



OPEN ACCESS

EDITED BY

Agnes Karmaus,
Integrated Laboratory Systems, Inc.,
United States

REVIEWED BY

Jui-Hua Hsieh,
National Institute of Environmental Health
Sciences (NIH), United States
Kimberly T. To,
Integrated Laboratory Systems, Inc.,
United States

*CORRESPONDENCE

Allan Peter Davis,
✉ apdavis3@ncsu.edu

RECEIVED 24 May 2024

ACCEPTED 26 June 2024

PUBLISHED 22 July 2024

CITATION

Wyatt B, Davis AP, Wieggers TC, Wieggers J,
Abrar S, Sciaky D, Barkalow F, Strong M and
Mattingly CJ (2024), Transforming
environmental health datasets from the
comparative toxicogenomics database into
chord diagrams to visualize
molecular mechanisms.
Front. Toxicol. 6:1437884.
doi: 10.3389/ftox.2024.1437884

COPYRIGHT

© 2024 Wyatt, Davis, Wieggers, Wieggers, Abrar,
Sciaky, Barkalow, Strong and Mattingly. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Transforming environmental health datasets from the comparative toxicogenomics database into chord diagrams to visualize molecular mechanisms

Brent Wyatt¹, Allan Peter Davis^{1*}, Thomas C. Wieggers¹,
Jolene Wieggers¹, Sakib Abrar¹, Daniela Sciaky¹, Fern Barkalow¹,
Melissa Strong¹ and Carolyn J. Mattingly^{1,2}

¹Department of Biological Sciences, North Carolina State University, Raleigh, NC, United States, ²Center for Human Health and the Environment, North Carolina State University, Raleigh, NC, United States

In environmental health, the specific molecular mechanisms connecting a chemical exposure to an adverse endpoint are often unknown, reflecting knowledge gaps. At the public Comparative Toxicogenomics Database (CTD; <https://ctdbase.org/>), we integrate manually curated, literature-based interactions from CTD to compute four-unit blocks of information organized as a potential step-wise molecular mechanism, known as “CGPD-tetramers,” wherein a chemical interacts with a gene product to trigger a phenotype which can be linked to a disease. These computationally derived datasets can be used to fill the gaps and offer testable mechanistic information. Users can generate CGPD-tetramers for any combination of chemical, gene, phenotype, and/or disease of interest at CTD; however, such queries typically result in the generation of thousands of CGPD-tetramers. Here, we describe a novel approach to transform these large datasets into user-friendly chord diagrams using R. This visualization process is straightforward, simple to implement, and accessible to inexperienced users that have never used R before. Combining CGPD-tetramers into a single chord diagram helps identify potential key chemicals, genes, phenotypes, and diseases. This visualization allows users to more readily analyze computational datasets that can fill the exposure knowledge gaps in the environmental health continuum.

KEYWORDS

data visualization, chord diagram, r, molecular mechanisms, environmental health, database, tetramers

1 Introduction

Chemicals are important environmental factors that can affect human health, yet the molecular mechanisms connecting an exposure to an adverse outcome often remain unknown, reflecting knowledge and information gaps in the continuum (Thessen et al., 2020). The Comparative Toxicogenomics Database (CTD; <https://ctdbase.org/>) is an innovative public resource that provides manually curated information from the scientific literature to harmonize and centralize cross-species data for chemical exposures and their interactions with genes, phenotypes (non-disease biological

