國家衛生研究院
National Health Research Institutes

BioCompLab.
https://cwtung.nhri.edu.tw/

ChemDIS-ZF

Inferring Potential Effects of PFASs via a Novel Chemical-phenotype Inference System ZFinfer

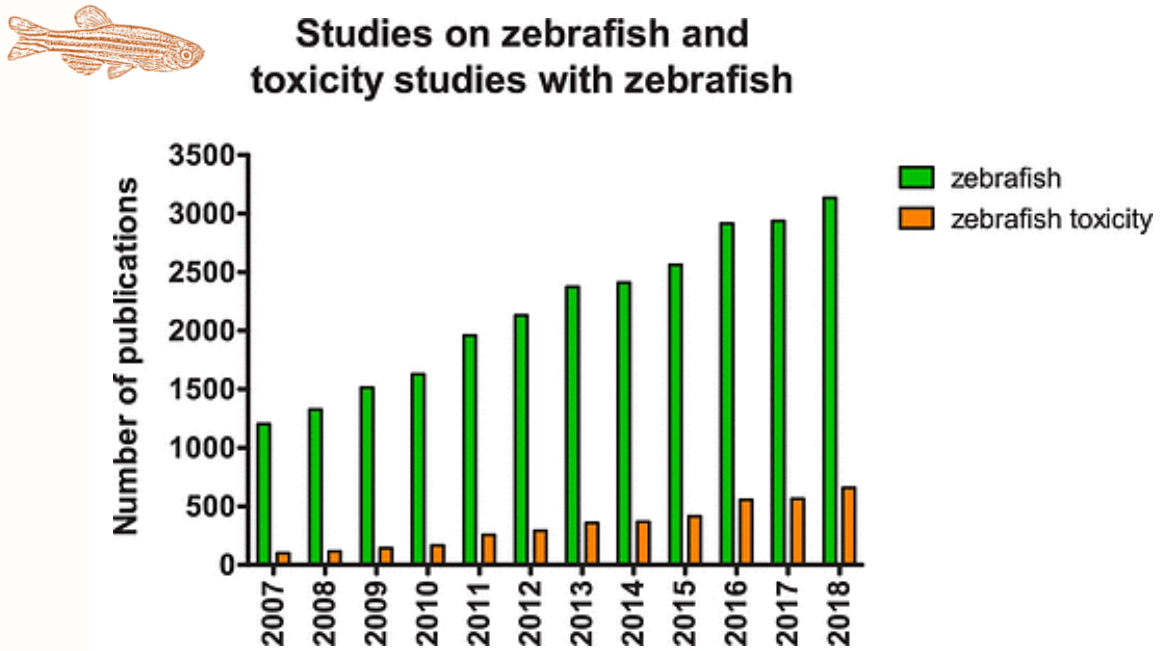
Hung-Lin Kan, Shan-Shan Wang, Chun-Wei Tung

