The Monarch Initiative

Translating Human to Models and Back Again: Phenotype Ontologies for Data Integration and Discovery

Melissa Haendel, PhD

April 20, 2016

NHGRI Genomic Medicine IX

@monarchinit @ontowonka

haendel@ohsu.edu

Global Alliance for Genomics & Health

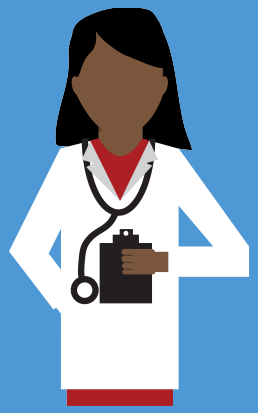
Undiagnosed Diseases Network

IRDiRC International Rare Diseases Research Consortium

Oregon Health & Science University

The genome is sequenced, but…

?



OMIM

3,435

Mendelian Diseases with no known genetic basis

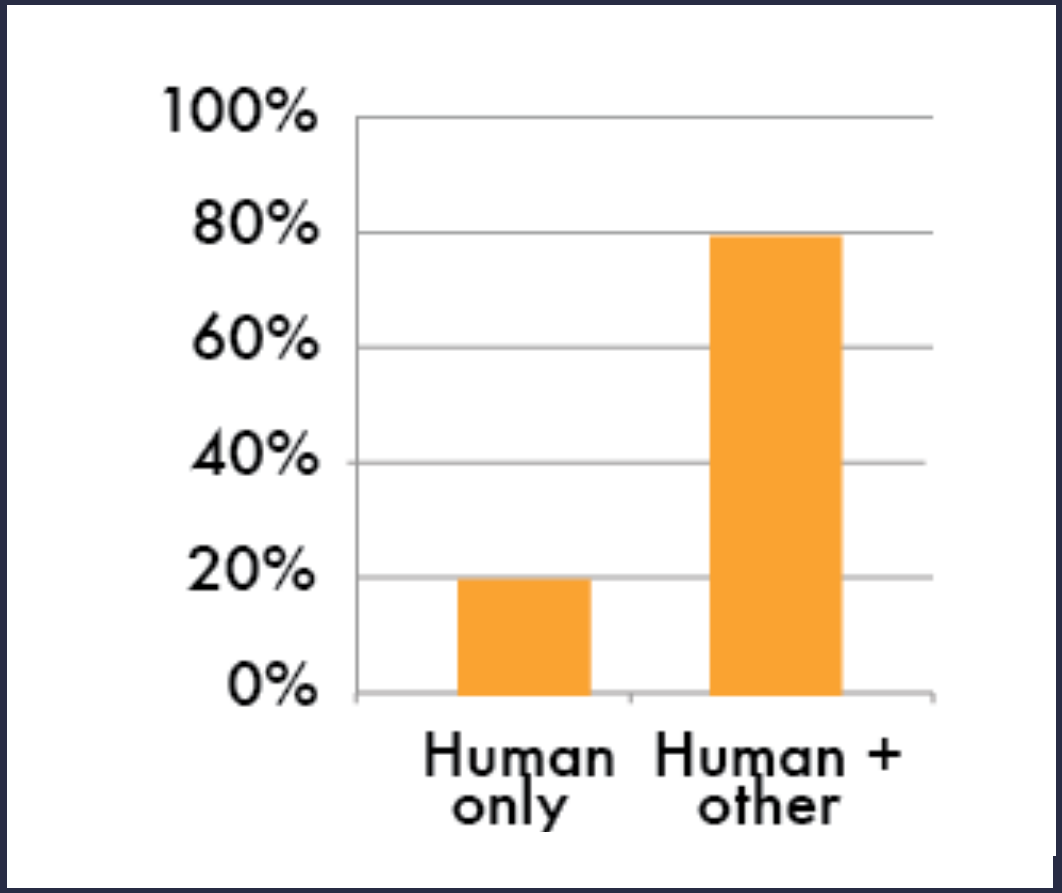
ClinVar

66,396

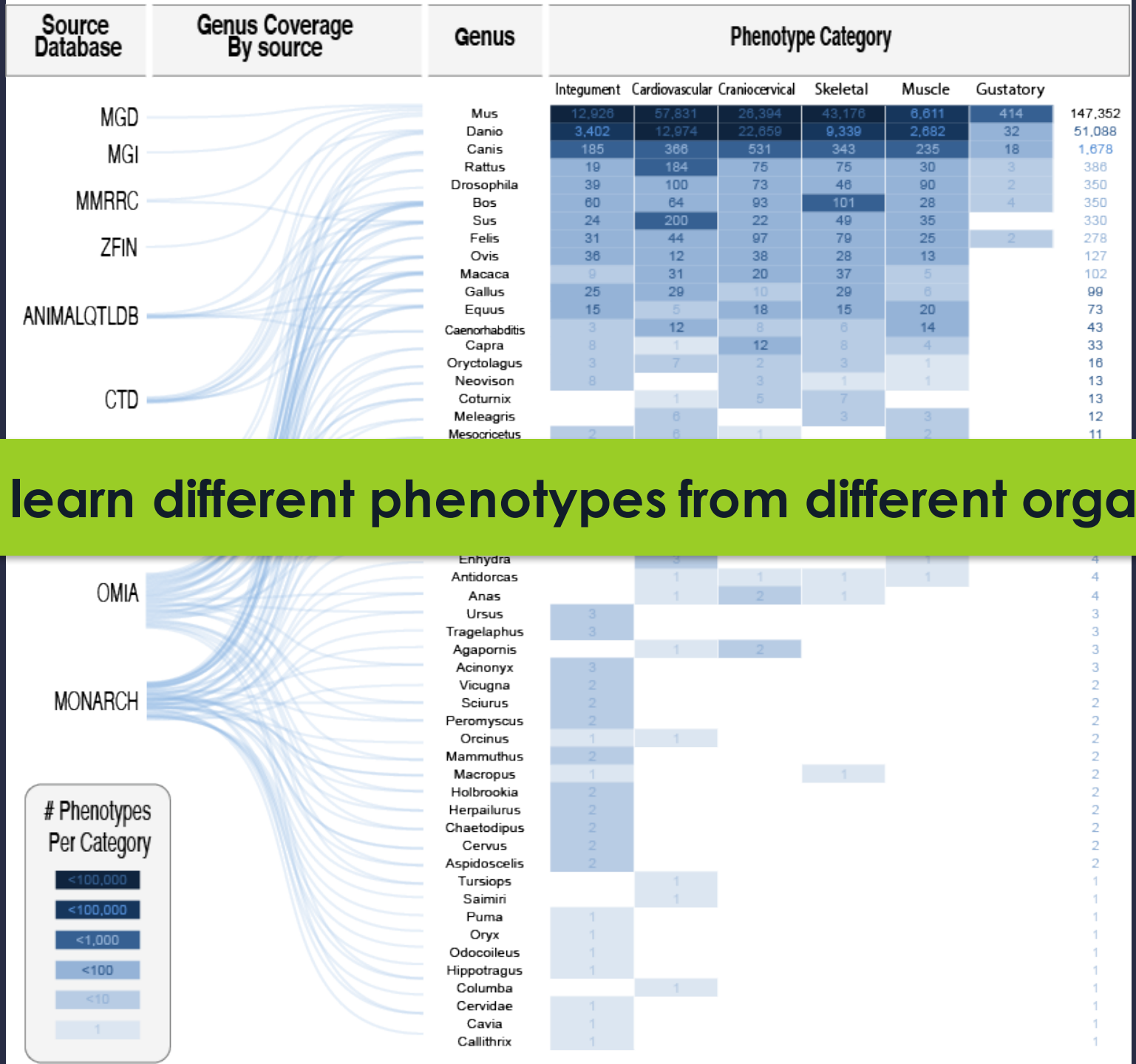
Variants with no known pathogenicity

…we still don't know very much about what it does

Filling the G2P knowledge gap from other organisms

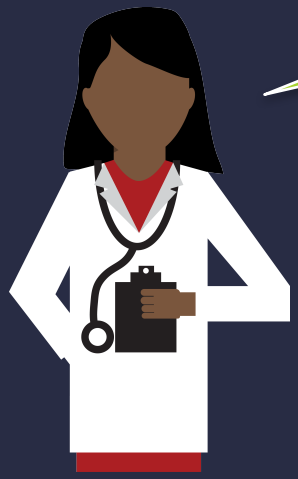


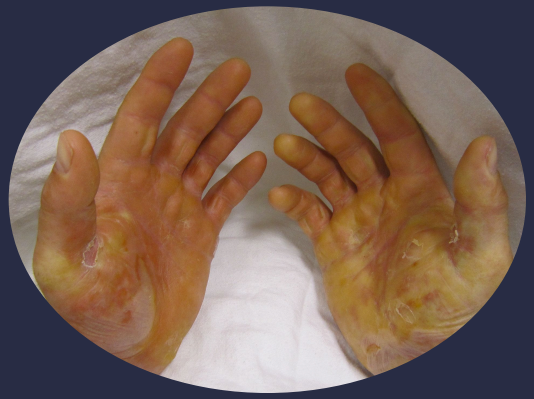
Other= rat, fly, worm, mouse, zebrafish



We learn different phenotypes from different organisms

Palmoplantar hyperkeratosis





Ulcerated paws

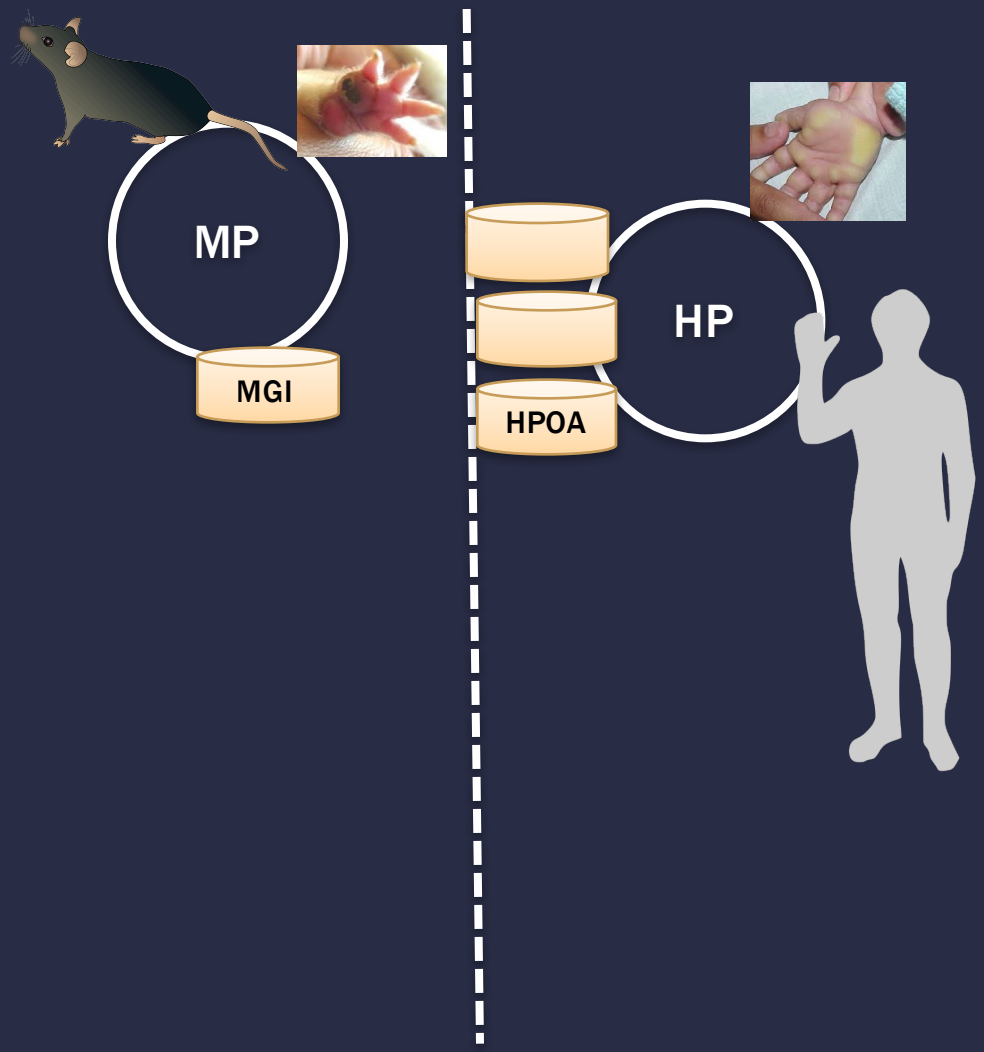
Thick hand skin

Image credits:

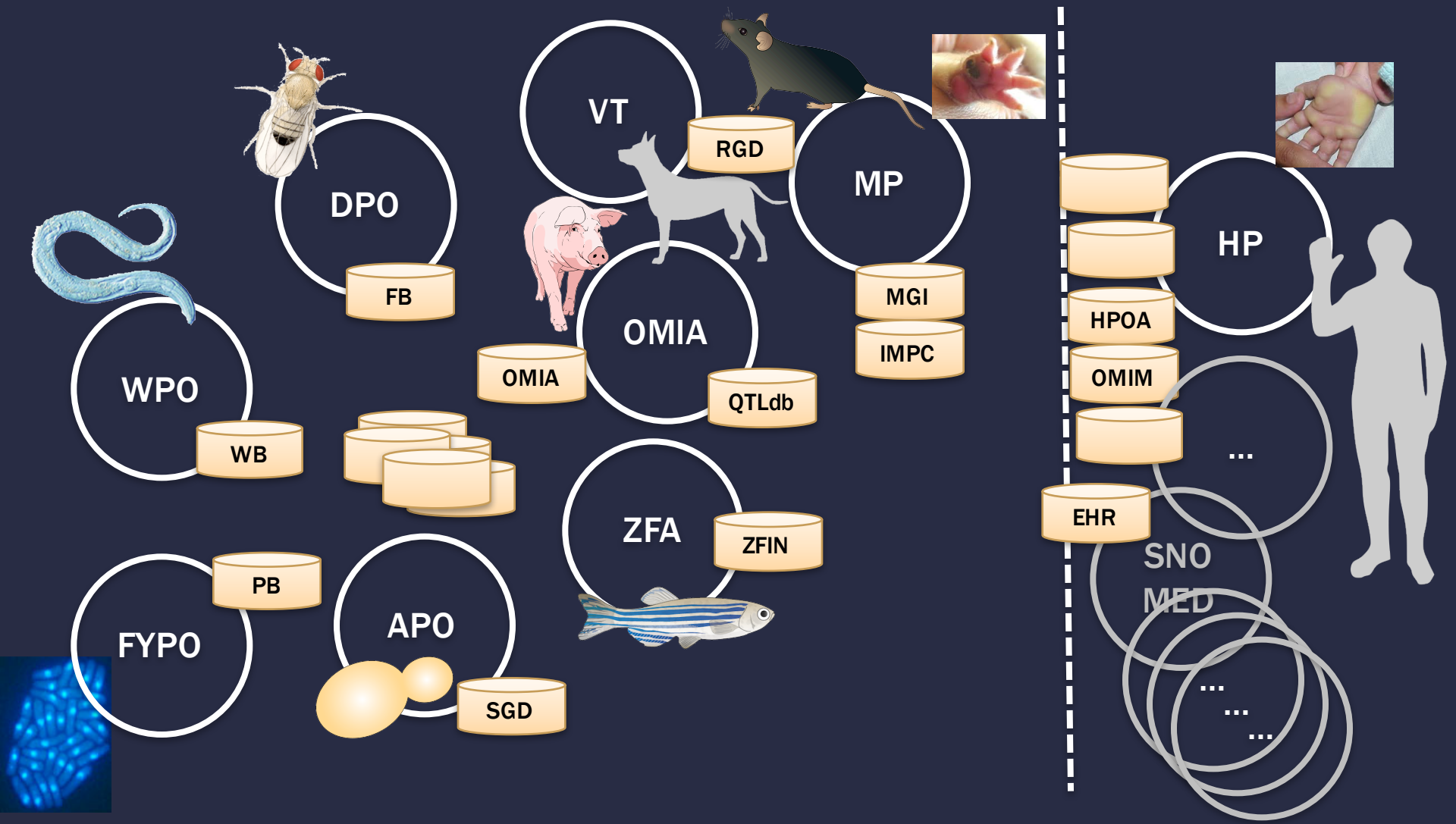
"HandsEBS" by James Heilman, MD - Own work. Licensed under CC BY-SA 3.0 via Commons – https://commons.wikimedia.org/wiki/File:HandsEBS.JPG#/media/File:HandsEBS.JPG

http://www.guinealynx.info/pododermatitis.html

Challenge: Each data source uses their own vocabulary/ontology



Challenge: Each data source uses their own phenotype vocabulary/ontology



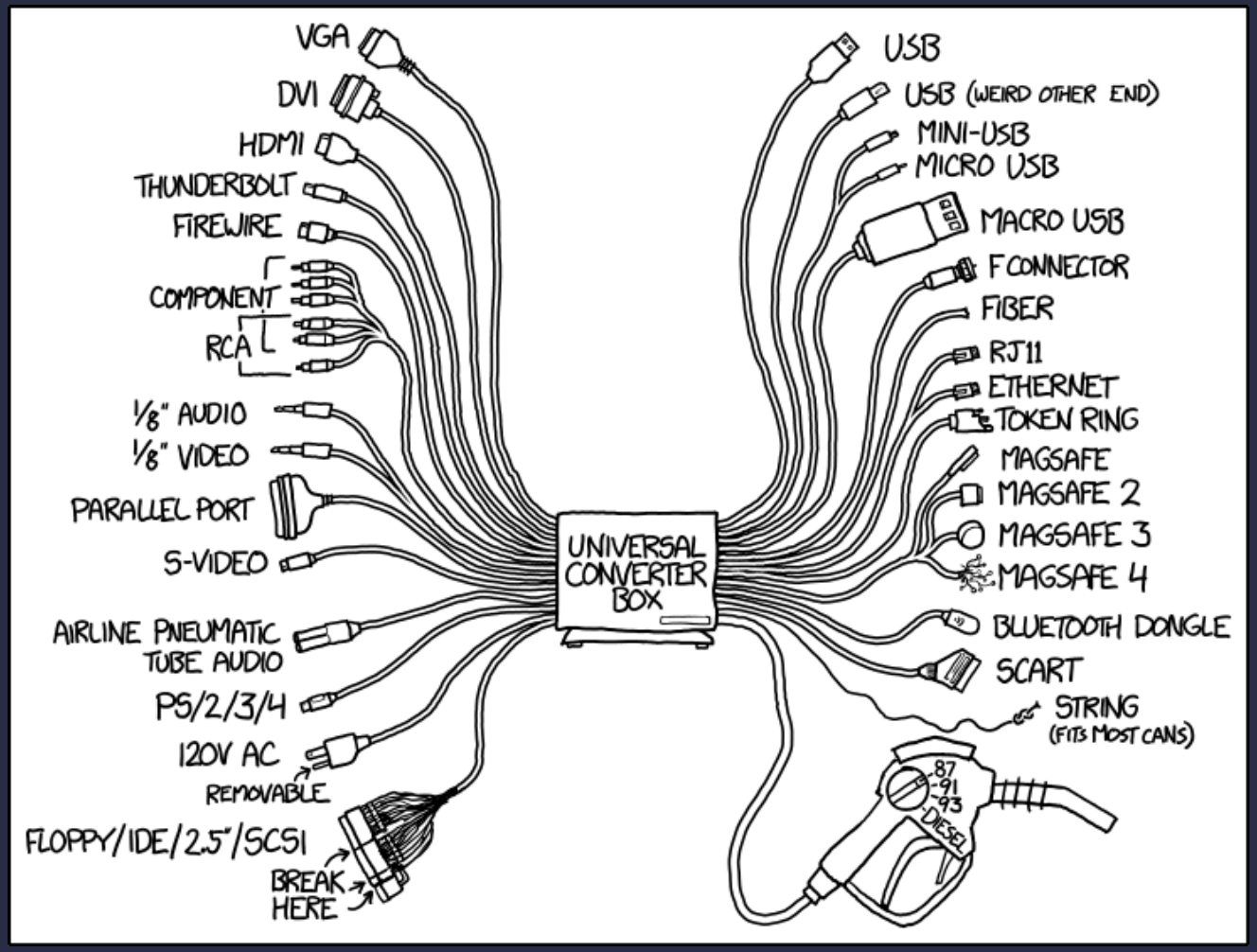
Can we help machines understand phenotype terms?