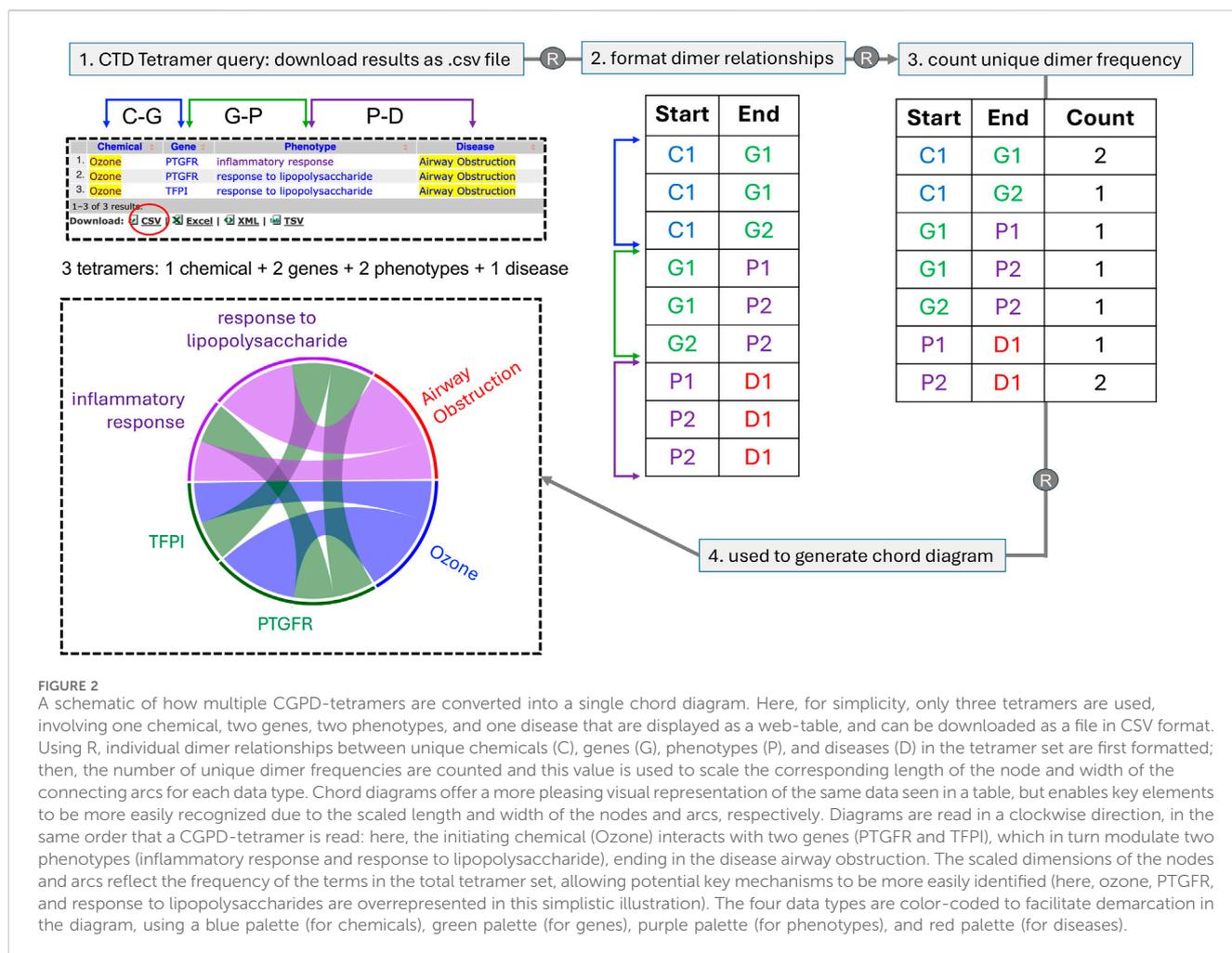


FIGURE 2

A schematic of how multiple CGPD-tetramers are converted into a single chord diagram. Here, for simplicity, only three tetramers are used, involving one chemical, two genes, two phenotypes, and one disease that are displayed as a web-table, and can be downloaded as a file in CSV format. Using R, individual dimer relationships between unique chemicals (C), genes (G), phenotypes (P), and diseases (D) in the tetramer set are first formatted; then, the number of unique dimer frequencies are counted and this value is used to scale the corresponding length of the node and width of the connecting arcs for each data type. Chord diagrams offer a more pleasing visual representation of the same data seen in a table, but enables key elements to be more easily recognized due to the scaled length and width of the nodes and arcs, respectively. Diagrams are read in a clockwise direction, in the same order that a CGPD-tetramer is read: here, the initiating chemical (Ozone) interacts with two genes (PTGFR and TFPI), which in turn modulate two phenotypes (inflammatory response and response to lipopolysaccharide), ending in the disease airway obstruction. The scaled dimensions of the nodes and arcs reflect the frequency of the terms in the total tetramer set, allowing potential key mechanisms to be more easily identified (here, ozone, PTGFR, and response to lipopolysaccharides are overrepresented in this simplistic illustration). The four data types are color-coded to facilitate demarcation in the diagram, using a blue palette (for chemicals), green palette (for genes), purple palette (for phenotypes), and red palette (for diseases).

grow in length and width, it becomes visually apparent which components occur most frequently (Figure 2). This approach allows all the CGPD-tetramers from a table to be combined and visualized in a single graph, facilitating the identification of potentially key mechanistic elements connecting a chemical exposure to a disease endpoint.

The chord diagrams are compact, elegant representations of complex tetramer datasets. To generate these diagrams, we make use of R, a widely used and publicly available programming environment that can be run on all major operating systems, including Windows, MacOS, and Linux. The R-based script and instructions that we provide are easy to learn, so that even novice users can quickly begin creating chord diagrams illuminating CTD tetramer results.

2 Methods

2.1 CTD data version and analysis

Analysis was performed using CTD public data available (revision 17,266; March 2024). CTD is updated with new content on a monthly basis (<https://ctdbase.org/about/dataStatus.go>); consequently, query results described in this text may vary over time.

