Deep Learning Profile QSAR Modeling to Impute In Vitro Assay Results and Predict Chemical Carcinogenesis Mechanisms

Alexandre Borrel[1*], Agnes L. Karmaus[1], Getachew Tedla[1], Kamel Mansouri[2], Tom Luechtefeld[3], Ruth Lunn[4], Amy Wang[4], David Allen[1], Nicole Kleinstreuer[2]

[1]Inotiv, RTP, NC, United States; [2]NIH/NIEHS/DTT/NICEATM, RTP, NC, United States; [3]Insilica LLC, Bethesda, MD, United States; [4]NIH/NIEHS/DTT/IHAB, RTP, NC, United States,

**Presenting author*

Carcinogenesis is a multistep process in which normal cells acquire properties that allow them to form benign tumors or malignant cancers. These properties have been associated with 10 well-established hallmarks of cancer. The concept of key characteristics of carcinogens (KCC) has also been developed to describe 10 properties that are shared by viruses and chemicals that induce human cancers. Quantitative structure-activity relationship (QSAR) models that rely on structural or physicochemical properties to predict carcinogenesis potential endpoints usually perform poorly, likely because they lack sufficient information on the complex mechanisms involved in carcinogenicity. We combined a novel imputation profile QSAR modeling approach with modern machine learning to analyze data on 10,000 Tox21/ToxCast chemicals and 2,000 in vitro assay endpoints associated with KCC. Because limited experimental data were available, we filled data gaps by imputing assay results for the Tox21/ToxCast inventory using structural and physicochemical properties and deep learning. Imputed in vitro assay results were enriched using data in the BioBricks platform, which compiles toxicity-relevant databases into a harmonized easily accessible format. This enrichment allowed us to include additional information such as protein target binding or assay results to the model. Finally, various machine learning approaches including a multitask deep learning model were applied to predict each chemical's likelihood of inducing cancer based on the imputed in vitro data. Results included output metrics on the quality of imputation, defined by grouping of assays, and performance computed per chemical. Project was funded by NIEHS under Contract No. HHSN273201500010C.