Human phenotype

"Palmoplantar hyperkeratosis"

I have absolutely no idea what that means

Ontologies serve as a bridge



http://xkcd.com/1406/

Decomposition of complex concepts allows interoperability

Human phenotype

"Palmoplantar hyperkeratosis"

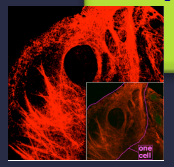
=

PATO

increased

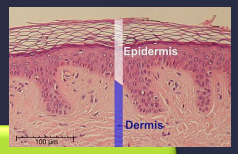
GO

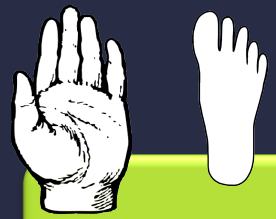
keratinization



Uberon

Stratum corneum layer of skin



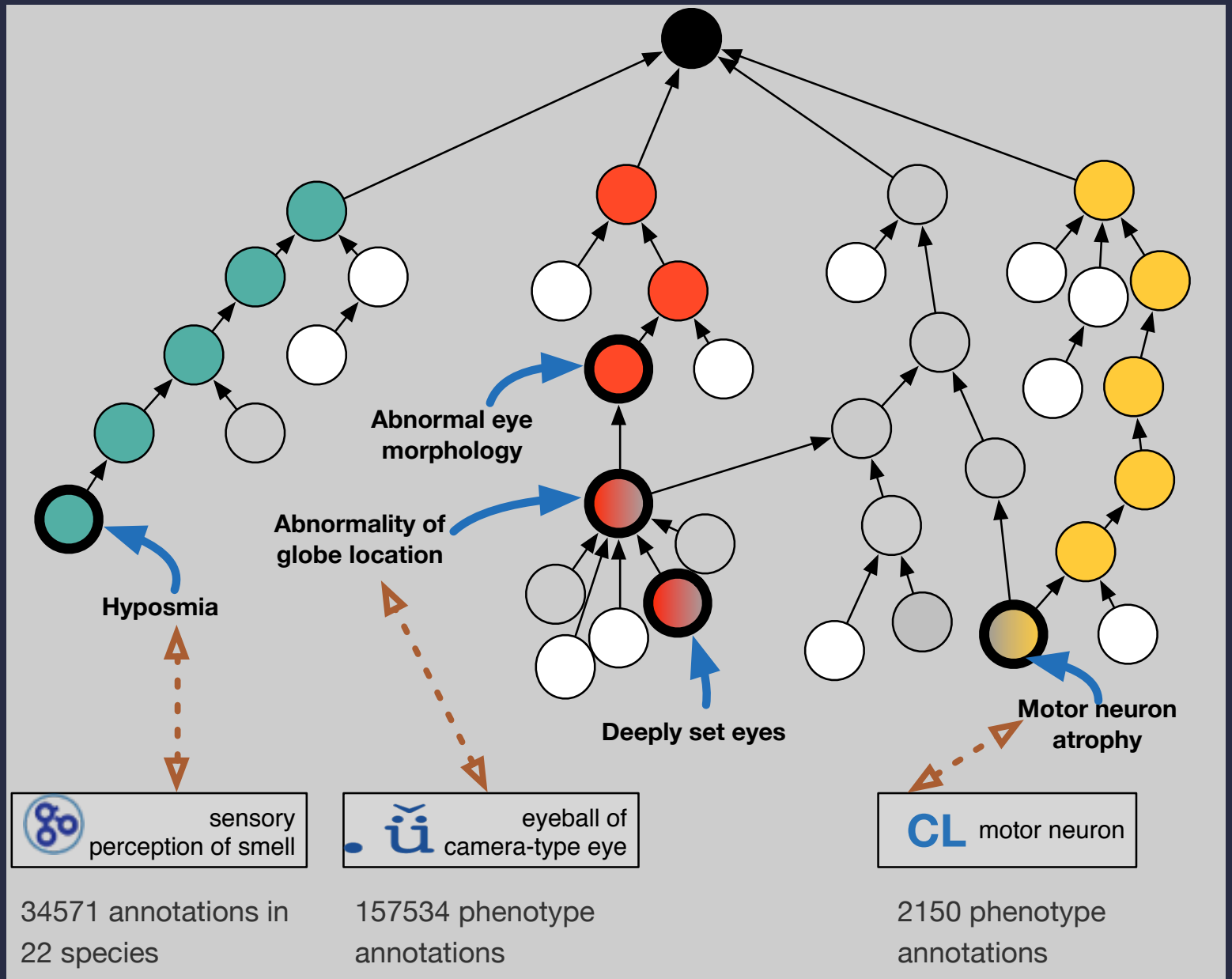


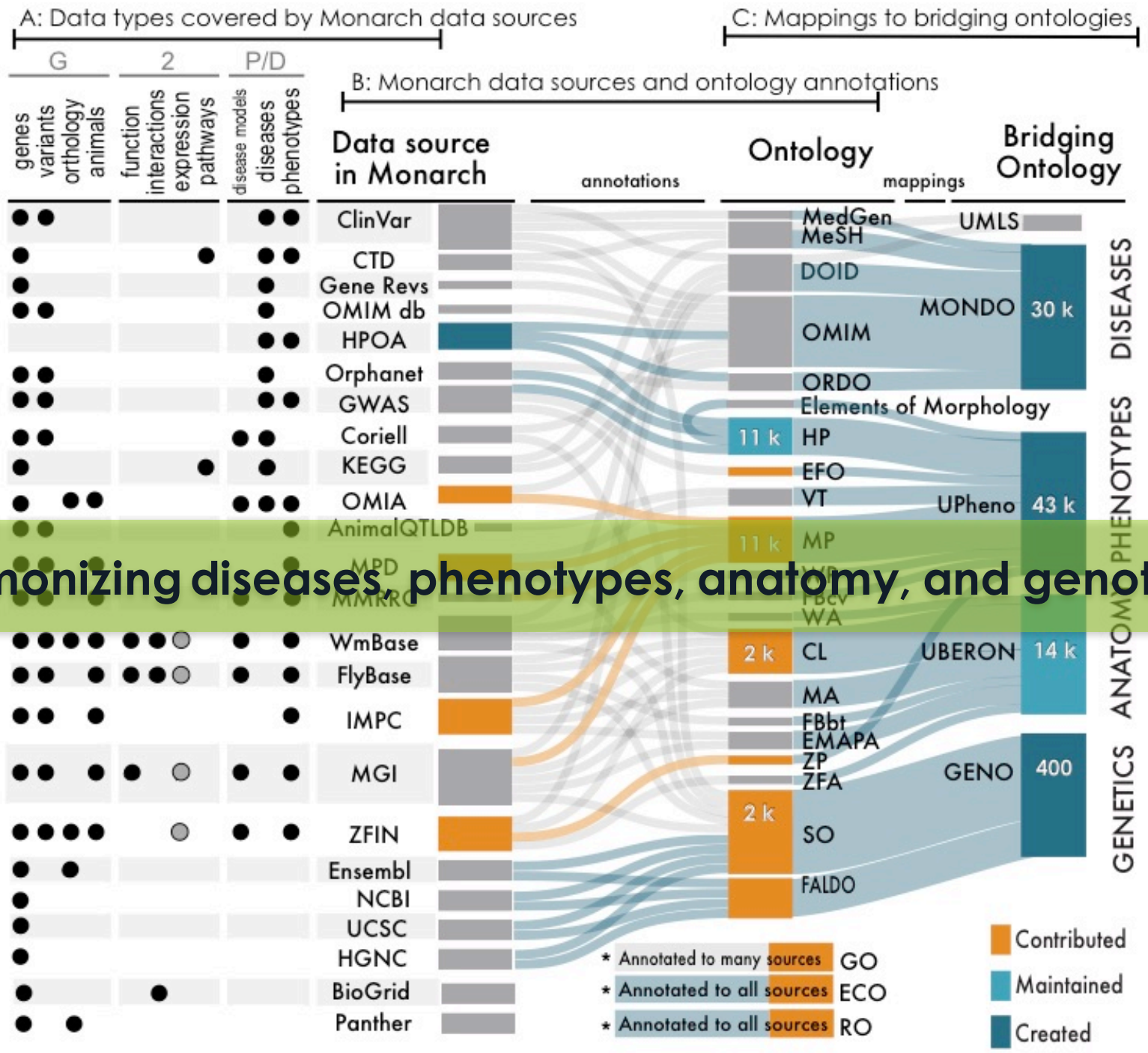
Autopod

Species neutral ontologies, homologous concepts

Mungall, C. J., Gkoutos, G., Smith, C., Haendel, M., Lewis, S., & Ashburner, M. (2010). Integrating phenotype ontologies across multiple species. *Genome Biology*, **11(1), R2. doi:10.1186/gb-2010-11-1-r2**

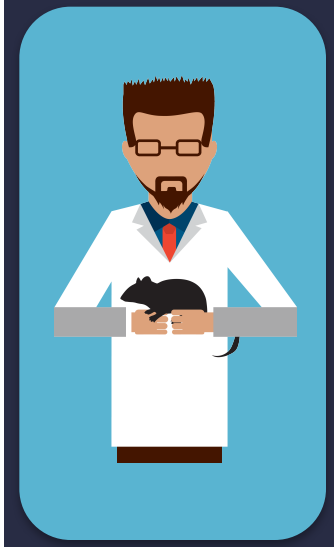
The Human Phenotype Ontology





Harmonizing diseases, phenotypes, anatomy, and genotypes

Phenotypic matchmaking across species



Gene Profile

Gene D
One or more mutations known to cause "Disease D"

homology

Patient A
gene panel/ exome/ whole genome

orthology

Gene M or Genotype M

