



Modeling Chemicals' Adverse Effects by high-throughput transcriptomics

Stefano Monti

Division of Computational BioMedicine, BU School of Medicine
Department of Biostatistics, BU School of Public Health
Bioinformatics Program, Boston University

Exposures and Latent Disease Risk: Session IV - Moving Forward
Sponsored by: NIEHS Superfund Research Program



Boston University
Superfund Research Program



National Institute of
Environmental Health Sciences
Superfund Research Program



NIH LINCS
PROGRAM



EVANS CENTER
FOR INTERDISCIPLINARY
BIOMEDICAL RESEARCH



FIND
THE
CAUSE
Breast Cancer Foundation

Outline

Overview of the experimental and computational approaches we have developed and applied to model environmental chemicals and to predict their long-term adverse effects from short-term transcriptomics assays.

Vignettes from two studies

- ❖ **The Carcinogenome Project:** Predicting Long-Term Chemical Carcinogenicity and Genotoxicity from Short-Term Assays.
- ❖ **The Adipogenome Project:** Genomics Characterization of Adipocyte Dysregulation by Environmental and Therapeutic Perturbagens.

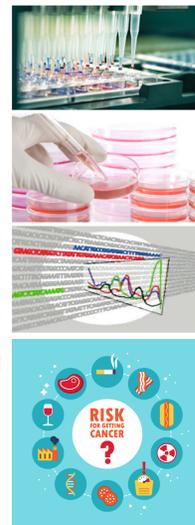
Challenges and lessons learned.

Modeling Chemical Carcinogenicity

Development of
analytical/experimental framework to
predict **long-term cancer risk**
from exposure to chemical compounds
using **genomic assays** and **computation**



Cancer Prevention



Modeling Chemical Adverse Effects

Development of
analytical/experimental framework to
predict **long-term disease risk**
from exposure to chemical compounds
using **genomic assays** and **computation**



Disease Prevention



Chemical Exposure Understudied



- Constant exposure to pesticides, industrial pollutants, consumer products and drugs ~ **85,000**
- Less than 2% of all chemical compounds have been systematically tested ~ **1,500**
- Mixtures of compounds challenging to evaluate ~ **10⁹-10¹²**

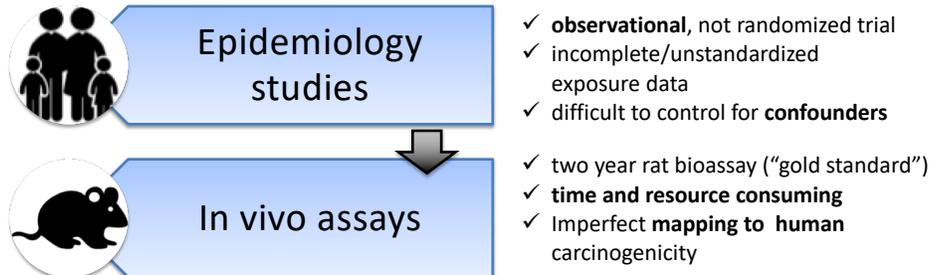
Carcinogenicity Testing approaches



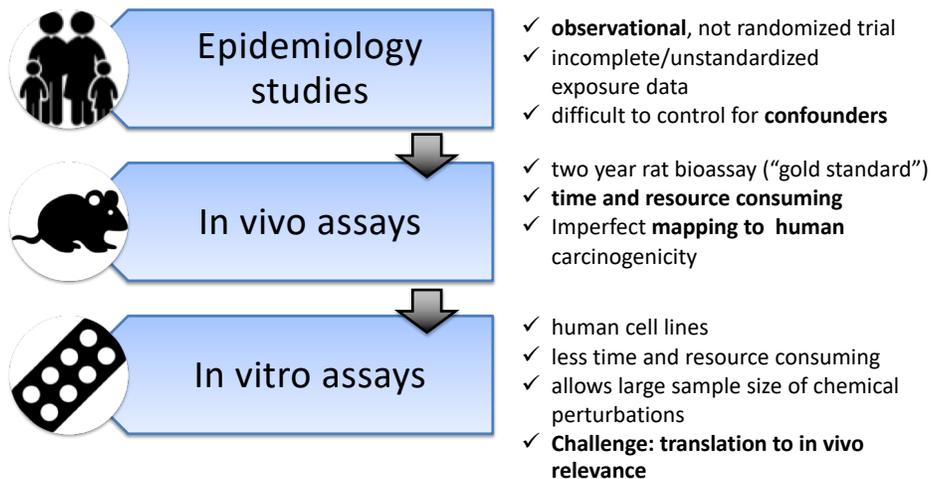
Epidemiology studies

- ✓ **observational**, not randomized trial
- ✓ incomplete/unstandardized exposure data
- ✓ difficult to control for **confounders**

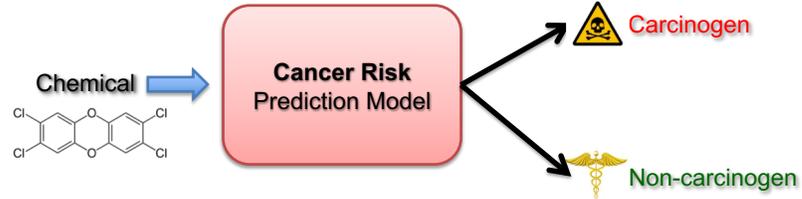
Carcinogenicity Testing approaches



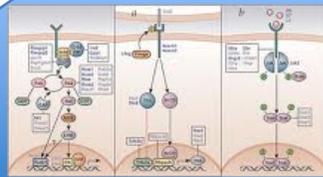
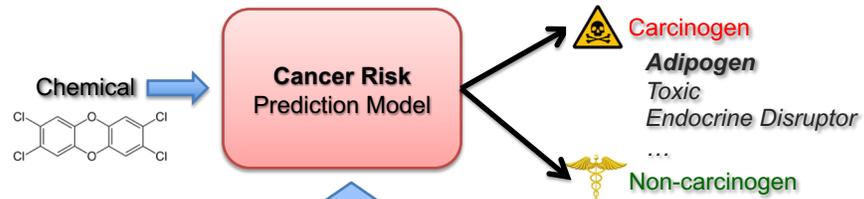
Carcinogenicity Testing approaches



The Quest for a Chemical Carcinogenicity "Crystal Ball"



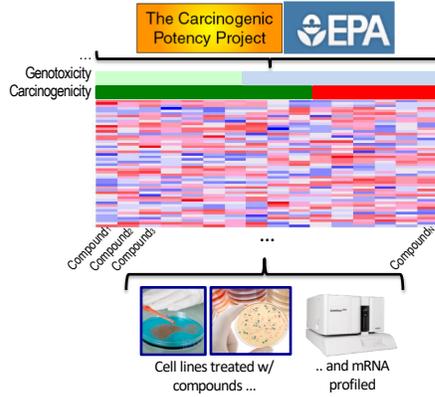
The Quest for a Chemical Carcinogenicity "Crystal Ball"



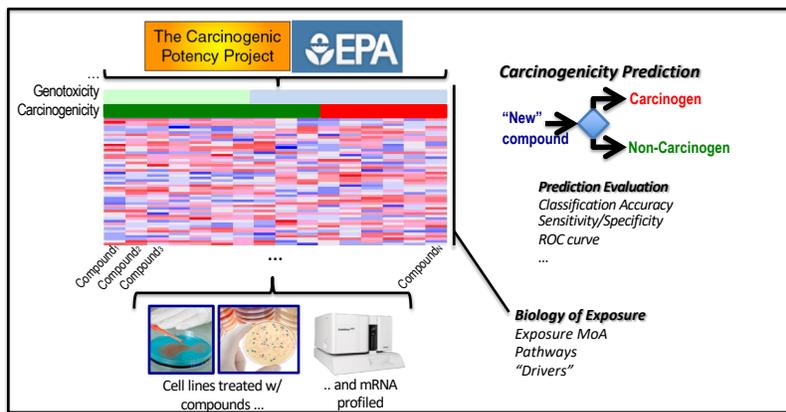
- ✓ Pathways affected
- ✓ Driving genetic alterations
- ✓ Biomarkers
- ✓ ...

