THE APPLICATION OF THE HUMAN PHENOTYPE ONTOLOGY

II International Summer School
RARE DISEASE AND ORPHAN DRUG REGISTRIES

Melissa Haendel

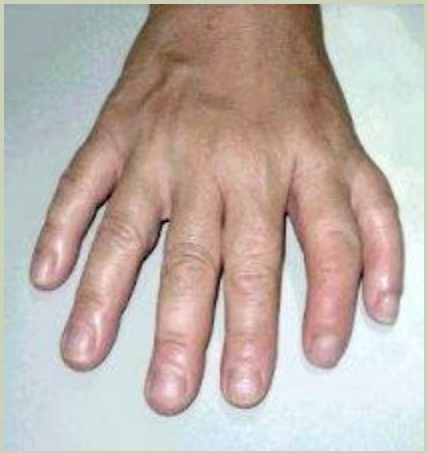
Sept 19th, 2014

OUTLINE

- Why phenotyping is hard
- About Ontologies
- Diagnosing known diseases
- Getting the phenotype data
- How much phenotyping is enough?
- Model organism data for undiagnosed diseases

PHENOTYPIC BEGINNINGS

http://rundangerously.blogspot.com/2010/06/inconvenient-summer-head-cold.html



http://www.vetnext.com/images

http://commons.wikimedia.org/wiki/File:CleftLip1.png

Phenotyping SEEMS like a simple task, but there are shades of grey and nuances that are difficult to convey.

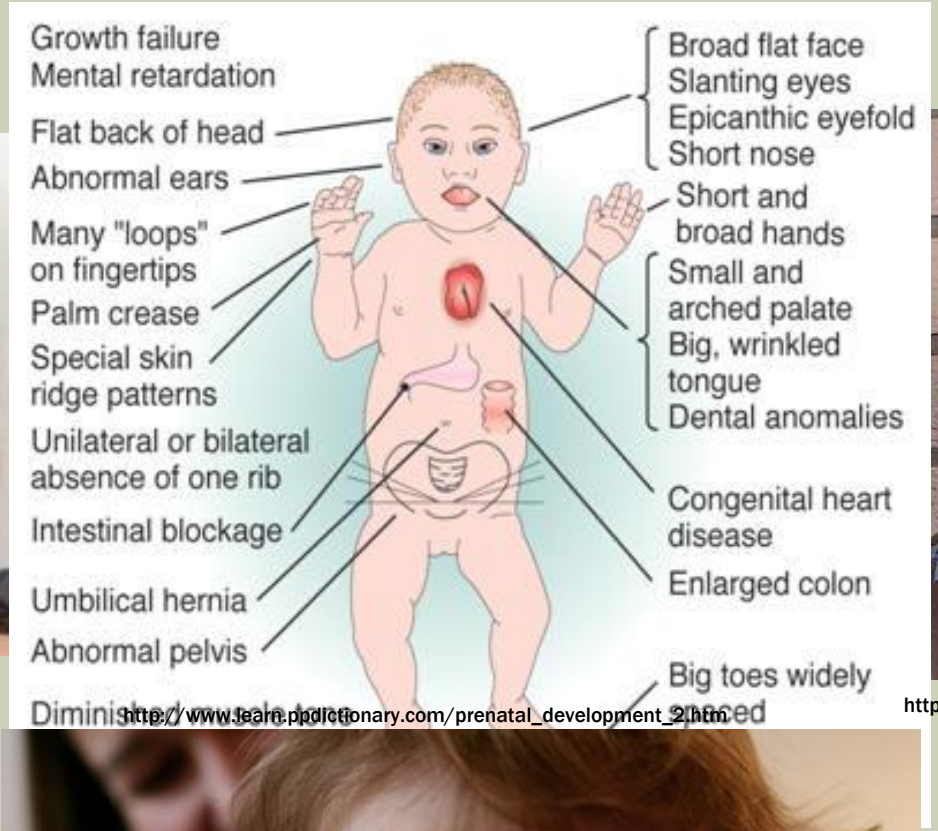
http://www.pyroenergen.com/articles07/downs-syndrome.htm

http://anthro.palomar.edu/abnormal/abnormal_4.htm

http://www.theguardian.com/commentisfree/2009/oct/27/downs-syndrome-increase-terminations

THE CONSTELLATION OF PHENOTYPES SIGNIFIES THE DISEASE – A 'PROFILE'

http://www.pyroenergen.com/articles07/downs-syndrome.htm



http://www.learn.ppdictionary.com/prenatal_development_2.htm

http://anthro.palomar.edu/abnormal/abnormal_4.htm

http://www.theguardian.com/commentisfree/2009/oct/27/downs-syndrome-increase-terminations

CLINICAL PHENOTYPING

Often free text or checkboxes

Dysmorphic features
- df
- dysmorphic
- dysmorphic faces
- dysmorphic features

Congenital malformation/anomaly:
- congenital anomaly
- congenital malformation
- congenital anamoly
- congenital anomly
- congential anomaly
- congentital anomaly
- cong. m.
- cong. Mal
- cong. malfor
- congenital malform
- congenital m.
- multiple congenital anomalies
- multiple congenital abormalities
- multiple congenital abnormalities

Examples of lists:
* dd. cong. malfor. behav. pro.
* dd. mental retardation
* df< delayed puberty
* df&lt
* dd df mr
* mental retar.short stature

Phenotypic description (Clinical symptoms)

Behavior, Cognition and Development
- □ Global development delay
- □ Fine motor delay □ Gross motor delay
- □ Language delay
- □ Learning disability
- □ Mental retardation
  - □ Mild
  - □ Moderate
  - □ Severe
- □ Attention deficit hyperactivity disorder
- □ Autism
- □ Pervasive developmental delay
- □ Psychiatric disorders (Specify below)
- □ Other: _____

Neurological
- □ Hypotonia
- □ Seizures
- □ Ataxia
- □ Dystonia
- □ Chorea

Cardiac
- □ ASD
- □ VSD
- □ AV canal defect
- □ Coarctation of aorta
- □ Tetralogy of fallot
- □ Other: _____

Craniofacial
- □ Craniosynostosis
- □ Cleft lip □ Cleft palate
- □ Microretrognathia □ Retrognathia
- □ Facial dysmorphism (Specify below)
- □ Other: _____

Eye Defects
- □ Blindness
- □ Coloboma
- □ Epicanthus
- □ Eyelid abnormality (Specify bellow)
- □ Other: _____

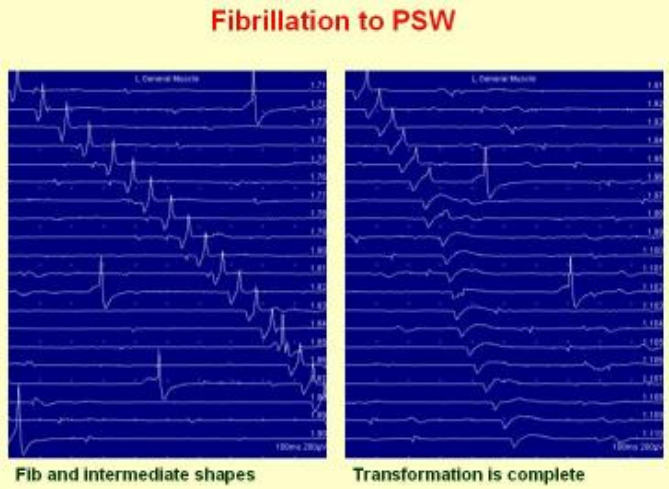
SEARCHING FOR PHENOTYPES USING TEXT ALONE IS INSUFFICIENT

OMIM Query	# Records
“large bone”	785
“enlarged bone”	156
“big bone”	16
“huge bones”	4
“massive bones”	28
“hyperplastic bones”	12
“hyperplastic bone”	40
“bone hyperplasia”	134
“increased bone growth”	612

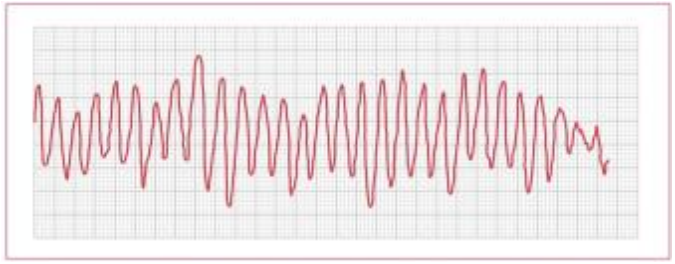
OMIM
Online Mendelian Inheritance in Man

Johns Hopkins University

TERMS SHOULD BE WELL DEFINED SO THEY GET USED PROPERLY



fibrillation . . .



fibrillation . . .

muscle fibrillation $=$ fibrillation $\neq$ fibrillation $=$ ventricular fibrillation

We need to capture synonyms and use unique labels

SO WHAT IS THE PROBLEM?

- Obviously similar phenotype descriptions mean the same thing to you, but not to a computer:
  - generalized amyotrophy
  - generalized muscle, atrophy
  - muscular atrophy, generalized

- Many publications have little information about the actual phenotypic features seen in patients with particular mutations

- Databases cannot talk to one another about phenotypes

OUTLINE

- Why phenotyping is hard
- About Ontologies
- Diagnosing known diseases
- Getting the phenotype data
- How much phenotyping is enough?
- Model organism data for undiagnosed diseases

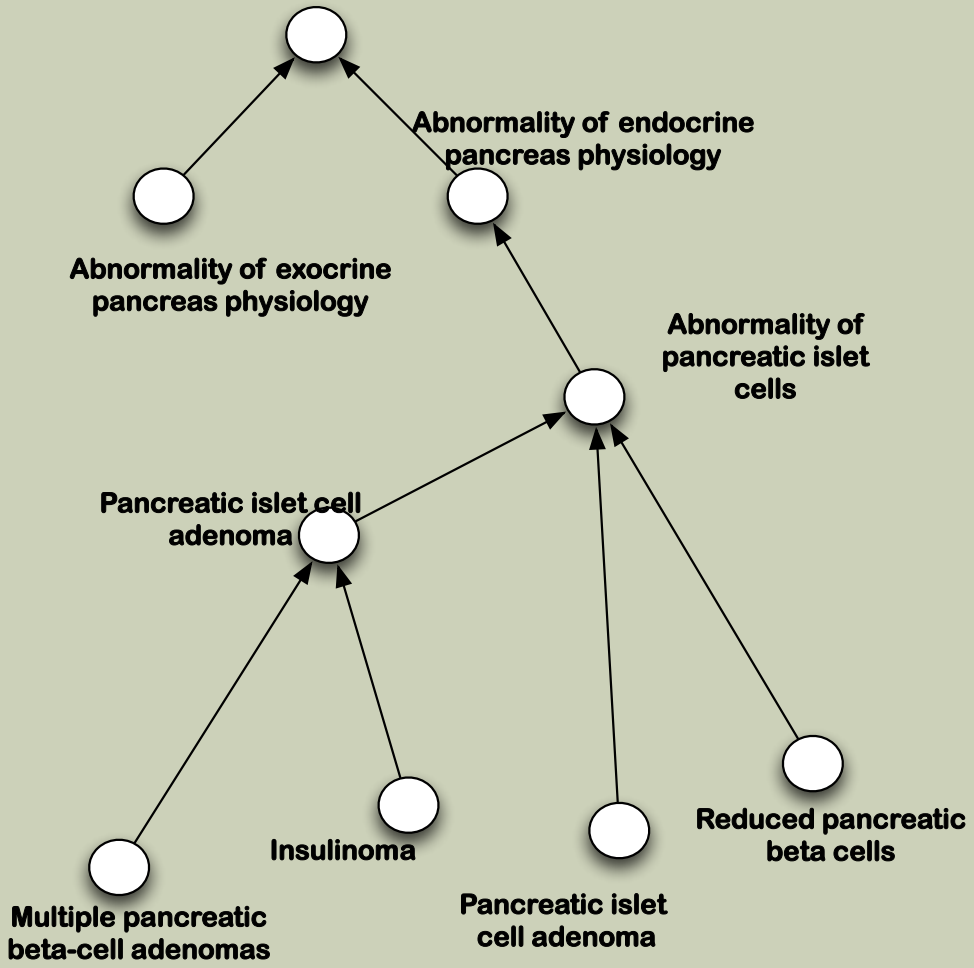
ONTOLOGIES CAN HELP.

A controlled vocabulary of logically defined, inter-related terms used to annotate data

- Use of common or logically related terms across databases enables integration
- Relationships between terms allow annotations to be grouped in scientifically meaningful ways
- Reasoning software enables computation of inferred knowledge
- Some well known ontologies are SNOMED-CT, Foundational Model of Anatomy, Gene Ontology, Linnean Taxonomy of species

OTHER COMMON USES OF ONTOLOGIES

HUMAN PHENOTYPE ONTOLOGY



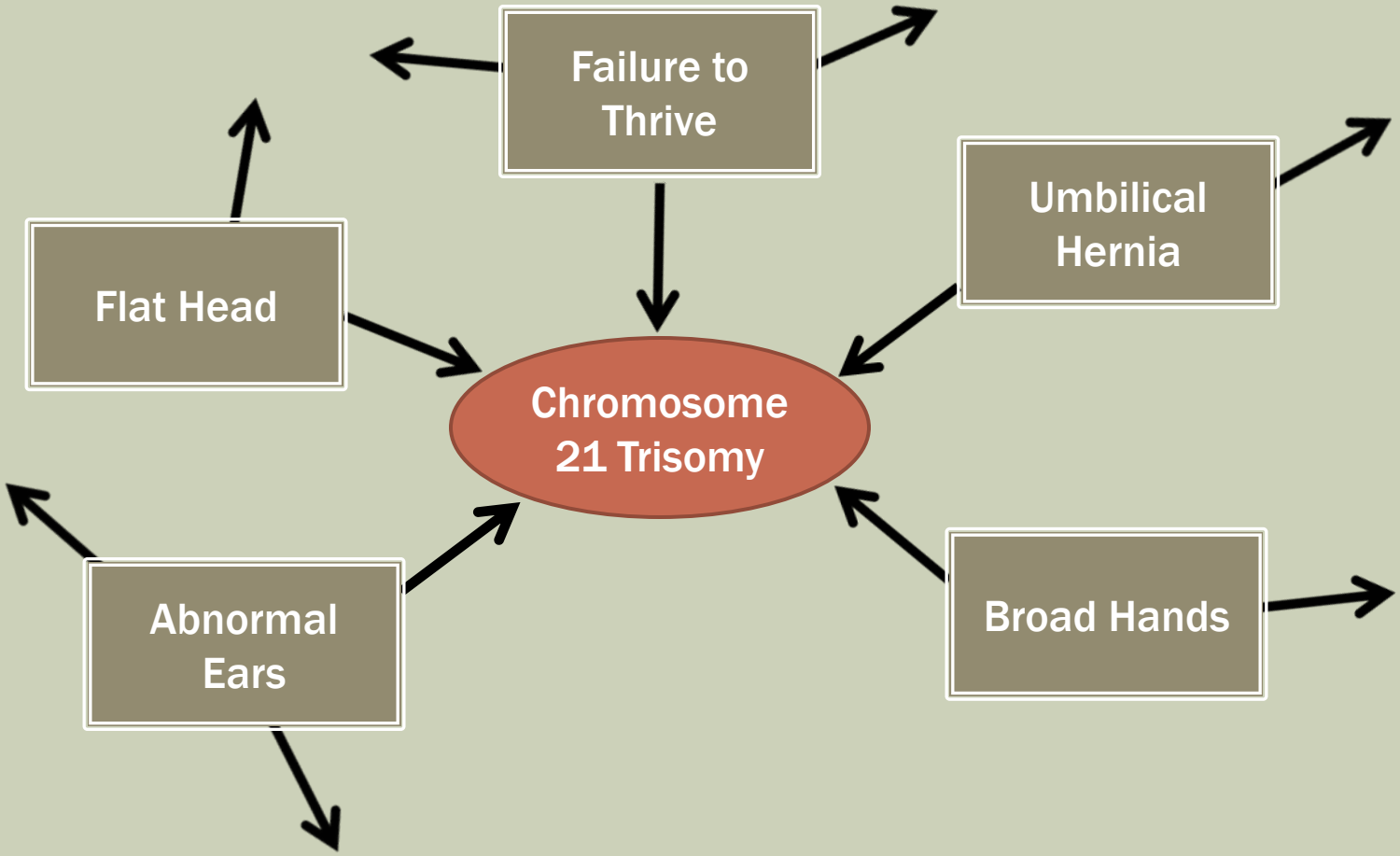
Used to annotate:

