Project summary

My project is concerned with multi-omics data integration in cancer for the purpose of drug response modelling and prediction. The multi-omics include gene expression, mutation and copy number variants corresponding to publicly available cancer cell lines (CTRP, GDSC, CCLE). We will combine these with patient tumor samples from TCGA as well as tumor samples collected at Mount Sinai (e.g.., from ovarian and myeloma cancer patients). The data processing pipeline will include multiple steps of data filtering scaling and normalization, computation of kernel matrices between samples and subsequently the training of a multi-label classifier to model the relationship between samples’ omics data and their response to a battery of compounds and CRISPR screens. The training phase will be followed by prediction of novel sample-drug interactions as well as prediction of missing drug responses for the available samples. We will also carry feature selection to determine the major contributors to drug response.

Project significance

For many cancers, resistance to established chemotherapy is on the grow and there is an urgent need for new compounds and especially the repurposing of existing cancer drugs for new targets. By some estimates 1 in 5 existing cancer drugs could be effective in cancers others than the ones they are currently used for. Repurposing requires a deep understanding of the molecular underpinnings of cancer by computationally exploiting the wealth of omics data already collected and available as well as the results of hundreds of in vitro compound screens.

Computational approach

This is a multi-step computational pipeline using state-of-the-art statistical and machine learning tools for data filtering, normalization, kernel matrix computations as well as state-of-the-art classifier algorithms to model the response of cancer samples to hundreds of compounds.