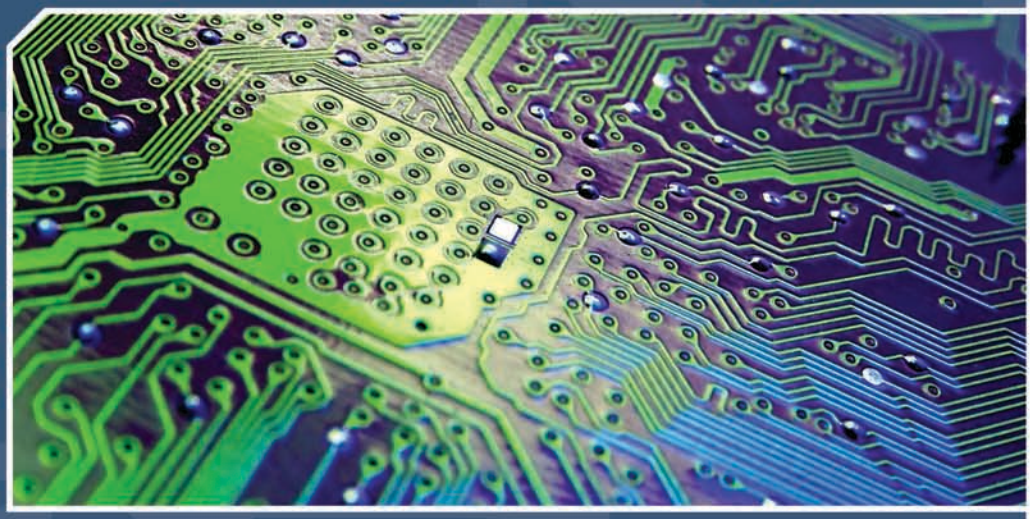
EPA
United States
Environmental Protection
Agency

Proceedings of the Computational Toxicology Centers Science To Achieve Results (STAR) Progress Review Workshop



OCTOBER 1, 2009

U.S. EPA, MAIN CAMPUS, BUILDING C
109 TW ALEXANDER DRIVE
RESEARCH TRIANGLE PARK, NC

Office of Research and Development
National Center for Environmental Research

Table of Contents

U.S. Environmental Protection Agency (EPA), Office of Research and Development (ORD),
National Center for Environmental Research (NCER)

Computational Toxicology Centers STAR Progress Review Workshop

U.S. Environmental Protection Agency
Main Campus, Building C, Auditorium C111A/B
109 TW Alexander Drive
Research Triangle Park, NC 27711

Thursday, October 1, 2009

Agenda

8:00 a.m. – 8:30 a.m.	**Registration**
8:30 a.m. – 9:00 a.m.	**Welcome, Introduction, and Review of Meeting Goals** *Robert Kavlock, EPA, ORD, and Deborah Segal, EPA, ORD, NCER*
9:00 a.m. – 10:00 a.m.	**Carolina Center for Computational Toxicology** *Ivan Rusyn, University of North Carolina*
10:00 a.m. − 10:15 a.m.	**Collaborative Work With EPA** *Ann Richard, EPA, National Center for Computational Toxicology (NCCT)*
10:15 a.m. – 10:30 a.m.	**Break**
10:30 a.m. – 11:30 a.m.	**New Jersey Environmental Bioinformatics and Computational Toxicology Center** *William Welsh, University of Medicine and Dentistry of New Jersey*
11:30 a.m. − 11:45 a.m.	**Collaborative Work With EPA** *Susan Euling, EPA, National Center for Environmental Assessment (NCEA)*
11:45 a.m. – 12:30 p.m.	**Lunch** (On Your Own)
12:30 p.m. – 1:30 p.m.	**Carolina Environmental Bioinformatics Research Center** *Fred Wright, University of North Carolina*
1:30 p.m. − 1:45 p.m.	**Collaborative Work With EPA** *Richard Judson, EPA, NCCT*
1:45 p.m. – 2:45 p.m.	**The Texas-Indiana Virtual STAR Center: Data-Generating *In Vitro* and *In Silico* Models of Developmental Toxicity in Embryonic Stem Cells and Zebrafish** *Maria Bondesson Bolin, University of Houston*
2:45 p.m. − 3:00 p.m.	**Collaborative Work With EPA** *Thomas Knudsen, EPA, NCCT*
3:00 p.m. – 3:30 p.m.	**A Proposal from the European Commission's Complementary Research Program** *Bart van der Burg, BioDetection Systems B.V.*
3:30 p.m. – 4:15 p.m.	**Discussion on Research Needs** *Chair: Maggie Breville, EPA, ORD*
4:15 p.m.	**Adjournment**

Carolina Center for Computational Toxicology

EPA Grant Number: R833825

Investigators:

1. Ivan Rusyn	E-mail: iir@unc.edu
2. Timothy Elston	E-mail: telston@amath.unc.edu
3. Shawn Gomez	E-mail: smgomez@unc.edu
4. Mayetri Gupta	E-mail: gupta@bios.unc.edu
5. Andrew Nobel	E-mail: nobel@stat.unc.edu
6. Wei Sun	E-mail: wsun@bios.unc.edu
7. Alex Tropsha	E-mail: alex_tropsha@email.unc.edu
8. Simon Wang	E-mail: wangx@email.unc.edu
9. Fred A. Wright	E-mail: fwright@bios.unc.edu

Current Investigators:

1. Ivan Rusyn	E-mail: iir@unc.edu
2. Timothy Elston	E-mail: telston@amath.unc.edu
3. Shawn Gomez	E-mail: smogomez@unc.edu
4. Alex Tropsha	E-mail: alex_tropsha@email.unc.edu
5. Fred A. Wright	E-mail: fwright@bios.unc.edu
6. Karin Yeatts	E-mail: karin_yeatts@unc.edu

Institution:
1. University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, 27599

EPA Project Officer:
Project Period: April 1, 2008 through March 31, 2012
Project Amount: $3,400,000

RFA:

Research Category:

Description:

Objective:
The objective of this proposal is to create The Carolina Center for Computational Toxicology. We present a clear plan for an effective, broad and interdisciplinary effort to devise novel tools, methods and knowledge that will utilize publicly available data to assist the regulatory agencies and the greater environmental health sciences community in protecting the environment and human health.

Approach:
The Center will apply knowledge and expertise of the individual investigators and teams to develop complex predictive modeling solutions that span from mechanistic- to discovery-based efforts. The Center will be divided into three Research Projects and an Administrative Core Unit. To balance the research needs detailed in the Funding Opportunity EPA-G2007-STAR-D1 and maximize the interactions

within the Center and between the Center and the larger environmental health community, the following sub-disciplines were recognized as critical to the Center: 1) Biomedical modeling of chemical-perturbed networks (Project 1, PIs Gomez and Elston), 2) Toxico-genetic modeling (Project 2, PIs Wright and Rusyn), and 3) Chem-informatics (Project 3, PI Tropsha). Overall, we chose a bottom-up approach to predictive computational modeling of adverse effects of toxic agents. Our emphasis spans from the fine-scale predictive simulations of the protein-protein/-chemical interactions in nuclear receptor networks (Project 1), to mapping chemical-perturbed networks and devising modeling tools that can predict the pathobiology of the test compounds based on a limited set of biological data (Project 1), to building tools that will enable toxicologists to understand the role of genetic diversity between individuals in responses to toxicants (Project 2), to unbiased discovery-driven prediction of adverse chronic *in vivo* outcomes based on statistical modeling of chemical structures, high-throughput screening and the genetic makeup of the organism (Project 3). The Administrative Core Unit provides administrative and programming staff in support of the entire Center, is responsible for ensuring that Center objectives and goals are being met, and provides oversight for each for the Projects. A detailed Quality Management Plan ensures that the research and data management will be conducted with integrity and adhering to appropriate data interchange standards. The plans for Public Outreach will ensure that the activities of the Center are translated into useable information and materials for the public and policy makers.

Expected Results:

The Center will advance the field of computational toxicology through the development of new methods and tools, as well as through collaborative efforts. In each Project, new computer-based models will be developed and published that represent the state-of-the-art. The tools produced within each project will be widely disseminated, and the emphasis will be placed on their usability by the risk assessment community and the investigative toxicologists alike. The synthesis of data from a variety of sources will move the field of computational toxicology from a hypothesis-driven science toward a predictive science.

Environmental Bioinformatics and Computational Toxicology Center

EPA Grant Number: R832721

Investigators:
1. William J. Welsh E-mail: welshwj@umdnj.edu
2. Panos G. Georgopoulos E-mail: panosg@fidelio.rutgers.edu

Current Investigators:
1. William J. Welsh E-mail: welshwj@umdnj.edu
2. Ioannis Androulakis E-mail: yannis@rci.rutgers.edu
3. Christodoulos Floudas E-mail: floudas@titan.princeton.edu
4. Panos G. Georgopoulos E-mail: panosg@fidelio.rutgers.edu
5. Marianthi Ierapetritou E-mail: marianth@sol.rutgers.edu
6. Herschel Rabitz E-mail: hrabitz@princeton.edu
7. Weida Tong E-mail: weida.tong@fda.hhs.gov

Institution:
1. University of Medicine and Dentistry of New Jersey, Newark, New Jersey, 07101
2. Princeton University, Princeton, New Jersey, 08544
3. Rutgers University, New Brunswick, New Jersey, 08901

Current Institution:
1. Princeton University, Princeton, New Jersey, 08544
2. Rutgers University, New Brunswick, New Jersey, 08901
3. U.S. Food and Drug Administration, Silver Spring, Maryland, 20993
4. University of Medicine and Dentistry of New Jersey, Newark, New Jersey, 07101

EPA Project Officer:
Project Period: October 1, 2005 through September 30, 2010

Project Amount: $5,422,135

RFA:
Research Category:
Description:

Objective:
The Research Center will bring together a team of computational scientists, with diverse backgrounds in bioinformatics, cheminformatics and enviroinformatics, from UMDNJ, Rutgers, and Princeton Universities, and the USFDA's Center for Toxicoinformatics. This team will address, in a systematic and integrative manner, multiple elements of the toxicant *Source-to-Outcome sequence (Investigational Area 1*, as identified in the RFA) as well as develop cheminformatics tools for toxicant characterization

(*Investigational Area 2, Predictive Models for Hazard Identification*). The computational tools to be developed through this effort will be extensively evaluated and refined through collaborative applications involving Center scientists as well as colleagues from the three universities and USEPA; particular emphasis will be on methods that enhance current quantitative risk assessment practices and reduce uncertainties.

Approach:
The proposed Center will address a wide range of issues in Investigational Areas 1 and 2 and, furthermore, will pursue complementary applications in risk assessment (Investigational *Area 3*). This will be achieved with the requested resources, by building upon a variety of methods and software systems recently developed at UMDNJ, Rutgers, Princeton (with funding from USEPA, USDOE, NIH and NSF), and USFDA. Research activities over the proposed 5-year effort will be organized in five projects; each project will develop a set of “stand-alone" components addressing specific problems of computational toxicology. Furthermore, Research Project 1 will provide an integrative framework for Investigational Area 1 while Project 4 will address the core issues of Area 2. Extensive interaction as well as public outreach and training activities will constitute essential elements of the Center and will be tightly interwoven with the research activities.

Expected Results:
Research Project 1 (Development and Application of a Dose-Response Information Analysis [DORIAN] System) will provide an integrative framework for the outcomes of the other projects. This framework will include the following components: a web-accessible Environmental Bioinformatics Knowledge Base (EBKB) that will provide a user-oriented interface to an extensive set of information and modeling resources; the ebTrack integrated analysis system that will include linkages to multiple (public and commercial) computational and database systems; Bayesian computational tools for characterizing and reducing uncertainties in mechanistic modeling of toxicity pathways; diagnostic computational tools for sensitivity and stability analysis of mechanistic models and statistical methods for data analysis; and enhanced tools for quantitative risk assessment (QRA) applications (e.g. for cross-species extrapolation, chemical mixtures, and dose-response).
Research Project 2 (Hepatocyte Metabolism Model for Xenobiotics) will develop tools for identifying maximally informative sets of toxicologically relevant genes; tools for analysis of toxicologically relevant regulatory networks; an expanded version of the Rutgers hepatocyte metabolism model that will incorporate transformations of xenobiotics; and tools for the analysis of transcriptional regulation that will allow assessing changes in hepatocyte phenotypic phase space.
Research Project 3 (Tools for Optimal Identification of Biological Networks) will develop efficient identification tools for inferring biological network structure from available laboratory data; optimization tools for extracting quantitative information of biological system parameters (rate constants, diffusion coefficients, binding affinities, etc.); global sensitivity analysis tools for identifying most effective molecular targets or pathways of biological networks and for guiding the design of laboratory experiments; and optimal feedback control tools for inferring networks with feedback loops.

Research Project 4 (Cheminformatics Tools for Toxicant Characterization) will develop an integrative hierarchical decision-forest framework for toxicant characterization that encompasses several novel technologies, including the Shape Signatures tool that rapidly matches organic and organometallic chemicals with each other or, alternatively, against target receptor sites/subsites; the Polynomial Neural Network (PNN) that automatically generates physically-intuitive linear or non-linear QSAR models; and virtual high-throughput screening (vHTS) methods that predict ligand binding affinity and provide mechanistic information (toxicity pathways).

Research Project 5 (Optimization Tools for In Silico Proteomics) will customize computational methods for protein structure prediction and de novo protein design, with specific focus on the important families of Glutathione Transferases (GST) (cytosolic, mitochondrial and microsomal GST); develop and implement computational methods for elucidating the topology of signal transduction networks and addressing uncertainties in experimental data and models; and develop de novo computational proteomics methods for peptide and protein identification via tandem mass spectroscopy.

Carolina Environmental Bioinformatics Center

EPA Grant Number: R832720

Investigators:

1. Fred A. Wright	E-mail: fwright@bios.unc.edu
2. Kenneth J. Galluppi	E-mail: galluppi@unc.edu
3. Lawrence Kupper	E-mail: kupper@bios.unc.edu
4. Stephen J. Marron	E-mail: marron@email.unc.edu
5. Jan F. Prins	E-mail: prins@cs.unc.edu
6. Ivan Rusyn	E-mail: iir@unc.edu
7. David Stotts	E-mail: stotts@cs.unc.edu
8. David Threadgill	E-mail: dwt@med.unc.edu
9. Alex Tropsha	E-mail: alex_tropsha@email.unc.edu

Current Investigators:

1. Fred A. Wright	E-mail: fwright@bios.unc.edu
2. Rosann Farber	E-mail: rosann.farber@pathology.unc.edu
3. Leonard McMillan	E-mail: mcmillan@cs.unc.edu
4. Ivan Rusyn	E-mail: iir@unc.edu
5. Alex Tropsha	E-mail: alex_tropsha@email.unc.edu

Institution:
1. University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, 27599

EPA Project Officer:
Project Period: October 1, 2005 through September 30, 2010
Project Amount: $4,494,117

RFA:

Research Category:
Description:

Objective:
The Carolina Environmental Bioinformatics Research Center brings together multiple investigators and disciplines, combining expertise in biostatistics, computational biology, chem-informatics and computer science to advance the field of Computational Toxicology.

The objective of this proposal is to create an Environmental Bioinformatics Research Center with broad-ranging capability to enhance and advance the field of Computational Toxicology. The Center will develop novel analytic and computational methods, create efficient user-friendly tools to disseminate the methods to the wider community, and will apply the computational methods to data from molecular toxicology and other studies.

Approach:
Effort will be divided into three Research Projects and an Administrative Unit. Each Research Project is further divided into Functional Areas consisting of Analysis, Methods Development, and Tools Development. Project 1 (Biostatistics in Computational Biology) will provide biostatistical support to the Center, performing analysis and developing new methods in collaboration with EPA personnel and the computational toxicology community. Project 2 (Chem-informatics) will coordinate the compilation and mining of data from relevant external databases and perform analysis and methods development for investigating Quantitative Structure-Activity Relationships with burgeoning high-throughput chem-informatics data. In addition, Project 2 will develop computational tools to perform these tasks. Project 3 (Computational Infrastructure for Systems Toxicology) will create a framework for merging data from various –omic technologies in a systems biology approach. The investigation of rodent liver toxicity is used as a driving biological problem, inspiring new methods and architectures for data storage. Finally, Project 3 will provide programming support for the further development of tools arising from Projects 1 and 2. The Administration Core provides and staff and support to the Center, is responsible for ensuring that Center objectives and goals are being met, and provides oversight for each for the Functional Areas. A detailed Quality Management Plan ensures that the research and data management will be conducted with integrity and adhering to appropriate data interchange standards. The plans for Public Outreach and Translation Activity will ensure that the activities of the Center are translated into useable information and materials for the public and policy makers.

Expected Results:
The Center is expected to advance the field of computational toxicology through the development of new methods and tools, as well as through direct collaborative efforts with EPA and other environmental scientists. In each Project, we expect that new methods will be developed and published that represent the state-of-the-art. The tools developed within each project will be widely disseminated, and will be useful both to trained bioinformatics scientists and bench scientists. The synthesis of data from a variety of sources will move the field of computational toxicology from a hypothesis-driven science toward a predictive science. Each Project is goal-oriented, with criteria for success that will be reviewed by the Scientific Advisory Committee.

The Texas-Indiana Virtual STAR Center; Data-Generating *in vitro* and *in silico* Models of Developmental Toxicity in Embryonic Stem Cells and Zebrafish

EPA Grant Number: 83428901

Investigators:
1. Prof. Jan-Åke Gustafsson (Contact PI) E-mail: jgustafsson@uh.edu
2. Prof. Richard H. Finnell E-mail: rfinnell@ibt.tamhsc.edu
3. Prof. James A. Glazier E-mail: glazier@indiana.edu

Institutions:
1. University of Houston, Department of Biology and Biochemistry, Houston, Texas, 77204
2. The Texas A&M Institute for Genomic Medicine, Texas A&M University/Texas A&M Health Science Center, Houston, Texas, 77030
3. Indiana University, Department of Physics, Bloomington, Indiana, 47405-7003

EPA Project Officer: (leave blank)
Project Period:
Project start: November 1, 2009
Project end: October 31, 2012

Project Amount: $3,190,993

RFA: (leave blank)

Research Category: (leave blank)

Description

Objectives/Hypothesis:

As chemical production increases worldwide, there is increasing evidence as to their hazardous effects on human health at today's exposure levels, which further implies that current chemical regulation is insufficient. Thus, a restructuring of the risk assessment procedure will be required to protect future generations. Given the very large number of man-made chemicals and the likely complexity of their various and synergistic modes of action, emerging technologies will be required for the restructuring. The main objective of the proposed multidisciplinary Texas Indiana Virtual STAR (TIVS) Center is to contribute to a more reliable chemical risk assessment through the development of high throughput *in vitro* and *in silico* screening models of developmental toxicity. Specifically, the TIVS Center aims to generate *in vitro* models of murine embryonic stem cells and zebrafish for developmental toxicity. The data produced from these models will be further exploited to produce predictive *in silico* models for developmental toxicity on processes that are relevant also for human embryonic development.

Approach:

The project is divided into three Investigational Areas; zebrafish models, murine embryonic stem cells models and in silico simulations. The approaches are to:

1. Generate developmental models suitable for high throughput screening. Zebrafish developmental models (transgenic GFP/EGFP/RFP models of crucial steps in development) and embryonic stem cell (ESC) differentiation models (transgenic beta-geo models of crucial steps in differentiation) will be generated. Important morphology features and signaling pathways during development will be documented. The impact of environmental pollutants on development and differentiation will be assessed in the models. Finally, the models will be refined for high throughput screening and automation.
2. Generate a computational model that faithfully recreates the major morphological features of normal wild-type zebrafish development (ie-segmentation into somites, proper patterning of vascular and neural systems) and the differentiation to three primitive layers (endoderm, mesoderm and ectoderm) in mouse embryonic stem cells. The data for simulations are produced from developed high information content zebrafish and ESC models. Once a working model of normal development has been generated, we will carry out a directed series of parameter sweeps to try to create developmental defects *in silico.* We will compare the results of computationally created defects with experimentally-generated defects in zebrafish and embryonic stem cells. Best matches between the two datasets will suggest hypotheses about possible mechanisms by which defects occur.
3. Perform proof-of-concept experiments of the *in vitro* and *in silico* test platforms with a blind test of chemicals.

Techniques will be molecular biology techniques on zebrafish and ESC models, such as cloning, imaging, in vitro differentiation and in vitro exposure studies, and in silico mathematical simulations.

Expected Results (Outputs/Outcomes):

In collaboration with other initiatives taken in the field of chemical safety, our generated results and models will contribute to large screening effort to prioritize chemicals for further risk assessment. We will specifically contribute with:

- 9 transgenic fish lines validated for toxicity screening
- 16 embryonic stem cell models validated for toxicity screening
- High information content models on development and differentiation to produce data for in silico simulations, within the project and elsewhere
- Computational models for developmental toxicology of normal development and of mechanisms by which chemical perturbations cause experimentally-observed developmental defects
- Information on developmental toxicity on 39 compounds

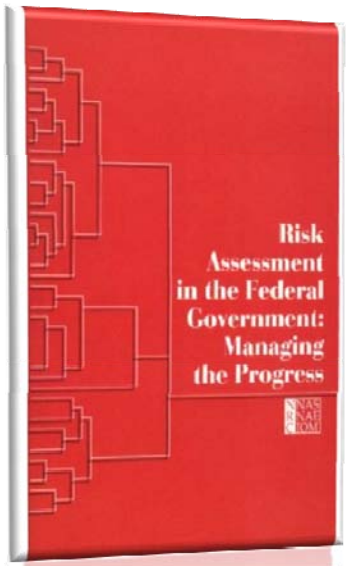
All the data produced in this project will be released to public databases. The developed models will be automated for high throughput screening.

Supplemental Keywords:
Risk assessment, effects, dose-response, teratogen, organism, cellular, infants, chemicals, toxics, aquatic ecosystem protection, pollution prevention, green chemistry, public policy, environmental chemistry, biology, physics, genetics, mathematics, modeling, measurement methods

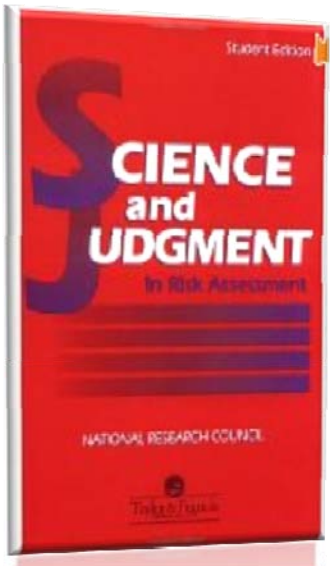
Chemical Substance *In Vitro*/*In Silico* Screening System To Predict Human and Ecotoxicological Effects (Chem*Screen*)

Bart van der Burg, BioDetection Systems, Amsterdam, The Netherlands (Coordinator)

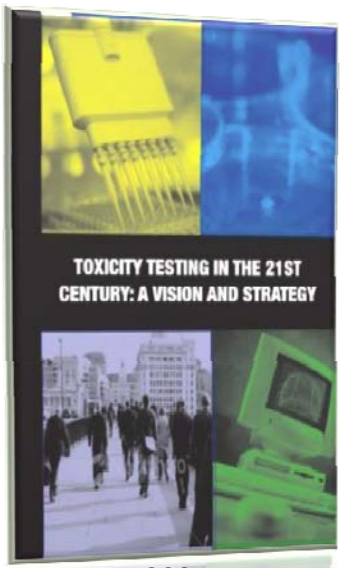
The current system of risk assessment of chemicals is complex, very resource-intensive, and extremely time-consuming. Because of this, there is a great need to modernize this process. However, this is not feasible without alternative, integrated testing strategies in which chemical characteristics are used to more advantage, and where costly and time-consuming animal tests are replaced to a large extent by more rapid, cheap, and ethically less controversial methods. This is particularly needed for reproductive toxicity testing of chemicals. Reproductive toxicity is important to assess both human and environmental toxicity and uses the most animals in toxicity testing. Unfortunately, there are very few alternative methods. The EU project Chem*Screen* is a partnership between nine European institutes and companies from five different countries. It aims to generate alternative methods and place the tests in a more general innovative animal-free testing strategy. For this, we will generate a simple rapid screening system, which aims at widespread implementation within the tight time schedule of the REACH program. It will be a flexible tool that can be adapted and used for applications beyond the scope of REACH and in the post-REACH period. It will use *in silico* methods for prescreening chemicals for all relevant toxic effects. When found positive, this will be followed by further *in silico* and *in vitro* tests, most of which are available already. To fill the gap of suitable alternative methods for reproductive toxicity testing, we will use a novel high-throughput approach combining *in silico/in vitro* methods. In this approach, we will combine knowledge of critical processes affected by reproductive toxicants with knowledge on the mechanistic basis of such effects. Straightforward data interpretation and decision trees will be developed in which all information on the potential toxicity of a chemical is considered. In this way, we will provide a cost-effective means to generate a basic set of data on toxicological properties of chemicals and a decision tool to assess if further testing of chemicals is required.



1983



1996



2007

2008

EMERGING SCIENCE FOR ENVIRONMENTAL HEALTH DECISIONS

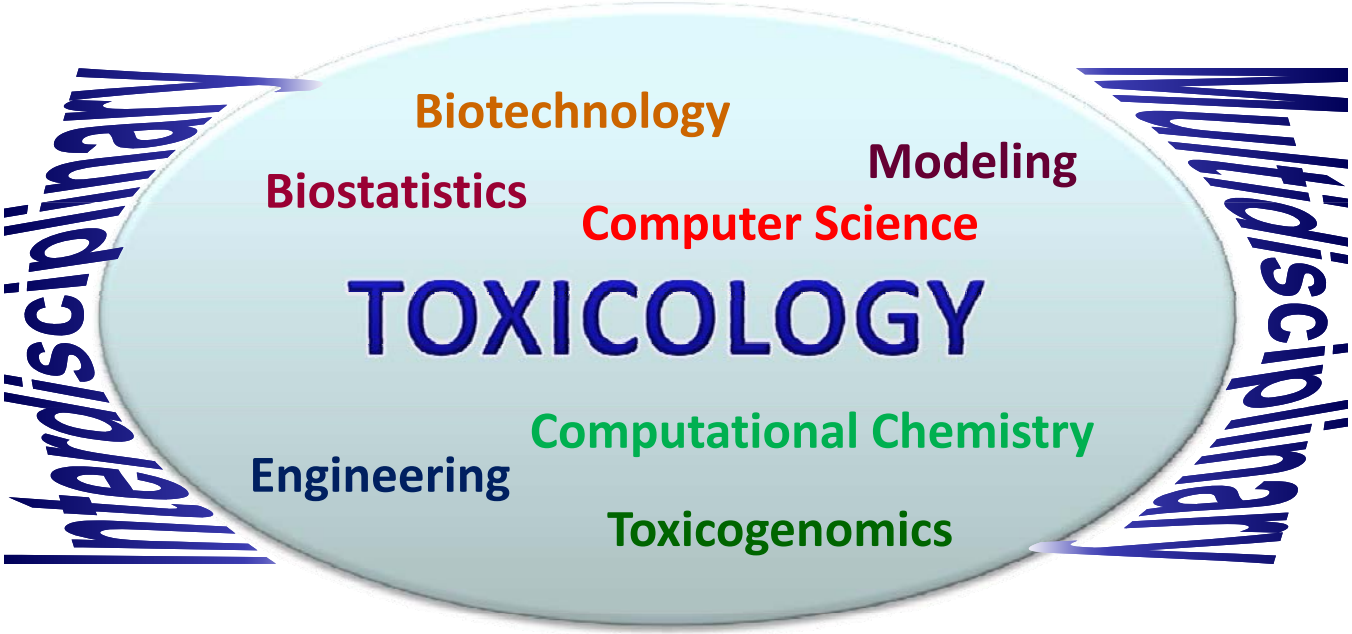
Computational Toxicology: From Data to Analyses to Applications

SEPTEMBER 21-22, 2009 ▪ WASHINGTON, DC

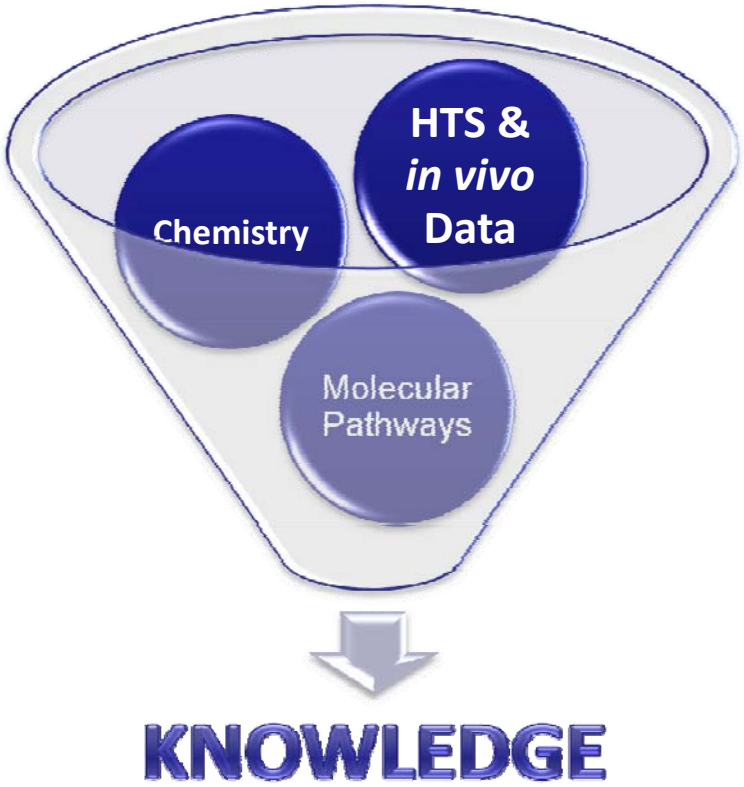
LECTURE ROOM ▪ NAS BUILDING ▪ 2101 CONSTITUTION AVENUE, NW (*NOT* 500 FIFTH STREET)

Computational Toxicology:

a sub-discipline of toxicology that aims to use the mathematical, statistical, modeling and computer science tools to better understand the mechanisms through which a given chemical induces harm and, ultimately, be able to predict adverse effects of the toxicants on human health and/or the environment



EMERGING SCIENCE FOR ENVIRONMENTAL HEALTH DECISIONS

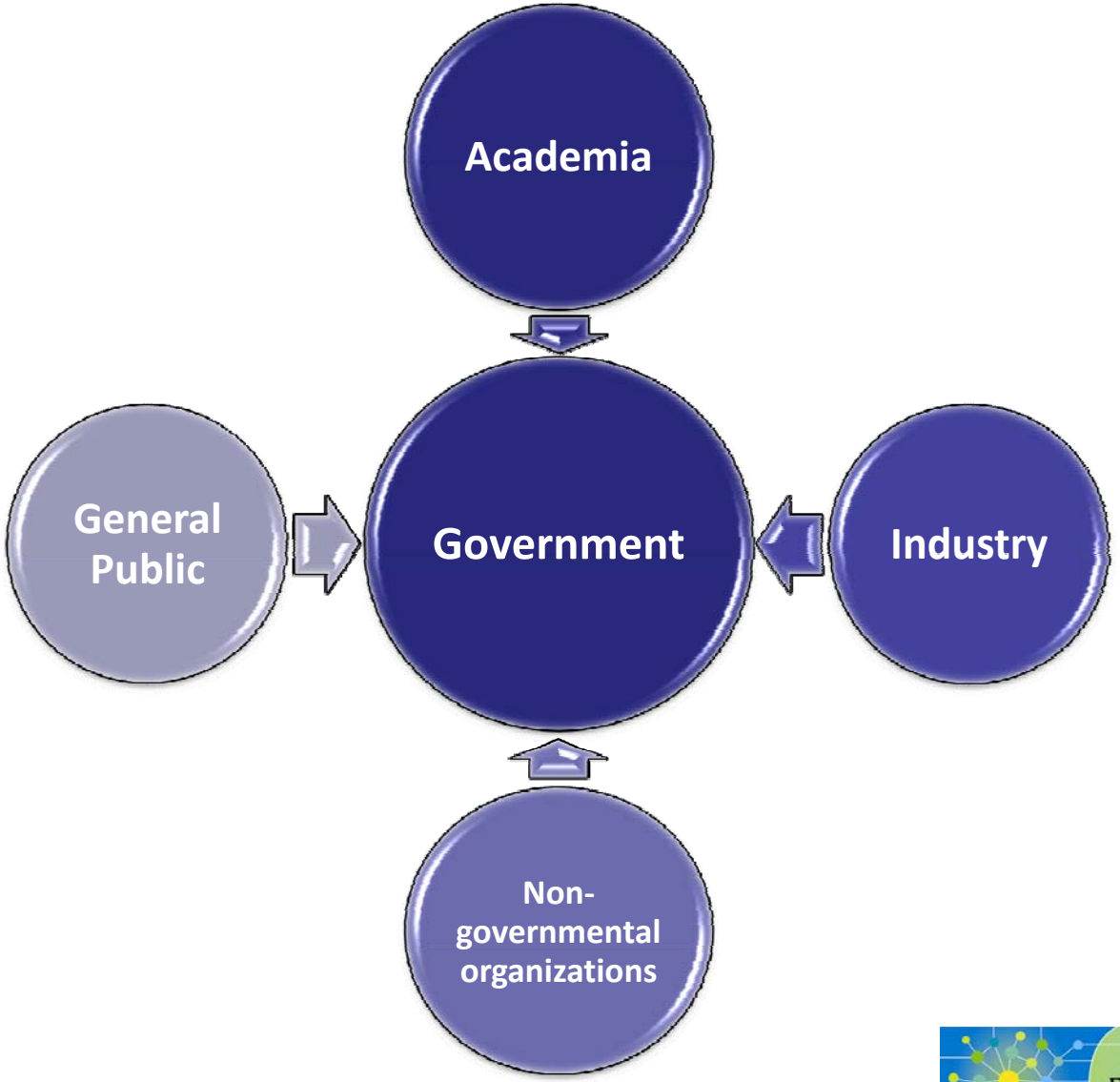


Computational Toxicology:

- **Relies on high-throughput and high-content screening assays to provide unparalleled level of detail for chemical and molecular interactions, cellular pathways and tissue-level processes**
- **Provides a novel framework for the *in silico* modeling and simulation to validate and predict key aspects of both the physiology and toxicant-induced pathology**
- **Enables fundamental understanding of the complex relationships across biological systems and supports a scientifically sound process of projecting human health risks posed by chemicals**

EMERGING SCIENCE FOR ENVIRONMENTAL HEALTH DECISIONS

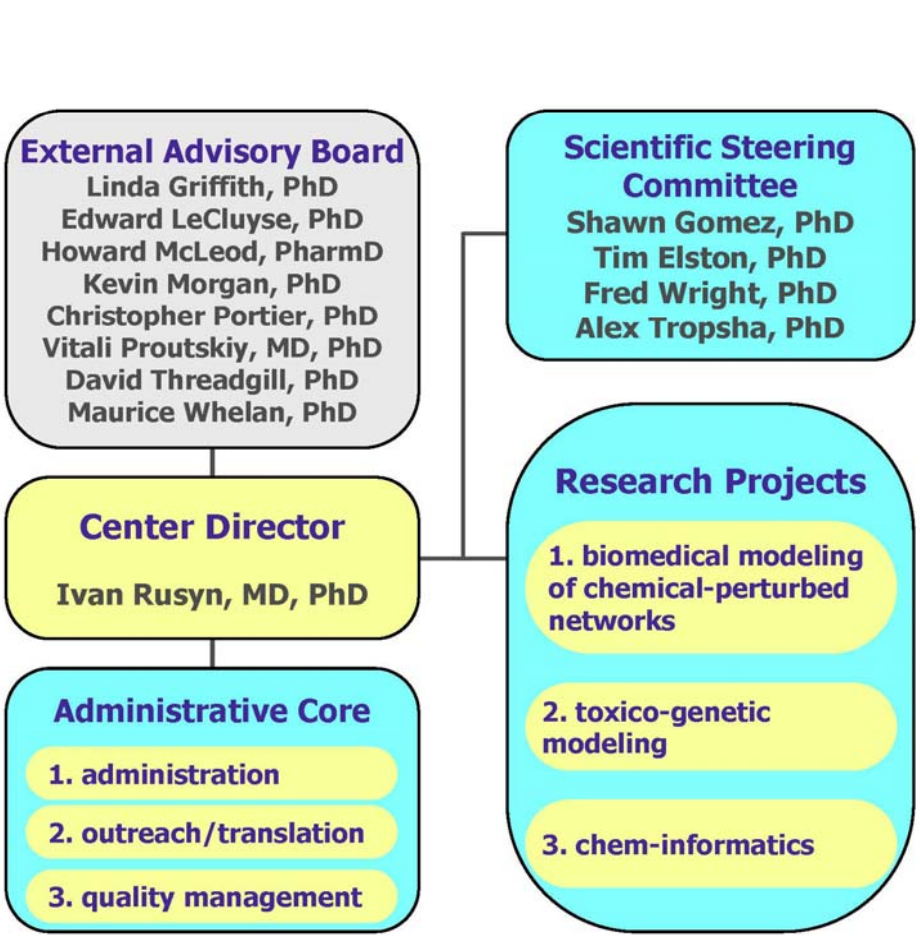
Computational Toxicology: Stakeholders

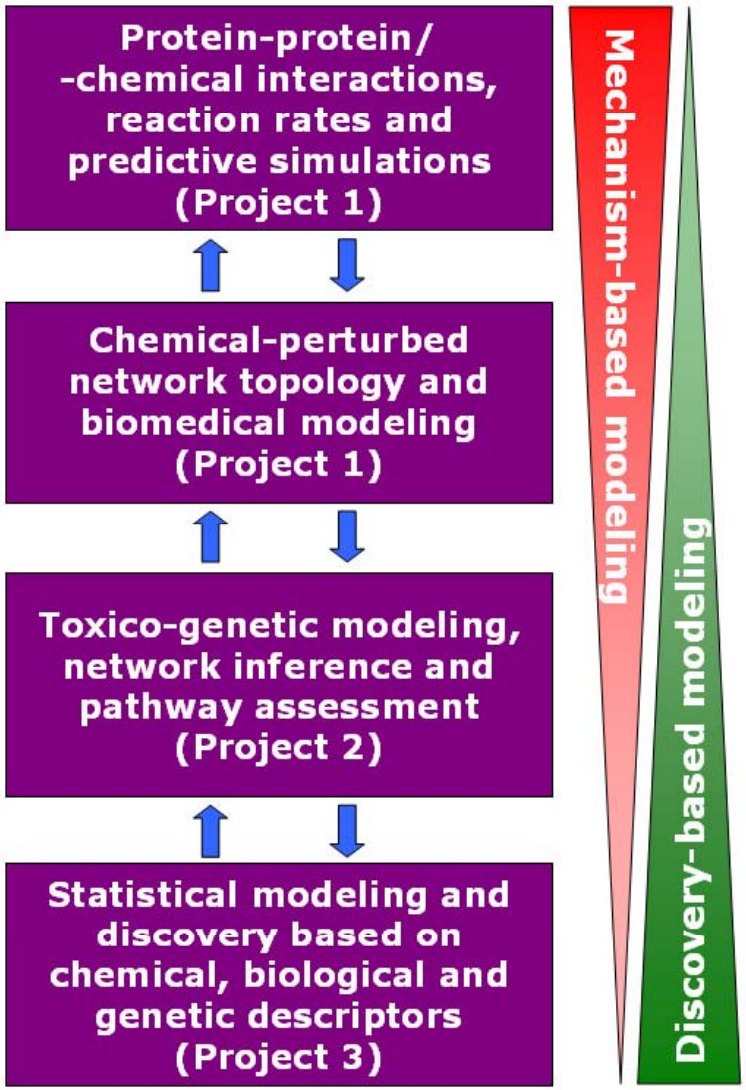


EMERGING SCIENCE FOR ENVIRONMENTAL HEALTH DECISIONS

Carolina Center for Computational Toxicology

Organizational Structure





Carolina Center for Computational Toxicology

Administrative Core

Administration Function:

- Project and budget management
- Communications
- Reporting to EPA and UNC
- Organization of the annual EAB meetings

Integration Function

- Promoting interactions within the Center
- Promoting interactions with EPA/NCCT and other partners
- Facilitating scientific interactions between Projects

Public Outreach/Translation Function

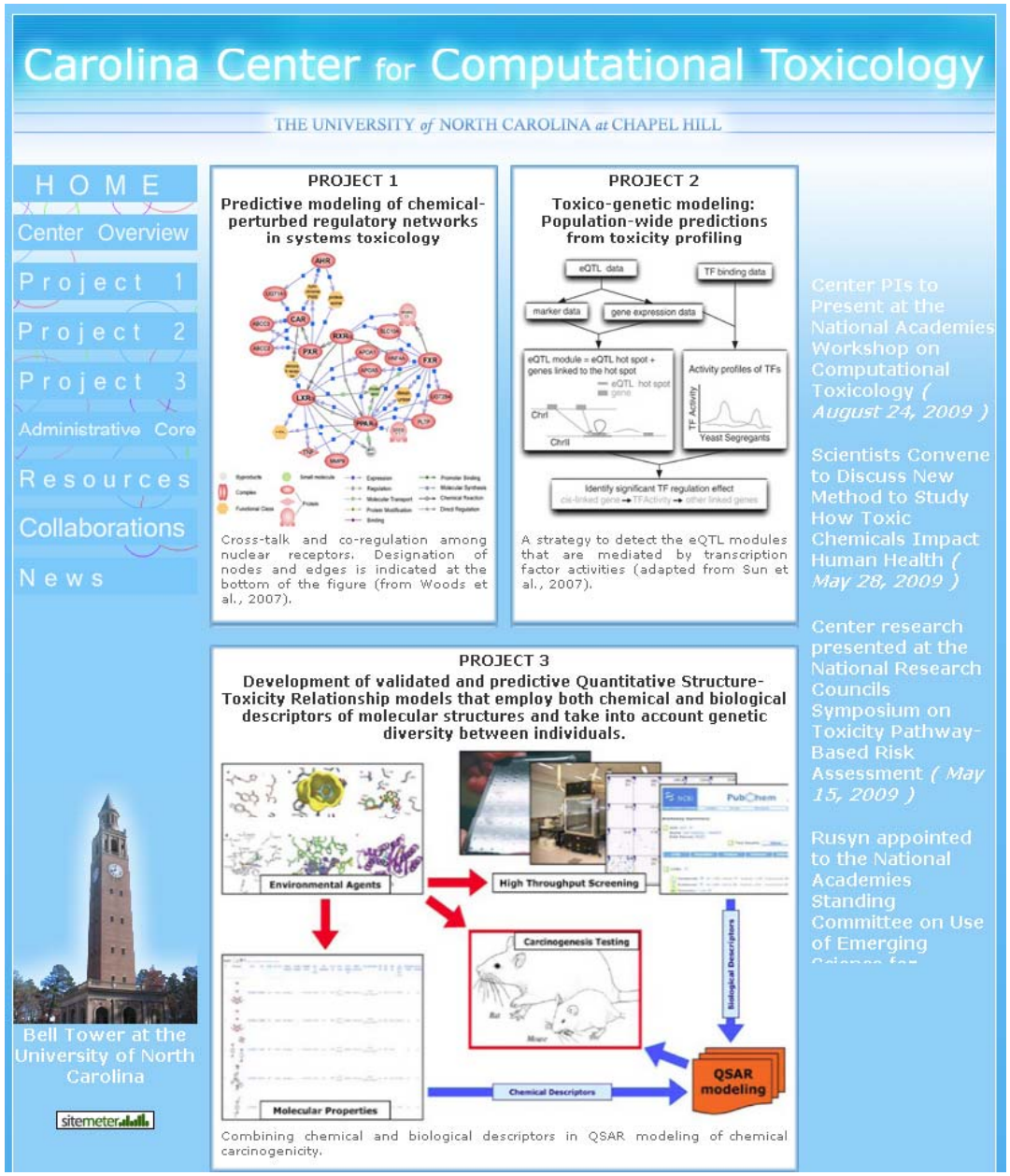
- Created Center website: http://comptox.unc.edu
- Implementing bioinformatics and chemo-informatics tools into GUI-enabled software
- Conducting joint research meetings with EPA/NCCT
- Presenting at the state, national and international scientific meetings

Quality Management Function

- Center-wide quality management plan developed and approved by the EPA
- Quality assurance project plans developed and annual audits performed for Year 1
- Remedial actions will be completed by November 01, 2009

In Step With the US EPA Guidance: Commitment to Transparency

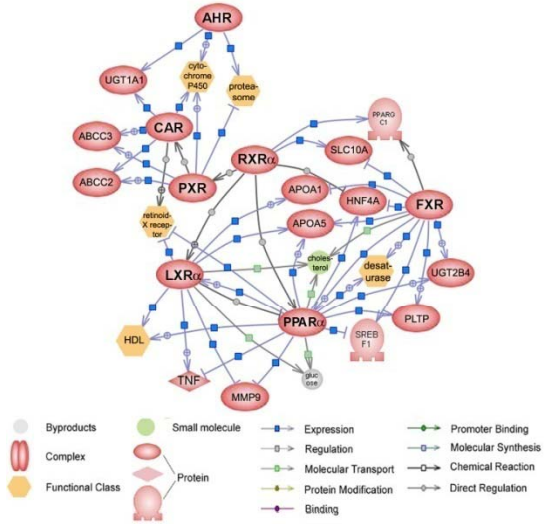
comptox.unc.edu



Carolina Center for Computational Toxicology

Project 1

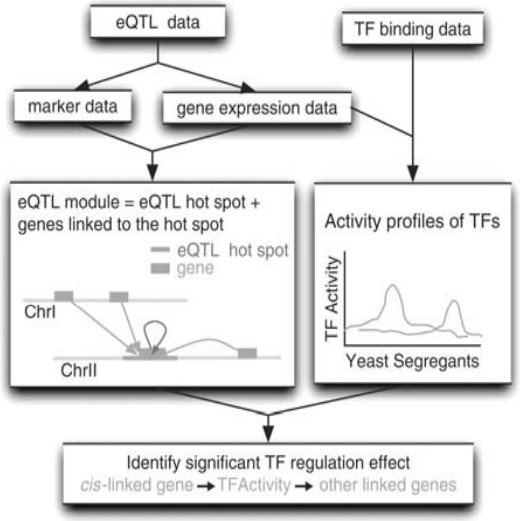
Predictive modeling of chemical-perturbed regulatory networks in systems toxicology



v-Liver
ToxCast
ToxRefDB

Project 2

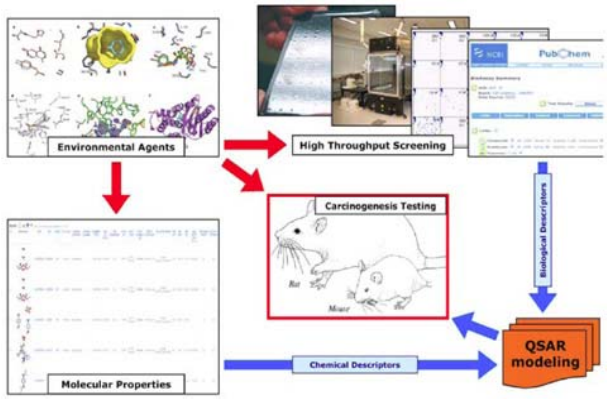
Toxico-genetic modeling: Population-wide predictions from toxicity profiling



