

OpenTox USA 2013

# CEBS DATABASE: A STRUCTURE FOR DATA INTEGRATION ACROSS STUDIES

Asif Rashid

Contract No. HHSN273201000063U



# Who am I?

- Federal Contractor (Lucky one who could work during Government shutdown)
- Employee of Vistronix, Inc.
- Program Manager - NTP Computer and User Support Contract
- CEBS is supported by this contract beside other products in NTP
- Husband and Father of three beautiful sons

# Chemical Effects in Biological Systems (CEBS)

- ◎ Relational public database
  - User interface
  - FTP site with data domain “bundles”
- ◎ Private database + file system (Bins)
- ◎ CEBS: <http://cebs.niehs.nih.gov>

# Agenda

- Three examples of CEBS usage
- Meta data and data dictionary
- CEBS data in RDF

# Example 1

- See all studies in CEBS related to a test article (Toluene)
- How studies are organized in CEBS
- How to access study-level data



## Research

Resources for Scientists

Databases

Alu Pairs Database

Biomarkers of Oxidative Stress Study

**Chemical Effects in Biological Systems (CEBS)**

Citing CEBS

Terms of Use

Genetic Alterations in Cancer (GAC)

Human DNA Polymerase Gamma Mutation Database

Microarray Group cDNA Clone Search

SNPinfo Web Server

Spin Trap Database

All Scientists

All Research Groups

## Chemical Effects in Biological Systems (CEBS)

The CEBS database houses data of interest to environmental health scientists. CEBS is a public resource, and has received depositions of data from academic, industrial and governmental laboratories. CEBS is designed to display data in the context of biology and study design, and to permit data integration across studies for novel meta analysis.

**Note:** Users who are using JAWS as accessibility tool are advised to install scripts for using Adobe Flex applications with JAWS before using this application. These scripts have been developed by Adobe for Flex application. You can get the scripts from [here](#) (Get JAWS scripts for Flex). Once you have installed these scripts, press the insert key and q together to start the form mode in JAWS. There are number of sound hints available for alerts for opening of dialog boxes and data download. Use Tab key to navigate.

Open CEBS

Contact CEBS

Download Data



A<sup>+</sup> A<sup>-</sup> Share

### RELATED LINKS

>> [NIEHS Software and Online Tools Updates](#)

## Download Instructions

This FTP site allows you to download study data in CEBS to your computer. You can access it through your browser by clicking the [Download](#) link or entering the URL <ftp://157.98.192.110/ntp-cebs/datatype/>

IF YOU ARE A MAC USER, use Firefox to download data from the FTP site.

This FTP site does not work well with most versions of Safari. Most likely, a Safari user will be able to see the folders or file icons but be unable to open or download the files.

DO NOT try to enter a username or password at the FTP Download login screen.

IT DOES NOT REQUIRE A USERNAME OR PASSWORD.

If your browser shows a login screen on this site asking for a username and password, do not type anything in those boxes.

- If you are using Safari (not recommended) as your browser, click the radio button next to guest and then click connect.
- If you are using other browsers: Close, then Restart your browser, and Clear the Browser Cache before clicking the download link.

If you still see the login screen, please contact [fostel@niehs.nih.gov](mailto:fostel@niehs.nih.gov).

## Contact

Jennifer Fostel, Ph.D.

# CEBS Home Page

A+ A- Share

CEBS Home

Open My Data

Chemical Effects in Biological Systems

## Search Study

Search studies using test article, CEBS accession number, NTP Study ID, NTP Study Type, CEBS data domain or institution.

Test Article

NTP Study ID

CEBS accession number

NTP Study Type

CEBS data domain

Institution

## Search Across Studies

Search across the database for subjects meeting your criteria, filtering by:

- **Characteristics** (of studies, participants, stressor, or protocols)
- **Response** (of individual assays or pathology diagnosis)
- **Test Articles** (Chemicals or non-chemical stressors)

Combining search results and then viewing or downloading the results.

Go

## Data Status (Last updated on May 2012)

Data Source	Total	Searchable Pathology
GeneTox	6306	no pathology
Non-NTP studies	96	25 with pathology
NTP Bioassay - from CBDS	1315	1315
NTP Bioassay - from TDMS	1312	441
NTP ImmunoTox	89	no pathology
NTP Special studies	470	no pathology
Tox21 phase 1	15	no pathology

## Workflows

Tox21 (Phase I)

Explore Conclusion

Site Association

## Help

[Identify and download the list of all studies producing a positive result in the Ames test](#)

[Add selected search result data to the Workspace](#)

[Download additional depositor data](#)

[Accessing Tox21 Phase 2 workflow](#)

[Depositing Data to CEBS](#)

[Tox21 Phase I: Filtering assay results by activity call](#)

[Tox21 Phase I: Find all data for an assay](#)

# Select Test Article

CEBS Home

Open My Data

## Search Study

Search studies using test article, CEBS accession number, NTP Study ID, NTP Study Type, CEBS data domain or institution.

### Test Article

A. Enter part of test article or cas number

Submit

B. Select from list of matches and click on search button

CAS Number	Test Article Name	Matched Attribute	Attribute Value	# Study(s)
602-01-7	2,3-Dinitrotoluene	NTP_USE	Intermediate in the manufacture of toluene diisocyanate; plasticizer in propellants	16
584-84-9	2,4-Toluene diisocyanate	STRESSOR_NAME	Toluene-2,4-diisocyanate	19
108-88-3	Toluene	STRESSOR_NAME	Toluene	24
6369-59-1	2,5-Toluenediamine sulfate	STRESSOR_NAME	2,5-Toluenediamine sulfate	8
TOLUENES	Toluenes (2,4-diaminotoluene, p-nitrotoluene, 4-chloro-o-phenylenediamine	CAS_NUMBER	TOLUENES	0
25013-15-4	Vinyl toluene	STRESSOR_NAME	Vinyl toluene	23

