

# CEBS—Chemical Effects in Biological Systems: a public data repository integrating study design and toxicity data with microarray and proteomics data

Michael Waters<sup>1</sup>, Stanley Stasiewicz<sup>1</sup>, B. Alex Merrick<sup>1</sup>, Kenneth Tomer<sup>1</sup>, Pierre Bushel<sup>1</sup>, Richard Paules<sup>1</sup>, Nancy Stegman<sup>1</sup>, Gerald Nehls<sup>1</sup>, Kenneth J. Yost<sup>2</sup>, C. Harris Johnson<sup>2</sup>, Scott F. Gustafson<sup>2</sup>, Sandhya Xirasagar<sup>2</sup>, Nianqing Xiao<sup>2</sup>, Cheng-Cheng Huang<sup>2</sup>, Paul Boyer<sup>2</sup>, Denny D. Chan<sup>2</sup>, Qinyan Pan<sup>2</sup>, Hui Gong<sup>2</sup>, John Taylor<sup>3</sup>, Danielle Choi<sup>4,5</sup>, Asif Rashid<sup>4</sup>, Ayazaddin Ahmed<sup>6</sup>, Reese Howle<sup>6</sup>, James Selkirk<sup>1</sup>, Raymond Tennant<sup>1</sup> and Jennifer Fostel<sup>4,\*</sup>

<sup>1</sup>NIEHS, National Center for Toxicogenomics, PO Box 12233, Research Triangle Park, NC 27709, <sup>2</sup>Science Applications International Corporation, 1710 SAIC Drive, McLean, VA 22101, <sup>3</sup>Large Scale Biology Corporation, 3333 Vaca Valley Parkway, Vacaville, CA 95688, <sup>4</sup>Lockheed Martin Information Technologies, PO Box 12233, Research Triangle Park, North Carolina 27709, <sup>5</sup>Research Triangle Institute, PO Box 12194, Research Triangle Park, NC 27709 and <sup>6</sup>Alpha Gamma Technologies, Inc., 4700 Falls of Neuse Road, Suite 350, Raleigh, NC, 27609, USA

Received June 28, 2007; Revised August 30, 2007; Accepted September 11, 2007

## ABSTRACT

CEBS (Chemical Effects in Biological Systems) is an integrated public repository for toxicogenomics data, including the study design and timeline, clinical chemistry and histopathology findings and microarray and proteomics data. CEBS contains data derived from studies of chemicals and of genetic alterations, and is compatible with clinical and environmental studies. CEBS is designed to permit the user to query the data using the study conditions, the subject responses and then, having identified an appropriate set of subjects, to move to the microarray module of CEBS to carry out gene signature and pathway analysis. Scope of CEBS: CEBS currently holds 22 studies of rats, four studies of mice and one study of *Caenorhabditis elegans*. CEBS can also accommodate data from studies of human subjects. Toxicogenomics studies currently in CEBS comprise over 4000 microarray hybridizations, and 75 2D gel images annotated with protein identification performed by MALDI and MS/MS. CEBS contains raw microarray data collected in accordance with MIAME guidelines and provides tools for data selection, pre-processing and analysis resulting in annotated lists of genes of interest. Additionally, clinical chemistry and histopathology

findings from over 1500 animals are included in CEBS. CEBS/BID: The BID (Biomedical Investigation Database) is another component of the CEBS system. BID is a relational database used to load and curate study data prior to export to CEBS, in addition to capturing and displaying novel data types such as PCR data, or additional fields of interest, including those defined by the HESI Toxicogenomics Committee (in preparation). BID has been shared with Health Canada and the US Environmental Protection Agency. CEBS is available at <http://cebs.niehs.nih.gov>. BID can be accessed via the user interface from <https://dir-apps.niehs.nih.gov/arc/>. Requests for a copy of BID and for depositing data into CEBS or BID are available at <http://www.niehs.nih.gov/cebs-df/>.

## INTRODUCTION

CEBS (Chemical Effects in Biological Systems) is a public repository for toxicogenomics data developed by the National Center for Toxicogenomics (NCT) within the National Institute of Environmental Health Science (NIEHS). Development of CEBS began in 2002 (1) and focused first on capture of microarray and proteomics data. The CEBS SysBio Object Model (2), based on MIAME (3) and MIAPE Standard (4), was used for this

\*To whom correspondence should be addressed. Tel: +1 919 541 5055; Email: [fostel@niehs.nih.gov](mailto:fostel@niehs.nih.gov)

portion of the development of CEBS. CEBS1 was released in August 2003, followed by the start of development of CEBS2. The aim of the second stage of CEBS development was to integrate study design and toxicological assay data with the omics data captured in CEBS. Thus, the CEBS SysTox Object Model (5) and the CEBS Data Dictionary (CEBS-DD) (6) were developed to permit accurate management of study data. CEBS2 was released in November 2006.

As of July 2007, there are 27 toxicogenomics studies in CEBS. A 'study' refers to an observational or perturbational experiment carried out over a defined timeline to understand a biological system, address a scientific question and/or to generate hypotheses. Of the 27 studies, 22 are of rat, 4 are of mouse and 1 is of *Caenorhabditis elegans*. Twenty-six of the studies have associated microarray data, one has proteomics data. Companies which have published data in CEBS include Iconix Biosciences (<http://www.iconixpharm.com/>), Johnson & Johnson (7,8), Pfizer Inc. (9) and Sankyo Co., Ltd (10). Other data in CEBS have been submitted by researchers at the National Cancer Institute (11,12), the University of Tennessee (13–15), the University College of London (16,17), the HESI Toxicogenomics Committee (18,19) and the Toxicogenomics Research Consortium (submitted). Additional data have been deposited in CEBS from in-house studies carried out at the NIEHS (20) and at the National Toxicology Program (NTP) (21,22).

CEBS can store data from studies of laboratory animals, cultured cells or humans. Most studies in CEBS contain observations or measurements made of the study subjects and of specimens such as blood or tissue sections derived from these subjects. The objective of CEBS is to permit the user to integrate various data types and studies. The CEBS user can select groups of subjects drawn from different studies, based on subject responses or study conditions. Once the subjects are selected, any associated microarray data can be analyzed to produce lists of annotated genes that can shed light on the biological and toxicological processes occurring in the subjects.