ACToR
ToxCast
ToxRefDB

Project 3

Development of validated and predictive Quantitative Structure-Toxicity Relationship models that employ both chemical and biological descriptors of molecular structures and take into account genetic diversity between individuals



DSSTox
ACToR
ToxCast
ToxRefDB

PROJECT 1

Predictive modeling of chemical-perturbed regulatory networks in systems toxicology

Shawn Gomez – co-PI
Assistant Professor, Department of Biomedical Engineering, UNC-Chapel Hill

Timothy Elston – co-PI
Professor, Department of Pharmacology, UNC-Chapel Hill

- Develop and apply data-driven methods for the inference and high-level modeling of regulatory network response to chemical perturbation
- Develop mechanistic models of nuclear receptor function
- Integrate and deploy high-, and low-level modeling tools

Major Interactions with the US EPA

- Exploring toxicity modeling (mechanistic, dose-response, etc.): with Rory Connolly (EPA-NHEERL)
- Extension and integration of mechanistic metabolism and other models: work relevant to the v-Liver Project, Imran Shah (EPA-NCCT)
- ToxCAST: with Richard Judson (EPA-NCCT)

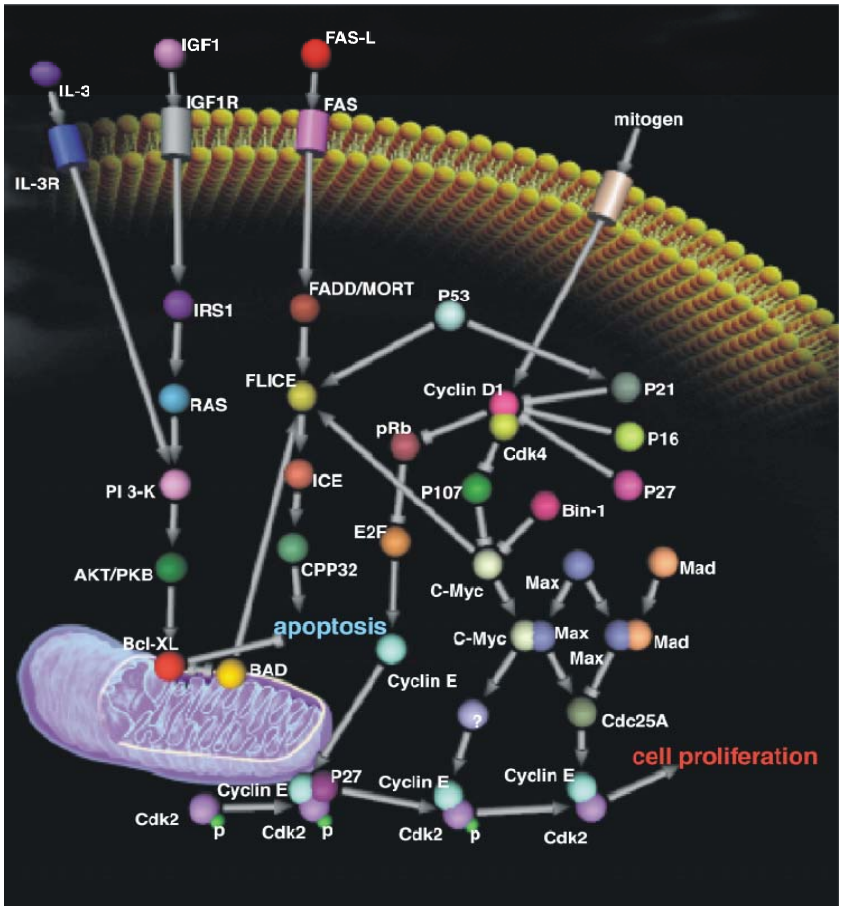
Inference & Modeling of Biological Networks

Short term:

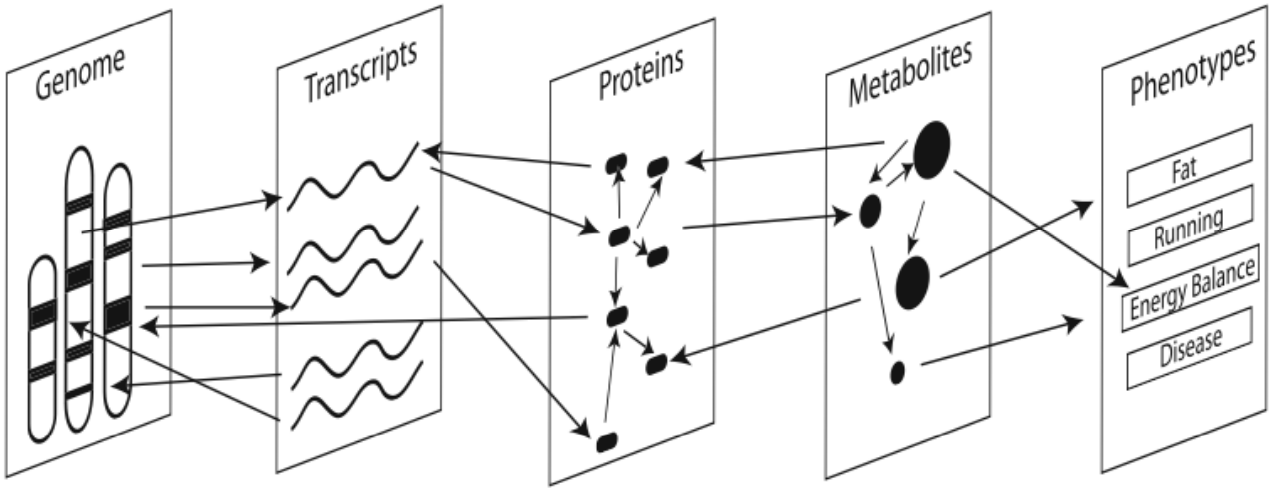
- Tool in data analysis and interpretation
- Help establish biological-chemical context

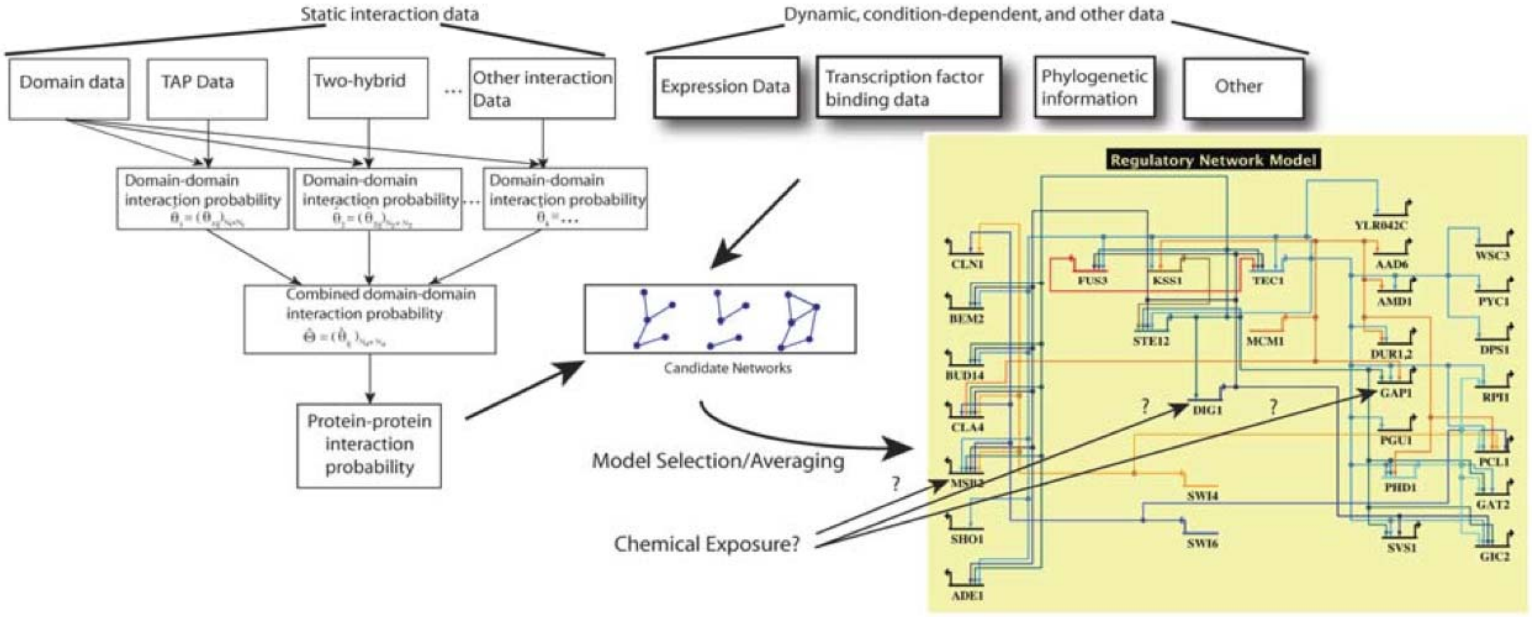
Long term:

- Components to systems - simplistic wiring
- Framework for understanding systems properties, pathways and cross-talk,...
- Basis for mechanistic models

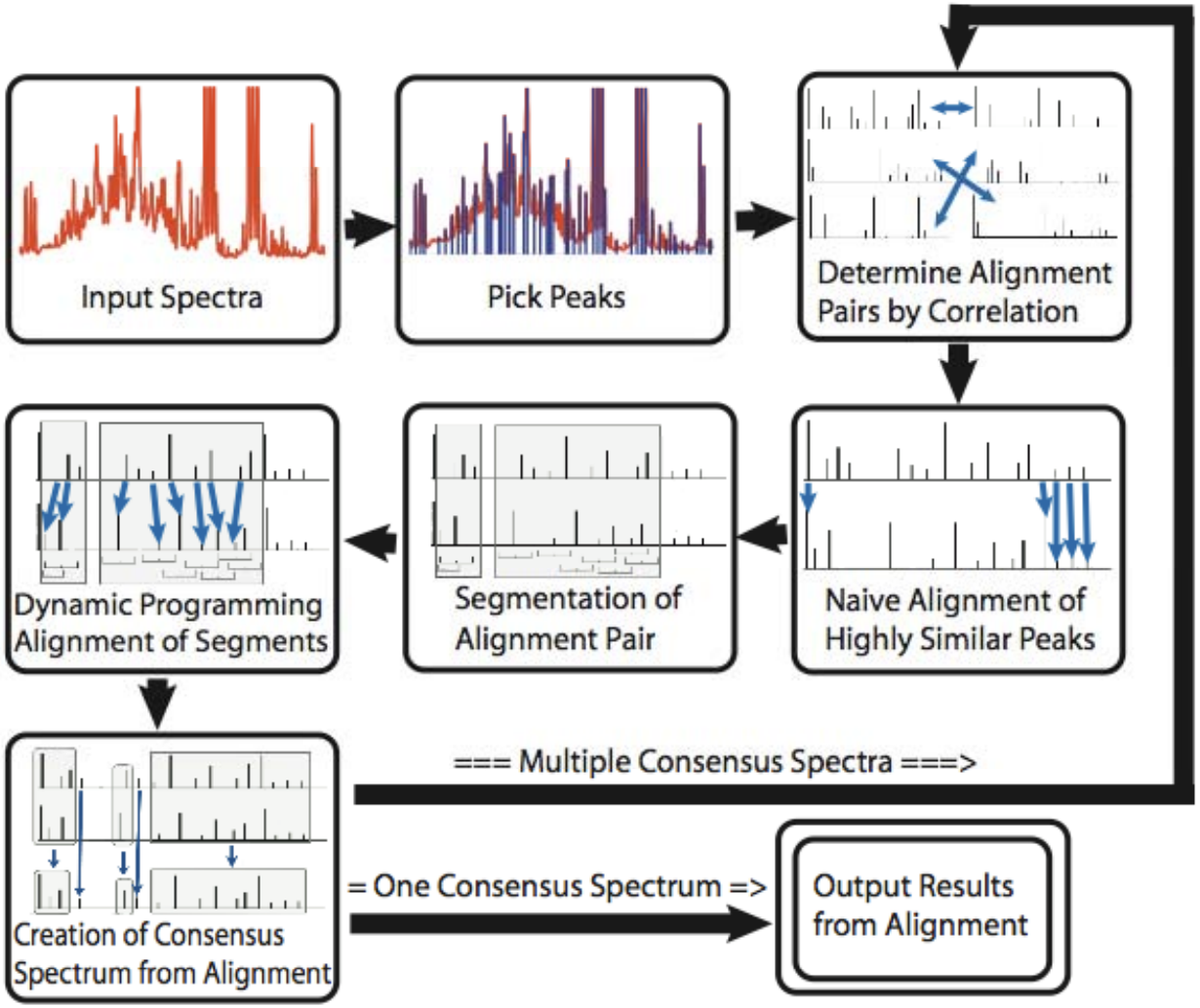


Challenge #1: Data Integration



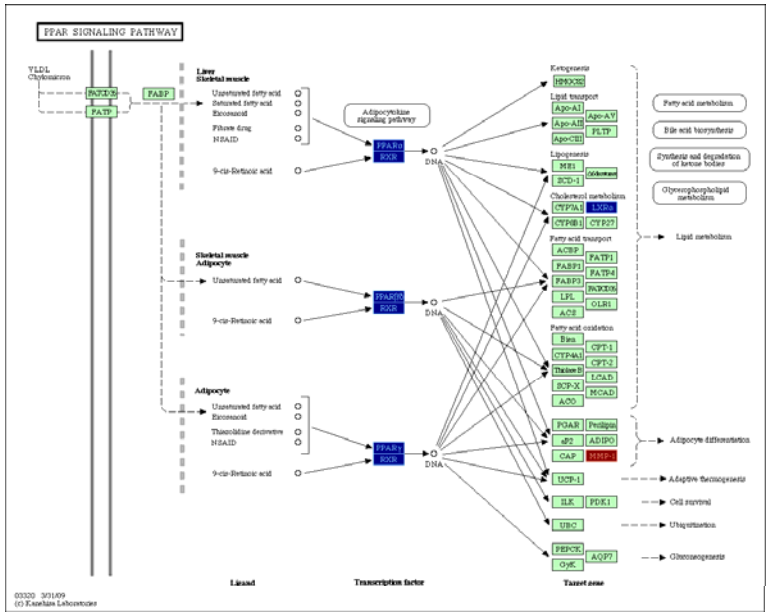


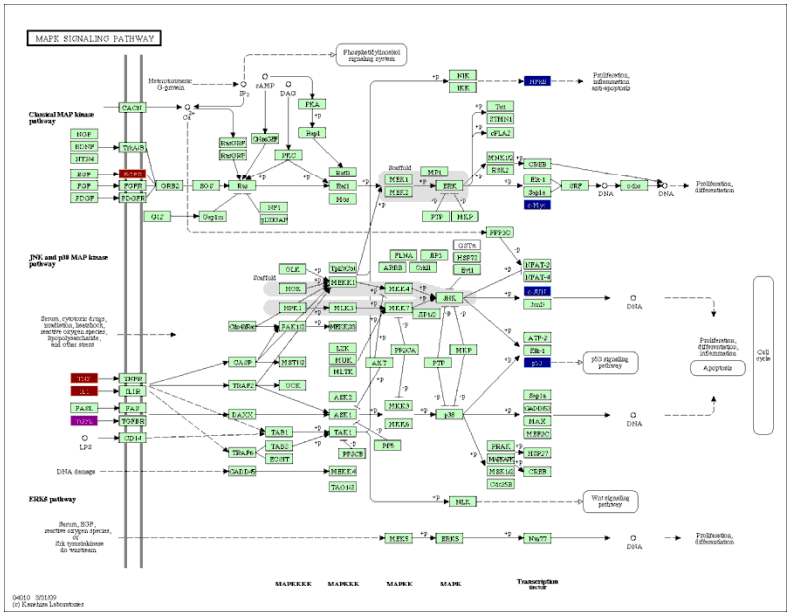
PCANS - NMR spectra alignment



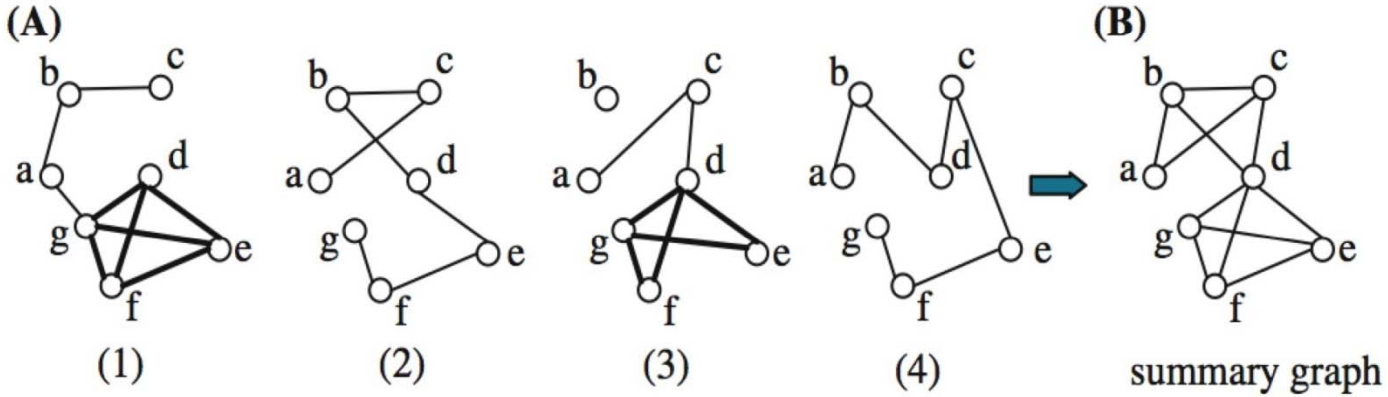
Staab J et al. (2009) BMC Bioinformatics (In revision)

Network Context: Traditional ways to create networks





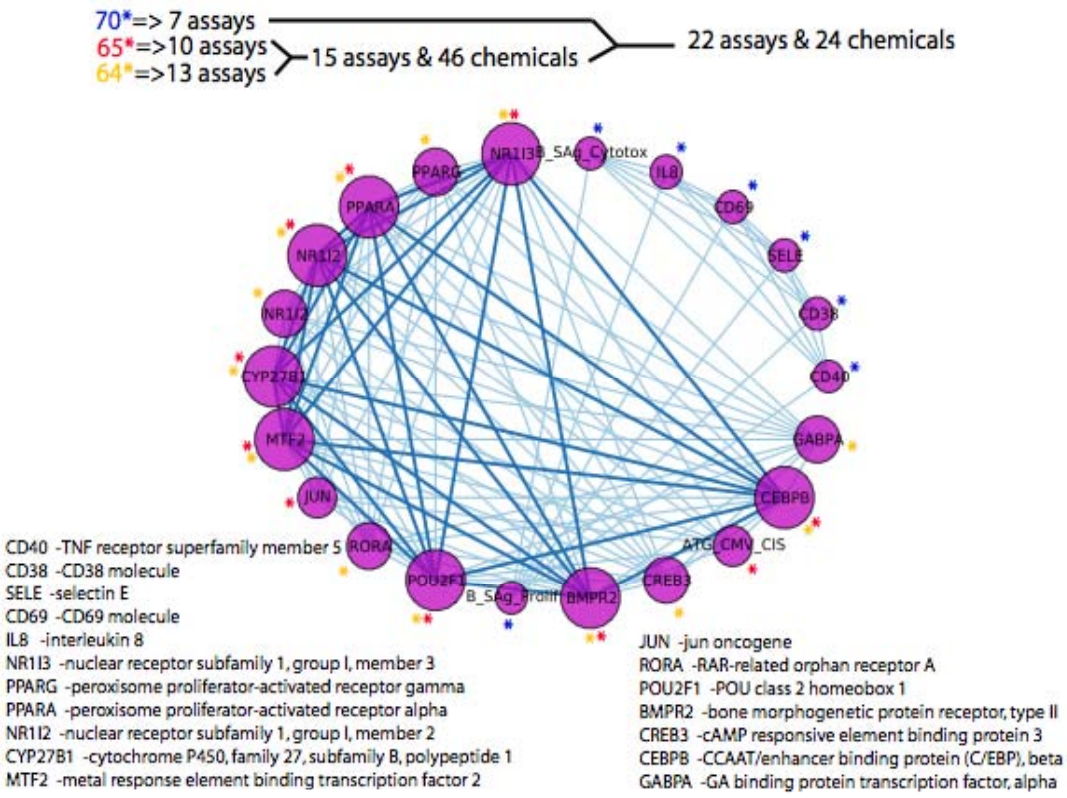
NEMO (Yan et al., 2007): frequent dense vertex set mining algorithm



Network Context: Subgraph Mining

v-Liver
ToxCast
ToxRefDB

"Functional Module" in Subgraph Mining



- Mines binary data to find all frequent 'dense' sub-graphs (cliques)
  - Nodes: Assay
  - Edges: Set of 'Active' Chemicals shared between Nodes
  - Finds all unique subgraphs for a minimum frequency of 'Active' chemicals
- Differs from Hierarchical clustering by focusing on subsets of the data
- Useful for defining composite assays that might be more predictive
- Useful for associating Assay/Chemical combinations to endpoints

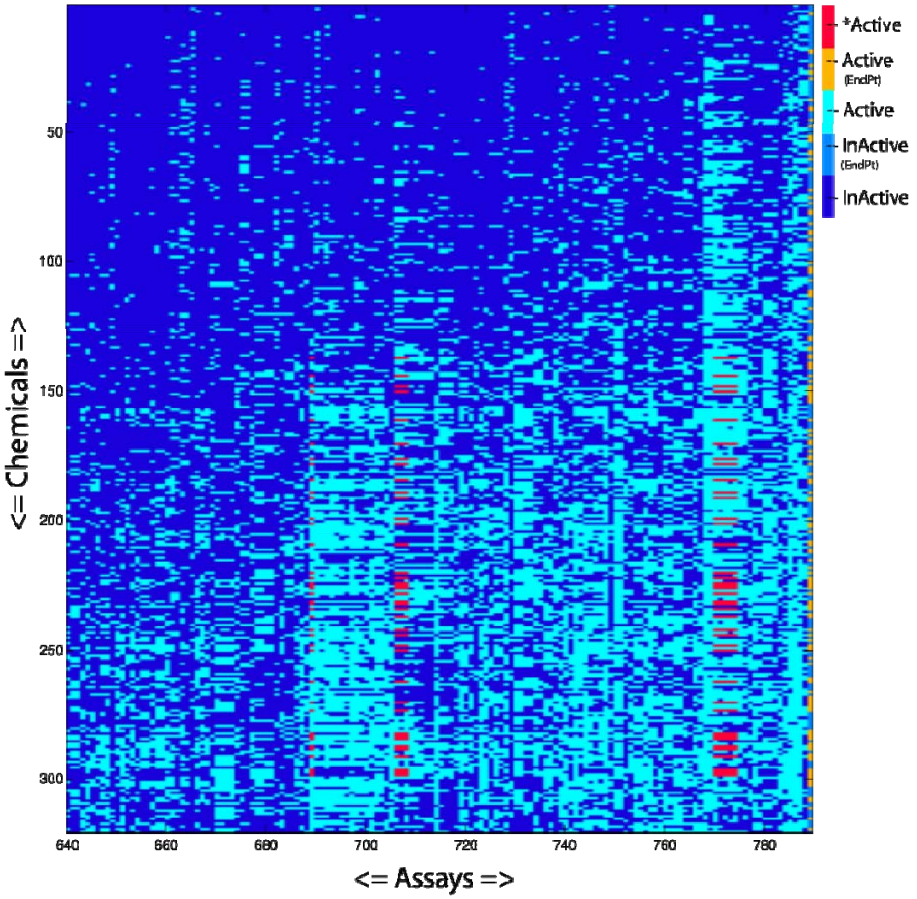
Network Context: Subgraph Mining

v-Liver
ToxCast
ToxRefDB

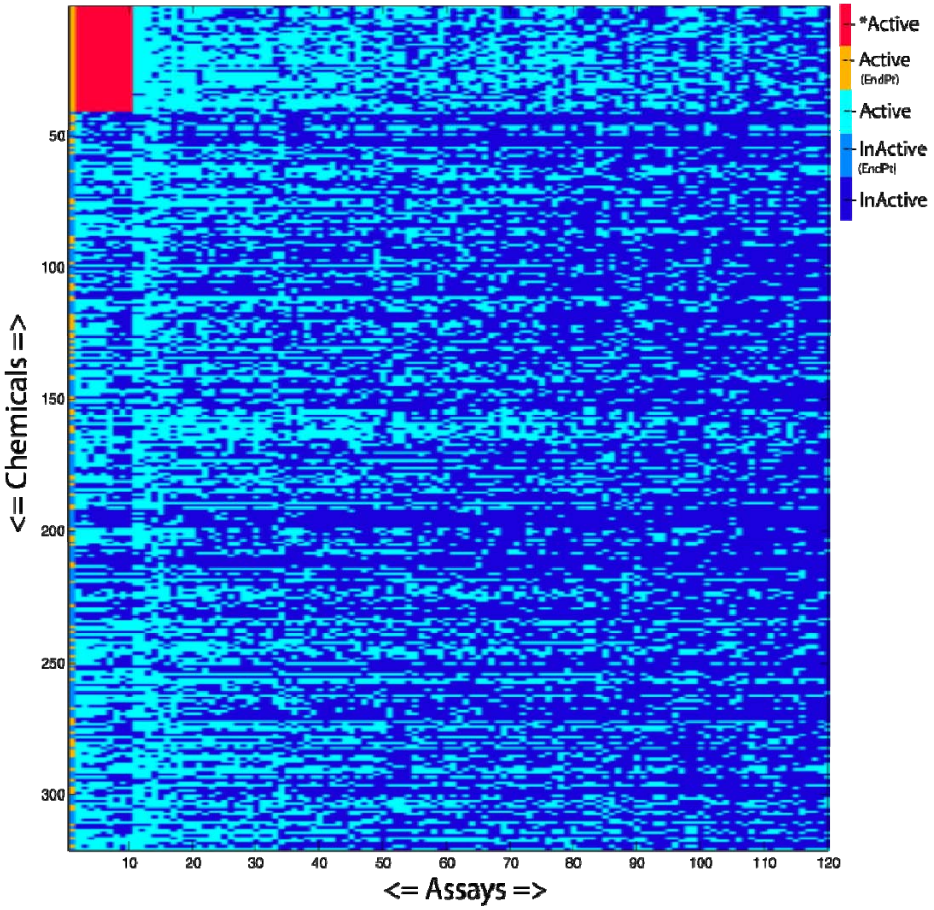
Endpoint: RatLiver_AnyLesion **Minimum Frequency: 40 chemicals (~30%)**

Module Found: 10 Assays for a set of 41 chemicals (*Active)

2D Hierarchical Clustering



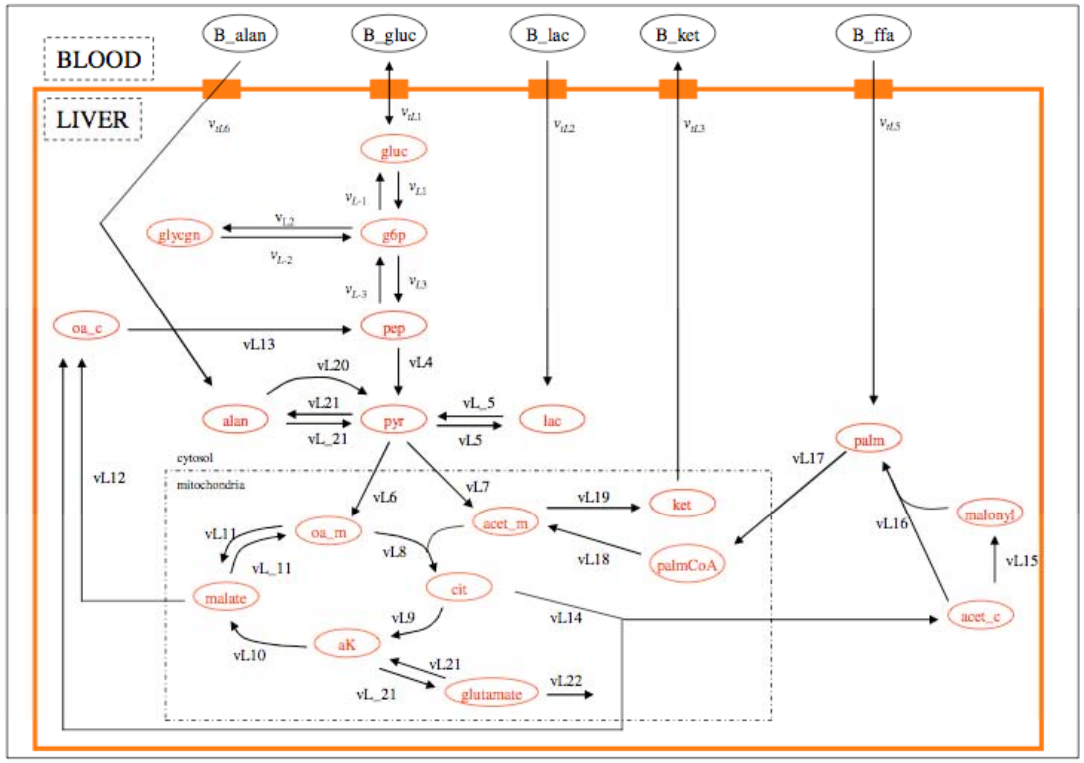
Subgraph Mining

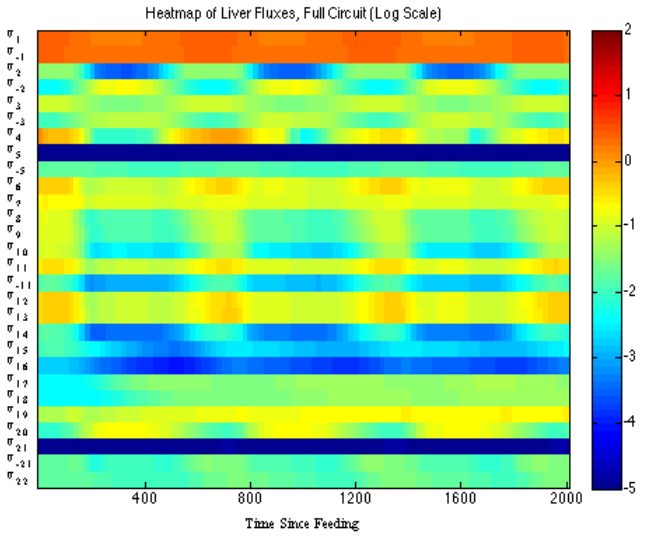


Development of a mechanistic model of cellular metabolism:

predicting changes in metabolic flux

v-Liver
ToxCast
ToxRefDB





PROJECT 2

Toxico-genetic modeling: Population-wide predictions from toxicity profiling

Fred Wright – co-PI
Professor, Department of Biostatistics, UNC-Chapel Hill

Ivan Rusyn – co-PI
Assoc. Prof., Dept. of Environmental Sciences & Engineering, UNC-Chapel Hill

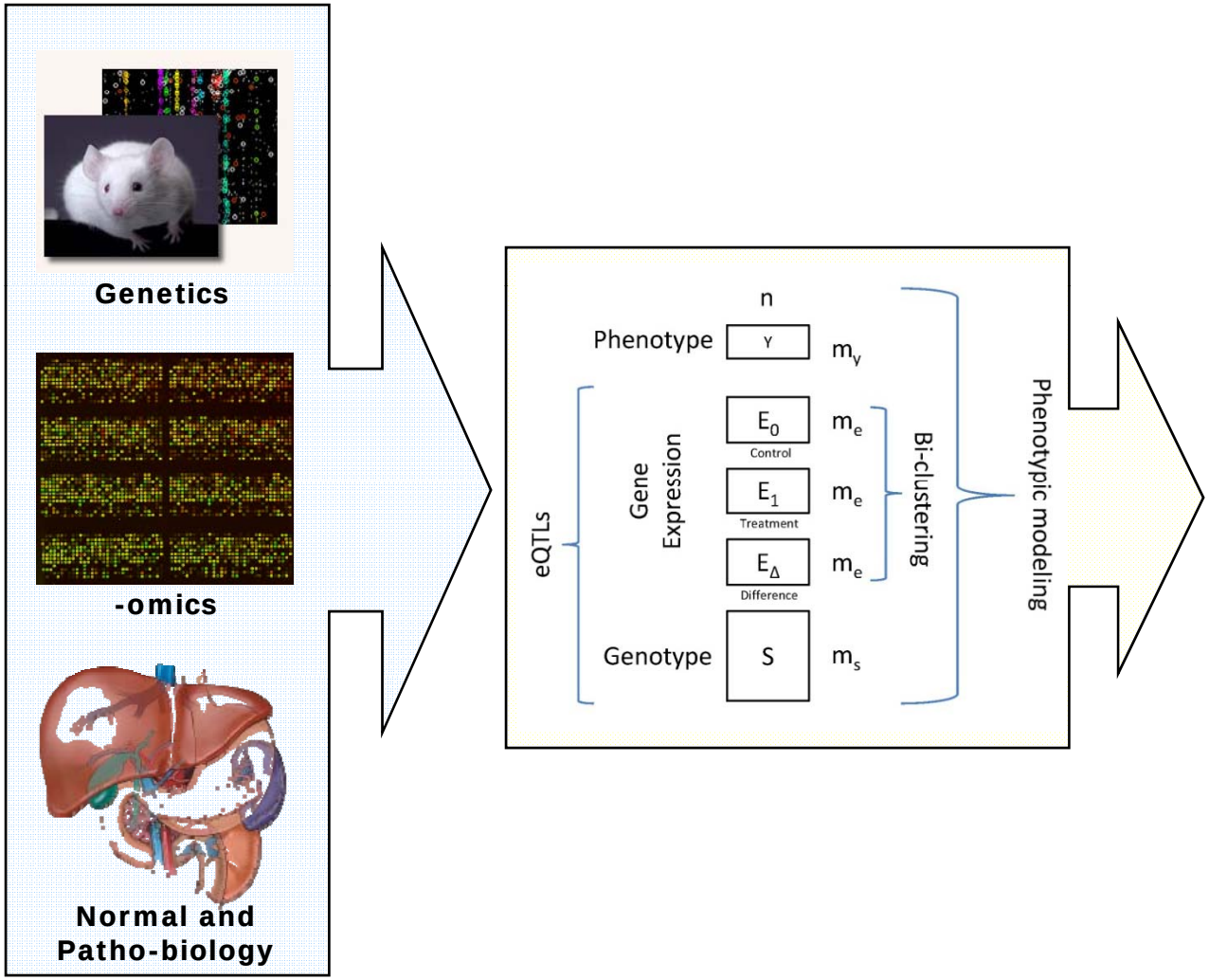
- Develop toxicogenetic expression Quantitative Trait Loci (eQTL) mapping tools, perform transcription factor network inference and integrative pathway assessment
- Perform toxicogenetic modeling of liver toxicity in cultured mouse hepatocytes
- Discover chemical-induced regulatory networks using population-based toxicity phenotyping in human cells

Major Interactions with the US EPA

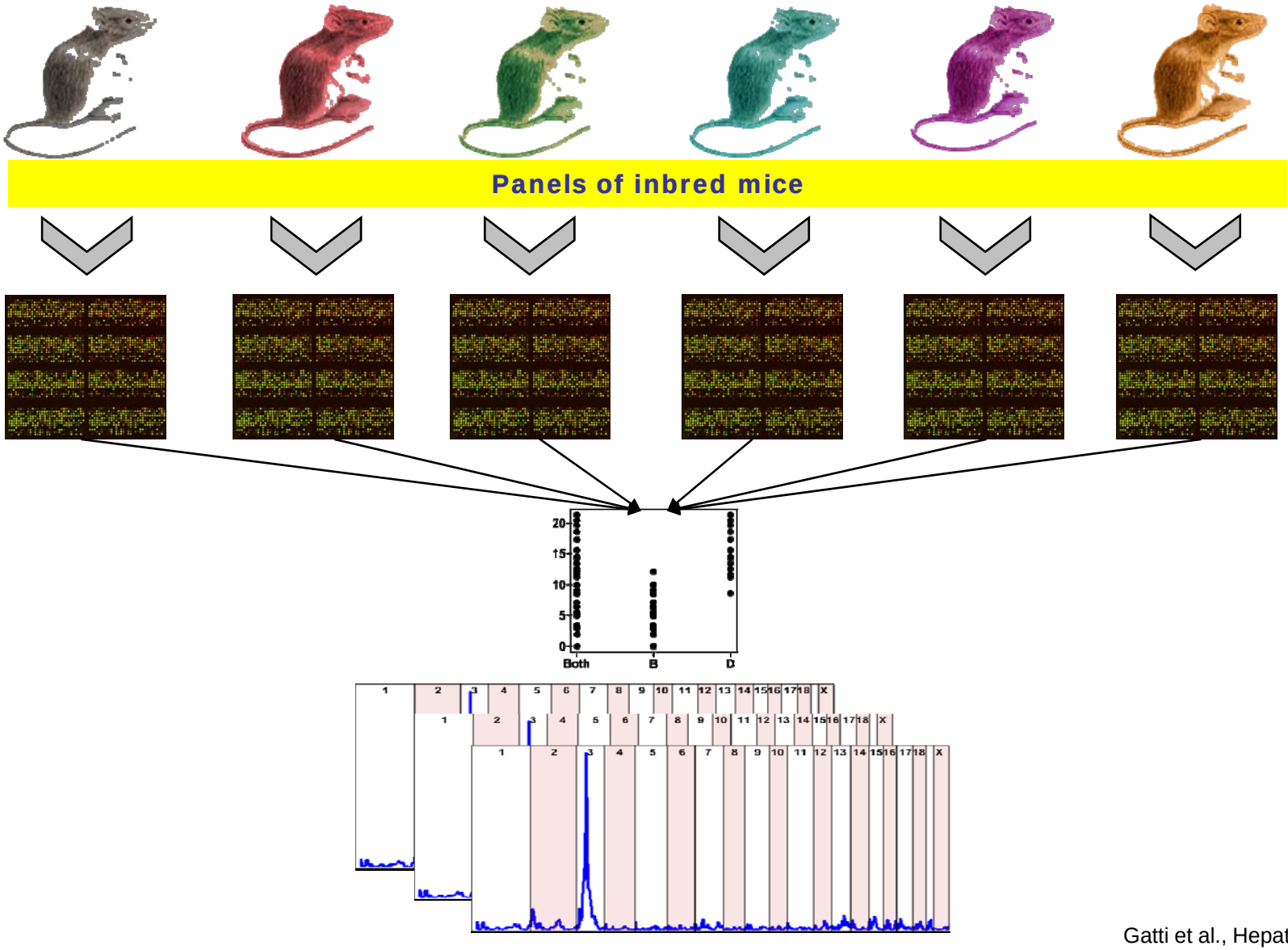
- Developing *in vitro* tools which will enable testing for inter-individual susceptibility: with David Dix (EPA-NCCT) and other Tox21 partners
- Developing statistical methodology and computational tools capable of processing higher-order multi-dimensional data: work relevant to future ToxCAST efforts and current Tox21 datasets
- ToxCAST: with Richard Judson (EPA-NCCT)

Population-wide predictions from toxicity profiling: linking toxicology with -omics and genetics

Data | **Analysis** | **Knowledge**

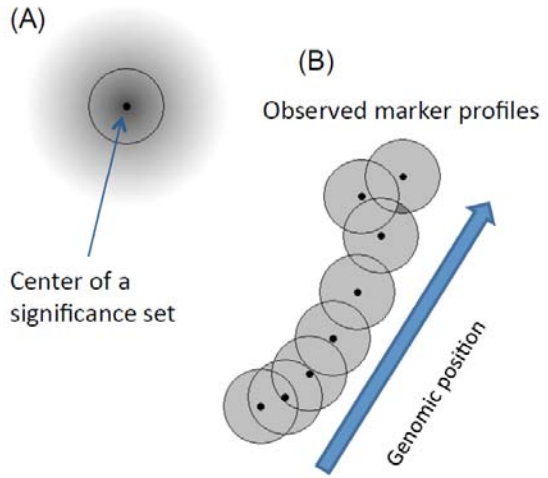


Genome-level analysis of genetic regulation of liver gene expression networks (eQTL mapping)

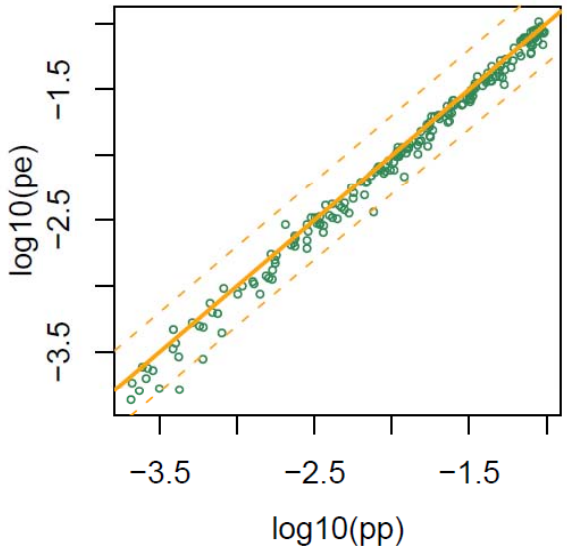


Gatti et al., Hepatology, 2007

Specific Objective 1: Development of Fast and Efficient Toxicogenetic Expression Quantitative Trait Loci (eQTL) Mapping Tools



Permutation p-value estimation



Fast methods to perform p-value-based eQTL inference

A geometric view of permutation p-values

- For each transcript, we imagine a hypersphere in the vicinity of the most significant possible genotype profile
- Permutations correspond to rotations of sets of observed genotypes within the space
- Significance thresholds determined by “volume” of space occupied by observed genotypes

Specific Objective 1: Development of Fast and Efficient Toxicogenetic Expression Quantitative Trait Loci (eQTL) Mapping Tools

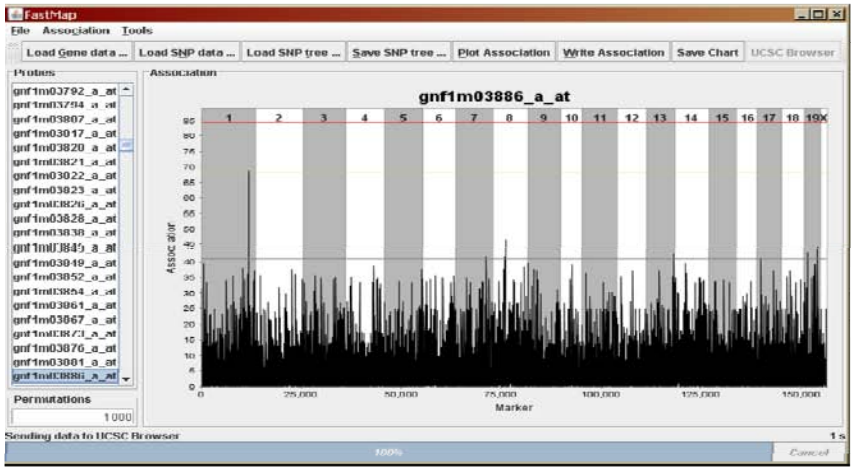
BIOINFORMATICS **ORIGINAL PAPER** *Vol. 25 no. 4 2009, pages 482–489*
doi:10.1093/bioinformatics/btn648

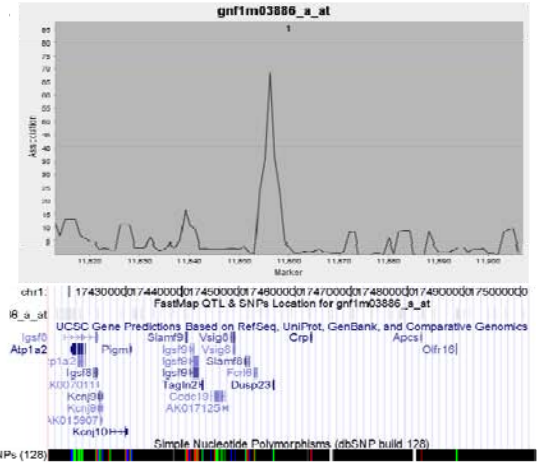
Gene expression

FastMap: Fast eQTL mapping in homozygous populations

Daniel M. Gatti[1,†], Andrey A. Shabalin[2,†], Tieu-Chong Lam[1], Fred A. Wright[3], Ivan Rusyn[1,*] and Andrew B. Nobel[2,3,*]