Search

Cancel




# Investigations using that Test Article

CEBS Home   My Workspace   Open My Data

All Data

Show Details   Add to Workspace

Investigation/Study	Accession Number	
▶ HTS Cell Viability Studies of NTP1408 Compound Library	013-00001-0000-00...	Natio
▼ NTP Investigation of Toluene	002-02916-0000-00...	Natio
📄 Cytogenetic study of Toluene in Chinese Hamster Ovary Cell Chromosome Aberrations Test	002-02916-0002-00...	
📄 Cytogenetic study of Toluene in Chinese Hamster Ovary Cell Sister Chromatid Exchange Test	002-02916-0001-00...	
📄 Effects of Toluene in Mouse Lymphoma Study 434629	002-02916-0009-00...	
📄 Genetic Toxicity Evaluation of Toluene (technical)(108-88-3) in Micronucleus Study 044232 on B6C3F1 Mice	002-02916-0003-00...	
📄 Genetic Toxicity Evaluation of Toluene in Salmonella/E.coli Mutagenicity Test or Ames Test Study 132615	002-02916-0004-00...	
📄 Toxicity Evaluation of Toluene (108-88-3) on B6C3F1 Mouse	002-02916-0006-00...	
📄 Toxicity Evaluation of Toluene (108-88-3) on F 344/N Rat	002-02916-0005-00...	



# 2-year rat study

Toxicity Evaluation of Toluene (108-88-3) on F 344/N Rat

Characteristics

Design

Timeline

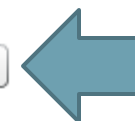
Data

**Accession Number:** 002-02916-0005-0000-5

**Study Conclusion:**

**Stressor Name:** No stressor

More



Press More button to see other study stressors

**Study Variables:** DOSE|DOSE\_UNIT|TREATMENT\_GROUP\_TYPE

Characteristic	Characteristic Value
INSTITUTION	NTP
LABORATORY	IRDC
NTP_CHEMTRACK_NUM	C07272B
NTP_STUDY_TYPE	General Toxicology - Long Term
NTP_TDMS_NUMBER	05701-02
NTP_TOXICOLOGY_TYF	Long-Term
START_DATE	09/27/82

**Publications:**

Title	PubMed

# Design tab – explains the groups and the comparators

Toxicity Evaluation of Toluene (108-88-3) on F 344/N Rat

Characteristics

Design

Timeline

Data

TRIAL	TREATMENT_GROUP_TYPE	DOSE_UNIT	DOSE
▶ <a href="#">GROUP5</a>	Test Chemical	PPM	600
▶ <a href="#">GROUP3</a>	Test Chemical	PPM	1200
▶ <a href="#">GROUP6</a>	Test Chemical	PPM	1200
▶ <a href="#">GROUP1</a>	Vehicle Control	PPM	0
▶ <a href="#">GROUP2</a>	Test Chemical	PPM	600
▶ <a href="#">GROUP4</a>	Vehicle Control	PPM	0

# Timeline Tab – gives the study events and protocols

Toxicity Evaluation of Toluene (108-88-3) on F 344/N Rat

Characteristics Design **Timeline** Data

Protocol   day:	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
<a href="#">STRESSOR PROTOCOL</a>				✓																			
<a href="#">PREPARATION</a>																							
<a href="#">OBSERVATION</a>	✓					✓		✓	✓				✓		✓	✓					✓	✓	✓
<a href="#">DISPOSITION</a>																							
<a href="#">CARE</a>				✓																			

Toxicity Evaluation of Toluene (108-88-3) on F 344/N Rat

Characteristics Design **Timeline** Data

705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720	721	722	723	724	725	726	727	728	729	730	731	732	733
	✓		✓		✓	✓		✓		✓		✓		✓			✓	✓			✓			✓	✓	✓	✓	✓
	✓		✓	✓	✓	✓	✓	✓		✓		✓		✓			✓	✓			✓			✓	✓	✓	✓	✓
																								✓	✓		✓	✓

# Timeline Tab – gives the study events and protocols

**Toxicity Evaluation of Toluene (108-88-3) on F 344/N Rat**

Characteristics Design **Timeline** Data

Protocol	day:	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
<a href="#">STRESSOR PROTOCOL</a>					✓																		
<a href="#">PREPARATION</a>																							
<a href="#">OBSERVATION</a>		✓					✓		✓	✓				✓		✓	✓				✓	✓	✓
<a href="#">DISPOSITION</a>																							
<a href="#">CARE</a>					✓																		

Protocol Type/Name/Attributes	Value
▼ CHEM_STRESSOR_PROTOCOL	
▶ Toluene-0	
▼ Toluene-1200	
VEHICLE NAME	AIR
IS VEHICLE CONTROL	F
ROUTE ADMIN	RESPIRATORY EXPOSURE WHOLE BODY
DOSE PER ADMIN UNIT	PPM
DOSE PER ADMIN	1200
STRESSOR NAME	Toluene
▼ Toluene-600	
DOSE PER ADMIN UNIT	PPM

# Study data by “domains”

Chemical Effects in Biological Systems (CEBS)

## Study Information

CEBS Accession Number:	002-02916-0005-0000-5
Chemical Name:	Toluene
CASRN:	108-88-3
NTP Study Type:	General Toxicology - Long Term
NTP Study ID:	C07272B

## View Study Details

Animal Number	Dose (PPM)	Species	Strain	Sex	Removal Reason
GROUP1_CAGE2_M_16561	0	Rat	F 344/N	Male	Terminal Sacrifice

## Organ and Accountable Site Status

Organ	Status
	MISSING
	Organ
Adrenal gland	
	NORMAL
	Organ
Blood vessel	
Bone	
Bone marrow	
Brain	
Epididymis	
Esophagus	
Intestine, large	
Intestine, small	
Islets	
Mammary gland	
Parathyroid gland	
Salivary gland	
Skin	
Stomach	
Thymus	
Thyroid	
Trachea	
Urinary bladder	

## Observations

Organ	Organ Site	Observation	Qualifiers/Severity
Adrenal gland	Cortex	Hyperplasia	Mild
Heart		Cardiomyopathy	Mild Chronic
Kidney	Renal Tubule	Pigmentation	Mild
Liver	Bile Duct	Hyperplasia	Mild
Lung	Arteriole	Mineralization	Mild
Lymph node	Mandibular	Hyperplasia	Mild Lymphoid
Nose	Olfactory Epi	Degeneration	Mild
Pancreas	Acinus	Atrophy	Minimal
Pituitary gland	Pars Distalis	Hyperplasia	Mild
Preputial gland		Inflammation	Mild Chronic
Prostate		Inflammation	Mild Chronic Active
Spleen		Pigmentation	Mild
Stomach	Forestomach	Hyperplasia	Mild Squamous
Testes		Atrophy	Mild

