If we cannot count rare disease patients, they will not count

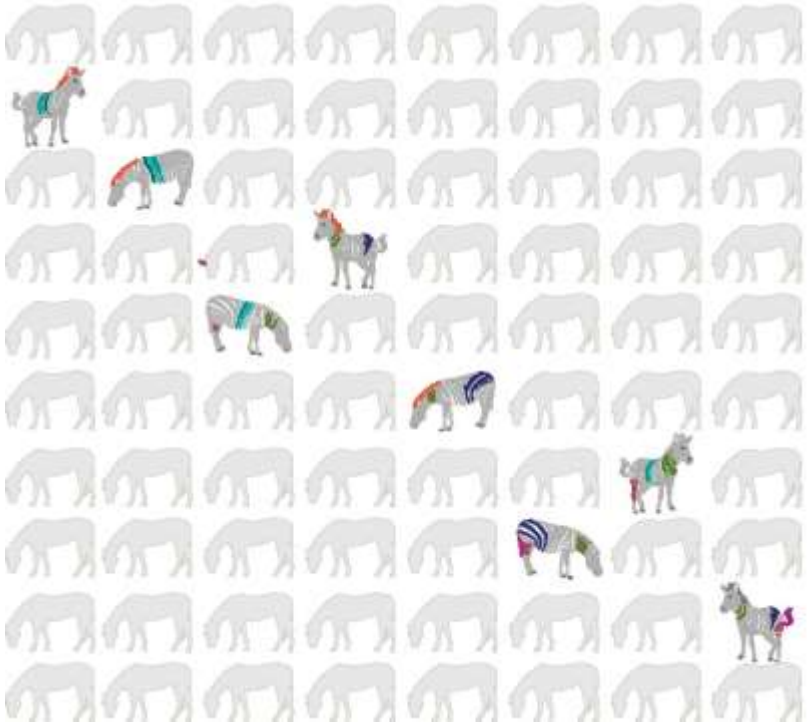
Melissa Haendel, PhD, FACMI

University of Colorado Anschutz Medical Campus

Gold Lab Symposium
20 May, 2022 | These slides:

1/10 americans (400M globally) affected

each patient's characteristics are akin to the variation in zebra stripes



The case for open science: rare diseases
https://academic.oup.com/jamiaopen/article/3/3/472/5904414

"When you hear hoof beats, think horses, not zebras."

"There are ~7000 rare diseases"

- per Orphan Drug Act (1983);

This number is demonstrably wrong.

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

We don't have the same criteria for "rare" 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.

We add new diseases all the time, but don't update the number:

- New rare diseases are discovered every week by organizations such as the Undiagnosed Disease Network
- The literature and public databases abound with new weekly entries
- N-of-1s are matched, defining new diseases

We don't define or IDENTIFY diseases in the same way

- Dozens of terminologies and disease registries exist, each with their own identifier systems or lack thereof
- Rare diseases are often not included in standard clinical terminologies (such as ICD)
- Fundamentally, the definition of a rare disease and how to model it computationally has remained more an art than a science, preventing interoperability

Jessica

6,414,934
variants in Jessica's genome

677,556
are rare

AATACTAA
TTGTAGCCAA
CTGAGTCATTG
GATAACTGATTC
ATAATCCATGCA
TAATCCCTTA
ATGGC

2,826
predicted to cause change in a protein

67
different to her parents

PanelApp

1
was in a gene listed in PanelApp

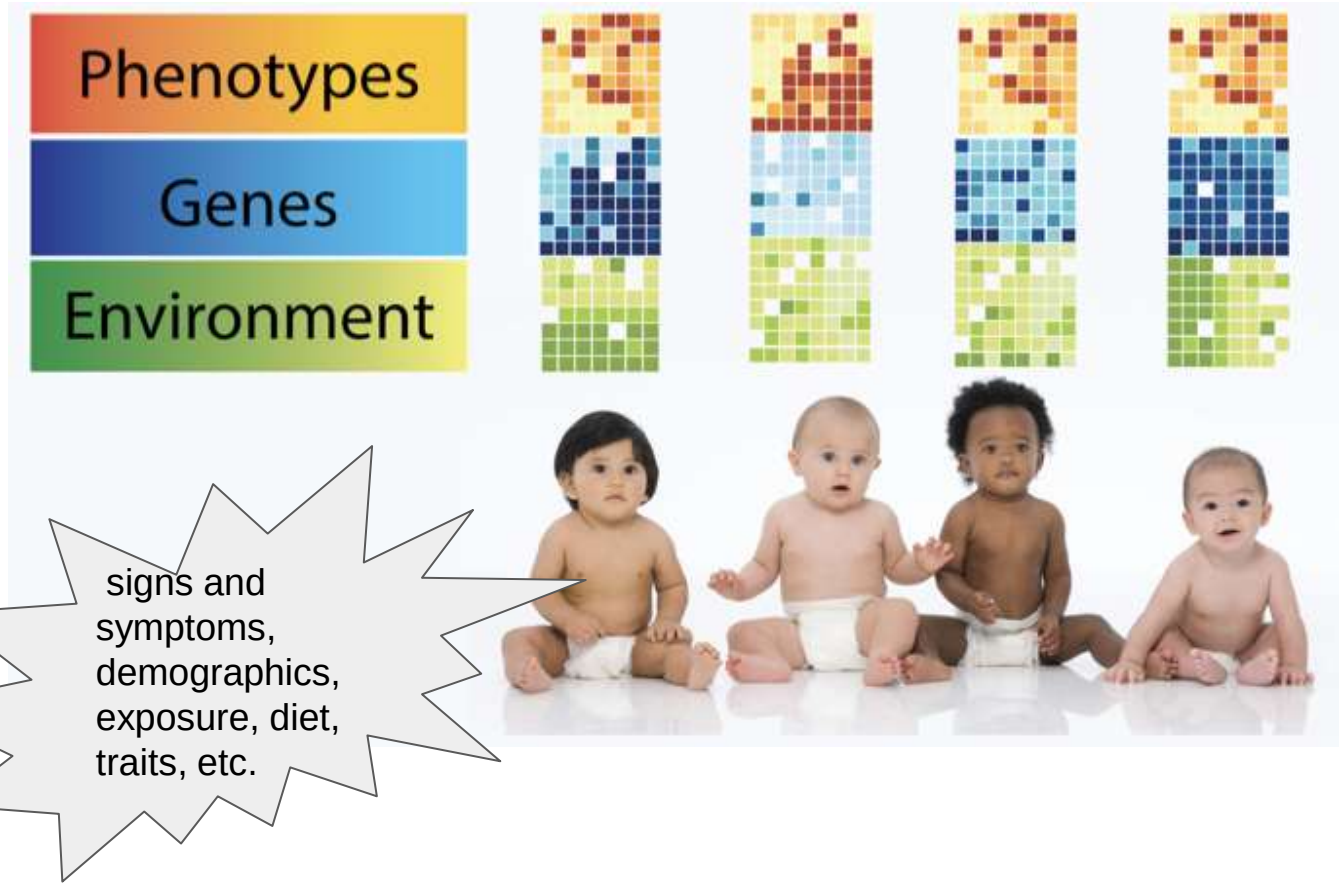
- Jessica (age 4) has a rare condition which causes epilepsy, affects her movement and developmental delay. Standard genetics tests negative.
- De novo deletion in *SLC2A1* identified as the cause of her Glut 1 deficiency syndrome
- Now being successfully treated with a ketogenic, low-carb diet
- Low risk for future pregnancies

Genomics england

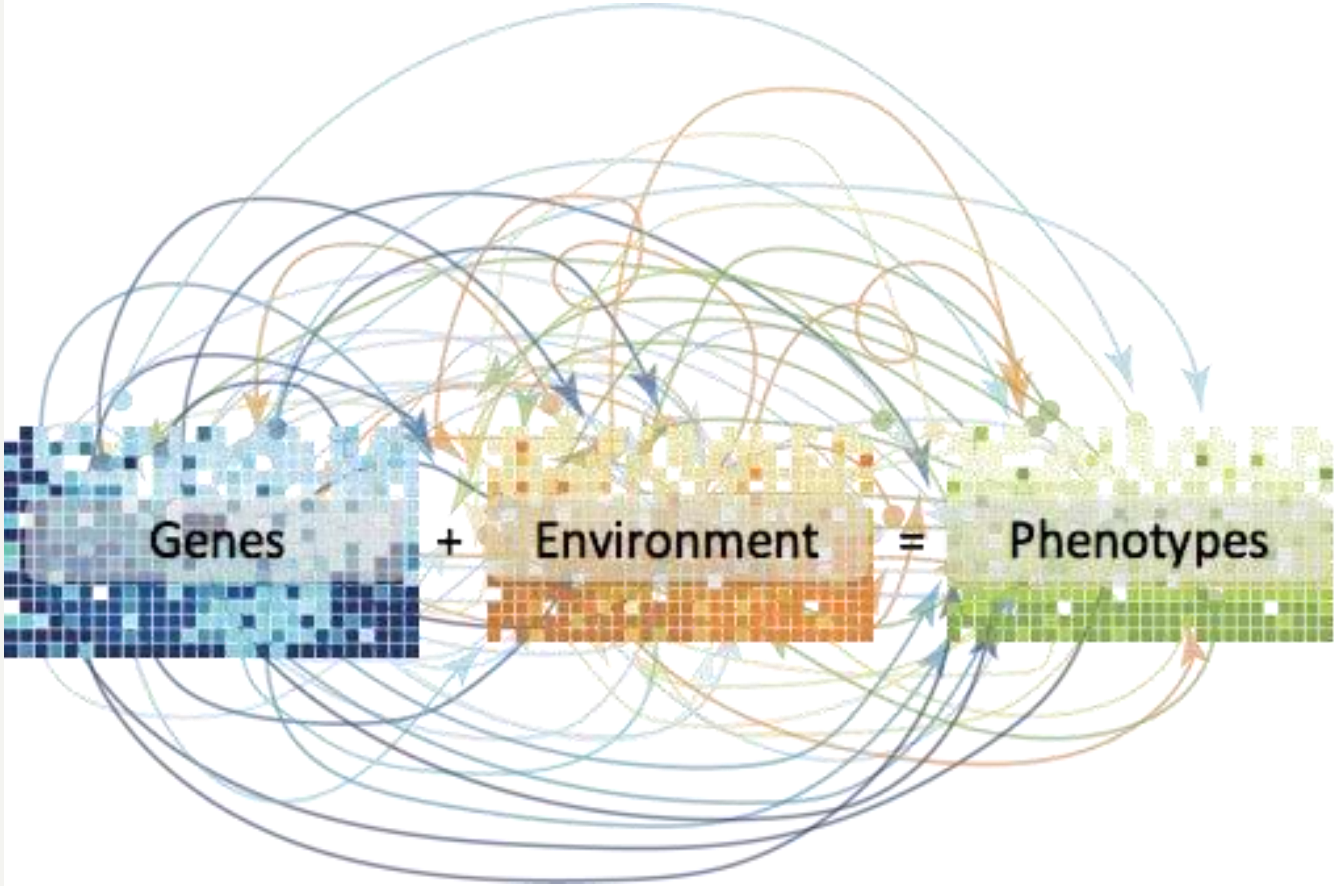
It can take patients 4-7 yrs to get a diagnosis

Every person is an n-of-one disease (this is the promise of precision medicine)

The question is what are meaningfully groupings of patients?



