Uncovering chemical toxicity by high-throughput transcript profiling: linking molecular targets to potential adverse effects

Chris Corton

Center for Computational Toxicology and Exposure
US-Environmental Protection Agency
Research Triangle Park, NC

Physicians Committee for Responsible Medicine Summer School on Innovative Approaches in Science, RTP, NC June, 2022

Disclaimer

- The views expressed are those of Dr. Chris Corton and do not reflect US-EPA policy or product endorsement by the US-EPA.

Outline of Objectives

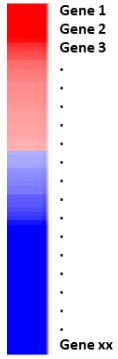
- How high-throughput transcript profiling is carried out as a method to comprehensively assess the effects of chemicals on biological systems
  - Different platforms for assessing genome scale gene expression changes
  - High-throughput transcript profiling (HTTr)

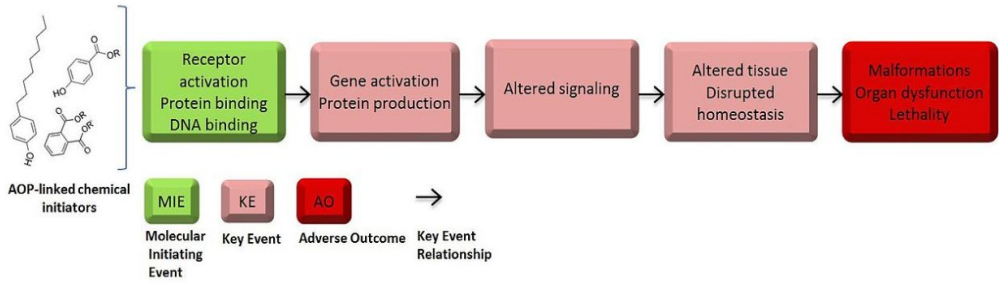
- How to identify the molecular targets of chemicals
  - Hypothesis generating tools
  - Gene expression biomarkers
  - How to
    - Identify predictive gene sets
    - Characterize the gene sets
    - Determine predictive accuracy
    - Use in screening chemicals

- How to link the alterations in molecular targets to potential adverse events.
  - Use of the adverse outcome framework

Definitions

Treated vs. Control





Gene Expression Comparison (Differentially Expressed Genes)

- List of statistically-filtered genes derived from a comparison between treated and control groups

Gene Expression Biomarker

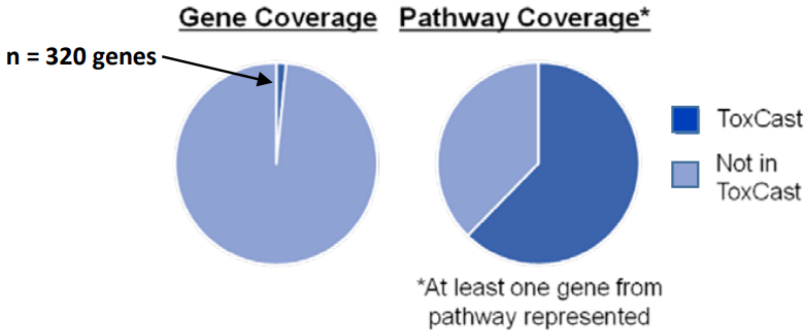
- List of genes and associated fold-change values or ranks
- Measures a molecular initiating event or key event in an adverse outcome pathway using transcript profiling

Adverse Outcome Pathway

- Structured representation of biological events leading to adverse effects; relevant to risk assessment
- A series of causally connected key events (KE) between two points — a molecular initiating event (MIE) and an adverse outcome (AO) that occur at a level of biological organization relevant to risk assessment

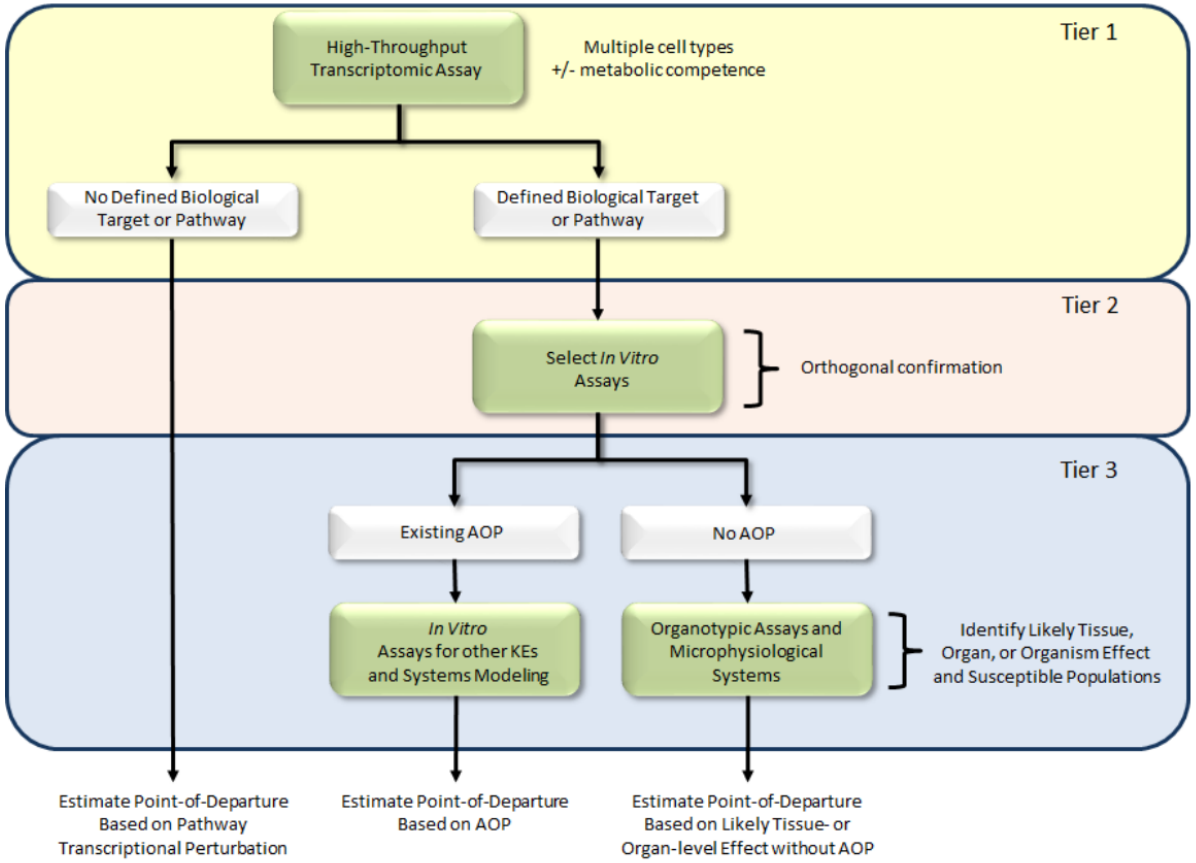
High throughput toxicity testing

- ToxCast assays cover many genes and pathways, but do not provide complete coverage of biological space.



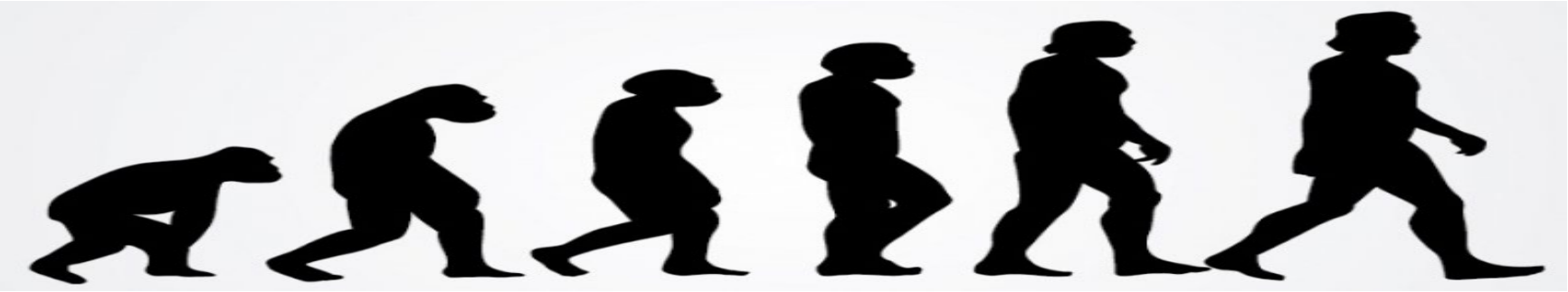
- **USEPA Strategic Vision and Operational Roadmap:**
  - Tier 1 strategy must cast the broadest net possible for capturing hazards associated with chemical exposure.
  - Global gene expression provides a robust and comprehensive evaluation of chemically induced changes in biological processes.
  - Increasing efficiency and declining cost of generating whole transcriptome profiles has made high-throughput transcriptomics (HTTr) a practical option for determining bioactivity thresholds in *in vitro* models.

A strategic vision and operational road map for computational toxicology at the U.S. Environmental Protection Agency



Joshua Harrill, EPA

Evolution of gene expression profiling

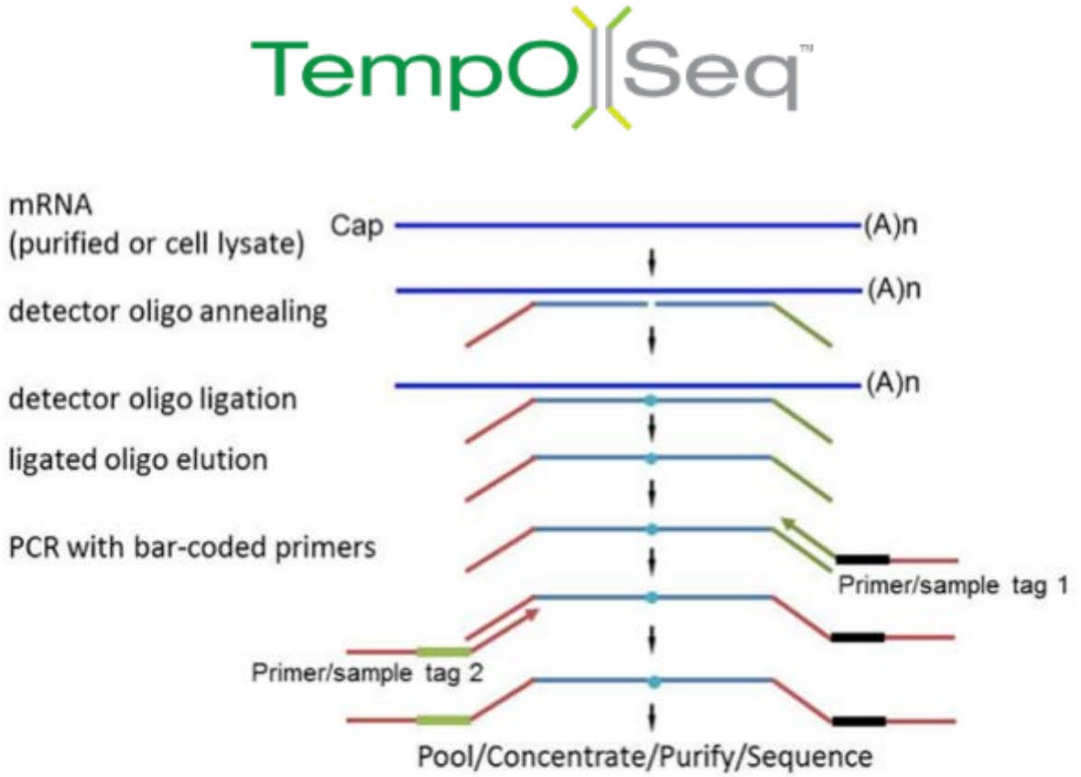


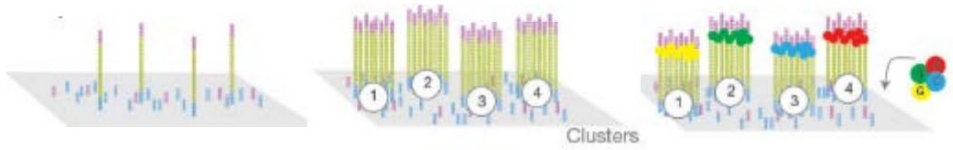
Key Driver – lower costs of profiling

- 1992: Differential display; Liang and Pardee Science. 257(5072):967-71
- 1995: Two-color microarrays; Schena et al. Science. 270(5235):467-70
- Late 1990s: Agilent and Affymetrix arrays – full genome analysis
- 2010s: RNA-Seq
  - Not amenable to high-throughput
- 2017: PLATE-Seq Bush et al. Nat Commun. 8(1):105
  - 96 samples processed simultaneously; uses cell lysates and oligo(dT)-coated plates; 2M reads/sample; ~$15/sample
- 2017: Tempo-Seq; BioSpyder; Yeakley et al. PLoS One. 12(5):e0178302
  - 384 samples processed simultaneously; uses cell lysates; full-genome and 1500+ platforms
- 2018: DRUG-Seq; Ye et al. Nat Commun. 9(1):4307
  - 384 samples processed simultaneously; uses cell lysates; 2M reads/sample; ~$2-4/sample

BioSpyder TempO-Seq

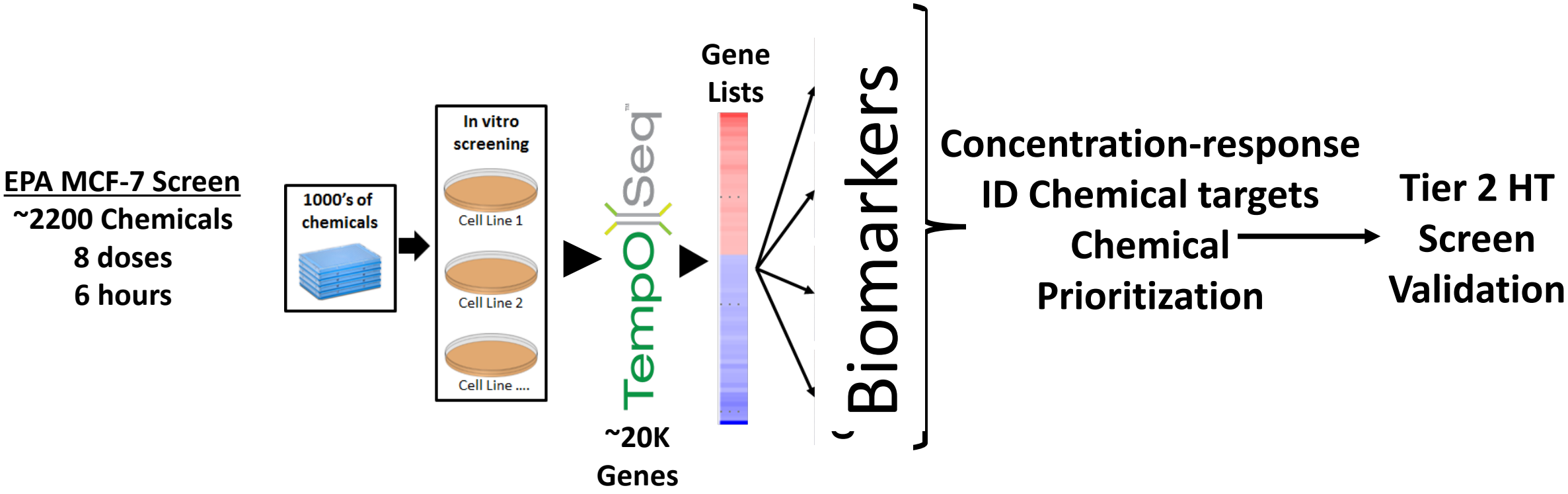
- Targeted RNA-Seq technology
- Whole transcriptome assay provides output on > 20,000 transcripts.
- Requires very low input (< 10 pg total RNA).
- Performed on “standard” PCR and Next Gen Sequencers.
- Compatible with purified RNA or cell lysates.





