

# **The Monarch Initiative**

**By: Dina Bashoura and Nicole Kalcic**

Loyola Marymount University  
BIOL/CMSI 367-01: Biological Databases  
October 2017

# About The Monarch Initiative

**Introduction**

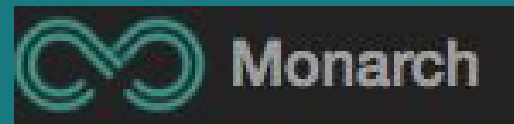
**Overview of  
Database**

**Scientific  
Quality of  
Database**

**General  
Utility of  
Database to  
Scientific  
Community**

**Conclusion**

- **Integrate, align, and re-distribute cross-species gene, genotype, variant, disease, and phenotype data**
- **Facilitate identification of animal models of human disease through phenotypic similarity**
- **Enable quantitative comparison of cross-species phenotypes**
- **Improve ontologies to better curate genotype-phenotype data**



# How Information Is Retrieved

Introduction

Overview of  
Database

Scientific  
Quality of  
Database

General  
Utility of  
Database to  
Scientific  
Community

Conclusion

- Secondary source with information curated from databases, spreadsheets, text files, web API's and others like...
  - Oregon Health & Science Univeristy
  - Lawrence Berkely National Laboratory
  - Jackson Laboratory
  - Charité - Universitätsmedizin Berlin
  - Queen Mary of London
  - Garvan Institute
  - Renaissance Computing Institute
  - University of Pittsburgh

**Introduction**

**Overview of  
Database**

**Scientific  
Quality of  
Database**

**General  
Utility of  
Database to  
Scientific  
Community**

**Conclusion**

# Funding

- **The Monarch Initiative found funding for the database through...**
  - **National Institutes of Health Office of the Director Grant**
  - **National Institutes of Health Undiagnosed Diseases Program**



# Model Organisms

Introduction

Overview of  
Database

Scientific  
Quality of  
Database

General  
Utility of  
Database to  
Scientific  
Community

Conclusion

Model organisms used for comparison:

- Human
- Mouse
- Worm
- Fly
- Zebrafish
- Dog
- Cow



# Authors Claims on Coverage

Introduction

Overview of  
Database

Scientific  
Quality of  
Database

General  
Utility of  
Database to  
Scientific  
Community

Conclusion

- Phenotypic coverage of all 19,008 human protein-coding genes
- 51% of human protein-coding genes have at least 1 phenotype association in humans
- 58% have orthologs with causal phenotypic associations reported in at least one non-human model
- 40% have annotations in both human and non human orthologs