We need to leverage ALL biological knowledge about the relationships



Turning information into meaning

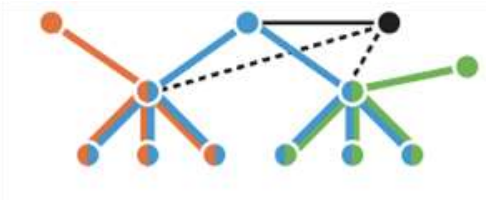
Flat List

e.g. , Problem lists, drugs

Taxonomy

e.g., ICD

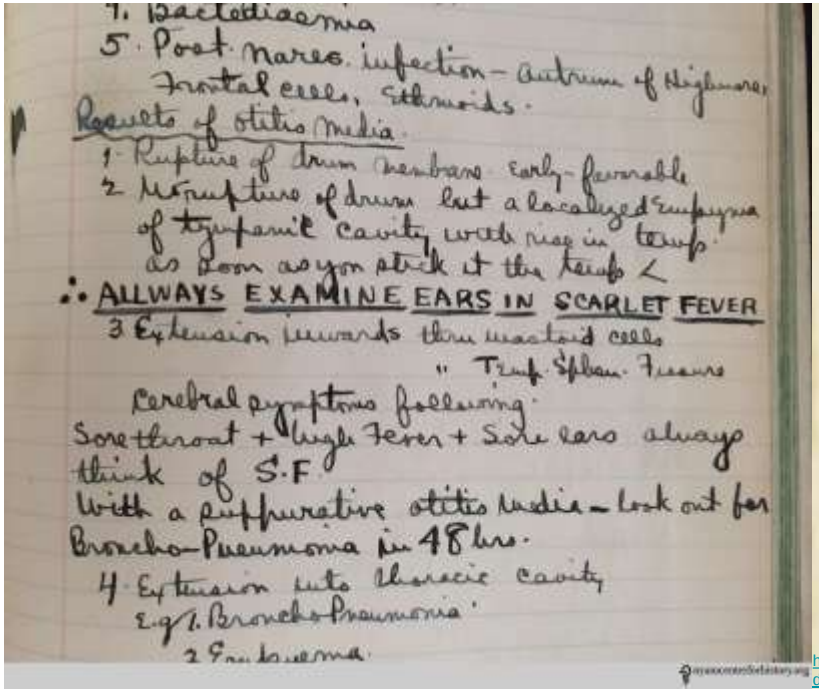
Graph

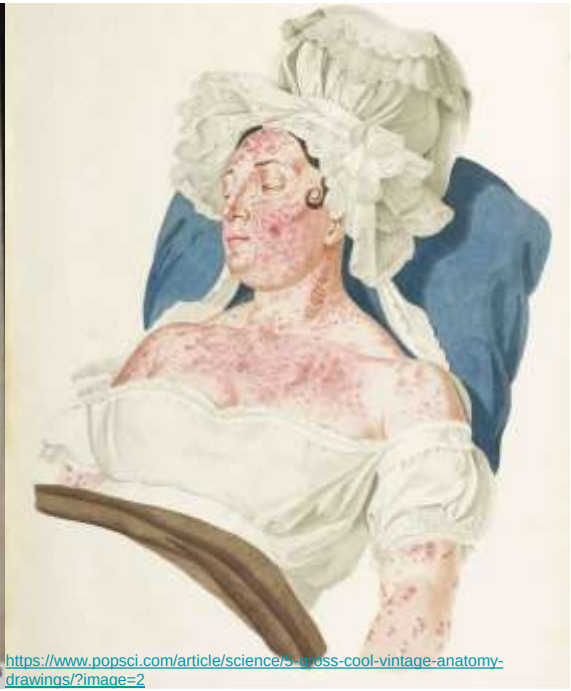


We are mostly here

We need to be here

The challenges start with the basics: Phenotyping





https://www.popsci.com/article/science/5-gross-cool-vintage-anatomy-drawings/?image=2

Patients and families should be empowered to work as partners...

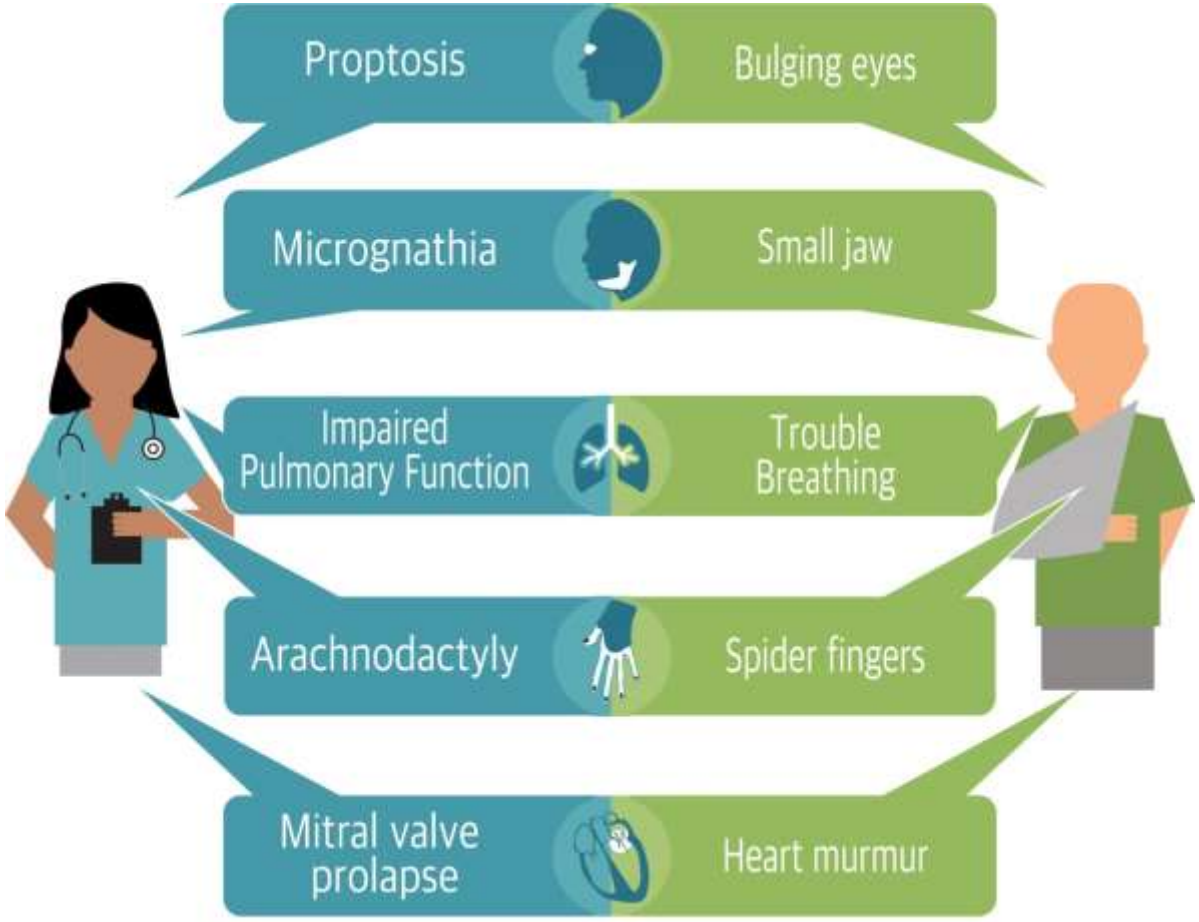
The time is now!

- Provider time constraints
- Care fragmentation
- Knowledge democratization
- Tools and standards emerging to help

Answer:

Give patients the ability to describe their conditions in their own words while leveraging formal vocabularies used by clinicians.

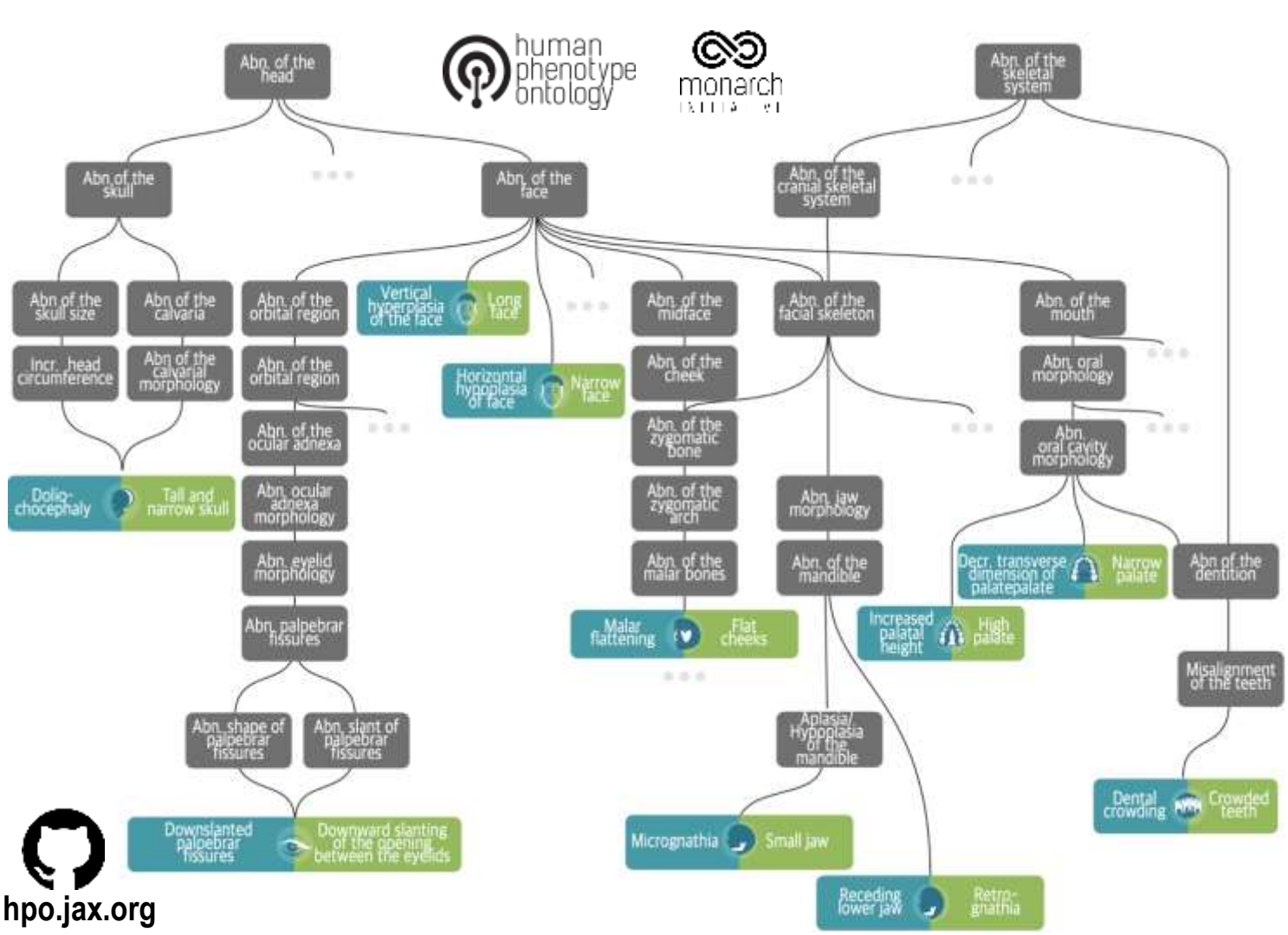
human phenotype ontology



Nature Genetics 50, 474–476 (2018)

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.



When describing a patient, how many terms can you take away and still diagnose the disease?

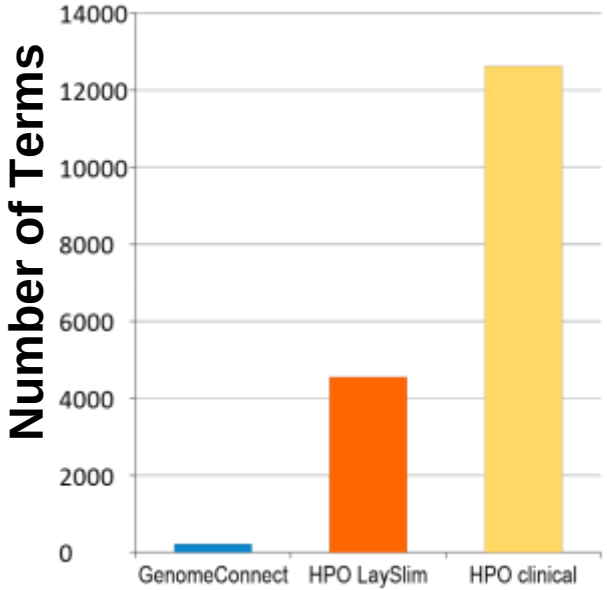
An example comparison of simulated profiles relative to gold standard

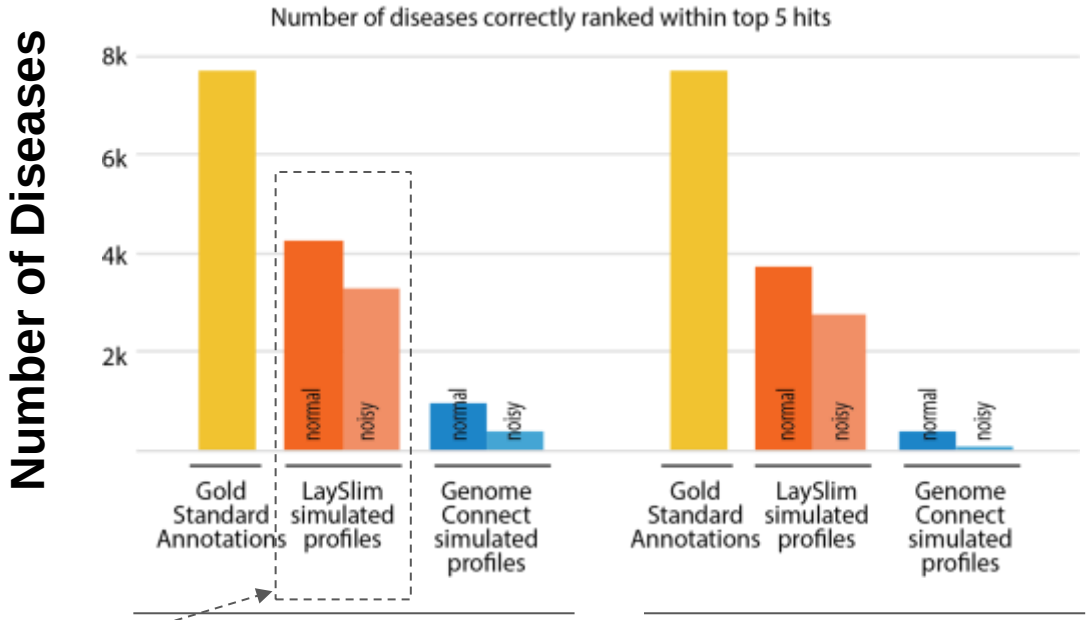
Legend

- Perfect Match
- Fuzzy Match
- No Match

doi:10.6084/m9.figshare.5513356.v2

What is the diagnostic power of the layperson HPO?