# Study data by “domains”

## Toxicity Evaluation of Toluene (108-88-3) on F 344/N Rat

### Toxicity Evaluation of Toluene (108-88-3) on F 344/N Rat

Characteristics Design Timeline **Data**

#### HISTOPATHOLOGY

Group	Subject	Collection Time (day)	COMMENTS	ORGAN	PRIMARY ORGAN	SUB TOPOGRAPHY	ORIGINAL ...	SEVERITY
GROUP1_CAGE2_M	GROUP1_CAGE2_M_16561	730		Adrenal gla		Cortex	Hyperplasia	Mild
GROUP1_CAGE2_M	GROUP1_CAGE2_M_16561	730		Testes		Interstit Ce	Hyperplasia	Mild
GROUP1_CAGE2_M	GROUP1_CAGE2_M_16561	730	Chronic	Kidney			Nephropath	Moderate
GROUP1_CAGE2_M	GROUP1_CAGE2_M_16561	730		Kidney		Renal Tubu	Pigmentatio	Mild
GROUP1_CAGE2_M	GROUP1_CAGE2_M_16561	730		Liver			Focal Cellul	
GROUP1_CAGE2_M	GROUP1_CAGE2_M_16561	730		Liver		Bile Duct	Hyperplasia	Mild
GROUP1_CAGE2_M	GROUP1_CAGE2_M_16561	730	Chronic	Liver		Portal	Inflammatory	Mild
GROUP1_CAGE2_M	GROUP1_CAGE2_M_16561	730		Lung			Congestion	Mild
GROUP1_CAGE2_M	GROUP1_CAGE2_M_16561	730	Lymphocyte	Lung			Infiltration	Mild
GROUP1_CAGE2_M	GROUP1_CAGE2_M_16561	730	Chronic Act	Lung		Interstitium	Inflammatory	Minimal
GROUP1_CAGE2_M	GROUP1_CAGE2_M_16561	730		Lung		Arteriole	Mineralizati	Mild
GROUP1_CAGE2_M	GROUP1_CAGE2_M_16561	730	Lymphoid	Lymph nodi		Mandibular	Hyperplasia	Mild
GROUP1_CAGE2_M	GROUP1_CAGE2_M_16561	730		Nose		Olfactory Ep	Degenerati	Mild
GROUP1_CAGE2_M	GROUP1_CAGE2_M_16561	730		Nose		Respirat Ep	Degenerati	Mild
GROUP1_CAGE2_M	GROUP1_CAGE2_M_16561	730		Nose		Lumen	Hemorrhag	Moderate

# Example 2

- ◎ Find data for male rats with liver injury
- ◎ Example of searching across studies using CEBS
  - Search using pathology results, also get all data from these animals
- ◎ Example of limitations of data in CEBS
  - Legacy NTP terminology



# Searching across studies

## Search Study

Search studies using test article, CEBS accession number, NTP Study ID, NTP Study Type, CEBS data domain or institution.

Test Article

NTP Study ID

CEBS accession number

NTP Study Type

CEBS data domain

Institution

## Search Across Studies

Search across the database for subjects meeting your criteria, filtering by:

- **Characteristics** (of studies, participants, stressor, or protocols)
- **Response** (of individual assays or pathology diagnosis)
- **Test Articles** (Chemicals or non-chemical stressors)

Combining search results and then viewing or downloading the results.

  
Go

## Data Status (Last updated on May 2012)

Data Source	Total	Searchable Pathology
GeneTox	6306	no pathology
Non-NTP studies	96	25 with pathology
NTP Bioassay - from CBDS	1315	1315
NTP Bioassay - from TDMS	1312	441
NTP ImmunoTox	89	no pathology
NTP Special studies	470	no pathology
Tox21 phase 1	15	no pathology

## Workflows

Tox21 (Phase I)

Explore Conclusion

Site Association

## Help

[Identify and download the list of all studies producing a positive result in the Ames test](#)[Add selected search result data to the Workspace](#)[Download additional depositor data](#)[Accessing Tox21 Phase 2 workflow](#)[Depositing Data to CEBS](#)[Tox21 Phase I: Filtering assay results by activity call](#)[Tox21 Phase I: Find all data for an assay](#)

## Characteristics

## Response

## Test Articles

## A. Select the attribute

Attributes	Description
▶  STUDY	
▼  PARTICIPANT	
AGE_UNIT	The unit of test subjects' age.
DOSE_DURATION_UNIT	Unit for duration of dosing for given participant
DOSE_UNIT	Unit for dose given to the participant
GENUS	A taxonomic classification of living organisms that is superior to a species. (source: C. Zwickl, Lilly)
ORGAN_NAME	Name of organ
SEX	The designation of gender of the subject
SPECIES	A taxonomic classification of living organisms that is inferior to genus and superior to a subspecies, strain or variety.
SPECIES_COMMON_NAME	Common name for the species of lab animal or microbe used in the study
STRAIN	A taxonomic classification of living organisms within a particular species, characterized by some particular quality.
▶  STRESSOR	
▶  PROTOCOL	

## B. Select value from pull-down list

Rattus



Search

## Searches

Search SEX=Male

Search GENUS=Rattus

Show Studies

Clear Search(s)

Add to Workspace

Advanced Options

[Characteristics](#)[Response](#)[Test Articles](#)**A. Select the assay domain****B. Select the diagnosis****C. Select the organ****D. Select the organ part****E. Select the sub topography****F. Select the cells****G. Select the distribution****H. Select the severity****Searches**