## MATERIALS AND METHODS

### Scope and Utility of CEBS

CEBS is the first public repository designed to integrate toxicological, histopathological and other biological measures with omics data. A number of other databases, for instance the Gene Expression Omnibus (GEO) (23,24), capture microarray data and information about the sample treatment. The ArrayExpress database (25) captures observations and measures taken on the study subject concurrently with preparation of the tissue for microarray analysis (26). A distinguishing feature of CEBS is that the data captured from toxicogenomics studies includes observations made of the subject throughout the study timeline, potentially both before and after a specimen was taken for toxicological, histopathological or other biological analysis. Since the descriptions of the protocols used in the study and associated analyses are captured using controlled vocabularies rather than in free

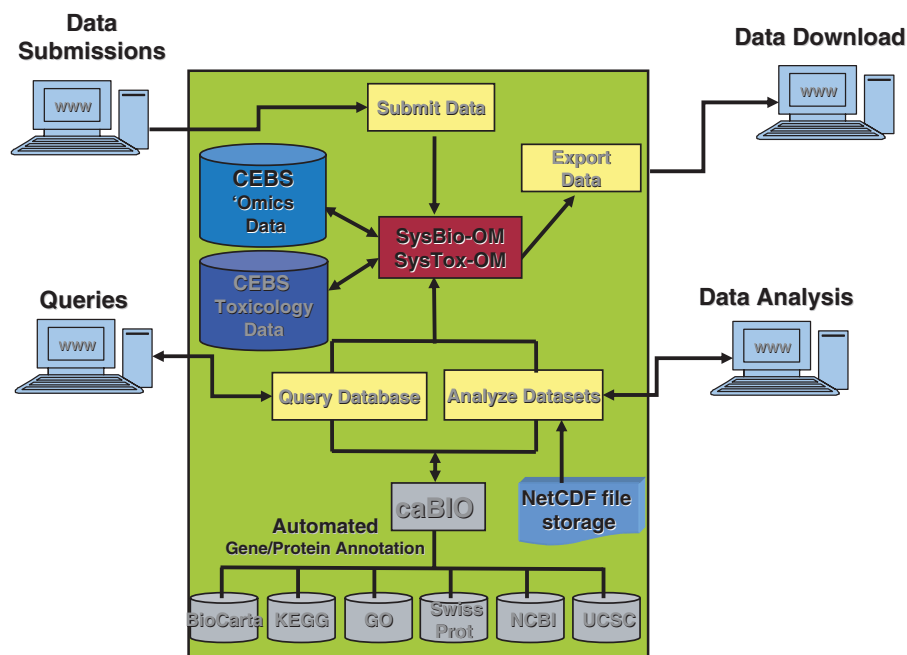
text form, these data are available for effective filtering and query. These protocols, measures and temporal events are useful in anchoring the transcriptomics or proteomics profile displayed by the specimen within the time- and dose-dependent biological responses seen in the study. Thus CEBS supports phenotypic anchoring (27–31), defined as the linking of microarray or proteomics data with a pathophysiological phenotype.

CEBS includes both microarray and proteomics data. Microarray data in CEBS includes 965 hybridizations to Affymetrix arrays, 924 to Agilent arrays and 1810 to custom format microarrays. Data can be combined within a given microarray platform for cross-study analysis in CEBS. Thus, data from rats exposed to any of a number of different chemical agents can be selected using a CEBS query tool and the microarray data can be compared to identify differentially responsive gene products. The proteomics data include downloadable 2D gel images, and MALDI and MS/MS spectra used to identify peptide spots from the gels. Intensity levels of both identified and unidentified spots are also available in CEBS, and can be browsed for association with time- and dose-response.

While the transcriptomics data in CEBS are captured using MIAME guidelines, at this time there is no widely accepted public standard for the exchange and capture of study design and toxicity assay data. A number of efforts are underway to create such a standard. Recently the result of a consensus about the minimal information to include was reported (32). In addition, a format for data exchange is being developed by the Standard for Exchange of Non-clinical Data (SEND) Consortium (<http://www.cdisc.org/standards/index.html>) and an ontology for describing a biomedical investigation, which would include a toxicology study, is under development by the OBI (Ontology for Biomedical Investigations) Working Group (<https://wiki.cbil.upenn.edu/obiwiki/index.php/HomePage>). CEBS will support these standards as they are developed.

Most institutions engaged in toxicology and toxicogenomics studies use in-house data repositories that are tailored to the study designs and regimens used in the institution [c.f. dbZach (33) and EDGE (34)]. In contrast, because it is a public resource, CEBS must be able to manage data from a variety of sources, reflecting a wide range of experimental organisms and study designs. Additionally, CEBS can manage data from experimental animals, from *in vitro* cells in culture, from human studies and from experiments with model organisms such as *C. elegans*. Each depositor to CEBS to date has used a different data experimental design, reflecting different means to care for the subjects, different treatment regimens, varying measurements taken and so forth.

The problems associated with managing various data streams have been addressed by the creation of BID, the Biomedical Investigation Database, which is based on the CEBS data dictionary (CEBS-DD). The CEBS-DD describes the incoming data based on alignment with public standards and proprietary data formats. At present, data submissions to CEBS are handled by collaboration between the depositor and the CEBS curation staff, because, to date, each depositor has used



**Figure 1.** Architecture of CEBS. The green box represents CEBS, and shows the relationships between CEBS components: the SysBio and SysTox Object models, internal databases and file structure and the caBIO annotation engine and external annotation resources. User access is represented using blue terminals.

a different format. Information about data deposition is available at the CEBS Development Forum (<http://www.niehs.nih.gov/research/resources/databases/cebs/forum>).

