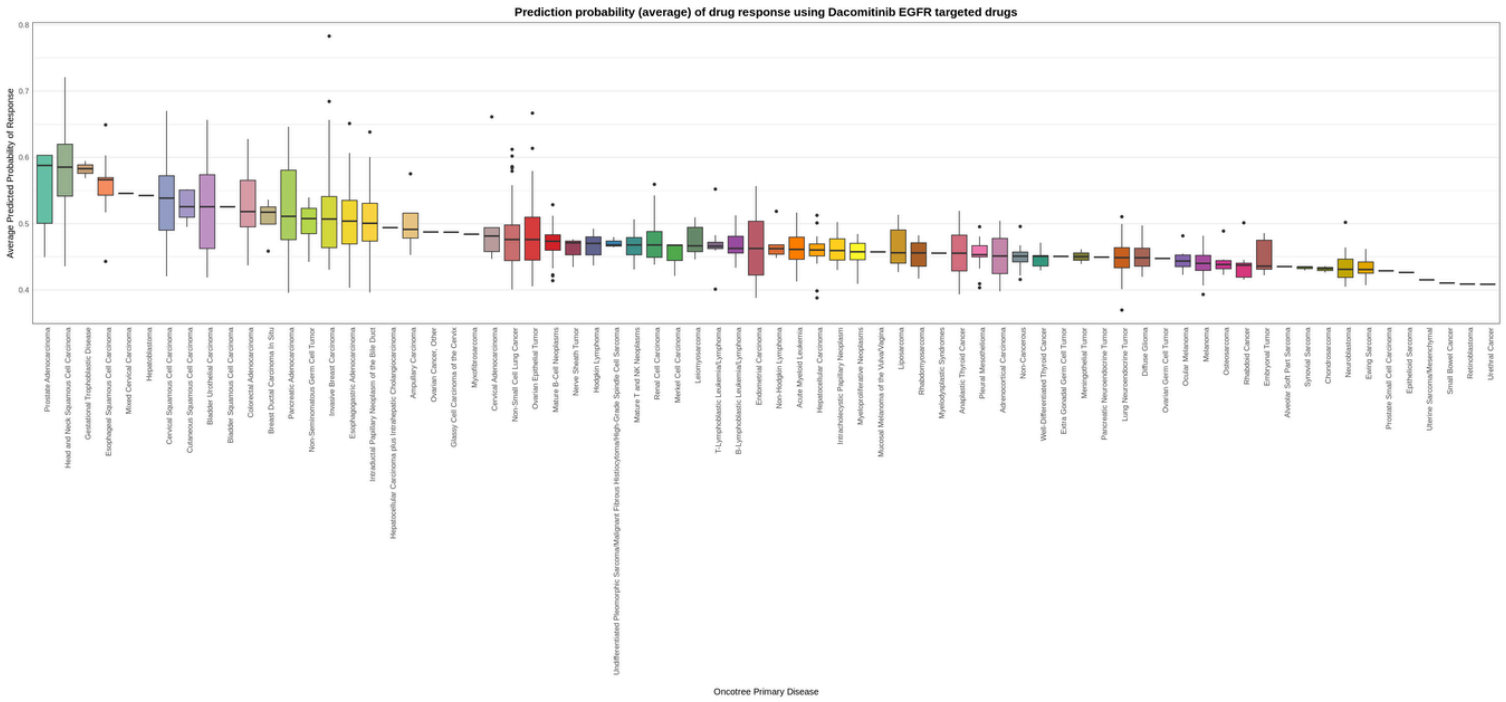

Figure 4: Average Prediction of the Model on Cell-Lines Grouped by Disease.

Discussion

This study aimed to develop a new model for drug repurposing using multi-omics data and to evaluate its potential with the PRISM dataset from DepMap, extending its application beyond the widely used GDSC database. We developed ReMAP (Repurposing through Multi-Omics Analysis and Prediction) and trained it on PRISM data, focusing on all EGFR-targeted drugs and a subset of the top 5 EGFR drugs.

Inspired by the MOLI framework, we trained ReMAP using an ensemble of drugs targeting a single gene—in this case, EGFR. Our dataset included 39 such drugs, following the similar strategy. Using ReMAP, we demonstrated the potential of Dacomitinib in treating additional cancers beyond its current approvals. ReMAP achieved performance levels comparable to those in the original study, which is an encouraging and promising outcome for drug repurposing.

An important aspect of our study is the potential application of ReMAP to proprietary patient multi-omics profiles. Training the model on in-house data can improve performance by tailoring it to specific patient cohorts, thus enhancing clinical relevance. Additionally, our experiment with Dacomitinib revealed indications of sensitivity in Head & Neck Squamous Cell Cancer and prostate adenocarcinoma.This is consistent with prior studies demonstrating Dacomitinib's efficacy in inhibiting proliferation in various cancer types, including head and neck cancer[23,24],pancreatic adenocarcinoma,[25]and castration-resistant prostate cancer,[26]all by targeting EGFR receptors. Furthermore, ReMAP predicted

Dacomitinib's efficacy in esophageal squamous cell carcinoma, which, to our surprise, is already in a phase II clinical trial.[27] These findings validate ReMAP's predictions and underscore its potential clinical impact.

