

Integrating Network Analysis And Machine Learning To Elucidate Chemical-Induced Pancreatic Toxicity In Zebrafish Embryos



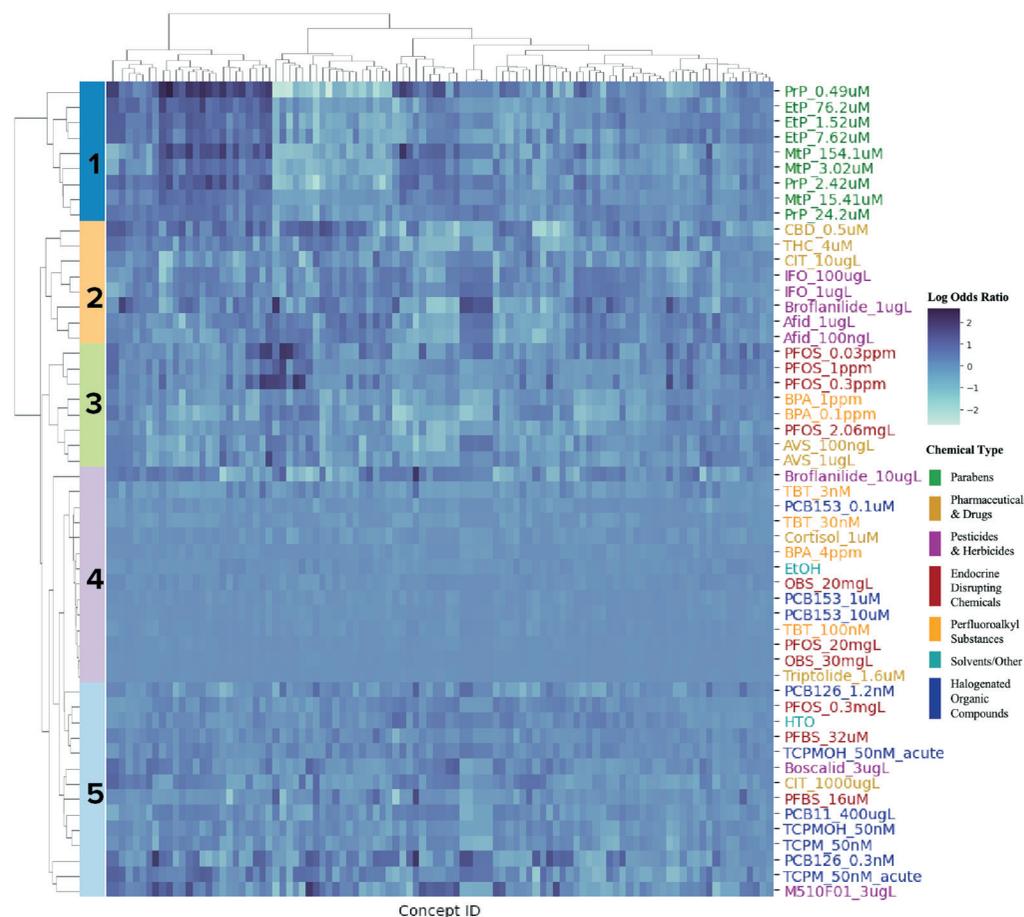
Exposure to environmental contaminants during development has been linked to birth defects and diseases later in life. To address the need for rapid toxicity assessment, zebrafish (*Danio rerio*) are used as a high-throughput model for studying embryonic development. This study focuses

on chemical-induced toxicity to the pancreas, using RNA sequencing of zebrafish embryos exposed to 53 chemical compounds, including pesticides, pharmaceuticals, and endocrine disruptors. Differential gene expression and clustering analysis identified five distinct chemical groups affecting pancreatic pathways. Gene co-expression network analysis highlighted key driver genes within these clusters, providing insights into potential biomarkers of chemical-induced pancreatic toxicity, such as disruptions in insulin regulation, glucose homeostasis, and pancreatic development. Machine learning (random forest) achieved 74% accuracy in predicting pancreatic toxicity, identifying key chemical properties influencing toxicity such as monoisotopic mass and biodegradation half-life. These findings indicate that specific chemical properties may reliably predict the characteristic of pancreatic toxicity, providing insights into how different exposures lead to similar biological outcomes. Our integrative approach combining transcriptomics, network analysis, and machine learning offers valuable insights into the relationship between chemical properties and organ-specific toxicity, with potential applications for other organ systems.

