Explainable Artificial Intelligence Models for Developmental and Reproductive Toxicity Prediction using ToxCast/Tox21 and OECD TG 414 Data in Adverse Outcome Pathway Framework
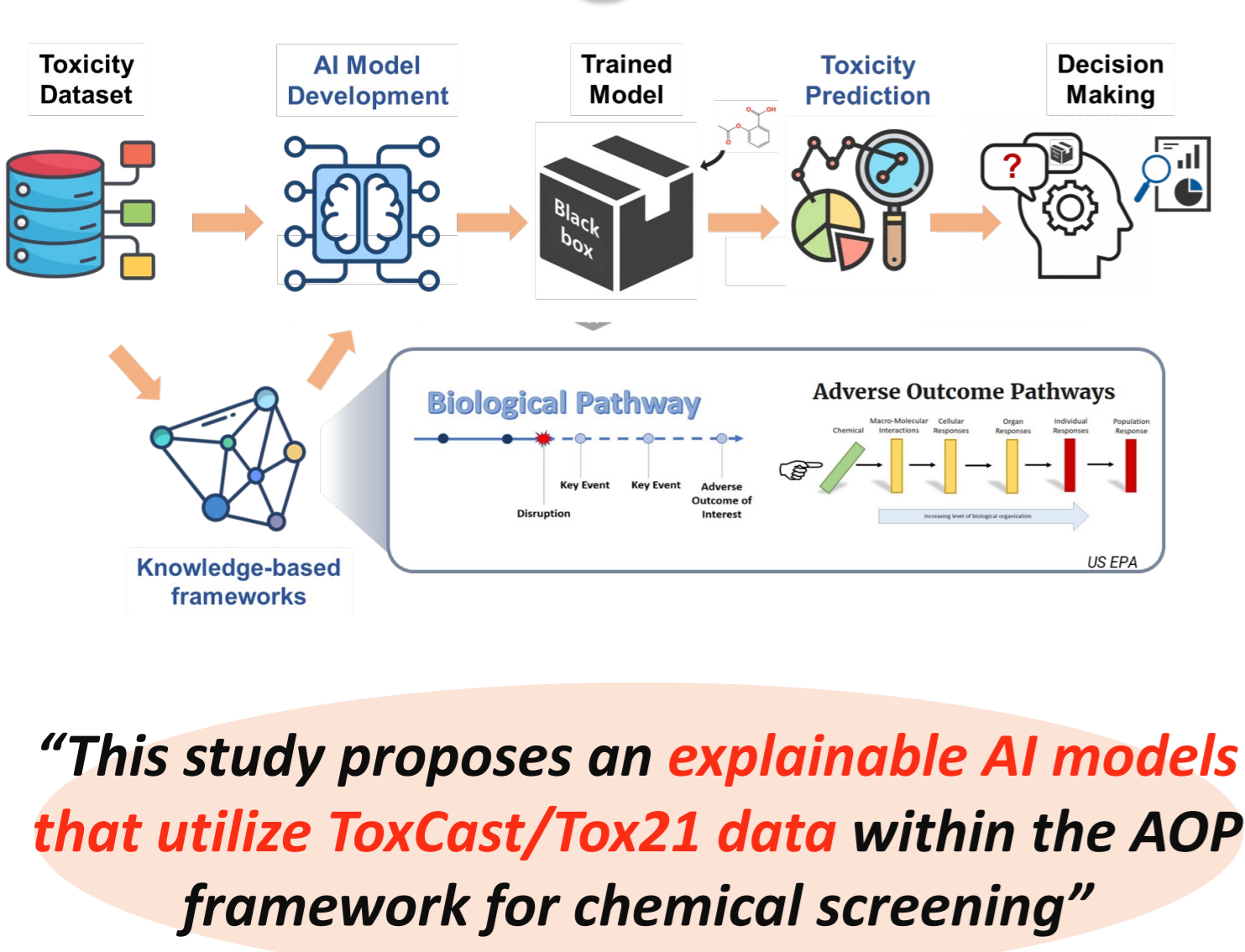
TEAM UOS ToxBAI (Donghyeon Kim, Siyeol Ahn, Jiyong Jeong, and Jinhee Choi*)