www.biospyder.com www.illumina.com

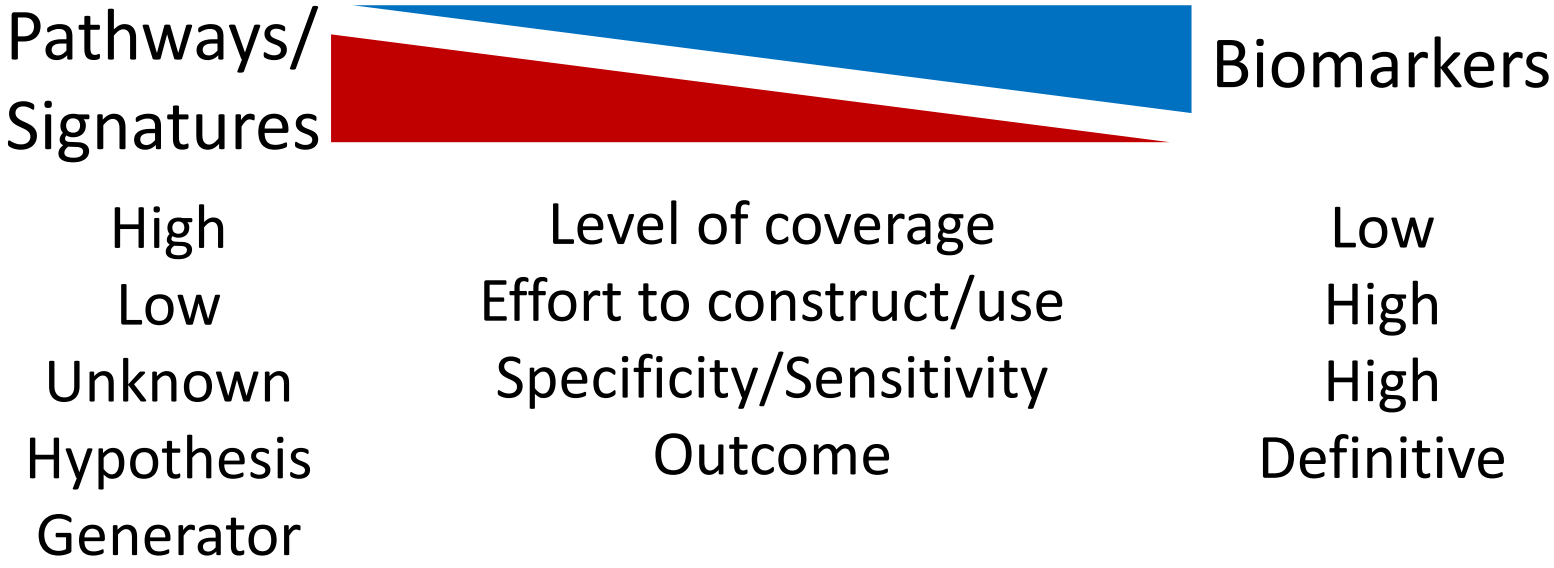
Using gene expression biomarkers to identify molecular targets of chemicals in transcriptomic studies



- Use predictions for
  - Chemical prioritization as part of Tier 1 screening
- Followed up with short-term tests in organotypic cultures or animals

Strategies for identifying molecular targets of chemicals in gene expression profiles: Pathways vs. Biomarkers

- Biomarker defined as “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention.” (1998, the National Institutes of Health Biomarkers Definitions Working Group)
  - Very few examples of well characterized gene expression biomarkers
  - No examples of gene expression biomarkers accepted by regulatory agencies for toxicity testing
- Pathways/signatures are often used to interpret gene expression
  - Gene Set Enrichment Analysis (GSEA)
  - Ingenuity Pathway analysis (IPA)
- Pathways/signatures and biomarkers are complimentary approaches



- A gene expression biomarker is a short list of genes and associated fold-change values that are used to predict the activity of a factor important in mediating effects of chemicals

Biomarkers that predict key events in human cells in vitro

United States Environmental Protection Agency

Endocrine disruption

- Ryan et al. (2016). Moving Toward Integrating Gene Expression Profiling Into High-Throughput Testing: A Gene Expression Biomarker Accurately Predicts Estrogen Receptor α Modulation in a Microarray Compendium. Toxicol Sci. 151(1):88-103.
- Androgen receptor: Rooney et al. (2018). Identification of Androgen Receptor Modulators in a Prostate Cancer Cell Line Microarray Compendium. Toxicol Sci. 166:146-162.

DNA Damage Response

- Corton et al. (2018). Using a gene expression biomarker to identify DNA damage-inducing agents in microarray profiles. Environ Mol Mutagen. 59:772-784.
- Cho et al. (2019). Assessment of the performance of the TGx-DDI biomarker to detect DNA damage-inducing agents using quantitative RT-PCR in TK6 cells. Environ Mol Mutagen. 60:122-133.
- Corton JC, Witt KL, Yauk CL. (2019). Identification of p53 Activators in a Human Microarray Compendium. Chem Res Toxicol. 32(9):1748-1759.

Epigenetic effects

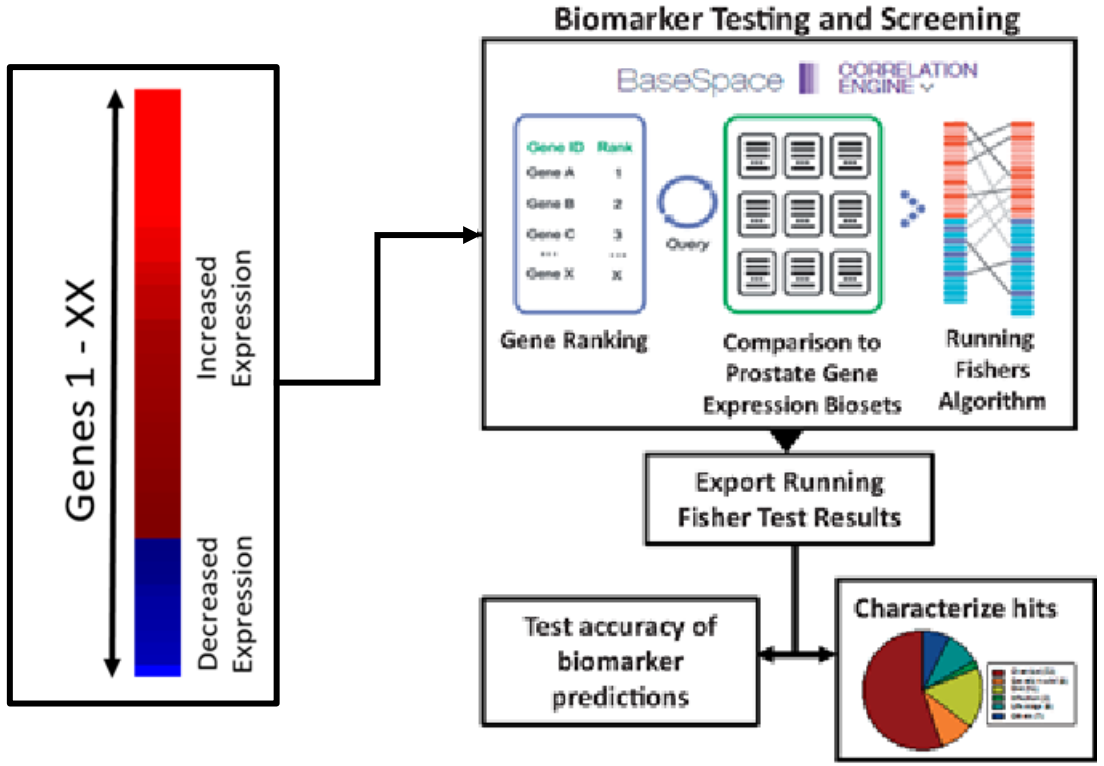
- Corton et al. A Gene Expression Biomarker Identifies Inhibitors of Two Classes of Epigenome Effectors in a Human Microarray Compendium. Submitted.

Stress factors

- Cervantes PW, Corton JC. (2021). A Gene Expression Biomarker Predicts Heat Shock Factor 1 Activation in a Gene Expression Compendium. Chem Res Toxicol. 2021 34(7):1721-1737.
- Jackson AC, Liu J, Vallanat B, Jones C, Nelms MD, Patlewicz G, Corton JC. (2020). Identification of novel activators of the metal responsive transcription factor (MTF-1) using a gene expression biomarker in a microarray compendium. Metallomics. 12(9):1400-1415.
- Korunes KL, Liu J, Huang R, Xia M, Houck KA, Corton JC. (2022). A gene expression biomarker for predictive toxicology to identify chemical modulators of NF-κB. PLoS One. 17(2):e0261854.
- Rooney JP, Chorley B, Hiemstra S, Wink S, Wang X, Bell DA, van de Water B, Corton JC. (2020). Mining a human transcriptome database for chemical modulators of NRF2. PLoS One. 15(9):e0239367.

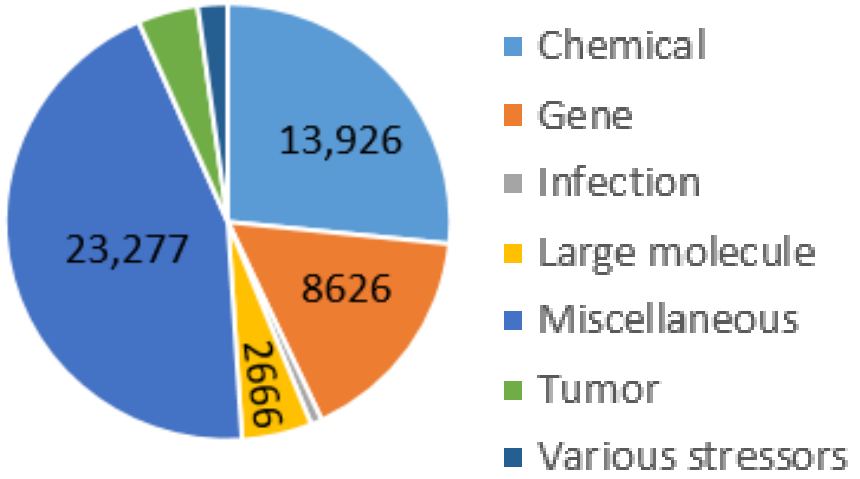
CENTER FOR COMPUTATIONAL TOXICOLOGY AND EXPOSURE

Comparing gene lists in BaseSpace Correlation Engine



- Utilize Illumina's BaseSpace Correlation Engine
- Contains ~130,000 microarray comparisons of statistically significant genes
- Valuable computational tools
- Compares all microarray comparisons to each other in a pairwise fashion using a Running Fisher test
- For each pair-wise comparison: generates the number of overlapping genes, correlation direction and p-value

- ~51,600 microarray comparisons in human database



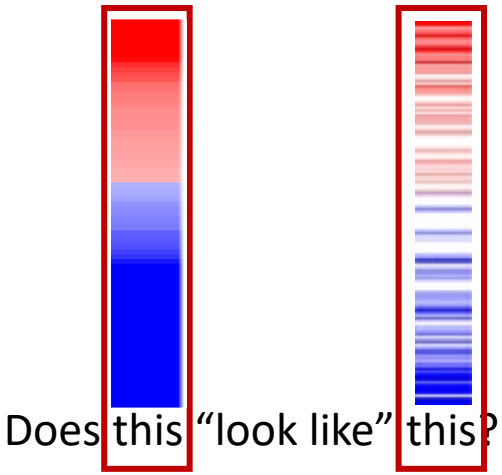
- Includes ~1950 chemicals
- ~8600 perturbations of ~1700 genes (knockdowns, overexpression, mutants)

- Greatly accelerated construction and analysis of biomarkers

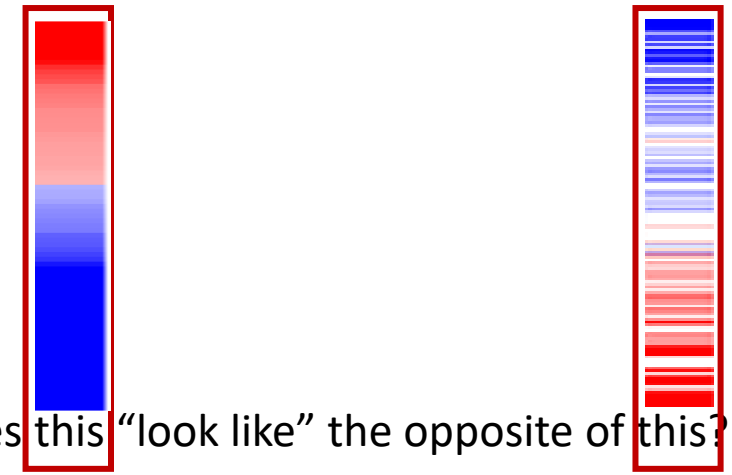
Derived from Rooney et al. Toxicol Sci. 166:146-162

Correlation analysis using the Running Fisher Test

- Identification of factors (chemicals, hormones, diets, genes, etc.) that “look” like your gene list



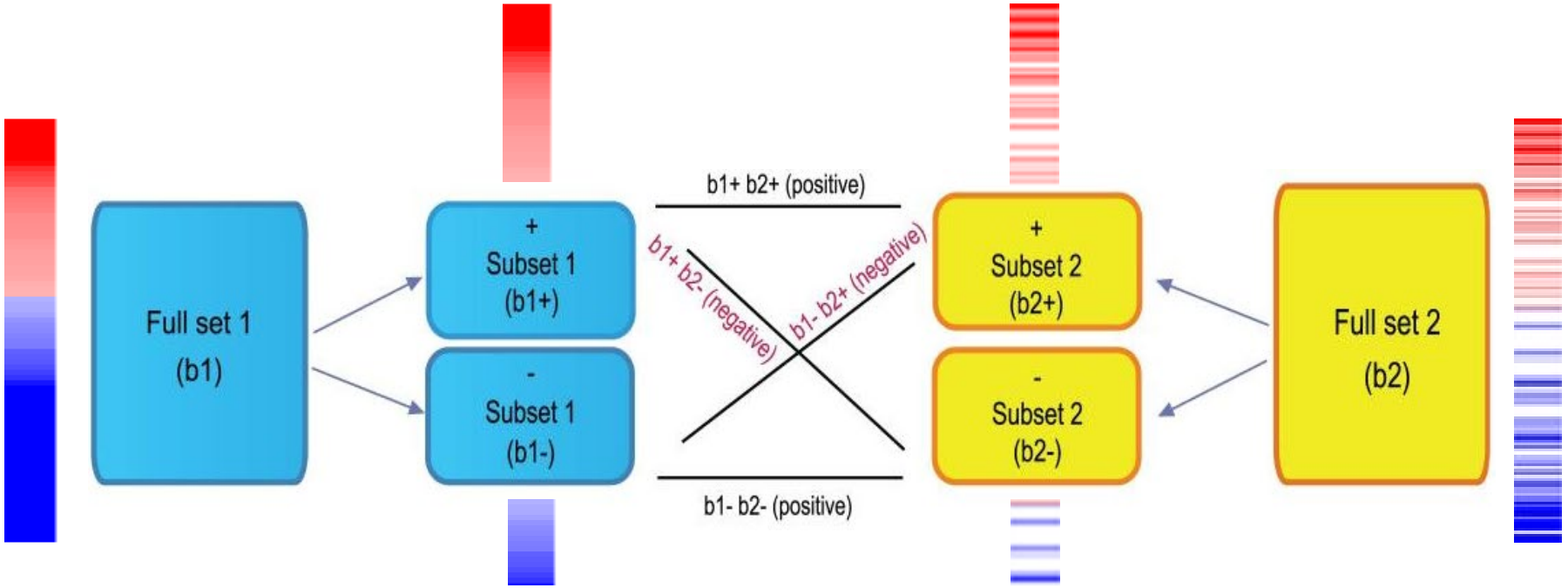
Does this “look like” this?



Does this “look like” the opposite of this?