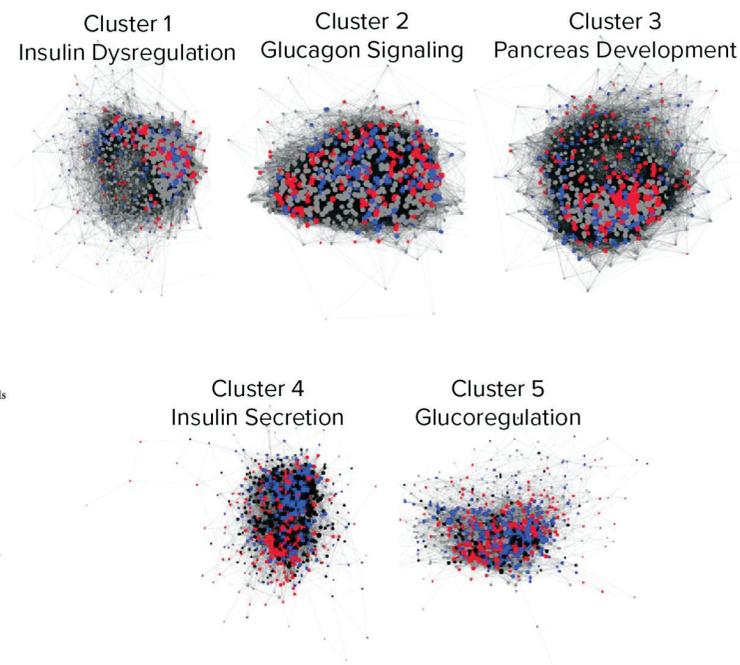
Ashley Schwartz, Karilyn E. Sant and Uduak Z. George

This research is supported by the SDSU University Graduate Fellowship, the National Science Foundation grant (DUE-1930546), the San Diego ARCS Foundation Scholar Award, the National Institute of Health (National Institute of Diabetes and Digestive and Kidney Diseases) award (1R21DK134931-01 to U.Z.G and K.E.S.), the National Institute of Environmental Health Sciences (K01ES031640 to K.E.S.), the National Science Foundation Faculty Early Career Development (CA-REER) award (DMS2240155 to U.Z.G.), and the Computational Science Research Center (CSRC) at San Diego State University

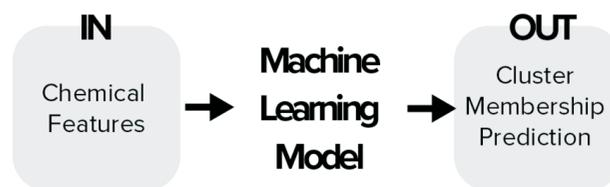
Heatmap of Zebrafish Pancreas Pathway Enrichment Response to 53 Chemical Exposures



Pancreas Gene Co-Expression Networks for each Individual Cluster



Machine Learning Model Performance and Feature Importance



The machine learning model predicts the cluster membership, and therefore the cluster impacts on the pancreas, based solely on chemical features.

Model	Accuracy		Precision		Recall	
	Mean	SD	Mean	SD	Mean	SD
Random Forest	0.74	0.19	0.69	0.25	0.74	0.19
XGBoost	0.70	0.18	0.65	0.23	0.70	0.18
Support Vector Machine	0.69	0.14	0.65	0.18	0.69	0.14
Multiclass Logistic Regression	0.73	0.16	0.70	0.19	0.73	0.16

