mondo

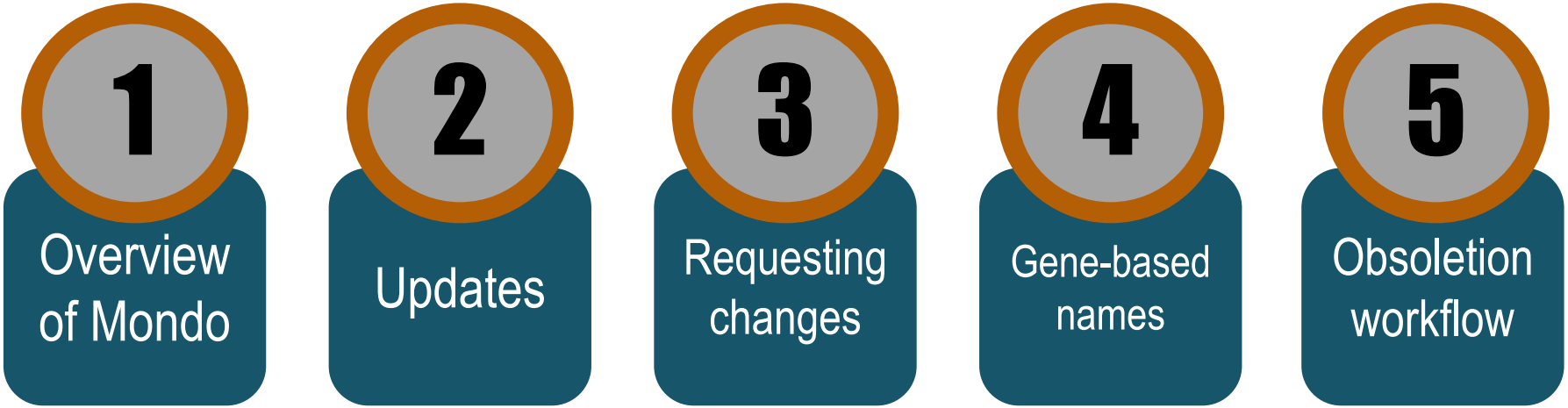
THE WORLD'S DISEASE CONCEPTS, UNIFIED

Nicole Vasilevsky, University of Colorado

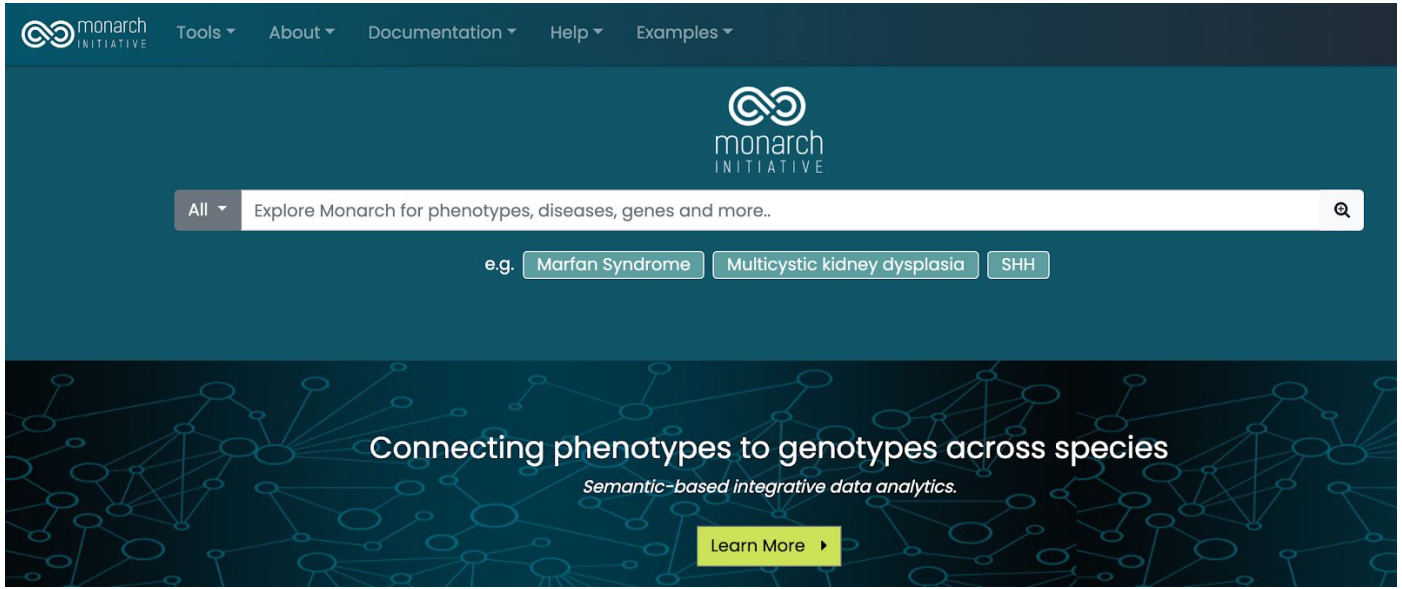
ClinGen Retreat, June 24, 2021

These slides:
https://bit.ly/clingen-ws

Overview


1
Overview of Mondo
2
Updates
3
Requesting changes
4
Gene-based names
5
Obsoletion workflow

Monarch: connecting diseases, phenotypes, and genes



https://monarchinitiative.org/

https://bit.ly/clingen-ws

Mondo development team

Curation

Nicole Vasilevsky

Sabrina Toro

Kallia Panoutsopoulou

Development/Ontology engineering

Chris Mungall

Nico Matentzoglu

Dazhi Jiao

Medical experts

Peter Robinson

Ada Hamosh

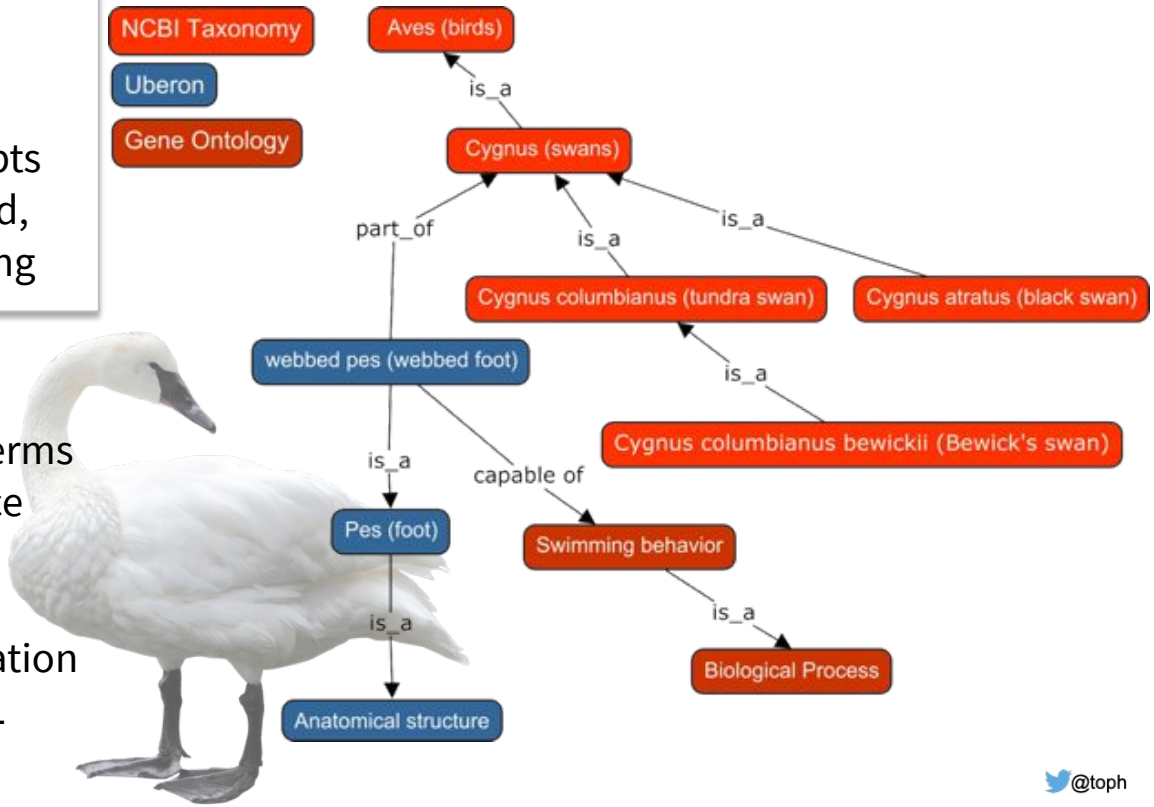
What is an ontology?