# Authors Claims on Coverage

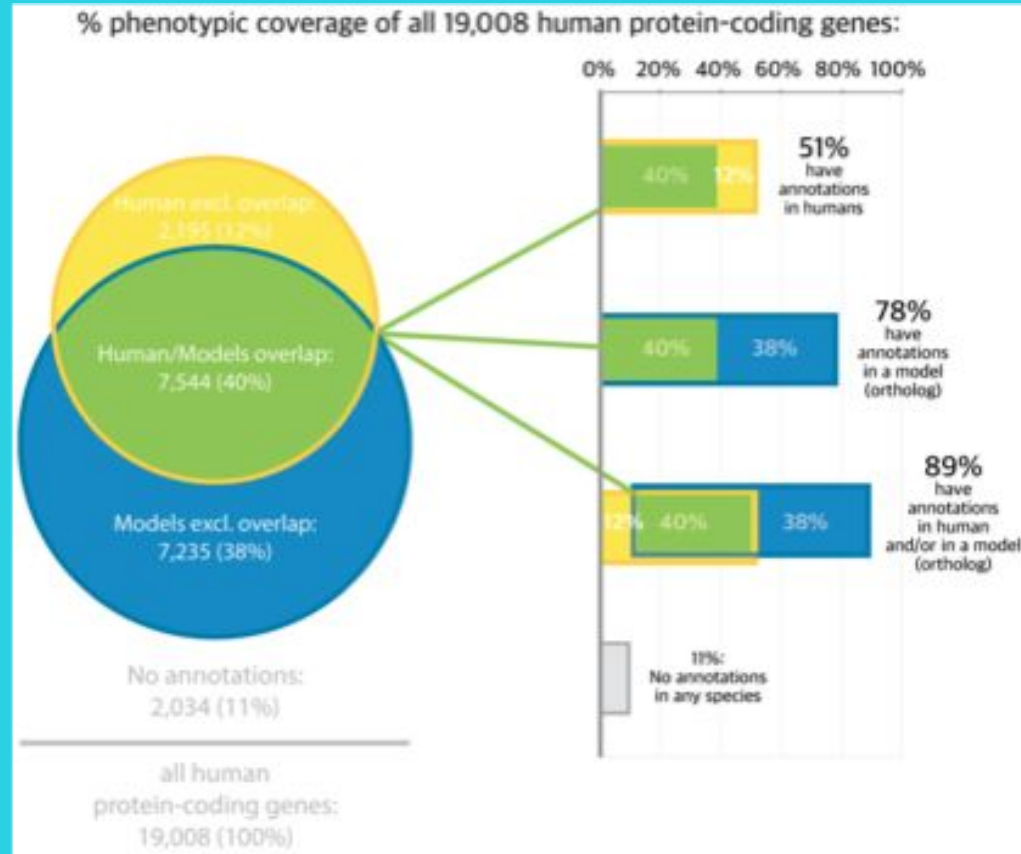
Introduction

Overview of Database

Scientific Quality of Database

General Utility of Database to Scientific Community

Conclusion



# Live Demonstration

Introduction

Overview of  
Database

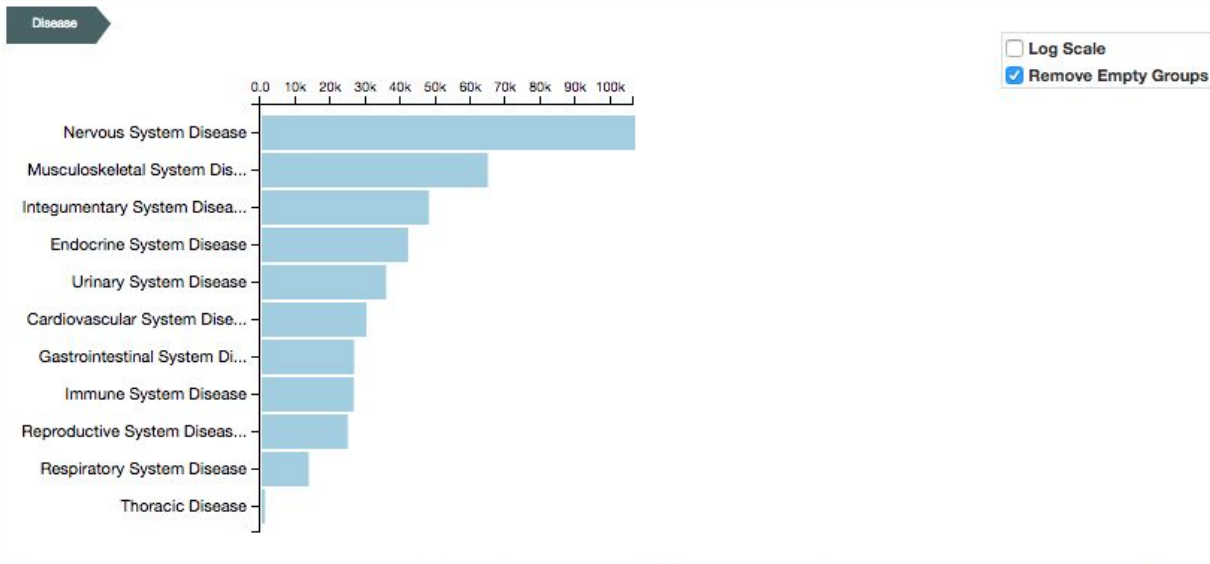
Scientific  
Quality of  
Database

General  
Utility of  
Database to  
Scientific  
Community

Conclusion

## Explore Diseases

### Number of Disease Phenotype Associations



This *interactive graph* displays the number of disease-phenotype associations per major disease category. You can explore the distribution of disease subtypes by clicking on the Y axis labels, or visit the disease page itself by clicking on the corresponding bar. You can navigate back to less-specific disease categories by clicking on the breadcrumbs listed in the top left-hand corner. Organisms can be added and removed from the graph by clicking on the legend labels.



