PRECISION
TOX

Hazard
Using New Approach Methodologies

Jiarui (Albert) Zhou | University of Birmingham

ASPIS

UNIVERSITY OF BIRMINGHAM

ECETOC Workshop on Quantitative Response-Response Relationships (qAOPs)

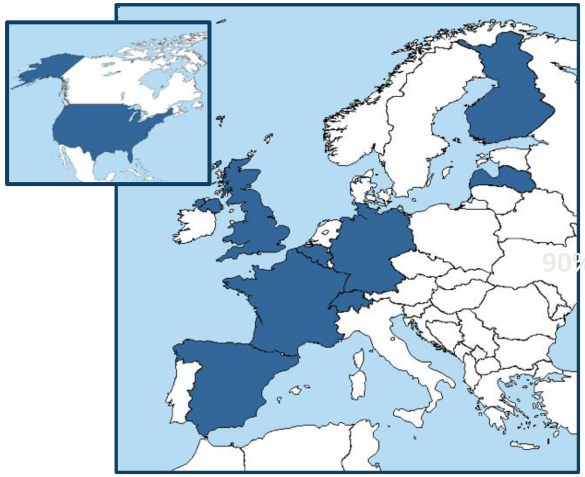
18-19 October 2022, Brussels & Online

Precision Toxicology

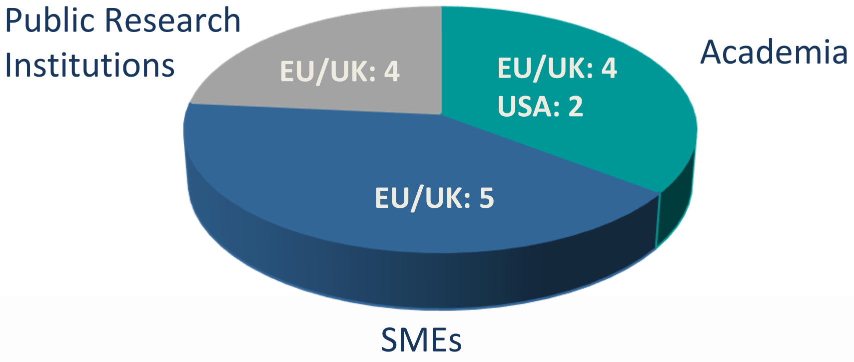
- Aims at better protecting the health of people and the environment by establishing **new approach methodologies (NAMs)** for chemical safety testing
- Establishes a new regulatory paradigm of detecting toxicity using molecular biology for greater certainty at predicting which chemicals cause harm while avoiding traditional animal testing
- Accomplish this goal by identifying **molecular key event (KE) biomarkers**, predictive of chemically induced adverse health effects in humans, that feed directly into regulatory and industry practice

Participants of *PrecisionTox*

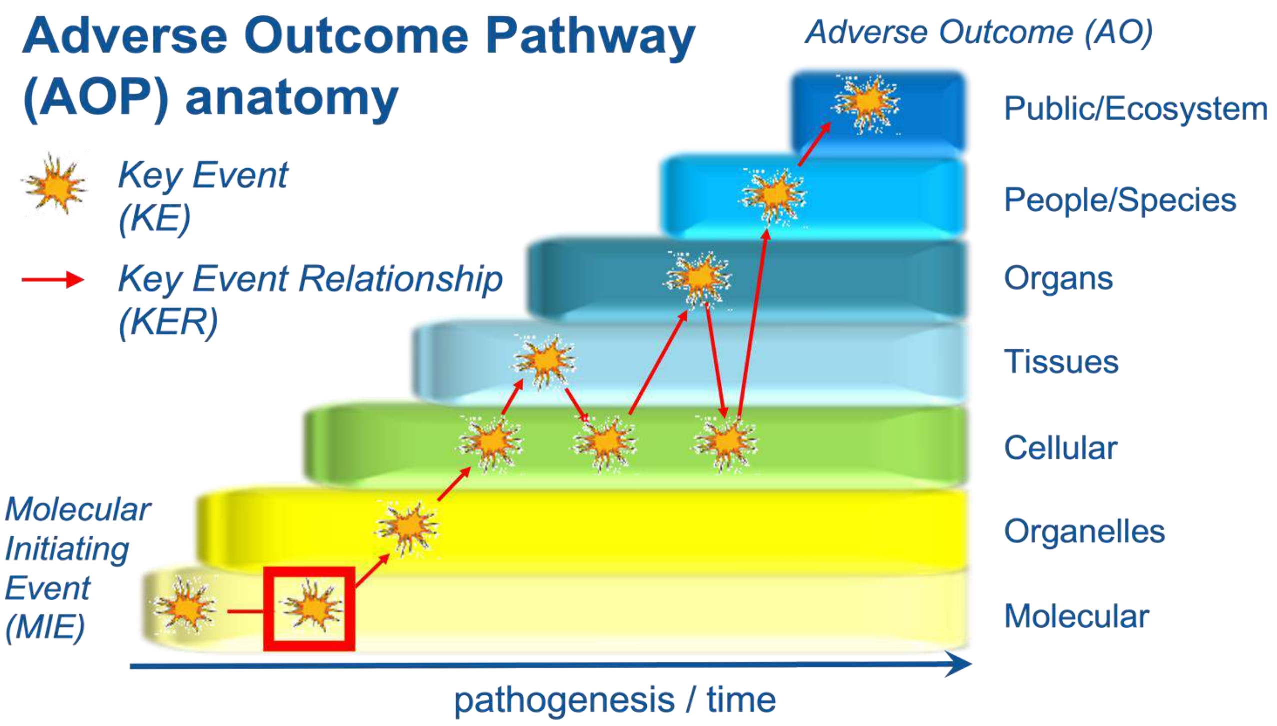
15 participating organizations across 8 countries