on·tol·o·gy
/änˈtäləjē/

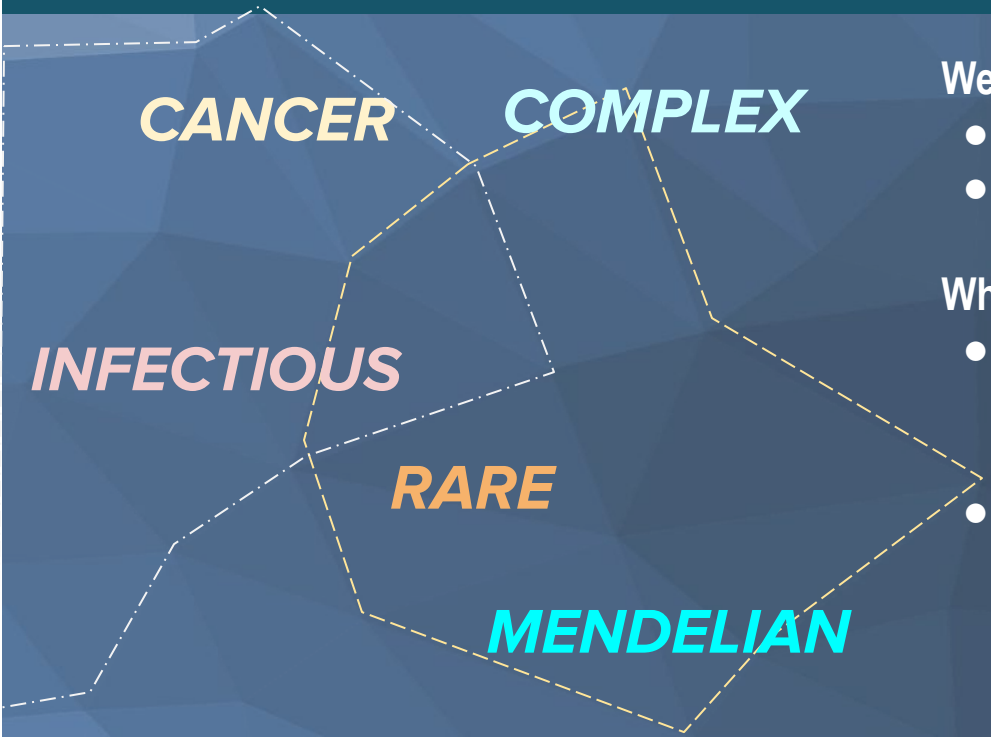
A knowledge classification of a domain, where the relationships between concepts are formally defined and logically related, which allows for computational reasoning

Key Features:

- Terms are defined
- Semantics - relationships between terms are defined, allowing logical inference and sophisticated data queries
- Terms are arranged in a hierarchy
- Expressed in a knowledge representation language such as RDFS, OBO, or OWL



What is the most clinically useful way to define and group diseases?

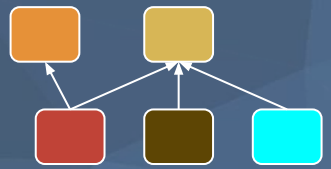


We needed:

- disease concepts spanning multiple categories
- a systematic way of relating these concepts

Why not just use mappings?