- Correlation can be determined computationally using the Running Fisher test in BSCE

Computing directionality and final correlation scores between two gene lists



- Score(b1, b2) = sum(b1+b2+, b1+b2-, b1-b2+, b1-b2-)
- Running Fisher Test p-value
- Direction of the correlation

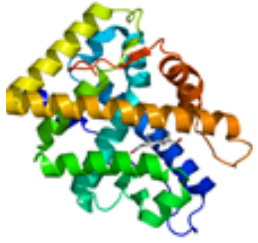
- **The Running Fisher test p-value is a useful metric of correlation between gene sets**

Adapted from Kuperschmidt et al. (2010) PLoS One

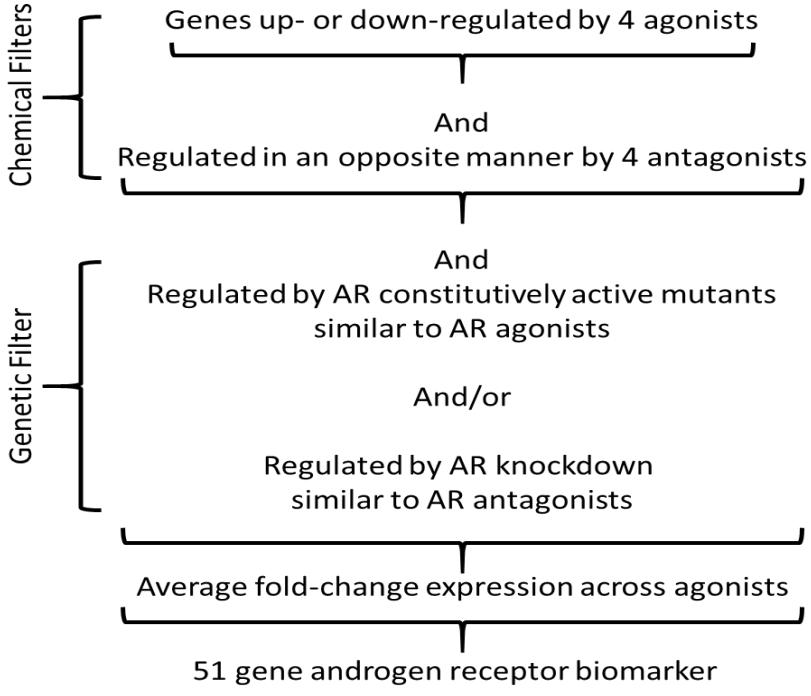
Construction of an AR biomarker – use of gene perturbation comparisons

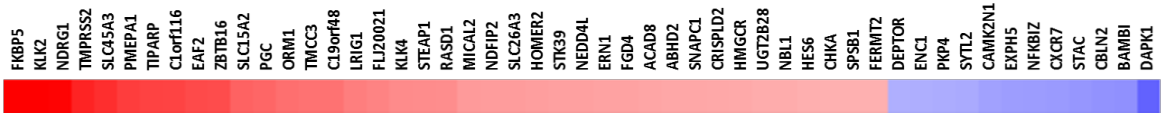
- AR activation is a key driver in androgen-dependent prostate cancer
- Focused on developing methods for predicting AR modulation in AR positive prostate cancer cell lines

Androgen Receptor



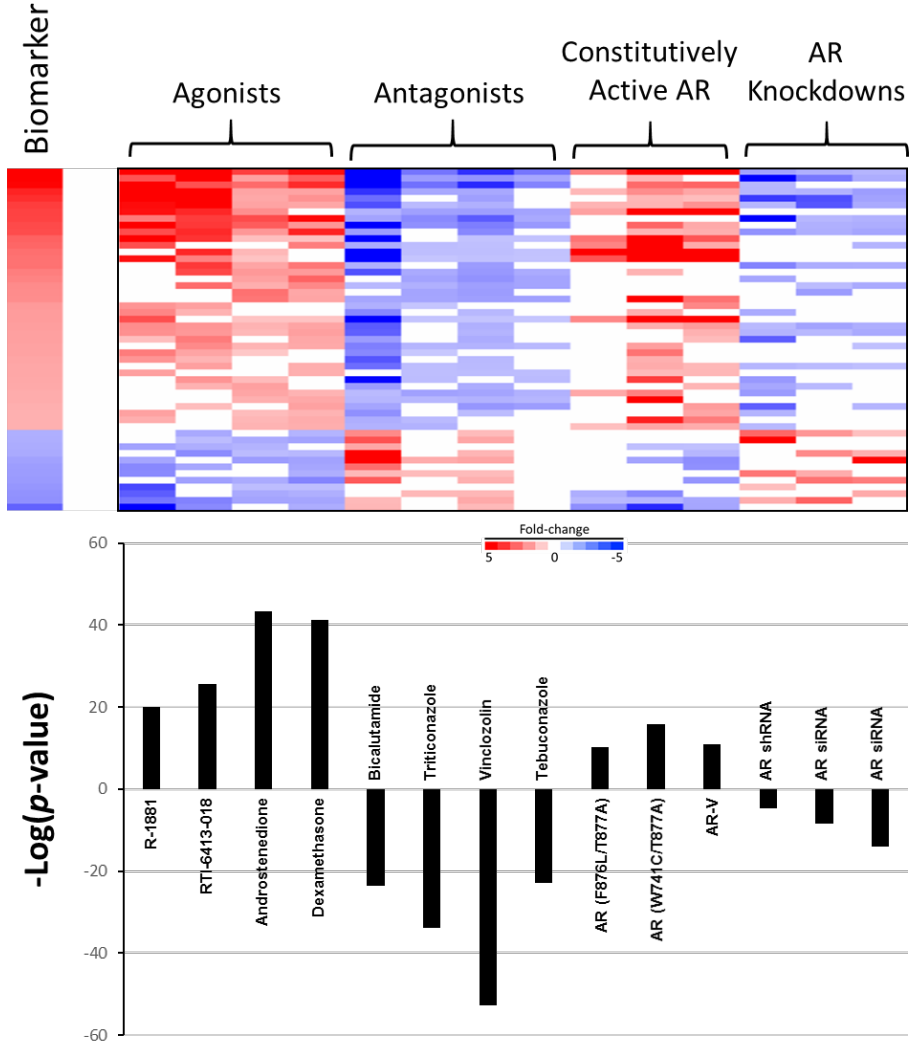
Filters used to identify AR biomarker genes





Constitutively active AR mutants
AR (F876L)
AR (W741C/T877A)
AR-V7

AR knockdowns



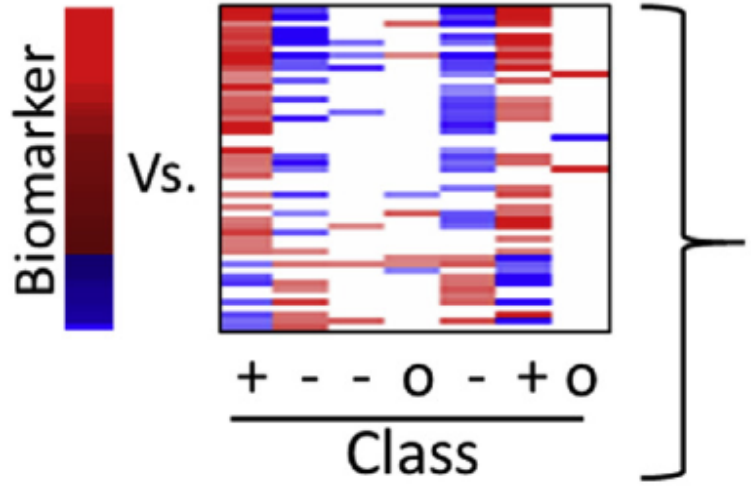
- Consistent activation or suppression across the biosets
- ~90% of the genes are direct targets of AR as determined by post-hoc analysis of ChIP-Seq studies

Rooney et al. (2018) Toxicol Sci. 166:146-162

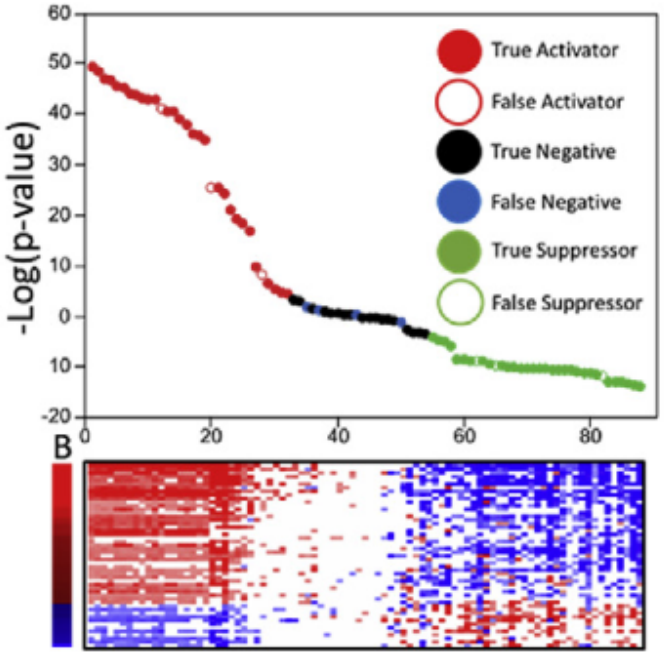
Determination of biomarker accuracy

Accuracy Determination

Comparison of biomarker to chemical profiles with known outcomes



Ranking by Correlation

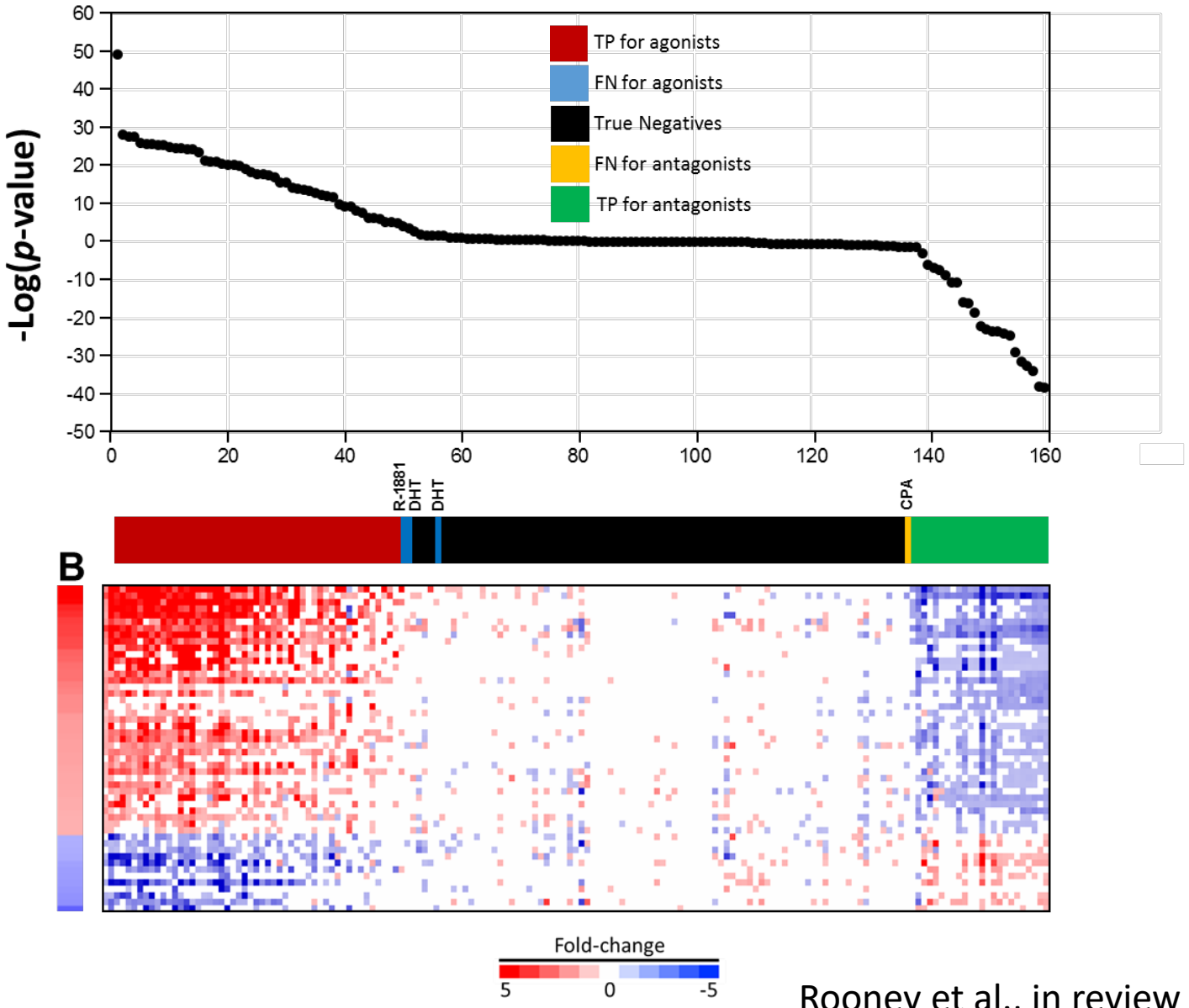


Accuracy Determination

- Sensitivity
- Specificity
- Positive predictive value
- Negative predictive value
- Balanced accuracy

The biomarker predicts AR activation and suppression

- 163 biosets from prostate cancer cells treated with 98 chemicals with known effects on AR
- Classification of activation or suppression required a threshold p-value ≤ 10^{-4}
- For activation, the AR biomarker had a sensitivity of 94% and a specificity of 100%, with a balanced accuracy of 97%
- For suppression, the AR biomarker had a sensitivity of 96% and a specificity of 100%, with a balanced accuracy of 98%
- There were few chemicals in this analysis that were environmentally relevant

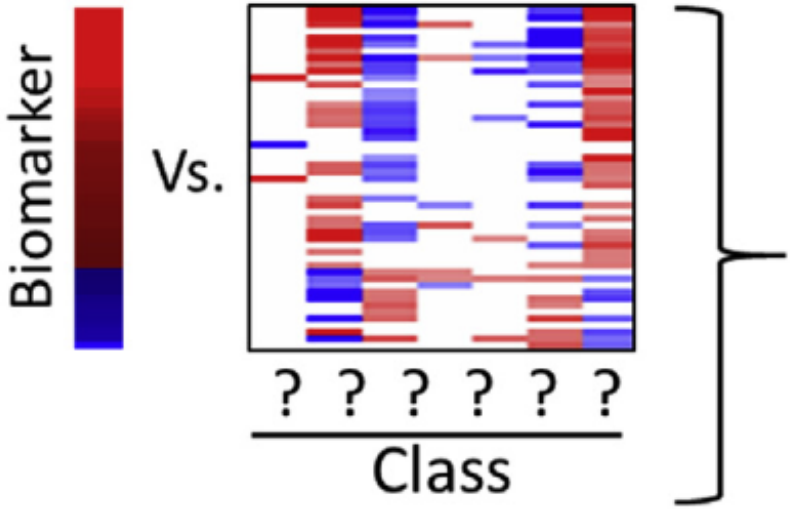


Rooney et al., in review

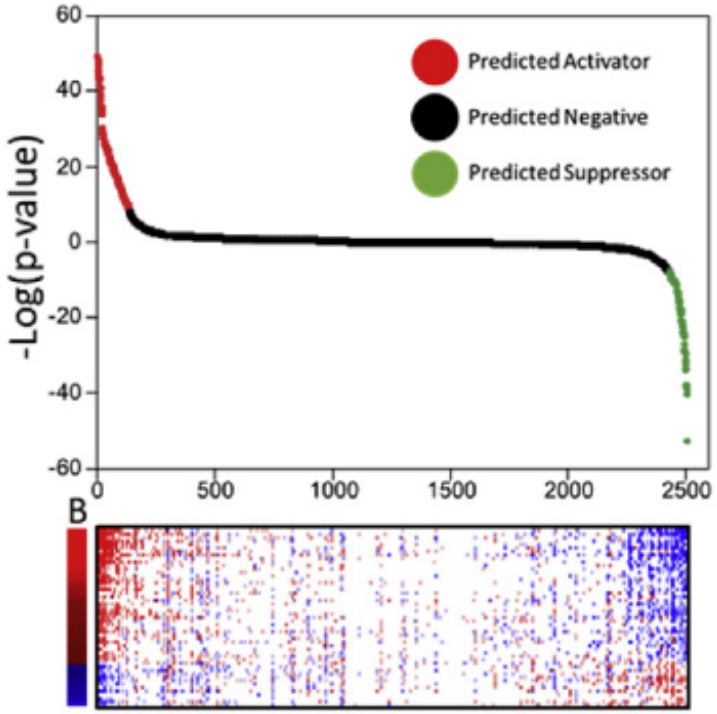
Use of biomarkers in HTTr chemical screening

