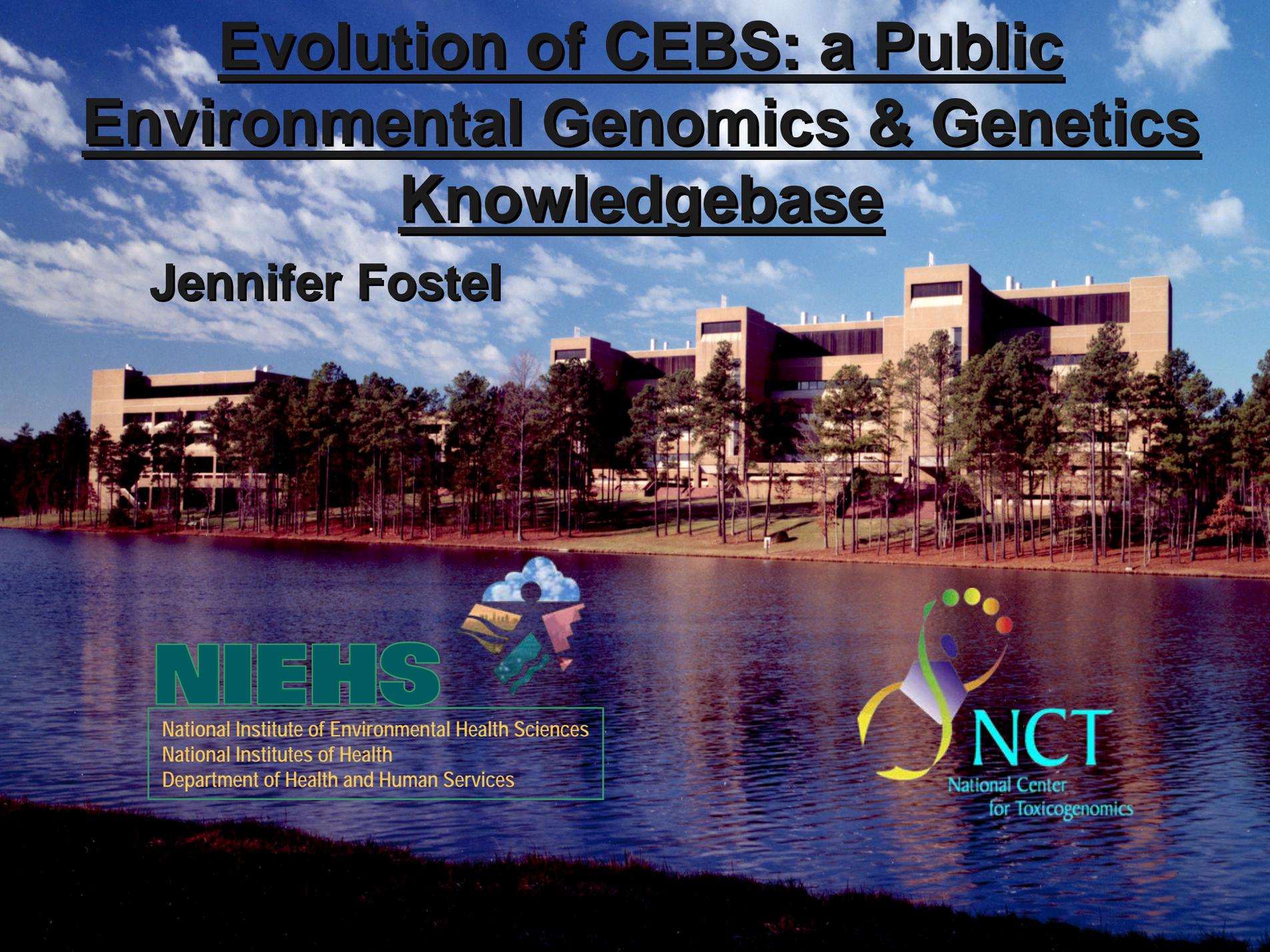# Evolution of CEBS: a Public Environmental Genomics & Genetics Knowledgebase

Jennifer Fostel

NIEHS

National Institute of Environmental Health Sciences
National Institutes of Health
Department of Health and Human Services

NCT
National Center for Toxicogenomics

CEBS: Chemical Effects in Biological Systems

- AIMS:

  - Create a public reference database of environmental chemicals/stressors and their effects on disease / human risk assessment

  - Develop descriptive data compendia on toxicologically important genes, groups of genes, SNPs, mutants, and biological phenotypes relevant to human health and environmental disease

  - Enable hypothesis-driven and discovery research in environmental genetics and genomics

  - Create a knowledgebase capable of deriving new knowledge from the literature and data in CEBS

  - Develop ARC in collaboration with the NCTR National Center for Toxicological Research; use ARC to load data into CEBS and prototype new features

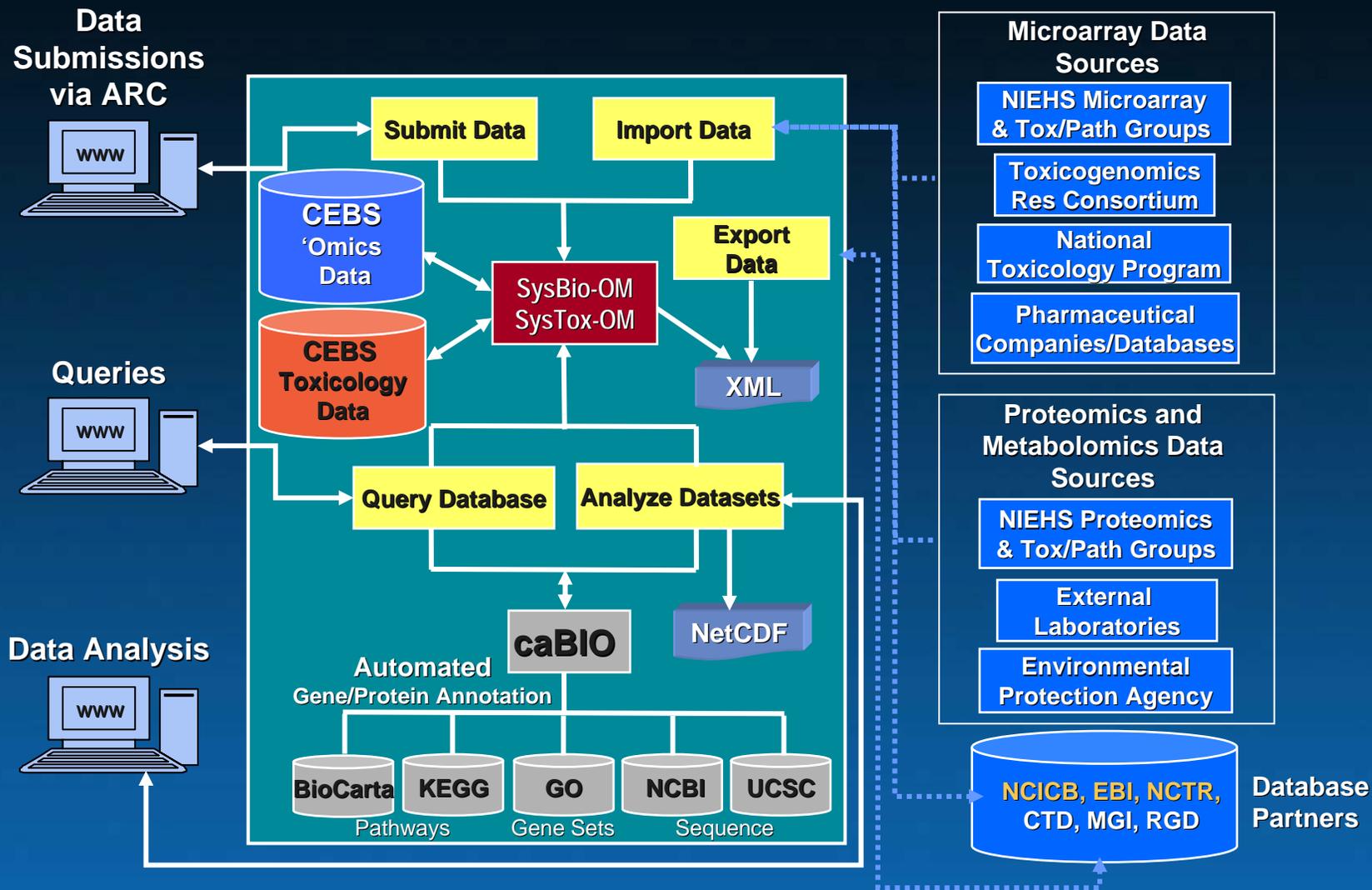# Toxicogenomics: Environmental Genomics and Genetics
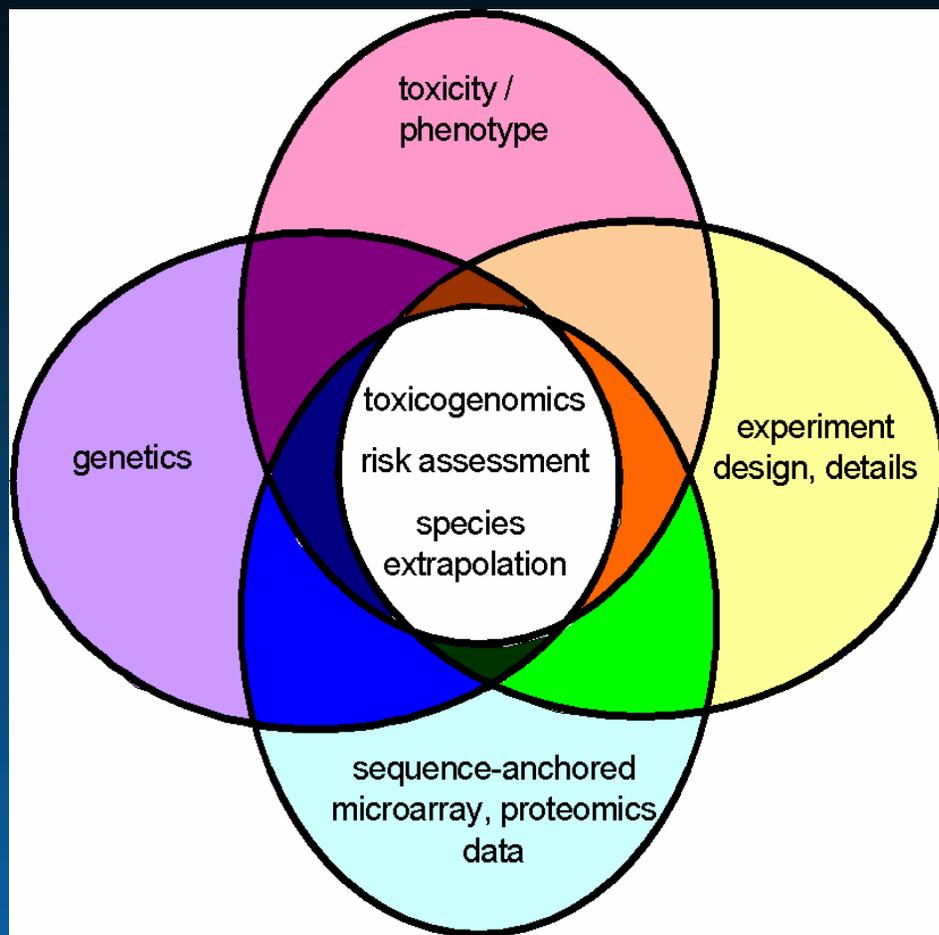
The study of the response of a genome to environmental stressors and toxicants,

interpreted in the context of conventional toxicology and the study design

CEBS status and links

- Project anticipated finish: 2012

- CEBS v1.6 (microarray database w/ example toxicity data)
  - http://cebs.niehs.nih.gov/

- Standardized concepts CEBS data dictionary
  - http://toxsci.oxfordjournals.org/cgi/content/abstract/kfi315v1
  - http://www.niehs.nih.gov/cebs-df/index.cfm

- ARC
  - https://dir-apps.niehs.nih.gov/arc/
  - http://www.niehs.nih.gov/cebs-df/index.cfm

- CEBS v2.0 (integrated toxicogenomics database)
  - In beta testing
  - http://www.niehs.nih.gov/cebs-df/index.cfm

CEBS Architecture   (CEBS v1.6.1)

Microarray - Statistical Analysis

>> **Minimum Fold Change:** 2

>> **Statistical Significance:** (only applicable when comparing multiple samples/arrays)

Step 1. Choose the following test for each gene:

- ⦿ t test: Welch's two sample t test
- ○ Wilcoxon test: Mann-Whitney test

Step 2. Choose a *p*-value threshold and a multiple testing procedure to apply:

- ⦿ Directly use single gene test *p*-values at threshold of 0.05
- ○ Control False Discovery Rate (*FDR*) with adjusted *p*-value below 0.1
  Method to control *FDR*:
  - ⦿ Benjamini & Hochberg step-up procedure
  - ○ Benjamini & Yekutieli step-up procedure
- ○ Control Family-wise Type-I Error Rate (*FWER*) with adjusted *p*-value below 0.1
  Method to control *FWER*:
  - ⦿ Holm step-down procedure
  - ○ Sidak single-step procedure
  - ○ Sidak step-down procedure
  - ○ Bonferroni
- ○ Do not use the *p*-values for gene selection