- **Patients**
- **Disorders**
- **Genotypes**
- **Genes**
- **Sequence variants**

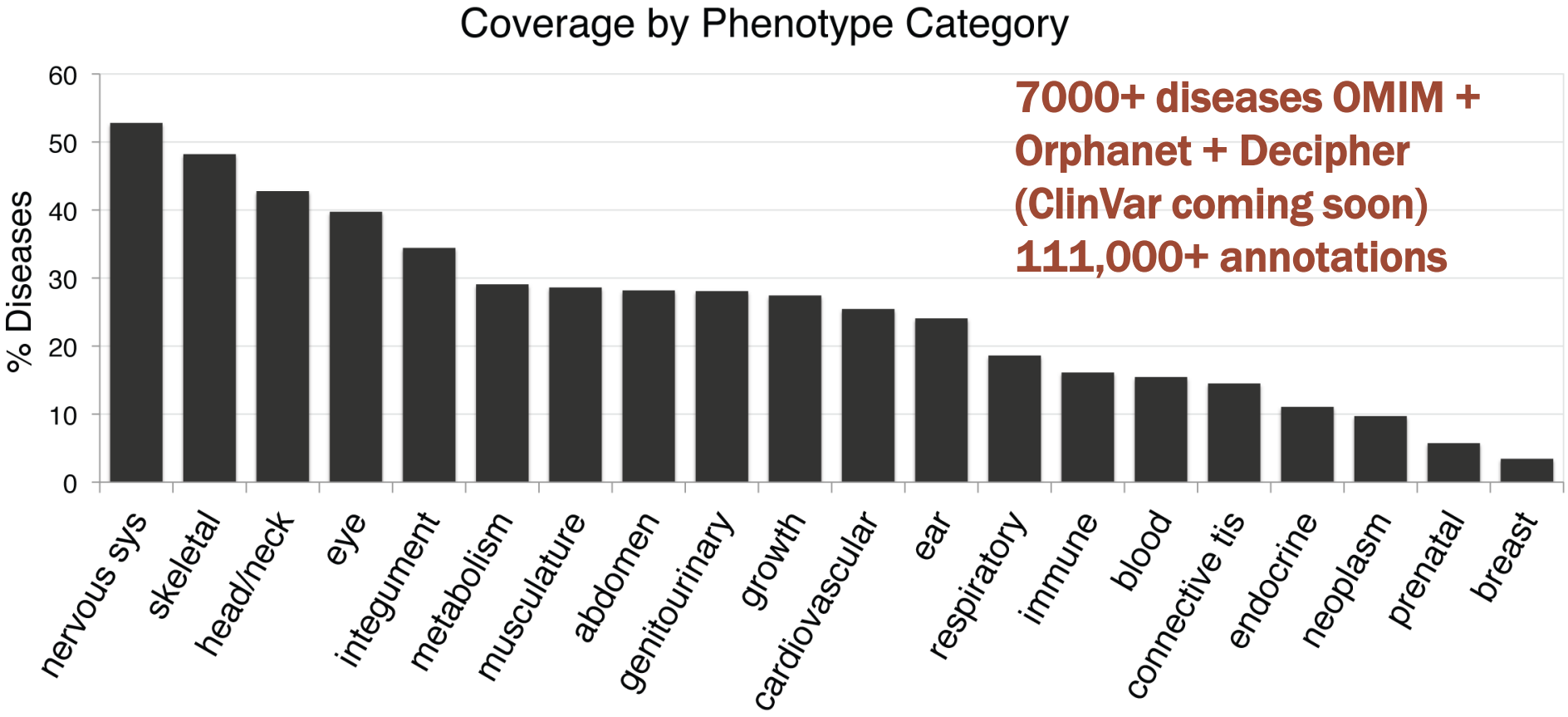
Mappings to SNOMED-CT, UMLS, MeSH, ICD, etc.

Köhler et al. **The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data.** Nucleic Acids Res. 2014 Jan 1;42(1):D966-74.

USING A CONTROLLED VOCABULARY TO LINK PHENOTYPES TO DISEASES

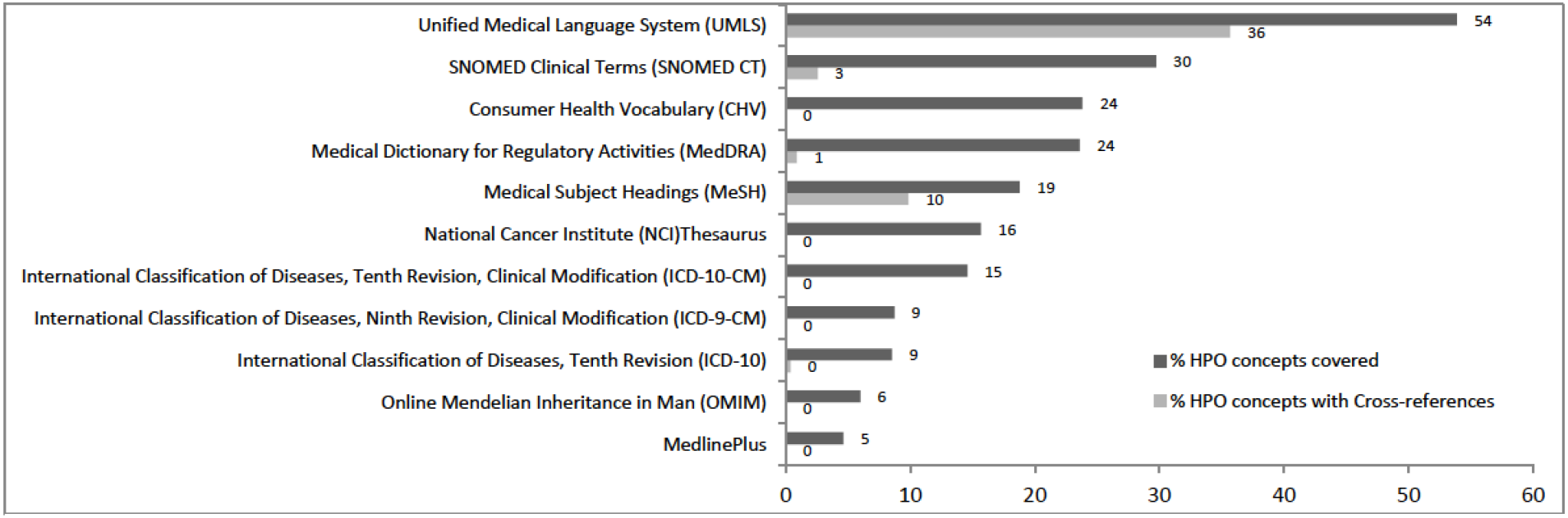


SURVEY OF ANNOTATIONS IN DISEASE CORPUS



Phenotype annotations are unevenly distributed across different anatomical systems

HOW DOES HPO RELATE TO OTHER CLINICAL VOCABULARIES?

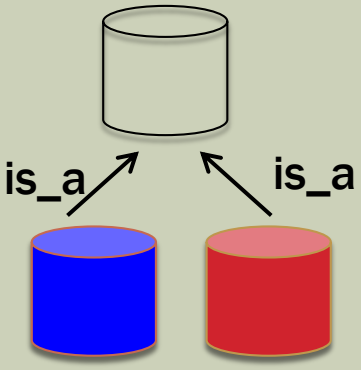


Winnenburg and Bodenreider, ISMB PhenoDay, 2014

LOGICAL TERM DEFINITION

Definitions are of the following Genus-Differentia form:

X = a Y which has one or more differentiating characteristics.
where X is the is_a parent of Y.



Definition of a cylinder:
Surface formed by the set of lines perpendicular to a plane, which pass through a given circle in that plane.

Definition: Blue cylinder = Cylinder that has color blue.

Definition: Red cylinder = Cylinder that has color red.

ABOUT REASONERS

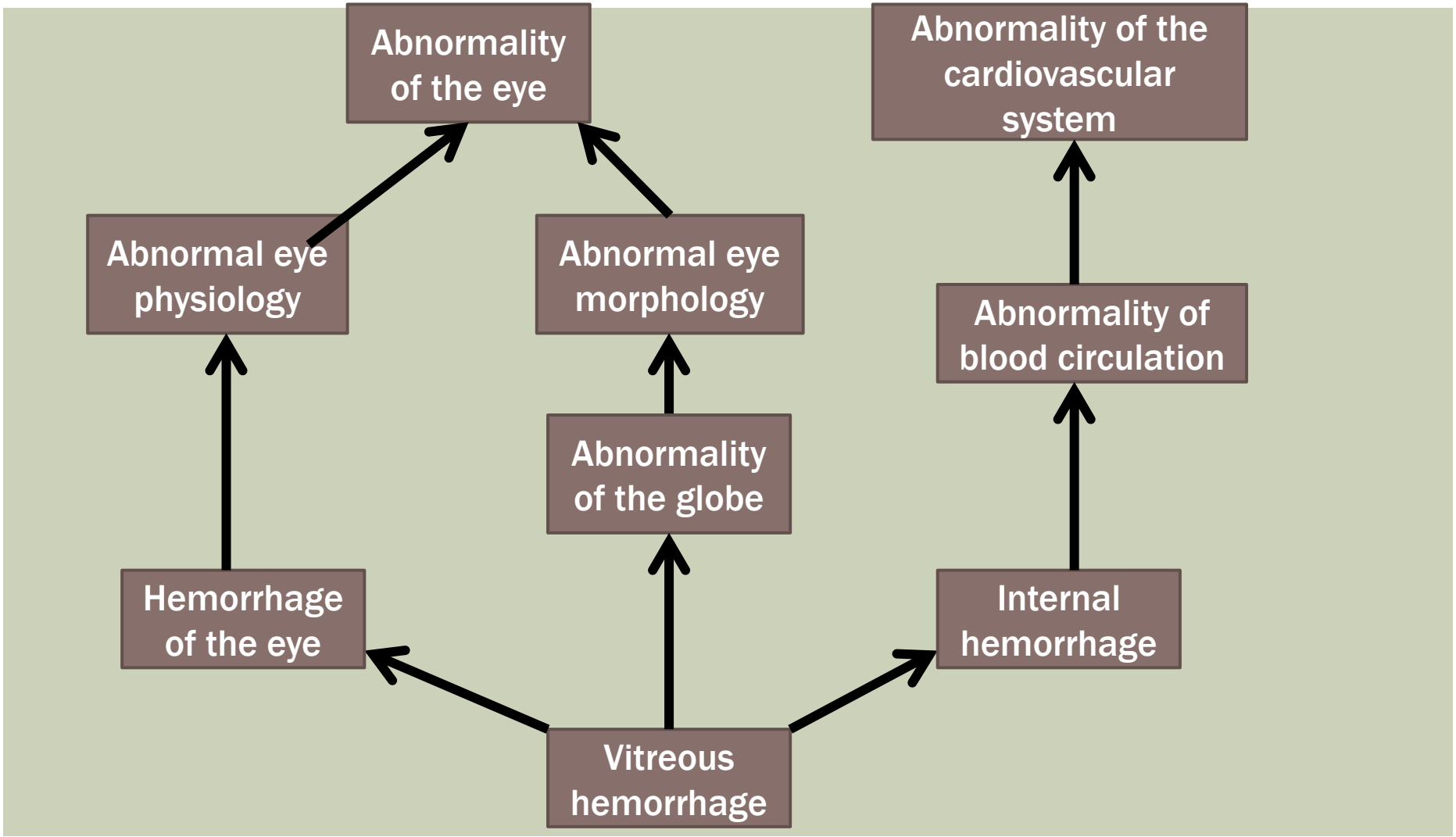
A piece of software able to infer logical consequences from a set of asserted facts or axioms.

They are used to check the logical consistency of the ontologies and to extend the ontologies with "inferred" facts or axioms

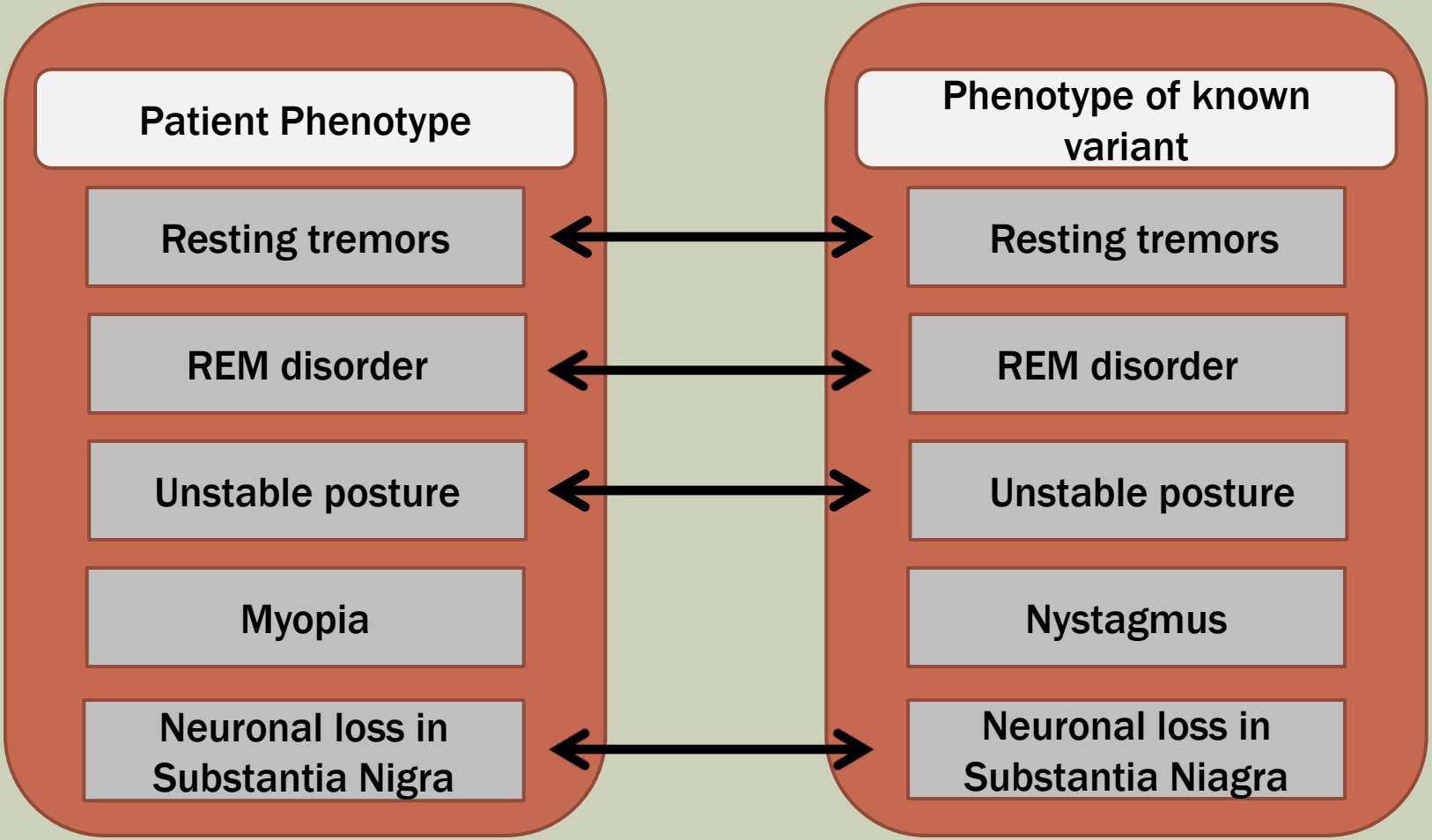
For example, a reasoner would infer:

