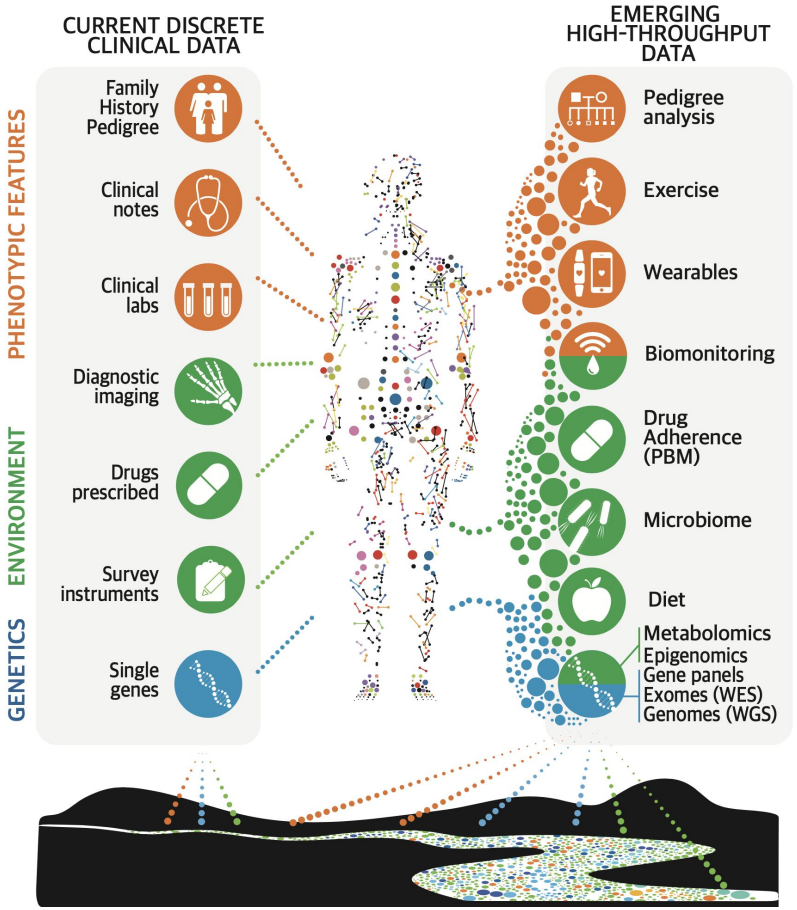
Harmonizing and sharing phenotypes across organisms for diagnostics and mechanism discovery