Data Viewer Options

- ⦿ View Expression Report for All Differentially Expressed Genes
- ○ Perform Gene Category Analysis by BioCarta Pathways
- ○ Perform Gene Category Analysis by KEGG Pathways
- ○ Perform Gene Category Analysis by Gene Ontology (GO)

Gene Ontology for Significantly Expressed Genes

Aspect Type: Biological Process Granularity Level: Level 3 Submit for Analysis

- Total Number of GO Categories: 52
- *This Table is Sortable by Clicking on a Column Header*
- Click on the Following Links for Data Export: Export Report Export Sample Data

GO category	Total	Up	Down	Change	Enrichment	Fisher exact test p-value	Expression report	External links
metabolism	3037	60	73	133	1.54544	0.0	View Details	AmiGO | QuickGO
cellular physiological process	4586	83	94	177	1.36202	1.0E-5	View Details	AmiGO | QuickGO
embryo implantation	5	0	2	2	14.11578	0.00758	View Details	AmiGO | QuickGO
morphogenesis	1034	21	22	43	1.46755	0.01282	View Details	AmiGO | QuickGO
cell differentiation	380	9	10	19	1.76447	0.01797	View Details	AmiGO | QuickGO
response to stimulus	853	20	16	36	1.48936	0.02053	View Details	AmiGO | QuickGO
organ development	894	18	19	37	1.46053	0.02369	View Details	AmiGO | QuickGO

Visualization on KEGG Pathways

Visualization on BioCarta Pathways

Genes On Chip
Expression +/-
Reset
Pathway Summary Report

Annotation for Individual Expressed Genes

Gene Information For: Rn Aldh2 Aldehyde dehydrogenase 2
Sequence ID: NM_032416

Database Links

UniGene LocusLink   DTP SNPViewer Assemblies SNPs SNP500Cancer

Libraries and Tissues (from EST Data)

CEBS v1.6.1 – Public Dataset Characteristics

Industry data:
- Microarray data for multiple doses
- ~ 100 reference compounds
- Sankyo (phenobarb) and Johnson & Johnson (library)

Government data:
- Microarray data for multiple doses, times
- NIEHS DIR, NCT and NTP studies
- 6 hepatotoxicants
- Clin chem & histopath data for acetaminophen

Academic data:
- Microarray data
- Library of untreated recombinant mouse strains (UTenn)

Core CEBS Concept: Anchoring

- CEBS reference anchors:

    – genome (sequence of microarray probes)

    – genotype (strain, SNP, genotype of stubjects)

    – phenotype (pathology findings, clinical measures, anatomy)

    – stressor IDs (DSSTox chemical IDs, gene / disease ontologies)

    – study design / investigation descriptions

Data streams to CEBS
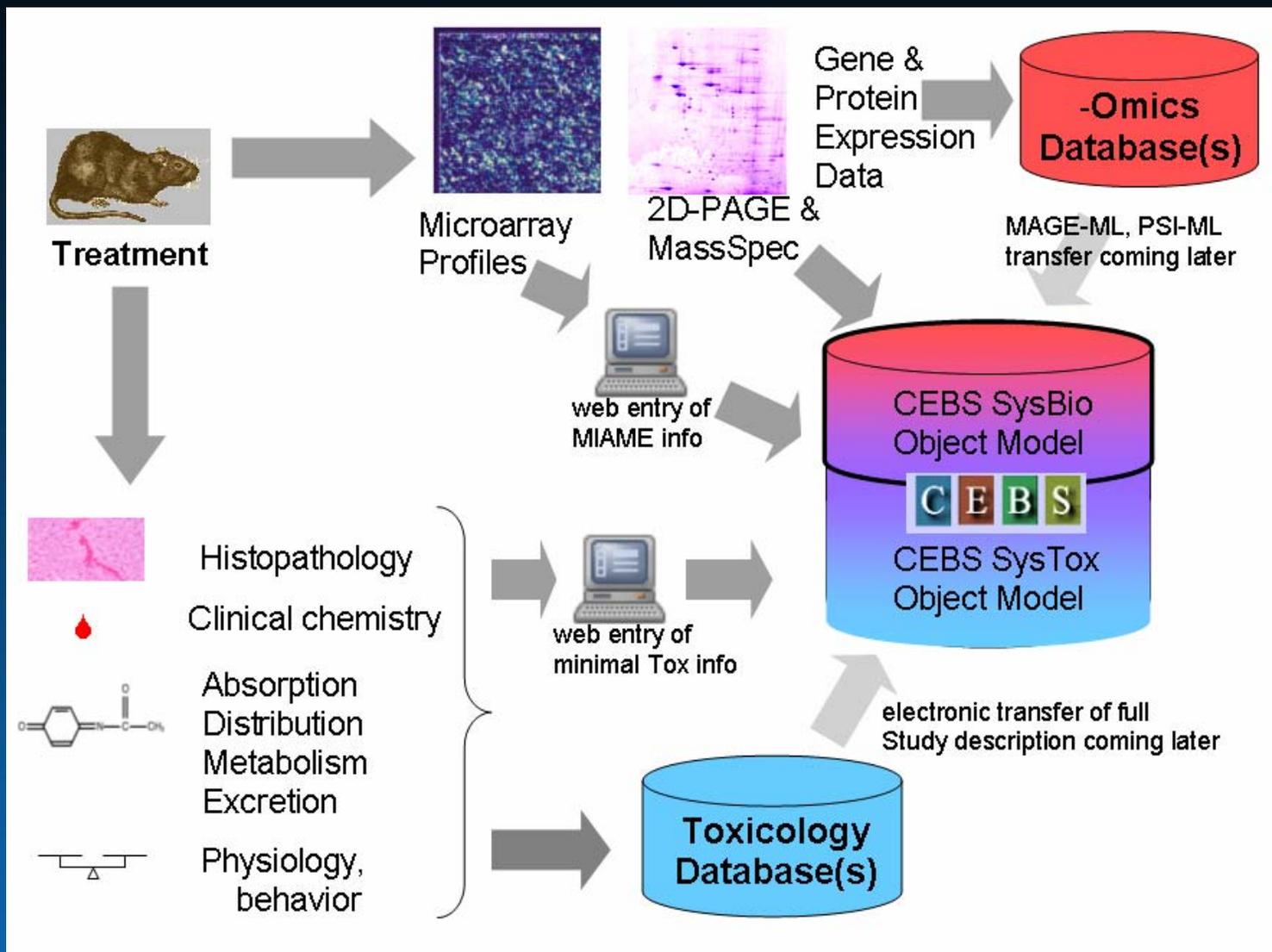
Data types

- Microarray, etc.
  - Exchange format well established
  - MIAME convention, et al.

- Clinical chemistry, hematology, measurements
  - Generally a spreadsheet or export file
  - Terminology straightforward to harmonize

- Histopathology, observations
  - Images, descriptions, spreadsheets, etc.
  - Lexicons; vocabulary not constrained

Need biological context to manage the data

- Which data came from a given subject

- When measurements were made relative to other Study events

- How and when Subjects were treated, observed, cared for and sacrificed / exited the Study

- Which Subjects are biological replicates

- Which experimental factors apply to a given Subject

- Characteristics of the Subjects and Stressors

Clock Time
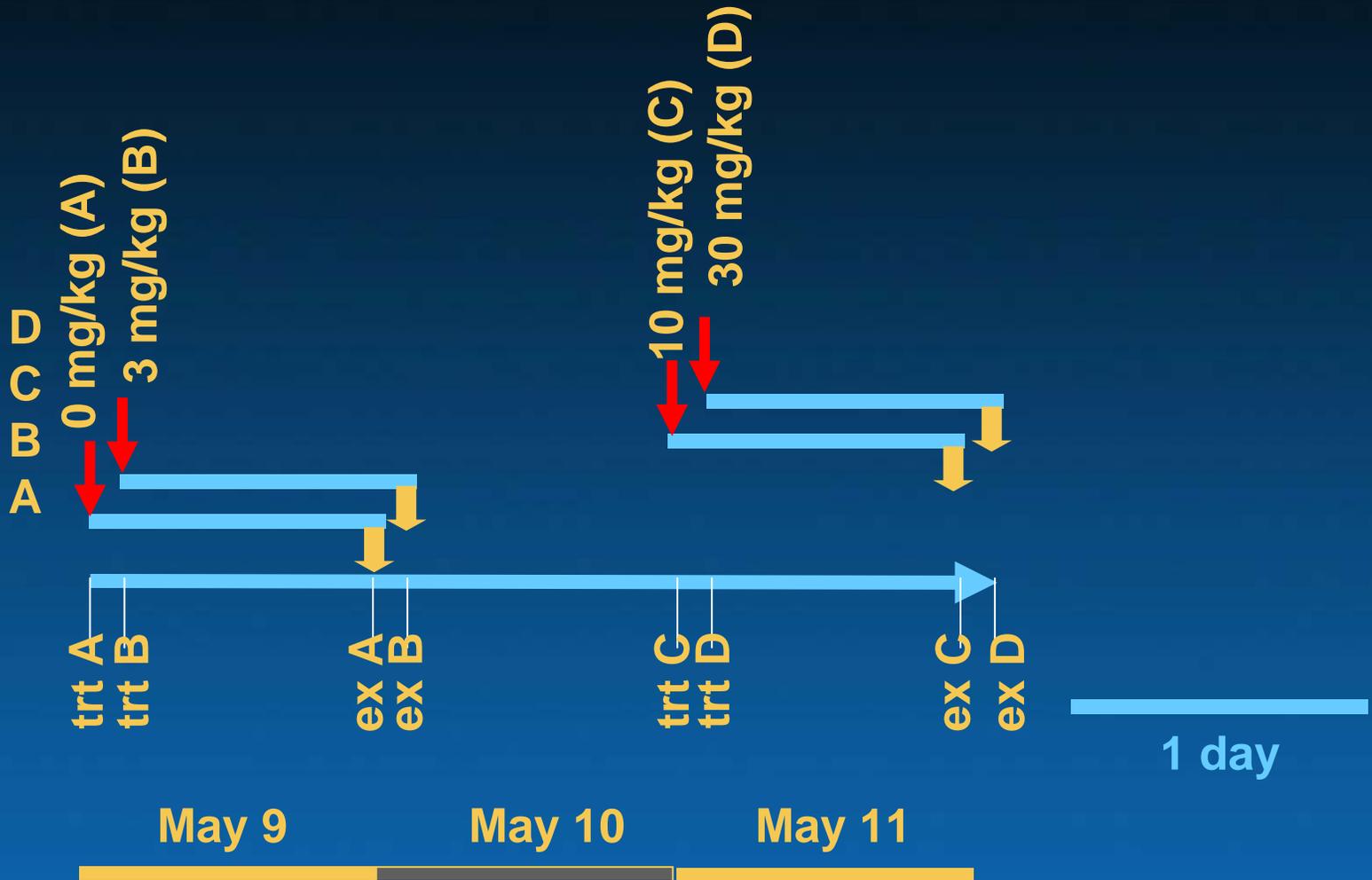
Study Time

study treatment applied to subject

1 day

subject exits the study

for each of four groups:
0 mg / kg - > A
3 mg / kg -> B
10 mg / kg -> C
30 mg / kg -> D

Study Components ("metadata")

- Used to describe and manage data

- Stressors

- Subjects and groups

- Timeline

- Protocols

- Study Details

Motivation for CEBS-Data Dictionary

- Public repository => multiple study designs

- Many factors differ / which are important?