Distribution of participants



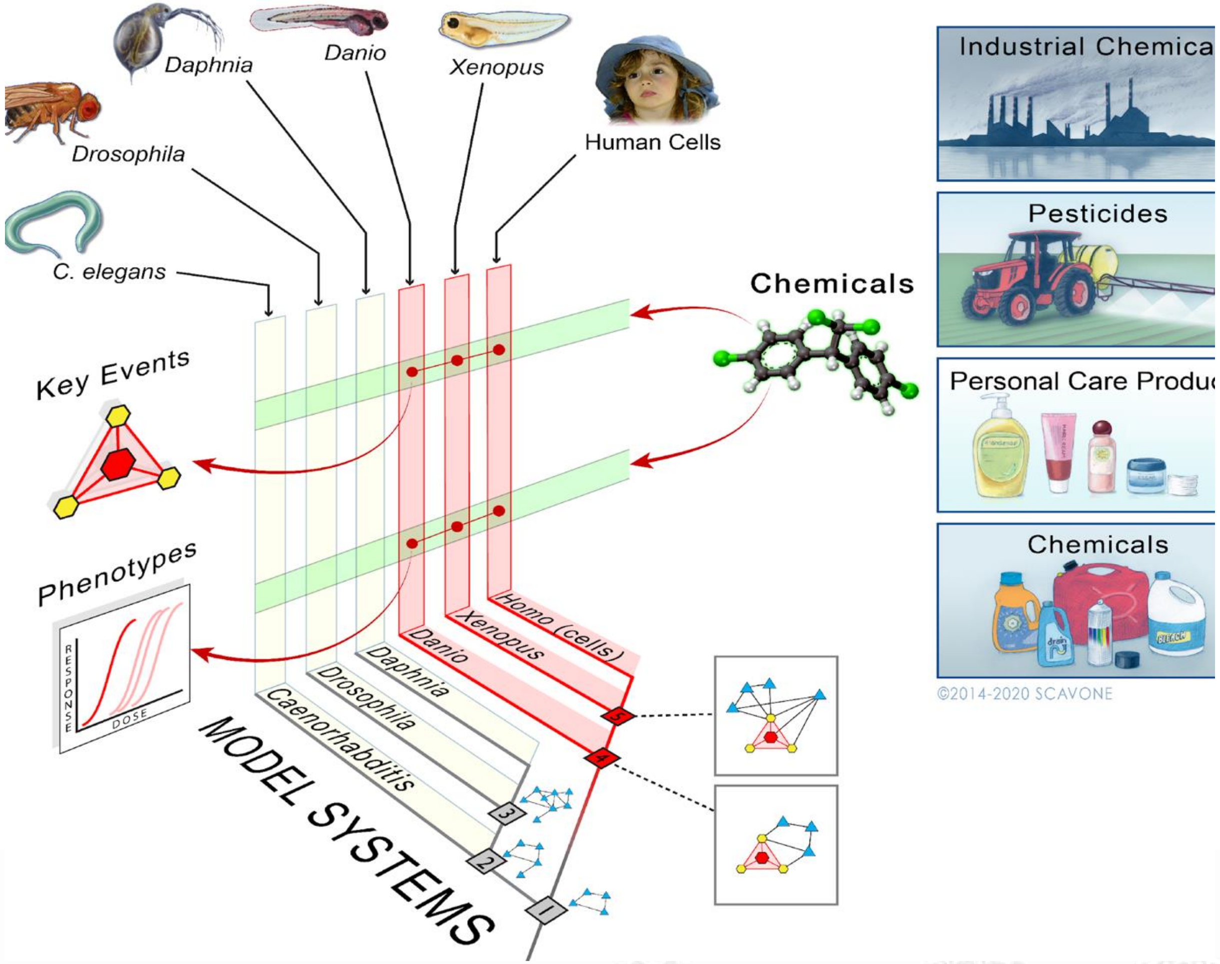
Molecular Key Event Biomarkers



- **mKE biomarkers** are quantifiable molecular indicator of a toxicological response predictive of the adverse outcome induced by exposure to a chemical substance
- Discovered via *in-silico* molecular biomarkers defined as a sparse network of interacting genes and their metabolic products

Linking Human Toxicology and Ecotoxicology

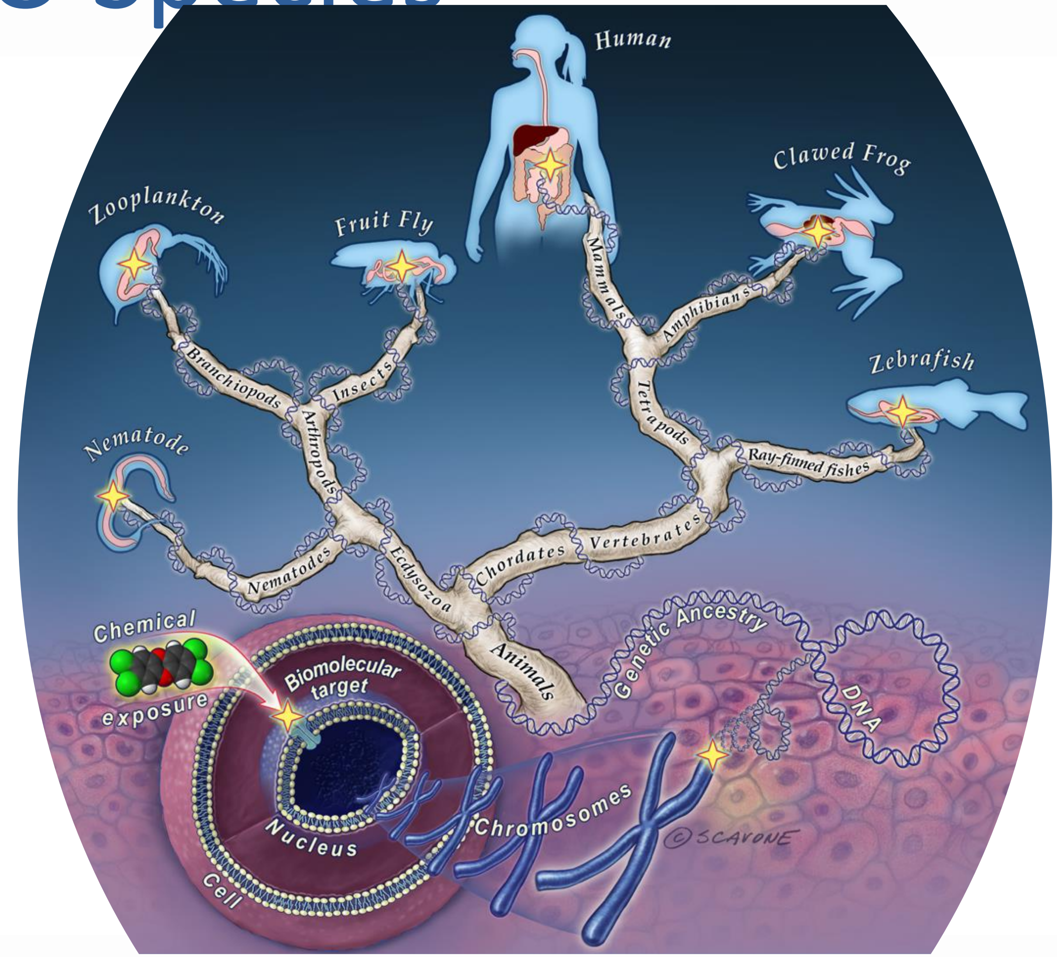
- **Challenge**: whole-organism testing is crucial, but no single model is a perfect human surrogate.
- **Solution**: evolutionarily diverse non-sentient organisms plus human cell line (PhyloToxicology)
    - Evolutionary origins of the interactome including toxicology-relevant networks
    - Induced by chemicals (chemical responsiveness)
    - Indicative of similar adversity (hazard relatedness)
    - Shared by multiple species (evolutionarily conserveness)
- Identification of molecular biomarkers of a chemical hazard