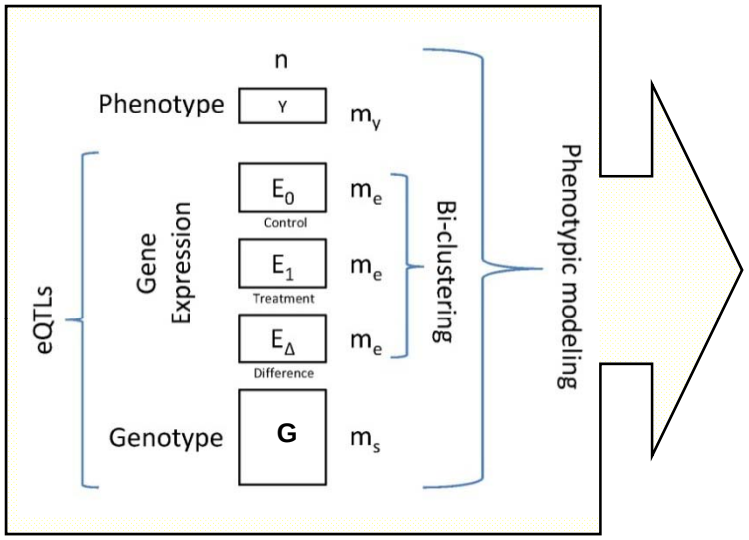
[1]Department of Environmental Sciences and Engineering, [2]Department of Statistics and Operations Research, and [3]Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina 27599, USA





- Java-based GUI which runs on a standard desktop PC
- Amenable to "proprietary" data
- Single marker or k-SNP window association mapping
- Permutation-based significance testing of the eQTLs
- Extended options for export of data/images and a link to UCSC genome browser

Specific Objective 1: Development of Fast and Efficient Toxicogenetic Expression Quantitative Trait Loci (eQTL) Mapping Tools



eQTL Studies 2.0

We should care about disease phenotype (susceptibility)

- Most genotype-transcript correlations are incidental
- We are interested in a small number of SNPs and transcripts with effects on phenotype
- This may be viewed as a huge variable selection problem

$$Y = \text{expression} + \text{genotype} + \text{genotype} \times \text{expression}$$

Specific Objective 1: Development of Fast and Efficient Toxicogenetic Expression Quantitative Trait Loci (eQTL) Mapping Tools

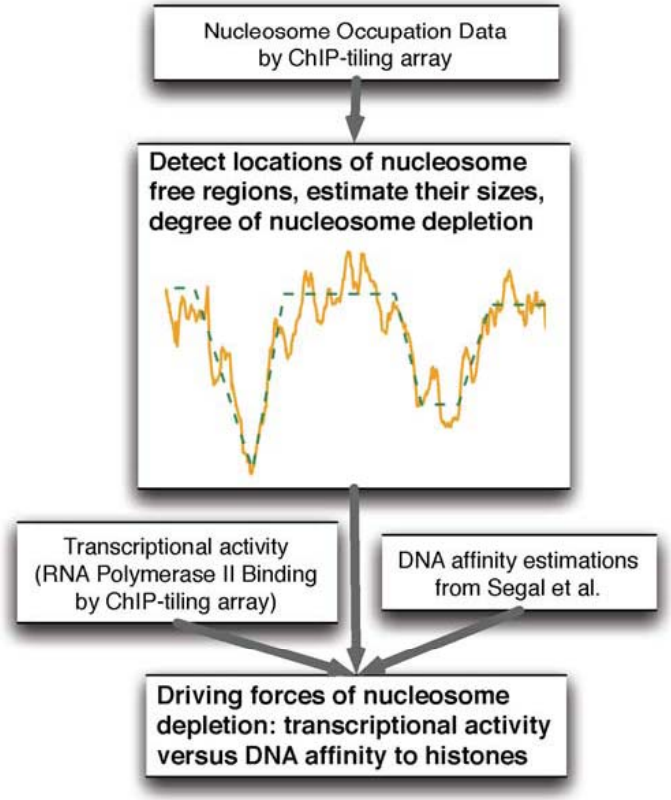
Understanding genomic context for expression

OPEN ACCESS Freely available online PLoS one

Dissecting Nucleosome Free Regions by a Segmental Semi-Markov Model

Wei Sun[1,2], Wei Xie[3,4], Feng Xu[3], Michael Grunstein[3]*, Ker-Chau Li[4,5]*

1 Department of Biostatistics, Carolina Center for Genome Science, University of North Carolina, Chapel Hill, North Carolina, United States of America, 2 Department of Genetics, Carolina Center for Genome Science, University of North Carolina, Chapel Hill, North Carolina, United States of America, 3 Department of Biological Chemistry, University of California Los Angeles, Los Angeles, California, United States of America, 4 Department of Statistics, University of California Los Angeles, Los Angeles, California, United States of America, 5 Institute of Statistical Science, Genomics Research Center, Academia Sinica, Taipei, Taiwan



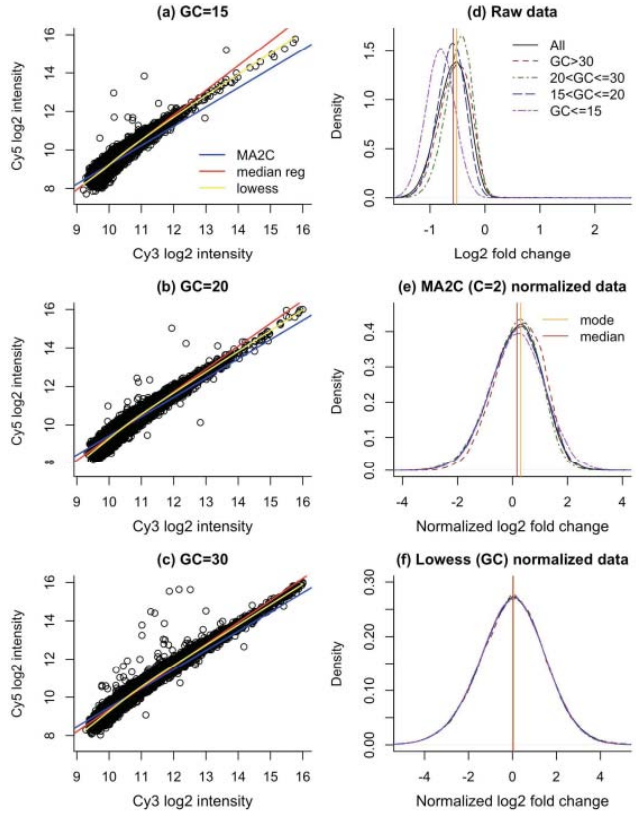
BMC Bioinformatics BioMed Central

Methodology article Open Access

Improved ChIP-chip analysis by a mixture model approach

Wei Sun*[1], Michael J Buck[2], Mukund Patel[3] and Ian J Davis*[3,4]

Address: [1]Department of Biostatistics, Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, [2]Department of Biochemistry, Center of Excellence in Bioinformatics and Life Sciences, State University of New York at Buffalo, Buffalo, NY, USA, [3]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA and [4]Department of Pediatrics, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA



PROJECT 3

Development of validated and predictive Quantitative Structure-Toxicity Relationship models that employ both chemical and biological descriptors of molecular structures and take into account genetic diversity between individuals

Alexander Tropsha – PI

Chair, Division of Medicinal Chemistry & Natural Products, UNC-Chapel Hill

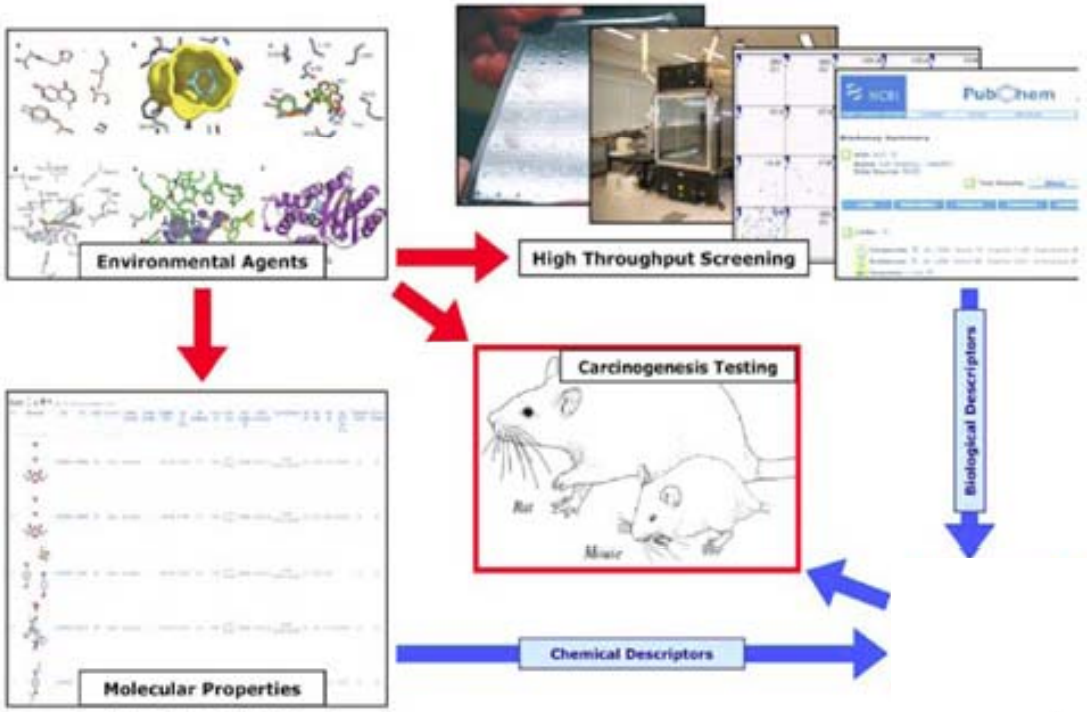
- Develop rigorous end point toxicity predictors based on the QSAR modeling workflow and conventional chemical descriptors
- Develop novel computational toxico-genomic models based on combined chemical and biological descriptors through QSAR modeling workflow
- Develop novel computational toxico-genetic models based on combined genetic, chemical and toxicity descriptors through QSAR-like modeling workflow

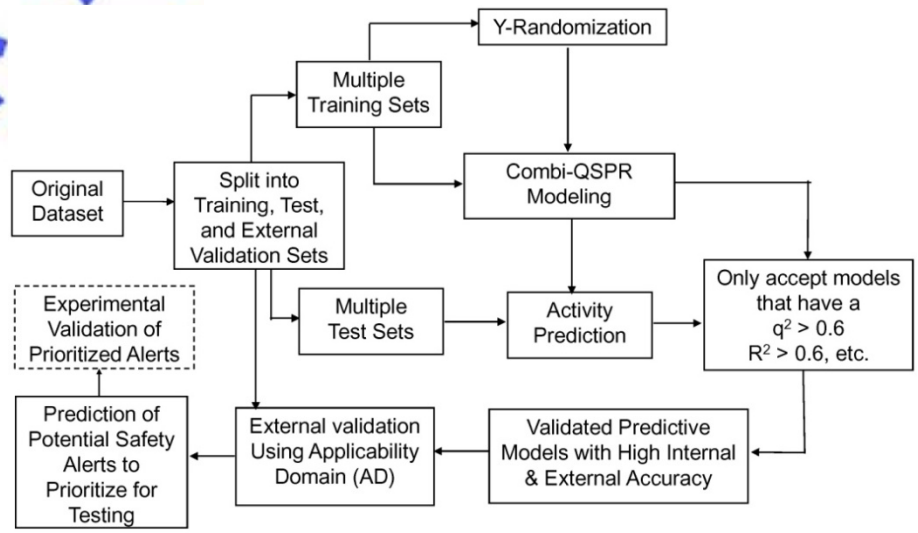
Major Interactions with the US EPA

- **Integrating chemical descriptors into DSSTox: with Ann Richard (EPA-NCCT)**
- **ToxCAST, ToxRefDB and ACToR data analysis: with Richard Judson (EPA-NCCT)**

Predictive Quantitative Structure-Toxicity Relationship Modeling

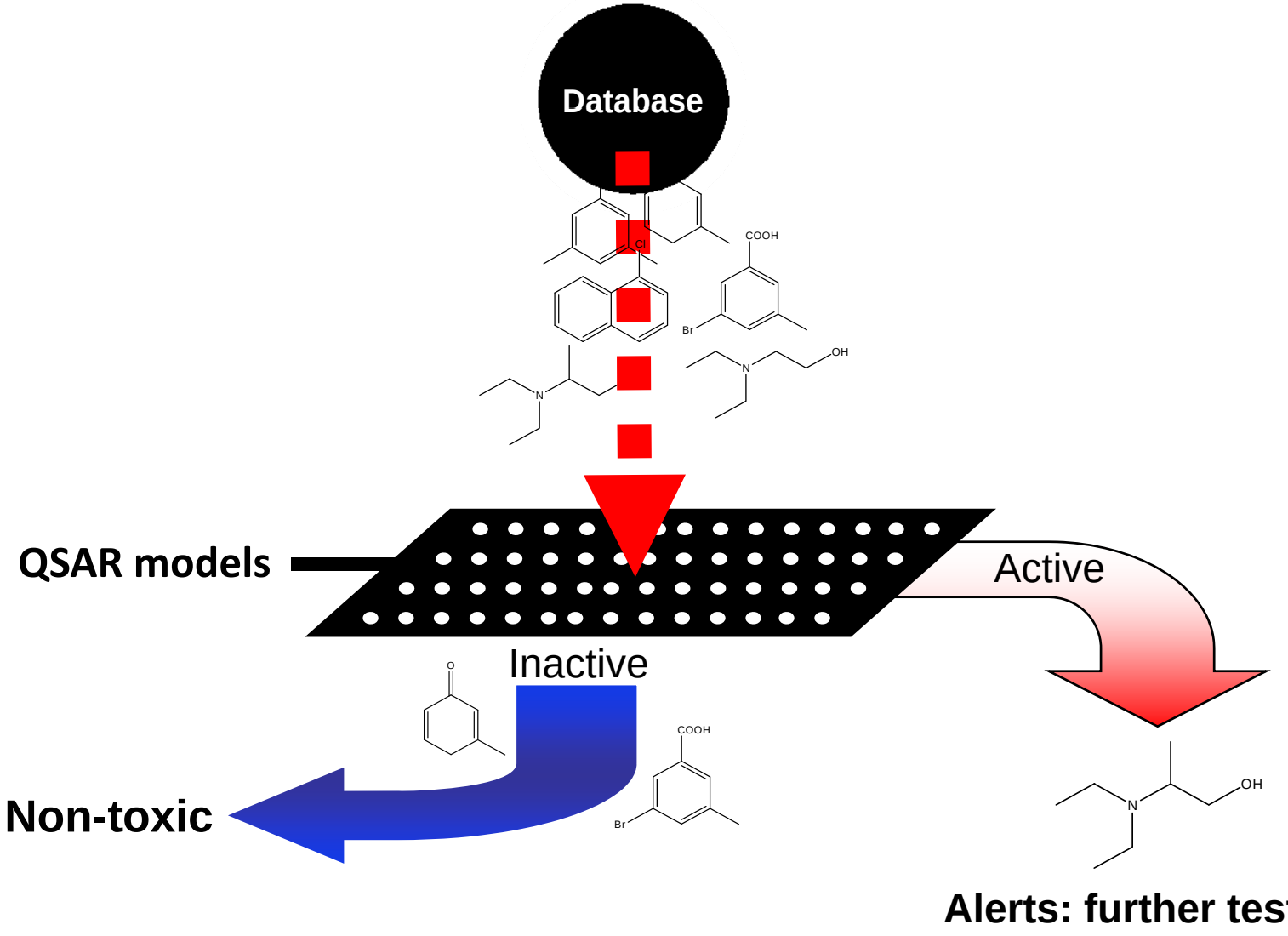
DSSTox
ACToR
ToxCast
ToxRefDB





Compound prioritization using the ensemble of QSAR models

DSSTox
ACToR
ToxCast
ToxRefDB



Data Curation

- ***In-vitro* assays: 524 → 353**
  - Remove one of two highly correlated (R^2 >0.95) assays and low-variance (<4 non-zero entries) assays

- **Chemicals: 320 → 228**
  - duplicate structures, mixtures, inorganic compounds, macromolecules were removed
  - Kept only those for which ***in-vivo*** data is available (i.e. chronic mouse toxicity)

Focusing on a small subset of data: Chronic Mouse Toxicity

- Continuity (overlaps with previous ToxRefDB data)
- Manageable (has only 7 *in-vivo* assays)
- 3 assays with the highest fraction of actives chosen for initial studies:

  CHR_Mouse_LiverProliferativeLesions (87 actives)

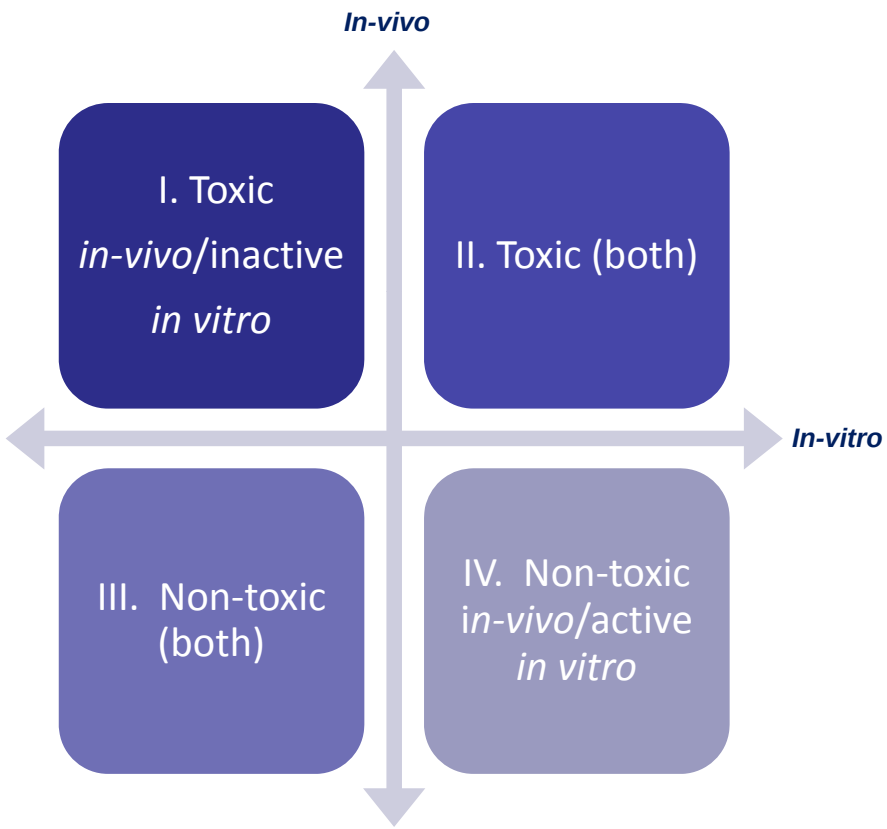
  CHR_Mouse_LiverTumors (68 actives)

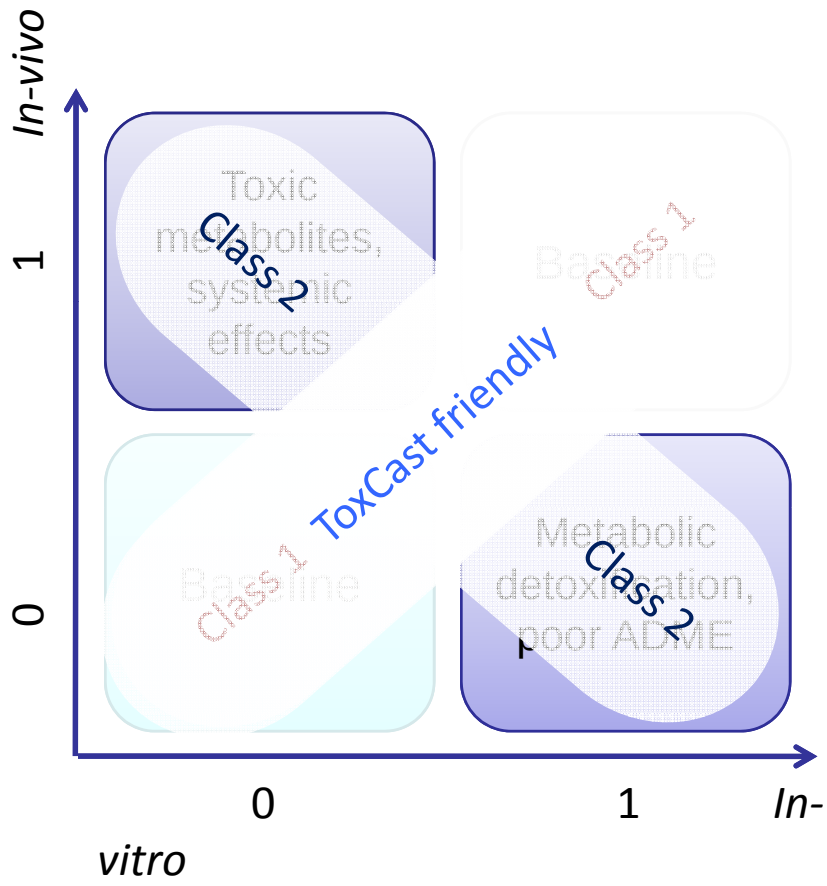
  CHR_Mouse_Tumorigen (88 actives)

Data partitioning based on *in vitro-in vivo* correlations as part of the QSAR Modeling workflow

DSSTox
ACToR
ToxCast
ToxRefDB

For <u>each</u> In-vitro vs. *In-vivo* profile (3 x 353 = 1059 combinations):



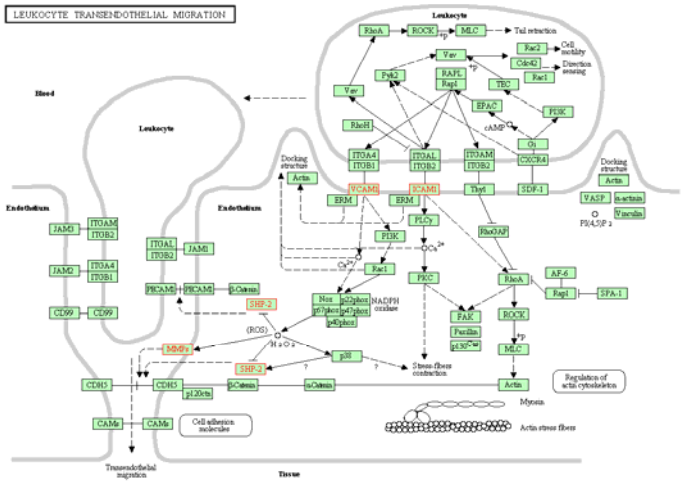


Binary classification QSAR for “baseline” (II & III) vs. off-line (I & IV) using chemical descriptors only

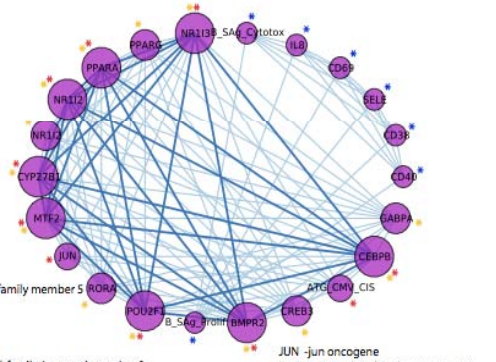
Developing Novel Bio-Descriptors

DSSTox
ACToR
ToxCast
ToxRefDB

Pathway-derived



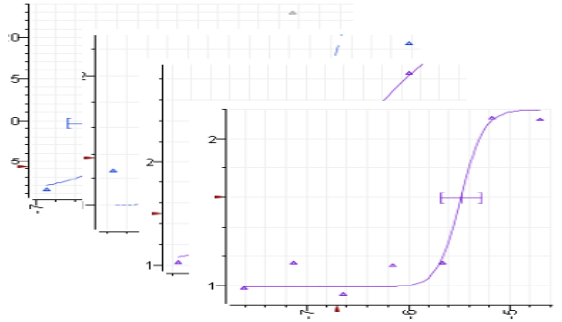
70*=> 7 assays
65*=>10 assays
64*=>13 assays
15 assays & 46 chemicals
22 assays & 24 chemicals

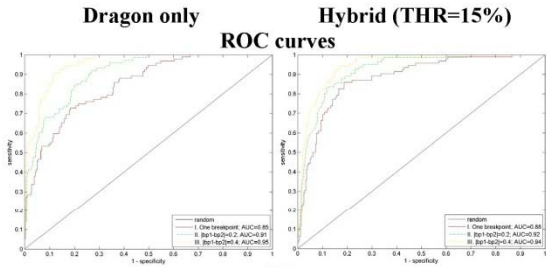


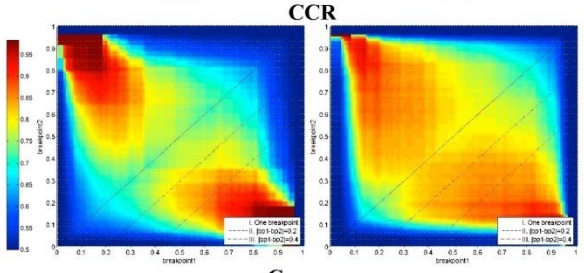
CD40 -TNF receptor superfamily member 5
CD38 -CD38 molecule
SELE -selectin E
CD69 -CD69 molecule
IL8 -interleukin 8
NR1I3 -nuclear receptor subfamily 1, group I, member 3
PPARG -peroxisome proliferator-activated receptor gamma
PPARA -peroxisome proliferator-activated receptor alpha
NR1I2 -nuclear receptor subfamily 1, group I, member 2
CYP27B1 -cytochrome P450, family 27, subfamily B, polypeptide 1
MTF2 -metal response element binding transcription factor 2

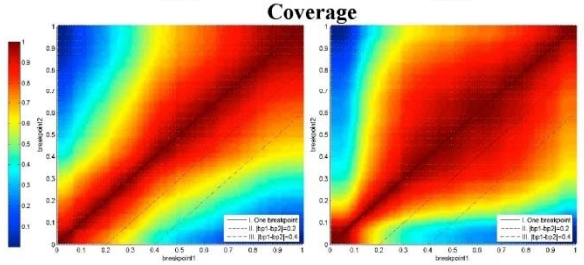
JUN -jun oncogene
RORA -RAR-related orphan receptor A
POU2F1 -POU class 2 homeobox 1
BMPR2 -bone morphogenetic protein receptor, type II
CREB3 -cAMP responsive element binding protein 3
CEBPB -CCAAT/enhancer binding protein (C/EBP), beta
GABPA -GA binding protein transcription factor, alpha

Dose-response-derived









- Focus on accurate prediction of external datasets is much more critical than accurate fitting of existing data:
  - consensus (collaborative!) prediction using all acceptable models
  - experimental validation of a small number of computational hits
  - outcome: decision support tools in selecting future experimental screening sets
- Neither cheminformatics nor HTS and –omics data alone is insufficient to achieve the desired accuracy of the end point property prediction
  - Integration of **chem**informatcs and **bio**informatics: predictive model s of selected endpoints using integrated short term biological profiles (biodescriptors) and chemical descriptors for compound subsets
  - New computational approaches (e.g., hybrid and hierarchical QSAR)
  - Interpretation of significant chemical and biological descriptors

Center publications in Year 1

- Choi K, and Gomez SM. (2009) BMC Bioinformatics (In revision)
- Staab J et al. (2009) BMC Bioinformatics (In revision)
- Gatti DM et al. (2009) Bioinformatics 4:482-489
- Sun W et al. (2009) PLoS One 4:e4721
- Zhu H et al. (2009) Envr Health Persp 117:1257-1264
- Gatti, DM et al. (2009) Mamm Genome 20:437-454
- Harrill AH et al. (2009) Tox Sci 110:235-243
- Sun W, and Wright FA (2009) Ann Appl Stat (accepted)
- Sun W et al. (2009) BMC Bioinformatics 10:173
- Zhu H et al. (2008) Environ. Health Persp 116: 506-513
- Zhu H et al. (2009) Chem Res Tox (In revision)
- Artemenko AG et al. (2009) Chem Res Tox (In revision)

Short-Term Goals for Year 2

Project 1:

- Continue in depth analysis of ToxCast Phase I data;
- Further refine the methods for integration across data types;
- Investigate the applicability of the metabolism model as a tool for the prediction of the effects of chemical perturbation of metabolic pathways;
- Integration of the eQTL analyses/approaches with the network-focused methodologies (with Proj. 2);
- Establish the network context for QSAR (with Proj. 3).

Project 2:

- Continue development of FastMap software;
- Construct transcription regulation networks in the Bayesian framework by combining eQTLs, nucleosome occupancy, and transcriptional regulation data;
- Complete characterization of the mouse hepatocyte cultures and perform experiments with key toxicants;
- Complete GWAS analyses of the HapMap lymphoblast cell viability and apoptosis data and correlate the toxicity endpoints with basal gene expression profiles.

Project 3:

- Complete the analysis of ToxCast data;
- Continue to explore other datasets that provide both *in vivo* and *in vitro* data for chemicals;
- Build models that could be used by EPA to prioritize the selection of ToxCast Phase 2 compounds.

Carolina Environmental Bioinformatics Research Center: *Collaborative work with EPA*

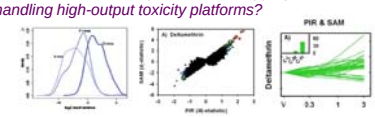
October 1, 2009 *Presented by Ann M. Richard*

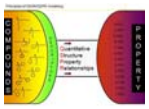
This work was reviewed by EPA and approved for presentation but does not necessarily reflect official Agency policy. Mention of trade names or commercial products does not constitute endorsement or recommendation by EPA for use.

NC Bioinformatics STAR Center

Project 1: Biostatistics for Computational Toxicology

What statistical techniques are most appropriate for handling high-output toxicity platforms?



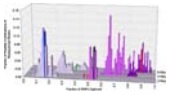


Project 2: Cheminformatics

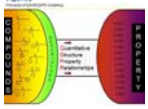
How can biological information & in vitro HTS data be incorporated into QSAR models?

Project 3: Computational Infrastructure for Systems Toxicology

What computational tools are necessary for these and related questions arising in model organism toxicity research?



NC Bioinformatics STAR Center

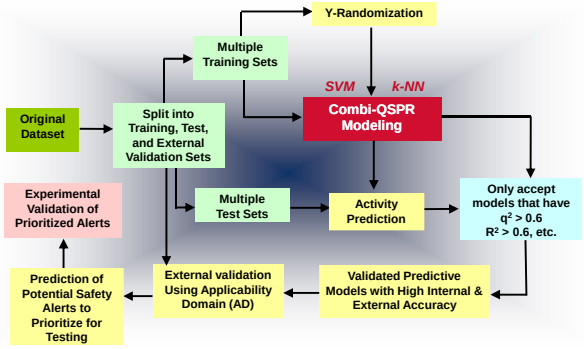


Project 2: Cheminformatics

How can biological information & in vitro HTS data be incorporated into QSAR models?

- Ability to generate thousands of QSAR descriptors representing categories of structure-based computed properties (DRAGON):
  - *Electronic, topological, constitutional, geometrical*
  - *feature counts, functional groups, 2Dfingerprints, etc.*
- Sophisticated QSAR workflow:
  - *kNN & sphere exclusion methods*
  - *Randomized y variable test*
  - *External test set validation*
  - *Consensus models*

Predictive QSAR Workflow*



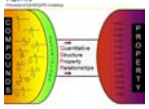
**Tropsha, A.,* Golbraikh, A. Predictive QSAR Modeling Workflow, Model Applicability Domains, and Virtual Screening. Curr. Pharm. Des., 2007, 13, 3494-3504.*

Chemical Descriptors (DRAGON):

- Computed from 2D molecular structures provided in DSSTox SDF files
- Can use selected categories, or all
- Provide different representations of chemical space in relation to activity
- Different degrees of interpretability

1	constitutional descriptors	48
2	topological descriptors	119
3	walk and path counts	47
4	connectivity indices	33
5	information indices	47
6	2D autocorrelations	96
7	edge adjacency indices	107
8	Burden eigenvalue descriptors	64
9	topological charge indices	21
10	eigenvalue-based indices	44
11	Randic molecular profiles	41
12	geometrical descriptors	74
13	RDF descriptors	150
14	3D-MoRSE descriptors	160
15	WHIM descriptors	99
16	GETAWAY descriptors	197
17	functional group counts	154
18	atom-centered fragments	120
19	charge descriptors	14
20	molecular properties	29
21	2D binary fingerprints	780
22	2D frequency fingerprints	780
	TOTAL	3224

NC Bioinformatics STAR Center

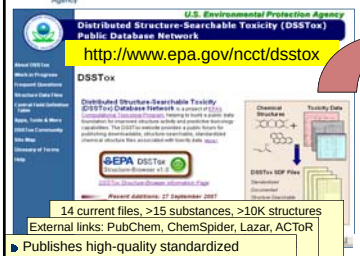


Project 2: Cheminformatics

How can biological information & in vitro HTS data be incorporated into QSAR models?

- QSAR models based on DSSTox published data files and structure-inventories
- Share processed data files and calculated descriptors with EPA researchers for public release
- Coauthored publications:
  - *Zhu H, Rusyn I, Richard A, Tropsha A. (2008) Use of cell viability assay data improves the prediction accuracy of conventional quantitative structure-activity relationship models of animal carcinogenicity. Environ. Health Perspect. **116**: 506-513.*
  - *Zhu H, Ye L, Richard A, Golbraikh A, Rusyn I, Tropsha A. (2009) A Two-step Hierarchical Quantitative Structure Activity Relationship Modeling Workflow for Predicting in vivo Chemical Toxicity from Molecular Structure. Environ. Health Perspect. **117**:1257-1264.*

DSSTox: Distributed Structure-Searchable Database Network Project…& NC CEBC



http://www.epa.gov/ncct/dsstox

EPA

- CPDBAS (Rodent carcinogenicity data); NTPHTS cytotox assays
- ZEBET acute tox data (to be published)
- ToxCast Phase I chemical inventory (TOXCST)
- ToxRefDB in vivo endpoints for modeling

14 current files, >15 substances, >10K structures

External links: PubChem, ChemSpider, Lazar, ACToR

- Publishes high-quality standardized structure-data (SD) files pertaining to toxicology:
  - *EPA, HPV-IS, IRIS, NTP, FDA, NCBI, EBI…*
- SAR-ready summary tox data for modeling
- Public substance/structure ID registry system
- Public forum for SAR file/data sharing
- DSSTox Structure-Browser

- Processed data sets (ZEBET acute tox)
- Calculated chemical descriptors (DRAGON) for ToxCast inventory

UNC CEBC

kNN Consensus QSAR Modeling of NTP-HTS Data

NTP HTS 1408 | **Carcinogenic Potency Database 1481**

270

Chemical Descriptors Only
(9 models > 0.7 validation cutoff)

Chemical + 7 HTS "Descriptors"
(34 models > 0.7 validation cutoff)

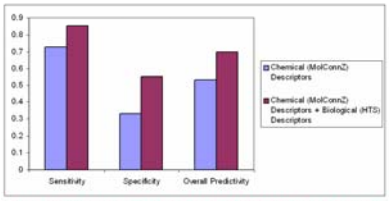
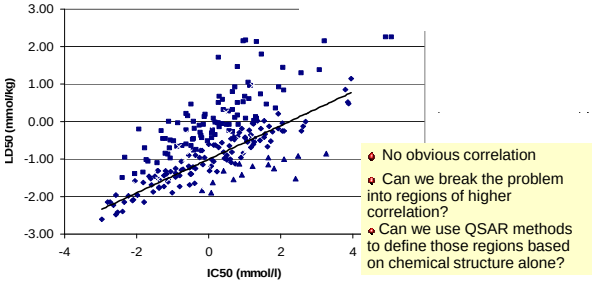


Figure 3. Comparison of the results from kNN QSAR models using two types of descriptors.

UNC CEBC

*Zhu H, Rusyn I, Richard A, Tropsha A. (2008) EHP **116**: 506-513.*

Can *in vitro* IC50 data be used to inform development of model for *in vivo* Rat Oral LD50?



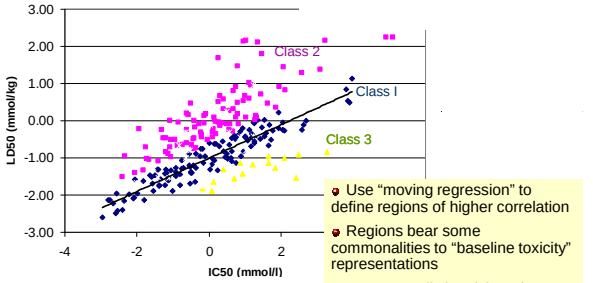
- No obvious correlation
- Can we break the problem into regions of higher correlation?
- Can we use QSAR methods to define those regions based on chemical structure alone?

Office of Research and Development
National Center for Computational Toxicology

8

*Zhu H, Ye L, Richard A, Golbraikh A, Rusyn I, Tropsha A. (2009) EHP **117**:1257-1264.*

Can *in vitro* IC50 data be used to inform development of model for *in vivo* Rat Oral LD50?

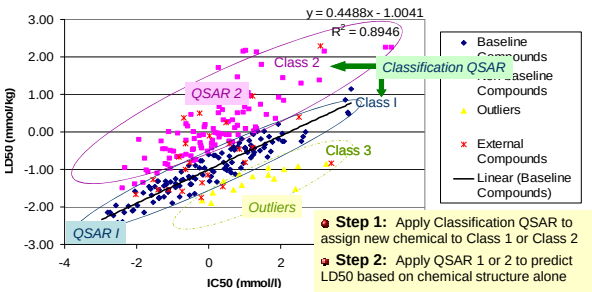


- Use "moving regression" to define regions of higher correlation
- Regions bear some commonalities to "baseline toxicity" representations
- Attempt to distinguish regions based on chemical structure alone

Office of Research and Development
National Center for Computational Toxicology

*Zhu H, Ye L, Richard A, Golbraikh A, Rusyn I, Tropsha A. (2009) EHP **117**:1257-1264.*

Can *in vitro* IC50 data be used to inform development of model for *in vivo* Rat Oral LD50?

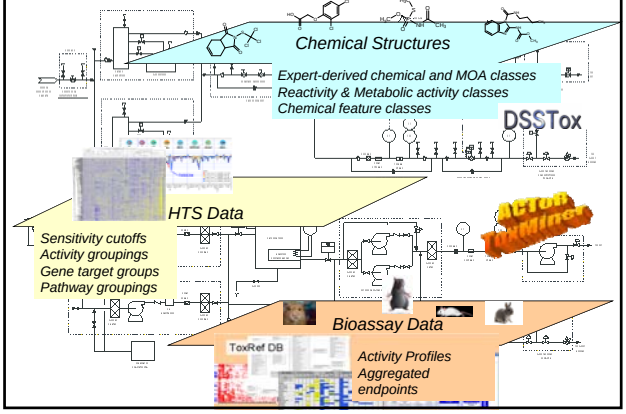


- **Step 1:** Apply Classification QSAR to assign new chemical to Class 1 or Class 2
- **Step 2:** Apply QSAR 1 or 2 to predict LD50 based on chemical structure alone
- **Step 3:** Validate approach with external data

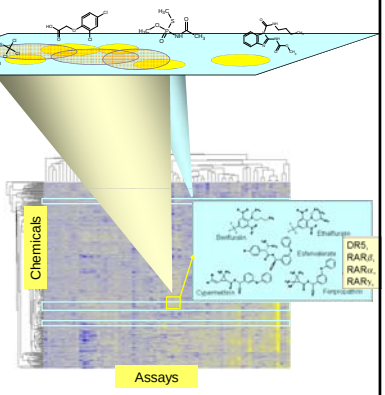
IC50 used to inform construction of QSARs, but not needed for prediction

*Zhu H, Ye L, Richard A, Golbraikh A, Rusyn I, Tropsha A. (2009) EHP **117**:1257-1264.*

ToxCast & Tox21: High-Multi-Dimensional Data



Structure Class vs Bioactivity Class



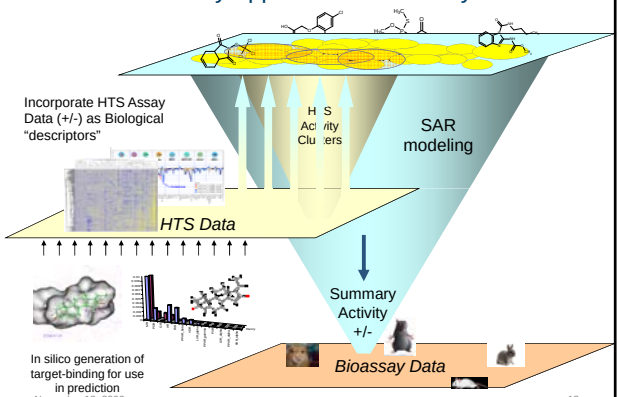
Chemical structure class:

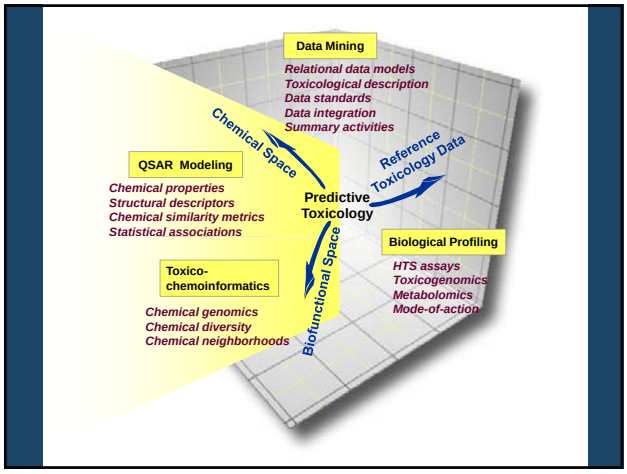
- Cluster according to activity and mechanism
- Differences in activity profiles can discriminate within structure class

Bioactivity profile class:

- Can project onto multiple chemical classes
- Potentially broader coverage of chemical space
- Implies mechanistic similarity

Structure-Activity Approaches to Toxicity Prediction





ebCTC environmental bioinformatics and Computational Toxicology Center

Environmental Bioinformatics and Computational Toxicology Center (ebCTC): Research in Multiscale Modeling of the Effects of Environmental Toxicants

William J. Welsh and Panos G. Georgopoulos

www.ebCTC.org

Presented at USEPA Computational Toxicology Centers Progress Review Workshop
Research Triangle Park, NC - October 1, 2009

Funded by the USEPA Science to Achieve Results (STAR) Grant RD-83272101

Consortium Members

UMDNJ Robert Wood Johnson Medical School
- Computational Chemodynamics Laboratory, Environmental & Occupational Health Sciences Institute
- Department of Environmental & Occupational Medicine
- Department of Pharmacology
- Informatics Institute

Rutgers The State University of New Jersey
- Department of Biomedical Engineering
- Department of Chemical & Biochemical Engineering
- Department of Environmental Sciences
- Department of Statistics

Princeton University
- Computer Aided Systems Laboratory, Department of Chemical Engineering
- Department of Chemistry
- Program in Applied and Computational Mathematics

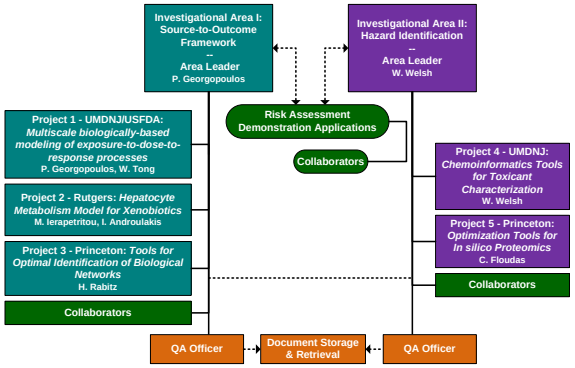
FDA U.S. Food and Drug Administration
- Center for Toxicoinformatics, National Center for Toxicological Research

ebCTC objectives and general approach

- **Objectives**
  - To address toxicant *Source-to-Outcome Continuum* through development of an integrated, modular, computational framework
  - To develop predictive cheminformatics tools for *Hazard Identification and Toxicant Characterization*
  - To demonstrate the above tools through applications in *Quantitative Risk Assessment*
- **General Approach**
  - A computational/engineering/systems perspective
    - *utilizing a team of computational scientists and engineers, with diverse backgrounds in bioinformatics, cheminformatics, and enviroinformatics*
  - New framework and tools build upon an extensive base of past developments
  - The research effort emphasizes interaction and collaboration
    - *among participating scientists in the STAR Bioinformatics Centers*
    - *with USEPA centers and laboratories*
    - *with other centers and institutes of excellence*

ebCTC research activities organized in 2 research areas and 5 projects

- Each project is developing a set of "stand-alone" components addressing specific CT problems
- Research Project 1 provides an integrative framework for Investigational Area 1
- Project 4 addresses the core issues of Area 2



The ebCTC research integration plan is consistent with the "Vision and Strategy" outlined by NAS in 2007 for Toxicity Testing in the 21st Century

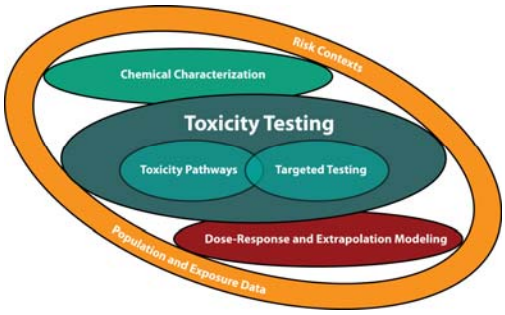


Figure adapted from NAS (2007) *Toxicity Testing in the 21st Century: A Vision and a Strategy*

The ebCTC research integration plan is consistent with the "Vision and Strategy" outlined by NAS in 2007 for Toxicity Testing in the 21st Century

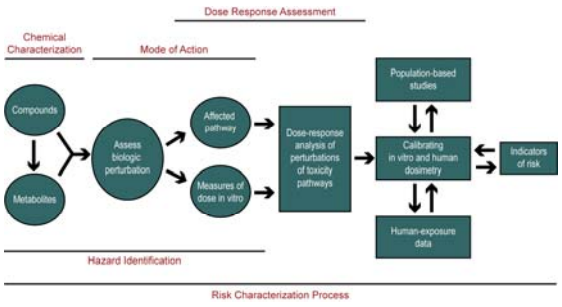
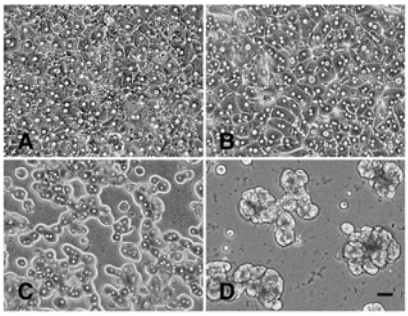