The Monarch Initiative

Monica Munoz-Torres, PhD., Anne Thessen, PhD., Melissa Haendel, PhD. FMCA

3 August, 2021

These slides bit.ly/mi-risk-s3



Clinical Phenotyping & Diagnostics

Monarch Initiative

monarchinitiative.org

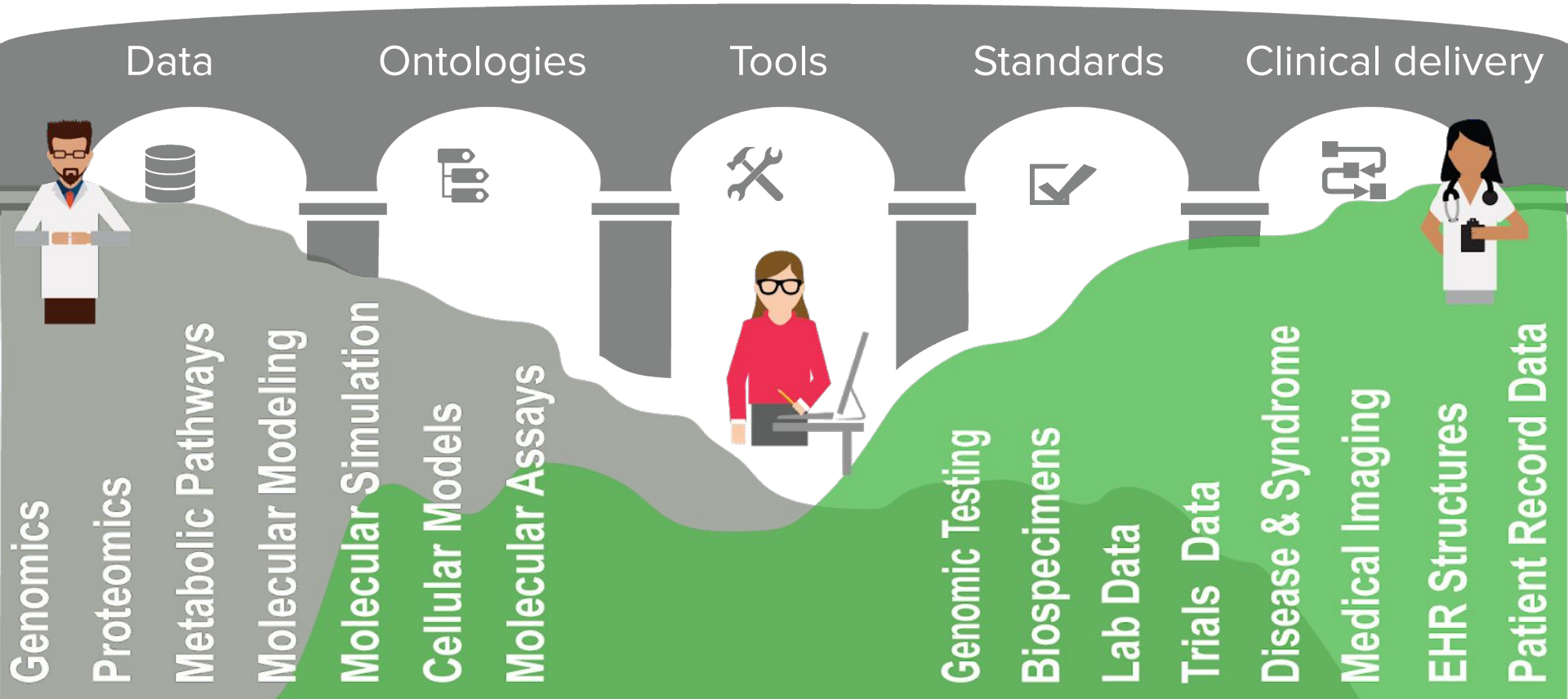
Unified Human Data

Development of ontologies and design patterns

Tools, web services, visual widgets

E.g.,: Phenopackets

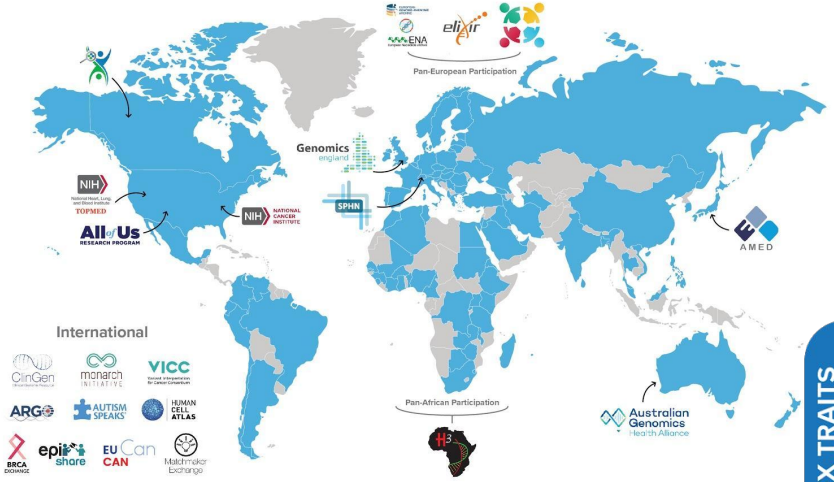
A bridge to the clinic

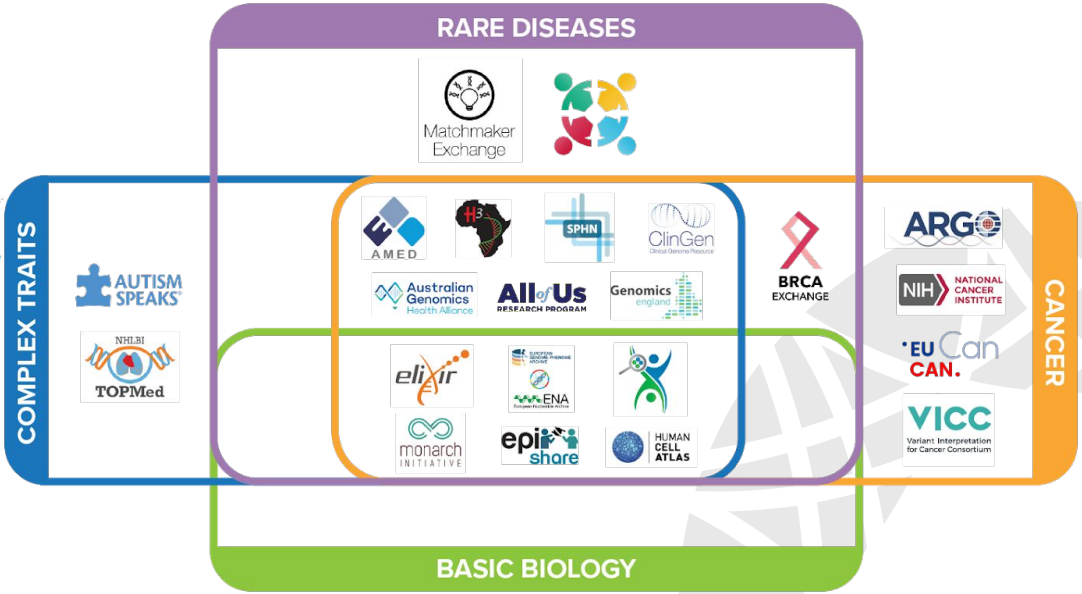


Global Alliance for Genomics & Health (GA4GH)

www.ga4gh.org

GA4GH aims to accelerate progress in genomic science and human health by developing standards and framing policy for responsible genomic and health-related data sharing.