In silico Screening

Comparison of biomarker to uncharacterized chemicals



Ranking by Correlation

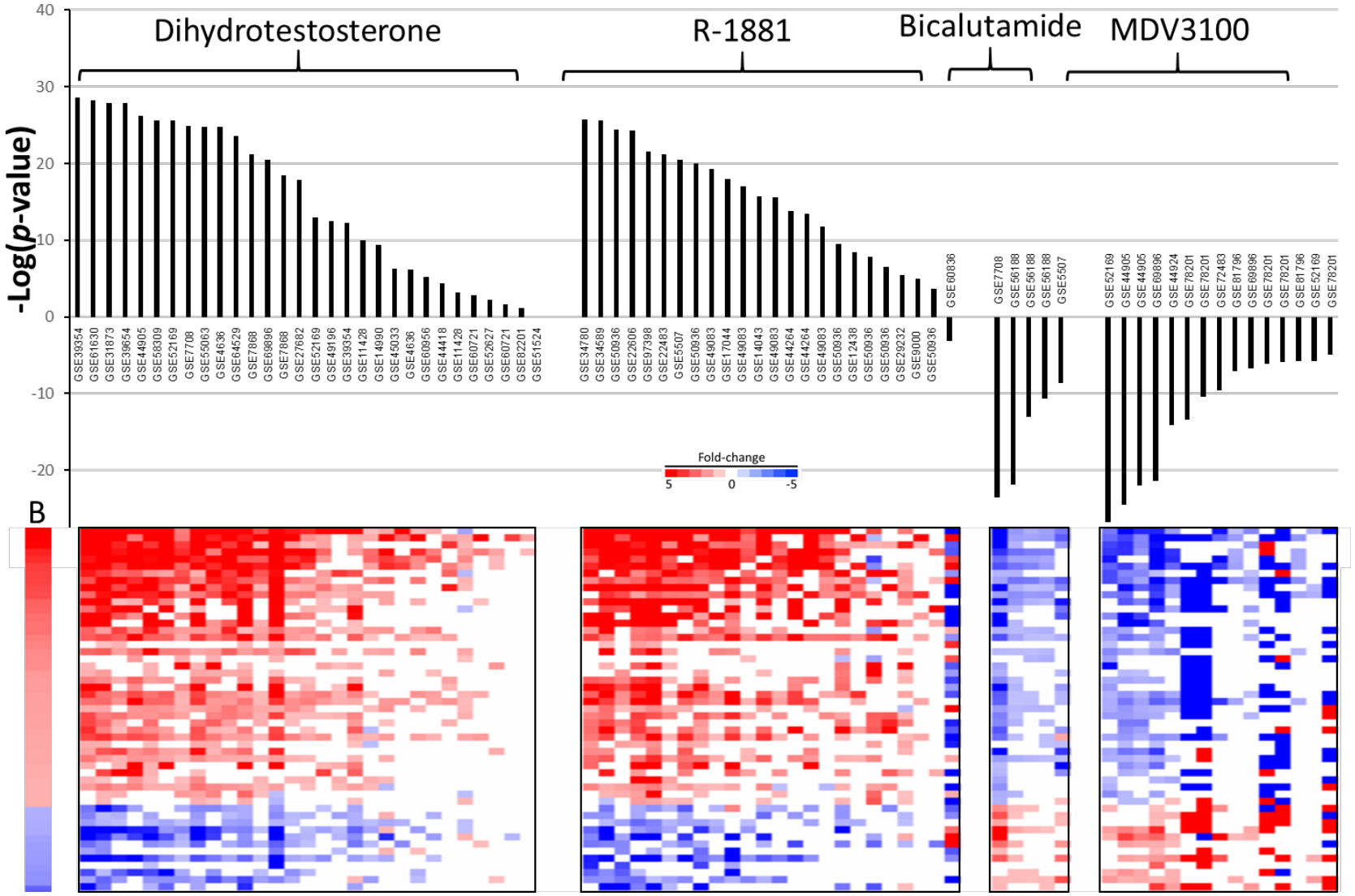


Characterize Hits

- Confirm positives
- Determine mechanism of modulation

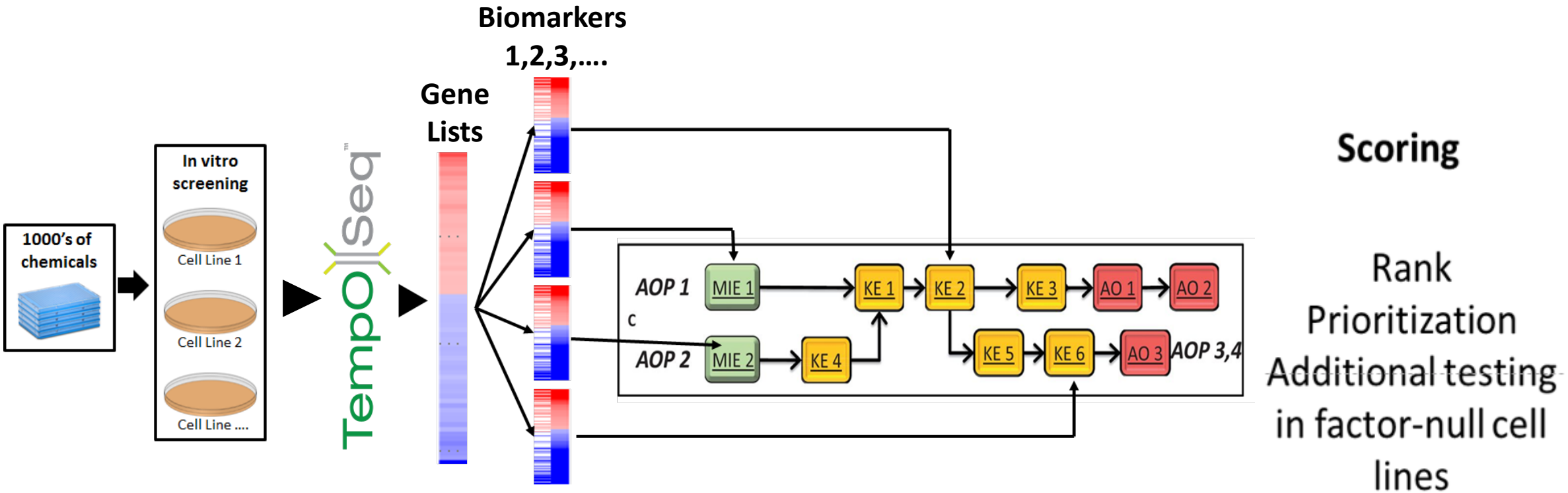
The AR biomarker identifies AR agonists and antagonists

- Examined prototypical AR agonists or antagonists
- Consistent activation or suppression of biomarker responses
- Expression of the biomarker genes reflects the biomarker activation or suppression



Rooney et al., in review

Using gene expression biomarkers to identify molecular targets of chemicals in transcriptomic studies



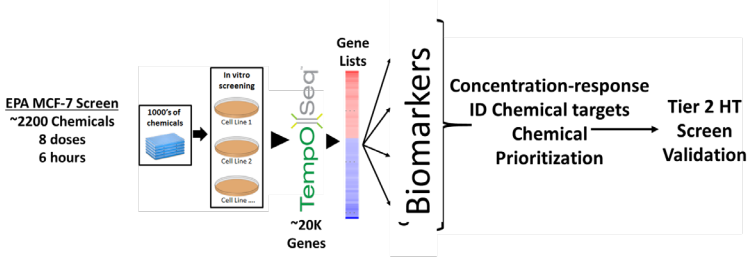
- Use predictions for
  - Chemical prioritization as part of Tier 0 screening
  - Predict molecular initiating events and key event perturbations in adverse outcome pathways
- Followed up with short-term tests in knockout/knockdown cell lines, organotypic cultures or animals
- **Ultimate Goal: Move from hypothesis generation to final predictions to minimize further testing**

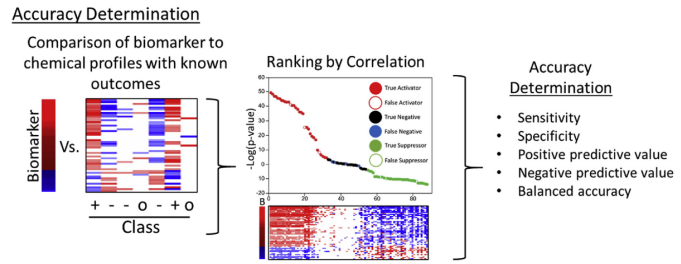
What you now know!

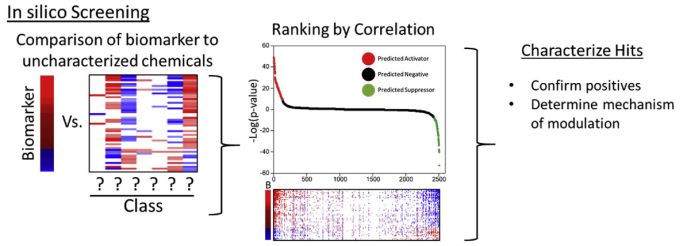
- How high-throughput transcript profiling is carried out as a method to comprehensively assess the effects of chemicals on biological systems
  - Different platforms for assessing genome scale gene expression changes
  - High-throughput transcript profiling (HTTr)

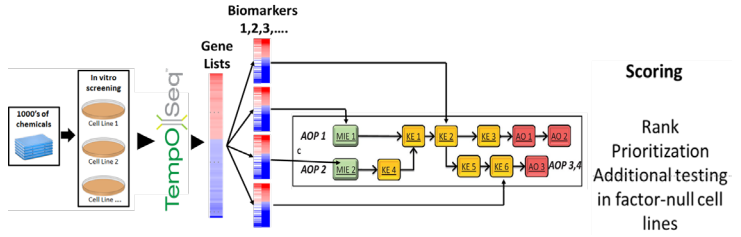
- How to identify the molecular targets of chemicals
  - Hypothesis generating tools
  - Gene expression biomarkers
  - How to
    - Identify predictive gene sets
    - Characterize the gene sets
    - Determine predictive accuracy
    - Use in screening chemicals

- How to link the alterations in molecular targets to potential adverse events.
  - Use of the adverse outcome framework









United States Environmental Protection Agency

Questions?

Chris Corton: corton.chris@epa.gov

Supporting Materials

- **Adverse Outcome Pathways**
  - Link to Wiki: https://aopwiki.org/
  - General reviews of AOPs
    - Carusi et al. (2018) Sci Total Environ. 628-629:1542.
    - Ankley and Edwards (2018) Curr Opin Toxicol. 9:1.
    - Leist et al. (2017) Arch Toxicol. 91:3477.
    - Vinken et al. (2017) Arch Toxicol 91:3697.
    - Ankley et al. (2010) Environ Toxicol Chem. 29:730.
  - Using AOPs to help guide building predictive assays
    - Coady et al. (2019) Integrated Environmental Assessment and Management 15:633.
    - Wang et al. (2019) Environ Int 126:377.

- **General papers and reviews on the construction and use of gene expression biomarkers**
  - Li et al. (2017) Proc Natl Acad Sci U S A. 114:E10881-E10889.
  - Corton et al. (2019) Toxicol Appl Pharmacol. 380:114683.
  - Corton (2019) Current Opinion in Toxicol 18:54.

- **Construction and use of rat liver gene expression biomarkers**
  - Rooney et al. (2018) Toxicol Appl Pharmacol. 356:99.

Biomarkers that predict key events in human cells in vitro

Endocrine disruption

- Ryan et al. (2016). Moving Toward Integrating Gene Expression Profiling Into High-Throughput Testing: A Gene Expression Biomarker Accurately Predicts Estrogen Receptor α Modulation in a Microarray Compendium. Toxicol Sci. 151(1):88-103.
- Androgen receptor: Rooney et al. (2018). Identification of Androgen Receptor Modulators in a Prostate Cancer Cell Line Microarray Compendium. Toxicol Sci. 166:146-162.

DNA Damage Response

- Corton et al. (2018). Using a gene expression biomarker to identify DNA damage-inducing agents in microarray profiles. Environ Mol Mutagen. 59:772-784.
- Cho et al. (2019). Assessment of the performance of the TGx-DDI biomarker to detect DNA damage-inducing agents using quantitative RT-PCR in TK6 cells. Environ Mol Mutagen. 60:122-133.
- Corton JC, Witt KL, Yauk CL. (2019). Identification of p53 Activators in a Human Microarray Compendium. Chem Res Toxicol. 32(9):1748-1759.

Epigenetic effects

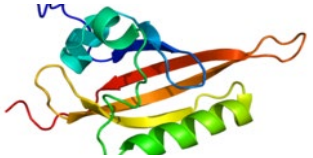
- Corton et al. A Gene Expression Biomarker Identifies Inhibitors of Two Classes of Epigenome Effectors in a Human Microarray Compendium. Submitted.

Stress factors

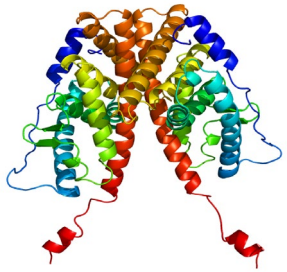
- Cervantes PW, Corton JC. (2021). A Gene Expression Biomarker Predicts Heat Shock Factor 1 Activation in a Gene Expression Compendium. Chem Res Toxicol. 2021 34(7):1721-1737.
- Jackson AC, Liu J, Vallanat B, Jones C, Nelms MD, Patlewicz G, Corton JC. (2020). Identification of novel activators of the metal responsive transcription factor (MTF-1) using a gene expression biomarker in a microarray compendium. Metallomics. 12(9):1400-1415.
- Korunes KL, Liu J, Huang R, Xia M, Houck KA, Corton JC. (2022). A gene expression biomarker for predictive toxicology to identify chemical modulators of NF-κB. PLoS One. 17(2):e0261854.
- Rooney JP, Chorley B, Hiemstra S, Wink S, Wang X, Bell DA, van de Water B, Corton JC. (2020). Mining a human transcriptome database for chemical modulators of NRF2. PLoS One. 15(9):e0239367.

Biomarkers that predict key events in the livers of mice and rats

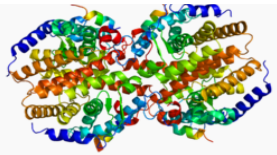
AhR



Estrogen Receptor α

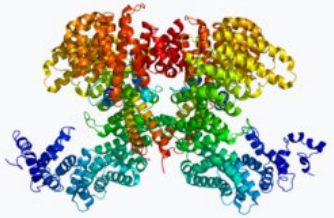


CAR

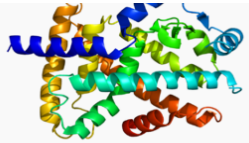


NRF2

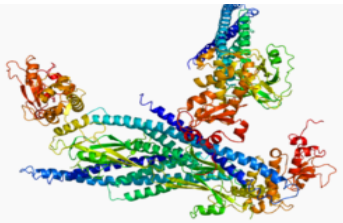
SREBP



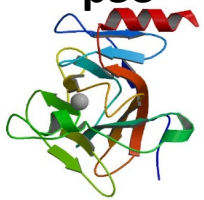
PPARα



STAT5b



p53



- Oshida et al. (2015). Identification of Modulators of the Nuclear Receptor Peroxisome Proliferator-Activated Receptor α (PPARα) in a Mouse Liver Gene Expression Compendium. PLoS One. 10(2):e0112655.
- Oshida et al. (2015). Identification of Chemical Modulators of the Constitutive Activated Receptor (CAR) in a Mouse Liver Gene Expression Compendium. Nuclear Receptor Signaling. 13:e002.
- Oshida et al. (2015). Screening a Mouse Liver Gene Expression Compendium Identifies Effectors of the Aryl Hydrocarbon Receptor (AhR). Toxicology. 336:99-112.
- Oshida et al. (2015). Disruption of STAT5b-Regulated Sexual Dimorphism of the Liver Transcriptome by Diverse Factors Is a Common Event. PLoS One. 11(3):e0148308.
- Oshida et al. (2015). Chemical and Hormonal Effects on STAT5b-Dependent Sexual Dimorphism of the Liver Transcriptome. PLoS One. 2016 11(3):e0150284.
- Rosen et al. (2017). PPARα-independent transcriptional targets of perfluoroalkyl acids revealed by transcript profiling. Toxicology. 387:95-107.
- Rooney et al. (2017). Genomic Effects of Androstenedione and Sex-Specific Liver Cancer Susceptibility in Mice. Toxicol Sci. 160(1):15-29.
- Rooney et al. (2018) Activation of Nrf2 in the liver is associated with stress resistance mediated by suppression of the growth hormone-regulated STAT5b transcription factor. PLoS One. 13(8):e0200004.
- Rooney et al. (2018). Activation of CAR leads to activation of the oxidant-induced Nrf2. Toxicol Sci. 167:172-189.
- Rooney et al. (2018). Adverse outcome pathway-driven identification of rat liver tumorigens in short-term assays. Toxicol Appl Pharmacol. 356:99-113.
- Corton (2019). Frequent Modulation of the Sterol Regulatory Element Binding Protein (SREBP) by Chemical Exposure in the Livers of Rats. Comput. Toxicol. 10:113-129.

