The dangers of bad mappings: How imprecise and incomplete mappings can cost lives

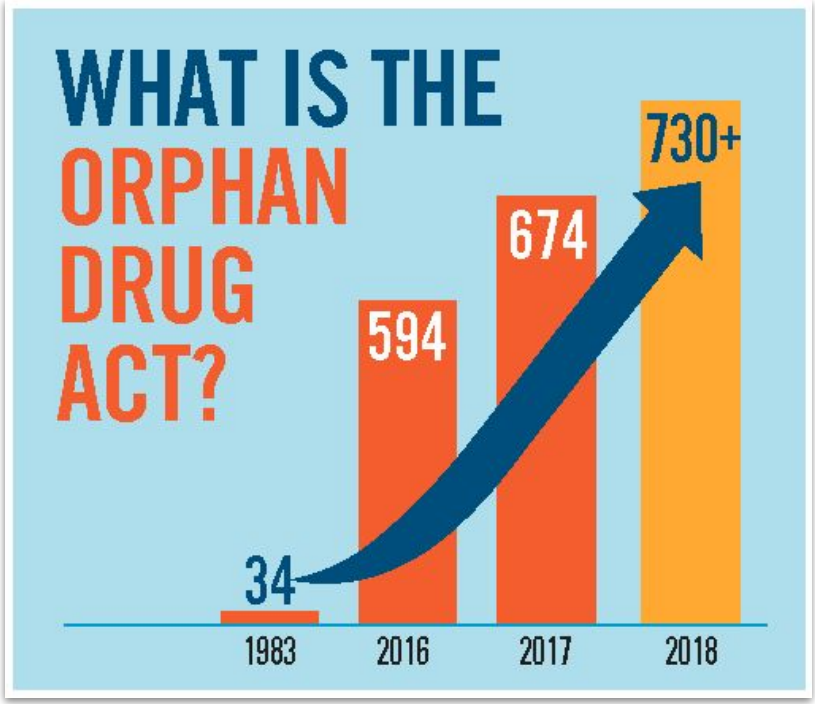
WSBO-2021: Workshop on Synergizing Biomedical Ontologies

July 14th, 2021

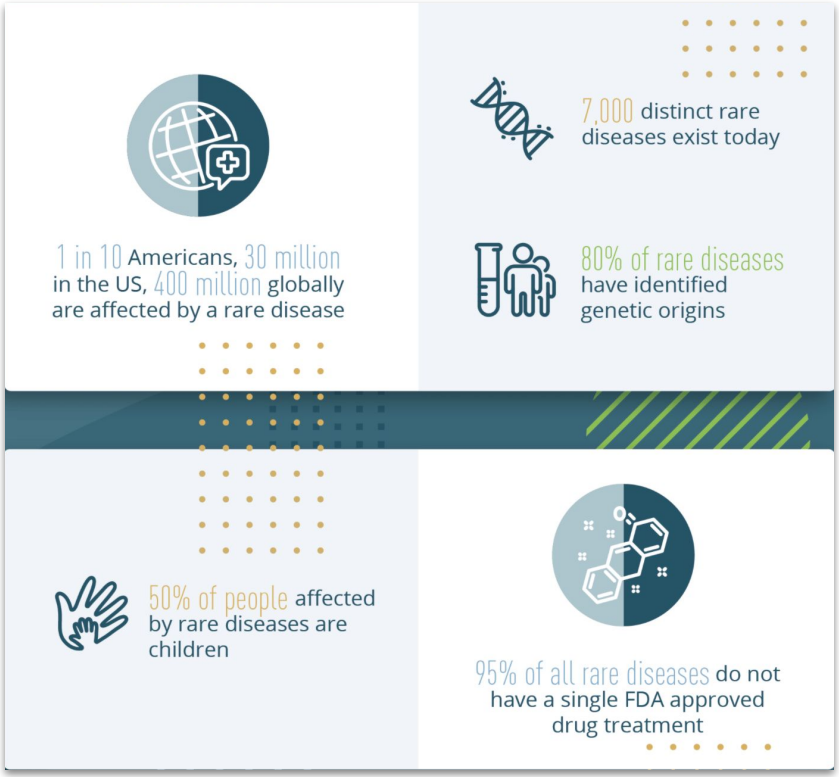
These slides: bit.ly/wsbo-raredisease

@ontowonka

Orphan Drug Act of 1983



https://rarediseases.org/orphan-drug-act-resolution-introduced-in-congress



https://globalgenes.org/rare-facts

"Nearly 1 out of every 10 Americans, lives with at least 1 of more than 7,000 known rare diseases"

Why the number of rare diseases is hard to determine (and is not 7000)

Criteria for “rare” vary around the world:

- 1983: From the Orphan Drug Act: A rare disease affects fewer than 200,000 people
- 2000: European Union considers a disease to be rare when it affects fewer than 1 in 2,000 people

New diseases discovered all the time, but # not updated:

- New rare diseases are discovered every week by organizations such as the Undiagnosed Disease Network
- The literature, OMIM, etc. abound by new weekly entries
- N-of-1s are matched, defining new diseases in systems like the Matchmaker Exchange

Disease definitions vary around the world

- Dozens of terminologies and disease registries exist
- These are often not included in clinical terminologies (such as ICD) commonly used in EHRs
- Fundamentally, the definition of a rare disease and how to model it computationally has remained more an art than a science

Why do I care about this?

Not being able to count the number of rare diseases means we don’t have clear definitions

This makes it very hard to build diagnostic tools and reveal mechanisms

→ Yes, bad mappings can cost lives!