©2014-2020 SCAVONE

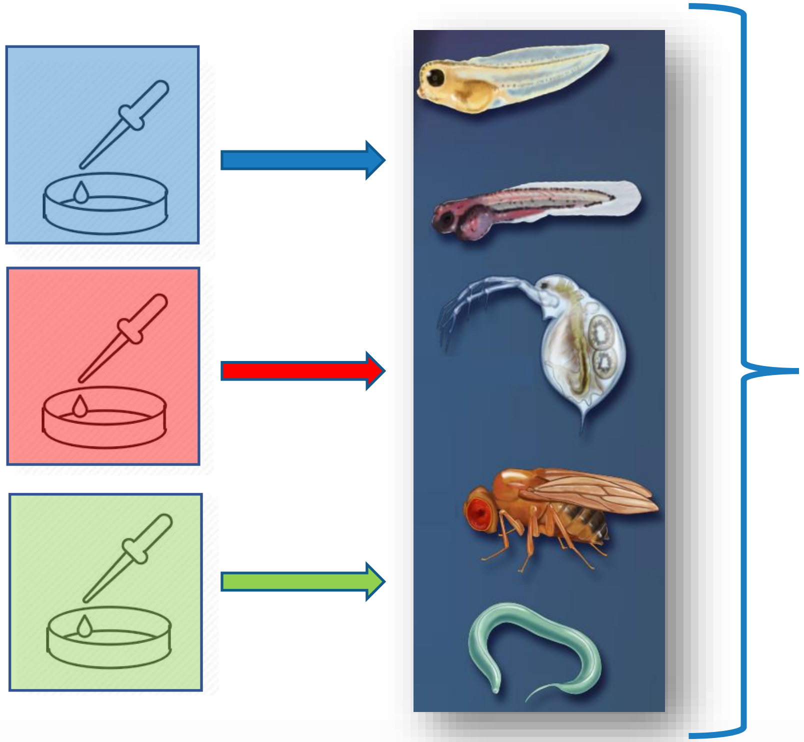
Replacing Mammals by NSO Species

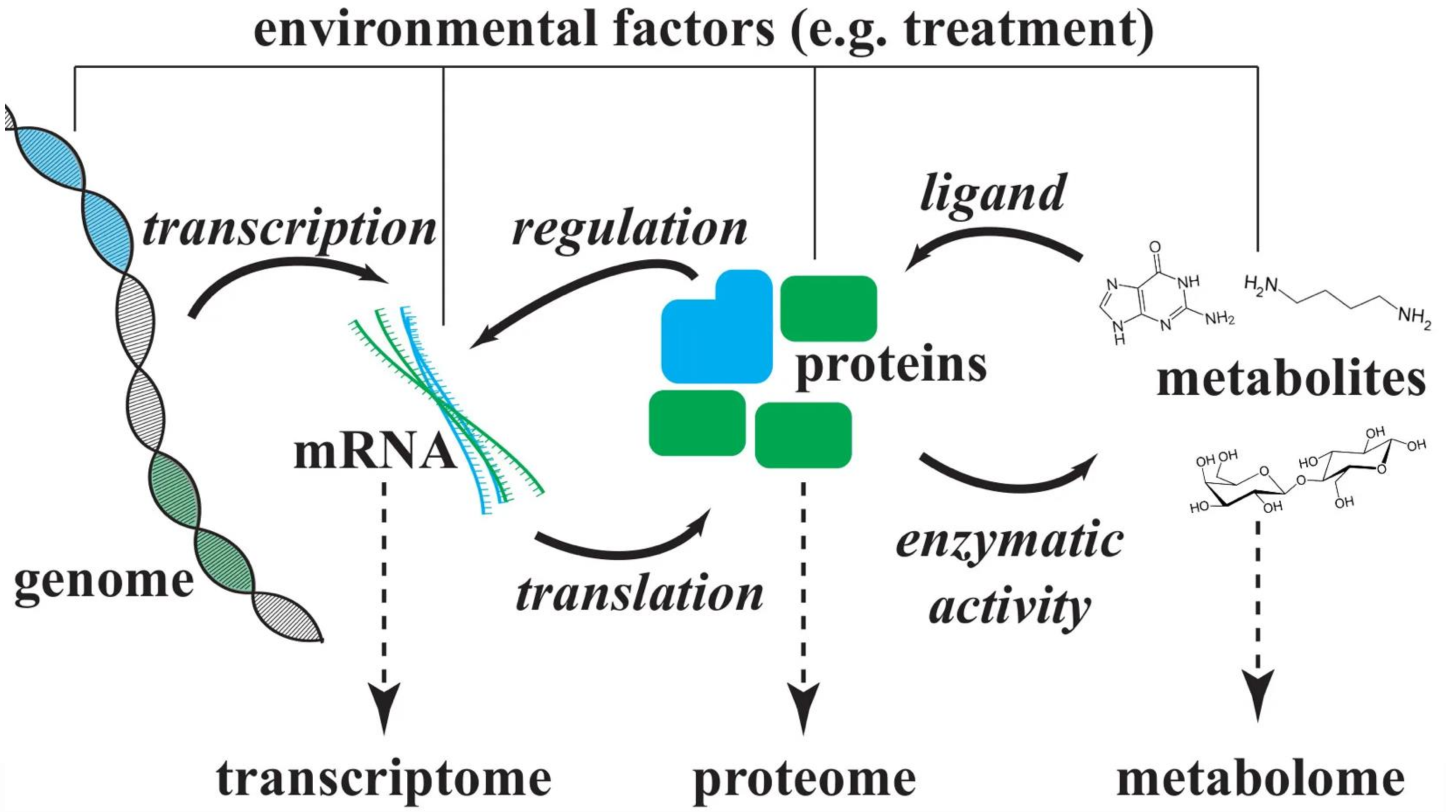
- Systematic use of distantly related non-sentient organism (NSO) species from across the tree of life
  - *Daphnia magna* (water flea)
  - *Drosophila melanogaster* (fruit fly)
  - *Caenorhabditis elegans* (nematode)
  - *Danio rerio* (zebra fish; embryo)
  - *Xenopus laevis* (frog; embryo)
  - *Homo sapiens* (human; cell-line)
- 250 21,00chemicals from various chemical classes, producing 0 samples for transcriptomics and metabolomics
- NAMs for chemical toxicity testing: combination of machine learning paradigms for multi-omics integration and mKE biomarkers identification



