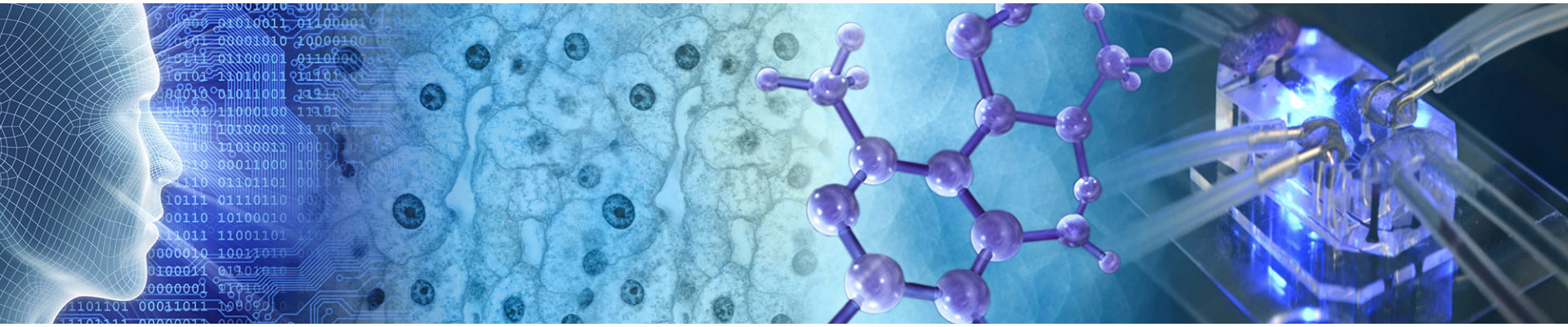
NTP
National Toxicology Program



Balancing Machine Learning and Mechanistic Modeling

Nicole C. Kleinstreuer
NICEATM Deputy Director

19th September, SACATM, Washington DC

Expert-driven vs. Data-driven

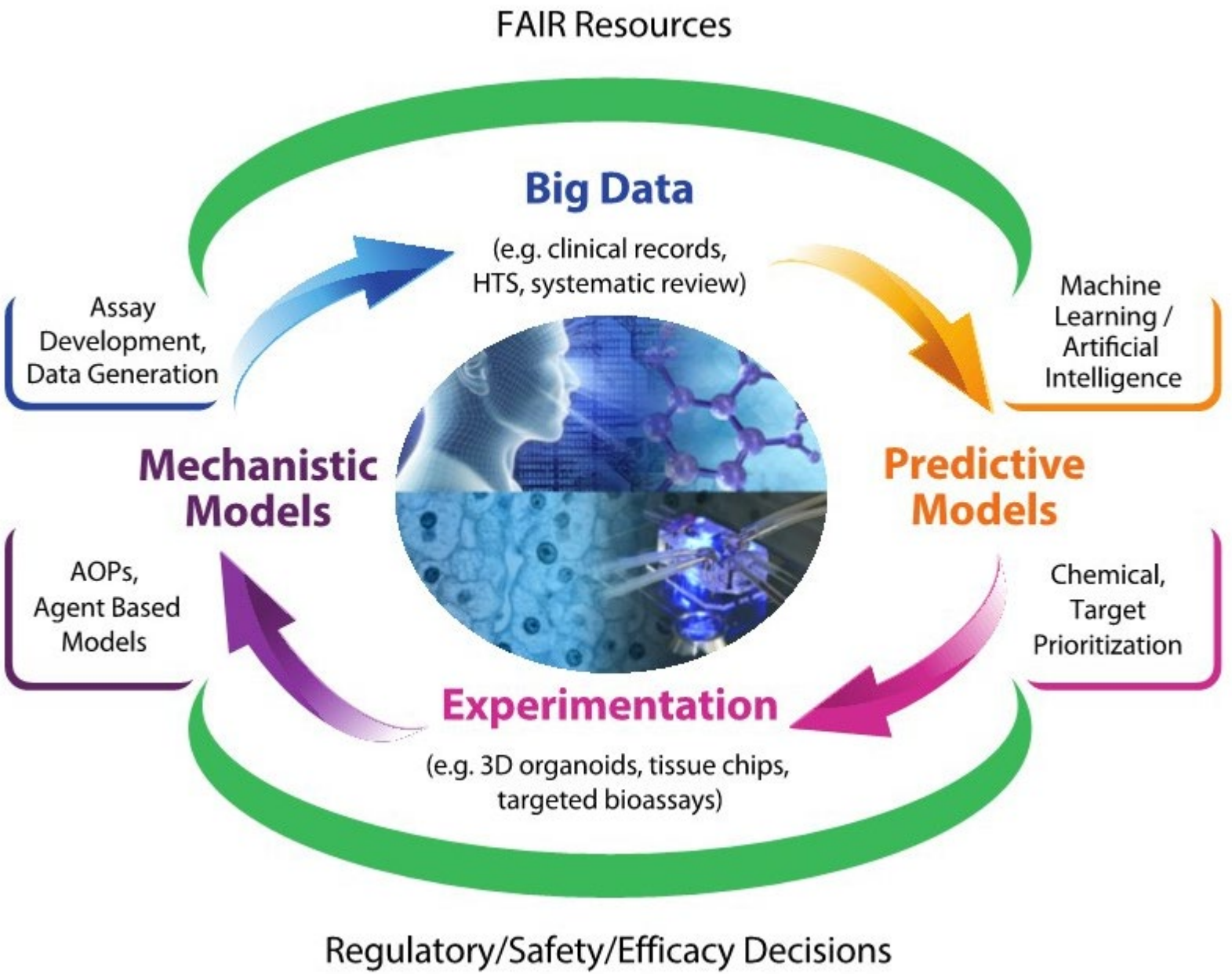
- Two “competing” approaches to modern toxicology/drug discovery:

1) Build testing strategies/models based exclusively on existing biological knowledge

2) Generate as much data as possible and let the machines sort it out

- Success lies in leveraging both approaches
- BUT this requires appropriate toolkits, resources and support infrastructure

Predictive Toxicology Vision



Making Data Systems and Resources FAIR

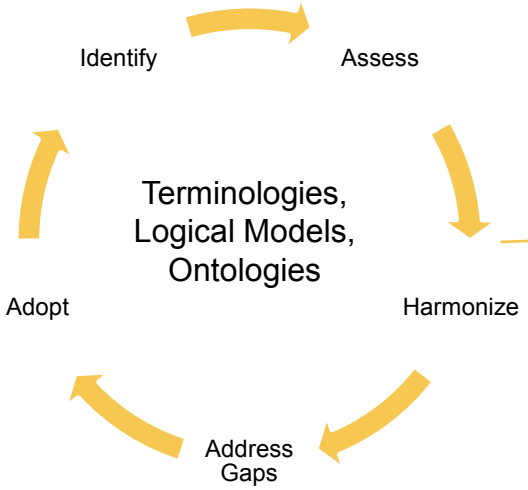
FAIR PRINCIPLES	
Findable	A data object should be uniquely and persistently identifiable.
Accessible	Data is accessible by authorized users (human and machine) through a well-defined protocol.
Interoperable	(Meta) data assigned to the data object is syntactically parse-able and semantically machine accessible.
Reusable	Data objects must comply with the above three principles and sufficiently documented to allow integration/linkage with other data sources.