- Many terminologies / ontologies / lists include mappings
  - These can be used to cross-walk
- Problems:
  - Often mutually inconsistent
  - N^2 sets of mappings!
  - Not 1:1 equivalents



ANNUAL REVIEWS
For Librarians & Agents For Authors
JOURNALS A-Z JOURNAL INFO

Home / Annual Review of Biomedical Data Science / Volume 1, 2018 / Haendel, pp 305-331

A Census of Disease Ontologies

Annual Review of Biomedical Data Science
Vol. 1:305-331 (Volume publication date July 2018)
First published as a Review in Advance on May 9, 2018
https://doi.org/10.1146/annurev-biodatasci-080917-013459

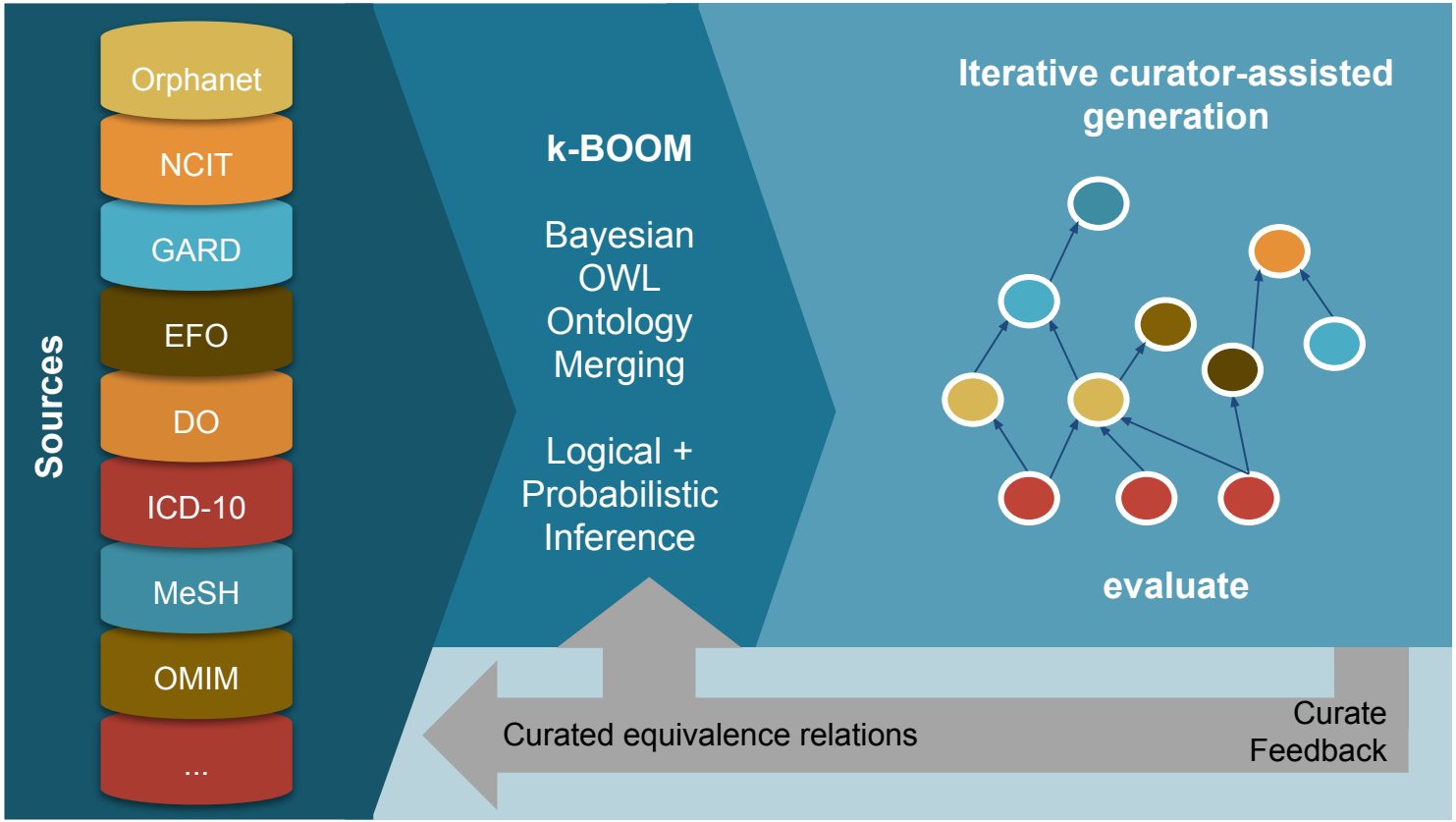
Melissa A. Haendel,[1,2] Julie A. McMurry,[1] Rose Relevo,[1] Christopher J. Mungall,[3] Peter N. Robinson,[4] and Christopher G. Chute[5]

[1]Department of Medical Informatics and Clinical Epidemiology, Oregon Health and Science University, Portland, Oregon 97239, USA; email: haendel@ohsu.edu
[2]Linus Pauling Institute, Oregon State University, Corvallis, Oregon 97331, USA
[3]Environmental Genomics and Systems Biology, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA
[4]The Jackson Laboratory, Farmington, Connecticut 06032, USA
[5]School of Medicine, School of Public Health, and School of Nursing, Johns Hopkins University, Baltimore, Maryland 21205, USA