Understand Why

Experimental Design Overview

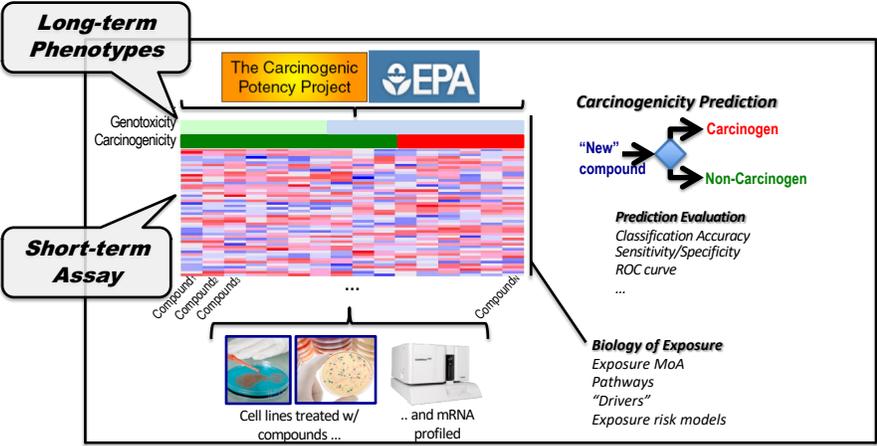


Experimental Design Overview



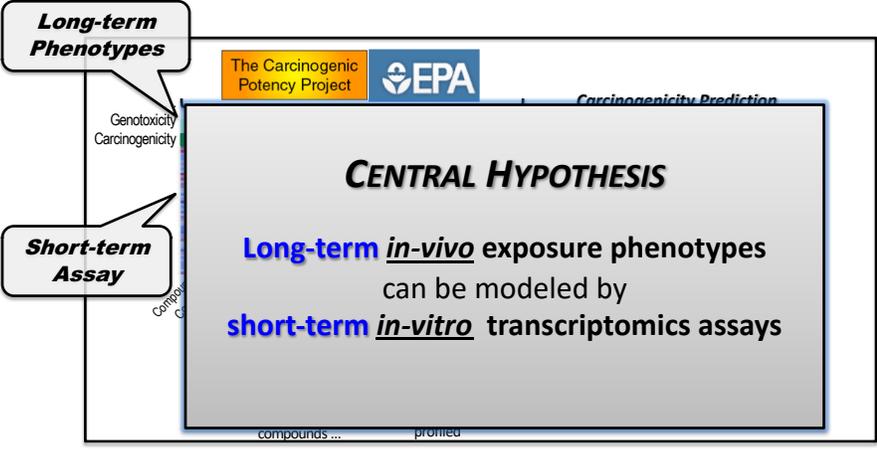
*Project relies on high-throughput, cost-effective gene expression assays
Luminex-1000 (L1000) @ Broad Institute
(or highly multiplexed RNA-sequencing)*

Experimental Design Overview



Project relies on high-throughput, cost-effective gene expression assays
Luminex-1000 (L1000) @ Broad Institute
(or highly multiplexed RNA-sequencing)

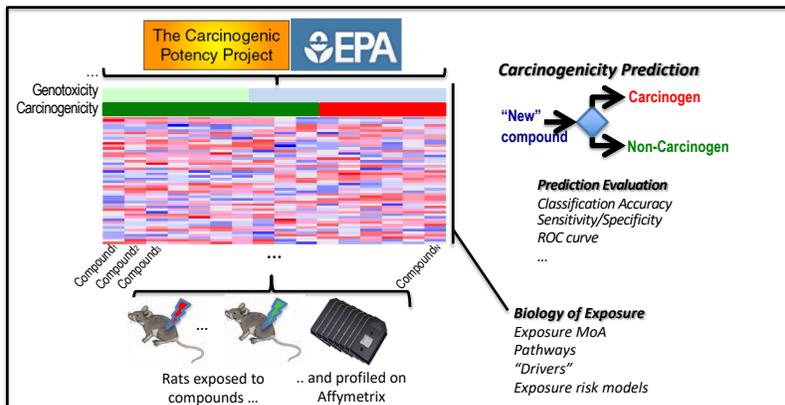
Experimental Design Overview



Project relies on high-throughput, cost-effective gene expression assays
Luminex-1000 (L1000) @ Broad Institute
(or highly multiplexed RNA-sequencing)

Can Carcinogenicity be Predicted from GEP?

the answer from short-term in-vivo (rat-based) assays



DrugMatrix TG-GATES



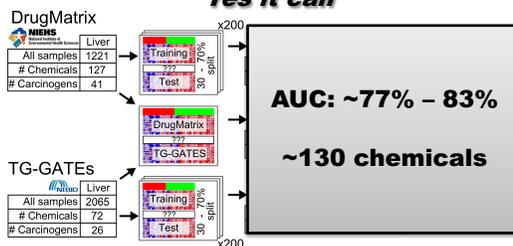
1000s of profiles, 100s of chemicals

Gusenleitner et al., 2014

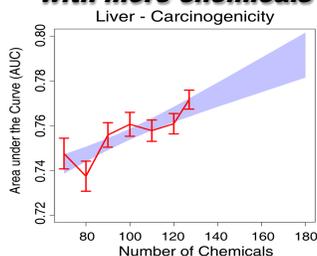
Can Carcinogenicity be predicted from GEP?

The DrugMatrix/TG-GATES answer

Yes it can



Prediction can be improved with more chemicals



Helps elucidate modes of action

MoA's

- DNA damage
- Oxidative Stress
- Altered metabolism
- Proteasome
- ...

Gusenleitner et al., 2014

Carcinogenicity Prediction is tissue-specific

Classification Results

Tissue Agnostic
AUC: 65%



Tissue Specific (Liver)
AUC: 77-83%



Gusenleitner et al., 2014

Home About Contact

<https://carcinogenome.org>

The Carcinogenome Project

Chemical Carcinogenicity Screening using high-throughput transcriptomics assays

Project	Cell line	Description
Liver Carcinogenicity	HEPG2	This experiment uses 330 selected chemicals for in-vivo liver carcinogenicity testing, including 128 liver carcinogens, 168 non-carcinogens, and 34 miscellaneous chemicals (e.g. nuclear receptor ligands). Chemical carcinogenicity and genotoxicity annotations are based on the Carcinogenicity Potency Database (CPDB), which is the result of tissue-specific long-term animal cancer tests in rodents. In the liver carcinogenome project, HepG2 (liver) cells are exposed to each individual chemical for 24 hours and their gene expression is profiled on the L1000 platform. Each chemical is assayed at 6 doses (2 fold dilutions starting from the highest concentration of 40uM or 20uM) with triplicate profiles generated for each dose...
Breast Carcinogenicity	MCF10A	This experiment uses 345 selected chemicals for breast carcinogenicity testing, including 120 breast carcinogens, 114 non-carcinogens, and 68 miscellaneous chemicals (e.g. nuclear receptor ligands, BU SRP chemicals, lung carcinogens). Chemical carcinogenicity and genotoxicity annotations are based on the Carcinogenicity Potency Database (CPDB), which is the result of tissue-specific long-term animal cancer tests in rodents, or breast carcinogens published from Rudel et. al., 2007. In the CRCGN project, MCF10A (breast epithelial) cells are exposed to each individual chemical for 24 hours and their gene expression is profiled on the L1000 platform. Each chemical is assayed at 3 doses (3 fold dilutions starting from the highest concentration of 100uM, with the exception of selected BUSRP chemicals...