Search SEX=Male

Search GENUS=Rattus

Search RESPONSE where HISTOPATHOLOGY obser





# The CEBS Workspace

CEBS Home My Workspace Open My Data

My Workspace

Source	Study title or Search name
From Search	male rats with liver toxicity

Help

	# Studies	# Groups	# Subje...
	14	198	911



Show Details

 Build Data Matrix

 Microarray Analysis

 Visual Data Mining


Test Articles List

Study Information

# Details for studies in selected list

Study Information for male rats with liver toxicity

Accession #	Study Title	Conclusion	Start Date	Species Co	Sex	Strain	Study Factor	CEBS URL
008-00001-	A 1-, 3- & 7-		1/1/00		Male	Sprague-Da	Dose Time	<a href="http://tools">http://tools</a>
008-00001-	Clofibrate:		1-23-01		Male	Sprague Da	Dose Time	<a href="http://tools">http://tools</a>
008-00001-	Methapyrile		1-23-01		Male	Sprague-Da	Dose Time	<a href="http://tools">http://tools</a>
008-00001-	Clofibrate: (		1/1/00		Male	Sprague Da	Dose Time	<a href="http://tools">http://tools</a>
001-00001-	Application			Rat	Male	F/344	DOSE TIME	<a href="http://tools">http://tools</a>
001-00001-	Application			Rat	Male	F/344	DOSE TIME	<a href="http://tools">http://tools</a>
001-00001-	Application			Rat	Male	F/344	DOSE TIME	<a href="http://tools">http://tools</a>
001-00001-	Application			Rat	Male	F/344	DOSE TIME	<a href="http://tools">http://tools</a>
001-00001-	Application			Rat	Male	F/344	DOSE TIME	<a href="http://tools">http://tools</a>
001-00001-	Application			Rat	Male	F/344	DOSE TIME	<a href="http://tools">http://tools</a>
001-00001-	Application			Rat	Male	F/344	DOSE TIME	<a href="http://tools">http://tools</a>
001-00001-	Application			Rat	Male	F/344	DOSE TIME	<a href="http://tools">http://tools</a>

 Download  Close

# The CEBS Workspace

The screenshot shows the top navigation bar with three buttons: "CEBS Home", "My Workspace" (highlighted), and "Open My Data". Below this is a section titled "My Workspace" containing a table with two columns: "Source" and "Study title or Search name".

Source	Study title or Search name
From Search	male rats with liver toxicity

A summary statistics table with a "Help" link in the top right corner. The table has four columns: an unlabeled column, "# Studies", "# Groups", and "# Subje...".

	# Studies	# Groups	# Subje...
	14	198	911



A horizontal row of six buttons: "Show Details", "Build Data Matrix", "Microarray Analysis", "Visual Data Mining", "Test Articles List", and "Study Information".



# Chemicals in selected studies

Test Articles List for male rats with liver toxicity

Test Article	Cas Number	Is Vehicle Control?
1,2-Dichlorobenzene	95-50-1	False
1,4-Dichlorobenzene	106-46-7	False
Acetaminophen	103-90-2	False
Bromobenzene	108-86-1	False
Clofibrate	637-07-0	False
Diquat dibromide	85-00-7	False
Methapyrilene	91-80-5	False
Monocrotaline	315-22-0	False
N-Nitrosomorpholine	59-89-2	False
Thioacetamide	62-55-5	False

Download Close

# The CEBS Workspace

CEBS Home My Workspace Open My Data

My Workspace

Source	Study title or Search name
From Search	male rats with liver toxicity

Help

	# Studies	# Groups	# Subje...
	14	198	911



Show Details

 Build Data Matrix

 Microarray Analysis

 Visual Data Mining

Test Articles List

Study Information



# Quick review of clinical chemistry.....

Visual Data Mining

ALBUMIN

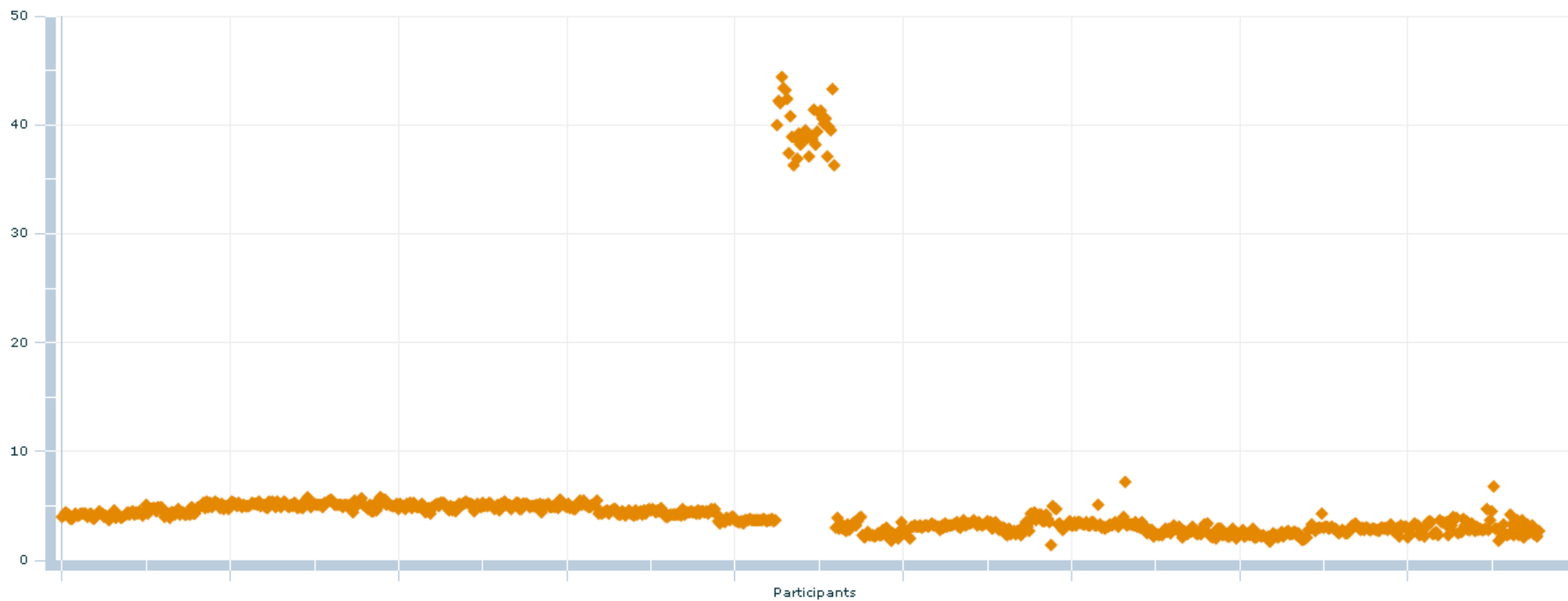


Table Plot values Options

# Example 3

- ⦿ Visual datamining over all data in a domain
- ⦿ This workflow is implemented for Tox21 phase 1 data
  - 1408 compounds
  - 117 assays
- ⦿ The Tox21 phase I data was just made public, and this change has not yet been implemented in CEBS

