CEBS update: curated toxicology database with enhanced tools for data integration

Cari Martini¹, Ying Frances Liu¹, Hui Gong¹, Nicole Sayers¹, German Segura¹ and Jennifer Fostel²,*

¹ASRC Federal, 430 Davis Dr., Suite 400, Morrisville, NC 27560, USA and ²Division of the National Toxicology Program, National Institute of Environmental Health Sciences, PO Box 12233, Research Triangle Park, NC 27709, USA

Received September 09, 2021; Revised October 05, 2021; Editorial Decision October 06, 2021; Accepted November 01, 2021

ABSTRACT

The Chemical Effects in Biological Systems database (CEBS) contains extensive toxicology study results and metadata from the Division of the National Toxicology Program (NTP) and other studies of environmental health interest. This resource grants public access to search and collate data from over 10 250 studies for 12 750 test articles (chemicals, environmental agents). CEBS has made considerable strides over the last 5 years to integrate growing internal data repositories into data warehouses and data marts to better serve the public with high quality curated datasets. This effort includes harmonizing legacy terms and metadata to current standards, mapping test articles to external identifiers, and aligning terms to OBO (Open Biological and Biomedical Ontology) Foundry ontologies. The data are made available through the CEBS Homepage (https://cebs.niehs.nih.gov/cebs/), guided search applications, flat files on FTP (file transfer protocol), and APIs (application programming interface) for user access and to provide a bridge for computational tools. The user interface is intuitive with a single search bar to query keywords related to study metadata, publications, and data availability. Results are consolidated to single pages for each test article with NTP conclusions, publications, individual studies, data collections, and links to related test articles and projects available together.

INTRODUCTION

The Chemical Effects in Biological Systems (CEBS) database (https://cebs.niehs.nih.gov/cebs/; also: https://doi.org/10.22427/NTP-DATA-1) is a public resource housing toxicological environmental health data. CEBS is the primary repository for the Division of the National Toxicology Program (NTP), covering a wide array of toxicology data from studies conducted at NTP over a 45-year period. Additionally, CEBS includes a few depositions from academic, industrial, and other government laboratories; streamlining the capability to serve as a repository for data from these sources is a future direction for CEBS.

The first version of CEBS was released in 2003 as a repository for toxicogenomics data produced by the National Center for Toxicogenomics and National Institute of Environmental Health Science (NIEHS) (1,2). When CEBS was moved to NTP the use of CEBS focused on allowing public access to published NTP data tables, and this is still the major use. Over 2000 users access CEBS each business day to review or download summary data files from NTP reports and publications.

The CEBS database has expanded in the years since to include over 10 250 studies covering a variety of data including general toxicology, carcinogenicity, genetic toxicology and gene expression. More data types are in the pipeline and will be released when reported by NTP. NTP added the CEBS Publications application to serve public environmental health data reported in journal articles and other sources for user review and download (3).

Test articles (chemicals, environmental agents) are nominated for testing by NTP and a series of related studies may be conducted to uncover the effects of exposure. The results of these tests include summary and conclusion level results from NTP decision-making committees along with results of statistical analyses and the individual animal or in vitro assay data. Currently, CEBS houses data for over 12 750 test articles. This number reflects NTP testing program studies, other investigations deposited in CEBS, and large-scale open resources such as the toxicological and expression data reported in DrugMatrix (4,5) with the aim to build a broad and rich resource.

Our primary focus at this time is on extending CEBS from primarily serving summary data pages for individual NTP reports to providing an integrated data resource for knowledge discovery, integrating data from various studies and test types into a single, easily queried resource. This work...
Table 1. Toxicology study types added to CEBS

| Area of toxicology   | Studies added                                                                 | Reference |
|---------------------|-------------------------------------------------------------------------------|-----------|
| Developmental and reproductive toxicology | ● 9 Prenatal developmental toxicity studies  
● 3 Reproductive assessment through continuous breeding studies  
● 6 Modified one generation studies | (6,7)     |
| Chemical disposition | ● 74 Toxicokinetic studies  
● 141 ADME (absorption, distribution, metabolism, and excretion) studies | (8)       |
| Genetic toxicology  | ● 2 in vitro Erythrocyte micronucleus assays  
● 59 Comet DNA damage assays | (9)       |
| Immunotoxicology    | ● 3 Hypersensitivity studies | (10)      |
| Expression profiling | ● 4 Short term expression studies with pathway annotation | (11)      |

is still in progress; towards this goal we have published the CEBSR (CEBS Reporting) Data Warehouse for evaluation by the public. Before we could integrate our data streams into CEBSR we needed to harmonize the various data feeds in the CEBS system and to map legacy terminology and metadata to the current NTP lexicon so that data can be easily found with a search. These standard terms and metadata are then annotated with external identifiers and ontology terms to improve accessibility to the data.