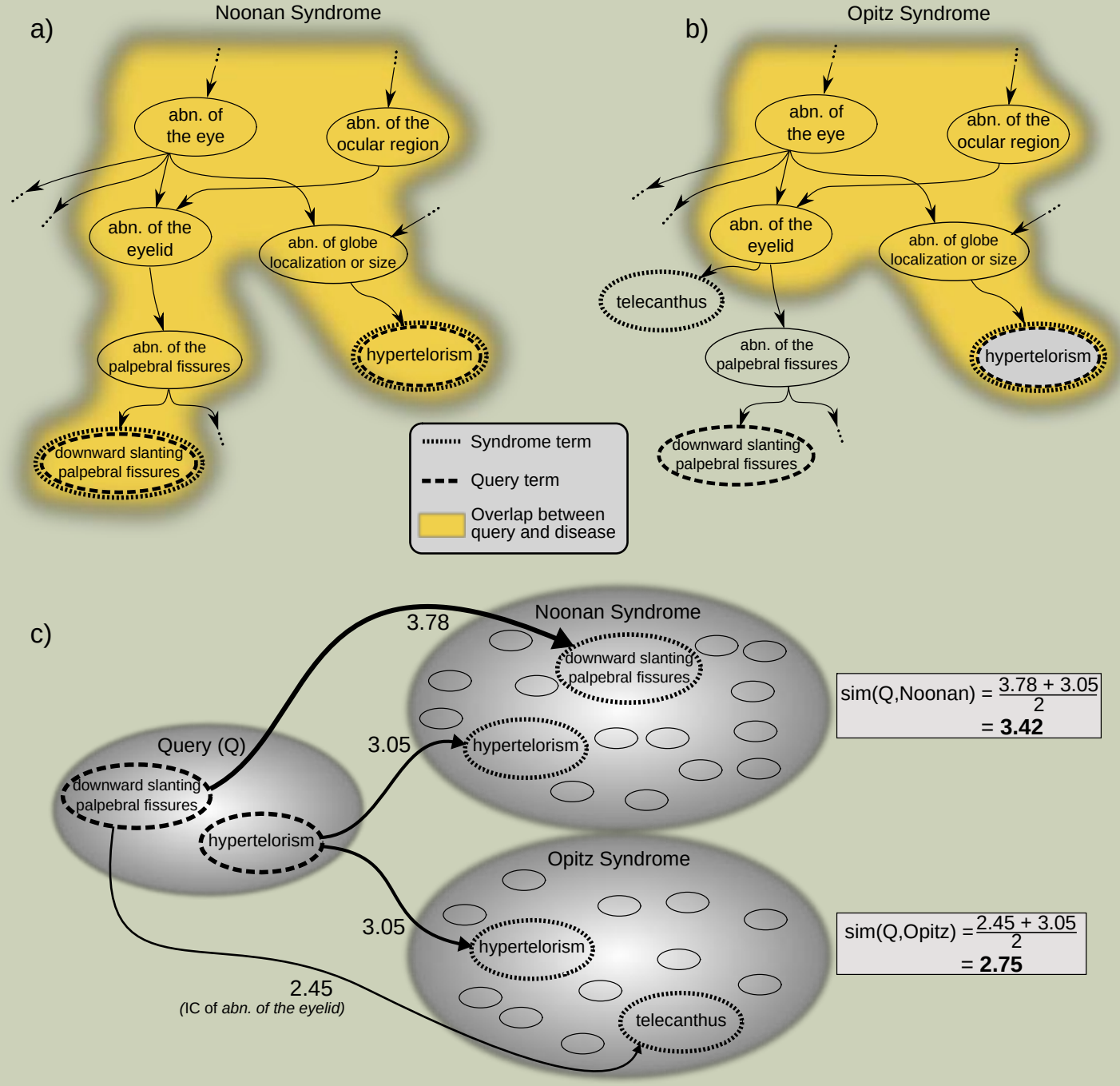
Major premise: All mortals die.
Minor premise: Some men are mortals.
Conclusion: Some men die.

PHENOTYPES CAN BE CLASSIFIED IN MULTIPLE WAYS

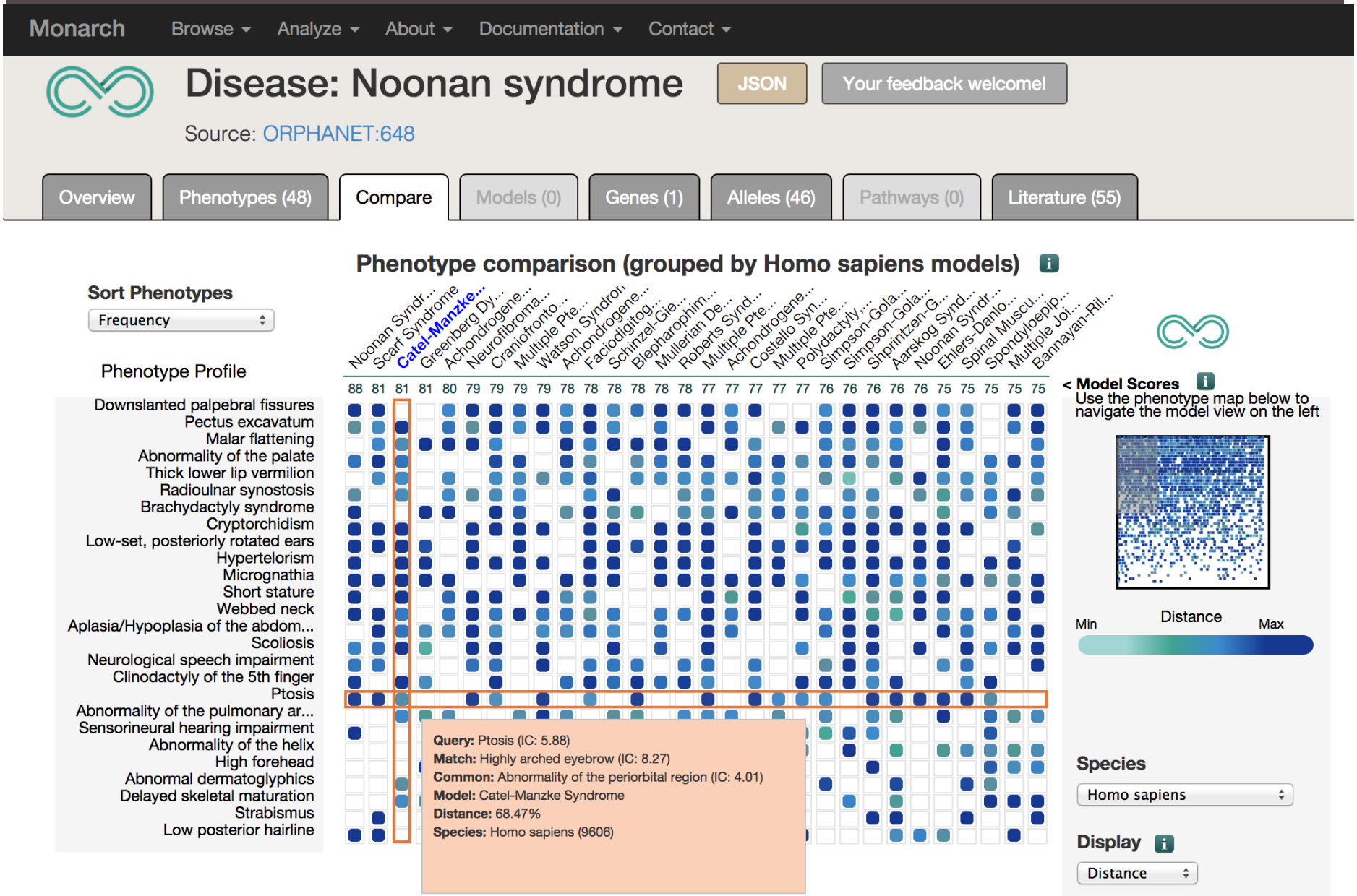


PHENOTYPE MATCHING




a)
Noonan Syndrome
abn. of the eye
abn. of the ocular region
abn. of the eyelid
abn. of globe localization or size
abn. of the palpebral fissures
hypertelorism
downward slanting palpebral fissures
b)
Opitz Syndrome
abn. of the eye
abn. of the ocular region
abn. of the eyelid
abn. of globe localization or size
telecanthus
abn. of the palpebral fissures
hypertelorism
downward slanting palpebral fissures
Syndrome term
Query term
Overlap between query and disease
c)
Noonan Syndrome
downward slanting palpebral fissures
hypertelorism
3.78
3.05
Query (Q)
downward slanting palpebral fissures
hypertelorism
sim(Q,Noonan) = (3.78 + 3.05) / 2 = 3.42
Opitz Syndrome
3.05
hypertelorism
2.45
(IC of abn. of the eyelid)
telecanthus
sim(Q,Opitz) = (2.45 + 3.05) / 2 = 2.75

COMPARATIVE VISUALIZATONS



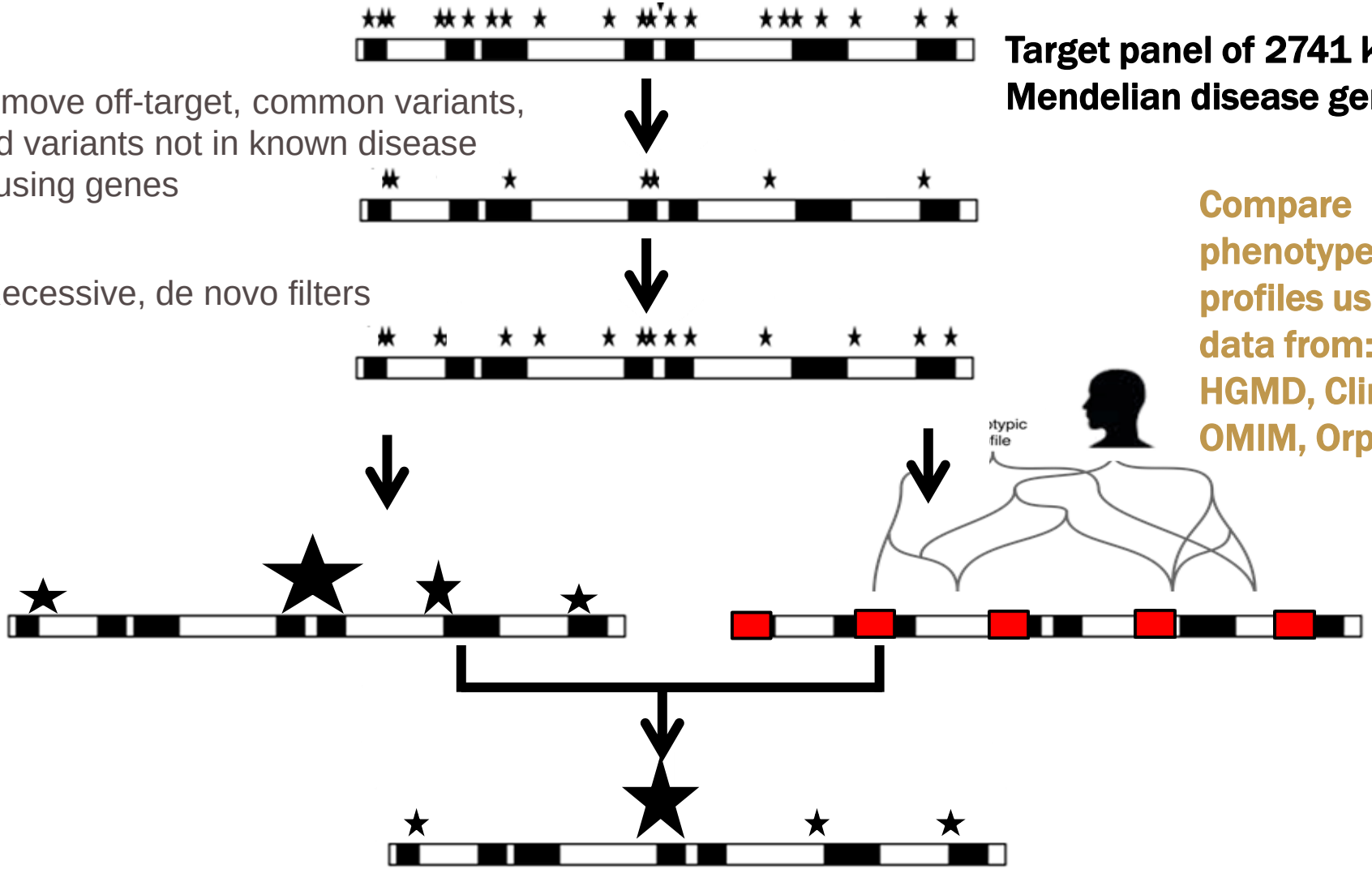
OUTLINE

- Why phenotyping is hard
- About Ontologies
- **Diagnosing known diseases**
- Getting the phenotype data
- How much phenotyping is enough?
- Model organism data for undiagnosed diseases

THE YET-TO-BE DIAGNOSED PATIENT

- Known disorders not recognized during prior evaluations?
- Atypical presentation of known disorders?
- Combinations of several disorders?
- Novel, unreported disorder?