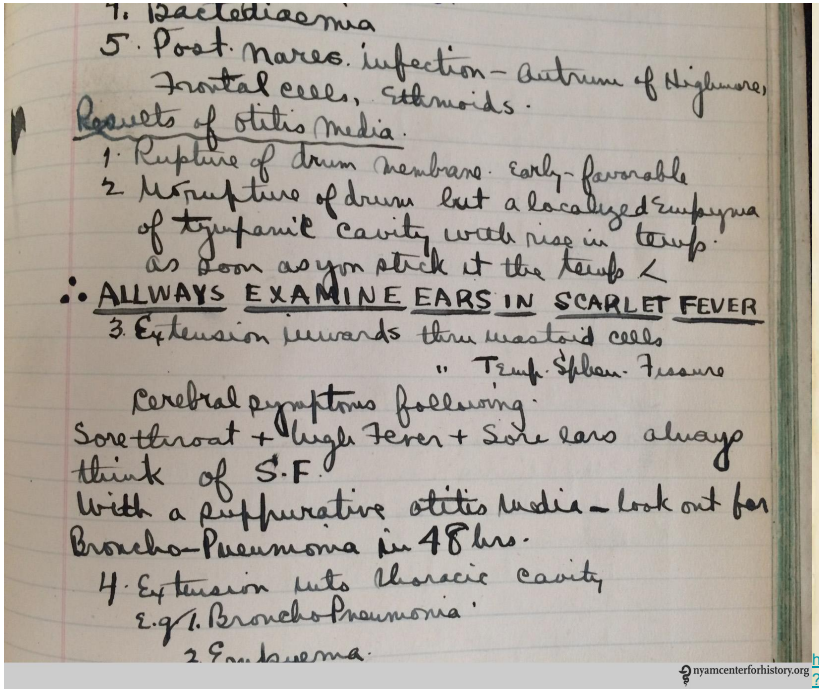
Jessica's story:
Evidence-based diagnosis requires evidence

Jessica (aged 4) has a rare condition which causes epilepsy, affects her movement and developmental delay. Standard genetics tests negative.

To solve her case requires the ability to compare Jessica to a multitude of other available data, both from humans and from other animals.

Genomics england

Patient phenotyping is ubiquitous, but hard to document well



https://www.popsci.com/article/science/3-gross-cool-vintage-anatomy-drawings/?image=2

We have a long way to go in how we computably define diseases

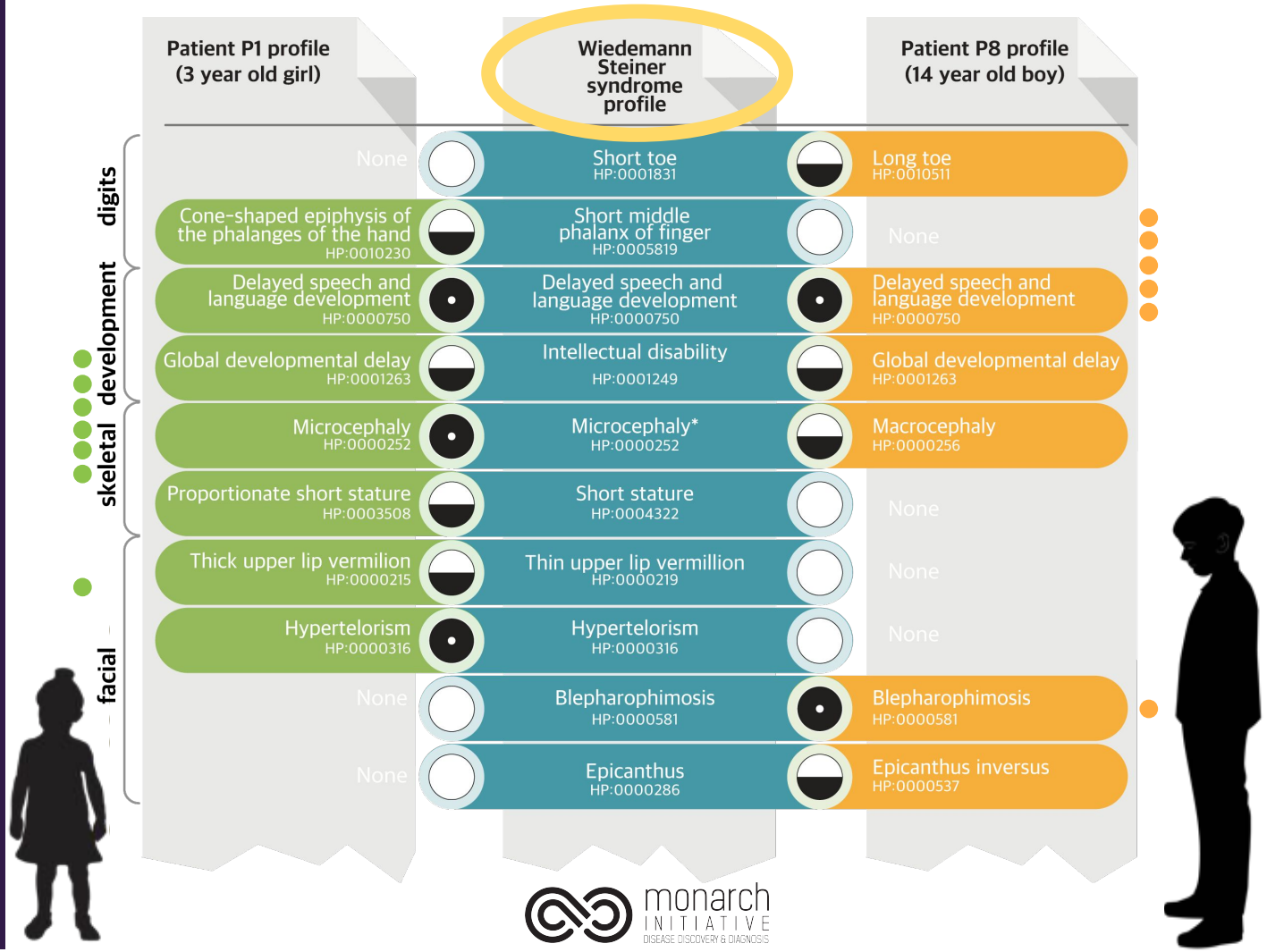
Fuzzy Phenotype Matching

Not same variant, but same disease and gene, KMT2A.