Most of the data in CEBS today are still served from the CEBS database which provides the underlying data to the user in several ways, through (i) the homepage search for chemical-by-chemical access to data, (ii) through large datasets made available via FTP (file transfer protocol) and (iii) searching and filtering subsets of data through guided search applications and via application programming interfaces (APIs).

DATA CONTENTS

CEBS has grown in the five years since our last update to include five new areas, as shown in Table 1.

An additional 72 studies from general toxicology and carcinogenicity studies have been added (12,13). NTP scientists are continuing to publish on these and other existing data types adding to the collection of these data in CEBS. In addition to data pages related to NTP reports, CEBS now serves data pages containing supplemental data, software, and documentation for manuscripts and special investigations through the CEBS Publications tool; 125 data pages have been made available over the last five years in CEBS Publications. This resource allows authors to share curated 508-compliant data supporting environmental health science and translational research for scientists to access.

In previous iterations of CEBS, the SIFT (Simple Investigation Formatted Text) was required to deposit data as described in our 2007 and 2017 updates. CEBS still accepts this format, but we have also started to accept deposits in the form of flat files that contain collections of data with summary or raw data values and curated metadata. These collections are reviewed for accuracy, integrated into the CEBS repository by associating the results by test article and other metadata, and made available in the NTP Data Collections guided search application (https://cebs.nih.gov/datasets/). In addition to the guided search, the full contents of each data collection are made available via flat files on the FTP and are updated on a regular basis as new data are published; as we develop the planned CEBS Data Ecosystem these files will be moved to that resource for the public to access. Currently, over 30 collections of data are available in this resource (see Supplemental File 1 – CEBS Data Collections).

NTP is responding to requests for summaries of data growth by adding the date of data publication to the report year currently stored in CEBS database; this will permit the growth of CEBS data to be monitored and reported.

USING CEBS

There are three major users of CEBS, as diagrammed in Figure 1. The Supplemental Materials contain additional details of these user experiences.

CEBS was moved to NTP in order to support public access to data underlying NTP reports (use case 1). This user is interested in reviewing the data in a NTP report, for instance TOX-96 (14), data page (15), or other NTP data tables for a chemical in the report.

Other users wish to download all the NTP data for a test type (use case 2), for example results from studies of the genotoxic effects of various chemicals, and upload to their tool for knowledge discovery. These data downloads are available from CEBS via anonymous FTP and also available through HealthData.gov (16).

Other users are seeking to interact with the integrated data in CEBS (use case 3). Recently, NTP has dedicated efforts to merge data from legacy NTP studies into the CEBSR Data Warehouse, described more fully in Data Models, and to make the data searchable by means of an API. The user can also access selected subsets of CEBSR data via the CEBS Data Collections application. An example of using the CEBS Data Collections user interface is described in Supplemental Materials Use Case 3A. In this example, the user can filter to retrieve all (570) significantly changed results for male rats in a family of seven related chemicals; prior to this the user was limited to obtaining these data on a study-by-study basis via the data page. The user can then elect to filter further in CEBS or download for use with tools of their choice.

DATA MODELS

Our planned future state will be to house all data and metadata from CEBS in a new CEBS Data Warehouse, a central integrated repository fed from various source streams, both legacy and evolving. The original data files will be maintained in the CEBS Data Ecosystem. We plan to create CEBS Data Marts containing curated and standardized
Using CEBS (additional details of each use case are provided in Supplemental Materials). The user can 1) review files supporting NTP reports, 2) download pre-defined bulk data for a given data domain or 3) interact with CEBS data using a CEBS Guided Search or an API.