PhenIX EXOME ANALYSIS



Target panel of 2741 known Mendelian disease genes

Remove off-target, common variants, and variants not in known disease causing genes

Recessive, de novo filters

Compare phenotype profiles using data from: HGMD, Clinvar, OMIM, Orphanet

http://compbio.charite.de/PhenIX/

Zemojtel et al. Sci Transl Med 3 September 2014: Vol. 6, Issue 252, p.252ra123

PHENIX PERFORMANCE TESTING

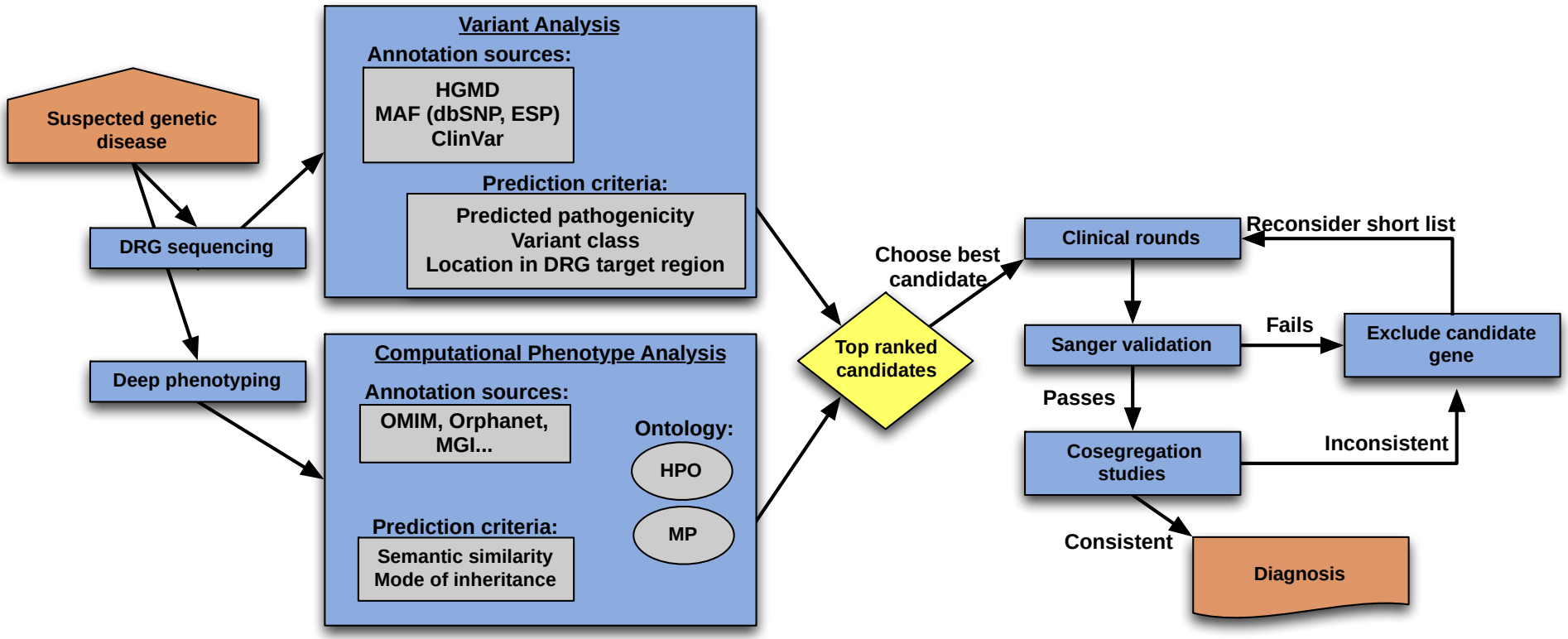
Figure removed due to restrictions. Please see the paper:
http://stm.sciencemag.org/content/6/252/252ra123.full

Simulated datasets created by spiking DAG panel generated VCF file with the causative mutation removed

CONTROL PATIENTS WITH KNOWN MUTATIONS

Inheritance	Gene	Average Rank
AD	ACVR1, ATL1, BRCA1, BRCA2, CHD7 (4), CLCN7, COL1A1, COL2A1, EXT1, FGFR2 (2), FGFR3, GDF5, KCNQ1, MLH1 (2), MLL2/KMT2D, MSH2, MSH6, MYBPC3, NF1 (6), P63, PTCH1, PTH1R (2), PTPN11 (2), SCN1A, SOS1, TRPS1, TSC1, WNT10A	1.7
AR	ATM, ATP6V0A2, CLCN1 (2), LRP5, PYCR1, SLC39A4	5
X	EFNB1, MECP2 (2), DMD, PHF6	1.8

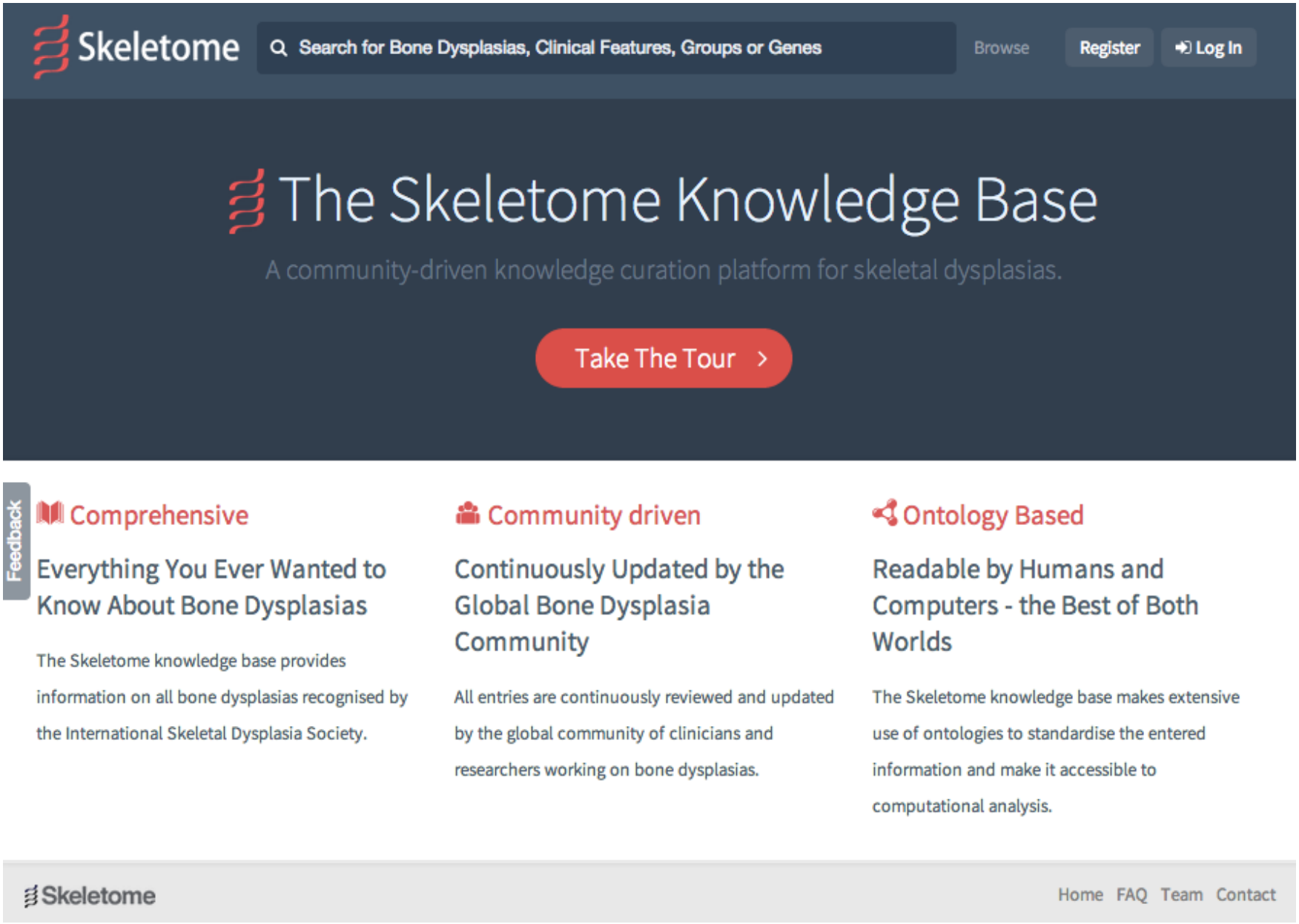
WORKFLOW FOR CLINICAL EXOME ANALYSIS



PHENIX HELPED DIAGNOSE 11/40 PATIENTS

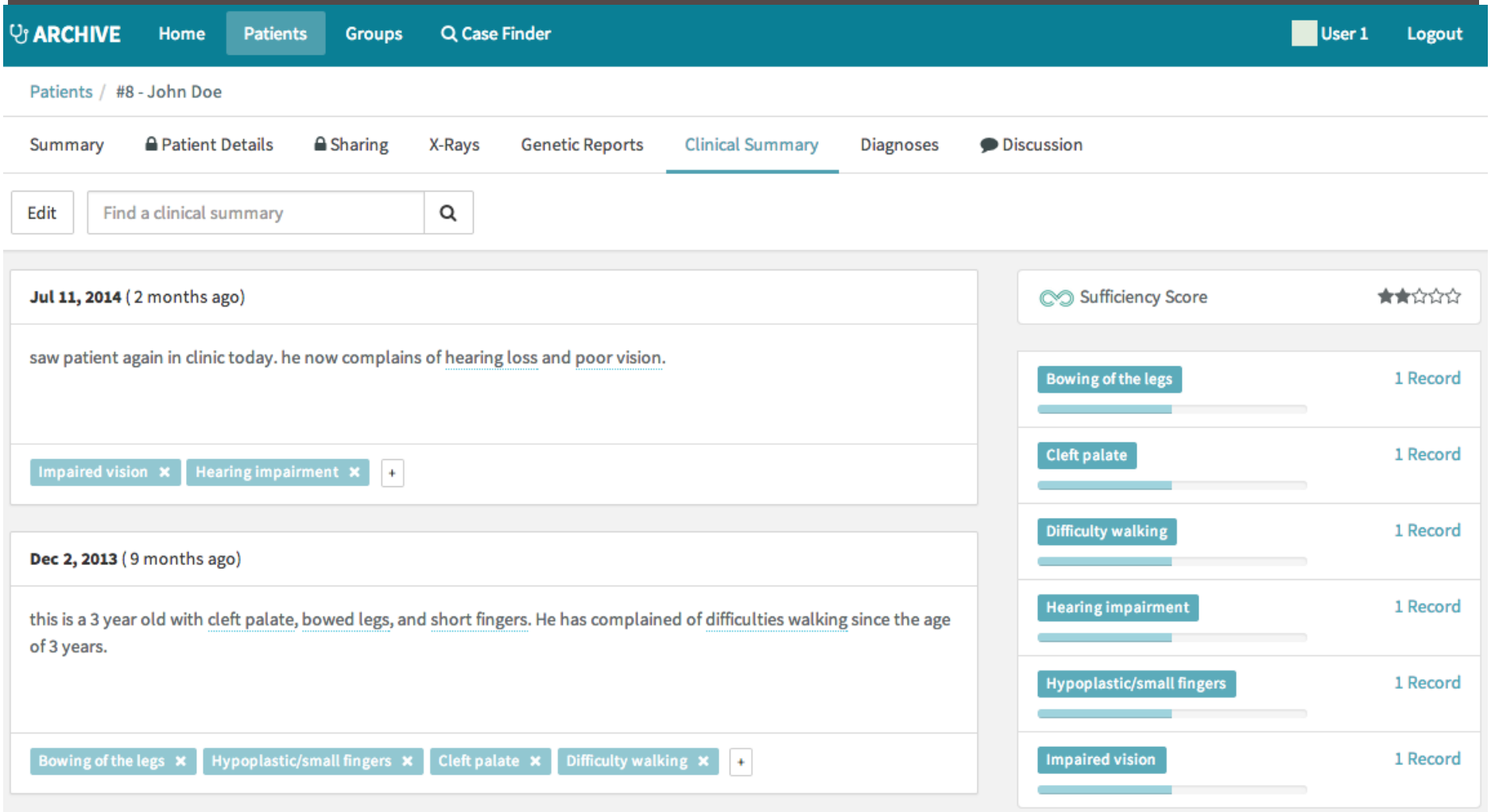
ID	Age, Sex	Presentation	Gene	Rank	Diagnosis
P1	3y (f)	Intellectual disability + multiple congenital anomalies	*MLL*	2	Wiedemann-Steiner syndrome (*39*)
P2	5y (f)				
P3	6y (f)				
P4	Death (f)				
P5	6m (f)				
P6	Fetus (m) Death gestat				
P7	7y (m)				
P8	14y (r				
P9	6y (f)				
P10	4 betwe and 7y				
P11	3y (m				

global developmental delay (HP:0001263)
delayed speech and language development (HP:0000750)
motor delay (HP:0001270)
proportionate short stature (HP:0003508)
microcephaly (HP:0000252)
feeding difficulties (HP:0011968)
congenital megaloureter (HP:0008676)
cone-shaped epiphysis of the phalanges of the hand (HP:0010230)
sacral dimple (HP:0000960)
hyperpigmentated/hypopigmentated macules (HP:0007441)
hypertelorism (HP:0000316)
abnormality of the midface (HP:0000309)
flat nose (HP:0000457)
thick lower lip vermilion (HP:0000179)
thick upper lip vermilion (HP:0000215)
full cheeks (HP:0000293)
short neck (HP:0000470)



University of Queensland, University of Sydney

SKELETOME PATIENT ARCHIVE



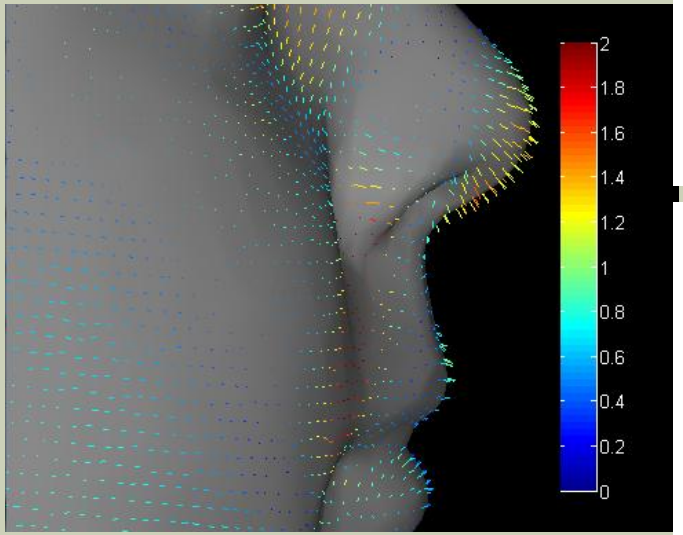
- Integration with the HPO, Orphanet, and Monarch Initiative
- Automated phenotyping from clinical summaries
- Collaborative diagnosis

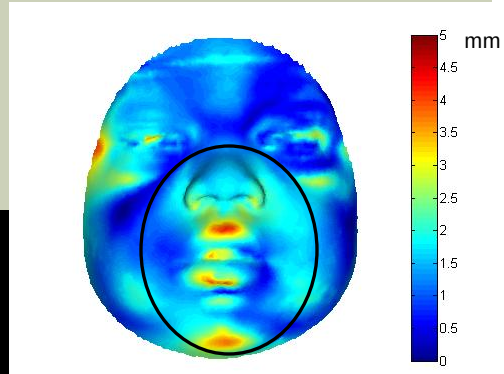
University of Queensland, University of Sydney

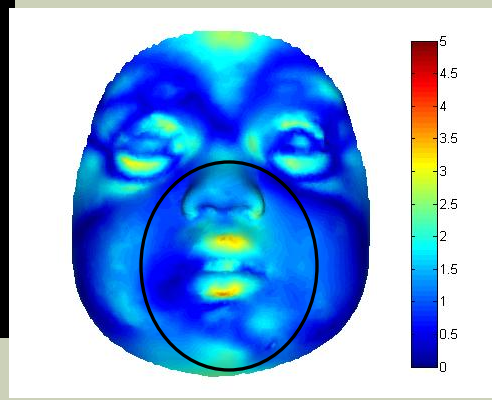
THE FACES OF RARE DISEASES

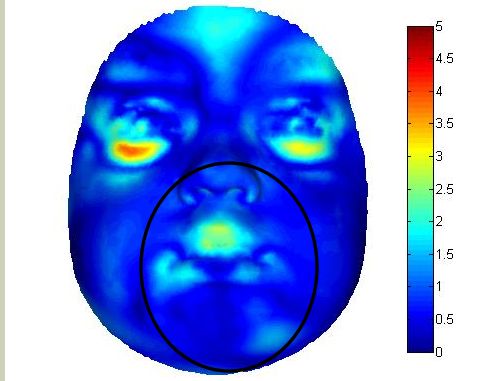
Non-invasive, non-irradiating deeply precise 3D facial analysis

- Screening and diagnosis
- Treatment monitoring
- Surgical planning and audit
- Genotype-phenotype correlation
- Cross-species comparisons
- Face to text conversion for text mining