No one system fit the bill

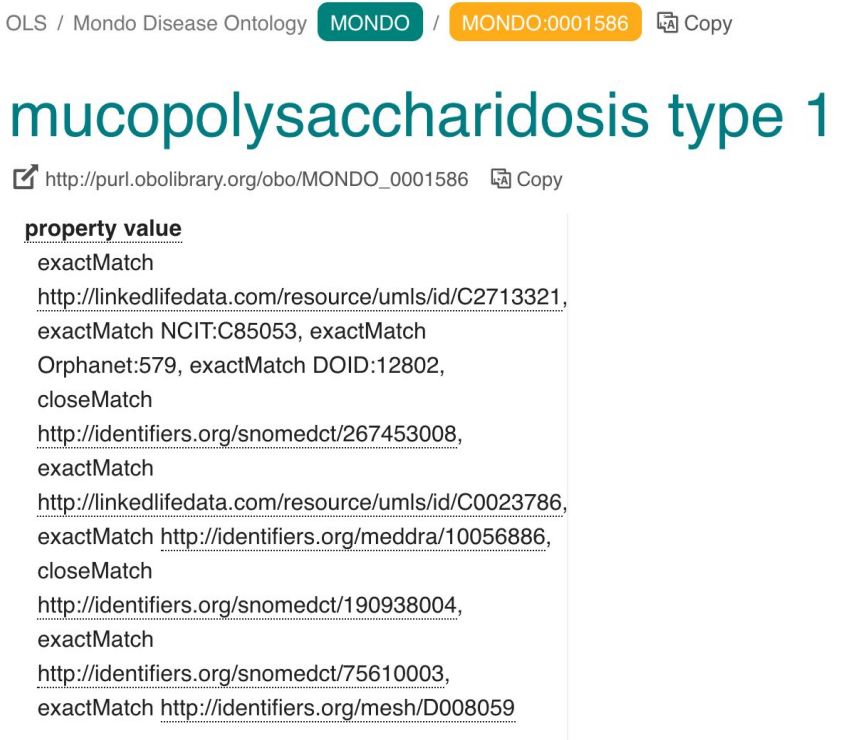
- There are a wealth of disease resources
  - *specialized*
    - (e.g OMIM: Mendelian)
  - *generalized*
    - (e.g. MESH, SNOMED)
- No single resource sufficient
  - *specialized*
    - did not include concepts we needed
  - generalized
    - lacked sufficient depth/precision in key domains

Evidence-based merging of equivalent classes



Relationships to other resources

- In obo format version and in OLS these are database cross references (‘xrefs’)
- In owl version we use explicit *logical axioms*, e.g:
  - **equivalentTo**
  - **relatedTo**



Mondo status

MONDO IDs assigned and tracked for each concept

Use of standard ontology engineering practices

Periodically aligned and synced with existing resources

Released monthly (obo, owl, json)

61 releases

Total number of classes **24,297**

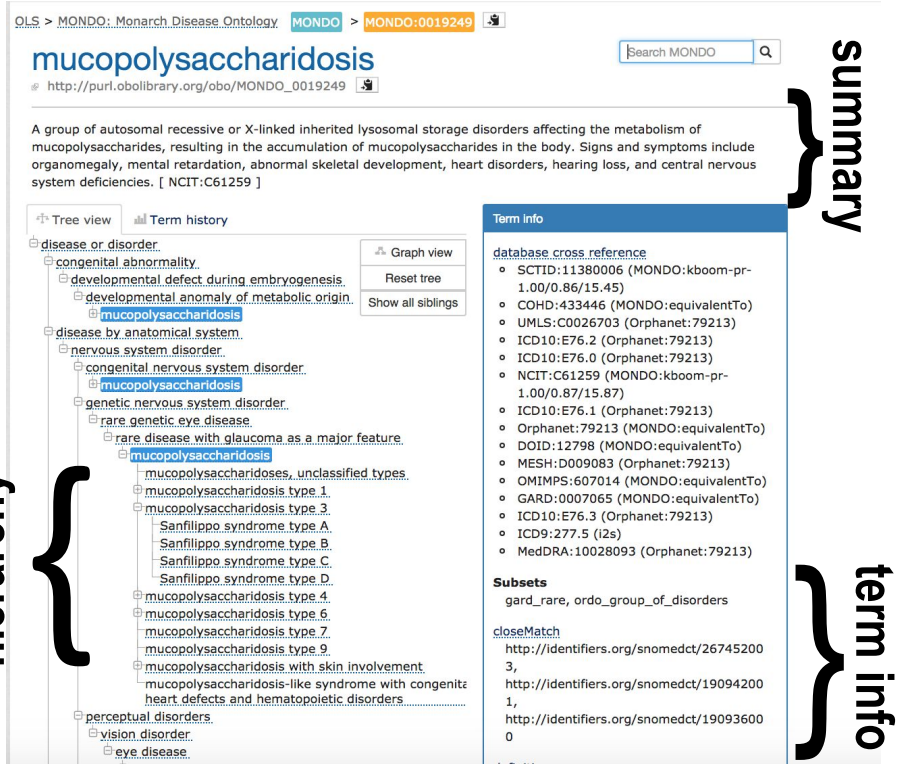
Unique classes (non-obsolete) **22,332**

Classes with definitions **15,284** (68%)

Classes without definitions **7,048** (32%)

Synonym Type	Count
Has exact synonym	17,815
Has related synonym	11,663
Has broad synonym	281
Has narrow synonym	298
No synonym	4,489
Total	**34,546**

Where to view Mondo



https://www.ebi.ac.uk/ols/ontologies/mondo

OBO Foundry
obofoundry.org/ontology/mondo

Ontology Lookup Service
ebi.ac.uk/ols/ontologies/mondo

GitHub
github.com/monarch-initiative/mondo
>1300 issues reported

Protege
download OWL file from GitHub

Mondo community

GitHub
github.com/monarch-initiative/mondo
>1500 issues reported

Mondo users list
https://groups.google.com/forum/#!forum/mondo-users

Email
nicole@tislab.org

Community developed

Weekly Calls

Thursdays, 10am PT/1pm ET
Zoom

Workshops

Next workshop July 13, 2021
10am-12pm PT
1pm-3pm ET

Revised top level hierarchy

Goal:
Revise the **top-level classification** in a way that is more intuitive for **browsing the hierarchy**