  - feed, housing, timing, …

- CEBS will serve pathologists, bioinformaticians, toxicologists, chemists, risk assessors, scientific community, general public – different types of queries

- Capture as much as possible in queryable state

Repositories for Toxicity Data

- **In Vivo Data Warehouse** Lilly Greenfield Research Lab
- **TDMS and ClinChem DB** US National Toxicity Program

Study Results
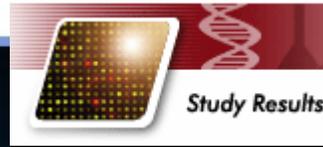
Formats for Exchange of Toxicity Data with Regulators

- **SEND** Standard for Exchange of Nonclinical Data CDISC, Pharmquest

CDISC

- **21CFR part11-compliant repositories,** e.g. Tox/Path from Xybion Medical Systems

Xybion Medical Systems

Other Toxicogenomics Database

- **MIAME/Tox & Tox/ArrayExpress**
  HESI Committee: Application of Genomics to Mechanism based Risk Assessment
  US National Center for Toxicogenomics
  The EMBL European Bioinformatics Institute

EMBL-EBI

ILSI — International Life Sciences Institute

Toxicity Data Indexed by Chemical Structure

- **DSSTox, US EPA**
- **ToxML, LIST Consortium**

Leadscope

COMPUTATIONAL TOXICOLOGY

C E B S

CHEMICAL EFFECTS in BIOLOGICAL SYSTEMS

Data Dictionary

LIMS systems

- MAPS (NIEHS)
- TSP (US EPA)

NIEHS National Institute of Environmental Health Sciences

Lilly

CEBS Data Dictionary – Toxicological terms & synonyms from public data formats

One entry in the CEBS-DD

Content of Row in CEBS-DD	Name of Column in CEBS-DD
Required	CEBS-minimal flag
Study	CEBS entity
Study title	CEBS term
The descriptive experiment title for the Stu	CEBS term definition
text	CEBS expected content
Study	SysTox table name
StudyTitle	SysTox Field names
	SysTox table name
	SysTox Field names
tProject	TSP Table Name
	TSP Link ID
	TSP LookUp Table
ProjectName	TSP Field
TEST_ARTICLES	NTP_DATA Table
TEST_ARTICLE_NO	NTP_DATA link
CT_CHEMTRACK_DATA	TDMSE Table
TDMS_STUDY_NO	TDMSE link
	TDMSE Table 2
	TDMSE Table 2 link
	TDMSE TABLE 2 target field
	TDMSE Table 3 / comments
	TDMSE TABLE 3 or comments
All	LEADSCOPE WORKBOOK
StudyTitle	LEADSCOPE FIELD
The StudyTitle of the report	LEADSCOPE Description
Protocol	Xybion-module
Protocol Information (general information)	Xybion-table
ZT.GEN	Xybion-code
Study Title	Xybion-ID
SS	SEND DOMAIN
SSTESTCD	SEND LABEL
STTITL	TERM
SSORRES	RESULTS
Study TitleCharQualifierTitle of study. Example: "91-Day Feeding Study with Compound XYZ in Fischer 344 Rats".	SEND variable label or usage notes
STUDY	Lilly-entity
STUDY NAME	Lilly-entity/attribute
The precise and unambiguous label or specification used to identify a particular (i.e. instance of) STUDY; the title of a STUDY.	Lilly-definition

CEBS-DD: Common Nomenclature Informed by Different Formats, Different Aims and Interests

Term = "Study Title"
Belongs to "Study"
Definition: "The descriptive title for the Study"

Synonyms are found in:
○ TDMS: STUDY_NO in CHEMTRACK_DATA
○ Lilly: Study Name in STUDY
      (Toxicity Data Repositories)

○ Xybion Path/Tox: Study Title in ZT.GEN
○ SEND: STTITL, SSORRES in SS
      (Exchange Formats for Regulators)

● Tox-ML: Study Title in all workbooks
      (Chemical Structure Index)

○ TSP: Project name, in tProject
      (example LIMS system)

Stressor

- CEBS includes interventional and observational Studies

- All Stressor protocols need stressor name, dose, dose units,

  frequency of dosing, frequency units

- Stressor Types:

  - Chemical Stressor

  - Environmental Stressor

  - Genetic Stressor

  - Disease Stressor

  - Mechanical / Surgical Stressor

  - Nanoparticle Stressor

Subjects and Groups

- Experimental Subjects
  - can be lab animals, humans, in vitro cultures, etc.

- Subject Group
  - group of biological replicates
  - established by the experimental factors

- Specimen
  - part of a Subject; no longer part of the Study timeline
  - can produce a subSpecimen

- Pool
  - made from Specimens or subSpecimens

Timeline

- Linear arrangement of events that happened during the Study, includes "History"

- Can use either Study time or Clock time

- Link Protocol, Subject Group and Event

- Scheduled events

  – planned by the PI; apply to Subject Groups

- Unscheduled events

  – planned by human Subjects or experienced by Subjects in environmental studies; apply to individual Subjects

  – Concomitant Medication, Substance Use

Protocols

- Stressor application
  - Each Stressor type has a protocol type

- Care
  - husbandry, culture, clinical care

- Disposition
  - euthanasia, harvest, study exit

- Specimen preparation and preservation

- Observation

- Assay (performed outside the Study timeline)

What is minimal information for a study?

- CEBS-DD aims to be "maximal"

- recommended for "minimal":
  - study timeline
  - timing of treatment, disposition events
  - subject IDs and groups / experimental factors
  - phenotype (e.g. pathology or clinical chemistry)
  - species, strain, sex

ARC = ArrayTrack & CEBS

- Integration of ArrayTrack microarray domain and CEBS-DD

- Flexible: can add new terms and concepts

- Load data from multiple sources, formats

- Visualize and curate data

- Pipeline data to CEBS using standard format

- Prototype new tools and capabilities at NIEHS

- Integrated into ArrayTrack analysis; running at NCTR

ARC home page
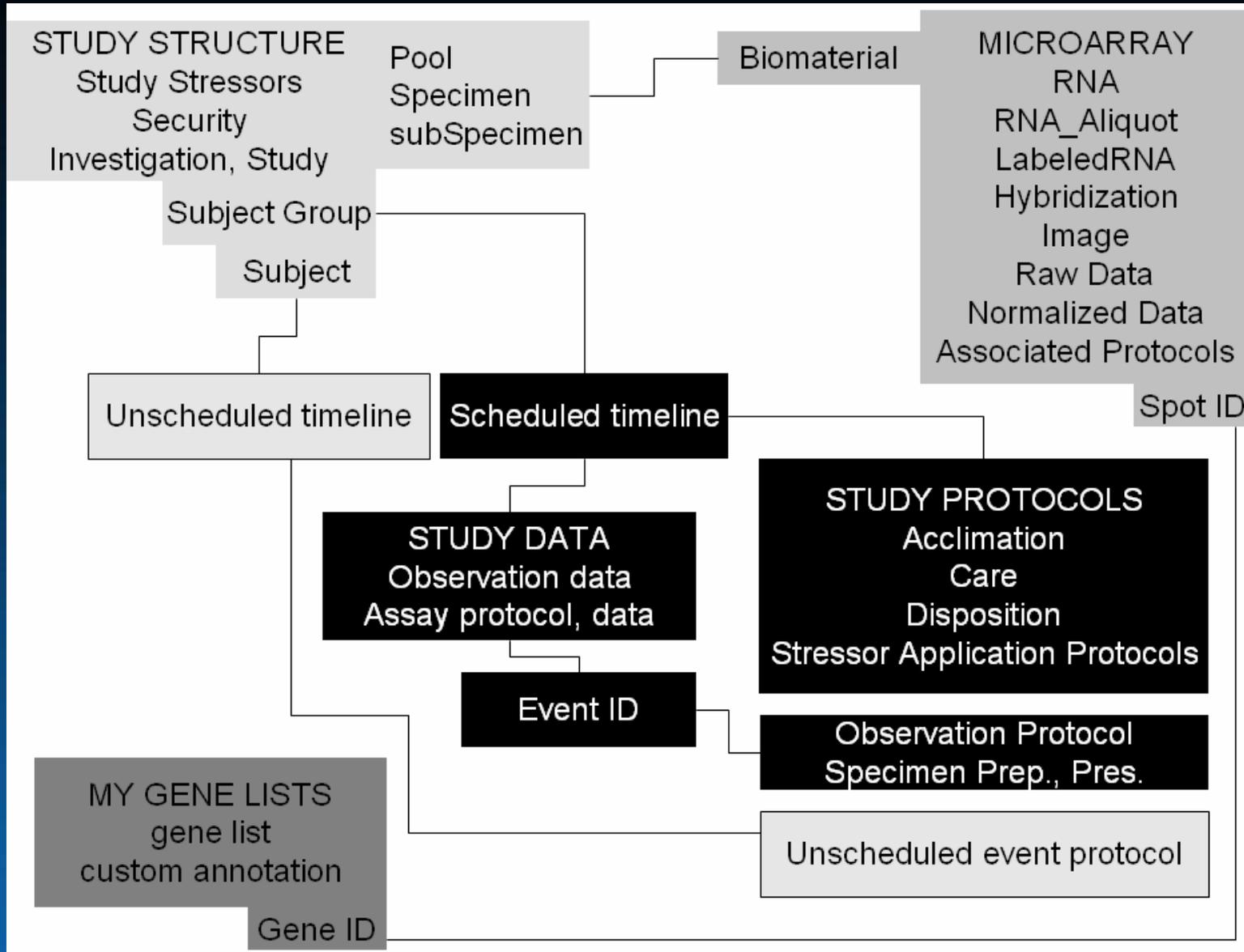
Search ARC by study characteristics;
use to verify data entry as well as to support user query

List of studies in ARC

View of Study details within ARC

Study Details

Study Characteristics

CEBS Accession #	001-00001-0003-000-5
Study Title	Application of 1,2-dichlorobenzene to F344 rats via oral gavage to evaluate acute toxicity
Species	Rat
Strain	F344/N
Study Discipline	Acute toxicology
Stressor Type	Chemical
Stressor Name	1,2-Dichlorobenzene
Start Date	2004-09-14 00:00:00.0
Expected/observed Target Organ	Liver
Expected/observed pathology, toxicity	
Pharmacological Action	
Study Description	This study will examine the gene expression, clinical chemistry and pathology profile in the liver and kidney of rats 6, 24, and 48 hours after exposure to a single dose of 1, 2-dichlorobenzene.
Publication citation	
Study Number (Depositor)	N114-303 (NCT028)
Subgroup factor	
Reason for species	The rat was selected because of its widespread use in toxicology studies, and along with data from a pilot study conducted at ILS, there exists previous data in F344 rats for the toxic effects of 1,2-dichlorobenzene on hepatic and/or renal cells.
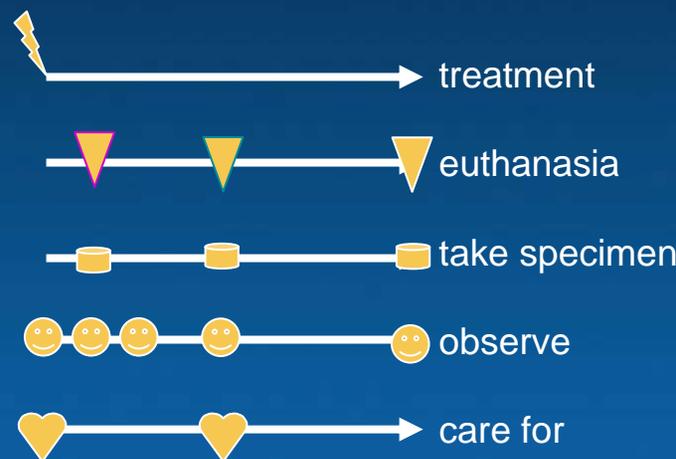
Close

View of study design and study timeline

Study Design - (Bromobenzene-Rat-F344/N-2004)

Group Treatments

Factor 2 - Dose (mg/kg)	Factor 1 - Time (hour)		
	6	**24**	**48**
0	Vehicle 6 hour	Vehicle 24 hour	Vehicle 48 hour
25	25mg/kg 6 hour	25mg/kg 24 hour	25mg/kg 48 hour
75	75mg/kg 6 hour	75mg/kg 24 hour	75mg/kg 48 hour
250	250mg/kg 6 hour	250mg/kg 24 hour	250mg/kg 48 hour

Close

Study Timeline

Study Timeline

Investigation Title: Molecular characterization of phenotypic response to parallel acute exposure to one of 8 hepatotoxicants to F344 rats (Accession # 001-00001-0001-000-3)
Study Title: Application of bromobenzene to F344 rats via oral gavage to evaluate acute toxicity (Accession # 001-00001-0002-000-4)

Study Duration (hour) ▶		1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
Event types ▼ / Phases ▶		No Phase defined
Treatment	Protocol	
Observation	Protocol ☺	☺ ☺
Care	Protocol ♥	
Disposition	Protocol	✖ ✖
Specimen Prep & Assays	Protocol	✎ ✎

Note: Click on protocol to see protocol details.

Access protocol details from the study timeline

Study Timeline

Study Timeline

Investigation Title: Molecular characterization of phenotypic response to parallel acute exposure to one of 8 hepatotoxicants to F344 rats (Accession # 001-00001-0001-000-3)

Study Title: Application of bromobenzene to F344 rats via oral gavage to evaluate acute toxicity (Accession # 001-00001-0002-000-4)

Study Duration (hour) ▶	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
Event types ▼ / Phases ▶	No Phase defined
Treatment	Protocol
Observation	Protocol
Care	Protocol
Disposition	Protocol
Specimen Prep & Assays	Protocol

Note: Click on protocol to see protocol details.