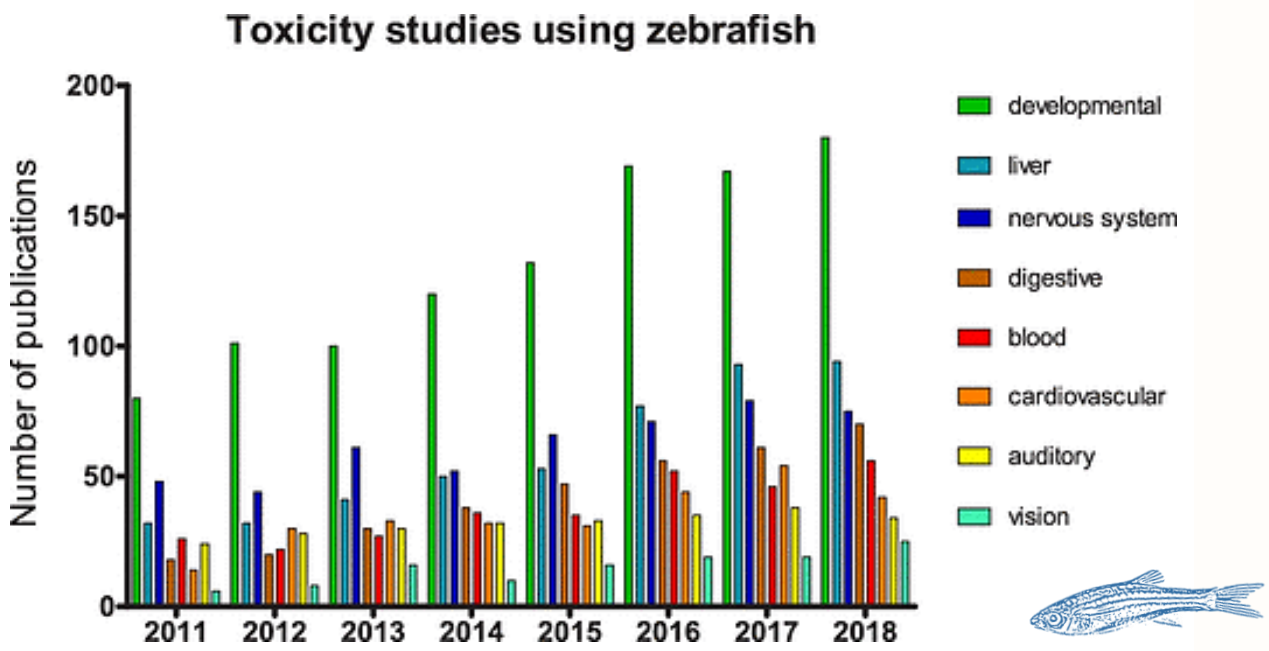
Institute of Biotechnology and Pharmaceutical Research,
National Health Research Institutes, Taiwan

Zebrafish in Drug Discovery and Toxicology Studies

- Comparison to the human reference genome shows that approximately **70%** of human genes have at least one obvious zebrafish orthologue.
- Of the genes bearing morbidity descriptions, **82%** can be related to at least one zebrafish orthologue.

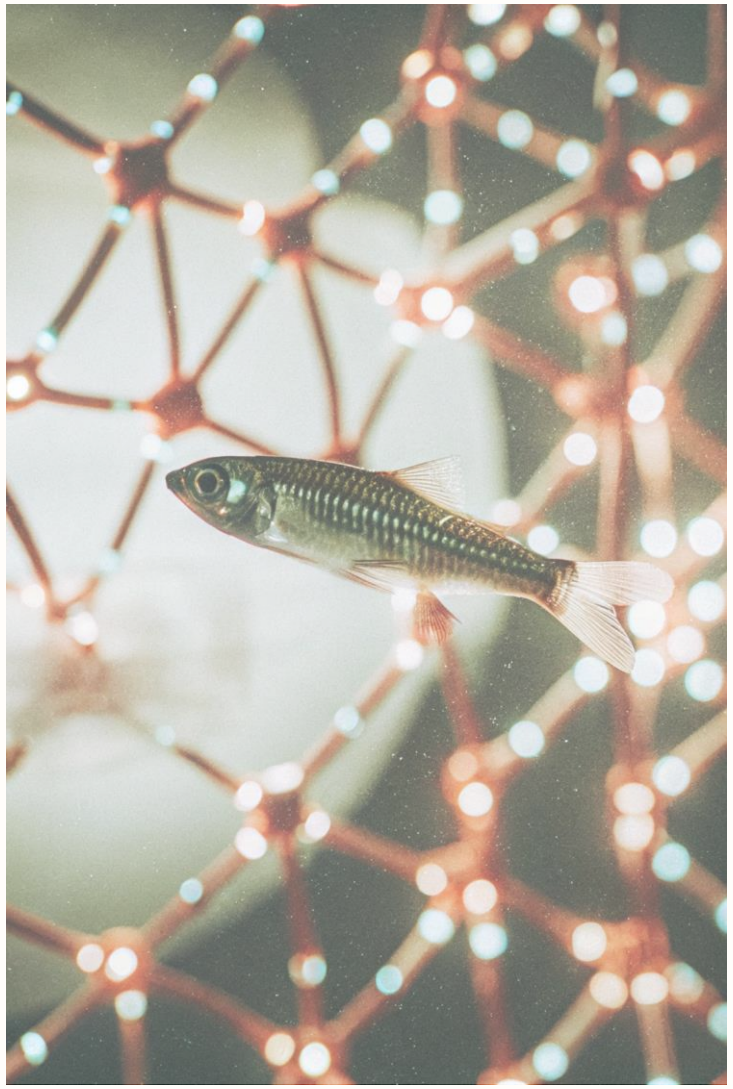
Howe, K. et al. *Nature.* **496**, 7446 (2013)





Cassar, S. et al. *Chemical Research in Toxicology.* **33**, 95–118 (2020).