Figure adapted from NAS (2007) *Toxicity Testing in the 21st Century: A Vision and a Strategy*

ebCTC pursues an integrative multiscale research approach (from molecules to cells to tissues to organs to organisms to populations) recognizing the importance of processes/signals at all levels of biological organization

Primary cultures of human hepatocytes maintained for 6 days under different matrix conditions



Cell morphology depends on the topology and composition of the matrix environment

(A) Rigid collagen, type I, substratum with no overlay.
(B) Rigid collagen, type I, substratum with a Matrigel overlay.
(C) Gelled collagen, type I, substratum with a collagen overlay.
(D) Matrigel substratum. Bar 50 µm

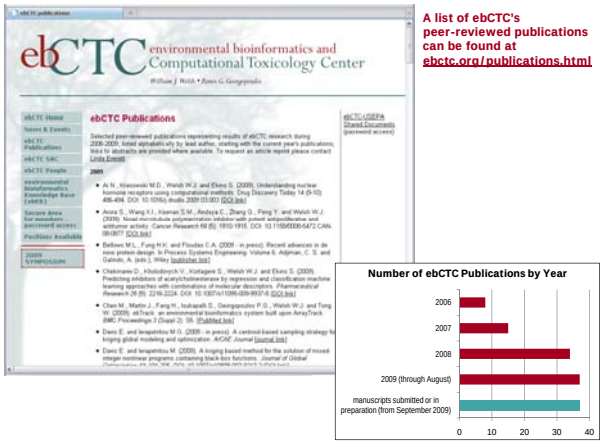
From Hamilton et al. (2001) *Cell Tissue Res* 306:85–99

A representative sample of USEPA/ebCTC project interactions and collaborations

- Toxicogenomic analysis of phthalate exposure data
  S. Euling, B. Benson, W. Chiu, L.E. Gray, S. Hester, C. Keshava, N. Keshava, S. Makris, C. Thompson, V.Wilson (USEPA);
  I. Androulakis, M. Ovacik, M. Ierapetritou (ebCTC)
- Toxicogenomic analysis of conazole exposure data and in vitro species extrapolation in primary hepatocytes
  S. Hester, D. Wolf, W. Ward (USEPA);
  M. Ierapetritou, P. Georgopoulos, I. Androulakis, W. Welsh, V. Iyer (ebCTC)
- Development of integrated PBPK/PD models for Arsenic and its compounds
  E. Kenyon, H. El Masri (USEPA);
  S. Isukapalli, P. Georgopoulos, C. Brinkerhoff, A. Sasso, S. Stamatelos, I. Androulakis, M. Ovacik (ebCTC)
- Computational tools for reconstructing exposures from biomarkers
  M. Tornero-Velez, C. Dary, D. Vallero, L. Reiter (USEPA);
  S. Isukapalli, P. Georgopoulos, A. Sasso (ebCTC)
- Incorporating the effects of aging on Physiologically Based Toxicokinetic (PBTK) models
  M. Tornero-Velez, M. DeVito, E. Kenyon, M. Evans (USEPA);
  P. Georgopoulos, S. Isukapalli, A. Sasso (ebCTC)
- Optimal analysis of proteomic data
  M. Hemmer, C. Walker (USEPA);
  C. Floudas (ebCTC)

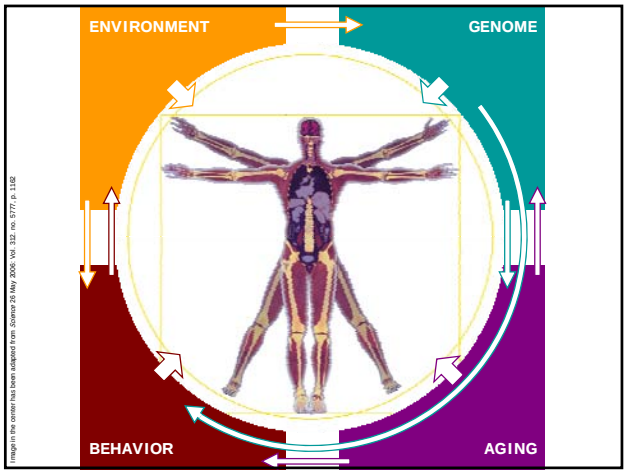
- Computational modeling of cellular signaling pathways: Implications for dose-response (modular infrastructure for virtual organs – liver, skin)
  I. Shah, R. Judson (USEPA); P. Georgopoulos, S. Isukapalli, M. Ierapetritou, I. Androulakis, C. Brinkerhoff, C. Roth, W. Welsh, H. Rabitz (ebCTC)
- Interactions of ToxCast chemicals with liver Nuclear Receptors
  R. Judson, D. Dix, I. Shah (USEPA); S. Mani, W. Welsh (ebCTC)

A list of ebCTC's peer-reviewed publications can be found at ebctc.org/publications.html



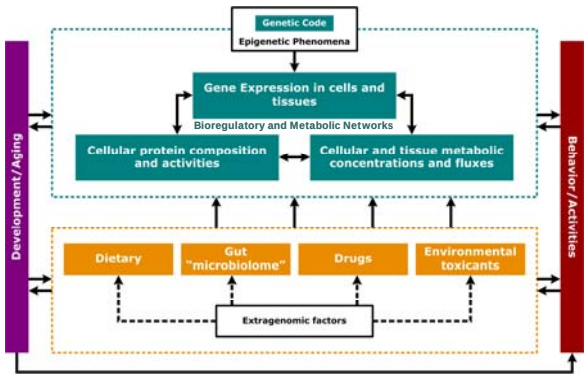
Research Area I:

A Source-to-Outcome Framework to Support Risk Characterization

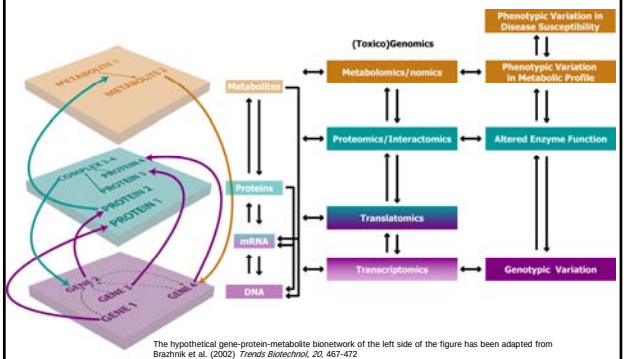


A general mathematical framework for environmental health risk analysis must consider multiscale bionetwork dynamics

(spanning the genome, transcriptome, proteome, metabolome, cytome, physiome) linked with the dynamics of environmental ("extragenomic") stressor networks

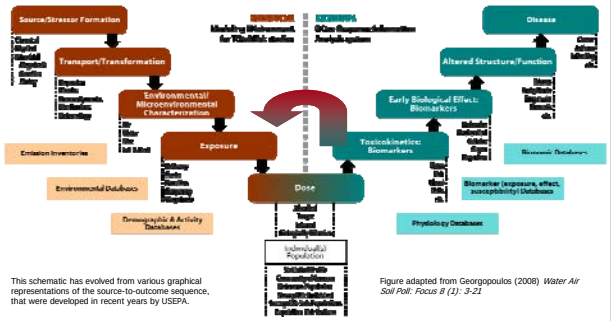


Connecting genotypes with phenotypes to assess toxicokinetic and toxicodynamic variability - and associated disease susceptibility - to environmental agents requires integrating data/information across multiple biological levels



The hypothetical gene-protein-metabolite bionetwork of the left side of the figure has been adapted from Brazhnik et al. (2002) *Trends Biotechnol, 20,* 467-472

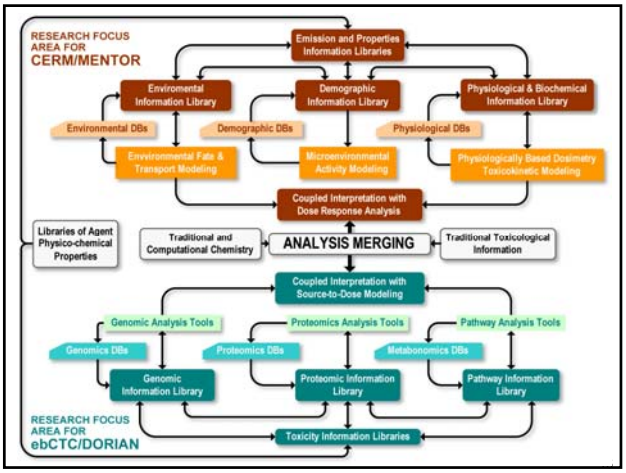
Research to address the toxicant "Source-to-Effect Continuum" through development of an integrated, modular, computational framework



This schematic has evolved from various graphical representations of the source-to-outcome sequence, that were developed in recent years by USEPA.

Figure adapted from Georgopoulos (2008) *Water Air Soil Poll: Focus 8 (1): 3-21*

CERM: Center for Exposure and Risk Modeling
MENTOR: Modeling ENvironment for TOtal Risk studies (development started in 1993 with CDC funding; USEPA funding commenced in 1998)
ebCTC: environmental bioinformatics and Computational Toxicology Center
DORIAN: DOse-Response Information Analysis system (development started in 2006 with USEPA STAR funding; consortium of UMDNJ-RWJMS with Rutgers, Princeton and USFDA)



A general Bayesian framework for exposure reconstruction from inversion of biomarker data (individual and population)

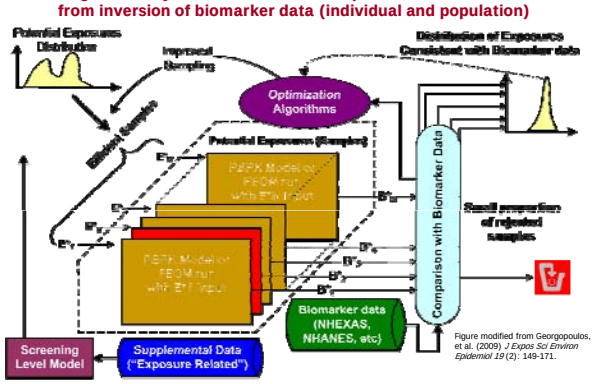


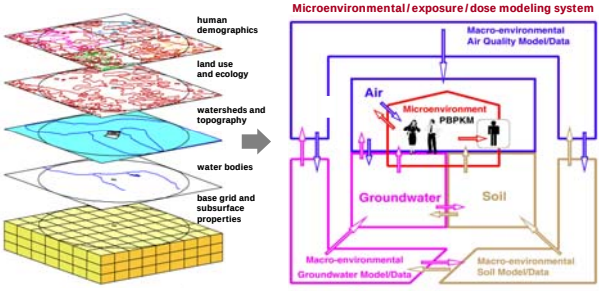
Figure modified from Georgopoulos, et al. (2009) *J Expos Sci Environ Epidemiol 19* (2): 149-171.

Novel methods have been developed that allow the systematic construction of Fast Equivalent Operational Models (FEOMs); these include the Stochastic Response Surface Method (SRSM) and the High Dimensional Model Representation (HDMR)

A "sample" of on-going applications within Research Area I of ebCTC (including various Risk Assessment Demonstration applications)

Air Contaminant Applications	Multimedia Applications
▪ urban/local/personal scale inhalation exposures to complex mixtures of co-occurring ozone, PM, other criteria pollutants, and air toxics, ▪ exposures to contaminant releases from forest and urban fires, ▪ exposures to contaminant releases from chemical facility accidents, ▪ exposures to bioaerosols (ranging from anthrax spores to birch and ragweed pollen), ▪ etc.	▪ exposures to mixtures of metals and metalloids (Hg, Cd, Cu, As, etc.) and their compounds, ▪ exposures to pesticides (organophosphates, conazoles), ▪ exposures to organic solvents, ▪ exposures to water chlorination by-products, ▪ exposures to phthalates, ▪ exposures to PCBs and dioxin-like compounds, ▪ exposures to CWAs, ▪ etc.

MENTOR employs an "anthropocentric" (person-oriented) approach, linking multiple scales of macroenvironmental and local models and information with microenvironmental conditions and human activities in time/space

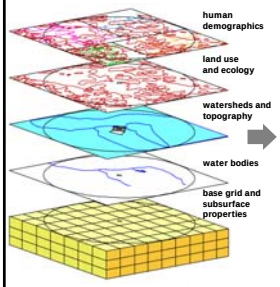


Source: 3MRA User Guide 2002

Source: Georgopoulos et al., ES&T, 1997, 31(1)

DORIAN aims to provide multiscale integration of toxicokinetic and toxicodynamic processes (cell to organism) for dose-to-response studies

MENTOR employs an "anthropocentric" (person-oriented) approach, linking multiple scales of macroenvironmental and local models and information with microenvironmental conditions and human activities in time/space

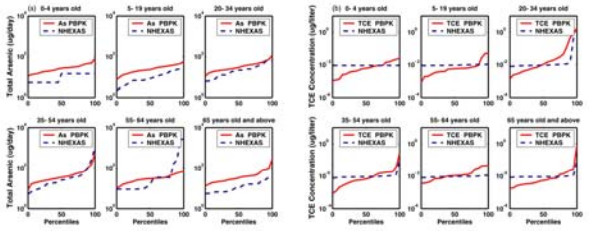


Source: 3MRA User Guide 2002

Human activities determine pathways of exposure

Animation source: EA Games - The Sims™

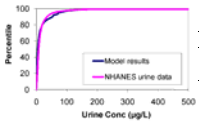
Example: Cumulative distributions of total As (left) and TCE (right) in urine from MENTOR predictions for Franklin County, OH compared with the measurements from NHEXAS-V (corresponding percentiles) for different age groups



Left panel from: Georgopoulos, et al. (2008) *J Expo Sci Environ Epidemiol 18 (5): 462-476. Right panel from unpublished data*

10^{-2} µg/L is the detection limit for TCE in the NHEXAS measurements

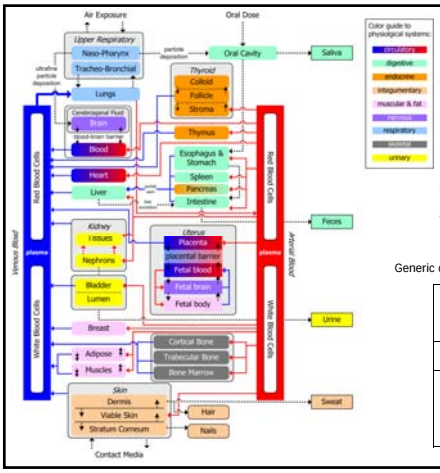
Comparison of total As concentration levels in urine samples of the NHANES population with corresponding MENTOR predictions



Comparison of total As in Urine (µg/L) from NHANES data and PBPK model

	N	Mean	Std	50th	25th	75th	95th
PBPK model	2487	18.32	46.86	8.1	4.7	16.1	58.9
Measured conc.	3166	18.06	42.12	4.89	2.5	14.64	74.84

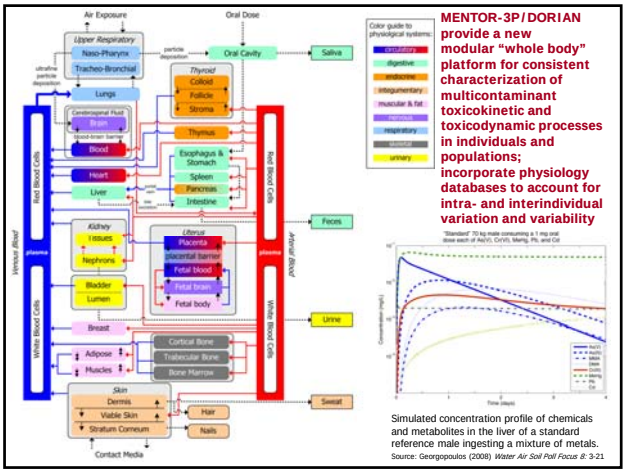
From: Xue, et al. (2009). Manuscript submitted to *Environ Health Perspect*



MENTOR-3P/DORIAN provide a new modular "whole body" platform for consistent characterization of multicontaminant toxicokinetic and toxicodynamic processes in individuals and populations; incorporate physiology databases to account for intra- and interindividual variation and variability

Generic compartmental substructure

capillary	blood cells
	plasma
interstitial space	
tissue cells	nonspecific binding
	specific binding

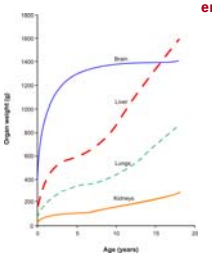


MENTOR-3P/DORIAN provide a new modular "whole body" platform for consistent characterization of multicontaminant toxicokinetic and toxicodynamic processes in individuals and populations; incorporate physiology databases to account for intra- and interindividual variation and variability

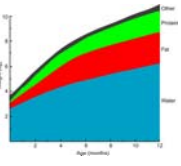
Simulated concentration profile of chemicals and metabolites in the liver of a standard reference male ingesting a mixture of metals.

Source: Georgopoulos (2008) *Water Air Soil Poll Focus 8*: 3-21

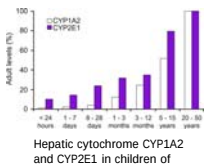
Individual and population human biology (physiology and biochemistry) changes non-uniformly with development, aging, disease, drug treatment, diet, environmental exposures, etc.



Organ weight from birth to adolescence in boys (based on Haddad et al. 2001)



Weights of water, fat, protein, and other components as a function of age, from birth to one year of age. [Figure reproduced from Fomon (1966) with permission from W.B. Saunders Co.]

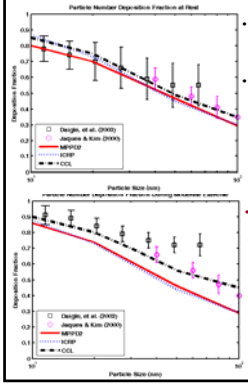


Hepatic cytochrome CYP1A2 and CYP2E1 in children of various age groups as a percentage of adult weights (from Cresteil, 1998).

Example references:

- WHO (2006). *Principles for Evaluating Health Risks in Children Associated with Exposure to Chemicals.* World Health Organization. Environmental Health Criteria 237;
- USEPA (2006) *Use of PBPK models to quantify the impact of human age and interindividual differences in physiology and biochemistry pertinent to risk;*
- Thompson, et al. (2009) *J Toxicol Environ Health B 12*: 1-24.

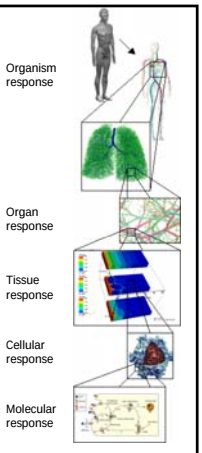
MENTOR/DORIAN offes a "whole organism" modular toxicokinetic/toxicodynamic platform for incorporating organ/tissue representations at various levels of detail (on-going projects focus on lung, skin and liver)



- Data from Jaques & Kim (2000) and Daigle, et al. (2003) studies at rest and during moderate exercise
- Experimental data compared to model predictions using MPPD2, ICRP, and (HUMTRN-derived) module of MENTOR-3P; experimental conditions used as model inputs

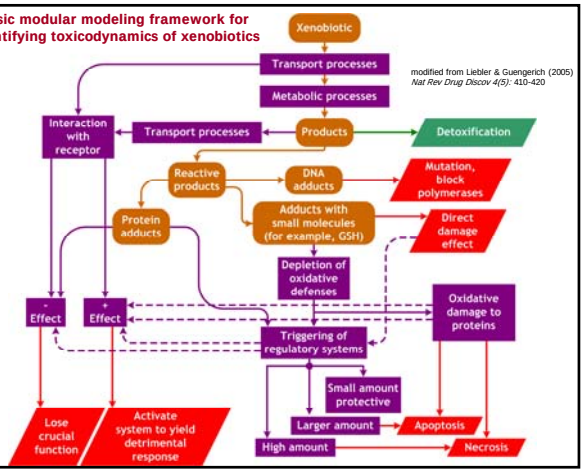
- Incorporation of "virtual organs" in MENTOR will support the **evolution from Physiologically Based Pharmacokinetic models to integrative Physiologically Based Pharmacokinetic/Pharmacodynamic models**

From Georgopoulos (2008) *Water Air Soil Poll: Focus 8* (1): 3-21



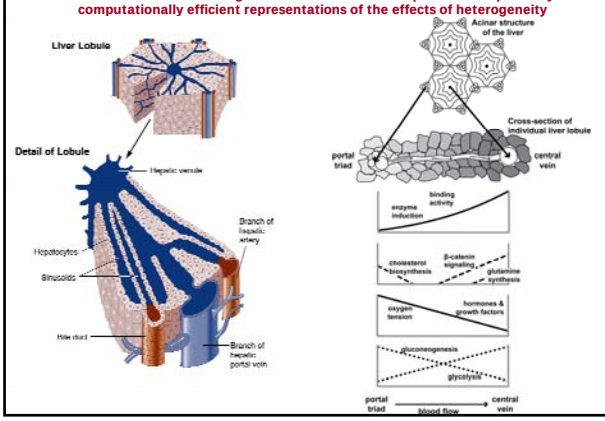
(Graphics from Physiome Project)

Basic modular modeling framework for quantifying toxicodynamics of xenobiotics



modified from Liebler & Guengerich (2005) *Nat Rev Drug Discov 4(5):* 410-420

Lung and skin models are critical for assessing exposure and intake/uptake; liver is critical for biotransformation and elimination of xenobiotics: recent/current liver modeling efforts in MENTOR development focus primarily on computationally efficient representations of the effects of heterogeneity



An overview of different mathematical descriptions of the liver for simulating toxicokinetics and toxicodynamics

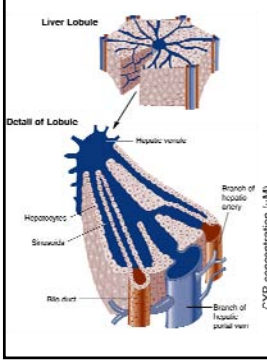
Model type	Schematic depiction	Assumptions/Applicability	References
One-Compartment Models			
Well Stirred (CSTR)		Well mixed (both macro- and micro-mixing). Uniform metabolic and biochemical properties throughout the liver.	Abu-Zahra and Pang (2000) Drug Metab Dispos 28, 807-813; Ridgway, et al. (2003) Journal of Biological Physics 29, 1-21
Plug flow (PFR)		Flow is uniform with no mixing and metabolism is fast.	Anissimov, et al. (1997) J Theor Biol 188, 89-101; Ridgway, et al. (2003) Journal of Biological Physics 29, 1-21
Dispersion flow		Highly non-uniform flow patterns; incomplete mixing. Uniform chemical and biochemical properties.	Anissimov, et al. (1997) J Theor Biol 188, 89-101; Ridgway, et al. (2003) Journal of Biological Physics 29, 1-21
Distribution-Based Models			
RTD-based Circulatory models		Non-mechanistic study of distribution of toxicants through residence time analysis.	Pang, et al. (2007) Aaps Journal 9, E268-E283; Weiss (1999) Eur J Pharm Sci 7, 119-27; Weiss, et al. (2007) J Pharm Sci 96, 913-926
Statistical distribution based model		Representation of heterogeneity through a statistical distribution. Useful when heterogeneity is due to limited set of factors	Ierapetritou, et al. (2009) Clinical and Translational Science (in press),
Stochastic/fractal models		Heterogeneity in the liver modeled through stochastic terms or fractal descriptions.	Forte, et al. (2002) Physical Review E 66; Pang, et al. (2007) Aaps Journal 9, E268-E283; Dirkesen and de Gaetano (2005) Bull Math Biol 67, 547-561
Multi-Compartment Models			
CSTRs in series		Multiple regions of the liver with each region well mixed. Flow is uniform and from one region to the next.	Abu-Zahra and Pang (2000) Drug Metab Dispos 28, 807-813; Anissimov and Roberts (2002) J Pharmacokinet Pharmacodyn 29, 131-56; Ridgway, et al. (2003) Journal of Biological Physics 29, 1-21
Multi-zonal (multi-compartmental) model		Multiple regions of the liver with different uptake and metabolic properties; metabolism occurring in deep tissue.	Abu-Zahra and Pang (2000) Drug Metab Dispos 28, 807-813; Andersen, et al. (1997) Toxicol Appl Pharmacol 144, 135-144
Back-mixing plus fixed lag times and slow perfused sinusoids		Zonal model with significant back mixing.	Anissimov and Roberts (2002) J Pharmacokinet Pharmacodyn 29, 131-56
Compartmental model with cellular compartments		Zonal model with significant back mixing, with variation across bulk tissue, deep tissue, and cellular space.	Anissimov and Roberts (2002) J Pharmacokinet Pharmacodyn 29, 131-56; Tanaka, et al. (1999) J Pharmacokinet Biopharm 27, 597-623

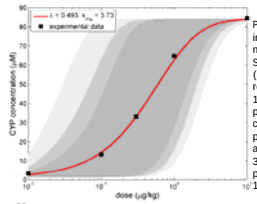
An overview of different mathematical descriptions of the liver for simulating toxicokinetics and toxicodynamics (continued)

Discrete, Agent-Based Models			
Agent-based		Bottom-up synthetic, non-mechanistic"description of multi-level processes within the liver; computationally and data intensive. * Mechanistic is defined here as derived from first principles of thermodynamics and kinetics (i.e. conservation laws and constitutive equations).	Hunt, et al. (2006) J Pharmacokinet Pharmacodyn 33, 737-72; Yan, et al. (2008) Pharm Res 25, 1023-36; Kawazu, et al. (2006) Conf Proc IEEE Eng Med Biol Soc 1, 4163-6
"Higher Dimensional" Models			
Continuous interconnected tubes		Variation in uptake and metabolic properties across the cross-section and along the direction of uniform flow.	Anissimov, et al. (1997) J Theor Biol 188, 89-101; Banks, et al. (1998) Mathemat Comput Mod 28, 9-29; Chalhoub, et al. (2007) Ann Biomed Eng 35, 474-491
Distributed zones		Variation in uptake and metabolic properties across the cross-section and along the direction of non-uniform flow.	Anissimov, et al. (1997) J Theor Biol 188, 89-101
Discrete interconnected tubes		Same as distributed zones but with intermittent mixing.	Anissimov, et al. (1997) J Theor Biol 188, 89-101
Fluid mechanics modeling of liver lobules		Computational fluid dynamics based, detailed realistic modeling of individual liver lobules; computationally and data intensive.	Rani, et al. (2006) J Biomech 39, 551-63

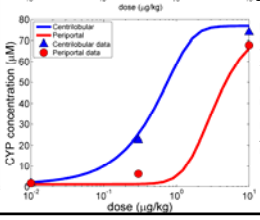
Current liver modeling efforts in MENTOR/DORIAN development focus primarily on computationally efficient representations of the effects of heterogeneity

Example: Modeling variable CYP induction by TCDD via a Distribution-Based Model (Preliminary/unpublished results from Isukapalli et al., 2009)





Predictions of CYP induction compared to measurements of Santosefano *et al.* (1998). Shaded areas represent the 5/95, 13/87 and 25/75th percentiles. (The centrilobular and periportal zones are assumed to be 27 and 39% of the liver, as per Andersen *et al.* 1997).

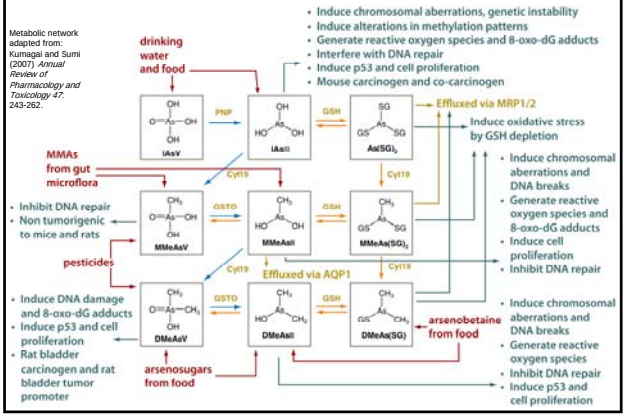


Predictions of CYP induction compared to data of Santosefano *et al.* (1999) derived from mRNA measurements.

Model parameters: λ= 1.41, K_{mx} = 3.4, n = 4.55.

Example of on-going research: Modeling sources, transport, biotransformations, and toxicodynamics of As species in the human body

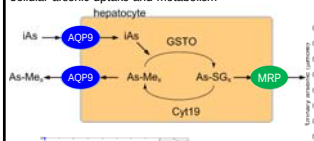
Metabolic network adapted from: Kumagai and Sumi (2007) *Annual Review of Pharmacology and Toxicology 47:* 243-262.

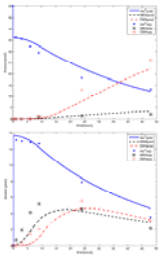


Research in-progress: "reconciliation" of biotransformation and transport of As modeled at both the individual hepatocyte and the whole organ level

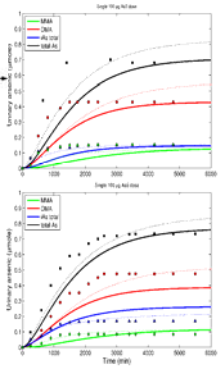
(preliminary/unpublished results from Stamatelos *et al.*, 2009 and from Sasso *et al.*, 2009)

Cellular arsenic uptake and metabolism





Time course prediction of As methylation in human hepatocytes vs data from Styblo et al. 1999



PBTKM predictions of cumulative urinary arsenic excretion in human volunteers after exposure to iAs

Solid lines: predictions of the MENTOR/DORIAN multiscale whole-body PBTK model

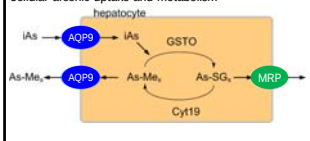
Dashed lines: predictions of the MENTOR-3P implementation of the El-Masri et al. PBTKM

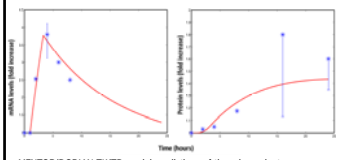
Data (from El-Masri & Kenyon 2008): controlled exposure experiments of humans ingesting 100 µg AsIII (top) and 100 µg AsV (bottom)

Research in-progress: "reconciliation" of biotransformation and transport of As modeled at both the individual hepatocyte and the whole organ level

(preliminary/unpublished results from Stamatelos *et al.*, 2009)

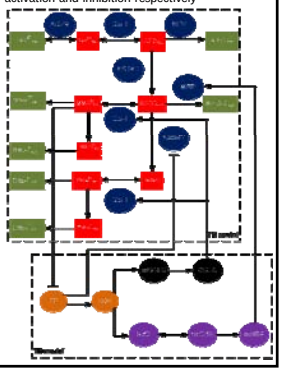
Cellular arsenic uptake and metabolism





MENTOR/DORIAN TK/TD model predictions of time-dependent transcription (left) and translation (right) of GCLC enzyme in murine hepatocytes. The measured values represent the fold increase of mRNA and protein levels of the enzyme after exposure to 10 µM of iAsIII. The data depict average values ±SEM of at least three independent experiments from Thompson et al. 2009

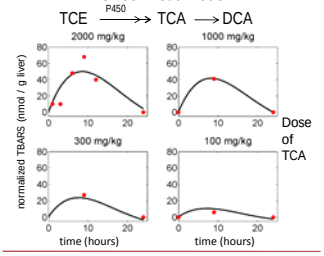
Schematic of component interactions of the TK/TD model; arrow and hammerhead indicate activation and inhibition respectively



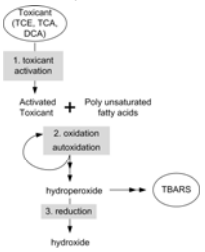
Modeling quantitative metrics of oxidative stress from exposure to TCE

(preliminary/unpublished results from Brinkerhoff et al., 2009)

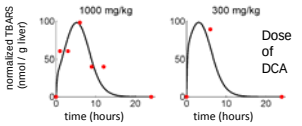
Toxicokinetic model



Toxicodynamic model



TBARS = thiobarbituric acid-reactive substances



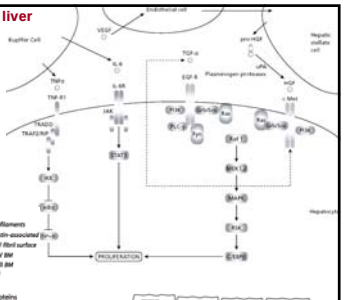
Toxicodynamic equation

$$\frac{d[TBARS]}{dt} = \frac{V_{max}[Toxicant]}{K_m + [Toxicant]} - (k_1[TBARS] + k_2 * TBARS_{AUC})$$

Experimental data of Larson & Bull, *Toxicol Appl Pharmacol (1992) 115: 268-277*

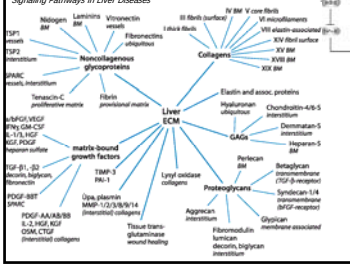
Intercellular interactions in the liver

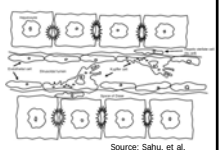
Source: Rodes et al. (2007) *Textbook of Hepatology - From Basic Science to Clinical Practice*, 3rd ed, Vols 1 & 2



Constituents of liver extracellular matrix

Source: Dufour & Clavien (2005) *Signaling Pathways in Liver Diseases*





Source: Sahu, et al. (2007) *Hepatotoxicity: From Genomics to In Vitro and In Vivo Models*

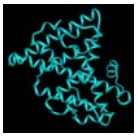
Research Area II:

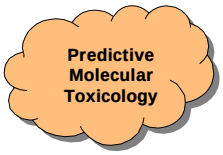
Hazard Identification

Molecular-scale methods

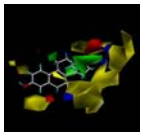
Characterize molecular components and interactions at a "local" (e.g. ligand-receptor) scale

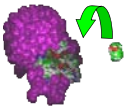
Receptor-based Approaches





Ligand-based Approaches (QSARs, etc)





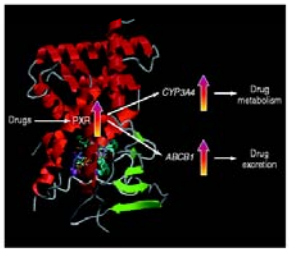
Virtual Screening; Data Analysis

Ligand-Receptor Interactions
Pregnane X Receptor (PXR)

- PXR modulates the transcription of metabolic enzymes and >36 other genes.
- PXR co-regulates the CYP3A4 metabolic gene and the ABCB1 "drug efflux" gene

[Synold, TW, et al., Nature Medicine **7**, 584-590 (2001).]

- Involved in many drug-drug interactions, giving rise to adverse drug effects.
- Many xenobiotics activate or repress the transcriptional machinery of PXR.
- Studies on PCBs show that the responsive active of PXRs to xenobiotics varies from species to species.



PXR ligands are pervasive and structurally diverse

- bile acids (bile salts, cholesterol metabolites)
- food ingredients, dietary supplements (e.g., isothiocyanate sulforaphane in broccoli)
- prescription drugs (e.g., statins, paclitaxel, antibiotics, azole antifungals, rifampicin)
- herbal components (e.g. hyperforin in St. John's Wort)
- environmental chemicals (EDCs, pesticides, plasticizers, PCBs, PBDEs)

PXR and Xenobiotics

Ory, DS. Nuclear receptor signaling in the control of **cholesterol** homeostasis: Have the Orphans Found a Home? *Circ. Res.* 95:660-670 (2004).

Tabb MM, Kholodovych V, Grün F, Zhou C, Welsh WJ, Blumberg B. Highly chlorinated **PCBs** inhibit the human xenobiotic response mediated by the steroid and xenobiotic receptor (SXR). *EHP* 112:163-169 (2004).

Yu S, Kong AN. Targeting carcinogen metabolism by dietary **cancer preventive compounds**. *Curr Cancer Drug Targets* 7(5):416-24 (2007).

Goetz AK, Dix DJ. Mode of action for reproductive and hepatic toxicity inferred from a genomic study of **triazole antifungals**. *Toxicol Sci.* 110(2):449-62 (2009).

Lin YS, Yasuda K, Assem M, Cline C, Barber J, Li CW, Kholodovych V, Ai N, Chen JD, Welsh WJ, Ekins S, Schuetz EG. The major human pregnane X receptor **(PXR) splice variant, PXR.2,** exhibits significantly diminished ligand-activated transcriptional regulation. *Drug Metab Dispos.* 37(6):1295-304 (2009).

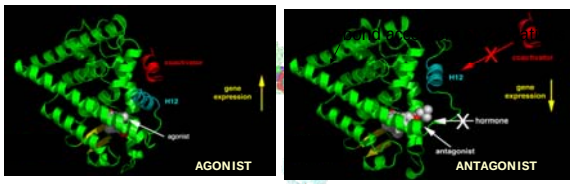
Kortagere S, Chekmarev D, Welsh WJ, Ekins S. Hybrid scoring and classification approaches to predict human pregnane X receptor **(PXR) activators**. *Pharm Res.* 26(4):1001-11 (2009).

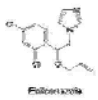
Unusual PXR antagonist binding site of conazoles

A series of conazoles antagonize PXR (10-20μM); mutagenesis data indicate that they bind to the outer surface of PXR---AF-2(H12) binding site

Huang et al., *Oncogene* 26: 258 (2007); Wang et al., *Clin Cancer Res* 13: 2488 (2007)

conventional structural model for nuclear receptor agonist and antagonist action



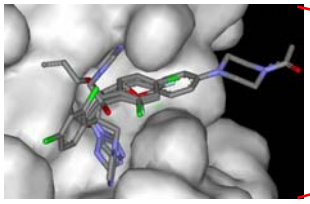


Hydrophobe / aromatic ring

Ekins, Welsh, et al., *Mol Pharmacol* 72:592–603 (2007).

Unusual PXR antagonist binding site of conazoles

- Using ligand-PXR docking simulations, we identified an alternative antagonist binding site anchored by Lys277 located in the AF-2 site
- Lys277 most likely serves as a "charge clamp" for interaction between the co-activator SRC-1 (His687) and PXR
- Conazoles compete with binding of co-activator SRC-1 to the AF-2 site



Ekins, Welsh, et al., Mol Pharmacol 72:592–603 (2007)

Methods Development for Data Analysis

Analysis of Toxcast 309 Data Set

Biological Spectra Analysis (BSA):
Link biological activity profiles to molecular structures

- Traditional (Q)SAR methods use the structure-based features (molecular descriptors) of a collection of chemicals to describe and compare their biological activities.

molecular structure ➡ bioactivity

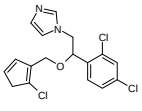
- In contrast, BSA uses the biological response profiles of the chemicals to describe and compare their molecular structures.

molecular structure ⬅ bioactivity

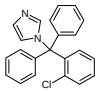
Fliri AF et al PNAS 12(2), 261-266 (2005)

Biological Activity Spectra (BSA)

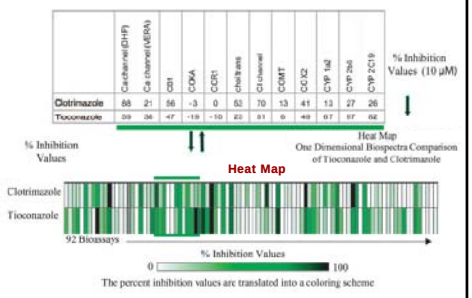
- depicted as a heat map -



Tioconazole



Clotrimazole



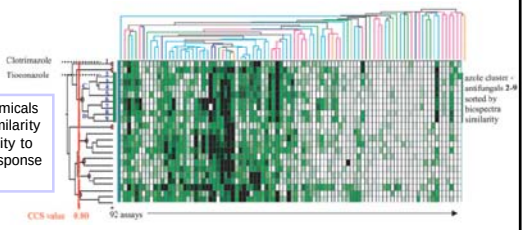
Fliri AF et al PNAS 2005, 12(2), 261-266

Heat map for a collection of chemicals and a panel of protein receptors

Two-way Hierarchical Clustering

clusters proteins based on similarity in their bioresponse profile

clusters chemicals based on similarity in their ability to induce bioresponse profile



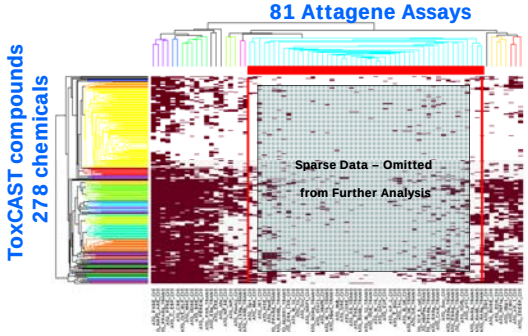
Fliri AF et al PNAS 2005, 12(2), 261-266

BSA study on assay data from Attagene, Inc.

- Transcription Activation (TA) assays
- 309 ToxCast chemicals @ 81 assays
- Reported LEL (lowest effective level) values from each assay
- *Inactive* chemical-assay combinations were assigned LEL =1000000

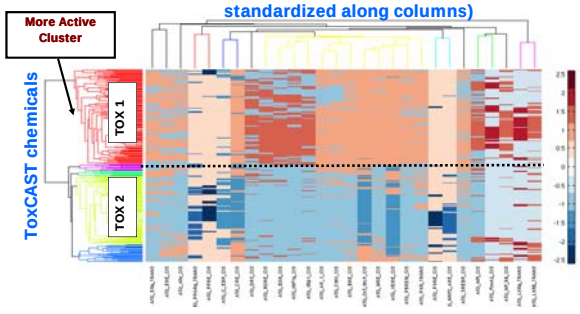
- Two-way hierarchical (UPGMA) clustering from Bioinformatics Toolbox v.3.1, MATLAB 7.6
- Analysis employed both Euclidean distance and Cosine metrics
- Assay results and calculated molecular descriptors were pre-processed using Unsupervised Forward Selection (UFS)

BSA study on Attagene Data



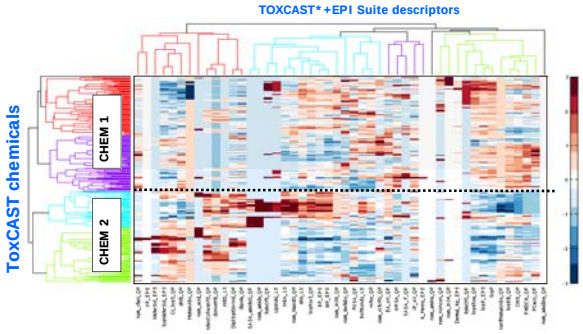
Dark regions - Compounds with measurable reported LEL (lowest effective level).
White regions- *Inactive* chemical-assay combinations (LEL =1000000).

Heat map for reduced set of 28 assays



Hierarchical clustering of ToxCast chemicals in the space of Attagene response biospectra reveals two major clusters **TOX1** (red) and **TOX2** (violet-green-lime-blue).

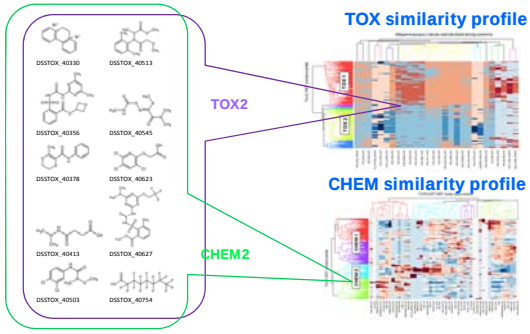
Heat map for space of chemical descriptors



Hierarchical clustering of ToxCast chemicals in chemical descriptor space reveals two major clusters **CHEM1** (red-violet) and **CHEM2** (blue-green).

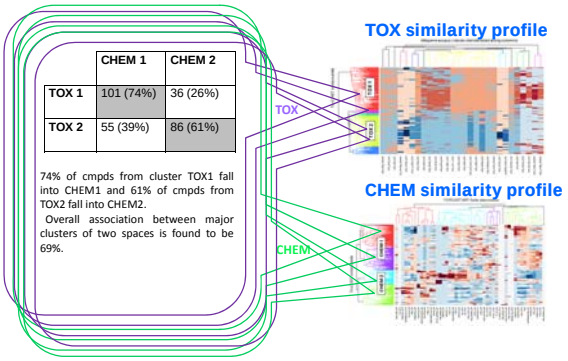
* TOXCAST: a combined set of Leadscope, QikProp and PhysChem derived descriptors

Connection between similarities in biospectra and chemical space



No obvious chemical similarities within individual subclusters.

Cross-mapping of TOX and CHEM spaces



	CHEM 1	CHEM 2
TOX 1	101 (74%)	36 (26%)
TOX 2	55 (39%)	86 (61%)

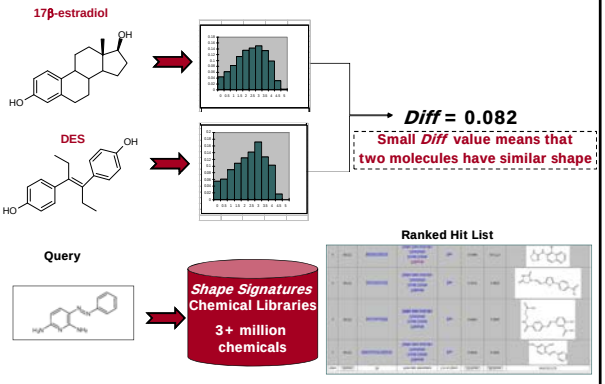
74% of cmpds from cluster TOX1 fall into CHEM1 and 61% of cmpds from TOX2 fall into CHEM2.

Overall association between major clusters of two spaces is found to be 69%.

Ligand-based Models, Rapid Virtual Screening & Chemical Prioritization

Shape Signatures

molecules are compared by subtracting their histograms



Shape-based QSAR Models for Toxicity Prediction

- **Cardiotoxicity**
  - hERG
  - $5HT_{2B}$

  Chekmarev DS, Kholodovych V, Balakin KV, Ivanenkov Y, Ekins S, Welsh WJ. Chem Res Toxicol. 21(6):1304-14 (2008).
- **Neurotoxicity**
  - blood-brain barrier (BBB) permeability

  Kortagere S, Chekmarev D, Welsh WJ, Ekins S. Pharm Res. 25(8):1836-45 (2008).
- **Hepatotoxicity**
  - PXR induction & repression

  Ai N, Krasowski MD, Welsh WJ, Ekins S. Drug Discov Today 14(9-10):486-94 (2009).

