

## Preview

# Artificial intelligence-assisted drug repurposing via “chemical-induced gene expression ranking”

Takaaki Masuda<sup>1</sup> and Koshi Mimori<sup>1,\*</sup><sup>1</sup>Department of Surgery, Kyushu University Beppu Hospital, Beppu, Japan\*Correspondence: [mimori.koshi.791@m.kyushu-u.ac.jp](mailto:mimori.koshi.791@m.kyushu-u.ac.jp)<https://doi.org/10.1016/j.patter.2022.100470>

Drug repurposing using artificial intelligence algorithms is a powerful technique that leverages existing datasets to find new medical applications for approved drugs. Pham et al. developed CIGER, a deep learning framework to overcome unreliable data in the datasets and present repositioned drugs against pancreatic cancer.

Drug development is a time-intensive process, and many challenges are exacerbated by high attrition rates and high financial costs. Drug repurposing, also known as drug repositioning, represents a strategy to manage the difficulties of traditional drug development through the redeployment of drugs with established records of safety and efficacy. In this strategy, new indications are developed through the identification of novel effects of approved drugs.<sup>1</sup> While the use of approved drugs has clear benefits, the discovery of these novel effects requires the deployment of sophisticated techniques.

Conventional methods for identifying potential drugs to be repositioned or for identifying new targets include experimental approaches, such as protein affinity chromatography. However, *in vitro* and *in vivo* approaches can be expensive and time consuming. Therefore, *in silico* methods of drug repurposing are considered promising strategies for target identification. *In silico* drug repurposing involves bioinformatic analyses of public datasets to systematically identify interaction networks that link drugs with potential targets. This approach has had some preliminary success because of the recent accumulation of large amounts of relevant data regarding protein structures, genomics, transcriptomics, proteomics, and pharmacophores, along with advancements in bioinformatic tools and computational science. These new bioinformatic techniques have proven to be particularly useful in the systematic identification of drug-target interaction networks, especially via the analysis of gene expression

profiles using large existing public datasets.<sup>2,3</sup>

The effectiveness of this approach is limited by several problems with currently available datasets. First, despite the comprehensive coverage of chemical information such as concentrations of chemicals and cell lines used in treatment experiments and gene expression patterns, key information is frequently missing from existing datasets. Second, gene expression data tend to be quite noisy due to experimental limitations.

To address these issues, Pham et al. developed chemical-induced gene expression ranking (CIGER): a neural network-based algorithm to predict the ranking of genes in gene expression profiles that would be produced upon the treatment of cells with chemical agents.<sup>4</sup> To predict the ranking of genes, CIGER leverages publicly available data and information regarding biologically relevant entities, including chemicals, cell lines, and genes.

As a proof of concept, Pham et al. employed CIGER to conduct a drug repurposing screen for pancreatic cancer, one of the leading causes of cancer-related deaths worldwide. This screen was based on the concept that treatment with effective drugs for particular diseases should lead to similar gene expression patterns. Therefore, drugs that may be applied to pancreatic cancer can be identified solely based on their initiation of favorable gene expression patterns.

In order to identify drugs to repurpose for use in pancreatic cancer, Pham et al. first trained CIGER with the LINCS L1000 dataset. This dataset arose from a National Institutes of Health project in which

approximately 100 cell lines were treated with over 50,000 chemicals, leading to the generation of more than 1,000,000 gene expression profiles that were used to identify relationships between gene expression profiles and molecular structures. Next, molecular structures were retrieved from the DrugBank database, which focuses on drug information that includes interactions, pharmacology, chemical structures, targets, and metabolism (<https://go.drugbank.com/>). These structures were entered into CIGER to predict the gene expression profiles that likely would correspond with the chemical structures. These profiles were compared with treatment profiles calculated from untreated samples with the target disease and those treated with bona fide drugs. CIGER then identified chemicals in DrugBank whose predicted gene expression patterns most closely matched those in treated samples.

Using this approach, Pham et al. identified several drug candidates for pancreatic cancer based on similarities in the rank orders of the gene expression profiles of pancreatic cancer cell lines treated with both metformin and vitamin C. This combination has been reported to suppress the progression of pancreatic ductal adenocarcinoma (PDAC) by restoring the activity of *TET2* and *GATA6*.<sup>5</sup> Similar to this work with CIGER, Zhu et al. also demonstrated a deep learning-based efficacy prediction system that identifies drug candidates by using changes in gene expression profiles in the diseased state from the LINCS L1000 dataset and gene set enrichment analysis.<sup>6</sup>



Thus, because of its capability of predicting rankings in chemical-induced gene expression profiles using only chemical structures, CIGER opens up new possibilities for gene expression-based drug repurposing screens. Most significantly, CIGER has the potential to identify repositioned drugs for which expression data are not available. Furthermore, it should be noted that this approach can readily be applied for finding treatments for multiple disease states, not only malignancies.