2.2 R environment

The freely available version 4.4.0 of R (<https://www.r-project.org>) was used (R Core Team, 2018). The R package data.table version 1.15.2 (<https://CRAN.R-project.org/package=data.table>) was used in preparing the data and the R package circlize version 0.4.16 (<https://CRAN.R-project.org/package=circlize>) was used to generate the chord diagrams (Gu et al., 2014). To format CGPD-tetramers and plot a basic chord diagram, download the *CTD-vizscript.R*, from GitHub (https://github.com/CTD-NCSU/Chord_Diagrams) or use the version included with a dataset of test tetramers (“test.csv”) for practice (Supplemental File S1). For CTD users unfamiliar with R, it is not necessary to download local versions of the software to generate the chord diagrams. Instead, a free, web-based version of R and RStudio environment from Posit Cloud (<https://posit.cloud>) may be used to generate the diagrams; users will need to first create a free account at Posit Cloud, and the steps required using the latter method are provided below.

2.3 Basic steps for CTD tetramer visualization using web-based Posit Cloud

Step 1: Download CTD Tetramers query results in CSV format (and rename the file to something more intuitive); start with a

dataset of no more than 1,500 tetramers (discussed below); alternatively, use the test sample CSV file provided as “test.csv” (Supplemental File S1), which contains 309 CGPD-tetramers composed of two chemicals (copper and iron), 61 genes, 87 phenotypes, and 1 disease (liver neoplasms).

Step 2: Download *CTD-vizscript.R* locally from GitHub (https://github.com/CTD-NCSU/Chord_Diagrams) or use the provided file (Supplemental File S1)

Step 3: Create a free account at Posit Cloud (<https://posit.cloud>) and log in to the “Your Workspace” web page.

Step 4: On the right-hand side of the workspace web page, click on the “New Project” button and select “New RStudio Project” from the drop-down menu. The browser will reload showing three panes: “Console” in the upper left, “Environment” in the upper right, and “Files” in the bottom right; browser screenshots for the subsequent steps are provided (Supplemental File S2). In the bottom right pane (“Files”), click on the “Upload” button to browse for and select the downloaded copy of *CTD-vizscript.R* file (panel A, Supplemental File S2); then upload the CSV formatted file of the CTD Tetramer query results that are to be visualized, or use the provided sample file “test.csv.” Both files will now be listed in the “Files” pane.

Step 5: Click on the *CTD-vizscript.R* file and the script will appear in a new upper left-hand pane (and “Console” will displace to the bottom left). An alert will pop up and ask to install *circlize* and *data.table*. Choose install (panel C, Supplemental File S2). In the “Console” window, you’ll see the program run until it is finished.

Step 6: Load *data.table* and *circlize* using the `library()` command to make them available for future steps. To do this, go to code lines 2–3 and execute the two lines of the script by using the cursor to physically highlight the two lines for `library(data.table)` and `library(circlize)` in the window pane, and then click “Run” button (panel D, Supplemental File S2).

Step 7: Under “#Import data,” at code line 6 assign the name of the CTD Tetramer query result file to the variable “tetramers” (panel E, Supplemental File S2). To do this, type in the exact name of the file, which is case-sensitive, surrounded by quotation marks, and suffixed with “.csv,” all enclosed with the parentheses. For example, if using the provided test file (test.csv), then set the line to read: `tetramers <- read.csv(“test.csv”).` The case-sensitive file name must be spelled exactly, contain the CSV suffix, be flanked by quotation marks, and enclosed in the parentheses.

Step 8: Using the cursor, highlight the line from step 7 (where the tetramer file name is entered on code line 6) along with the rest of the script to code line 60 (panel F, Supplemental File S2). Click the “Run” button.

Step 9: If everything runs successfully, the tool will generate a colored chord diagram in the lower right pane under “Plots” (panel I, Supplemental File S2). This process can take a few minutes, depending upon the total number of CGPD-tetramers that are being graphed. See below for common errors that might arise. To adjust the font size of the labels, go to code line 59 and find “`cex =`” near the end of the script under “## Add labels to diagram” (panel H, Supplemental File S2). “`cex`” stands for “character expansion” and is the parameter that changes the font size of the node labels. It has a default value of 1; to make the font size smaller, the parameter can be changed to a value between 0 and 1; for example, `cex = 0.5` will result in text that is 50% smaller than the default value. Changing the parameter to a value greater than 1 will result in larger text; for

example, `cex = 1.5` will result in text that is 50% larger than the default value. The greater the number of tetramers being plotted, the smaller the value of `cex` should be in order for the labels to be clearly separated and not overlapping. Additionally, or alternatively, for very large sets of tetramers where the node labels are difficult to resolve, a user might instead prefer to display only the node accession identifiers, as discussed below. After changing the value of `cex`, go to code line 45 and highlight the text beginning with `circos.clear` along with the remaining script, and click “Run.” This is necessary to clear out the old data and rerun the program using the newly entered values.

Step 10: The chord diagram can be saved as either a PDF or an image by clicking the “Export” button under the “Plots” tab of the lower right-hand pane (panel I, Supplemental File S2).

