Interpretable AI identifies protein complexes to predict chemoresistance

Xiaoyu Zhao[1], Akshat Singhal[2], Sungjoon Park[1], JungHo Kong[1,3],
Robin Bachelder[1], Trey Ideker[1,2,3,4,*]

[1] Division of Human Genomics and Precision Medicine, Department of Medicine, University of California, San Diego, La Jolla, California

[2] Department of Computer Science and Engineering, University of California, San Diego, La Jolla, California

[3] Moores Cancer Center, School of Medicine, University of California, San Diego, La Jolla, California

[4] Department of Bioengineering, University of California, San Diego, La Jolla, California

[*] Corresponding author: Trey Ideker (T.I.); E-mail: tideker@ucsd.edu

Abstract

DNA replication is central to cell division and proliferation, involving closely orchestrated functions among hundreds of proteins. Although the replication machinery is highly accurate, it faces challenges from both extrinsic and intrinsic factors. These challenges can result in stalled replication forks, occurrence of DNA breaks, reduced replication precision, and other factors collectively known as RS. A multitude of cancer therapeutics leverage replication stress to eliminate cancer cells, using a variety of replication stress-inducing (RSi) mechanisms. These include classic chemotherapeutic agents that directly affect DNA integrity (e.g., cisplatin and gemcitabine) and agents that target DNA polymerases or DNA damage response proteins (e.g., CD437 and olaparib). Treatment outcomes vary widely due to significant differences in drug sensitivity and resistance across tumors, motivating efforts to better understand response mechanisms. These efforts have led to an expanding catalog of genetic alterations associated with sensitivity or resistance to RSi drugs, including mutations in *BRCA1*, *BRCA2*, *ATM*, or *ATR*. This list is almost certainly not exhaustive, because the vast majority of genetic alterations identified are rare rather than common in tumors. Moreover, it is unclear how these multiple single-gene effects integrate to constitute an overall drug response. Here we develop and evaluate interpretable deep learning models of RSi drug response. Starting from the genetic alterations detected in a tumor sample, the models predict the sensitivity or resistance to specific RSi drugs. The architecture of these models is guided by the knowledge of cancer protein complexes in a cancer cell map called Nested Systems in Tumors (NeST), meaning their predictions can be interpreted mechanistically, and both single-drug and multitask (multidrug) models are evaluated. These models identify 41 molecular assemblies that integrate alterations in hundreds of genes for accurate drug response prediction. These cover roles in transcription, repair, cell-cycle checkpoints, and growth signaling, of which 30 are shown by loss-of-function genetic screens to regulate drug sensitivity or replication restart. The model translates to cisplatin-treated cervical cancer patients, highlighting an RTK–JAK–STAT assembly governing resistance. This study defines a compendium of mechanisms by which mutations affect therapeutic responses, with implications for precision medicine.