  Lin YS, Yasuda K, Assem M, Cline C, Barber J, Li CW, Kholodovych V, Ai N, Chen JD, Welsh WJ, Ekins S, Schuetz EG. Drug Metab Dispos. 37(6):1295-304 (2009).
- **Pesticides**
  - acetylcholinesterase inhibitors

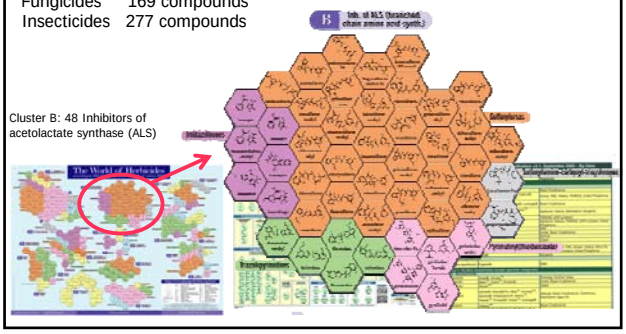
  Chekmarev D, Kholodovych V, Kortagere S, Welsh WJ, Ekins S. Pharm Res 26(9):2216-24 (2009).
- **Fungicides, Herbicides, Insecticides**

Analysis of Pesticides

Data on pesticides were collected from *The Pesticide Manual*:
http://www.pesticidemanual.com/index.htm

Herbicides	300 compounds
Fungicides	169 compounds
Insecticides	277 compounds

Cluster B: 48 Inhibitors of acetolactate synthase (ALS)



Analysis of pesticides

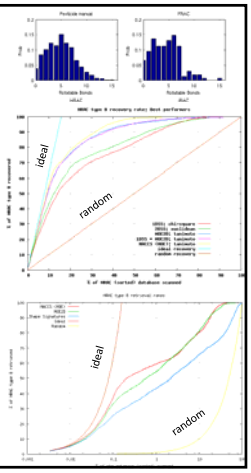
Molecular properties distribution:
MW, drug likeness, etc

Enrichment study on Herbicides
- variable descriptor-based techniques
  - Shape Signatures alone
  - Shape Signatures + MOE
  - Shape Signatures + MACCS keys

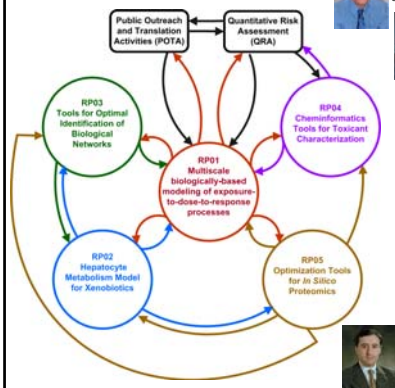
Retrieval study with 30,000 NCI
Cluster B is only 0.16% of entire set

Enrichment factor E > 250
E(random) = 1; E(ideal) = 600

$$E = \frac{\frac{H_a}{H_t}}{\frac{A}{D}}$$



Interactions/integration of ebCTC research projects




Dr. William Welsh
Center Director and
Project Principal Investigator (RP04)
UMDNJ-RWJ Medical School


Dr. Panos Georgopoulos
Associate Center Director and
Project Principal Investigator (RP01)
UMDNJ-RWJ Medical School


Dr. Weida Tong
Project Co-Principal
Investigator (RP01)
USFDA


Dr. Marianthi Ierapetritou
Project Principal
Investigator (RP02)
Rutgers University


Dr. Ioannis Androulakis
Project Co-Principal
Investigator (RP02)
Rutgers University


Dr. Herschel Rabitz
Project Principal Investigator
(RP03)
Princeton University

Dr. Christodoulos Floudas
Project Principal Investigator
(RP05)
Princeton University

Acknowledgments

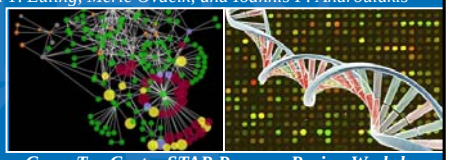
- **Funding**
  - USEPA funded environmental bioinformatics and Computational Toxicology Center (ebCTC) (STAR Grant RD-83272101)
- **Individuals**
  - Collaborators within UMDNJ-RWJMS, Rutgers, Princeton, USFDA as well as many other academic institutions, including the Albert Einstein College of Medicine, University of Pittsburgh, Mount Sinai Medical School, University of Montreal, etc.
  - Numerous collaborators within the USEPA National Center for Computational Toxicology and other USEPA Laboratories and Centers

Please visit our website
www.ebCTC.org for events, news, publications, contacts

EPA
United States Environmental Protection Agency

NJ Environmental Bioinformatics and Computational Toxicology Center – EPA Collaboration on *An Approach to Using Toxicogenomic Data in Risk Assessment: Dibutyl Phthalate (DBP) Case Study*

Susan Y. Euling, Meric Ovacik, and Ioannis P. Androulakis



Comp Tox Center STAR Progress Review Workshop
RTP, NC
October 1, 2009

Office of Research and Development
National Center for Environmental Assessment

STAR Bioinformatics Center/ebCTC Collaborators

❖Ioannis P. Androulakis

❖Meric Ovacik

at Rutgers University

2

HOW CAN GENOMIC DATA BE USED EFFECTIVELY IN RISK ASSESSMENT?

3

Strengths and Limitations of Using Omic Data in Risk Assessment

	STRENGTHS
Omics Technologies	Powerful because global Can identify: •Precursor events •Biomarker of effect •Biomarker of exposure •Mechanism
Data Analysis Tools (e.g., pathway analysis)	For microarray data: •Pathway analysis •Gene network analysis Can Identify: •Affected processes, pathways, & networks

PROJECT GOALS:

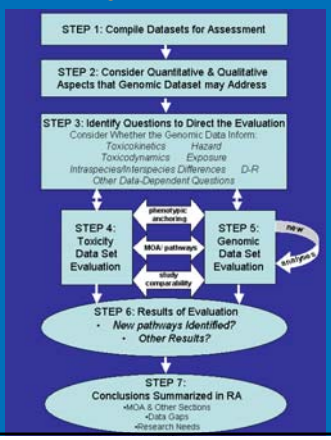
1. Develop an approach for using toxicogenomic data in risk assessment.
2. Perform a case study using this approach.

CASE STUDY SCOPE:

- Use an ongoing or completed assessment as starting point.
- Evaluate <u>available</u> data; not a data generation project
- Selected DBP for case study:
  - has a relatively large genomic data set and phenotypic anchoring for some of the observed gene expression changes.
  - case study is separate from IRIS assessment with a different purpose.

5

Approach for Evaluating Genomic Data in Health Assessments



6

Proposed DBP Mechanism of Action

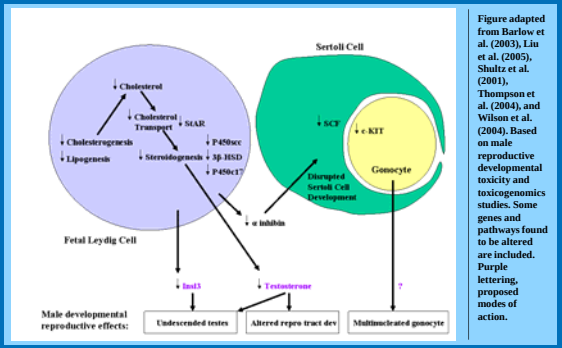


Figure adapted from Barlow et al. (2003), Liu et al. (2005), Shultz et al. (2001), Thompson et al. (2004), and Wilson et al. (2004). Based on male reproductive developmental toxicity and toxicogenomics studies. Some genes and pathways found to be altered are included. Purple lettering, proposed modes of action.

Toxicogenomics Dataset: Studies of Male Rat Tissues after in utero DBP Exposure

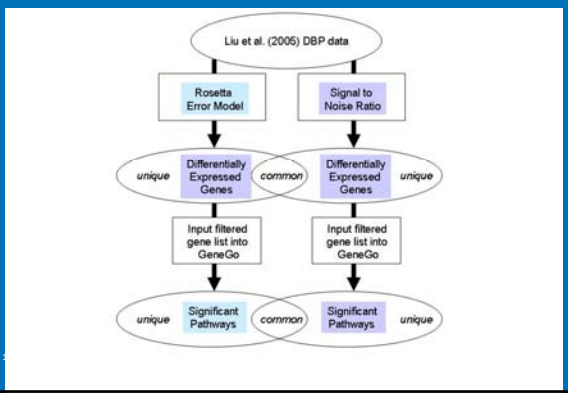
STUDY	DBP DOSE	TREATMENT INTERVAL	TOXICOGENOMIC METHOD		TISSUE COLLECTED
			MICROARRAYS (Platform)	RT-PCR	
Barlow et al., 2003	500 mg/kg/day	GD 12-19	No	Yes	Testis
Bowman et al., 2005	500 mg/kg/day	GD 12-19 or 19-21	Yes (Clontech cDNA arrays)	Yes	Wolffian ducts
Lehmann et al., 2004	0.1, 1.0, 10, 50, 100, or 500 mg/kg/day	GD 12-19	No	Yes	Testis
Liu et al., 2005	500 mg/kg/day	GD 12-19	Yes (Affymetrix GeneChip oligo arrays)	Yes	Testis
Shultz et al., 2001	500 mg/kg/day	GD 12-16, 12-19, or 12-21	Yes (Clontech cDNA arrays)	Yes	Testis
Thompson et al., 2004	500 mg/kg/day	GD 12-17, 18, or 19; 13-19, 14-19, 15-19, 16-19, 17-19, 18-19, or 19	No	Yes	Testis
Wilson et al., 2004	1000 mg/kg/day	GD 13-17	No	Yes	Testis
Thompson et al., 2005	500 mg/kg/day	0.5 – 24 hr on GD 18-19 or GD 19	Yes (Affymetrix GeneChip oligo arrays)	Yes	Testis
Plummer et al., 2007	500 mg/kg/day	GD 12.5-15.5; 12.5-17.5, or 12.5-19.5	Yes (Agilent 22K & 44K oligo arrays)	Yes	Testis: Whole, Seminiferous cord, and interstitial regions

Case Study Project: Pathway Analysis of Liu et al. Microarray Study

Issue:
- ❖Differentiating signal from noise in microarray studies

Explored use of:
- ❖Signal-to-noise ratio (SNR) method for identifying DEGs
- ❖DEG filter methods comparison: SNR to Rosetta Error Model (REM)

NEW ANALYSIS OF LIU et al. DATA: COMPARISON OF TWO STATISTICAL FILTERS



IN COMMON PROCESSES & PATHWAYS IDENTIFIED (SNR & REM)

BIOLOGICAL PROCESS	PATHWAYS
CELL ADHESION	Cytoskeleton remodeling; ECM remodeling; Endothelial cell contacts by junctional mechanisms; Ephrins signaling; Integrin inside-out signaling; Integrin outside-in signaling; Integrin-mediated cell adhesion; Reverse signaling by ephrin B
CELL SIGNALING	Activation of PKC via G-Protein coupled receptor; CCR3 signaling in eosinophils; ChREBP regulation pathway; G-Protein beta/gamma signaling cascades; G-Proteins mediated regulation p38 and JNK signaling; Regulation of actin cytoskeleton by Rho GTPases; Role of PKA in cytoskeleton reorganisation; ***Leptin signaling via JAK/STAT and MAPK cascades2***
DISEASE	NF-AT signaling in Cardiac Hypertrophy; NTS activation of IL-8 in colonocytes
GROWTH & DIFFERENTIATION	WNT signaling pathway; Regulation of acetyl-CoA carboxylase 2 activity in muscle; MAG-dependent inhibition of neurite outgrowth; EPO-induced Jak-STAT pathway; Angiotensin signaling via STATs; Angiotensin activation of ERK
HORMONES	Ligand-dependent activation of the ESR1/SP pathway
IMMUNE RESPONSE	CXCR4 signaling pathway; MIF - the neuroendocrine-macrophage connector
METABOLISM	***Androstenedione and testosterone biosynthesis and metabolism p.1; Cholesterol Biosynthesis; Cholesterol metabolism;*** dATP/dITP metabolism; dGTP metabolism; Estrone metabolism; Fructose metabolism; G-alpha(q) regulation of lipid metabolism; Gamma-aminobutyrate (GABA) biosynthesis and metabolism; Glutathione metabolism; Glycolysis and gluconeogenesis (short map); Glycolysis and gluconeogenesis p. 1; Glycolysis and gluconeogenesis p. 2; Histamine metabolism; Histidine-glutamate-glutamine and proline metabolism; Leucine, isoleucine and valine metabolism.p.2; Lysine metabolism; Mitochondrial ketone bodies biosynthesis and metabolism; Mitochondrial long chain fatty acid beta-oxidation; Mitochondrial unsaturated fatty acid beta-oxidation; Peroxisomal branched chain fatty acid oxidation; Phenylalanine metabolism; ***PPAR regulation of lipid metabolism; Propionate metabolism p.1; Propionate metabolism p.2***; Regulation of fatty acid synthesis: NLTP and EHHADH; Regulation of lipid metabolism by niacin and isoprenaline; ***Regulation of lipid metabolism via LXR, NF-Y and SREBP; Regulation of lipid metabolism via PPAR, RXR and VDR***; Serotonin - melatonin biosynthesis and metabolism; TCA; Triacylglycerol metabolism p.1; Tryptophan metabolism
TRANSCRIPTION	Transcription factor Tubby signaling pathways; Role of VDR in regulation of genes involved in osteoporosis; Brca1 as transcription regulator

Source: Sen and Hester

Exploratory Methods Development for Analysis of Genomic Data for Application to Risk Assessment

Issue:
- ❖For risk assessment, we're interested in affected pathways; traditional pathway analysis methods may lose gene and pathway information

Explored use of:
- ❖Pathway Activity Level method & utilized the results to build a gene network model.

12

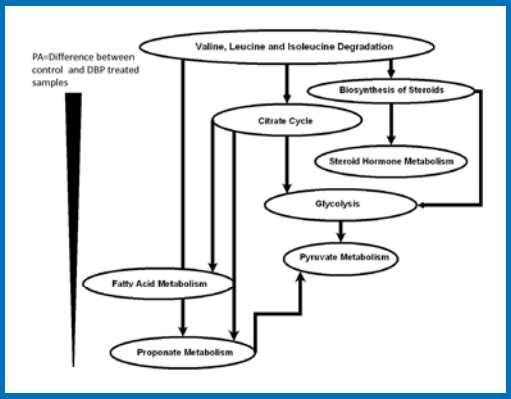
Pathway Activity Level Approach

- Adapted method of Tomfohr, J; Lu, J; Kepler, TB. (2005) Pathway level analysis of gene expression using singular value decomposition. BMC Bioinformatics 6:225.
- Identifies impact on a pathway without 1st identifying differentially expressed genes

Advantages:
- Considers all genes (whether DEG or not) in a pathway
- Can compare PA among pathways

13

Putative Metabolic Gene Network Based on the Pathway Activity Method: Liu et al. (2005) Data and KEGG Database



14

Source: I. Androulakis and M. Ovacik

Exploring Methods to Measure Interspecies Differences in Toxicodynamics

Issue:
- Need for approaches and metrics to extrapolate from animal model to human for risk assessment.

Explored use of:
- Utilizing available data to develop cross-species metrics for the biosynthesis of steroids pathway -

1) DNA sequence data: Compared predicted amino acid sequences of proteins

2) Enzyme presence data

15

Team Members

U.S. EPA
Susan Makris (NCEA, ORD)
Banalata Sen (formerly NCEA)
Andrea S. Kim (formerly NCEA)
Bob Benson (Region 8)
Channa Keshava (IRIS, NCEA, ORD)
Nagalakshmi Keshava (NCEA, ORD)
Susan Hester (NHEERL, ORD)
Vickie S. Wilson (NHEERL, ORD)
L. Earl Gray Jr. (NHEERL, ORD)
Chad Thompson (formerly NCEA)
Weihsueh Chiu (NCEA, ORD)

THE HAMNER INSTITUTES for HEALTH SCIENCES
Kevin W. Gaido

NIEHS
Paul M.D. Foster
Lori White

NCER STAR BIOINFORMATICS CENTER/ebCTC
Ioannis P. Androulakis (Rutgers)
Meric Ovacik (Rutgers)
Marianthi G. Ierapetritou (Rutgers)
Panos P. Georgopoulos (UMDNJ)
William Welsh (UMDNJ)

16

Final Report Available on the NCEA Website

An Approach to Using Toxicogenomic Data in U.S. EPA Human Health Risk Assessments: A Dibutyl Phthalate Case Study

Available at: http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=205303

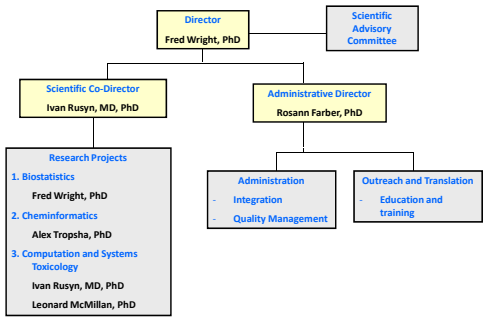
17

The Carolina Environmental Bioinformatics Center (CEBC)

- One of two EPA STAR Centers funded in November 2005, intended to extend capabilities in computational toxicology
- Specific capabilities highlighted included 'omics expertise and strengths in elucidating genetic variation
- Here we describe the Center and highlight recent collaborations

1

Organization

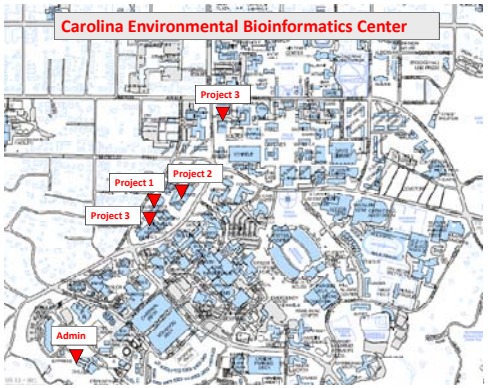


2

Organization

- Three major Research Projects: (1) Biostatistics, (2) Cheminformatics, and (3) Computational Infrastructure for Systems Toxicology
- Administrative Unit
- Outreach and Translational Activity (POTA)
- Each project includes direct collaboration with environmental scientists

3



Progress

- Publications
- Collaborations with environmental scientists
- UNC awarded a second STAR Center (2008), The Carolina Center for Computational Toxicology (CCCT, Ivan Rusyn, P.I.)
- software development, and web tools

5

Representative Joint Publications with EPA

•Harrill JA ,Li Z, Wright FA, Crofton KM. Transcriptional response of rat frontal cortex following acute exposure to the pyrethroid insecticide permethrin or deltamethrin. BMC Genomics, 2008 Nov 18;9(1):546
•Harrill JA, Li Z, Wright FA, Crofton K (2007). Transcriptional response of rat cerebrocortical tissue following acute exposure to the pyrethroid insecticide permethrin or deltamethrin, submitted.
•Judson R, Elloumi F, Setzer WR, Li Z , Shah I. (2008) A Comparison of Machine Learning Algorithms for Chemical Toxicity Classification Using a Simulated Multi-Scale Data Model BMC Bioinformatics, Vol. 9, 241.
•Li Z, Wright FA and Royland JE. Age-dependent Variability in Gene Expression in Fisher 344 Rat Retina. Toxicological Science, 2008 Nov 18;9(1):546.
•Zhu H, Rusyn I, Richard A, Tropsha A. Use of cell viability assay data improves the prediction accuracy of conventional quantitative structure-activity relationship models of animal carcinogenicity. Environ. Health Perspect. 2008; (116): 506-513.
•Zhu H, Tropsha A, Fourches D, Varnek A, Papa E, Gramatica P, Oberg T, Dao P, Cherkasov A, Tetko I V. Combinatorial QSAR Modeling of Chemical Toxicants Tested against Tetrahymena pyriformis. J. Chem. Inf. Model. 2008; (48): 766-784.
•Zhu H, Ye L, Richard A, Golbraikh A, Rusyn I, Tropsha A. A Two-step Hierarchical Quantitative Structure Activity Relationship Modeling Workflow for Predicting in vivo Chemical Toxicity from Molecular Structure. Environ. Health Perspect. Submitted.

Representative Joint Abstracts/Posters

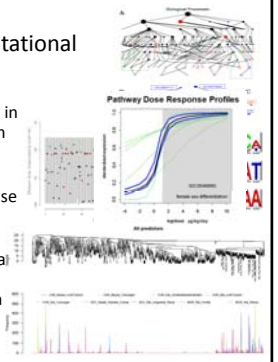
•Dix D, Judson R, Elloumi, F, Li Z, Wright FA. Reif, David, Rotroff, Daniel, Singh, Amar, Knudsen, Thomas, Houc k, Keith (2008). The Analysis of Genomic Dose-Response Data in the EPA ToxCast Program. SRA Annual Meeting, Boston, USA, submitted.
•Dix D, Judson R, Kavlock R, Li Z, Martin M, Richard A, Setzer W, Houck K (2007). EPA's ToxCast program for predicting hazard and prioritizing toxicity testing of environmental chemicals. International Science Forum on Computational Toxicology. Research Triangle Park, NC.
•Dix D, Martin MT, Li Z, Judson R, Houck KA (2007). Development of EPA's ToxCast program for prioritizing the toxicity testing of environmental chemicals. Annual Meeting of Society for Biomolecular Sciences, Montreal, Canada.
•Elloumi, F, Li, Z, Judson, Richard (2008). EPA Analysis for MAQC Toxicogenomics Datasets. The 8th MAQC Project Meeting Development and Validation of Predictive Models. March 24–26, 2008, Washington DC, USA.
•Kavlock RJ, Dix D, Judson R, Li Z, Martin M, Richard A, Setzer W and Houck K (2007). US EPA's ToxCast program for predicting hazard and prioritizing toxicity testing of environmental chemicals. 6th World Congress on Alternatives and Animal Use in the Life Sciences, Tokyo.
•Li Z, Wright FA and Royland JE (2007). Age-dependent heterogeneity of gene expression in Fisher 344 rat retina. The 46th Annual Meeting of the Society of Toxicology, Charlotte, NC.
•Li Z, Wright FA and Royland JE (2007). Significance Analysis of Variation Change in gene expression studies. 15th Annual International Conference on Intelligent Systems for Molecular Biology and 6th Annual European Conference on Computational Biology. Vienna, Austria.
•Rodgers AD, Zhu H, Rusyn I, Tropsha A. QSAR Modeling of Human Liver Adverse Effects Database Using kNN method. Abstract 244; presented at the 47th National Meeting of the Society of Toxicology, Seattle, WA, March 2008
•Wang K, Richard A, Rusyn I, and Tropsha A (2007). Toxico-Cheminformatics and QSPR Modeling of the Carcinogenic Potency Database. The 46th Society of Toxicity (SOT) Annual Meeting, Charlotte, NC, accepted.
•Zhang L, Zhu H, Rusyn I, Judson R, Dix D, Houck K, Martin M, Richard A, Kavlock R, and Tropsha A. Cheminformatics Analysis of EPA ToxCast Chemical Libraries to Identify Domains of Applicability for Predictive Toxicity Models and Prioritize Compounds for Toxicity Testing. Society of Toxicology Annual Meeting, Baltimore, MD. 2009.
•Zhu H, Rusyn I, Richard A, and Tropsha A (2007). The utilization of NTP-HTS data in predictive ADME/tox modeling. US EPA International Science Forum on Computational Toxicology, Research Triangle Park, NC.
•Zhu H, Wang K, Rusyn I, Richard A, and Tropsha A (2007). The utilization of NTP-HTS data in predictive ADME/tox modeling. The 46th Society of Toxicity (SOT) Annual Meeting, Charlotte, NC, accepted.
•Zhu H, Ye L, Rusyn I, Richard A, Golbraikh A, Tropsha A. Two-step Quantitative Structure Activity Relationship modeling of in vivo toxicity using in vitro cytotoxicity data. Abstract 245; presented at the 47th National Meeting of the Society of Toxicology, Seattle, WA, March 2008.
•Li Z, Elloumi F, Wright FA (2008). Chemical Toxicity Prediction for Toxicogenomics Studies. The 47th Annual Meeting of the Society of Toxicology. Seattle, USA.

While the CCCT is more highly focused on biology and mechanistic modeling, the CEBC focuses on discovery and obtaining valid statistical conclusions.

Discovery and Statistical Modeling

(1) Biostatistics in Computational Toxicology

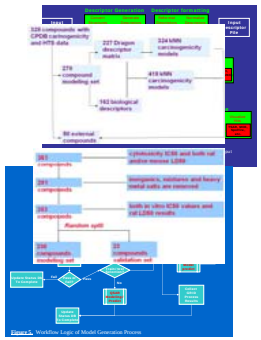
- Existing emphasis on strengths in microarray analysis, elucidation of networks/pathways, eQTL analysis
- New emphasis on dose-response testing, data mining, and penalized regression
- Analysis of ToxCast Phase I data from EPA and development of related methods will likely be a large portion of remaining activity



9

(2) Cheminformatics

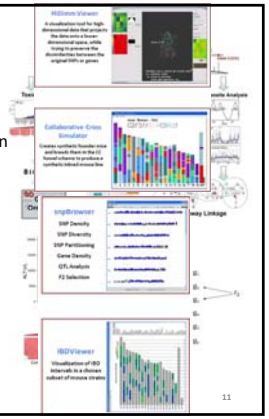
- seeks to establish a universally applicable and robust predictive toxicology modeling framework
- Focuses on Quantitative Structure Activity/Property Relationships (QSAR)
- Establishes a modeling workflow, toxicity prediction scheme and software development



10

(3) Computation and Systems Toxicology

- Uses model for toxicity profiling in multiple strains of mice to set up computational infrastructure
- Computational methods development
- Develops user-friendly software tools from methods in Projects 1 and 2



11

Project 1:
Biostatistics in Computational Toxicology

- Fred Wright, Ph.D. (P.I.) –statistical genetics, genomic analysis
- Andrew Nobel, Ph.D. – clustering, data dimensional reduction, genetic pathway analysis
- Other faculty have been phased out
- Zhen Li, M.S. – all of the above
- Partial postdoc and student positions

12

Project Objectives

•provide biostatistical support to the Center

•perform data analysis and develop methods

•collaborate with EPA and the computational toxicology community.

Recent Activities

- Direct collaborations and data analysis
  - Work with Project 2 investigators on toxicity prediction/data mining methods
  - Work with Project 3 investigators on rodent toxicity and eQTL mapping
  - Analysis of clinical toxicity and metabolomic data to explore a large number of prediction approaches
  - abstracts on ToxCast data and proposed analyses for prioritization of chemicals
  - Expression QTL mapping relevant to toxicity

14

At any one time, about 3 active analysis projects
-Collaborations inspire new methods development
-A recent example:

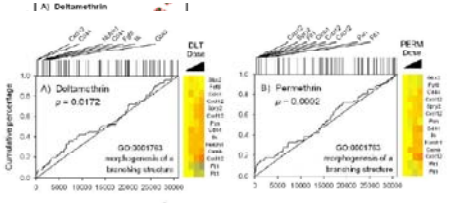
BMC Genomics

Research article Open Access

Transcriptional response of rat frontal cortex following acute *in Vivo* exposure to the pyrethroid insecticides permethrin and deltamethrin

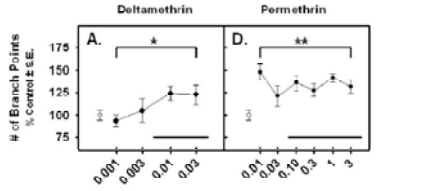
Joshua A Hamlin[1], Zhen Li[2], Fred A Wright[2], Nicholas M Radin[1], William R Mundy[3], Rogelio Tornero-Velez[4] and Kevin M Crofton[*3]

Our pathway analysis procedure SAFE used to identify pathways…



16

…followed by experimental evidence of pyrethroid effects on the total number of branch points in primary cortical cell cultures exposed to deltamethrin or permethrin.



Harrill et al., BMC Genomics. 2008 Nov 18;9:546.

17

This experience, in addition to exposure to dose-response data from NCCT personnel, got us thinking…

- Relatively few methods for dose-response that are tuned to gene expression studies
- Even fewer that consider "pathways" (gene sets)
- A primary challenge is maintaining appropriate type I error control for individual transcripts, whether parametric or not
- We would like methods to be fast, for permutation or bootstrapping.
- How to aggregate evidence across transcripts within a pathway?

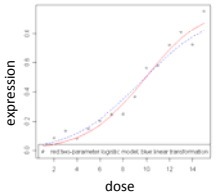
18

Dose response modeling for gene expression and pathways

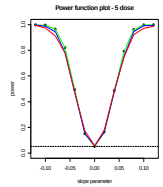
$$Y_{ij} = f(d_i, \theta) + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$

Y_{ij} is (continuous) response of the *j-th* subject on the *i-th* dose d_i; θ is the vector of parameters for the distribution f

We have performed extensive investigation of simple (approximate) two-parameter logistic fits, establishing reasonable false positive rates and power for small sample sizes



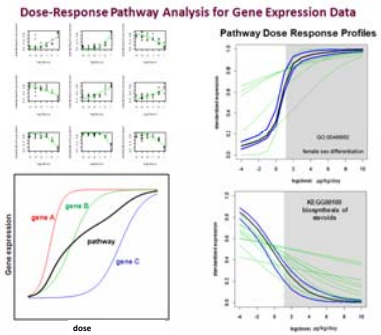
19



Power of simple dose-response approximations (linear after transformation)

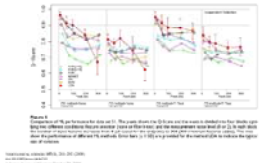
-- two parameter logistic model; -- linear model; -- fast sigmoid curve fitting

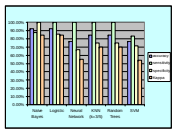
So that we can build on top of our existing gene expression pathway analysis software



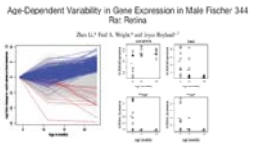
21

•Several other collaborative examples





Chemical Toxicity Prediction for Toxicogenomics Studies Using an Example Dataset. Li et al., SOT 2008.

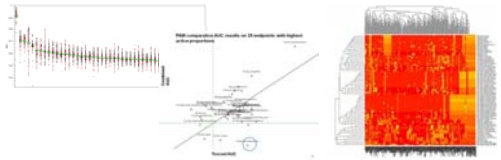


22

Prediction of *in vivo* toxicity endpoints from ToxCast™ Phase I data using a variety of machine learning approaches

Fred A. Wright[1,2], Zhen Li[1,2], Hanwen Huang[1], Arpita Ghosh[1,2], Wenjun Bao[3], Lili Li[3], Tzu-Ming Chu[3], Russ Wolfinger[3], Wei Sun[1,2], Fei Zou[1,2], Ivan Rusyn[2,4]

[1]Department of Biostatistics, [2]The Carolina Centers for Environmental Bioinformatics and Computational Toxicology, University of North Carolina, Chapel Hill, NC, [3]SAS Institute, Cary, NC, [4]Department of Environmental Sciences and Engineering, University of North Carolina.



Prediction of *in vivo* toxicity endpoints from ToxCast™ Phase I data using a variety of machine learning approaches

Fred A. Wright[1,2], Zhen Li[1,2], Hanwen Huang[1], Arpita Ghosh[1,2], Wenjun Bao[3], Lili Li[3], Tzu-Ming Chu[3], Russ Wolfinger[3], Wei Sun[1,2], Fei Zou[1,2], Ivan Rusyn[2,4]

[1]Department of Biostatistics, [2]The Carolina Centers for Environmental Bioinformatics and Computational Toxicology, University of North Carolina, Chapel Hill, NC, [3]SAS Institute, Cary, NC, [4]Department of Environmental Sciences and Engineering, University of North Carolina.

•309 chemicals

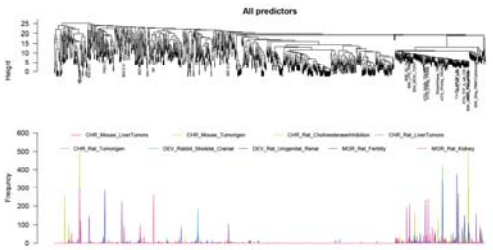
• over 70 toxicity endpoints to be predicted

• 600+ bioassay results

•1224 Dragon chemical descriptors provided by Drs. Hao Zhu and Alex Tropsha (Project 2) as additional toxicity predictors.

•Extensive work on cross-validation and ROC area under the curve (AUC) assessment of 84 (and now nearly 200) prediction models provides a global view of the strengths and weaknesses of various prediction approaches (details in talk at ToxCast Data Analysis Summit web site).

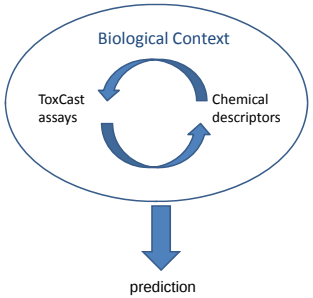
Both chemical descriptors and ToxCast assays remain in the predictive models

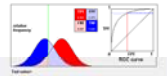


Even biologically naïve prediction models suggests improvement for several endpoints

Endpoint	percent AUC improvement ToxCast Phase I assays over chemical descriptors
CHR_Rat_LiverProliferativeLesions	18.2%
DEV_Rat_Urogenital_Ureteric	13.6%
CHR_Rat_LiverTumors	14.0%
DEV_Rat_Urogenital_Re.1l	8.3%
CHR_Rat_LiverNecrosis	9.6%
MGR_Rat_LiveBirthPND1	9.1%
DEV_Rat_Orofacial_CleftLipPalate	8.4%
CHR_Rat_TesticularAtrophy	8.0%
MGR_Rat_Fertility	6.7%
MGR_Rat_Prostate	4.5%
MGR_Rat_Epididymis	5.3%
MGR_Rat_Implantations	5.3%

A more refined view

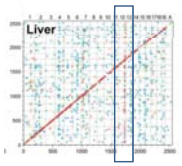




Sensitivity/specificity tradeoffs may be favorable, and better prediction for certain applicability domains

Additional methods development in Project 1 (one example)

- Methods for detecting true “trans-bands” in eQTL studies



“Real” or not?

Results appear highly unlikely to be due to chance, but can artificially result from transcript correlation

We have worked out permutation and analytic (matrix decomposition) methods to assess

GATTI, MAKI, CHESLER, ET AL.
HEPATOLOGY, Vol. 46, No. 2, 2007

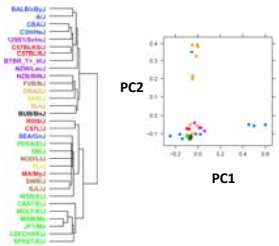
28

This work is part of a larger effort to get a statistically valid “snapshot” of eQTL data without the need for resampling.

Related efforts include

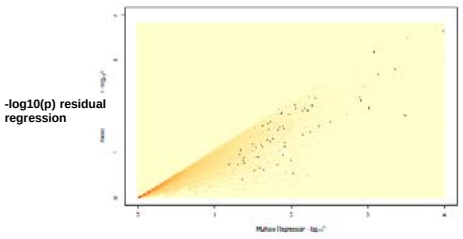
(i) transforming transcript data to handle outliers, which can be a problem for SNPs with low minor allele frequency

(ii) Principal component handling of stratification



29

PC-based stratification control for eQTL analysis, mouse data



Bottom line – these issues matter in identifying eQTLs, and therefore in elucidating genetic susceptibility

Project 2:
Cheminformatics

- Alex Tropsha, Ph.D. (P.I.) – Quantitative Structure Activity Relationship (QSAR) modeling, software tools for chemical descriptor-based prediction
- Hao Zhu, Ph.D. – QSAR modeling
- Additional postdoctoral researchers, research faculty, and students
- Leverages effort in the Laboratory for Molecular Modeling , School of Pharmacy, UNC

31

Project Objectives

•coordinates the compilation and mining of data from relevant external databases

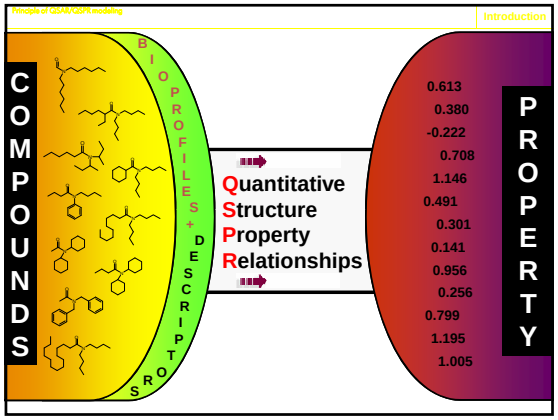
•performs analysis and methods development for building statistically significant and externally predictive Quantitative Structure-Activity Relationship models of chemical toxicology data

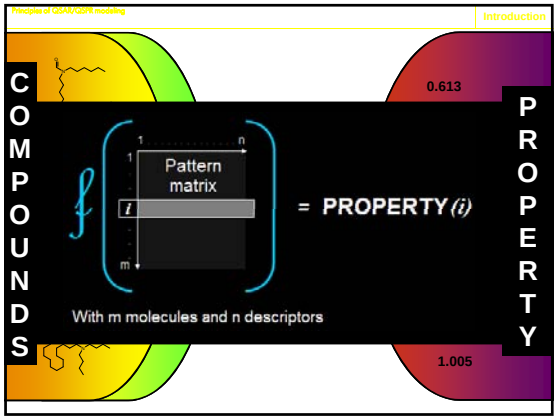
•Performs collaborative work within the Center and with EPA collaborators

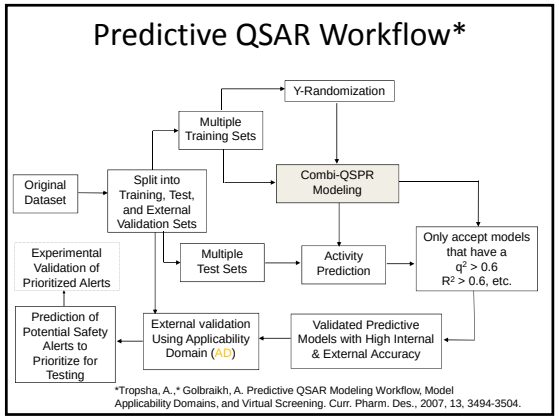
• Recent activity highlighted here

Improved quantitative models of chemical toxicity based on combined application of chemical and biological molecular descriptors

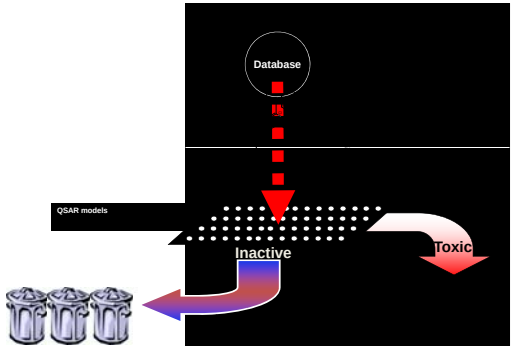
- Overall project vision: exploiting the entire structure – *in vitro* – *in vivo* continuum
- Predictive QSAR Modeling Workflow
- Applications
  - The use of hybrid chemical biological descriptors
  - novel data partitioning approach based on in vitro – in vivo correlations: Hierarchical QSAR modeling of rodent toxicity







Compound prioritization using QSAR models



Application I. Using Full High-Throughput Screening Dose Response Curves as Biological Fingerprints of Organic Compounds in QSAR Studies

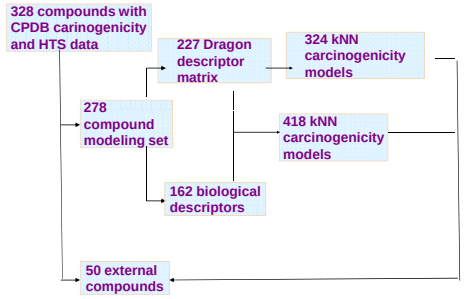
Zhu, Sedykh, et al, in preparation; EPA Collaborator: Ann Richard

Zhu H, Rusyn I, Richard A, Tropsha A. Use of cell viability assay data improves the prediction accuracy of conventional quantitative structure-activity relationship models of animal carcinogenicity. *Environ Health Perspect* 2008; (116): 506-513

Using HTS Dose Response Curve to Assist QSAR Modeling of Carcinogenicity

- Three types of descriptors:
  Chemical (300+); Biological (150+); Hybrid (400+)
- CPDB carcinogenicity data: 328 unique organic compounds with multi-cell carcinogenicity calls, 189 actives and 139 inactives

Using HTS data for Carcinogenicity Modeling



Prediction of the External Validation Set

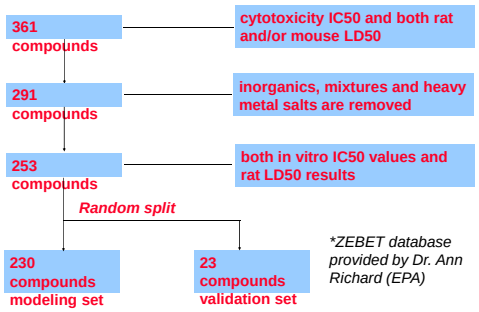
	kNN-Dragon	kNN-Hybrid
Sensitivity	69%	66%
Specificity	46%	56%
CCR	57%	62%
Coverage	72%	70%

average values after repeating the experiments 5 times.

Application II: A Two-step Hierarchical QSAR Modeling Workflow for Predicting *in vivo* Chemical Toxicity*

*Zhu, Rusyn, Wright, et al, EHP, 2009(8), 1257-64; in collaboration with Ann Richard, NCCT, US EPA

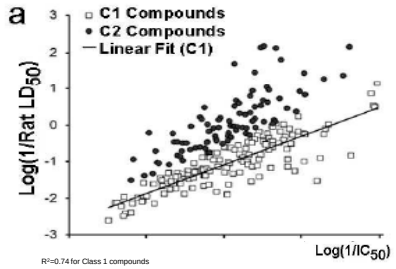
ZEBET Database* and Data Preparation



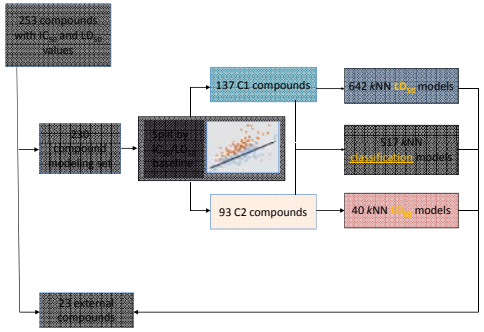
*ZEBET database provided by Dr. Ann Richard (EPA)

Data partitioning based on the moving regression approach

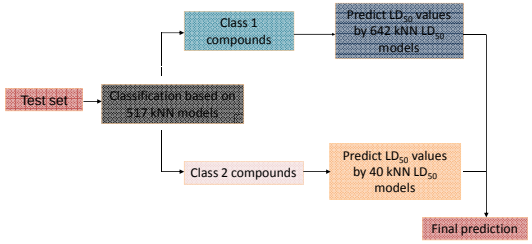
- IC50 vs. rat LD50 values



Modeling Workflow



Prediction Workflow



Classification of the Rat LD50 Values for the External Set of 23 Compounds

No AD:
Classification rate = 62%

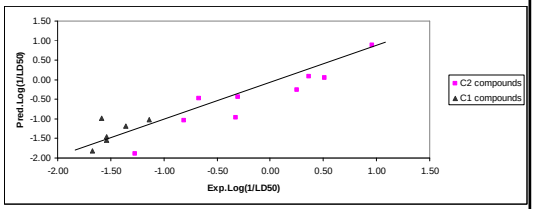
	Pred. C1	Pred. C2
Exp. C1	7	2
Exp. C2	6	5

With AD:
Classification rate = 78%

	Pred. C1	Pred. C2
Exp. C1	6	0
Exp. C2	4	5

Prediction of the Rat LD50 Values of the External 23 Compounds

- R^2=0.79, *MAE*=0.37, Coverage=74% (17 out of 23)

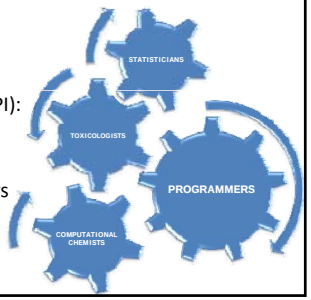


Future Studies

- Analyze models to identify significant assay-chemical combinations that are predictive of *in vivo* outcomes
- Explore the entire NTP dataset
- Apply model prospectively to prioritize new compounds for focused toxicity testing.

Project 3:
Computational Infrastructure for Systems Toxicology

- Ivan Rusyn (co-PI): toxicology, genomics
- Leonard McMillan (co-PI): computer science, GUI, software engineering
- Additional programmers and students



Project Objectives

- Develop and implement algorithms that streamline the analysis of multi-dimensional data streams in dose-response assessment and cross-species extrapolation
- Facilitate the development of a standard workflow for (i) analysis of the -omics data, (ii) linkages to classical indicators of adverse health effects, and (iii) integration with other types of biological information such as genome sequences and genetic differences between species
- Build web-based, open-source and user-friendly graphical interfaces associated with interoperable computational tools for data analysis that facilitate incorporation of new data streams into basic research and decision-making pipelines (methods from Projects 1 and 2)

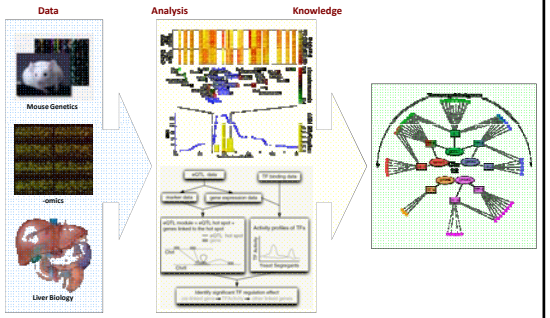
•has created a framework for handling emerging –omics data on genetic susceptibility in model organisms.