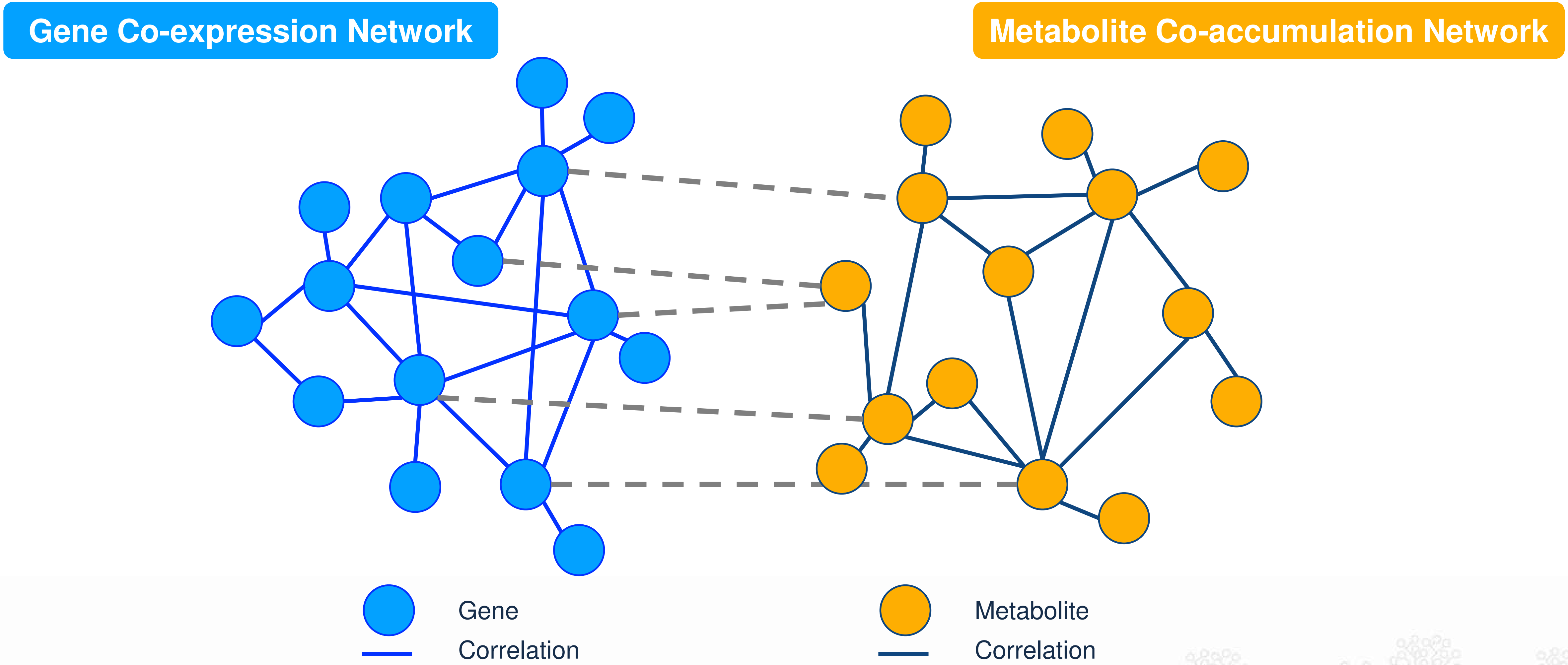
Step 1: Multi-Omics Molecular Data Production