Interested in Assay

Interested in Chemical

**A. Select the Study Year**

All ▼

**B. Select the species**

All ▼

**C. Select the cell line**

All ▼

**D. Select the assay target**

Cytotoxicity ▼

**E. Select the readout**

Luminescence ▼

**F. Select the analyte**

ATP ▼

**G. Select the assay**

H-4-II-E\_viability  
HUV-EC-C\_viability  
Mesenchymal cell\_viability  
N2a\_viability  
NIH 3T3\_viability  
Proximal tubule cell\_viability  
SH-SY5Y\_viability

**H. Select the conclusion**

NCGC Curve Class ▼

Next ▶

Show selected assay result in context of all HTS data

Select assay

-- Or --

Select chemical

View response  
in context of  
all assays and  
all chemicals

Select the  
response level  
of interest

Combine searches

Download

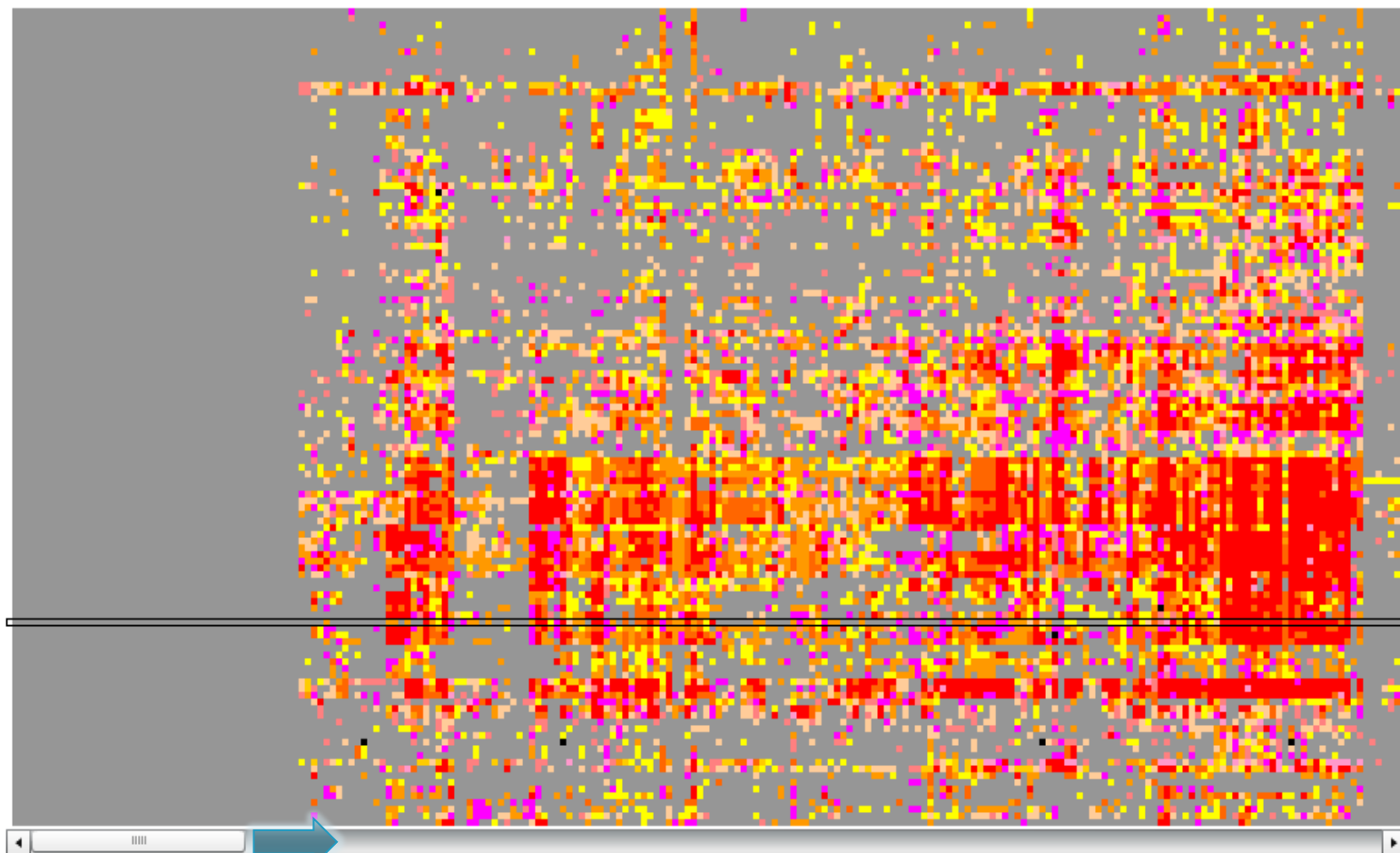
# Responses of all assays to all chemicals; target assay is highlighted

## Legend

NCGC Curve Class

- 1.1 Or 1.1
- 1.2 Or 1.2
- 1.3 Or 1.3
- 1.4 Or 1.4
- 2.1 Or 2.1
- 2.2 Or 2.2
- 2.3 Or 2.3
- 2.4 Or 2.4
- 3 Or 3
- 4
- 5

