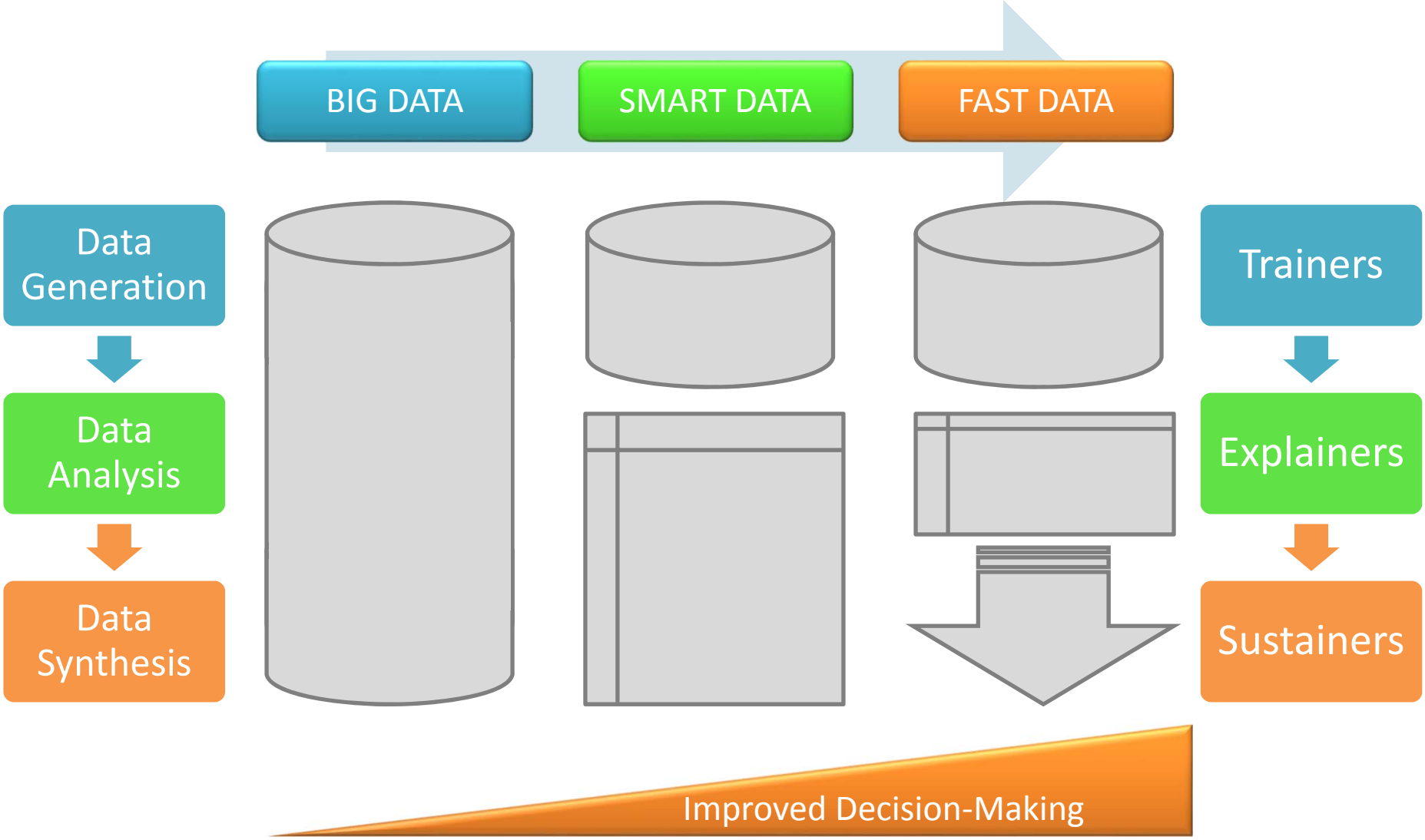
Quality Assessment of Big and Complex Data in Pharmaceutical Target and Chemical Safety Assessment

Matt Martin (Computational Toxicology Lead)