DOI: 10.1126/scitranslmed.3009262

Legend

- Perfect Match
- Fuzzy Match
- No Match



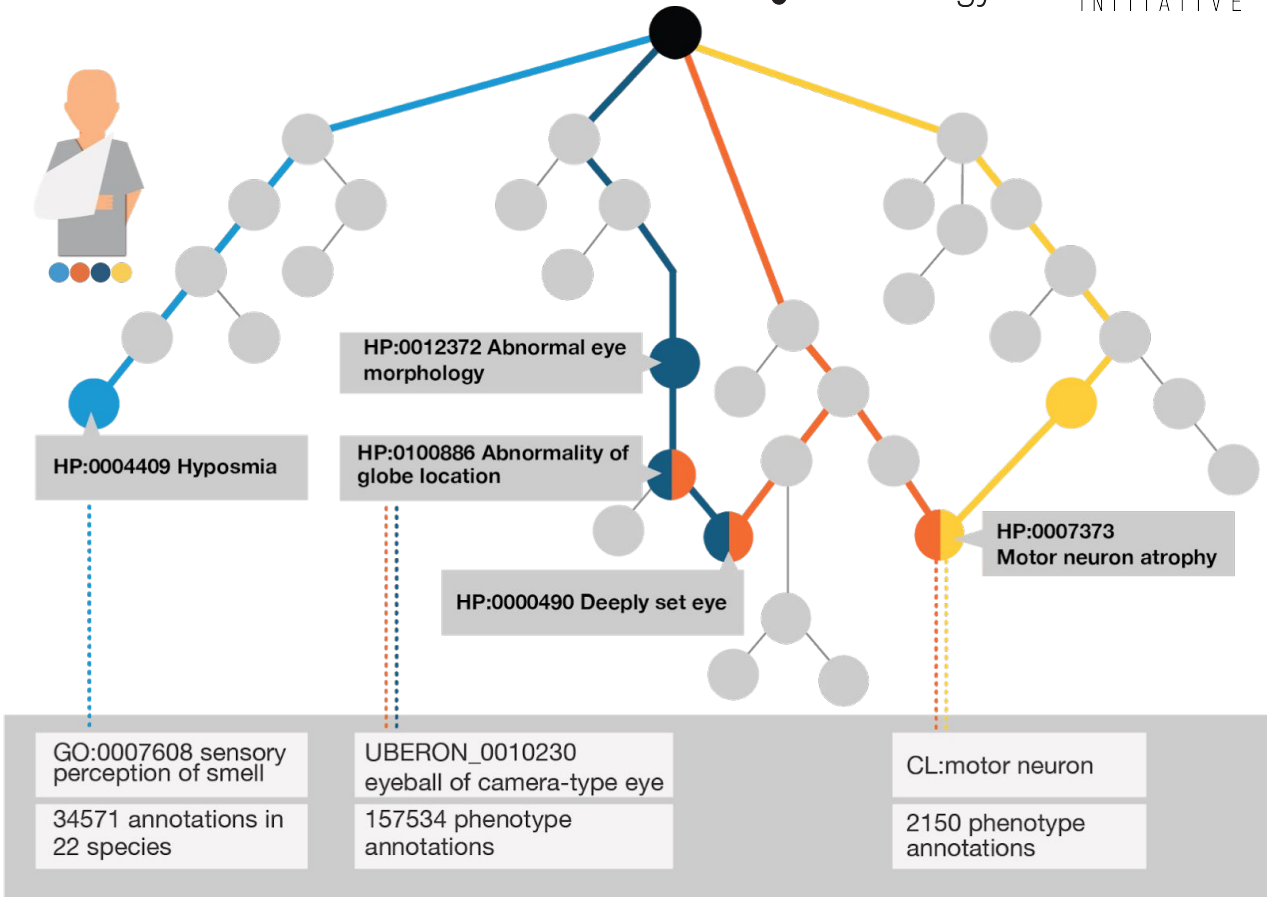
Human Phenotype Ontology (HPO)

- **Phenotyping terminology**
  >14,500 terms
- **Computational disease models**
  >190,000 disease-phenotype annotations
- **Widely adopted in rare disease genomic diagnostic tools**
  100,000 Genomes Project, SOLVE-RD, NIH-UDP, etc.