Facial Characteristic	*ID*	*Explanation*	*Technical Formulation*
Malar flattening	A	Change in surface curvature at the cheekbones	$\lvert C^{B}_{Cheekbones} - C^{A}_{Cheekbones} \rvert$
	B	Change in variation or dispersion in position (standard deviation) along the antero-posterior (sagittal) axis of the cheekbones	$\lvert \sigma^{B}_{Cheekbones} - \sigma^{A}_{Cheekbones} \rvert$
	C	Inward/outward movement of the medial part of the midface, in relation to the lateral part of the lower face	$\lvert N_{MidFaceMedial} - N_{LowerFaceLateral} \rvert$
Square/round face	A	Change in similarity between facial contours and a square (**Error! Reference source not found.**)	*Take a border projection of each shape transformation and compare it with a fitted square:* $\left\lvert \sqrt{\sum_{V \in border\ B} \left(\min_{V_i \in square\ B} \left(\Delta(P_i, P_j) \right) \right)^2} - \sqrt{\sum_{V \in border\ A} \left(\min_{V_i \in square\ A} \left(\Delta(P_i, P_j) \right) \right)^2} \right\rvert$ *The square fitted to each of the morphs is defined by placing the 2 vertical lines at the most lateral positions of the facial border (left and right) and the 2 horizontal lines at the most superior and inferior positions of the facial border.*
Micrognathia	A	Inward/outward movement of the chin	$\lvert N_{Chin} \rvert$
	B	Area increase/decrease of the chin	$\left\lvert -\log\left(\frac{S^{B}_{Chin}}{S^{A}_{Chin}} \right) \right\rvert$
	C	Change in distance between labiale inferius and gnathion, along the vertical (longitudinal) axis	$\lvert \Delta_y(LM^{B}_{Labiale\ Inferius}, LM^{B}_{Gnathion}) - \Delta_y(LM^{A}_{Labiale\ Inferius}, LM^{A}_{Gnathion}) \rvert$
Microcephaly	A	Change in size of the circle fitted to the surface of the forehead. The circle is placed in the transverse plane, halfway the glabella and the top of the face (**Error! Reference source not found.**)	*Take the intersection plane with the facial mesh (through point halfway between glabella and top of head, parallel to the XZ plane), fit a circle through the intersection points, and take radius as*

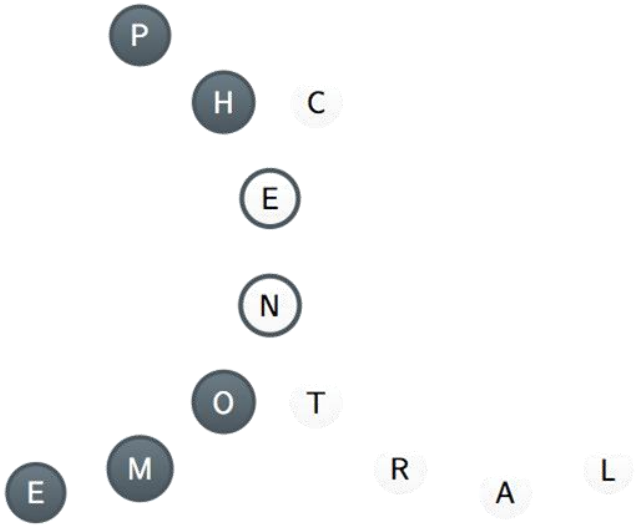
RD Connect

University of Western Australia

OUTLINE

- Why phenotyping is hard
- About Ontologies
- Diagnosing known diseases
- **Getting the phenotype data**
- How much phenotyping is enough?
- Model organism data for undiagnosed diseases

PHENOTIPS



PHENOTIPS & PHENOMECENTRAL

Founding consortia

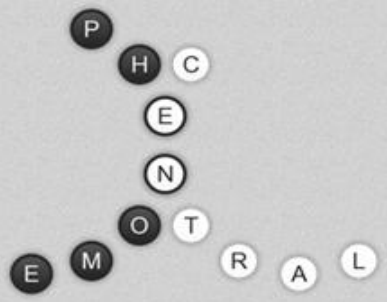
CARE forRARE

RD Connect

NIH Undiagnosed Diseases Program

USING ONTOLOGIES IN THE CLINIC

- Ontologies are large (HPO has > 10,000 terms) and difficult to navigate
- Mapping data to an ontology post-visit is time consuming and prone to error
- Best time to phenotype using ontologies is during the patient visit
- Goals of PhenoTips
  - Make deep phenotyping simple
  - Make it “faster than paper”



PhenomeCentral

An integrated portal for sharing and searching patient phenotype data for rare genetic disorders.

Sign up Login

PhenomeCentral is a Matchmaker

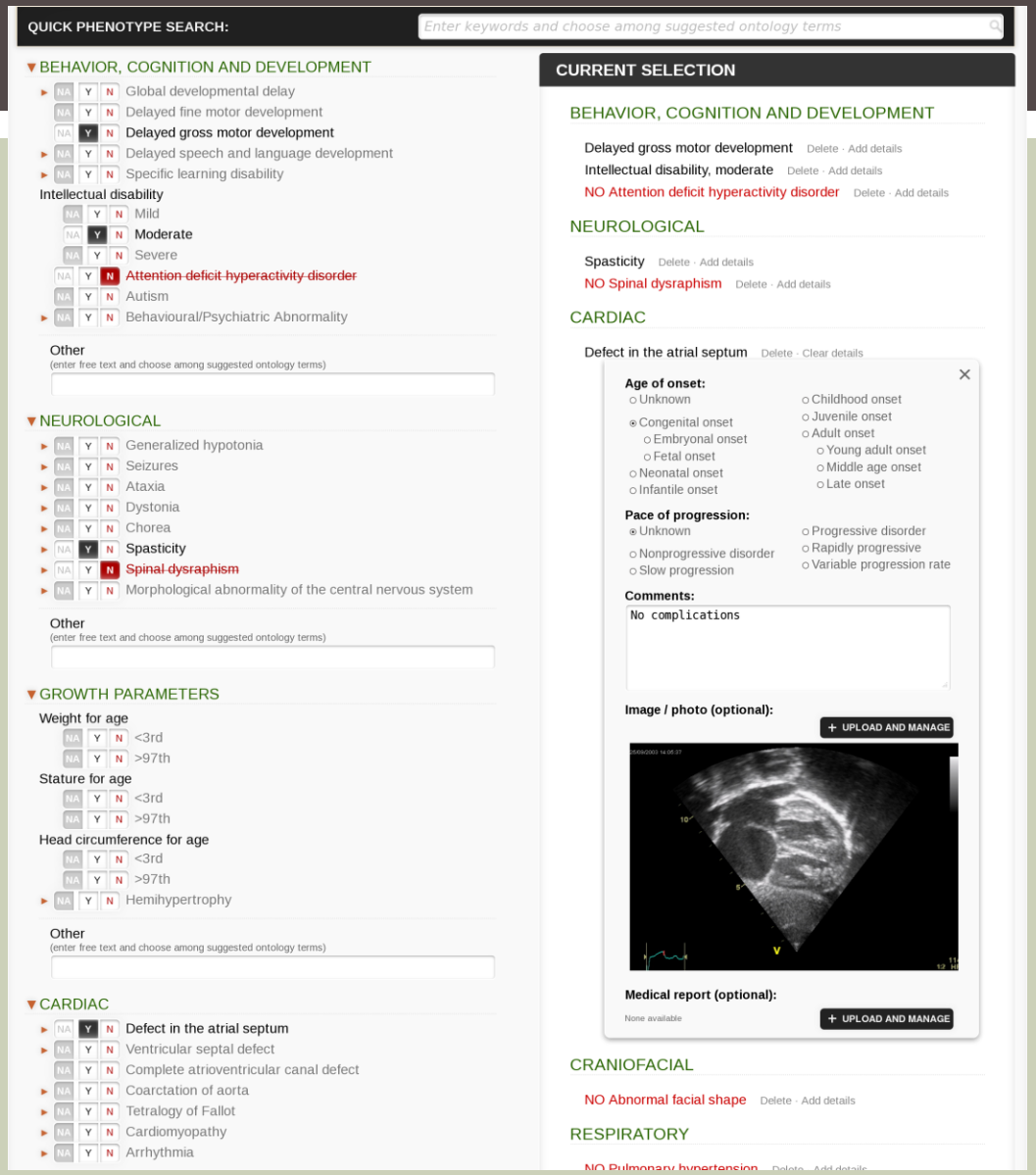
- Lets you know about other similar patients
- Lets you easily connect with other users

Each Patient Record can be:

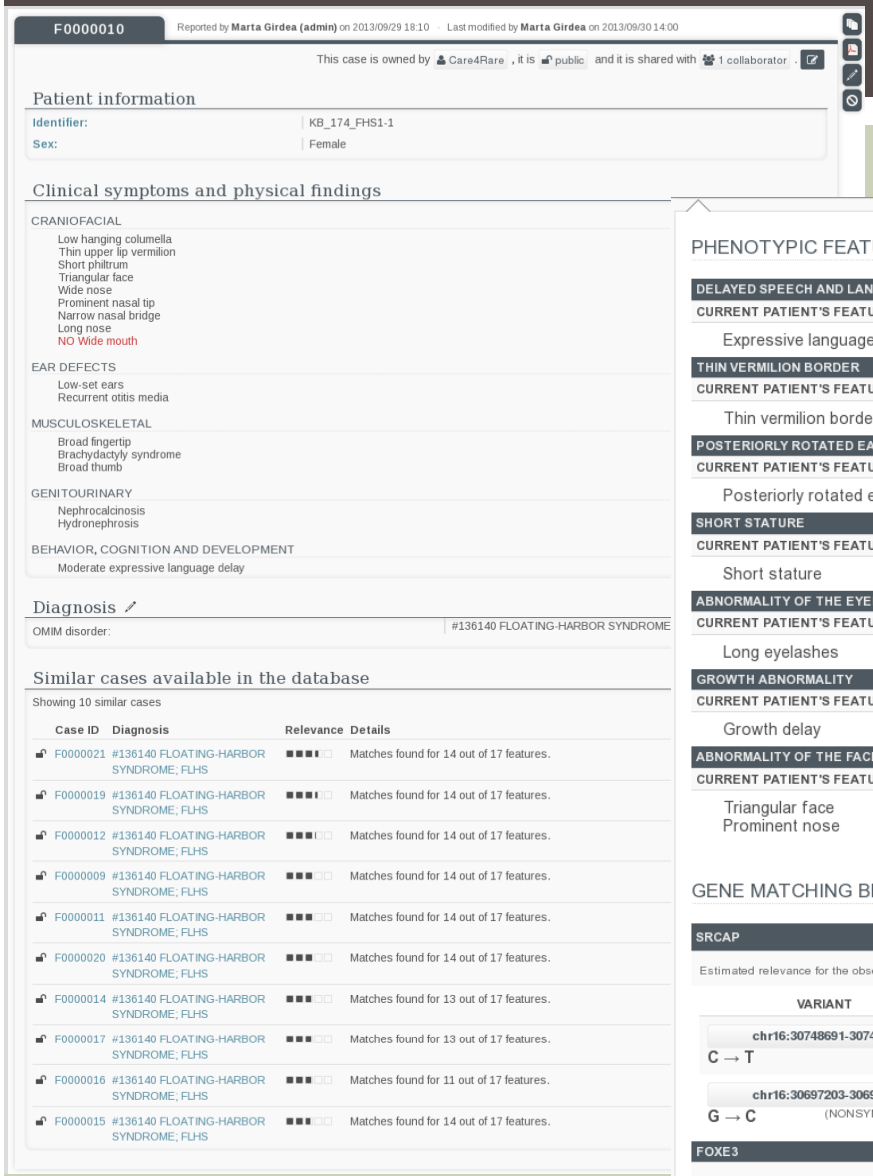
- **Public** – Anyone can see the record
- **Private** – Only specified users/consortia can see the record
- **Matchable** – The record cannot be seen, but can be "discovered" by users who submit similar patients

STEP 1: ADD PATIENT

- Can use the interface built into PhenomeCentral
- Can export data directly from a local PhenoTips instance
- Add a vcf file (or list of genes)
- Set each record as Private, Public or **Matchable**



STEP 2: SEE PATIENTS SIMILAR TO YOURS



F0000010

Reported by **Marta Girdea (admin)** on 2013/09/29 18:10 · Last modified by **Marta Girdea** on 2013/09/30 14:00

This case is owned by Care4Rare , it is public and it is shared with 1 collaborator .

Patient information

Identifier:	KB_174_FHS1-1
Sex:	Female

Clinical symptoms and physical findings

CRANIOFACIAL
- Low hanging columella
- Thin upper lip vermilion
- Short philtrum
- Triangular face
- Wide nose
- Prominent nasal tip
- Narrow nasal bridge
- Long nose
- NO Wide mouth

EAR DEFECTS
- Low-set ears
- Recurrent otitis media

MUSCULOSKELETAL
- Broad fingertip
- Brachydactyly syndrome
- Broad thumb

GENITOURINARY
- Nephrocalcinosis
- Hydronephrosis

BEHAVIOR, COGNITION AND DEVELOPMENT
- Moderate expressive language delay

Diagnosis

OMIM disorder: #136140 FLOATING-HARBOR SYNDROME

Similar cases available in the database

Showing 10 similar cases

Case ID	Diagnosis	Relevance	Details
F0000021	#136140 FLOATING-HARBOR SYNDROME; FLHS		Matches found for 14 out of 17 features.
F0000019	#136140 FLOATING-HARBOR SYNDROME; FLHS		Matches found for 14 out of 17 features.
F0000012	#136140 FLOATING-HARBOR SYNDROME; FLHS		Matches found for 14 out of 17 features.
F0000009	#136140 FLOATING-HARBOR SYNDROME; FLHS		Matches found for 14 out of 17 features.
F0000011	#136140 FLOATING-HARBOR SYNDROME; FLHS		Matches found for 14 out of 17 features.
F0000020	#136140 FLOATING-HARBOR SYNDROME; FLHS		Matches found for 14 out of 17 features.
F0000014	#136140 FLOATING-HARBOR SYNDROME; FLHS		Matches found for 13 out of 17 features.
F0000017	#136140 FLOATING-HARBOR SYNDROME; FLHS		Matches found for 13 out of 17 features.
F0000016	#136140 FLOATING-HARBOR SYNDROME; FLHS		Matches found for 11 out of 17 features.
F0000015	#136140 FLOATING-HARBOR SYNDROME; FLHS		Matches found for 14 out of 17 features.

PHENOTYPIC FEATURES BREAKDOWN

DELAYED SPEECH AND LANGUAGE DEVELOPMENT 36%

CURRENT PATIENT'S FEATURES	OTHER PATIENT'S FEATURES
Expressive language delay	Delayed speech and language development

THIN VERMILION BORDER 36%

CURRENT PATIENT'S FEATURES	OTHER PATIENT'S FEATURES
Thin vermilion border	Thin upper lip vermilion

POSTERIORLY ROTATED EARS 32%

CURRENT PATIENT'S FEATURES	OTHER PATIENT'S FEATURES
Posteriorly rotated ears	Low-set, posteriorly rotated ears

SHORT STATURE 25%

CURRENT PATIENT'S FEATURES	OTHER PATIENT'S FEATURES
Short stature	Severe short stature

ABNORMALITY OF THE EYELID 24%

CURRENT PATIENT'S FEATURES	OTHER PATIENT'S FEATURES
Long eyelashes	Blepharophimosis

GROWTH ABNORMALITY 20%

CURRENT PATIENT'S FEATURES	OTHER PATIENT'S FEATURES
Growth delay	Decreased body weight

ABNORMALITY OF THE FACE 12%

CURRENT PATIENT'S FEATURES	OTHER PATIENT'S FEATURES
Triangular face Prominent nose	Dental malocclusion Wide mouth Microdontia

GENE MATCHING BREAKDOWN

SRCAP HIDE VARIANTS...