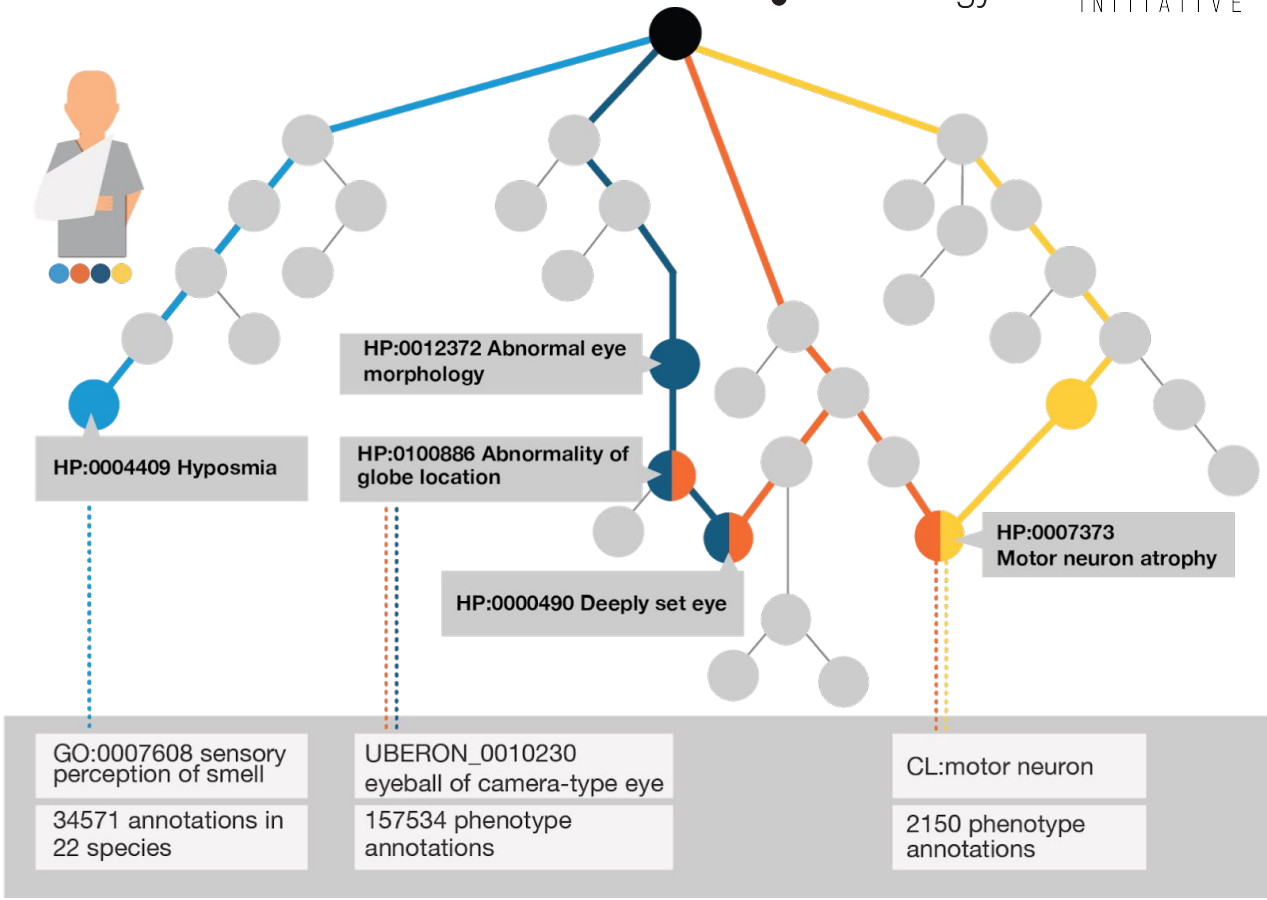
Global Alliance
for Genomics & Health

Collaborate. Innovate. Accelerate.

Human Phenotype Ontology (HPO)

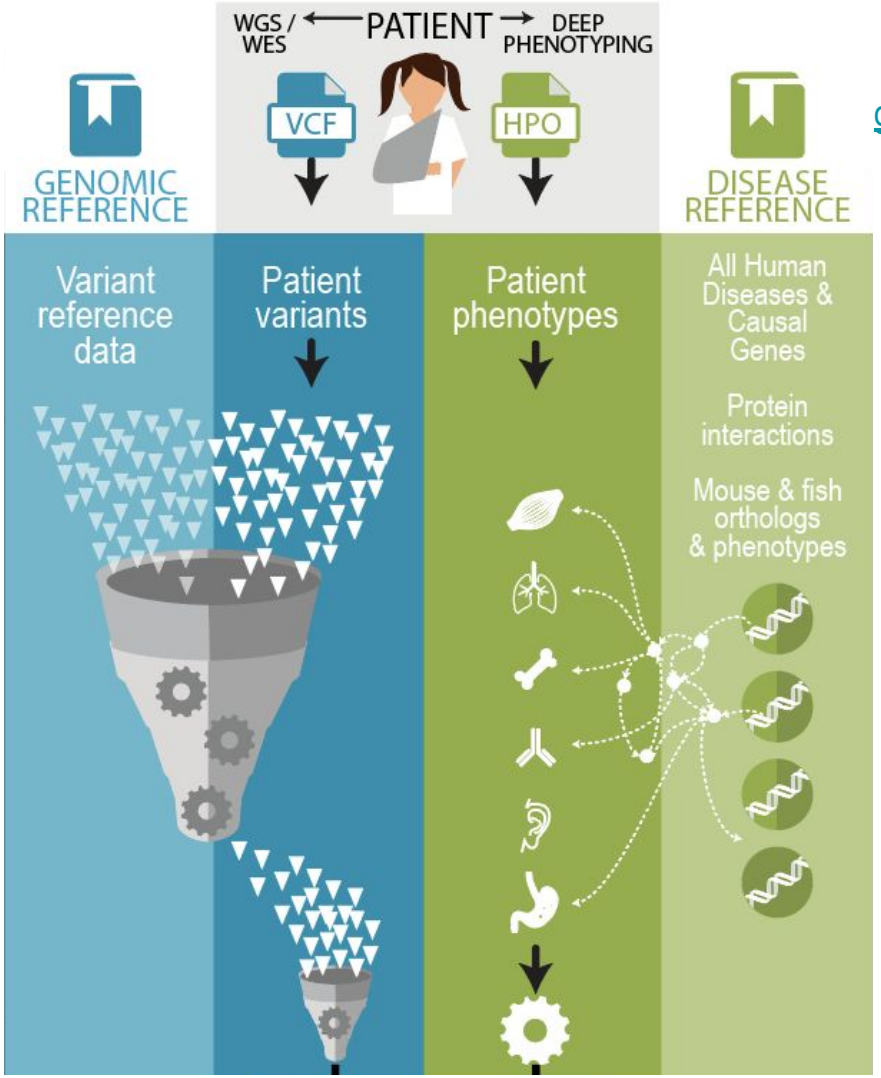
- **Phenotyping terminology**
  >14,500 terms
- **Computational disease models**
  >190,000 disease-phenotype annotations (associations)
- **Widely adopted in rare disease genomic diagnostic tools**
  100,000 Genomes Project, SOLVE-RD, NIH-UDP, etc.