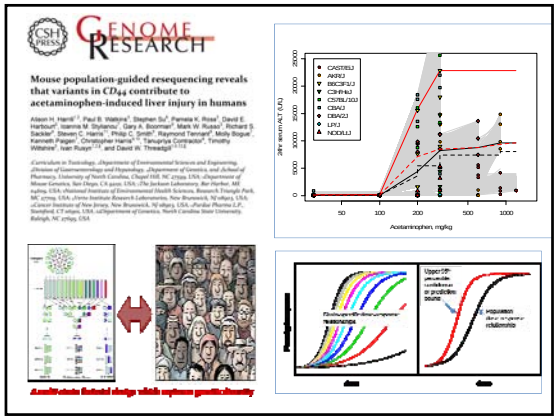
•provides programming expertise to create graphical tools that are used by partners within the Center and in collaboration with EPA personnel and other environmental scientists

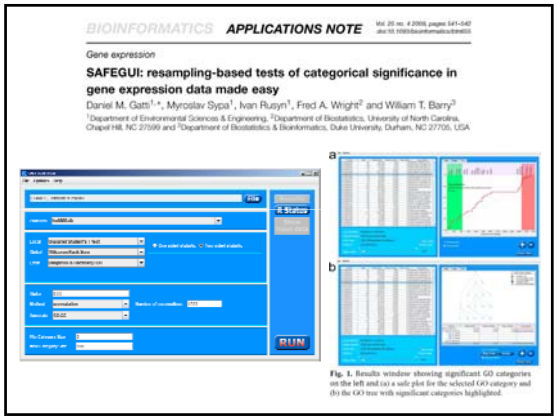
•strengthens and advances the field of computational toxicology through direct partnerships and the dissemination of tools used by both bioinformatics and bench scientists.

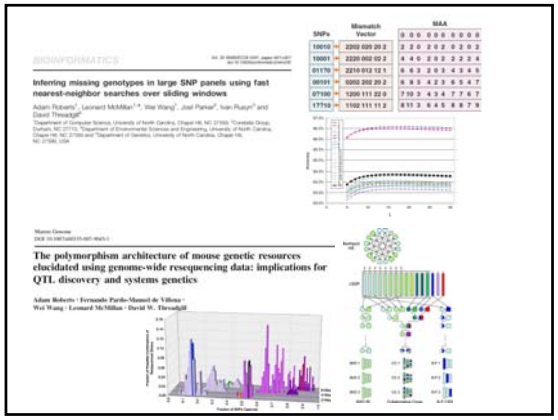
Staff Locations:
Facilitating Inter-Disciplinary Interactions

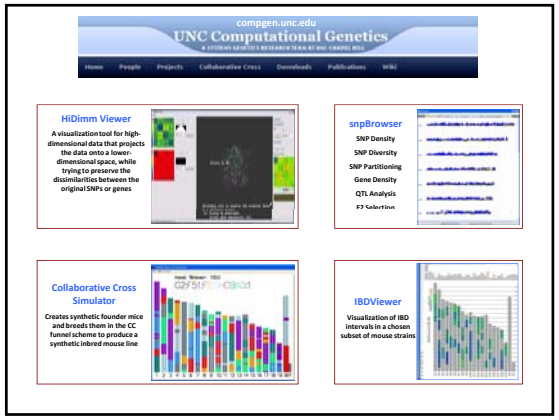
Driving Biological problem:
Population-wide predictions from toxicity profiling











The next year – Project 1

- Finish methodology for open projects and collaboration
- Finish dose-response pathway analysis method
- ToxCast data analysis – bring to intermediate conclusion
- ToxCast – go deeper, in terms of choices of endpoints, sensitivity vs. specificity, domains of applicability

61

The next year – Project 2

•Continuing work on QSAR modeling of multiple animal toxicity endpoints

•Developing novel QSAR methodology by using in vitro biological information to model in vivo toxicity endpoints

• QSTR modeling of nanotoxicology data.

•For all of these activities we on data collected under the ToxCast, DSSTox, and other projects.

62

The next year – Project 3

•Continuing integration/support of tools from other CEBC projects

•continued programming and algorithmic I

•improvements to algorithms in tools and applications

•development of specific data-mining algorithms for genomic databases

•continued biology-driven research that generates appropriate datasets for testing and implementing novel computational and biostatistical approaches.

63

Center-wide

•Emphasis on training other scientists in tools developed

•Bringing open source code and methods to new stage in evolution

64

EXTRA

50 endpoints selected based on the rank of the frequency of "actives"

Endpoint	Frequency
DEV_Rat_Skeletal_Axial	111
DEV_Rabbit_PregnancyRelated_Mate	109
MGR_Rat_Liver	104
CHR_Rat_Tumorigen	97
CHR_Mouse_LiverProliferativeLesi	93
CHR_Mouse_Tumorigen	92
DEV_Rat_General_FetalWeightReduc	87
MGR_Rat_Kidney	74
CHR_Mouse_LiverTumors	72
DEV_Rabbit_PregnancyRelated_Embr	70
MGR_Rat_ViabilityPND4	68
CHR_Mouse_LiverHypertrophy	66
CHR_Rat_LiverHypertrophy	65
CHR_Rat_LiverProliferativeLesion	65
DEV_Rabbit_Skeletal_Axial	55
DEV_Rat_PregnancyRelated_EmbryoF	55
DEV_Rabbit_General_FetalWeightRe	49
DEV_Rat_PregnancyRelated_Materna	49
DEV_Rat_Skeletal_Appendicular	47
CHR_Mouse_KidneyPathology	45
CHR_Rat_CholinesteraseInhibition	45
MGR_Rat_LitterSize	43
--------	--------
DEV_Rat_Orofacial_JawHyoid	12

67

Chemical Toxicity Prediction for Toxicogenomics Studies Using an Example Dataset

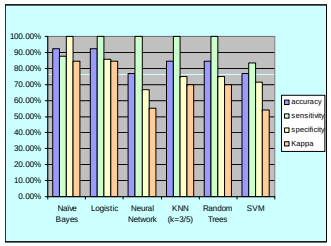
Zhen Li, Fathi Elloumi, Fred A Wright

Goal: to predict potential toxicity for 13 chemicals

Chemicals (13 in total and single dose):
1,5-Naphthalenediamine, 2,3-benzofuran, 4-Nitroanthranilic,N-(1-naphthyl)ethylenediamine dihydrochloride, benzene, coumarin, pentachloronitrobenzene, 2,2-bis(bromomethyl)-1,3- propanediol, 1,2-dibromoethane, 2-chloromethylpyridine hydrochloride, N-methylolacrylamide, diazinon and malathion.

Controls: corn oil, water, rodent chow

Performance of different classifiers



68

Project 2 Acknowledgements

Principal Investigator
Alexander Tropsha

Research Professors
Clark Jeffries, Alexander Golbraikh, Hao Zhu, Simon Wang

Postdoctoral Fellows
Georgiy Abramochkin, Lin Ye, Denis Fourches

Visiting Research Scientist
Aleks Sedykh

Adjunct Members
Weifan Zheng, Shubin Liu

Collaborators:
UNC: I. Rusyn, F. Wright, S. Gome
EPA: T. Martin, D. Young
A. Richard, R. Judson, D. Dix, R. Kavlock

Graduate Research Assistants
Christopher Grulke, Nancy Baker, Kun Wang, Hao Tang, Jui-Hua Hsieh, Rima Hajjo, Tanarat Kietsakorn, Tong Ying Wu, Liying Zhang, Melody Luo, Guiyu Zhao, Andrew Fant

Research Programmer
Theo Walker

System Administrator
Mihir Shah

MAJOR FUNDING
NIH
- P20-HG003898
- R21GM076059
- R01-GM66940
- R0-GM068665

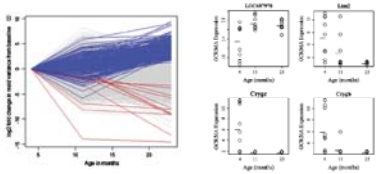
EPA (STAR awards)
- RD832720
- RD833825

TOXICOLOGICAL SCIENCES 107(1), 281–292 (2009)
doi:10.1093/toxsci/kfn215
Advance Access publication October 20, 2008

Age-Dependent Variability in Gene Expression in Male Fischer 344 Rat Retina

Zhen Li,* Fred A. Wright,* and Joyce Reyland†,‡

Needed method for quantifying changes in variability with high statistical power



70

Project Objectives, cont.

- Provide an interdisciplinary computer science resource to the environmental sciences and toxicology community
- Longer-term objectives include new software engineering methods for better execution and maintenance of above, and sharing and disseminating results

Overview of Carolina Center for Computational Toxicology STAR Program

October 1, 2009

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY

COMPUTATIONAL TOXICOLOGY

This work was reviewed by EPA and approved for presentation but does not necessarily reflect official Agency policy. Mention of trade names or commercial products does not constitute endorsement or recommendation by EPA for use.

The EPA's Task

$H_2C{=}O$ → **RfD**

EPA nomination held up amid debate over formaldehyde risks

September 24, **2009**

Environmental Protection Agency Administrator Lisa Jackson visited Sen. David Vitter, R-La., in his office Thursday to ask him to release his hold on the nomination of Paul Anastas to be the EPA's assistant administrator in charge of its Office of Research and Development. Vitter wants the EPA to agree to have the National Academy of Sciences review its assessment of the risks posed by **formaldehyde**, which is best known to folks in the Gulf Coast because of **respiratory complaints** lodged by people who lived in FEMA trailers with elevated levels of formaldehyde.

BRITISH MEDICAL JOURNAL 15 MARCH 1975

MEDICAL MEMORANDA

Formalin Asthma in Hospital Staff

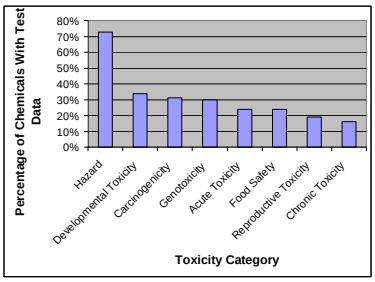
D. J. HENDRICK, D. J. LANE

British Medical Journal, 1975, 1, 607-608

Few cases of airways obstruction attributable to inhaled formaldehyde have been reported, though it has been suggested that the presence of formaldehyde contributes to the aggravation of chest diseases caused by air pollution (Kotin and Falk, 1964). Occupational "formalin asthma" has been described in a worker in a match factory (Vaughan, 1939) and in workers employed in the tanning and rubber industries (Popa *et al.*, 1969). We report here the use of inhalation provocation tests to investigate the relevance of inhaled formalin fumes to airways obstruction in two hospital staff members continually exposed to this substance in the course of their work.

Many Chemicals, Little Data

- Tox21 Priority List: ~19,000 Chemicals
- EPA, NTP, NCGC, FDA, OECD



Chemical Classes of Interest
High Production Volume
Medium Production Volume
Pesticidal Actives
Antimicrobial Actives
Pesticide Inerts
Drinking Water Contaminants
Air Pollutants
Superfund Priority Chemicals
Food Additives
Fragrances
Drugs
Green Chemicals
Persistent Chemicals
Active Metabolites

Toxicity Areas
-Carcinogens
-Genotoxicants
-Developmental Toxicants
-Developmental Neurotoxicants
-Reproductive Toxicants
-Endocrine
-Immunotoxicants

Computational Toxicology

- Addresses Issues of Too Many Chemicals, Too Little Data
- Priority Areas of Research and Method Development
  - –Prioritization
  - –Mechanism of Action
  - –Dose-Response Modeling
  - –Susceptible Populations

Carolina Computational Toxicology Center STAR Program: Project 1

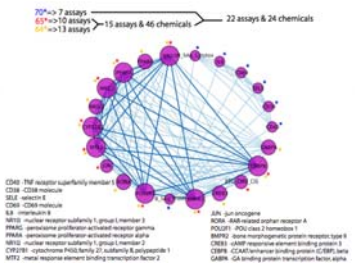
- Develop and apply data-driven methods for the inference and high-level modeling of regulatory network response to chemical perturbation
- Develop mechanistic models of nuclear receptor function
- Integrate and deploy high-, and low-level modeling tools

- Prioritization
- **Mechanism of Action**
- Dose-Response Modeling
- Susceptible Populations

Mechanism of Action

Network Context: Subgraph Mining

"Functional Module" in Subgraph Mining



- Mines binary data to find all frequent 'dense' sub-graphs (cliques)
  - Nodes: Assay
  - Edges: Set of 'Active' Chemicals shared between Nodes
  - Finds all unique subgraphs for a minimum frequency of 'Active' chemicals
- Differs from Hierarchical clustering by focusing on subsets of the data
- Useful for defining composite assays that might be more predictive
- Useful for associating Assay/Chemical combinations to endpoints

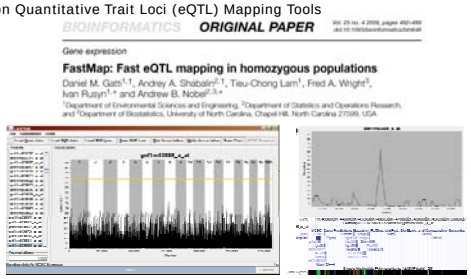
Carolina Computational Toxicology Center STAR Program: Project 2

- Development of Fast and Efficient Toxicogenetic Expression Quantitative Trait Loci (eQTL) Mapping Tools
- Discovery of the chemical-induced regulatory networks using the population-based toxicity phenotyping in human cells

- Prioritization
- **Mechanism of Action**
- Dose-Response Modeling
- **Susceptible Populations**

Susceptible Populations
Mechanism of Action

Specific Objective 1: Development of Fast and Efficient Toxicogenetic Expression Quantitative Trait Loci (eQTL) Mapping Tools

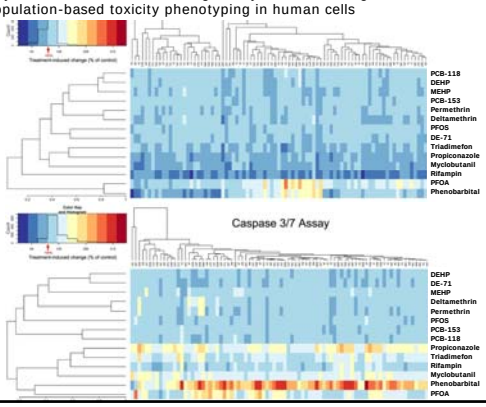


- Java-based GUI which runs on a standard desktop PC
- Amenable to "proprietary" data
- Single marker or k-SNP window association mapping
- Permutation-based significance testing of the eQTLs
- Extended options for export of data/images and a link to UCSC genome browser

---

Susceptible Populations
Mechanism of Action

Specific Objective 3:
Discovery of the chemical-induced regulatory networks using the population-based toxicity phenotyping in human cells



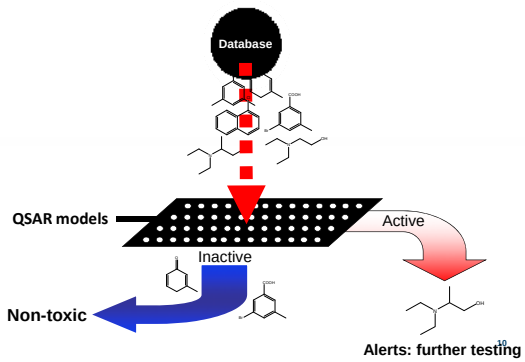
Carolina Computational Toxicology Center STAR Program: Project 3

- Develop rigorous end point toxicity predictors based on QSAR modeling workflow and conventional chemical descriptors
- Develop novel computational toxico-genomic models based on combined chemical and biological descriptors through QSAR modeling workflow
- Develop novel computational toxico-genetic models based on combined genetic, chemical and toxicity descriptors through QSAR-like modeling workflow

- **Prioritization**
- **Mechanism of Action**
- Dose-Response Modeling
- Susceptible Populations

Prioritization

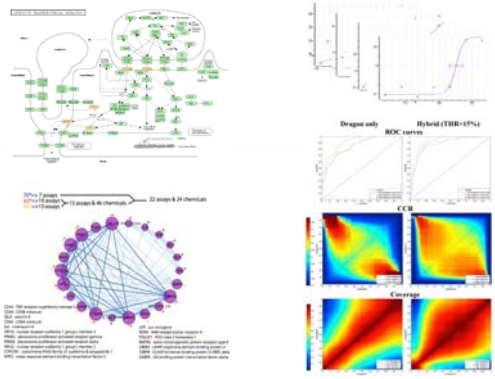
Compound prioritization using the ensemble of QSAR models



Mechanism of Action

Developing Novel Bio-Descriptors

Pathway-derived | **Dose-response-derived**



Summary

- Carolina Center for Computational Toxicology is developing promising new approaches to address EPA CompTox research areas of:
  - Prioritization
  - Mechanism of Action
  - Susceptible Populations
- Can some of these methods be extended to help understand dose-response relationships?

The Texas-Indiana Virtual STAR center;

Data-Generating *in vitro* and *in silico* Models of Developmental Toxicity in Embryonic Stem Cells and Zebrafish

Jan-Åke Gustafsson, Richard H. Finnell and James A. Glazier

University of Houston, Texas A&M, Indiana University

November 2009-October 2012

Background

Birth defects

Birth defects affect about one in every 33 babies born in the United States each year (3%) (6% worldwide). They are the leading cause of infant deaths, accounting for more than 20% of all infant deaths. Babies born with birth defects have a greater chance of illness and long term disability than babies without birth defects.

Heart defects: 1 in every 100 to 200 babies

Neural tube defects: defects of the spine (spina bifida) and brain (anencephaly). 1 of 1,000 pregnancies (2.6/1000 worldwide)

Orofacial clefts: include cleft lip, cleft palate, and combined cleft lip and cleft palate.

1 in 700 to 1,000 babies

Reasons

Genetic and environmental factors

Methyl mercury:

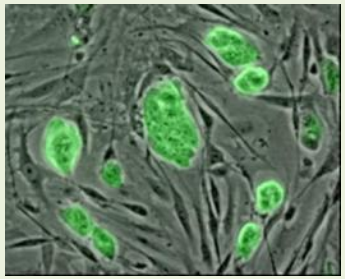
The birth defects are small head size, cerebral palsy, developmental delay and/or mental retardation, blindness, muscle weakness, and seizures.

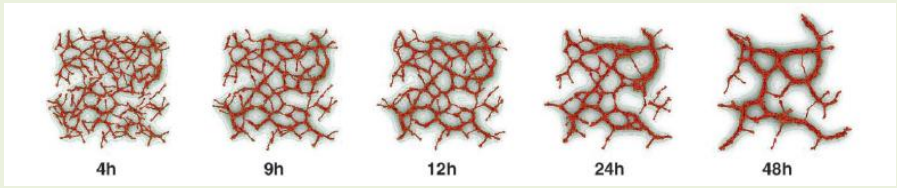
Knowledge gap!

Research objective

New screening models for developmental toxicity

From Biological Models of Developmental Toxicity to Computer Simulations





Main research goals

1. Generate developmental models based on mouse embryonic stem cells and zebrafish suitable for high-throughput screening.

2. Generate high-information-content models on development and differentiation using mouse embryonic stem cells and zebrafish.

3. Develop computational models for developmental toxicity with the ultimate aims of first recreating normal development (in wild-type) and then classifying possible mechanisms by which chemical perturbations cause experimentally observed developmental defects.

4. Perform proof-of-concept experiments of the *in vitro* and *in silico* test platforms with a blind test of chemicals.

Investigational Areas

Three Investigational Areas:

1. Zebrafish as a model to elucidate the morphological and mechanistic effects of environmental pollutants.

   PI Jan-Åke Gustafsson

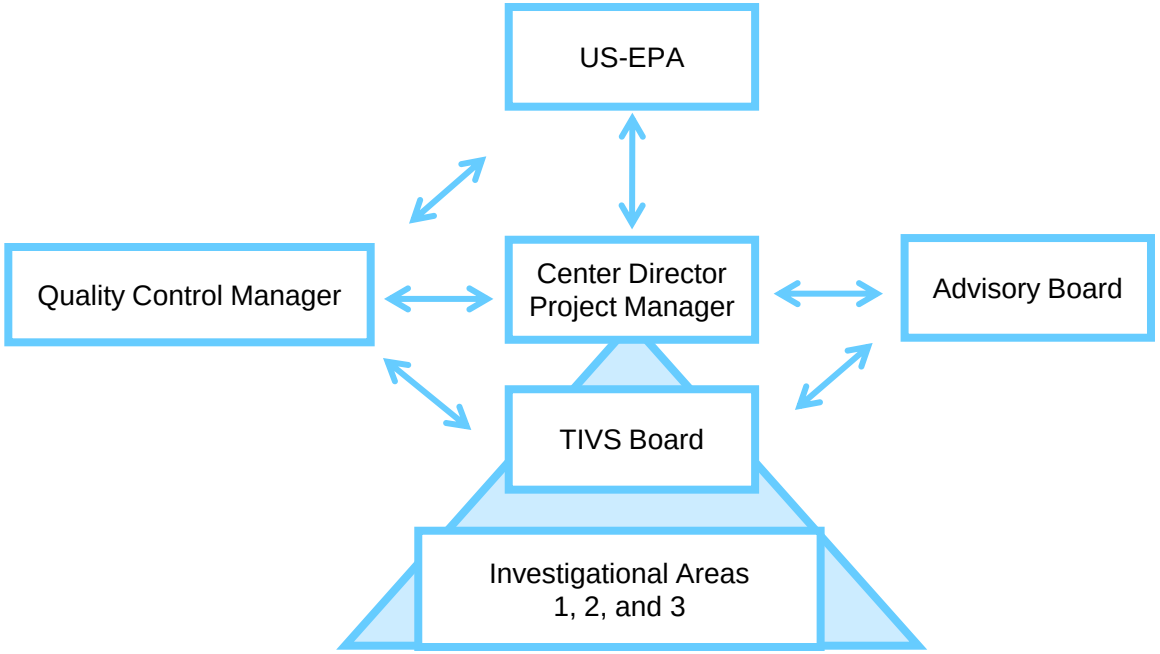
2. The effects of environmental contaminants on mouse embryonic stem cell differentiation.

   PI Richard H. Finnell

3. Development of computer simulations facilitating assessment of toxicity based on perturbed development in zebrafish and mouse embryonic stem cells.

   PI James A. Glazier

Management



TIVS board
1 representative from each IA
Main decisions

Center Director/Project Manager
Operational management
Reporting
Fiscal responsibility

Quality Control Manager
Donald P. McDonell, Duke University

Advisory board
Advice and Evaluate
George Daston, Procter and Gamble
Nadine Peyrieras, CNRS, Paris
Helen Håkansson, Karolinska Institutet, Stockholm
Menghang Xia, NCGC, NIH
Bart van der Burgh, ChemScreen (EC-funded project on ENV.2009.3.3.1.1)
STAR Center representatives

Teaching and information

Courses

Three courses for PhD students and post docs:

1. Zebrafish development
2. Embryonic stem cells
3. Computer simulations

Posted on our website
www.cnrcs.uh.edu/TIVS-Center

Information

Develop public web
Internal web
Meetings, workshops, newsletters

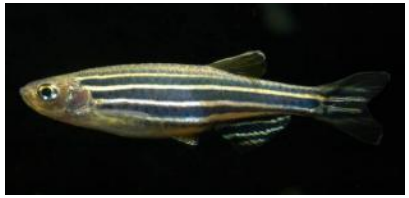
Collaboration with stakeholders and other projects

OECD, WHO, ChemTRUST
STAR Centers
Chemscreen, Cascade, Crescendo, Ceasar, Carcinogenomics, SafeFoods, Rainbow, RA-Courses, **TRISK**

Zebrafish as a model to elucidate the morphological and mechanistic effects of environmental pollutants

Zebrafish, Danio rerio

- Small size, small test volumes
- Transparent embryos/fish
- External rapid embryonic development
- Hundreds of eggs weekly/pair
- Genome sequenced,
75% of genes have human homologues
- Conserved developmental processes and
signaling pathways
- Many mutants
- Morpholino knockdown
- Cost efficient
- Adaptable to medium to high through put screening



Generation of screening models for teratogens

10 transgenic fish expressing fluorescent markers to follow development and patterning.

Endpoints:

- Gastrulation and early embryonic cell movements
- Patterning of CNS and neurogenesis
- Hematopoiesis and angiogenesis
- Yolk utilization and morphological effects on somitogenesis

Morphology and GFP/RFP expression will be recorded during normal development.
Is development changed by teratogenic chemicals?

Scale up and automate for high throughput screening

Transgenic fish for screening

	Gene	HTTA	Reporter Status	Readout	Start time of expected expression (hpf)
1	*goosecoid*	Early patterning, epiboly, early cell movements and developmental delay	RFP- to be made	Time of appearance/disappearance, Spatial distribution of expression domain, intensity of expression	3.5 hpf
2	*dharma*	Early patterning, epiboly, early cell movements and developmental delay	GFP – to be made	Time of appearance/disappearance, Spatial distribution of expression domain, intensity of expression	3.5 hpf
3	*bmp2b*	Patterning (anterior-posterior symmetry), early cell movements	GFP-to be made	Total length of expression domain, Time of appearance/disappearance, Spatial distribution of expression domain, intensity of expression	1 cell stage 0 hpf (maternal contributed)
4	*wnt8*	Patterning (anterior-posterior symmetry), early cell movements	GFP-to be made	Total length of expression domain, Time of appearance/disappearance, Spatial distribution of expression domain, intensity of expression	1 cell stage 0 hpf (maternal contributed)
5	*bmp4*	Patterning (left-right symmetry)	GFP-to be made	Total length of expression domain, Time of appearance/disappearance, Spatial distribution of expression domain, intensity of expression	10 hpf
6	*ngn1*	Neurogenesis, Axon guidance, early, developmental delay	GFP/RFP-available	Time of expression, region of expression, intensity, cell numbers, axonal length and pathfiinding	10 hpf
7	*fli1*	Angiogenesis and blood vessel remodeling, heart morphology and function	EGFP-available with us	Time of expression, region of expression, intensity, angiogenesis, blood flow, heart size, rate of heart beat, number and size of trunk vessels	11 hpf
8	*flk1*	Angiogenesis and blood vessel remodeling, heart morphology and function. Expressed in tip cells.	GFP-available with us	Time of expression, region of expression, intensity, angiogenesis, blood flow, heart size, rate of heart beat, number and size of trunk vessels	11 hpf
9	*Unc5b*	Blood vessel formation, expressed in tip cells at the forefront of arterial and venous sprouts.	RFP-to be made	Time of expression, region of expression, intensity, angiogenesis	9hpf
10	*unc45b*	Muscle development and somitogenesis	GFP-available	Somite formation, somite size, time of appearance, muscle formation, intensity, spontaneous movements, time and region of appearance	9hpf

Generation of high-information-content models

- Somite formation
- Blood-vessel formation
- Axonal pathfinding

Map expression of crucial factors
Adhesion factors
Repulsion factors

Immunostaining, *In situ* hybridization

Knockdown of crucial factors

Morpholino knockdown

Simulations *in silico*

Test chemicals

37 CERCLA chemicals known or expected to be teratogens and associated with developmental malformations

Rank number indicates the potential threat to human health of these environmental pollutants as determined by ATSDR and the EPA.

Abbreviations:
Chemical Abstracts Service (**CAS**), central nervous system (**CNS**), gastrointestinal (**GI**), genitourinary (**GU**), musculoskeletal (**MS**).

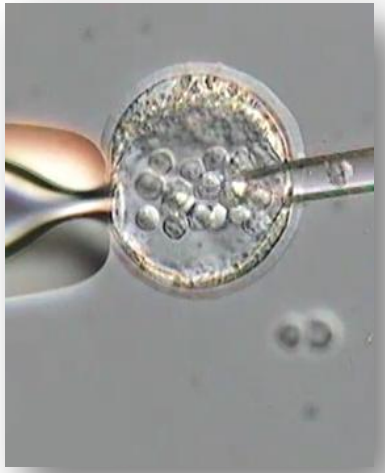
(The Comprehensive Environmental Response, Compensation, and Liability Act, **CERCLA**)

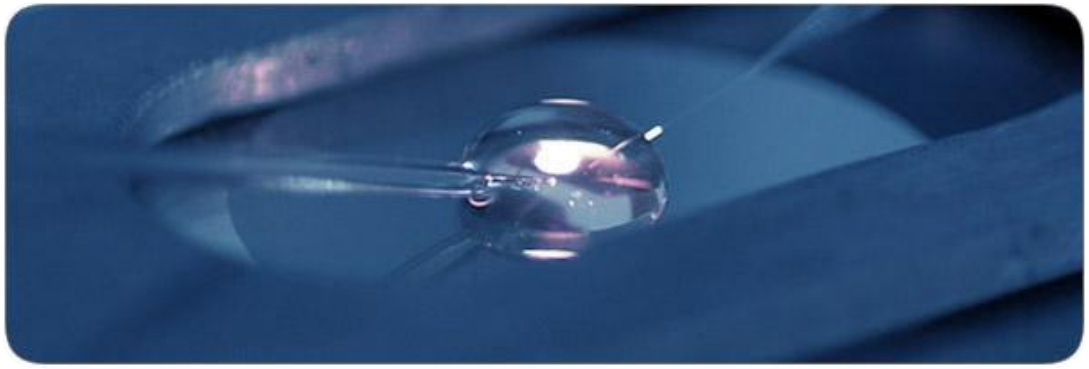
RANK	SUBSTANCE NAME	CAS #	CNS	Eye	Heart	GI	GU	MS
1	arsenic	007440-38-2	+	+	+	+	+	+
2	lead	007439-92-1	+	+	-	+	+	-
3	mercury	007439-97-6	+	+	+	+	+	+
4	vinyl chloride	000075-01-4	+	+	-	-	-	+
5	polychlorinated biphenyls	001336-36-3	+	+	+	+	+	+
6	benzene	000071-43-2	+	+	+	+	+	+
7	cadmium	007440-43-9	+	+	+	-	+	+
11	chloroform	000067-66-3	+	+	-	-	+	+
16	trichloroethylene	000079-01-6	-	-	-	+	+	+
17	dieldrin	000060-57-1	-	+	+	-	+	+
24	aldrin	000309-00-2	-	-	-	-	+	-
45	pentachlorophenol	000087-86-5	-	-	-	-	+	-
47	carbon tetrachloride	000056-23-5	+	+	+	+	+	+
53	nickel	007440-02-0	+	-	-	-	+	+
54	endosulfan	000115-29-7	+	+	+	-	+	+
61	methoxychlor	000072-43-5	-	+	-	-	+	-
71	toluene	000108-88-3	-	+	-	-	+	-
78	naphthalene	000091-20-3	+	+	+	-	+	+
80	methylene chloride	000075-09-2	+	+	-	-	+	-
84	hydrazine	000302-01-2	+	+	+	-	+	+
93	hexachlorobenzene	000118-74-1	+	+	+	-	+	-
94	2,4-dinitrotoluene	000121-14-2	-	-	-	-	-	+
145	parathion	000056-38-2	-	-	-	-	+	-
147	selenium	007782-49-2	-	+	-	-	+	-
176	carbon disulfide	000075-15-0	-	-	-	-	-	+
182	phenol	000108-95-2	+	-	-	-	+	+
189	carbon monoxide	000630-08-0	+	+	+	-	+	-
209	2,4-dichlorophenol	000120-83-2	-	-	-	-	+	-
224	arsenic trioxide	001327-53-3	+	+	+	+	+	+
240	dichlorvos	000062-73-7	-	+	-	-	+	-
241	sodium arsenite	007784-46-5	+	+	+	+	+	+
244	formaldehyde	000050-00-0	+	-	-	-	-	-
250	diuron	000330-54-1	+	-	-	-	+	+
264	methyl parathion	000298-00-0	-	+	-	-	+	-
271	styrene	000100-42-5	+	+	+	-	+	+
272	carbaryl	000063-25-2	-	-	-	-	+	-
274	acrylonitrile	000107-13-1	+	-	-	-	-	-

Mouse embryonic stem cells as a model to elucidate the morphological and mechanistic effects of environmental pollutants

House Mouse, *Mus musculus*
- Mouse genes (99%) have homologues in humans
- Relatively short gestational age

Mouse Embryonic Stem Cells
- Small size, small test volumes
- Conserved developmental processes and signaling pathw
- Mimic *in vivo* development
- Amendable to genetic manipulation
- Cost efficient
- Adaptable to medium to high through put screening





Embryonic Stem Cell Differentiation

In the Beginning…
ES cells must be isolated and maintained or else…
ES cells differentiate into epiblast

Epiblast gives rise to
embryoid body &
germ layer cells

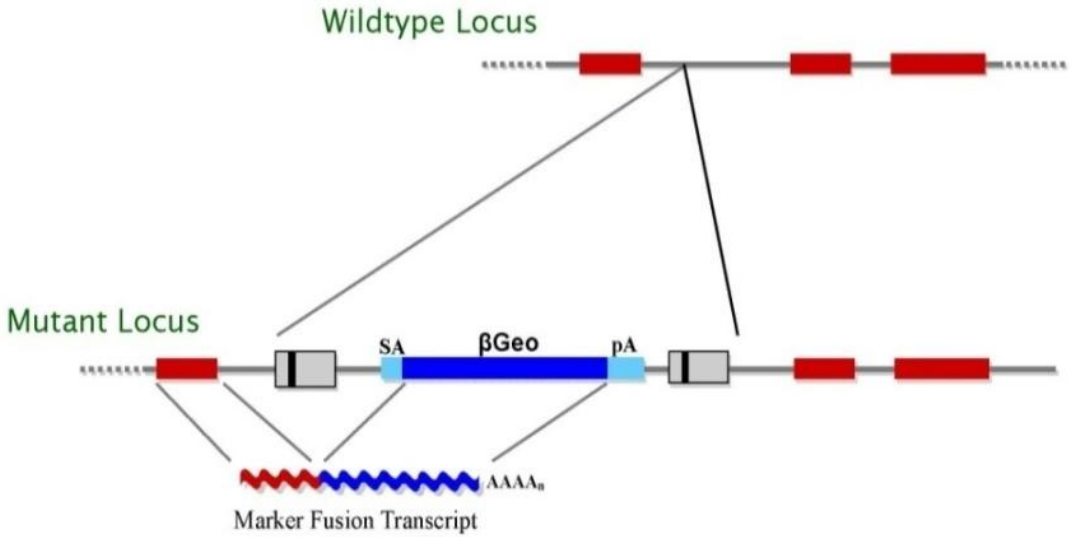
Germ layer cells
differentiate into specific cell types

Genetic Manipulation of Mouse ES Cells: Gene Trap

C57Bl/6 Gene Trap Library

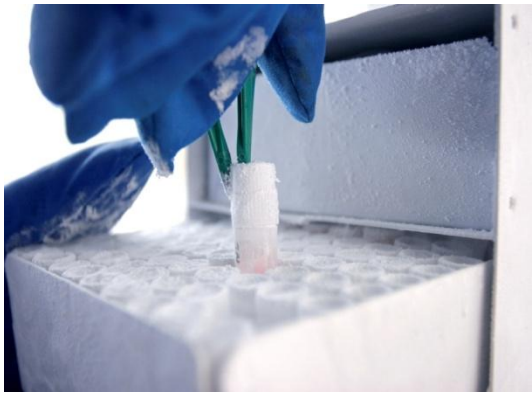
- Retrovirus inserts transgenic construct
- > 350,000 ES cell clones produced
- > 10,000 genes contain inserts
- ROSA β-geo gene trap vector (marker)

Retroviral gene trapping vector



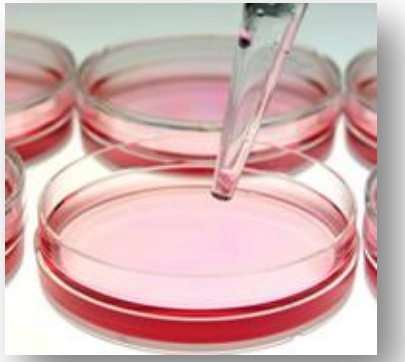
Selection and Generation of ES Based Screening Models

16 transgenic mouse ES cells expressing a reporter (β-geo) thawed and cultured:



Selected Genes:
Follow developmental and patterning processes.

Including:
Gastrulation and early embryonic cell movements
Patterning of CNS and neurogenesis
Hematopoiesis and angiogenesis



Expected Reults:
Documentation of morphology and β-geo expression during:

- normal development
- teratogenic chemical exposure

Scale up and automate for high throughput screening

Selected Transgenic β-geo Mouse ES cells for Screening

Gene	Name	Function/Expression
Nodal	Nodal	Interacts with type I receptor complexes: ALK4 and ALK7, and type II receptors: activin receptor 2a or 2b
Wnt3	wingless-related MMTV integration site 3	Wnt signaling ligand
Fgf4	fibroblast growth factor 4	FGF signaling ligand
Gsc	Goosecoid	homeodomain transcription factor, executer of cell migration during **gastrulation**
Cdh1	cadherin 1 (E-cadherin)	calcium ion-dependent cell adhesion molecule in **epithelial cells**
Pou5f1	POU domain, class 5, transcription factor 1	regulation of **pluripotency** during normal development
Meox1	mesenchyme homeobox 1	homeobox gene expressed in mesoderm of **primitive streak and somites**
Bmp4	bone morphogenetic protein 4	**bone and cartilage** development
Mapt	tau	**neuronal microtubule** associated protein
Syn1	synapsin I	**synaptic vesicle** glycoprotein present in cells involved in synaptic transmission
ABCG2	ATP-binding cassette superfamily G member 2	stem cell and hematopoietic **stem cell** marker
Tie1	tyrosine kinase with immunoglobulin-like and EGF-like domains 1	angiopoietin receptors and e**ndothelial** marker
Pcam1	platelet/endothelial cell adhesion molecule 1	cell adhesion molecule and **endothelial** marker
GATA3	GATA binding protein 3	transcription factor in **myocytes**
Mef2a	Myocyte-Specific Enhancer Factor 2a	transcription factor in **myocytes**
Myl2	myosin light chain 2V	regulatory light chain associated with **cardiac** myosin beta

Alternative Transgenic β-geo Mouse ES cells for Screening

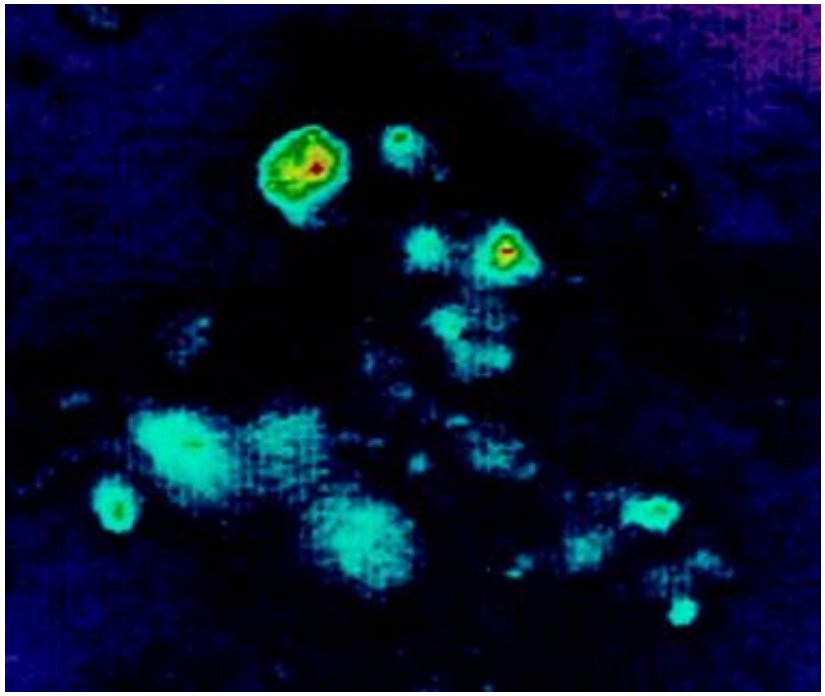
In the event that selected clones do not pass quality control, or are not responsive to chemical insults, alternative gene/clones are also available, *e.g.* :

Gene	**Name**	**Function/Expression**
Smad1	MAD homolog 1	Proteins that modulate the activity of TGF β ligands
Prdm14	PR-domain containing protein 14	Functions in PGC specification
Spred	Sprouty-related protein with an EVH1 domain	Regulates Ras-ERK signaling pathway
Zic family	Zinc finger protein of the cerebellum	Neural development
Zic2		
Zic5		
VEGF	Vascular endothelial growth factor	VEGF signaling ligand
Vegfb		
Vegfc		
Notch1	Notch gene homolog 1	Functions in vascular remodeling during development

Generation of High Throughput/Information Content Models

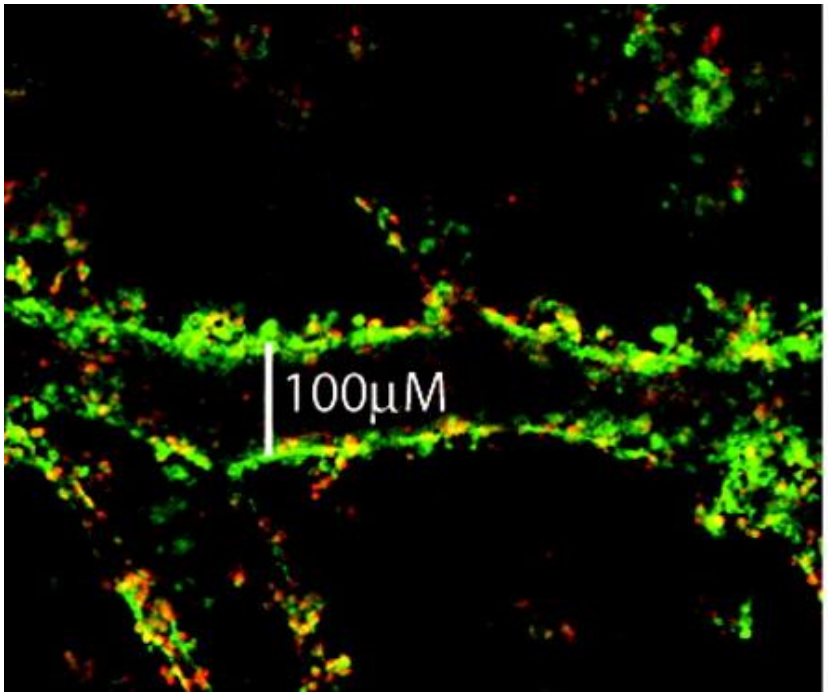
Detection of transgenic ES cell β-geo (*lacZ)* expression:
*In Vivo (*ImaGene Green, Invitrogen)

Imagene Green staining of ES cell-derived spontaneously contracting cardiac myocytes



Circulation Research. 1996;78:547-552.

Imagene Green and propidium iodide staining of *in vitro* endothelial differentiation

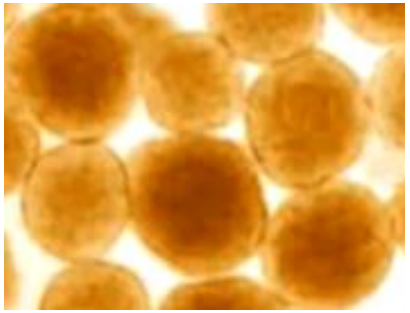


Arteriosclerosis, Thrombosis, and Vascular Biology. 2004;24:691

Generation of High Throughput/Information Content Models

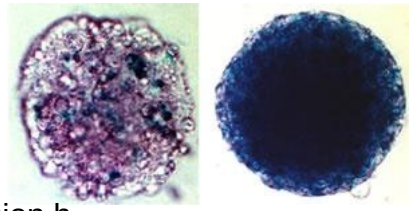
ES cells

Standardized Embryoid Body Production



Application of teratogen or test chemical(s)

Differentiatation & Detection of β-geo expression



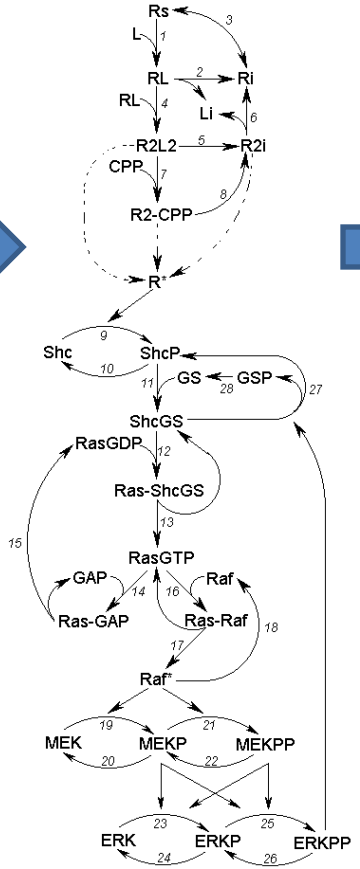
www.cmhd.ca/genetrap/database/search_expression.html

Simulations *in silico*

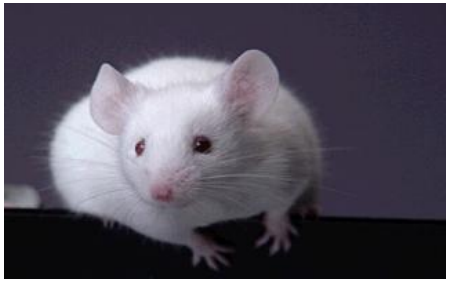
Development of computer simulations facilitating assessment of toxicity based on perturbed development in zebrafish and mouse embryonic stem cells

Multi-cell modeling provides a platform to go from molecule to cell behavior to development.

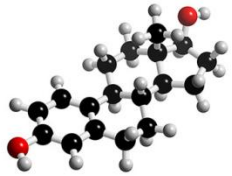
http://nomadlife.org/dna.jpg



http://www.stanford.edu/group/Urchin/LP/
[Lauren Palumbi]



http://www.kvarkadabra.net/images/articles/Regeneracija-organov_1_original.jpg



lowestrogensymptoms.com

http://amazingphotos4all.blogspot.com/2009_03_01_archive.html

Multi-cell Modeling as a Bridge from *in vitro* to Organ/Organism

Still a huge gap between level of molecular data and observed developmental patterns.

Multi-cell Models separate two questions:

- How do molecular processes drive cell phenomenology?
- How does cell phenomenology drive tissue-level patterning?

Why useful?

- Brute force (molecule→organism) computationally intractable.
- Allows focus on key molecular pathways. And cell-cell interaction mechanisms.
- Most mammalian cells are fairly limited in their behaviors, simplifying model construction.
- Rapidly developing tools and standards.