Disease D

Patient A Phenotype Profile

Mouse M

Phenotype Profile

Disease D		Patient A Phenotype Profile		Mouse M
Hypoplasia of the frontal lobes	Aplasia/ Hypoplasia of the cerebrum	Micro-cephaly		
Distal lower limb amyotrophy	Muscular atrophy	Generalized amyotrophy	Abnormality of muscle morphology	dystrophic muscle
Optic nerve hypoplasia	Abnormal optic nerve morphology	Optic atrophy	Abnormal retina morphology	optic nerve swelling
Lyssen-cephaly	Abnormal cortical gyration	Poly-microgyria	Abnormal cerebral cortex morphology	thin cerebral cortex
Contractures of the large joints	Flexion contracture	Contractures of the joints of the lower limbs	Abnormal joint mobility	campto-dactyly
Focal seizures	Seizures	Seizures	Seizures	abnormal spike wave discharge
Decreased body weight	Decreased body weight	Decreased body weight	Decreased body weight	Decreased body weight
		Hypo-thermia	Hyper-thermia	decreased body temp.

Closest term in common

HP Terms | Bridging ontology term | MP Terms

Putting deep phenotyping to use

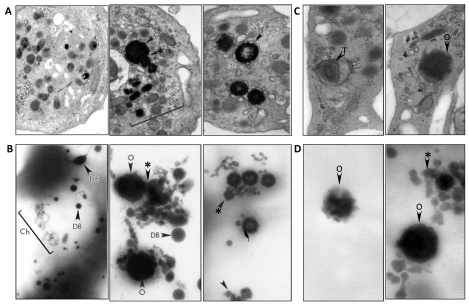
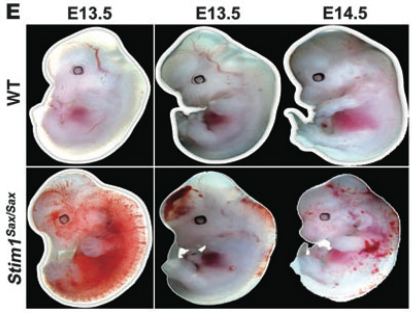
http://bit.ly/stim1paper