hpo.jax.org



Combining genomic and phenomic data improves variant prioritization for diagnosis

doi: 10.1038/gim.2015.137

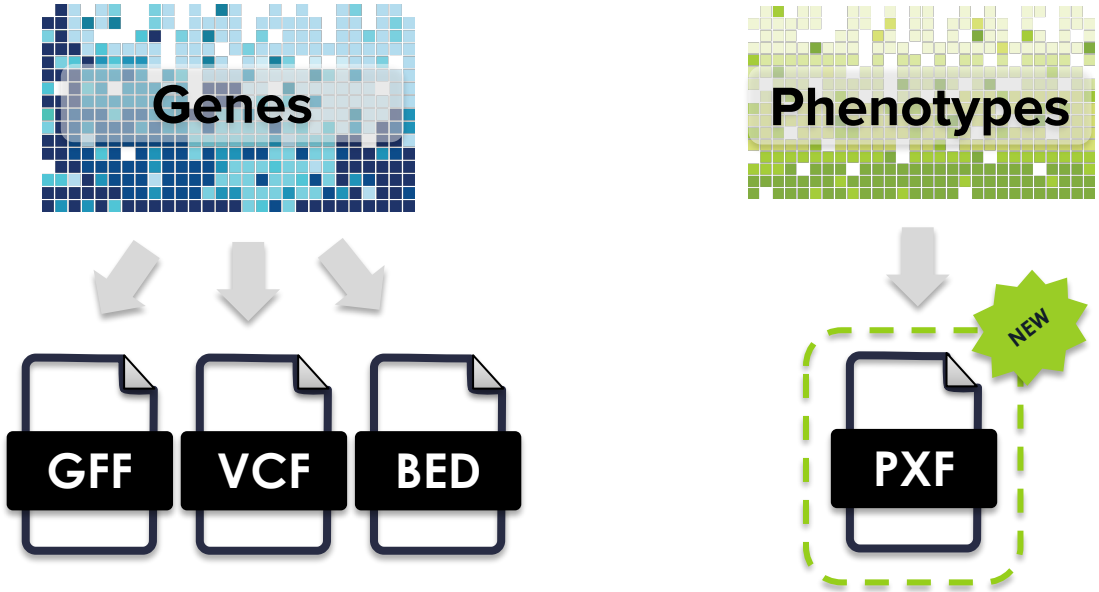


EXOMISER

github.com/exomiser/Exomiser

monarch INITIATIVE

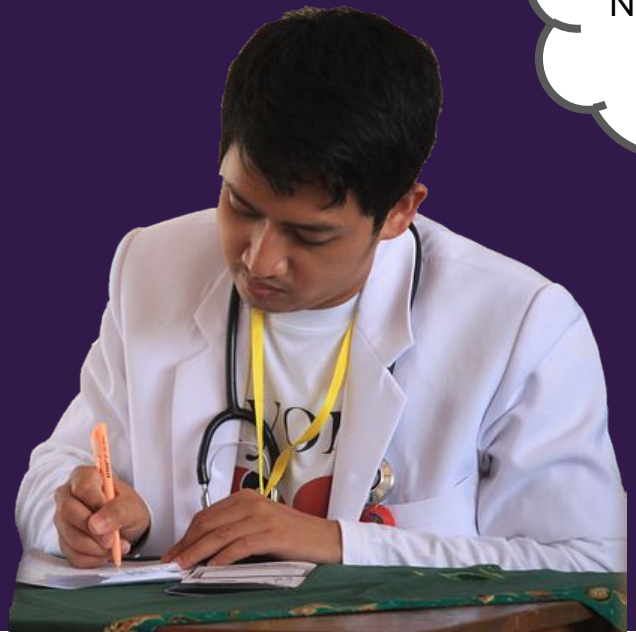
Standard exchange formats exist for sequence/genomes but not for phenotypes



We need a standard way to share case-level phenotypic information that is not free text, a candidate diagnosis proxy, nor full EHR data exported via PDF

Phenopackets can help us do better

Craniosynostosis
Brachydactyly
Proptosis
Broad thumb...



