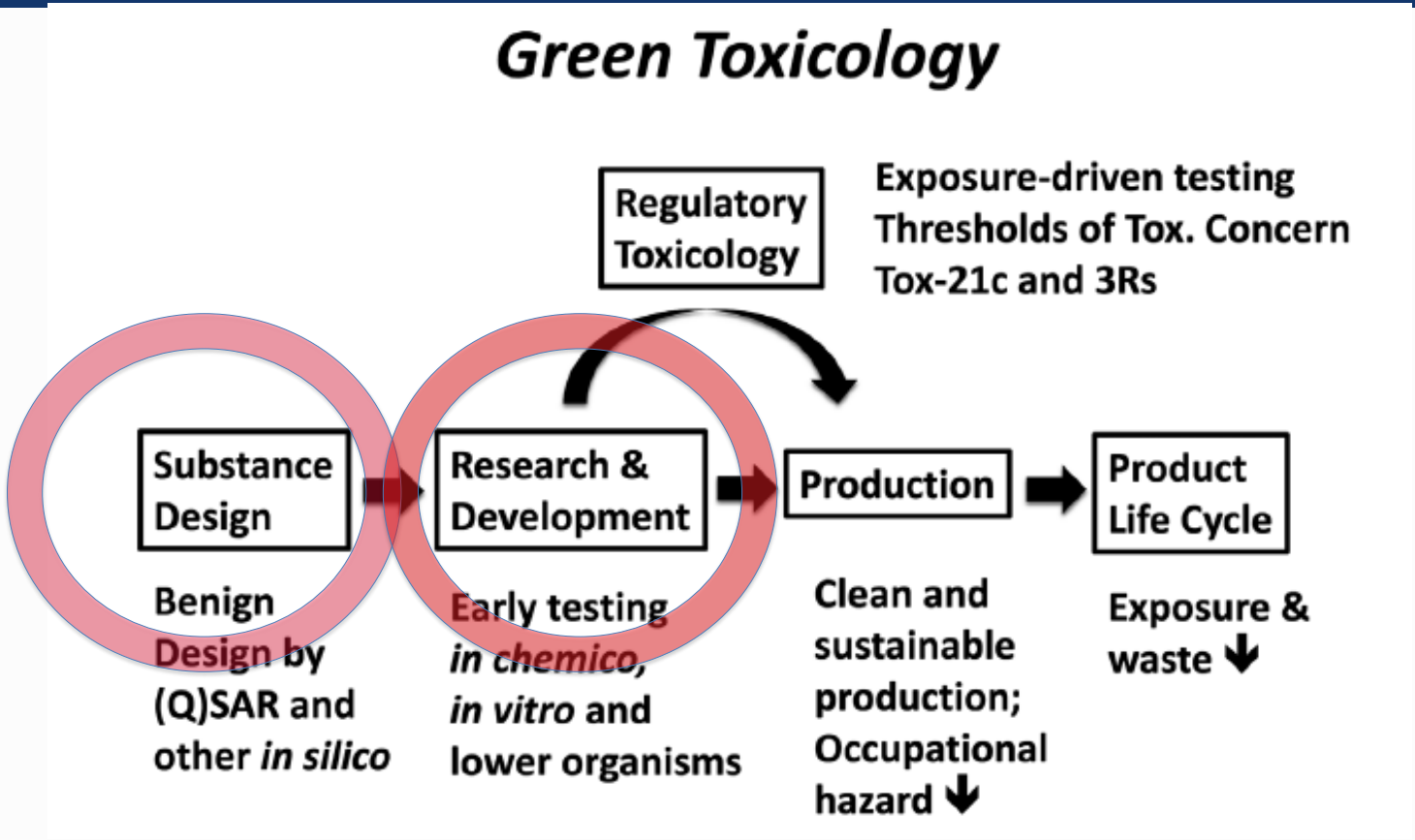
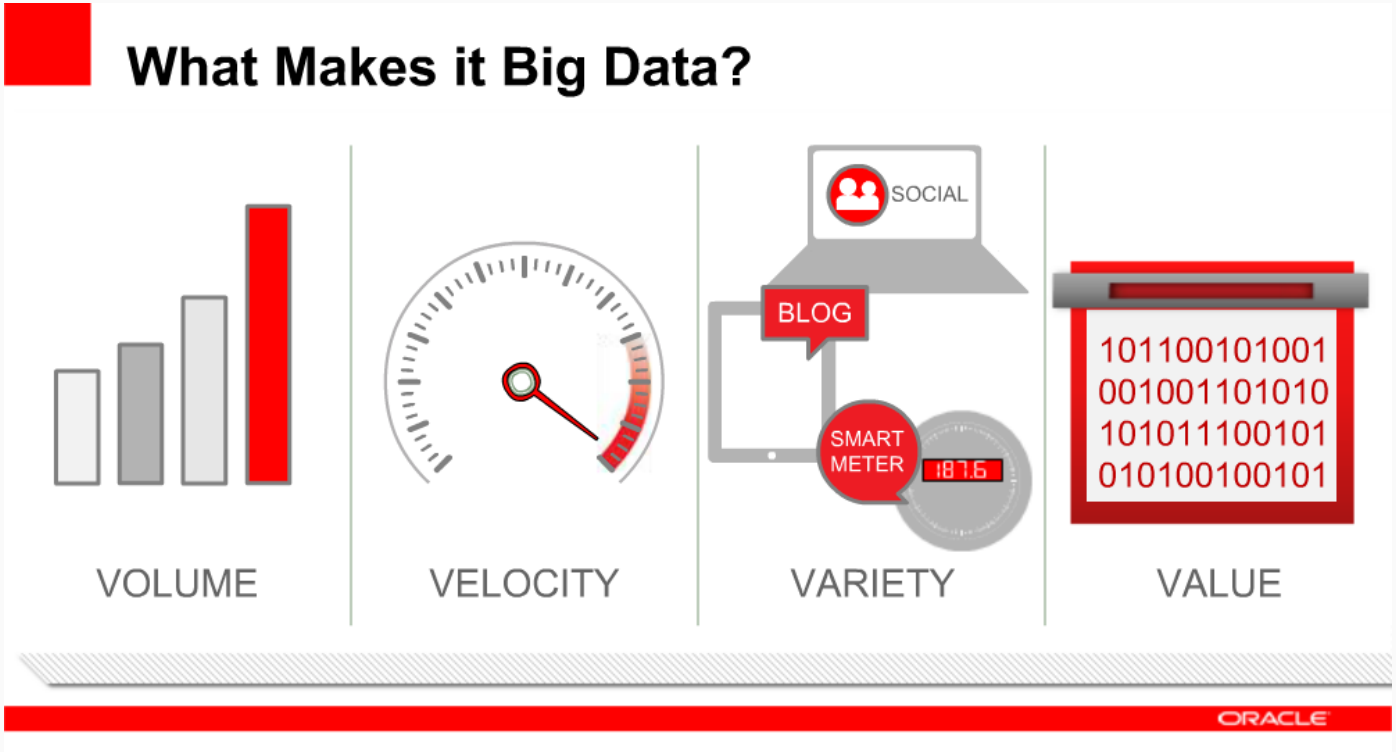