Data Inputs for Multi-cell Modeling

Organ/Organism level:

Qualitative selection of model developmental systems.

Quantitative study of normal and perturbed development of these.

Cell tracking (*in vivo*).

Expression mapping (*in vivo* and *in vitro*).

Identification of key ECM & extracellular signals (*in vivo* and *in vitro*).

Cell level:

Qualititative identification of key cell types.

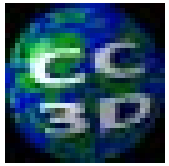
Quantitative descriptions of their phenomenology *in vivo* and *in vitro.*

Molecular level:

Qualitative identification of key regulatory pathways (*in vivo* and *in vitro*).

Quantitative description these pathways and their perturbations (*in vitro)*.

CompuCell3D (Indiana University, Bloomington)
Multi-Cell Modeling Environment



Open-Source, Multi-Platform Simulation Environment:

- **Simulations Based on Cell Behaviors**
- **Simulation Specification in High-Level Language (CC3DML, Python)**
- **Fast Simulation Development**
- **Reuse of Simulation Components**
- **Connects to Systems Biology Workbench for Pathway Modeling**

http://www.compucell3d.org/

Systems Biology Workbench (U. Washington, Seattle)
Reaction-Kinetics Modeling Environment

Open-Source, Multi-Platform Simulation Environment:

- **Simulations Based on Molecular Reactions**
- **Simulation Specification in High-Level Language**
- **Fast Simulation Development**
- **Reuse of Simulation Components**
- **Connects to CompuCell3D for Multi-Cell Modeling**

http://www.sys-bio.org/

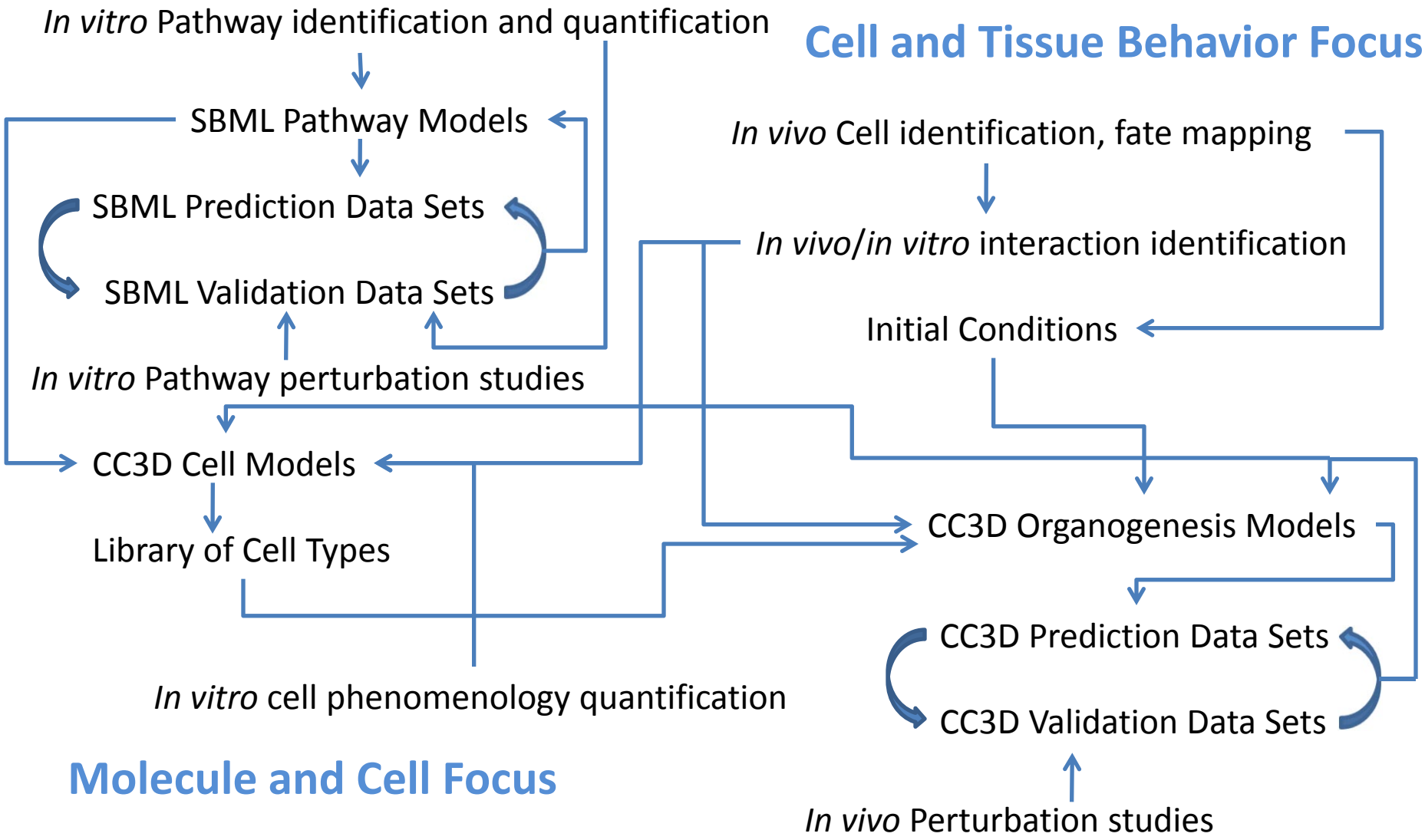
Cell Behavior Ontology/ Cell Behavior Model Specification Language (Under Development)

Community-Oriented Language Development

- **Implementation-Independent Specification of Multi-Cell Models**
- **Improved Annotation of Microscopy Data for High-Throughput Experiments and Model Generation**
- **Unification of SBML and CC3DML**

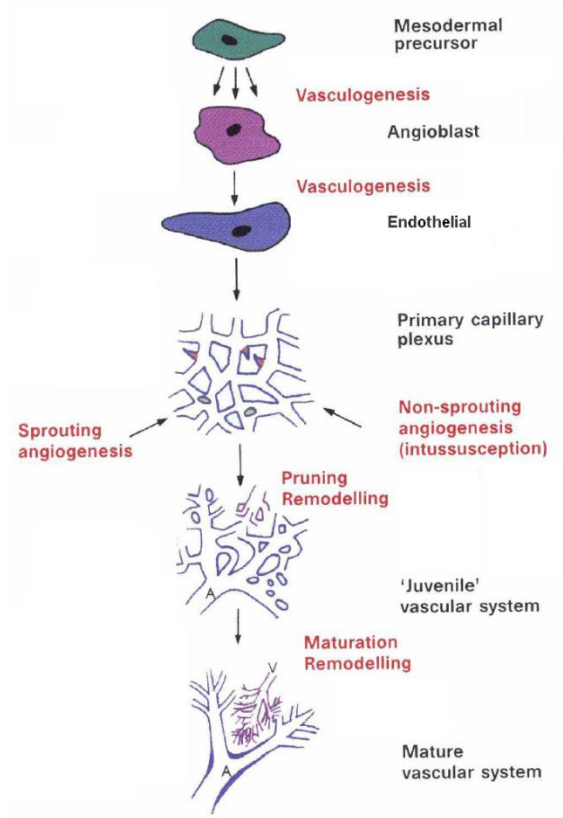
http://bioportal.bioontology.org/ontologies/39336

Information Flow



Existing CC3D Applications (I) Role of VE-Cadherin in Angiogenesis

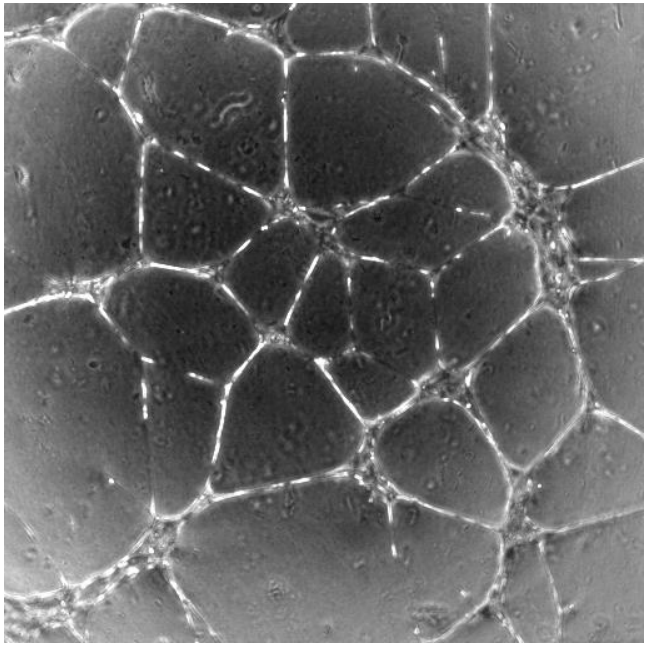
- Vasculogenesis
  - The formation of early vascular plexus from in situ differentiated **Endothelial Cells** (***ECs***)



Werner Risau, *Nature* **386**, 671 - 674 (1997)

- Angiogenesis
  - The formation of new blood vessels from pre-existing ones
    - Sprouting Angiogenesis
    - Non-sprouting Angiogenesis (**Intussusceptive angiogenesis**)

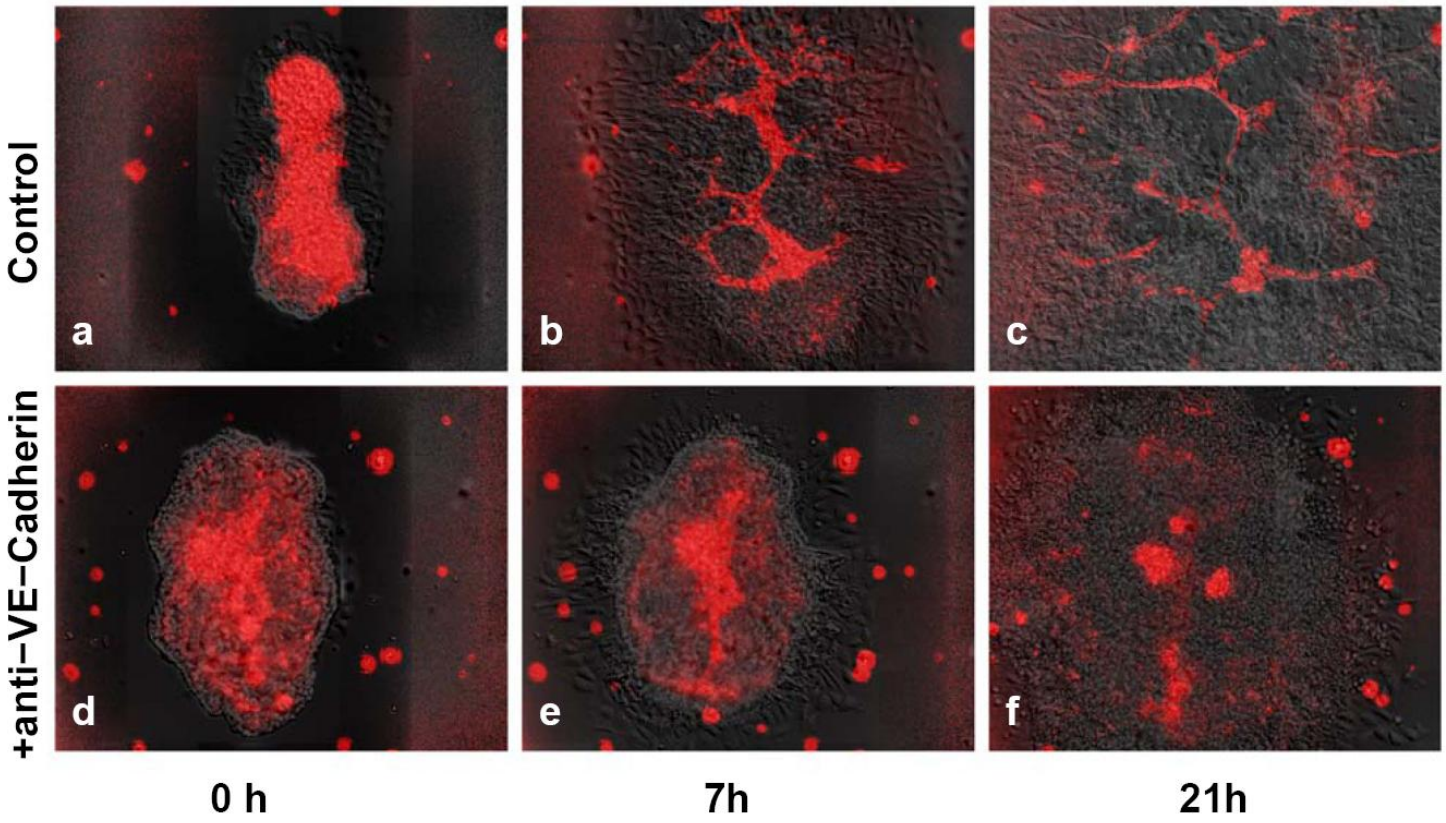
***In Vitro* HUVEC Model**



D. Ambrosi et al., Phys. Rev. Letters **90**, 118101

Existing Applications (I) Role of VE-Cadherin in Angiogenesis

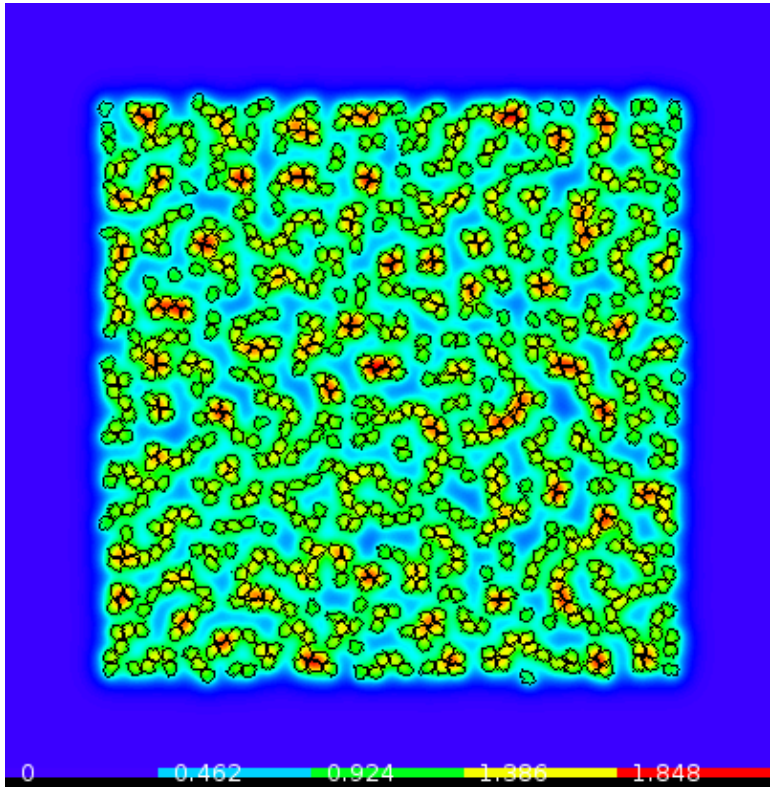
- **VE-Cadherin** (an adhesion molecule) clusters at adherens junctions between endothelial cells and **suppresses chemotaxis** at cell-cell interfaces



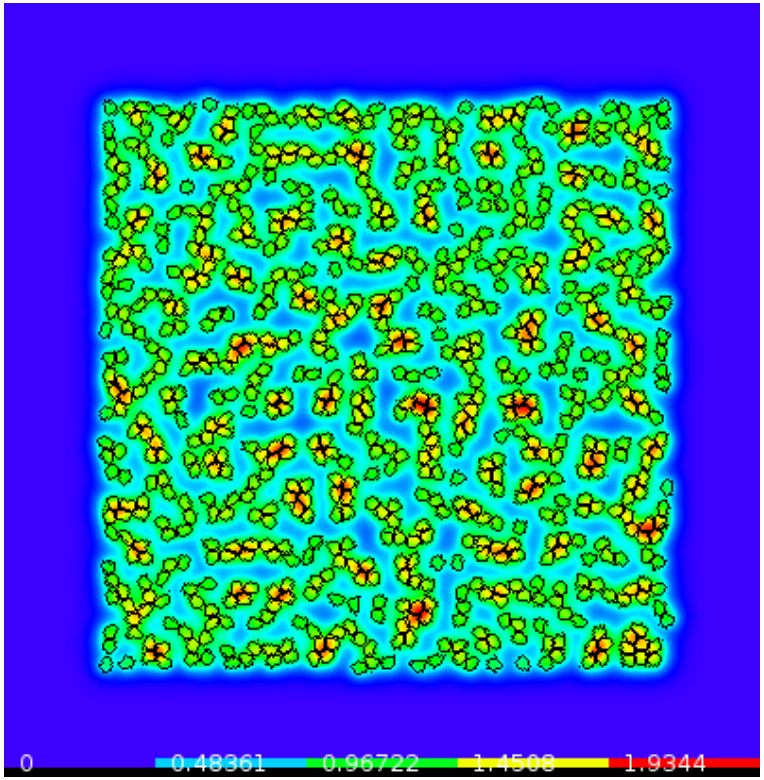
Anti-VE-cadherin antibody inhibits *de novo* blood-vessel growth in mouse allantois cultures. (Roeland M. H. Merks , Erica D. Perryn , Abbas Shirinifard, and James A. Glazier, *PLoS Computational Biology* 2008)

Existing Applications (I) Role of VE-Cadherin in Angiogenesis

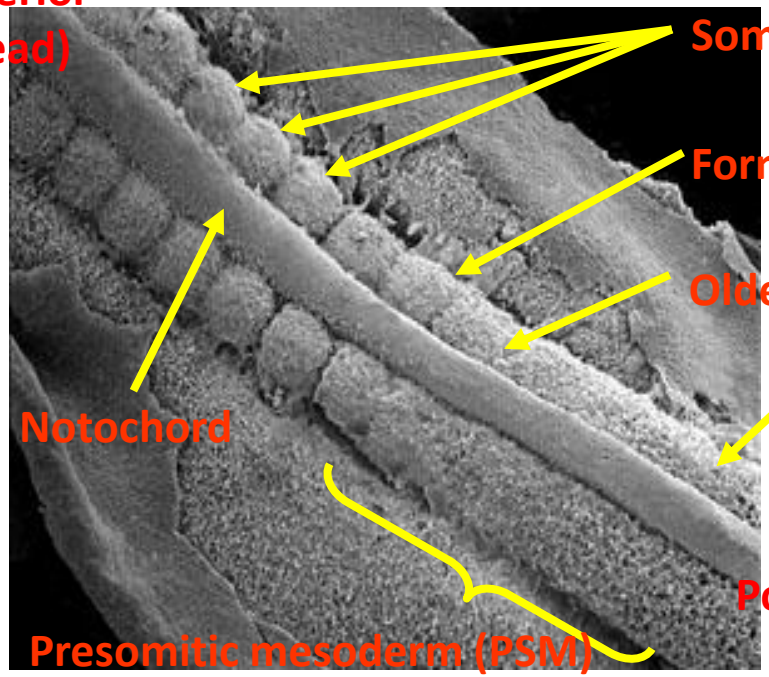
Wild Type Simulation

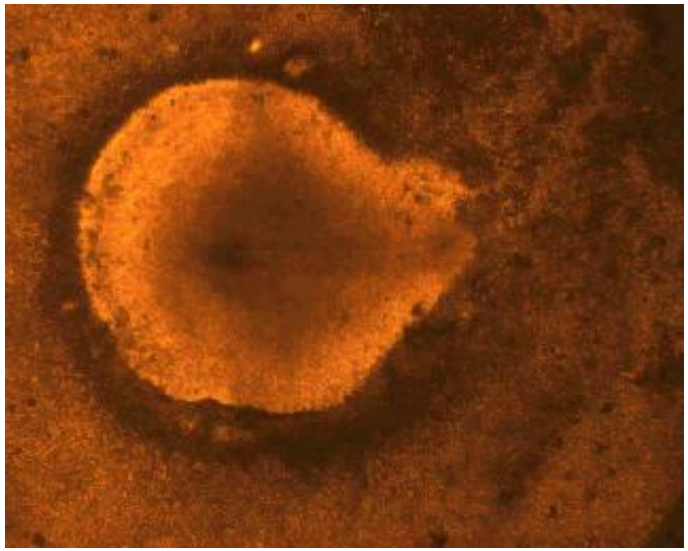


VE-Cadherin Knockout Simulation

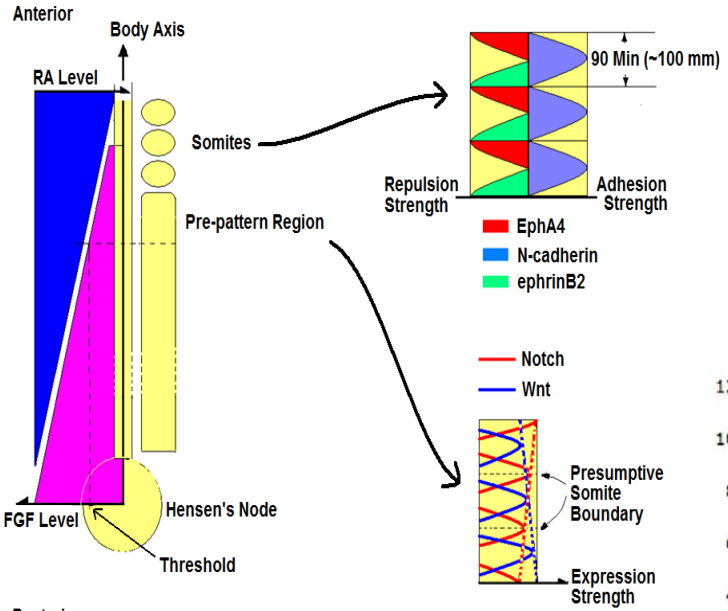


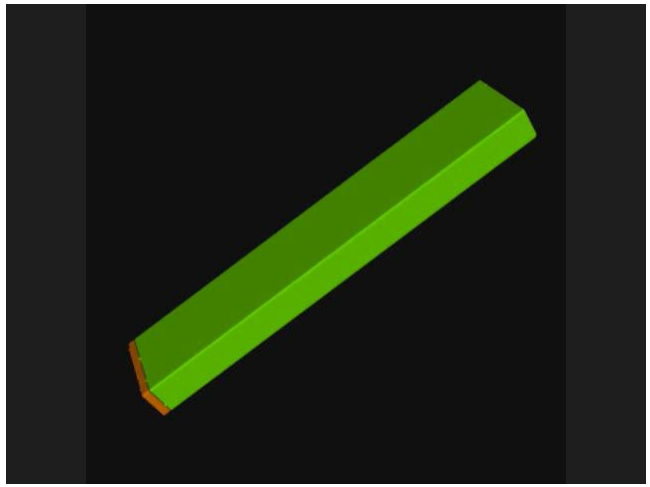
Existing Applications (II) Role of N-Cadherin in Somitogenesis

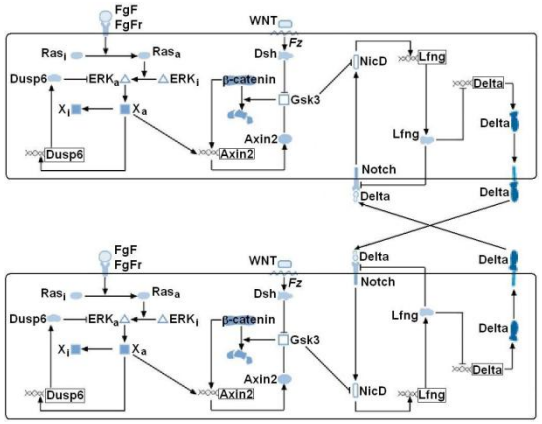




Existing Applications (II) Role of N-Cadherin in Somitogenesis

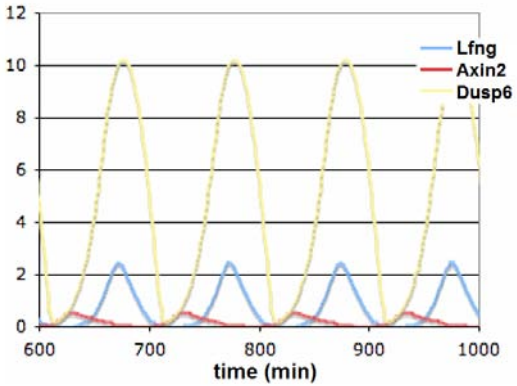


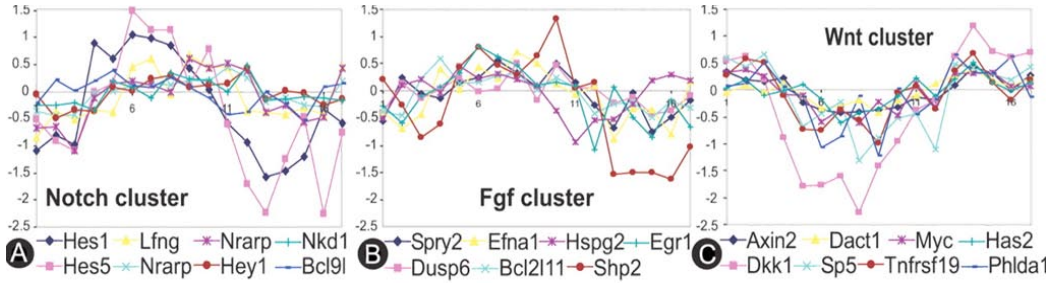




(Lewis *et al*. 2003)

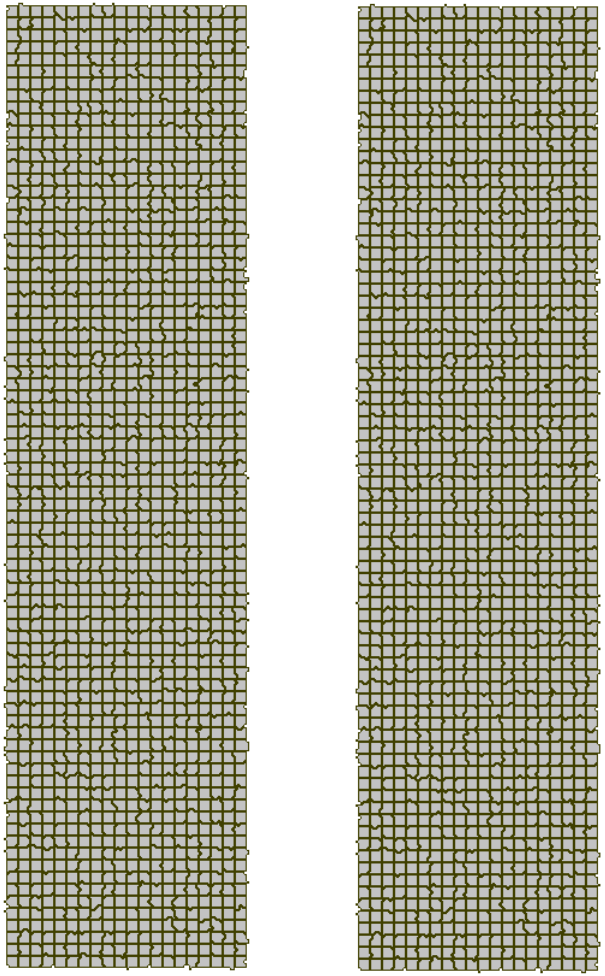
(Goldbeter & Pourquié 2008)

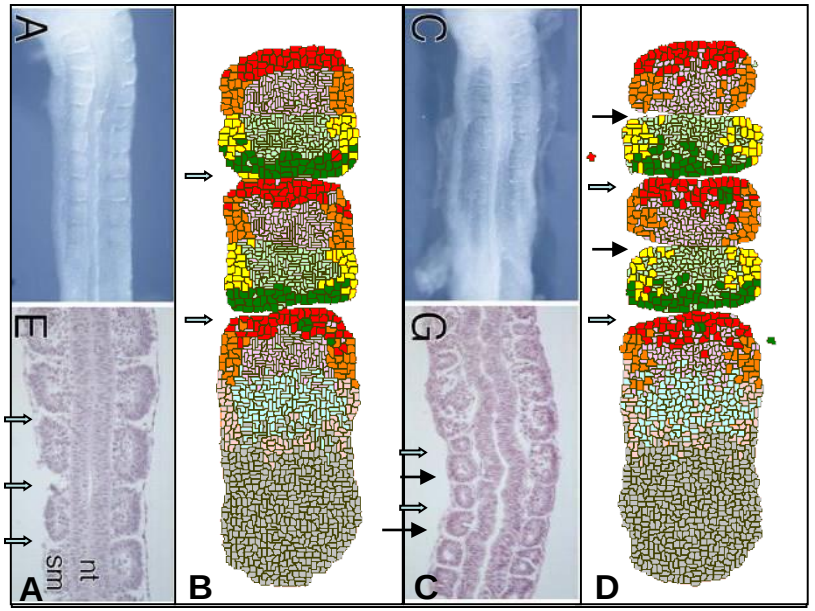




Dequeant *et al*. 2006 (microarray time series of mRNA in mouse)

Existing Applications (II) Role of N-Cadherin in Somitogenesis





N-cadherin knockout

Multi-Cell Modeling as a Predictive Tool

Multi-cell modeling in CompuCell3D+SBW will integrate molecular, cellular and whole-organ level data to predict developmental effects of pathway disruption.

Allows construction of standard libraries for reuse of information.

Lack quantitative experimental data to build/validate simulations:

- Cell Tracking
- Mechanics
- Pathways
- Interactions
- Morphology

TIVS will provide these data.

Collaborations: Texas-Indiana Virtual STAR Center

Thomas B. Knudsen, PhD
National Center for Computational Toxicology

Disclaimer: views are those of the presenter and do not necessarily reflect Agency policy

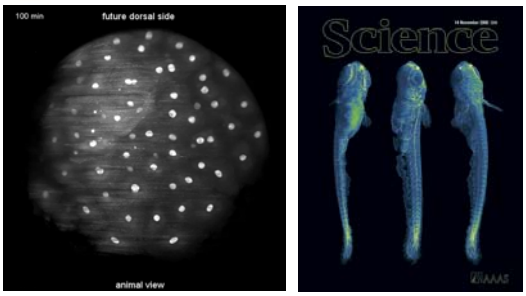
Developmental Toxicity

- RFA-EPA-G2008-STAR-W:
  Computational Toxicology Research Centers: in vitro and in silico models of developmental toxicity pathways
- exposures that perturb biological events during formative stages of the reproductive cycle affecting:
  - embryo and fetal development
  - postnatal development
  - fertility and reproduction
  - children's health

Some key research issues …

1. TIMING: morphogenesis and differentiation require precisely timed genetic signals and responses
2. SENSITIVITY: metabolic and regulatory pathways are prone to genetic errors and environmental disruptions
3. COMPLEXITY: simple lesions propagated to complex phenotypes or complex lesions → simple phenotypes
4. MATERNAL FACTORS: impact of maternal exposure biology during prenatal and lactational stages

Cellular dynamics



Zebrafish tracked with H2B-EGFP
by DSLM at 90s intervals over 18h

Source: Keller et al. (2008) Science 322: 1065-69

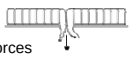
Fundamental processes

Core developmental processes
- patterning (sets up future events)
- timing (clocks and oscillators)
- differentiation (cell diversification)
- morphogenesis (tissue organization)

Cellular primitives
- growth (proliferation)
- death (apoptosis)
- differentiation (function)
- adhesion (DAH)
- shape (geometry)
- motility (cell migration)
- ECM (remodeling)

Morphogenetic movements
- folding
- epiboly

- convergent extension
- branching morphogenesis
- cell condensation
- cell sorting
- trans-differentiation
- cavitation
- involution



- tractional forces

After: Bard (2005) J Anat 206: 1 - 16

TIVS Project 1: zebrafish development

Zebrafish as a model to elucidate the morphological and mechanistic effects of environmental pollutants

- Padilla (EPA): pathways linking to developmental & neurodevelopmental endpoints
- data sharing: same compounds to confirm (+)ves and (-)ves across labs and fish strains
- resource sharing: reporter fish lines, existing (vegF) and new (STAR), for functional analysis of specific pathways

Developmental signaling pathways

PHASE	SIGNALING PATHWAY
Early development	Wnt family (Wnt1 ... 19) TGF-beta (BMP, Nodal, Activin, TGF-b) Hedgehog family (Shh, Ihh, Dhh) Receptor tyrosine kinase (EGF, FGF, IGF, EPH) Notch-Delta (Not-Dl) cytoplasmic tyrosine kinase (cytokine)
Middle development	TLR NF-KB (IL-1) Nuclear hormone receptors (steroid, retinoid) Apoptosis Receptor phosphotyrosine phosphatase
Late development	Receptor guanylate cyclase Nitric oxide receptors G-protein coupled receptors Integrin-mediated cell adhesion Cadherin-mediated cell adhesion Gap junction Ligand-gated cation channels
Intracellular checkpoints	Unfolded protein response (damaged proteins) DNA damage response (mutagenic signals)

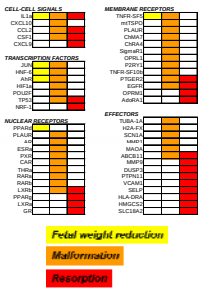
SOURCE: National Research Council, 2000

TIVS project 2: Embryonic Stem Cells

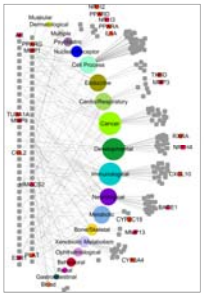
The effects of environmental contaminants on mouse embryonic stem cell differentiation

- Hunter (EPA): pathways that control cell signaling and specify cell fate
- data sharing: same compounds to confirm (+)ves and (-)ves across labs and ES lines (human, murine)
- resource sharing: TIGM gene trap resources for functional analysis of specific pathways; genomic profiling

Profiling DevTox target-pathways

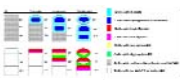


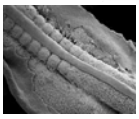
SOURCE: T Knudsen, NCCT

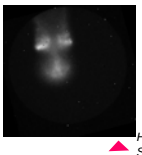


SOURCE: H Mortensen, NCCT

TIVS project 3: Agent-based models





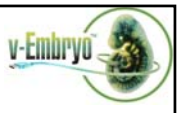


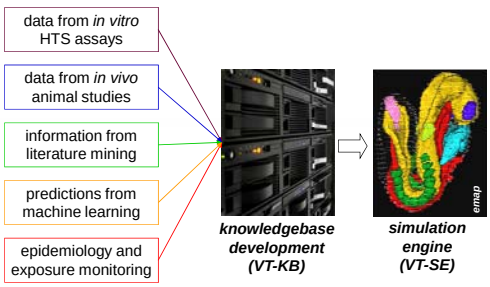
Development of computer simulations facilitating assessment of toxicity based on perturbed development in zebrafish and mouse embryonic stem cells

In silico model, CompuCell3D software
SOURCE: Glazier et al. (2008) Cur Top Dev Biol 81:205

Hes1-EGFP time-lapse (3h) clock-wavefront
SOURCE: Masamizu et al. (2006) PNAS USA 103:1313-18

Toward a Virtual Embryo





Opportunities for collaboration

- testing chemicals using developmentally-competent *in vitro* assays (ES cell and ZF embryos) and targets
- use predictive associations from ToxCast™ HTS data to build hypotheses about mechanisms of action
- studies to generate data testing hypotheses and improving predictive models
- improve virtual tissue models to a level that can help prioritize chemicals for quantitative risk assessment

An introduction to

ChemScreen

Bart van der Burg

Outline

- What is Chem*Screen*?
- Background
- Approach

Chemical substance in vitro/in silico screening system to predict human- and ecotoxicological effects

- EU framework program 7 (FP7)
- Collaborative project
- 9 partners from 5 countries
- Not yet started
- 1 Month after signing of the contract: December 1 2009?
- 4 years program, with majority of practical work in the first three years

100,106 chemicals on market in 1981 ("existing substances"); 1% tested on hazardous properties

EU White Paper: Strategy for a future Chemicals Policy, 2001

Most of 100,000 chemicals on market largely untested: REACH

- Registration Evaluation Authorisation of CHemicals program to catch up
- Start: June 1, 2007

General features REACH

- Supply chain to provide data
- Shift responsibilities from authorities towards industry
- Registration all compounds >1 ton/year
  - At central European Chemicals Agency (ECHA)
  - Data sharing obligatory (One Substance One Registration:OSOR)
  - Substance Information Exchange Forum (SIEF)
- Evaluation dossiers by ECHA/public authorities
  - May request additional data, with animal testing to the absolute minimum
- Authorisation required for harmful compounds taking into account risk, benefits, alternatives, etc.

Which prioritized effects in REACH?

- CMRs: Carcinogenic, mutagenic or toxic to reproduction
- PBTs: Persistent, bio-accumulative and toxic
- vPvBs: Very persistent, very bio-accumulative

How many chemicals?

All chemicals >1 tons per year: ~30,000

Estimated costs REACH:

- Costs: 2.8 – 5.2 bn € (EU) (Hartung 2009: x6)
- Carcinogenicity, Mutagenicity and Reproductive toxicity (CMR): ca 90% costs

Estimated benefit:

- Health improvement: 50 bn € (EU)

Phases

Registration:

Pre-registration:	June 2008
Higher risk (e.g. CMR):	December 2010
> 1000 tons (HPV):	December 2010
Remaining >100 tons:	June 2013
Remaining >1 tons:	June 2018

Evaluation:

Completed:	June 2022

Majority of testing 2011-2017

When traditional animal tests are used progress of REACH will be seriously hampered by:

1. Ethics: resistance to the excessive use of animals.
2. Costs: particular those linked to labour intensive animal testing
3. Capacity: lack of capacity to carry out these tests.
4. Speed: the use of the same traditional methods will not allow major advances in speed of the process to be made

>In order to be successful cost-effective, rapid in vitro tests need to be adopted

Incentives use of alternative (non-vertebrate) tests in REACH:

- Agency (ECHA) will publish test proposals (by chemical manufacturers) and invites third parties to submit alternative proposal
- Explicit allowance for alternative to in vivo tests, including in vitro and non-testing methods (QSAR, grouping, exposure,read across)
- Accepts “suitable methods”
- Regular reporting by Agency and Commission on use of alternative methods

ChemScreen — Background

Why reprotox?

- Prioritised in REACH
- Reproductive toxicity is important to assess both human and environmental toxicity
- Uses the most animals in toxicity testing
- Unfortunately, there are very few alternative methods

ChemScreen — Approach

Our approach:

- Identify sensitive parameters for reproductive toxicity
- Identify critical mechanisms involved in perturbation of these parameters
- Build high throughput system using this modules
- Expand step-wise
- Integrate with bioinformatics/data interpretation
- Build integrated testing strategies, including non-testing methods

ChemScreen — Approach

Work packages

1.Establish *in silico* prescreening and toxicity prediction methods prioritizing *in vitro* toxicity testing (WP1, leading partner; DTU)

2.Establish a database and an *in silico* prescreen to identify potential reproductive toxicants (WP2, FhG)

3.Establishment of sensitive parameters and a medium throughput 'minimal essential' *in vitro* assay panel (WP3, RIVM)

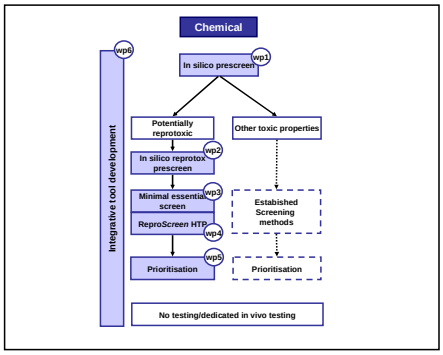
4.Establish a high throughput mechanistic pathway screen, for reproductive toxicants (WP4, EKUT)

5.Integrative methods to predict *in vivo* reprotoxicity allowing informed decisions on prioritization for eventual further testing (WP5, TNO)

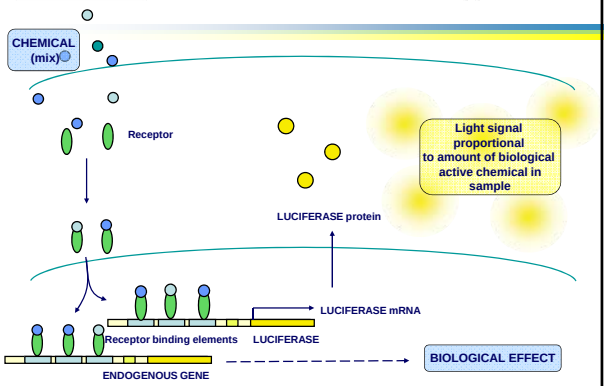
6.Integration into one user-friendly tool (WP6, P&GEN)

7.Dissemination (WP7, BDS)

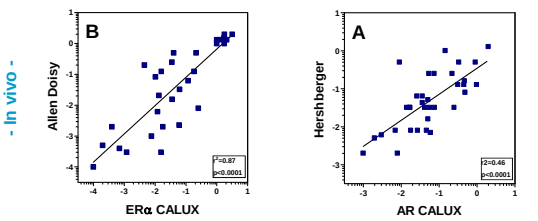
ChemScreen — Approach



ChemScreen — Approach



ChemScreen — Approach



Sonneveld et al.2006 Toxicol. Sci., 89:173-87

ChemScreen — Approach

Overestimation power animal data: poor predictions

TABLE 4. How Some Human Teratogens Have Been Discovered

Agent or Drug	Major Means of Discovery: Human Epidemiology Studies	Alert Physician or Scientist; Cluster	Animal Studies	In Vitro Studies	
Rubella		+++			
Aminopterin		++			
Anticonvulsants		+++			
Hydantoins 1963					
Trimethadione 1970		++			
Valproic acid 1982	+++				+
Vitamin A, 1953			+++		+
Isotretinoin 1983		++			+
Etretinate 1984		++			+
PCBs 1968		++			+
Coumarin 1968		++			
Alcohol 1967		++			
Lithium 1970		++			
Diethylstillbesterol 1971		++			+
Penicillamine 1971		++			
Misoprostol		++			
Trimethorpin	+				
Chorionic villous sampling		++			

\+ =strong nuclear receptor ligands

Brent (2004) Pediatrics 113, 984

ChemScreen — Approach

Screening systems

- Panel (15-50) reporter gene assays in human cells (nuclear receptors, dioxin receptor, signaling/stress /developmental pathways)
- Reporter gene assays in mouse ES cells (ReProGlow; developmental pathways)
- Wildtype ES/transcriptomics
- Metabolising cell systems
- Zebrafish/transcriptomics
- Others for critical endpoints reprotoxicity (e.g. spermatogenesis)

ChemScreen — Approach

In silico tools

- Exposure module
- Toxicity screening tool (>70 QSARs)
- In vivo reprotoxicity database (FeDTex, RepDose)
- Automated decision tool

ChemScreen — Partners

BioDetection Systems (BDS)	Bart van der Burg	Netherlands
Fraunhofer Institute for Toxicology and Experimental Medicine (FhG)	Inge Mangelsdorf	Germany
Netherlands Organization for Applied Scientific Research (TNO)	Dinant Kroese	Netherlands
Simpple (SIM)	Eduard Pauné	Spain
National Institute for Public Health and the Environment (RIVM)	Aldert Piersma	Netherlands
Danish Technical University Food Institute (DTU)	Jay Niemalä	Denmark
Procter & Gamble Eurocor (P&GEN)	Joanna Jaworska	Belgium
Eberhard Karls University of Tübingen (EKUT)	Michael Schwarz	Germany
University of Konstanz (UKON)	Daniel Dietrich	Germany

ChemScreen — Results

- **Sorry, no results yet!**

U.S. Environmental Protection Agency
Office of Research and Development
National Center for Environmental Research
Computational Toxicology Centers Science To Achieve Results (STAR) Progress Review Workshop

U.S. Environmental Protection Agency
Research Triangle Park, NC

October 1, 2009

MEETING SUMMARY

OVERVIEW

The U.S. Environmental Protection Agency (EPA) Office of Research and Development's (ORD) National Center for Environmental Research (NCER) Computational Toxicology Centers Science To Achieve Results (STAR) Progress Review was held October 1, 2009, in Research Triangle Park, North Carolina. The workshop was sponsored by ORD's NCER. Scientists from academia, government, and nongovernmental organizations assembled to discuss recent computational toxicology research and plan for future needs. The meeting provided an opportunity for grantees in the EPA-funded STAR Program to present their research and interact with EPA staff and others conducting computational toxicology research. Approximately 60 individuals attended the meeting.

Welcome, Introduction, and Review of Meeting Goals
Deborah Segal, EPA, ORD, NCER; and Robert Kavlock, EPA, ORD, National Center for Computational Toxicology (NCCT)

Ms. Deborah Segal explained that ORD provides leadership in science and conducts the majority of EPA's research and development. NCER is ORD's extramural research arm, with a research budget of $440 million, of which $65.5 million is allocated for competitive extramural grants and fellowships, such as the STAR, Small Business Innovation Research (SBIR), and Greater Research Opportunities (GRO) Programs. ORD works with other EPA offices to select research topics for the STAR Program, which was established in 1995 as part of a reorganization of ORD. STAR aims to include the country's universities and nonprofit centers in EPA's research program to ensure the highest quality science in areas of highest risk and greatest importance to the Agency. STAR issues approximately 25 Requests for Applications (RFAs) and awards approximately $65 to $100 million annually.

The STAR Research Program in Computational Toxicology aims to integrate computational methods and advanced molecular biology techniques and develop the use of computational approaches to provide tools for quantitative risk assessment and more efficient strategies for prioritizing chemicals for screening and testing. Five RFAs have been issued under this program. A new RFA is in development for Fiscal Year 2010.

Dr. Robert Kavlock noted that the grand challenge is predicting human toxicity, moving from exposure conditions to impacts on molecular targets that result in cell changes and ultimately in toxicity to the organism. Tools that allow scientists to interrogate different levels of this biological complexity now are being released. These range from high-throughput screening biochemical assays to cell-based assays to modeling systems. The STAR Center researchers presenting at this progress review are actively involved in various phases of this work.

