Predictive Model of Reproductive Toxicity from ToxCast High Throughput Screening

Matthew Martin (EPA/ORD/NCCT)
CPCP – May 2011


Overview

- The Problem
  - Human reproductive impairment & the environment
  - Current testing paradigm & in vivo toxicology
- A Solution
  - In Vitro assays & high-throughput screening
  - Bioinformatics & predictive models

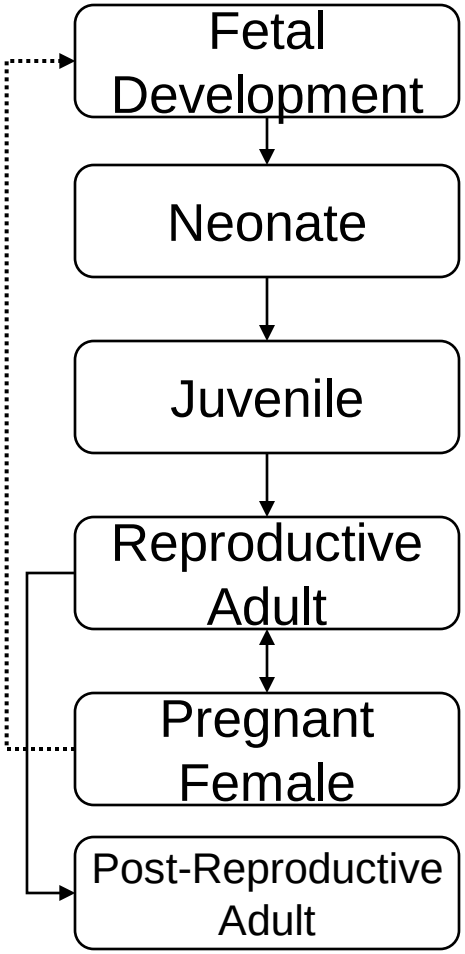
Human Reproductive Impairment
&
The Environment

Reproductive Impairment & Toxicity

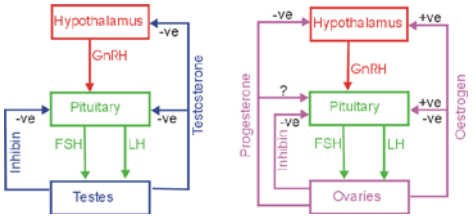
- What?
  - Environmental impact on and causes of reproductive impairment
- Why?
  - 1 in 7 couples experience infertility
  - Miscarriage occurs in a reported 10-25% of all pregnancies
  - Increased focus on Testicular Dysgenesis Syndrome (TDS) & Endometriosis
  - Links between chemical exposure and reproductive anomalies
- How?
  - Chemical toxicity studies in rodent
  - Human epidemiology studies
  - Worker exposure studies
  - In Vitro bioactivity profiling?

Reproductive Impairment & Toxicity

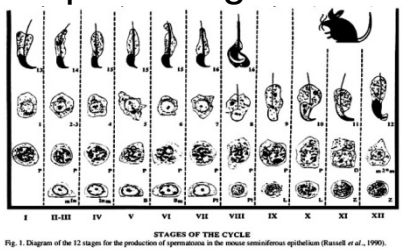
- Fetal Development
  - **Survival**
  - **Malformations (e.g. TDS)**
  - **Future reproductive health**
- Neonate
  - **Survival**
  - **Reproductive organ**
- Juvenile
  - **Sexual development**
  - **Pubertal delay/acceleration**
- Reproductive Adult
  - **Gametogenesis**
  - **Menstrual Cycle**
  - **Neuromuscular**
- Pregnant Female
  - **Failed pregnancy**
  - **Lactation effects**
- Post Reproductive Adult
  - **Neoplasia/Cancers**