Construction of an Androgen Receptor Biomarker

A

Chemical Filters

Genes up- or down-regulated by 4 agonists

And

Regulated in an opposite manner by 4 antagonists

Genetic Filter

And

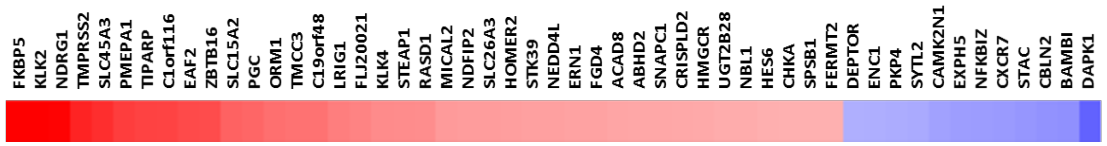
Regulated by AR constitutively active mutants similar to AR agonists

And/or

Regulated by AR knockdown similar to AR antagonists

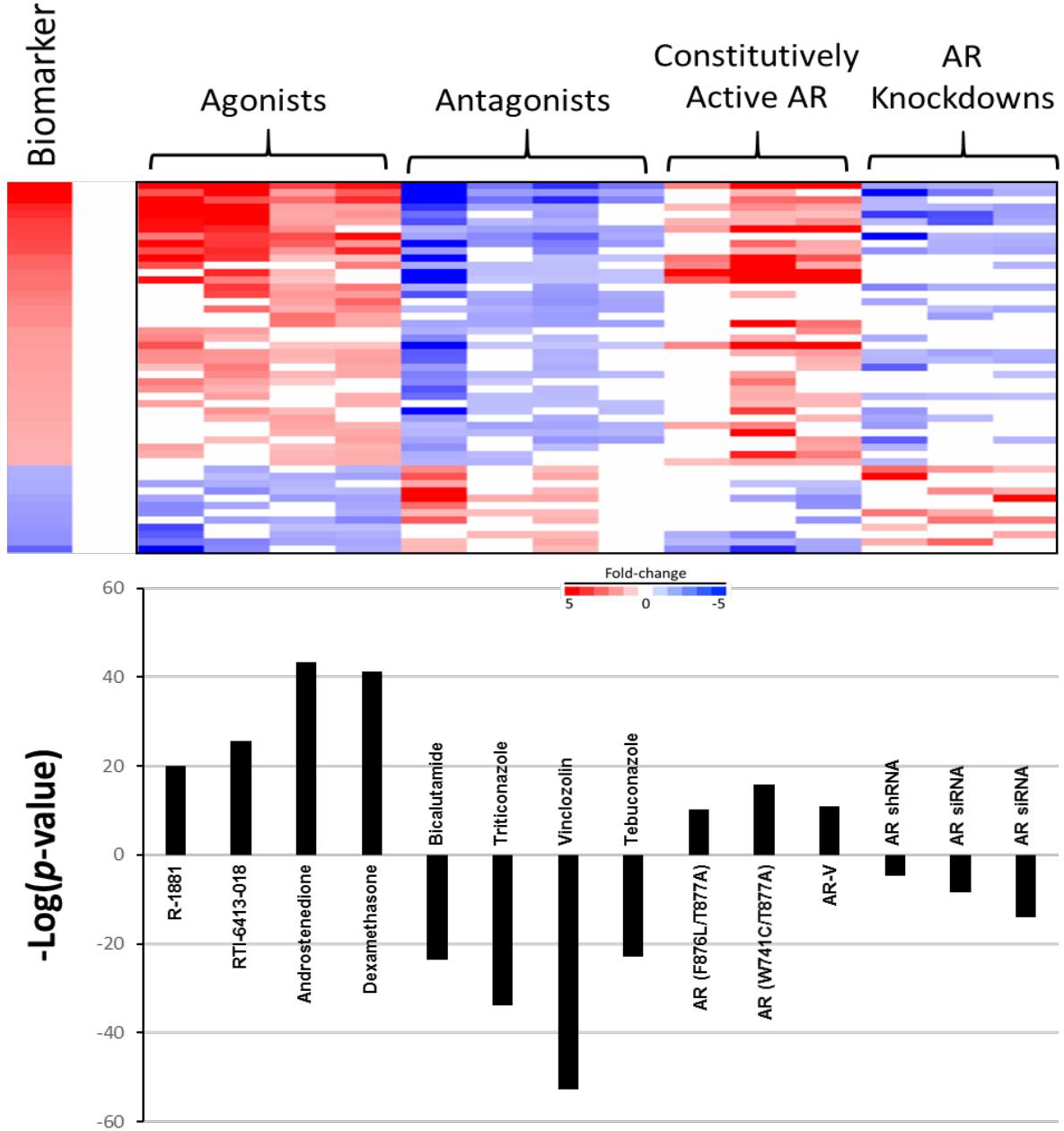
Average fold-change expression across agonists

51 gene androgen receptor biomarker

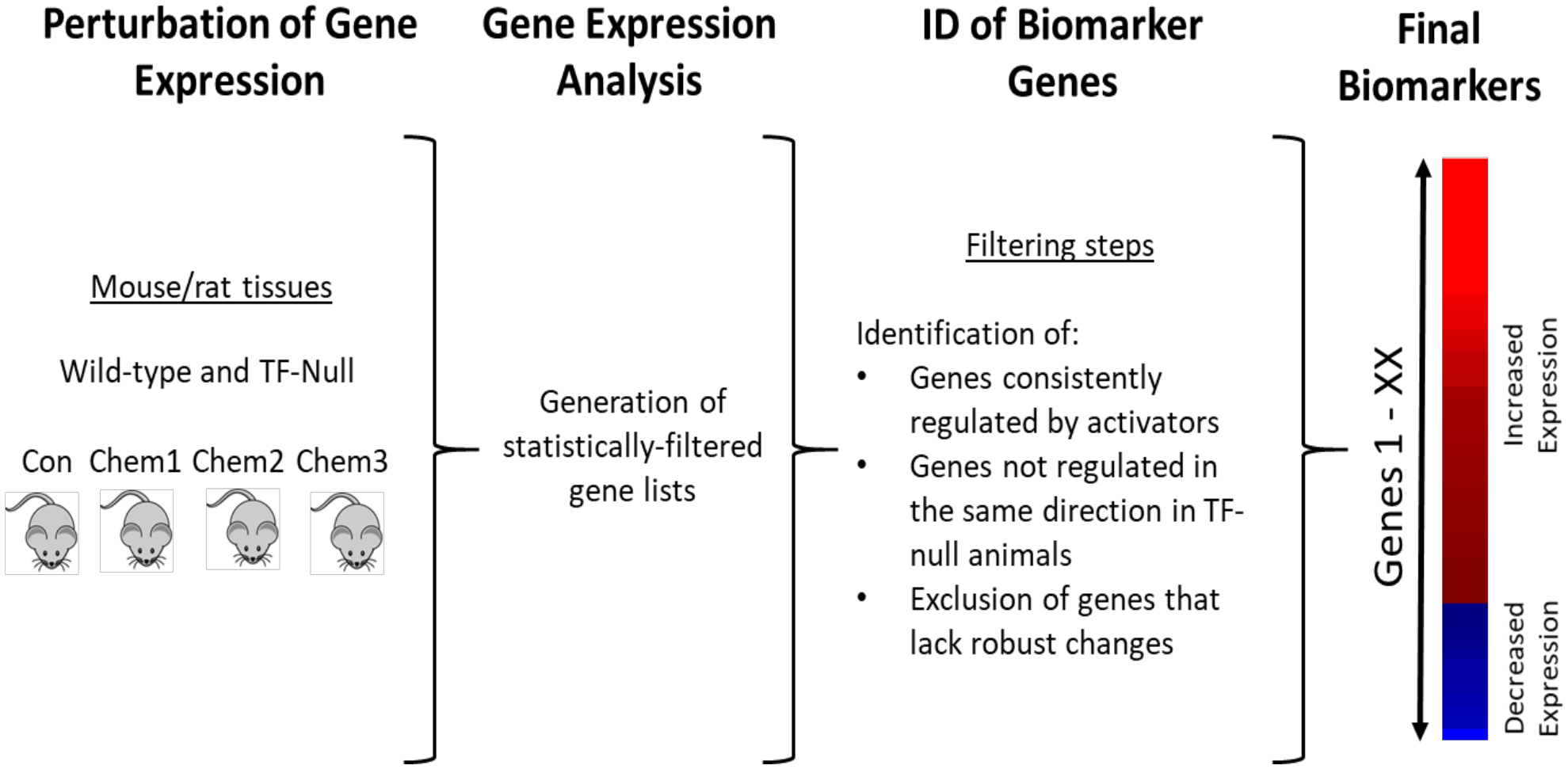


From Rooney et al. (2018) ToxSci. In press.

B



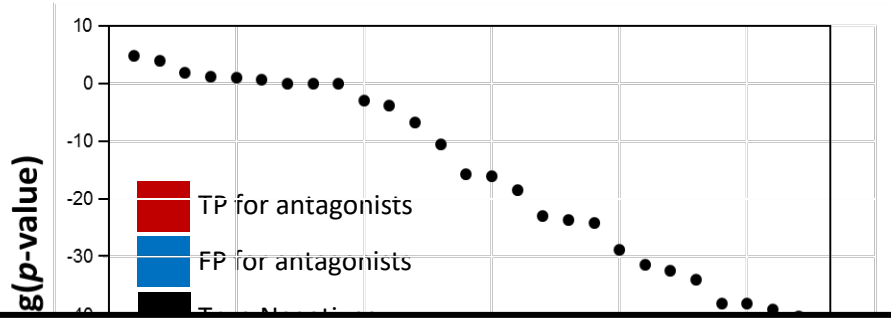
Construction of biomarkers from microarray data generated in animal tissues



From Corton (2019) Current Opinion in Toxicol 18:54

The AR biomarker accurately replicates the ToxCast AR pathway model

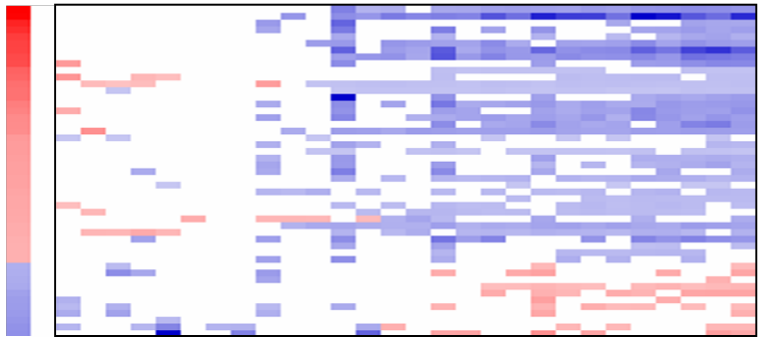
- The ToxCast AR pathway model uses 11 ToxCast HT assays to identify AR actives (Kleinstreuer et al., Chem Res Toxicol. 2017 Apr 17;30(4):946-964)
  - In examining the 1855 ToxCast chemicals, most of the AR hits (87%) were anti-androgens
- Set out to examine a set of antiandrc
  the study to determine if the biomar
  replicate the results of the AR pathw
- Prostate cancer cell line LAPC-4 cells
  exposed to 28 chemicals in antagonis
  - Chemical+R1881(0.33nM) vs. R1881(0.33nM)
  - Exposed cells for 6 hrs
  - Biological replicates (cells exposed on three separate days)
- Examined gene expression using Illumina bead arrays
- Gene expression analyzed using Partek Genomics



	Suppression
True positives	16
True negatives	10
False positives	1
False negatives	1
Sensitivity	0.941
Specificity	0.909
Positive predictive value	0.941
egative predictive value	0.909
alanced accuracy	**0.925**

Take home message:

- The AR biomarker and computational methods can replicate the accuracy of the ToxCast AR pathway model



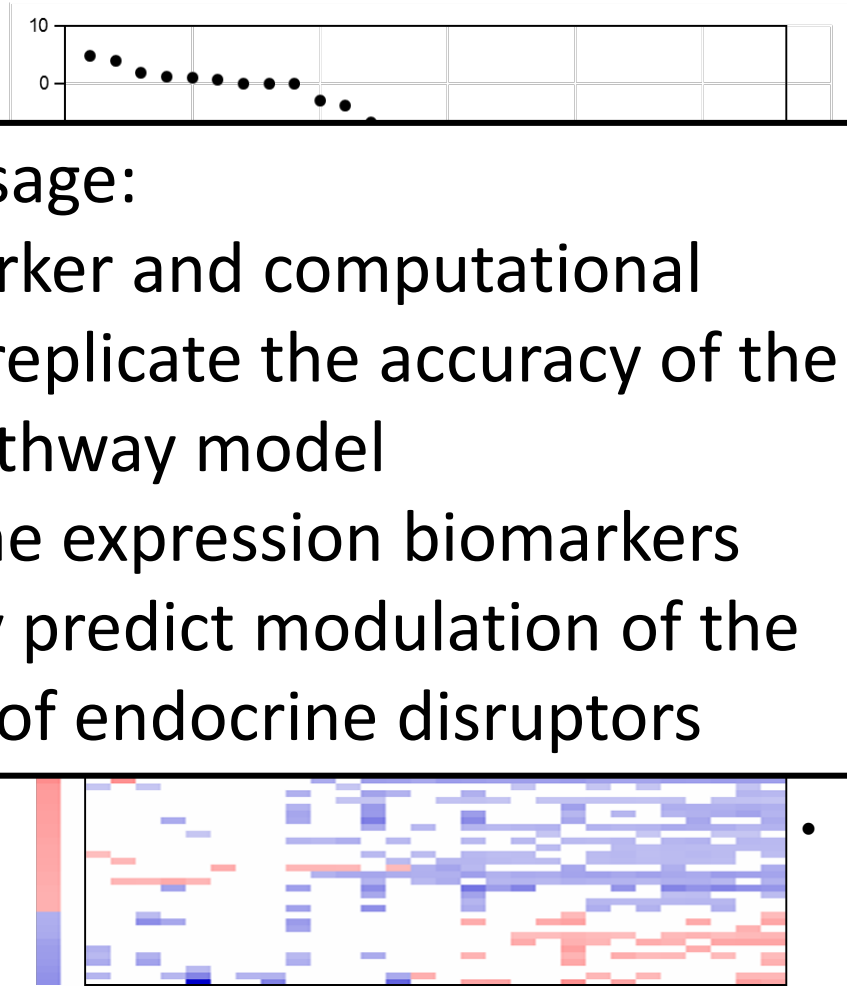
“false positive”
steride is a known anti-androgen but not identified using the ToxCast model

- Incorporating profiles from genetic perturbations into biomarkers may help to increase accuracy of predictions

Rooney et al., in review

The AR biomarker accurately replicates the ToxCast AR pathway model

- The ToxCast AR pathway model uses 11 ToxCast HT assays to identify AR actives (Kleinstreuer et al., Chem Res Toxicol. 201
  - In examining the 185 most of the AR hits (androgens
- Set out to examine a set o the study to determine if replicate the results of the
- Prostate cancer cell line L exposed to 28 chemicals i
  - Chemical+R1881(0.3 R1881(0.33nM)
  - Exposed cells for 6 hrs
  - Biological replicates (cells exposed on three separate days)
- Examined gene expression using Illumina bead arrays
- Gene expression analyzed using Partek Genomics



	Suppression
True positives	16
True negatives	10
False positives	1
False negatives	1
Sensitivity	0.941
Specificity	0.909
Positive predictive value	0.941
Negative predictive value	0.909
Balanced accuracy	**0.925**

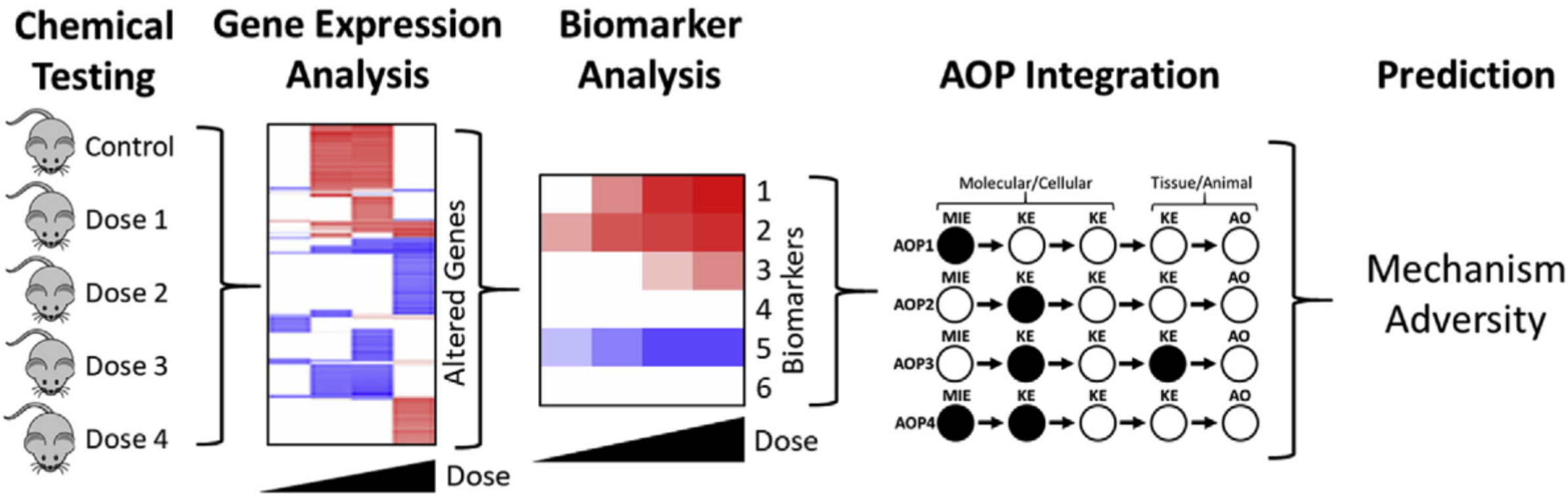
Take home message:

- The AR biomarker and computational methods can replicate the accuracy of the ToxCast AR pathway model
- Show that gene expression biomarkers can accurately predict modulation of the major targets of endocrine disruptors

he "false positive" nasteride is a known anti- ndrogen but not identified sing the ToxCast model

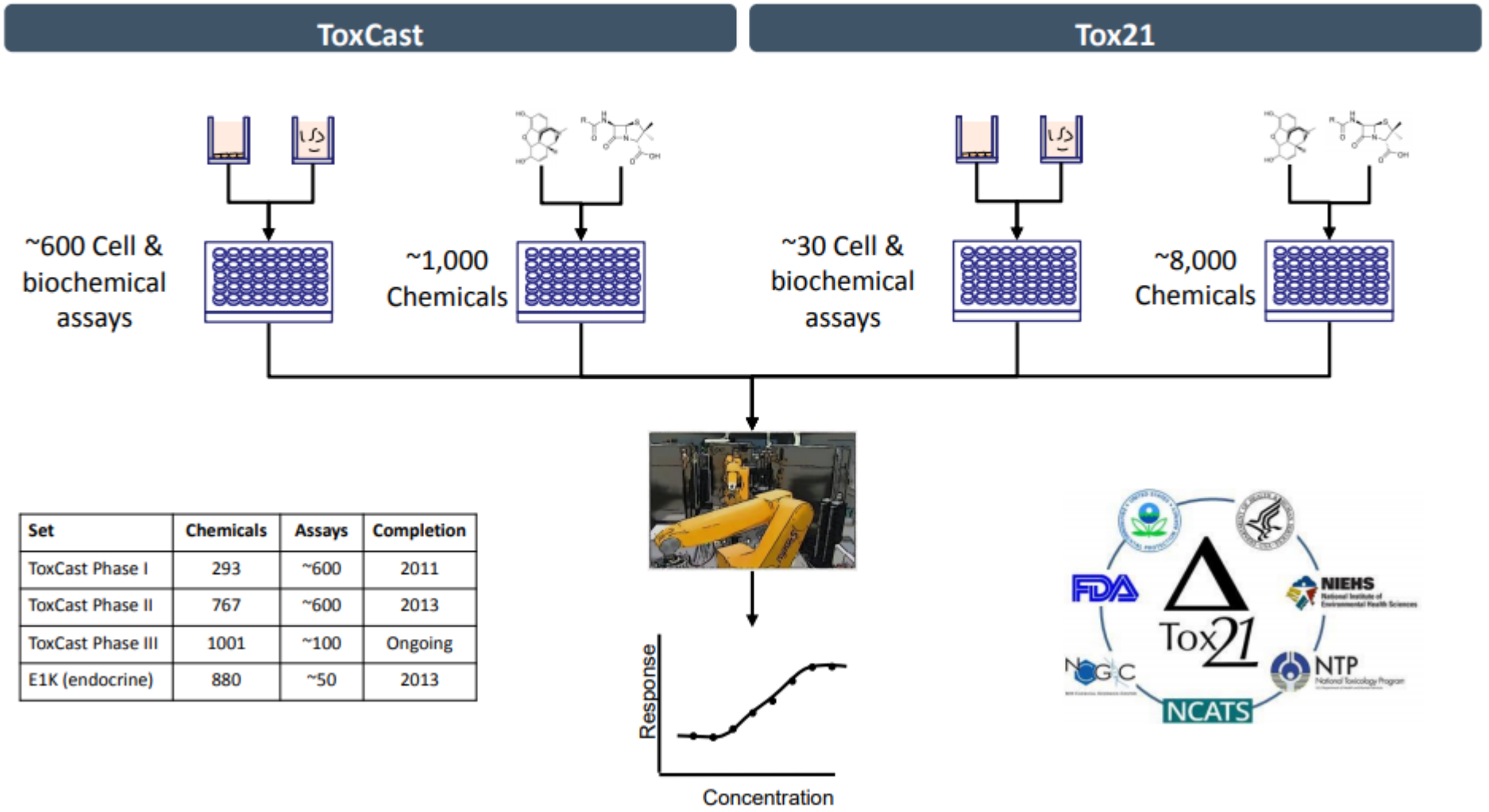
- Incorporating profiles from genetic perturbations into biomarkers may help to increase accuracy of predictions

Rooney et al. Toxicological Sciences. In press.


Chemical Testing
Gene Expression Analysis
Biomarker Analysis
AOP Integration
Prediction
Control
Dose 1
Dose 2
Dose 3
Dose 4
Altered Genes
Dose
1
2
3
4
5
6
Biomarkers
Dose
Molecular/Cellular
Tissue/Animal
MIE
KE
KE
KE
AO
AOP1
AOP2
AOP3
AOP4
Mechanism
Adversity

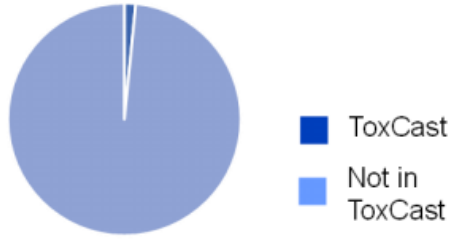
High throughput toxicity testing

ToxCast and Tox21 High-Throughput Screening



Set	Chemicals	Assays	Completion
ToxCast Phase I	293	~600	2011
ToxCast Phase II	767	~600	2013
ToxCast Phase III	1001	~100	Ongoing
E1K (endocrine)	880	~50	2013

Gene Coverage



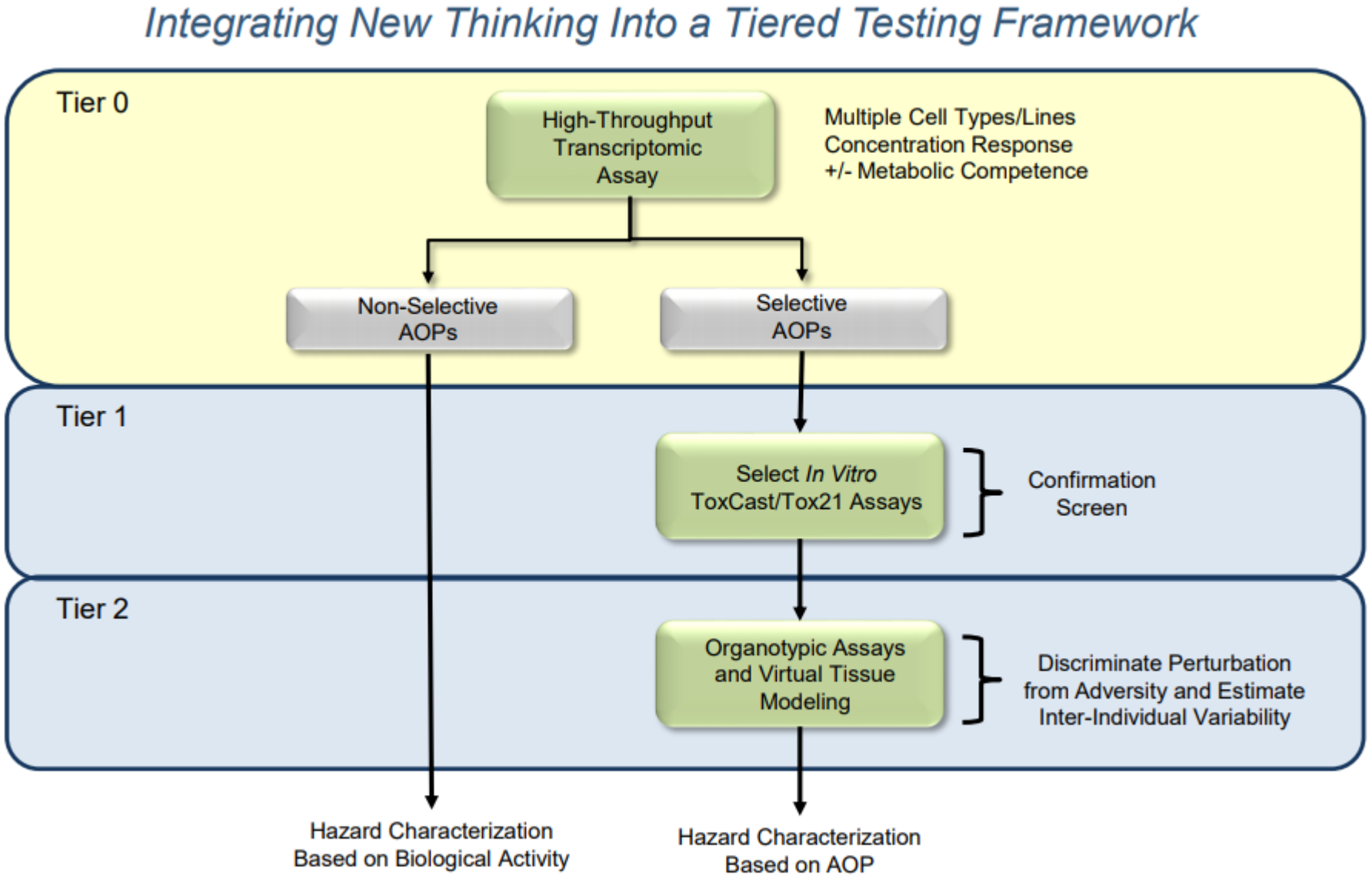
Pathway Coverage*

*At least one gene from pathway represented

What are we missing?

From Thomas, http://qsb.webcast.fi/e/echa/echa_2016_0419_workshop_day2_part2/ECHA_workshop_day2.pdf

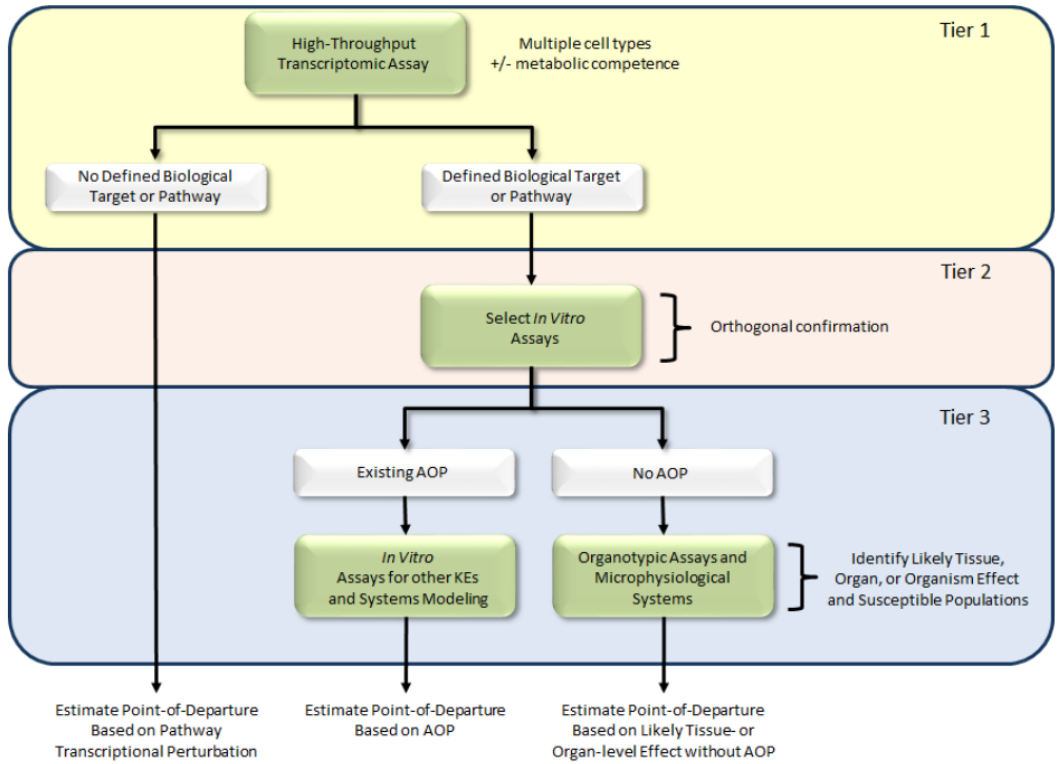
High throughput toxicity testing



From Thomas, http://qsb.webcast.fi/e/echa/echa_2016_0419_workshop_day2_part2/ECHA_workshop_day2.pdf

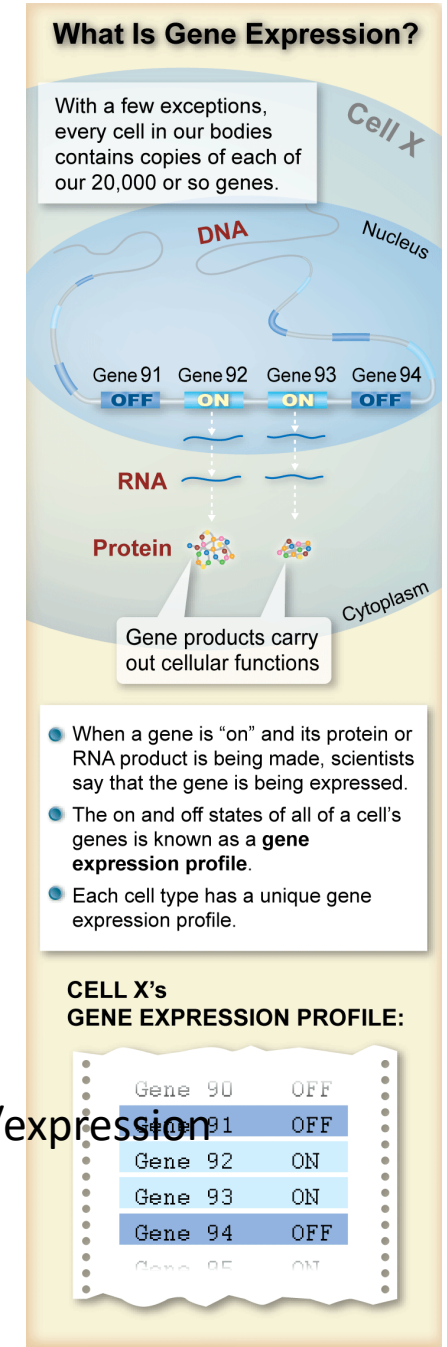
- A flexible, portable and cost efficient platform to comprehensively evaluate the potential biological pathways and processes impacted by chemical exposure
  → High-throughput transcriptomics (HTTr)
- Identify the concentration at which biological pathways/processes begin to be impacted
- Assign putative biological targets for chemicals