Zebrafish as an Alternative Model



Zebrafish is a useful model organism for toxicological research due to their **small size**, **fast reproduction**, and **genetic similarity** to humans.

However, as environmental pollutants increase, it becomes difficult to identify all hazards using zebrafish models alone.

***In silico* models** can assist in **identifying** **chemical priorities** for further experimental evaluation and provide insights into the underlying mechanisms.

Lin *et al. Journal of Cheminformatics* (2024) 16:91
https://doi.org/10.1186/s13321-024-00891-4

Journal of Cheminformatics

RESEARCH Open Access

A novel multitask learning algorithm for tasks with distinct chemical space: zebrafish toxicity prediction as an example

Run-Hsin Lin[1,2], Pinpin Lin[3], Chia-Chi Wang[4] and Chun-Wei Tung[1,2*]

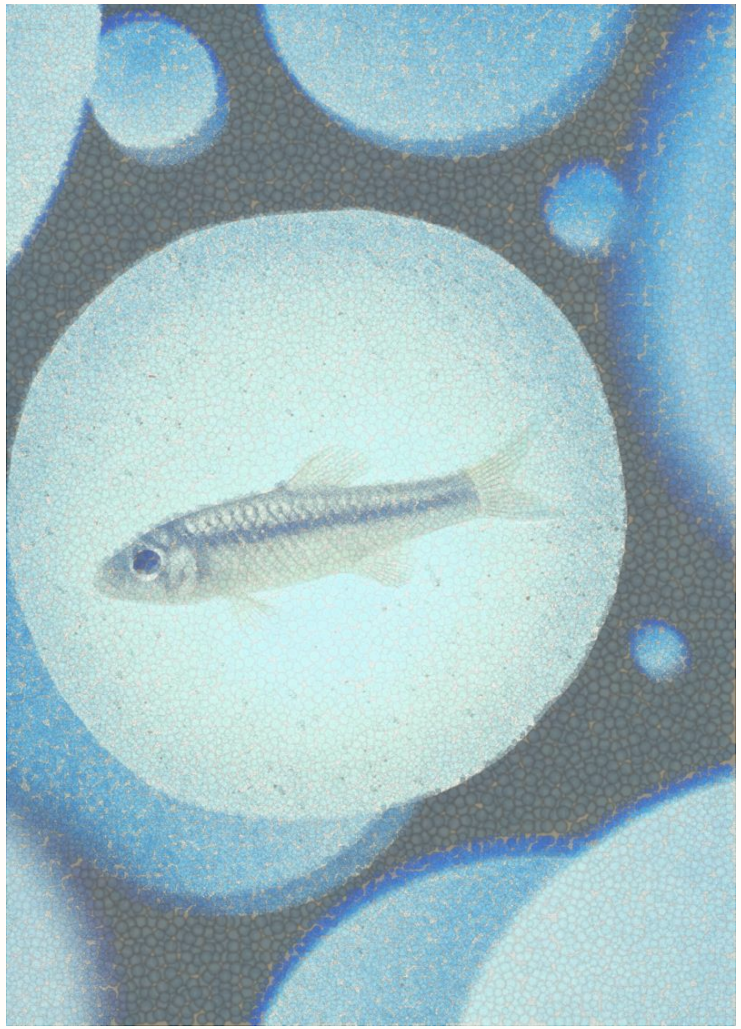
Zebrafish model

Bridging Data Gaps

The **virtual zebrafish model** speeds up **toxicity assessments** and **drug development**. It still has certain limitations, including the lack of:

- **Lack of Mechanistic Insights**
  Mechanistic insights into how chemicals influence the overall effects on zebrafish.
- **Insufficient Linkage to Toxicity Endpoints**
  A clear linkage between the overall effects caused by chemicals and specific toxicity endpoints.
- **Potential Disease Outcomes**
  Diseases that may result from toxic responses caused by chemicals.