Chemical Stressor

Protocol Name	Chemical name	Purity (%)	Vehicle name	Dose per admin (Unit)	Route admin
Corn oil vehicle	Corn oil		Corn oil	()	Oral gavage
1,2-Dichlorobenzene low dose	1,2-Dichlorobenzene		Corn oil	15 (mg/kg)	Oral gavage
1,2-Dichlorobenzene mid dose	1,2-Dichlorobenzene		Corn oil	150 (mg/kg)	Oral gavage
1,2-Dichlorobenzene high dose	1,2-Dichlorobenzene		Corn oil	1500 (mg/kg)	Oral gavage

Details of subjects in one of the groups are available from the study design grid (click on group name)

Study Design - (Epidermal abrasion-Mouse-TgAC-2003)

Group Treatments

Strain and abrasion		Time (day)				
		3	**5**	**9**	**18**	**30**
	Abraded Tg.AC	Abraded Tg.AC 3-days	Abraded Tg.AC 5-days	Abraded Tg.AC 9-days	Abraded Tg.AC 18-days	Abraded Tg.AC 30-days
	Abraded FVB/N	Abraded FVB/N 3-days	Abraded FVB/N 5-days	Abraded FVB/N 9-days	Abraded FVB/N 18-days	Abraded FVB/N 30-days

Close

Legends:

Control Group - Group

Normal Group - Group

Group Subjects List

Group Subjects

Group Name: Abraded Tg.AC 5-days
Comparitor Name: Abrasion_D5FA
Factor 1: 5 day
Factor 2: Abraded Tg.AC
Genus: Mus
Species: musculus (Mouse)
Strain: Tg.AC

Subject ID	Sex	Exceptions/Comments
Abrasion_D5TA-1	Female	
Abrasion_D5TA-2	Female	
Abrasion_D5TA-3	Female	
Abrasion_D5TA-4	Female	
Abrasion_D5TA-5	Female	

Detail | Back | Close

Users can also query by subject characteristics

Search study by subject characteristics

Search Options

Categories	Fields	Available Values
Groups	Select Field	
Subjects	Select Field	
Specimen	Select Field	
Collected Study Data		
Clinical Chemistry	Select Field	
Hematology	Select Field	
Histopathology	Select Field	
Observations	Select Field	

Add

Search Result

10 Studies/50 Groups/200 Subjects are found with above criteria   Search Result >>

Study Histopathlogy Data

Study = Evaluation of the acute toxicity of carbon tetrachloride administered via oral gavage in male F344 rats

Show Images   Close

Specimen Name	Organ	Diagnosis	Description	Severity	Distribution	Infiltration Cell Type

Close

NB: possible to link to image at MRPath

Retrieval of clinical chemistry data: flexible sorting / filtering screen

Study Clinical Chemistry Data

Filter Options

Study	Select Values	**Group**	Select Values
Specimen	Select Values	**Test name**	Select Values

Graph   Close

Specimen Name	Test Name	Test Value	Comments
Study - Application of bromobenzene to F344 rats via oral gavage to evaluate acute toxicity			
Group - High Dose 24-hours			
BB_29_serum	Alanine Aminotransferase (ALT)	5720 (U/L)	
BB_29_serum	Albumin	1.4 (g/dl)	
BB_29_serum	Alkaline Phosphatase	(U/L)	Quantity not sufficient.
BB_29_serum	Aspartate Aminotransferase (AST)	44800 (U/L)	
BB_29_serum	Cholesterol	40 (mg/dl)	
BB_29_serum	Creatine Kinase	(U/L)	Quantity not sufficient.
BB_29_serum	Creatinine	0.5 (mg/dl)	
BB_29_serum	Direct Bilirubin	(mg/dl)	Quantity not sufficient.
BB_29_serum	Lactate Dehydrogenase (LDH)	80000 (U/L)	
BB_29_serum	Serum Urea Nitrogen (BUN)	29 (mg/dl)	
BB_29_serum	Sorbitol Dehydrogenase (SDH)	(U/L)	Quantity not sufficient.
BB_29_serum	Total Bile Acids	155.9 (mMol/L)	
BB_29_serum	Total Bilirubin	(mg/dl)	Quantity not sufficient.
BB_29_serum	Total Protein	6.2 (g/dl)	
BB_29_serum	Triglycerides	91 (mg/dl)	
BB_30_serum	Alanine Aminotransferase (ALT)	5160 (U/L)	
BB_30_serum	Albumin	5 (g/dl)	
BB_30_serum	Alkaline Phosphatase	429 (U/L)	
BB_30_serum	Aspartate Aminotransferase (AST)	41300 (U/L)	
BB_30_serum	Cholesterol	35 (mg/dl)	
BB_30_serum	Creatine Kinase	257 (U/L)	
BB_30_serum	Creatinine	0.3 (mg/dl)	
BB_30_serum	Direct Bilirubin	0.3 (mg/dl)	
BB_30_serum	Lactate Dehydrogenase (LDH)	78100 (U/L)	
BB_30_serum	Serum Urea Nitrogen (BUN)	24 (mg/dl)	
BB_30_serum	Sorbitol Dehydrogenase (SDH)	340 (U/L)	
BB_30_serum	Total Bile Acids	140.4 (mMol/L)	
BB_30_serum	Total Bilirubin	1.9 (mg/dl)	
BB_30_serum	Total Protein	7.2 (g/dl)	
BB_30_serum	Triglycerides	222 (mg/dl)	
BB_31_serum	Alanine Aminotransferase (ALT)	4780 (U/L)	
BB_31_serum	Albumin	4.8 (g/dl)	
BB_31_serum	Alkaline Phosphatase	387 (U/L)	
BB_31_serum	Aspartate Aminotransferase (AST)	40300 (U/L)	
BB_31_serum	Cholesterol	42 (mg/dl)	

Filter rapidly to see subjects / tests of interest

Study Clinical Chemistry Data

Filter Options

Study	Bromobenzene-Rat-F344/N-2004 ▼	Group	Select Values ▼
Specimen	Select Values ▼	Test name	Aspartate Aminotransferase (AST) ▼

[Graph] [Close]

Specimen Name	Test Name	Test Value	Comments
Study - Application of bromobenzene to F344 rats via oral gavage to evaluate acute toxicity			
Group - High Dose 24-hours			
BB_29_serum	Aspartate Aminotransferase (AST)	44800 (U/L)	
BB_30_serum	Aspartate Aminotransferase (AST)	41300 (U/L)	
BB_31_serum	Aspartate Aminotransferase (AST)	40300 (U/L)	
BB_32_serum	Aspartate Aminotransferase (AST)	44400 (U/L)	
Group - High Dose 48-hours			
BB_45_serum	Aspartate Aminotransferase (AST)	5700 (U/L)	
BB_46_serum	Aspartate Aminotransferase (AST)	4360 (U/L)	
BB_47_serum	Aspartate Aminotransferase (AST)	4220 (U/L)	
BB_48_serum	Aspartate Aminotransferase (AST)	3540 (U/L)	
Group - High Dose 6-hours			
BB_13_serum	Aspartate Aminotransferase (AST)	149 (U/L)	
BB_14_serum	Aspartate Aminotransferase (AST)	179 (U/L)	
BB_15_serum	Aspartate Aminotransferase (AST)	85 (U/L)	
BB_16_serum	Aspartate Aminotransferase (AST)	124 (U/L)	
Group - Low Dose 24-hours			
BB_21_serum	Aspartate Aminotransferase (AST)	81 (U/L)	
BB_22_serum	Aspartate Aminotransferase (AST)	114 (U/L)	
BB_23_serum	Aspartate Aminotransferase (AST)	106 (U/L)	
BB_24_serum	Aspartate Aminotransferase (AST)	104 (U/L)	
Group - Low Dose 48-hours			
BB_37_serum	Aspartate Aminotransferase (AST)	78 (U/L)	
BB_38_serum	Aspartate Aminotransferase (AST)	97 (U/L)	
BB_39_serum	Aspartate Aminotransferase (AST)	83 (U/L)	
BB_40_serum	Aspartate Aminotransferase (AST)	107 (U/L)	
Group - Low Dose 6-hours			
BB_5_serum	Aspartate Aminotransferase (AST)	82 (U/L)	
BB_6_serum	Aspartate Aminotransferase (AST)	67 (U/L)	
BB_7_serum	Aspartate Aminotransferase (AST)	67 (U/L)	
BB_8_serum	Aspartate Aminotransferase (AST)	85 (U/L)	
Group - Mid Dose 24-hours			
BB_25_serum	Aspartate Aminotransferase (AST)	129 (U/L)	
BB_26_serum	Aspartate Aminotransferase (AST)	130 (U/L)	
BB_27_serum	Aspartate Aminotransferase (AST)	508 (U/L)	
BB_28_serum	Aspartate Aminotransferase (AST)	830 (U/L)	
Group - Mid Dose 48-hours			

Plot clinical chemistry values; X-axis uses either experimental factor
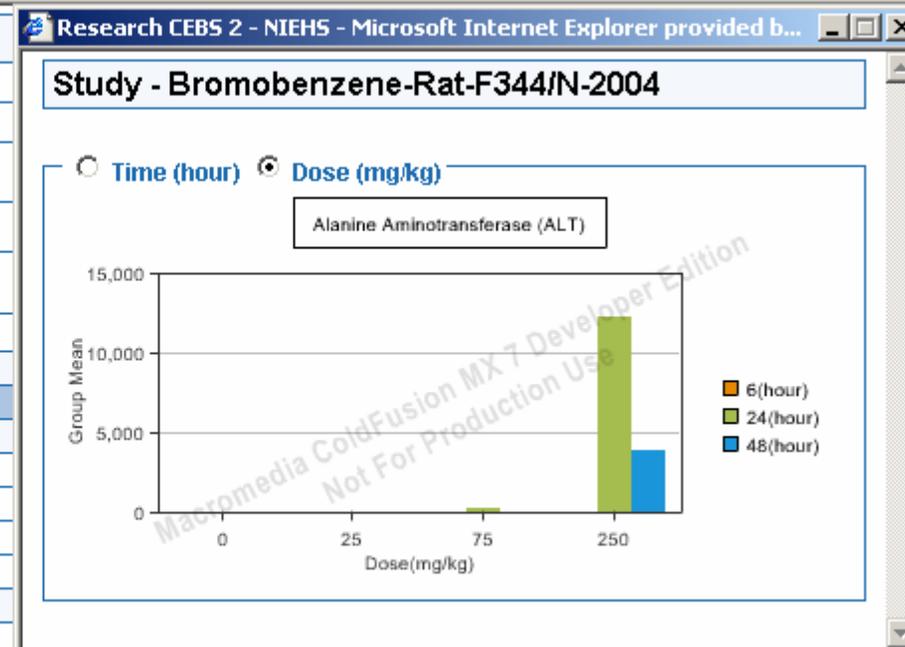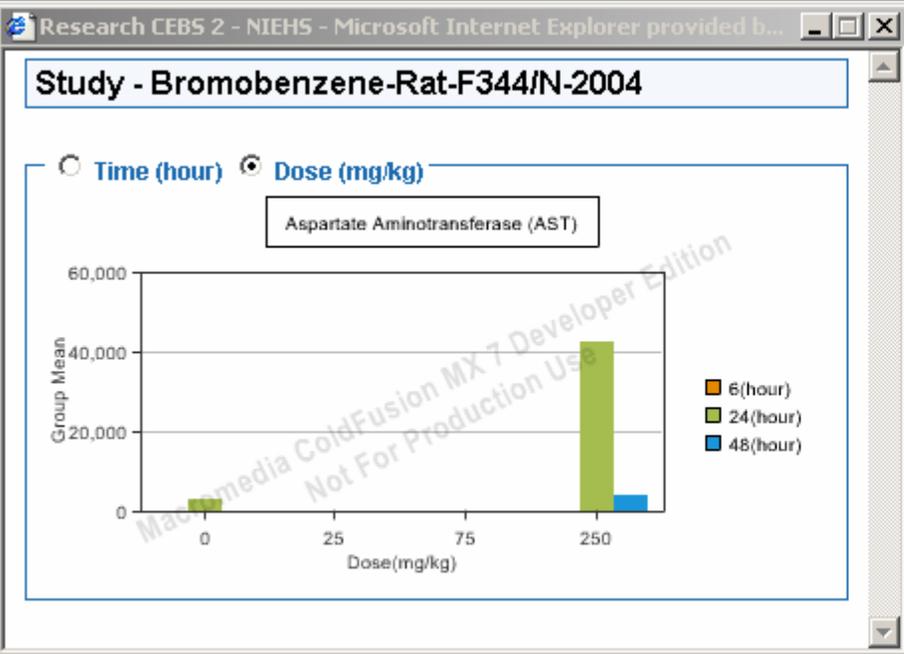
With login one can export files from ARC

Research CEBS

Home

- Home
- Search
- Data Import
- Data Export
- My Genes List
- Logout
- Contact Us

Data Export - Export Data from Research CEBS

- **CEBS required file format**
  Tox data
- **CEBS required file format**
  Microarray data

Export files required by CEBS (2 clicks – create & export)

Data Export - CEBS Data Format (Tox Data)

Export Options

Investigation	Allyl alcohol: Responses over time to different dose levels in male F344/N rats.
Study	Evaluation of Acute Toxicity of Allyl Alcohol Via Oral Administration in Male F344 Rats

Files to Export - [Select All]