250 Chemicals **NSO Species**

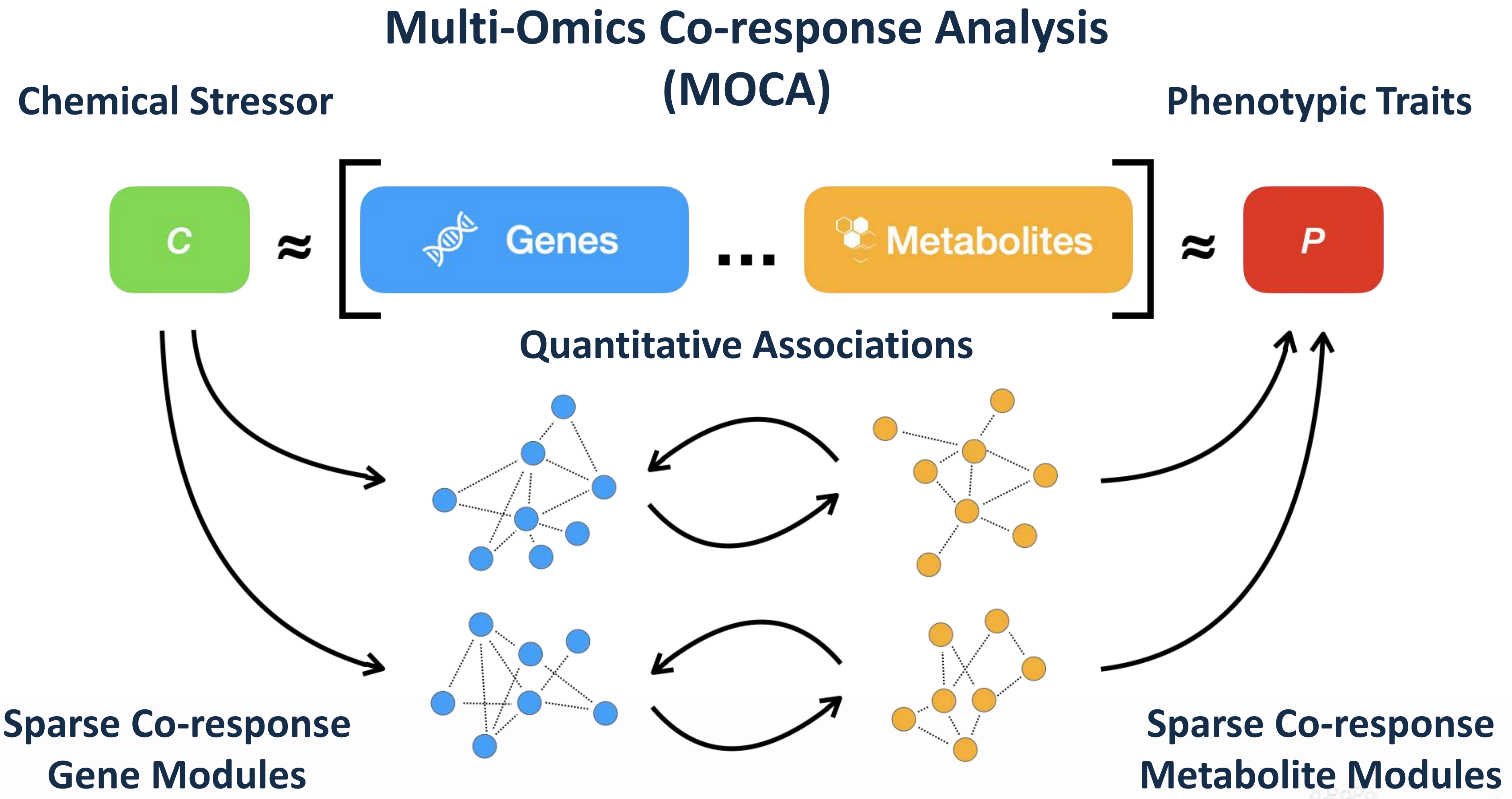




Step 2: Biomolecular Network Construction



Step 3: Association Quantification



Multi-Omics Co-response Analysis (MOCA)