https://www.force11.org/fairprinciples

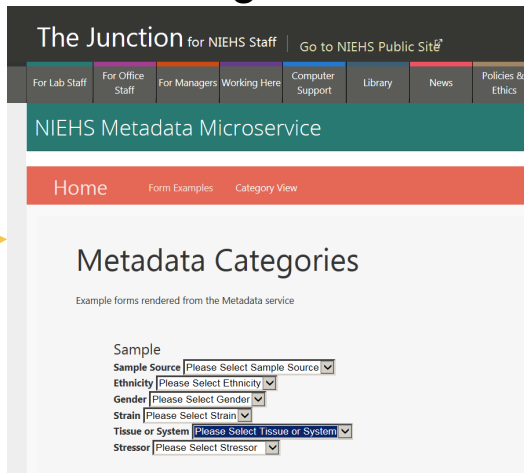
Wilkinson et al. 2016

NIEHS Data Commons – initial stage

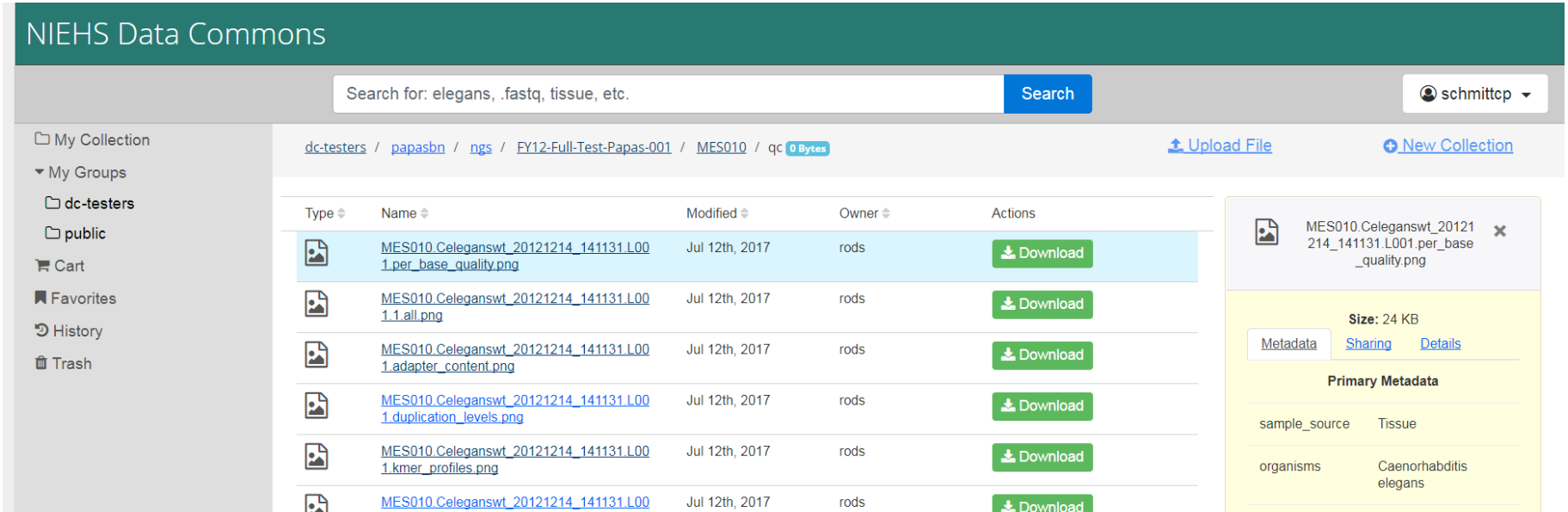
- Internal research data & metadata
- Capture, access, control, search, and sharing
- Engage external stakeholders



Live catalog

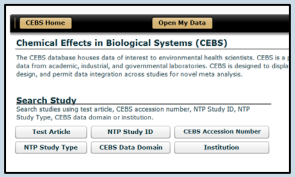


Courtesy of C. Schmitt



Interoperability Across Systems

Consistent & compatible web-APIs

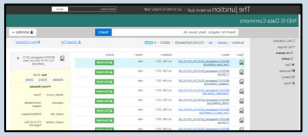


CEBS

Integrated Chemical Environment

ICE

Others…



Data Commons

NIEHS Data Systems

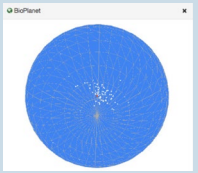
Consistent data set access & retrieval

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY

EPA Dashboard

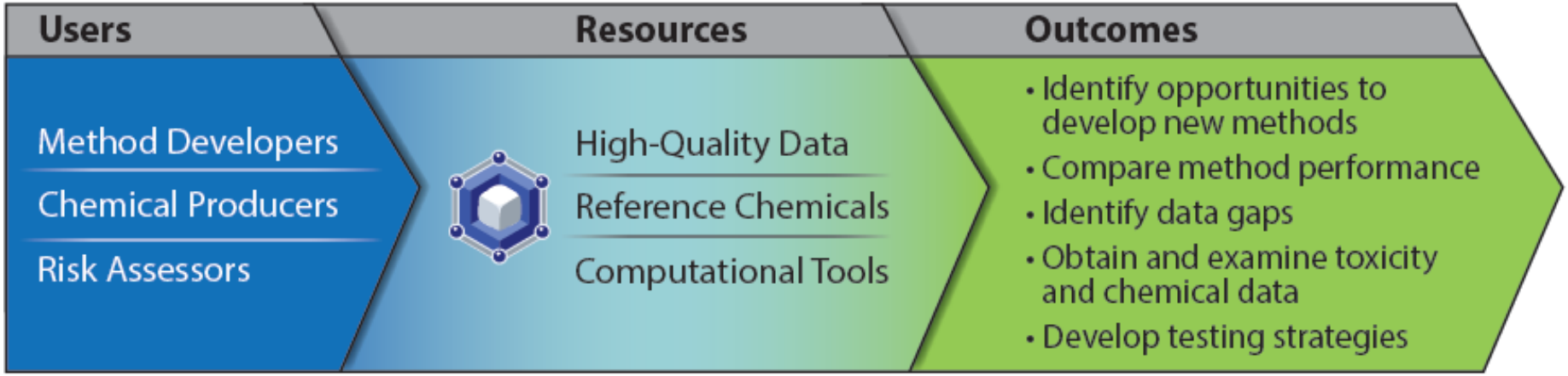
PubChem

Many others…



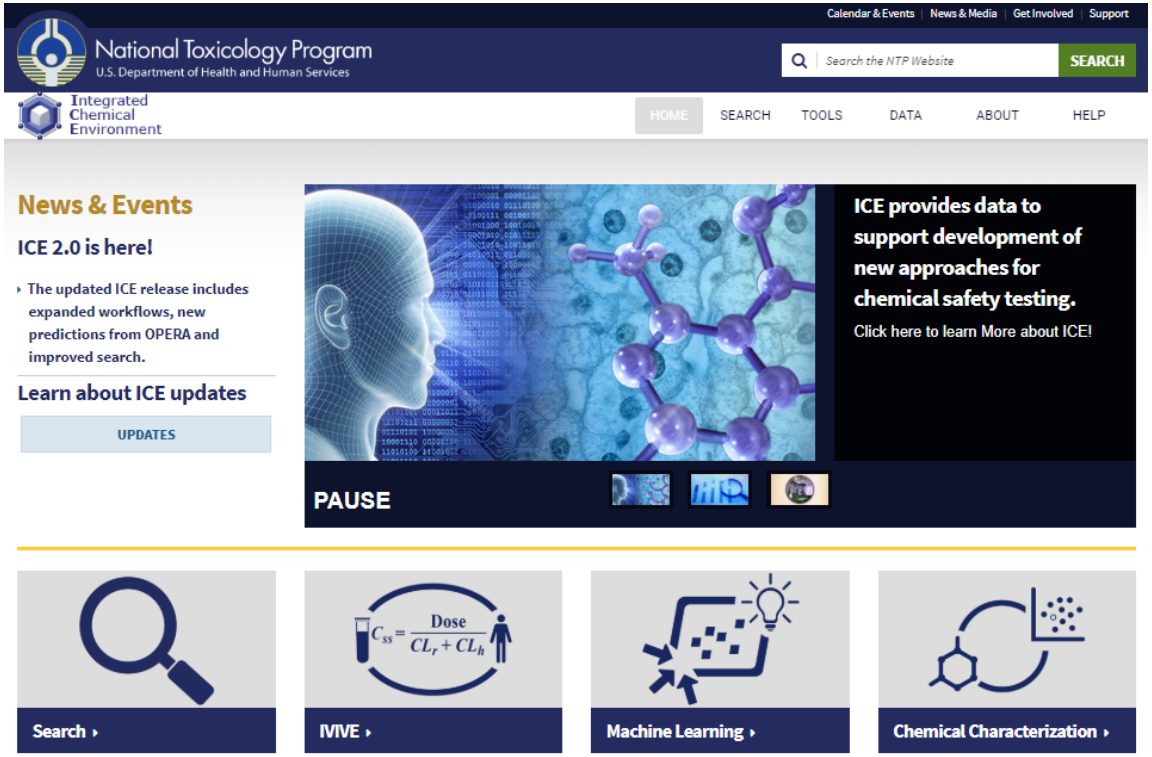
NCATS BioPlanet

Integrated Chemical Environment: ICE



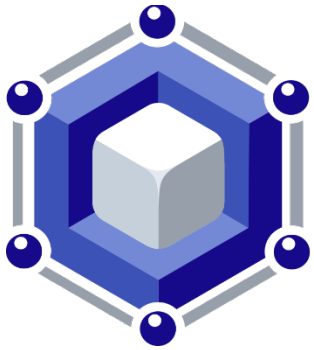
Bell et al. 2017 EHP

https://ice.ntp.niehs.nih.gov/



Goals of ICE

- Uphold FAIR principles for ICCVAM Data
- Provide intuitive access to high quality (curated) data and tools to support:
  - chemical evaluations,
  - data integration,
  - informatics analyses, and
  - model development
- Enable wider community to engage in the use of alternative and computational approaches for assessing chemical safety



Core Trustworthy Data Repository Certification

31 October 2019: Release of Final Core Trustworthy Data Repositories Requirements 2020–2022

- R0: Context
- R1: Mission/Scope
- R2: Licenses
- R3: Continuity of Access
- R4: Confidentiality/Ethics
- R5: Organizational Infrastructure
- R6: Expert Guidance
- R7: Data Integrity and Authenticity
- R8: Appraisal
- R9: Documented Storage Procedures
- R10: Preservation Plan
- R11: Data Quality
- R12: Workflows
- R13: Data Discovery and Identification
- R14: Data Reuse
- R15: Technical Infrastructure
- R16: Security

https://www.coretrustseal.org/

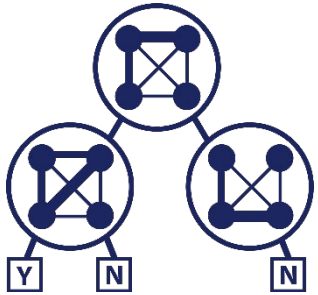
What goes into ICE?