The Carcinogenome Project: Developed by Monti Lab at Boston University
2017

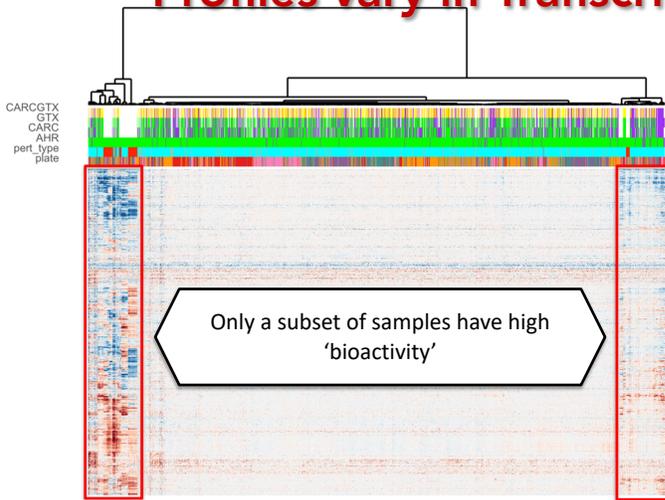
The Carcinome Project

In-vitro Carcinogenicity Profiling

Profiled >330 chemicals (~6,000 profiles) in liver cell lines with "liver-specific" carcinogenicity annotation

	Cell line	Chemical Type	# Chemicals	# Profiles
24h 6 doses 3 replicates 1 cell type	HEGP2	Liver carcinogens	131	2358
		Non-carcinogens	172	3096
		Others (BUSRP)	33	594
		Total	336	6048
24h 3 doses 3 replicates 2 cell type	MCF-10A, MCF-10A P53-	Breast carcinogens	120	2160
		Non-carcinogens	114	2052
		Others (BUSRP)	68	1224
	Total	302	5436	
	MCF10A & HEPG2	breast carcinogens + others	115	2070

Profiles vary in Transcriptional Bioactivity

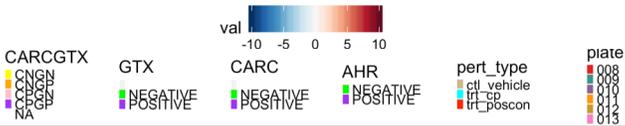


Transcriptional Activity Score (TAS)

summary of signature strength (SS_{ngene}) and replicate correlation (CC_{q76})



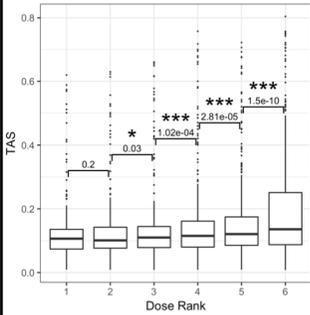
adapted from Lev Litichevskiy @ Broad



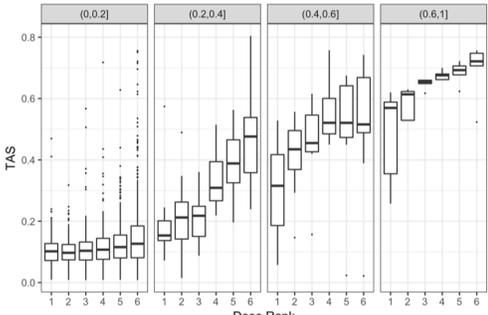
Bioactivity vs. Carcinogenicity

no significant association

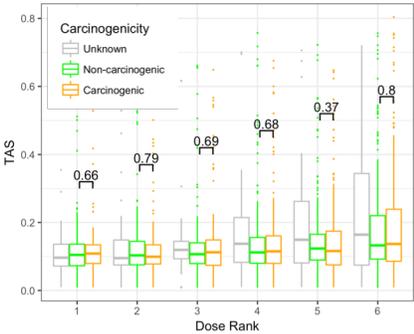
Bioactivity (TAS) increases with Dose



Dose response steeper with higher Bioactivity (TAS)



Carcinogenicity not associated with Bioactivity (TAS)



Difference between Acute vs. Chronic Response

short-term chemical perturbation with minimal transcriptional response cannot be assumed "safe"

Carcinogenicity Prediction

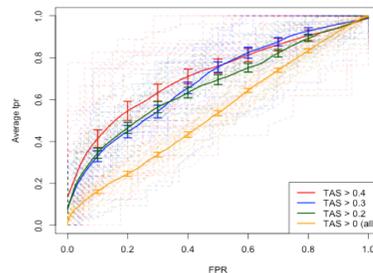
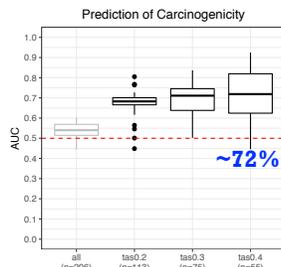
accuracy improves with higher bioactivity

Carcinogenicity

Carcinogenicity Prediction

accuracy improves with higher bioactivity

Carcinogenicity AUC			
data	median	mean	se
all	54	53.9	0.9
tas > 0.2	68.3	66.9	1.6
tas > 0.3	71.1	69.1	1.8
tas > 0.4	71.9	72.2	2.7

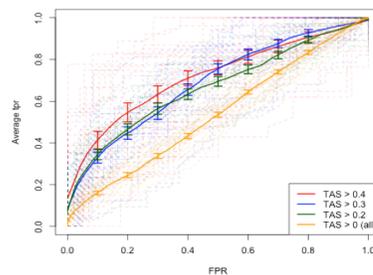
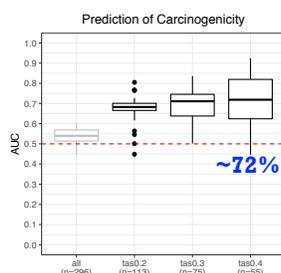


Carcinogenicity

Carcinogenicity Prediction

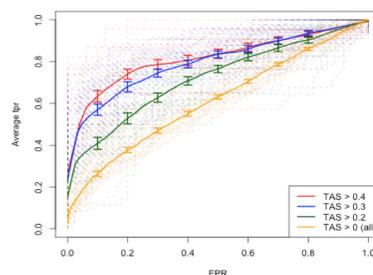
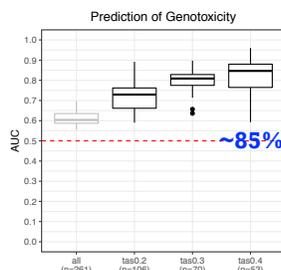
accuracy improves with higher bioactivity

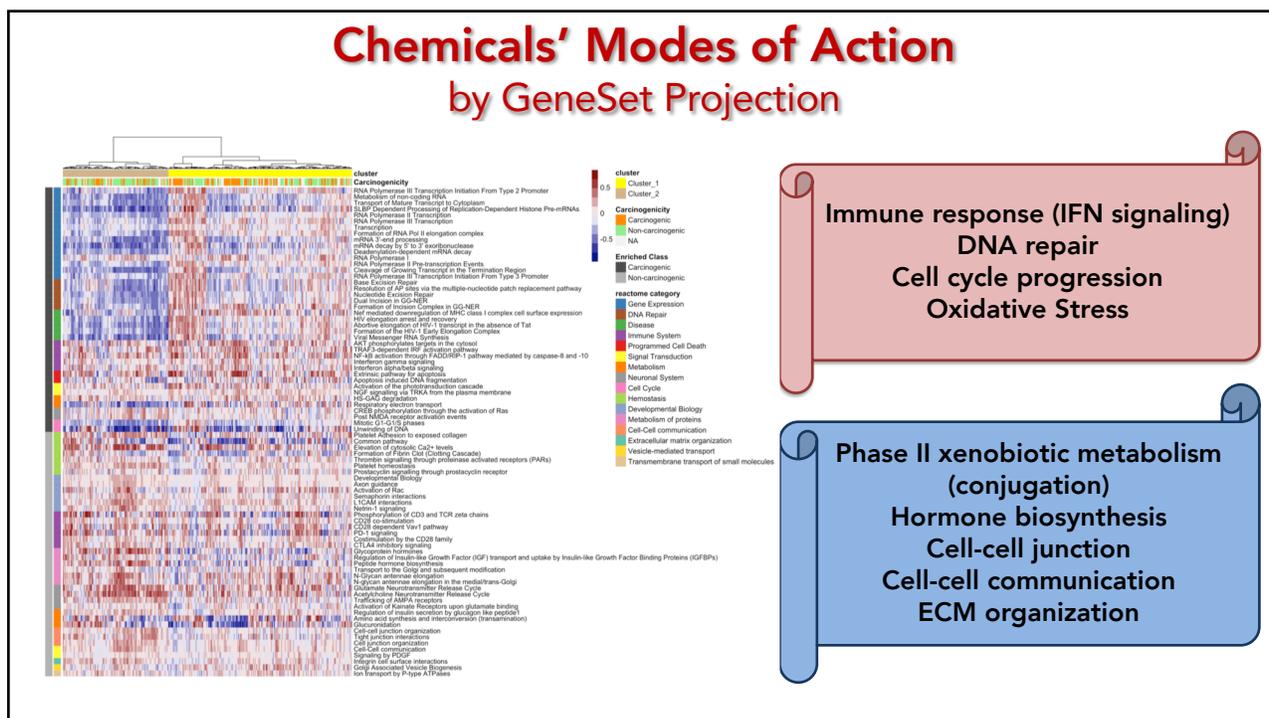
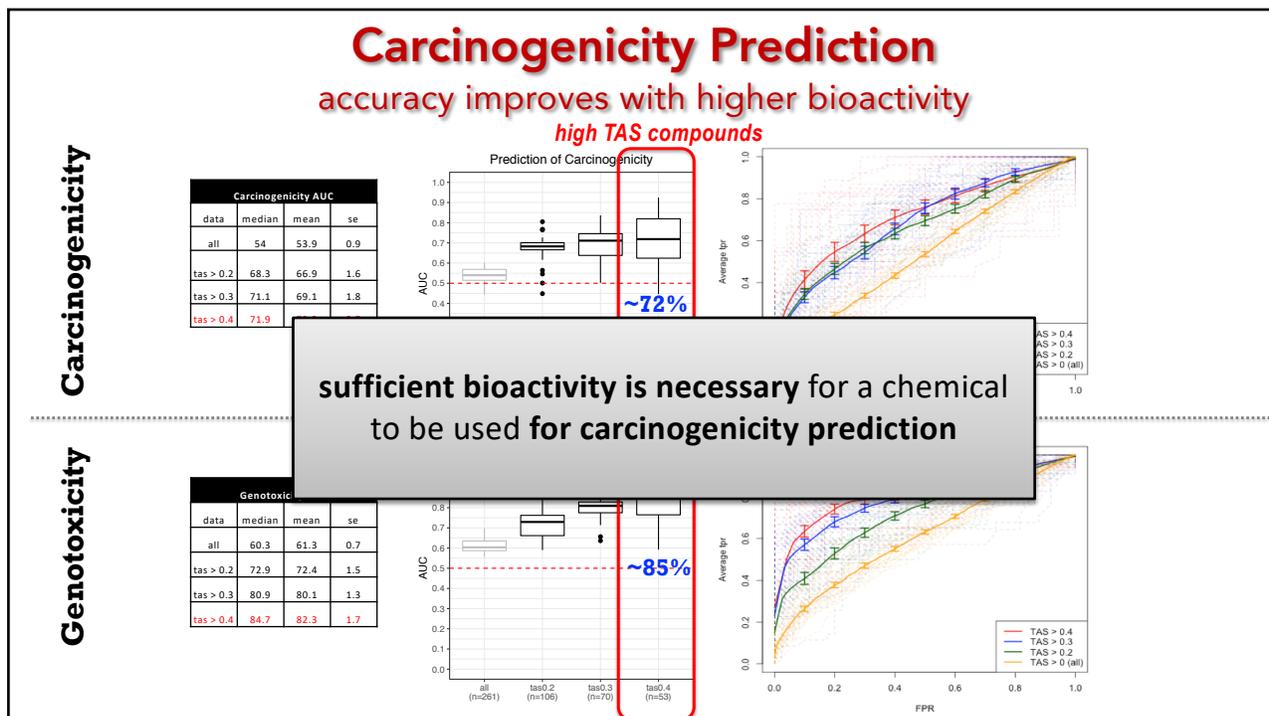
Carcinogenicity AUC			
data	median	mean	se
all	54	53.9	0.9
tas > 0.2	68.3	66.9	1.6
tas > 0.3	71.1	69.1	1.8
tas > 0.4	71.9	72.2	2.7