Figure 1. Moving towards this future state, we have created the public CEBSR Data Warehouse in response to stakeholder requests to have summary data and statistical results available for query. CEBSR is a public resource which allows the user access to the summary data from CEBS, and supports comparison of statistically significant responses across test articles, and filtering by species, study type and other metadata, and download of the results. CEBSR currently contains statistically analyzed results for histopathology and numeric endpoints including clinical pathology, organ weights, and reproductive endpoints. CEBSR contains the results of standard NTP statistical analysis (trend/pairwise significance) as well as computed values for NEL (no effect level) and LEL (lowest effect level) for all numeric endpoints. BMD (benchmark dose) (17–20) values are also available for histopathology data. The public can access CEBSR through an API and run pre-defined queries or manipulate query templates to retrieve any or all information in CEBSR. A description of CEBSR, API access, and query templates are available (21,22) and Supplemental Materials Use Case 3C. The CEBS Data Warehouse (now under development) will build on lessons learned from CEBSR.

To support the effort of combining legacy data systems the former CEBS Data Dictionary has been expanded to cover histopathology data (organs, lesions), include more endpoint names and metadata and to start to add harmonized terms for metadata such as study type and data domain. This is renamed to NTP Data Dictionary. We noticed that legacy terms can be difficult to find when searching since the older terms have been replaced over time. In collaboration with members of the Comparative and Molecular Pathogenesis Branch at NTP we are building the NTP Data Dictionary to map legacy terms in CEBS to curated and harmonized terms and to external ontologies and lexicons. The relationship between originally deposited terms or values with the curated and harmonized...
Figure 2. CEBSR Data Pipeline. CEBS supports active data streams which are stored in relational databases and maintains several legacy systems. These data are moved using ETL (extract transform load) processes and through curation efforts to the current CEBSR Data Warehouse. Subsets of the data are moved to individual Data Marts for specific data types or to allow users to answer targeted questions with refined metadata. There are APIs available for the data marts and select CEBS guided search applications are powered by the data marts.

Table 2. CEBSR integrated data sets available from CEBS data collections. Supplemental File 1 – CEBS Data Collections contains the list of all current CEBS data collections

| Name of data collection | Description |
|-------------------------|-------------|
| CEBSR: Statistically Analyzed NTP Pathology Lesions (with incidence, BMD, NEL, and LEL) | Histopathology responses that changed significantly, with comparative statistics (283 832 records) |
| CEBSR: Control NTP Pathology Lesions | Histopathology responses from all control groups, most useful to assess historical incidence (655 738 records) |
| CEBSR: NTP Clinical Pathology Data | Group-level data from clinical pathology measurements; comparative statistics are included (112 308 records) |
| CEBSR: NTP Clinical Pathology Control Data | Group-level data from clinical pathology measurements of control animals, useful to assess historical range in response (20 049 records) |
| NTP Individual Animal Pathology Lesions | Individual animal data with any histopathological finding, with legacy and standard term (3 879 434 records, extracted from CEBS not CEBSR) |

dictionaries will be maintained in the CEBS Data Warehouse. The NTP Data Dictionary is available at https://doi.org/10.22427/NTP-DATA-002--00092-0001--0000-4. This page contains the links between legacy and standard terminology as well as the links between standard terminology for histopathology and clinical pathology to external ontologies.

The original depositions of data in CEBS continue to use the toxicology data model as reported in the first CEBS *Nucleic Acids Research* (NAR) description (1), then termed Biomedical Investigation Database (BID) architecture. A study in CEBS is annotated with experimental details, metadata at the study, group, and subject level, and an association is maintained between every data point from one subject. This is the data model for toxicology studies used in CEBS and has stood the test of time. One improvement upon the data design includes the ability to capture developmental and reproductive toxicology data where animals are treated differently across generations. The system now maintains the relationship between dams and pups, fetuses, and litters. Data are primarily reported on the dam level for pups and litters, and if pups are continued on the study or weaned additional data are collected against the individuals. This relationship provides metadata required to account for F1 generation litter effects (animals in the same litter may be more alike). All data and metadata described for a subject, group, and study are retained when these data are moved to the CEBSR Data Warehouse. Additionally, a unique identifier (UID) is associated with all CEBSR records to allow cross reference to all data for a study, group, or subject and further locate all metadata for a data point.