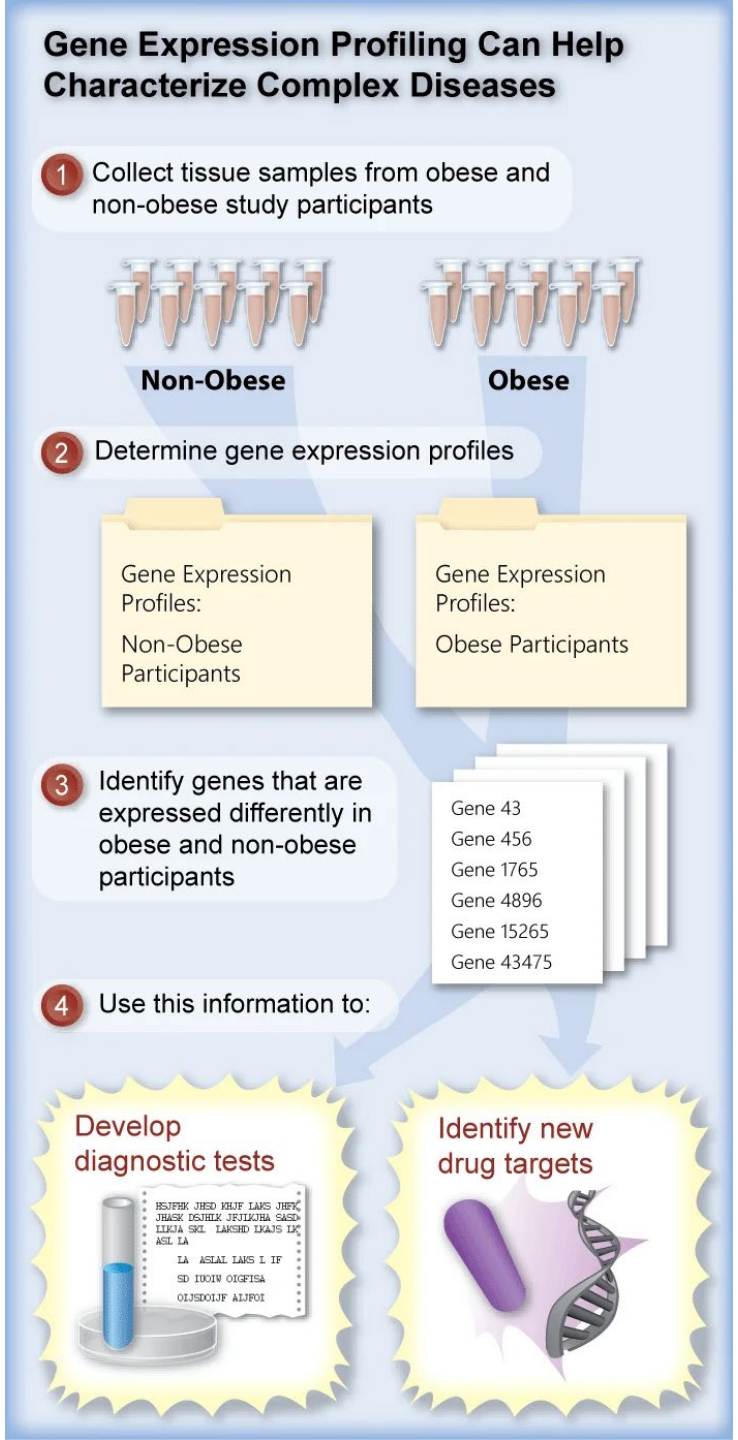
A strategic vision and operational road map for computational toxicology at the U.S. Environmental Protection Agency [DRAFT]



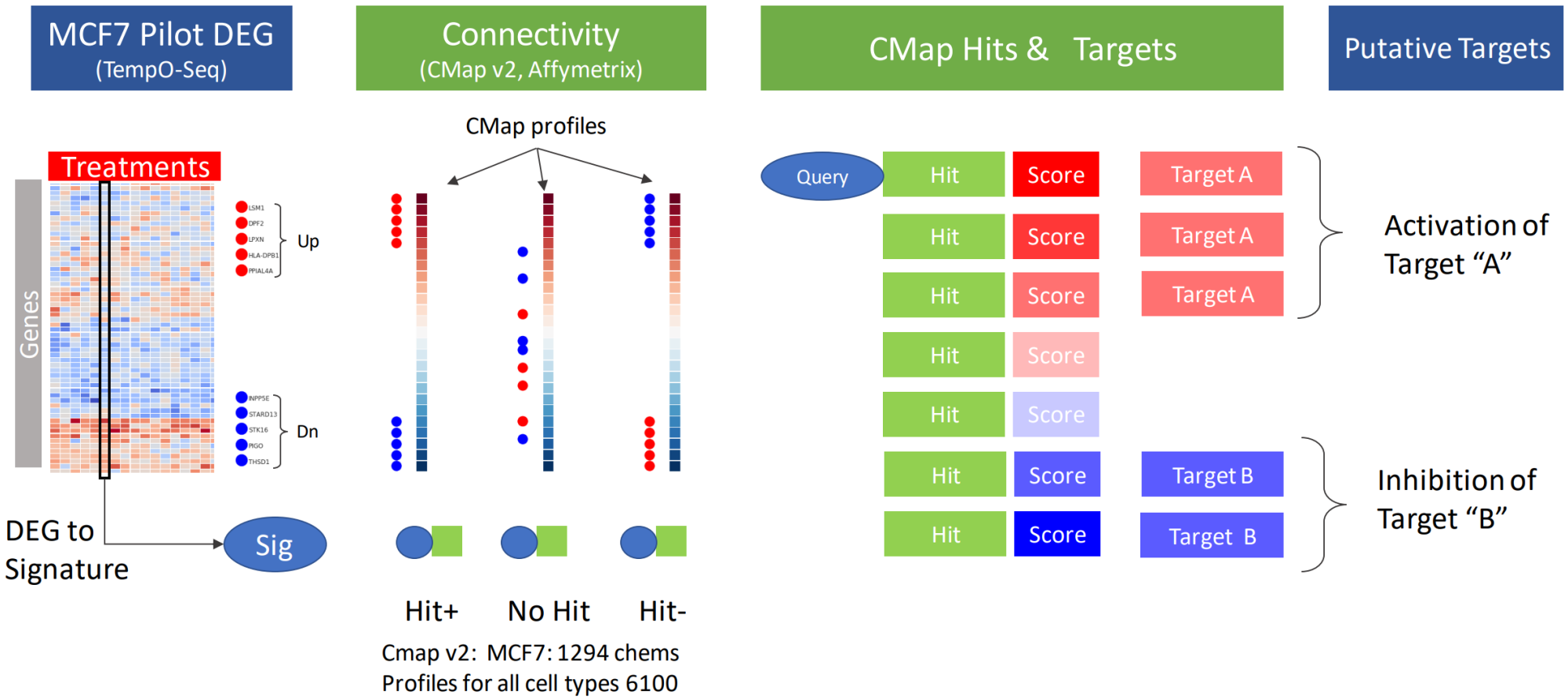
R. Thomas

https://learn.genetics.utah.edu/content/science/expression

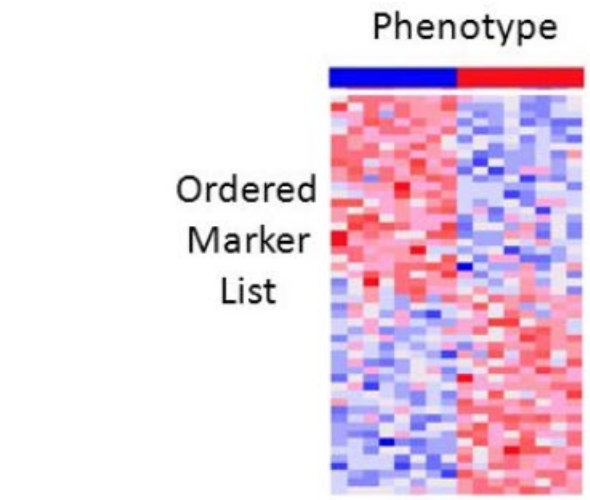




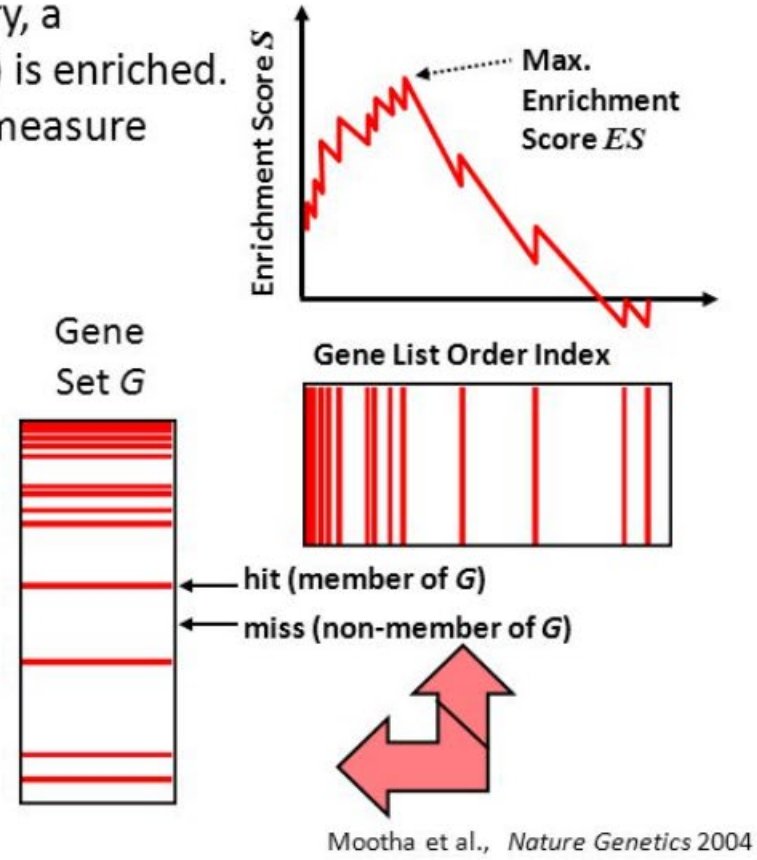
Putative Targets by Gene Set Connectivity



- Rank genes according to their "correlation" with the class of interest.
- Test if a gene set (e.g., a GO category, a pathway, a different class signature) is enriched.
- Use *Kolmogorov-Smirnoff* score to measure enrichment.



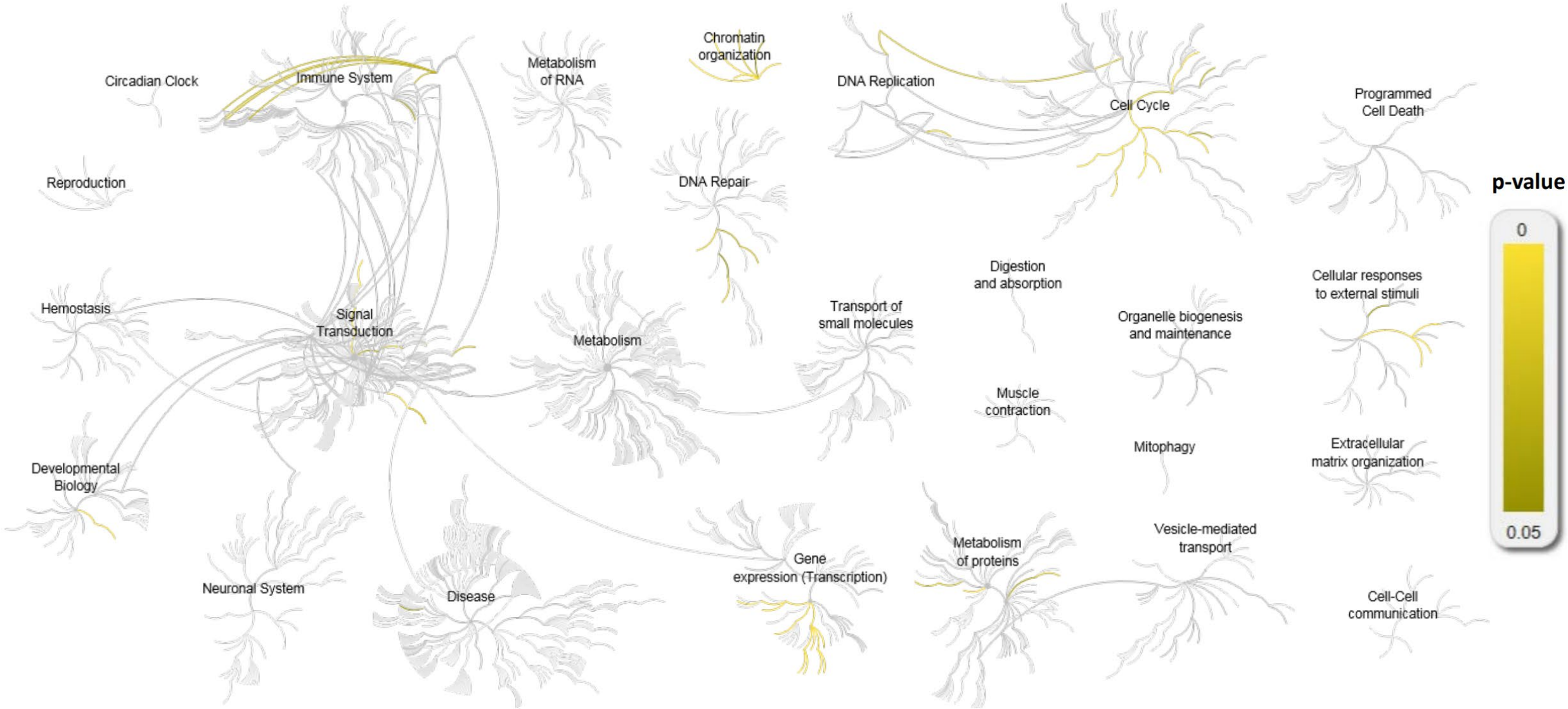
Subramanian et al., PNAS 2005



Mootha et al., *Nature Genetics* 2004

https://www.gsea-msigdb.org/gsea/index.jsp

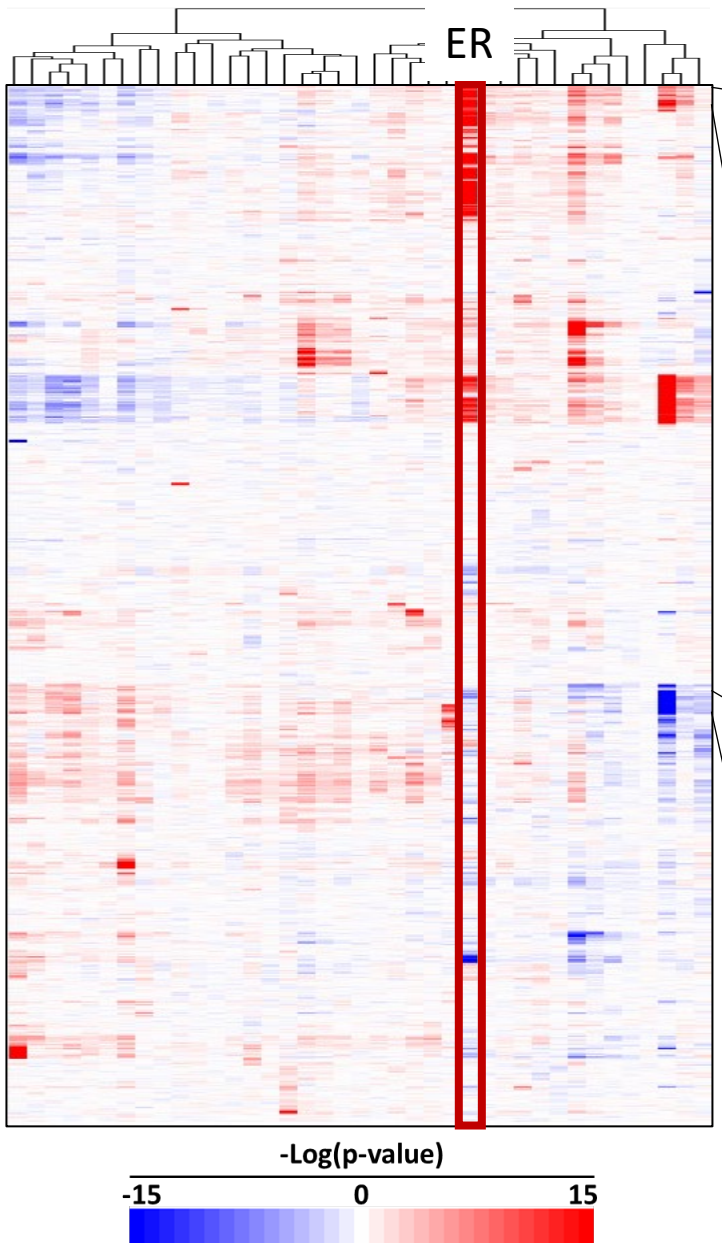
Network Mapping [Clomiphene Citrate]

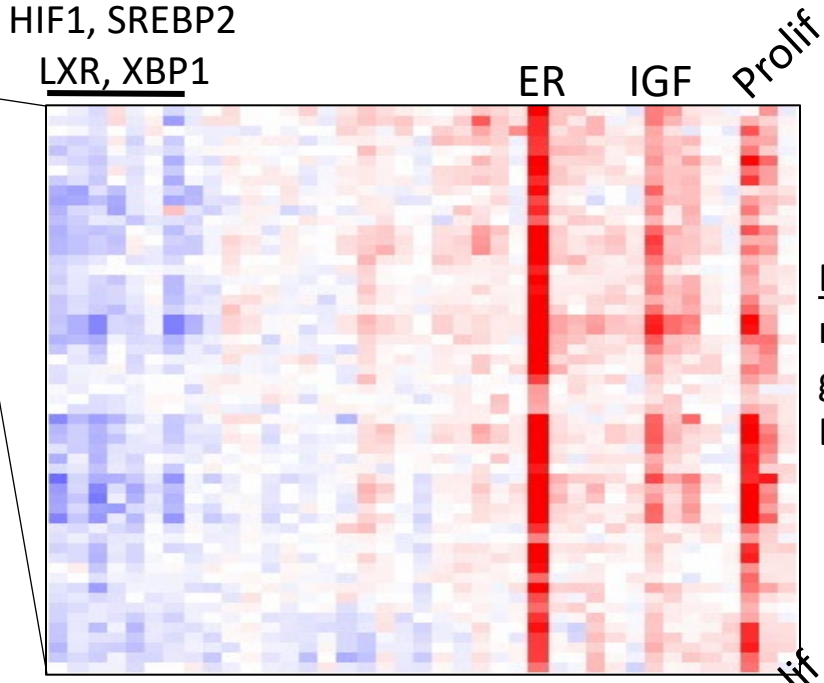


- Reactome (v60) Pathway Hierarchy → Overlaid with enrichment scores based on probes with acceptable BMD model fit
- Highlights different areas of biology affected by a chemical