Estimated relevance for the observed phenotype in the **current patient**: 60%

Estimated relevance for the observed phenotype in the **other patient**: 65%

VARIANT	ESTIMATED HARMFULLNESS
chr16:30748691-30748691 C → T (STOPGAIN)	95%
chr16:30697203-30697203 G → C (NONSYNONYMOUS)	57%

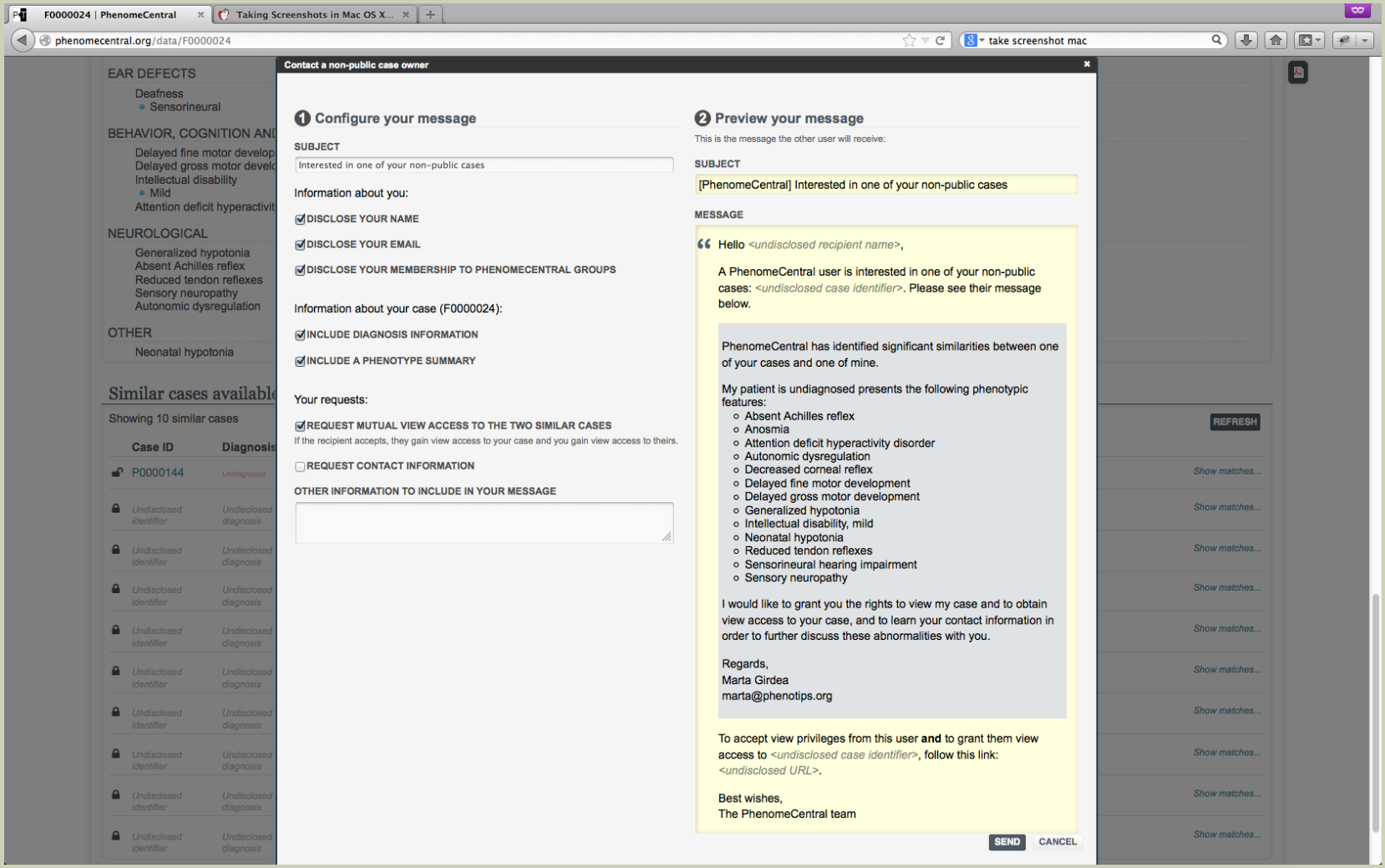
VARIANT	ESTIMATED HARMFULLNESS
chr16:30751917-30751917 G → A (FS_INSERTION)	95%

FOXE3

Estimated relevance for the observed phenotype in the **current patient**: 57%

Estimated relevance for the observed phenotype in the **other patient**: 50%

STEP 3: CONTACT THE SUBMITTER OF THE OTHER DATASET



OUTLINE

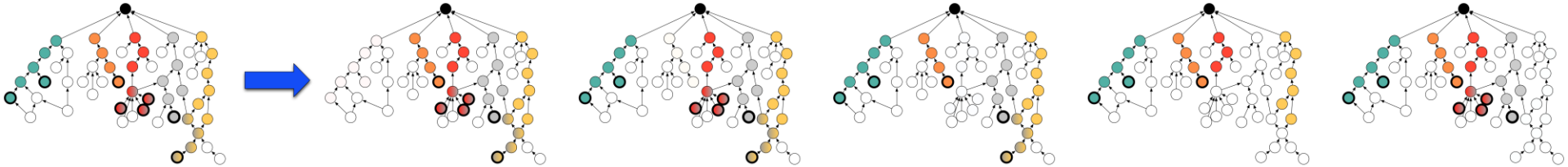
- Why phenotyping is hard
- About Ontologies
- Diagnosing known diseases
- Getting the phenotype data
- **How much phenotyping is enough?**
- Model organism data for undiagnosed diseases

HOW MUCH PHENOTYPING IS ENOUGH?

- How many annotations…?
- How many different categories?
- How many within each?

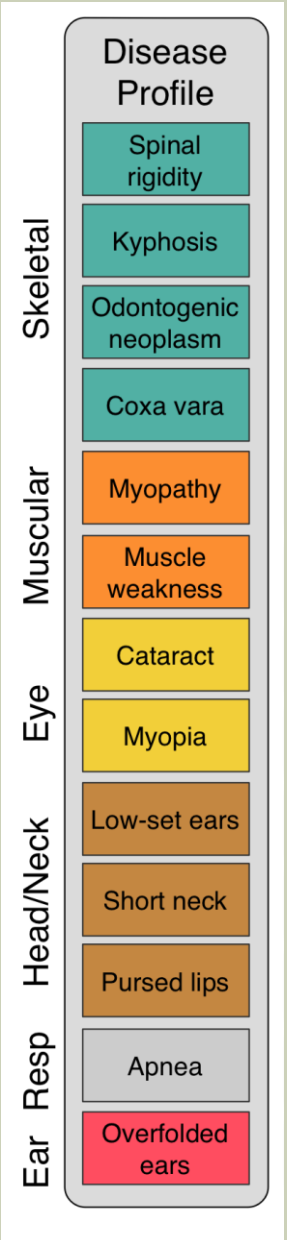
Not everything that counts can be counted and not everything that can be counted counts -Albert Einstein

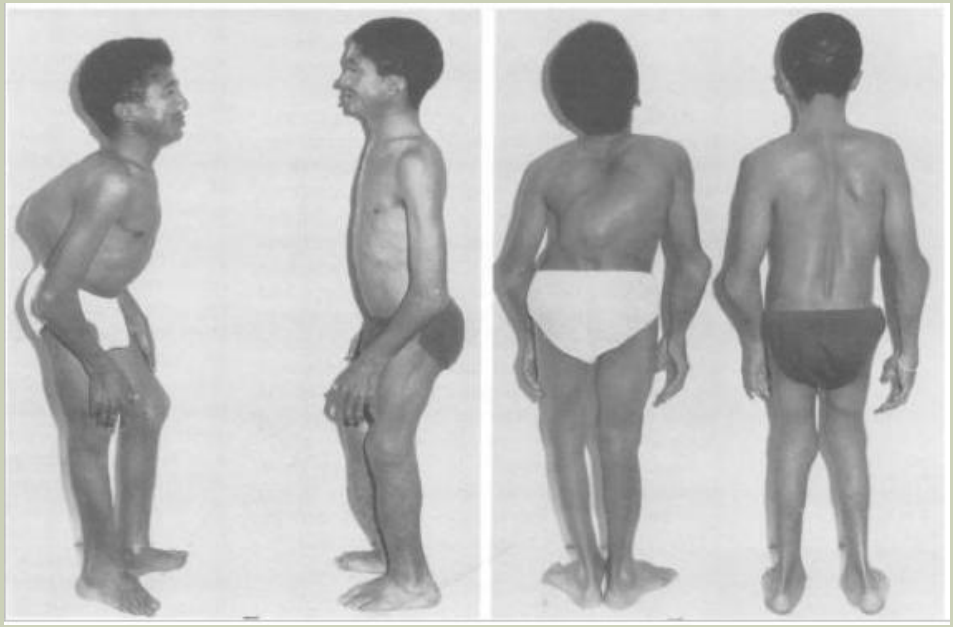
METHOD: DERIVE BY CATEGORY REMOVAL



- Remove annotations that are subclasses of a single high-level node
- Repeat for each 1° subclass

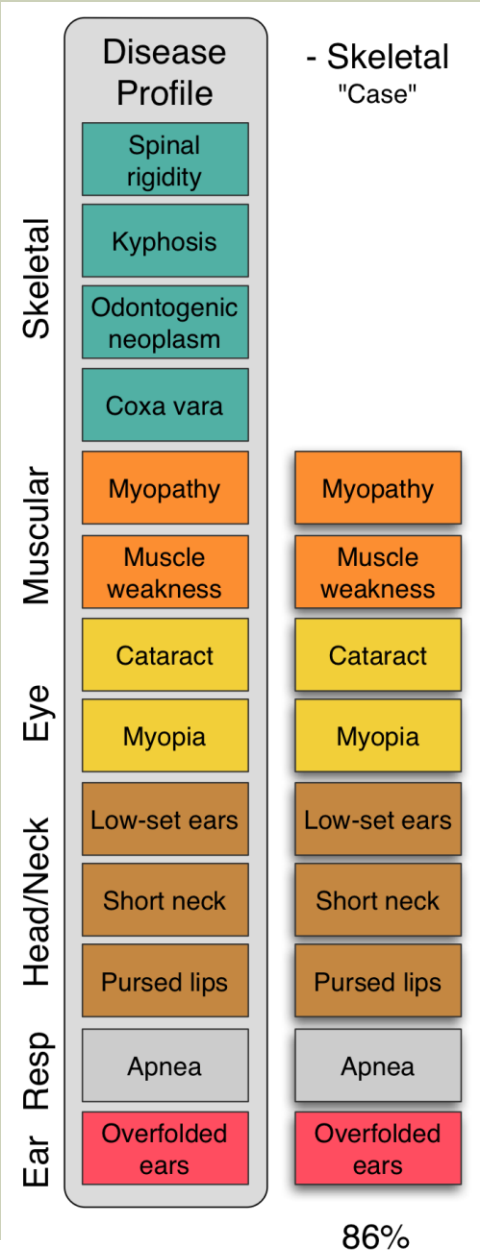
Example: Schwartz-jampel Syndrome, Type I to test influence of a single phenotypic category



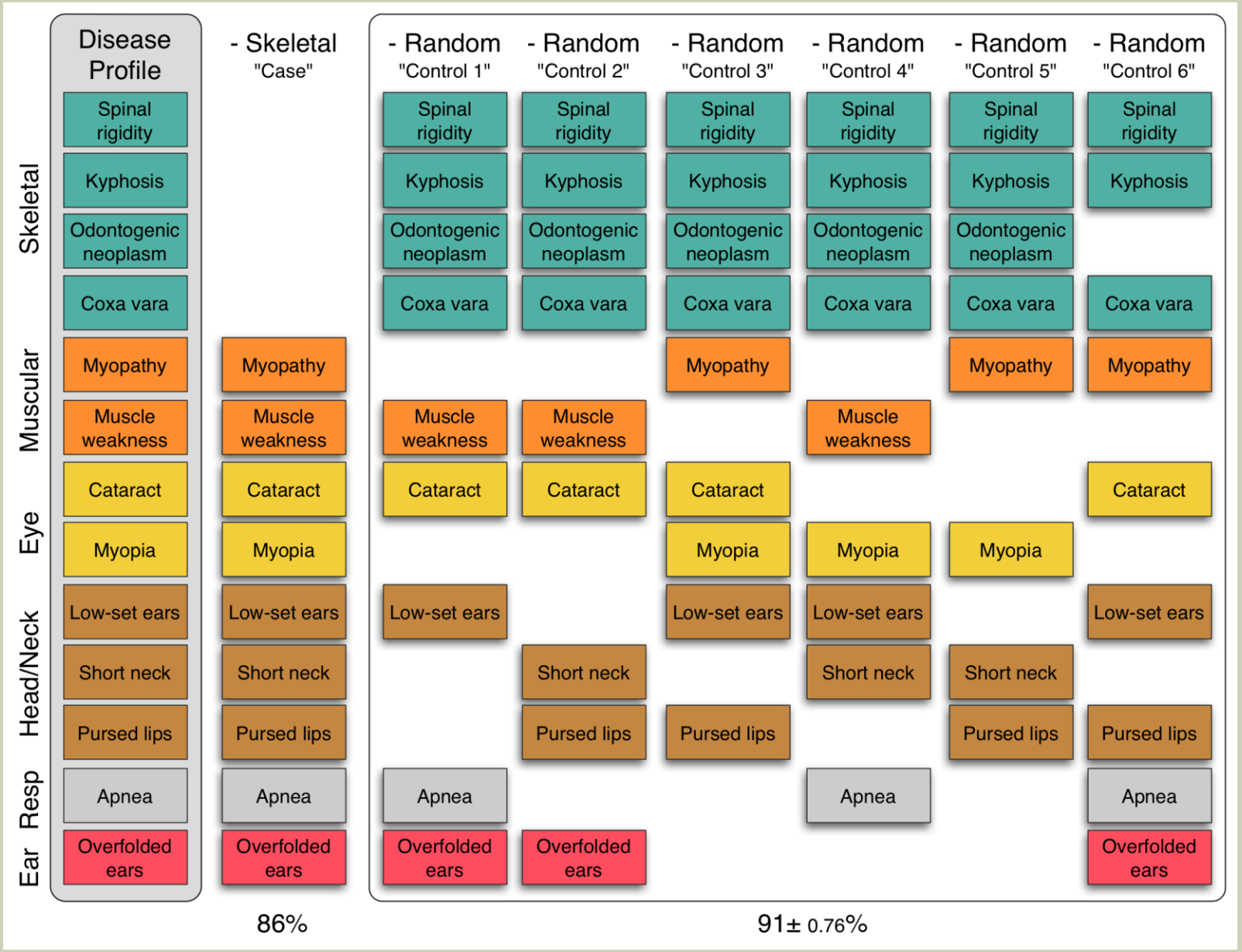


Example: Schwartz-jampel Syndrome

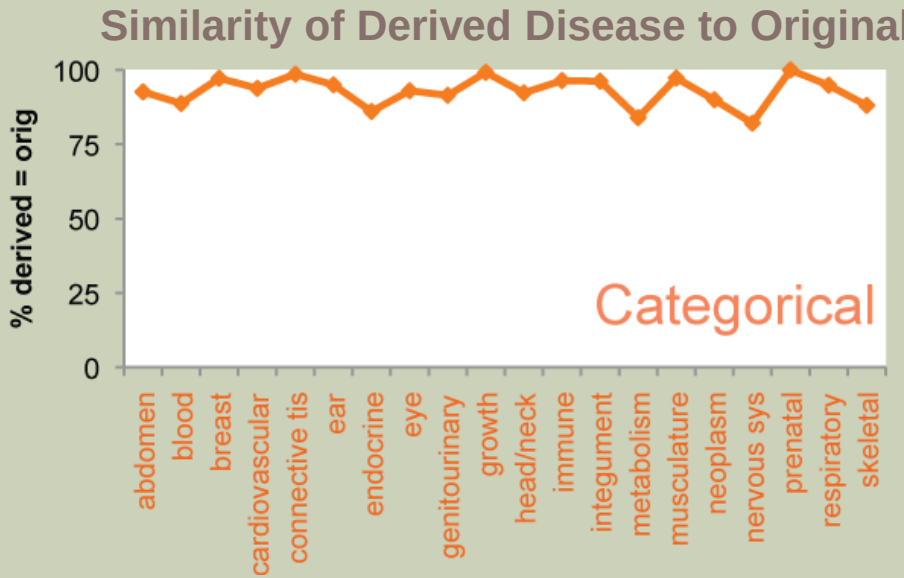
derivations to test influence of a single phenotypic category