2.4 Fixing common errors that might be encountered

Common error 1: An error that states Error in `setDT(cgpd)`: could not find function “`setDT`” indicates the *data.table* package is not loaded. Highlight the line that says `library(data.table)` near the top of the script (code line 2) and click “Run.” This will make the function “`setDT`” available.

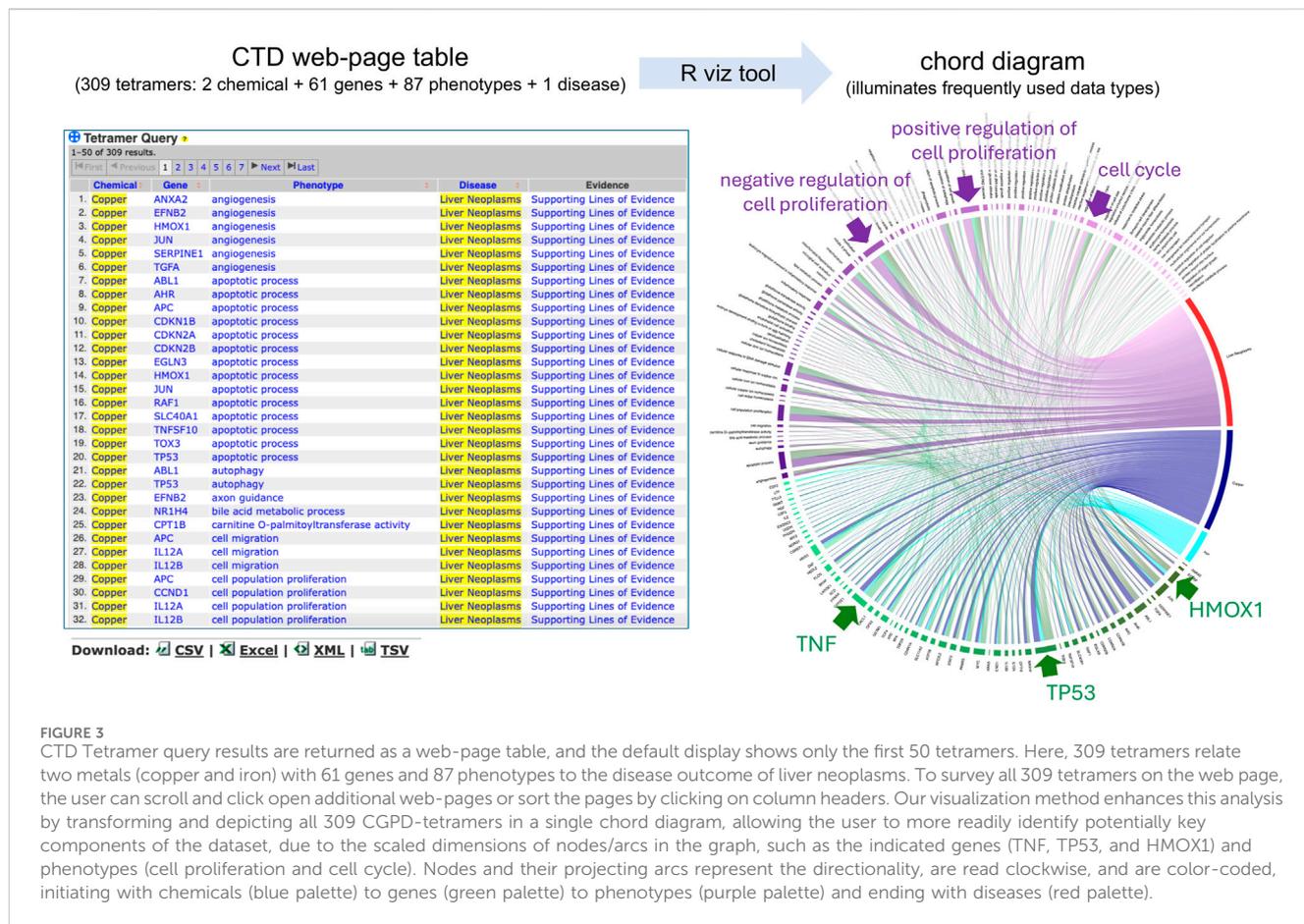
Common error 2: An error that states any of the following are not found, “`circos.clear`,” “`circos.par`,” “`chordDiagram`,” or “`circos.track`,” indicates the *circlize* package is not loaded. Highlight code line 3 that says `library(circlize)` and click “Run.” This will make those functions available.

Common error 3: An error that states “Maybe your “`gap.degree`” is too large so that there is no space to allocate sectors” indicates that there is a very large number of tetramers. First, highlight the code line 45 that says `circos.clear()` and click “Run.” Next, go to code line 47 that says `circos.par(gap.degree = 1)` and change `gap.degree` to a number less than 1 (panel G, Supplemental File S2). This is a setting for the gaps between neighboring nodes. Larger values for `gap.degree` will create larger spacing between nodes and smaller values for `gap.degree` will create smaller spacing between nodes. Once `gap.degree` is changed to a smaller value, highlight the line along with the rest of the script and click “Run.” If the error occurs again, repeat the process of running `circos.clear`, changing `gap.degree` to an even smaller number and rerunning the script.