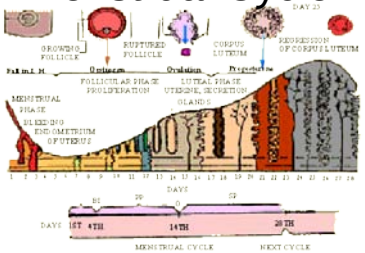
HPG-Axis



Spermatogenesis

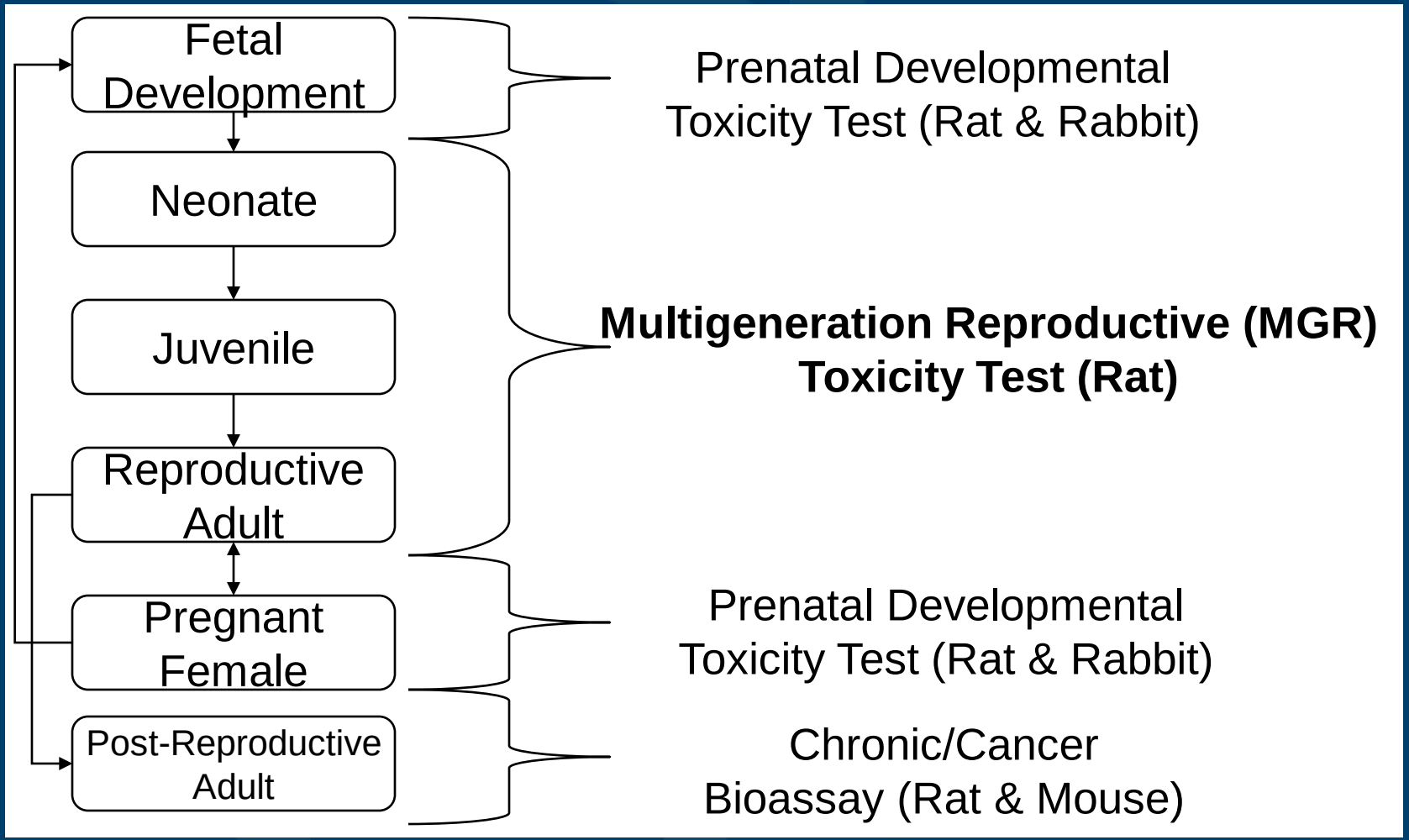


Menstrual Cycle



Current Testing Paradigm
&
Reference In Vivo Reproductive Study Data

Current Testing Paradigm



Reproductive Toxicity Testing

- Limited regulatory capacity to require or request reproductive testing
  - Primarily food-use pesticides require study through FIFRA/FQPA
  - In 30 years, ~500-1000 chemicals tested in MGR study
  - >10,000 chemicals currently in the environment
- Repro testing for REACH compliance
  - 70% of costs
  - 90% of animal use
  - Similar to eventual TSCA reform?
- Gross endpoints measured/investigated
  - Lack of mechanistic information
  - Non-gender specific
- Limited effort in predictive modeling of reproductive toxicity
  - Complexity of study design & endpoints
  - Lack of high quality reference information

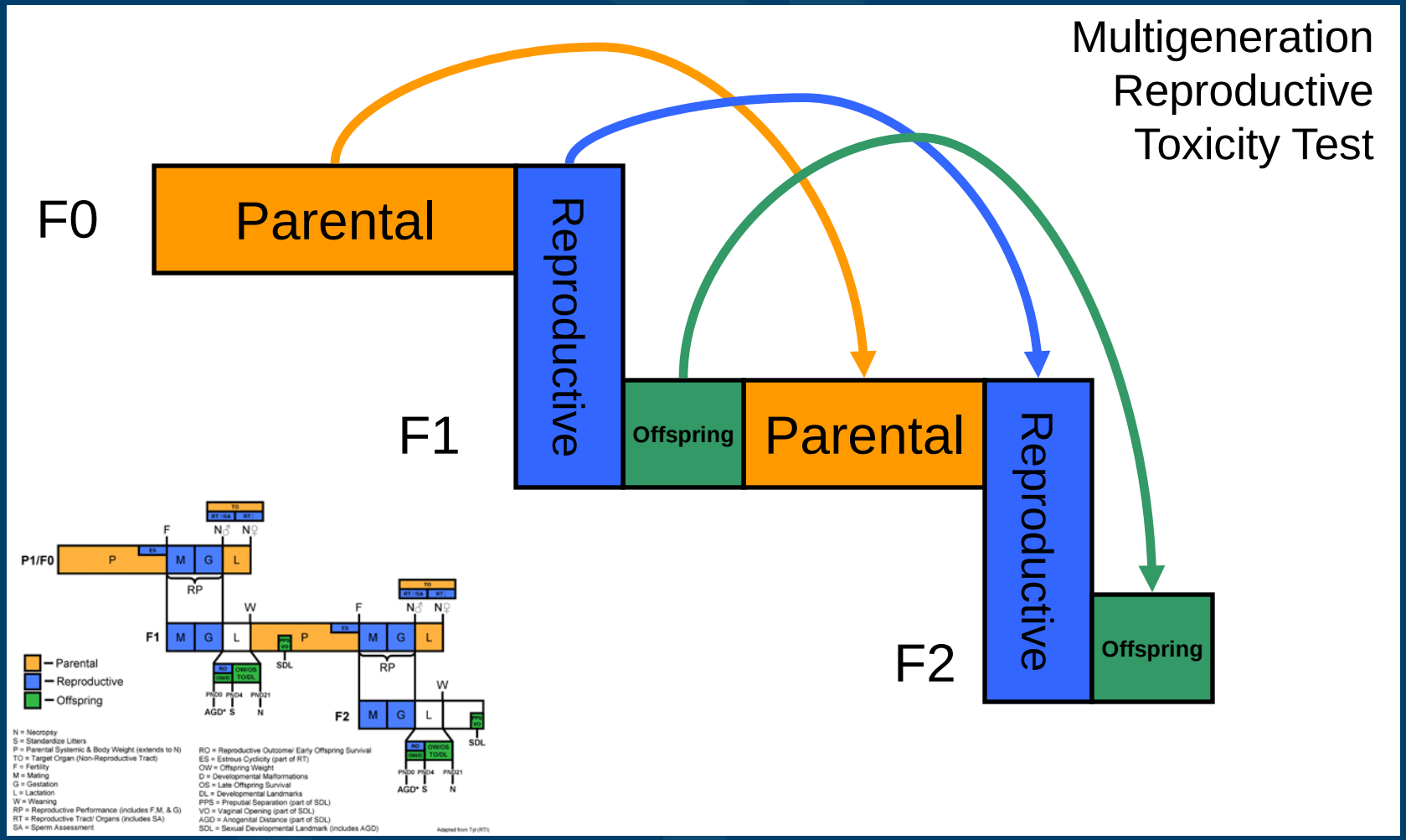
FIFRA: Federal Insecticide, Fungicide, and Rodenticide Act