EXOMISER

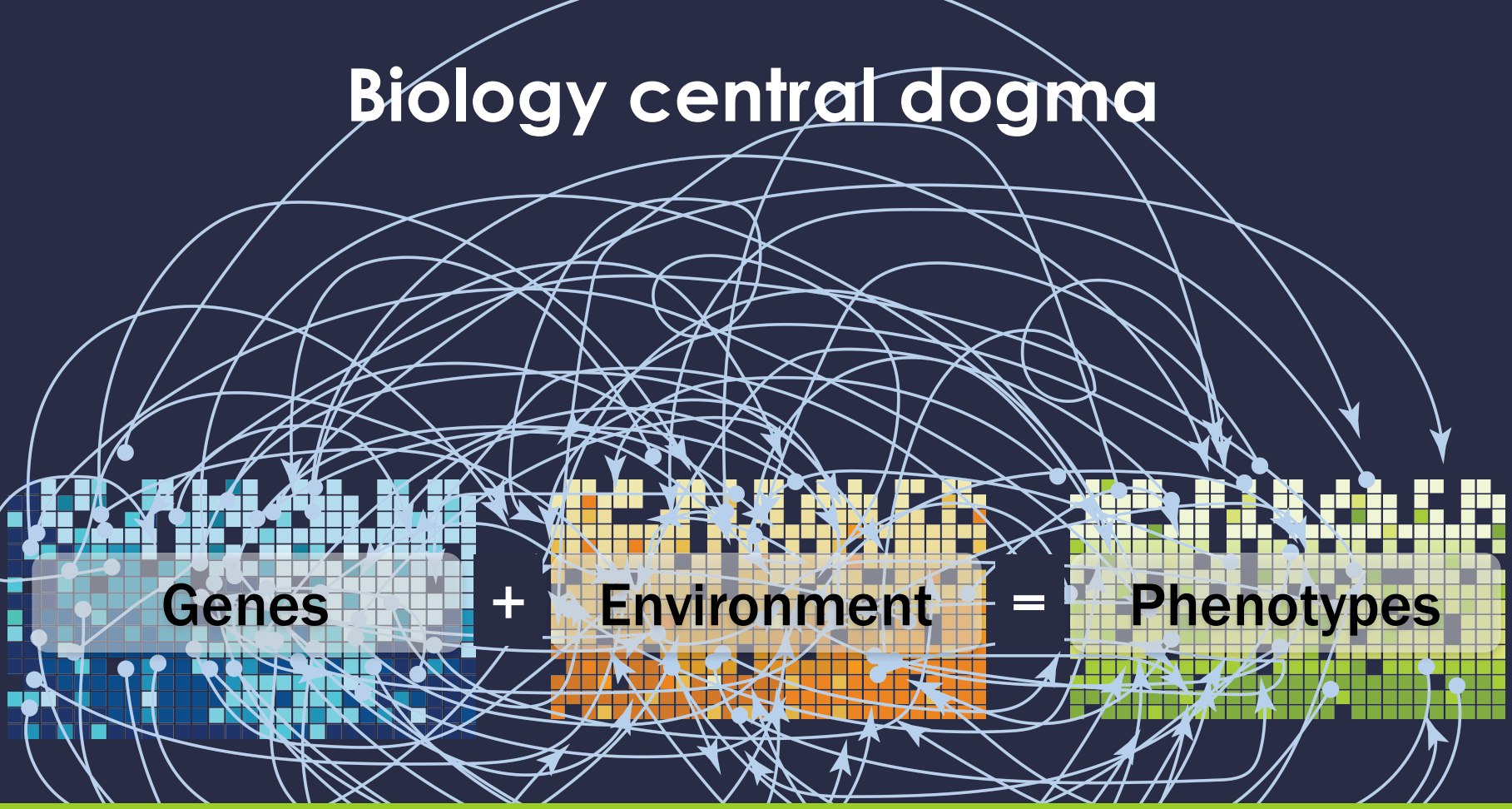
http://bit.ly/exomiser

Ranked STIM-1 variant maximally pathogenic
based on cross-species G2P data,
in the absence of traditional data sources

monarch INITIATIVE

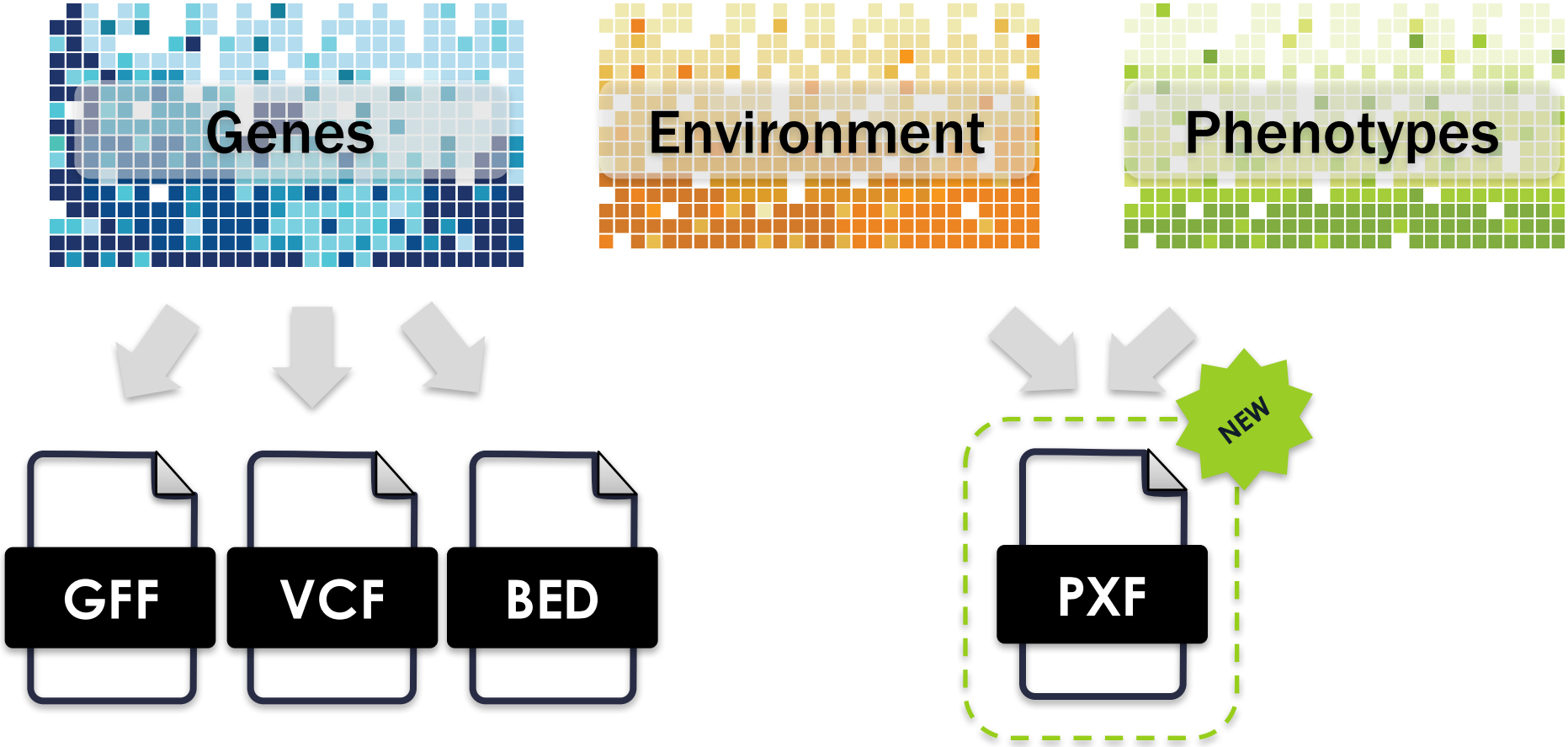
	NIH Undiagnosed Diseases Program	MGI Stim1$^{Sax/Sax}$	OMIM	dbSNP Short Genetic Variations / ClinVar CAGGC TGAGG
Phenotypic profile			N/A	
Genes	Heterozygous, missense mutation STIM-1	Heterozygous, missense mutation STIM-1	N/A	