☑ Investigation	☑ Study	☑ Phase	☑ Person
☑ StudyPartyRole	☑ GroupCharacteristics	☑ Group	☑ ChemicalStressorProtocol
☑ ChemicalStressorCharacteristics	☑ GeneticStressorProtocol	☑ GenenticStressorCharacteristics	☑ EnvironmentStressorProtocol
☑ CellCultureProtocol	☑ AnimalCareProtocol	☑ DispositionProtocol	
☑ SpecimenPreparationProtocol	☑ SpecimenPreservationProtocol	☑ ChemicalStressorTreatment	☑ GeneticStressorTreatment
☑ EnvironmentStressorTreatment	☑ CellCulture	☑ AnimalCare	☑ DispositionEvent
☑ SpecimenPreparation	☑ GroupChemicalStressorTreatment	☑ GroupGeneticStressorTreatment	☑ GroupEvironStressorTreatment
☑ GroupCellCulture	☑ GroupAnimalCare	☑ GroupDisposition	☑ GroupSpecimenPreparation
☑ GroupObservation	☑ Subject	☑ SpecimenProtocolSubject	☑ SpecimenPool
☑ SubSpecimen	☑ SpecimenCharacteristics	☑ ObservationProtocol	☑ ObservationEvent
☑ ClinicalChemistryTestProtocol	☑ HematologyTestProtocol	☑ PathologyTestProtocol	☑ SubjecrClinicalObservation
☑ Subject			
☑ Specim			

Data Export - CEBS Data Format (Microarray)

Export Options

Investigation	Allyl alcohol: Responses over time to different dose levels in male F344/N rats.
Study	Evaluation of Acute Toxicity of Allyl Alcohol Via Oral Administration in Male F344 Rats

Files to Export - [Select All]

|			
☑ ExperimentInformation	☑ RNAToSpecimen	☑ RNA	☑ RNALabeling
☑ Hybridization	☑ Hyb & Array Design Files		

Export

Zip file ready for CEBS
(study description, subjects & protocols, tox data, microarray experiment description and samples)

ExperimentInformation.txt
Hybridization.txt
RNA.txt
RNALabeling.txt
RNAToSpecimen.txt

File	Date	Size	Ratio	Packed	Type
AnimalCare.txt	9/16/2005 8:48 AM	331	49%	168	te...
AnimalCareProtocol.txt	9/16/2005 8:48 AM	1,149	52%	554	te...
CellCulture.txt	9/16/2005 8:48 AM	228	47%	120	te...
CellCultureProtocol.txt	9/16/2005 8:48 AM	591	55%	263	te...
ChemicalStressorCharacteristics.txt	9/16/2005 8:48 AM	894	59%	368	te...
ChemicalStressorProtocol.txt	9/16/2005 8:48 AM	2,432	79%	501	te...
ChemicalStressorTreatment.txt	9/16/2005 8:48 AM	706	70%	210	te...
ClinicalChemistryTestProtocol.txt	9/16/2005 8:48 AM	843	68%	270	te...
DispositionEvent.txt	9/16/2005 8:48 AM	736	74%	188	te...
DispositionProtocol.txt	9/16/2005 8:48 AM	623	56%	275	te...
EnvironmentalStressorProtocol.txt	9/16/2005 8:48 AM	795	57%	342	te...
EnvironmentalStressorTreatment.txt	9/16/2005 8:48 AM	252	50%	125	te...
GeneticStressorCharacteristics.txt	9/16/2005 8:48 AM	173	35%	113	te...
GeneticStressorProtocol.txt	9/16/2005 8:48 AM	564	53%	266	te...
GeneticStressorTreatment.txt	9/16/2005 8:48 AM	242	50%	122	te...
Group.txt	9/16/2005 8:48 AM	2,813	83%	489	te...
GroupAnimalCare.txt	9/16/2005 8:48 AM	947	85%	146	te...
GroupCellCulture.txt	9/16/2005 8:48 AM	32	0%	34	te...
GroupCharacteristics.txt	9/16/2005 8:48 AM	4,787	90%	480	te...
GroupChemicalStressorTreatment.txt	9/16/2005 8:48 AM	1,124	84%	182	te...
GroupDisposition.txt	9/16/2005 8:48 AM	1,418	88%	167	te...
GroupEnvironStressorTreatment.txt	9/16/2005 8:48 AM	42	0%	42	te...
GroupGeneticStressorTreatment.txt	9/16/2005 8:48 AM	36	0%	38	te...
GroupObservation.txt	9/16/2005 8:48 AM	6,582	93%	453	te...
GroupSpecimenPreparation.txt	9/16/2005 8:48 AM	2,012	92%	164	te...
HematologyTestProtocol.txt	9/16/2005 8:48 AM	883	64%	318	te...
Investigation.txt	9/16/2005 8:48 AM	273	28%	196	te...
ObservationEvent.txt	9/16/2005 8:48 AM	991	73%	269	te...
ObservationProtocol.txt	9/16/2005 8:48 AM	686	67%	226	te...
PathologyTestProtocol.txt	9/16/2005 8:48 AM	238	40%	143	te...
Person.txt	9/16/2005 8:48 AM	288	54%	132	te...
Phase.txt	9/16/2005 8:48 AM	174	40%	105	te...
SpecimenCharacteristics.txt	9/16/2005 8:48 AM	130,966	97%	4,033	te...
SpecimenClinicalChemTestRes.txt	9/16/2005 8:48 AM	237	50%	119	te...
SpecimenHematologyTestRes.txt	9/16/2005 8:48 AM	262,720	96%	11,713	te...
SpecimenPathologyTestRes.txt	9/16/2005 8:48 AM	497	41%	293	te...
SpecimenPool.txt	9/16/2005 8:48 AM	1,609	87%	202	te...
SpecimenPreparation.txt	9/16/2005 8:48 AM	567	69%	177	te...
SpecimenPreparationProtocol.txt	9/16/2005 8:48 AM	776	63%	291	te...
SpecimenPreservationProtocol.txt	9/16/2005 8:48 AM	1,424	55%	640	te...
SpecimenProtocolSubject.txt	9/16/2005 8:48 AM	129,345	96%	5,453	te...
Study.txt	9/16/2005 8:48 AM	1,992	53%	937	te...
StudyPartyRole.txt	9/16/2005 8:48 AM	134	32%	91	te...
Subject.txt	9/16/2005 8:48 AM	33,637	96%	1,331	te...
SubjectClinicalObservations.txt	9/16/2005 8:48 AM	278	53%	131	te...
SubjectOrganData.txt	9/16/2005 8:48 AM	206	47%	109	te...
SubSpecimen.txt	9/16/2005 8:48 AM	184	60%	73	te...

Experimental design: Acetaminophen (APAP)

(Low) SubToxic
150 mg/kg/day

(Hi) Toxic, Recoverable
1500 mg/kg/day

0   6   24   48

Time (Hrs)
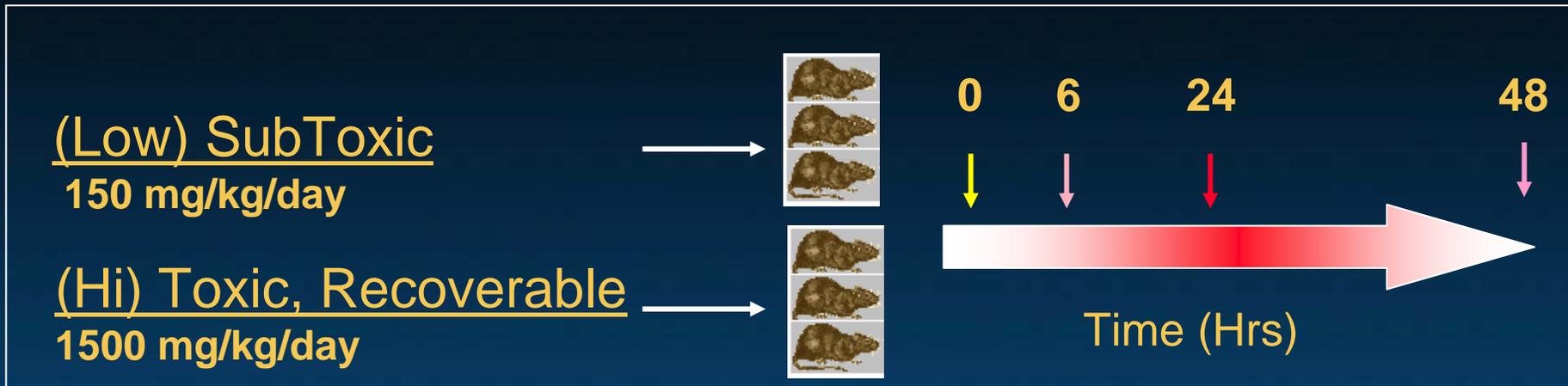
• Acetaminophen is well studied; both therapeutic and toxic effects

• Liver toxicity is a common response in rodents and humans

• Metabolism is similar in rodents and humans

• Opportunities for clinical investigation

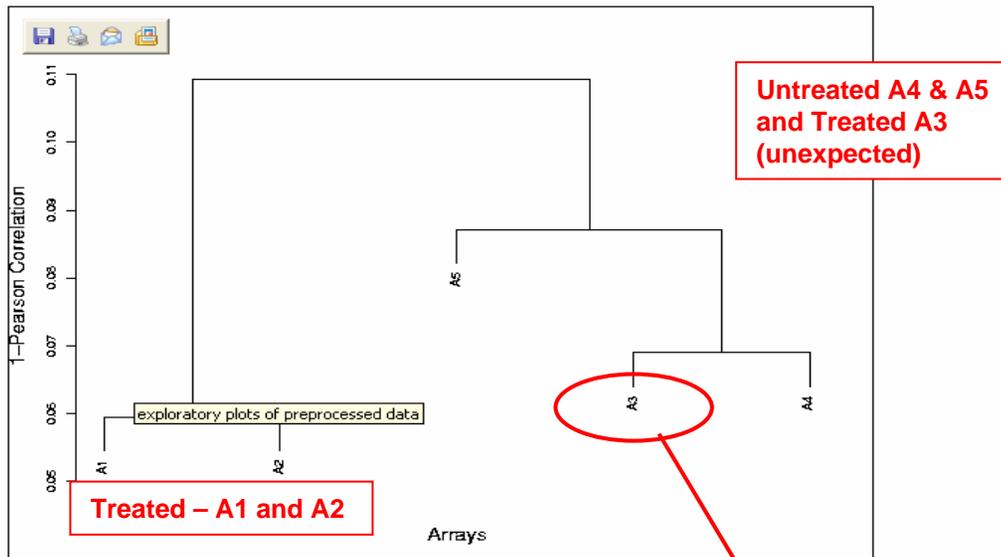# Example: use of CEBS v1.6.1 to explore acetaminophen toxicity and expression response

Acetaminophen study: rats treated with various doses and sacrificed at 6, 24, or 48 hours post dose

- Select 1500 mg/kg dose level, 24 hours post dose

- Affymetrix microrrays; 3 treated, 2 control rats

- Check microarray behavior

- Check clinical chemistry and histopathology

- Explain behavior of the three treated rats

Clustering shows two groups of arrays (treated rat 3020 [A3] clusters with controls)

Clustering of Arrays

This plot displays the relative similarities between arrays (based on Pearson correlation of global expression), and hierarchical clustering of arrays.

Untreated A4 & A5 and Treated A3 (unexpected)

Treated – A1 and A2

exploratory plots of preprocessed data

See Legend for Array Information
Back to Preprocessing | Continue Microarray Analysis

Legend for Array Information

Label	Experiment ID	Array Name	Sample Name
A1	522398544	1500mg_Acetaminophen_24h_Male_Rat_3018_206559593	1500mg_APAP_24hr_3018
A2	522398544	1500mg_Acetaminophen_24h_Male_Rat_3019_206559594	1500mg_APAP_24hr_3019
A3	522398544	1500mg_Acetaminophen_24h_Male_Rat_3020_206559595	1500mg_APAP_24hr_3020
A4	522398544	1500mg_Acetaminophen_24h_Male_Rat_Pool_206771750	1500mg_APAP_24hr_Pool_3012_301
A5	522398544	1500mg_Acetaminophen_48h_Male_Rat_Pool_206559665	1500mg_APAP_48hr_Pool_3000_300

What is the reason for (apparent) misclassification?

- Technical effect in microarray?

- Biological effect in the subject?