FQPA: Food Quality and Protection Act

REACH: Registration, Evaluation, Authorisation and Restriction of Chemical substances

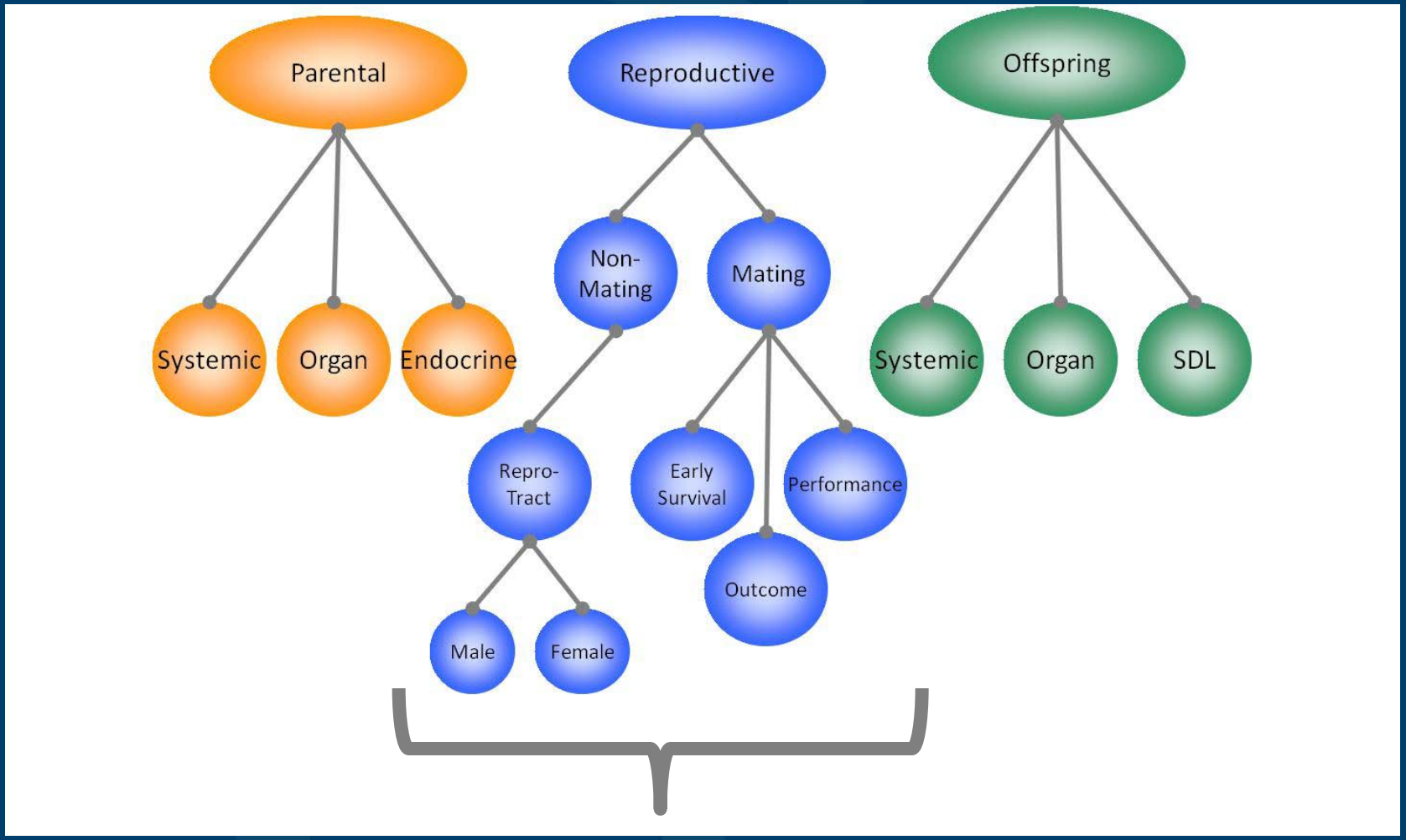
TSCA: Toxic Substance Control Act

ToxRefDB: Capturing & Simplifying the MGR Study Design



F0: 1st Generation Parental Animals
F1: 1st Generation Offspring & 2nd Generation Parental Animals
F2: 2nd Generation Offspring Animals