2.5 Customizing chord diagrams

To enhance the readability of large chord diagrams, it is sometimes preferable to shorten the node term label to an accession identifier (ID), as this further compacts the diagram by reducing the text. A label can be switched from “node term” to “node ID” easily by adding “.ID” as a suffix the end of the Chemical, Gene, Phenotype, or Disease variables (code lines 10–13) in the section under “# Prep data” (panel J, Supplemental File S2). Highlight the script beginning with this line along with the remaining script and click “Run.” Additional customization options can be found in the *circlize* documentation at https://jokergoo.github.io/circlize_book/book/index.html.

The default selection of color palettes for CTD tetramer chord diagrams are blue (chemical), green (gene), purple (phenotype), and



red (disease). If necessary, users can change these colors using code lines 37–40 (panel K, [Supplemental File S2](#)). Each line of script for each of the four variables contains two colors. The script will create a palette from the first color listed to the second color. To change the color, simply edit the name in the code line, making sure the new color name is in quotation marks and the two colors are separated by a comma. There are many R color cheat sheets that can be found with an internet search to aid in choosing colors, including those that are specifically designed for users with color-vision deficiencies ([Dahl et al., 2022](#)).

A summary of the entire CTD-vizscript code is provided with information boxes showing the exact code lines that users can edit to customize the chord diagram output (panel L, [Supplemental File S2](#)).

3 Results

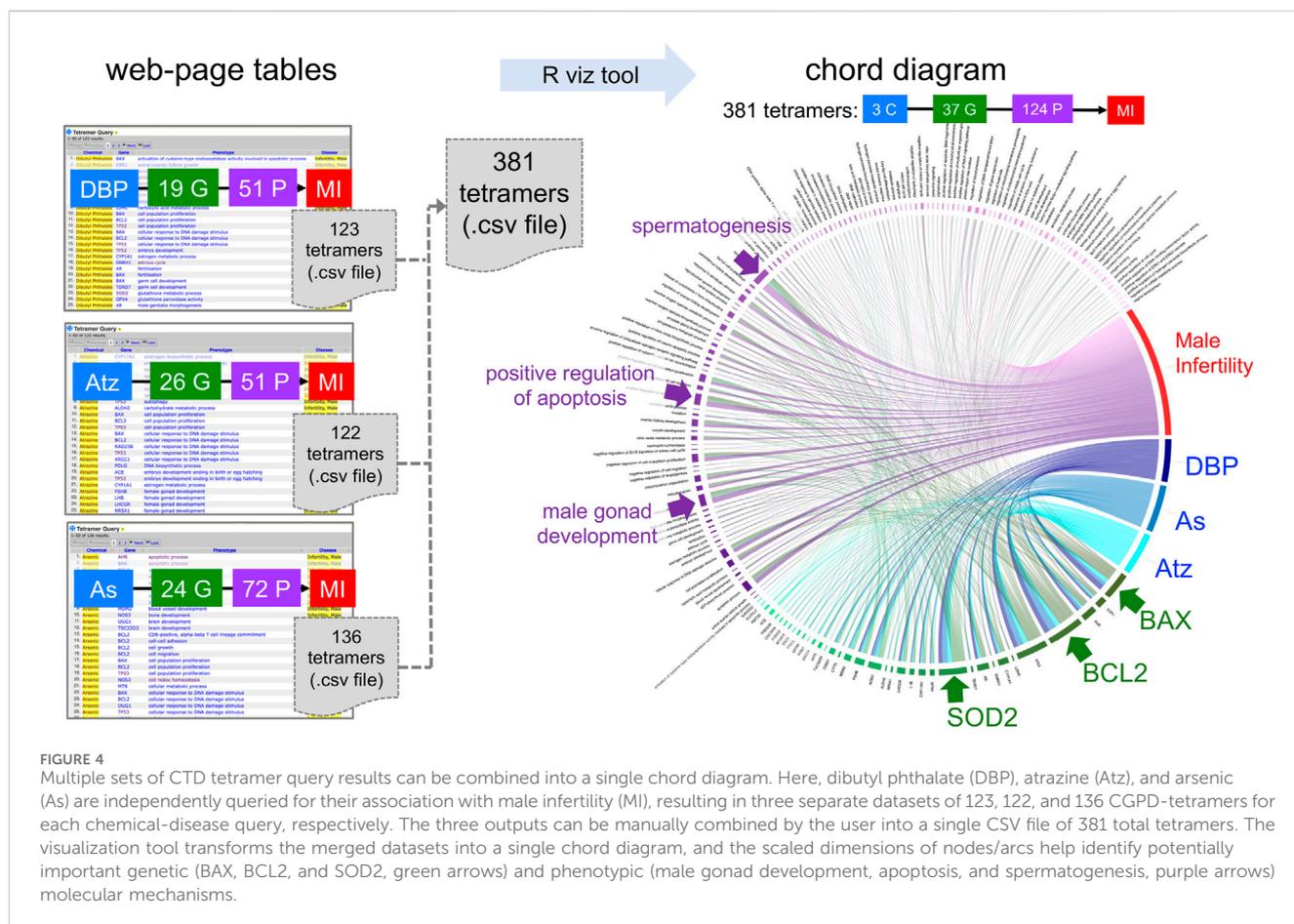
3.1 Visualizing a single dataset of tetramers