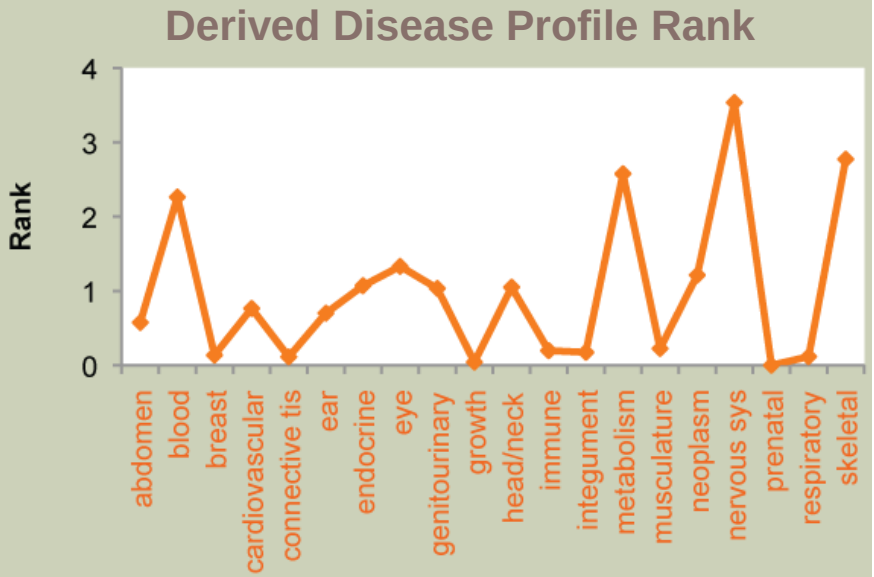


Schwartz-jampel Syndrome derivations



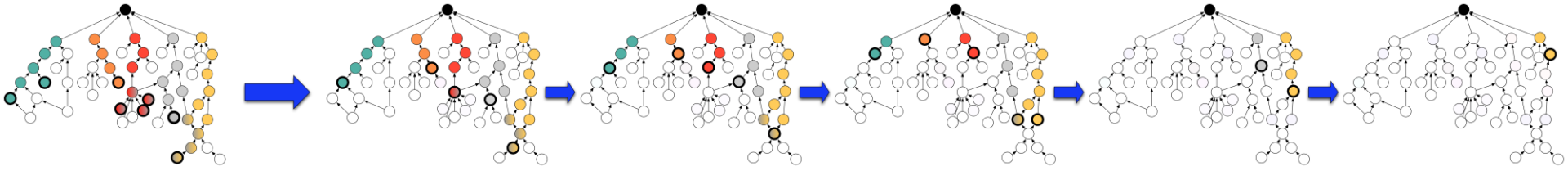
SEMANTIC SIMILARITY ALGORITHMS ARE ROBUST IN THE FACE OF MISSING INFORMATION





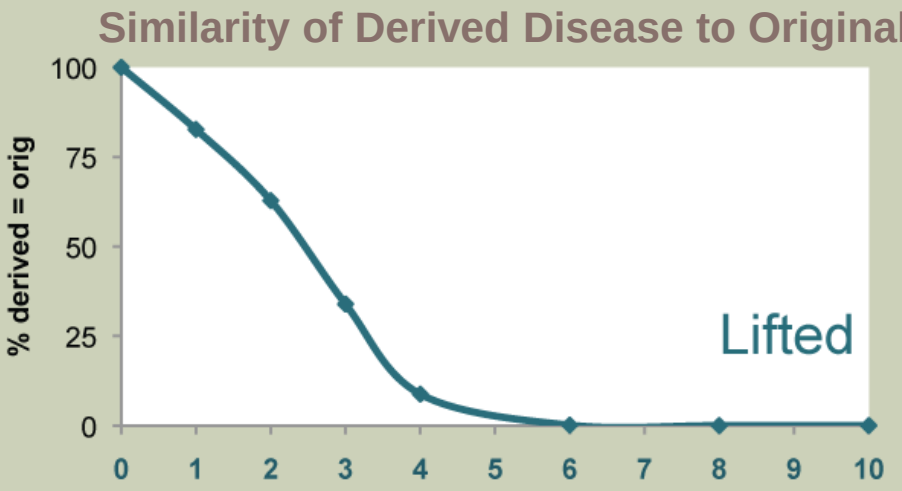
- (avg) 92% of derived diseases are most-similar to original disease
- Severity of impact follows proportion of phenotype

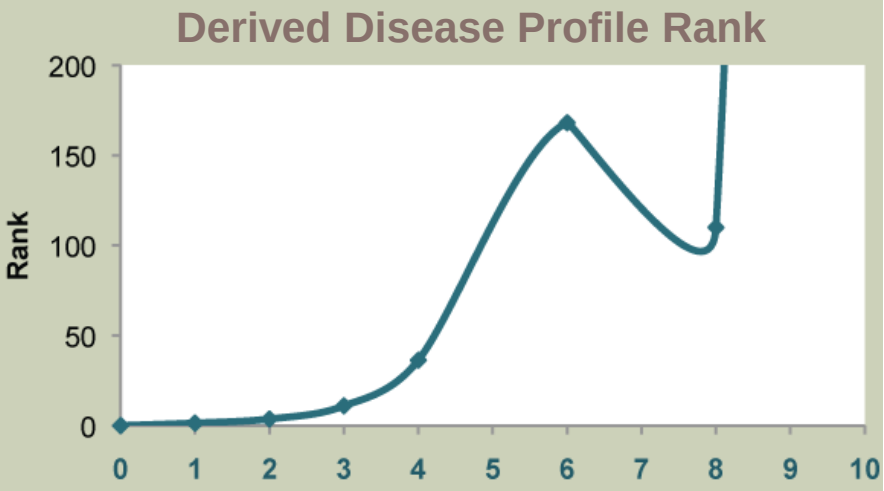
METHOD: DERIVE BY LIFTING



- Iteratively map each class to their direct superclass(es)
- Keep only leaf nodes

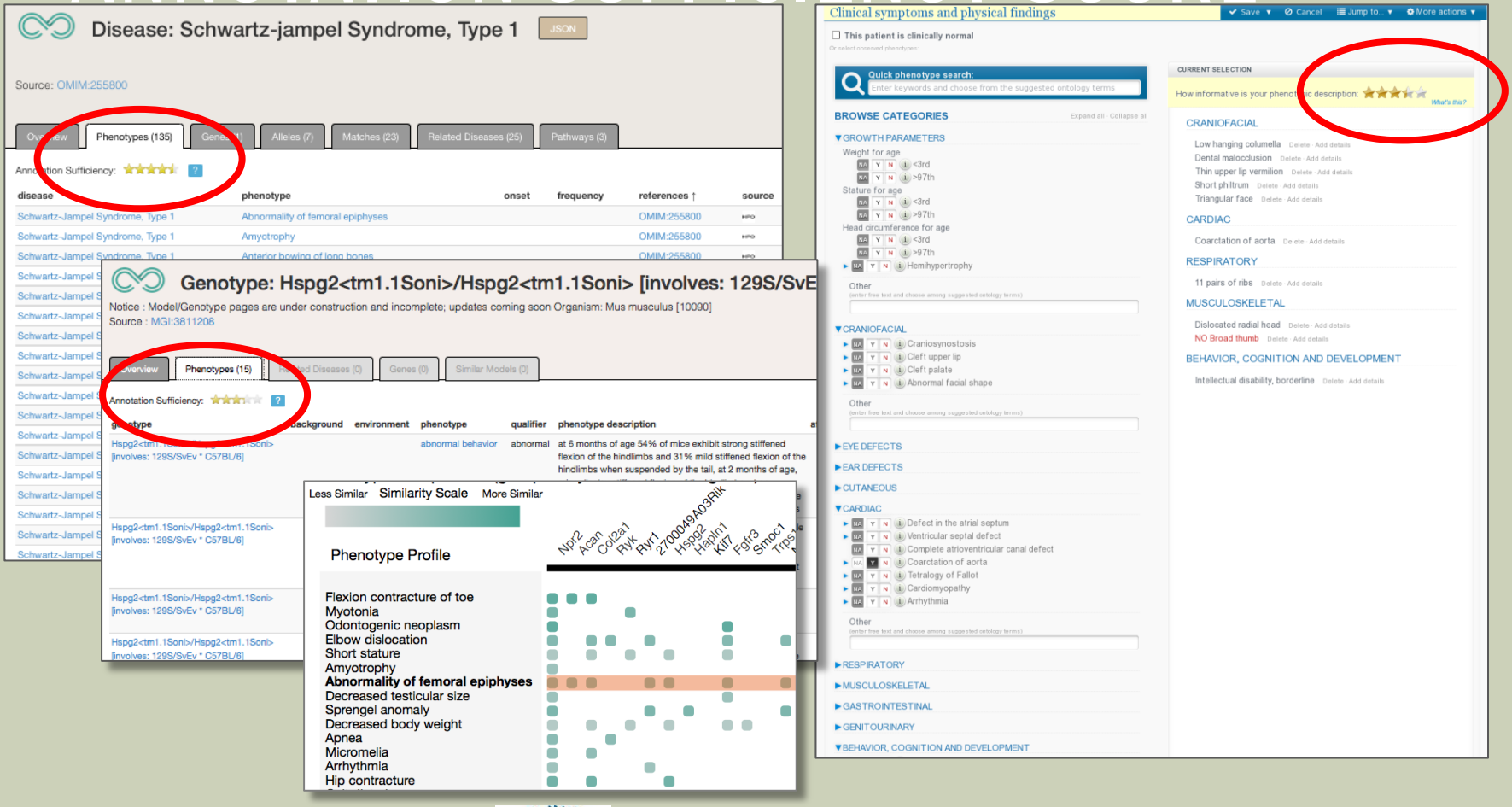
SEMANTIC SIMILARITY ALGORITHMS ARE SENSITIVE TO SPECIFICITY OF INFORMATION





- Severity of impact increases with more-general phenotypes

ANNOTATION SUFFICIENCY SCORE



NATIONAL INSTITUTES OF HEALTH

PHENOTIPS

http://monarchinitiative.org/page/services

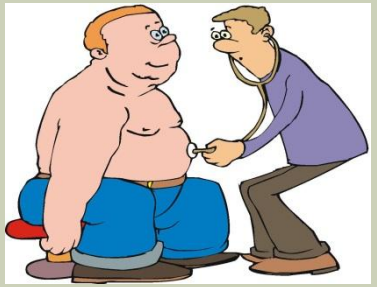
OUTLINE

- Why phenotyping is hard
- About Ontologies
- Diagnosing known diseases
- Getting the phenotype data
- How much phenotyping is enough?
- **Model organism data for undiagnosed diseases**

WHAT TO DO WHEN WE CAN'T DIAGNOSE WITH A KNOWN DISEASE?

MODELS RECAPITULATE VARIOUS PHENOTYPIC ASPECTS OF DISEASE

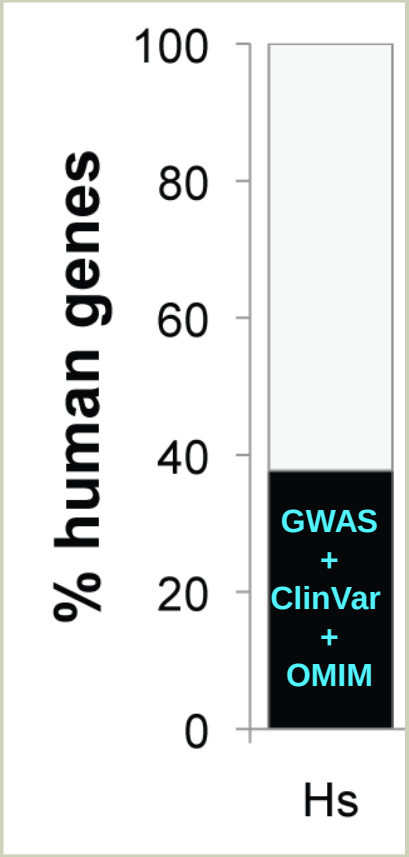
GENOTYPE

$kcnj11^{c14/c14}$; $insr^{t143/+}$(AB)	B6.Cg-Alms1$^{foz/fox}$/J	ALSM1(NM_015120.4) [c.10775delC] + [-] ?
		
increased weight, adipose tissue volume, glucose homeostasis altered	increased food intake, hyperglycemia, insulin resistance	obesity, diabetes mellitus, insulin resistance

PHENOTYPE

HOW MUCH PHENOTYPE DATA?

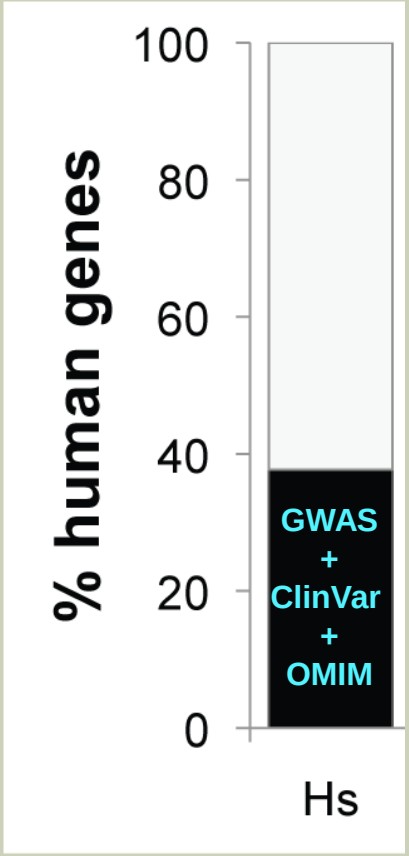
- Human genes have poor phenotype coverage



HOW MUCH PHENOTYPE DATA?

- Human genes have poor phenotype coverage

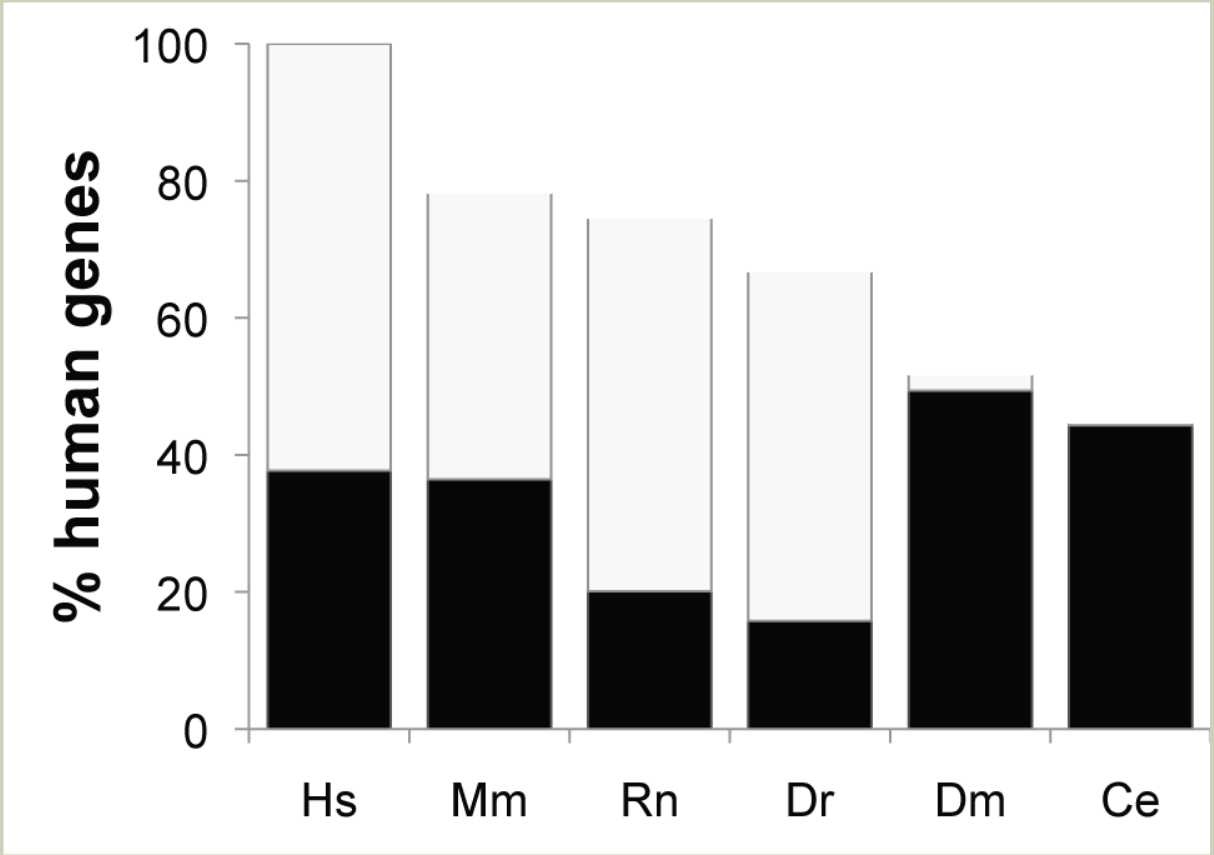
  What else can we leverage?