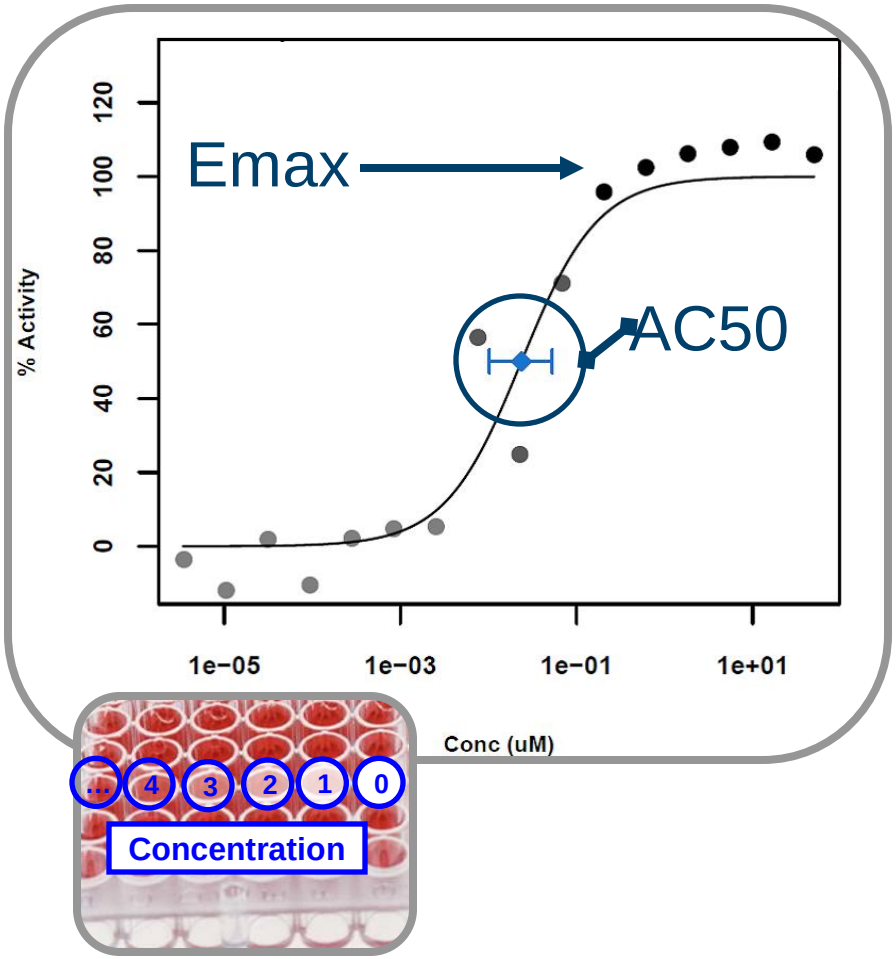
Reproductive Toxicity as an Endpoint for Predictive Modeling

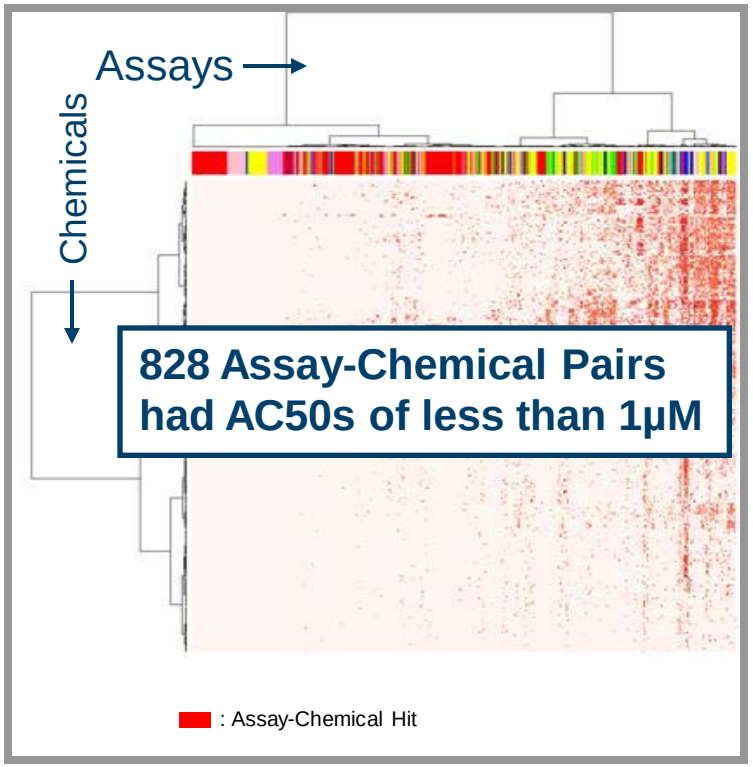


98 total chemicals w/ reproductive LOAEL (rLOAEL ≤ 500 mg/kg/day)
86 from **256** acceptable studies in ToxRefDB
12 from **39** unacceptable studies in ToxRefDB

