

# The Comparative Toxicogenomics Database, a Database for Scientific Professionals



**Zach Van Ysseldyk** LMU College of Business Administration  
Department of AIMS

**Katie Wright** LMU Seaver College of Science and Engineering  
Department of Chemistry and Biochemistry

**BIOL/CMSI 367-01: Biological Databases**  
**October 3, 2017**

<http://ctdbase.org/>

# Outline

- 1. A Publicly Funded Healthcare Database**
- 2. Useful Information in Narrow Field**
- 3. Generally User-Friendly, Although Inconsistent**
- 4. Useful for Professionals or Those Familiar with the Field**

# Outline

- 1. A Publicly Funded Healthcare Database**
- 2. Useful Information in Narrow Field**
- 3. Generally User-Friendly, Although Inconsistent**
- 4. Useful for Professionals or Those Familiar with the Field**

# A Healthcare Database

1. **Focuses on how environmental toxins affect human health<sup>1</sup>**
  - **chemical-gene/protein interactions, chemical-disease, and gene-disease relationships**
2. **Manually Curated by an in house staff of 5 database personnel<sup>2</sup>**
3. **Mainly logs and curates secondary data<sup>3</sup>**

[1] <http://ctdbase.org/about/>

[2] <http://ctdbase.org/about/personnel.jsp>

[3] <http://ctdbase.org/about/dataStatus.go>

# In-House Maintenance

**CTD is a public database maintained by a team of 10 individuals, coming from:**

- North Carolina State University**
- Mount Desert Island Biological Laboratory**
- Remotely located biocurators<sup>4</sup>**

[4] <http://ctdbase.org/help/faq/?p=5800169>

# Publicly Funded

- **They are primarily funded through the National Institute of Environmental Health Sciences.**
- **Grant specificities can be found on the CTD's funding page.<sup>5</sup>**

# Outline

1. A Publicly Funded Healthcare Database
- 2. Useful Information in Narrow Field**
3. Generally User-Friendly, Although Inconsistent
4. Useful for Professionals or Those Familiar with the Field

# Vast Data Coverage in Narrow Field

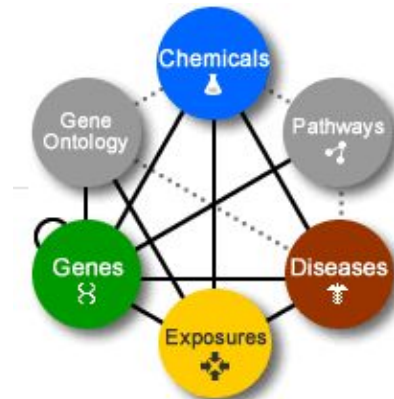
- **There are 34,572,272 unique records.<sup>7</sup>**
- **CTD claims the integrity of secondary sources harvested by manual curation.**
- **They do not make claims for absolute coverage of toxicogenomic information.**
- **CTD has no specific species.**

[7] <http://ctdbase.org/about/dataStatus.go>



# Many Interconnecting Topics

- This information can be utilized to learn how the below categories affect each other:
  - Chemicals
  - Pathways
  - Diseases
  - Exposures
  - Genes
  - Gene Ontology



# Relevant and Frequently Updated Data

- **Useful because toxicogenomics help us to understand and eventually cure diseases.**
- **CTD is the only comparative toxicogenomics database - however primary sources can be found elsewhere.**
- **CTD is celebrating their 10th anniversary, however went online November 2, 2011.**
- **CTD tries to update at the beginning of each month.**
- **The last update was August 24th.**

# Outline

1. A Publicly Funded Healthcare Database
2. Useful Information in Narrow Field
3. **Generally User-Friendly, Although Inconsistent**
4. Useful for Professionals or Those Familiar with the Field

# Links to Other Databases

1. **BioGRID** (3.4.151 release)
2. **ChemIDplus**<sup>®</sup> (as of August 18, 2017)
3. **DrugBank** (as of August 18, 2017)
4. **GO** (as of August 18, 2017)
5. **KEGG** (as of August 18, 2017)
6. **MeSH**<sup>®</sup> (2016 MeSH release)
7. **NCBI Gene** (as of August 18, 2017)
8. **NCBI Taxonomy** (as of August 18, 2017)
9. **PubMed**<sup>®</sup> (as of August 18, 2017)
10. **Reactome** (as of August 18, 2017)

# Data Browsing Inconsistent Across Data Type

- **Genes: Overall Inconvenient**
  - large data set
  - Listed alphabetically
- **Chemicals & Diseases: well organized**
- **Uncomfortable for users without biological background**
- **Exposure studies: no browsing capabilities**

Ancestors ? Top ↑

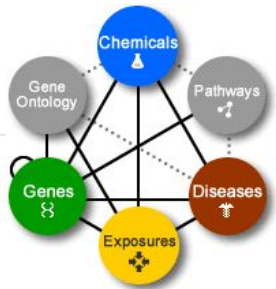
1. Chemicals ← Organic Chemicals ⌘ ⬆ ⬇ ← Carboxylic Acids ⌘ ⬆ ⬇ ← Acids, Acyclic ⌘ ⬆ ⬇ ← Sugar Acids ⌘ ⬆ ⬇ ← Ascorbic Acid ⌘ ⬆ ⬇
2. Chemicals ← Organic Chemicals ⌘ ⬆ ⬇ ← Carboxylic Acids ⌘ ⬆ ⬇ ← Hydroxy Acids ⌘ ⬆ ⬇ ← Sugar Acids ⌘ ⬆ ⬇ ← Ascorbic Acid ⌘ ⬆ ⬇
3. Chemicals ← Carbohydrates ⌘ ⬆ ⬇ ← Sugar Acids ⌘ ⬆ ⬇ ← Ascorbic Acid ⌘ ⬆ ⬇