BID is built in Oracle with a Cold Fusion interface, and, because it is used to load study data into CEBS, it contains essentially the same content as CEBS does. In addition, BID can easily be modified to contain additional data fields, as requested by users. Thus, as part of the collaboration with the HESI Toxicogenomics Committee, BID was extended to include PCR data and to capture additional fields describing subject handling during the study. Additionally, the BID interface was extended to permit query by the users of these fields. These data will be available to the public via BID once the Committee releases them. At the present time BID is a data management tool, permitting access to data and download capabilities.

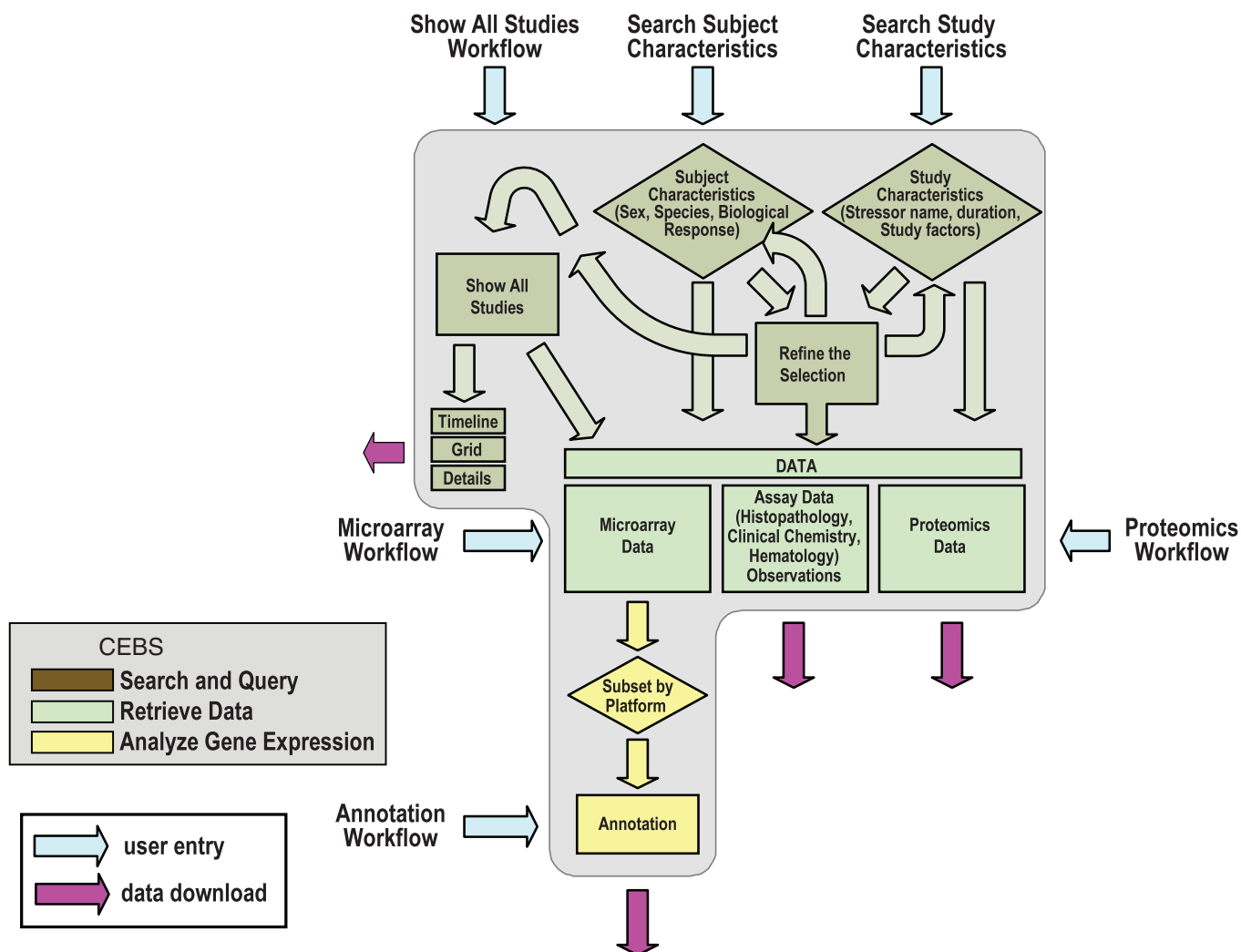
### CEBS functionality

The architecture of CEBS is shown in Figure 1, which displays the relationships between CEBS components: the CEBS SysBio and SysTox Object Models; the Oracle databases handling study design, assay data and metadata for microarray and proteomics data; the caBIO annotation engine developed by the National Cancer Institute's Center for Bioinformatics (NCICB). Microarray data files are stored as netCDF files within CEBS. Access to the data in CEBS is via the SysTox Browser (<http://cebs.niehs.nih.gov/>). CEBS was moved to the NIEHS at the end of 2006 from its developmental location at SAIC. Prior to implementing CEBS at NIEHS, a series of stress tests were run to determine whether the workloads could

be supported with the infrastructure and to identify any potential bottlenecks. This involved simulating up to 100 concurrent users in various typical functional scenarios.

The CEBS user can follow various workflows within CEBS. These include: Show All Studies; Search by Study Characteristics; Search by Subject Characteristics; Browse Proteomics Data; Analyze Microarray Data Workflow; Annotate Gene List. The CEBS user can also combine aspects of these different workflows to customize their exploration and use of the data in CEBS (Figure 2). Users can combine elements of different workflows to customize their queries and use of CEBS as diagrammed in Figure 2. Additionally, users can download data and annotation in different formats at various points throughout the workflows.

'Show All Studies Workflow' permits the user to see a list of all studies and investigations in CEBS (Figure 3A). An investigation refers to a self-contained scientific enquiry, which can be composed of several studies. A study is, as defined earlier, an observational or perturbational experiment carried out over a defined timeline to understand a biological system, address a scientific question and/or to generate hypotheses. The data type(s) associated with each study in CEBS are shown with icons next to the title, and the user can quickly retrieve any data associated with the study using links on the page. The Study Timeline, Study Details and Study Group Grid are also accessible from this page. Study Timeline provides a graphical representation of the timeline of the Study, showing when treatment was applied, when observations and husbandry occurred, and other important events that occurred during the Study (Figure 3B). The Study Group Grid provides a rapid



**Figure 2.** Workflows in CEBS. The gray box represents CEBS. Search and query workflows are shown in tan, data in green and analysis and annotation in yellow. Entry points for users are shown with blue arrows, and data download points with magenta arrows.

overview of the study subjects, and permits the user to unambiguously identify relevant groups of biological replicates for analysis and comparison (Figure 3C).