Previous classification

disease or disorder
- Mendelian disease
- acute disease
- cell proliferation disorder
- congenital abnormality
- connective tissue disease
  - Rheumatologic disorder
- disease by anatomical system
- disease by subcellular system affected
- disorder by anatomical region
- disorder involving pain
- disorder of development or morphogenesis
- iatrogenic disease
- infectious disease
- inflammatory disease
- injury
- non-human animal disease
- nutritional or metabolic disease
- perinatal disease
- post-infectious disorder
- pregnancy disorder
- puerperal disorder
- radiation or chemically induced disorder
- regional odontodysplasia
- segmental odontomaxillary dysplasia
- serpinopathy
- syndromic disease
- systemic or rheumatic disease
- viral disease or post-viral disorder

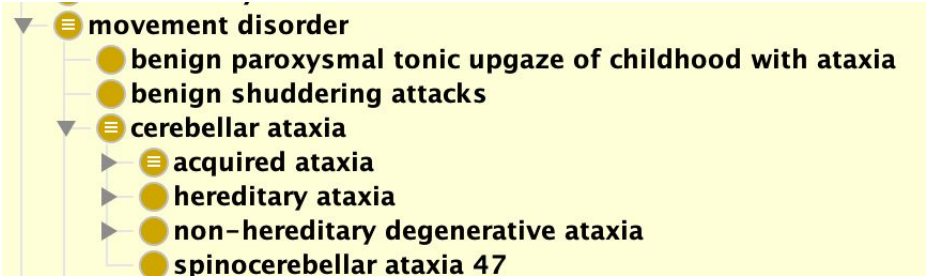
Revised classification

disease or disorder
- cell proliferation disorder *
- disorder of development or morphogenesis
- **'emergency medicine condition' ***
  - injury
  - radiation or chemically induced disorder
- endocrine system disease
- iatrogenic disease
- **idiopathic disease**
- infectious disease
  - post-infectious disorder
  - viral disease or post-viral disorder
- inflammatory disease
- Mendelian disease
  - heritable connective tissue disease
- metabolic disease
- non-human animal disease
- nutritional disease
- **obstetric disorder**
  - pregnancy disorder
  - puerperal disorder
- syndromic disease
- systemic or rheumatic disease
  - post-infectious disorder
  - Rheumatologic disorder

* to be renamed
Blue : Split class
Bold: new class

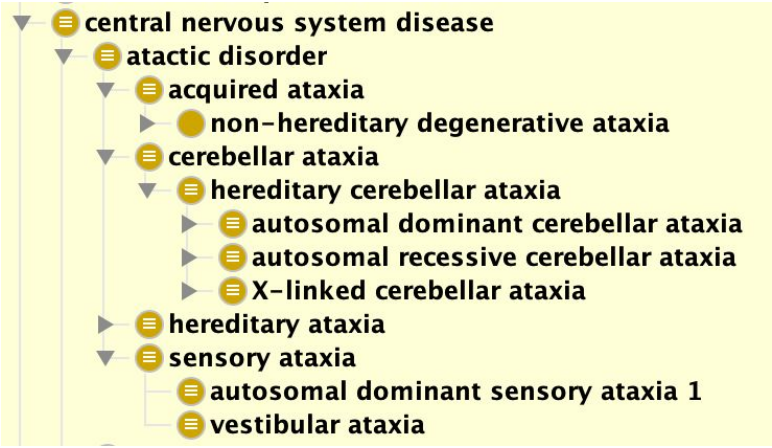
Atactic disorder

Previous classification



- not all ataxias are cerebellar

Revised classification

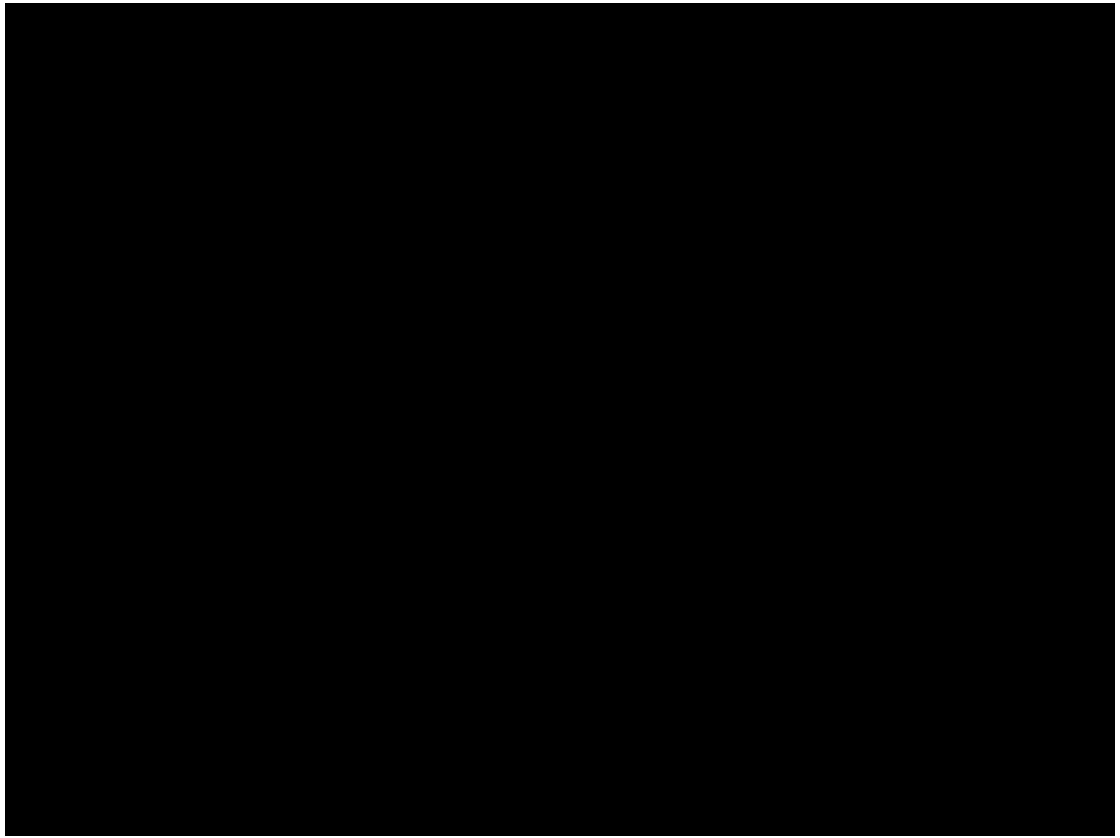


- added grouping class (atactic disorder)
- classification by acquired/hereditary
- classification by mechanism (cerebellar, sensory, vestibular)