A variety of reports guide the Computational Toxicology Research Program (CTRP), including the National Academy of Sciences 2007 report, *Toxicity Testing in the 21st Century: A Vision and a Strategy.* Other reports that have informed the Program in terms of the challenges of the current testing paradigm and the opportunities available to use innovative technologies to address these important issues include *Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment*; *Phthalates and Cumulative Risk Assessment: The Task Ahead*; and *Science and Decisions: Advancing Risk Assessment*. *Toxicity Testing in the 21st Century: A Vision and a Strategy* discusses biological processes and the changes caused by exposure. At lower doses, cellular changes begin to manifest, but there still is an adaptive response. At higher doses, the result can be cell injury and morbidity and mortality. Understanding and developing assays for signaling systems involved in the induction of toxicities will help researchers to better understand toxicity.

The CTRP's mission is to integrate modern computing and information technology with molecular biology to improve Agency prioritization of data requirements and risk assessment of chemicals. The Program provides decision-support tools for high-throughput screening, risk assessment, and risk management and is committed to transparency and public release of all data. The Program operates under tight deadlines, initially given 5 years to prove that this type of approach is effective. The recently completed Board of Scientific Counselors (BOSC) review recommends that the Program be renewed for an additional 5 years.

The Program supports EPA's strategic plan by focusing on its goals of identifying and screening toxicity pathways, conducting toxicity-based risk assessment, and providing the information to EPA's regulatory arm. EPA Administrator Lisa Jackson's priorities include managing chemical risks; she has stressed the importance of assessing and managing risks of chemicals in consumer products, the workplace, and the environment as well as the importance of protecting vulnerable subpopulations. The *Essential Principles for Reform of Chemicals Management Legislation* includes the review of chemicals against safety standards based on sound science, reflecting the risk-based criteria protective of human health and the environment. An initial list of chemicals that EPA is considering for action plan development under these principles includes bisphenol A, perfluorinated chemicals, and phthalates.

Computational toxicology research is conducted via the NCCT, ORD projects, and the Computational Toxicology STAR Centers. The STAR Centers are housed at the New Jersey Environmental Bio-informatics and Computational Toxicology Center, Carolina Environmental Bioinformatics Research Center, Carolina Center for Computational Toxicology, and Texas-Indiana Virtual STAR Center. Implications for success include additional closing of the toxicological information gap, providing mode of action information to risk assessment, more effectively using animal and human resources related to the evaluation of hazard and risk, and performing ancillary applications related to mixtures, chirals, nano-materials, green chemistry, and lot variations. This meeting will provide an opportunity for introductions, reflections on the work accomplished to date, integration of the work, and discussion of next steps.

Carolina Center for Computational Toxicology
Ivan Rusyn, University of North Carolina

Computational toxicology is a synthesis of chemistry, high-throughput screening, *in vivo* data, and molecular pathways to generate new knowledge. With increasing amounts of data becoming available, risk assessors now are better able to understand the risks to human health and the environment. As it is an interdisciplinary science, computational toxicology represents a tremendous opportunity for incorporating other disciplines into traditional toxicology research and for training new researchers. Researchers need to recognize that this should not be simply an academic exercise; it is very important that the value and the early results of computational toxicology research be communicated to the general public, industry, and other stakeholders.

The Carolina Center for Computational Toxicology consists of an administrative core and three research projects and is directed by an internal steering committee assisted by an external advisory board. The administrative core serves a number of functions, including management, integration, public outreach/translation, and quality control. Project 1 is focused on predictive modeling of chemical-perturbed regulatory networks in systems toxicology. Objectives of this project include: developing and applying data-driven methods for the inference and high-level modeling of regulatory network response to chemical perturbation, developing mechanistic models of nuclear receptor function, and integrating and deploying high- and low-level modeling tools. Interactions with EPA have been centered on exploring toxicity pathways, extending and integrating mechanistic metabolism and other models, and working with ToxCast™ data. For inference and modeling of biological networks, short-term goals include developing tools for data analysis and interpretation and helping to establish the biological-chemical context in high-throughput screening assay datasets. Long-term goals include developing components to systems (simplistic wiring); developing a framework for understanding systems' properties, pathways, and cross-talk; and providing a basis for mechanistic models. The first major challenge of this project involves the integration of different types of data, from genome data to phenotype data. The individual data streams are not well-defined, and the network context can be viewed in a number of different ways. A software package that will stratify data for subgraph mining to study various pathways is under development; this is an innovative approach, as it can define composite assays that will be more predictive than individual assays. Also under development is a mechanistic model of cellular metabolism that will predict changes in metabolic flux.

Project 2 is focused on toxicogenetic modeling: population-wide predictions from toxicity profiling. This project is exploring the promises and challenges of incorporating the knowledge of interindividual genetic variability as an important dimension of toxicity testing. Objectives of the project include developing toxicogenetic expression quantitative trait loci (eQTL) mapping tools; performing transcription factor network inference and integrative pathway assessment; performing toxicogenetic modeling of liver toxicity in cultured mouse hepatocytes; and discovering chemical-induced regulatory networks using population-based toxicity phenotyping in human cells. Interactions with EPA have included developing and testing novel *in vitro* tools that will enable testing for interindividual susceptibility, developing statistical methodology and computational tools capable of processing higher order multidimensional data, and working on future ToxCast™ efforts and current Tox21 datasets. This project is combining multiple streams of data and adding a level of genetic variability. One basic idea for combining genetic diversity and biology is through eQTL mapping. The challenge, however, is determining true genetic susceptibility and doing so in a timely fashion. This project also aims to understand whether the type of mapping used can determine how genetic polymorphisms can control the molecular pathways perturbed by environmental exposures. Another aim is to understand genomic context for expression.

Project 3 is focused on the development of validated and predictive quantitative structure-toxicity relationship models that employ chemical and biological descriptors of molecular structures and take into account genetic diversity among individuals. Objectives of the project are to develop rigorous endpoint toxicity predictors based on the quantitative structure-activity relationship (QSAR) modeling workflow and conventional chemical descriptors, develop novel computational models based on combined chemical and biological descriptors through QSAR modeling workflow, and develop novel computational toxicogenetic models based on combined genetic, chemical, and toxicity descriptors through QSAR-like modeling workflow. Interactions with EPA have focused on integrating chemical descriptors into the Distributed Structure-Searchable Toxicity (DSSTox) Database Network, ToxCast™, Toxicity Reference Database (ToxRefDB), and Aggregated Computational Toxicology Resource (commonly known as ACToR) data analysis. This project integrates chemical descriptors and high-throughput screening biological descriptors with the QSAR modeling paradigms to predict animal *in vivo* endpoints and, hopefully, human disease endpoints. This work has shown that a focus on accurate prediction of external datasets is much more critical than accurate fitting of existing data. Also, cheminformatics, high-

throughput screening, nor omics data alone is sufficient to achieve the desired accuracy of the endpoint property prediction.

In the first year of the Center's operation, 12 research papers have been produced and are in various stages of the publication process. Project 1 short-term goals for Year 2 are to continue in-depth analysis of ToxCast™ Phase I data, further refining the methods for integration across data types, investigate the applicability of the metabolism model as a tool for the prediction of the effects of chemical perturbation of metabolic pathways, integrate the eQTL analyses/approaches with the network-focused methodologies, and establish the pathway-based biological network context for QSAR. Project 2 short-term goals for Year 2 are to continue development of FastMap software; construct transcription regulation networks in the Bayesian framework by combining eQTLs, nucleosome occupancy, and transcriptional regulation data; complete characterization of the mouse hepatocyte cultures and perform experiments with key toxicants; and complete genome-wide association studies of the HapMap lymphoblast cell viability and apoptosis data and correlate the toxicity endpoints with basal gene expression profiles. Project 3 short-term goals are to complete the analysis of the ToxCast™ data, continue to explore other datasets that provide both *in vivo* and *in vitro* data for chemicals, and build models that could be used by EPA to prioritize the selection of ToxCast™ Phase II compounds.

Dr. Kavlock asked whether the researchers had identified gaps in pathway coverage for which new assays are needed. Dr. Rusyn responded that for Project 1, the focus is on current ToxCast™ assays, whereas Project 2 is searching for the genes and pathways that are most susceptible to interindividual variability; after those genes and pathways are identified, the next step will be to consider the assays needed.

Dr. David Dix referred to Project 2, asking if there was value in focusing on more specific molecular endpoints. He asked Dr. Rusyn for his thoughts on moving this type of approach forward. Dr. Rusyn stated that some of the Center's work has involved taking a leap of faith and moving forward with the most commonly used assays; he would like to complete this analysis before determining the next steps.

A participant noted that dose-response information for individual assays was missing and asked whether the researchers had considered using a composite dose-response. Dr. Rusyn replied that the current binary classification does not necessarily take into account all of the dose-response information. Dose responses differ between different datasets, making it difficult to align the information. The Center is testing a number of different approaches to determine the meaning of the dose-response information. Dr. Rusyn welcomed suggestions on the best features of dose-response to study.

Collaborative Work With EPA
Richard Judson, EPA, ORD, NCCT

EPA studies individual chemicals and determines maximum safe doses for human exposure. The Tox21 Priority List includes 19,000 chemicals, and there is an enormous data gap for many of these chemicals, so it is imperative that the testing be prioritized and performed in a timely manner. Priority areas for research methodology and development include prioritization, mechanism of action determination, dose-response modeling, and susceptible populations.

The Carolina Center for Computational Toxicology's Project 1 is developing and applying data-driven methods for the inference and high-level modeling of regulatory network response to chemical perturbation, developing mechanistic models of nuclear receptor function, and developing methods for integrating and deploying high- and low-level modeling tools. An important issue for NCCT has been selection of assays to be developed for ToxCast™ and Tox21. The Carolina Center's work will help EPA with this task. Project 2 is developing fast and efficient toxicogenetic eQTL mapping tools and working to better understand chemical-induced regulatory networks using population-based toxicity phenotyping in human cells. The Carolina Center is in the early stages of this work. Project 3 is developing rigorous endpoint toxicity predictors based on QSAR modeling workflow using conventional chemical descriptors.

In addition, the Center is developing novel computational toxicogenomic models based on combined chemical and biological descriptors. This project is addressing mechanism of action and should help EPA to prioritize chemicals for further study. In summary, the Carolina Center is developing promising new approaches to address EPA computational toxicology research areas of prioritization, mechanism of action determination, and susceptible population study methodology. The question is whether some of these methods can be extended to help understand dose-response relationships.

New Jersey Environmental Bioinformatics and Computational Toxicology Center
Panos Georgopoulos and William Welsh, University of Medicine and Dentistry of New Jersey

The objectives of the New Jersey Environmental Bioinformatics and Computational Toxicology Center are to address the toxicant source-to-outcome continuum through the development of an integrated modular computational framework, develop predictive cheminformatics tools for hazard identification and toxicant characterization, and demonstrate the above tools through applications in quantitative risk assessment. The Center takes a computational/engineering/systems perspective, utilizing a team of computational scientists and engineers with diverse backgrounds in bioinformatics, cheminformatics, and enviroinformatics. New frameworks and tools build on an extensive base of past developments. This research effort emphasizes interaction and collaboration among participating scientists in the STAR Bioinformatics Centers and with EPA centers and laboratories and other centers and institutes of excellence. The research is divided into two major areas. Investigational Area I focuses on a source-to-outcome framework to support risk characterization, and Investigational Area II focuses on hazard identification. There are three projects under Investigational Area I. The first project involves multiscale biologically based modeling of exposure-to-dose-to-response processes, the second project involves hepatocyte metabolism modeling for xenobiotics, and the third project focuses on tools for optimal identification of biological networks. Under Investigational Area II, a fourth project develops cheminformatics tools for toxicant characterization, and a fifth project develops optimization tools for *in silico* proteomics. The Center's research integration plan is consistent with the 2007 NAS report, *Toxicity Testing in the 21st Century: A Vision and a Strategy*. The Center pursues an integrative multiscale research approach—from molecules to cells to tissues to organs to organisms to populations—recognizing the importance of processes/signals at all levels of biological organization. Additionally, the Center's close interaction with EPA has resulted in several publications.

Dr. Georgopoulos described Investigational Area I in further depth, noting that computational toxicology emphasizes chemicals, pathways, and toxicity, but it also must inform the science of risk assessment. In addition to biology, risk also depends on the environment, behavior, and time (development and aging).
A general mathematical framework for environmental health risk analysis must consider multiscale bionetwork dynamics (spanning the genome, transcriptome, proteome, metabolome, cytome, and physiome) linked with the dynamics of environmental stressor networks in food, air, water, and soil. The Center has studied how these networks are coupled with the regulatory and metabolic bionetworks using complex, multiscale modeling. Dr. Georgopoulos displayed a graphic depicting the sequence from source/stressor formation to dose to toxicokinetic effects to modifications of the environmental agent by the organism to biological effects to health outcomes. This includes a key element that is missing from most representations of source-to-effect continuum approaches; this element allows using biological data and biomarkers to evaluate assessments of exposure, locate source contributions, and perform account-ability studies. Thus, a general Bayesian framework is being developed to reconstruct exposure from inversion of biomarker data for individuals and populations.

The Modeling ENvironment for TOtal Risk Studies (MENTOR) employs an anthropocentric (person-oriented) approach, linking multiple scales of macroenvironmental and local models and information with microenvironmental conditions and human activities in time/space. It has been applied to study exposures to a wide variety of contaminants in different media (e.g., metals, dioxins and polychlorinated biphenyls, air toxics), selecting in particular arsenic and trichloroethylene (TCE) as "model contaminants" for

comprehensive source-to-dose-to-response studies. These studies showed close agreements of predictions with measurements of population biomarkers. The Center is working to further refine the MENTOR system and integrate it with the Dose-Response Information Analysis (DORIAN) system.

MENTOR with Physiologically Based Pharmacokinetic Modules for Populations (MENTOR-3P) combined with the DORIAN system provides a new modular "whole body" platform for consistent characterization of multicontaminant, toxicokinetic, and toxicodynamic processes in individuals and populations. This approach incorporates physiology databases to account for intra- and interindividual variation and variability. Major ongoing research efforts of MENTOR/DORIAN focus on a library of software modules for "virtual organs" (with primary focus on the liver) that account for heterogeneities (in metabolism and biological response) within an organ. One case study focused on the spectrum of cytochrome P450 induction by dioxin within the liver and was able to account for and explain observed biochemical variability. Research in progress is using arsenic and TCE as model contaminants and aims to reconcile the biotransformation and transport at both the individual hepatocyte and the whole-organ scales, as well as on modeling quantitative metrics of oxidative stress resulting from exposure to these contaminants. The computational models are being used in collaborations with scientists from EPA to study issues of sensitivity analyses and effects of aging and assess population exposures from biomarkers.

Dr. Welsh further described Investigational Research Area II, noting that in any multiscale enterprise, molecular scale must be addressed, for which there are three different approaches. Receptor-based approaches study the protein structure of a receptor associated with a pathway or some aspect of a toxicological event. Ligand-based approaches seek to gather data about the ligands to determine commonalities among the ligands that give rise to a certain biological effects. The third approach is virtual screening.

Receptor-based approaches figure prominently in computational toxicology. Pregnane X receptor (PXR) is a hepatic nuclear receptor that is responsible, along with other nuclear receptors and proteins, for modulating a number of metabolic enzymes and more than 36 other genes. PXR ligands are pervasive and structurally diverse. They come from dietary products and supplements, hormones, prescription drugs, herbal components, and environmental chemicals. Thus, humans are exposed to PXR ligands constantly. Published experimental data show that when certain conazoles bind to PXR, they turn off the transcriptional machinery. Based on this observation, the researchers performed computational docking studies that show that the conazoles do not competitively bind with the agonist site but instead appear to bind on an outer surface. This is an important finding that can inform the development of new hypotheses.

Analysis of the ToxCast™ 309 dataset helped the researchers to develop and adapt various new computational models for data analysis. Traditional QSAR techniques use the structure-based features (molecular descriptors) of a collection of chemicals to describe and compare their biological activities. Biological spectra analysis is a new technique that uses the biological response profiles of the chemicals to describe and compare their molecular structures. Panels of chemicals and protein receptors were assayed and the numerical values depicted as heat intensity bars. Chemicals were clustered based on similar abilities to induce a biological response across all of the proteins. Proteins were clustered based on similarities in their bioresponse profiles. Ultimately, cross-mapping of the toxicological and chemical similarity profiles showed that 74 percent of the compounds from the TOX1 cluster also were in the CHEM1 cluster, and 61 percent of the compounds from the TOX2 cluster also were in the CHEM2 cluster. Overall association between the major clusters of the two spaces was found to be 69 percent.

The Center also has developed a novel technique for comparing molecules. Shape signatures compare molecules by subtracting their histograms. A software program sketches the molecule, and a special algorithm converts three-dimensional molecules into small, compact representations based on the molecular shape and surface charge distribution, the two features predominantly associated with receptor ligand binding. The shape signatures of different molecules then can be compared. The smaller the

difference between the histograms, the more similar the molecules. The Center has created a shape signature library that houses more than 3 million compounds. A number of shape-based QSAR models for toxicity prediction have been developed.

New Jersey Environmental Bioinformatics and Computational Toxicology Center – EPA Collaboration on an Approach to Using Toxicogenomic Data in Risk Assessment: Dibutyl Phthalate Case Study

Susan Euling, EPA, ORD, National Center for Environmental Assessment (NCEA)

How can genomic data be used effectively in risk assessment? Collaboration between mathematicians and biologists is needed to answer this question. Genomics technologies are powerful because they are global or genome-wide and toxicogenomic data can identify precursor events, biomarkers of effect or exposure, and mechanisms and modes of action. Strengths of microarray data include the ability to identify pathways, build gene networks, and identify affected processes, pathways, and networks. Challenges include the size and complexity of the datasets and the fact that statistical cutoffs do not necessarily indicate biological significance. Limitations of using toxicogenomics technologies have included reproducibility issues, the need to link affected pathways and genes to an adverse outcome, and the cost involved in performing dose-response microarray studies.

The overall project goals were to develop an approach for using toxicogenomic data in risk assessment and perform a case study using this approach. Dibutyl phthalate (DBP) was selected for the case study because it has a relatively large genomic dataset, and there is phenotypic anchoring for a number of the observed gene expression changes. There are two well-characterized modes of action for DBP responsible for the male reproductive developmental effects: a decrease in Insl3 and a decrease in fetal testicular testosterone. Questions were identified to direct the DBP case study evaluation. The questions were whether the toxicogenomic data could inform additional modes and mechanisms of action for the DBP male reproductive developmental effects and whether the genomic dataset could inform interspecies differences in the reduced testicular testosterone mode of action. To explore modes of action, the consensus pathways were identified from two different pathway analysis approaches for a selected microarray study of testes after *in utero* DPB exposure.

There is concern that the traditional method of first identifying differentially expressed genes and then as a second step performing pathway mapping might result in a loss of information. Thus, the STAR Center collaborators took a different approach to identify significantly affected pathways, considering all of the genes in the pathway and calculating a pathway activity level for different pathways. Advantages of this approach include the consideration of all genes in a pathway and the ability to compare activity among pathways.

Methods to inform interspecies differences in mode of action were explored. There is a need for approaches and metrics to extrapolate from animal model findings to humans for risk assessment. Available data were used to develop cross-species metrics for the biosynthesis-of-steroids pathway, one of the pathways that underlies the decrease in fetal testicular testosterone mode of action. Three different data sources were used to assess rat-to-human pathway similarity, and results showed approximately 85 percent similarity using any of these three approaches. A remaining issue in applying any or all of these methods to risk assessment is determining whether these are “low” or “high” degrees of similarity. This issue can be explored further to develop a basis for comparison.

Case study findings include the identification of additional functions (e.g., cell adhesion) and pathways (e.g., Wnt signaling) affected after *in utero* DBP exposure that may inform modes of action responsible for the “unexplained” endpoints. Hypothesis testing studies are needed. Other accomplishments include the development of a systematic approach for evaluating toxicogenomics data for use in future risk assessments; the development and exploration of the application of microarray analytical methods to risk

assessment including the pathway activity method, the gene network model over time, and the exploration of methods to assess cross-species conservation on a given pathway; and the identification of research needs for toxicity and genomics studies for use in risk assessment.

Recommendations based on the case study are to evaluate genomic and other gene expression data for consistency of findings across studies for affected genes and pathways, perform benchmark dose response modeling when high-quality reverse transcriptase-polymerase chain reaction data are available for genes known to be in the causal pathway for a mechanism of action or outcome, and perform new analysis of genomic data if re-analysis is expected to yield new information useful to risk assessment.

Dr. Kavlock asked the STAR Center researchers in general whether STAR funding had been useful in obtaining other grant funding, including stimulus funding. The consensus among the group was that the STAR funding had been useful for leveraging additional funding.

Carolina Environmental Bioinformatics Research Center
Fred Wright, University of North Carolina

The Carolina Environmental Bioinformatics Center (CEBC) was funded to extend capabilities in computational toxicology. Specific capabilities include omics expertise and strengths in elucidating genetic variation. The Center's three research projects focus on biostatistics, cheminformatics, and computational infrastructure for systems toxicology; each project collaborates directly with environmental scientists. The Center also includes an administrative unit and an outreach and translational activity unit. The Center has collaborated extensively with EPA; seven joint papers are in various stages of publication, and 14 joint abstracts/posters have been accepted at scientific meetings. Whereas the Carolina Center for Computational Toxicology is more highly focused on biology and mechanistic modeling, the CEBC focuses on discovering and obtaining valid statistical conclusions.

Project 1, the biostatistics in computational toxicology project, includes an emphasis on strengths in microarray analysis, elucidation of networks/pathways, and eQTL analysis. There is a new emphasis on dose-response testing, data mining, and penalized regression. Analysis of ToxCast™ Phase I data from EPA and development of related methods likely will be a large portion of the remaining activity. Project objectives include providing biostatistical support to the Center, performing data analysis and developing methods, and collaborating with EPA and the computational toxicology community. Recent activities include direct collaborations via data analysis work with Project 2 investigators on toxicity prediction and data mining methods and work with Project 3 investigators on rodent toxicity modeling. In addition, the project is performing analysis of clinical toxicity and metabolomic data to explore a large number of prediction approaches, analysis of ToxCast™ data, and expression QTL mapping relevant to toxicity. Collaborations have inspired the development of new methods. For example, CEBC scientists worked with EPA scientists on a microarray dose-response study. This work led to new considerations for using dose-response data; there currently are relatively few methods for dose-response that are tuned to gene expression studies and even fewer that consider pathways (gene sets). An important question that arose from this work was how to aggregate evidence across transcripts within a pathway. For dose-response modeling for gene expression and pathways, the researchers have performed extensive investigation of simple (approximate) two-parameter logistic fits, establishing reasonable false positive rates and power for small sample sizes. A new tool that will perform dose-response pathway analysis for gene expression data is under development. Other collaborations with EPA include comparing machine learning algorithms in a simulated model for chemical toxicity and various efforts to predict chemical toxicity. Another example of methods development is the work on methods for detecting true trans-bands in eQTL studies and consideration of the importance of PC-based stratification control for eQTL analysis. In the next year, Project 1 will focus on completing the methodology for open projects and collaboration, completing the dose-response pathway analysis method, bringing the ToxCast™ data analysis to an

intermediate conclusion, and deepening the ToxCast™ data analysis in terms of choices of endpoints, sensitivity versus specificity, and domains of applicability.

The objectives of Project 2 (cheminformatics) include coordinating the compilation and mining of data from relevant external databases, performing analysis and methods development for building statistically significant and externally predictive QSAR models of chemical toxicology data, and performing joint work within the Center and with EPA collaborators. Under this project, one subproject works to improve quantitative models of chemical toxicity through the use of hybrid chemical and biological descriptors. The Center is working with EPA scientists, using high-throughput screening dose-response curves to assist QSAR modeling of carcinogenicity. In this work, more than 300 chemical descriptors, 150 biological descriptors, and 400 hybrid descriptors are being used to predict carcinogenicity. Also under development is a two-step hierarchical QSAR modeling workflow for predicting *in vivo* chemical toxicity. Future studies include analyzing the models to identify significant assay-chemical combinations that are predictive of *in vivo* outcomes, exploring the entire National Toxicology Program (NTP) dataset, and applying modeling prospectively to prioritize new compounds for focused toxicity testing. In the next year, Project 2 will focus on continuing work on QSAR modeling of multiple animal toxicity endpoints and developing novel QSAR methodology by using *in vitro* biological information to model *in vivo* toxicity endpoints. For all of these activities, the project will continue to use data collected under ToxCast™, DSSTox, and other EPA projects.

Project 3, the computational infrastructure for systems toxicology project, is using a model for toxicity profiling in multiple strains of mice to inform and develop an appropriate computational infrastructure, with a focus on computational methods development and the development of user-friendly software tools from methods in Projects 1 and 2. Project objectives include developing and implementing algorithms that aid the analysis of multidimensional data streams in dose-response assessment and cross-species extrapolation; facilitating the development of a standard workflow for analysis of the omics data, linkages to classical indicators of adverse health effects, and integration with other types of biological information such as genome sequences and genetic differences between species; and building Web-based open source and user-friendly graphical interfaces associated with interoperable computational tools for data analysis that facilitate the incorporation of new data streams into basic research and decision-making pipelines (methods from Projects 1 and 2). This project has created a framework for handling emerging omics data on genetic susceptibility in model organisms, provides programming expertise to create graphical tools that are used by partners within the Center and in collaboration with EPA personnel and other environmental scientists, and works to strengthen and advance the field of computational toxicology through direct partnerships and the dissemination of tools used by both bioinformatics and bench scientists. The driving biological problem is how to make population-wide predictions from toxicity profiling. Efforts toward integrating varying types of biological information have been informed by examples such as the study of the genetic factors underlying interindividual susceptibility to acetaminophen toxicity. In this unique human-to-mouse-to-human work, the researchers have shown that the power of mouse genetics can be extremely useful in discovering susceptibility genes, even when human data are available from very small cohorts. Project 3 also is developing software tools, including a graphical interface for the Significance Analysis of Function and Expression (SAFE) software, which assesses the significance of biological categories in microarray studies while properly accounting for the effects of correlations among genes. Investigators in this project also are key players in the integration of existing and new tools into the Predictive Toxicology Web Portal (http://ceccr.unc.edu). Papers on the algorithms used are in various stages of publication. In the next year, Project 3 will continue integration/support of tools from other CEBC projects, continue programming and algorithmic developments, further improve algorithms in tools and applications, develop specific data-mining algorithms for genomic databases, and continue biology-driven research that generates appropriate datasets for testing and implementing novel computational and biostatistical approaches.

Across the Center, there will be more emphasis on dissemination of information and training other scientists in the use of the tools developed and on bringing open source code and methods to a new stage in their evolution.

Collaborative Work With EPA
Ann Richard, EPA, ORD, NCCT

The work of the CEBC and the Carolina Center for Computational Toxicology overlaps nicely in terms of methods development and moving that work into predictive model-building. The recent BOSC review emphasized the importance of EPA maintaining an ongoing dialogue with academia; the biostatistics capability at the University of North Carolina has brought high standards of statistical analysis to help EPA evaluate the new data streams arriving via ToxCast™ and Tox21.

CEBC's cheminformatics project has the ability to generate thousands of QSAR descriptors representing categories of structure-based computed properties (DRAGON), and the project has developed a sophisticated predictive QSAR workflow. EPA defines the problems and provides data and guidance on how to approach these problems. The DRAGON descriptors include many categories of chemicals and different ways of describing these chemicals, which allows for flexibility in determining how best to approach a problem. CEBC has developed QSAR models based on DSSTox-published data files and structure inventories. The processed data files and calculated descriptors then are shared with EPA researchers for public release. EPA and CEBC have co-authored several publications.

DSSTox has published structure annotated toxicity data, which have been used in the cheminformatics work. A major objective of this project is to try to curate quality structure annotation and publish datasets that provide representations of activity that are particularly amenable to structure activity modeling. EPA's contribution to the Project 2 work has been through the ToxCast™ Phase I Chemical Inventory and the ToxRefDB *in vivo* endpoints for modeling. CEBC used this information to process datasets (ZEBET Acute Tox) and to calculate chemical descriptors (DRAGON) for the ToxCast™ Inventory. CEBC's cheminformatics project overlapped published data for 1,408 compounds from the NTP High-Throughput Screening Program with data from a carcinogenicity potency database. The aim was to determine the ability of the NTP high-throughput assays to predict carcinogenicity. Data generated to date show that the *in vitro* assays used have some ability to enhance modeling capabilities. The idea is that if *in vitro* assays that presumably are unrelated to the endpoint can enhance modeling, *in vitro* assays that are related to the endpoint should prove even more useful.

For years, there has been an effort to replace *in vivo* assays with *in vitro* screening methods. Many efforts have been made to correlate *in vitro* half-maximal inhibitory concentration (IC_{50}) with *in vivo* rat oral median lethal dose (LD_{50}), but none have been successful. It is important to consider new ways of incorporating the IC_{50} data. Two key questions arose: Can the problem be broken into regions of higher correlation? Can QSAR methods be used to define those regions based on chemical structure alone? Moving regression was used to define regions of higher correlation, and a classification QSAR was applied to assign the chemicals to one of three groups. The LD_{50} then was predicted for each group.

The Texas-Indiana Virtual STAR Center: Data-Generating* In Vitro *and* In Silico *Models of Developmental Toxicity in Embryonic Stem Cells and Zebrafish
Maria Bondesson Bolin, University of Houston; Richard Finnell, Texas A&M University; James Glazier, Indiana University

Approximately one in every 33 U.S. infants has a congenital anomaly. Heart defects are the most common anomalies; others include neural tube defects and orofacial clefts. Although the causes of congenital anomalies are both genetic and environmental, there is major concern about environmental compounds as causative agents. In some cases, it is known that specific compounds cause anomalies. For

example, methyl mercury and other heavy metals have been shown to be teratogenic. There remains, however, a large knowledge gap in terms of which compounds cause congenital anomalies.

The Center's objective is to develop new screening models for developmental toxicity. The aim is to move from biological models of developmental toxicity to computer simulations. The main research goals are to generate developmental models based on mouse embryonic stem cells and zebrafish suitable for high-throughput screening, generate high-information content models on development and differentiation using mouse embryonic stem cells and zebrafish, develop computational models for developmental toxicity with the aim of first re-creating normal development (in wild-type) and then classifying possible mechanisms by which chemical perturbations cause experimentally observed developmental defects, and perform proof-of-concept experiments of the *in vitro* and *in silico* test platforms with a blind test of chemicals.

The project has been divided into three investigational areas: (1) zebrafish as a model to elucidate the morphological and mechanistic effects of environmental pollutants, (2) the effects of environmental contaminants on mouse embryonic stem cell differentiation, and (3) the development of computer simulations facilitating assessment of toxicity based on perturbed development in zebrafish and mouse embryonic stem cells. Courses on zebrafish development, embryonic stem cells, and computer simulations for doctoral students and postdoctoral fellows have been developed. The Center regularly collaborates with stakeholders and other researchers. For all three projects, 37 chemicals that are known or expected to be teratogenic have been chosen for study. The chemicals have been ranked by potential threat to human health as determined by the Agency for Toxic Substances and Disease Registry and EPA.

The first investigational area uses zebrafish models to elucidate the morphological and mechanistic effects of environmental pollutants. Zebrafish were chosen for a number of reasons: they are small, embryos are transparent, fish can be transparent, they experience rapid external embryonic development and produce hundreds of eggs weekly, the genome is homologous to humans, the developmental pathways between fish and mammals are similar, many zebrafish mutants exist, it is relatively easy to knock down gene expression in zebrafish, and they are cost-efficient and adaptable to medium- to high-throughput screening.

Transgenic fish embryos will be produced, with the transgenes marking certain cell types during development. The Center plans to construct 10 transgenic fish expressing fluorescent markers to follow development and patterning. The endpoints include gastrulation and early embryonic cell movements, patterning of the central nervous system and neurogenesis, hematopoiesis and angiogenesis, and yolk utilization and morphological effects on somitogenesis. Morphology and green fluorescent protein/red fluorescent protein expression will be recorded during normal development, and the embryos will be treated with different toxicants to determine whether development is altered by teratogenic chemicals. At the end of the project, the goal is to scale up and automate for high-throughput screening. High-information content models based on the transgenic fish will be developed.

The second investigational area uses mouse embryonic stem cells as a model to elucidate the morphological and mechanistic effects of environmental pollutants. A recently created gene trap library contains more than 350,000 embryonic stem cell clones and between 10,000 and 13,000 inactivated genes. The aim is to use these embryonic stem cell resources to study specific markers of differentiation and patterns to determine how environmental agents affect development. Genes have been selected primarily based on their role in early embryonic development patterning, particularly those involving gastrulation and cell movements. Expected results include documentation of morphology and β-geo expression during normal development and teratogenic chemical exposure.

The third investigational area will develop computer simulations facilitating the assessment of toxicity based on perturbed development in zebrafish and mouse embryonic stem cells. This work still is in its

infancy. The major research question is how to move from cell phenomenology to tissue-level patterning and structure. The goal is to build simulations that address some of these missing pieces in the understanding of increased developmental defects. Other related projects include the CompuCell3D Multi-Cell Modeling Environment project, which is developing an open source, multiplatform modeling environment that allows the building of multicell simulations of developmental phenomena and diseases. Another project is the Systems Biology Workbench (SBW) Reaction-Kinetics Modeling Environment, which is a standard for performing reaction kinetic modeling of subcellular regulatory metabolic networks. Multicell modeling in CompuCell3D and SBW will integrate molecular-, cellular-, and whole-organ-level data to predict developmental effects of pathway disruption.

The BOSC review that took place immediately prior to this workshop included discussion of platforms for many possible directions of approaching toxicological development. Questions yet to be answered include: Is it more useful to develop organ systems that already are NCCT foci (e.g., liver, limb, gastrulation) or novel ones (e.g., vasculogenesis)? What are the best ways to integrate the biological approach of the Texas-Indiana Virtual STAR (TIVS) Center team with the prioritization outcomes for EPA? Should the focus be on one or two classes of perturbation agents? Should the focus be on tool/data or model development?

Collaborative Work With EPA
Thomas Knudsen, EPA, ORD, NCCT

The RFA under which the TIVS Center was funded encouraged a different approach that studies exposures that perturb biological events during formative stages of the reproductive cycle affecting embryo and fetal development, postnatal development, fertility and reproduction, and children's health. Some key research issues illustrate the complexity of this work, including timing of cellular interactions, sensitivity of these systems, complexity of interactions, and maternal influence. Many developmental patterns can be tracked and studied *in vitro* to define how chemicals disrupt fundamental control in patterning, timing, differentiation, and morphogenesis.

TIVS will use *in vitro* models (zebrafish embryos, mouse embryonic stem cells) and *in silico* computational models to elucidate the morphological and mechanistic effects of environmental pollutants. For the research on zebrafish embryos, there are a number of opportunities for connections between EPA and TIVS, including data and resource sharing to evaluate developmental signaling pathways. Some chemical effects in this system already are represented in ToxCast™ datasets, and the pathways identified by this project will be added. TIVS can help EPA to prioritize the most important pathways in developmental toxicity. For the research on the effects of environmental contaminants on mouse embryonic stem cell differentiation, there also are opportunities for collaboration and data and resource sharing. The gene trap studies are a nice adjunct to the zebrafish project. It will be important for TIVS and the other three STAR centers to collaborate and share data analysis methods.

NCCT is interested in moving predictive capacity of ToxCast™ chemicals to developmental impacts. The development of computer simulations facilitating the assessment of toxicity based on perturbed development in zebrafish and mouse embryonic stem cells will provide a means of incorporating chemical and biological information into systems complex enough to be relevant but not so complex that they are intractable. NCCT is interested in developing virtual embryo systems to validate or invalidate predictions generated by researchers. The opportunities for collaboration with TIVS include merging data from developmentally competent *in vitro* assays with cellular and molecular assay targets, using predictive associations from ToxCast™ high-throughput screening data to build hypotheses about mechanisms of action, conducting studies to generate data testing hypotheses and improving predictive models, and improving virtual tissue models to a level that can help prioritize chemicals for quantitative risk assessment.

A Proposal From the European Commission's Complementary Research Program
Bart van der Burg, BioDetection Systems B.V.

The chemical substance *in vitro/in silico* screening system to predict human and ecotoxicological effects (ChemScreen) is a collaborative project involving nine partners in five countries. It has not yet begun but will span 4 years, with the majority of the practical work to be performed within the first 3 years.

Most of the 100,000 chemicals currently on the market are largely untested. To address this, the Registration Evaluation Authorisation of Chemicals (REACH) Program began in June 2007. Under REACH, industry is responsible for providing data on chemicals. For compounds manufactured or imported in quantities greater than 1 ton, manufacturers and importers must register the compounds with the European Chemicals Agency (ECHA). ECHA also may request additional data as needed. Authorization is required for harmful compounds. Approximately 30,000 chemicals are covered by REACH. Prioritized effects under REACH include chemicals that are carcinogenic, mutagenic, or toxic to reproduction; chemicals that are persistent, bioaccumulative, and toxic; and chemicals that are very persistent and very bioaccumulative. It is estimated that REACH will cost between 2.8 and 5.2 billion Euros during the course of 11 years, but REACH is estimated to save 50 billion Euros over 30 years as a result of health improvements.

When traditional animal tests are used, the progress of REACH will be seriously hampered by ethics, costs, capacity, and speed. To be successful, cost-effective, rapid *in vitro* tests need to be adopted. REACH offers incentives for the use of alternative (nonvertebrate) tests. ECHA publishes test proposals (by chemical manufacturers) and invites third parties to submit alternative proposals. There is an explicit allowance for alternatives to *in vivo* tests, including *in vitro* and nontesting methods (QSAR, grouping, exposure, read across). ECHA accepts the use of suitable methods and regularly reports on the use of alternative methods. ChemScreen focuses on reproductive toxicity because it is important for assessing both human and environmental toxicity, and its prioritization under REACH. Reproductive toxicity uses the most animals in toxicity testing, and unfortunately there are few alternative methods.

The ChemScreen approach is to identify sensitive parameters for reproductive toxicity, identify critical mechanisms involved in perturbation of these parameters, build a high-throughput system using these modules, expand step-wise, integrate with bioinformatics/data interpretation, and build integrated testing strategies, including nontesting methods. Work under ChemScreen will be divided as follows: (1) establishment of *in silico* prescreening and toxicity prediction methods prioritizing *in vitro* toxicity testing, (2) establishment of a database and an *in silico* prescreen to identify potential reproductive toxicants, (3) establishment of sensitive parameters and a medium-throughput minimal essential *in vitro* assay panel, (4)establishment of a high-throughput mechanistic pathway screen for reproductive toxicants, (5) development of integrative methods to predict *in vivo* reprotoxicity allowing informed decisions on prioritization for eventual further testing, (6) integration into one user-friendly tool, and (7) dissemination.

Receptor gene assays that have been shown to reasonably predict the *in vivo* potency of compounds will be used in ChemScreen. Dr. van der Burg displayed a table from a review in *Pediatrics* showing that most compounds that are teratogenic in humans were not identified in animal studies. Screening systems will include a panel of 15 to 50 reporter gene assays in human cells (nuclear receptors, dioxin receptor, signaling/stress/developmental pathways); reporter gene assays in mouse embryonic stem cells (ReProGlow, developmental pathways); wild-type embryonic stem/transcriptomics, metabolizing cell systems, zebrafish/transcriptomics; and others for critical reprotoxicity endpoints (e.g., spermatogenesis). *In silico* tools include an exposure module, a toxicity screening tool, *in vivo* reprotoxicity databases, and an automated decision tool.

Discussion on Research Needs
Maggie Breville, EPA, ORD

Ms. Maggie Breville facilitated the discussion on computational toxicology research needs, using the questions on the handout that was distributed to participants entitled "Research Needs to Advance the Field of Computational Toxicology."

Question 1. Can the same techniques used by ToxCast™ to identify chemicals with a high likelihood of being harmful also be used to identify and/or inform the design of safe chemicals that can be manufactured and used (i.e., green chemistry)? What additional research is needed to make this happen?

Dr. van der Burg said that there is a great opportunity to develop cost-effective screening methods. There also is a danger if a certain screening method is relied on too much, as one method may not be able to identify all toxic chemicals. Research should focus on cost-effectiveness. Dr. Ann Richard noted that it should be recognized that green chemistry products are safer alternatives; a battery of high-throughput screening on both chemicals and alternatives would be helpful. A participant added that green chemistry could help guide modeling efforts to develop safer alternatives. Dr. Dix said that green chemistry needs to be repositioned to serve as a resource for chemical screening and testing.

Question 2. What type of information can we expect toxicity signatures developed through ToxCast™ and other computational methods to provide regarding dose-response, chronic exposures, and potency?

There were no comments on this question. Participants were asked to send their answers via e-mail to Ms. Segal.

Question 3. Is the ultimate goal of computational toxicology research to develop a virtual organism?

Dr. Kavlock said that the ultimate goal is to protect human health and the environment. A virtual organism is a tool to achieve this goal, but it is not the ultimate goal. Dr. Glazier noted that a virtual organism could be used to address Question 2. Dr. Kavlock added that more sophisticated tools are needed to address toxicology in the risk assessment context.

Question 4. For results developed using computational techniques to be used in risk assessments, what research and regulatory questions need to be answered?

A participant observed that this is an important question for regulatory decision-making support. This issue has been addressed in Europe. It might be helpful for EPA to develop guidelines for computational science, especially in terms of metrics and asking questions such as, "Why are we doing this?" Dr. Richard said that to incorporate all of the methods in ToxCast™ there must be a standard for methods development. Dr. Glazier noted that the answer depends on the goal. If the goal is to replace *in vivo*/*in vitro* with *in silico* models, then there will be a different set of false positives and negatives than with *in vivo*/*in vitro*. In the medical device field, changing methods opens up legal liability issues because even if the new method results in fewer false positives and negatives, some people who would not have been hurt by the older method are inevitably hurt by the new method. Dr. Glazier noted that when using new approaches, researchers must be prepared for misses. A participant noted that some models and data are not suited for regulatory purposes.

Question 5. What additional research needs should be addressed?

Dr. Kavlock said that a point raised in the BOSC review was that the current system under which EPA manages the STAR Centers does not encourage or allow the renewal of the Centers. Much time and effort are spent developing synergism and tools, but then it all comes to an end. The BOSC suggests examining ways to evaluate the success of the STAR Centers and keep the research moving forward via renewal of

the grants. A participant noted that there is some redundancy between the Centers and asked if it would be possible to have an annual retreat for the Center leaders to allow for more collaboration. Ms. Segal said that the Centers are asked to set aside funding for attending the progress reviews; having retreats in place of progress reviews is an alternative that could be considered in the future.

Ms. Breville thanked the presenters and attendees for their contributions to the workshop and the support contractors for their logistical assistance. She adjourned the meeting at 4:29 p.m.

U.S. Environmental Protection Agency (EPA) and National Center for Environmental Research (NCER)

Computational Toxicology Centers STAR Progress Review Workshop
October 1, 2009

EPA Main Campus
Building C, Auditorium C111A/B
109 TW Alexander Drive
Research Triangle Park, NC 27711

POST PARTICIPANTS LIST

Cal Baier-Anderson
Environmental Defense Fund

Marianne Barrier
U.S. Environmental Protection Agency

Timothy Barzyk
U.S. Environmental Protection Agency

Maria Bondesson Bolin
University of Houston

Carole Braverman
U.S. Environmental Protection Agency

Maggie Breville
U.S. Environmental Protection Agency

Kelly Chandler
U.S. Environmental Protection Agency

Chris Corton
U.S. Environmental Protection Agency

Alva Daniels
U.S. Environmental Protection Agency

Sally Darney
U.S. Environmental Protection Agency

Jimena Davis
U.S. Environmental Protection Agency

Rob DeWoskin
U.S. Environmental Protection Agency

David Dix
U.S. Environmental Protection Agency

Peter Egeghy
U.S. Environmental Protection Agency

Susan Euling
U.S. Environmental Protection Agency

Richard Finnell
Institute of Biosciences and Technology

Elaine Francis
U.S. Environmental Protection Agency

Panos Georgopoulos
University of Medicine and Dentistry of New Jersey

James Glazier
Indiana University

Najwa Haykal-Coates
U.S. Environmental Protection Agency

David Herr
U.S. Environmental Protection Agency

Ross Highsmith
U.S. Environmental Protection Agency

Keith Houck
U.S. Environmental Protection Agency

Elaine Cohn Hubal
U.S. Environmental Protection Agency

Sid Hunter
U.S. Environmental Protection Agency

Lora Johnson
U.S. Environmental Protection Agency

Bonnie Joubert
U.S. Environmental Protection Agency

Richard Judson
U.S. Environmental Protection Agency

Robert Kavlock
U.S. Environmental Protection Agency

Thomas Knudsen
U.S. Environmental Protection Agency

Robert MacPhail
U.S. Environmental Protection Agency

Matthew Martin
U.S. Environmental Protection Agency

Catherine McCollum
University of Houston

Larry McMillan
U.S. Environmental Protection Agency

Leonard Mole
U.S. Environmental Protection Agency

Holly Mortensen
U.S. Environmental Protection Agency

Michael Morton
U.S. Environmental Protection Agency

Stephanie Padilla
U.S. Environmental Protection Agency

Heidi Paulsen
U.S. Environmental Protection Agency

James Rabinowitz
U.S. Environmental Protection Agency

Joseph Retzer
U.S. Environmental Protection Agency

Ann Richard
U.S. Environmental Protection Agency

Michael Rountree
Rountree Consulting

Ivan Rusyn
University of North Carolina at Chapel Hill

Paul Schlosser
U.S. Environmental Protection Agency

Deborah Segal
U.S. Environmental Protection Agency

Imran Shah
U.S. Environmental Protection Agency

Linda Sheldon
U.S. Environmental Protection Agency

Amar Singh
Lockheed Martin Contractor

Richard Spencer
Lockheed Martin Contractor

Bart van der Burg
BioDetection Systems

John Vandenberg
U.S. Environmental Protection Agency

Vikrant Vijay
National Institutes of Health

John Wambaugh
U.S. Environmental Protection Agency

William Ward
U.S. Environmental Protection Agency

William Welsh
University of Medicine and Dentistry of New Jersey

ClarLynda Williams-Devane
U.S. Environmental Protection Agency

Maritja Wolf
Lockheed Martin Contractor

Fred Wright
University of North Carolina at Chapel Hill

Contractor Support

Ramona Spencer
The Scientific Consulting Group, Inc.