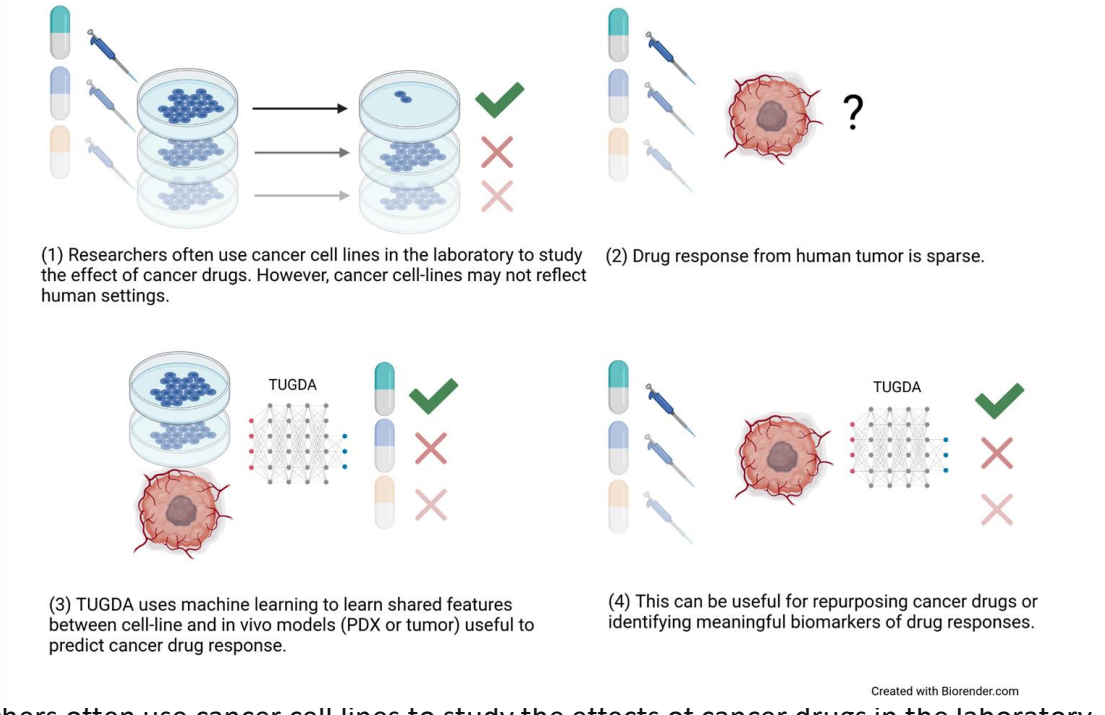
MACHINE-LEARNING METHOD TO TRANSFER KNOWLEDGE FROM CELL-LINE MODELS TO PATIENTS AND PREDICT CANCER DRUG RESPONSE



Researchers often use cancer cell lines to study the effects of cancer drugs in the laboratory. While some core biological processes are conserved, cell lines are not expected to capture key aspects relevant to human tumour response (e.g, tumour heterogeneity and microenvironment), and this imposes limitations in the translation of knowledge to patients. Despite the availability of sufficient data and models to predict drug response using cell lines, the utility in clinical settings is still limited.

To address this gap, we leverage a machine learning paradigm known as transfer learning, which seeks to extrapolate knowledge across domains. We developed a new approach called TUGDA, which unifies information available across various cancer drugs and different biological settings (cell lines, mouse models and patient tumour data).

Extensive benchmarking of TUGDA against state-of-the-art approaches showed that TUGDA is more robust in transferring knowledge across domains. It can simultaneously predict response for multiple drugs with the ability to identify shared features between *in vivo* and *in vitro* data that are predictive of drugs mechanisms. TUGDA relies on a set of more realistic assumptions (e.g, assumes that some drugs working on cell lines might not work in humans) that can overcome previous methods. It can be used by biologists and clinicians for repurposing cancer drugs or identifying meaningful biomarkers of drug responses.

TUGDA thus brings us one step closer in a vision of integrated datasets across drugs and biological domains to robustly guide cancer treatments, and identify new mechanisms and biomarkers. The initial version of these ideas was accepted in 2020 in the premier artificial intelligence conference AAAI, and its extension and expanded validation was recently accepted at top-tier computational biology conference ISMB (journal version of the paper is made available at *Bioinformatics*).

This work was led by Mr. Rafael Peres da Silva (Ph.D. Student at Department of Computer Science at the National University of Singapore, and affiliated with GIS) in collaboration with Dr. Chayaporn Suphavilai (Postdoctoral Fellow at GIS) and Dr. Niranjan Nagarajan (Associate Director and Senior Group Leader in GIS, and Associate Professor in the Department of Medicine and Department of Computer Science at the National University of Singapore).