- From the tox data for the study, select animals with elevated blood liver enzymes

- plus their biological replicates

- (can include / exclude comparator groups)

Anchor phenotype in biological process

pre-toxic

peak toxicity

recovery

signal intensity

ALT = 250

time
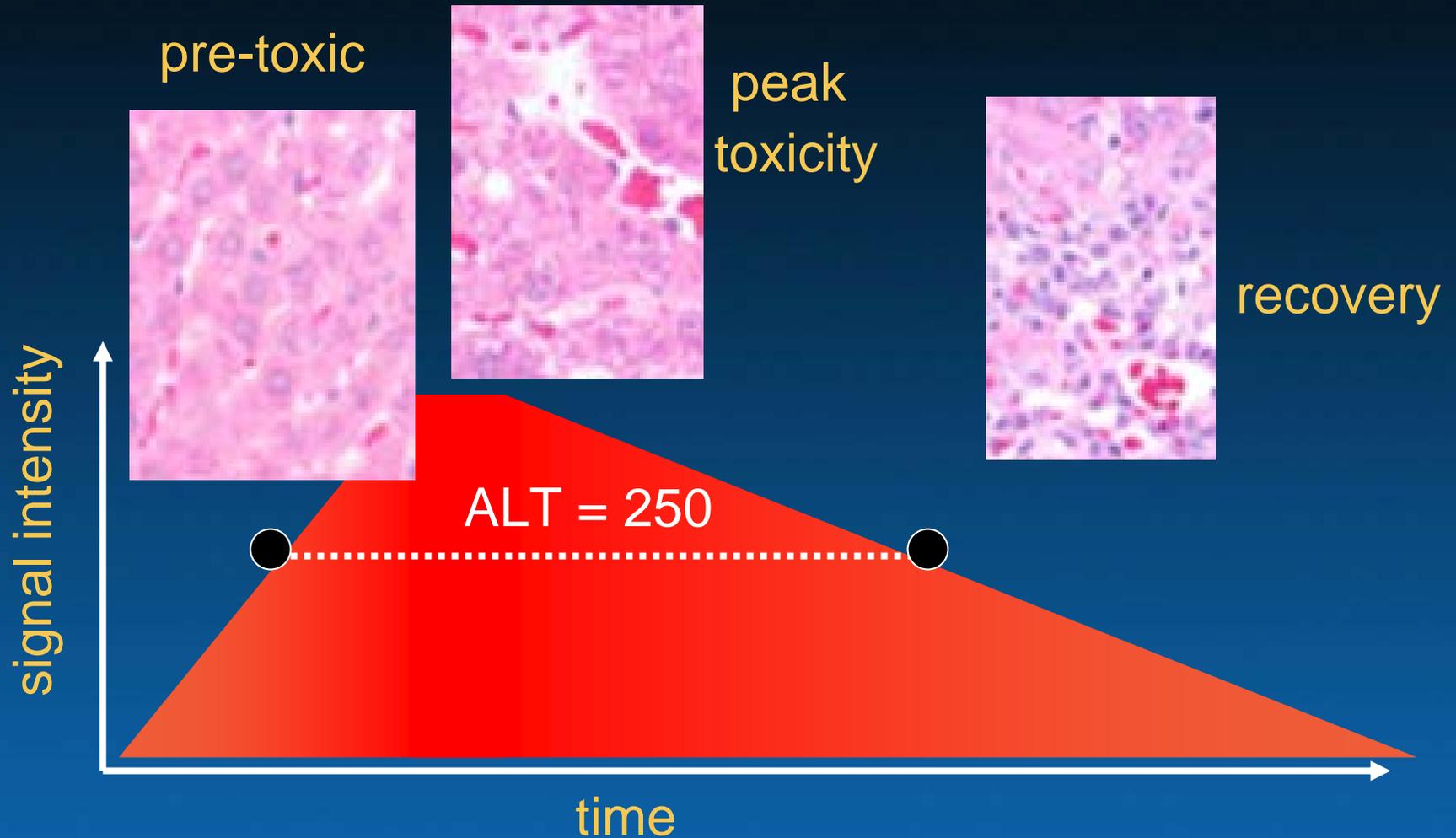
Clinical Chemistry

Clinical Chemistry Results for Selec

The clinical chemistry test results for the animals in selected group(s) are displayed in this page. For v may not have complete sets of test results for clinical chemistry; some of the animals may have no clini only the tests that have test values will be displayed; if multiple groups are selected, the tests displayed

Study Number:	NCT008						
Group Name:		APAP 1500mg/kg 24h					
Intervention:		Acetaminophen; 1500.0 mg/kg; 24.0 Hr					

Animal Id	total protein	bile acids	total bilirubin	ALT	SDH	ASP	AST
	(g/dl)	(µM/l)	(mg/dl)	(units/l)	(units/l)	(units/l)	(units/l)
APAP 1500mg/kg 24h:3018	7.4	254.0	0.6	1600.0	310.0	266.0	2030.0
APAP 1500mg/kg 24h:3019	7.0	294.0	0.3	3690.0	360.0	229.0	5540.0
APAP 1500mg/kg 24h:3020	7.6	40.4	0.7	80.0	7.9	238.0	104.0
APAP 1500mg/kg 24h:3021	7.0	59.3	0.5	8600.0	124.0	227.0	7400.0
APAP 1500mg/kg 24h:3022	7.6	423.0	0.5	6940.0	316.0	285.0	13120.0
APAP 1500mg/kg 24h:3023	7.1	63.4	0.5	233.0	77.0	227.0	377.0

Group Name:		APAP 2000mg/kg 24h					
Intervention:		Acetaminophen; 2000.0 mg/kg; 24.0 Hr					

Pathology for same animals

Animal ID:	APAP 1500mg/kg 24h:3018				
Organ	Lesion Name	Histological Site	Lesion Present	Severity	Infiltration Cell Type
Liver	Degeneration	Centrilobular	Yes	Moderate	
Liver	Glycogen depletion		No		
Liver	Hypertrophy	Centrilobular	No		
Liver	Infiltration	Centrilobular	Yes	Mild	Mononuclear
Liver	Necrosis	Centrilobular	Yes	Mild	
Liver	Regeneration		No		
Liver	Congestion	Hepatic Sinusoid	No		
Animal ID:	**APAP 1500mg/kg 24h:3019**				
Organ	Lesion Name	Histological Site	Lesion Present	Severity	Infiltration Cell Type
Liver	Congestion	Hepatic Sinusoid	No		
Liver	Degeneration	Centrilobular	Yes	Mild	
Liver	Glycogen depletion		No		
Liver	Hypertrophy	Centrilobular	No		
Liver	Infiltration	Centrilobular	Yes	Mild	Mononuclear
Liver	Necrosis	Centrilobular	Yes	Moderate	
Liver	Regeneration		No		
Animal ID:	**APAP 1500mg/kg 24h:3020**				
Organ	Lesion Name	Histological Site	Lesion Present	Severity	Infiltration Cell Type
Liver	Congestion	Hepatic Sinusoid	No		
Liver	Degeneration	Centrilobular	Yes	Minimal	
Liver	Glycogen depletion		No		
Liver	Hypertrophy	Centrilobular	No		
Liver	Infiltration	Centrilobular	Yes	Minimal	Mononuclear
Liver	Necrosis	Centrilobular	No		
Liver	Regeneration		No		

Housing and feed regimen impact the study

Clinical Chemistry Results for Selec

The clinical chemistry test results for the animals in selected group(s) are displayed in this page. For v
may not have complete sets of test results for clinical chemistry; some of the animals may have no clini
only the tests that have test values will be displayed; if multiple groups are selected, the tests displaye

Study Number:	NCT008						
Group Name:		APAP 1500mg/kg 24h					
Intervention:		Acetaminophen; 1500.0 mg/kg; 24.0 Hr					

Animal Id	total protein	bile acids	total bilirubin	ALT	SDH	ASP	AST
	(g/dl)	(µM/l)	(mg/dl)	(units/l)	(units/l)	(units/l)	(units/l)
APAP 1500mg/kg 24h:3018	7.4	254.0	0.6	1600.0	310.0	266.0	2030.0
APAP 1500mg/kg 24h:3019	7.0	294.0	0.3	3690.0	360.0	229.0	5540.0
APAP 1500mg/kg 24h:3020	7.6	40.4	0.7	80.0	7.9	238.0	104.0
APAP 1500mg/kg 24h:3021	7.0	59.3	0.5	8600.0	124.0	227.0	7400.0
APAP 1500mg/kg 24h:3022	7.6	423.0	0.5	6940.0	316.0	285.0	13120.0
APAP 1500mg/kg 24h:3023	7.1	63.4	0.5	233.0	77.0	227.0	377.0

Group Name:		APAP 2000mg/kg 24h
Intervention:		Acetaminophen; 2000.0 mg/kg; 24.0 Hr

Prototyping future ARC capabilities

- Comparing a "hidden" compound with public data in CEBS using phenotype to match the two compounds

- Integrating microarray data and proteomics data from two Studies of the same compound, with very similar experimental designs
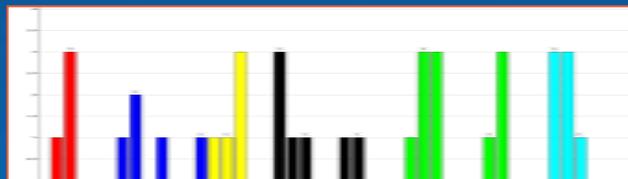
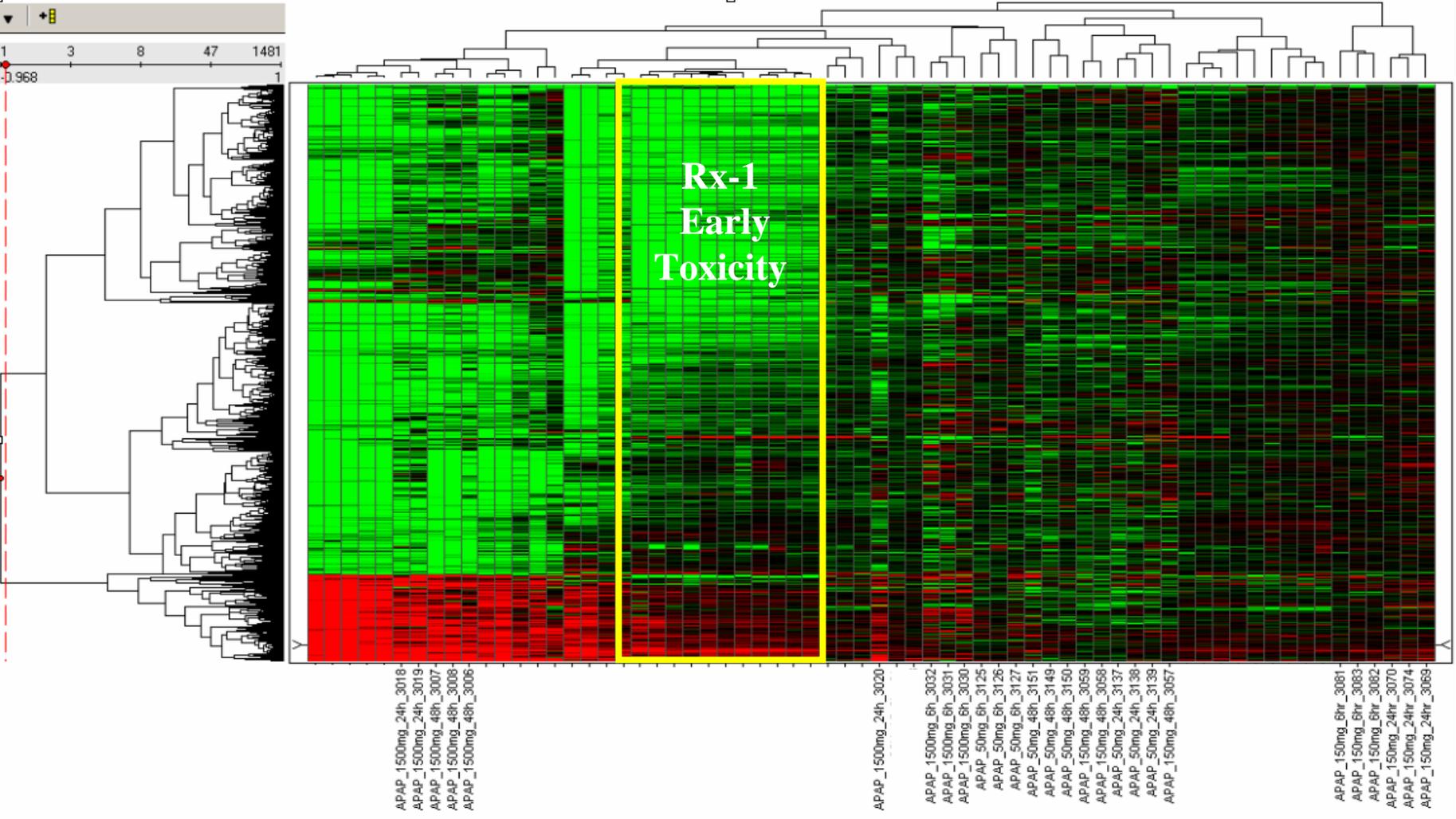# Experimental design: APAP and Compound X
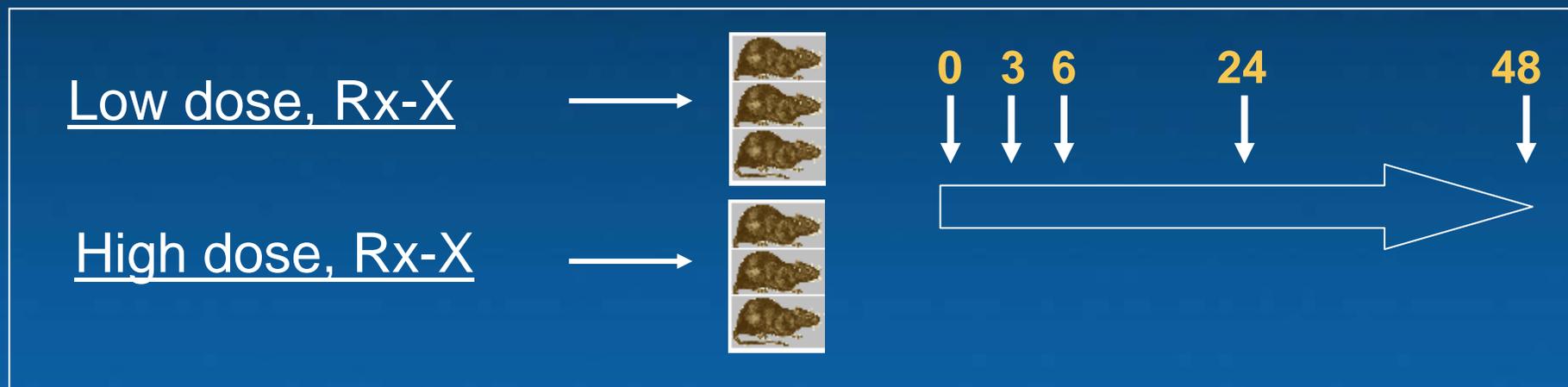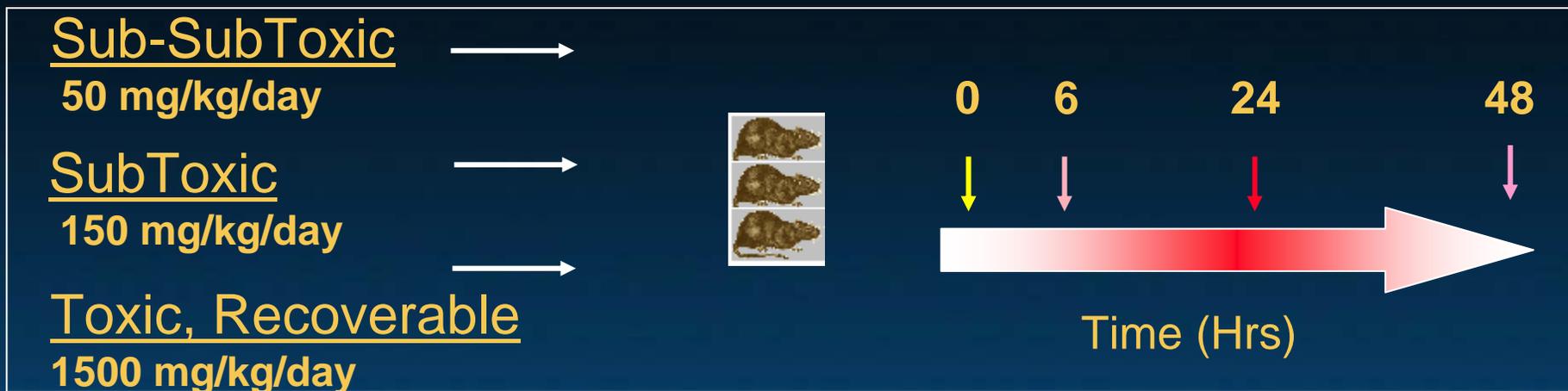
UNSUPERVISED CLUSTERING