Were they NOT observed?

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

How severe are these?

Are some more severe than others?

When were they first observed?

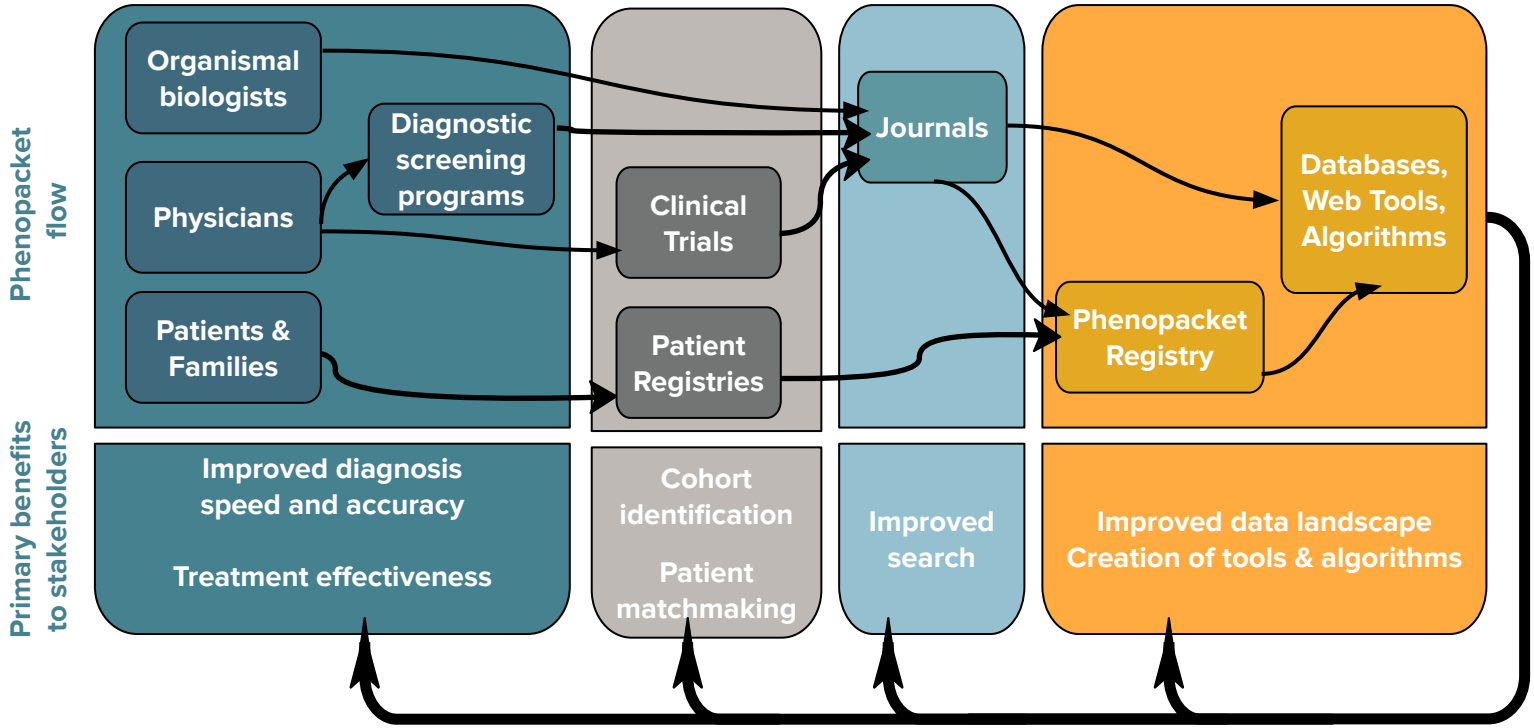
monarch INITIATIVE

Global Alliance for Genomics & Health

Collaborate. Innovate. Accelerate.

The Phenopacket Ecosystem

users and use cases

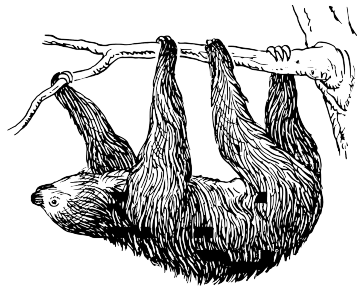
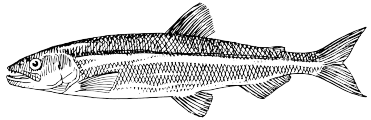


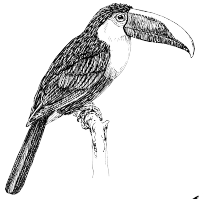
phenopackets.org

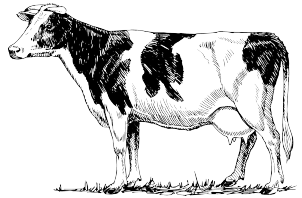
https://github.com/phenopackets

Global Alliance for Genomics & Health

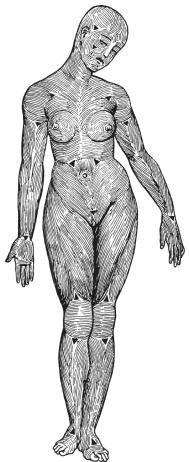
monarch INITIATIVE

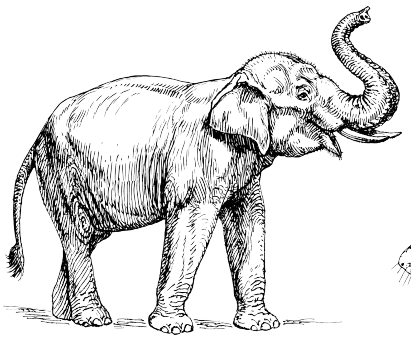



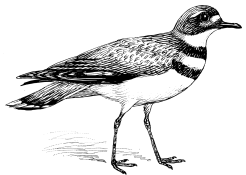
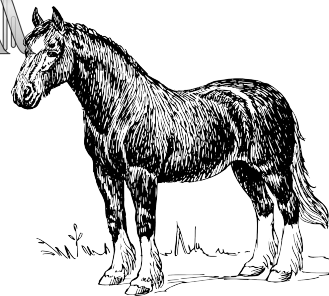