When any profile is further constrained by deleted terms or less specific terms general, the diagnostic power reduces as expected.

Algorithm: Boqa

Algorithm: Phenodigm

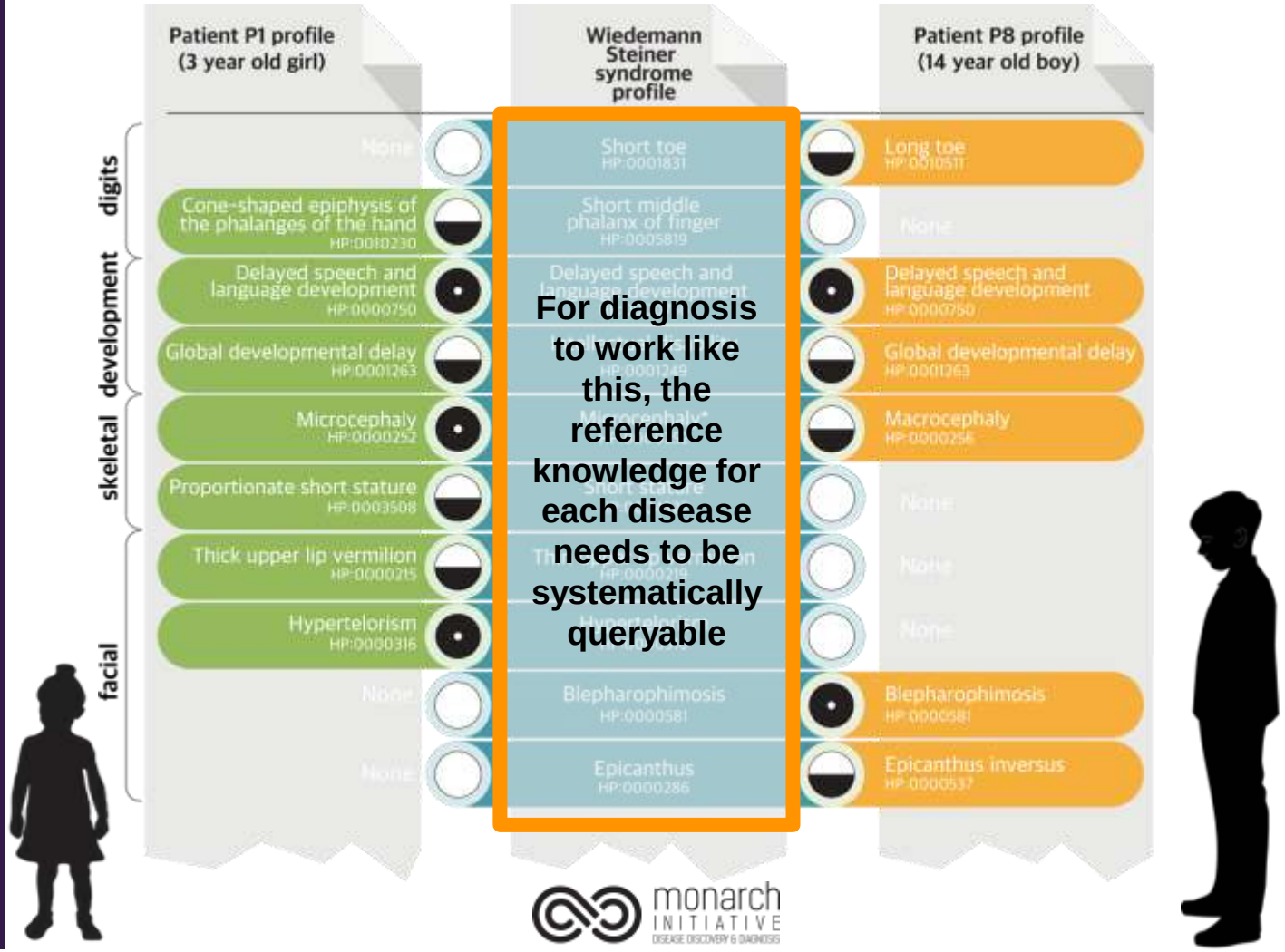
Phenotype-assisted diagnosis in action

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

DOI: 10.1126/scitranslmed.3009262

Legend

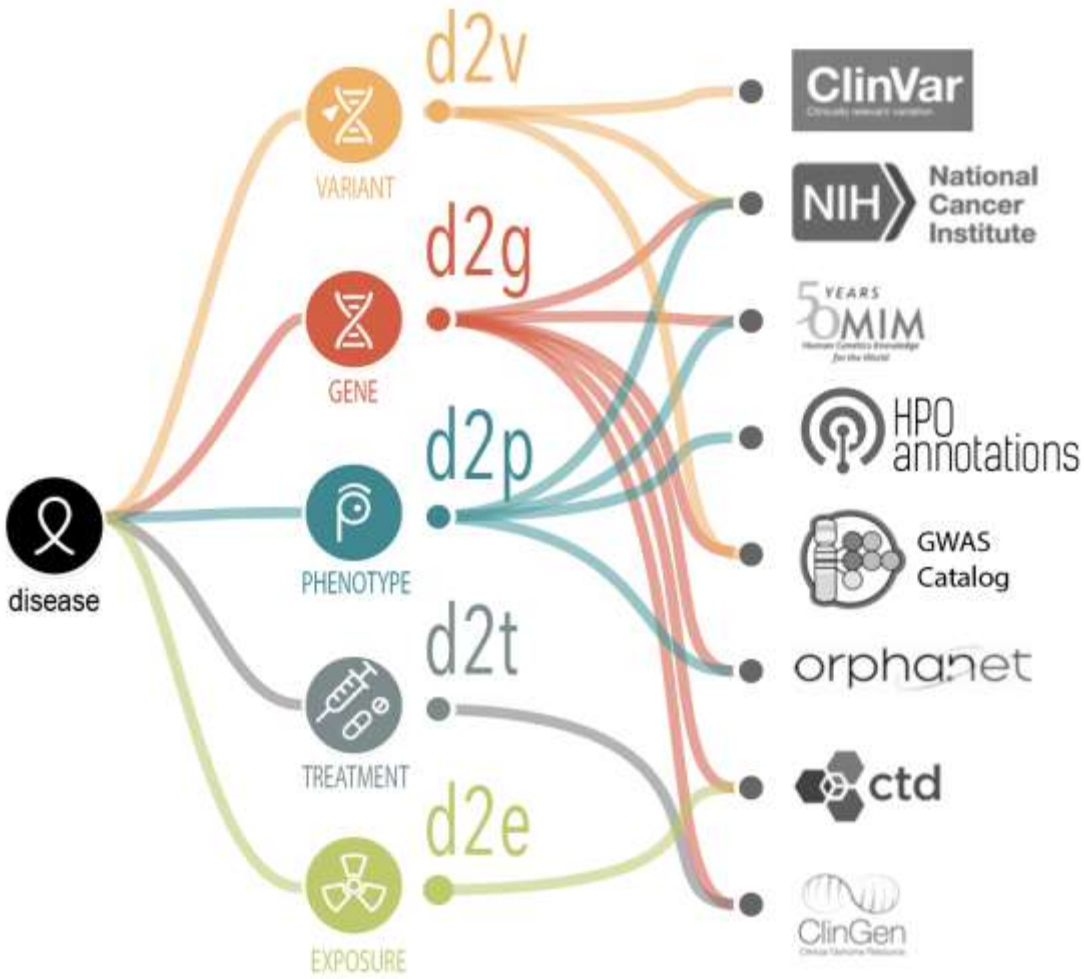
- Perfect Match
- Fuzzy Match
- No Match



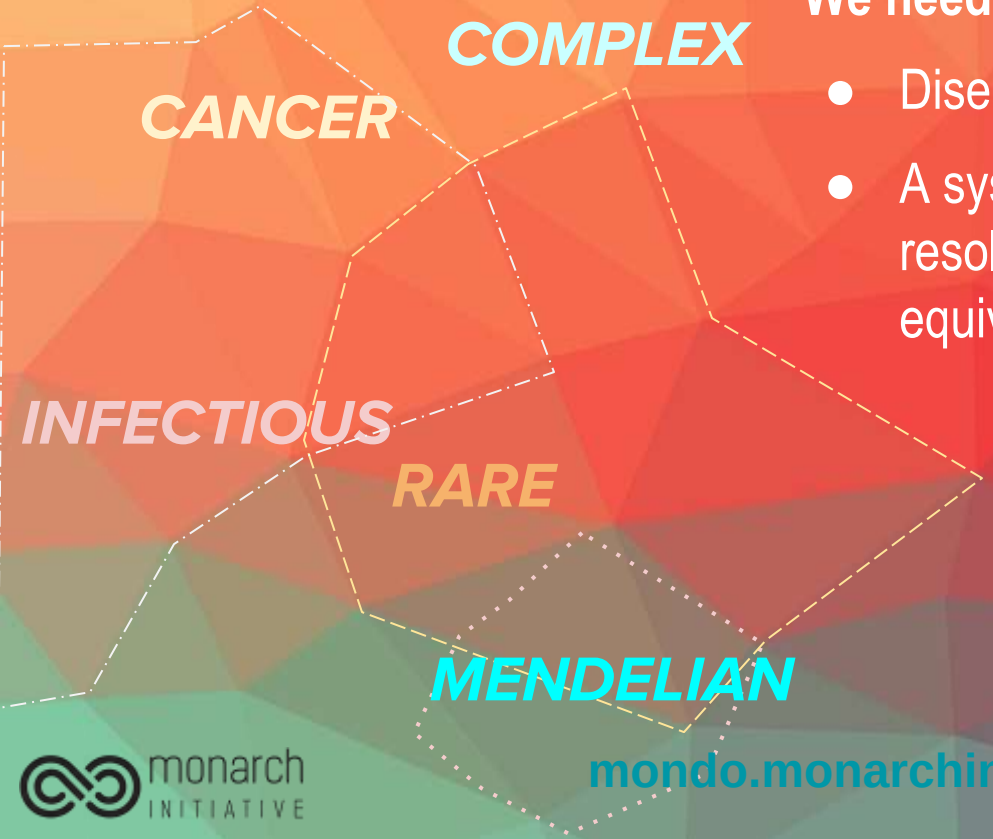
But ...

Assembling a comprehensive source of disease definitions (and annotations) is hard.

Different communities annotate different relationships, at different levels of granularity and using different vocabularies



The knowledge about diseases come from different sources / communities, often siloed by the type of disease

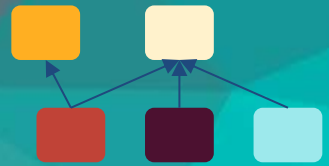


We needed:

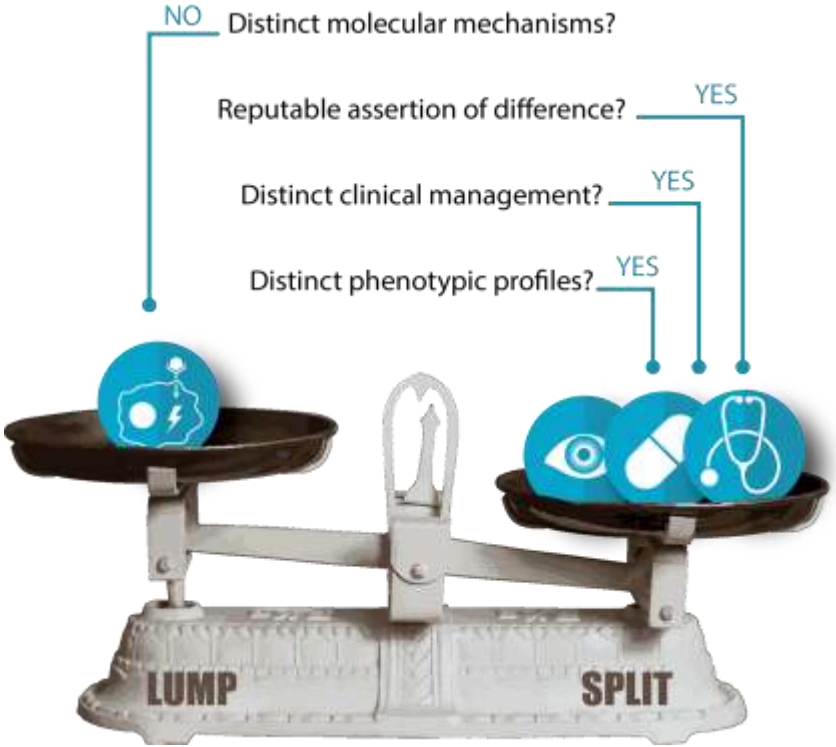
- Disease concepts spanning multiple categories
- A systematic way of relating these concepts and resolving inconsistencies, especially where 1:1 equivalence not possible

monarch INITIATIVE

mondo.monarchinitiative.org



Defining diseases: Lumping and splitting



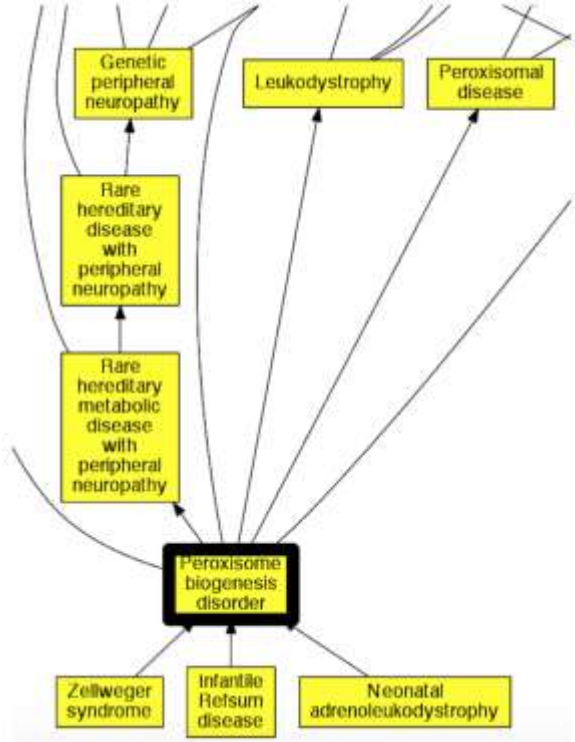
bit.ly/lump-split

ClinGen

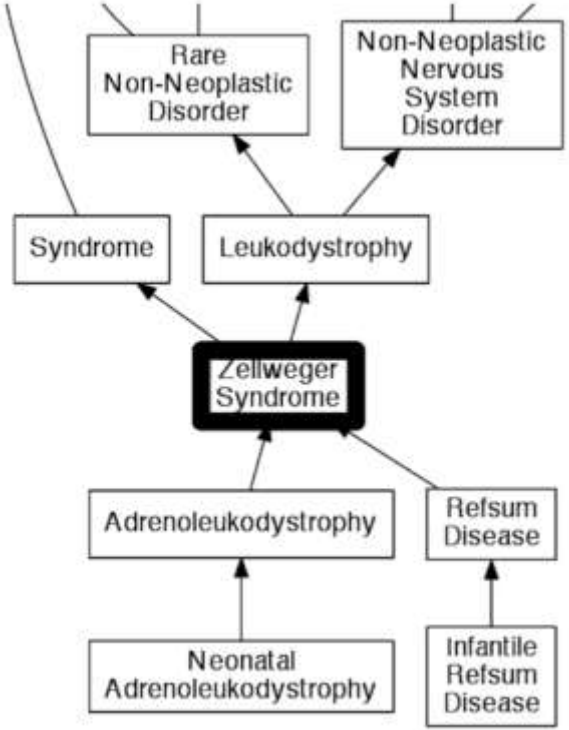
ACMG
American College of Medical Genetics and Genomics

Are they they same disease?

Orphanet



NCIt



DO

```
is_a DOID:0014667 ! disease of metabolism
  is_a DOID:655 ! inherited metabolic disorder
   is_a DOID:906 ! peroxisomal disease
    is_a DOID:0050444 ! infantile Refsum disease
    is_a DOID:0050452 ! mevalonic aciduria
    is_a DOID:0050797 ! peroxisomal acyl-CoA oxidase deficiency
    is_a DOID:0060602 ! alpha-methylacyl-CoA racemase deficiency
    is_a DOID:0090031 ! D-bifunctional protein deficiency
    is_a DOID:2582 ! acatalasia
    is_a DOID:905 ! Zellweger syndrome ***
```

- Different labels
- Different parents
- Different children
- Different synonyms
- Different text definitions

Are they equivalent?

https://github.com/monarch-initiative/mondo/issues/61

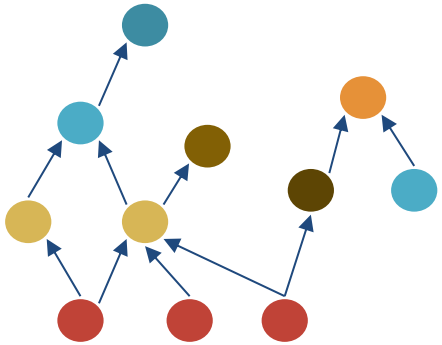
Evidence-based, curated merging of equivalent disease concepts

kBOOM

Bayesian OWL Ontology Merging

Logical + Probabilistic Inference

iterative curator-assisted equivalence inference



evaluate

mondo

Curated Equivalence Relations

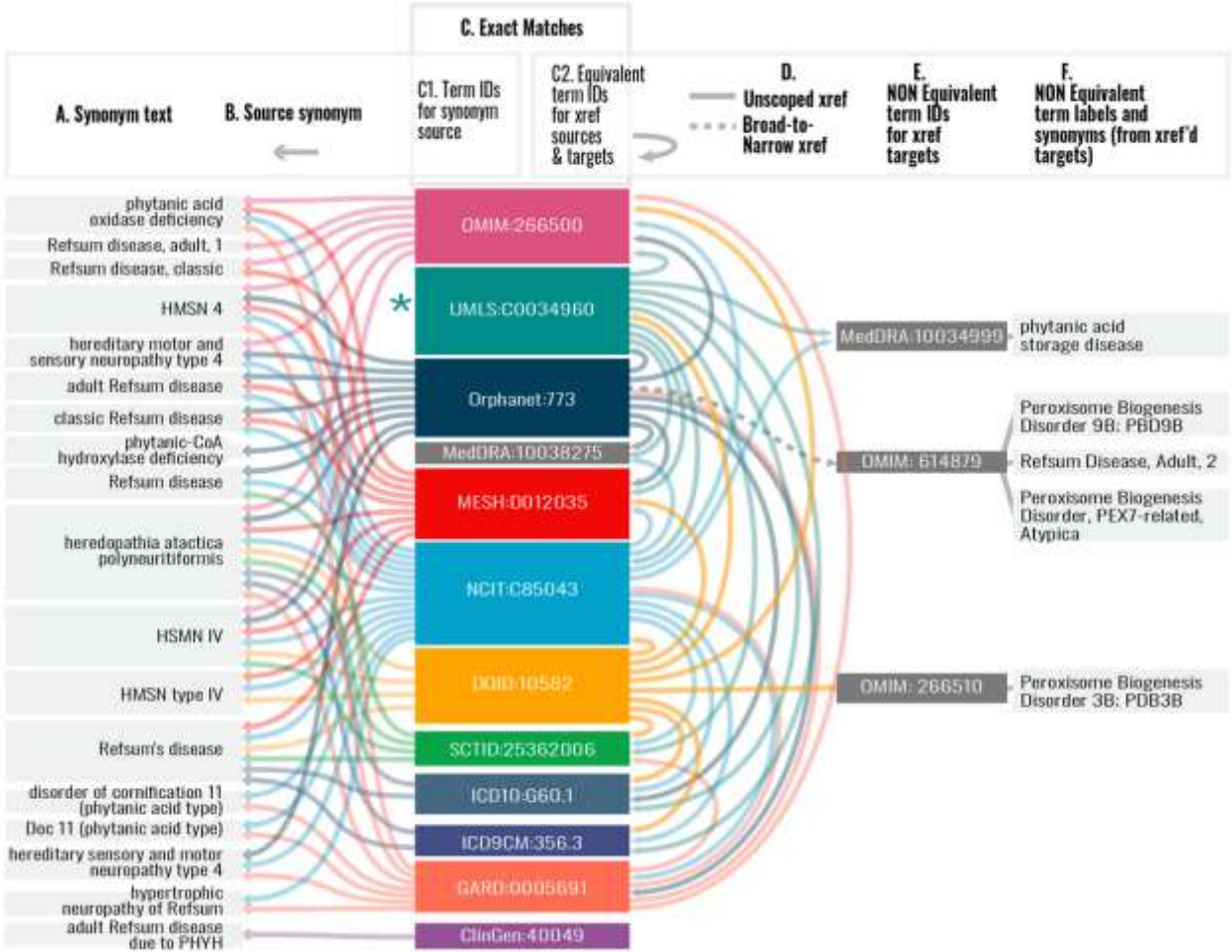
curate

feedback

doi.org/10.1101/048843

Aligning disease knowledge across sources and tracking provenance:

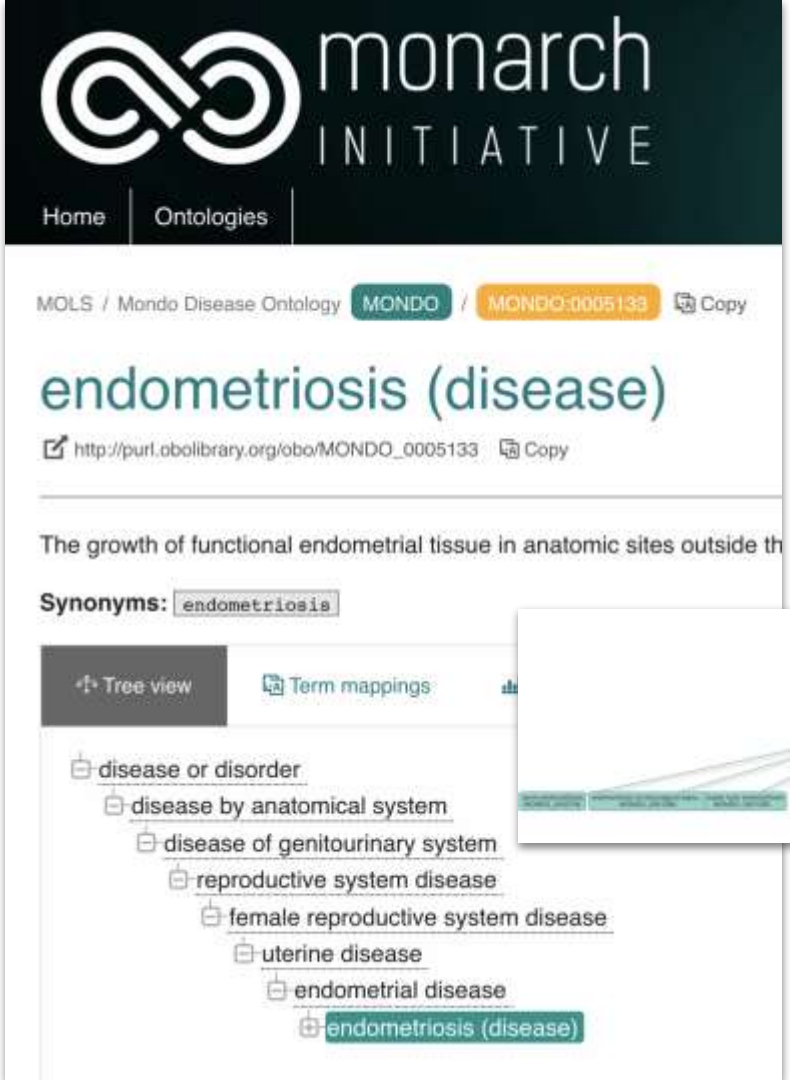
Mondo concept for adult Refsum disease (MONDO:0009958)