School of Environmental Engineering, University of Seoul, 163 Seoulsiripdae-ro, Dongdaemun-gu, Seoul 02504, Republic of Korea
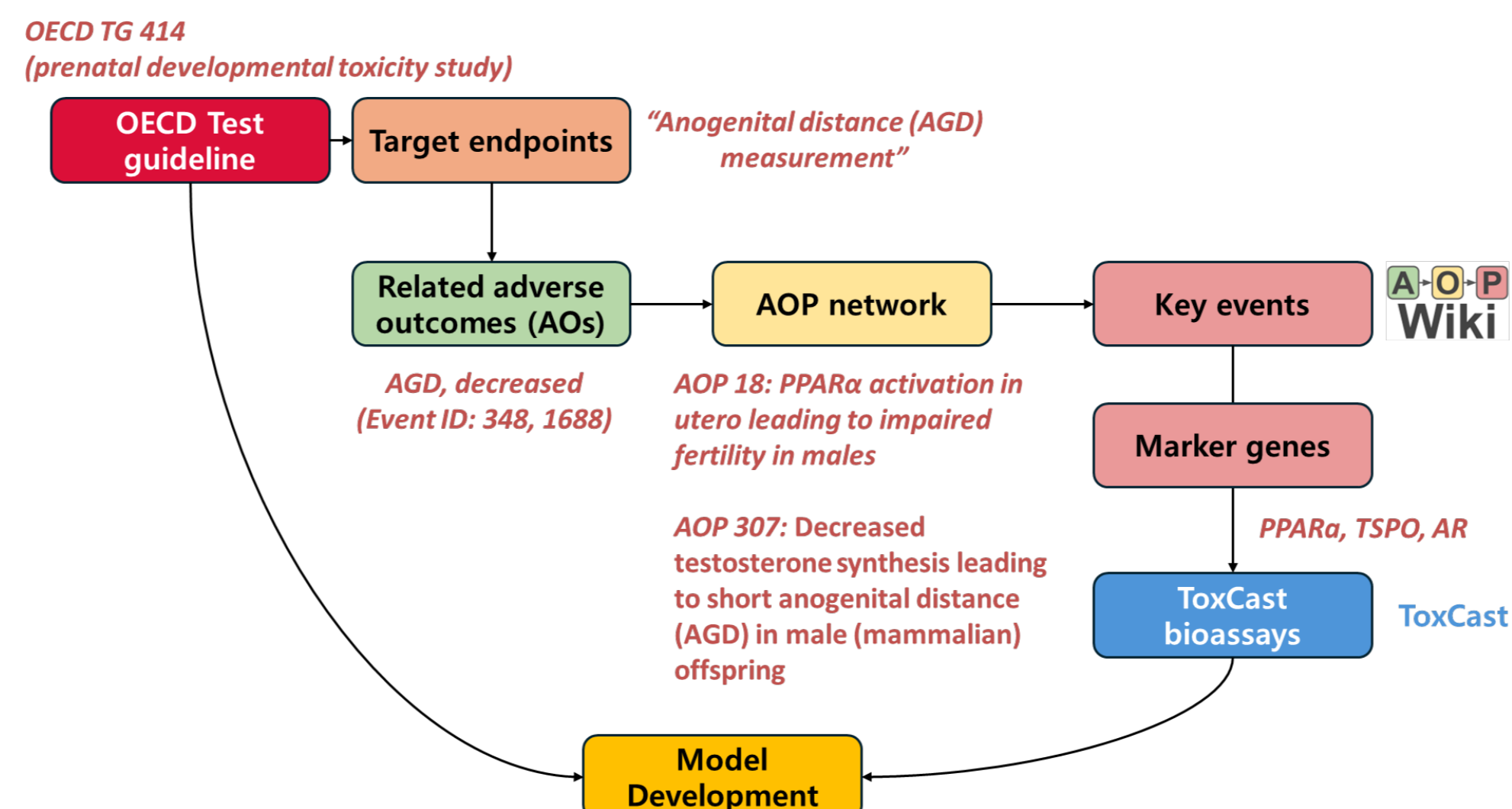
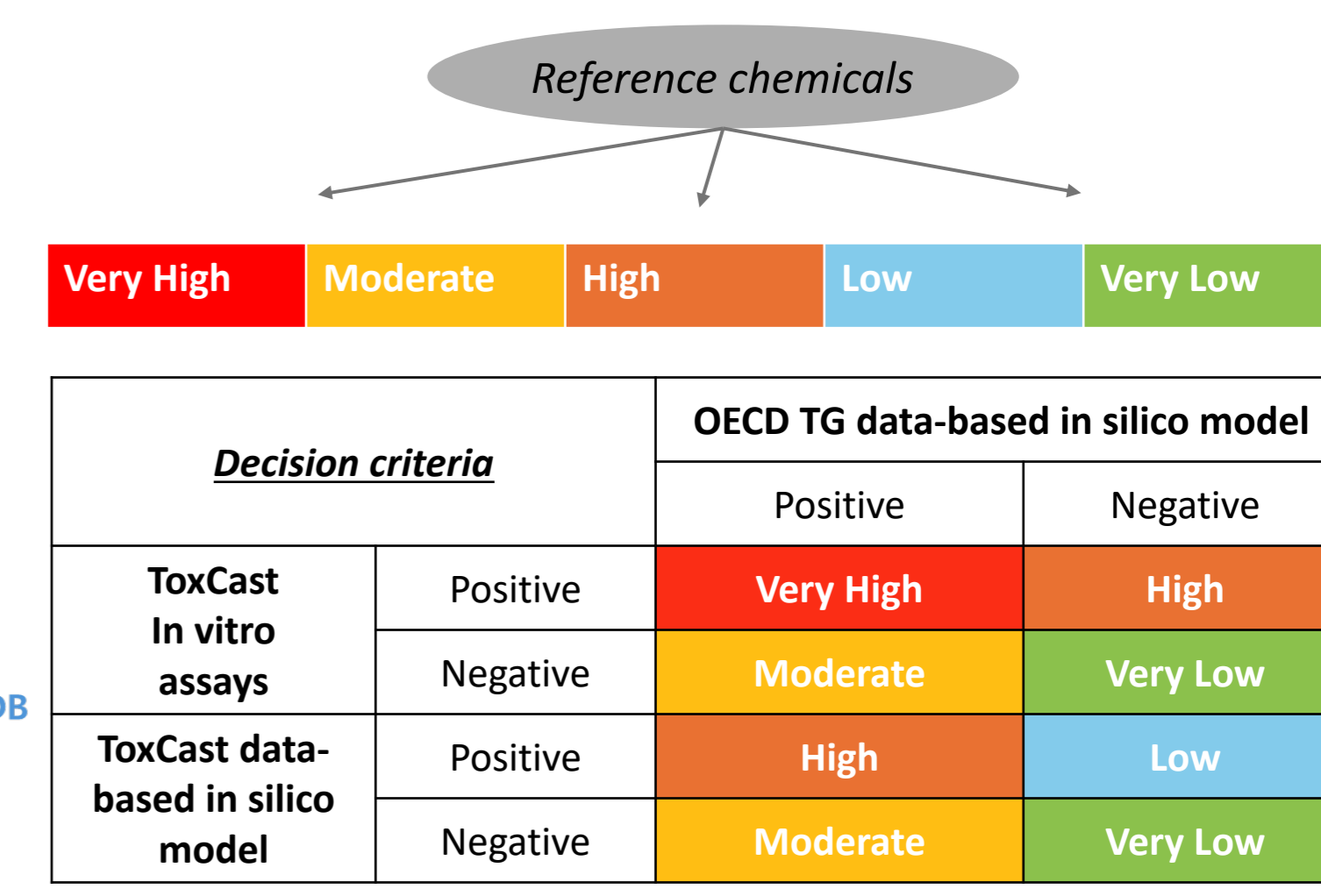*Correspondence: jinhchoi@uos.ac.kr

