Artificial intelligence and proteomics: predicting genome-wide perturbation response

David Futorian, Gali Arad, Nitzan Simchi, Oren Fischman, Kirill Pevzner | Protai, Tel Aviv, Israel

Introduction

Despite recent advances in precision medicine, accurate prediction of the response of a patient to drug treatment remains challenging. As proteins and Post Translational Modifications (PTM's) best represent the functional state of cells, we utilize these modalities in our overarching goal to build a general purpose response prediction model.

In this work, we use Deep Learning to combine high resolution mass spectrometry proteomics, PTMs, and additional omics to predict cell line response to CRISPR knockout and drug screens.

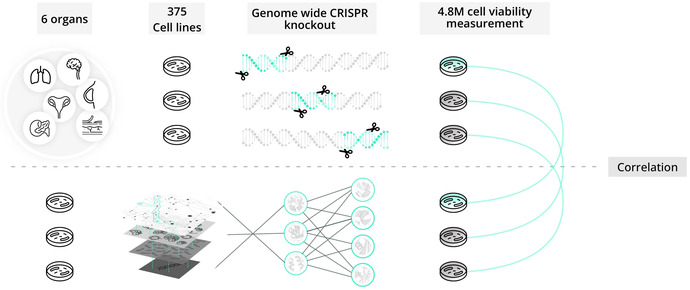


Figure 1: Experimental and computational process for CRISPR KO prediction

Methods

We re-analyzed the raw mass spectrometry proteomic data of 375 cancer cell lines provided by the Cancer Cell Line Encyclopedia (CCLE[1]). Our reanalysis resulted in identification of over 12,000 proteins as well as 16,000 phosphorylation sites (>1000 per sample). Experimental outcome of genome wide CRISPR screen of the same cell lines was obtained from the DepMap[2] portal.

We built a Graph Neural Network (GNN) model based on protein-protein interactions (STRING[3] - human only high confidence protein network). We used proteomics, transcriptomics, and pathway-annotation as input features to predict a CRISPR knockout dependency score (Fig. 2).

Model design follows protein interaction network

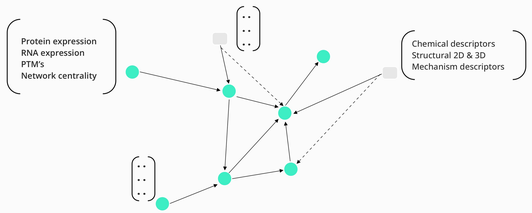
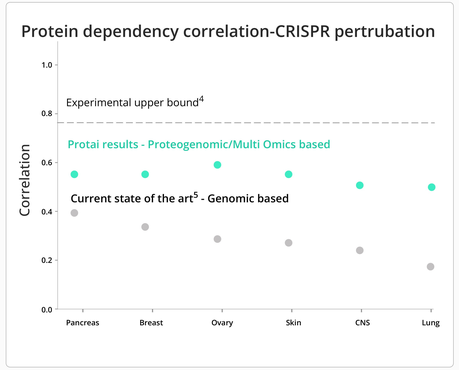
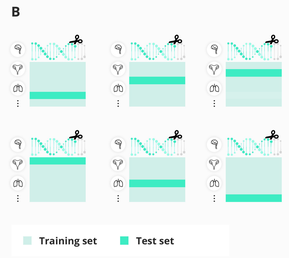


Figure 2: GNN model architecture with proteins and drugs (nodes) and PPI interactions (edges)

Results

Protai platform outperforms current state of the art predictions of CRISPR knockouts





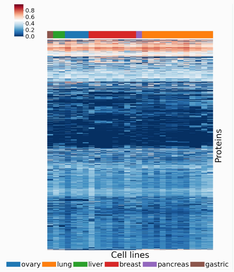
Data description:

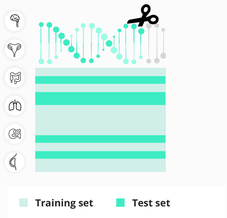
→ 6 organ systems
→ 375 cell lines
→ ~20,000 genes
→ 4.8M cell death experiments[2]

Figure 3: Dependency score correlation between DepMap experimental and predicted results

The model was then evaluated by calculating Pearson correlation coefficients between the DepMap dependency score and the predicted model score across all genes per cell line. Significant improvement is observed when the model's results are compared to recently published results of a similar prediction tool[5] using gene expression and somatic mutations. The comparison was done using a one-tissue-holdout as a test set.

High accuracy of dependency score prediction across 95% of genes



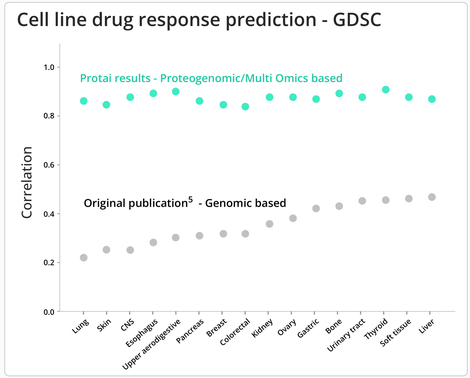


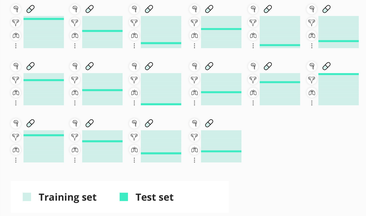
The heatmap in Figure 4 shows the dependency score error per gene in each cell line.

Evaluation of the error, demonstrates that the model achieves good precision on more than 95% of the genes, consistently across the different tissues.

Figure 4: CRISPR KO prediction performance

Using the same approach, we applied our model to predict cell-line/drug treatment-response (IC50) based on the GDSC[6] drug screen dataset.
Our model (light-green) shows a major improvement across all tissues when compared to a genomics-based model (gray) in a one-tissue-holdout method (Fig. 5).





Data description:

→ 16 organ systems
→ 212 cell lines
→ 360 drugs
→ 76K IC50 experiments[6]

Figure 5: IC50 score correlation between GDSC experimental and predicted results

Conclusions and future research

Aiming to develop general purpose multidrug response predictor, we demonstrate initial results on the potential of proteomics and PTMs to predict preclinical cell line response to treatment with CRISPR and drugs.

Our future research will explore the utility of this approach for clinical questions of response prediction, identify new therapeutic targets, clinical trial optimization and patient stratification use cases.

References

1. Nusinow, David P., et al. "Quantitative proteomics of the cancer cell line encyclopedia." Cell 180.2 (2020): 387-402.
2. Tsherniak, Aviad, et al. "Defining a cancer dependency map." Cell 170.3 (2017): 564-576.
3. Szklarczyk, Damian, et al. "The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/. measurement sets." Nucleic acids research 49.D1 (2021): D605-D612.
4. Dempster, Joshua M., et al. "Agreement between two large pan-cancer CRISPR-Cas9 gene dependency data sets." Nature communications 10.1 (2019): 1-14.
5. Ma, Jianzhu, et al. "Few-shot learning creates predictive models of drug response that translate from high-throughput screens to individual patients." Nature Cancer 2.2 (2021): 233-244.
6. Iorio F et al. A landscape of pharmacogenomic interactions in cancer. Cell 166, 740–754 (2016)
7. Ghandi M et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. Nature 569, 503–508 (2019)
8. Meyers RM et al. Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. Nat. Genet 49, 1779–1784 (2017)