<http://ctdbase.org/detail.go?type=chem&acc=D001205#descendants>

# It is a user-friendly database

We found it easy to navigate the data due to:

- Broad to specific design
- Uses legend icons for quick help 🌐 🧬 🧪
- Clickable image navigates user to desired data category
- Easiness to query data
- Provided help bar with video tutorials on how to navigate webpage<sup>6</sup>



**1 Select your input type**

- Chemicals (MeSH® names, synonyms, or IDs, or CAS RNs) ?
- Diseases (MeSH or OMIM names, synonyms, or IDs) ?
- Genes (NCBI symbols or IDs) ?
- Gene Ontology terms (GO names, synonyms, or IDs) ?
- Pathways (KEGG or REACTOME names or IDs) ?
- References (PubMed® IDs or DOIs) ?

**2 Provide query terms (up to 4,000)**

Return-, tab- or |-delimited

Or **upload** a tab-separated file:  
Choose File No file chosen  
Identifiers column: 1

**3 Choose data to download**

**Data**

**Chemical-gene interactions** ?

Curated

Type(s):

- glycosylation
- N-linked glycosylation
- O-linked glycosylation
- hydroxylation
- lipidation
- farnesylation

**Chemical associations** ?

Curated

**Gene associations** ?

Curated

**Disease associations** ?

All

Curated

Inferred

**Pathway associations** ?

Inferred

Enriched (recommended)

**Gene Ontology associations** ?

Enriched (recommended)

All





Download Clear

# Downloading Data

- Many Standard file formats available
- One-click download
- Also possible to download query results

44. Learning Disorders	1 chemical: methylmercuric chloride	3.27	1
45. Weight Gain	1 chemical: methylmercuric chloride	3.25	1
46. Cognition Disorders	1 chemical: methylmercuric chloride	3.19	1
47. Prenatal Exposure Delayed Effects	1 chemical: methylmercuric chloride	3.19	1
48. Hypertension	1 chemical: methylmercuric chloride	2.64	1

1-48 of 48 results.

Download:  CSV |  Excel |  XML |  TSV

<http://ctdbase.org/detail.go?type=gene&acc=189099&view=disease>

# Outline

1. A Publicly Funded Healthcare Database
2. Useful Information in Narrow Field
3. Generally User-Friendly, Although Inconsistent
4. **Useful for Professionals or Those Familiar with the Field**



# Personal Recommendations

**As an inexperienced user, I would not recommend this database for someone who has very little knowledge about comparative toxicogenomics, or biology.**

# By Professionals, for Professionals

- Curated by  
doctoral level staff
- Very specific  
information

## Principal Investigator



**Carolyn J. Mattingly, PhD**  
Principal Investigator, Associate  
Professor  
[cjmattin@ncsu.edu](mailto:cjmattin@ncsu.edu)

## Curation



**Allan Peter Davis, PhD**  
Biocuration Project Manager  
[apdavis3@ncsu.edu](mailto:apdavis3@ncsu.edu)



**Robin Johnson, PhD**  
Scientific Curator  
[rjohns3@ncsu.edu](mailto:rjohns3@ncsu.edu)



**Daniela Sciaky, PhD**  
Scientific Curator  
[dsciaky@ncsu.edu](mailto:dsciaky@ncsu.edu)



**Cynthia J. Grondin, PhD**  
Scientific Curator  
[cjgrondin@ncsu.edu](mailto:cjgrondin@ncsu.edu)



**Kelley Lennon-Hopkins, DSc**  
Scientific Curator

<http://ctdbase.org/about/personnel.jsp>

# Summary

- 1. NIEHS Funded Healthcare Database**
- 2. Useful Information in Field of Toxicogenomics**
- 3. Generally User-Friendly, Although Inconsistent**
- 4. Designed for Professionals or Highly Knowledgeable Amateur Scientists**

# References

**Davis AP, Grondin CJ, Johnson RJ, Sciaky D, King BL, McMorran R, Wiegers J, Wiegers TC, Mattingly CJ. The Comparative Toxicogenomics Database: update 2017. Nucleic Acids Res. 2016 Sep 19;[Epub ahead of print]. Retrieved September 29, 2017.**

**LMU BioDB 2017. (2017). Week 5. Retrieved September 29, 2017, from [https://xmlpipedb.cs.lmu.edu/biodb/fall2017/index.php/Week\\_5](https://xmlpipedb.cs.lmu.edu/biodb/fall2017/index.php/Week_5).**

# Acknowledgments

- **Loyola Marymount University, Seaver College of Science and Engineering**
- **The Departments of Biology and Computer Science**
- **Professor Dionisio and Professor Dahlquist**
- **Students of BIOL/CMSI 367-01: Biological Databases**

**Questions?**