Background

"This study proposes an *explainable AI models that utilize ToxCast/Tox21 data* within the AOP framework for chemical screening"

NAM Approach

PART1. Development of AOP-based prediction models

AOP 18: PPARα activation in utero leading to impaired fertility in males

AOP 307: Decreased testosterone synthesis leading to short anogenital distance (AGD) in male (mammalian) offspring

PART2. Application to chemical screening

Reference chemicals

| Very High | Moderate | High | Low | Very Low |

Decision criteria		OECD TG data-based in silico model	
		Positive	Negative
ToxCast In vitro assays	Positive	Very High	High
	Negative	Moderate	Very Low
ToxCast data-based in silico model	Positive	High	Low
	Negative	Moderate	Very Low

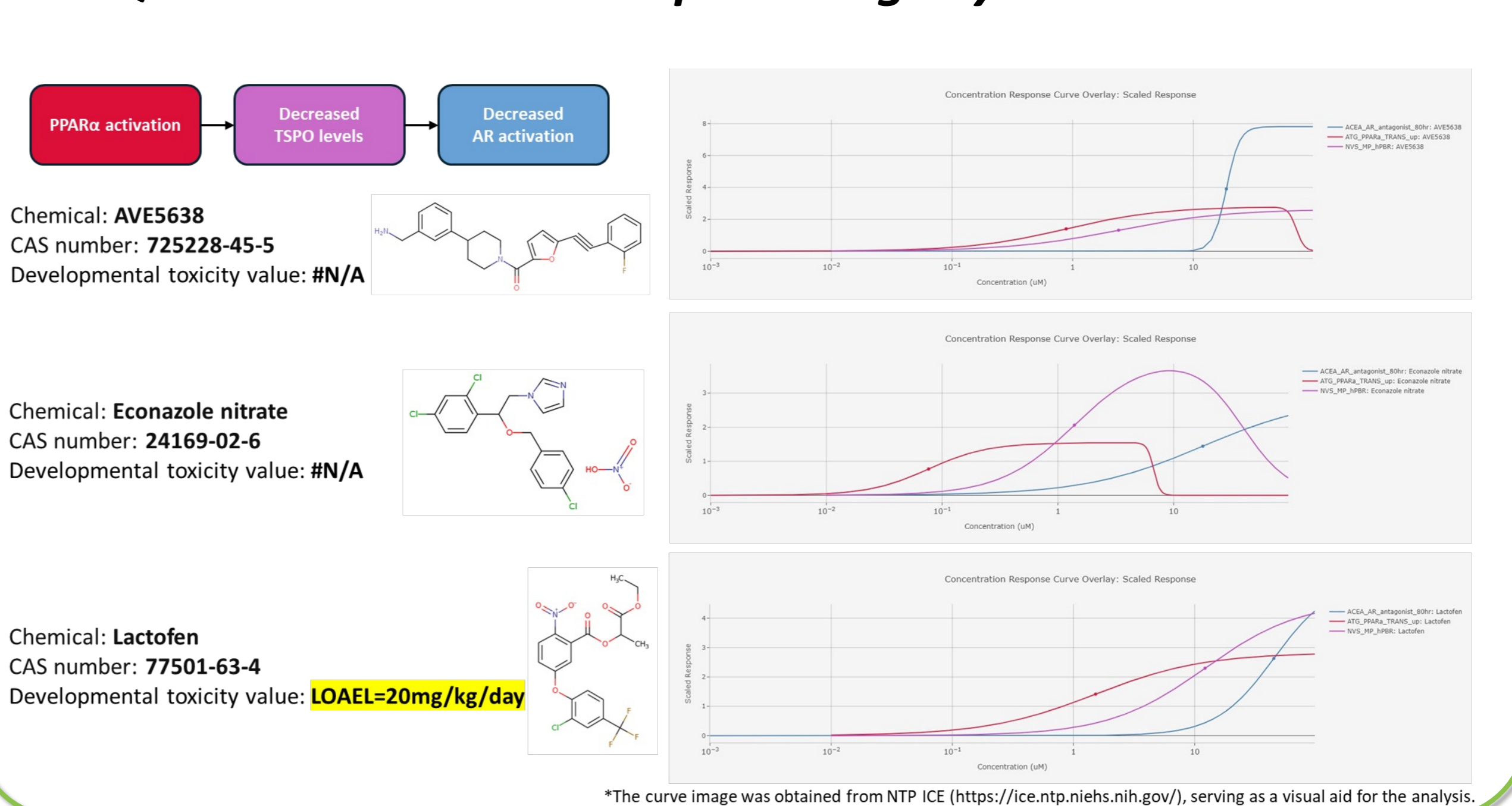## NAM-based Solution

1. In vitro and in vivo assays for DART related AOPs
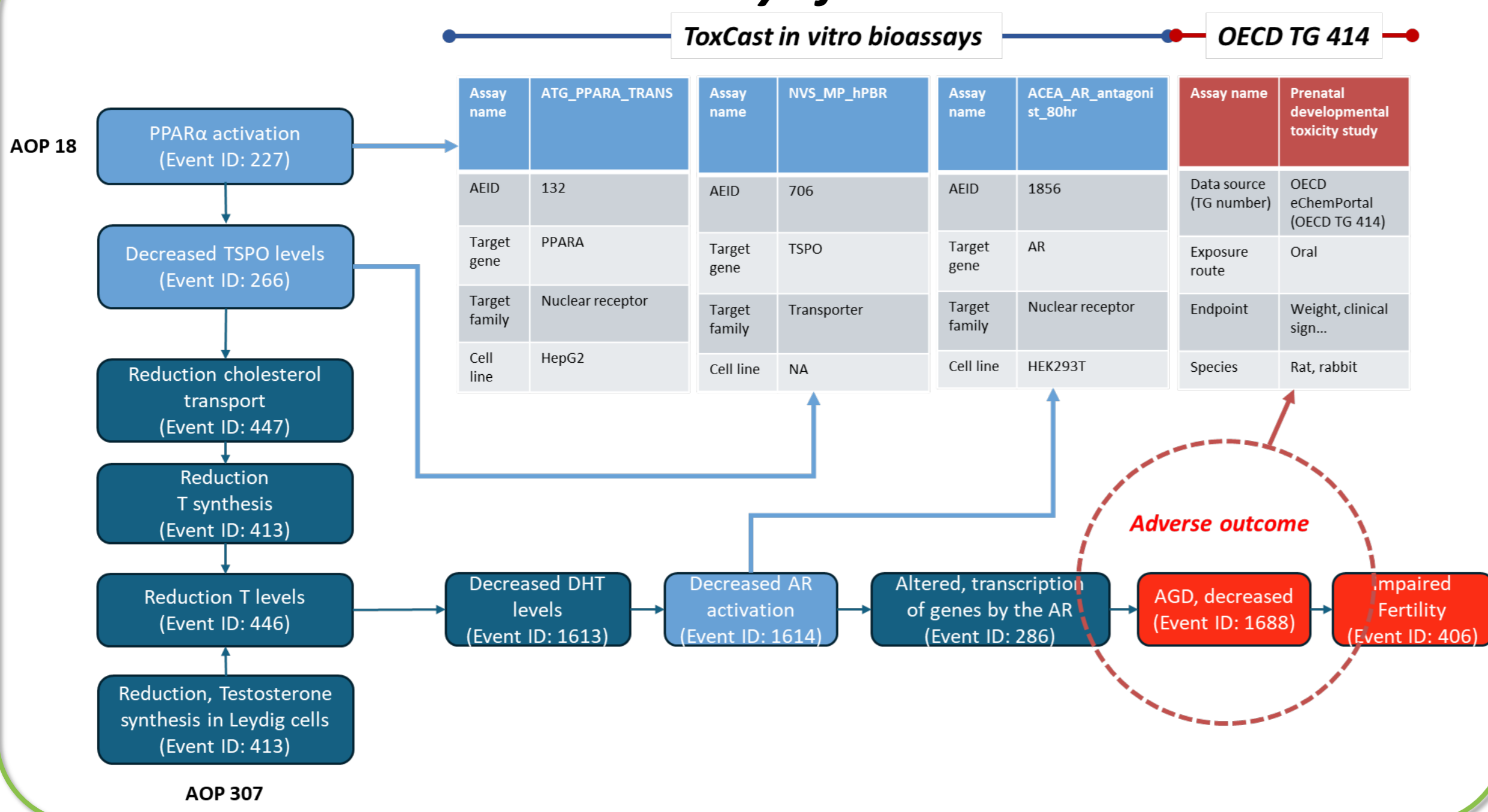
2. Quantitative relationships among key events in ToxCast

Chemical: AVE5638
CAS number: 725228-45-5
Developmental toxicity value: #N/A

Chemical: Econazole nitrate
CAS number: 24169-02-6
Developmental toxicity value: #N/A

Chemical: Lactofen
CAS number: 77501-63-4
Developmental toxicity value: LOAEL=20mg/kg/day

*The curve image was obtained from NTP ICE (https://ice.ntp.niehs.nih.gov/), serving as a visual aid for the analysis.