ReMAP architecture employs a late integration approach to multi-omics data. This method contrasts with early integration, where data from different omics layers are combined at the initial stage of analysis. Early integration allows for capturing interactions between different data types from the outset but can lead to challenges such as high dimensionality and potential loss of data specificity. In contrast, late integration processes each omics data type independently before combining the results, preserving the unique characteristics of each dataset and often resulting in improved model performance, particularly in complex biological systems like cancer. The ReMAP architecture leverages this late integration strategy, enabling more nuanced predictions by separately modeling each omics layer before integration. This approach aligns with our findings, where despite the challenges posed by the limited gene set and the absence of certain drugs in the training data, ReMAP demonstrated robust performance. The distinction between early and late integration highlights the importance of choosing the appropriate integration strategy based on the specific goals and challenges of the analysis. As with MOLI, it can be inferred that ReMAP will also demonstrate superior prediction accuracy in external validations when compared to single-omics and early integration multi-omics methods. This improved performance is likely due to MOLI's and ReMAP's ability to better capture complex biological interactions through late integration of multi-omics data.

Moreover, our approach to identifying new indications for existing drugs can significantly reduce preclinical costs and expand therapeutic options. Providing new indications for already tested drugs accelerates the drug repurposing process, offering potential treatments for cancers with limited therapeutic options. This highlights the practical utility of ReMAP in reducing the time and cost associated with drug development. The relatively low AUC performance on the validation PDX-Cetuximab data can be attributed to a smaller number of cell lines and genes. Future efforts should focus on enhancing the model's performance by incorporating a larger number of genes, increasing the diversity of training datasets, and exploring advanced techniques to handle high-dimensional multi-omics data. Further validation using additional datasets and clinical data will be crucial to establish the model's robustness and clinical applicability.

This case study demonstrates the power of integrating multi-omics data with advanced machine learning for drug response prediction, showcasing how AI-driven approaches can accelerate drug repurposing in oncology. By bridging computational prediction and clinical application, models like ReMAP offer a systematic and efficient method for discovering new cancer treatments. However,

the challenge of generalizing models across diverse datasets due to high dimensionality remains significant. Our research highlights the potential of using PRISM data for constructing multi-omics models, offering valuable insights for drug repurposing and generating leads for further investigation.
In conclusion, this study demonstrates the feasibility and potential benefits of using deep learning models like ReMAP or MOLI for drug repurposing. Our approach shows promise in predicting drug responses and identifying new therapeutic indications, paving the way for more efficient and cost-effective cancer treatments.

References

1. Tambuyzer, E., Vandendriessche, B., Austin, C. P., Brooks, P. J., Larsson, K., Miller Needleman, K. I., ... & Prunotto, M. (2020). Therapies for rare diseases: therapeutic modalities, progress and challenges ahead. Nature Reviews Drug Discovery, 19(2), 93-111.

2. Kort, E., & Jovinge, S. (2021). Drug repurposing: claiming the full benefit from drug development. Current Cardiology Reports, 23(6), 62.

3. Haendel, M., Vasilevsky, N., Unni, D., Bologa, C., Harris, N., Rehm, H., ... & Oprea, T. I. (2020). How many rare diseases are there?. Nature reviews drug discovery, 19(2), 77-78.

4. Boycott, K. M., Lau, L. P., Cutillo, C. M., & Austin, C. P. (2019). International collaborative actions and transparency to understand, diagnose, and develop therapies for rare diseases. EMBO Molecular Medicine, 11(5), e10486.

5. Roessler, H. I., Knoers, N. V., van Haelst, M. M., & van Haaften, G. (2021). Drug repurposing for rare diseases. Trends in pharmacological sciences, 42(4), 255-267.

6. Pushpakom, S., Iorio, F., Eyers, P. A., Escott, K. J., Hopper, S., Wells, A., ... & Pirmohamed, M. (2019). Drug repurposing: progress, challenges and recommendations. Nature reviews Drug discovery, 18(1), 41-58.

7. Barnhill, R. L., & McDougall, A. C. (1982). Thalidomide: use and possible mode of action in reactional lepromatous leprosy and in various other conditions. Journal of the American Academy of Dermatology, 7(3), 317-323.

8. Singhal, S., Mehta, J., Desikan, R., Ayers, D., Roberson, P., Eddlemon, P., ... & Barlogie, B. (1999). Antitumor activity of thalidomide in refractory multiple myeloma. New England Journal of Medicine, 341(21), 1565-1571.