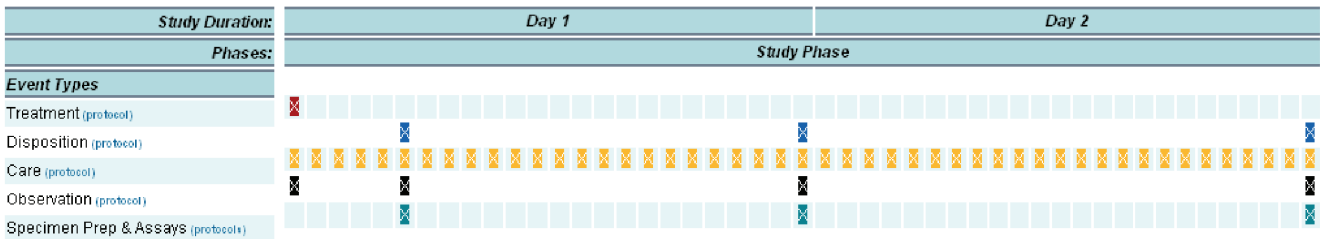
- (i) 'Search by Study Characteristics Workflow' permits the user to identify all studies with particular characteristics, such as the duration of the study, the species used, the stressor used (e.g. 'show me all studies using acetaminophen'), and study factor(s) (e.g. studies using time course or dose-response or genotype as study variables). Once the subset of studies is identified, the user can jump to 'Show all Studies' and browse the subset selected, examine timelines and retrieve data.
- (ii) 'Search by Subject Characteristics Workflow' permits the CEBS user to select individual Subjects by their strain or species, age, sex or biological response as revealed in assay data from these subjects. For example, the user can select all animals with a particular grade of a given histopathology finding, plus their comparators, and move into the microarray analysis section of CEBS to identify the genes that are differentially responsive between the two sets.
- (iii) 'Browse Proteomics Data' permits the user to select from five proteomics experiments in CEBS and explore changes in protein profiles in response to treatments in a time- and dose-dependent manner. These studies include subcellular fraction analysis of proteomes such as preparations of nuclei, enriched cytosol and enriched microsomes. Actual and reference spot lists with pI and apparent MW, spectra from particular spots chosen for identification, and protein identifications can all be browsed or downloaded in commonly used formats. Images of the 2D gels used to separate the proteins, and the mass spectrometry spectra can be downloaded.
- (iv) 'Analyze Microarray Data Workflow' permits the user to list all the microarray experiments in CEBS or to enter this Workflow with either a subset of Studies or a list of individual hybridizations. The user can select one or more studies from a given

**A Investigation**

Study	Accession	PI	Institution	Details	Data
<input type="checkbox"/> + 2004-dose-and time-responses to acute administration of phenobarbital	004-00001-0010-000-6	Naoki Kiyosawa	Sankyo Co., Ltd.		MA DB
<input type="checkbox"/> + 2004-gene expression in C.elegans Dauer larvae and long-lived daf-2 mutants	003-00002-0010-000-6	Joshua McElwee	University College London		MA DB
<input type="checkbox"/> - 2005-acute dose- and time-effects of hepatotoxicants	001-00001-0020-000-4	Ray Tennant	National Center for Toxicogenomics		MA CS FA DB
<input type="checkbox"/> + 1,2-Dichlorobenzene-Rat-F344/N-2004	001-00001-0021-000-5	Ray Tennant		Details Design Timeline	MA CS FA DB
<input type="checkbox"/> + 1,4-Dichlorobenzene-Rat-F344/N-2004	001-00001-0022-000-6	Ray Tennant		Details Design Timeline	MA CS FA DB
<input type="checkbox"/> + Bromobenzene-Rat-F344/N-2004	001-00001-0028-000-2	Ray Tennant		Details Design Timeline	MA CS FA DB
<input type="checkbox"/> + Diquat dibromide-Rat-F344/N-2004	001-00001-0025-000-9	Ray Tennant		Details Design Timeline	MA CS FA DB
<input type="checkbox"/> + Galactosamine-Rat-F344/N-2004	001-00001-0024-000-8	Ray Tennant		Details Design Timeline	MA CS FA DB
<input type="checkbox"/> + Monocrotaline-Rat-F344/N-2004	001-00001-0023-000-7	Ray Tennant		Details Design Timeline	MA CS FA DB
<input type="checkbox"/> + N-nitrosomorpholine-Rat-F344/N-2004	001-00001-0026-000-0	Ray Tennant		Details Design Timeline	MA CS FA DB
<input type="checkbox"/> + Thioacetamide-Rat-F344/N-2004	001-00001-0027-000-1	Ray Tennant		Details Design Timeline	MA CS FA DB
<input type="checkbox"/> + 2005-dose-and time-responses to acute administration of acetaminophen	002-00001-0010-000-4	Rick Irwin	National Toxicology Program		MA CS FA
<input type="checkbox"/> + Hyperoxia_Mouse_2004	005-00003-0010-000-8	Steven Kleeberger	National Institute of Environmental Health Sciences		MA

**B**

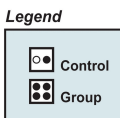
<b>Investigation Title:</b>	Molecular characterization of phenotypic response to parallel acute exposure to one of 8 hepatotoxicants to F344 rats
<b>Study Title:</b>	Application of 1,2-dichlorobenzene to F344/N rats via oral gavage to evaluate acute toxicity



**C**

<b>Investigation Title:</b>	Molecular characterization of phenotypic response to parallel acute exposure to one of 8 hepatotoxicants to F344 rats
<b>Study Title:</b>	Application of bromobenzene to F344/N rats via oral gavage to evaluate acute toxicity