Validation Studies

Databases

Published Data

Computational Models



What data are currently in ICE?

Toxicity endpoint/ Data source	Assays	# of chemicals
Acute Oral Toxicity	Acute oral toxicity	10,348
Skin Sensitization	DPRA, hCLAT, KeratinoSens, LLNA, human potency, etc	578
Skin Irritation	Acute skin irritation/corrosion, 4h HPT	120
Eye Irritation	Acute eye irritation/corrosion (e.g, Draize eye), Vitrigel	183
Endocrine	AR/ER Pathway Models, Uterotrophic, AR/ER binding	1903
cHTS	ToxCast and Tox21 assays	9076
OPERA predictions	BP, HLC, KOA, BCF, LogP, MP, MW, VP, WS	705,666
Formulation data	Acute 6-pack	298 (747 formulations)

Call for Data

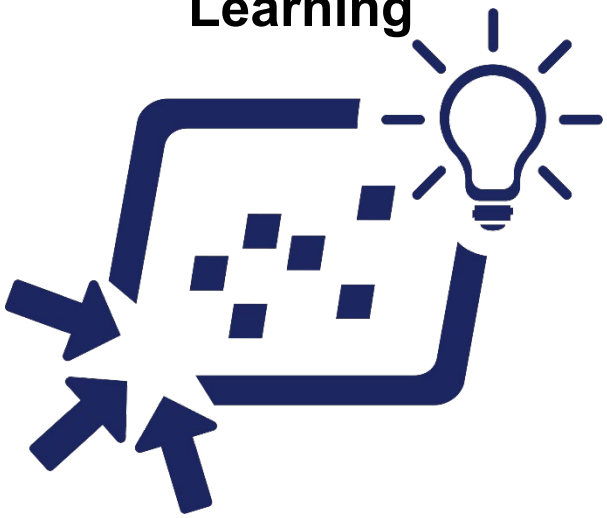
Currently of interest:

- *In vivo* data
  - Collections of data generated using regulatory guideline-like studies
  - Acute inhalation, skin and eye irritation/corrosion
- Toxicokinetic data
  - Collections of *in vivo* measurements
  - Data from *in vitro* assays aimed at informing modeling of chemical ADME within the body

ICE-support@niehs.nih.gov

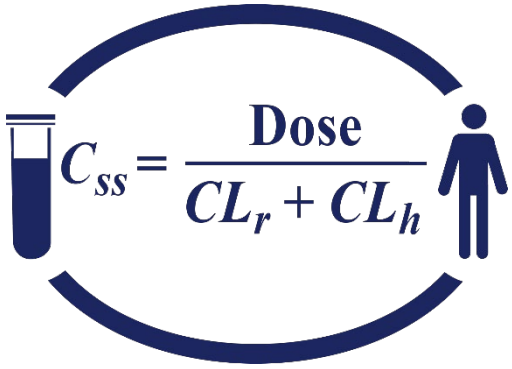
Tools

Machine Learning

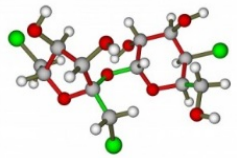


Chemical Space Characterization

IVIVE



Experimental measurement



QSAR or QPPR prediction

Hepatic clearance, fraction unbound to plasma protein (f_u), K_p, gut absorption, etc

One-compartment PK or PBPK models

Plasma concentration at dose of 1 mg/kg/day (C_{ss}, $C_{ss} * f_u$ or C_{max})

$$C_{ss} = \frac{\text{Dose}}{CL_r + CL_h}$$

In Vitro HTS Assays

Activity concentration (AC)

Equivalent

In Vivo Effects

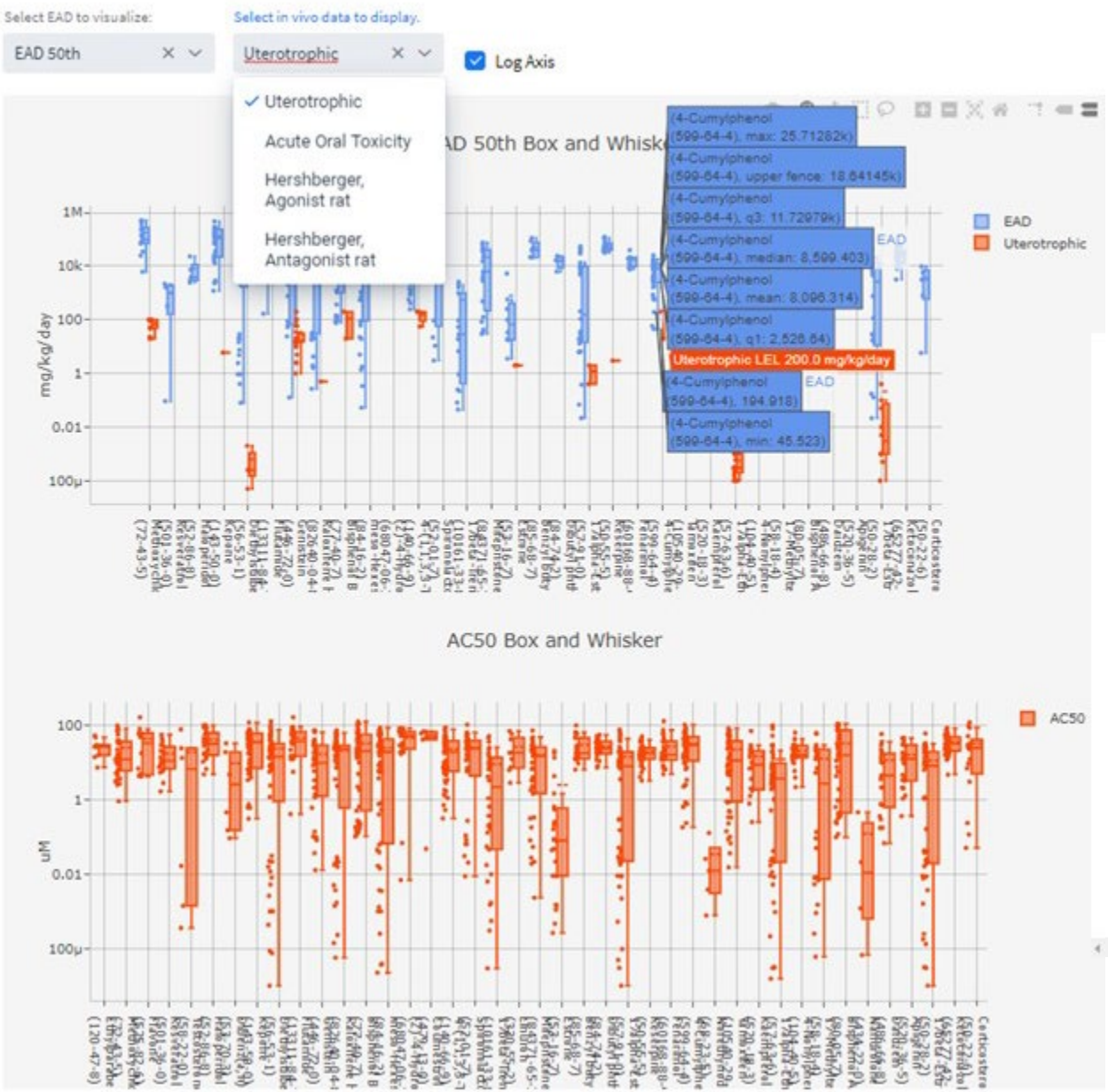
Target plasma concentration

Reverse dosimetry

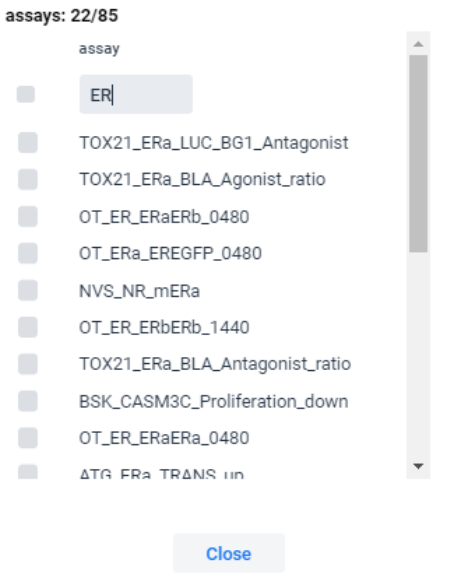
Range of equivalent administered doses (EADs) (mg/kg/day)

$$AC * \left(\frac{1 \text{ mg/kg/day}}{\text{plasma concentration at dose of1 mg/kg/day}} \right)$$

Overlay In Vivo Data on IVIVE Results



- Overlay in vivo data
- Update the plots with assay filtering
- Download data to explore locally



Predicting Key Toxicokinetic Parameters

Multiple Machine Learning Models (SVM, DNN, XGB, etc.)