APAP Toxicity
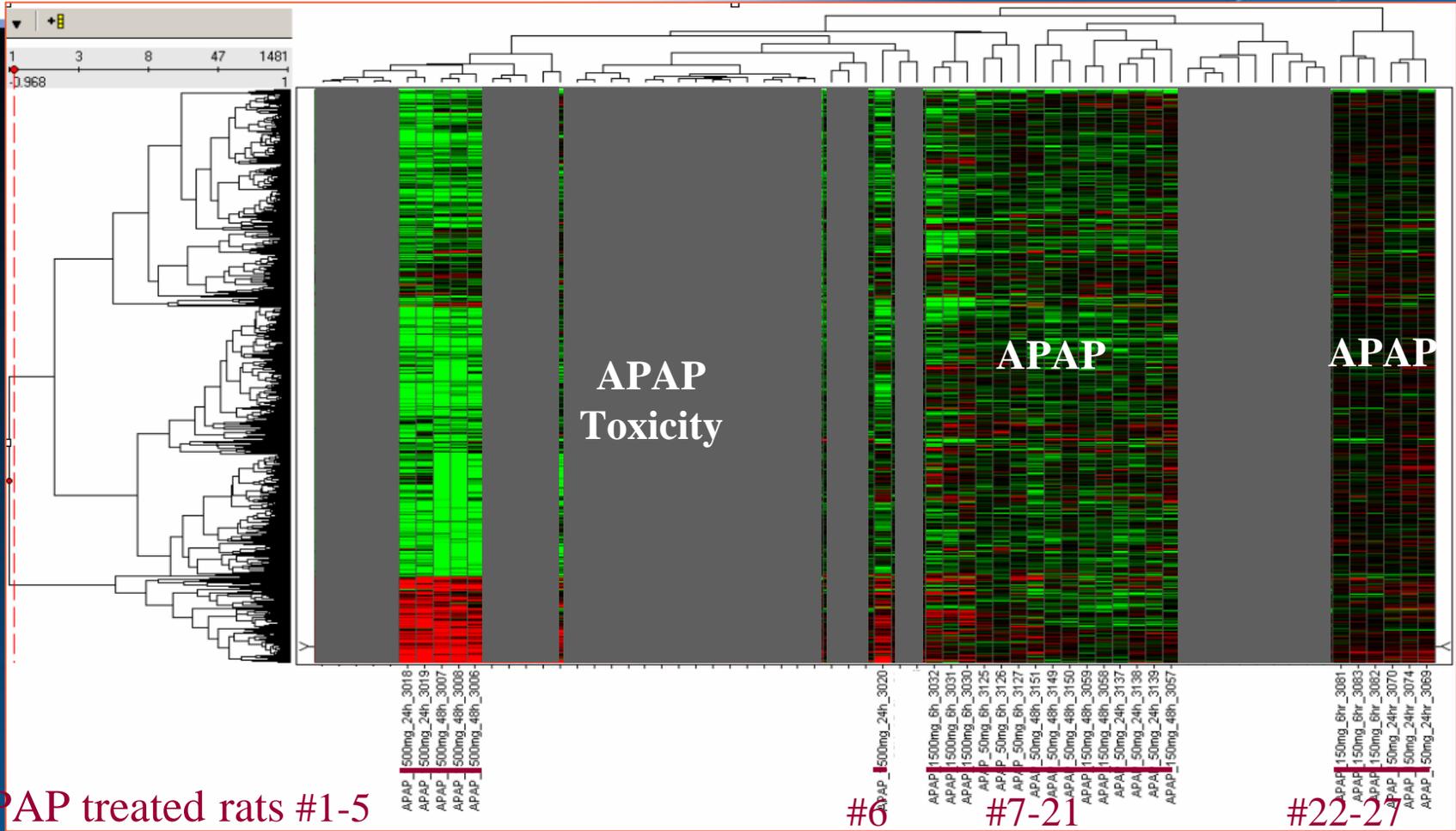
APAP

APAP

APAP treated rats #1-5   #6   #7-21   #22-27

Toxicity Phenotypes

No Histo- or Clinical Pathology Observed
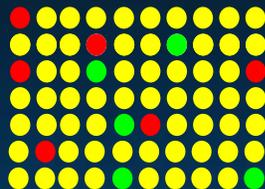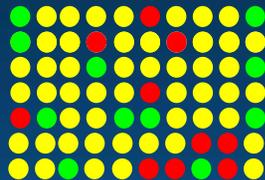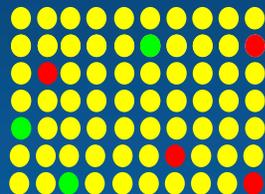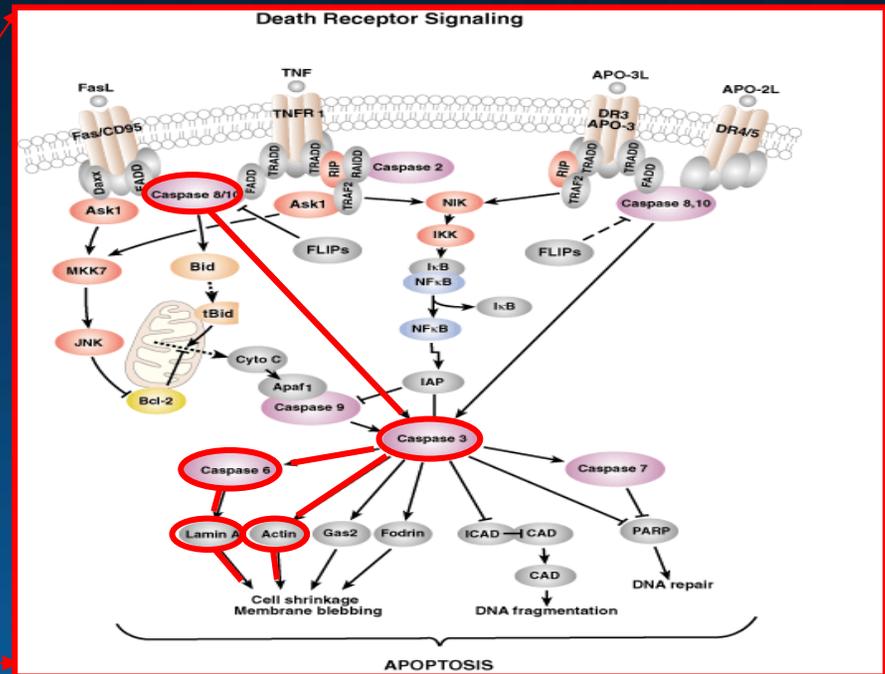
Integrating across Studies

(Low) SubToxic
150 mg/kg/day
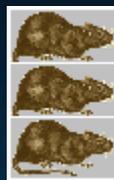
(High) Toxic, Recoverable
1500 mg/kg/day

0   6   24   48

Time (Hrs)
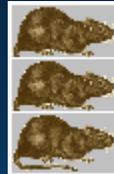
- APAP microarray Study
  - 30XX animal IDs
- APAP proteomics Study
  - animals 1 - 65

- Differences:
  - individual animals
  - vehicle
  - feed regimen
  - housing
  - specimen preparation

Data

- Microarray:

  - Agilent rat genome arrays

  - use manufacturer's normalization

  - microarray data from liver (target organ)

- Proteomics:

  - use intensity data following alignment with Master Spot List

  - convert to ratios to match microarray data

  - data from liver and serum handled separately

Align by dose and time
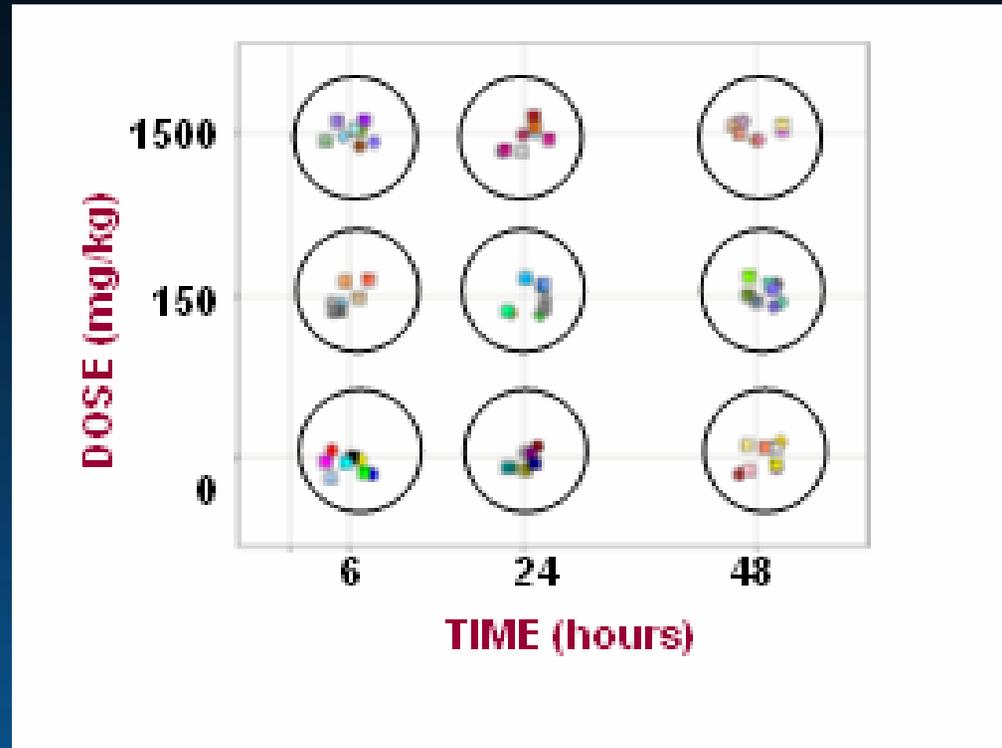
|	DT - X - 6	DT- X - 24	DT – X - 48
DT – 0 – X	3075-3077, 3024-3026, 1 – 5	3064, 3065, 3067, 3012-3014, 26-30	3051-3053, 3000-3002, 51-55
DT – 150 – X	3081 – 3083, 6 - 10	3069, 3070, 3074, 31-35	3057-3059, 56-60
DT – 1500 – X	3030-3032, 11-15	3057 – 59, 56 - 60	3006-3008, 61-65