WEB INTERFACE UPDATES

The CEBS Homepage (https://cebs.niehs.nih.gov/cebs/) was redesigned in 2017 with improvements to enhance access to the data. CEBS Homepage has a search bar for users seeking data from a report, chemical, type of data, or Chemical Abstract number. There are three other categories of tools on CEBS Homepage: NTP Guided Searches (see New Guided Searches), Additional Resources (links to the APIs, CEBS documentation and access to pre-decisional data), and Bulletins which contain FAQs and other references. CEBS data pages themselves have been updated to give a high-level overview of all results available in the database for a given chemical or publication. This is illustrated in Supplemental Materials – Example Test Article Page.
This new Test Article view gives users a ‘one stop shop’ for the information related to a chemical in order to minimize the need to conduct multiple queries, for example NTP evaluations, reports, other publications. All studies of the chemical are accessible from this page, including links to NTP Data Collections with records for the test article (clicking on this takes users to the collection filtered to the test article result), and links to related pages (test articles, NTP projects, or publications). These pages are used by over 2000 users each workday.

The CEBS Homepage is now served from the National Cancer Institute (NCI) Amazon Cloud (AWS). This move allows the application to provide flexible resource levels to support periods of high demand. Stress tests were constructed from this baseline to estimate the resources needed to support this baseline average of CEBS users where the resources were modulated to levels of low / normal / high / very high loads to determine the response rates of the different AWS server sizes to these different loads. NCI uses a scalable server model to allow us to upgrade server instances with more memory when the application is performing slowly due to higher-than-normal use. This model is helpful to maintain good performance when new data of interest are published or immediately following presentations on the tool that may result in unusually high traffic. Resources can be reduced when no longer needed.

NEW GUIDED SEARCHES

Guided search applications in CEBS provide users direct access to curated datasets to query and filter and download data of interest, using a user interface optimized to highlight important information to users (Supplemental Materials Use Case 3B). The DNA Damage Classification Tool permits users to integrate their datasets in real time. The data underlying those applications are also available through CEBS Data Collections.

Two new guided searches have been added since our last update. First, the DNA Damage Classification Tool (https://cebs.niehs.nih.gov/tgxddi/). This tool allows users to upload their own test microarray data to compare their gene signature against the TGx-DDI (toxicogenomic DNA damage inducing) biomarker developed by Health Canada and HESI (Health and Environmental Sciences Institute) (23–26). The application classifies uploaded data into likely DNA damaging mutagens or non-DNA damaging mutagens. The user uploads their microarray data and receives a similarity score to the two classes. Results include fold change, gene cluster, and chemical cluster analysis as well as heatmap and cluster plot images that give the comparison of the user’s uploaded data to the gene signature.

The second new guided search is the Toxicogenomic Benchmark Dose Response application (https://cebs.niehs.nih.gov/tgxbenchmarkdoseresponse/) built to visualize expression results. This tool serves the public with two interactive ways to explore a chemical/tissue gene expression response: (i) in an accumulation plot highlighting the most affected biological processes as identified using GO (Gene Ontology) biological pathways and (ii) a filterable table with the details of the most affected biological processes, with overall up or down regulation. Source data files, coordinate files for the accumulation plots, and useful links are also available from each page.

FAIR DATA STANDARDS

FAIR guiding principles were developed to provide a guideline for data management teams to enhance the access to data for computational tools and enhance reusability of data for the scientific community (27). The CEBS database is governed under the FAIR principles to make our data more Findable, Accessible, Interoperable, and Reusable.

CEBS data are findable

Data in CEBS are made findable with the addition of DOIs (digital object identifiers). A DOI is an immutable identifier that can be assigned at various levels of digital information, CEBS has DOIs for the data found in homepage test article pages, data supporting publications, and collections of data in NTP Data Collections. These identifiers are registered with an external source (CrossRef) which ensures the information are persistent and unique. We have used the DOIs in this publication to ensure persistence of the links.

Another criterion for ‘findable’ data is the availability of rich, clearly defined, registered metadata. As described in Data Models, the CEBS team along with NTP scientists re-curated terms and endpoint names used in legacy studies and aligned to modern terminology. Additionally, we harmonized terms used to describe study, subject, and group metadata. The units used in assays can change with the depositor or data stream so we identified units used now by NTP and aligned any units used in CEBS by adding the conversion equations to CEBS. Numeric data in CEBS are deposited with their original units and then mapped to CEBS standard units for the assay or measurement so that all data can be presented in the same units without requiring user adjustment. The IDs used in CEBS warehouse permit traceability from the values with standard units to the unit and values in the original deposition.