- ADME properties
  - Plasma fraction unbound (FuB)
  - Intrinsic clearance (Clint)
- Tissue partition coefficient inputs
  - pKa
  - Log D

Mansouri et al. 2019 Journal of Cheminformatics in press

https://github.com/NIEHS/OPERA

Global Collaborative Projects

Applying machine learning to predict endpoints of regulatory importance

CERAPP

Collaborative Estrogen Receptor Activity Prediction Project (2015/16)

CoMPARA

Collaborative **M**odeling **P**roject for **A**ndrogen **R**eceptor **A**ctivity **(2017/18)**

CATMoS

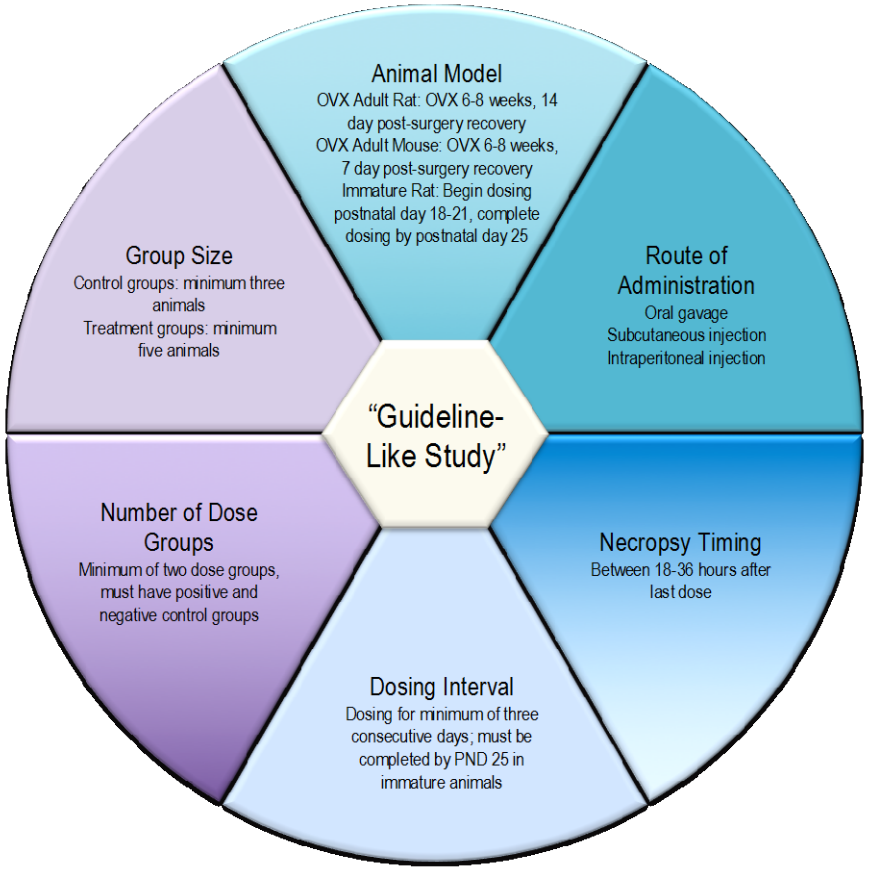
Collaborative **A**cute **T**oxicity **Mo**deling **S**uite **(2018/19)**

Endocrine Disruptor Screening Program (EDSP)

ICCVAM Acute Systemic Toxicity Workgroup

Mansouri et al. 2016 EHP 124:1023–1033
Mansouri et al. 2019 under revision at EHP
Kleinstreuer et al. 2018 Comp Tox; Mansouri et al. 2019 in prep

Manually Identifying Reference Data

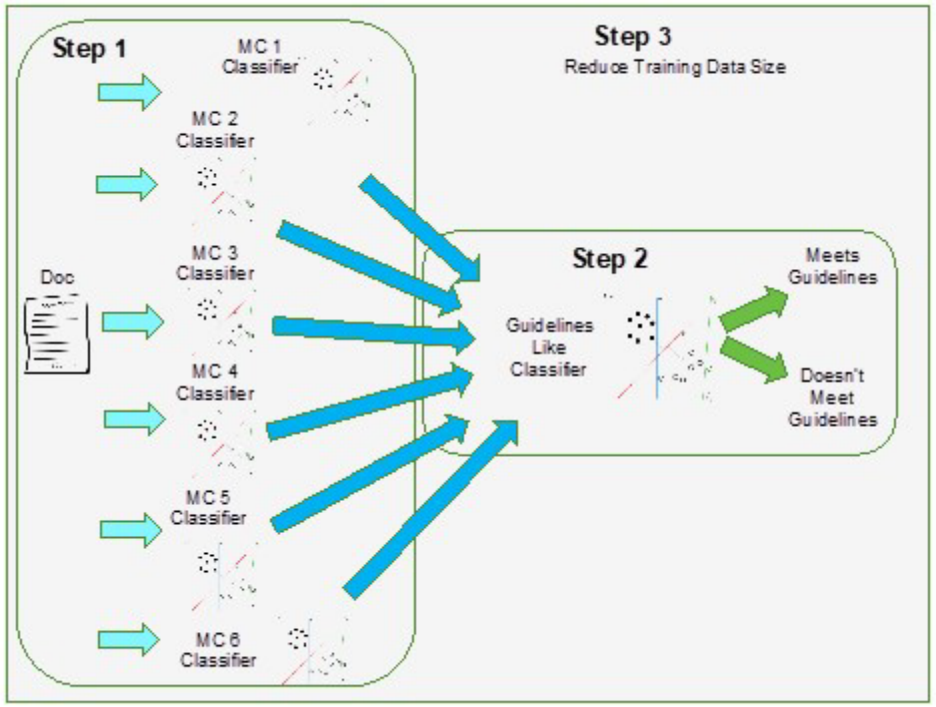


Ex: Uterotrophic Database

- Systematic literature search of publically available data (e.g. PubMed, Scopus)
- Identify chemical activities measured in "guideline-like" uterotrophic studies
- Identify a subset of *in vivo* reference chemicals
  - Active chemicals verified in ≥2 independent studies
  - Inactive chemicals verified in ≥2 independent studies (with no positive results in any study)

Kleinstreuer et al. EHP (2015)

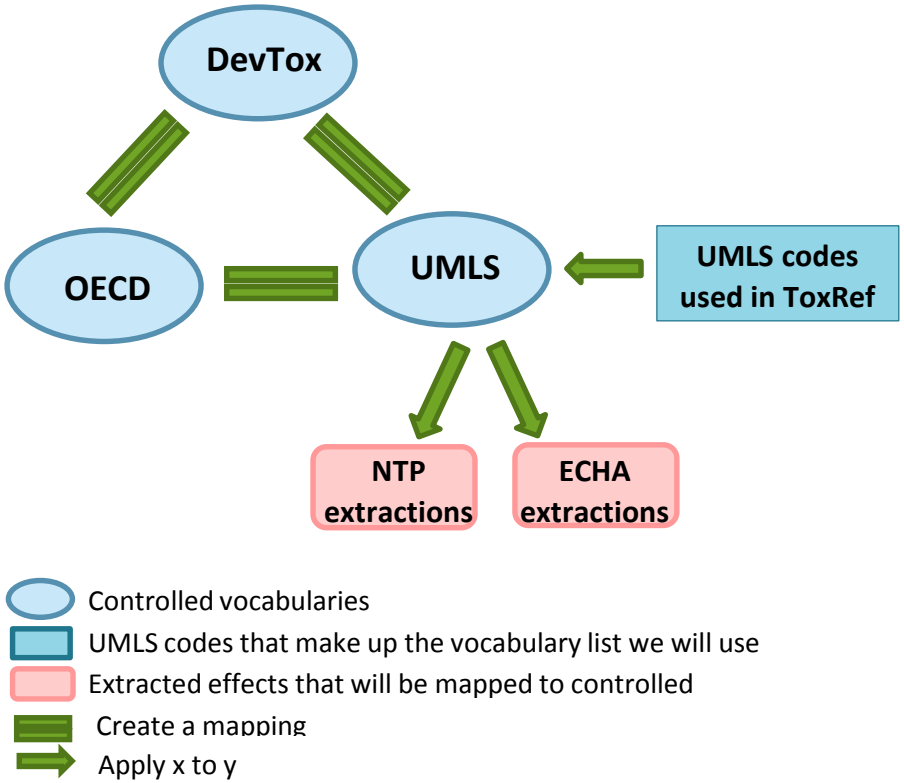
Automating Reference Data Identification



- Project with Oak Ridge National Labs (ORNL) and FDA CFSAN to apply text-mining (NLP) approaches & ML to identify high-quality data
- Semi-automated retrieval and evaluation of published literature (trained on uterotrophic database)
- Apply to developmental toxicity studies (with ICCVAM DARTWG)
  - Define literature search keywords, identify corpus
  - Extract/characterize study protocol details from regulatory guidelines: minimum criteria
  - Apply ML algorithms to identify high-quality studies, expert check

Study Extractions and Endpoint Mapping

- Extract study details from prenatal developmental toxicity guideline studies
  - NTP legacy studies
  - ECHA submissions (expert reviewed for quality)
- Map results to controlled vocabularies/ontologies
  - UMLS (ToxRefDBv2.0)
  - EPA/BfR DevTox DB
  - OECD Harmonized Templates