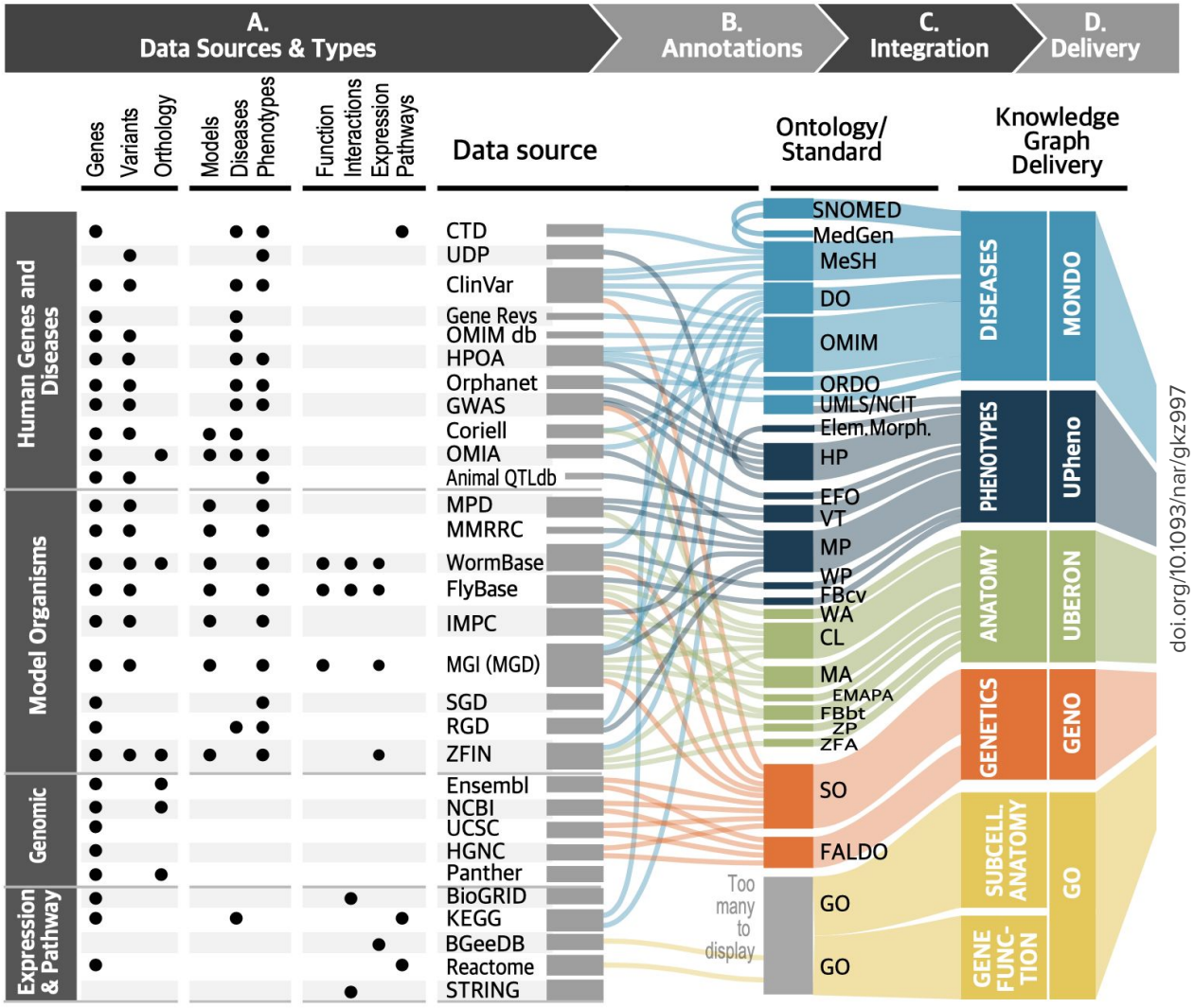
hpo.jax.org

human phenotype ontology

monarch INITIATIVE

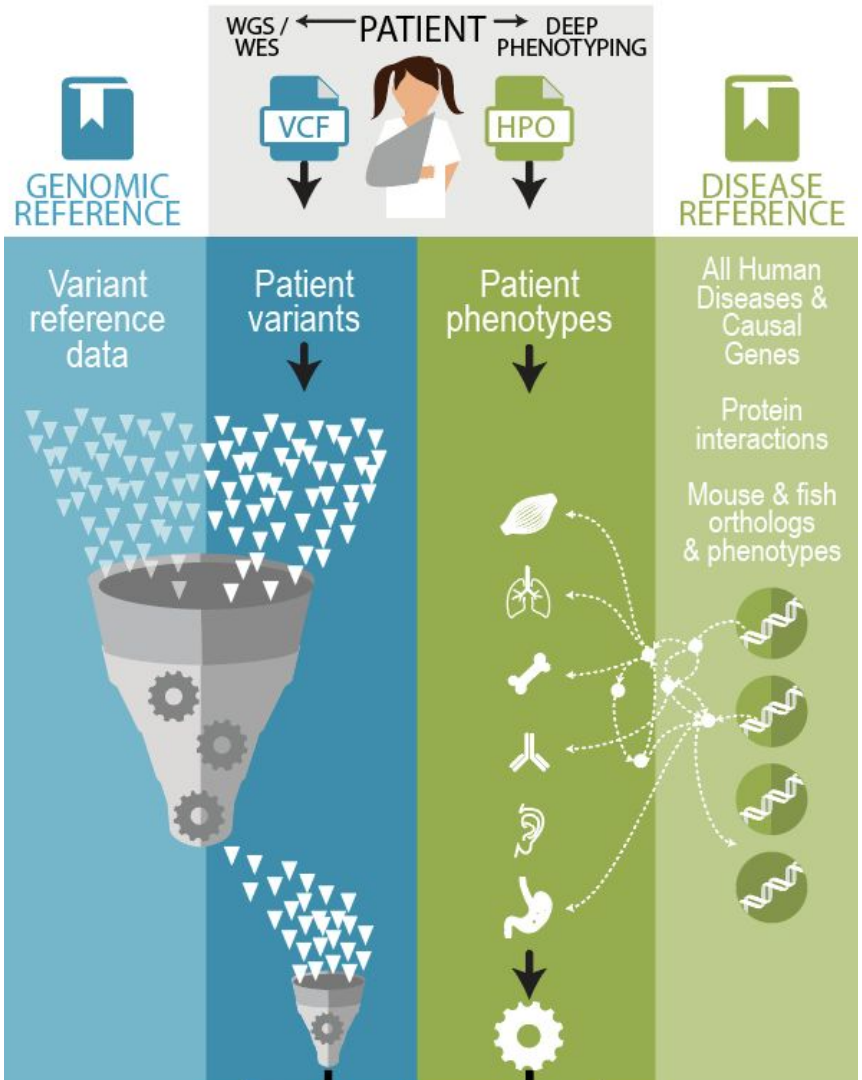


Data integration across sources and data types involves complex identifier management



Combining evidence from genomic, and phenomic, and other data improves variant prioritization for diagnosis

doi: 10.1038/gim.2015.137

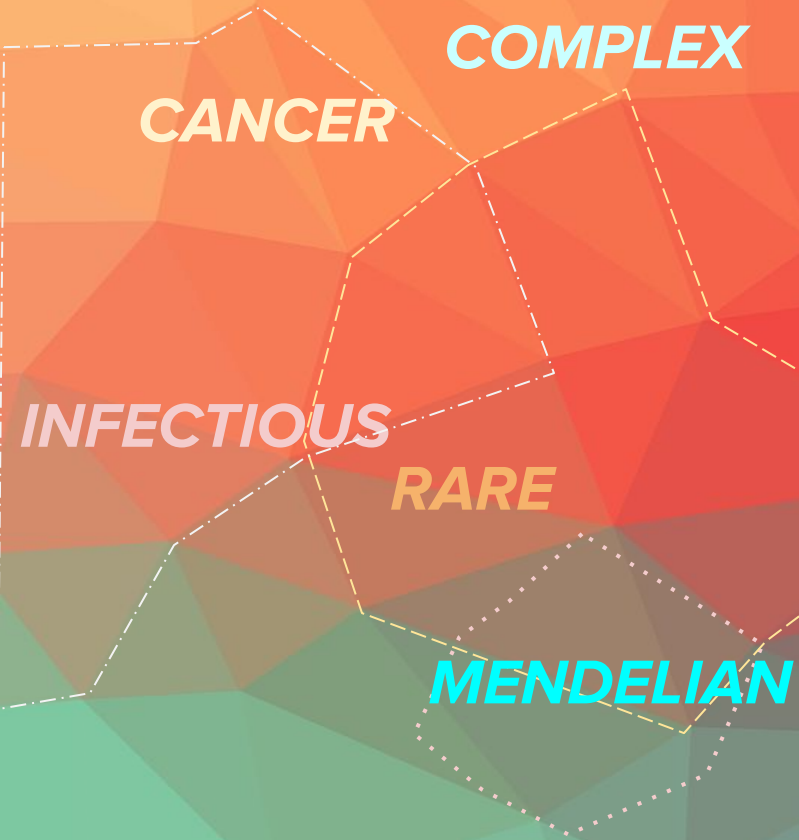


EXOMISER

But:

Combining data relating to different disease terminologies is hard

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

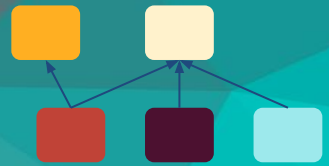


We need:

- 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



Assessment of synonyms, hierarchies, and mappings across ontology sources for example diseases EDS and Pancreatic cancer

Source ontology	General information			Ehlers–Danlos syndrome				Pancreatic cancer			
	Has definition[b]	Hierarchy type	Language translation	Xrefs	Synonyms	Parents	Descendants	Xrefs	Synonyms	Parents	Descendants
DO	Yes	Single	No	6/11	0/2	1/1	6/6	4/11	0/9	2/2	37/37
ICD-10	~Yes[c]	Single	No	–	–	1/1	–	–	–	1/1	–
ICD-11	Yes	Multi	Planned	–	1/2	2/2	26/26	–	2/7	3/3	23/23
MedDRA	No	Multi	Yes	–	–	2/2	–	–	–	1/2	–
MeSH	~Yes[d]	Single	Yes	–	22/23[e]	3/4	16/21	–	12/25[e]	3/3	14/14
MonDO	Yes	Multi	Planned	–	0/2	7/7	36/36	–	8/8	2/2	44/44
NCIt	Yes	Multi	No	–	–	2/2	7/7	–	3/3	2/2	68/68
OMIM[f]	Yes	Flat	No	–	7/7	–	–	2/2	–	–	–
Orphanet	No	Multi	No	2/5	–	6/6	24/24	5/5	0/2	7/7	10/10
SNOMED	N	Multi	Yes	–	2/11	8/8	21/21	–	5/5	3/3	47/47
UMLS	No	Multi	Yes	–	27/46	2/2	7/8	–	50/62	–	6/6

Wide heterogeneity in:

- placement of diseases hierarchically, and therefore meaning,
- mapping to other diseases,
- the number and typing of synonyms

2018 statistics shown from this manuscript:
http://bit.ly/annual-reviews-ontologies

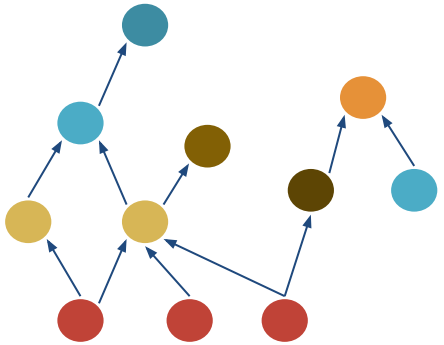
Evidence-based merging of equivalent disease concepts

k-BOOM

Bayesian OWL Ontology Merging

Logical + Probabilistic Inference

iterative curator-assisted equivalence inference



evaluate

mondo

Curated Equivalence Relations

curate

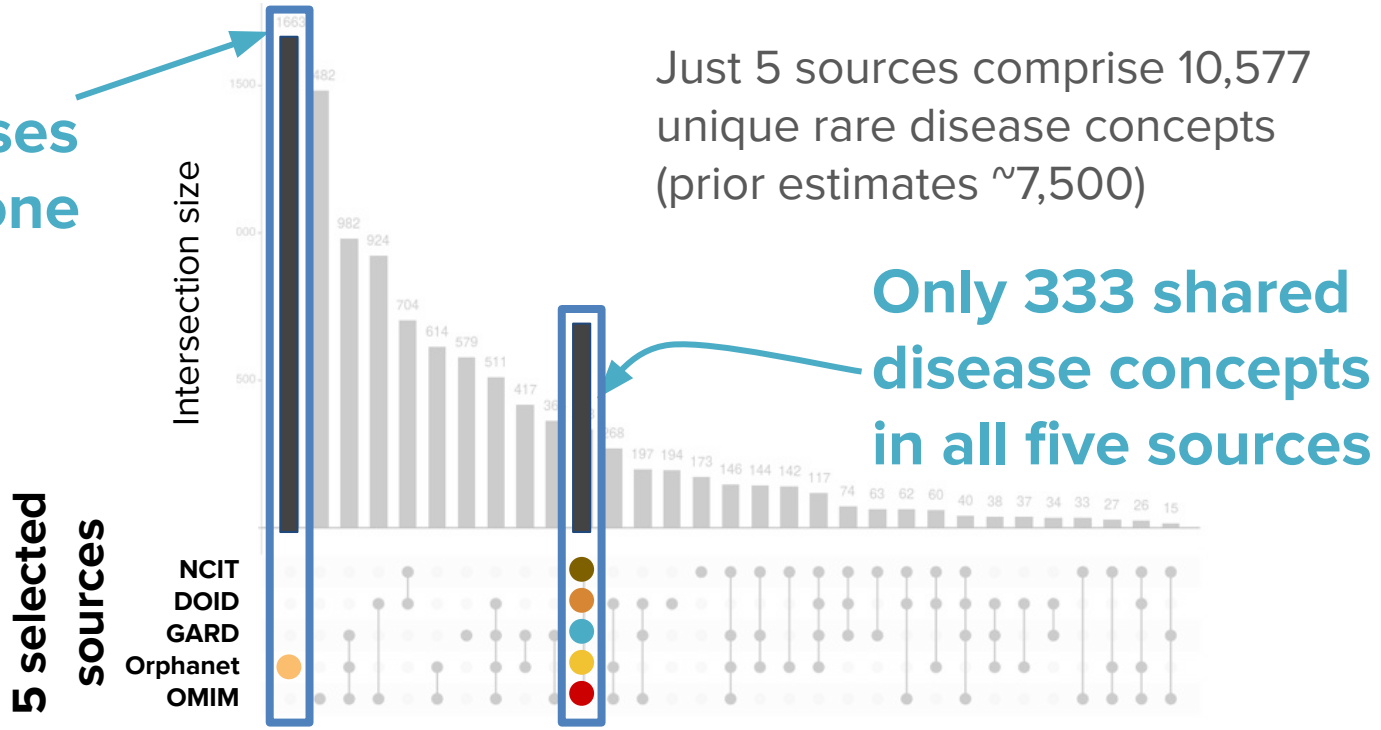
feedback

doi.org/10.1101/048843

kboom

Now we can start to answer the question: How many rare diseases are there?