CEBS data are accessible

Data accessibility in CEBS has come a long way in the last five years. Data are still available in flat files but the addition of APIs to CEBSR provides computational tools a more straightforward approach to linking or ingesting CEBS data into applications and larger data repositories. Data for each data domain in CEBS are also aggregated into a flat file which is accessible from the CEBS FTP site (https://cebs.niehs.nih.gov/ftp/) or HealthData.gov (16) (Supplemental Materials Use Case 2).

The NTP Data Collections application also provides increased availability for users by providing direct access to the data to permit the user to ask questions of integrated datasets. These resources are more straightforward for users to find now with the homepage redesign (Supplemental Materials – Example Test Article Page). CEBS Homepage now integrates all results into the CEBS test article pages, over 9000 web pages containing data and other links for each test article studied by NTP. These test article pages contain the
NTP conclusion and other decisions, links to NTP publications on that test article, the data for each study performed by NTP, as well as links to relevant CEBS data collections, NTP research projects, and other CEBS resources. The test article pages also contain links to the protocol for each NTP study type which contains more details on how the assays are conducted, along with references to the specification documentation that labs use to collect the information.

Lastly, CEBS data will remain accessible through the support of NTP and NIEHS and will continue to grow and improve.

CEBS data are interoperable and reusable

The interoperability and reusability of CEBS data also received a major upgrade with extensive efforts to link NTP data to external authorities. Test articles studies by NTP have been aligned to EPA DTXSIDs, the substance identifiers provided by the USEPA (28,29). In addition, NTP supports the TRUST principles for digital repositories (33): Transparency, Responsibility, User focus, Sustainability and Technology, which provide a baseline for digital repository trustworthiness. The plan is to make CEBS fully TRUST-worthy over the next few years. CEBS already meets the standards for Sustainability and Technology since it is hosted in a government data center and has support from NIEHS and NTP.

Since our last update in Nucleic Acids Research (3) CEBS has grown with the addition of 25% more studies and 15% more test articles available in the repository. In addition to the expanded data availability, CEBS has greatly improved accessibility to the data so as to allow the public to ask more targeted questions of highly curated NTP toxicology datasets. Additionally, the CEBS team is striving to enhance user experience by creating APIs that allow users computational access to large datasets of NTP data. Recognizing not all of our users are familiar or comfortable with large data handling tools we have built CEBS Data Collections to allow users to browse the data in CEBS Data Marts and access the extensive toxicology data housed in the NTP databases for knowledge-discovery and applications to improve human health. Taken together, the growing knowledge base developed and cultivated in CEBS continues to further our mission to improve toxicology data availability and accessibility for the scientific community.

DATA AVAILABILITY

The public repository is available from the NTP website: https://cebs.niehs.nih.gov/cebs/, also accessible via https://doi.org/10.22427/NTP-DATA-1. The database is registered with NIEHS and accessioned under NTP-DATA-1.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank NTP for data, direction and support, and the CEBS curation and development teams who are stewards for the data, supported the use of the tools and created novel applications to support the mission of the NTP. We are also grateful for very helpful comments from anonymous reviewers.

FUNDING

ASRC Federal under National Institute of Environmental Health Sciences contracts ‘Support for Toxicological Data for the National Toxicology Program’ [HHSN31620120054W and HHSN273201000063U GS06F0629Z]. Funding for open access charge: NIEHS.
Conflict of interest statement. None declared.

REFERENCES

1. Waters, M., Boorman, G., Busnel, P., Cunningham, M., Irwin, R., Merrick, A., Olden, K., Paules, R., Selkirk, J., Statiecwicz, S. et al. (2003) Systems toxicology and the Chemical Effects in Biological Systems (CEBS) knowledge base. *EHP Toxicogenomics*, 111, 15–28.

2. Waters, M., Stassiewicz, S., Merrick, B.A., Tomer, K., Busnel, P., Paules, R., Stegman, N., Nehls, G., Yost, K.J., Johnson, C.H. et al. (2008) CEBS—chemical effects in biological systems: a public data repository integrating study design and toxicity data with microarray and proteomics data. *Nucleic Acids Res.*, 36, D982–D990.

3. Lea, I.A., Gong, H., Paleja, A., Rashid, A. and Fostel, J. (2017) CEBS: a comprehensive annotated database of toxicological data. *Nucleic Acids Res.*, 45, D964–D971.