Assays



Chemicals

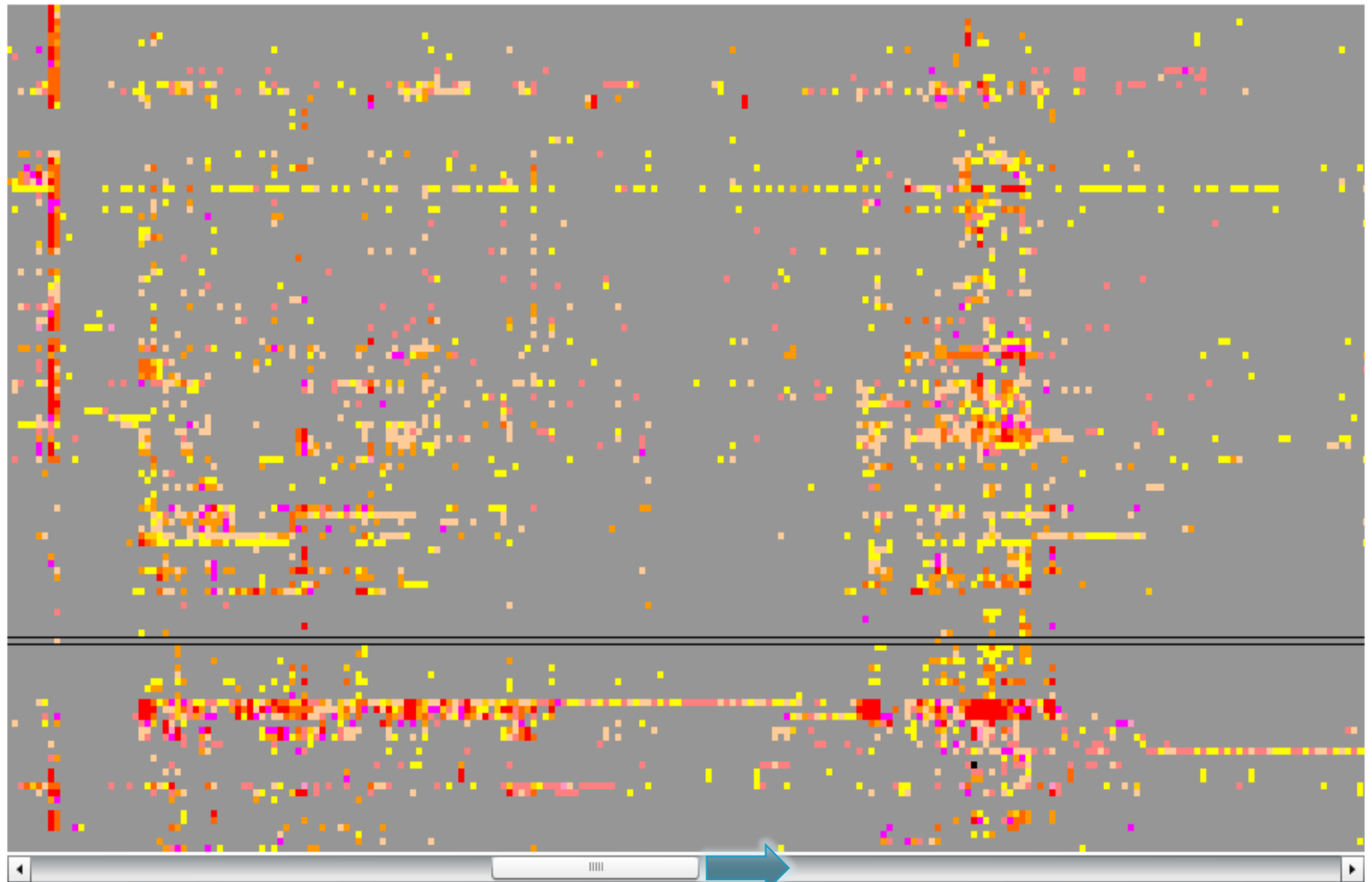
# Clustered bi-directionally to group assays and chemicals on the basis of response

## Legend

NCGC Curve Class

- 1.1 Or 1.1
- 1.2 Or 1.2
- 1.3 Or 1.3
- 1.4 Or 1.4
- 2.1 Or 2.1
- 2.2 Or 2.2
- 2.3 Or 2.3
- 2.4 Or 2.4
- 3 Or 3
- 4
- 5

Assays



Chemicals

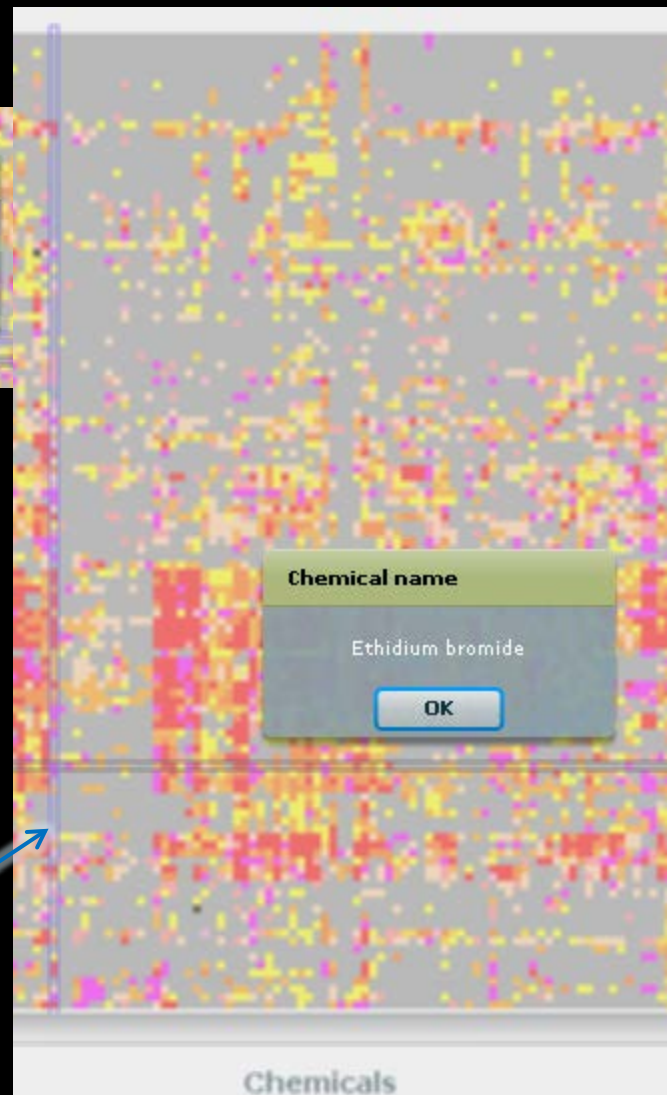


# Additional functionality: look up chemical or assay.



Assay Lookup

Click to get "cursor"  
To highlight assay of interest



Chemical Lookup

# After reviewing responses select response level(s) of interest:

Select NCGC curve class

-1.1 Or 1.1

-1.2 Or 1.2

-1.3 Or 1.3

-1.4 Or 1.4

-2.1 Or 2.1

-2.2 Or 2.2

-2.3 Or 2.3

-2.4 Or 2.4

-3 Or 3

4

5

Select

list of chemicals producing that response:

Assay	Chemical	Single Channel
▼ Assay=Mesenchymal cell_viability, Curve class=		
Mesenchymal cell_viability	Dimenhydrinate	-1.2
Mesenchymal cell_viability	13-cis-Retinal	-1.1
Mesenchymal cell_viability	2,2',4'-Trichloroacetophenone	-1.1
Mesenchymal cell_viability	2-Octyl-3-isothiazolone	-1.1
Mesenchymal cell_viability	3,3'-Dichlorobenzidine dihydrochloride	-1.2
Mesenchymal cell_viability	4'-(Chloroacetyl)acetanilide	-1.1
Mesenchymal cell_viability	4,4'-Thiobis(6-tert-butyl-m-cresol)	-1.1
Mesenchymal cell_viability	4-(Hydroxyphenyl)retinamide	-1.1
Mesenchymal cell_viability	4-Hydroxy-tamoxifen	-1.1
Mesenchymal cell_viability	5-Azacytidine	-1.2
Mesenchymal cell_viability	6-Mercaptopurine monohydrate	-1.2
Mesenchymal cell_viability	6-Thioguanine	-1.2
Mesenchymal cell_viability	C.I. Acid red 114	-1.2
Mesenchymal cell_viability	CI Basic red 9 monohydrochloride	-1.1
Mesenchymal cell_viability	Captan	-1.1
Mesenchymal cell_viability	Captan	-1.1

# Look up chemical to find the assays with response of selected level(s):

Select NCGC curve class

- 1.1 Or 1.1
- 1.2 Or 1.2
- 1.3 Or 1.3
- 1.4 Or 1.4
- 2.1 Or 2.1
- 2.2 Or 2.2
- 2.3 Or 2.3
- 2.4 Or 2.4
- 3 Or 3
- 4
- 5

Select

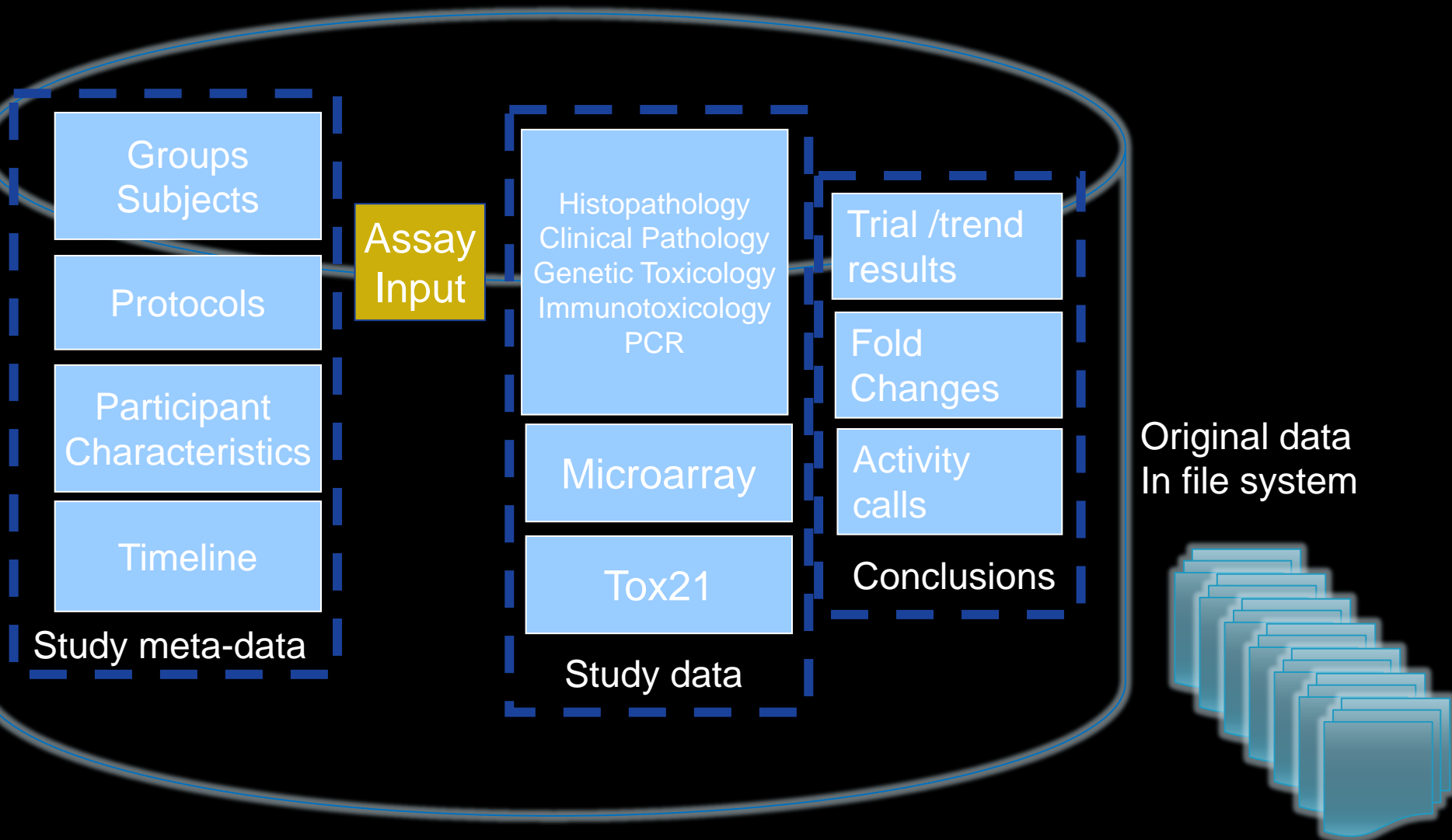
Assay	Chemical	Single Channel
▼ Chemical=Malachite green oxalate, Curve clas		
HUV-EC-C_viability	Malachite green oxalate	-1.1
DNA_repair_915_p2	Malachite green oxalate	-1.1
BJ_viability	Malachite green oxalate	-1.1
N2a_viability	Malachite green oxalate	-1.1
DNA_repair_1782_p2	Malachite green oxalate	-1.1
Estrogen_receptor_alpha_antagonist	Malachite green oxalate	-1.1
HepG2_viability	Malachite green oxalate	-1.1
MRC-5_viability	Malachite green oxalate	-1.1
Proximal tubule cell_viability	Malachite green oxalate	-1.1
endotoxin-viability-ntp-o3	Malachite green oxalate	-1.1
endotoxin-antagonist-ntp-o2	Malachite green oxalate	-1.1
Hek293_viability	Malachite green oxalate	-1.1