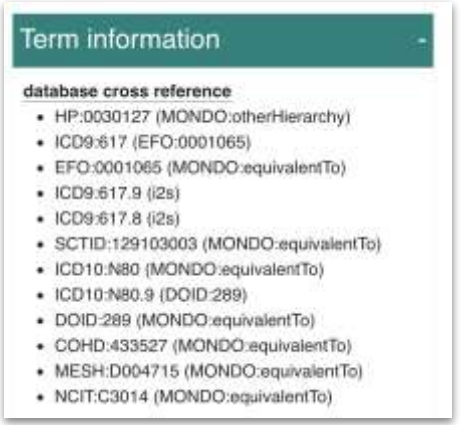


Mondo

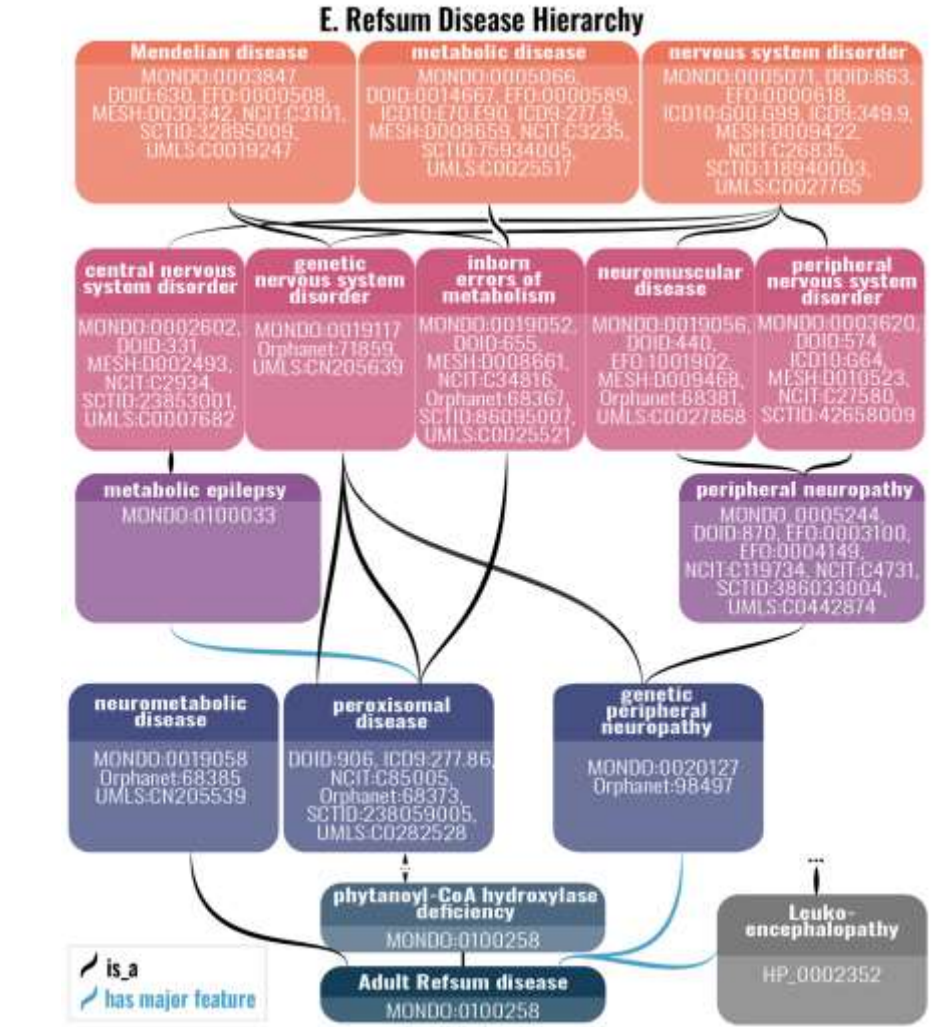
A curated, evidence-based merging to harmonize diseases and phenotypes across sources

mondo.monarchinitiative.org



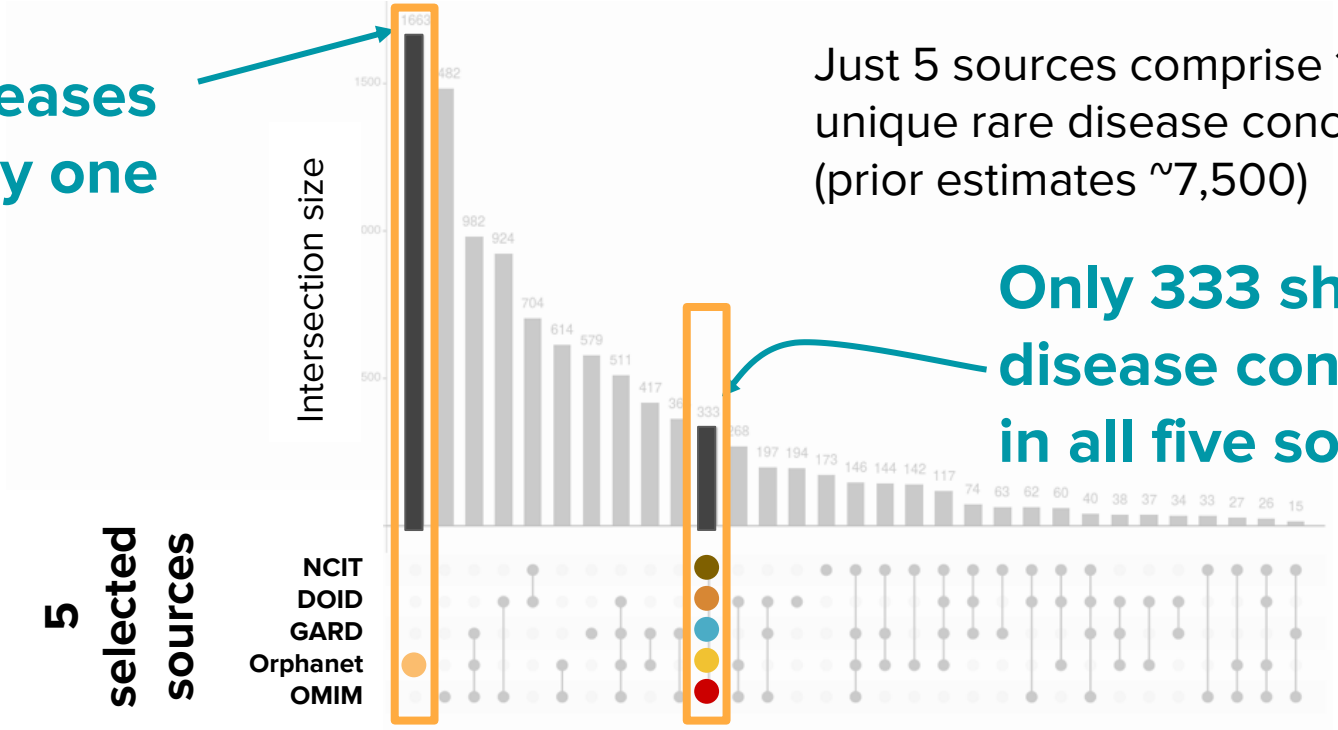


Hierarchical classification of adult Refsum disease



If rare diseases are not counted, rare disease patients will not count

Many diseases are in only one source



Just 5 sources comprise 10,443 unique rare disease concepts (prior estimates ~7,500)

Only 333 shared disease concepts in all five sources

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

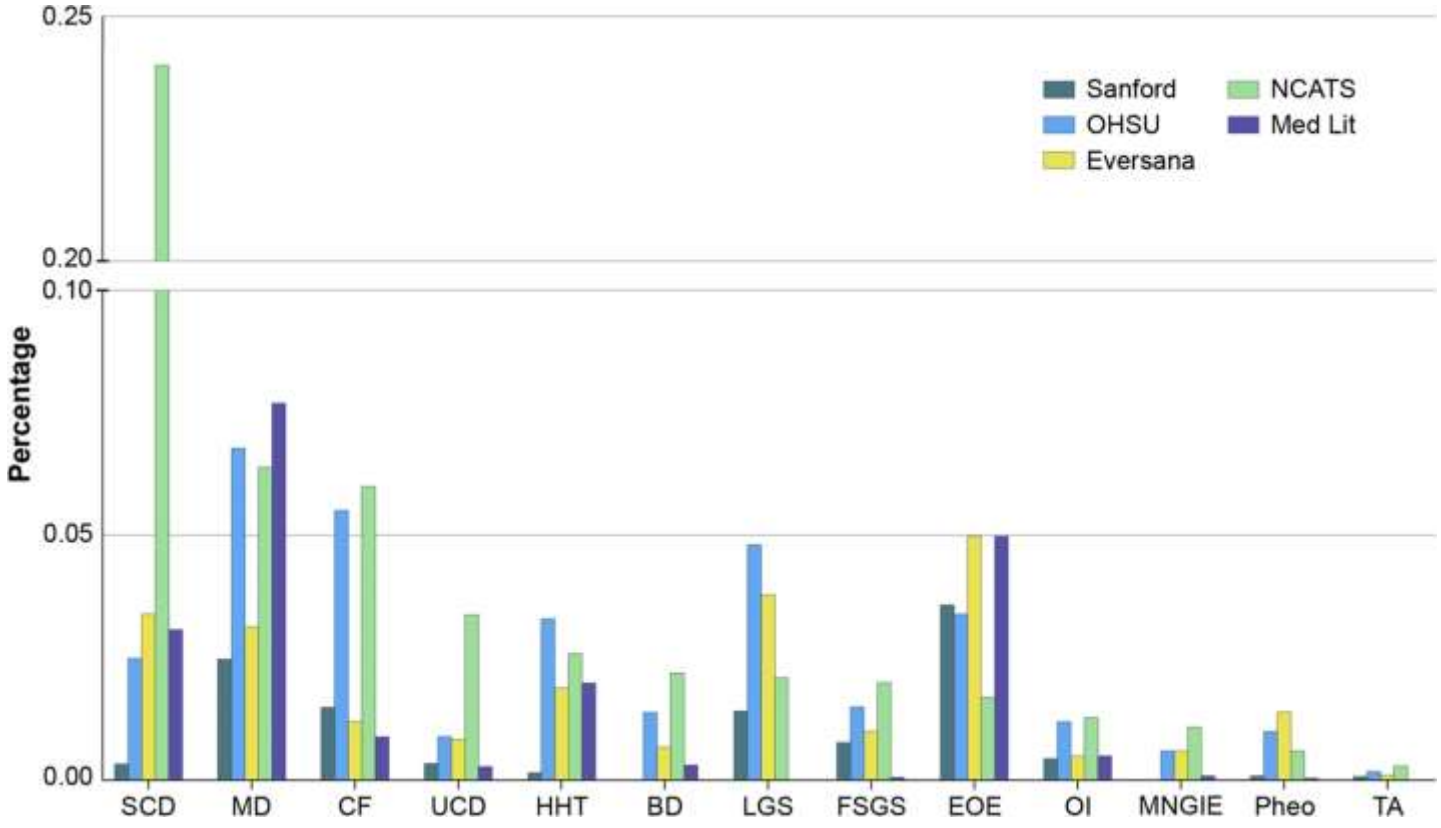
How many undiagnosed patients are there in your hospital?

From EHR data, we can estimate rare disease prevalence (for those with ICD codes) in a single system.

Results often at odds with the published estimates

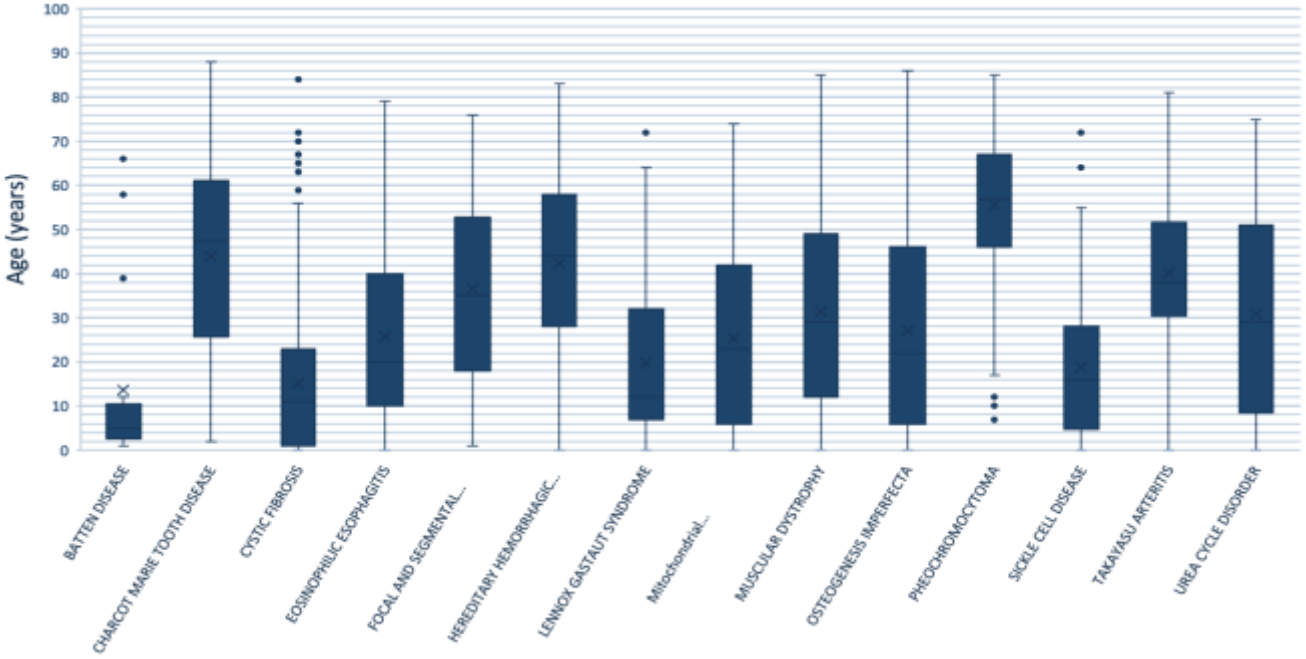
Rare Disease	Patient Count	OHSU Patient Prevalence (1,039,213 total patient count)	Published estimate of population prevalence
Cystic Fibrosis	567	0.0151%	0.0090%
Sickle Cell Disease	259	0.0250%	0.0308%
Muscular Dystrophy	709	0.0680%	0.0769%
Lennox Gastaut Syndrome	503	0.0480%	0.0001%
Urea Cycle Disorder	93	0.0090%	0.0029%
Takayasu's Arteritis	24	0.0020%	0.0002%
Pheochromocytoma	100	0.0100%	0.0005%
Hereditary Hemorrhagic Telangiectasia	341	0.0330%	0.0200%
Osteogenesis Imperfecta	122	0.0120%	0.0050%
Eosinophilic Esophagitis	354	0.0340%	0.0500%
Batten Disease	145	0.0140%	0.0030%
Focal and Segmental Glomerulosclerosis	153	0.0150%	0.0007%
Mitochondrial Neurogastrointestinal Encephalopathy	66	0.0060%	0.0010%