# Green Toxicology: Using Big Data to Guide Design Safer Chemicals

Alexandra Maertens, Ph.D.

# Green Toxicology

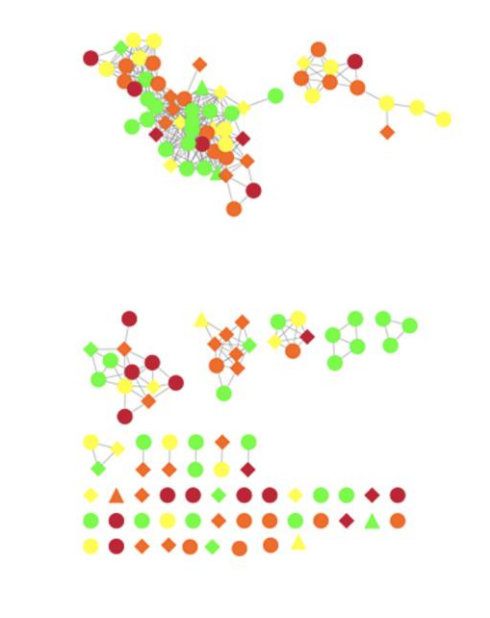


# Big Data


What Makes it Big Data?
SOCIAL
BLOG
SMART METER
187.6
101100101001
001001101010
101011100101
010100100101
VOLUME
VELOCITY
VARIETY
VALUE
ORACLE

# Big Data. . . for a Toxicologist

2015



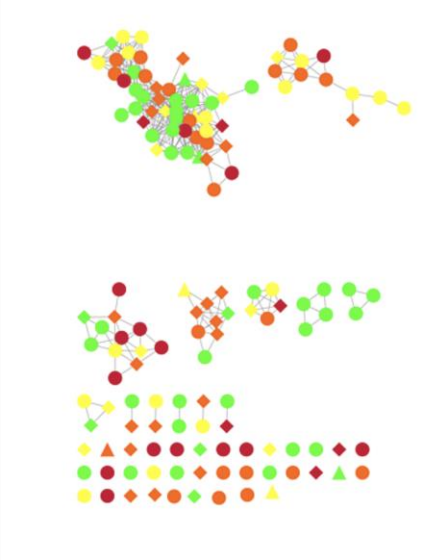
- More information is not always better; model improved when we used feature elimination
- Model improved when dose-response information included
- Variable ranking indicated in vitro assays provided information

**Probabilistic hazard assessment for skin sensitization potency by dose–response modeling using feature elimination instead of quantitative structure–activity relationships**

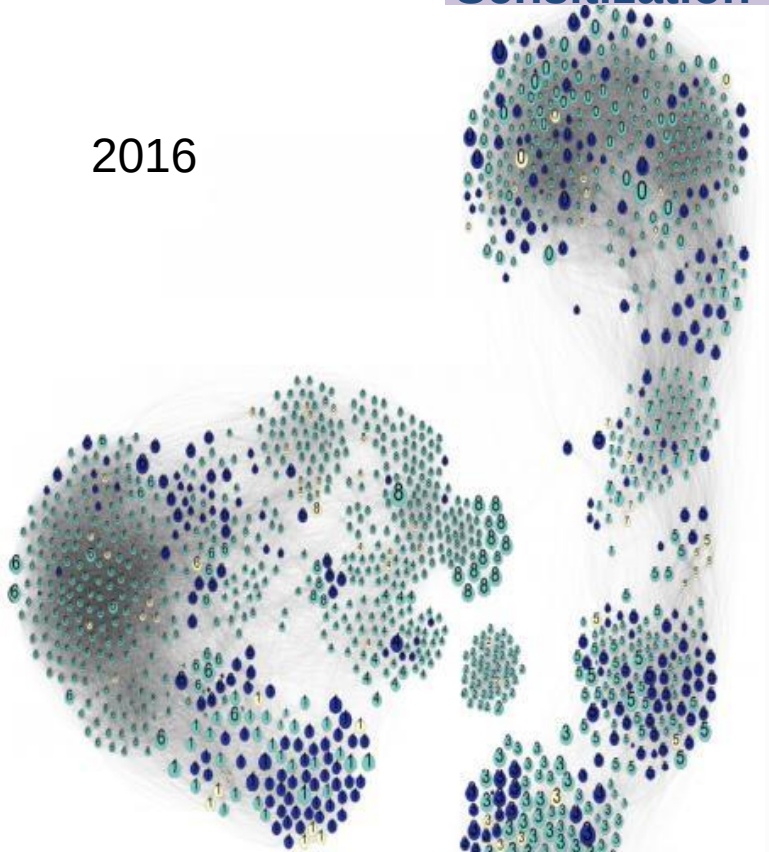
Thomas Luechtefeld[a†], Alexandra Maertens[a†], James M. McKim[b], Thomas Hartung[a,c*], Andre Kleensang[a] and Vanessa Sá-Rocha[a,d]