https://github.com/monarch-initiative/mondo/issues/2866

How to request new terms (Mondo IDs) & changes

1. GitHub tracker: New issue
2. Pick appropriate template
3. Fill in the information that is requested on the template below each header
4. Please include:
    a. A definition in the proper format
    b. Sources/cross references for synonyms
    c. Your ORCID or the URL for your ClinGen working group
    d. Add any additional comments at the end
5. Nicole will automatically be tagged
6. Please email Nicole or comment on the ticket (Nicole will be emailed) if you have any additional questions or need the ticket is high priority



Recommendations for GitHub tickets/new term requests

We appreciate your contributions to extending and improving Mondo

General Recommendations:

1. New term requests should not match existing terms or synonyms
2. Write a concise definition in the definition field. More info about writing definitions is here
3. Synonyms - please provide a source/cross-reference
4. Check OMIM for children classes

Formatting:

1. Preferred term labels should be lowercase (unless it is a proper name or abbreviation)
2. Write the request below the prompts on the template so the Markdown formatting displays properly
3. Synonyms should be lowercase (with exceptions above)
4. Definition source - if from PubMed, please use the format PMID:XXXXXX (no space)
5. Include the Mondo ID and label for the parent term
6. List the children terms with Mondo ID and label in a bulleted list

Writing Ontology Definitions

Guidelines for writing definitions in ontologies

Selja Seppälä
Ph.D., University of Geneva, Switzerland
Postdoctoral Associate, University of Florida, Gainesville, United States.
http://seljaseppala.wordpress.com
E-mail: sseppala@ufl.edu

Alan Ruttenberg
M.S. Massachusetts Institute of Technology
Director of Clinical Data Exchange, School of Dental Medicine,
University at Buffalo, United States.
http://alan.ruttenbergs.com
E-mail: alanruttenberg@gmail.com

Barry Smith
Ph.D., University of Manchester, United Kingdom.
Professor of Philosophy, University at Buffalo, Buffalo, United States.
http://ontology.buffalo.edu/smith/
E-mail: phismith@buffalo.edu

STRUCTURE OF A TEXTUAL DEFINITION

When used to refer to the natural language text of a definition, the term 'definition' itself can denote different forms: a sentence and a sentence fragment. Broadly, a definition has the canonical form X is a Y that Zs as in example (**1**) adapted from the definition of 'ligament' (synonym of 'skeletal ligament') in the Uberon multi-species anatomy ontology (UBERON).

(1) A **ligament** is a dense regular connective tissue connecting two or more adjacent skeletal elements.

Definitions in this form have a three-part structure:

1. a **definiendum** [*X*], i.e., the defined term;
2. a **definiens** [*a Y that Zs*], i.e., the part that expresses the definition content and that is called a **definition** in dictionaries;
3. a **copula** [*is*] that expresses an equivalence between definiendum and definiens.

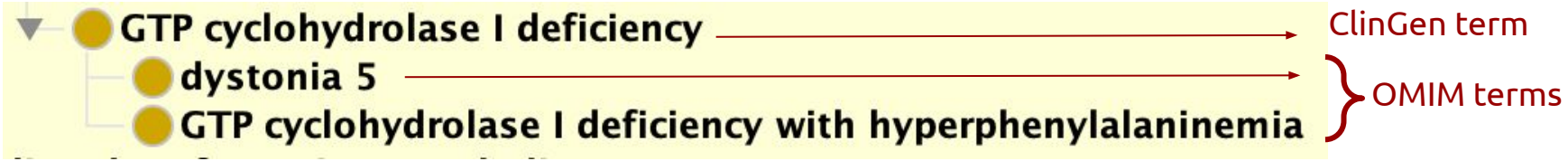
https://philpapers.org/archive/SEPGFW.pdf

https://bit.ly/clingen-ws

Gene-based names

For example, 'GTP cyclohydrolase I deficiency'

We are able to accomodate gene-based names



Synonym types

Scope	**Exact**	**An exact match**	E.g. hereditary Wilms' tumor exact synonym: familial Wilms' tumor
	Narrow	**A more specific term**	E.g. asthma narrow synonym: exercise-induced asthma
	Broad	**A more general term**	E.g. autoimmune hepatitis broad synonym, autoimmune liver disease
	Related	**A word of phrase has been used synonymously with the primary term name in the literature, but the usage is not strictly correct**	E.g. AGAT deficiency related synonym: disorder of glycine amidinotransferase activity
Type	**Excluded**	Some synonyms are annotated with EXCLUDE, e.g. "NOS" (not otherwise specified) synonyms. It is useful to have these in the edit version, but these are filtered on release.	
	Deprecated	We may also mark synonyms with DEPRECATED. E.g. all occurrences of "mental retardation" should be "intellectual disability"	

https://mondo.readthedocs.io/en/latest/editors-guide/f-entities/#synonyms

Clingen preferred synonyms

craniosynostosis 4

http://purl.obolibrary.org/obo/MONDO_0010929 Copy

has_exact_synonym [type: xsd:string]

ERF-related craniosynostosis

database_cross_reference [type: xsd:string]

PMID:23354439

database_cross_reference [type: xsd:string]

https://clinicalgenome.org/affiliation/40059/

has_synonym_type

'clingen preferred'