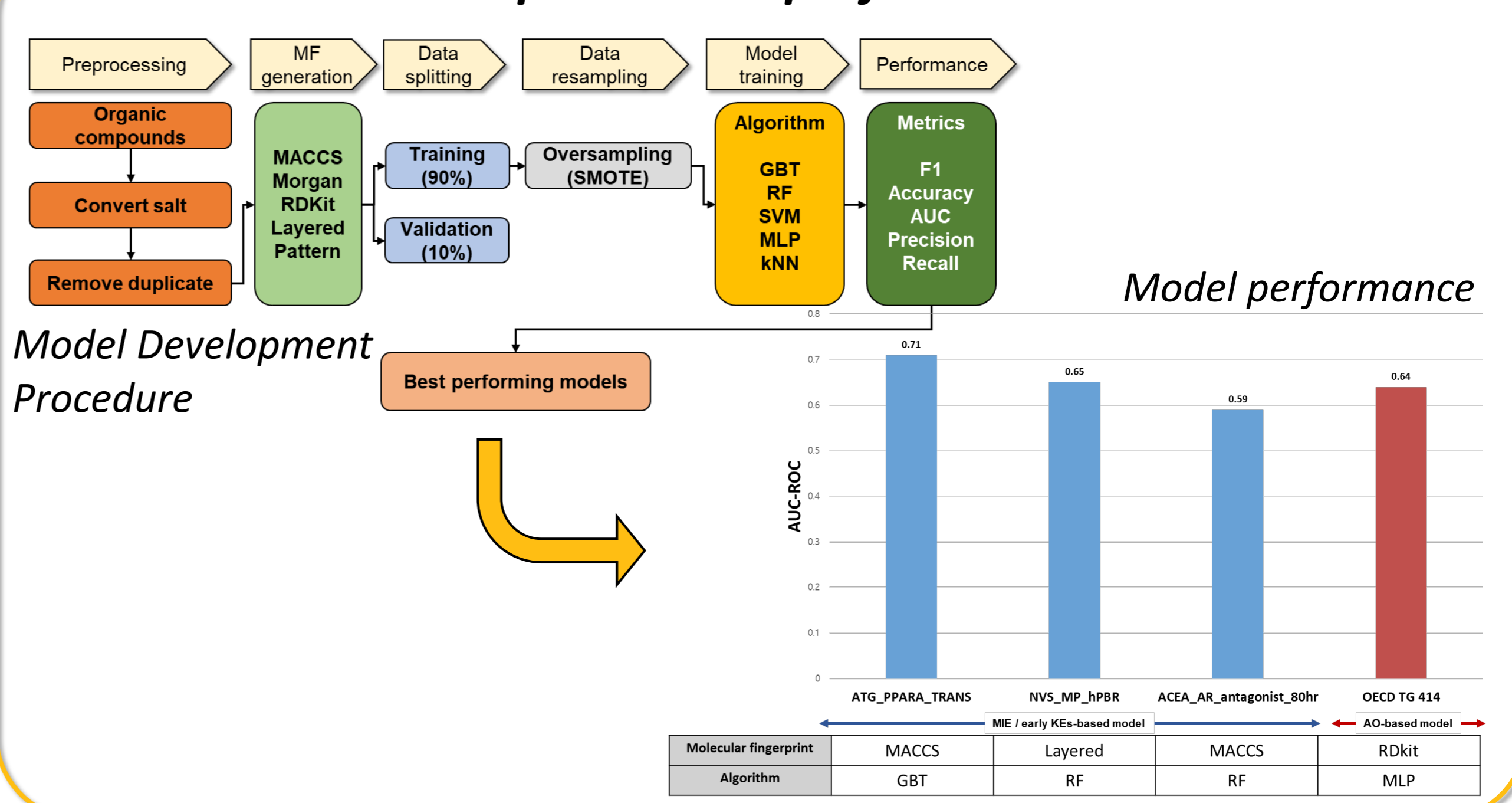### 3. Summary of the training dataset for model development