4. Gaater, B., Tugendreich, S., Pearson, C.L., Ayanoglu, E., Baumbueter, S., Bostian, K.A., Brady-L., Brown, J.L., Calvin, J.T., Day, G-J. et al. (2005) Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. *J. Biotechnol.*, 119, 219–244.

5. Roter, A.H. (2005) Large-scale integrated databases supporting drug discovery. *Curr. Opin. Drug Discov. Devel.*, 8, 309–315.

6. National Toxicology Program (2021) Developmental & Reproductive Toxicity. National Toxicology Program (NTP), Research Triangle Park, NC (USA), https://ntp.niehs.nih.gov/whatstestudy/testpgm/devrepro/index.html, (05 October 2021, date last accessed).

7. National Toxicology Program (2021) Specifications for the Conduct of Studies to Evaluate the Reproductive and Developmental Toxicity of Chemical, Biological and Physical Agents in Laboratory Animals for the National Toxicology Program (NTP). National Toxicology Program (NTP), Research Triangle Park, NC (USA), https://ntp.niehs.nih.gov/ntp/test_info/finalntp_reprospect Kasım2011_508.pdf, (05 October 2021, date last accessed).

8. National Toxicology Program (2021) Chemical Disposition & Toxicokinetics. National Toxicology Program (NTP), Research Triangle Park, NC (USA), https://ntp.niehs.nih.gov/whatstestudy/testpgm/chemdisp/index.html, (05 October 2021, date last accessed).

9. National Toxicology Program (2021) Genetic Toxicology. National Toxicology Program (NTP), Research Triangle Park, NC (USA), https://ntp.niehs.nih.gov/whatstestudy/testpgm/ genetic/index.html, (05 October 2021, date last accessed).

10. National Toxicology Program (2021) Levels of Evidence Criteria for NTP Immunotoxicology. National Toxicology Program (NTP), Studies Research Triangle Park, NC (USA), https://ntp.niehs.nih.gov/ntp/test_info/germolecitoxcriteriaasdfinal_s08.pdf, (05 October 2021, date last accessed).

11. National Toxicology Program (2021) NTP Research Report on National Toxicology Program Approach to Genomic Dose-Response Modeling. National Toxicology Program (NTP), Research Triangle Park, NC (USA), https://ntp.niehs.nih.gov/ntp/results/pubs/rr/reports/rr05_508.pdf, (05 October 2021, date last accessed).

12. National Toxicology Program (2021) Toxicology / Carcinogenicity. National Toxicology Program (NTP), Research Triangle Park, NC (USA), https://ntp.niehs.nih.gov/whatstestudy/testpgm/cartox/index.html, (05 October 2021, date last accessed).

13. National Toxicology Program (2021) Specifications for the Conduct of Studies to Evaluate the Toxic and Carcinogenic Potential of Chemical, Biological and Physical Agents in Laboratory Animals for the National Toxicology Program (NTP). National Toxicology Program (NTP), Research Triangle Park, NC (USA), https://ntp.niehs.nih.gov/ntp/test_info/finalntp_toxarspecsan2011.pdf, (05 October 2021, date last accessed).

14. National Toxicology Program (2021) Abstract for TOX-96. National Toxicology Program (NTP), Research Triangle Park, NC (USA), https://ntp.niehs.nih.gov/publications/reports/tox/tox00/tox96index.html?utm_source=direct&utm_medium=prod&utm_campaign=ntpgolinks&utm_term=tox096eabs, (05 October 2021, date last accessed).

15. National Toxicology Program (2021) TOX-96: Toxicity Report Tables & Curves. National Toxicology Program (NTP), Research Triangle Park, NC (USA), https://cebs.niehs.nih.gov/cebs/data/publication/TOX-96, (05 October 2021, date last accessed).

16. HealthData.gov (2021) *Chemical Effects in Biological Systems* (CEBS). https://healthdata.gov/dataset/Chemical-Effects-in-Biological-Systems-CEBS-pdfs-qnic, (05 October 2021, date last accessed).

17. U.S. Environmental Protection Agency (2012) Benchmark Dose (BMD) Technical Guidance Document. U.S. Environmental Protection Agency, Washington, DC, https://www.epa.gov/risk/benchmark-dose-technical-guidance, (05 October 2021, date last accessed).

18. U.S. Environmental Protection Agency (2005) *Guidelines for Carcinogen Risk Assessment*. https://www.epa.gov/risk/guidelines-carcinogen-risk-assessment, (05 October 2021, date last accessed).