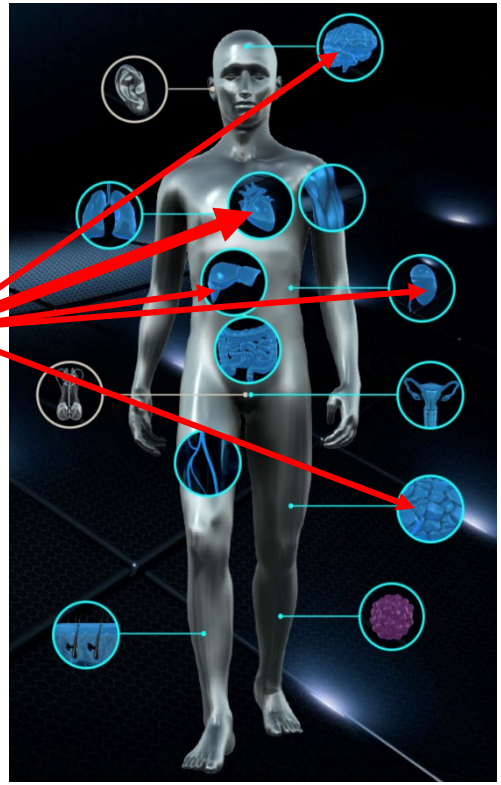


Flipping the Paradigm: Mechanistic Screening

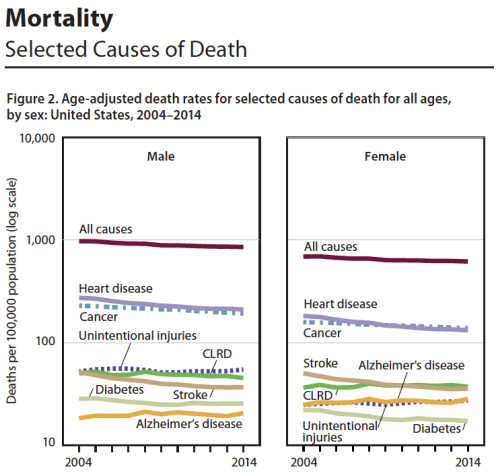
Initial Focus

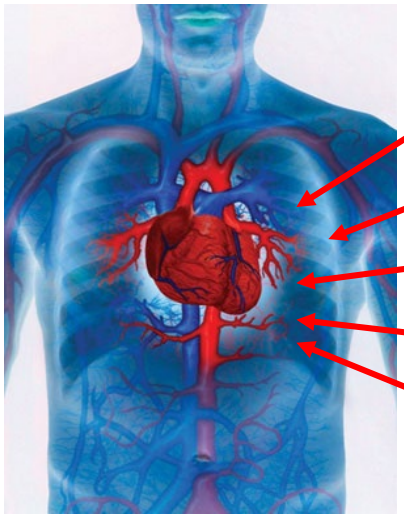
X lbs./yr. commercial production

Agent Y



https://ncats.nih.gov/tissuechip





Agent A

Agent B

Agent C

Agent Y

Agent Z

Courtesy of B. Berridge

Cardiovascular Health Effects Strategy

Contributing projects
Partners and opportunities

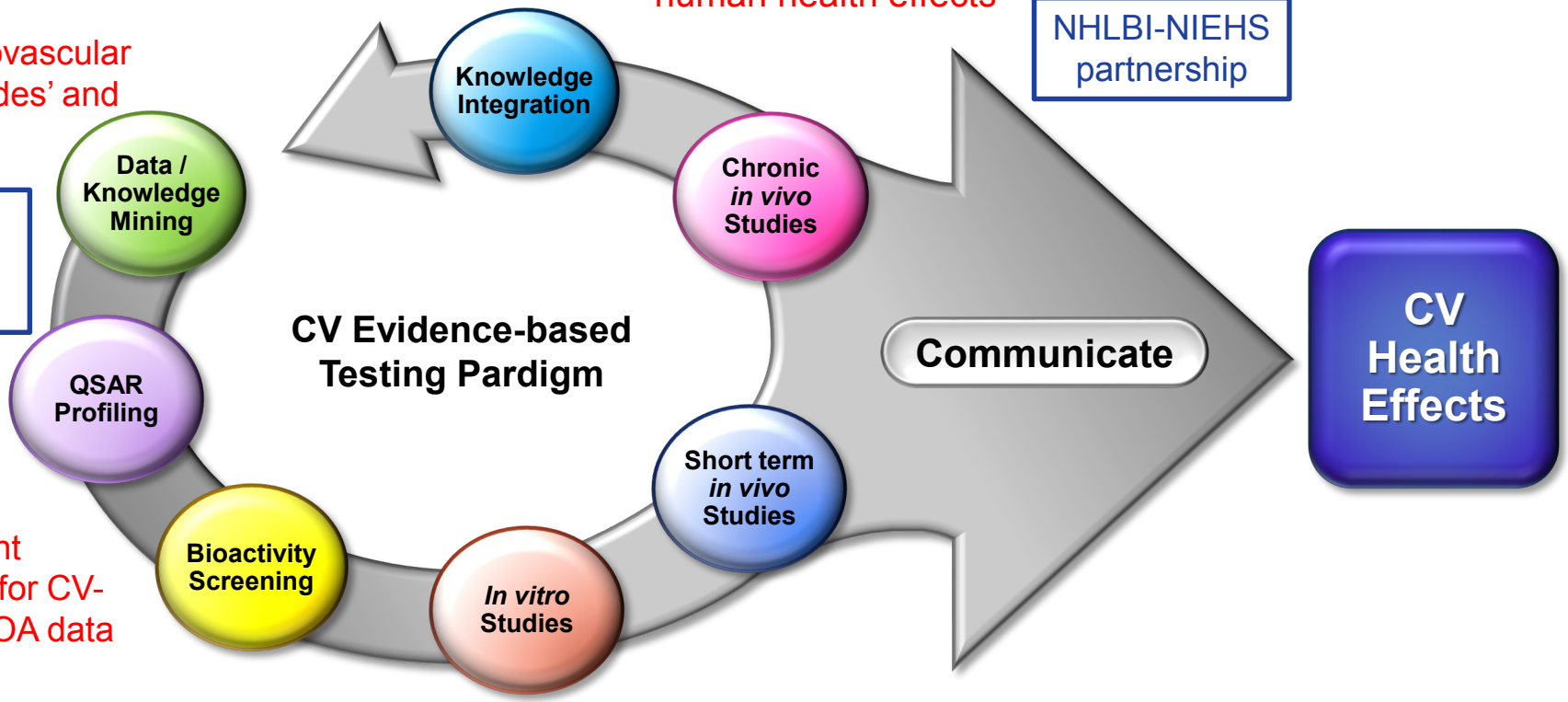
Regulatory and public health stakeholders

Identify translational biomarkers of chronic CV human health effects

NHLBI-NIEHS partnership

Map cardiovascular 'failure modes' and agents

FDA/NTP PostDoc Fellows



Mine current databases for CV-relevant MOA data

HESI Cardiac Safety Technical Committee

Define primary mechanistic screening strategy

Define and test Tier II 3D cardiomyocyte and vascular modeling systems

NCATS, IQ, NCTR partnerships

Adapted from B. Berridge - NTP BSC Presentation- Dec. 2018

Cellular Events Linked to CV Failure Modes

Drug actions on human receptors, ion channels, cellular processes

Potency + Exposure (dose, time)

1° Failure modes

- **Δ Vasoactivity**
- **Δ Action potential**
- **Δ Inotropy**
- **Cardiomyocyte/ myocardial injury**
- **Valvular injury/proliferation**
- **Endothelial injury/coagulation**

Systemic_Hypertension
- Antagonism at α2-receptors,
- Rebound phenomenon (α2-agonists and beta-blockers)
- Agonism at glucocorticoid receptors
- Inhibition of VEGF pathway,
- Inhibition of monoamine oxidases (MAO)
- 11β-hydroxysteroid dehydrogenase type 2 inhibition

Systemic Hypotension
- Antagonism at α1-receptors,
- Ca2+ channel blockade
- Opening of K+ channels,
- Inhibition or renin–angiotensin–aldosterone axis
- Agonism at β2 receptors
- Agonism at α2-receptors,
- Agonism at I1-receptors,
- Stimulation of cGMP Synthesis,
- Inhibition of phosphodiesterase 5

Left ventricular (LV) dysfunction/heart failure
- Ca2+ channel blockade
- Na+ channel blockade,
- Antagonism at β1-receptors
- Anthracyclines, cyclophosphamide (high dose), taxanes
- HER2 signaling inhibition
- VEGF signaling inhibition,
- tyrosine kinase inhibition (multikinase drugs),
- Proteasome inhibition

Bradydysrhythmias
- Ca2+ channel blockade,
- Na+ channel blockade,
- Blockade of If current
- Antagonism at β1-receptors,
- Agonism at M-receptors
- Agonism at I1-receptors,
- Agonism at sphingosine-1-phosphate receptor,