9. Gupta, S. C., Sung, B., Prasad, S., Webb, L. J., & Aggarwal, B. B. (2013). Cancer drug discovery by repurposing: teaching new tricks to old dogs. Trends in pharmacological sciences, 34(9), 508-517.

10. Lotfi Shahreza, M., Ghadiri, N., Mousavi, S. R., Varshosaz, J., & Green, J. R. (2018). A review of network-based approaches to drug repositioning. Briefings in bioinformatics, 19(5), 878-892.

11. Zhao, K., & So, H. C. (2019). Using drug expression profiles and machine learning approach for drug repurposing. Computational methods for drug repurposing, 219-237.

12. Tanoli, Z., Vähä-Koskela, M., & Aittokallio, T. (2021). Artificial intelligence, machine learning, and drug repurposing in cancer. Expert opinion on drug discovery, 16(9), 977-989.

13. Wang, Y., Aldahdooh, J., Hu, Y., Yang, H., Vähä-Koskela, M., Tang, J., & Tanoli, Z. (2022). DrugRepo: a novel approach to repurposing drugs based on chemical and genomic features. Scientific Reports, 12(1), 21116.

14. Wang, C., Lye, X., Kaalia, R., Kumar, P., & Rajapakse, J. C. (2021). Deep learning and multi-omics approach to predict drug responses in cancer. BMC bioinformatics, 22(Suppl 10), 632.

15. Cancer Cell Line Encyclopedia Consortium, & Genomics of Drug Sensitivity in Cancer Consortium. (2015). Pharmacogenomic agreement between two cancer cell line data sets. Nature, 528(7580), 84-87.

16. Yang, W., Soares, J., Greninger, P., Edelman, E. J., Lightfoot, H., Forbes, S., ... & Garnett, M. J. (2012). Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic acids research, 41(D1), D955-D961.

17. Ding, Z., Zu, S., & Gu, J. (2016). Evaluating the molecule-based prediction of clinical drug responses in cancer. Bioinformatics, 32(19), 2891-2895.

18. Geeleher, P., Cox, N. J., & Huang, R. S. (2014). Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. Genome biology, 15, 1-12.

19. Graim, K., Friedl, V., Houlahan, K. E., & Stuart, J. M. (2018). PLATYPUS: A multiple—view learning predictive framework for cancer drug sensitivity prediction. In Biocomputing 2019: Proceedings of the pacific symposium (pp. 136-147).

20. Iorio, F., Knijnenburg, T. A., Vis, D. J., Bignell, G. R., Menden, M. P., Schubert, M., ... & Garnett, M. J. (2016). A landscape of pharmacogenomic interactions in cancer. Cell, 166(3), 740-754.

21. Sharifi-Noghabi, H., Zolotareva, O., Collins, C. C., & Ester, M. (2019). MOLI: multi-omics late integration with deep neural networks for drug response prediction. Bioinformatics, 35(14), i501-i509.

22. Knijnenburg, T. A., Klau, G. W., Iorio, F., Garnett, M. J., McDermott, U., Shmulevich, I., & Wessels, L. F. (2016). Logic models to predict continuous outputs based on binary inputs with an application to personalized cancer therapy. Scientific reports, 6(1), 36812.

23. Ather, F., Hamidi, H., Fejzo, M. S., Letrent, S., Finn, R. S., Kabbinavar, F., ... & Wong, S. G. (2013). Dacomitinib, an irreversible Pan-ErbB inhibitor significantly abrogates growth in head and neck cancer models that exhibit low response to cetuximab. PloS one, 8(2), e56112.

24. Williams, J. P., Kim, I., Ito, E., Shi, W., Yue, S., Siu, L. L., ... & Liu, F. F. (2014). Pre-clinical characterization of Dacomitinib (PF-00299804), an irreversible pan-ErbB inhibitor, combined with ionizing radiation for head and neck squamous cell carcinoma. PLoS One, 9(5), e98557.

25. Momeny, M., Esmaeili, F., Hamzehlou, S., Yousefi, H., Javadikooshesh, S., Vahdatirad, V., ... & Ghaffari, S. H. (2019). The ERBB receptor inhibitor dacomitinib suppresses proliferation and invasion of pancreatic ductal adenocarcinoma cells. Cellular Oncology, 42, 491-504.

26. Jathal, M. K., Steele, T. M., Siddiqui, S., Mooso, B. A., D'Abronzo, L. S., Drake, C. M., ... & Ghosh, P. M. (2019). Dacomitinib, but not lapatinib, suppressed progression in castration-resistant prostate cancer models by preventing HER2 increase. British journal of cancer, 121(3), 237-248.

27. Kim, H. S., Kim, S. M., Kim, H., Pyo, K. H., Sun, J. M., Ahn, M. J., ... & Cho, B. C. (2015). Phase II clinical and exploratory biomarker study of dacomitinib in recurrent and/or metastatic esophageal squamous cell carcinoma. Oncotarget, 6(42), 44971.