19. Crump, K. (2002) Critical issues in benchmark calculations from continuous data. *Crit. Rev. Toxicol.*, 32, 133–153.

20. Crump, K. (1984) A new method for determining allowable daily intakes. *Fundam. Appl. Toxicol.*, 4, 854–871.

21. National Toxicology Program (2021) CEBSR API. National Toxicology Program (NTP), Research Triangle Park, NC (USA), https://manticore.niehs.nih.gov/cebssearch/cebsr, (05 October 2021, date last accessed).

22. National Toxicology Program (2021) CEBSR Data Warehouse API Documentation and Examples. National Toxicology Program (NTP), Research Triangle Park, NC (USA), https://doi.org/10.22427/NTP-DATA-002-00091-0001-0000-3, (05 October 2021, date last accessed).

23. Buick, J.K., Moffat, I., Williams, A., Swartz, C.D., Recio, L., Hyduke, D.R., Li, H.-H., Fornace, A.J. Jr, Aubrecht, J. and Yauk, C.L. (2015) Integration of metabolic activation with a predictive toxicogenomics signature to classify genotoxic versus nongenotoxic chemicals in human TK6 cells. *Environ. Mol. Mutagen.*, 56, 520–534.

24. Cho, E., Buick, J.K., Williams, A., Chen, R., Li, H.-H., Corton, J.C., Fornace, A.J. Jr, Aubrecht, J. and Yauk, C.L. (2019) Assessment of the performance of the TGx-DDI biomarker to detect DNA damage-inducing agents using quantitative RT-PCR in TK6 cells. *Environ. Mol. Mutagen.*, 60, 122–133.

25. Jackson, M.A., Yang, L., Lea, I., Rashid, R., Kuo, B., Williams, A., Yauk, C.L. and Fostel, J. (2017) The TGx-28.65 biomarker online application for analysis of transcriptomics data to identify DNA damage-inducing chemicals in human cell cultures. *Environ. Mol. Mutagen.*, 58, 529–535.

26. Li, H.H., Hyduke, D.R., Chem, R., Heard, P., Yauk, C.L., Aubrecht, J. and Fornace, A.J. Jr (2015) Development of a toxicogenomics signature for genotoxicity using a dose-optimization and informatics strategy in human cells. *Environ. Mol. Mutagen.*, 56, 505–519.

27. Wilkinson, M.D., Dumont, M., D’Ambrosio, I.J., Appleton, G., Axtor, M., Baak, A., Blomberg, N., Boiten, J-W., Bonino da Silva Santos, L., Bourne, P.E. et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018.

28. U.S. Environmental Protection Agency (2021) CompTox Chemicals Dashboard. U.S. Environmental Protection Agency (EPA), Research Triangle Park, NC (USA). https://comptox.epa.gov/dashboard, (05 October 2021, date last accessed).

29. Williams, A.J., Gruke, C.M., Edwards, J., McEachran, A.D., Mansouri, K., Baker, N.C., Patlewicz, G., Shah, I., Wambaugh, J.F., Jusdon, R.S. and Richard, A.M. (2017) The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. *J. Cheminform.*, 9, 61.

30. Open Biological and Biomedical Ontology (2021) *The Open Biological and Biomedical Ontology (OBO) Foundry*. OBO Technical Working Group, http://www.obofoundry.org/, (05 October 2021, date last accessed).

31. National Toxicology Program (2021) DNTP Pathology Dictionary. National Toxicology Program (NTP), https://cebs.niehs.nih.gov/cebs/paper/15179, (05 October 2021, date last accessed).

32. Collapudi, B.B., Lynch, A.M., Helfich, R.H., Dertinger, S.D., Dobrovolsky, V.N., Froetschl, R., Horibata, K., Kenyon, M.O.,
Kimoto, T., Lovell, D. P. et al. (2015) The in vivo Pig-a assay: a report of the International Workshop on Genotoxicity Testing (IWGT) Workgroup. *Mutat Res. Genet. Toxicol. Environ. Mutagen.*, 783, 23–35.

Lin, D., Crabtree, J., Dillo, I., Downs, R. R., Edmunds, R., Giaretta, D., De Gusti, M., L’Hours, H., Hugo, W., Jenkyns, R. et al. (2020) The TRUST Principles for digital repositories. *Scientific Data*, 7, 144.