

Comparative Toxicogenomics Database's 20th anniversary: update 2025

Allan Peter Davis^{1,*}, Thomas C. Wiegiers¹, Daniela Sciaky¹, Fern Barkalow¹, Melissa Strong¹, Brent Wyatt¹, Jolene Wiegiers¹, Roy McMorran¹, Sakib Abrar¹ and Carolyn J. Mattingly^{1,2}

¹Department of Biological Sciences, 3510 Thomas Hall, 112 Derieux Place, North Carolina State University, Raleigh, NC 27695, USA

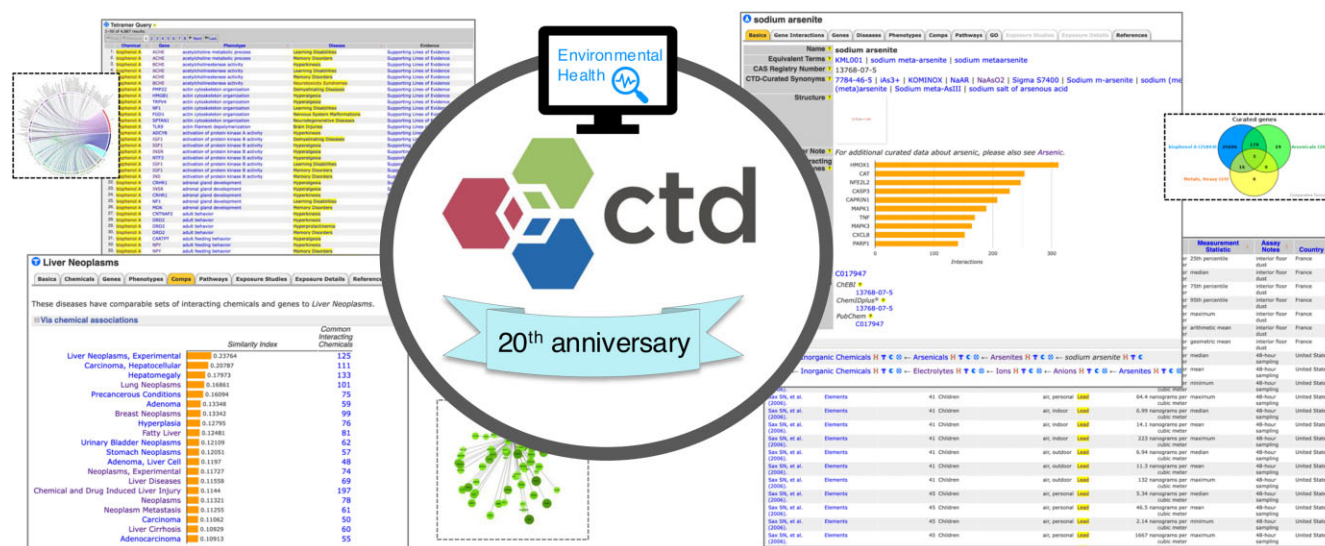
²Center for Human Health and the Environment, 850 Main Campus Drive, North Carolina State University, Raleigh, NC 27695, USA

*To whom correspondence should be addressed. Tel: +1 919 515 5705; Fax: +1 919 515 3355; Email: apdavis3@ncsu.edu

Abstract

For 20 years, the Comparative Toxicogenomics Database (CTD; <https://ctdbase.org>) has provided high-quality, literature-based curated content describing how environmental chemicals affect human health. Today, CTD includes over 94 million toxicogenomic connections relating chemicals, genes/proteins, phenotypes, anatomical terms, diseases, comparative species, pathways and exposures. In this 20th year anniversary update, we reflect on CTD's remarkable growth and provide an overview of the increased data content and new features, including enhancements to the curation workflow (e.g. new exposure curation tool and expanded use of natural language processing), added functionality (e.g. improvements to CTD Tetramers and Pathway View tools) and significant upgrades to software and infrastructure. Linking lab-based core curation with real-world human exposure curation via the use of controlled vocabularies facilitates analysis of content across the entire environmental health continuum, from molecular toxicological mechanisms to the population level, and vice versa. The 'prototype database' originally described in 2004 has evolved into a premier, sophisticated, highly cited and well-engineered knowledgebase and discoverybase that is utilized by scientists worldwide to design testable hypotheses about environmental health.