The Value of Cross-Species Research





Other species aren't just relevant; each has unique phenotypes

- The dog's retina has area centralis (analogous to the human macula) & fovea-like region, similar to humans; useful to study naturally occurring cone diseases

- Aged cats are natural models of Alzheimer's Disease: they form Abeta oligomers, neurofibrillary tangles, and have neuronal loss

- Naked Mole Rats don't get cancer

- Armadillos are a natural host of *M. leprae*, the mycobacterium that causes leprosy (only one besides humans)

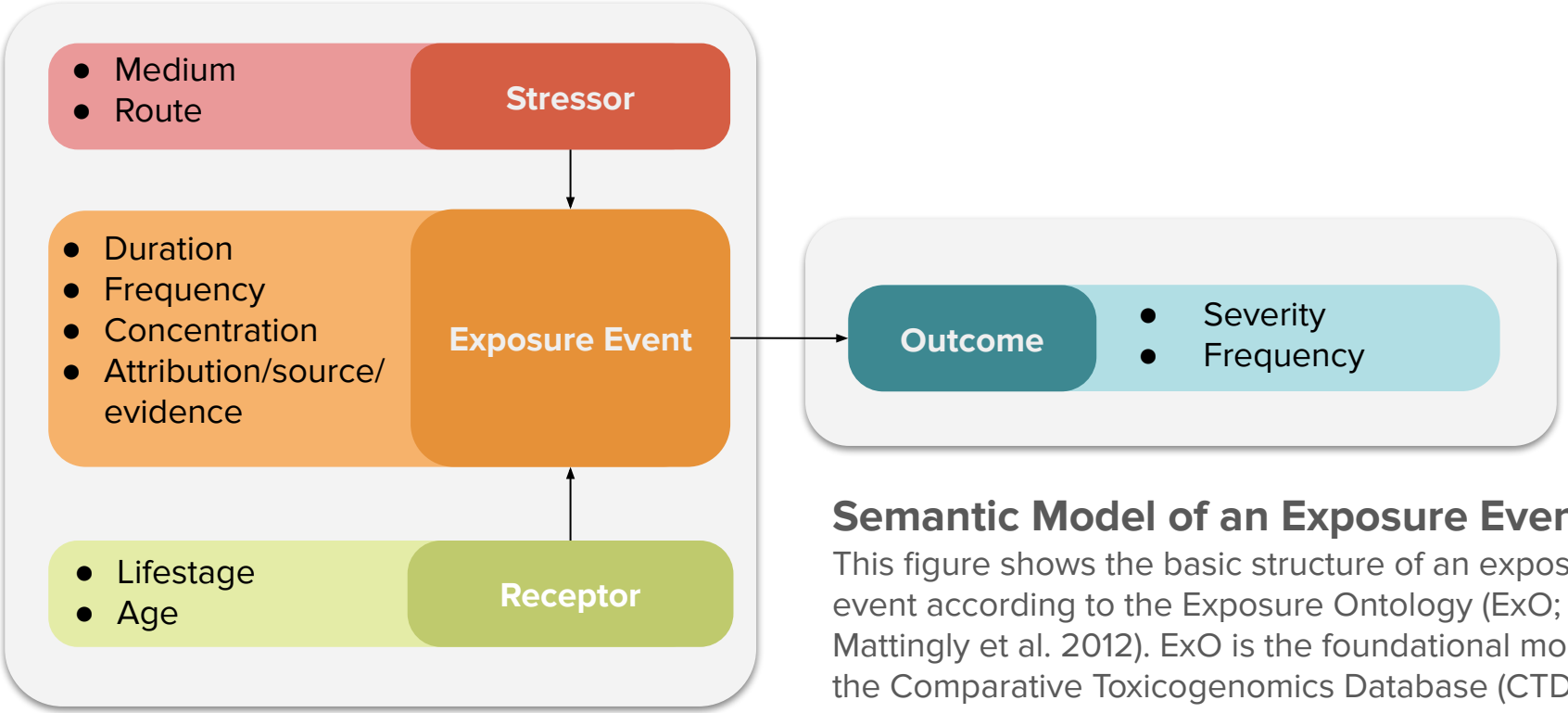
- Tree shrews' glioblastomas are morphologically & genetically similar to humans (& more similar than mouse models)

- Great pond snails are models of inflammation-mediated memory dysfunction, and show evidence of spontaneous neural tissue regeneration after injury