Align by toxicology (phenotype)
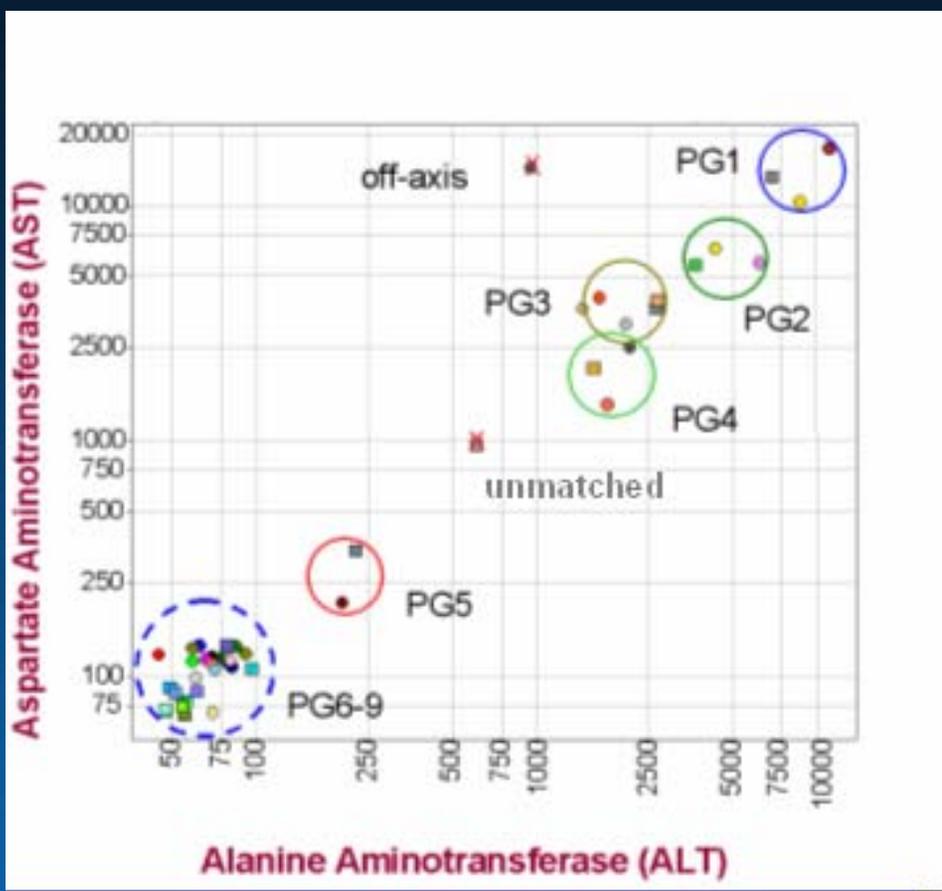
PG-group	Rats	ALT (U/ml)	AST (U/ml)
PG-1	38, 39, 3022	8945.3	13626.7
PG-2	37, 40, 3019	4758.7	4075.3
PG-3	61, 62, 63, 3006, 3007	2124.8	3688.4
PG-4	64, 65, 3018	1851.7	1986.7
PG-5	14, 3074	213	273.9
PG-6	6-10, 13, 15, 31, 3069	70	125.7
PG-7	12, 32-35, 57, 58, 60, 3031	80.2	113.5
PG-8	59, 3030, 3057, 3083	55.4	89.5
PG-9	56, 3032, 3058, 3059, 3070, 3081, 3081	54.6	72.6

Data following integration

- six "dose-time" groups of individual animals

- nine "phenotypially anchored" groups of animals

- Ensure each group contains animals with proteomics data and animals with microarray data

- merge microarray and proteomics data from animals within each group to create one "virtual" rat per group
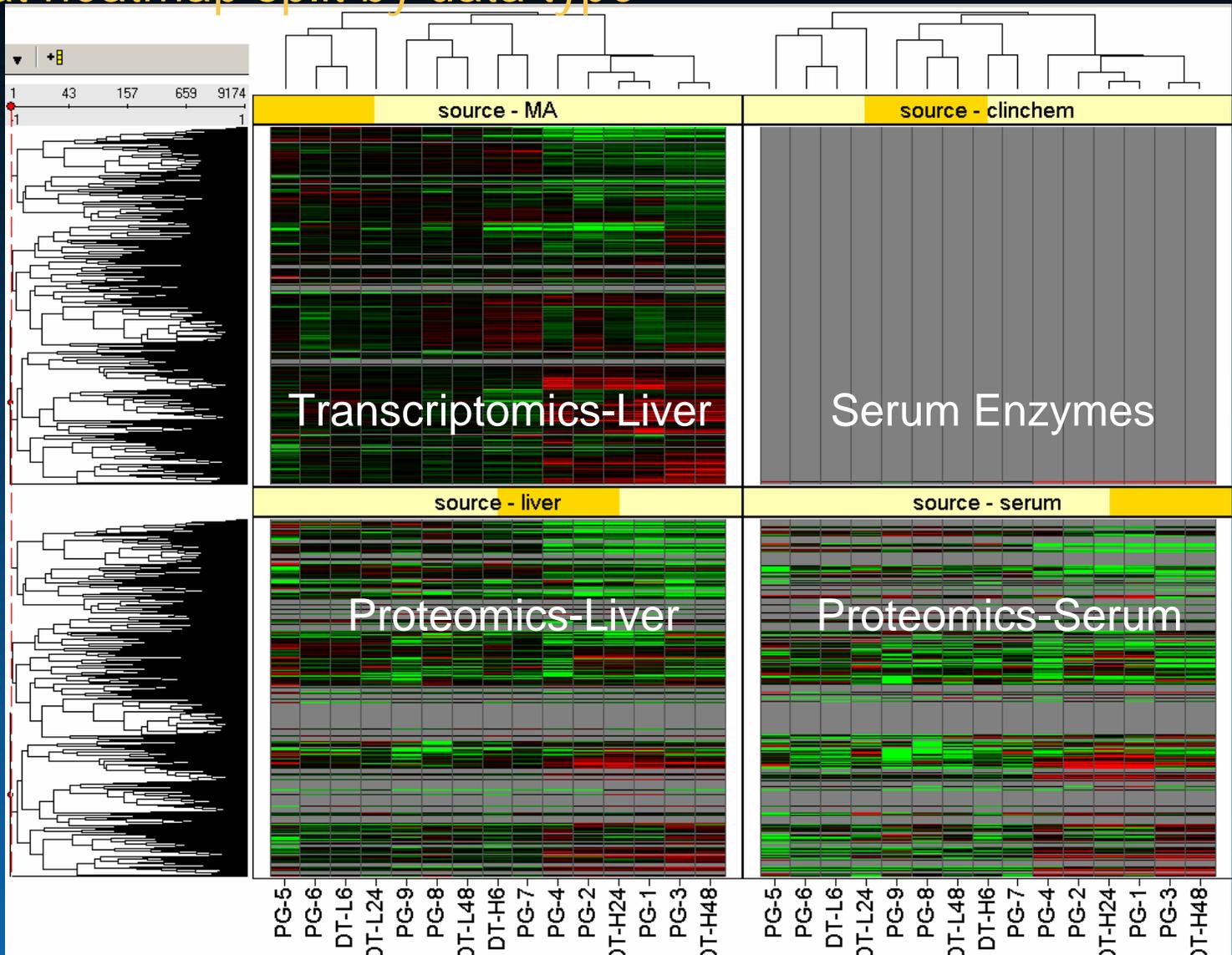
- cluster data from the 15 virtual rats

Clustered data from virtual rats

All Data Types

Log ratios
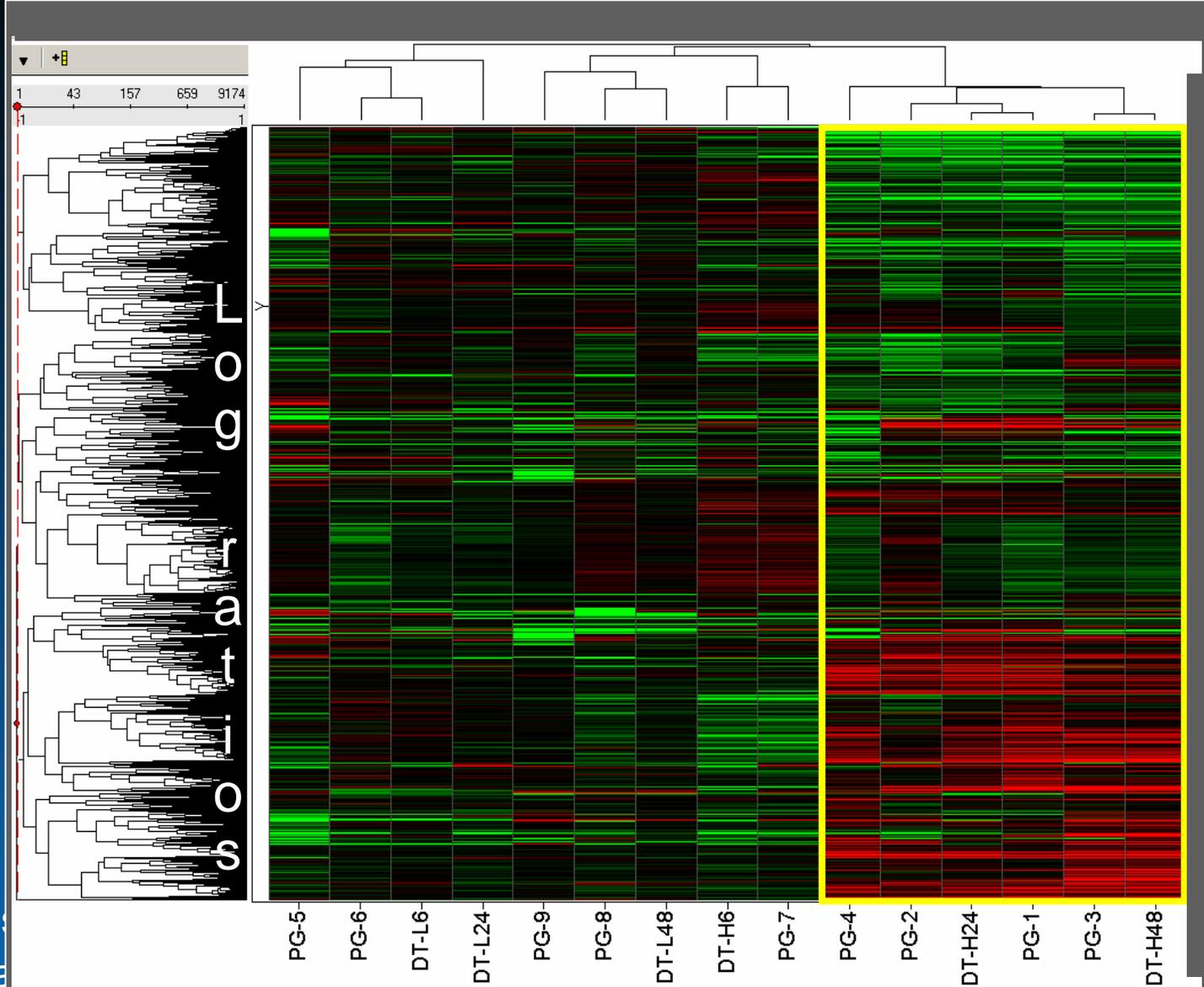
half-log increase
half-log decrease

Virtual rat heatmap split by data type

Summary

- CEBS http://cebs.niehs.nih.gov/

- ARC https://dir-apps.niehs.nih.gov/arc/

- CEBS Developmental Forum http://www.niehs.nih.gov/cebs-df/index.cfm

- Study design, phenotypic anchoring

- Integrated database permits

  - selection of 'omics data by Study characteristics

  - selection by 'omics data by Subject characteristics

- Case studies

  - Identify and explain outlier Subjects (actually done in CEBS)

  - Merge "secret" data with well-characterized public data

  - Integrate microarray and proteomics data

The NCT Retreat - 2004

The Toxicogenomics Research Consortium (TRC)
Science Applications International Corporation (SAIC)
Icoria (formerly Paradigm Genetics)
Lockheed Martin Information Technology (LMIT)
Alpha-Gamma Technologies, Inc. (AGTI)