# Big Data. . . For a Toxicologist

2015



2016



# Big Data . . . For a Toxicologist

***Skin sensitization:***
***Simple classification by nearest neighbor***

| Min. Similarity | Chemicals | Accuracy |
|---|---|---|
| 0.95 | 525 | 0.92 |
| 0.9 | 1189 | 0.85 |
| 0.85 | 1738 | 0.84 |
| 0.75 | 2288 | 0.80 |

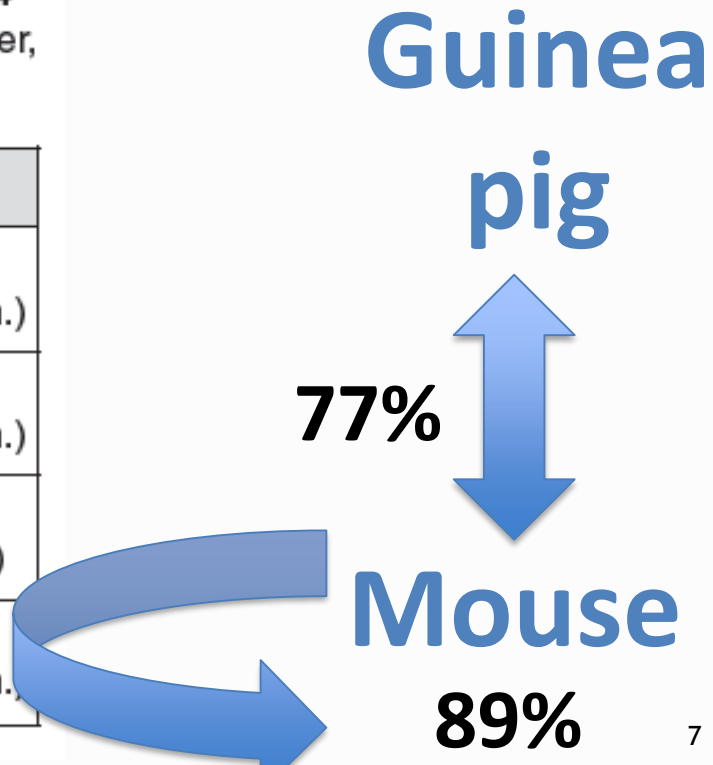
**Accuracies better than different animal TG against each other**

From: Luechtefeld et al. Predicting Skin Sensitization with REACH Dataset

# Big Data: Better Data

**Tab. 1: Classification agreement on chemicals with at least two sensitization studies in REACH dossiers from 2008-2014**
Studies found by searching for all studies with *studytype* = Buehler, GPMT, Patch-Test or LLNA.

| | Buehler | GPMT | Patch-test | LLNA |
|---|---|---|---|---|
| **Buehler** | 95.1% (344 chem.) | 91.8% (364 chem.) | 87.8% (58 chem.) | 76.8% (212 chem.) |
| **GPMT** | | 93% (624 chem.) | 90.5% (107 chem.) | 77.4% (403 chem.) |
| **Patch-test** | | | 92.1% (24 chem.) | 78.3% (40 chem.) |
| **LLNA** | | | | 88.5% (296 chem.) |



# Big Data: Public Data is Useful Data

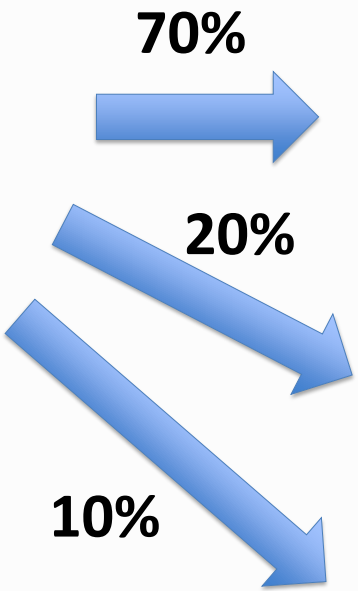
## *Repetitions of Draize rabbit eye test (TG 405)*

| Repeats | Number substances | Example ECNumber |
|---|---|---|
| 90 | 2 | 613-683-0,295-445-2 |
| 45 | 69 | 940-595-2,295-431-6 |
| 18 | 1 | 931-203-0 |
| 15 | 2 | 700-762-0,692-840-5 |
| 13 | 38 | 934-268-3,931-515-7 |
| 12 | 2 | 918-317-6,500-513-4 |
| 11 | 2 | 931-700-2,226-109-5 |
| 10 | 2 | 232-395-2,939-581-9 |
| 9 | 1 | 267-291-6 |
| 8 | 27 | 940-730-5,940-728-4 |
| 7 | 32 | 940-727-9,940-726-3 |
| 6 | 75 | 931-745-8,300-226-2 |
| 5 | 56 | 939-578-2,939-575-6 |
| 4 | 135 | 939-693-8,939-621-5 |
| 3 | 254 | 939-715-6,939-688-0 |
| 2 | 593 | 208-778-5,941-224-7 |
| 1 | 2388 | 293-029-5,273-224-1 |