Chemical–Disease Inference System for Zebrafish



ChemDIS-ZF: Chemical–Disease Inference System for Zebrafish

• Extensive Database

MongoDB.

Chemical-Protein Interaction data

STITCH

419,328 Chemicals
23,180 Zebrafish Proteins

Gene-Phenotype/Disease Interaction data

ZFIN

38,116 GO + Zebrafish anatomy and development ontology (ZFA)
65,800 Zebrafish Phenotype Ontology (ZP)
5,227 Disease Ontology (DO)

Gene-Pathway Interaction data

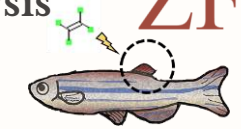
reactome KEGG

12,511 KEGG Pathways
161,192 Reactome Pathways

Gene Ontology (GO): Gene functions, including cellular component, molecular function, and biological processes.
Zebrafish Anatomy Ontology (ZFA): Zebrafish Genes to Anatomical System.
Zebrafish Phenotype Ontology (ZP): Consists of structural or process terms affected by specific genes and quality terms from the Phenotype Quality Ontology.

• Enrichment Analysis System

Phenotype enrichment analysis ZFinfer



GO analysis GENEONTOLOGY Unifying Biology

Pathway analysis reactome KEGG

Disease inference DISEASE ONTOLOGY

All the enrichment analyses were based on **hypergeometric tests** with multiple testing corrections using the **Benjamini–Hochberg method**.

System Validation (ex: Phenotype enrichment analysis - ZFinfer)

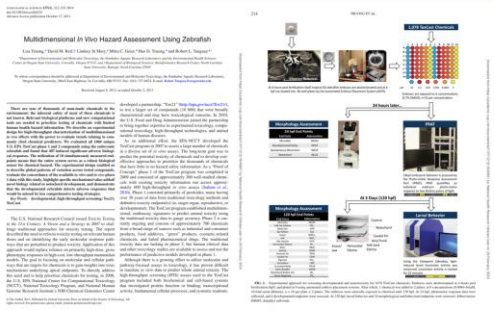
EPA United States Environmental Protection Agency

777

ToxCast Chemicals

Chemicals from the ToxCast used for validation.

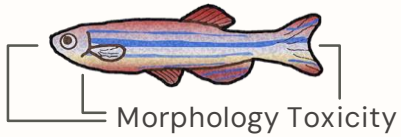
Multidimensional in vivo hazard assessment using zebrafish



Lowest Effect Level in Zebrafish

Chemical-induced Toxicity References for validation.

17 Morphology Toxicity by Lowest Effect Level



ZFinfer

Phenotype Enrichment Analysis Tool

17 Morphology Toxicity Endpoints related phenotypes

Endpoint	Description
YSE	Yolk Sac Edema
AXIS	Axis
EYE	Eye
SNOU	Snout
JAW	Jaw
OTIC	Otic
HEART	Heart
BRAIN	Brain
SOMI	Somite
PFIN	Pectoral Fin
CFIN	Caudal Fin
PIG	Pigmentation
CIRC	Circulation
TRUN	Trunk Length
SWIM	Swim Bladder
NC	Notochord Distortion
TR	Touch Response

189 Phenotypes (GO + ZFA)
3,542 Phenotypes (ZP)

Performance Measurement

- Sensitivity
- Specificity
- Accuracy
- Balanced Accuracy
- Positive Predictive Value
- Negative Predictive Value

System Performances

0.37

Average Sensitivity

Demonstrated sensitivity in 17 morphology endpoints.

- Sensitivity in critical morphological endpoints

0.72

Demonstrated average sensitivity in Heart, Trunk length, Eyes, and Brain.

- Sensitivity in different toxicity levels

0.49

Demonstrated average sensitivity in Strong Toxicity (<1 μM).