Biology central dogma



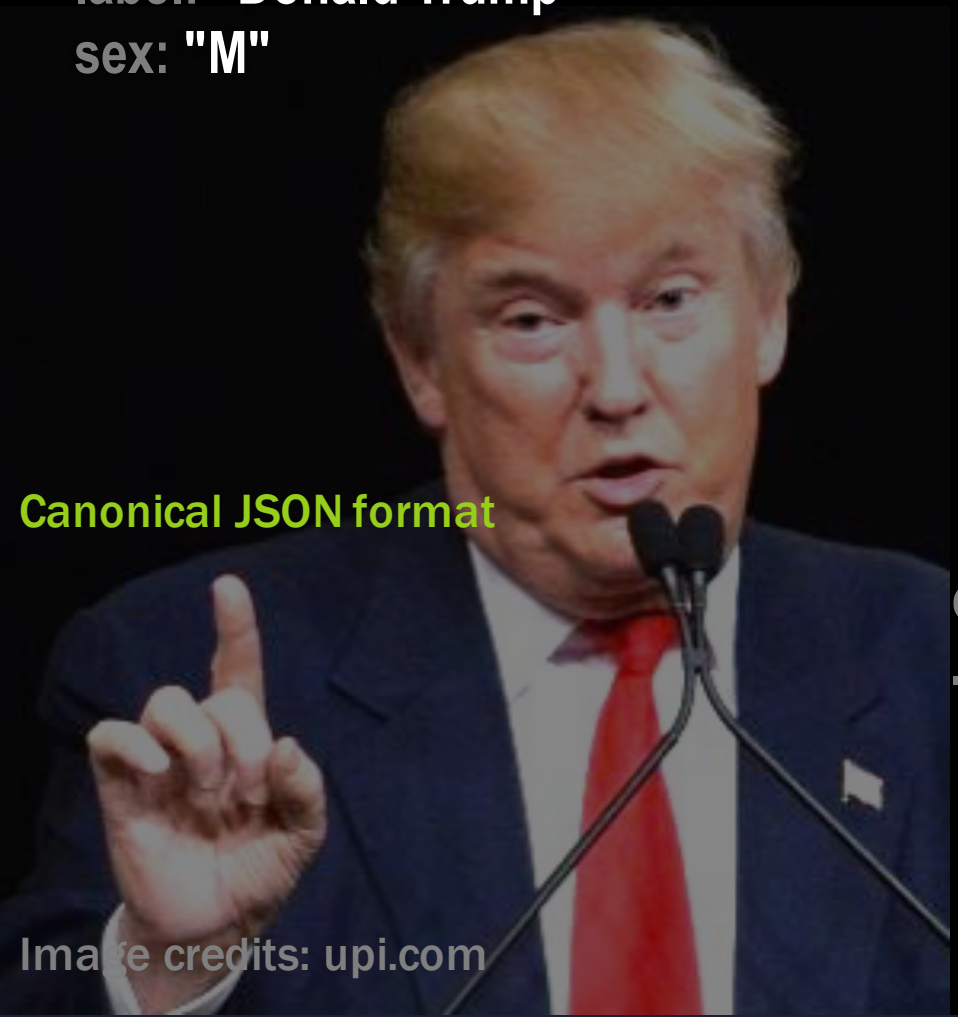
Standards for encoding and exchanging data must be up to these challenges.

Standard exchange formats exist for genes … but for phenotypes? Environment?



What does a PhenoPacket look like?

```
title: "age of onset example"
persons:
- id: "#1"
  label: "Donald Trump"
  sex: "M"
```