Many diseases are in only one source



Just 5 sources comprise 10,577 unique rare disease concepts (prior estimates ~7,500)

Only 333 shared disease concepts in all five sources

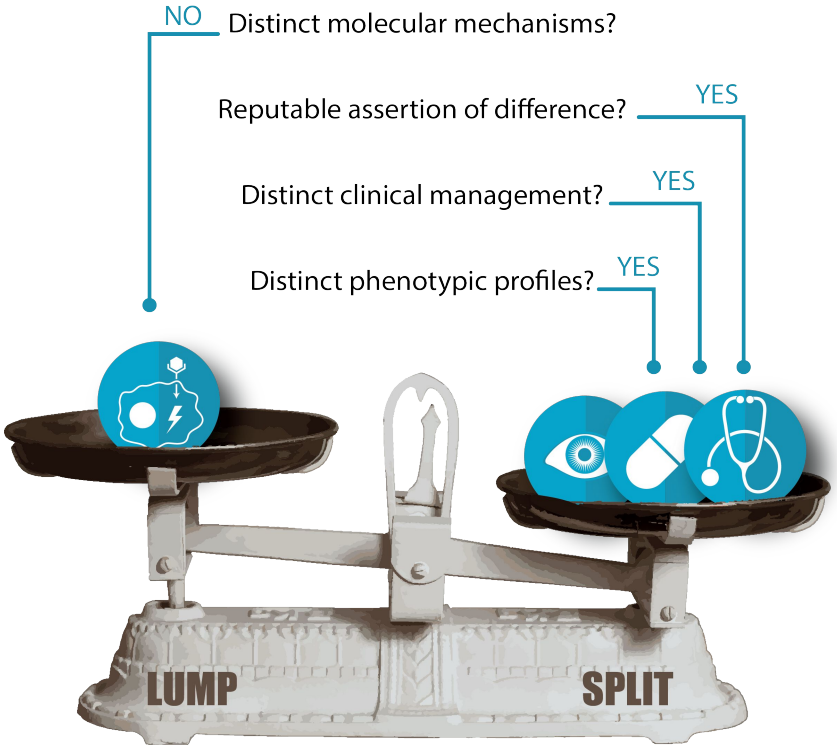
monarch INITIATIVE

Nature Reviews Drug Discovery (bit.ly/nature-rare-diseases)

Mapping itself will not solve the problem. It’s what we attach to those disease classes that matters.

Defining diseases: systematically lump and split to achieve disease harmony

bit.ly/lump-split

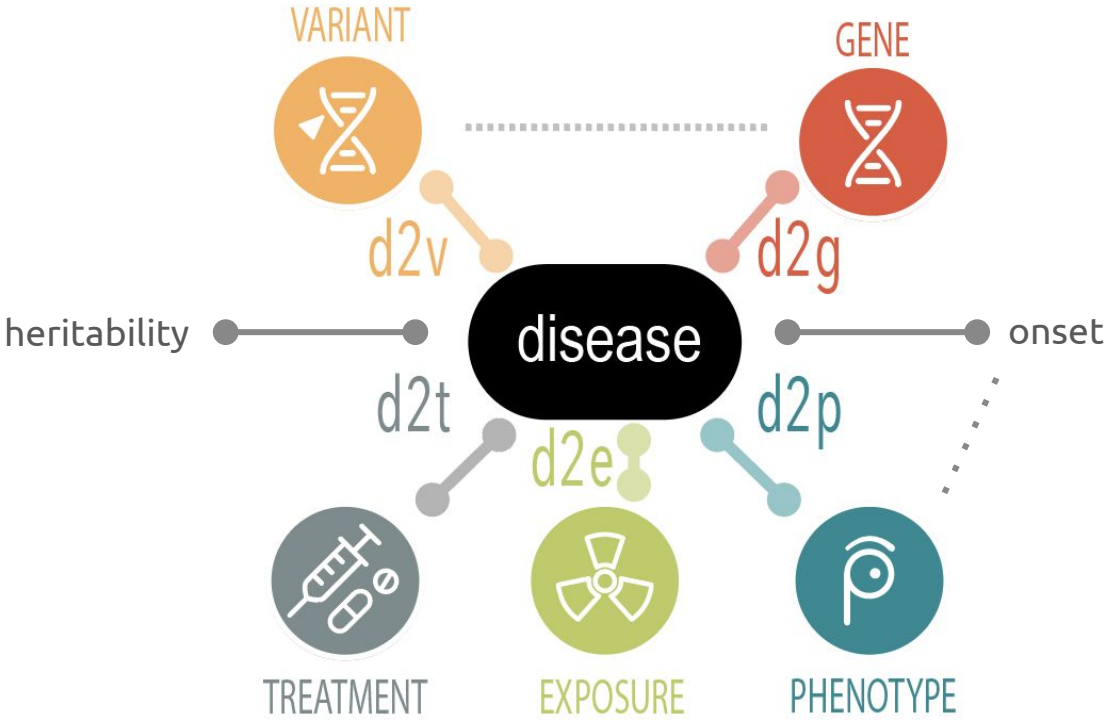


bit.ly/mondo-io

ClinGen
Clinical Genome Resource

ACMG
American College of Medical Genetics and Genomics

Each disease Identifier is associated with a “model” of that disease



Mondo design principles

Dead Simple Ontology Design Patterns (DOSDP)

Pre-made patterns that specify:

- label
- text definition
- synonyms
- *logical definition based on disease attributes*

Design patterns allow for consistency amongst terms and consistent, automated classification of the hierarchy

Disease realized in response to exposure

pattern_name: disease realized in response to environmental exposure

description: This pattern is used for a disease, where the cause of the disease is an exposure to an environmental stimulus (using ECTO exposure terms). Note that this pattern does not include infectious disease or classes that would include an organism, virus or viroid. Rather it includes exposures to chemicals (including drugs), or mixtures.