Term information -

database cross reference

- UMLS:C3806917 (OMIM:600775)
- OMIM:600775 (MONDO:equivalentTo)

clingen preferred

ERF-related craniosynostosis [

http://www.ncbi.nlm.nih.gov/pubmed/23354439

https://clinicalgenome.org/affiliation/40059/]

abbreviation

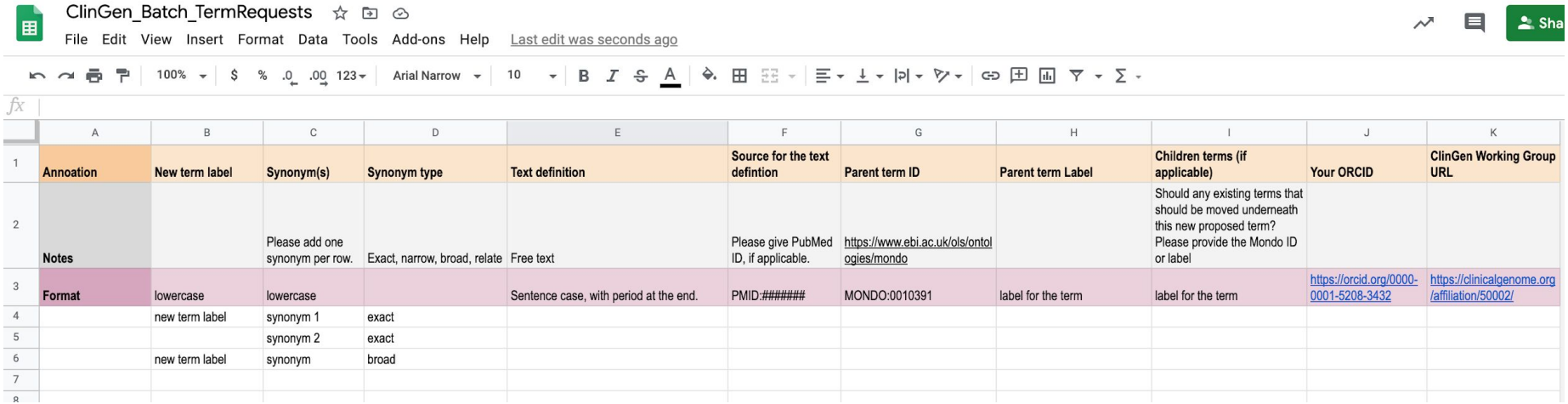
CRS4 [MONDO:Lexical

https://omim.org/entry/600775]

https://mondo.readthedocs.io/en/latest/editors-guide/f-entities/#synonyms

Batching requests

Created a google spreadsheet with the metadata we need to capture:
https://bit.ly/clingen-batch