Graphical abstract



CTD's 20th year anniversary

The Comparative Toxicogenomics Database (CTD; <https://ctdbase.org>) was launched 20 years ago (12 November 2004) as a novel 'prototype' public resource that imported and integrated data from the scientific literature relating nucleotide/protein sequences, environmental toxicants and taxa to explore toxicogenomic relationships from comparative species (1). Since that humble beginning, CTD has evolved and flourished into a highly cited, robust and so-

phisticated digital ecosystem. As of August 2024, CTD includes 3.8 million direct interactions manually curated from >149 000 scientific articles (Figure 1A), interrelating >17 700 chemicals, 55 400 genes, 6700 phenotypes, 7200 diseases, 214 000 exposures and 980 anatomical terms for over 630 species. These direct interactions are integrated (see below) to generate an additional 48 million inferred relationships (Figure 1B). Combining these data with imported gene Gene Ontology annotations (2), gene pathways from

Received: August 30, 2024. Editorial Decision: September 19, 2024. Accepted: September 25, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

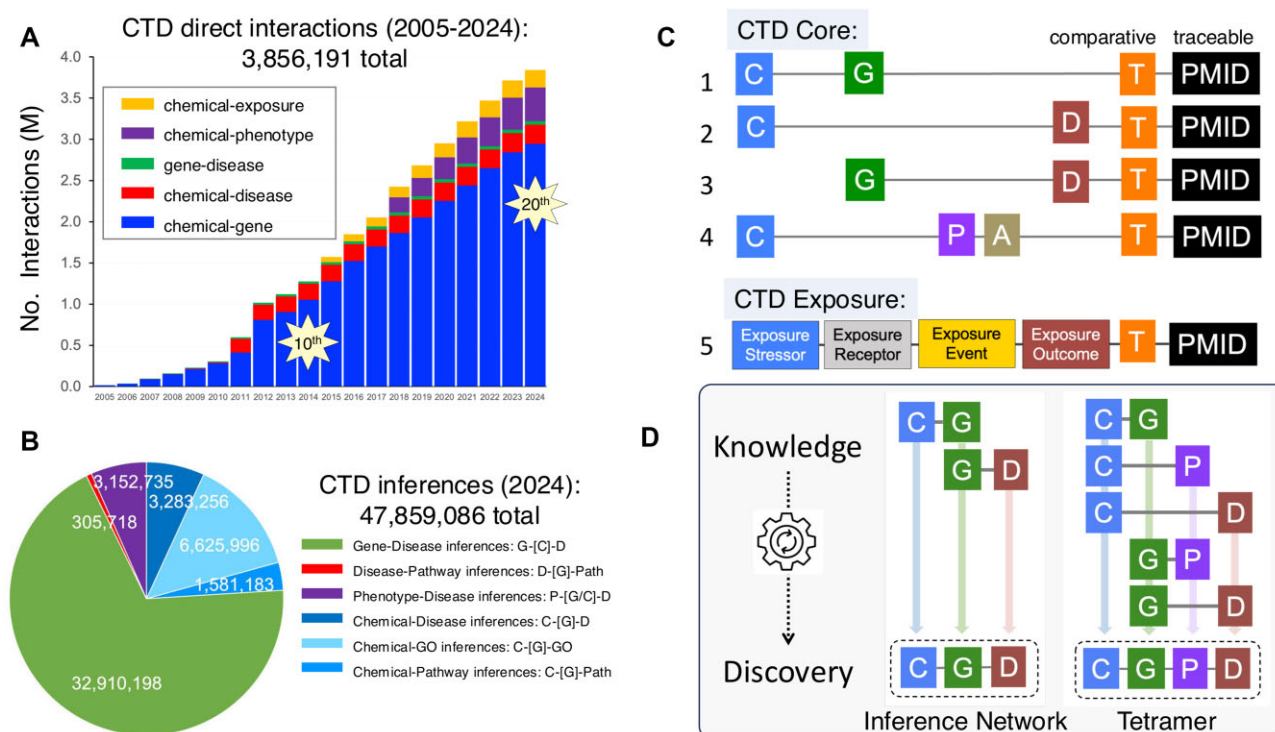


Figure 1. (A) Over the last 20 years, CTD has grown to >3.8 million manually curated direct interactions (y-axis in millions) and has expanded with new content (chemical-exposure statements in 2015 and chemical-phenotype interactions in 2018). CTD's 10th and 20th anniversaries are indicated, but counts for year 2024 are incomplete. (B) Integration of the 3.8 million direct interactions generates almost 48 million inferences, the majority of which (over 32 million) are for gene-disease inferences via a shared chemical: G-[C]-D. Phenotype-disease inferences can be made via either shared genes or shared chemicals and are reported together, as such: P-[G/C]-D. CTD is updated on a monthly basis and the most current counts are available at <https://ctdbase.org/about/dataStatus.go>. (C) CTD biocurators manually curate five types of interactions from the scientific literature. CTD Core captures four relationships between chemicals (C), genes (G), phenotypes (P), anatomical terms (A) and diseases (D). Interactions are contextualized with taxa (T), making the database comparative, and are traceable to source articles by PubMed identifiers (PMID). For simplicity, the 50 predicates and 15 qualifiers are not shown. In CTD Exposure, biocurators construct detailed statements using 20 controlled vocabularies to link 33 distinct data concepts for the four exposure categories relating real-world environmental measurements: exposure stressor, exposure receptor, exposure event and exposure outcome. (D) CTD uses data integration to transform curated knowledge into potential discoveries. Chemical-gene and gene-disease curated interactions are combined to generate 'Inference Networks' that connect chemicals to diseases by inferred gene intermediates. Similarly, five curated interactions are integrated to compute a 'Tetramer' that links a chemical, gene, phenotype and disease in a potential stepwise mechanistic pathway.

Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome (3,4) and gene-gene interactions from BioGRID (5) results in 94 million toxicogenomic relationships in CTD (<https://ctdbase.org/about/dataStatus.go>). This vast data landscape helps address knowledge gaps and enables construction of testable mechanistic hypotheses for environmental health. Here, we reflect on the historical uniqueness, complexity and utility of CTD as well as present our most recent updates.

One of CTD's first milestones was to design a curation paradigm that enabled biocurators to read the literature and manually code interactions using formatted, structured notation and controlled vocabularies and ontologies to capture any type of a chemical-induced biological effect (6,7). In the CTD Core module, biocurators collect four types of direct interactions from scientific articles: chemical-gene/protein interactions, chemical-phenotype interactions, chemical-disease associations and gene-disease associations (Figure 1C). This paradigm is distinct in three important ways. First, CTD does not limit data collection to any single organism, but captures toxicogenomic information for all Eumetazoa. Comparing toxicology data across diverse species was recognized early on as a critical factor to better understand toxicity mechanisms, susceptibility and drug safety

(8,9); however, when CTD was launched in 2004, most publicly curated data were siloed in independent single model organism databases, making cross-species comparison cumbersome. CTD addresses that shortcoming by capturing toxicogenomic data for any eumetazoan species encountered in an article. Second, CTD eschews the use of free text, requiring annotations to be constructed using controlled vocabularies and ontologies that align with CTD's data FAIRness policy (<https://ctdbase.org/about/ctdDataFairness.jsp>); this ensures that chemical, gene, phenotype, disease, anatomy, interaction and taxa information are standardized across all curated articles, no matter where or when the paper was originally published (e.g. the oldest paper curated in CTD is from 1947). Third, CTD interactions are formatted using a sentence structure (e.g. 'bisphenol A results in increased cleavage of CASP3 protein in *Mus musculus*'). Here, controlled terms are organized in a structured format to create a chemical-induced biological statement, including a subject (bisphenol A), predicate (increased cleavage) and direct object (CASP3 protein), all contextualized with a taxon (*M. musculus*). This paradigm ensures that cross-species heterogeneous data now become standardized, harmonized, centralized and interoperable, as well as facilitates the emergence of systems toxicology networks (10).

Ten years later (March 2015), CTD increased its complexity, uniqueness and utility by launching another novel module called CTD Exposure (11). This module was designed to help characterize the human exposome, report real-world chemical stressor and biomarker measurements, and address the scientific community's need to couple the exposome concept to mechanistic toxicology (12). In CTD Exposure, biocurators critically read complex exposure articles and utilize over 20 vocabularies to construct a detailed statement (composed of up to 33 unique data fields) interrelating an exposure stressor, exposure receptor, exposure event and exposure outcome (Figure 1C), which reflect the four main concepts of the exposure ontology (13). This module reports the real-world measurements of environmental toxicants and biomarkers with their associated adverse outcomes, collecting detailed information about the chemical stressor and its source, the receptor demographics (e.g. cohort size, age distribution, gender, smoking status, race and ethnicity), the event particulars (e.g. biomarker levels, assayed medium, statistics, time frame, country, state and project title) and potential outcomes (e.g. phenotypes, diseases and statistical strength). Importantly, we use the same controlled vocabularies (chemical, gene, phenotype, anatomy and disease) for both CTD Core and CTD Exposure curation efforts, enabling seamless integration and complementarity. CTD Core curation (from laboratory experiments) can help inform potential mechanistic pathways for environmental health, and CTD Exposure content (from real-world reporting) informs hypotheses for additional laboratory testing and validation. Together, the two interoperable modules empower users to explore environmental health content from populations to molecules, and vice versa.

CTD's second milestone was to develop data integration strategies that transform this curated knowledgebase into a toxicogenomic discoverybase (Figure 1D). Because CTD's manually curated interactions are constructed using controlled vocabularies, the datasets can now be integrated with one another to generate novel discoveries. For example, chemical-gene and gene-disease direct interactions are combined to produce 'inferences' following Swanson's ABC model for knowledge transfer (14): If chemical A interacts with gene B, and gene B is associated with disease C, then chemical A has an inferred relationship to disease C, via gene B (15). In a similar approach, five independently curated statements from CTD can be integrated to generate a 'CGPD tetramer': a mechanistic construct that links an initiating chemical with an intermediate gene to modulate a key biological phenotype that can play a role in a disease (16). These inferred and computationally derived relationships provide users with potential molecular mechanisms that can be assembled to construct chemical-disease pathways by filling the knowledge gaps between an exposure and outcome. Thus, CTD functions as both a knowledgebase (directly curated interactions from the literature) and a discoverybase (integration-derived predictions that can potentially fill knowledge gaps).

Paralleling this increased growth in curated content, CTD's software engineering, technological perspectives and infrastructure have also grown significantly over the last 20 years (Table 1). We have made considerable investments and upgrades in hardware and software development, with a more robust software engineering team.