Examples: chemically-induced disorder MONDO:0029001, alcohol amnestic disorder MONDO:0021702 (26 total)

classes:
disease: MONDO:0000001
exposure event: ExO:0000002

relations:
realized in response to: RO:0009501

vars:
disease: '"disease"'
exposure: '"exposure event"'

name:
text: '%s realized in response to %s'
vars:
- disease
- exposure event

https://github.com/monarch-initiative/mondo/tree/master/src/patterns/dosdp-patterns

Phenopacket schema for individual patients

- Craniosynostosis
- Brachydactyly
- Proptosis
- Broad thumb...

How are these linked to a patient?
To genomic info?
To samples?
To parents and siblings?

Were they NOT observed?

How severe are these?

Are some more severe than others?

When were they first observed?

monarch INITIATIVE

Global Alliance for Genomics & Health

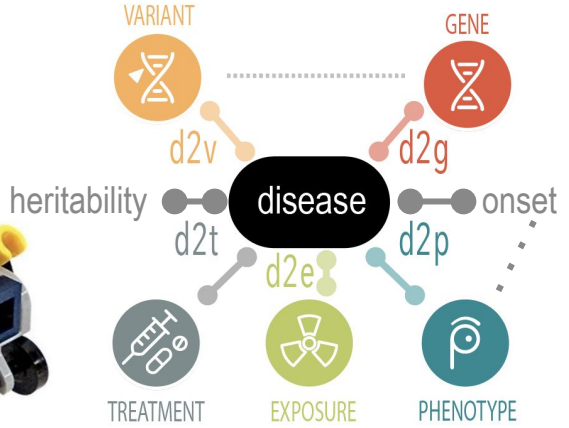
Collaborate. Innovate. Accelerate.

We need to match these!

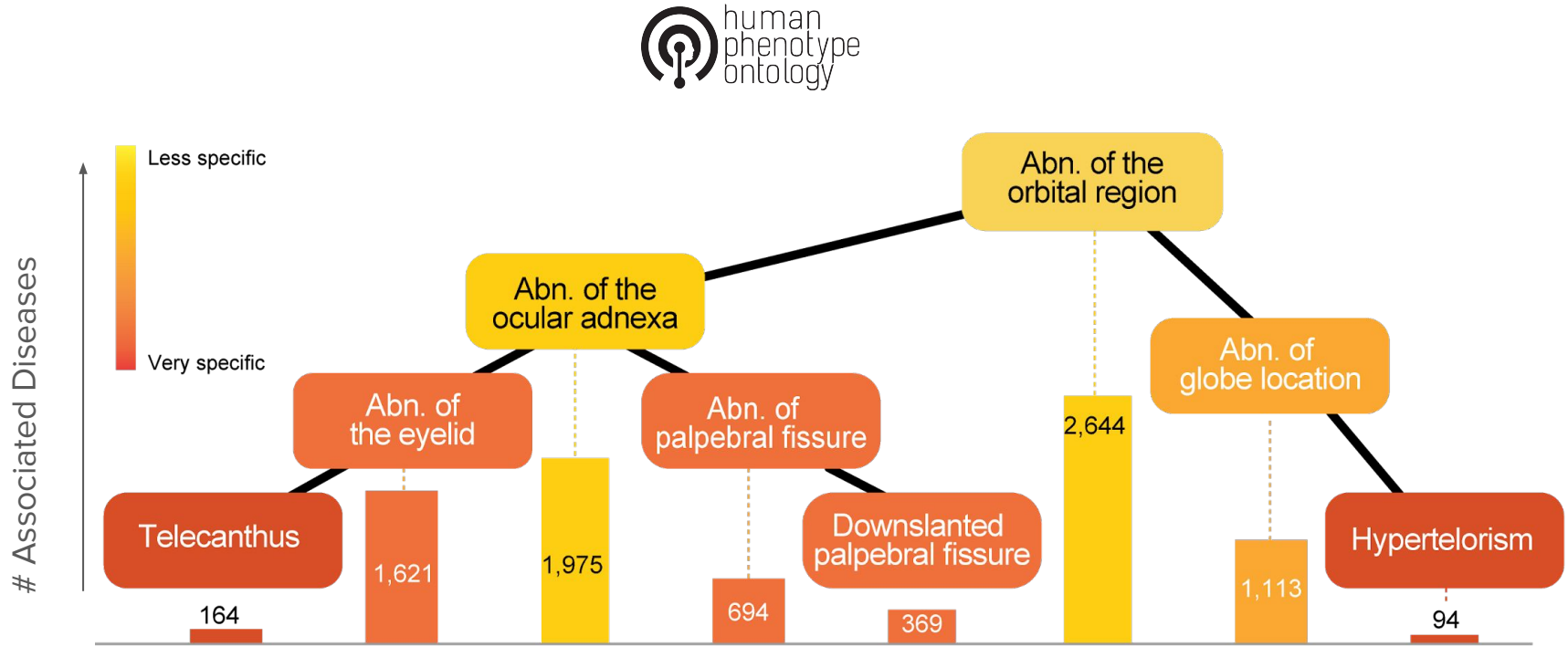
Individual patient


```
subject:
  id: "family"
proband:
  id: "14 year-old boy"
  subject:
    id: "14 year-old boy"
    timeAtLastEncounter:
      age:
        iso8601duration: "P14Y"
    sex: "MALE"
phenotypicFeatures:
- type:
    id: "HP:0009830"
    label: "Peripheral neuropathy"
  excluded: true
  evidence:
  - evidenceCode:
      id: "ECO:0000033"
      label: "author statement supported by traceable reference"
    reference:
      id: "PMID:30808312"
      description: "COL6A1 mutation leading to Bethlem myopathy with recurrent hematuria: a case report."
diseases:
  term:
    id: "MONDO:0008029"
    label: "Bethlem myopathy"
```