Predictive Model of Reproductive Toxicity

ToxCast Data Analysis

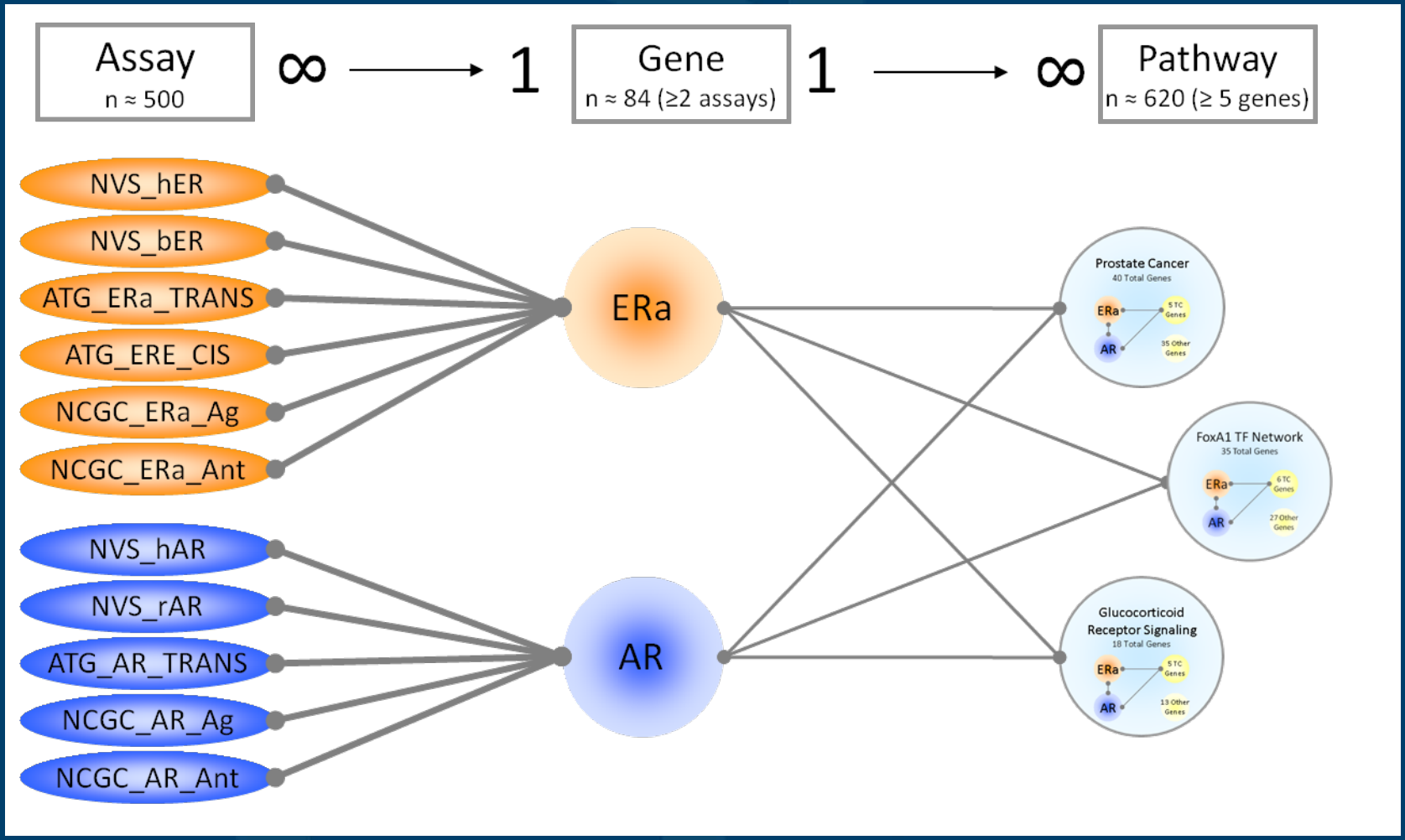




500 Assays X 320 Chemicals X {5-18} Concentrations X {1-3} Replicates X {1-3} Time Points ≈
3.2 Million Data Points

Emax: Maximal efficacy/response/activity
AC50: Concentration whereby 50% of maximal response was achieved
Conc (uM): Micromolar concentrations of chemical

Gene-level Analysis for Modeling



AssayScore = -log3(AC50/1000)
Negative = 0
Positive = {3-12}

GeneScore = Avg(AssayScore)
ERa = Estrogen Receptor Alpha
AR = Androgen Receptor

PathwayScore = Avg(GeneScore)