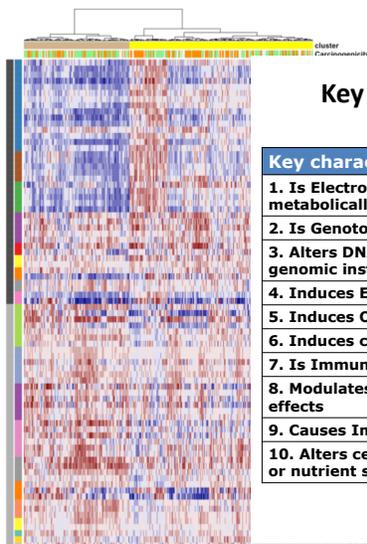
Genotoxicity

Genotoxicity AUC			
data	median	mean	se
all	60.3	61.3	0.7
tas > 0.2	72.9	72.4	1.5
tas > 0.3	80.9	80.1	1.3
tas > 0.4	84.7	82.3	1.7





Chemicals' Modes of Action by GeneSet Projection



Key Characteristics of Human Carcinogens

Key characteristic:

1. Is Electrophilic or can be metabolically activated ✓
2. Is Genotoxic ✓
3. Alters DNA repair or causes genomic instability ✓
4. Induces Epigenetic Alterations ✓
5. Induces Oxidative Stress ✓
6. Induces chronic inflammation ✓
7. Is Immunosuppressive ✓
8. Modulates receptor-mediated effects ✓
9. Causes Immortalization ✓
10. Alters cell proliferation, cell death, or nutrient supply ✓

Evidence that these characteristics are observed, especially in humans or as intermediate biomarkers in human specimens can provide biological plausibility for epidemiological findings and/or early warning if no epidemiology exists

Smith MT, Guyton KZ, Gibbons CF, Fritz JM, Portier CJ, Rusyn I, DeMarini DM, Caldwell JC, Kavlock RJ, Lambert PF, Hecht SS, Bucher JR, Stewart BW, Baan RA, Cogliano VJ and K Straif. *Env Health Persp.*, 124(6), 713, 2016.

adapted from M. Smith

response (IFN signaling)
DNA repair
cycle progression
Oxidative Stress

xenobiotic metabolism
(conjugation)
hormone biosynthesis
cell-cell junction
cell communication
ECM organization

Home About Contact

The Carcinogenome Project

Chemical Carcinogenicity Screening using high-throughput transcriptomics assays

<https://carcinogenome.org>

Project	Cell line	Description
Liver Carcinogenicity	HEPG2	This experiment uses 330 selected chemicals for in-vivo liver carcinogenicity testing, including 128 liver carcinogens, 168 non-carcinogens, and 34 miscellaneous chemicals (e.g. nuclear receptor ligands). Chemical carcinogenicity and genotoxicity annotations are based on the Carcinogenicity Potency Database (CPDB), which is the result of tissue-specific long-term animal cancer tests in rodents. In the liver carcinogenome project, HepG2 (liver) cells are exposed to each individual chemical for 24 hours and their gene expression is profiled on the L1000 platform. Each chemical is assayed at 6 doses (2 fold dilutions starting from the highest concentration of 40uM or 20uM) with triplicate profiles generated for each dose...
Breast Carcinogenicity	MCF10A	This experiment uses 345 selected chemicals for breast carcinogenicity testing, including 120 breast carcinogens, 114 non-carcinogens, and 68 miscellaneous chemicals (e.g. nuclear receptor ligands, BU SRP chemicals, lung carcinogens). Chemical carcinogenicity and genotoxicity annotations are based on the Carcinogenicity Potency Database (CPDB), which is the result of tissue-specific long-term animal cancer tests in rodents, or breast carcinogens published from Rudel et al., 2007. In the CRCGN project, MCF10A (breast epithelial) cells are exposed to each individual chemical for 24 hours and their gene expression is profiled on the L1000 platform. Each chemical is assayed at 3 doses (3 fold dilutions starting from the highest concentration of 100uM, with the exception of selected BUSRP chemicals...

The Carcinogenome Project: Developed by Monti Lab at Boston University
2017

Gusenleitner et al., *PLoS One* 2014

Mulas et al., *BMC Bioinformatics* 2017

Li et al., *Environmental Health Perspective* 2019

Outline

Overview of the experimental and computational approaches we have developed and applied to model environmental chemicals and to predict their long-term adverse effects from short-term transcriptomics assays.

Vignettes from two studies

- ❖ **The Carcinogenome Project:** Predicting Long-Term Chemical Carcinogenicity and Genotoxicity from Short-Term Assays.
- ❖ **The Adipogenome Project:** Genomics Characterization of Adipocyte Dysregulation by Environmental and Therapeutic Perturbagens.

Challenges and lessons learned.

Adipocytes Function

...

Maintaining Energy Homeostasis

“Obesity is a disorder of the energy homeostasis system, rather than just a passive accumulation of adipose, and **that environmental factors, including chemicals, confer obesity risk.**”

– Endocrine Society’s latest scientific statement

 Mitochondria  Nucleus  Lipid droplet

	Brown	White	Brite/beige
Primary Function	Thermogenesis Endocrine	Energy storage Endocrine	Thermogenesis? Endocrine?

adapted from [Guertin Lab @UMass](#)

Adipocytes Function

How is it affected by Exogenous Compounds?