To demonstrate how to transform a large dataset of CGPD-tetramers into a single chord diagram, we queried the CTD Tetramers tool (<https://ctdbase.org/query.go?type=tetramer>) using “metals” as the chemical input, selecting the hierarchy filter to include any descendant terms, and “liver neoplasms” as the disease outcome. This yields 309 CGPD-tetramers, composed of two chemicals (copper and iron), 61 genes, and 87 phenotypes connected to liver

neoplasms ([Figure 3](#)). The web-based table result displays only the first 50 of the 309 CGPD-tetramers (but it can be set to display 500). The tabular output is downloaded from the webpage as a CSV file, and then following the step-by-step directions (see [Methods](#)), the CTD Tetramer results are processed to generate a chord diagram that depicts all 309 CGPD-tetramers in a single image ([Figure 3](#)). Each chemical, gene, phenotype, and disease is represented as a colored node on the circumference of the circle, with chemicals shown in a blue palette, genes in a green palette, phenotypes in a purple palette, and diseases in a red palette. The arcs in the center of the circle connect the four nodes of each CGPD-tetramer, and are read in a clockwise manner: chemical (blue arc) to gene, gene (green arc) to phenotype, and phenotype (purple arc) to disease. The length of the node reflects the frequency that the specific data term occurs in the CGPD-tetramers and the width of the arc is proportional to the frequency that a dimer interaction occurs ([Figure 2](#)). Larger nodes/arcs readily highlight the more commonly used data types, and visually indicate potentially important molecular players, such as the genes HMOX1, TP53, and TNF, and the phenotypes cell proliferation and cell cycle for metal-associated liver neoplasms ([Figure 3](#)).

3.2 Combining and visualizing multiple datasets of tetramers

Environmental health is a complex, multifactorial system wherein outcomes can be the result of chemical mixtures and of multiple, independent chemical exposures from different sources



(Davis et al., 2023b). Thus, CTD users often query for and analyze different sets of CGPD-tetramers for different chemicals and then combine the data to look for overlaps between shared molecular mechanisms and diseases. To aid in such analysis, multiple tetramer query results can be combined into a single chord diagram. To merge query results, it is most straightforward to download the individual CSV files for each independent query, combine the files into a single CSV document (using Excel copy-and-paste), and then upload that single file as the data source for generating a chord diagram as described above.

For example, numerous environmental factors have been associated with male infertility, including phthalates, pesticides, and arsenic (Rodprasert et al., 2023). Three independent queries for CTD tetramers that relate the chemicals dibutyl phthalate, atrazine, and arsenic to male infertility can be combined into a single chord diagram to help identify potentially key genetic and phenotypic molecular mechanisms associated with these environmental stressors (Figure 4). In total, these three independent queries yield 381 CGPD-tetramers composed of 37 genes and 124 phenotypes connecting the three chemicals to male infertility. Visualization of this complex environmental health dataset illuminates candidate genes (BAX, BCL2, and SOD2) and phenotypes (male gonad development, apoptosis, and spermatogenesis) connecting the chemicals to male infertility, and thus may represent critical intermediate molecular mechanisms (Figure 4).

3.3 Limitations

An impediment for visualizing CGPD-tetramers as a chord diagram is spacing constraints that limit the comprehensibility of these diagrams. We have found that about 1,500 tetramers tend to represent a practical upper limit for figure clarity. For larger datasets, there are a few options. As described above, customizing the node displays either using the “cex” parameter (for font size) or changing the label from “term” to “node ID” will significantly reduce the text (panel J, Supplemental File S2), resulting in a more compact diagram, but still enabling nodes to be readily identified. As well, the starting number of CGPD-tetramers to be graphed can be reduced preemptively by adding additional search parameters in the query to narrow the dataset to focus on a more specific question. For example, a query for “liver neoplasms” results in over 16,700 CGPD-tetramers, and represents a very broad open-ended question with respect to environmental health. However, when “metals” is included as the chemical parameter in the search, the number reduces to 309 (Figure 3), and focuses on a more specific, addressable environmental health question. Alternatively, including phenotypes (e.g., “nervous system process,” “inflammatory response,” or “metabolic process,” with the filter option actively changed to include any descendant term) can also help focus questions and provide a more manageable number of CGPD-tetramers for analysis. Finally, CTD is also exploring ways to rank tetramers with a metric score as a means to prioritize the

query results with the CGPD-tetramers that have the most supporting lines of evidence from the literature. Currently, the default display is alphabetical on the web-page results. Ranking would allow users to select tetramer subsets (instead of using the entire output) to include in their chord diagram.