Tachydysrhythmias
- Blockade of hERG channels
- Agonism at β-receptors
- Antagonism at M-receptors
- Anhibition of Na+/K+ pump,
- Na+ channel blockade

Myocardial ischemia
- Agonism at β1-receptors (direct effects or indirect effect via endogenous Catecholamines)
- Rebound phenomenon (nitrates, β-blockers)

Myocarditis
- Autoimmune reactions (e.g., clozapine)
- monoclonal antibodies targeting PD-1

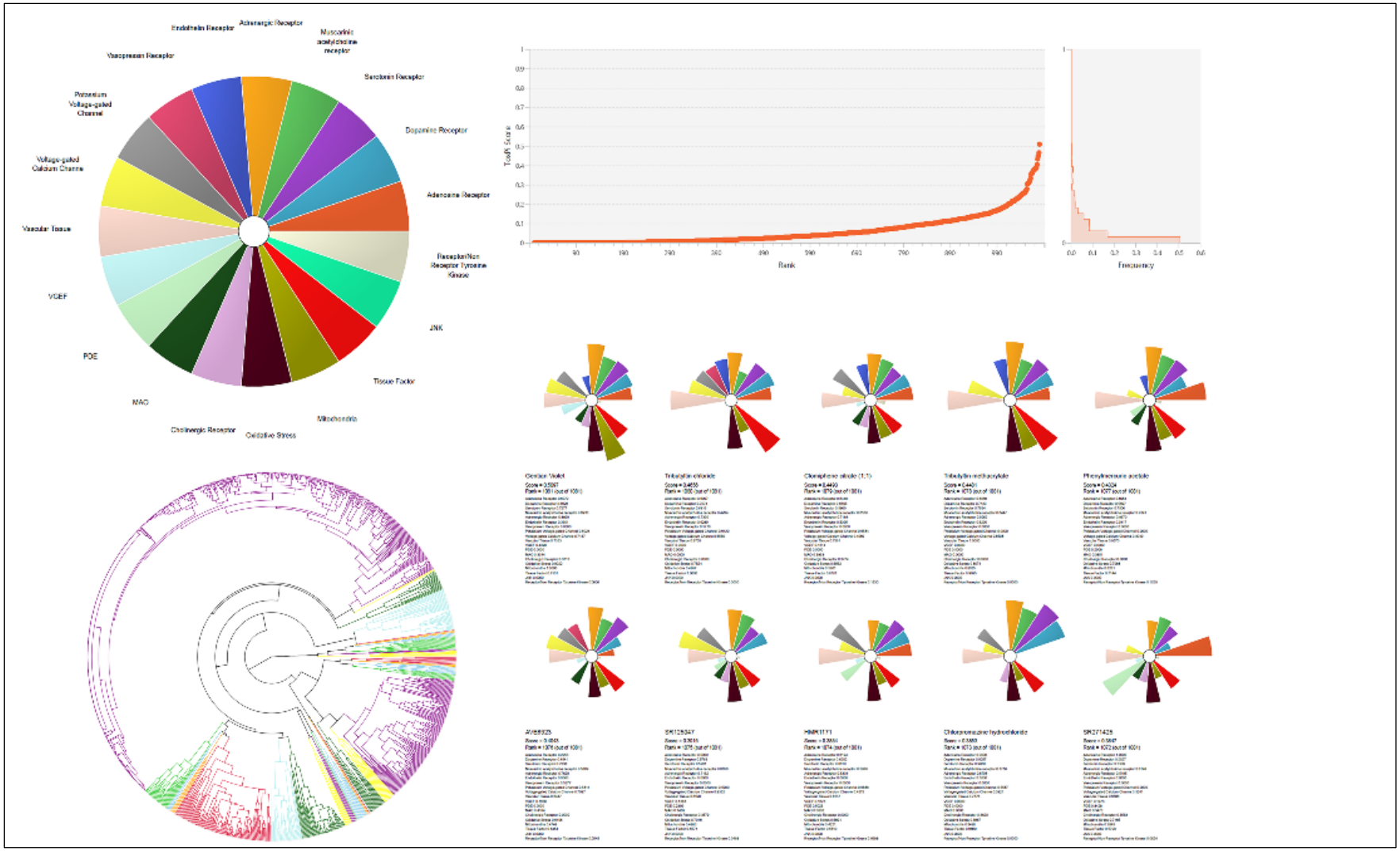
Impairment of cardiac valves
- Agonism at 5-HT2B receptors

Pericardial disease induction
- Immune reaction (e.g., drugs inducing lupus erythematosus)

Arterial
- Inhibition of cyclooxygenase 2
- VEGF targeting
- Agonism at erythropoietin receptors

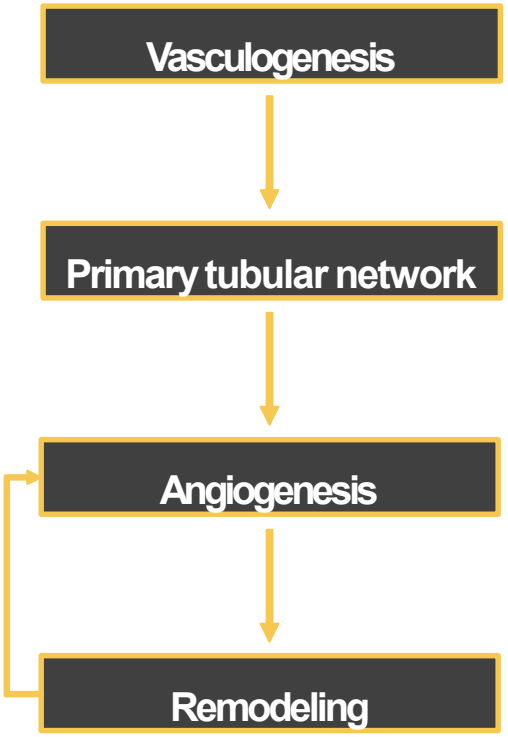
Venous
- Agonism at estrogenic receptors
- VEGF targeting
- Agonism at erythropoietin Receptors

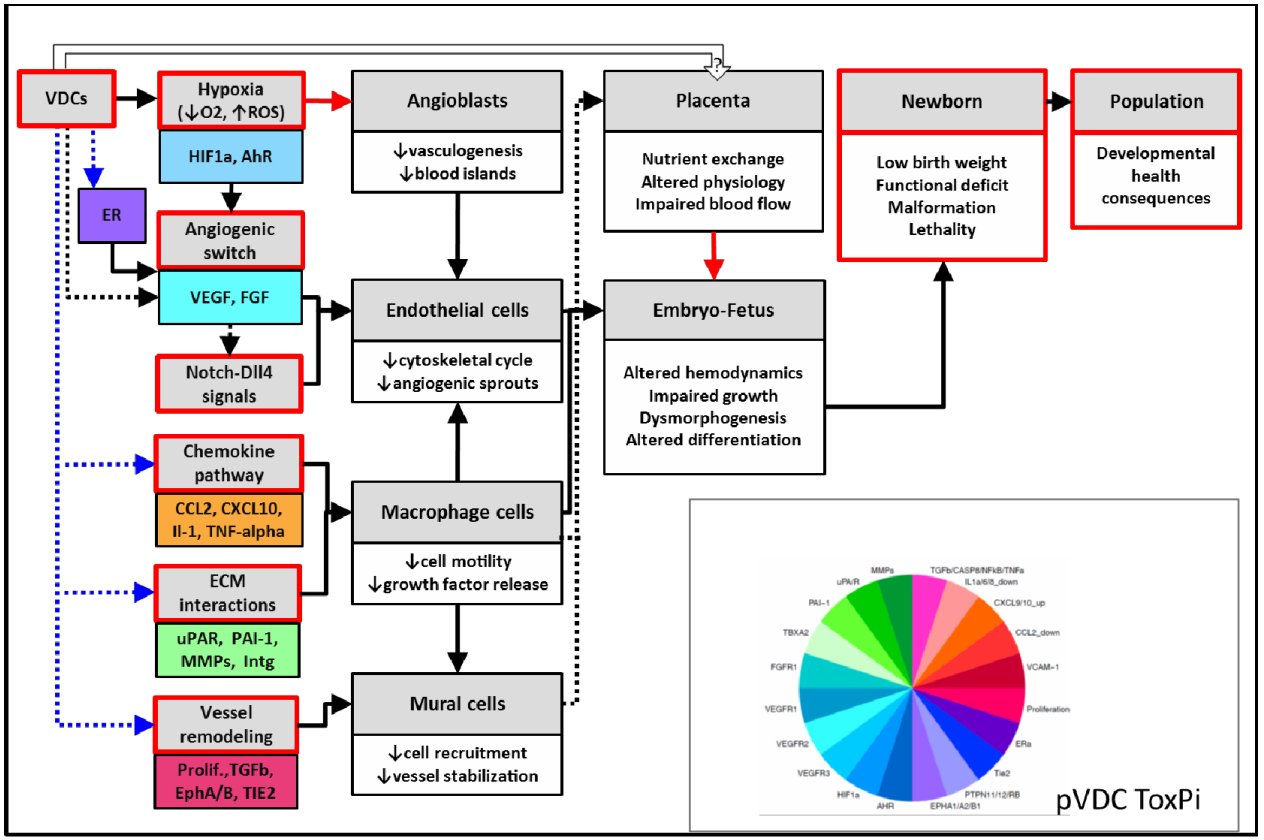
CardioToxPi: HTS Assay Mapping



Vascular Development & Disruption

Adverse Outcome Pathway (AOP)