Chemical Exposure → *Disruption of Metabolic Balance?*

	Brown	White	Brite/beige
Primary Function	Thermogenesis Endocrine	Energy storage Endocrine	Thermogenesis? Endocrine?

adapted from [Guertin Lab @UMass](#)

Adipogens

PPAR γ Activity Modifying Compounds

PPAR γ

Ligand Binding

Post-translational Modification

Int J Mol Sci. 2018 Jun; 19(6): 1738.
Published online 2018 Jun 12. doi: 10.3390/ijms19061738

The Adipogenome Project

Adipogens

Exogenous compounds that directly alter white adipocyte function via **modification of PPAR γ activity**

Project Goals

1. Create a **Classifier** to identify novel candidate adipogens
2. Create a **Taxonomy** to group chemicals based on their effects on PPAR γ 's transcriptome and downstream metabolic functions

Kim et al., *Arch Tox* 2018

Kim, Reed, et al., [biorXiv 519629](https://doi.org/10.1007/s00204-019-0210-0) (under 2nd review at EHP)

Experimental Design

PPAR γ Activity Modification



Manual Curation

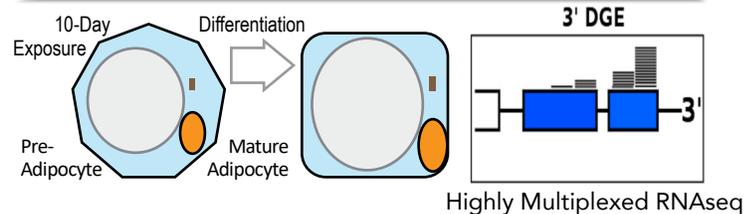
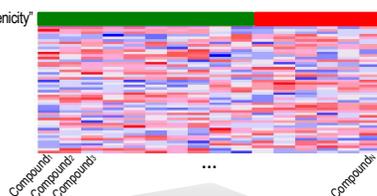
Cell Line

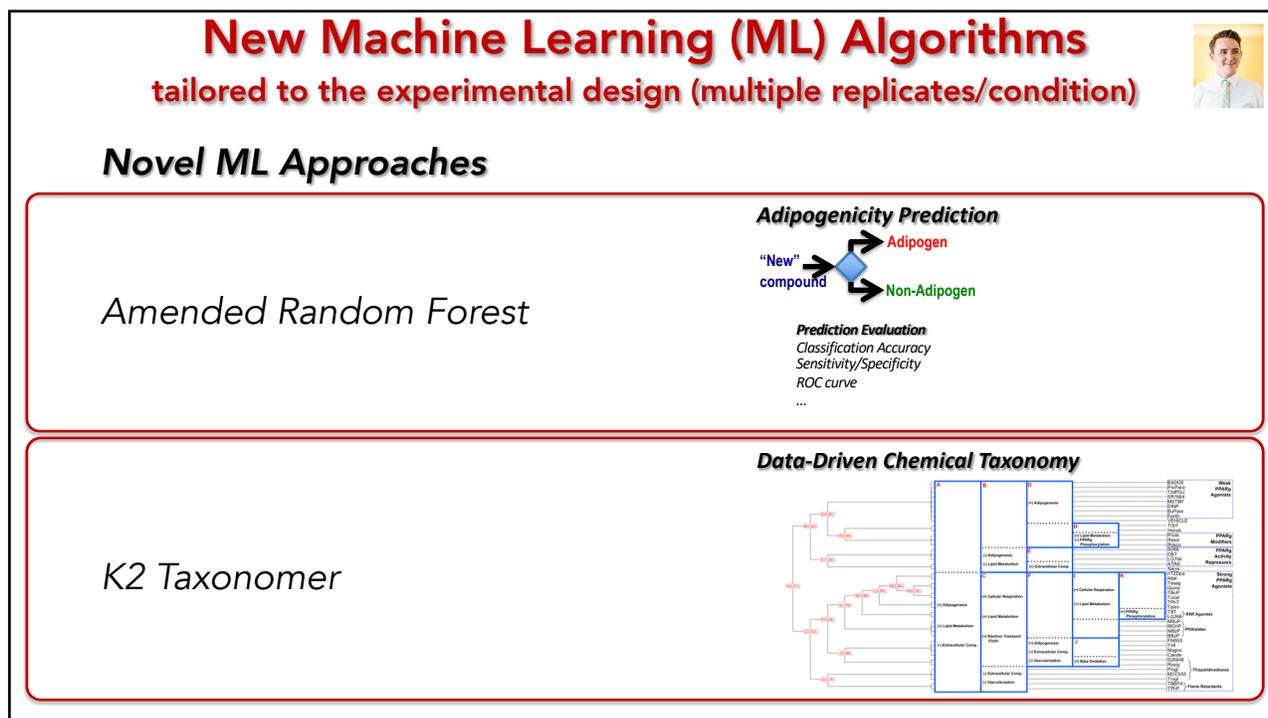
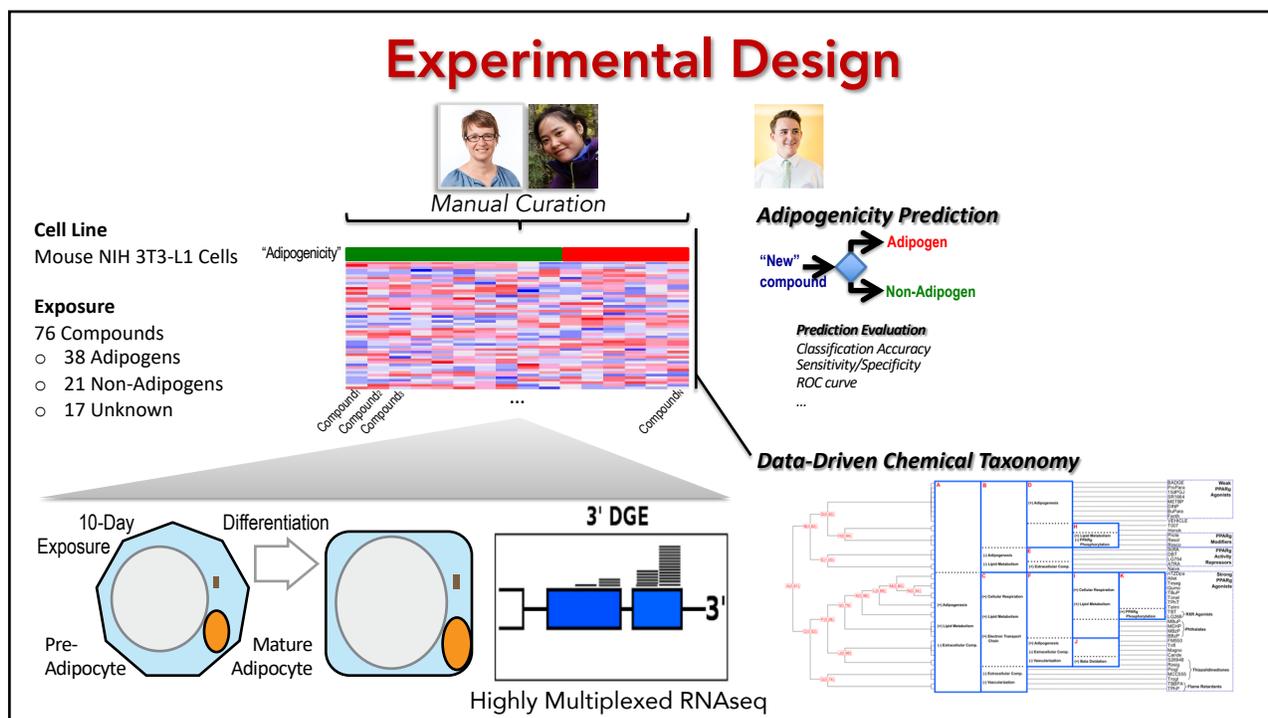
Mouse NIH 3T3-L1 Cells

"Adipogenicity"

Exposure

- 76 Compounds
 - 38 Adipogens
 - 21 Non-Adipogens
 - 17 Unknown





New Machine Learning (ML) Algorithms tailored to the experimental design

Novel ML Approaches

Amended Random Forest

Adipogenicity Prediction



Prediction Evaluation

Classification Accuracy

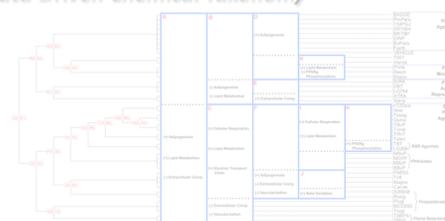
Sensitivity/Specificity

ROC curve

...