Another limitation is the use of colors to represent information, as this may be an issue for those with color-vision deficiencies. Users can customize the plot colors to provide accessibility for specific color-vision deficiencies (Dahl et al., 2022) or for aesthetic purposes (panel K, Supplemental File S2).

Finally, the obvious complicating factor to this approach is the manual effort required to download and transfer data files from CTD to an R environment in order to generate a chord diagram. Consequently, we plan to integrate similar functionality directly into CTD Tetramer query results pages in a future release.

4 Discussion

Advances are being made to provide clarity to complex environmental health data, including development and adoption of standardized vocabularies (Mattingly et al., 2016; Holmgren et al., 2023) and development of visual frameworks that display, organize, and integrate data from diverse sources. For example, the Aggregate Exposure Pathway (AEP) collates and structures chemical and biological knowledge at the population level to connect an environmental stressor with its source, medium, external/internal concentrations, and target site exposure (Teeguarden et al., 2016), while the Adverse Outcome Pathway (AOP) assembles information at the cellular level to link an initiating molecular event, a series of key processes, and an adverse outcome (Ankley and Edwards, 2018). Together, the AEP and AOP frameworks complement each other and connect population data to molecular data, respectively, helping to bridge the environmental health continuum (Hines et al., 2018).

To harmonize environmental health data, CTD embraces community-accepted FAIR controlled vocabularies and ontologies to curate the scientific literature for both population exposure science (Grondin et al., 2016) and molecular/cellular laboratory results (Davis et al., 2011). This imperative step enables disparate data to be integrated to generate inferred relationships that provide potential mechanistic solutions for recognized knowledge gaps (Davis et al., 2008). CGPD-tetramers are a novel derivative of this integration strategy, where the combined data types are structured as a step-wise mechanistic pathway: a chemical first interacts with a gene to then modulate a phenotype that may contribute to a final disease outcome (Davis et al., 2020). Users can easily generate CGPD-tetramers for any environmental chemical, gene, phenotype, or disease of interest at CTD (Davis et al., 2023b); however, these queries typically bring back hundreds of results, in a web-based table.

To help survey and distill these large environmental health datasets, we present a simple visualization method that transforms CGPD-tetramers into a chord diagram, allowing

all of the data to be condensed and viewed in a single, elegant figure. CGPD-tetramers are formatted and transformed using R. While our technique can be easily adopted by researchers already familiar with R, we provide step-by-step instructions for novice users, leveraging a free web-browser system that enables inexperienced users to generate publication-worthy chord diagrams for their own presentations. By their nature, chord diagrams show the interconnections between different data types and facilitate the visual identification of the most frequently used data components, allowing users to discern potentially important intermediate molecular mechanisms between an environmental exposure and a disease outcome. Graphing these complex, toxicology-related mechanistic pathways can help build, inform, or refine current AOPs (Davis et al., 2023b) and be used to provide visual mechanistic links to AEPs. Together, both the table-based version of CGPD-tetramers and the derived chord diagrams complement each other, with the former providing data in an easily sortable format and the latter providing a graphical view that helps to visually detect key data types.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

BW: Data curation, Investigation, Methodology, Validation, Writing—original draft, Writing—review and editing. AD: Conceptualization, Data curation, Investigation, Supervision, Validation, Visualization, Writing—original draft, Writing—review and editing. TW: Software, Writing—review and editing. JW: Software, Writing—review and editing. SA: Software, Writing—review and editing. DS: Data curation, Writing—review and editing. FB: Data curation, Writing—review and editing. MS: Data curation, Writing—review and editing. CM: Funding acquisition, Project administration, Writing—review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. National Institute of Environmental Health Sciences [U24 ES033155]. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Acknowledgments

We thank Dr. Antonio Planchart for testing the methods and critically reviewing the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/ftox.2024.1437884/full#supplementary-material>