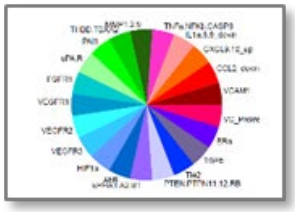


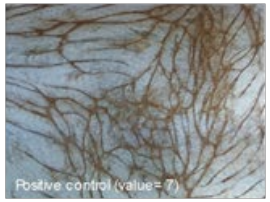


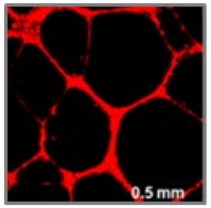
Knudsen and Kleinstreuer (2011) Birth Defects Res

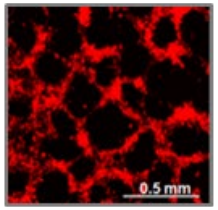
AOP43: one of 28 AOPs included in the OECD work plan with status 'open for citation & comment' https://aopwiki.org/wiki/index.php/Aop:43

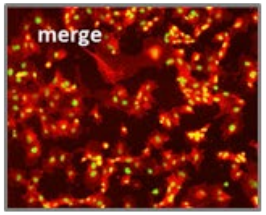
Mechanistic Models and Experimental Results

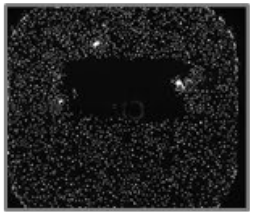


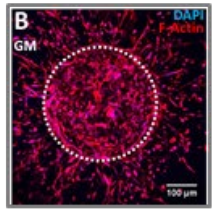


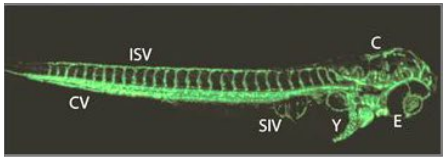


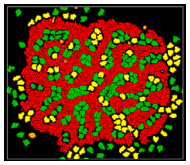


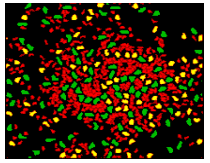


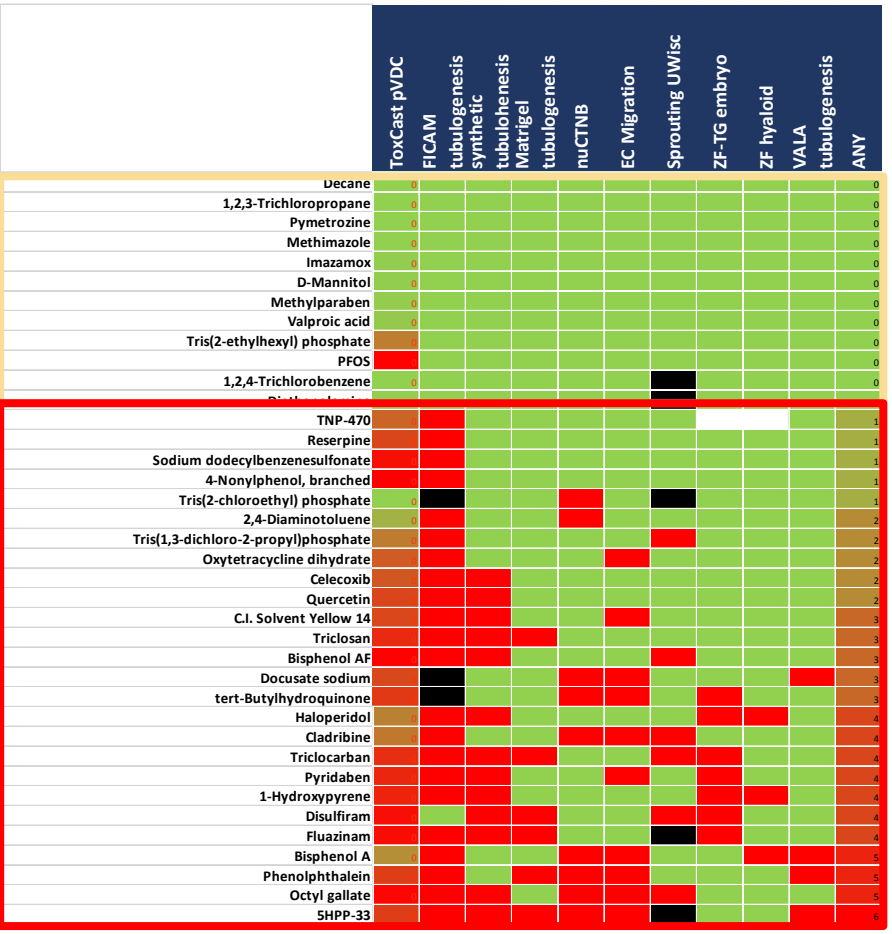












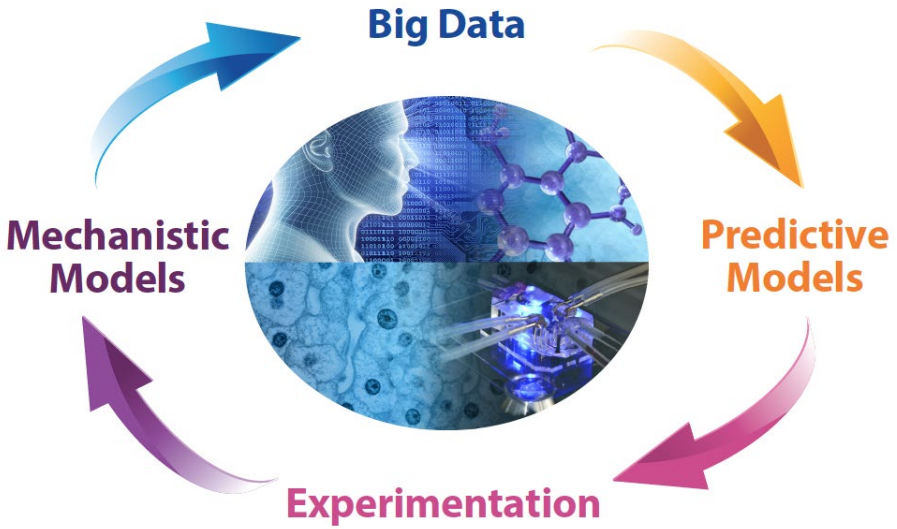
38 chemical test set: qualification of pVDC ToxPi across 9 endothelial behaviors

A pVDC score from ToxCast dataset (ToxPi)
B HUVEC tubulogenesis (FICAM)
C tubulogenesis in synthetic matrices
D tubulogenesis in Matrigel
E nuCTNB biomarker (EndMT)
F endothelial cell migration
G sprouting assay (iPSC-derived endothelial cells)
H reporter zebrafish (ISV outgrowth)
I reporter zebrafish (hyaloid vascular network)
J HUVEC tubulogenesis (VALA)
K ANY (B to J)