K2 Taxonomer

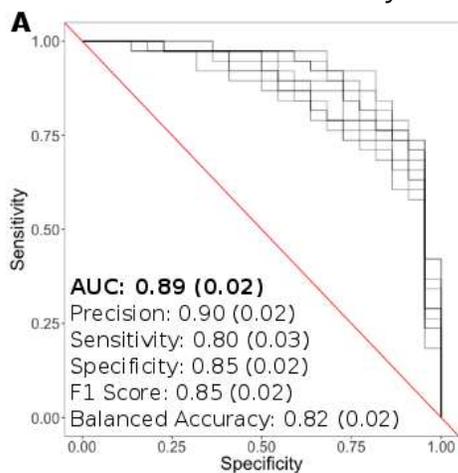
Data-Driven Chemical Taxonomy



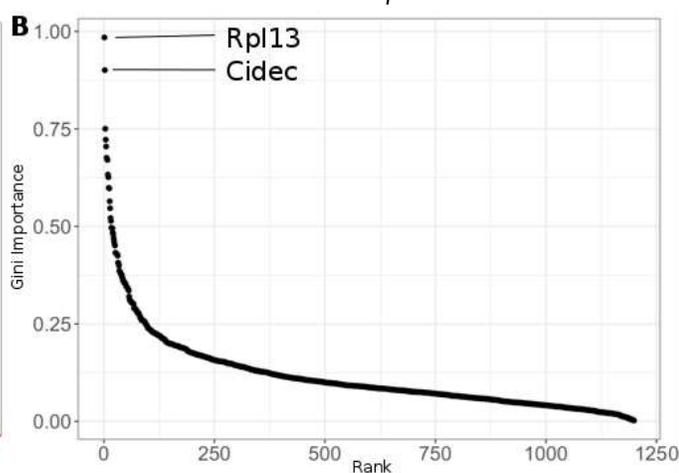
PPAR γ Activity Modifier Classification Results

by Amended Random Forest

Predictive Accuracy



Gene "Importance"



PPAR γ Activity Modifier Classification Results of 17 unknown compounds

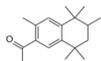
Final Prediction Results

Chemical Name	Abbreviated Name	Known Source/Use	PPAR γ Modifier Voting
d-cis,trans-Allethrin	Allethrin	Insecticide	0.91
→ Tonalid	Tonalid	Musk (fragrance)	0.90
→ Quinoxifen	Quinoxifen	Fungicide	0.90
Fenthion	Fenthion	Insecticide	0.88
2,4,6-Tris(tert-butyl)phenol	TTBP	Antioxidant (industrial)	0.80
Prallethrin	Prallethrin	Insecticide	0.78
Tebuconazole	Tebucon	Fungicide	0.78
Fludioxonil	Fludiox	Fungicide	0.77
Tris(1,3-dichloro-2-propyl) phosphate	TDCPP	Flame retardant	0.76
Cyazofamid	Cyazofamid	Pesticide	0.72
Perfluorooctanoic acid	PFOA	Fluorosurfactant	0.59
Triphenyl phosphite	Triphen_Phosphite	Pesticide	0.57
Tris(1-chloro-2-propyl) phosphate	TCCP	Flame retardant	0.54
Triphenylphosphine oxide	Triphen_Phox_Ox	Crystallizing aid, byproduct	0.49
Diphenyl phosphate	DPP	Metabolite of TPhP	0.47
Dioctyl sulfosuccinate sodium	DOSS	Surfactant	0.41
Perfluorooctanesulfonic acid	PFOS	Fluorosurfactant	0.40

High Confidence Adipogens

PPAR γ Activity Modifier Classification Results novel adipogens that favor white adipogenesis

Tonalid (Fragrance)



- (Reiner and Kannan, 2006)
 - 48% of perfumes
 - 29% of body lotions/creams,
 - 75% of deodorants
 - 14% of shower gel/shaving creams
 - 33% of hair products
 - 31% of sanitation products

Quinoxifen (Fungicide)



- FortressTM, OrkaTM, LegendTM, QuintecTM
 - Grain, Hops, Grapes
 - Low Residue
 - Bioaccumulates in Fish

Functional analyses confirmed that Quino and Tonalid **induce white, but not brite, adipogenesis** in both mouse and human preadipocyte models

The Adipogenome Project

Adipogens

Exogenous compounds that directly alter white adipocyte function via modification of PPAR γ activity

Project Goals

1. Create a **Classifier** to identify novel candidate adipogens
2. Create a **Taxonomy** to group chemicals based on their effects on PPAR γ 's transcriptome and downstream metabolic functions

Kim et al., *Arch Tox* 2018

Kim, Reed, et al., [biorXiv 519629](https://doi.org/10.1186/s12929-018-0462-9) (under 2nd review at EHP)

New Machine Learning (ML) Algorithms tailored to the experimental design

Novel ML Approaches

Amended Random Forest

Adipogenicity Prediction

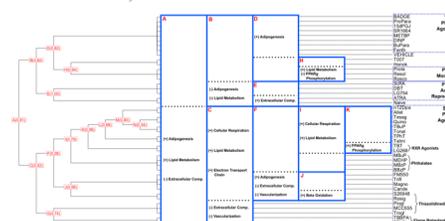


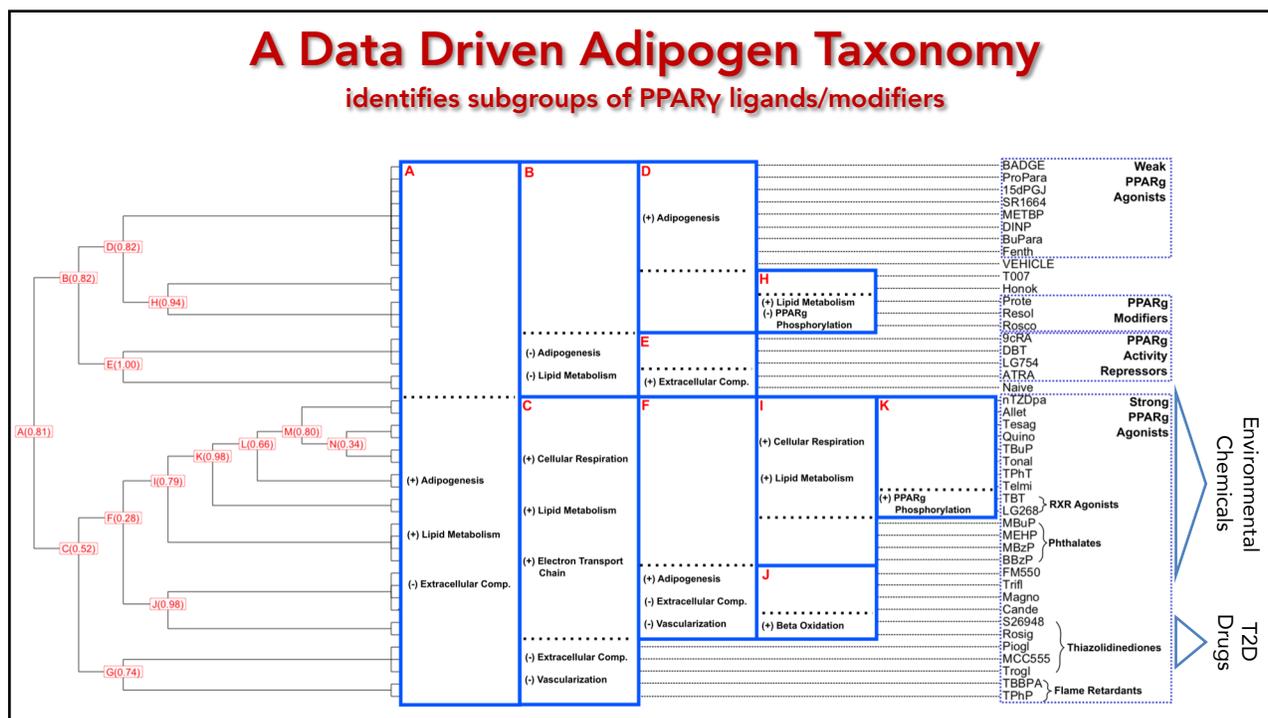
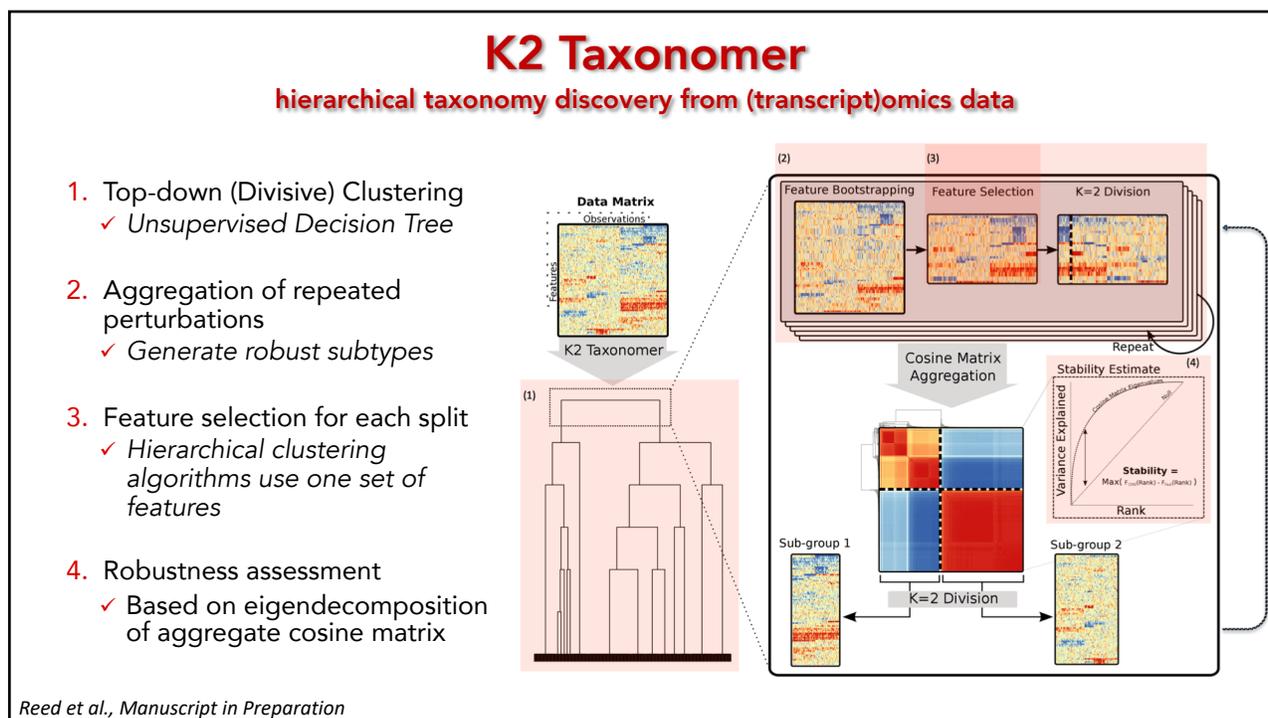
Prediction Evaluation