# Review of Content

Introduction

Overview of  
Database

Scientific  
Quality of  
Database

General  
Utility of  
Database to  
Scientific  
Community

Conclusion

Diseases described with phenotypes



Gene variations associated with certain diseases



Phenotype comparisons that suggest models



Models that suggest candidate genes

# Similar Databases?

Introduction

Overview of Database

Scientific Quality of Database

General Utility of Database to Scientific Community

Conclusion

- Some that provide cross-species examinations but none contain specific species used in this database
- Others don't have interactive graphs to display information in a clear and effective manner
- Monarch Initiative compares data already presented in database but WormBase compares DNA sequence you manually input into BLAST

Explore Worm Biology  
facilitating insights into nematode biology

control what you see on the page

skip tutorial

see a star? click on it to save to My Wormbase

News

Check out the [experimental evolution chapter in WormBook!](#) Fri, 20 Sep 2017  
Read about the the goals and major tools of experimental evolution with [C. elegans](#) and related species, in the chapter [Experimental Evolution](#)

Phenotype micropublication submission form now live! Fri, 01 Sep 2017  
Do you have phenotype results that you want to micropublish? A beta version of the phenotype submission form is now available online on [micropublication.org](#).

Gene name changes

Below are changes in gene names since the previous release WS259  
Gene name changes for each release since WS252 are archived [here](#)!

Genes with new primary names

Show	10	entries
New primary name	▲	G
act-1	W	Need help or have feedback...

Welcome to 4Dxpress

This database provides a platform to query and compare gene expression data during the development of the major model animals (zebrafish, drosophila, medaka, mouse). The high resolution expression data was acquired through whole mount *in situ* hybridisation, antibody or transgenic experiments.

Data was integrated from several species specific expression pattern databases, such as ZFIN, BDGP, OXD, MEPD as well as directly submitted by researchers of the participating groups at EMBL.

The name 4Dxpress stands for expression databases in 4D. The 4D (four dimensions) of 4Dxpress can be interpreted either as: 3 spatial dimensions plus time, or as 1 species 2 gene 3 developmental stage 4 anatomical structure

The major focus of this database lies in cross species comparison. To be able to compare gene expression patterns between different species, relationships are needed between:

genes	orthology mapping	The orthology mapping is taken from the Ensembl compara database.
developmental stage	We have mapped species specific developmental stages to a list of stages, which is common to all bilaterian animals: zygote, blastula, gastrula, neurula, segmentation-organogenesis, juvenile, adult.	
anatomical structures	The mapping of anatomical terms is an ambitious task, which has only been started. We will use lexical, orthology-structure and expression information to reveal relationships between the different species.	

# Accessing Sources

Introduction

Overview of  
Database

Scientific  
Quality of  
Database

General  
Utility of  
Database to  
Scientific  
Community

Conclusion

- Pages on genes/phenotypes/diseases/etc contain Uniform Resource Identifier links
- Thirty sources integrated into the database
- All sources are listed under “About” then “Sources”
- No restrictions on access to this database
- Bottom of webpage state a license under a Creative Commons Attribution 3.0 License

# Search and Download

**Introduction**

**Overview of  
Database**

**Scientific  
Quality of  
Database**

**General  
Utility of  
Database to  
Scientific  
Community**

**Conclusion**

- **Generally convenient to browse the data**
- **Single query will retrieve all relevant data from the system**
- **Some broken links under “The Monarch Flow Data”**
- **Downloading can be done by clicking on “About”, then “Data Downloads”**
- **Not the easiest process, as none of the files are .zip**

# Organization and Recommendation

Introduction

Overview of  
Database

Scientific  
Quality of  
Database

General  
Utility of  
Database to  
Scientific  
Community

Conclusion

- Naive user could easily navigate this website
- "About Monarch" page takes you through background, curation process of the data, what you should expect to find upon doing a search
- Web pages are organized with clear headings, use a design that is feels updated
- Designed with colors that help direct and engage a user

**Introduction**

**Overview of  
Database**

**Scientific  
Quality of  
Database**

**General  
Utility of  
Database to  
Scientific  
Community**

**Conclusion**

## **Conclusion**

- **Easy to maneuver**
- **Aesthetically pleasing**
- **Sufficient content**
- **Well organized**
- **Uses credible sources**
- **Professional and hobby database**

# Acknowledgements

- Dr. Dahlquist's Presentation Guidelines
- Presentation Rubric at LMU BIODB 2017
- Dr. Dahlquist and Dr. Dondi

# References

- "Monarch Initiative ." *Welcome to Monarch*, monarchinitiative.org/. Accessed Oct. 2017.
- Mungall, Christopher J., et al. "The Monarch Initiative: an integrative data and analytic platform connecting phenotypes to genotypes across species." *Nucleic Acids Research*, vol. 45, no. D1, 2016, doi:10.1093/nar/gkw1128.
- LMU BioDB 2017. (2017). Week 5. Retrieved October 2, 2017, from [https://xmlpipedb.cs.lmu.edu/biodb/fall2017/index.php/Week\\_5](https://xmlpipedb.cs.lmu.edu/biodb/fall2017/index.php/Week_5)