Rare disease prevalence varies across healthcare systems & public sources



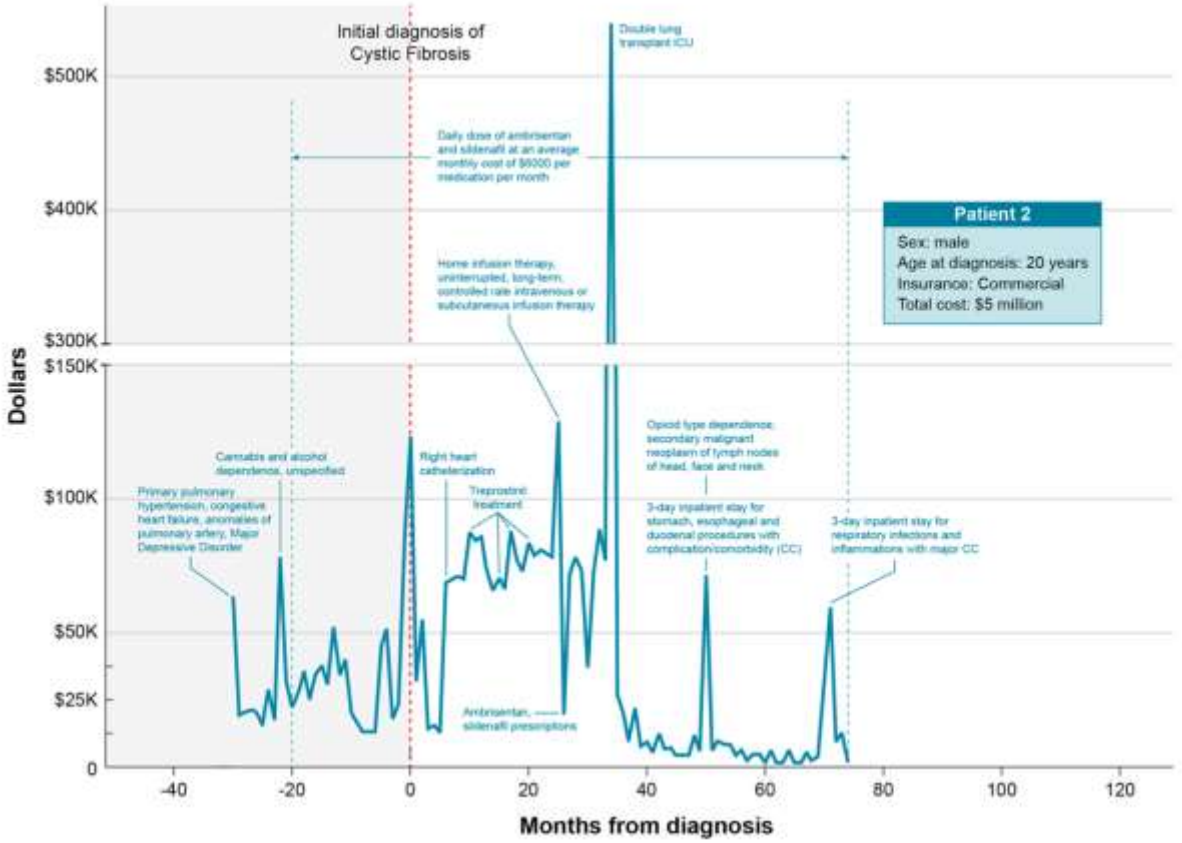
Differences in prevalence across healthcare systems are in part due to rarity and diagnostics, but also due to differences in coding

Age at First Diagnosis Varies by Rare Disease



Large spread of initial diagnosis is due to both challenges in diagnosis, but also in obtaining correct information from the EHR

Each patient is different: what are the patterns?



Patient diagnostic journey maps reveal patterns & differences across health care systems and diseases

Research | Open Access | Published: 22 October 2021

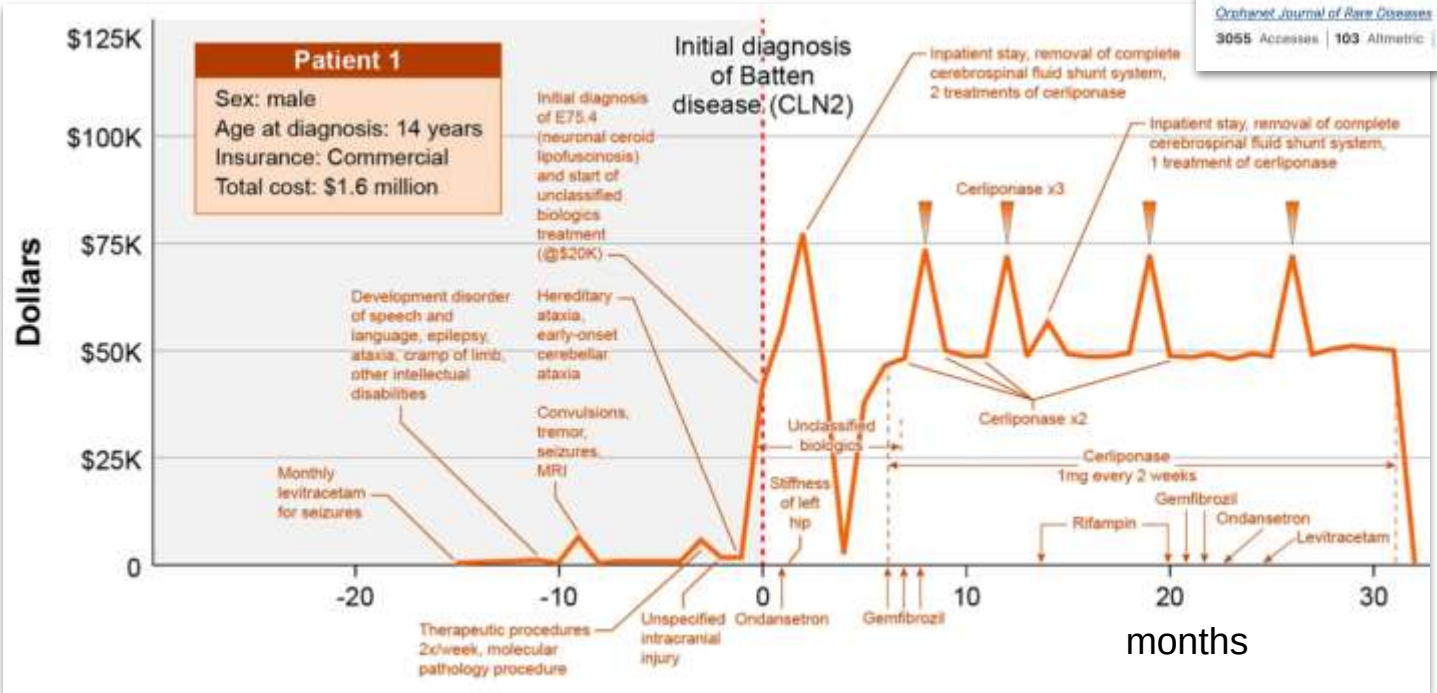
The IDeaS initiative: pilot study to assess the impact of rare diseases on patients and healthcare systems

Ainslie Tisdale, Christine M. Cutillo, Ramaa Nathan, Pierantonio Russo, Bryan Laraway, Melissa Haendel, Douglas Nowak, Cindy Hasche, Chun-Hung Chan, Emily Griese, Hugh Dawkins, Oodaye Shukla, David A. Pearce, Joni L. Rutter & Anne R. Pariser

Orphanet Journal of Rare Diseases **16**, Article number: 429 (2021) | Cite this article

3055 Accesses | 103 Altmetric | Metrics

bit.ly/ideas-initiative-rd



What about all the Rare Diseases without a ICD code?

We must get rare disease definitions into the EHR, but we have to start where we are

It's a Mad Max Sitch.

ICD-10

mondo

Image: https://www.trucks.com/2015/05/05/vehicles-of-mad-max/

Thank you!

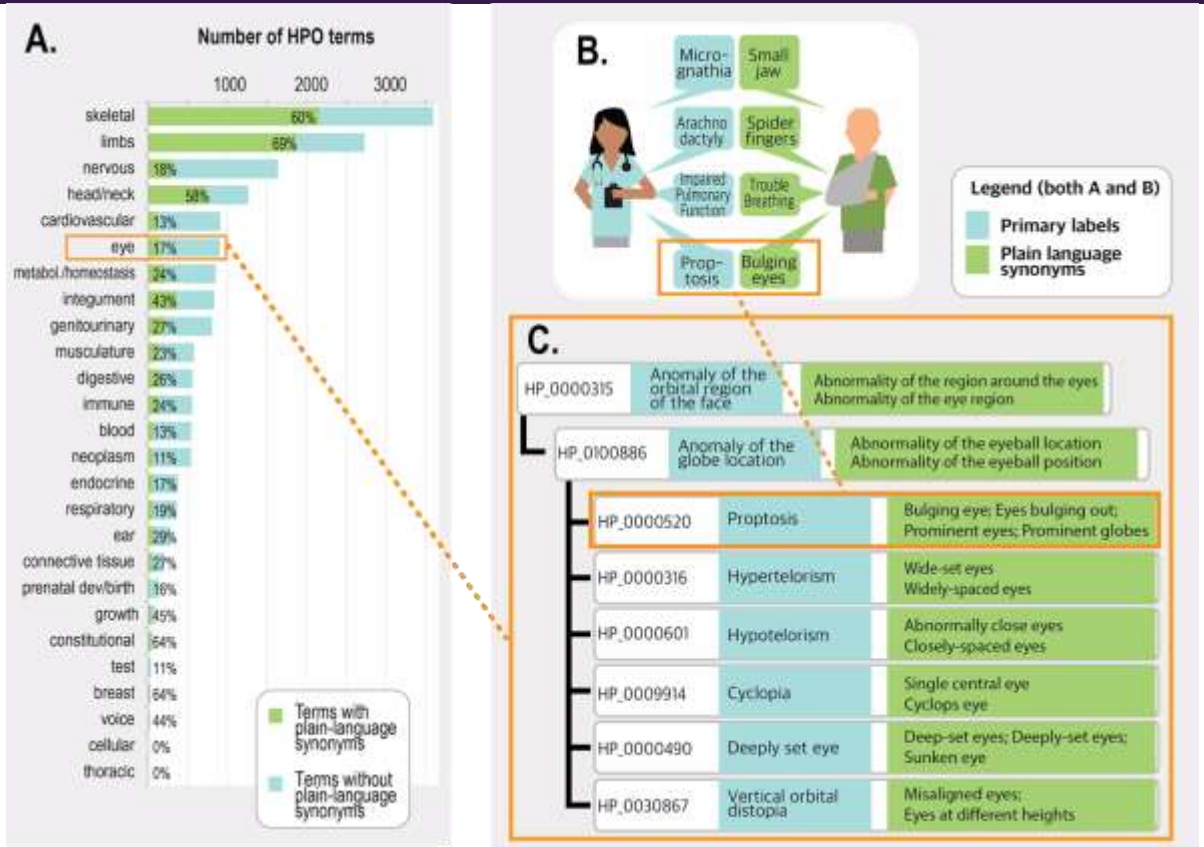
monarchinitiative.org

Peter Robinson (PI)
Chris Mungall (PI)
Bryan Laraway
Damian Smedley
David Osumi-Sutherland
Ada Hamosh
Chris Chute
Julie McMurry
Jules Jacobsen
Monica Munoz-Torres
Nomi Harris
Sebastian Koeller
Tim Putman
Anne Thessen
Harry Caufield
Harshad Hedge
Justin Reese
Kevin Schaper
Lauren Chan
Matt Brush
Nico Matentzoglu
Nicole Vasilevsky
Sabrina Toro
Sarah Gehrke
Seth Carbon
Shawn O'Neil
Sierra Moxon
Tudor Groza
Victoria Soesanto

Funded by NIH OD R24 OD011883, NHGRI CEGS RM1, U24HG011449

Extra

Plain language synonyms for patients to use



4887 of 13823 HPO terms have lay synonyms

Open access.
Open source.
Open science.

bit.ly/case-for-rd-open-science

The case for open science: rare diseases

Yaffa R Rubinstein ✉, Peter N Robinson, William A Gahl, Paul Avillach, Gar
Helene Cederroth, Rebecca M Goodwin, Stephen C Groft, Mats G Hansson,
Vojtech Huser, Deborah Mascalzoni, Julie A McMurry, Matthew Might, Chri
Barend Mons, Dina N Paltoo, Jonathan Pevsner, Manuel Posada, Alison P
Marco Roos, Tamar B Rubinstein ✉, Domenica Taruscio, Esther van Encke
Melissa A Haendel
Author Notes