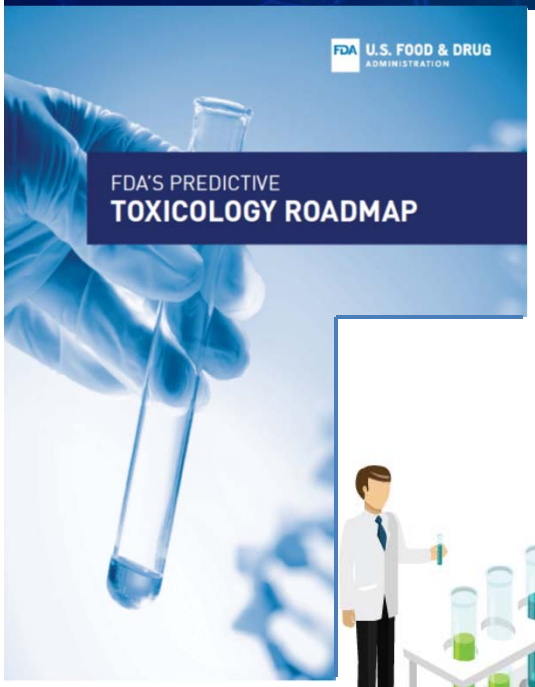
Pfizer WORLDWIDE RESEARCH & DEVELOPMENT

SCIENCE FOR LIFE-CHANGING IMPACT

Big Data to Smart Data to Fast Data



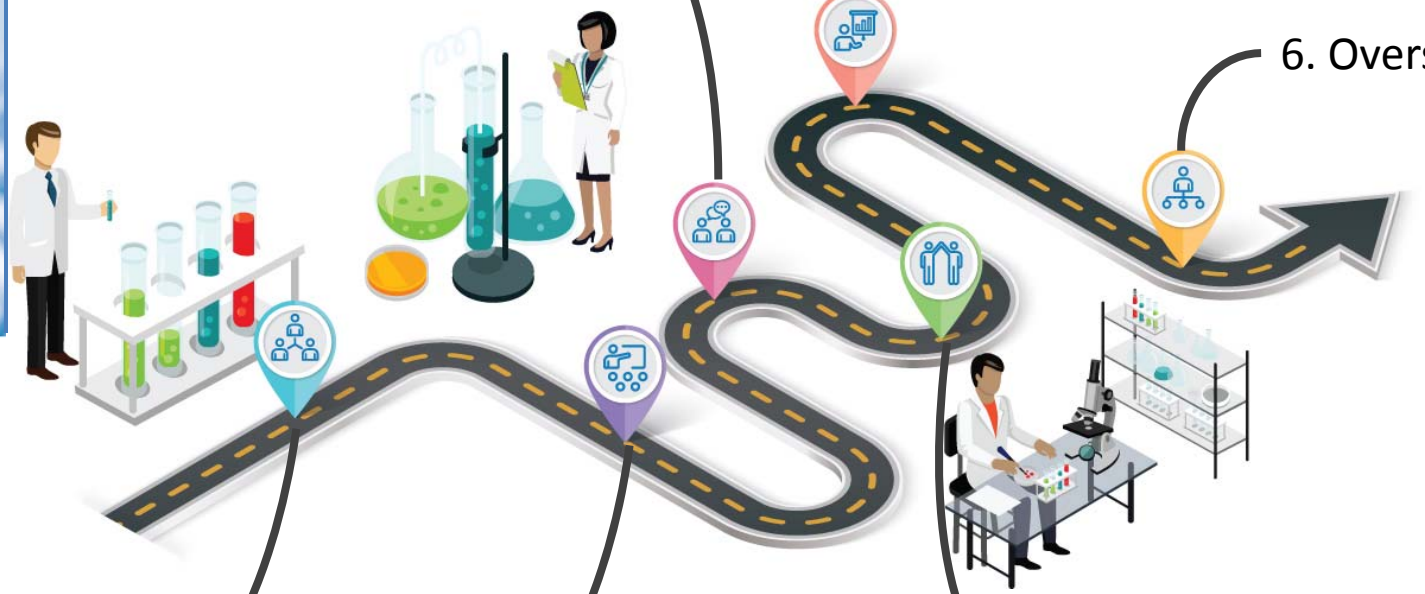
Preparing for a Changing Regulatory Landscape