References

- Ankley, G. T., and Edwards, S. W. (2018). The adverse outcome pathway: a multifaceted framework supporting 21st century toxicology. *Curr. Opin. Toxicol.* 9, 1–7. doi:10.1016/j.cotox.2018.03.004
- Chai, Z., Zhao, C., Jin, Y., Wang, Y., Zou, P., Ling, X., et al. (2021). Generating adverse outcome pathway (AOP) of inorganic arsenic-induced adult male reproductive impairment via integration of phenotypic analysis in Comparative Toxicogenomics Database (CTD) and AOP wiki. *Toxicol. Appl. Pharmacol.* 411, 115370. doi:10.1016/j.taap.2020.115370
- Dahl, E. M., Neer, E., Bowie, K. R., Leung, E. T., and Karstens, L. (2022). microshades: an R package for improving color accessibility and organization of microbiome data. *Microbiol. Resour. Announc.* 11, e0079522. doi:10.1128/mra.00795-22
- Davis, A. P., Murphy, C. G., Rosenstein, M. C., Wiegiers, T. C., and Mattingly, C. J. (2020). The Comparative Toxicogenomics Database facilitates identification and understanding of chemical-gene-disease associations: arsenic as a case study. *BMC Med. Genomics.* 1, 48. doi:10.1093/database/bar034
- Davis, A. P., Wiegiers, T. C., Grondin, C. J., Johnson, R. J., Sciaky, D., Wiegiers, J., et al. (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. doi:10.1093/toxsci/kfaa113
- Davis, A. P., Wiegiers, T. C., Johnson, R. J., Sciaky, D., Wiegiers, J., and Mattingly, C. J. (2023a). Comparative toxicogenomics database (CTD): update 2023. *Nucleic Acids Res.* 51, D1257–D1262. doi:10.1093/nar/gkac833
- 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 (Oxford)* 2011, bar034. doi:10.1093/database/bar034
- Davis, A. P., Wiegiers, T. C., Wiegiers, J. W., Wyatt, B., Johnson, R. J., Sciaky, D., et al. (2023b). 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. doi:10.1093/toxsci/kfad069
- Grondin, C. J., Davis, A. P., Wiegiers, T. C., King, B. L., Wiegiers, J. A., Reif, D. M., et al. (2016). Advancing exposure science through chemical data curation and integration in the Comparative Toxicogenomics Database. *Environ. Health Perspect.* 124, 1592–1599. doi:10.1289/ehp174
- Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). Circlize Implements and enhances circular visualization in R. *Bioinformatics* 30, 2811–2812. doi:10.1093/bioinformatics/btu393
- Hines, D. E., Edwards, S. W., Conolly, R. B., and Jarabek, A. M. (2018). A case study application of the Aggregate Exposure Pathway (AEP) and Adverse Outcome Pathway (AOP) frameworks to facilitate the integration of human health and ecological end points for cumulative risk assessment (CRA). *Environ. Sci. Technol.* 52, 839–849. doi:10.1021/acs.est.7b04940
- Holmgren, S., Bell, S. M., Wignall, J., Duncan, C. G., Kwok, R. K., Cronk, R., et al. (2023). Workshop report: catalyzing knowledge-driven discovery in environmental health sciences through a harmonized language. *Int. J. Environ. Res. Public Health.* 20, 2317. doi:10.3390/ijerph20032317
- Jeong, J., Kin, D., and Choi, J. (2023). Integrative data mining approach: case study with adverse outcome pathway network leading to pulmonary fibrosis. *Chem. Res. Toxicol.* 36, 838–847. doi:10.1021/acs.chemrestox.2c00325
- Kim, M., Kim, S. H., Choi, J. Y., and Park, Y. J. (2023). Investigating fatty liver disease-associated adverse outcome pathways of perfluorooctane sulfonate using a systems toxicology approach. *Food Chem. Toxicol.* 176, 113781. doi:10.1016/j.fct.2023.113781
- Mattingly, C. J., Boyles, R., Lawler, C. P., Haugen, A. C., Deary, A., and Haendel, M. (2016). Laying a community-based foundation for data-driven semantic standards in environmental health sciences. *Environ. Health Perspect.* 124, 1136–1140. doi:10.1289/ehp.1510438
- Nguyen, H. D., and Kim, M.-S. (2023). Interactions between cadmium, lead, mercury, and arsenic and depression: a molecular mechanism involved. *J. Affect. Disord.* 327, 315–329. doi:10.1016/j.jad.2023.02.013
- R Core Team (2018). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: <https://www.r-project.org/>.
- Rodprasert, W., Toppari, J., and Virtanen, H. (2023). Environmental toxicants and male fertility. *Best. Pract. Res. Clin. Obstet. Gynaecol.* 86, 102298. doi:10.1016/j.bpobgyn.2022.102298
- Teeguarden, J. G., Tan, Y.-M., Edwards, S. W., Leonard, J. A., Anderson, K. A., Gorley, R. A., et al. (2016). Completing the link between exposure science and toxicology for improved environmental health decision making: the Aggregate Exposure Pathway framework. *Environ. Sci. Technol.* 50, 4579–4586. doi:10.1021/acs.est.5b05311
- Thessen, A. E., Grondin, C. J., Kulkarni, R. D., Brander, S., Truong, L., Vasilevsky, N. A., et al. (2020). Community approaches for integrating environmental exposures into human models of disease. *Environ. Health Perspect.* 128, 125002. doi:10.1289/ehp7215
- Zhao, Y., Zhang, X., Zhang, Z., Huang, W., Tang, M., Du, G., et al. (2024). Hepatic toxicity prediction of bisphenol analogs by machine learning strategy. *Sci. Total Environ.* 934, 173420. doi:10.1016/j.scitotenv.2024.173420