- System Validation by ECOTOX Knowledgebase Data

EPA United States Environmental Protection Agency

51

Priority Pollutants

Pollutant chemicals from the USEPA priority pollutant list used for validation.

ECOTOX Knowledgebase

Chemical Toxicity Records

93%

Rediscovery Rate

Rediscovery rate in inferring relevant phenotypic effect groups from ECOTOX Knowledgebase.

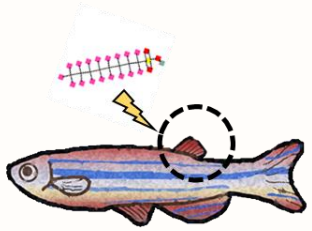
System Applications (ex: Phenotypes Inference of Environmental Pollutants)

Per- and poly-fluoroalkyl substances (PFAS) are widely used in various industries, and studies show that PFOS and PFOA can be toxic, but the effects of other PFAS are not well understood.

PFAS

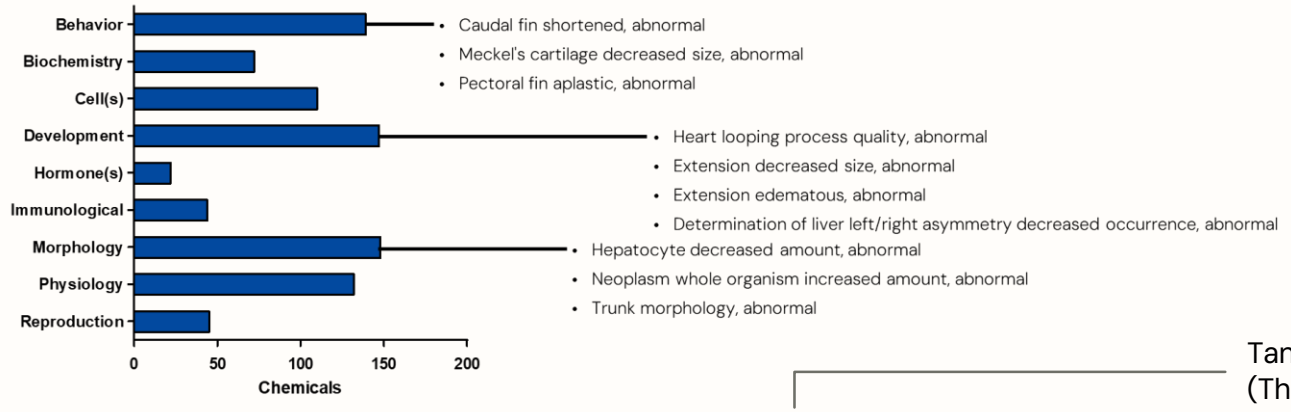
EPA PFAS Master List
(12,034 chemicals)

ZFinfer
Phenotype Enrichment Analysis Tool



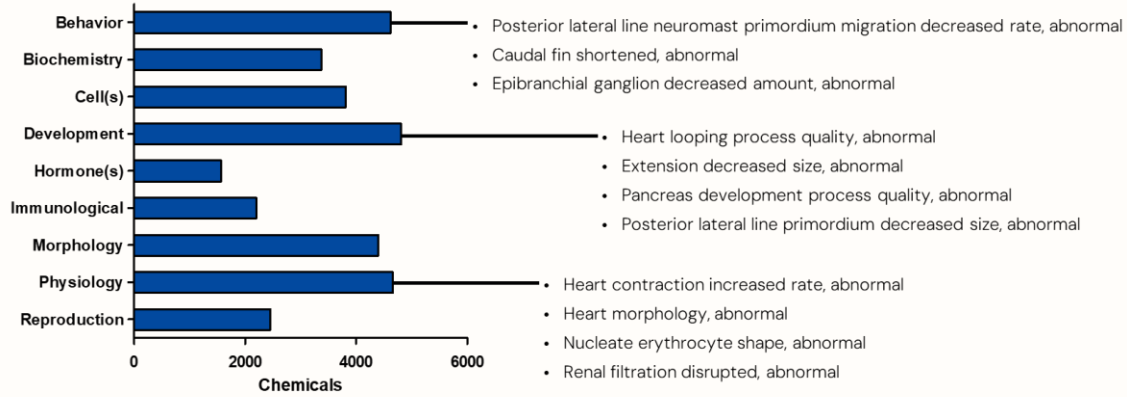
PFAS-induced Zebrafish phenotypes

164 chemicals (Phenotypes inferred by known protein-chemical interaction data)