3. Communication

5. Research
"Organs on a chip"
Alt. Test Methods
CompTox/Tox21
Read-Across

6. Oversight



4. Collaboration

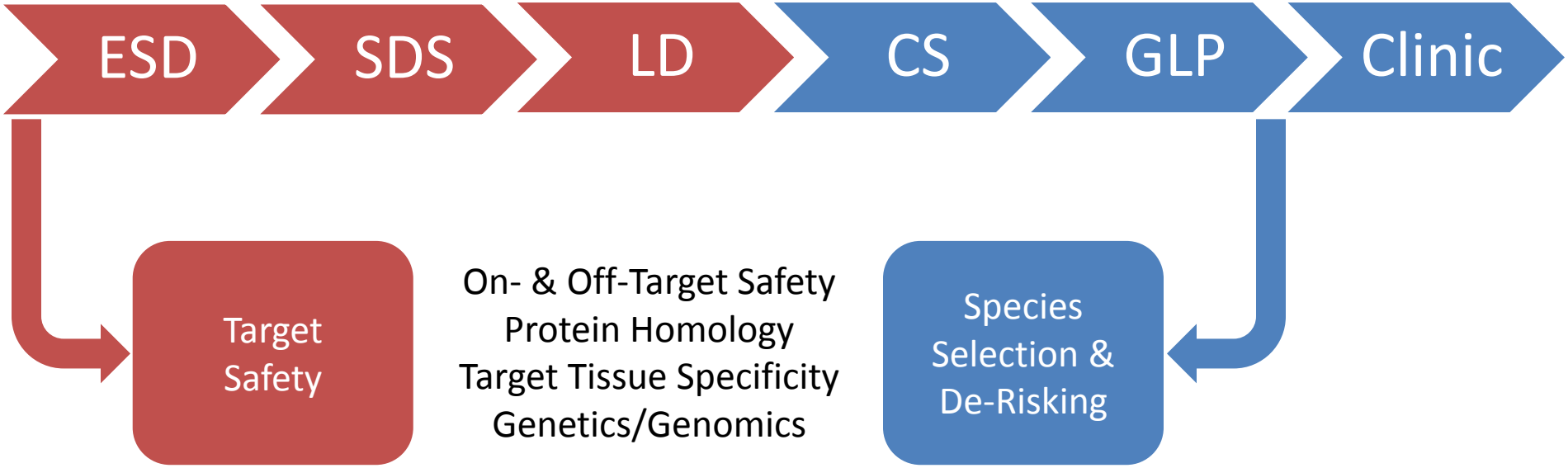
1. Organizing

2. Training

Computational Toxicology @ Pfizer

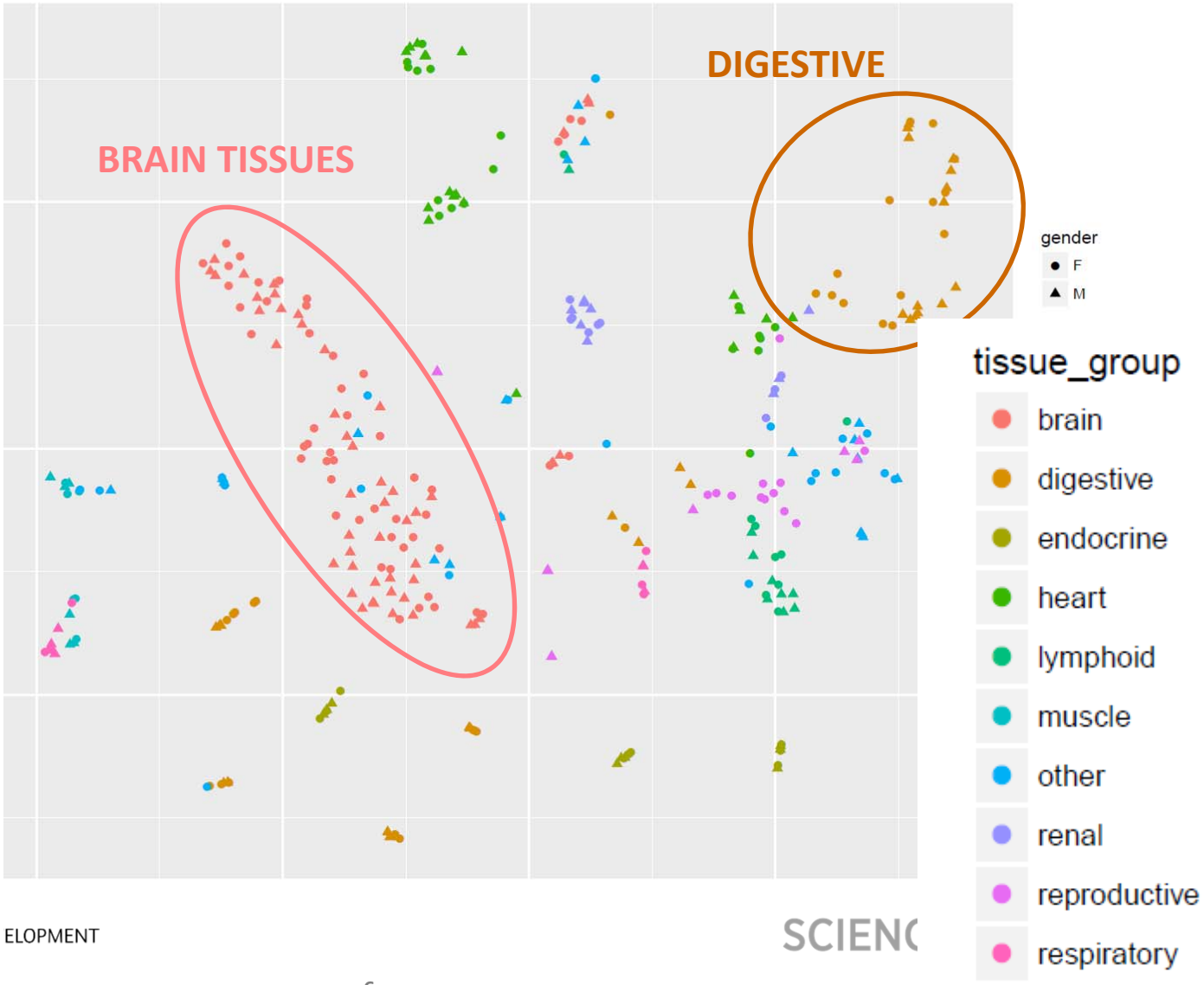
- **ESD** Target ID to Early Screen Development
- **SDS** Screening/Designed Synthesis
- **LD** Lead Development
- **CS** Candidate Seeking
- **GLP** Good Laboratory Practice Preclinical Studies
- **CLINIC** Clinical Trials to Post-Market