- Silkworms are a model for uric acid metabolism. Decreases in plasma uric acid are correlated with clinical progression of Parkinson's Disease

monarch INITIATIVE

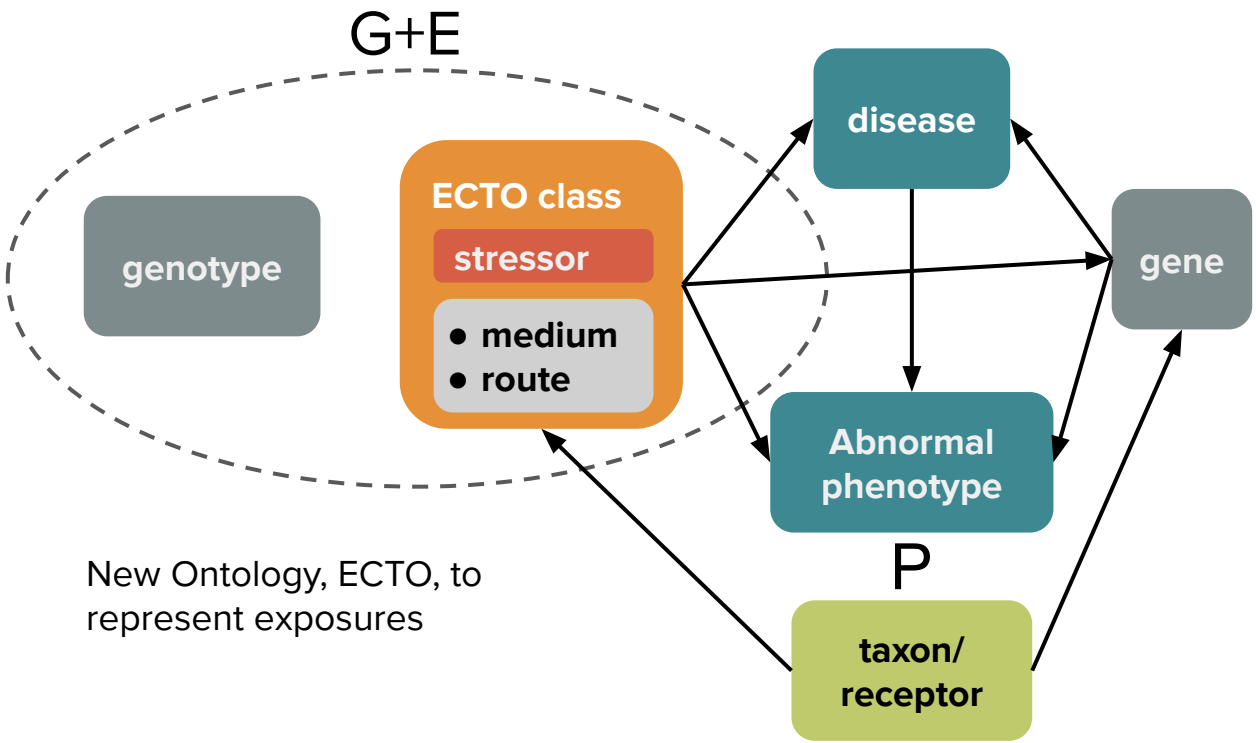
Environmental Exposures Modeling with ExO



Semantic Model of an Exposure Event.
This figure shows the basic structure of an exposure event according to the Exposure Ontology (ExO; Mattingly et al. 2012). ExO is the foundational model for the Comparative Toxicogenomics Database (CTD; Mattingly et al. 2006).