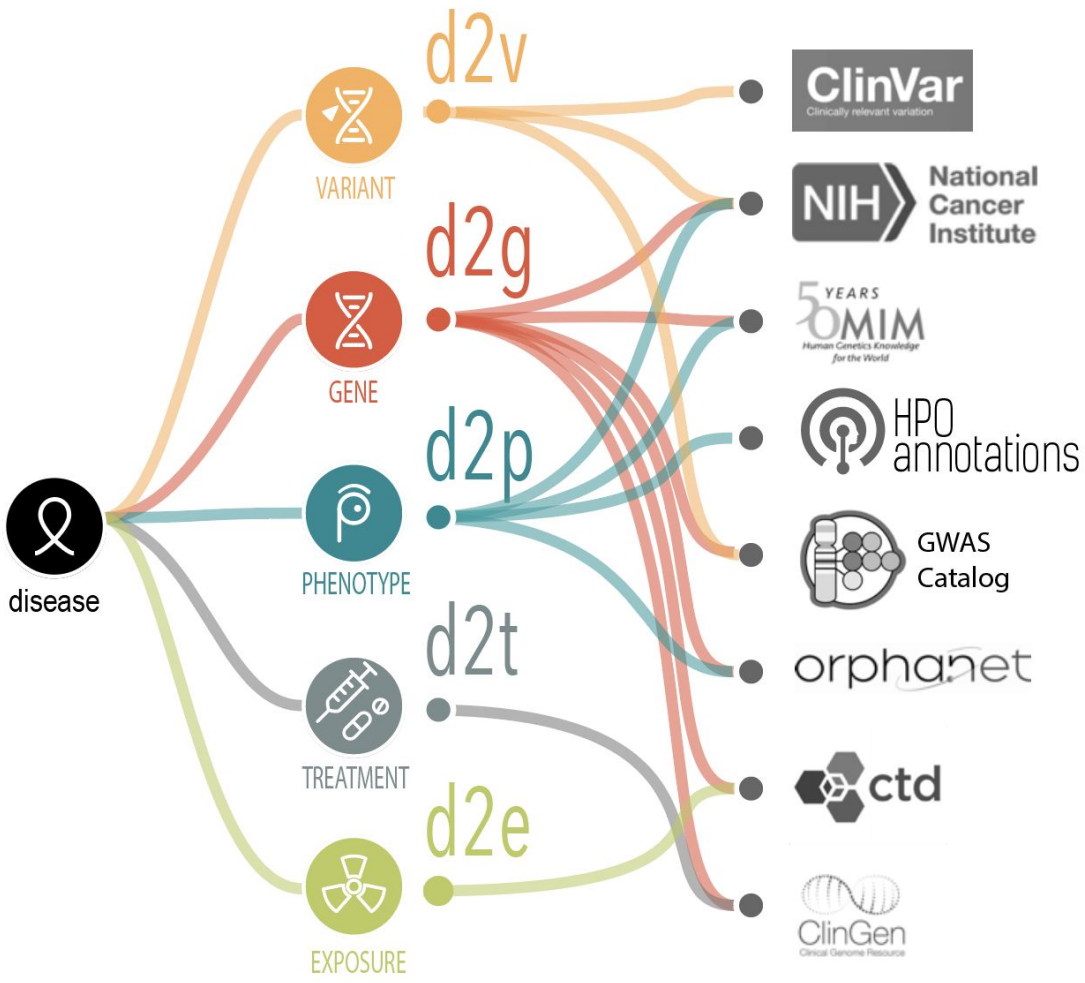

Generalized reference model of this specific disease



Each standardized disease has a phenotype profile



Different communities annotate different relationships, and this makes determining equivalency challenging



Map utility is in the eye of the beholder

Terrible Maps
@TerribleMaps

Duck

2:54 PM · Feb 20, 2020

35.9K 305 Copy link to Tweet

Take homes

- We need axiom and graph alignment to define rare diseases in the same way, around the world
- We need to model disease attributes in the same way, around the world
- We need to model individuals and share this information around the world, to be able to compare them against standard disease definitions
- **Good disease mappings will save lives :-)**

Join Us! Mondo community development

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)

62 releases

Where to view Mondo

OBO Foundry
obofoundry.org/ontology/mondo

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

GitHub
github.com/monarch-initiative/mondo

Weekly Calls

Thursdays, 10am PT/1pm ET
Zoom

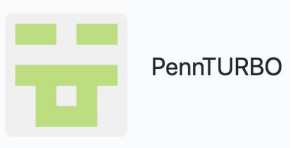
Mailing list:

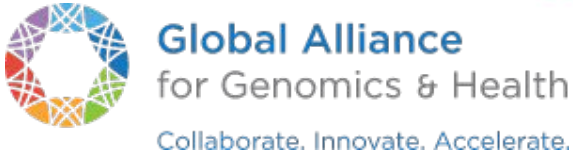
https://groups.google.com/forum/#!forum/mondo-users

Major changes (such as obsoletion candidates or new releases) are shared with the mailing list regularly

Special thanks to NHGRI for funding the Phenomics First Resource

Mondo Users & Contributors





MSEQDR- THE MITOCHONDRIAL DISEASE SEQUENCE DATA RESOURCE CONSORTIUM

MSeqDR – Mitochondrial disease diagnosis with new technology

Big thanks to our curators!

mondo
THE WORLD'S DISEASE CONCEPTS, UNIFIED

Monarch Initiative
Chris Mungall
Nicole Vasilevsky
Sabrina Toro
Leigh Carmody
Lauren Chan
Shahim Essaid
Nomi Harris
Nico Matentzoglu
Julie McMurry
Monica Munoz-Torres
Peter Robinson
Kent Shefchek
Anne Thessen
Deepak Unni

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
Deanne Taylor

CoRDS-Sanford
Austin Letcher

ClinGen
Larry Babb
Taylor Bingaman
Marina DiStefano
Jenny Goldstein
Shruthi Mohan
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
Kallia Panoutsopoulou

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

Johns Hopkins
Christopher Chute
Dazhi Jiao

NCIt
Gilberto Fragoso
Bron Kisler
Sherri De Coronado

NCBI
Donna Maglott
Megan Kane

NIH NCATS
Alice Chen
Eric Sid

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