Computational Toxicology @ Pfizer

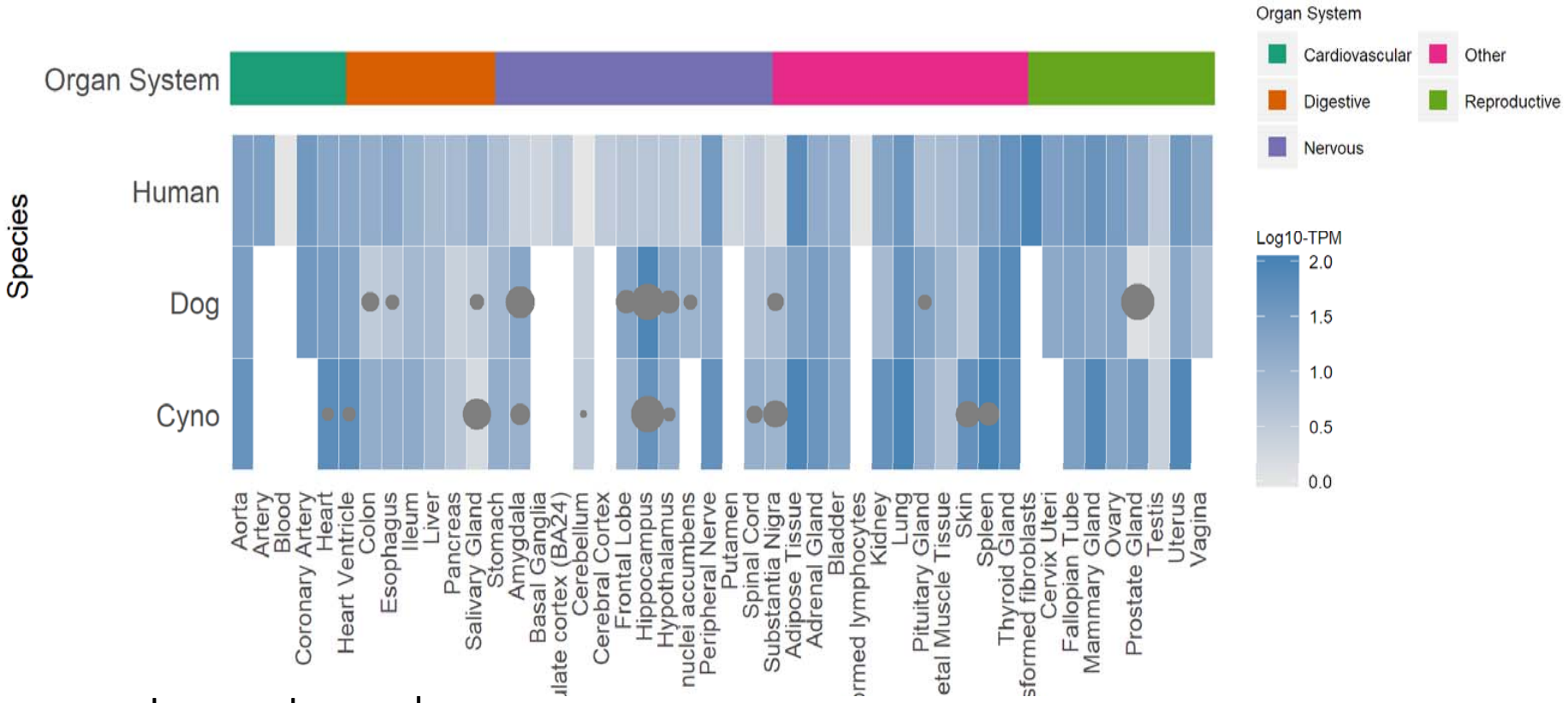


Target Tissue Specificity in Preclinical Species

- Developing comprehensive and highly comparable preclinical tissue maps (cyno, dog, rat, mouse)
- TBs of RNAseq data has been generated
- Calculating tissue specificity scores for all genes (e.g., Tau)
- Performing cross-species comparisons
- Improving Cynomolgus monkey genome annotation via resequencing effort