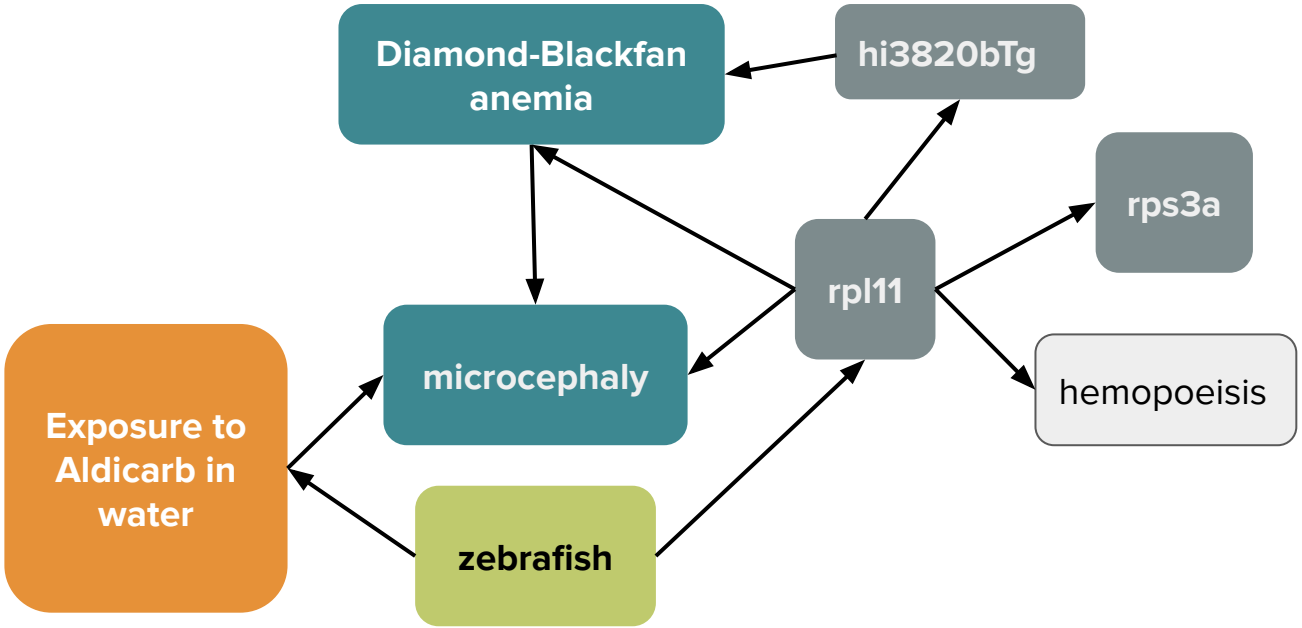
monarch INITIATIVE

Integration with the Monarch Knowledge Graph



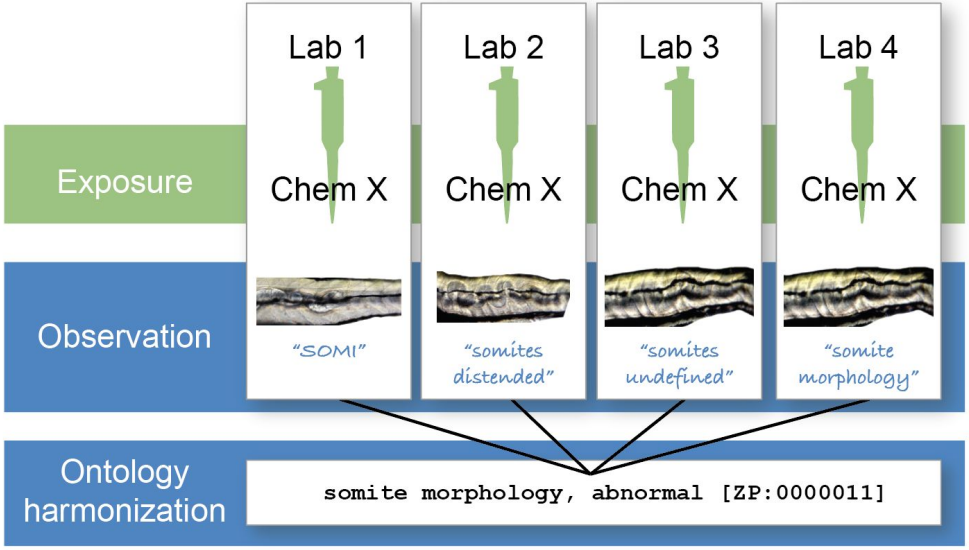
New Ontology, ECTO, to represent exposures

Integration with the Monarch Knowledge Graph



High Throughput Zebrafish Studies

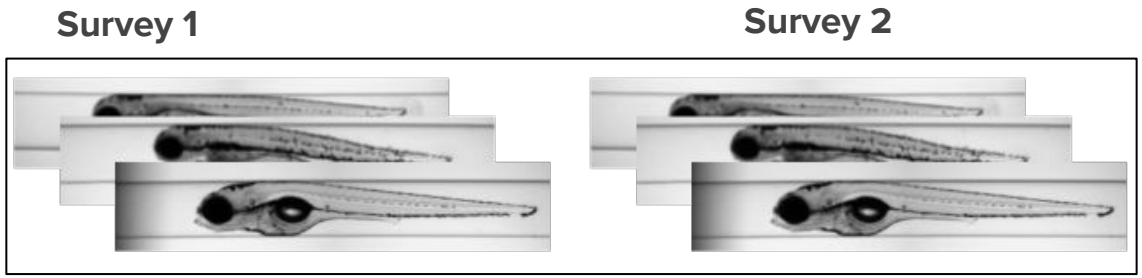
- High-throughput zebrafish studies
- Quickly finding endpoints
- Labs use their own vocabulary
- Hinders data integration
- Ontologies are a potential solution
- Can we standardize endpoint reporting in a useful way?



monarch INITIATIVE

High Throughput Zebrafish Studies

- Annotators used 1,748 unique terms to describe 48 ZPO phenotypes
- Abnormalities of the gut, pectoral fin, and otic vesicles were most heterogeneously described
- Abnormalities of the heart and yolk were most consistently identified



RESULTS:

- Does everyone mean the same thing when they use the same term? **NO**
  For example, the term 'heart edema' often lumped many heart deformities.

- Is everyone scoring the same images with the same endpoints? **NO**
  In survey 1, annotators did not even agree on which embryos were normal

- Does using the ontology improve consistency across labs? **YES**
  Agreement between annotators was better when ontology terms were provided (Fleiss' Kappa for Survey 1 = 0.07 and Survey 2 = 0.11)

monarch INITIATIVE

Thank you

monarchinitiative.org

Melissa Haendel (PI)
Peter Robinson (PI)
Chris Mungall (PI)
Damian Smedley (Site PI)
David Osumi-Sutherland (Site PI)
Julie McMurry
Jules Jacobsen
Kent Shefchek
Monica Munoz-Torres
Nomi Harris
Sebastian Koeller
Deepak Unni
Harshad Hedge
Justin Reese
Kevin Schaper
Lauren Chan
Matt Brush
Nicole Vasilevsky
Sabrina Toro
Seth Carbon
Shawn O'Neil
Sierra Moxon
Tudor Groza
Tudor Oprea

The Monarch Initiative is funded by NIH OD R24 OD011883

Additional support to TISLab: NHGRI RM1 HG010860, NCATS U24 TR002306