Chemical Stressor

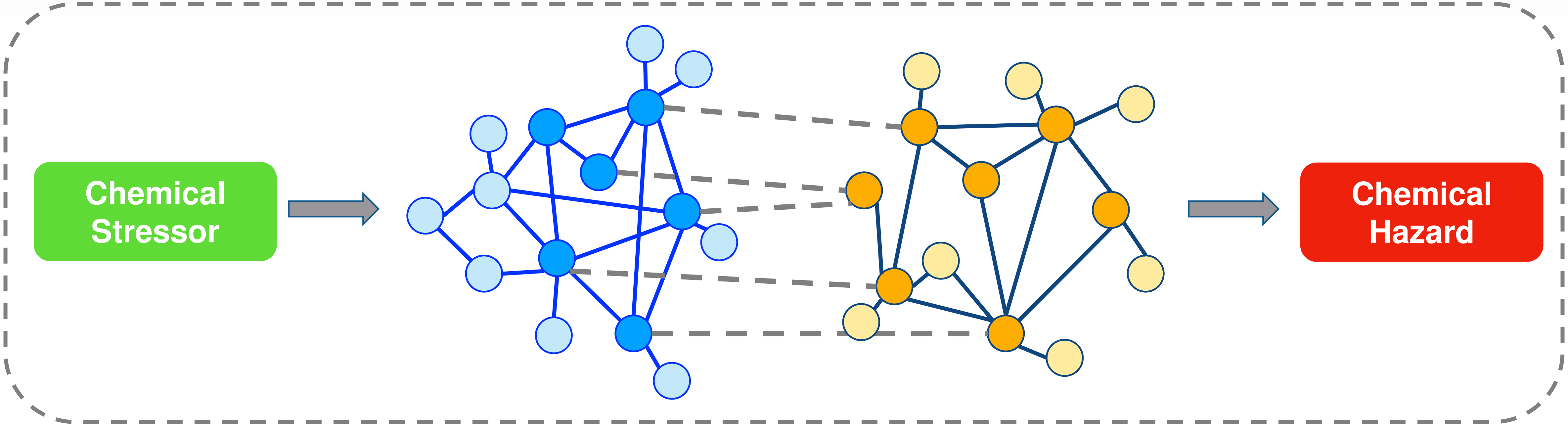
Phenotypic Traits

Quantitative Associations

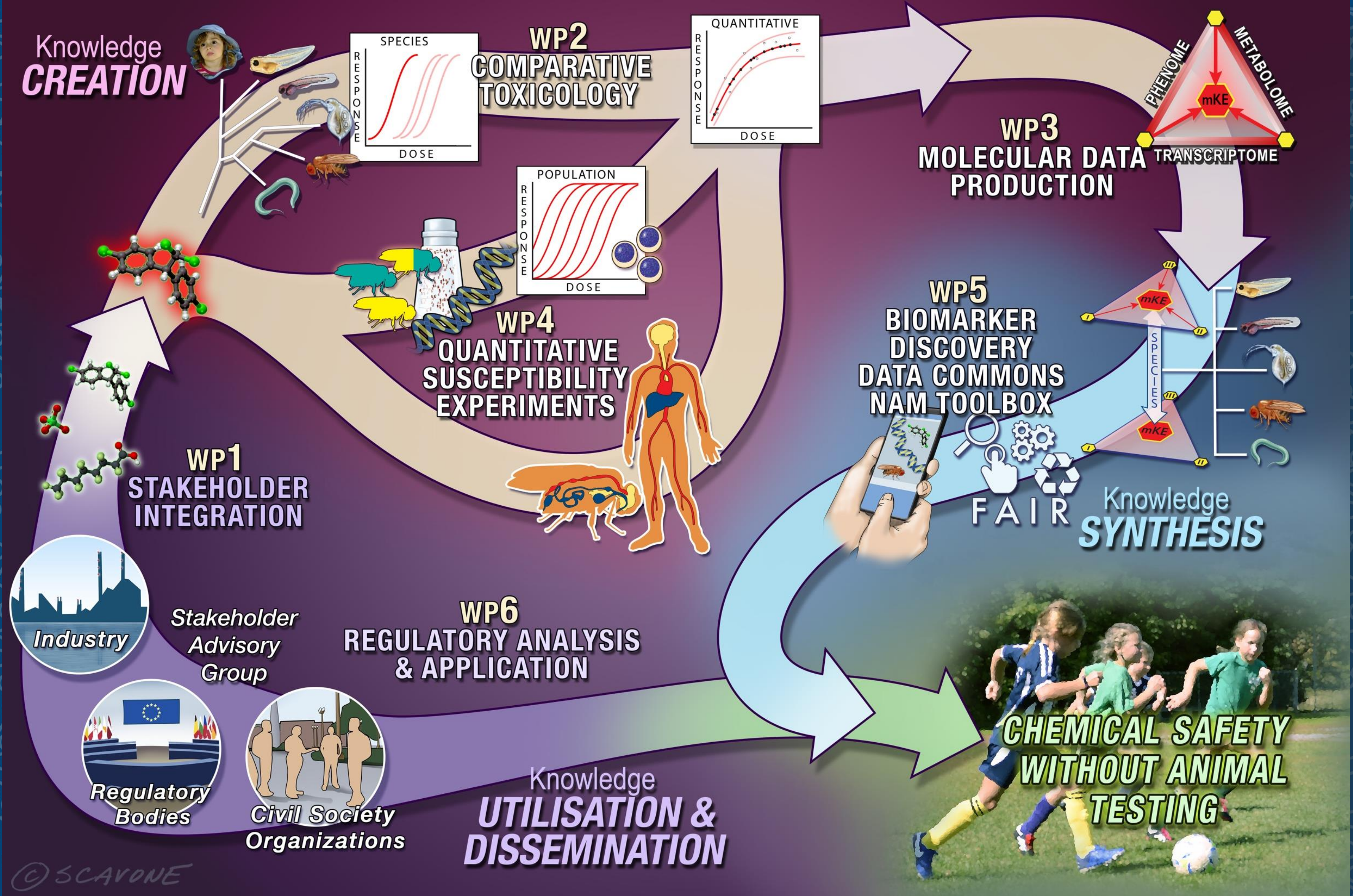
Sparse Co-response Gene Modules

Sparse Co-response Metabolite Modules

Step 4: Molecular Biomarker Identification



Mechanisms Explainability	Ortholog-based Annotation	Pathway Conservation
Adversity Predictability	Prediction Models	Probabilistic Learning