# Big Data: Public Data is Useful Data

**First test:**

**Eye corrosive**




**Second test:**

**corrosive**

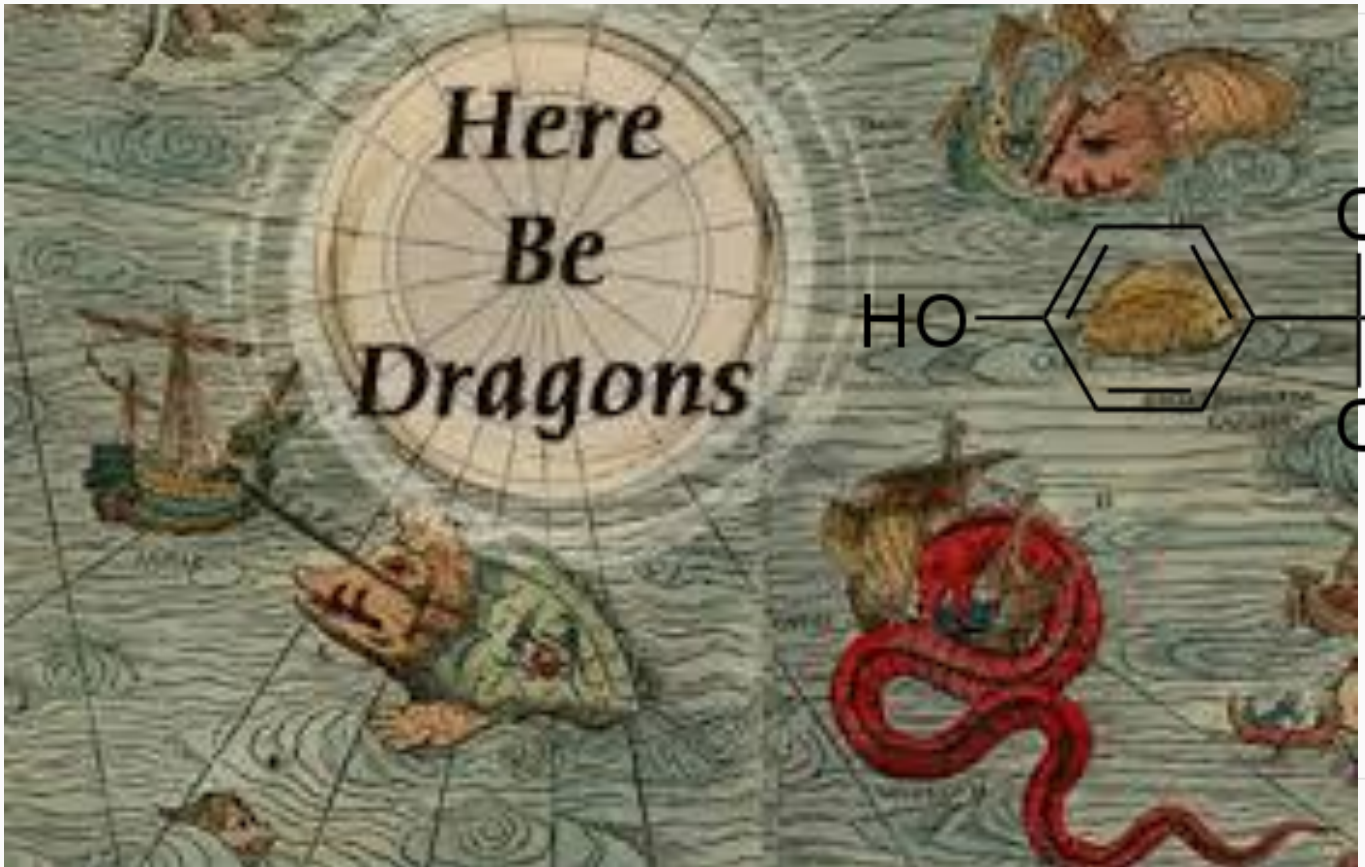
**mild**

**no effect**

**Analysis of 670 chemicals**

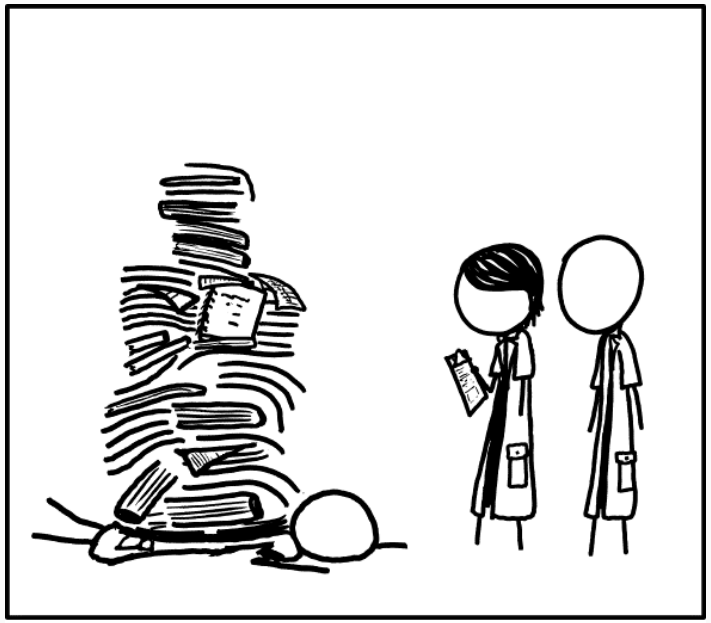
**ALTEX 2016, 33, 123-134**

# Mapping the Human Toxome



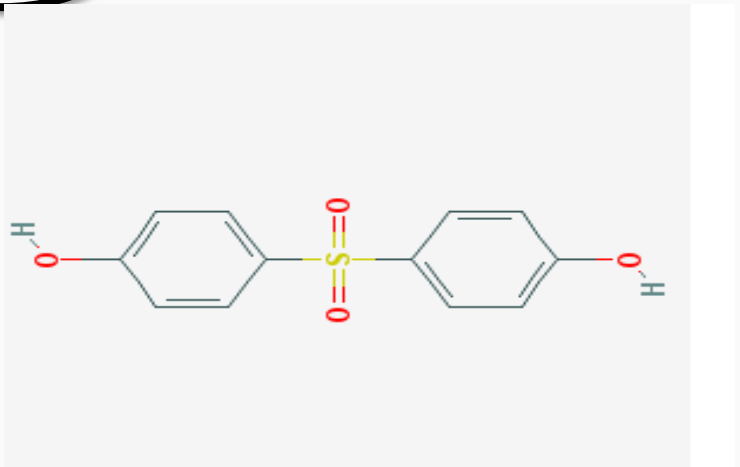
# Big Data. . . Of the Wrong Kind

- BPA has been subjected to multiple guideline studies in different species/strains
- Pubmed returned over 10,800 abstracts for BPA
- Comparative Toxigenomics Database shows BPA effects almost 2,000 genes when restricted to humans
- HSDB has over 79 studies laboratory animal studies