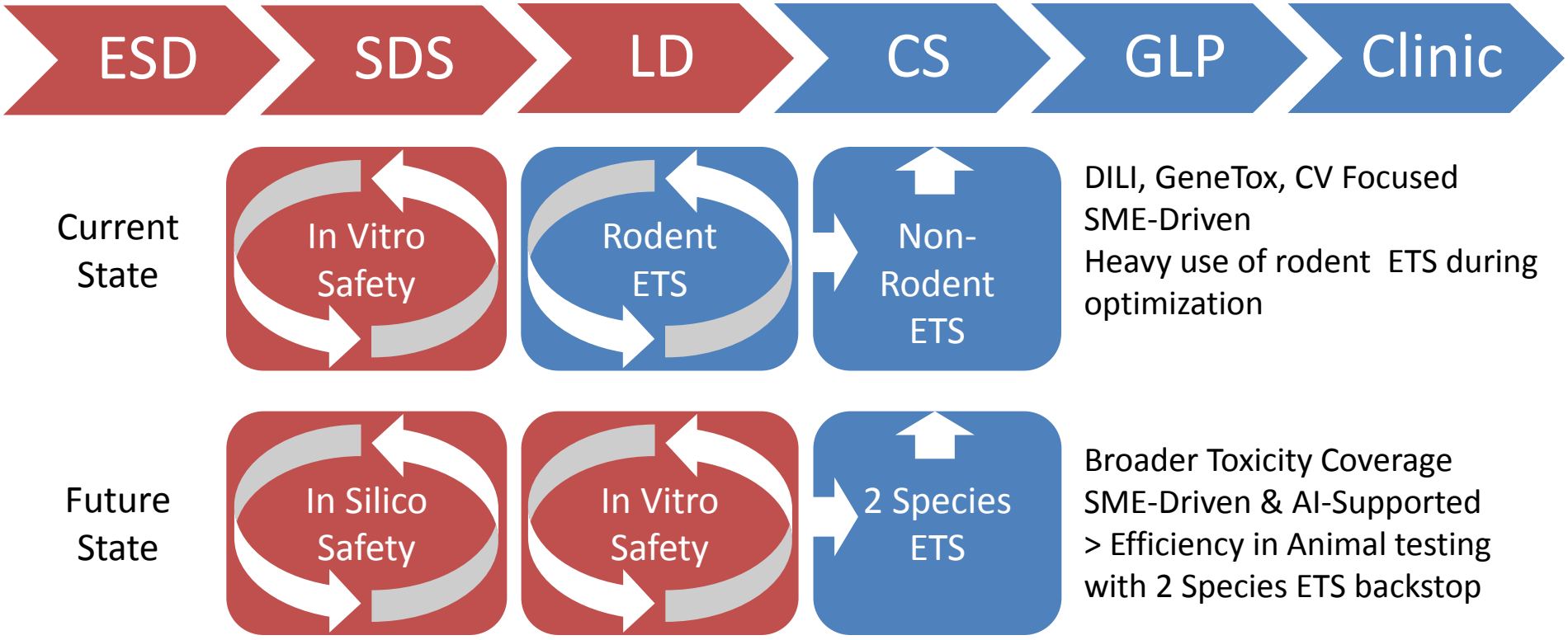
Target Tissue Specificity in Preclinical Species



<u>Lessons Learned</u>

- Need for automated NGS pipelines consistent across species
- Use data to ensure tissue/cell/species correctness
- Benchmark “quality” based on information gain as opposed to pure statistical indicators
- Need for uniform/automated analyses to interpret (not over-interpret) data

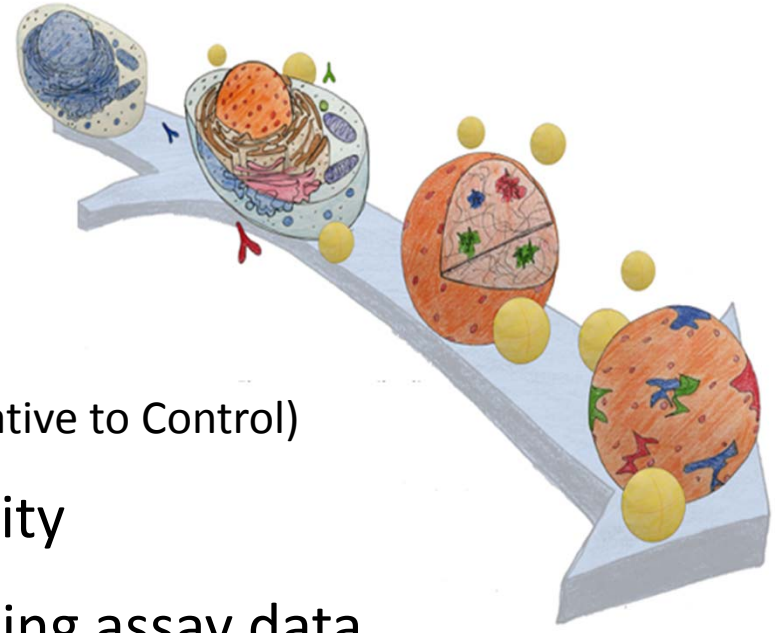
Computational Toxicology @ Pfizer



Early Toxicity Study (ETS)