Recently, a new artificial intelligence (AI)-based drug discovery platform, “lead identification with a graph-ensemble network for arbitrary targets by harnessing only underlying primary sequence” (LIGHTHOUSE) has been developed.<sup>7</sup> This algorithm is a deep learning mechanism that can predict the ability of chemicals to interact with proteins in the absence of structural information through the use of the search tool for interactions of chemicals (STITCH) dataset.<sup>8</sup> This dataset aggregates experimental data with predictive datasets from several algorithms that analyze protein-protein and protein-chemical interactions. LIGHTHOUSE has the potential to facilitate the identification of drug candidates for target proteins from a huge chemical space based only on the primary structures of target proteins.

In addition to potentially identifying repositioned drugs, AI could be applied to drug development by targeting genetic alterations. The *KRAS* gene is a representative oncogenic driver gene in human cancers.<sup>9</sup> Specifically, *KRAS* mutations are found in approximately 90% of

PDAC cases, and the G12D mutation in the *KRAS* protein is the dominant mutant subtype.<sup>10</sup> The high prevalence of gene alterations and their importance in tumor progression makes targeting them an attractive cancer therapy. AI-based drug development could pave the way for precision oncology by targeting genetic alterations in driver genes.

Drug repurposing is a rational approach to the development of medical treatments and saves both time and monetary costs, though it should be recognized that there are multiple practical limitations, especially regarding intellectual property rights and commercial considerations. Large and sophisticated datasets that integrate multiple sources of information about drug structures and interactions are needed in order to improve the performance of *in silico* approaches to drug repurposing, especially those utilizing AI algorithms.

#### ACKNOWLEDGMENTS

T.M. and K.M. are supported by the Japan Society for the Promotion of Science.

#### DECLARATION OF INTERESTS

The authors declare no competing interests.

#### REFERENCES

1. Malik, J.A., Ahmed, S., Jan, B., Bender, O., Al Hagbani, T., Alqarni, A., and Anwar, S. (2022). Drugs repurposed: An advanced step towards the treatment of breast cancer and associated challenges. *Biomed. Pharmacother.* *145*, 112375. <https://doi.org/10.1016/j.biopha.2021.112375>. PubMed.
2. Galati, S., Di Stefano, M., Martinelli, E., Poli, G., and Tuccinardi, T. (2021). Recent Advances in *In Silico* Target Fishing. *Molecules* *26*, 5124.

<https://doi.org/10.3390/molecules26175124>. PubMed.

3. Masuda, T., Tsuruda, Y., Matsumoto, Y., Uchida, H., Nakayama, K.I., and Mimori, K. (2020). Drug repositioning in cancer: The current situation in Japan. *Cancer Sci.* *111*, 1039–1046. <https://doi.org/10.1111/cas.14318>. PubMed.
4. Pham, T.H., Qiu, Y., Liu, J., Zimmer, S., O'Neill, E., Xie, L., et al. (2022). Chemical-induced Gene Expression Ranking and its Application to Pancreatic Cancer Drug Repurposing. *Patterns* *3*, 100441-1–100441-12.
5. Eyres, M., Lanfredini, S., Xu, H., Burns, A., Blake, A., Willenbrock, F., Goldin, R., Hughes, D., Hughes, S., Thapa, A., et al. (2021). TET2 Drives 5hmc Marking of GATA6 and Epigenetically Defines Pancreatic Ductal Adenocarcinoma Transcriptional Subtypes. *Gastroenterology* *161*, 653–668.e16. <https://doi.org/10.1053/j.gastro.2021.04.044>.
6. Zhu, J., Wang, J., Wang, X., Gao, M., Guo, B., Gao, M., Liu, J., Yu, Y., Wang, L., Kong, W., et al. (2021). Prediction of drug efficacy from transcriptional profiles with deep learning. *Nat. Biotechnol.* *39*, 1444–1452. <https://doi.org/10.1038/s41587-021-00946-z>. PubMed.
7. Shimizu, H., Kodama, M., Matsumoto, M., Orba, Y., Sasaki, M., Sato, A., Sawa, H., and Nakayama, K.I. LIGHTHOUSE illuminates therapeutics for a variety of diseases including COVID-19. Preprint at bioRxiv. 10.1101/2021.09.25.461785.
8. Szklarczyk, D., Santos, A., von Mering, C., Jensen, L.J., Bork, P., and Kuhn, M. (2016). STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data. *Nucleic Acids Res.* *44* (D1), D380–D384. <https://doi.org/10.1093/nar/gkv1277>. PubMed.
9. Huang, L., Guo, Z., Wang, F., and Fu, L. (2021). *KRAS* mutation: from undruggable to drug-gable in cancer. *Signal Transduct. Target. Ther.* *6*, 386. <https://doi.org/10.1038/s41392-021-00780-4>. PubMed.
10. Connor, A.A., and Gallinger, S. (2022). Pancreatic cancer evolution and heterogeneity: integrating omics and clinical data. *Nat. Rev. Cancer.* <https://doi.org/10.1038/s41568-021-00418-1> PubMed.