# Meta data and data dictionary

# How is CEBS organized?

- ◎ Meta data
  - Biological context, design, protocol and participant information
- ◎ Data
  - Based on assays (on specimen)
  - Or observations made during in-life phase
- ◎ Database organization
  - Study description and Assay domains
- ◎ Data format for loading
  - SIFT: Simple Investigation Formatted Text
- ◎ CEBS Data Dictionary

# CEBS database design



# SIFT: Standard format for loading CEBS

- ◎ Simple Investigation Formatted Text
- ◎ Syntax based on SOFT (GEO;  
*[www.ncbi.nlm.nih.gov/geo/](http://www.ncbi.nlm.nih.gov/geo/)*)
- ◎ Used to capture
  - Study meta-data
  - Study data
  - Data transformation
- ◎ Tools (java)
  - SIFT Loader (loads CEBS)
  - SIFT validator
  - CEBS SIFTBuilder
- ◎ Uses CEBS data dictionary to validate

# CEBS DD evolution

- Started in 2003
- Originally text based, now XML
- Based on definition tables in CEBS
- Correlation with other syntaxes used to build parsers for various data formats
  
- Currently aligned with
  - SEND (Standards for Exchange of Nonclinical Data)
  - OBI (Ontology for Biomedical Investigations)
- SIFT uses original MAGE-Tab for microarray data



# CEBS data in RDF

# Start with SIFT Tabular Data (implicit relationships)

Participant characteristics (meta data)

\$GROUP_NAME	\$COMPARTOR_NAME	\$IS_CONTROL_GROUP	\$COMPOUND	\$DOSE	\$DOSE_UNIT	\$ROUTE	\$SPECIES_COMMON_NAME	\$SEX	\$STRAIN
GROUP8	GROUP2		beta-Picoline	1250	mg/l	DOSED WATER	Mouse	FEMALE	B6C3F1
GROUP7	GROUP1		beta-Picoline	1250	mg/l	DOSED WATER	Mouse	MALE	B6C3F1
GROUP6	GROUP2		beta-Picoline	625	mg/l	DOSED WATER	Mouse	FEMALE	B6C3F1
GROUP5	GROUP1		beta-Picoline	625	mg/l	DOSED WATER	Mouse	MALE	B6C3F1
GROUP4	GROUP2		beta-Picoline	312.5	mg/l	DOSED WATER	Mouse	FEMALE	B6C3F1
GROUP3	GROUP1		beta-Picoline	312.5	mg/l	DOSED WATER	Mouse	MALE	B6C3F1
GROUP2		T	beta-Picoline	0	mg/l	DOSED WATER	Mouse	FEMALE	B6C3F1
GROUP1		T	beta-Picoline	0	mg/l	DOSED WATER	Mouse	MALE	B6C3F1

PARTICIPANT_NAME	ORGAN	HISTOPATHOLOGY
GROUP1_CAGE1_1	Lung	Hyperplasia
GROUP1_CAGE1_1	Nose	Metaplasia
GROUP1_CAGE3_3	Adrenal cortex	Adenoma
GROUP1_CAGE3_3	Nose	Atrophy
GROUP1_CAGE7_7	Nose	Hyperplasia

Histopathology observations (data)

# ...convert to RDF triples with explicit relationships in the data

```
:group1-cage1-1
  rdfs:label "GROUP1_CAGE1_1" ;      # This subject
  rdf:type obo:NCBITaxon_10090 ;    # is a mouse
  sift:member-of :group1-cage1 ;    # member of sub-group :group1-cage1
  bfo:0000159 obo:PATO_0000384 ;    # has quality at all times: male
  sift:dose :dose-0 ;               # has dose :dose-0
  sift:death-status :terminal-sacrifice ; # was sacrificed
  sift:death-date :study-day-731 . # died on day 731

:group1-cage1-1-lung
  rdf:type obo:MA_0000415 ;          # This is a mouse lung
  sift:part-of :group1-cage1-1 .    # part of :group1-cage1-1

:group1-cage1-1-lung-hyperplasia
  rdf:type obo:MPATH_134 ;           # This is a mouse hyperplasia
  obo:BFO_0000052 :group1-cage1-1-lung ; # in :group1-cage1-lung
  sift:severity sift:mild .         # has severity: mild
```

Create CEBS Application Ontology, Rules to capture SIFT relationships  
and tools to convert SIFT to ttl .....James A Overton

# Final aims

- RDF triple store
- This is Linked Data\*, structured and computer-readable
- Over which we can write SPARQL and other computational queries to produce summary reports
- Which will be available for use on the Semantic Web
- Which will be available for download
- Which users can query in new user interface

\*[Tim Berners-Lee](#) (2006-07-27). "[Linked Data—Design Issues](#)". [W3C](#). Retrieved 2010-12-18 at [http://en.wikipedia.org/wiki/Linked\\_data](http://en.wikipedia.org/wiki/Linked_data)

# Thanks

## ◎ NTP Team

Jennifer Fostel (COR), Beth Bowden , Michelle Hooth, Laura Hall, Raymond Tice, Scott Auerbach, Michael Devito

## ◎ Vistronix team

Julie Berke-Law, Phyllis Brown, Kevin Church, James Faison, Cari Favaro, Hui Gong, Mary Hicks, Isabel Lee, Justin Johnson, Jeremy McNabb, Robert Mullen, Anand Paleja, Veronica Pham, Mike Shaw

- James A Overton (CEBS Application Ontology)

Questions?