Knowledge CREATION

WP2 COMPARATIVE TOXICOLOGY

WP3 MOLECULAR DATA PRODUCTION

WP4 QUANTITATIVE SUSCEPTIBILITY EXPERIMENTS

WP5 BIOMARKER DISCOVERY DATA COMMONS NAM TOOLBOX

Knowledge SYNTHESIS

WP1 STAKEHOLDER INTEGRATION

Stakeholder Advisory Group

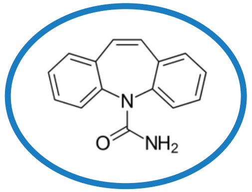
WP6 REGULATORY ANALYSIS & APPLICATION

Knowledge UTILISATION & DISSEMINATION

CHEMICAL SAFETY WITHOUT ANIMAL TESTING

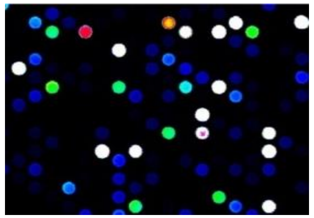
Pre-pilot Case Study

Exposure

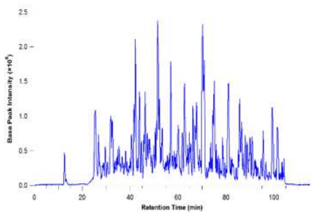


Daphnia magna

Bioassay

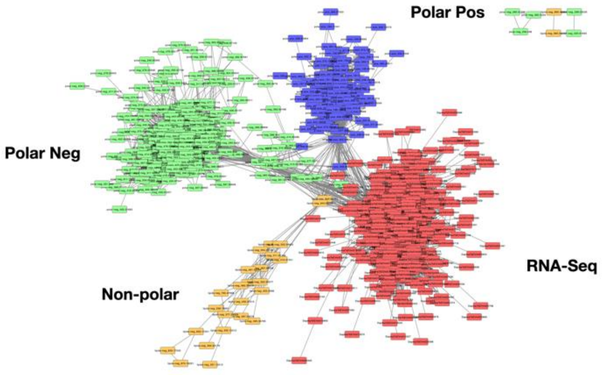


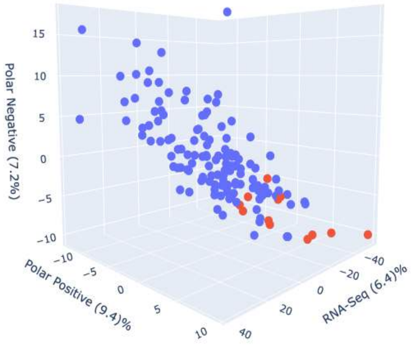
Transcriptome



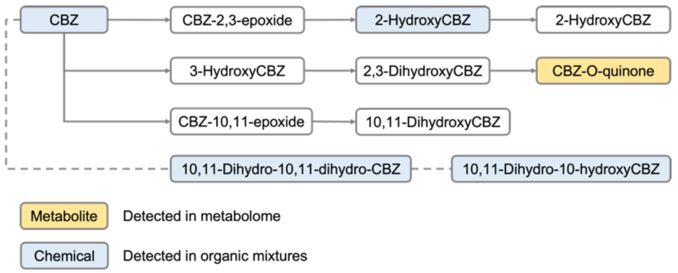
Metabolome

Molecular Biomarker Identification



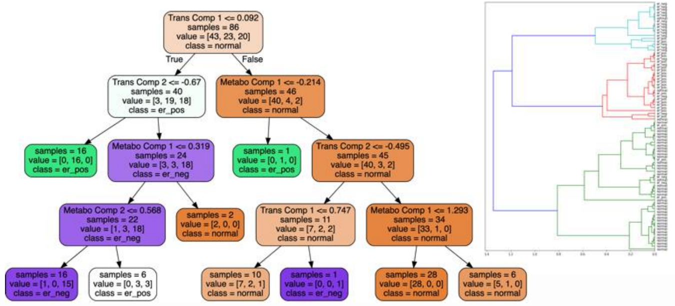


Mechanisms



CBZ
CBZ-2,3-epoxide
2-HydroxyCBZ
2-HydroxyCBZ
3-HydroxyCBZ
2,3-DihydroxyCBZ
CBZ-O-quinone
CBZ-10,11-epoxide
10,11-DihydroxyCBZ
10,11-Dihydro-10,11-dihydro-CBZ
10,11-Dihydro-10-hydroxyCBZ

Metabolite Detected in metabolome
Chemical Detected in organic mixtures



Prediction

PrecisionTox & qAOP

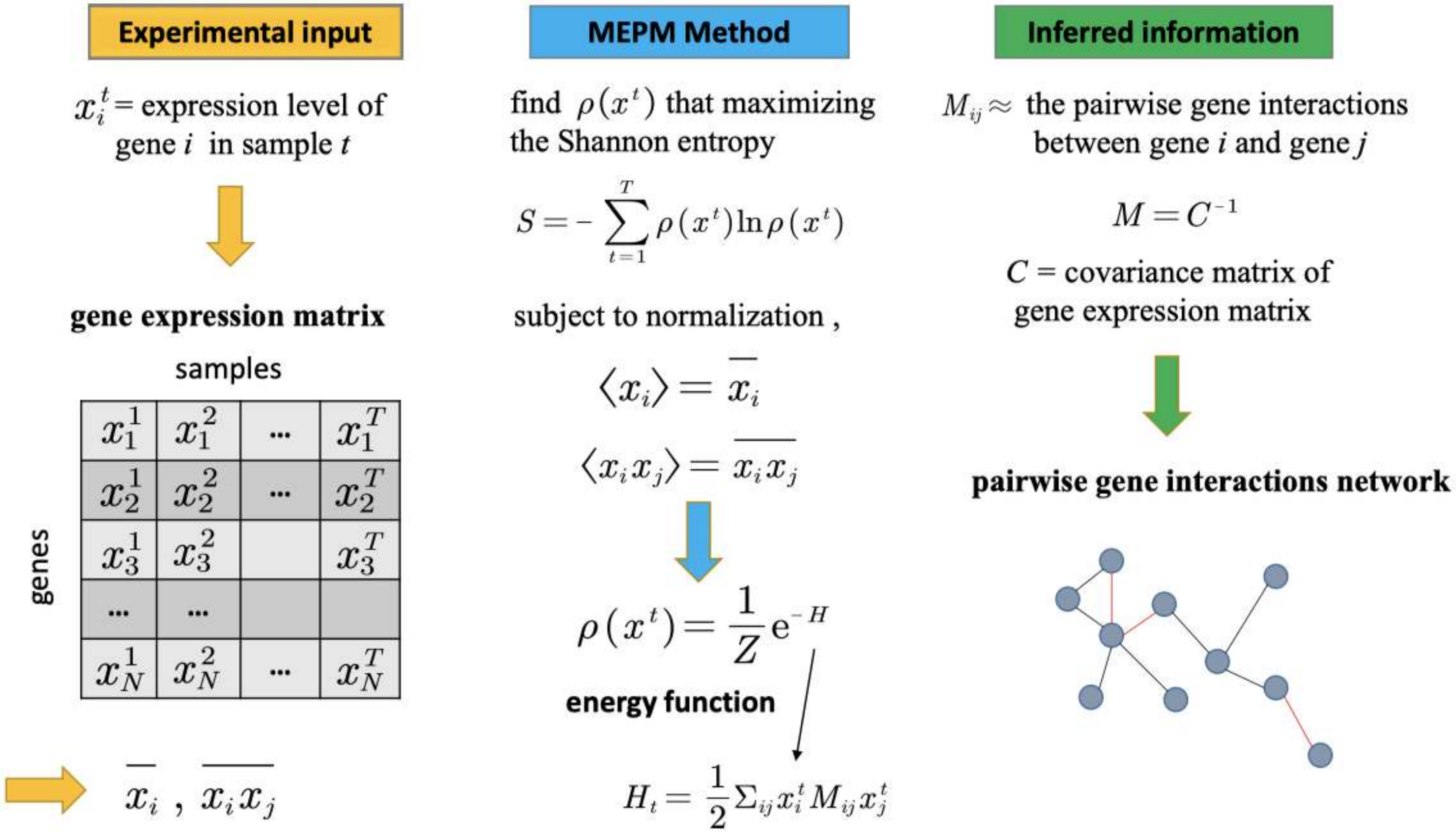
- PrecisionTox uses a systems biology strategy
  - Collects large sets of untargeted multi-omics data
  - Data-driven modelling using combination of machine learning paradigms
  - Assesses the quality of findings with validation experiments
  - Discovers both known unknowns and unknown unknowns
- Efforts in qAOP
  - **Central Database** provides chemical information, known molecular biomarkers and pathways, and their related AOP information
  - **Computational Workflow** for quantifying the molecular KE biomarker associations
- Missing components
  - (Un)certainty measurement
  - From correlation to causality