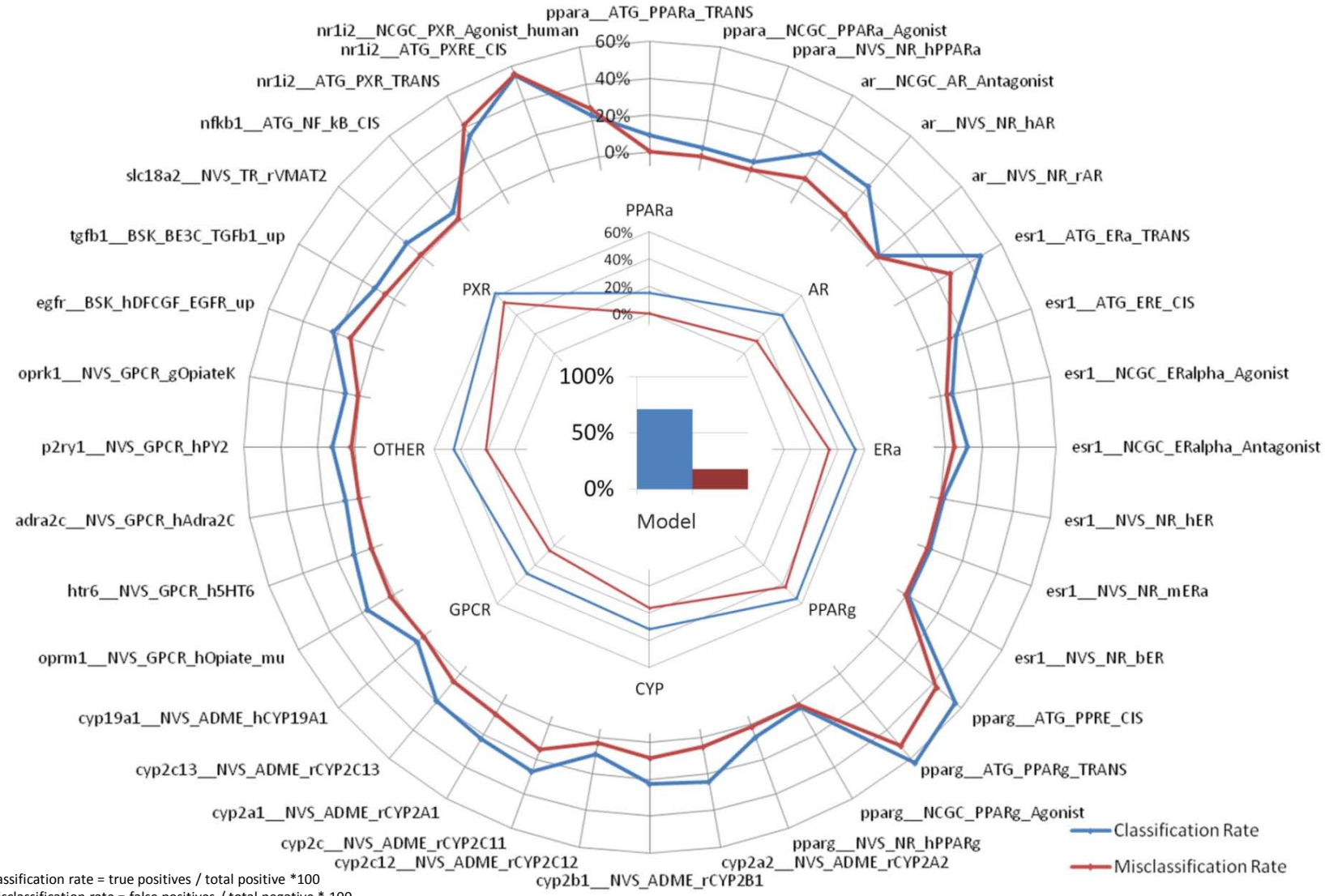
Model Development Steps

- Defined Class Data – In Vivo Reproductive Endpoint Positives and Negatives
- Defined Training Set
  - Acceptable In Vivo Study in ToxRefDB
  - >2% Active across ~500 Assays (Hit in 10/500 assays)
    - Filtered out potential confounders (e.g., insolubility, decomposition, etc.)
    - Similar profile of In Vivo Activity; Cannot assume negative for toxicity
- Performed Feature Selection
  - Statistical Association with Endpoint
  - Biologically relevant groupings (i.e., gene and gene sets)
- Developed Model
  - 5-Fold CV using LDA
  - Complete model on full training set (used downstream for model validation)
- Model Validation & Example Applications

Chemical Group Training Set Considerations

	In vitro Activity	Little to No *In vitro* Activity (<2% Active)	Total *In vivo* Chemical Counts
Acceptable Reproductive Study	206 (A)	50 (B)	256
Unacceptable Reproductive Study	31 (C)	8 (D)	39
No Reproductive Study Available	10 (E)	4 (F)	14
Total *In vitro* Chemical Counts	247	62	309