To aid users in exploring the CTD landscape, we provide a comprehensive suite of search menus (<https://ctdbase.org/search/>), analytical tools (<https://ctdbase.org/tools/>) and

Table 1. CTD computing upgrades from 2004 to 2024

Description	2004	2024	Fold increase (as applicable)
Operating system	Solaris 9	Rocky Linux 8.9	
Database	Oracle 10	PostgreSQL 14	
management system			
CPU cores	1	32	32
CPU clock	750 MHz	3.6 GHz	4.8
RAM	2 GB	768 GB	384
Storage (total)	432 GB	45 TB	107
Computer room (CR)	None (basement)	NCSU Eastern Data Center	
CR HVAC	None	Full	
CR fire suppression	None	Full	
CR wildlife	Spiders, snakes	None (that we know of)	

downloadable files (<https://ctdbase.org/downloads/>). Additionally, we have published 50 peer-reviewed publications (<https://ctdbase.org/about/publications/#ctdpubs>) and presented at 78 national and international conferences (<https://ctdbase.org/about/publications/#ctdpres>) to promote CTD's curation paradigms, tools, features, data content and functionalities, with numerous case studies to demonstrate how users can employ CTD to ask, address and answer scientific questions related to environmental health.

Over the last two decades, CTD has evolved into an authoritative 'golden set database' (17), acting as a comprehensive resource for the scientific community. Currently, CTD has over 8000 cumulative Google Scholar citations and a citation rate of three per day since 2023. Furthermore, 227 external databases link to or utilize CTD information (<https://ctdbase.org/about/publications/#use>), helping to further disseminate CTD data.

New features

Curation updates: natural language processing and new Exposure Curation Tool

CTD curation is limited to scientific publications indexed in PubMed (<https://pubmed.ncbi.nlm.nih.gov/>). To prioritize articles for curation, we use a complementary, two-pronged process (chemical-centric and targeted journals) that promotes data completeness and increases data currency, as previously described in detail (6,18,19). For a large corpus of selected articles, we also leverage our text-mining algorithm to score and rank the articles for curation, which we have shown increases both productivity and efficiency (19).

To further enhance and streamline curation, as well as to extend our use of text mining (20), we have now fully incorporated the natural language processing (NLP) capabilities provided by PubTator (21). CTD biocurators access scientific articles using the PubTator interface, where abstracts are illuminated with color-coded terms for chemicals, genes, diseases and species by PubTator's NLP (Figure 2). Additionally, these NLP-identified terms are summarized as a report and imported directly into the CTD Curation Tool data entry fields for biocurators to select using a pick-list when curating (Figure 2).

The CTD Exposure module continues to expand and now contains 214 000 exposure statements relating 1500 exposure stressors, 1400 human genes and 1000 exposure outcomes (500 phenotypes and 500 environmental diseases), manually