Measurement of (Un)certainty

- Probabilistic learning
  - Classical machine learning lacks the measurement to account for model (un)certainty and prediction confidence
  - Probabilistic learning provides the opportunity to incorporate prior knowledge and hypotheses
- Explainable machine learning
  - In contrast to the conventional “black box” models, the behaviour and predictions made by the explainable models are understandable to humans
  - Explainability provides inspectable predictions and insights of the underlying mechanisms

Causal Inference

- Probabilistic graphical model combines probabilistic learning, explainable ML, and causal inference
- Sparse genetic regulatory network (GRN) inference using maximum-entropy probability model and multi-objective memetic algorithm (MMA-MEPM)



Experimental input

x_i^t = expression level of gene i in sample t

gene expression matrix

samples

genes

x_1^1	x_1^2	...	x_1^T
x_2^1	x_2^2	...	x_2^T
x_3^1	x_3^2		x_3^T
...	...		
x_N^1	x_N^2	...	x_N^T

$\overline{x_i}$, $\overline{x_i x_j}$

MEPM Method

find $\rho(x^t)$ that maximizing the Shannon entropy

$$S = -\sum_{t=1}^{T} \rho(x^t) \ln \rho(x^t)$$

subject to normalization ,

$$\langle x_i \rangle = \overline{x_i}$$

$$\langle x_i x_j \rangle = \overline{x_i x_j}$$

$$\rho(x^t) = \frac{1}{Z} e^{-H}$$

energy function

$$H_t = \frac{1}{2} \Sigma_{ij} x_i^t M_{ij} x_j^t$$

Inferred information

$M_{ij} \approx$ the pairwise gene interactions between gene i and gene j

$$M = C^{-1}$$

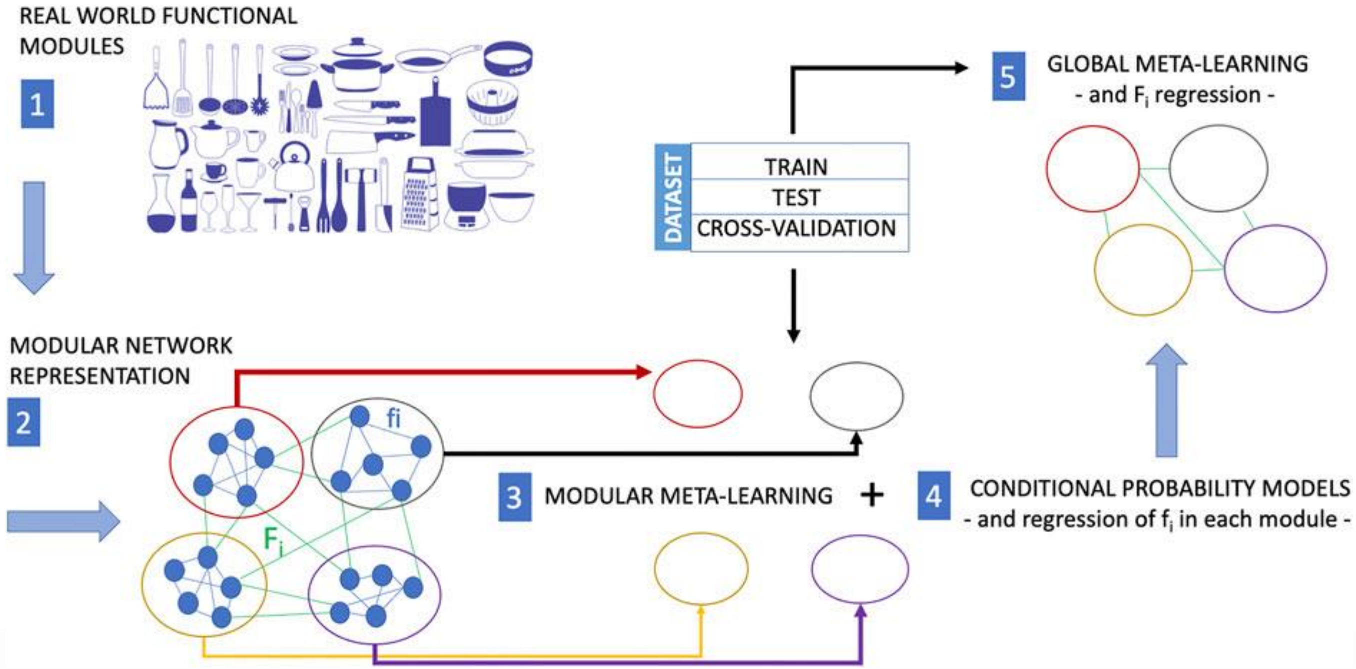
C = covariance matrix of gene expression matrix

pairwise gene interactions network

Fu, Y. (2022). Inferring sparse genetic regulatory networks based on maximum-entropy probability model and multi-objective memetic algorithm.

Causal Inference

- Modular meta-learning is particularly of interest due to the limited experimental data size and modular structure in biological networks



Lecca, P. (2021). Machine learning for causal inference in biological networks: Perspectives of this challenge. Frontiers in Bioinformatics, 1, 746712.

Acknowledgement

- Centre for Environmental Research and Advocacy (CERA)
- PrecisionTox Partners
- ASPIS Cluster & qAOP WG

UNIVERSITY OF BIRMINGHAM

PRECISION TOX

ASPIS

https://precisiontox.org

Thank You