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

A. Borrel[1*], A.L. Karmaus[1], G. Tedla[1], K. Mansouri[2], T. Luechtefeld[3], R. Lunn[4], A. Wang[4], D.G. Allen[1], N. Kleinstreuer[2]

[1]Inotiv, RTP, NC, United States; [2]NIH/NIEHS/DTT/PTB/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 various properties that allow them to form benign tumors or malignant cancers. These properties of cells have been associated with 10 well-established hallmarks of cancer (HMC). It has been further suggested that human carcinogens (e.g., chemicals, viruses) share one or more of 10 properties, namely key characteristics of carcinogens (KCC). QSAR models that rely on structural and/or physicochemical properties to predict carcinogenesis potential endpoints usually exhibit low performance, likely because they lack sufficient information on the complex mechanisms involved in carcinogenicity. We used a novel imputation profile QSAR modeling approach coupled with deep learning to analyze data on 10,000 Tox21/ToxCast chemicals and 2,000 in vitro assay endpoints subsetted by HMC and 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. In vitro assay results were enriched using data in the BioBricks platform (https://biobricks.ai/bricks/), which compiles toxicity-relevant databases into a harmonized easily accessible format. This enrichment allowed us to include additional information such as protein target binding in the model. Finally, multitask deep learning was applied to predict each chemical's likelihood of triggering cancer HMC and KCC based on the imputed and enriched in vitro data. Results included data on the quality of imputation, defined by grouping of assays, and performance computed per chemical. Project was funded by NIEHS under Contract No. HHSN273201500010C.