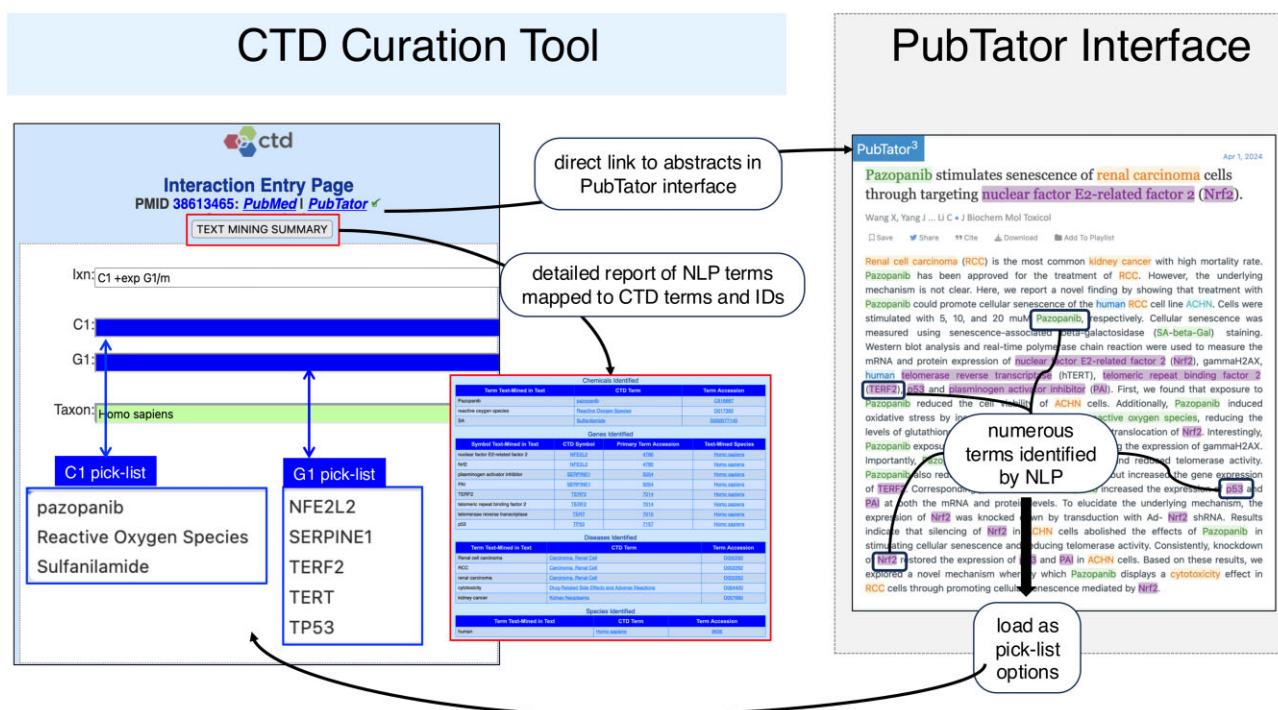


Figure 2. From the CTD Curation Tool, biocurators link to abstracts in the PubTator interface that display the color-coded terms for chemicals, genes, diseases and taxa identified by NLP. A detailed text-mining summary report shows the NLP terms mapped to their corresponding CTD controlled vocabularies and accession identifiers. Additionally, the CTD terms are loaded as pick-list options in the CTD Curation Tool [C1, chemicals; G1, genes; D1, diseases (not shown)] for biocurators to select when entering interactions in the tool, streamlining the curation process.

curated from over 3400 exposure articles (<https://ctdbase.org/about/dataStatus.go>). This module requires its own data collection platform, distinct from the CTD Core Curation Tool that has been operational since 2010 (6). We have now implemented a new CTD Exposure Curation Tool (ECT), which enables biocurators to directly submit their curated annotations derived from exposure articles (11). The new online ECT ensures both high-quality content and consistency as it automatically performs dozens of quality control checks in real time to validate data entry, enabling biocurators to immediately fix any detected error or conflicting value at the point of curation.

Tool updates: CTD tetramers and pathway view

CGPD tetramers (Figure 1D) provide insight into potential molecular mechanisms and can be assembled to construct complex chemical-induced pathways that have been used to help fill knowledge gaps and design adverse outcome pathways (AOPs) for a diversity of environmental health studies (22–26). In 2022, we released the online tool *CTD Tetramers* that empowers users to generate tetramers for any phenotype or disease of interest (16). We have now enhanced this tool (<https://ctdbase.org/query.go?type=tetramer>) in three significant ways. First, users now can include chemicals and/or genes in their tetramer queries, in addition to phenotypes or diseases. This allows users to search for targeted datasets for any of the four components of a tetramer (i.e. chemical, gene, phenotype and/or disease), or a combination of components, such as how heavy metals can be linked to Parkinson disease via a metabolism-related phenotype (Figure 3A). This enhancement provides greater querying power to discover specific sets of tetramers that can be used to help construct and refine chemical-induced pathways and AOPs. Second, we now

include an ‘Evidence’ column for each tetramer that shows the five supporting lines of evidence (and their source articles) used to generate the tetramer (Figure 3A). This transparency allows users to evaluate the strength of the evidence underlying each tetramer. Third, we provide a method to transform tabular tetramer query results into a novel chord diagram to help visualize and identify potential key molecular mechanisms (27). Chord diagrams provide an elegant visual representation of the tetramers connecting a chemical exposure to an adverse outcome, which can be used to understand the possible underlying mechanisms and guide AOP models.

Additionally, we have updated CTD’s Pathway View functionality, which uses web-based Cytoscape (28) for visualizing gene–gene networks for gene sets computed as either CTD ‘Gene Inference Networks’ (Figure 3B) or input directly using the CTD Set Analyzer tool (<https://ctdbase.org/tools/analyzer.go>). The modernized Pathway View maps are customizable with five new preset graphing layouts. In addition, users can drag nodes, remove node and edge labels, change node colors, zoom and pan, and export maps in two formats (PNG or JPG). This new functionality leverages the JavaScript implementation of the Cytoscape library, which enables layout management, viewport adjustments and other graph operations. To support additional physics simulation layouts with animation, two extension libraries have also been integrated into CTD’s Jakarta EE-based JSP/Servlet architecture.

Other enhancements: API and infrastructure upgrades

We improved the visibility and functionality of CTD’s application programming interfaces (APIs) to address the increased interest from external resources to retrieve CTD content computationally. An explicit link to our API is now found in the