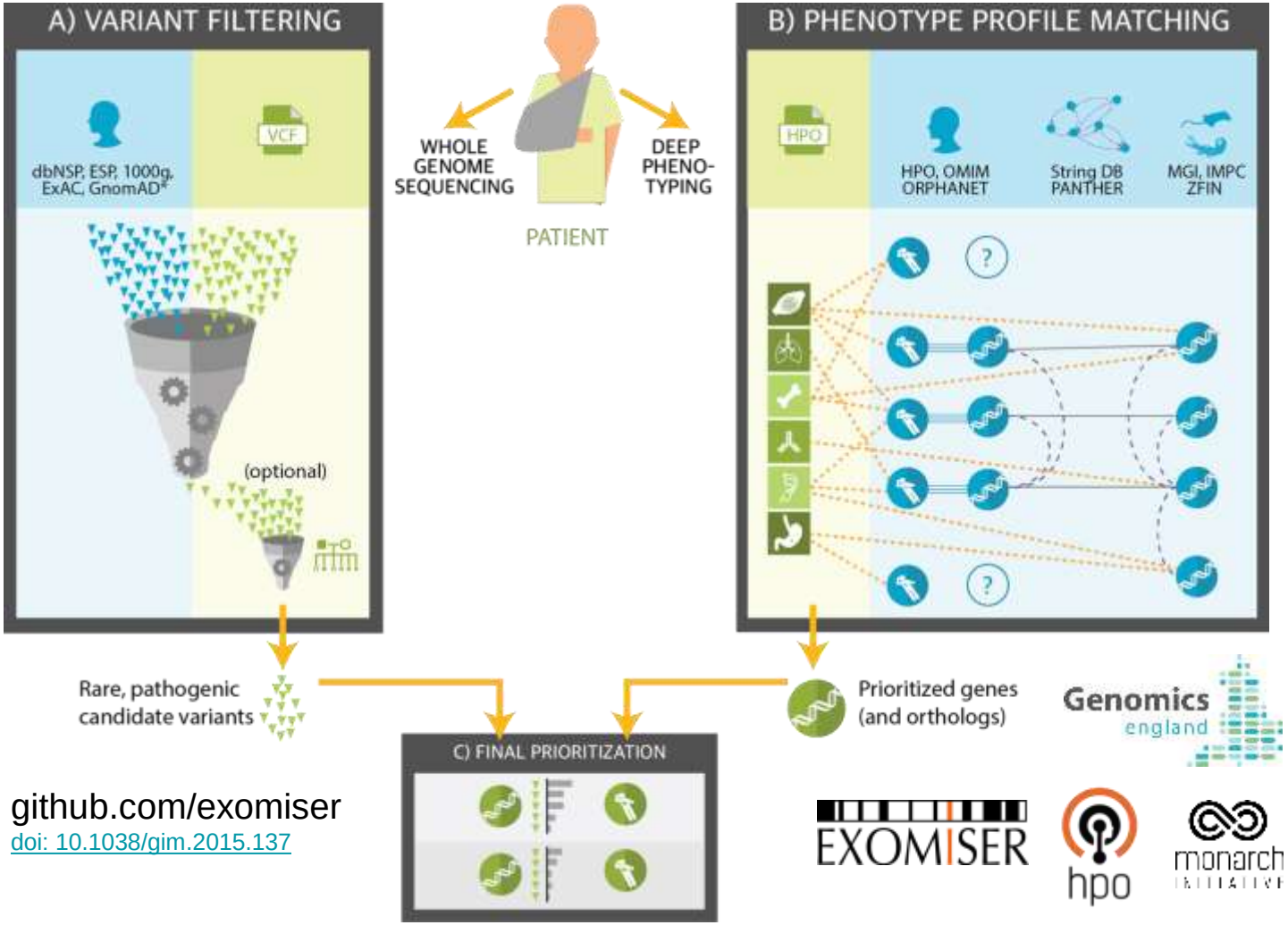
JAMIA Open, Volume 3, Issue 3, October 2020, Pages 472–486,
https://doi.org/10.1093/jamiaopen/ooaa030
Published: 11 September 2020 **Article history** ▾

PDF Split View Cite Permissions Share ▾

Abstract

The premise of Open Science is that research and medical managem
progress faster if data and knowledge are openly shared. The value
Science is nowhere more important and appreciated than in the rar
(RD) community. Research into RDs has been limited by insufficien

Combining genomic and phenomic data improves variant prioritization for diagnosis

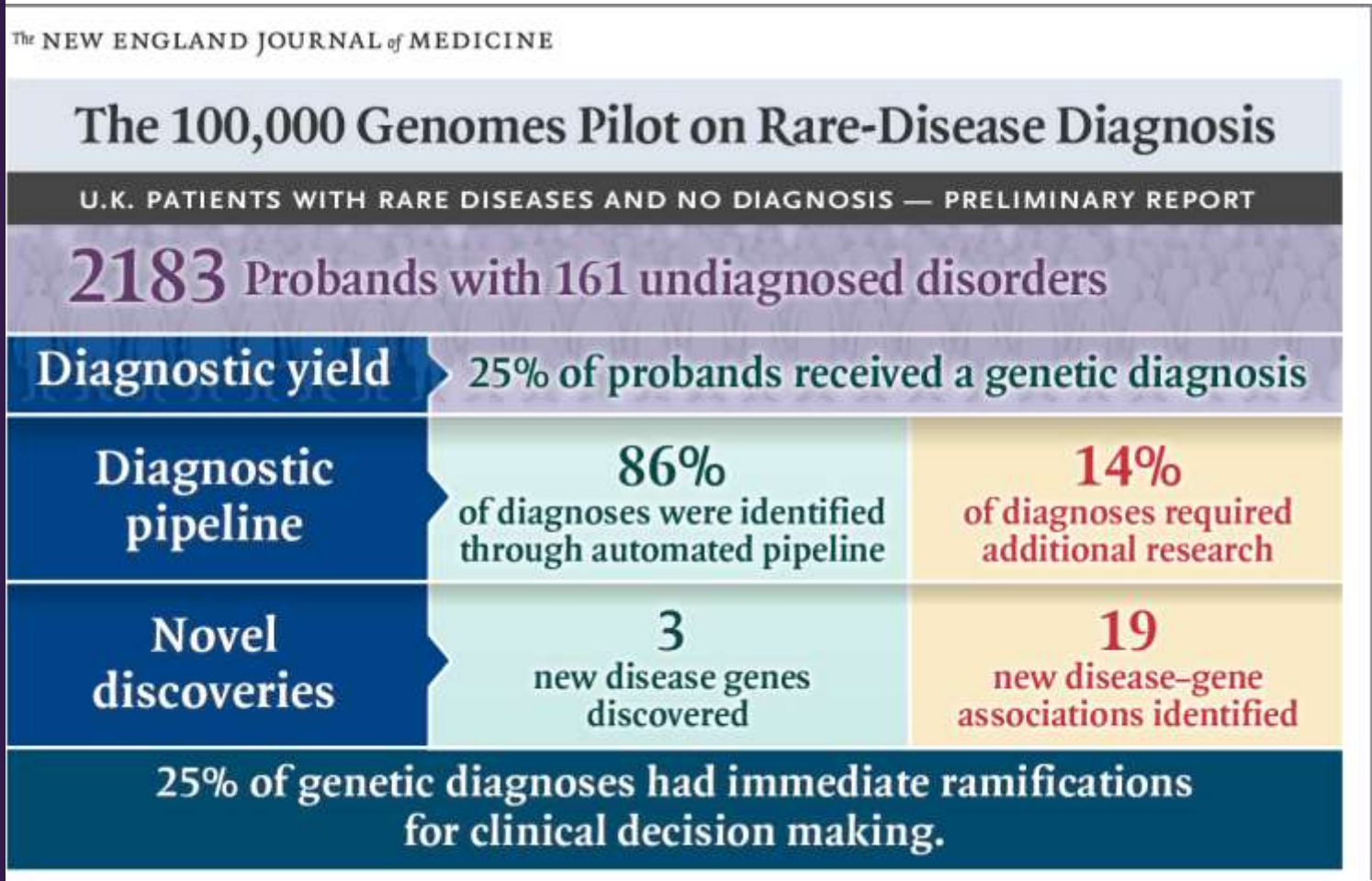


github.com/exomiser

doi: 10.1038/gim.2015.137

Evaluation of Exomiser in 100K genomes project:

Made possible by open science



November 11, 2021

https://www.nejm.org/doi/full/10.1056/NEJMoa2035790

Mondo community paper

with ClinGen OMIM, and the rest of the village!

https://www.medrxiv.org/content/10.1101/2022.04.13.22273750v3

medRxiv

THE PREPRINT SERVER FOR HEALTH SCIENCES

Cold Spring Harbor Laboratory BMJ Yale

Mondo: Unifying diseases for the world, by the world

Nicole A Vasilevsky, Nicolas A Matentzoglu, Sabrina Toro, Joseph E Flack IV, Harshad Hegde, Deepak R Unni, Gioconda F Alyea, Joanna S Amberger, Larry Babb, James P Balhoff, Taylor I Bingaman, Gully A Burns, Orion J Buske, Tiffany J Callahan, Leigh C Carmody, Paula Carrio Cordo, Lauren E Chan, George S Chang, Sean L Christiaens, Louise C Daugherty, Michel Dumontier, Laura E Failla, May J Flowers, H. Alpha Garrett Jr., Jennifer L Goldstein, Dylan Gration, Tudor Groza, Marc Hanauer, Nomi L Harris, Jason A Hilton, Daniel S Himmelstein, Charles Tapley Hoyt, Megan S Kane, Sebastian Köhler, David Lagorce, Abbe Lai, Martin Larralde, Antonia Lock, Irene López Santiago, Donna R Maglott, Adriana J Malheiro, Birgit H M Meldal, Monica C Munoz-Torres, Tristan H Nelson, Frank W Nicholas, David Ochoa, Daniel P Olson, Tudor I Oprea, David Osumi-Sutherland, Helen Parkinson, Zoë May Pendlington, Ana Rath, Heidi L Rehm, Lyubov Remennik, Erin R Riggs, Paola Roncaglia, Justyne E Ross, Marion F Shadbolt, Kent A Shefchek, Morgan N Similuk, Nicholas Sioutos, Damian Smedley, Rachel Sparks, Ray Stefancsik, Ralf Stephan, Andrea L Storm, Doron Stupp, Gregory S Stupp, Jagadish Chandrabose Sundaramurthi, Imke Tammen, Darin Tay, Courtney L Thaxton, Eloise Valasek, Jordi Valls-Margarit, Alex H Wagner, Danielle Welter, Patricia L Whetzel, Lori L Whiteman, Valerie Wood, Colleen H Xu, Andreas Zankl, Xingmin Aaron Zhang, Christopher G Chute, Peter N Robinson, Christopher J Mungall, Ada Hamosh, Melissa A Haendel

doi: https://doi.org/10.1101/2022.04.13.22273750

Mondo accommodates gene-based names AND groups terms based on gene-based etiology

For example, GTP cyclohydrolase I deficiency

- GTP cyclohydrolase I deficiency → ClinGen curation target
  - dystonia 5
  - GTP cyclohydrolase I deficiency with hyperphenylalaninemia

Mondo terms mapped to OMIM

128230

DYSTONIA, DOPA-RESPONSIVE; DRD

Alternative titles; symbols

DYSTONIA 5; DYT5

233910

HYPERPHENYLALANINEMIA, BH4-DEFICIENT, B; HPABH4B

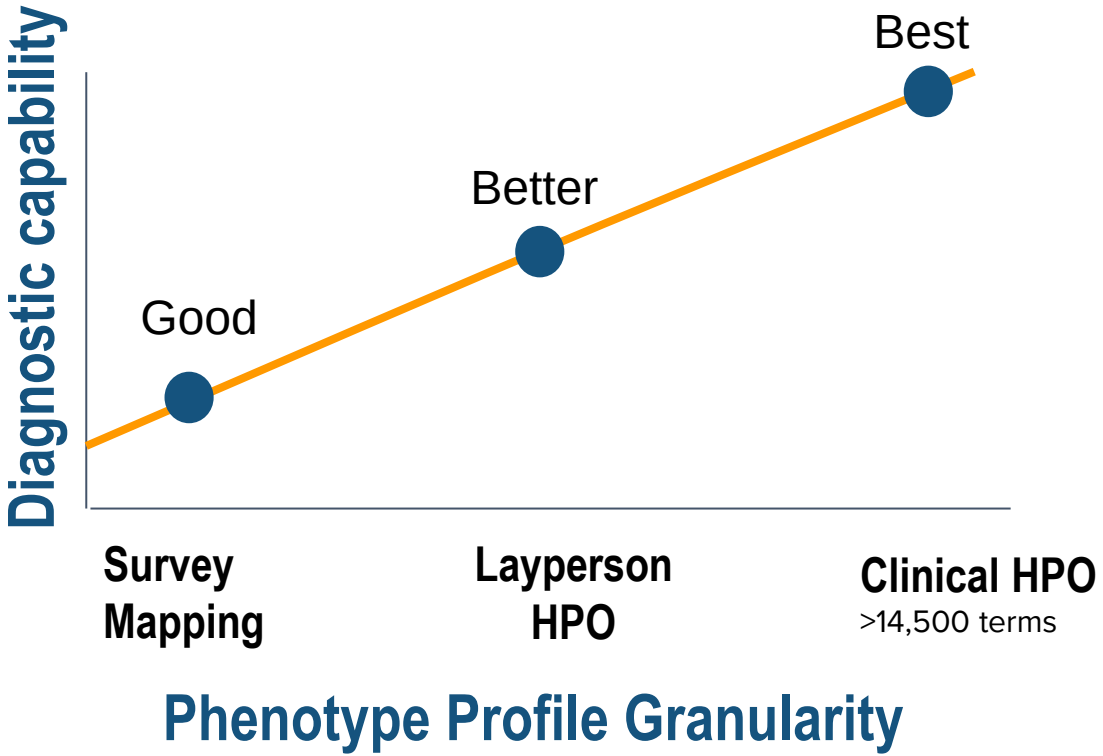
Alternative titles; symbols

HYPERPHENYLALANINEMIA, TETRAHYDROBIOPTERIN-DEFICIENT, DUE TO GTP CYCLOHYDROLASE I DEFICIENCY
GTP CYCLOHYDROLASE I DEFICIENCY

How diagnostically useful are the HPO terms generated by patients?