Source: XKCD

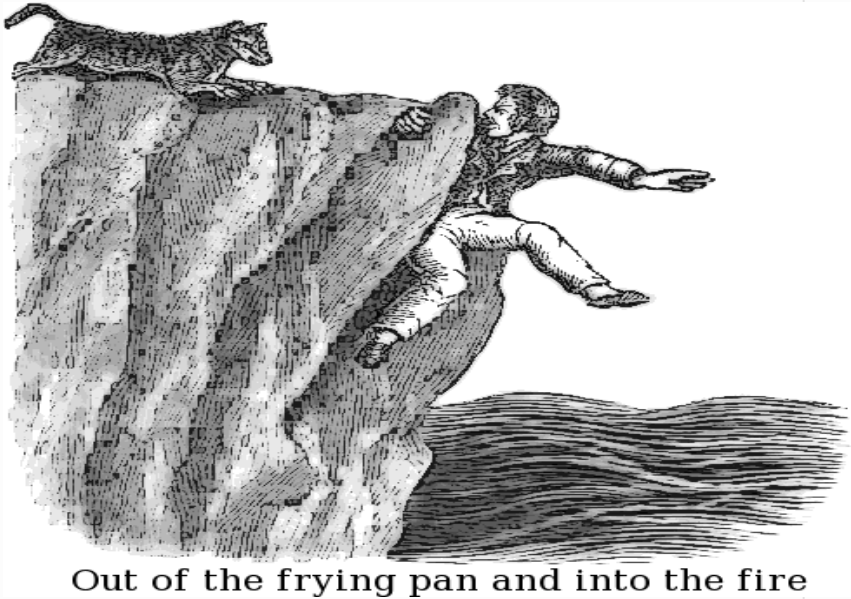
Aha! BPA binds to the Estrogen Receptor!
If I add a sulphonyl group, it will no longer bind!



Bisphenol Sulfate

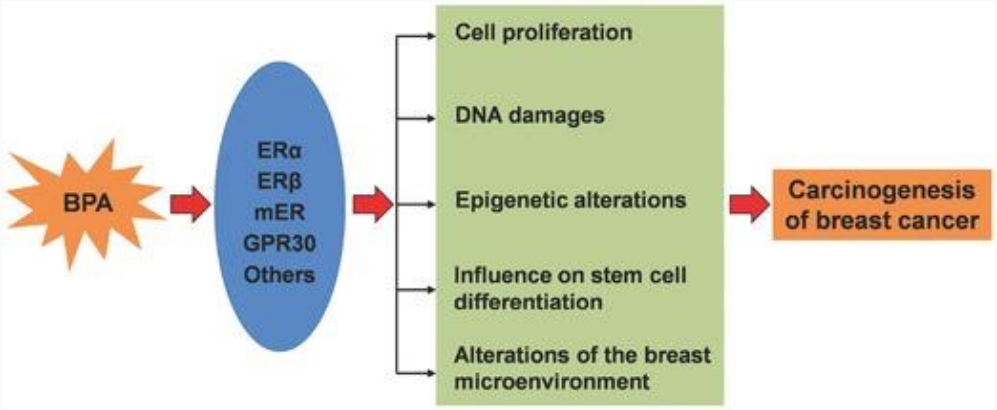
# Unfortunately. . .

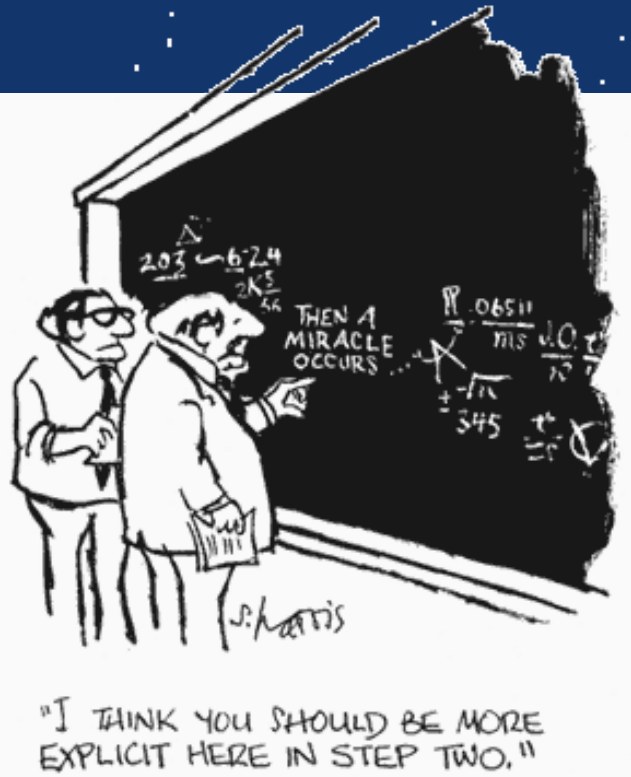
- turns out that BPA binds canonical Estrogen receptors but bind Estrogen-Related Receptor Gamma (ERR-Gamma) with higher affinity
- BPS binds ER-γ with higher affinity than BPA
- Unlikely to be detected by animal testing
- 1-10 Million tons of BPS are produced per year