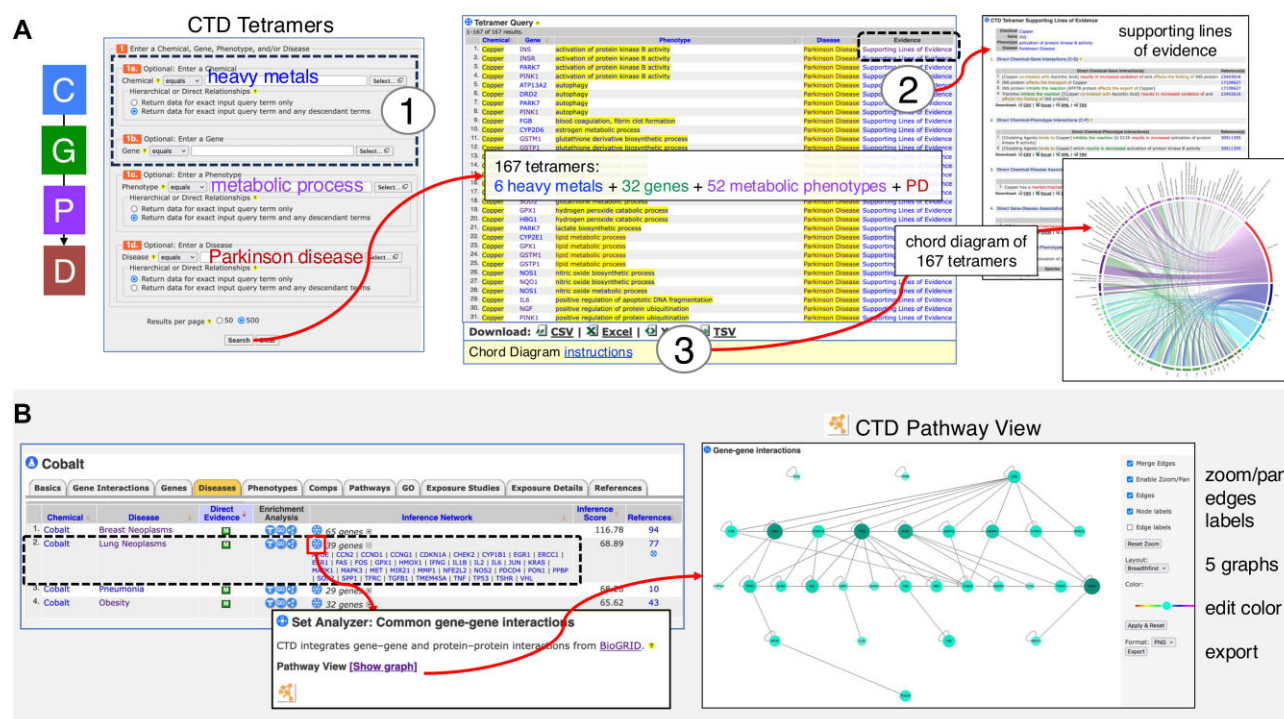


Figure 3. (A) Three enhancements have been added to the CTD Tetramers tool (<https://ctdbase.org/query.go?type=tetramer>) that is used to generate CGPD tetramers for constructing chemical–disease mechanistic pathways. Now users can include a chemical and/or gene (dotted black box), in addition to a phenotype and disease, to the query, enabling focused searches for specific molecular events (circle 1). Here, a multi-parameter query searches for CGPD–tetramers that can connect ‘heavy metals’ (chemical) to ‘Parkinson disease’ (disease) via an intermediate ‘metabolic process’ phenotype. Optional filters are selected to retrieve hierarchical descendant terms for both the chemical and phenotype terms. This query retrieves 167 CGPD–tetramers, composed of 6 heavy metals, 32 genes and 52 metabolic process phenotypes linked to Parkinson disease. A new ‘Evidence’ column (circle 2) provides the five supporting lines of evidence (and their source articles) used to construct the tetramer. The tabular output can be transformed into a chord diagram following a set of user-friendly instructions (circle 3); chord diagrams help visualize key mechanisms. **(B)** Gene–gene interaction maps can be easily generated from CTD ‘Gene Inference Networks’. Here, the metal cobalt is inferred to lung neoplasms via 39 genes (black dotted box). Clicking the tool icon (in front of the gene list) sends the genes to the Set Analyzer tool and constructs an interaction map (based upon BioGRID data), which can be seen by clicking the Pathway View link. The tool has been updated with new customizable features, including the ability to adjust edge and label displays, zoom/pan, change node colors, select from five graph layouts and export maps. Alternatively, users can input their own gene list using the Set Analyzer tool directly (<https://ctdbase.org/tools/analyzer.go>) to generate any gene–gene interaction map.

base menu options under the ‘Analyze’ tab of every CTD web page (<https://ctdbase.org/help/linking.jsp#batchqueries>). As well, the API has been updated to allow extraction of both phenotype-based curation and phenotype–disease inferences.

To address the increased content within CTD and provide the computational capacity necessary to support future needs, significant upgrades were made to CTD’s infrastructure. Existing computer hardware was replaced with three new servers that have dual 3.6 GHz 16-core processors and 768 GB memory capacity (Table 1). We also migrated to a new 45 TB NetApp all-flash (i.e. solid-state drive) storage array. These upgrades coincided with a physical move to NCSU’s new state-of-the-art facility Eastern Data Center with its much-improved 25 Gb/s network. In this updated infrastructure, we also migrated to new server operating systems, PostgreSQL database management systems, Tomcat application servers, and the latest versions of third-party libraries and software.

Summary

- Over the last 20 years, CTD has evolved into a premier toxicogenomic digital ecosystem, with new curation modules, features and tools to support the needs of the scientific community.