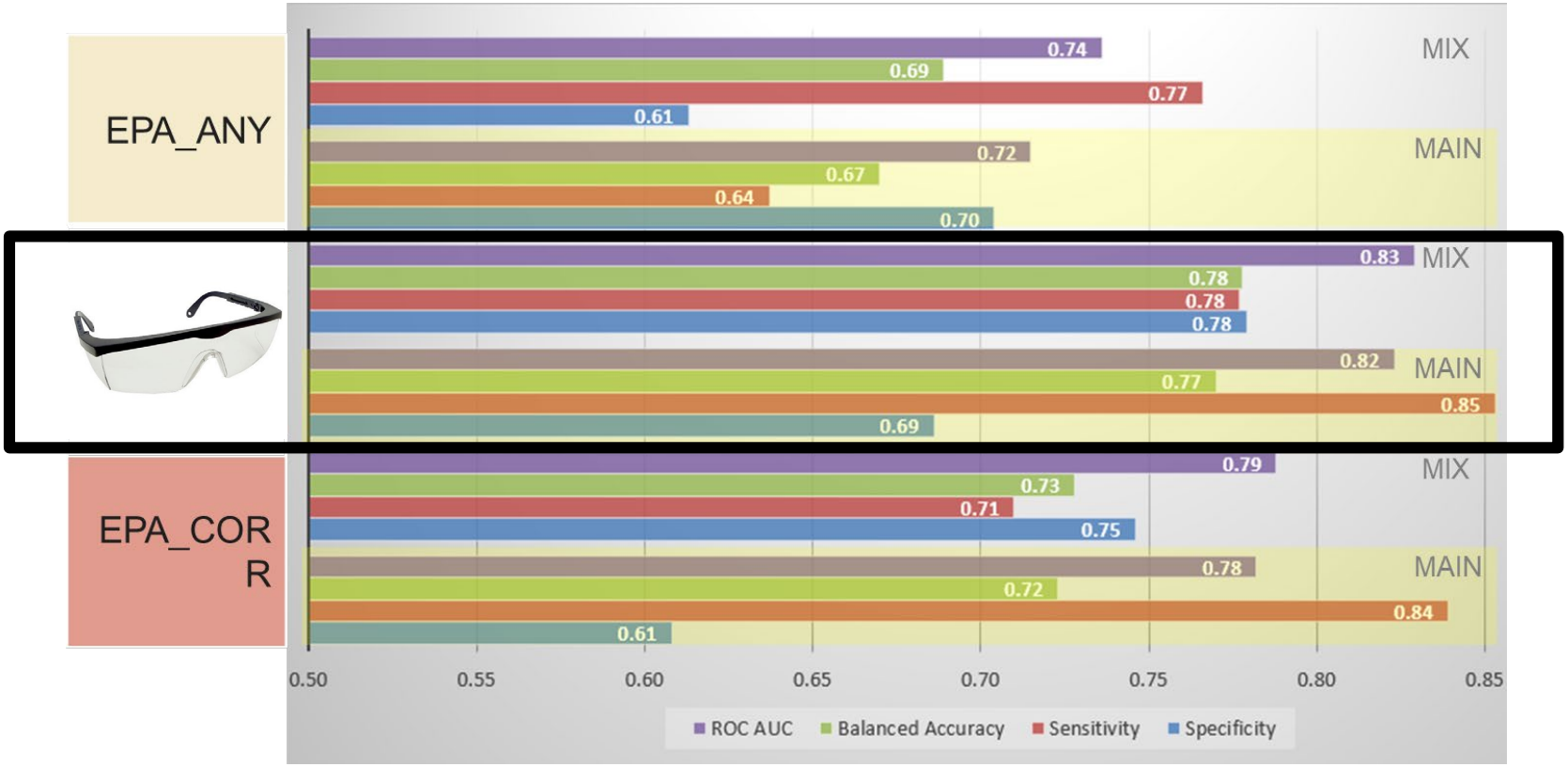
Saili et al 2019 Current Opinion in Toxicology

Predicting Toxicity of Mixtures

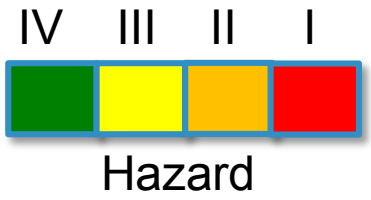
- How can we leverage machine learning, mechanistic modeling, and systems approaches to tackle complex problems such as predicting mixtures toxicity across heterogeneous populations?
- How do we build datasets that will allow models for mixtures toxicity against human health endpoints to be more effectively developed?



Ocular QSAR Mixture Models

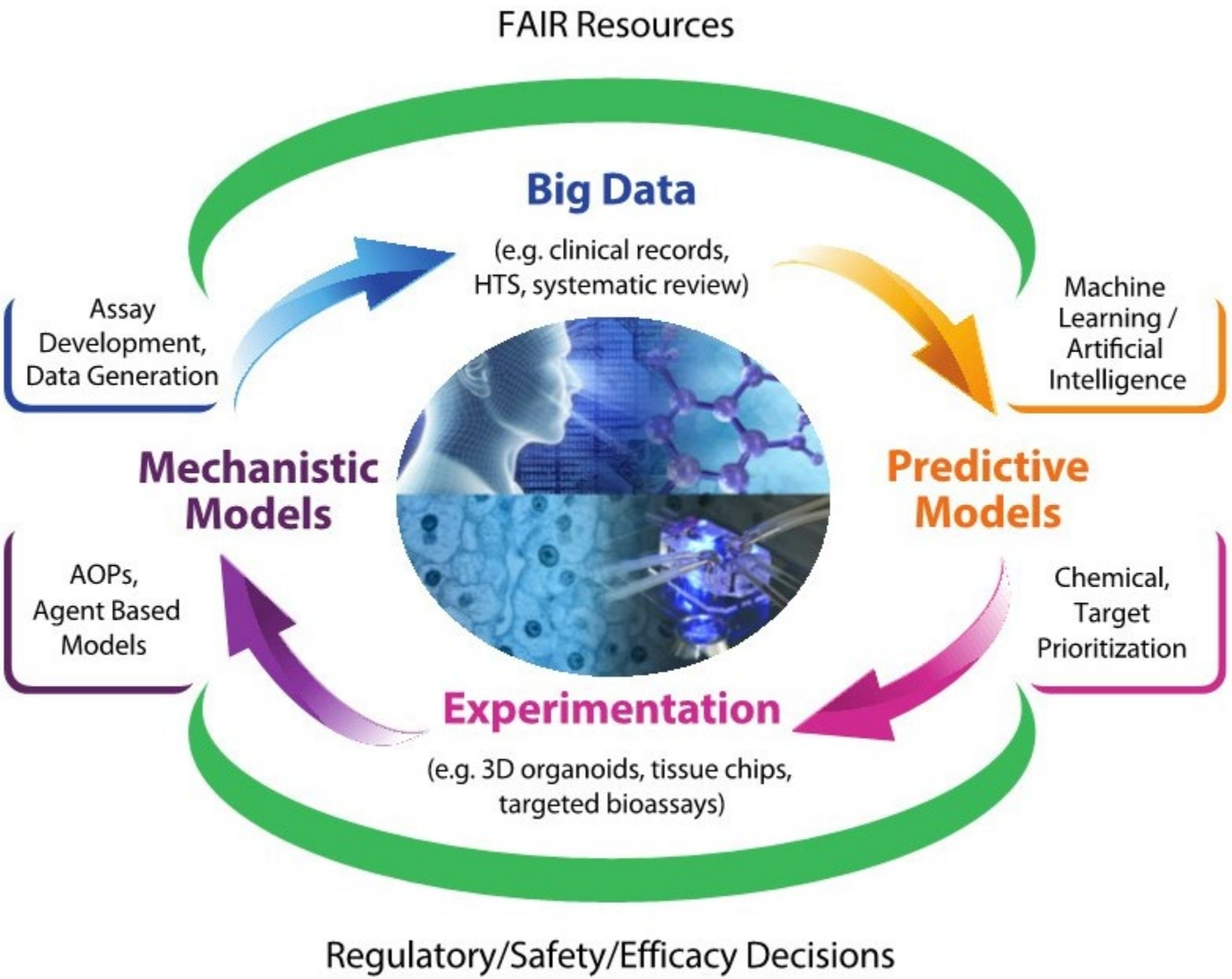


- EPA_ANY = Category I, II, III vs Category IV
- EPA_IRR = Category I/II vs Category III/IV
- EPA_ Corr = Category I vs Category II, III, IV



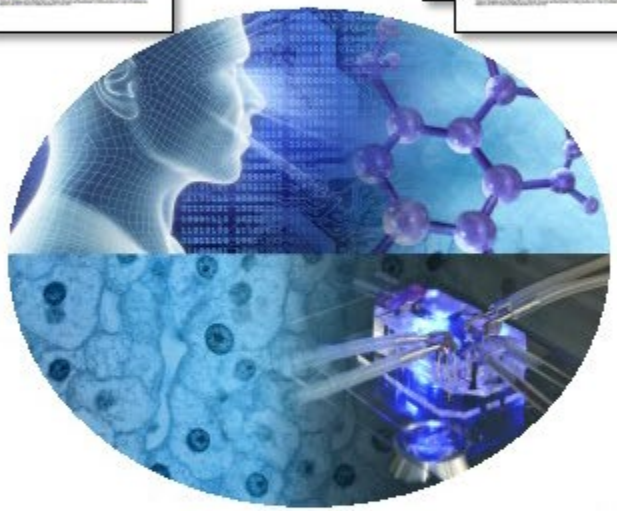
Sedykh et al. SOT 2019

Predictive Toxicology Vision



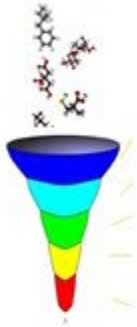
Predicting Human Toxicity

Big Data

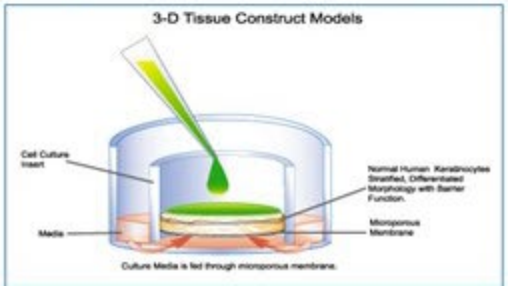


Mechanistic Models

Predictive Models



Experimentation



Acknowledgments

- ILS/NICEATM group
- Sciome collaborators
- ICCVAM agencies
- ICATM partners
- Brian Berridge (NTP)
- Shagun Krishna (Kelly)
- Robert Patton (ORNL)
- Jessica Wignall (ICF)
- Tom Knudsen (EPA)
- Todd Zurlinden (EPA)
- Kate Saili (EPA)
- Tony Williams (EPA)

Questions?