Classification Accuracy
Sensitivity/Specificity
ROC curve
...

K2 Taxonomer

Data-Driven Taxonomy





In-Silico Validation of (mouse-based) Taxonomy confirms its human relevance

Signature Projection
onto
METSIM

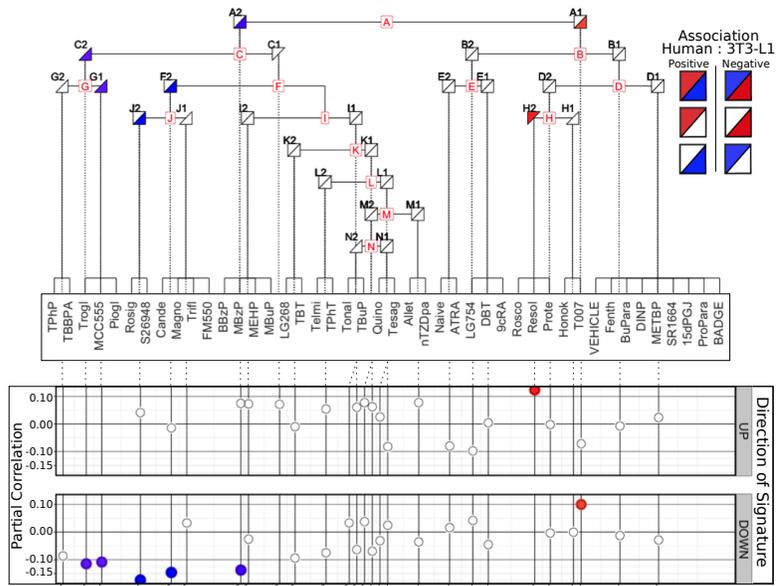
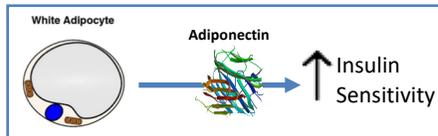
subcutaneous adipose tissues
770 subjects
12 clinical measurements

In-Silico Validation of Human Relevance in-vitro signatures significantly associated with clinical endpoints in primary tissues

Signature Projection
onto
METSIM

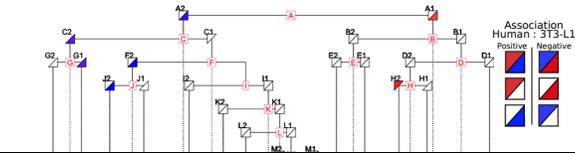
subcutaneous adipose tissues
770 subjects
12 clinical measurements

association with
Plasma Adiponectin

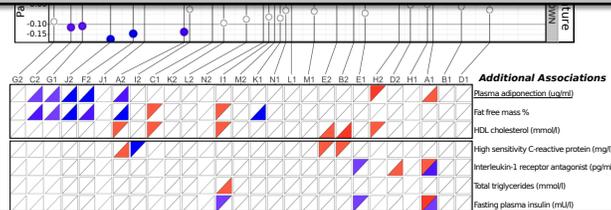


In-Silico Validation of Human Relevance

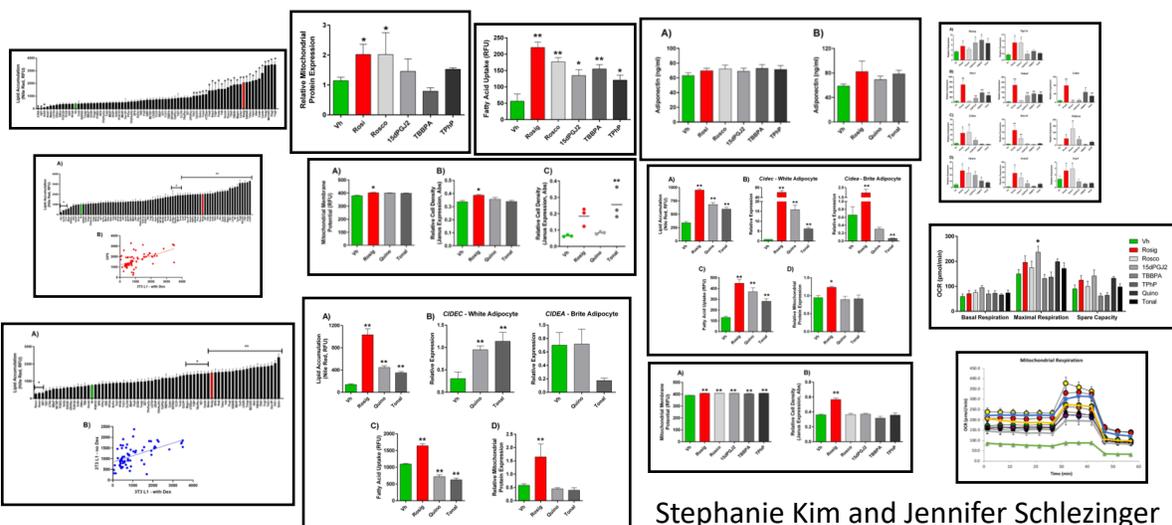
in-vitro signatures significantly associated with clinical endpoints in primary tissues



Mouse-based, in vitro-derived signatures capture salient functional aspects of healthy and unhealthy metabolic functions in human subjects