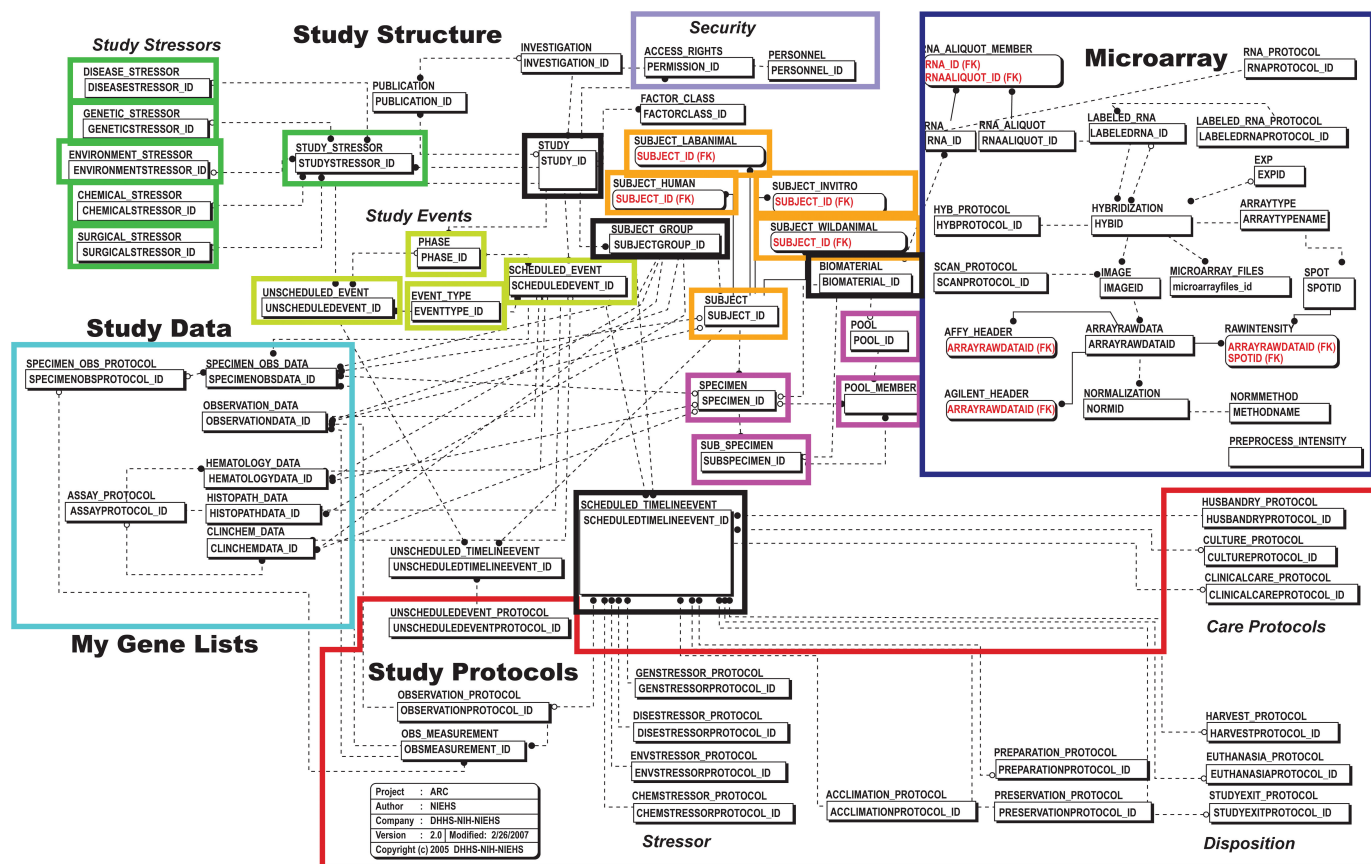
Dose	Time		
	6	24	48
0	Vehicle 6 hour	Vehicle 24 hour	Vehicle 48 hour
25	25mg/kg 6 hour	25mg/kg 24 hour	25mg/kg 48 hour
75	75mg/kg 6 hour	75mg/kg 24 hour	75mg/kg 48 hour
250	250mg/kg 6 hour	250mg/kg 24 hour	250mg/kg 48 hour



**Figure 3.** Screenshots of three novel CEBS displays, the Study list and disparate associated data types (A), the Study Timeline (Boxes with X indicates an event occurred at that time point. Details of protocols are accessible via links at left) (B) and the Study Group Grid (C).

microarray platform and view the details of the study, including access to the QC data when available. The next step in the analysis workflow is to pre-process the data using either default

settings or the normalization and thresholding options provided and then carry out a global visualization to check the distribution of data from individual arrays via box-and-whisker plots



**Figure 4.** BID architecture. Each table is labeled, and contains the primary and foreign keys to allow relationships to be easily discerned. The different modules of BID are enclosed in boxes. Study data tables are enclosed in aqua, study protocols in red, microarray workflow and data are in dark blue, study stressors in spring green, study events in olive green, study subject characteristics in orange, study specimens in magenta. Study tables, including publication and access rights, are not outlined.

and determine the suitability for further analysis by viewing a multi-dimensional scaling projection and a hierarchical clustering of all the selected hybridizations. Analysis routines are developed using the R open source statistical software and BioConductor. Following this step, the user can select comparisons between groups of arrays or within arrays (suited to two color hybridizations only). The user sets the test for significance from the menus provided, and retrieves a list of oligonucleotide probes passing the significance criteria with *P*-value if this option was selected. The list of probes is then passed to the Annotate Gene List option, described below.

- (v) 'Annotate Gene List Workflow' permits the user to add annotation to a list of genes entered or in a comma-delimited text file. The annotation is also applied to lists of genes generated from analysis within CEBS. CEBS utilizes the caBIO annotation engine from the NCICB, which has been modified to support rat and *C. elegans* caBIO provides links to annotation from BioCarta (<http://www.biocarta.com/>), the Gene Ontology GO (<http://www.geneontology.org/>), Kyoto Encyclopedia of Genes and Genomes KEGG (<http://www.genome.jp/kegg/>),

UniProt (for proteomics data, <http://www.pir.uni-prot.org/>), NCBI (<http://www.ncbi.nlm.nih.gov/>) and UCSC Genome (<http://genome.ucsc.edu/>) annotation sources. For gene lists computed within CEBS, a Fisher exact test is performed in order to identify significantly over- and under-represented biological categories associated with the identified genes. Significantly expressed genes by the microarray analysis may be projected onto BioCarta and KEGG pathway diagrams or the Gene Ontology within CEBS; genes with altered levels of product are indicated by different colors.