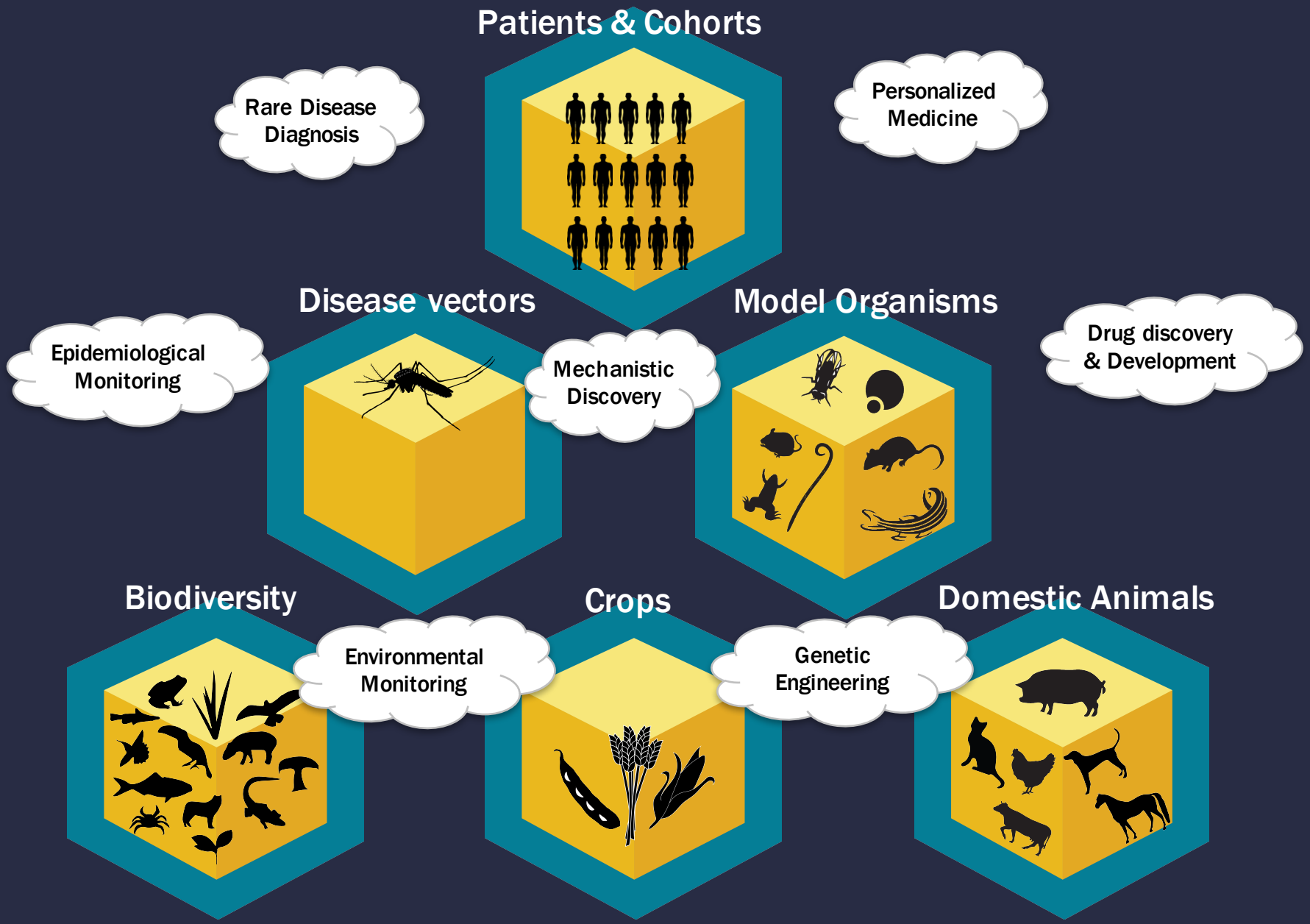


Canonical JSON format

Image credits: upi.com

```
phenotype_profile:
- entity: "person#1"
  phenotype:
    types:
    - id: "HP:0200055"
      label: "Small hands"
    onset:
      description: "during development"
      types:
        - id: "HP:0003577"
          label: "Congenital onset"
evidence:
- types:
  - id: "ECO:0000033"
    label: ”Traceable Author Statement"
  source:
  - id: "PMID:1"
```

If it is alive, it can be PhenoPackaged



Some biodiversity images adapted from http://i.vimeocdn.com/video/417366050_1280x720.jpg

Building a computable model for variant evidence

Provenance

- Materials & methods
- Agent(s) of evidence
- Agent(s) of claim
- Time and place

Evidence

- Data (eg: images, sequences)
- Evidence codes
- Publications
- Confidence (p-val, z-score)
- Summary figures
- Conclusions from previous studies
- Domain expert's knowledge

Claim

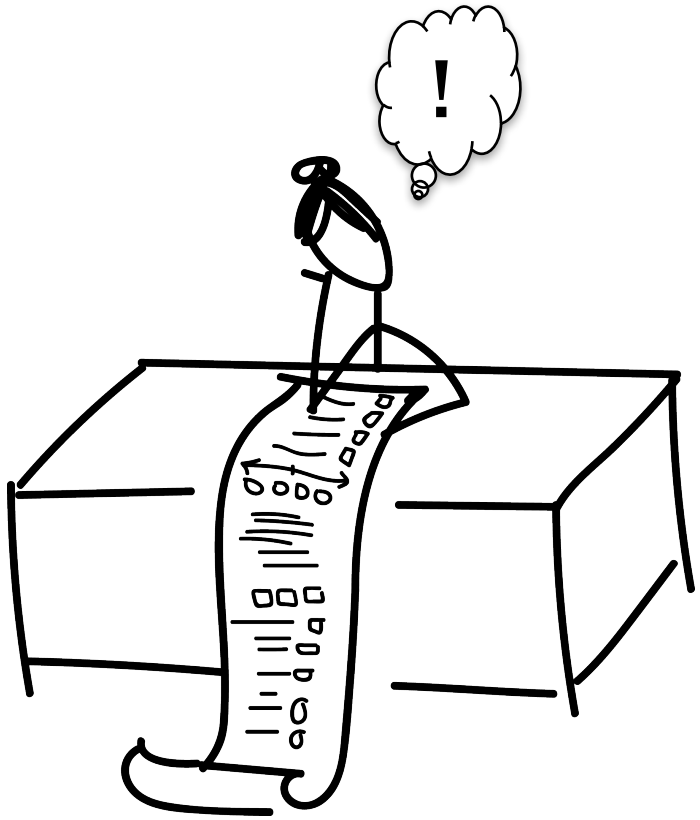
Causal relationships, hypothesized relationships, correlations etc.

Global Alliance for Genomics & Health

http://brcaexchange.org/

https://github.com/monarch-initiative/SEPIO-ontology

PHENOTYPING ISN'T FREE; SO HOW MUCH IS ENOUGH?



no	yes	
☐	☐	Enlarged ears
☐	☐	Hair on head
☐	☐	Enlarged lip
☐	☐	Increased skin pigmentation
☐	☐	Dark hair
☐	☐	Horns
☐	☐	Pointy ears
☐	☐	Blue skin

bit.ly/annotationsufficiency

Summary

- Deep phenotyping within and across species can aid diagnosis, discovery, and translational matchmaking
- An exchange standard is needed to facilitate distributed phenotype data sharing for patients and across species
- A computable G2P evidence model can aid variant interpretation

The Monarch Initiative

www.monarchinitiative.org

PDs: Melissa Haendel, Chris Mungall, Peter Robinson

Funding:

NIH Office of Director: 1R24OD011883; NIH-UDP: HHSN268201300036C, HHSN268201400093P; NCINCI/Leidos #15X143, BD2K U54HG007990-S2 (Haussler) & BD2K PA-15-144-U01 (Kesselman)