Actual mileage (for a given disease) may vary depending on its layperson-coverage of the corresponding phenotypes

60% of diseases are covered with HPO plain language subset terms



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

bit.ly/annual-reviews-ontologies

Total Costs of Rare Disease Vary By Disease and Health System

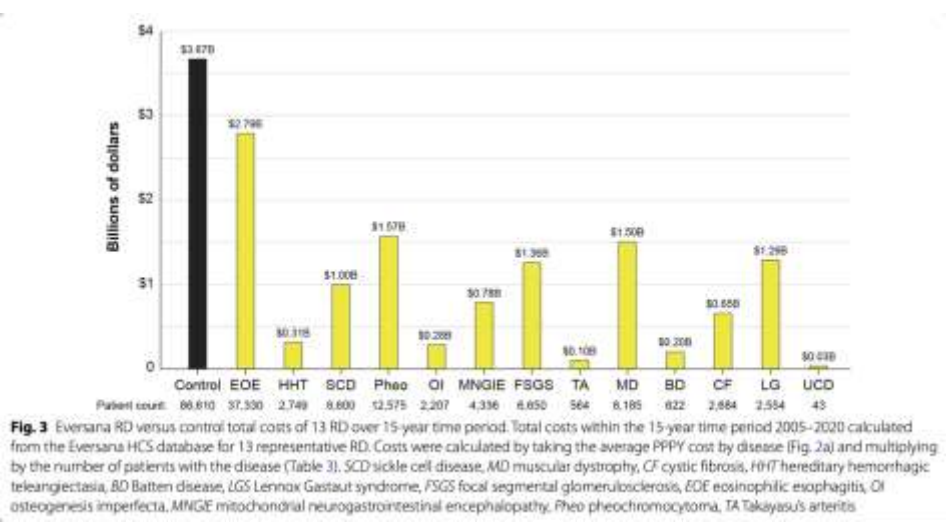


Fig. 3 Eversana RD versus control total costs of 13 RD over 15-year time period. Total costs within the 15-year time period 2005–2020 calculated from the Eversana HCS database for 13 representative RD. Costs were calculated by taking the average PPPY cost by disease (Fig. 2a) and multiplying by the number of patients with the disease (Table 3). *SCD* sickle cell disease, *MD* muscular dystrophy, *CF* cystic fibrosis, *HHT* hereditary hemorrhagic teleangiectasia, *BD* Batten disease, *LGS* Lennox Gastaut syndrome, *FSGS* focal segmental glomerulosclerosis, *EOE* eosinophilic esophagitis, *OI* osteogenesis imperfecta, *MNGIE* mitochondrial neurogastrointestinal encephalopathy, *Pheo* pheochromocytoma, *TA* Takayasu's arteritis

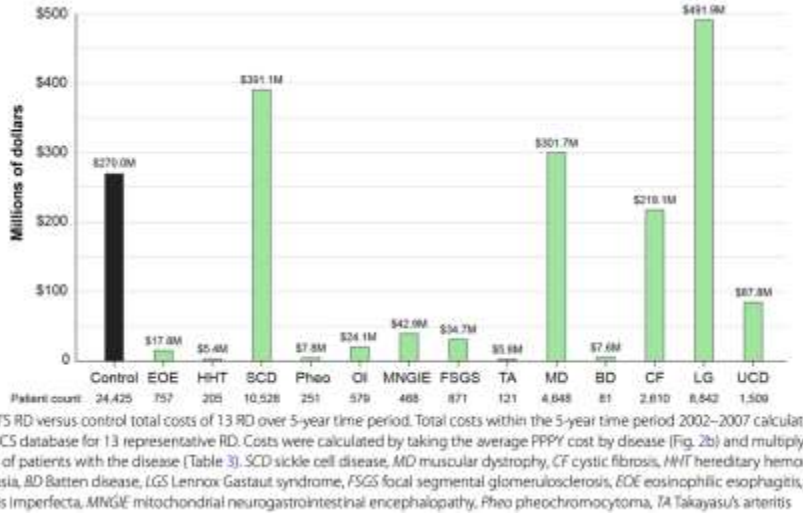


Fig. 4 NCATS RD versus control total costs of 13 RD over 5-year time period. Total costs within the 5-year time period 2002–2007 calculated from the NCATSHCS database for 13 representative RD. Costs were calculated by taking the average PPPY cost by disease (Fig. 2b) and multiplying by the number of patients with the disease (Table 3). *SCD* sickle cell disease, *MD* muscular dystrophy, *CF* cystic fibrosis, *HHT* hereditary hemorrhagic teleangiectasia, *BD* Batten disease, *LGS* Lennox Gastaut syndrome, *FSGS* focal segmental glomerulosclerosis, *EOE* eosinophilic esophagitis, *OI* osteogenesis imperfecta, *MNGIE* mitochondrial neurogastrointestinal encephalopathy, *Pheo* pheochromocytoma, *TA* Takayasu's arteritis

Total costs incurred at a organization are dependent on the size of the patient population.
While total costs for an individual rare disease cohort may not exceed an age-matched control cohort, collectively the rare disease cohort costs are greater than controls.

Rare Diseases Patients Have Increased Per-Patient Medical Costs

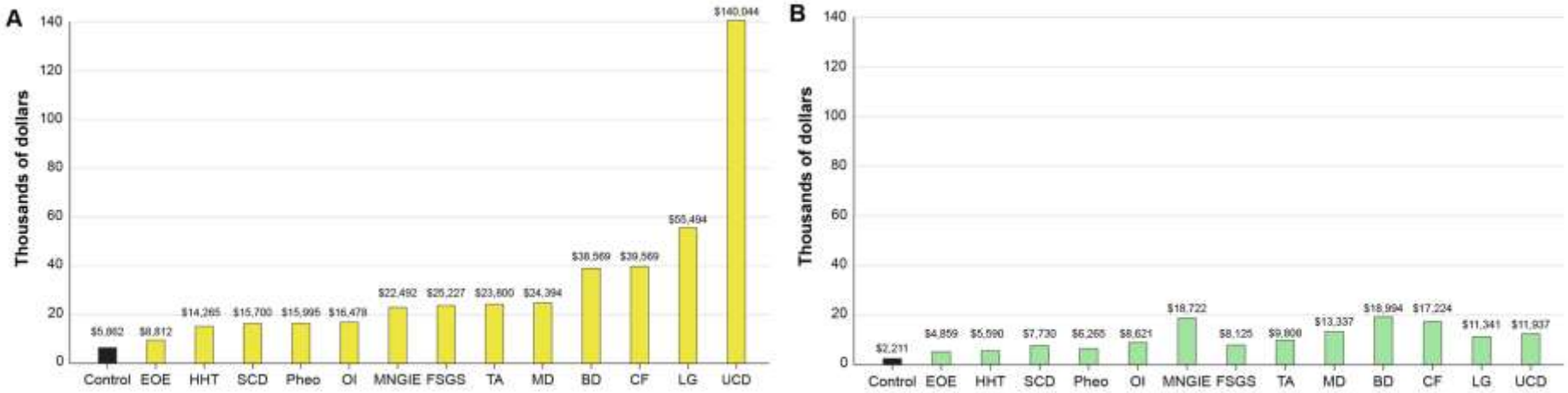
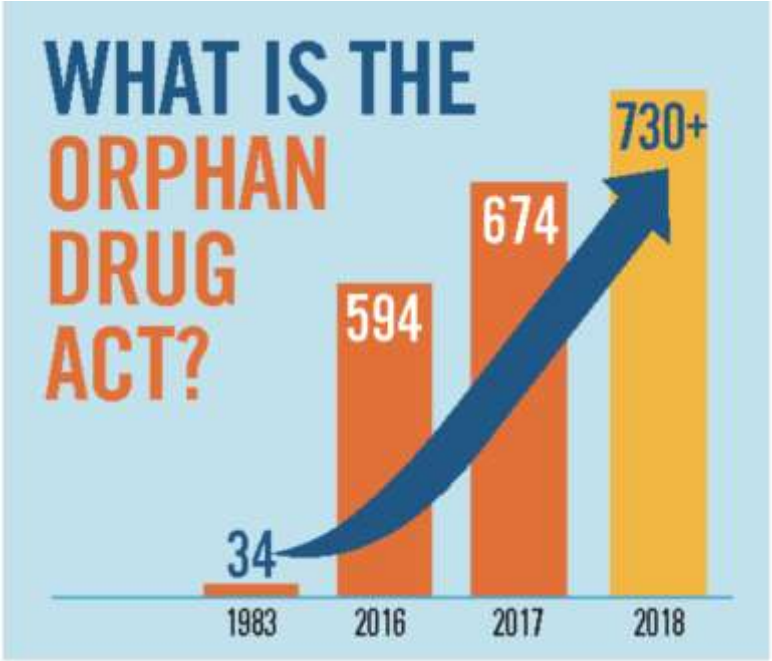


Fig. 2 PPPY cost of care of 13 RD versus control. Average per patient per year costs calculated within 2 different healthcare systems databases **A** Eversana and **B** NCATS, versus an age-matched control. *SCD* sickle cell disease, *MD* muscular dystrophy, *CF* cystic fibrosis, *HHT* hereditary hemorrhagic teleangiectasia, *BD* Batten disease, *LGS* Lennox Gastaut syndrome, *FSGS* focal segmental glomerulosclerosis, *EOE* eosinophilic esophagitis, *OI* osteogenesis imperfecta, *MNGIE* mitochondrial neurogastrointestinal encephalopathy, *Pheo* pheochromocytoma, *TA* Takayasu's arteritis

Rare disease patients have increased healthcare utilizations compared to control patients.
While costs are not available in N3C,
the healthcare utilizations (encounters, medications, procedures, etc.) *are* available.

Orphan Drug Act of 1983



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

https://globalgenes.org/rare-facts/

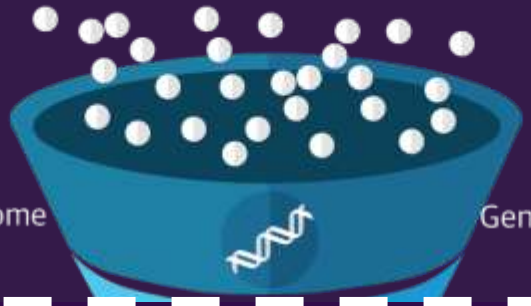
Rare Diseases collectively are an unrecognized public health crisis

Patient

Candidate Diseases

Population & Cohort

Exome/Genome



Genomic reference

Under-utilized data →
Loss of discriminatory power

Let's change this

Tisdale *et al. Orphanet J Rare Dis* (2021) 16:429
https://doi.org/10.1186/s13023-021-02061-3

Orphanet Journal of Rare Diseases

RESEARCH Open Access

The IDeaS initiative: pilot study to assess the impact of rare diseases on patients and healthcare systems

Ainslie Tisdale[1], Christine M. Cutillo[2], Ramaa Nathan[3], Pierantonio Russo[3,4], Bryan Laraway[5], Melissa Haendel[6], Douglas Nowak[7], Cindy Hasche[7], Chun-Hung Chan[8], Emily Griese[8,9], Hugh Dawkins[10], Oodaye Shukla[3], David A. Pearce[7,8,11], Joni L. Rutter[2] and Anne R. Pariser[1*]

NIH National Center for Advancing Translational Sciences
EVERSANA
SANFORD HEALTH
OHSU