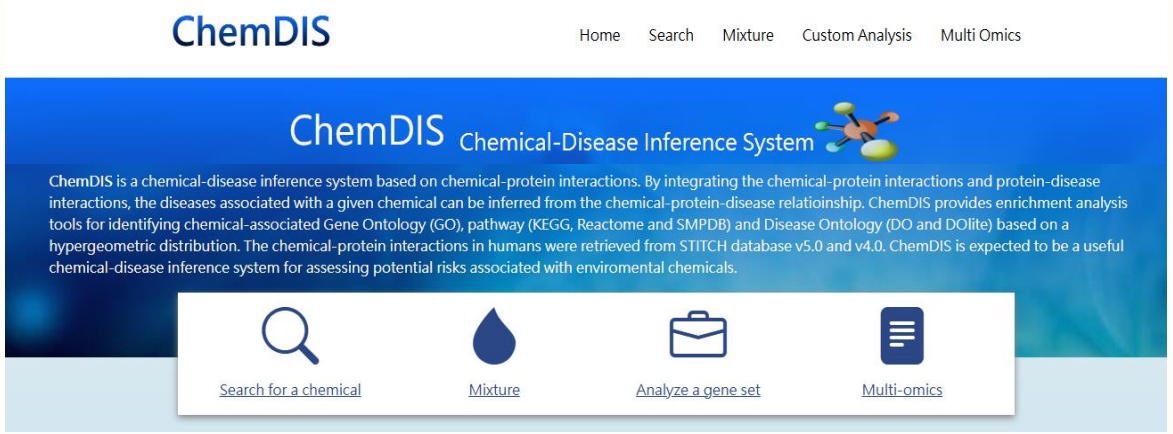
Tanimoto index on FP2
(Threshold: 0.65)

5,031 chemicals (Phenotypes inferred from Molecular Similarities of known chemicals)

