Drug target prediction for colorectal cancer by combining ontology and network approaches

Cui Tao[1]*, Jingchun Sun[1]*, W.Jim Zheng[1], Junjie Chen[2], Hua Xu[1#]

[1]Center for Computational Biomedicine, School of Biomedical informatics, University of Texas Health Science Center at Houston and [2]Department of Experimental Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, Houston, TX 77030, USA

Drug discovery is a time-consuming and expensive process, especially for complex diseases. In the last decade, target-based methods for drug discovery have become more common and effective comparing to traditional observation-based drug discovery. Recently, computational approaches for target prediction and drug repurposing have become more common and effective compared to traditional observation-based drug discovery. However, since the data about underlying molecular mechanisms of drugs distribute among different knowledge domains and different databases, it is very challenging to design effective strategies to discover novel drug targets and propose successful drug repurposing. To alleviate this problem, we propose a computational framework to integrate complex relationship among different types of data and infer the potential drug targets by using the semantic web technology, and to improve performance through network neighborhood effect modeling. In this study, we utilize the colorectal cancer (CRC) as a proof-of-concept use case to evaluate the approach.

We first constructed a CRC ontology including drugs, diseases, genes, pathways, SNPs, and their relations from the PharmGKB. The PharmGKB is a pharmacogenomics knowledge resource, which collects, curates, and disseminates knowledge about the impact of human genetic variation on drug responses through the following activities. On top of this PharmGKB ontology, we further specified drug target prediction ontology for CRC. A new CRC Drug class has been created to serve as the basis of our drug target inference. We further specified OWL DL (Description Logic) rules to infer possible CRC drug target genes. Starting from eight FDA-approved CRC drugs and the CRC ontology, we inferred 113 potential CRC drug targets. To prioritize the most promising targets among the ontology-driven CRC potential drug targets, we utilized their relationships with CRC-associated genes in the context of the human protein-protein interaction (PPI) network. Starting from these 113 potential targets, we ranked them based on the fraction of CRC-associated genes in their neighborhood at the first, second, and third shortest-path distances and then integrated the three sets of ranking scores using a robust rank aggregation (RRA) method. These CRC-associated genes were collected from the Cancer Gene Census, the Online Mendelian Inheritance in Man (OMIM), and the Genetic Association database (GAD). Among the 113 genes, 15 were selected as the promising drug targets based on their neighborhood of CRC disease genes in the context of one human PPI network. For example, three of them encode known CRC drug targets (*EGFR*, *TOP1*, and *VEGFA*). EGFR is targeted by the drugs cetuximab and panitumumab, TOP1 is targeted by the drug irinotecan, and VEGFA is targeted by the aflibercept and bevacizumab. Additionally, CCND1 (cyclin D1) is targeted by the drug arsenic trioxide, which is used to treat leukemia; and PTGS2 (prostaglandin-endoperoxide synthase 2) is targeted by multiple drugs such as lenalidomide, which is used for treating lymphoma, pomalidomide and thalidomide, both of which are used for treating multiple myeloma and other plasma cell neoplasms. Among the remaining genes, CCND1 is a well-recognized oncongene that is amplified and/or overexpressed in a substantial proportion of human cancers including colon cancer, prostate cancer and breast cancer. Therefore, it might be a promising anti-cancer therapeutic target. The gene PTGS2 encodes prostaglandin G/H synthase-2, which catalyses the first two steps in the metabolism of arachadonic acid. It is overexpressed in many types of cancer such as colon, stomach, breast, and lung. Additionally, it has three variations with pharmacogenomic significance (rs20417, rs5275, and rs689466). Therefore, its inhibition with drugs such as aspirin, celecoxib, and ibuprofen can be used in the prevention and treatment of cancer.

Therefore, in this study, we developed a unique computational framework to integrate the ontology technology and network neighborhood modeling for drug target prediction. The results demonstrate that this framework indeed identifies the novel targets. Besides, we see many opportunities to improve upon the basic design of this integration, including integrating more relationship from multiple data sources during the ontology construction, application of score strategies in the ontology reference, and the combination of more network properties into the gene ranking.