Feature Selection & Model Development

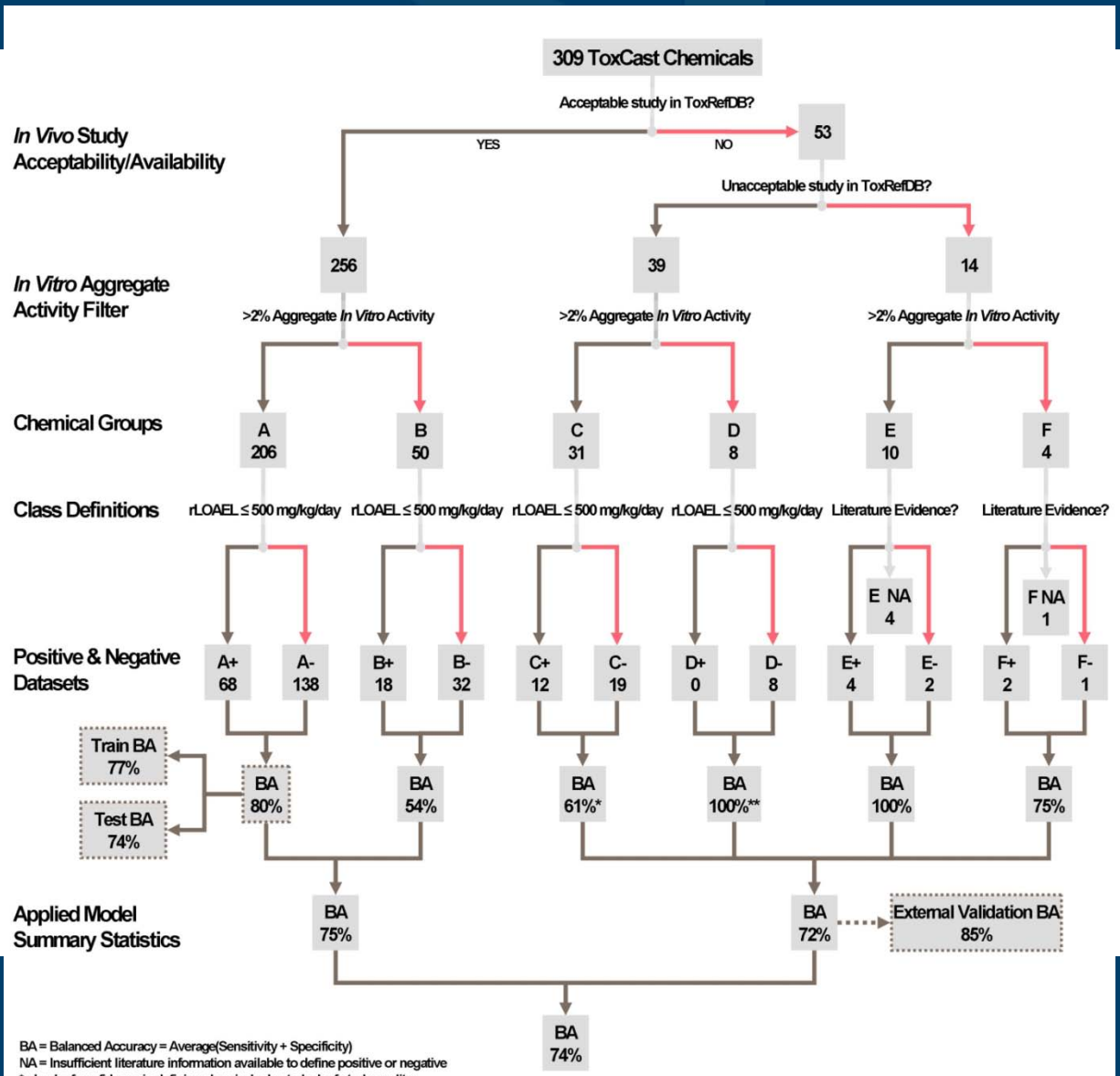


Classification rate = true positives / total positive *100
Misclassification rate = false positives / total negative * 100
Increased classification rate from assay to gene to model
Easier communication of model results based on gene or gene-set features

Model Summary Statistics

Cross-Validation Statistics		Full Model Statistics				Parameter Coefficients	
Learner	**LDA**	TP	**55**	F1	**73%**	**PPARα**	1.37
CV	**5-fold**	FP	**28**	RR	**6.3**	**AR**	0.98
No. F	**8**	FN	**13**	OR	**17**	**ERα**	0.45
Assays	**36**	TN	**110**	PPV	**66%**	**PPARγ**	0.23
BA Train	**77%**	SENS	**81%**	NPV	**90%**	**CYP**	0.28
SD Train	**2%**	SPEC	**80%**	Pred	**78%**	**GPCR**	0.5
BA Test	**74%**	BA	**80%**	P-Value	**4.2E-17**	**OTHER**	0.45
SD Test	**5%**	A	**80%**	Cutoff	**0.6**	**PXR**	-0.21