It

# BPA - Pathways and Adverse Outcomes

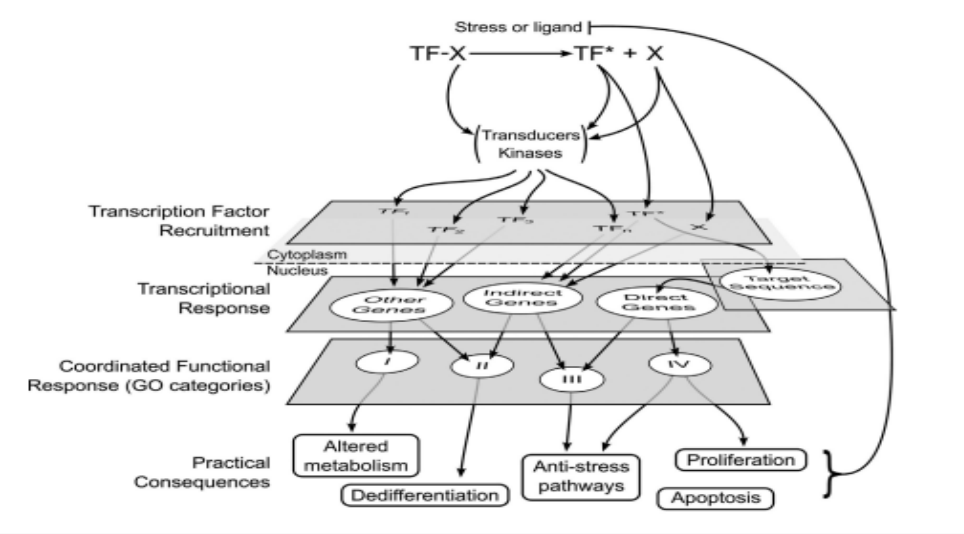



From:Low-Dose Bisphenol A Exposure: A Seemingly Instigating Carcinogenic Effect on Breast Cancer
Zhe Wang, 1 , 2 ,* Huiyu Liu, 3 ,* and Sijin Liu 1 ,*

# Mapping the Human Toxome

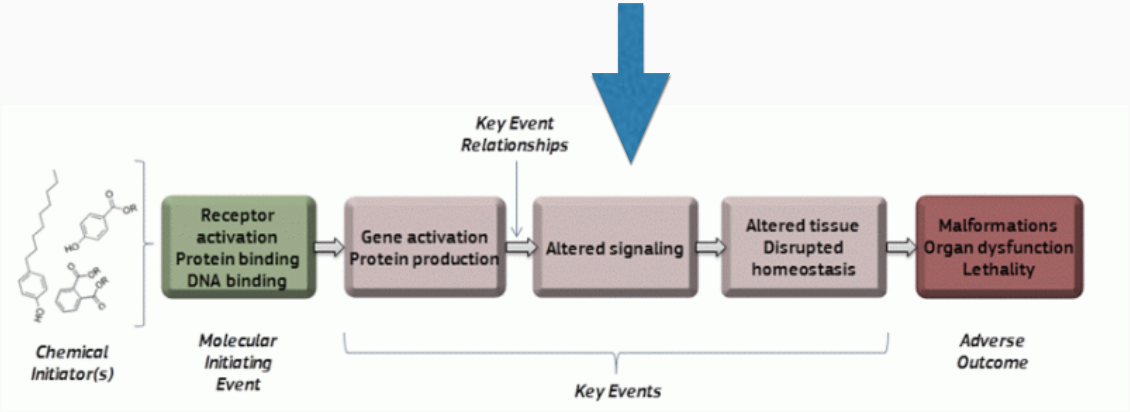
- Pathway of Toxicity is a molecular definition of the cellular processes shown to mediate adverse outcomes of toxicants.
- "Black box" animal models or simple molecular models towards understanding toxicity in it's full complexity
- A PoT helps to bring toxicology into the "systems biology"




From: Andersen, MA 2013

# Pathway of Toxicity

A **Pathway of Toxicity (PoT)** focuses on the molecular events within a cell - the networks of genes, proteins, or metabolites that are altered as a result of the molecular initiating event.



# Can we use *in vitro* data for a better understanding of molecular mechanisms?

- We used WGCNA and IDEA to extract a genetic regulatory network and possible transcription factors from cells treated with a variety of estrogen and BPA
- Goal: to understand which receptors and transcription factors are driving the biology

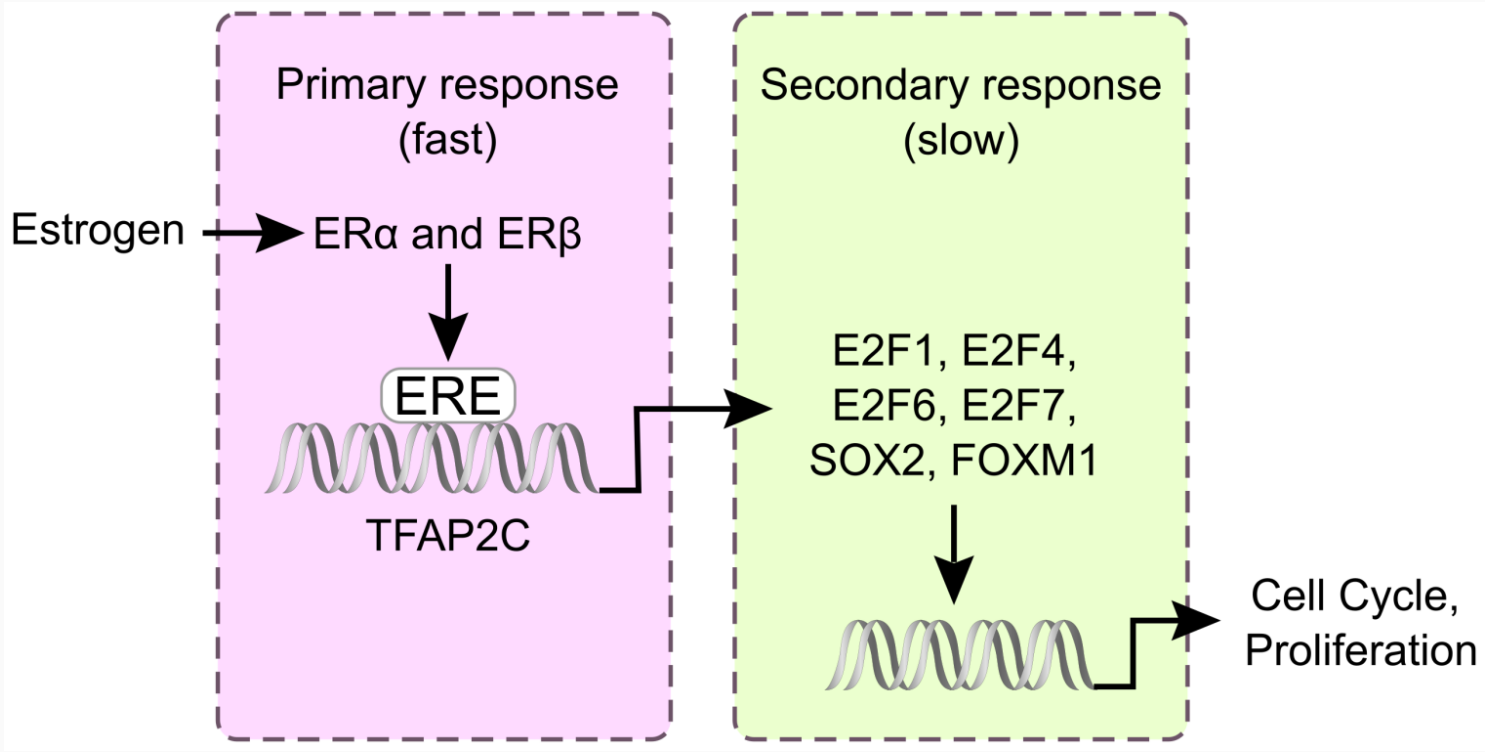
Correlation Based Approach using a “Guilt by Association Approach” to generate interaction partners from a microarray dataset