HTS DNA Damage Response (DDR) Assay

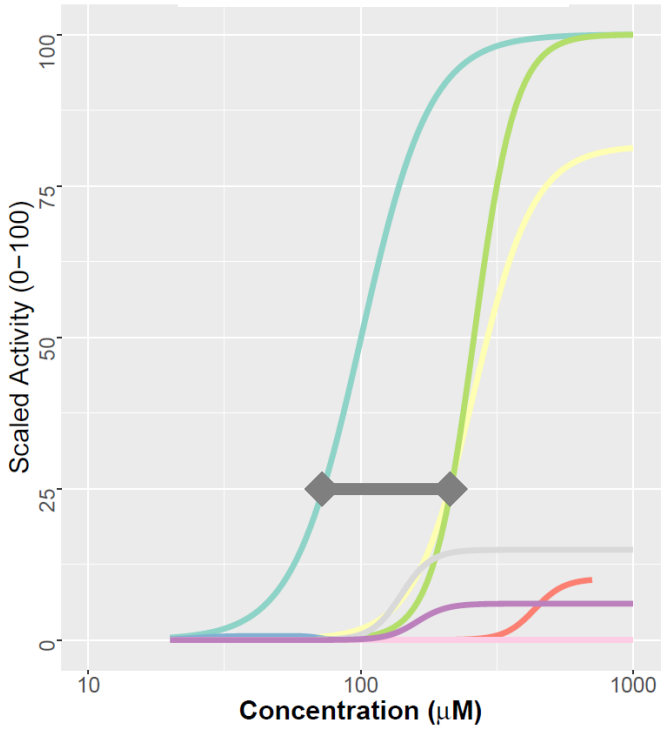
- Human lymphoblastoid TK6 cell line using flow cytometry
- **Earlier, faster, less compound with improved mechanistic and potency readouts compared to manual In Vitro Micronucleus Assay**
- 6 Markers of DNA Damage @ 4 & 24 hrs:
  - Double strand breaks (γ**H2AX**)
  - Mitotic Arrest (Phosphorylated Histone **H3**)
  - **Polyploidy** (DNA stain)
  - Apoptosis (Cleaved **PARP**)
  - DNA damage signaling (nuclear **P53**)
  - Culture Growth/**Cytotoxicity** (Cell Counts Relative to Control)
- Potential application to BM & GI toxicity
- Building predictive GeneTox model using assay data



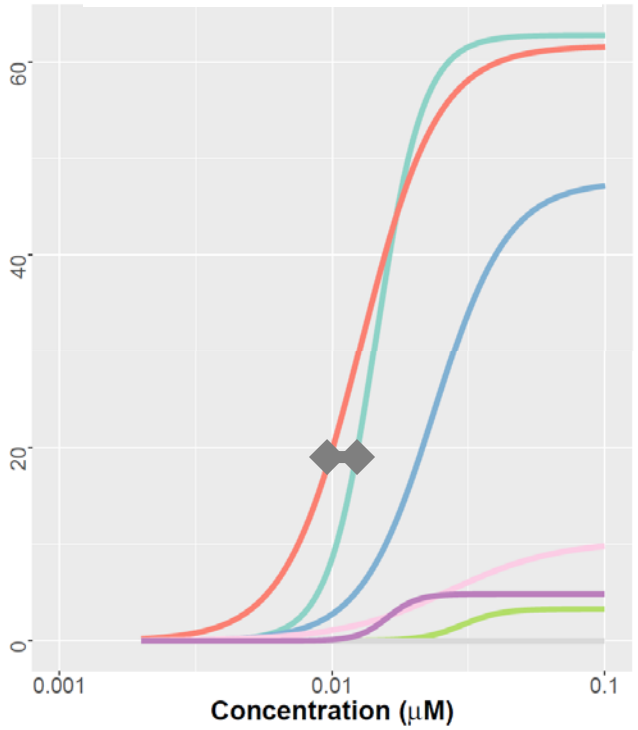
DDR Data Analysis & Model Inputs

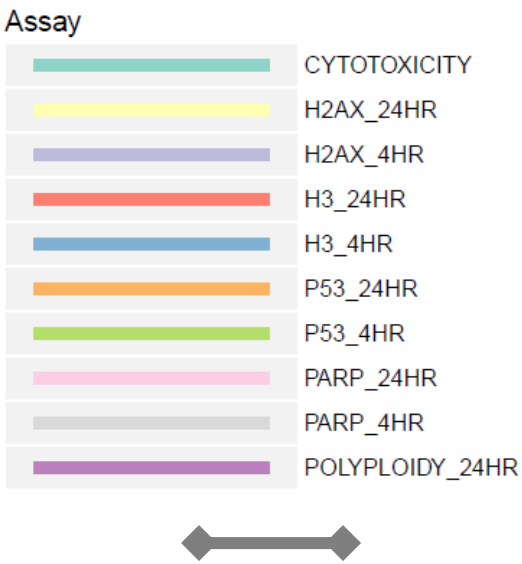
Many 1000s of complex DNA damage response data points reduced to maximal efficacy and relative (to cytotoxicity) potency estimates for model development

Non-genotoxicant



Genotoxicant (Aneugen)