- Manually curated CTD content increased by 13% to >3.8 million direct interactions since our last update; when integrated with other data types, CTD now generates over 94 million toxicogenomic relationships (<https://ctdbase.org/about/dataStatus.go>).
- Terms identified by NLP from PubTator are imported into the CTD Curation Tool to enhance curation processes.
- A new online ECT is now used in our curation process to increase efficiency and accuracy.
- CTD Tetramers tool was enhanced with features allowing for more sophisticated querying, increased transparency and traceability, and the ability to visualize results as chord diagrams.
- Pathway View tool used to map gene networks was updated with new functionalities.
- Visibility and function of CTD’s API data retrieval system were increased.
- Significant upgrades and investments to CTD software engineering and technical infrastructure are reported.

Data availability

CTD is publicly available at <https://ctdbase.org>. All aspects of CTD are freely available to non-commercial users; commer-

cial users are asked to purchase a license in order to download CTD data files to their systems. To cite CTD, please see <https://ctdbase.org/about/publications/#citing>. If interested in establishing links to CTD data, please notify us (<https://ctdbase.org/help/contact.go>) and follow the instructions at <https://ctdbase.org/help/linking.jsp>. External resources using CTD content are promoted at <https://ctdbase.org/about/publications/#use>.

Acknowledgements

We are grateful for all the contributions from former CTD team members over the last 20 years. We also thank our past and present internal advisors (Drs Antonio Planchart and David Reif) and our External Advisory Committee. Finally, we thank our loyal users and the numerous anonymous reviewers of our manuscripts for helpful feedback to improve the content and utility of CTD. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Funding

National Institute of Environmental Health Sciences [U24 ES033155]. Funding for open access charge: National Institute of Environmental Health Sciences [U24 ES033155].

Conflict of interest statement

None declared.