# Can we use *in vitro* data for a better understanding of molecular mechanisms?

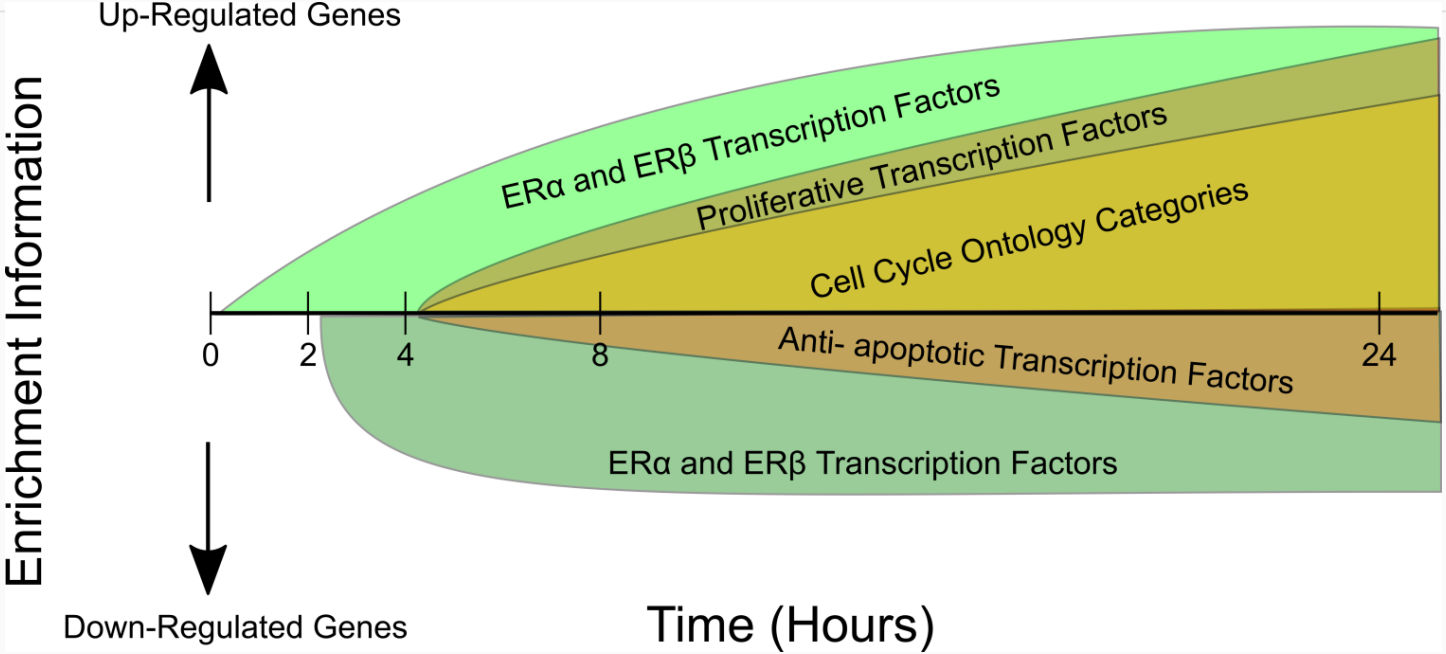
- We used WGCNA and IDEA to extract a genetic regulatory network and possible transcriptic factors from cells treated with a variety of estrogen and BPA
- Goal: to understand which receptors and transcription factors are driving the biology

Information Dependent Enrichment Analysis Uses Rank Order of Gene Expression and a Permutation Test to Establish Candidate Transcription Factors

# ESTROGEN

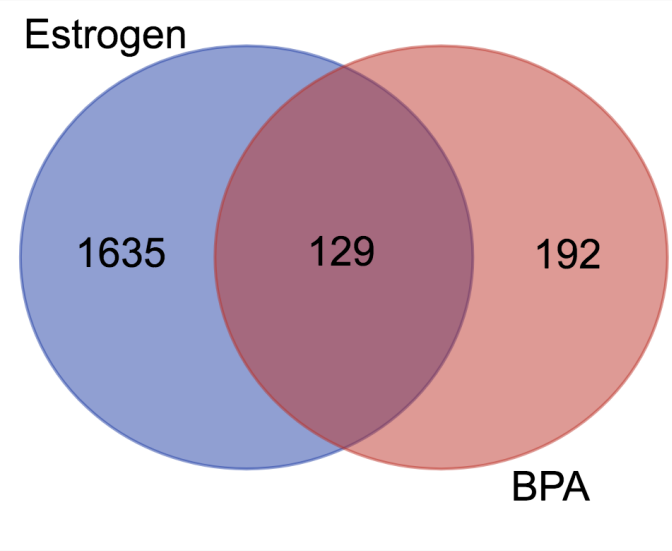

Primary response
(fast)
Secondary response
(slow)
Estrogen
ERα and ERβ
ERE
TFAP2C
E2F1, E2F4,
E2F6, E2F7,
SOX2, FOXM1
Cell Cycle,
Proliferation