Quantitative toxicity values → Binary class

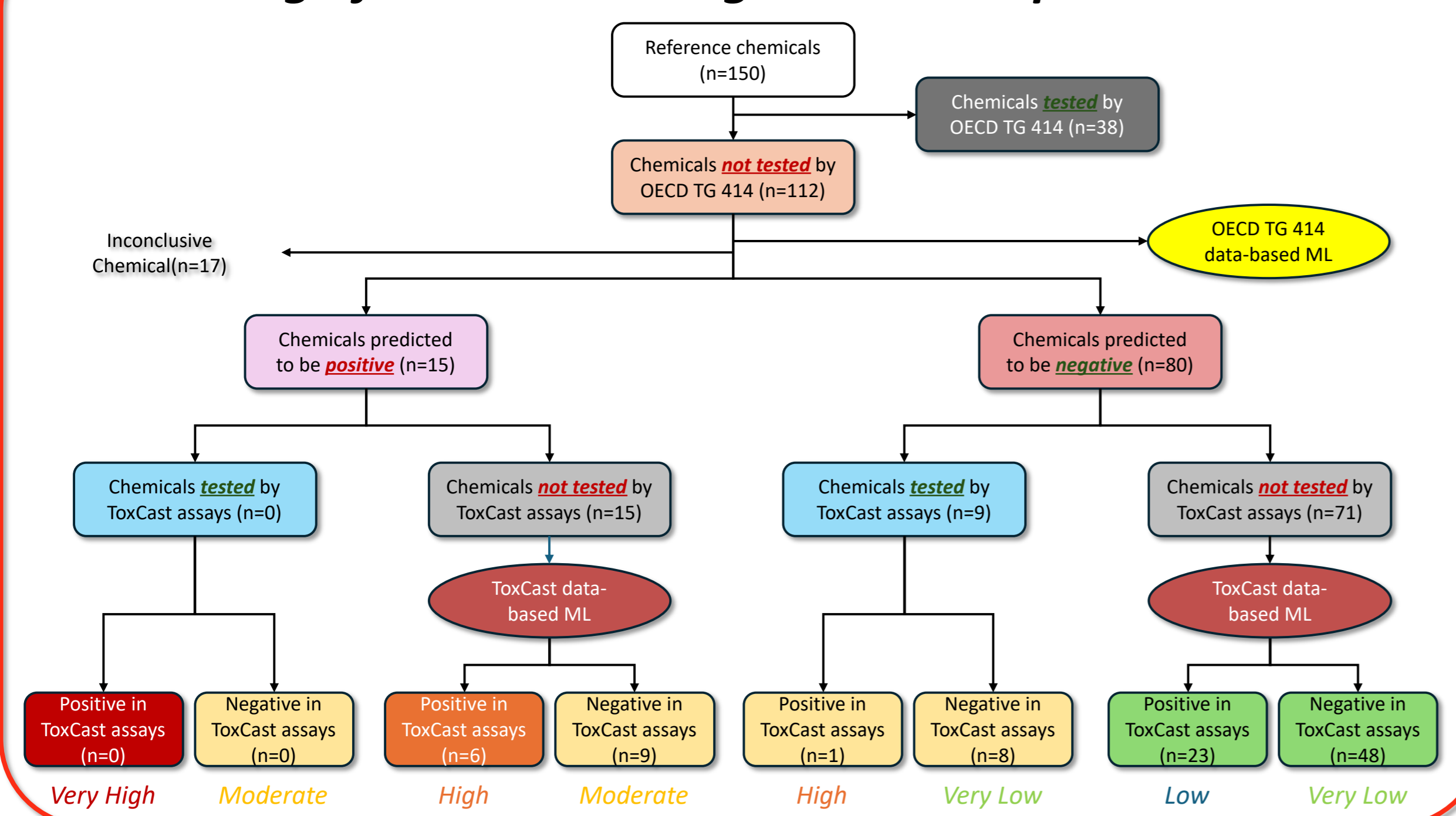In vitro ToxCast data — Hit-call dataset (*following ToxCast pipeline (tcpl)*)

In vivo Toxicity data from eChemPortalDB
- **Positive** class: Chemicals with LOAEL(C), LOEL(C)
- **Negative** class: Chemicals without LOAEL(C), LOEL(C) Chemicals with LOAEL ≥ 1000mg/kg/day