References

- Mattingly,C.J., Colby,G.T., Rosenstein,M.C., Forrest,J.N. and Boyer,J.L. (2004) Promoting comparative molecular studies in environmental health research: an overview of the Comparative Toxicogenomics Database (CTD). *Pharmacogenomics J.*, **4**, 5–8.
- Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T., *et al.* (2000) Gene Ontology: tool for the unification of biology. *Nat. Genet.*, **25**, 25–29.
- Kanehisa,M., Sato,Y., Furumichi,M., Morishima,K. and Tanabe,M. (2019) New approach for understanding genome variations in KEGG. *Nucleic Acids Res.*, **47**, D590–D595.
- Gillespie,M., Jassal,B., Stephan,R., Milacic,M., Rothfels,K., Senff-Ribeiro,A., Griss,J., Sevilla,C., Matthews,L., Gong,C., *et al.* (2022) The Reactome pathway knowledgebase 2022. *Nucleic Acids Res.*, **50**, D687–D692.
- Oughtred,R., Stark,C., Breitkreutz,B.-J., Rust,J., Boucher,L., Chang,C., Kolas,N., O'Donnell,L., Leung,G., McAdam,R., *et al.* (2019) The BioGRID interaction database: 2019 update. *Nucleic Acids Res.*, **47**, D529–D541.
- Davis,A.P., Wiegiers,T.C., Rosenstein,M.C., Murphy,C.G. and Mattingly,C.J. (2011) The curation paradigm and application tool used for manual curation of the scientific literature at the Comparative Toxicogenomics Database. *Database*, **2011**, bar034.
- Davis,A.P., Wiegiers,T.C., Wiegiers,J., Johnson,R.J., Sciaky,D., Grondin,C.J. and Mattingly,C.J. (2018) Chemical-induced phenotypes at CTD help to inform the predispose state and construct adverse outcome pathways. *Toxicol. Sci.*, **165**, 145–156.
- Oehme,F.W. (1970) Species differences: the basis for and importance of comparative toxicology. *Clin. Toxicol.*, **3**, 5–10.
- Mattingly,C.J., Rosenstein,M.C., Colby,G.T., Forrest,J.N. and Boyer,J.L. (2006) The Comparative Toxicogenomics Database (CTD): a resource for comparative toxicological studies. *J. Exp. Zool. A Comp. Exp. Biol.*, **305**, 689–692.
- Davis,A.P., Wiegiers,J., Wiegiers,T.C. and Mattingly,C.J. (2019) Public data sources to support systems toxicology applications. *Curr. Opin. Toxicol.*, **16**, 17–24.
- Grondin,C.J., Davis,A.P., Wiegiers,T.C., King,B.L., Wiegiers,J.A., Reif,D.M., Hoppin,J.A. and Mattingly,C.J. (2016) Advancing exposure science through chemical data curation and integration in the Comparative Toxicogenomics Database. *Environ. Health Perspect.*, **124**, 1592–1599.
- Barouki,R., Audouze,K., Becer,C., Blaha,L., Coumoul,X., Karakitsios,S., Klanova,J., Miller,G.W., Price,E.J. and Sarigiannis,D. (2022) The exposome and toxicology: a win-win collaboration. *Toxicol. Sci.*, **186**, 1–11.
- Mattingly,C.J., McKone,T.E., Callahan,M.A., Blake,J.A. and Cohen Hubal,E.A. (2012) Providing the missing link: the exposure science ontology ExO. *Environ. Sci. Technol.*, **46**, 3046–3053.
- Swanson,D.R. (1990) Medical literature as a potential source of new knowledge. *Bull. Med. Libr. Assoc.*, **78**, 29–37.
- Davis,A.P., Murphy,C.G., Rosenstein,M.C., Wiegiers,T.C. and Mattingly,C.J. (2009) The Comparative Toxicogenomics Database facilitates identification and understanding of chemical–gene–disease associations: arsenic as a case study. *BMC Med. Genomics*, **1**, 48.
- Davis,A.P., Wiegiers,T.C., Wiegiers,J., Wyatt,B., Johnson,R.J., Sciaky,D., Barkalow,F., Strong,M., Planchart,A. and Mattingly,C.J. (2023) CTD tetramers: a new online tool that computationally links curated chemicals, genes, phenotypes, and diseases to inform molecular mechanisms for environmental health. *Toxicol. Sci.*, **195**, 155–168.
- Galperin,M.Y., Fernandez-Suarez,X.M. and Rigden,D.J. (2017) The 24th annual *Nucleic Acids Research* database issue: a look back and upcoming changes. *Nucleic Acids Res.*, **45**, D1–D11.
- Davis,A.P., Johnson,R.J., Lennon-Hopkins,K., Sciaky,D., Rosenstein,M.C., Wiegiers,T.C. and Mattingly,C.J. (2012) Targeted journal curation as a method to improve data currency at the Comparative Toxicogenomics Database. *Database*, **2012**, bas051.
- Davis,A.P., Wiegiers,T.C., Johnson,R.J., Lay,J.M., Lennon-Hopkins,K., Saraceni-Richards,C., Sciaky,D., Murphy,C.G. and Mattingly,C.J. (2013) Text mining effectively scores and ranks the literature for improving chemical–gene–disease curation at the Comparative Toxicogenomics Database. *PLoS One*, **8**, e58201.
- Wiegiers,T.C., Davis,A.P., Cohen,K.B., Hirschman,L. and Mattingly,C.J. (2009) Text mining and manual curation of chemical–gene–disease networks for the Comparative Toxicogenomics Database (CTD). *BMC Bioinformatics*, **10**, 326.
- Wei,C.-H., Allot,A., Lai,P.-T., Leaman,R., Tian,S., Luo,L., Jin,Q., Wang,Z., Chen,Q. and Lu,Z. (2024) PubTator 3.0: an AI-powered literature resource for unlocking biomedical knowledge. *Nucleic Acids Res.*, **52**, W540–W546.
- Davis,A.P., Wiegiers,T.C., Grondin,C.J., Johnson,R.J., Sciaky,D., Wiegiers,J. and Mattingly,C.J. (2020) Leveraging the Comparative Toxicogenomics Database to fill in knowledge gaps for environmental health: a test case for air pollution-induced cardiovascular disease. *Toxicol. Sci.*, **177**, 392–404.
- Kim,J.W., Kim,H.S., Kim,H.R. and Chung,K.H. (2024) Next generation risk assessment of biocides (PHMG-p and CMIT/MIT)-induced pulmonary fibrosis using adverse outcome pathway-based transcriptome analysis. *J. Hazard. Mater.*, **476**, 134986.
- Vieira,L., Alves,M., Souza,T. and Farias,D. (2024) Revealing an adverse outcome pathway network for reproductive toxicity induced by atrazine, via oxidative stress. *Comput. Toxicol.*, **30**, 100317.
- Sahoo,A.K., Chivukula,N., Ramesh,K., Singha,J., Marigoudar,S.R., Sharma,K.V. and Samal,A. (2024) An integrative data-centric approach to derivation and characterization of an adverse outcome pathway network for cadmium-induced toxicity. *Sci. Total Environ.*, **920**, 170968.

26. Qin,S.-J., Zeng,Q.-G., Zeng,H.-X., Li,S.-P., Andersson,J., Zhao,B., Oudin,A., Kanninen,K.M., Jalava,P., Jin,N.-X., *et al.* (2024) Neurotoxicity of fine and ultrafine particulate matter: a comprehensive review using a toxicity pathway-oriented adverse outcome pathway framework. *Sci. Total Environ.*, **947**, 174450.
27. Wyatt,B., Davis,A.P., Wiegiers,T.C., Wiegiers,J., Abrar,S., Sciaky,D., Barkalow,F., Strong,M. and Mattingly,C.J. (2024) Transforming environmental health datasets from the Comparative Toxicogenomics Database into chord diagrams to visualize molecular mechanisms. *Front. Toxicol.*, **6**, 1437884.
28. Shannon,P., Markiel,A., Ozier,O., Baliga,N.S., Want,J.T., Ramage,D., Amin,N., Schwikowski,B. and Ideker,T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.