Poster Number 2

The Use of Small Training Data to Construct Multimodal Deep Learning Models for the Prediction of Solute Carrier Transporter Chemical Interactions

Sean Carey[1,2], Xia Wen[3], Oindrila Paul[3], Jacqueline B Tiley[4], Lauren M Aleksunes[3], Hao Zhu[1,2]

[1] Department of Chemistry and Biochemistry, Rowan University

[2] Center for Biomedical Informatics and Genomics, Tulane University

[3] Department of Pharmacology and Toxicology, Rutgers University

[4] Division of Pharmacotherapy and Experimental Therapeutics, University of North Carolina

Solute carrier (SLC) transporters play key roles in various physiological processes, supporting cellular homeostasis and metabolic function by mediating the cross-membrane transport of drugs and nutrients. Consequently, potentially harmful outcomes can arise when SLC-mediated transport is disrupted. Historically, machine learning (ML) approaches such as quantitative structure-activity relationship (QSAR) modeling have been successfully used to predict chemical-transporter interactions for well-known ATP-binding cassette (ABC) transporters. However, many clinically relevant SLC transporters remain understudied, such as OATP2B1, OCT3, with limited reported chemical interaction data. Data limitations, including scarce data and restricted chemical diversity, pose challenges for ML studies. In this study, we used advanced ML approaches to address conventional small-data challenges. To this end, we constructed a multimodal deep learning model integrating heterogenous chemical, protein, and experimental information to predict chemical-transporter interactions of SLC transporters with limited data. Public chemical bioactivity data against nine OATP, OCT, and OAT protein family transporters were collected. Our advanced multimodal approaches yielded comparable performance on SLC transporters with sufficient data and achieved meaningful improvements on small-data SLC transporters. External predictions of newly tested OATP2B1 and OCT3 inhibitors further illustrated the applicability of multimodal approaches for small-data tasks, correctly identifying 82% of the tested chemicals. These findings suggest that multimodal modeling strategies can facilitate the identification and evaluation of novel chemical interactions with small-data SLC transporters and additional understudied targets. This study demonstrates that advanced deep learning methodologies can achieve optimal performance on small-data targets by leveraging multimodal and multi-target data, despite scarcity in the current data landscape. Supported by NIH UC2HD113039.