# ESTROGEN

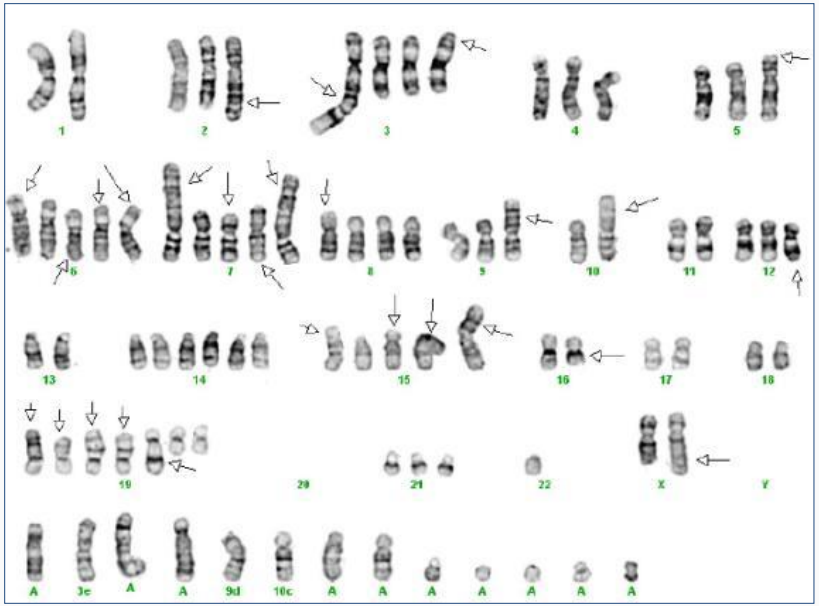

Up-Regulated Genes
Enrichment Information
ERα and ERβ Transcription Factors
Proliferative Transcription Factors
Cell Cycle Ontology Categories
Anti- apoptotic Transcription Factors
ERα and ERβ Transcription Factors
0
2
4
8
24
Down-Regulated Genes
Time (Hours)

# WGCNA - BPA

- Weak over-representation of canonical estrogen-receptor binding sites; several novel transcription factor unique to BPA
- Just under 10 percent of the genes in the data set were “open reading frames” - meaning they had no official gene name, and no annotations
- Several identified transcription factors (such as FIZ1) are poorly characterized - less than 8 PMIDs



# Use *In Vitro* Data Cautiously




**Karyotyping**

# MCF-7 Does It Really Capture Human Biology?

## *Extent of deviations from normal genome*

| Classification | Kilobases | Percentage of genome |
|---|---|---|
| Losses | 4587603 | 51.2% |
| Deletions | 667374 | 7.5% |
| Amplifications | 26904 | 0.3% |
| Gains | 2587093 | 28.9% |
| Normal | 871166 | 9.7% |
| Centromeres | 217339 | 2.4% |
| Total Abberations | 7868974 | 87.8% |
| All Entries | 8957479 | |

**SurePrint G3 ISCA CGH+SNP Microarray Kit, 4x180K**

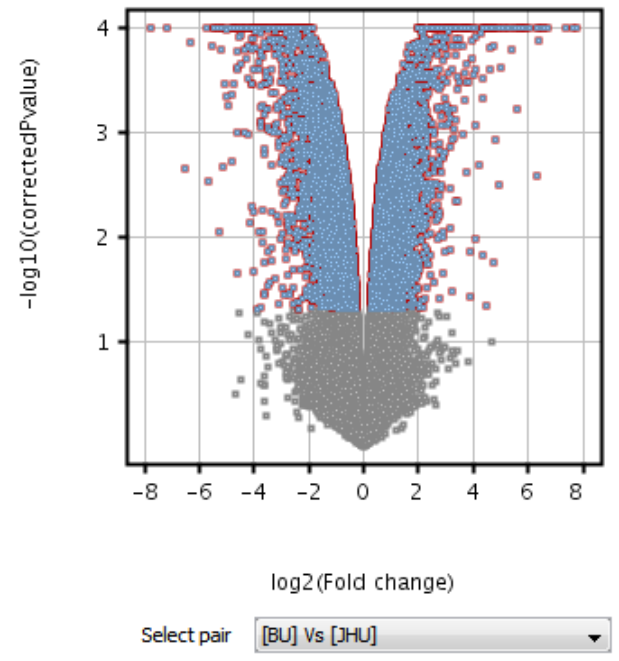
**115234 CGH features.2440 CGH replicate probes, 59647 SNP features**

**reference mapping: caucasian female human reference DNA**

**Kleensang et al., Nature Sci Rep, 2016**

# Use *In Vitro* Data Cautiously

## Comparison of MCF-7 in two laboratories




**Same batch from ATCC**

**Method transfer**

**Transcriptomics**

**negative controls, 4h, gene level, *n* = 3 / group**

# Conclusions

“The Universe is under no obligation to make sense to you.”

-Neil DeGrasse Tyson

# Conclusions

- The mechanism of toxicity is under no obligation to be simple for you -
  - Not necessarily mediated by a single receptor
  - Receptors may have different effects in different cells/organs
  - Dose-response may vary enormously depending on context - oxidative and metabolic state, epigenetic background, etc. . . . and your animal or *in vitro* model may be misleading
  - We need to stop “looking under the lampshade for the key, since that is where the light is” - we have to acknowledge how much of cell biology is unknown

# Mapping Mechanisms of Toxicology is Just a Big Data Problem

- For some endpoints, big data is already paying off - we have good models to guide molecular design and test quickly and without animals - skin sensitization, mutagenicity, etc.
- For more complicated endpoints, the map is still incomplete, but. . .
- The number of ways a chemical can interact with a cell *is finite* - it is “mappable”
- *In vitro* technologies are key, but we need to vastly expand our data sets and improve the way we approach our data sets, and respect the limits of our understanding