The architecture of BID is given in Figure 4. BID uses a workflow design, similar to the original MIAME/MAGE design. The BID database dependencies are set up to have the study defined prior to the subjects, and subjects and groups defined prior to specimens, and specimens defined prior to deposition of any associated data. Similarly, the microarray data storage portion of BID has been modified from the ArrayTrack (35) schema to model a hybridization workflow, capturing the links from a biomaterial to RNA to labeled RNA to hybridization to scanned data.

The majority of the work in BID has been in the area of study design and phenotypic data (primarily clinical chemistry and histopathology). Each subject type requires a spectrum of characteristics and protocols specific to that subject type. For example, if the subject is a lab animal, then the protocols are husbandry and euthanasia, whereas if the subject is a cell culture then the protocols are culture and harvest. Subject characteristics collected might be strain, sex, age, gut microflora characterization if the subject is a lab animal, while a cell culture might be characterized by cell cycle time, number of passages, ploidy, etc. This database design allows the user to focus quickly on relevant details both in data deposition and in querying. Similarly, details of stressor characteristics and protocols are specific for chemical, genetic and environmental stressors. A recent publication describes a checklist for the minimum information needed to interpret/exchange toxicology data, and BID adheres to this standard (32). In addition, the Ontology for Biomedical Investigations (OBI) Working Group (<https://wiki.cbil.upenn.edu/obiwiki/index.php/HomePage>) is developing an ontology to be used to annotate a biomedical investigation, as would be done for automatic depositions of data into CEBS.

#### Future directions

CEBS2 and BID are released, and can be accessed at <http://cebs.niehs.nih.gov> and <https://dir-apps.niehs.nih.gov/arc/>. Going forward we hope to integrate the best features of CEBS2 and BID and concentrate on three new areas as we develop CEBS3: facilitating data loading and exchange, addition of novel data modules, addition of enhanced analysis. CEBS3 will be made available to the public after development and testing.

*Integration of BID and CEBS2.* The architecture of CEBS3 will be streamlined to facilitate data entry and retrieval, and the workflow dependencies will become optional, so that data from observational and genetic studies may be more easily entered.

*Facilitating data loading and exchange.* Currently the biggest obstacle to increasing the content of CEBS is the unique formats used by each depositor, which reflects the lack of a common standard for formatting study data. There are efforts to address this need, led by the OBI working group and the CDISC/SEND projects. Once the data fields and format are addressed, then a minimum checklist for a study must be developed. Towards this, the MIBBI (Minimal Information about Biological and Biomedical Investigations) group has been formed, and a recent publication describing a checklist for a toxicology study written (32). Exchange formats for microarray data include the SOFT format (36), MAGE-tab (37) and MAGE-ML (38). BIO-tab is under development at the EBI to exchange functional genomics data ([http://nebc.nox.ac.uk/workshops/nebc\\_ebi/sansonne\\_biomap\\_info.pdf](http://nebc.nox.ac.uk/workshops/nebc_ebi/sansonne_biomap_info.pdf)). We are developing a format for study data termed SIFT (Simple Investigation Formatted Text), and will collaborate with BIO-tab as it is developed. Ideally, applications will be written to permit a user

to create a SIFT file for a study and associated data, verify the format, validate the contents and then transfer to NIEHS for automated loading into CEBS3, and rely on current microarray data formats for associated 'omics data.

*Addition of new data modules.* It is of interest to expand the capability of CEBS to be searched on the basis of chemical information, and towards this end we are collaborating with the NTP to permit views of short-term testing results obtained by the NTP Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM), and of high-throughput screening, and with the EPA to access a public chemical structure viewer. Additionally, we hope to use the standards developed by the MSI (Metabolomics Standards Initiative) to house experimental metabolomics data.

*Enhanced analytical features.* At the moment CEBS permits the user to identify genes with a significantly altered transcript levels, and to combine subjects from different studies if they were tested using the same microarray platform. However, the user must begin each analysis with raw data from the entire microarray. We plan to permit additional analytical tools, for example ANOVA and unsupervised pattern finding, and also store normalized data values so that the user does not need to re-analyze the array with each query. We anticipate that this will permit integration of data across microarray platform if the user chooses to do so.

#### SUMMARY

CEBS is a public repository integrating data describing study timeline and design, histopathological and biological measures and 'omics data. This permits the user to anchor 'omics data in the unfolding biological response pattern captured in the study data in CEBS. Users can access CEBS either by accessing 'omics data directly, or by way of the search and query workflows, using characteristics of studies or subjects to select 'omics data. To illustrate the various options available we have posted material at the CEBS Development Forum (<http://www.niehs.nih.gov/research/resources/databases/cebs/forum>) and as Supplementary Data here. CEBS integrates data from a number of contributors, making it possible to integrate disparate data and develop comprehensive answers to questions posed in the database. The BID data management tool is used to house data prior to loading into CEBS, and to expand the data management capabilities of the CEBS/BID system by permitting the user to deposit novel data types and attributes. The BID user interface permits the users to access the data in BID analogously to the searching capabilities of CEBS.

We anticipate that with publication of CEBS2 that more data will be contributed, making it possible to identify an ever-increasing number of gene signatures and mechanistic pathways and networks relevant to toxicogenomics. Instructions for CEBS contributors can be found at the CEBS Development Forum (<http://www.niehs.nih.gov/research/resources/databases/cebs/forum>).