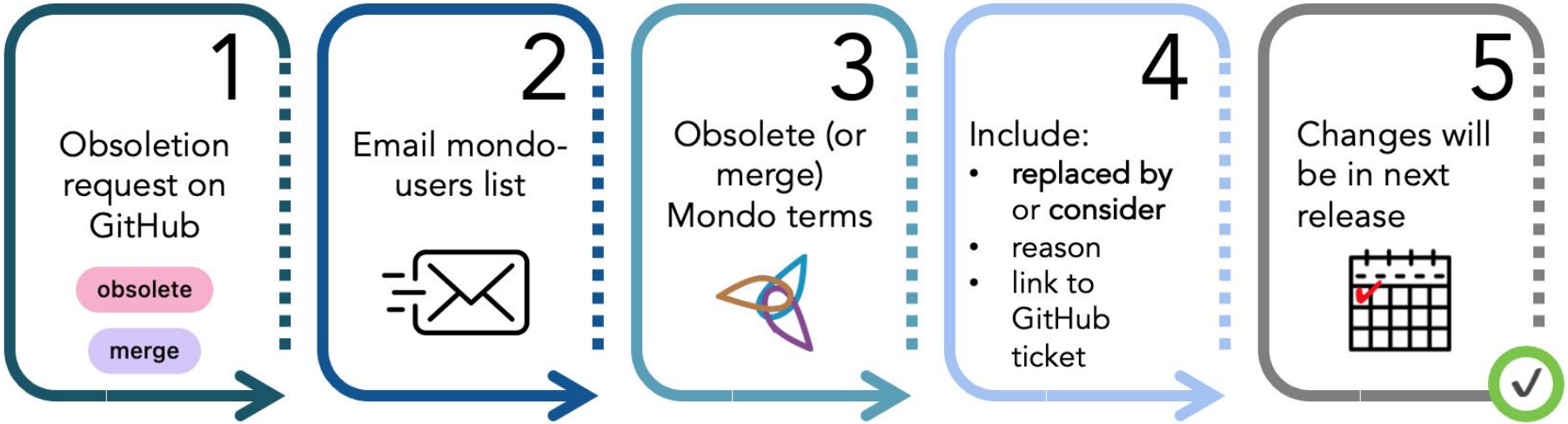


ClinGen_Batch_TermRequests

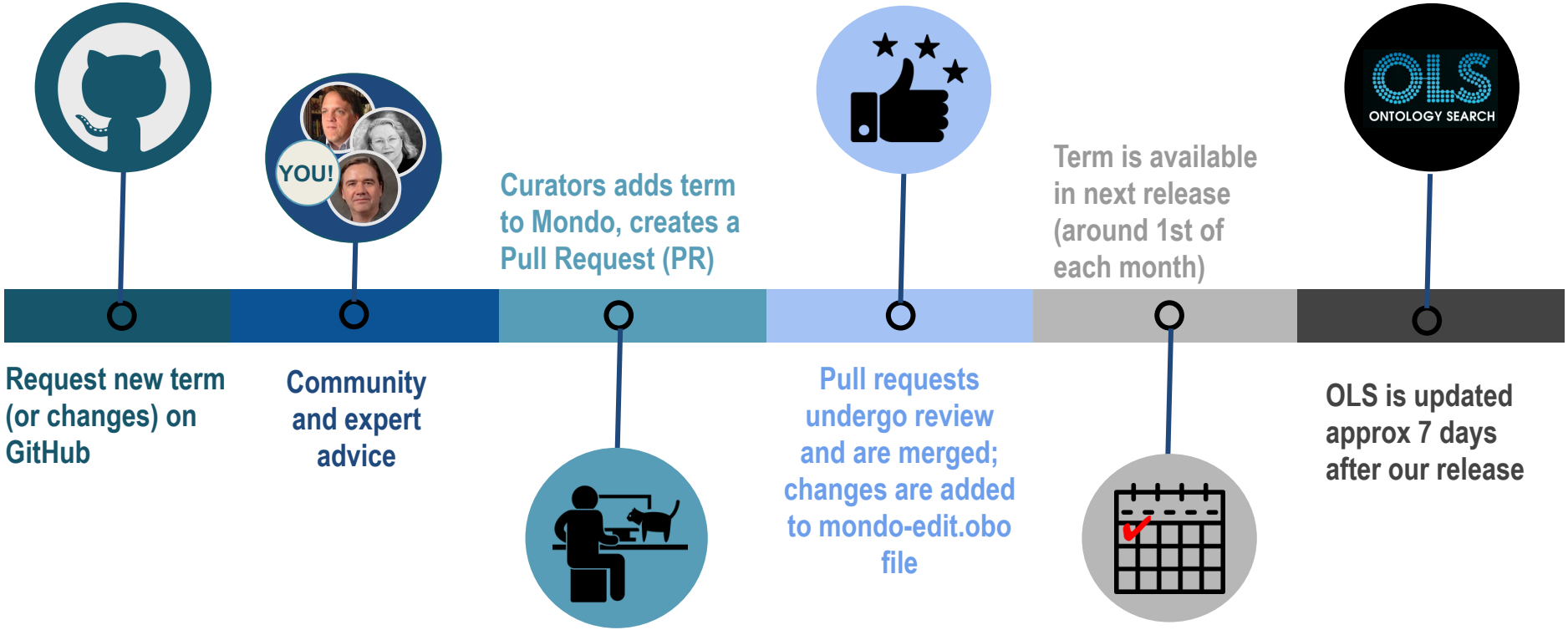
File Edit View Insert Format Data Tools Add-ons Help Last edit was seconds ago

	A	B	C	D	E	F	G	H	I	J	K
1	**Annoation**	**New term label**	**Synonym(s)**	**Synonym type**	**Text definition**	**Source for the text defintion**	**Parent term ID**	**Parent term Label**	**Children terms (if applicable)**	**Your ORCID**	**ClinGen Working Group URL**
2	**Notes**		Please add one synonym per row.	Exact, narrow, broad, relate	Free text	Please give PubMed ID, if applicable.	https://www.ebi.ac.uk/ols/ontologies/mondo		Should any existing terms that should be moved underneath this new proposed term? Please provide the Mondo ID or label		
3	**Format**	lowercase	lowercase		Sentence case, with period at the end.	PMID:#######	MONDO:0010391	label for the term	label for the term	https://orcid.org/0000-0001-5208-3432	https://clinicalgenome.org/affiliation/50002/
4		new term label	synonym 1	exact							
5			synonym 2	exact							
6		new term label	synonym	broad							
7											

Obsoletion workflow



Requesting changes to Mondo



mondo
THE WORLD'S DISEASE CONCEPTS, UNIFIED

University of Colorado Anschutz Medical Campus

Nicole Vasilevsky, CU
Email: nicole@tislab.org
Twitter: @n_vasilevsky
GitHub: nicolevasilevsky

Sabrina Toro, CU
Email: sabrina@tislab.org
GitHub: sabrinatoro

Nico Matentzoglu, Semanticly
Email: nicolas.matentzoglu@gmail.com
Twitter: @NicoMatentzoglu
GitHub: matentz

Chris Mungall, LBNL
Email: cjmungall@lbl.gov
Twitter: @chrismungall
GitHub: cmungall

Peter Robinson, Jackson Labs
Email: peter.robinson@jax.org
Twitter: @pnrobins
GitHub: pnrobinson

Melissa Haendel, CU
Email: melissa@tislab.org
Twitter: @ontowonka
GitHub: mellybelly

Funding:
Monarch is supported generously by a **NIH Office of the Director** Grant #5R24OD011883 and the Phenomics First Resource, **NIH-NHGRI** # 1 RM1 HG010860–01, a Center of Excellence in Genomic Science.

Big thanks to our contributors!

mondo
THE WORLD'S DISEASE CONCEPTS, UNIFIED

Broad Institute
Samantha Baxter
Andrew Grant
Jessica Hekman
Madeline Hughes
Kate Megquier
Kathy Reinold
Rebecca Siegert

CHOP
Colin Ellis
Allison Heath
Ingo Helbig
Avi Kelman

CoRDS-Sanford
Austin Letcher

ClinGen
Larry Babb
Taylor Bingaman
Marina DiStefano
Jenny Goldstein
Brooke Palus
Heidi Rehm
Erin Riggs
Tam Sneddon
Courtney Thaxton
Matt Wright
ClinGen Expert panels

EBI
Mélanie Courtot
Simon Jupp
David Osumi-Sutherland
Zoë Pendlington
Paola Roncaglia

GARD
Gioconda Alyea
PJ Brooks
Maria Della Rocca
Janine Lewis
Anne Pariser
Andrea Storm

Monarch Initiative
Melissa Haendel
Leigh Carmody
Shahim Essaid
Nomi Harris
Nico Matentzoglu
Julie McMurry
Moni Munoz-Torres
Peter Robinson
Kent Shefchek
Anne Thessen

NCIt
Gilberto Fragoso
Bron Kisler

NIH NCATS
Alice Chen
Eric Sid

NCBI
Donna Maglott

Johns Hopkins
Christopher Chute
Dazhi Jiao

NIH NHGRI
Robert Fullem
Morgan Similuk

NORD
Vanessa Boulanger

OMIM
Joanna Amberger
Ada Hamosh

Orphanet
Marc Hanauer
Annie Olry
Ana Rath

University of Colorado
Tiffany Callahan