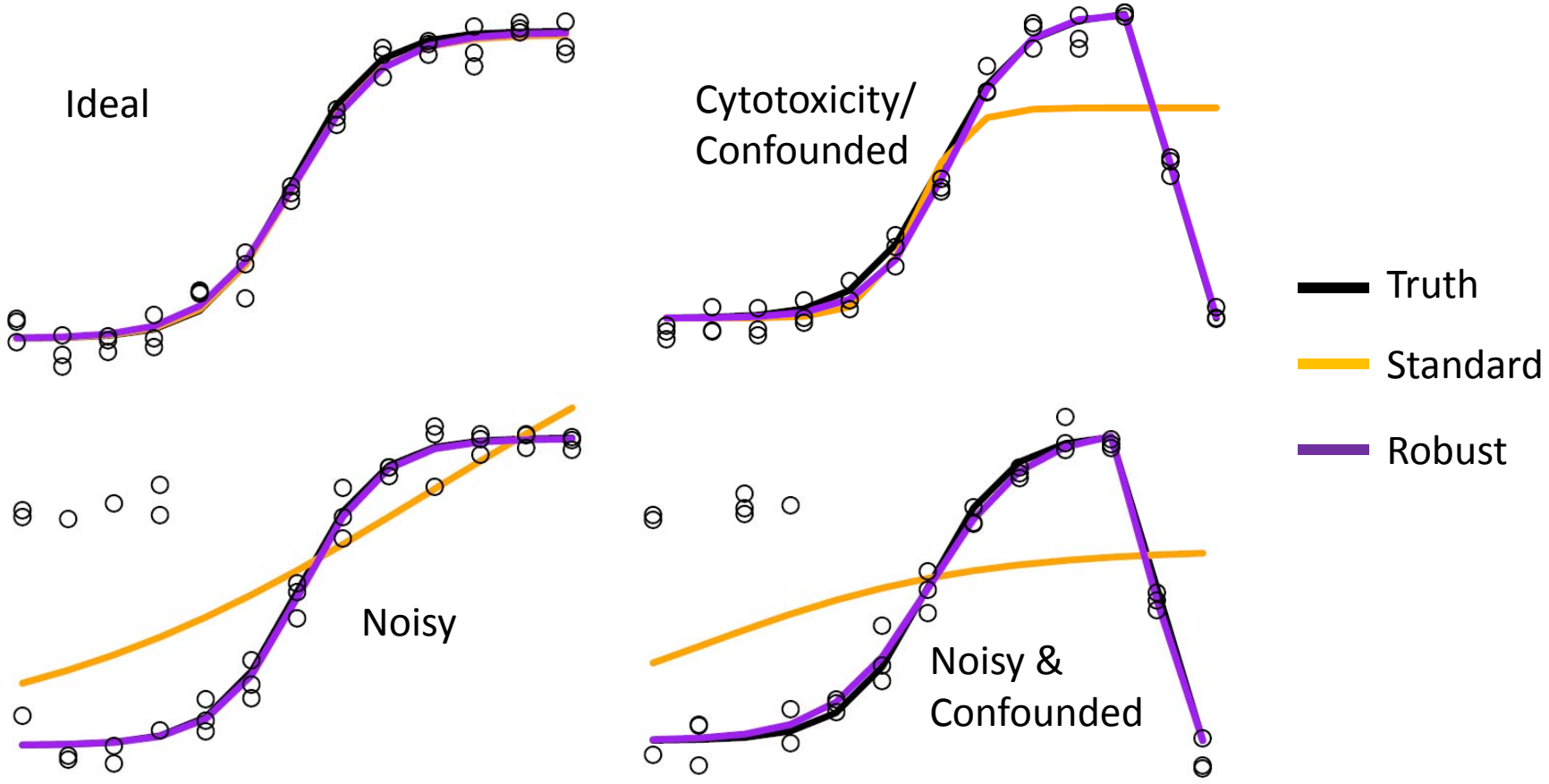




Cytotoxicity Potency Difference

Robust Concentration Response Analysis

Robust, Systematic, Flexible → 'Modelable', Comparable, Reusable

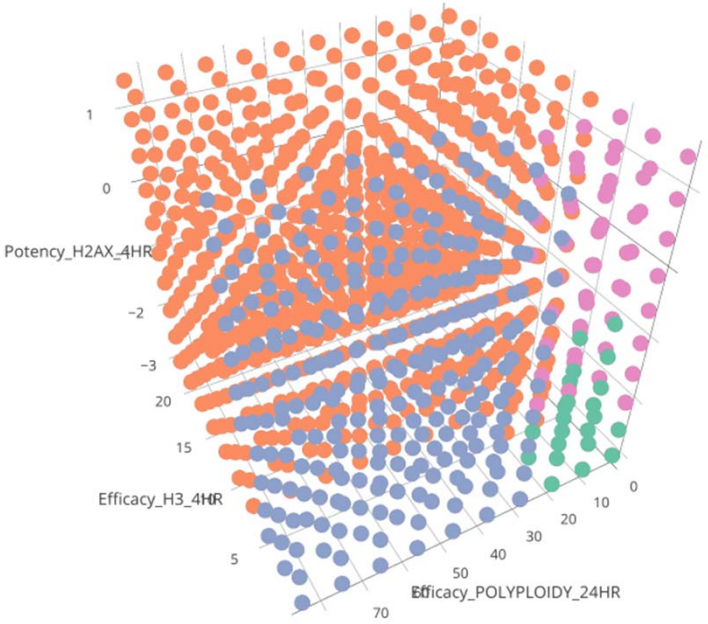


Ability to "re-fit" concentration-response data (internal and external)

https://github.com/USEPA/CompTox-ToxCast-tcpl

GeneTox Machine Learning Classification Model

- Random Forest model built on 58 well characterized chemicals
- Predicts **non-genotoxicity** vs genotoxicity
  - And class of genotoxicity (**aneugen – tubulin inhibitor**; **aneugen – aurora inhibitor**; **clastogen**)
  - >95% Accuracy
- Currently evaluating performance on internal Pfizer compounds