CEBS is available at <http://cebs.niehs.nih.gov>. BID can be accessed via the user interface from <https://dir-apps.niehs.nih.gov/arc/>.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We are grateful for the contributions of Sarah Bittenbender, Mark Jenkins, Tong Li, Scott McCrimmon, Sumeet Muju, Larry Schuler and Rona Zhou from the Science Applications International Corporation contract to the development of CEBS. This research was supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences. Funding to pay the Open Access publication charges for this article was provided by NIEHS Division of Intramural Research.

*Conflict of interest statement.* None declared.

## REFERENCES

- Waters, M., Boorman, G., Bushel, P., Cunningham, M., Irwin, R., Merrick, A., Olden, K., Paules, R., Selkirk, J. *et al.* (2003) Systems toxicology and the Chemical Effects in Biological Systems (CEBS) knowledge base. *EHP Toxicogenomics*, **111**, 15–28.
- Xirasagar, S., Gustafson, S., Merrick, B.A., Tomer, K.B., Stasiewicz, S., Chan, D.D., Yost, K.J.III, Yates, J.R.III, Sumner, S. *et al.* (2004) CEBS object model for systems biology data, SysBio-OM. *Bioinformatics*, **20**, 2004–2015.
- Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C.A. *et al.* (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat. Genet.*, **29**, 365–371.
- Orchard, S., Hermjakob, H., Julian, R.K.Jr, Runte, K., Sherman, D., Wojcik, J., Zhu, W. and Apweiler, R. (2004) Common interchange standards for proteomics data: Public availability of tools and schema. *Proteomics*, **4**, 490–491.
- Xirasagar, S., Gustafson, S.F., Huang, C.C., Pan, Q., Fostel, J., Boyer, P., Merrick, B.A., Tomer, K.B., Chan, D.D. *et al.* (2006) Chemical effects in biological systems (CEBS) object model for toxicology data, SysTox-OM: design and application. *Bioinformatics*, **22**, 874–882.
- Fostel, J., Choi, D., Zwickl, C., Morrison, N., Rashid, A., Hasan, A., Bao, W., Richard, A., Tong, W. *et al.* (2005) Chemical effects in biological systems—data dictionary (CEBS-DD): a compendium of terms for the capture and integration of biological study design description, conventional phenotypes, and 'omics data. *Toxicol. Sci.*, **88**, 585–601.
- McMillian, M., Nie, A.Y., Parker, J.B., Leone, A., Kemmerer, M., Bryant, S., Herlich, J., Yieh, L., Bittner, A. *et al.* (2004) Inverse gene expression patterns for macrophage activating hepatotoxicants and peroxisome proliferators in rat liver. *Biochem. Pharmacol.*, **67**, 2141–2165.
- McMillian, M., Nie, A.Y., Parker, J.B., Leone, A., Bryant, S., Kemmerer, M., Herlich, J., Liu, Y., Yieh, L. *et al.* (2004) A gene expression signature for oxidant stress/reactive metabolites in rat liver. *Biochem. Pharmacol.*, **68**, 2249–2261.
- Kramer, J.A., Curtiss, S.W., Kolaja, K.L., Alden, C.L., Blomme, E.A., Curtiss, W.C., Davila, J.C., Jackson, C.J. and Bunch, R.T. (2004) Acute molecular markers of rodent hepatic carcinogenesis identified by transcription profiling. *Chem. Res. Toxicol.*, **17**, 463–470.
- Kiyosawa, N., Tanaka, K., Hirao, J., Ito, K., Niino, N., Sakuma, K., Kanbori, M., Yamoto, T., Manabe, S. *et al.* (2004) Molecular mechanism investigation of phenobarbital-induced serum cholesterol elevation in rat livers by microarray analysis. *Arch. Toxicol.*, **78**, 435–442.
- Ross, D.T., Scherf, U., Eisen, M.B., Perou, C.M., Rees, C., Spellman, P., Iyer, V., Jeffrey, S.S., Van de Rijn, M. *et al.* (2000) Systematic variation in gene expression patterns in human cancer cell lines. *Nat. Genet.*, **24**, 227–235.
- Scherf, U., Ross, D.T., Waltham, M., Smith, L.H., Lee, J.K., Tanabe, L., Kohn, K.W., Reinhold, W.C., Myers, T.G. *et al.* (2000) A gene expression database for the molecular pharmacology of cancer. *Nat. Genet.*, **24**, 236–244.
- Zou, F., Gelfond, J.A., Airey, D.C., Lu, L., Manly, K.F., Williams, R.W. and Threadgill, D.W. (2005) Quantitative trait locus analysis using recombinant inbred intercrosses: theoretical and empirical considerations. *Genetics*, **170**, 1299–1311.
- Williams, R.W., Gu, J., Qi, S. and Lu, L. (2001) The genetic structure of recombinant inbred mice: high-resolution consensus maps for complex trait analysis. *Genome Biol.*, **2**, RESEARCH0046.
- Williams, R.W., Bennett, B., Lu, L., Gu, J., DeFries, J.C., Carosone-Link, P.J., Rikke, B.A., Belknap, J.K. and Johnson, T.E. (2004) Genetic structure of the LXS panel of recombinant inbred mouse strains: a powerful resource for complex trait analysis. *Mamm. Genome*, **15**, 637–647.
- McElwee, J.J., Schuster, E., Blanc, E., Thornton, J. and Gems, D. (2006) Diapause-associated metabolic traits reiterated in long-lived daf-2 mutants in the nematode *Caenorhabditis elegans*. *Mech. Ageing Dev.*, **127**, 458–472.
- McElwee, J.J., Schuster, E., Blanc, E., Thomas, J.H. and Gems, D. (2004) Shared transcriptional signature in *Caenorhabditis elegans* Dauer larvae and long-lived daf-2 mutants implicates detoxification system in longevity assurance. *J. Biol. Chem.*, **279**, 44533–44543.
- Kramer, J.A., Blomme, E.A., Bunch, R.T., Davila, J.C., Jackson, C.J., Jones, P.F., Kolaja, K.L. and Curtiss, S.W. (2003) Transcription profiling distinguishes dose-dependent effects in the livers of rats treated with clofibrate. *Toxicol. Pathol.*, **31**, 417–431.
- Amin, R.P., Vickers, A.E., Sistare, F., Thompson, K.L., Roman, R.J., Lawton, M., Kramer, J., Hamadeh, H.K., Collins, J. *et al.* (2004) Identification of putative gene based markers of renal toxicity. *Environ. Health Perspect.*, **112**, 465–479.
- Heinloth, A.N., Irwin, R.D., Boorman, G.A., Nettesheim, P., Fannin, R.D., Sieber, S.O., Snell, M.L., Tucker, C.J., Li, L. *et al.* (2004) Gene expression profiling of rat livers reveals indicators of potential adverse effects. *Toxicol. Sci.*, **80**, 193–202.
- Irwin, R.D., Boorman, G.A., Cunningham, M.L., Heinloth, A.N., Malarkey, D.E. and Paules, R.S. (2004) Application of toxicogenomics to toxicology: basic concepts in the analysis of microarray data. *Toxicol. Pathol.*, **32** (Suppl. 1), 72–83.
- Irwin, R.D., Parker, J.S., Lobenhofer, E.K., Burka, L.T., Blackshear, P.E., Vallant, M.K., Lebetkin, E.H., Gerken, D.F. and Boorman, G.A. (2005) Transcriptional profiling of the left and median liver lobes of male F344/n rats following exposure to acetaminophen. *Toxicol. Pathol.*, **33**, 111–117.
- Barrett, T. and Edgar, R. (2006) Gene expression omnibus: microarray data storage, submission, retrieval, and analysis. *Meth. Enzymol.*, **411**, 352–369.
- Edgar, R., Domrachev, M. and Lash, A.E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.*, **30**, 207–210.
- Rocca-Serra, P., Brazma, A., Parkinson, H., Sarkans, U., Shojatalab, M., Contrino, S., Vilo, J., Abeygunawardena, N., Mukherjee, G. *et al.* (2003) ArrayExpress: a public database of gene expression data at EBI. *C R Biol.*, **326**, 1075–1078.
- Mattes, W.B., Pettit, S.D., Sansone, S.A., Bushel, P.R. and Waters, M.D. (2004) Database development in toxicogenomics: issues and efforts. *Environ. Health Perspect.*, **112**, 495–505.
- Gottschalg, E., Moore, N.E., Ryan, A.K., Travis, L.C., Waller, R.C., Pratt, S., Atmaca, M., Kind, C.N. and Fry, J.R. (2006) Phenotypic anchoring of arsenic and cadmium toxicity in three hepatic-related cell systems reveals compound- and cell-specific selective up-regulation of stress protein expression: implications for fingerprint profiling of cytotoxicity. *Chem. Biol. Interact.*, **161**, 251–261.
- Luo, W., Fan, W., Xie, H., Jing, L., Ricicki, E., Vouros, P., Zhao, L.P. and Zarbl, H. (2005) Phenotypic anchoring of global gene expression profiles induced by N-hydroxy-4-acetylamino-biphenyl and benzo[a]pyrene diol epoxide reveals correlations between expression