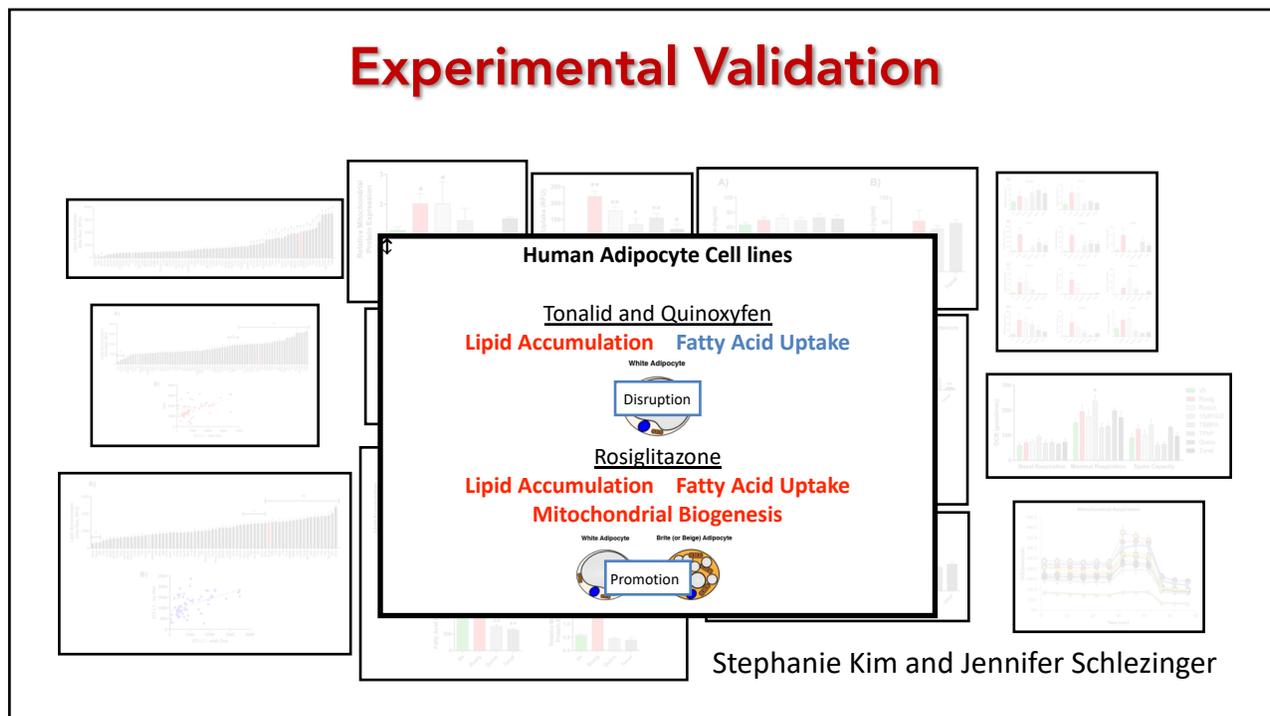


Experimental Validation



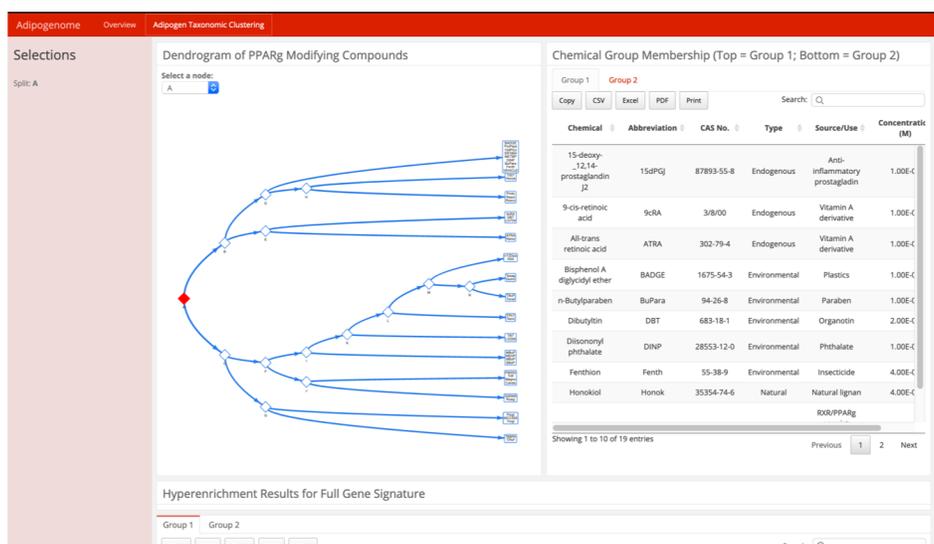
Stephanie Kim and Jennifer Schlezinger

Experimental Validation



The Adipogenome Portal

<https://montilab.bu.edu/adipogenome>



Summary

- ❖ Identified (new) PPAR γ agonists/modifiers
- ❖ Sorted agonists into likely white or brite adipogens
- ❖ Developed new classification and taxonomy discovery methods
- ❖ Computational & Experimental framework with general applicability to the classification of as-yet uncharacterized chemicals

Kim et al., *Arch Tox* 2018

Kim, Reed, et al., [biorXiv 519629](https://doi.org/10.1186/s12929-018-0450-0) (under review at EHP)

Bringing it altogether: The Xposome Portal

<https://montilab.bu.edu/Xposome>

Carcinome

(under construction)

Adipome

The Xposome Project

Chemical Screening using high-throughput transcriptomics assays

[Home](#) [About](#) [Contact](#) [Sign In](#)

Project	Cell line	Description
Liver Carcinogenicity	HEPG2	This experiment uses 330 selected chemicals for in-vivo liver carcinogenicity testing, including 128 liver carcinogens, 168 non-carcinogens, and 34 miscellaneous chemicals (e.g., nuclear receptor ligands). Chemical carcinogenicity and genotoxicity annotations are mainly based on the <i>Carcinogenicity Potency Database (CPDB)</i> , which is the result of tissue-specific long-term carcinogenicity tests in rodents. In this study, HepG2 (liver) cells were exposed to each individual chemical for 24 hours and their gene expression profiled on the L1000 platform. Each chemical was assayed at 6 doses (2-fold dilutions starting from the highest concentration of 40 μ M or 20 μ M) with triplicate profiles generated for each dose...
Breast Carcinogenicity	MCF10A	This experiment uses 345 selected chemicals for breast carcinogenicity testing, including 120 breast carcinogens, 114 non-carcinogens, and 68 miscellaneous chemicals (e.g., nuclear receptor ligands, <i>BUSRP</i> chemicals, lung carcinogens). Chemical carcinogenicity and genotoxicity annotations are based on the <i>Carcinogenicity Potency Database (CPDB)</i> , which is the result of tissue-specific long-term carcinogenicity tests in rodents, or breast carcinogens from [Rudel et al., 2007]. In this study, MCF10A (breast epithelia) cells were exposed to each individual chemical for 24 hours and their gene expression profiled on the L1000 platform. Each chemical was assayed at 3 doses (3-fold dilutions starting from the highest concentration of 100 μ M, with the exception of selected <i>BUSRP</i> chemicals)...
Adipogenicity	3T3-L1	This experiment uses 3T3-L1 cells differentiated in the presence of 77 chemicals, comprising 38 known PPAR γ ligands or modifying compounds, 22 negative controls, and 17 suspected PPAR γ ligands/modifiers. 3' digital gene expression profiling was performed on mouse NIH 3T3-L1 pre-adipocytes exposed to each of the chemicals over a 10-day period during differentiation...



...



Lessons Learned and Challenges

- ❖ "Logistics" (chemical procurement, profile generation, etc.)
- ❖ Dose for hazard determination: high enough to elicit "bioactivity"
- ❖ Acute vs. Chronic exposure
 - ✓ Low transcriptional response cannot be assumed "safe"
- ❖ Models more adequate for hazard prediction than MoA's
 - ✓ However, rich MoA information can still be parsed from data
- ❖ Wealth of results not adequately sharable through publications
 - ✓ Interactive online Portals a necessary complement
- ❖ Difficulty in funding these efforts
 - ✓ "hypothesis testing" bias
 - ✓ misplaced request for "in vivo validation" of results

Acknowledgements



Monti Lab

Amy Li
Eric Reed
Anthony Federico
Reina Chau
Adam Gower
Adam Labadorf



BU SPH

David Sherr
Jennifer Schlezinger
Stephanie Kim
Elizabeth Stanford

Broad Institute

Aravind Subramanian
Xiaodong Lu
Ted Natoli
Todd Golub
Josh Bittker

NTP/NIEHS

Scott Auerbach
Rick Paules
Ray Tice

BU Superfund

Mark Hahn
Jed Goldstone
John Stegeman
et al.

Tufts Medical

Paola Sebastiani
Gail Sonenshein
Charlotte Kuperwasser



Boston University
Superfund Research Program

Acknowledgements



NIH National Institute of
Environmental Health Sciences
Superfund Research Program



Boston University
Superfund Research Program

THE END