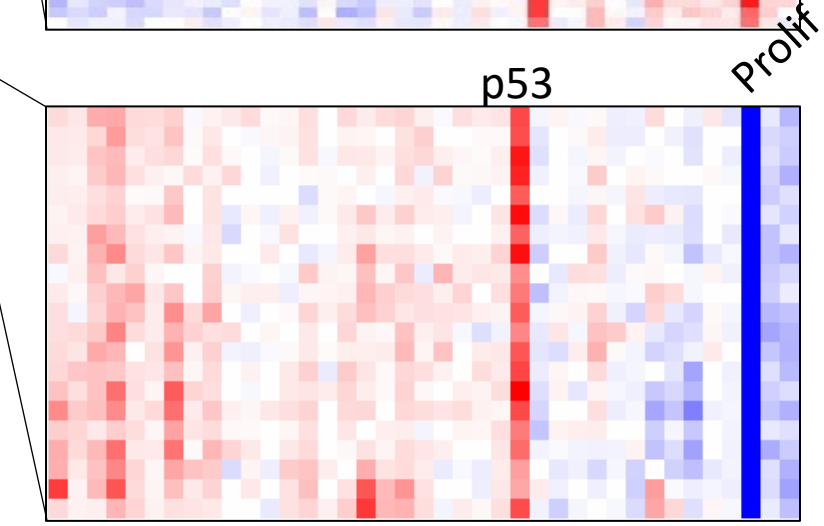
Behavior of biomarkers in MCF-7 cells

- Examined relationships between 2165 microarray comparisons in MCF-7 cells across 39 biomarkers
- Includes chemicals, various stressors, cytokines
- Two-dimensional hierarchical complete linkage clustering





ER activators: E2, nonylphenol, genistein, equol, BPA



p53 activators: nutlin-3, 5-fluorouracil, doxorubicin

The biomarker predicts AR activation and suppression

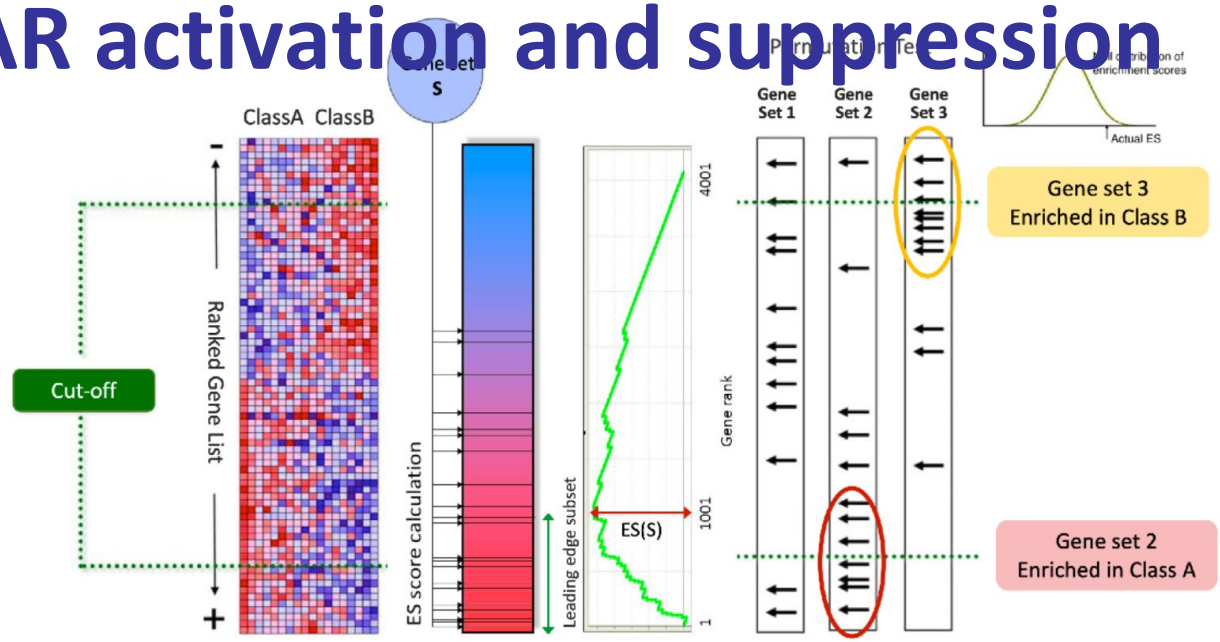
PNAS

Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles

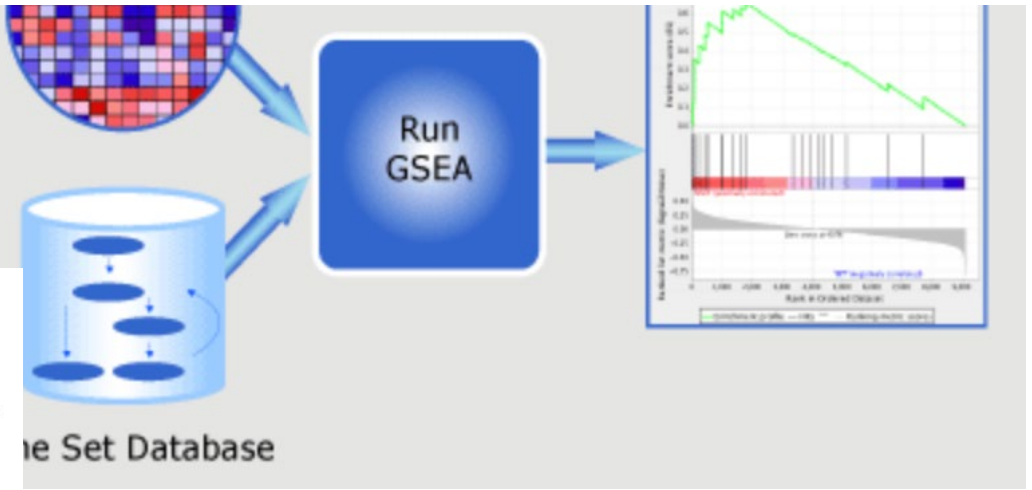
Aravind Subramanian[a,b], Pablo Tamayo[a,b], Vamsi K. Mootha[a,c], Sayan Mukherjee[d], Benjamin L. Ebert[a,e], Michael A. Gillette[a,f], Amanda Paulovich[g], Scott L. Pomeroy[h], Todd R. Golub[a,e], Eric S. Lander[a,c,i,j,k], and Jill P. Mesirov[a,k]

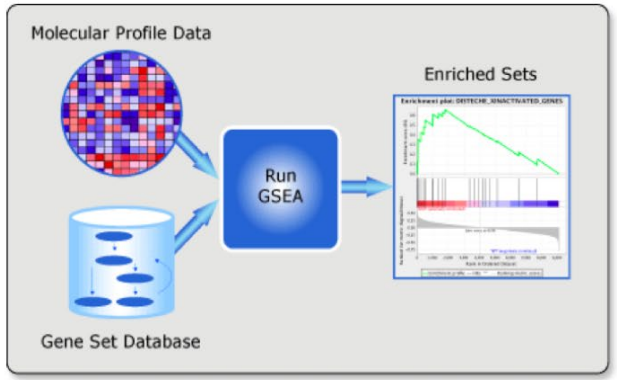
[a]Broad Institute of Massachusetts Institute of Technology and Harvard, 320 Charles Street, Cambridge, MA 02141; [c]Department of Systems Biology, Alpert 536, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02446; [d]Institute for Genome Sciences and Policy, Center for Interdisciplinary Engineering, Medicine, and Applied Sciences, Duke University, 101 Science Drive, Durham, NC 27708; [e]Department of Medical Oncology, Dana–Farber Cancer Institute, 44 Binney Street, Boston, MA 02115; [f]Division of Pulmonary and Critical Care Medicine, Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02114; [g]Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, C2-023, P.O. Box 19024, Seattle, WA 98109-1024; [h]Department of Neurology, Enders 260, Children's Hospital, Harvard Medical School, 300 Longwood Avenue, Boston, MA 02115; [i]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142; and [j]Whitehead Institute for Biomedical Research, Massachusetts Institute of Technology, Cambridge, MA 02142

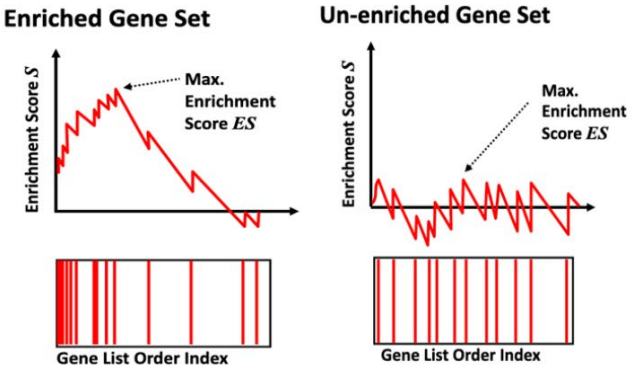
Contributed by Eric S. Lander, August 2, 2005



Gene Set Enrichment Analysis (GSEA) is a computational method that determines whether an a priori defined set of genes shows statistically significant, concordant differences between two biological states (e.g. phenotypes).



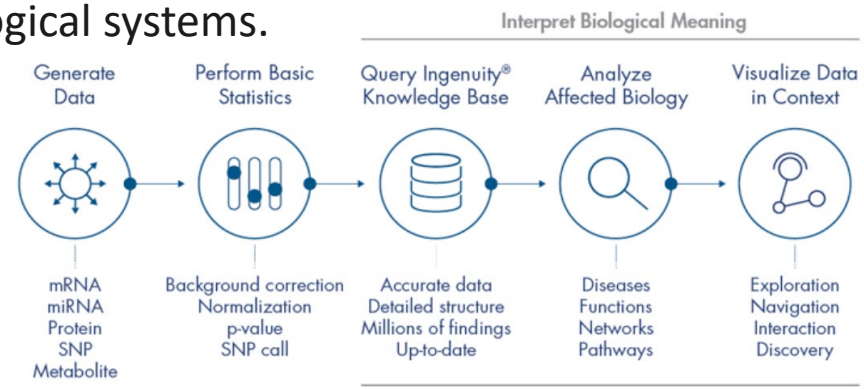


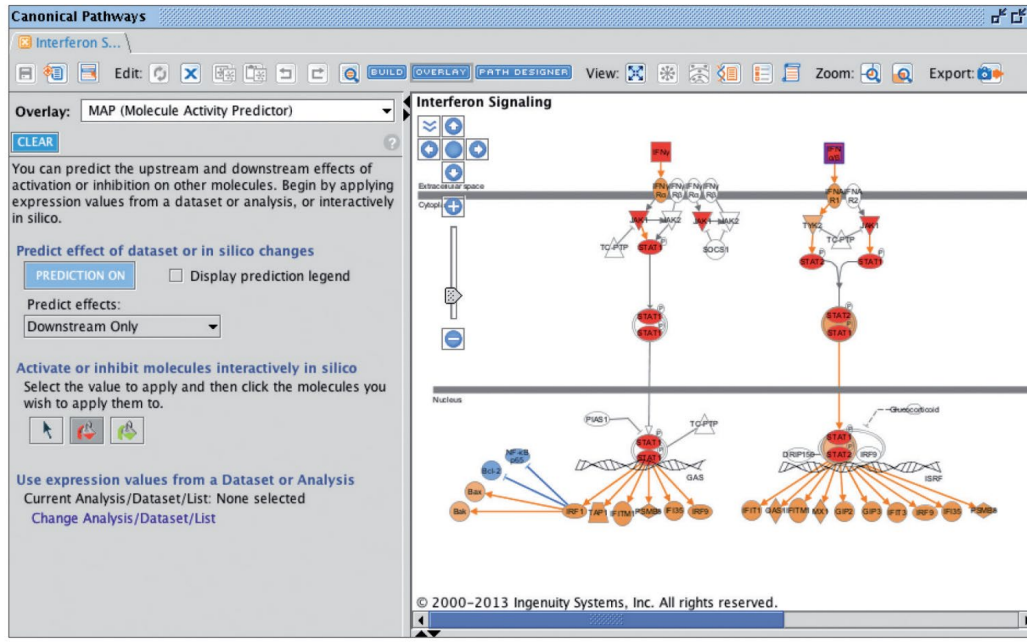


s://www.gsea-msigdb.org/gsea/index.jsp

The biomarker predicts AR activation and suppression

IPA is a web-based bioinformatics application that allows researchers to upload data analysis results from high-throughput experiments such as microarray and next generation sequencing for functional analyze, integration, and further understanding. This includes both microarray and RNA-Seq gene expression, miRNA, SNP, metabolomics, and proteomics data. In general, lists of genes or chemicals can be analyzed information on genes, proteins, chemicals, and drugs and allows intera biological systems.





Statistical analys

p-val
overl

z-sco

2. Measures the match between two analyses (Analysis Match)

- z-score ≥ 2 (match) and ≤ –2 (anti-match)

ACKNOWLEDGEMENTS

Environmental Protection Agency
John Rooney
Natalia Ryan
Brian Chorley
Thomas Hill

NIEHS
Nicole Kleinstreuer

Health Canada
Carole Yauk
Andrew Williams

University of Leiden
Bob van de Water
Steve Hiemstra

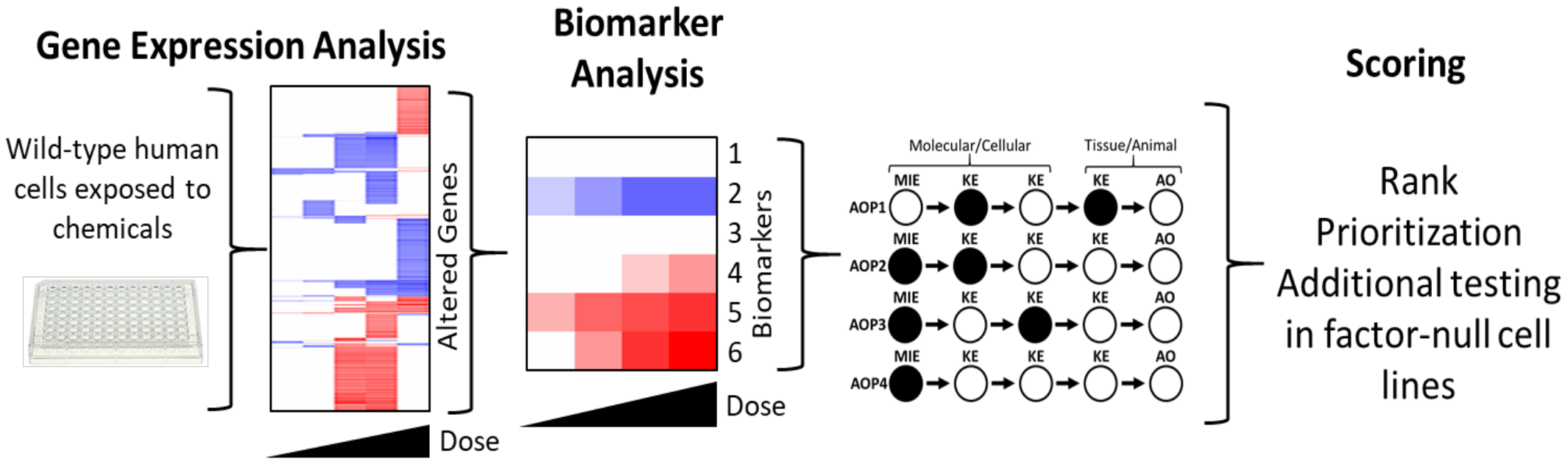
PamGene
Rinie van Beuningen
Rene Houtman

City of Hope Medical Center, Duarte
Shiuan Chen

Merck
Frank Sistare
Chunhua Qin

Support from EPA Chemical Safety for Sustainability Research Program

Putting biomarker predictions into networks of adverse outcome pathways



From Corton et al., submitted