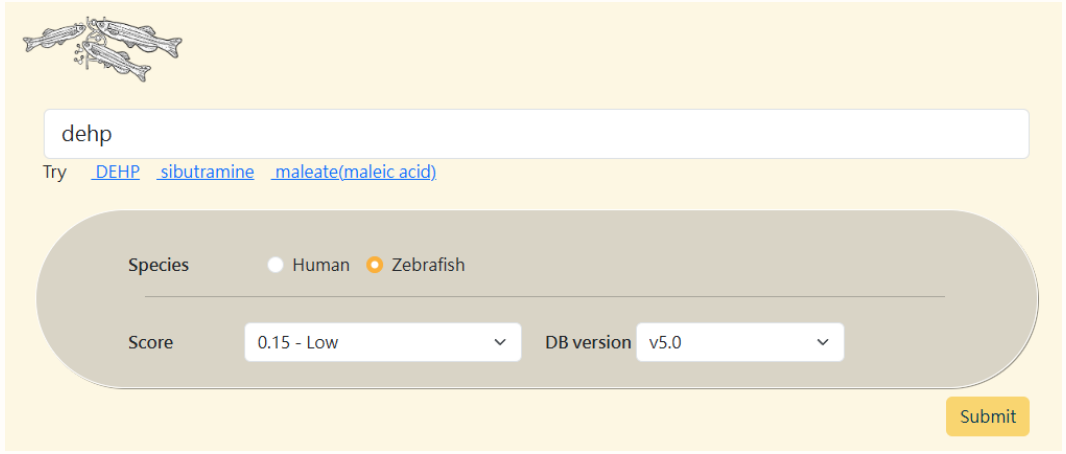


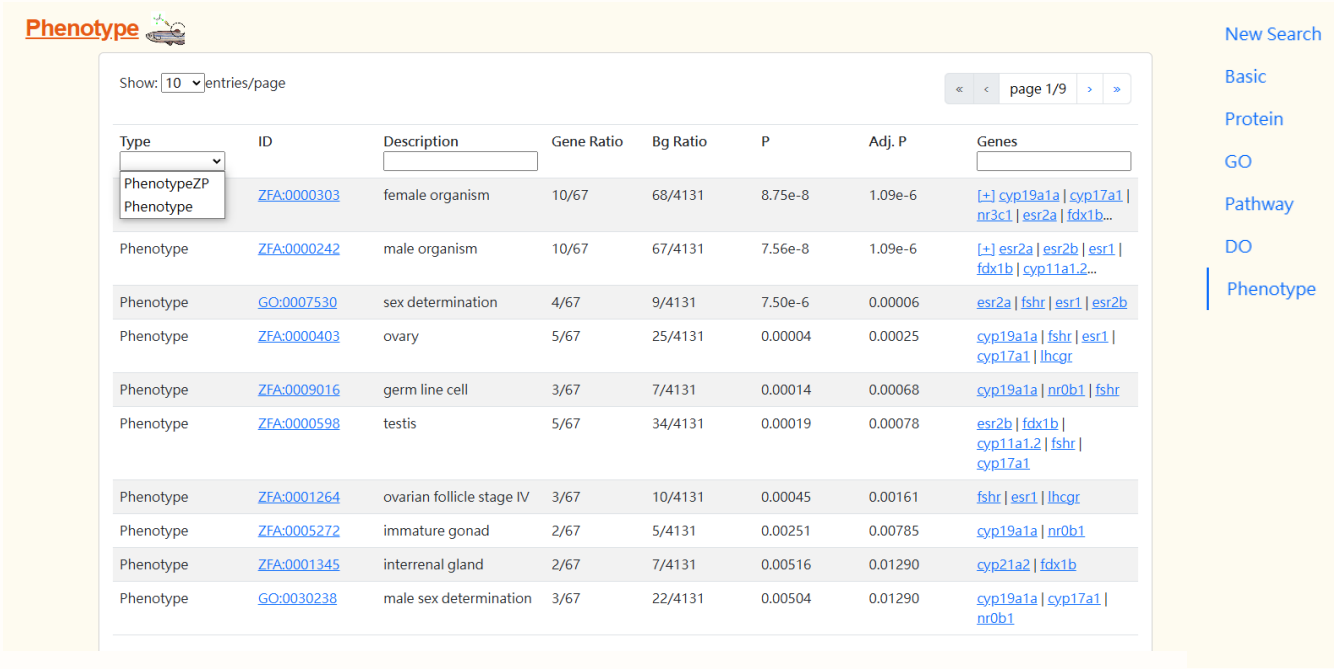
Web Tool

- ChemDIS Website



- Chemical Search





New Approaches: Mixtures

The **mixture** of chemicals raises concerns about **health** and **environmental effects**, as neglecting these interactions could result in underestimating risks.

Drakvik, E. et al. *Environment international.* **134** (2020)

Current **regulations** on unintentionally produced mixtures and emerging contaminants are **unclear** due to the random and unknown nature of these mixtures. To effectively address these challenges, we need to:

1. Improve our **understanding** of potential exposures to chemical mixtures in the environment.
2. Improve the **analysis** of toxicological **mechanisms** and **interaction** patterns of single and mixtures.
3. Identify specific mixtures that are required to be addressed with **priority**.

European Commission and Joint Research Centre (2018)

System Updates (In progress)