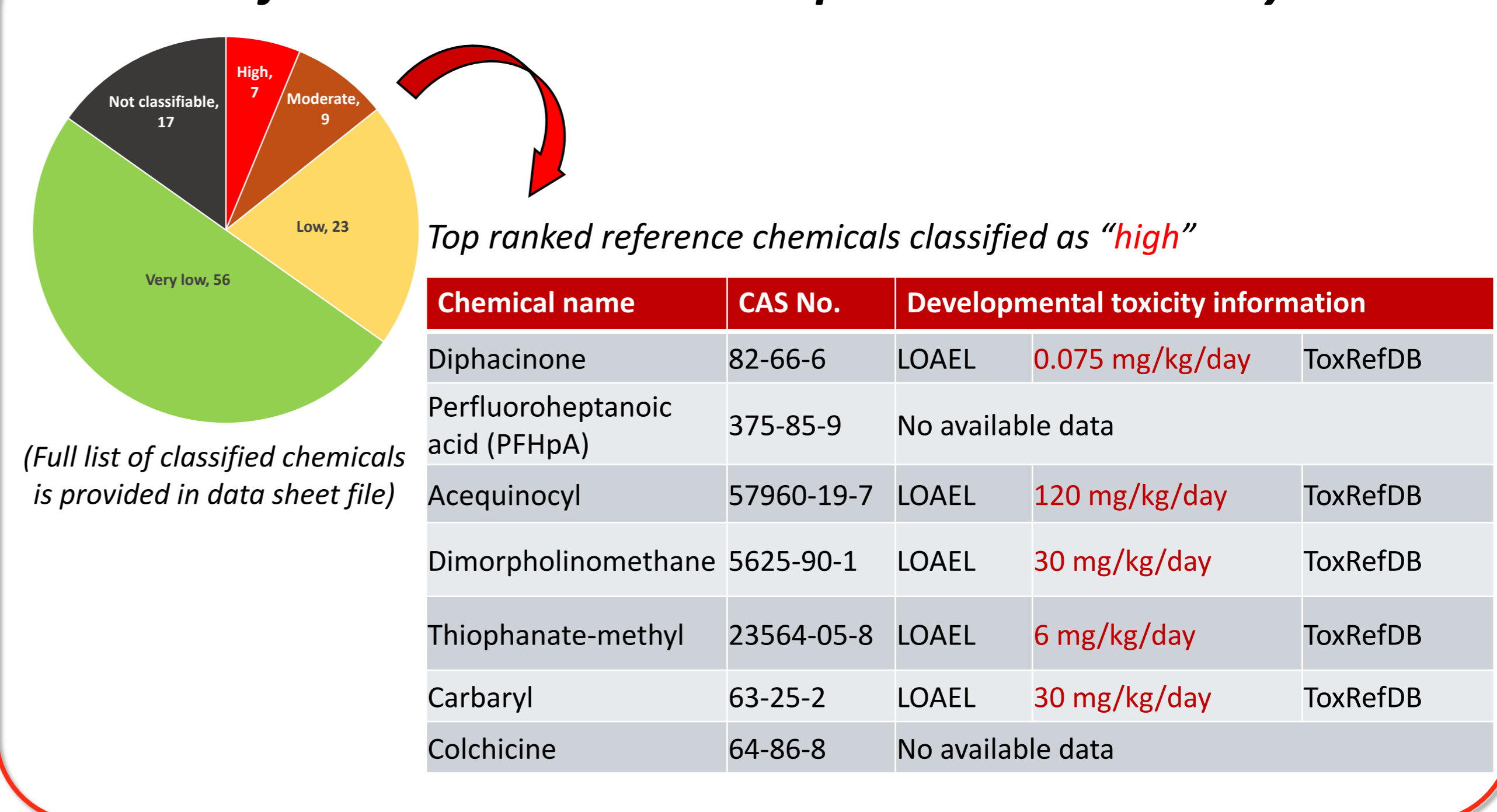| In vitro assay name		ToxCast AEID	No. of active	No. of positive	Active ratio (%)	Total number of data
MIE	ATG_PPARA_TRANS	132	269	4039	6.670	4308
KE1	NVS_MP_hPBR	706	298	456	65.35	754
KE2	ACEA_AR_antagonist_80hr	1856	704	1835	38.37	2539
In vivo assay name		OECD TG No.	No. of active	No. of positive	Active ratio (%)	Total number of data
AO	Prenatal developmental toxicity study	414	111	867	11.35	978

4. Model development and performance evaluation

Model Development Procedure

Model performance

Molecular fingerprint	MACCS	Layered	MACCS	RDkit
Algorithm	GBT	RF	RF	MLP
	MIE / early KEs-based model			AO-based model

5. Screening of chemicals using AOP-based prediction models

6. Classification results and comparison with toxicity data

(Full list of classified chemicals is provided in data sheet file)

Top ranked reference chemicals classified as "high"

Chemical name	CAS No.	Developmental toxicity information		
Diphacinone	82-66-6	LOAEL	0.075 mg/kg/day	ToxRefDB
Perfluoroheptanoic acid (PFHpA)	375-85-9	No available data		
Acequinocyl	57960-19-7	LOAEL	120 mg/kg/day	ToxRefDB
Dimorpholinomethane	5625-90-1	LOAEL	30 mg/kg/day	ToxRefDB
Thiophanate-methyl	23564-05-8	LOAEL	6 mg/kg/day	ToxRefDB
Carbaryl	63-25-2	LOAEL	30 mg/kg/day	ToxRefDB
Colchicine	64-86-8	No available data		

Limitations in the Applicability

1. Differences in their respective Domains of Applicability (DoA).
2. Limited quantitative understanding of relationships between the key events and adverse outcomes.
3. Inadequate coverage of biological targets and pathways in ToxCast assays

Acknowledgement