Model Results By Chemical Group



External Validation Set

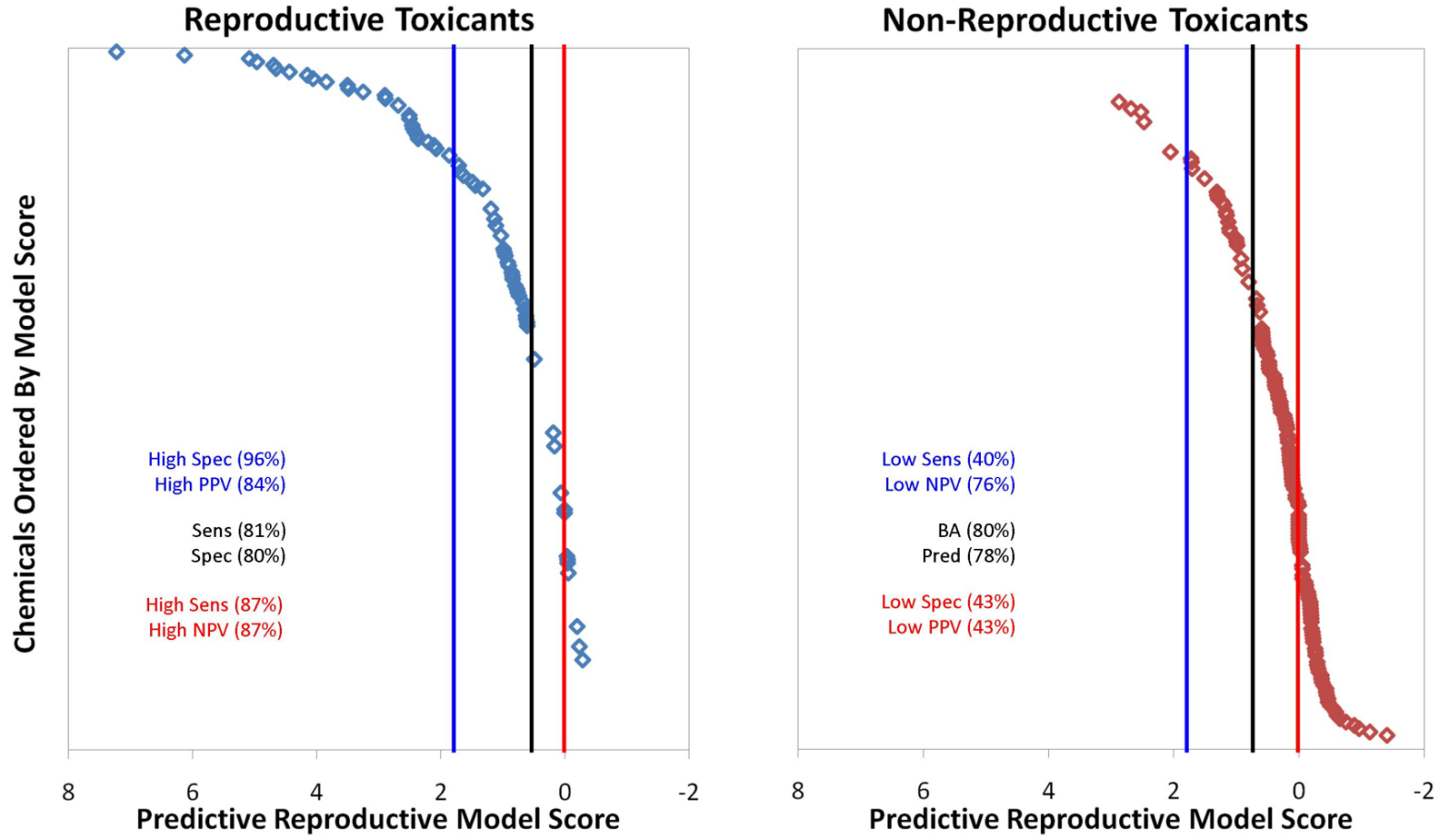
Chemical Group	Chemical Name	Evidence of Reproductive Toxicity	Predicted Reproductive Toxicant
E	HPTE	Yes	Yes
C	Fenitrothion	Yes	Yes
C	Prochloraz	Yes	Yes
E	Bromoxynil	Yes	Yes
E	Methoxychlor	Yes	Yes
C	Milbemectin	Yes	Yes
C	Metiram-zinc	Yes	Yes
C	Chlorsulfuron	Yes	Yes
F	Methyl cellusolve	Yes	Yes
C	Abamectin	Yes	Yes
C	Tebupirimfos	Yes	Yes
E	Alachlor	Yes	Yes
C	Tribufos	Yes	No
C	Spiroxamine	Yes	No
C	Tefluthrin	Yes	No
C	Disulfoton	Yes	No
C	Esfenvalerate	Yes	No
E	Methyl hydrogen phthalate	No	No
F	Monocrotophos	No	No
F	Dimethyl phthalate	No	No
E	Butralin	No	No
E	Clorophene	Unknown	Yes
E	Diniconazole	Unknown	Yes
E	Niclosamide	Unknown	Yes
F	Phenoxyethanol	Unknown	No
E	Symclosene	Unknown	No

Comparison to EU Repro C&L

Chemical Name	Predicted Positive	Repro C&L	Model Score
Bisphenol A	Yes	R62	6.1
Vinclozolin	Yes	R60&61	4.7
Flusilazole	Yes	R61	4.6
Linuron	Yes	R62&61	2.9
Myclobutanil	Yes	R63	2.4
Fenarimol	Yes	R62	2.5
Fentin	Yes	R63	3.5
Fluazifop-P-butyl	Yes	R63	1.7
Flumioxazin	Yes	R61	0.9
Cyproconazole	Yes	R63	1.2
Diethylhexyl phthalate (DEHP)	Yes	R60&61	0.9
Isoxaflutole	Yes	R63	0.6
Fluazifop-butyl	Yes	R61	1.0
Dibutyl phthalate	Yes	R62&61	0.8
Benomyl	No	R60&61	0.0
Diuron	No	---	0.4
Lindane	No	---	0.0
Propazine	No	---	-0.3
Propargite	No	---	-0.5

* Requires metabolic activation *

Chemical Testing Prioritization



Adjusting the model cutoff for specific applications can improve testing efficiency.

Conclusions

- Developed a robust, stable, and externally predictive model of reproductive toxicity
- Features and resulting model are easy to communicate and translate for decision making
- Model can readily be updated with new data or replacement of current assay data
- Model can be tuned to perform for different applications
- General model of reproductive toxicity, but provides mechanistic insight for targeted testing
- Recognized data gaps including metabolism, steroidogenesis, reactive chemicals
- Able to perform forward validation study using ToxCast Phase II data and chemical set