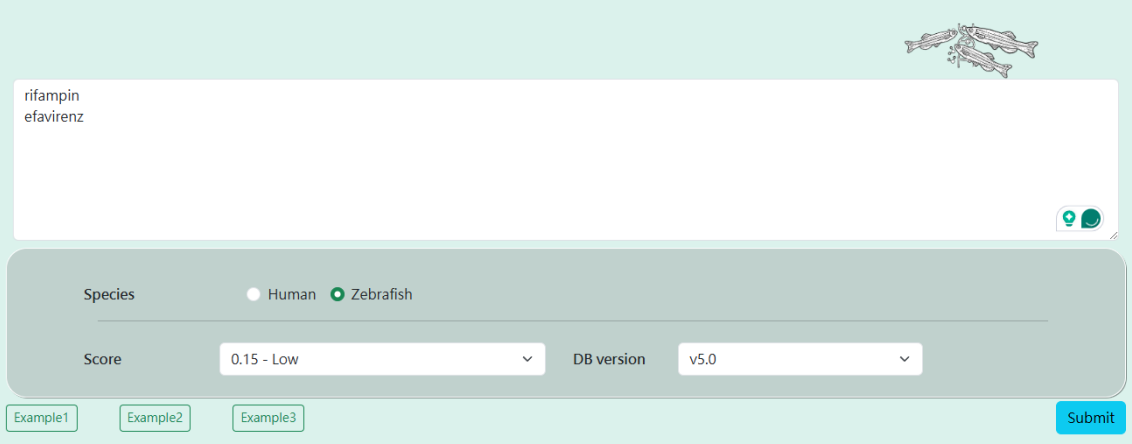
ChemDIS-ZF 2.0

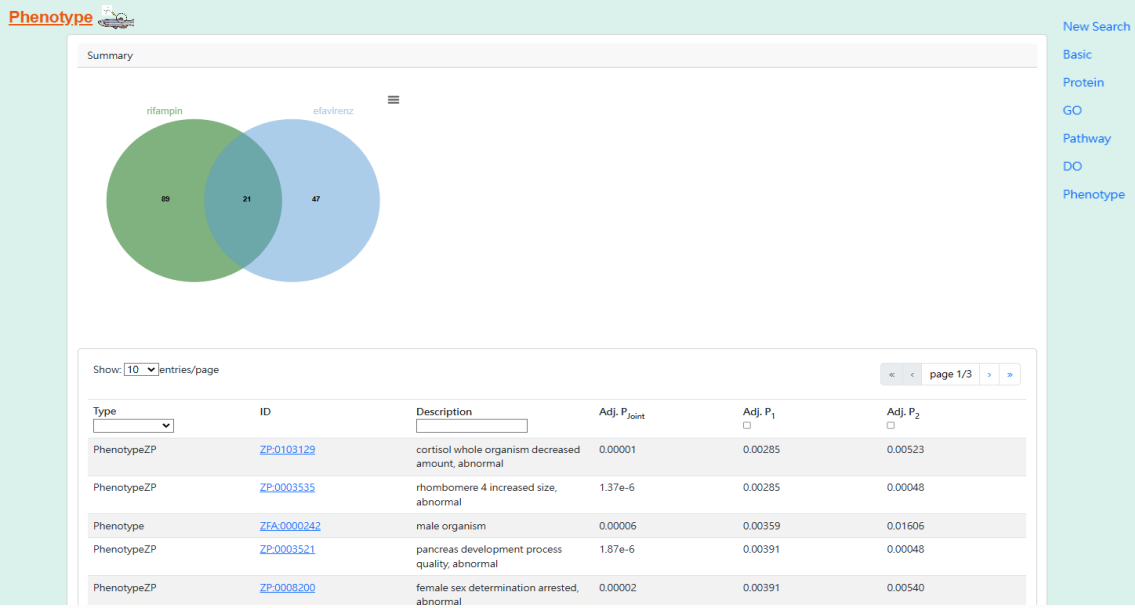
Toxicogenomic-based Mixture Effect Analysis System for Zebrafish

- **Mixture Effect Analysis**

Mixture effect enrichment analysis will be evaluated by using the overall and overlapped genes that interact with chemicals in mixtures.

- **Mixture chemicals Search**





Thanks for your attention

BioCompLab.
https://cwtung.nhri.edu.tw/

Acknowledgement

This study was supported by the National Science and Technology Council (formerly Ministry of Science and Technology) of Taiwan (MOST-110-2221-E-400-004-MY3, NSTC-112-2811-E-400-001, NSTC-113-2811-E-400-002).

Members involved in this research project:

Principal Investigator
Chun-Wei Tung, Ph.D.
Email: cwtung@nhri.edu.tw

Postdoctoral Research Fellow
Hung-Lin Kan, Ph.D.
Email: hlkan0422@nhri.edu.tw

Research Assistant
Shan-Shan Wang, M.S.

Research Assistant
Run-Hsin Lin, M.S.