Lessons Learned

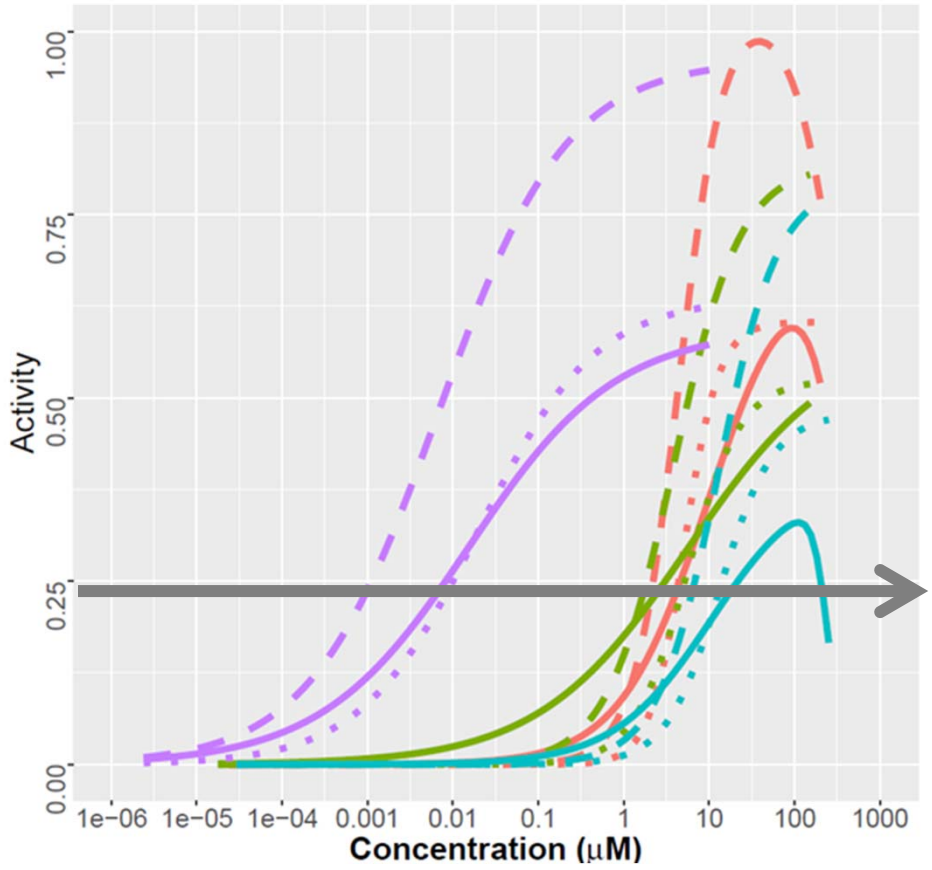
- Predictive models with good mechanistic underpinning do not need many chemicals for good predictivity
- Wet- and dry-Lab scientists need to work together closely
- Use of mechanistic assays can't explain mechanism in name only; data should do the talking
- Value is keeping all potency estimates, but making them relative to cytotoxicity

Complex/multi-parametric data generated, rigorously analyzed and distilled down to a single call in < 1 week

Computational Toxicology @ Pfizer

ESD → SDS → LD → CS → GLP → Clinic

Double Strand Breaks



Clinical Dose Selection & Mechanistic Differentiation

Lessons Learned

- Reanalyzing in vitro data in uniform manner increases comparability
- Using activity concentration at cutoff (ACC), for example, over AC50 or IC50 values is a less biased comparator

SCIENCE FOR LIFE-CHANGING IMPACT

From Big Data to Improved Decision Support

- **BIG DATA:** Generating large-scale tissue RNAseq tissue maps, HTS & whole transcriptomic in vitro profiling
- **SMART DATA**: Implemented high-throughput and automated system for robust concentration response analysis and safety classification & RNAseq pipeline with tissue specificity scoring
- **FAST DATA**: Simplified multi-parametric output to single call for rapid decision-making with ability to mechanistically interrogate & capacity to translate to other toxicities (e.g., DILI, CV, Kidney, Bone Marrow, GI)
- **QUALITY**: Automation and systemization of big/complex data inherently increases quality and utility of the data

Acknowledgements

DSRD

iTox: Anne Ryan, Yvonne Will, Mike Aleo, Marc Roy, Wendy Hu, Drew Burdick
Dennis Pelletier, Mark Gosink, Petra Koza-Taylor

GeneTox COE: Maik Schuler, Zhanna Sobol, Randy Spellman, Maria Engel, Krista Dobo

Safety Pharm COE: Steve Jenkinson

Jon Cook, Meg Driscoll, Mathy Nagappan

Medicine Design

Eric Watt, Chris Keefer, Cindy Li, Tristan Maurer