- profiles and mechanism of toxicity. *Chem. Res. Toxicol.*, **18**, 619–629.
29. Moggs, J.G., Tinwell, H., Spurway, T., Chang, H.S., Pate, I., Lim, F.L., Moore, D.J., Soames, A., Stuckey, R. *et al.* (2004) Phenotypic anchoring of gene expression changes during estrogen-induced uterine growth. *Environ. Health Perspect.*, **112**, 1589–1606.
30. Paules, R. (2003) Phenotypic anchoring: linking cause and effect. *Environ. Health Perspect.*, **111**, A338–A339.
31. Powell, C.L., Kosyk, O., Ross, P.K., Schoonhoven, R., Boysen, G., Swenberg, J.A., Heinloth, A.N., Boorman, G.A., Cunningham, M.L. *et al.* (2006) Phenotypic anchoring of acetaminophen-induced oxidative stress with gene expression profiles in rat liver. *Toxicol. Sci.*, **93**, 213–222.
32. Fostel, J.M., Burgoon, L., Zwickl, C., Lord, P., Corton, J.C., Bushel, P.R., Cunningham, M., Fan, L., Edwards, S.W. *et al.* (2007) Toward a checklist for exchange and interpretation of data from a toxicology study. *Toxicol. Sci.*, **99**, 26–34.
33. Burgoon, L.D., Boutros, P.C., Dere, E. and Zacharewski, T.R. (2006) dbZach: a MIAME-compliant toxicogenomic supportive relational database. *Toxicol. Sci.*, **90**, 558–568.
34. Hayes, K.R., Vollrath, A.L., Zastrow, G.M., McMillan, B.J., Craven, M., Jovanovich, S., Rank, D.R., Penn, S., Walisser, J.A. *et al.* (2005) EDGE: a centralized resource for the comparison, analysis, and distribution of toxicogenomic information. *Mol. Pharmacol.*, **67**, 1360–1368.
35. Tong, W., Cao, X., Harris, S., Sun, H., Fang, H., Fuscoe, J., Harris, A., Hong, H., Xie, Q. *et al.* (2003) ArrayTrack—supporting toxicogenomic research at the U.S. Food and Drug Administration National Center for Toxicological Research. *Environ. Health Perspect.*, **111**, 1819–1826.
36. Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Rudnev, D., Evangelista, C., Kim, I.F., Soboleva, A., Tomashevsky, M. *et al.* (2007) NCBI GEO: mining tens of millions of expression profiles – database and tools update. *Nucleic Acids Res.*, **35**, D760–D765.
37. Rayner, T.F., Rocca-Serra, P., Spellman, P.T., Causton, H.C., Farne, A., Holloway, E., Irizarry, R.A., Liu, J., Maier, D.S. *et al.* (2006) A simple spreadsheet-based, MIAME-supportive format for microarray data: MAGE-TAB. *BMC Bioinformatics*, **7**, 489.
38. Whetzel, P.L., Parkinson, H., Causton, H.C., Fan, L., Fostel, J., Fragoso, G., Game, L., Heiskanen, M., Morrison, N. *et al.* (2006) The MGED Ontology: a resource for semantics-based description of microarray experiments. *Bioinformatics*, **22**, 866–873.