HOW MUCH PHENOTYPE DATA?

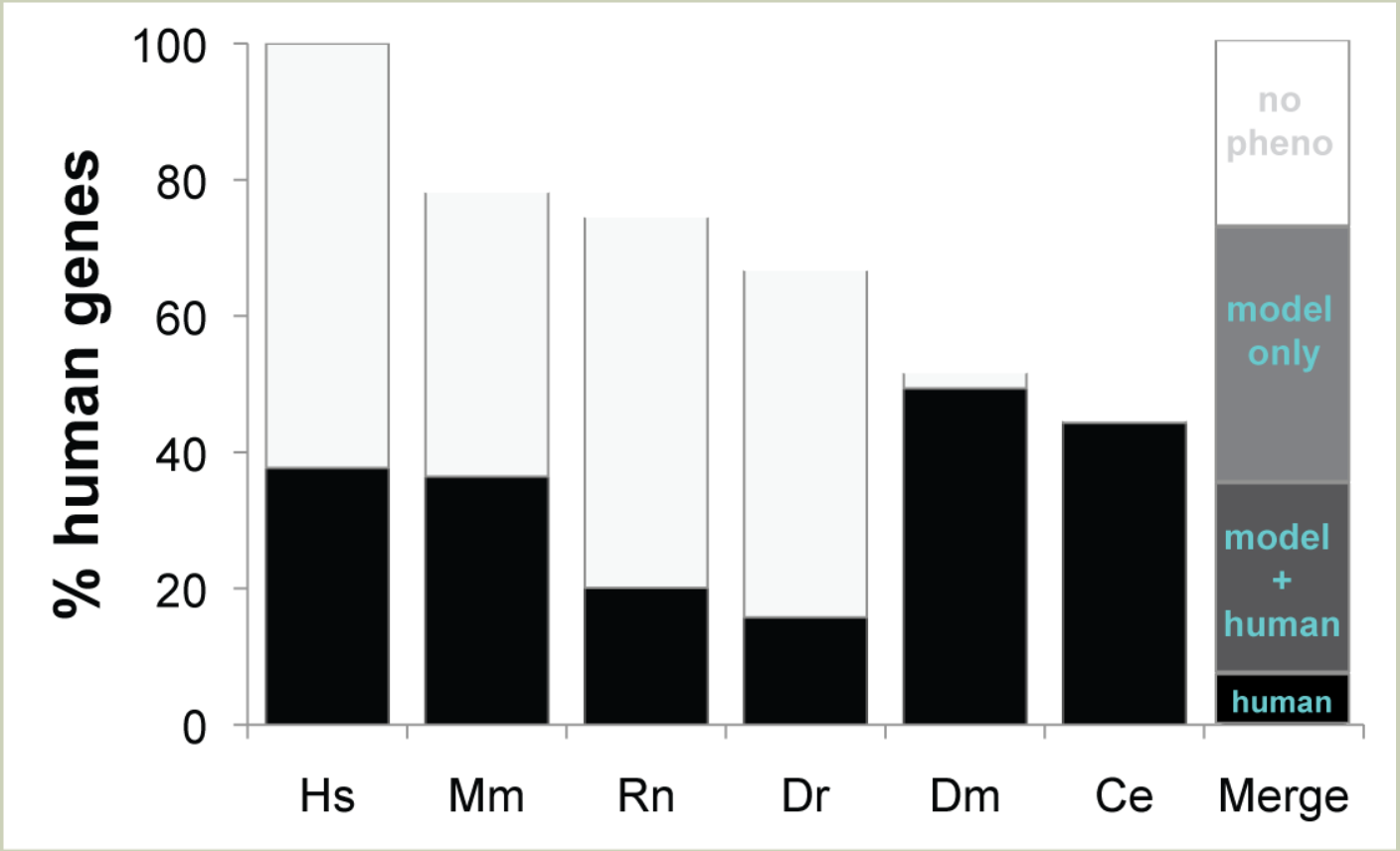
- Human genes have poor phenotype coverage

  What else can we leverage? …*animal models*



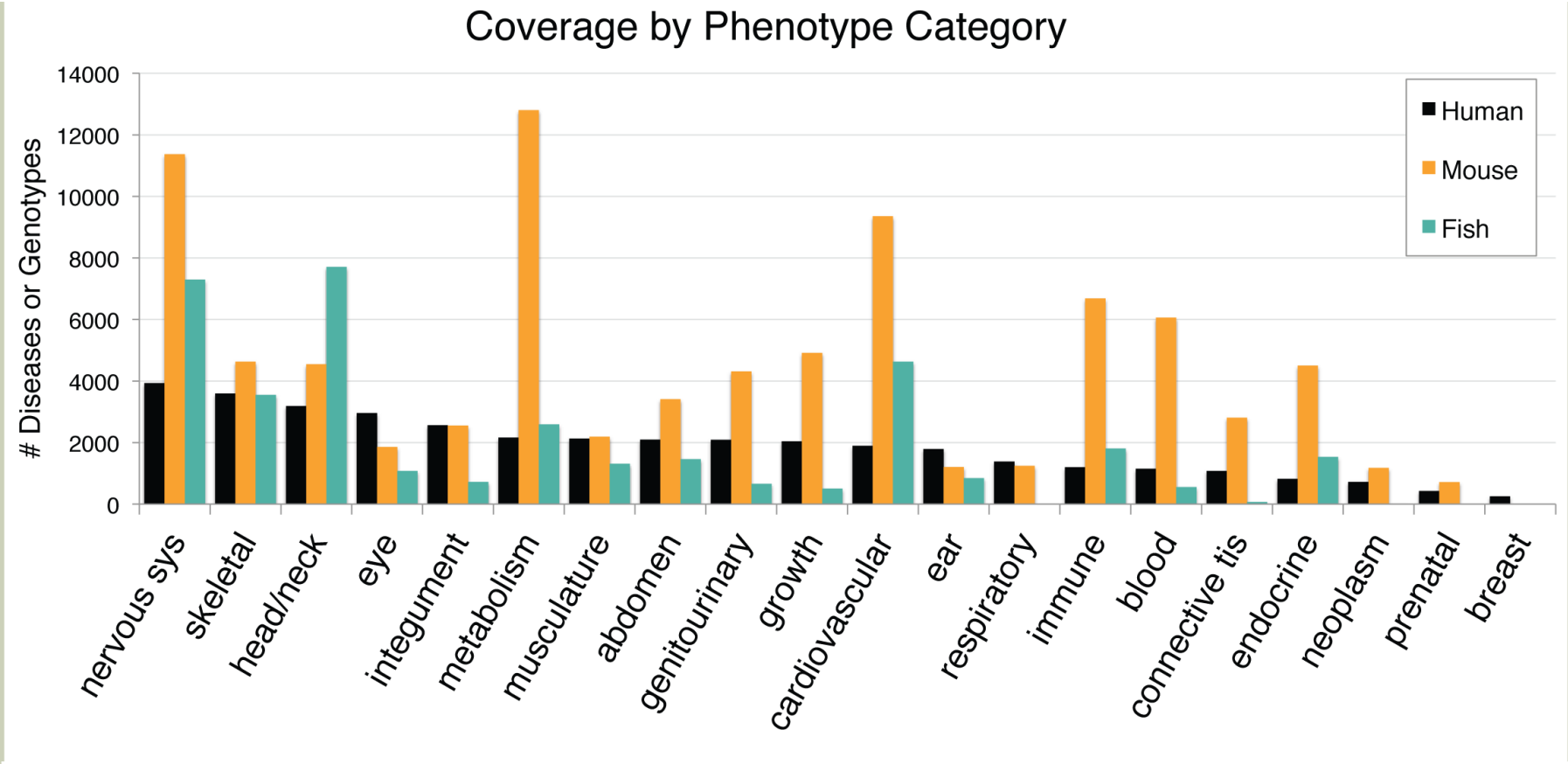
Orthology via PANTHER v9

COMBINED, HUMAN AND MODEL PHENOTYPES CAN BE LINKED TO >75% HUMAN GENES



Orthology via PANTHER v9

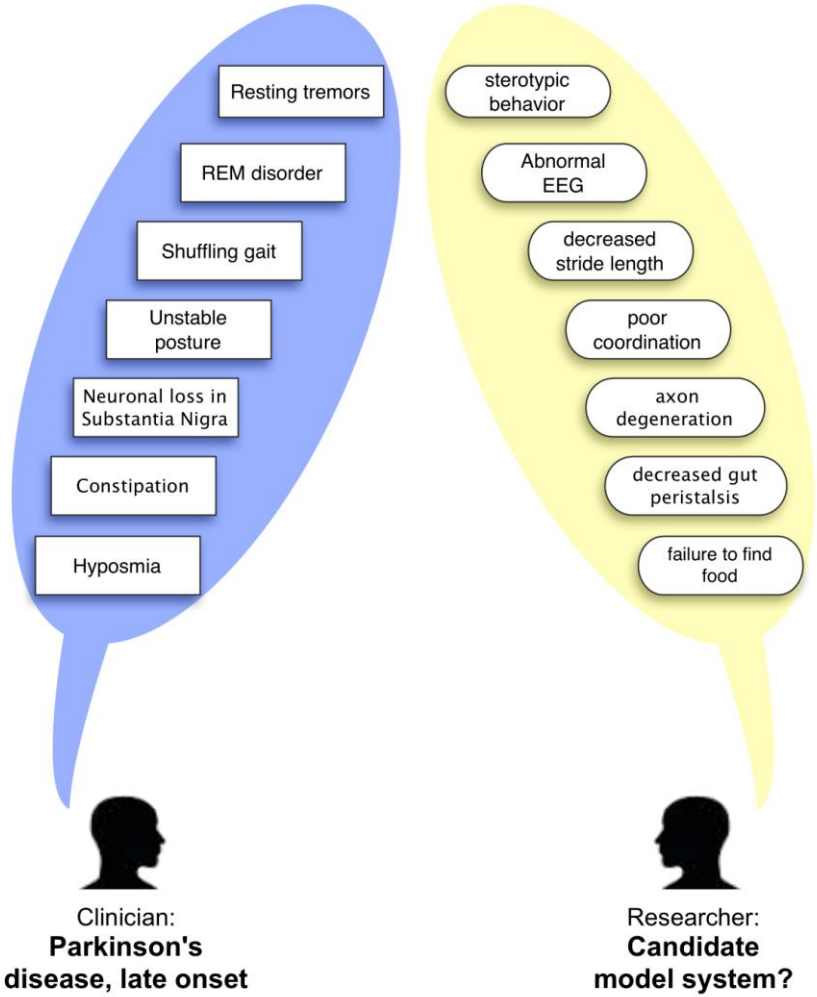
EACH MODEL CONTRIBUTES DIFFERENT PHENOTYPES



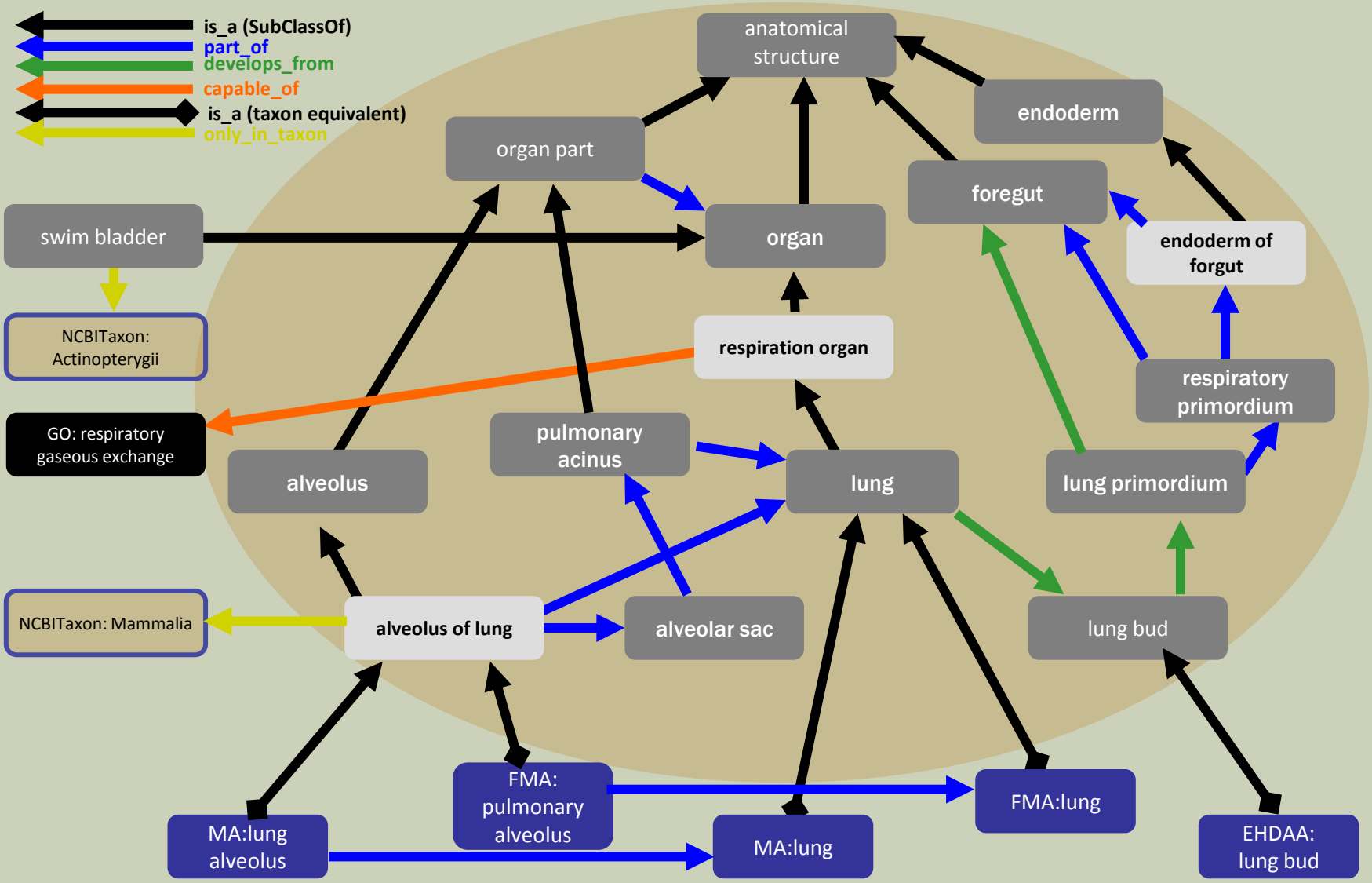
Data from MGI, ZFIN, & HPO, reasoned over with cross-species phenotype ontology

https://code.google.com/p/phenotype-ontologies/

PROBLEM: CLINICAL AND MODEL PHENOTYPES ARE DESCRIBED DIFFERENTLY

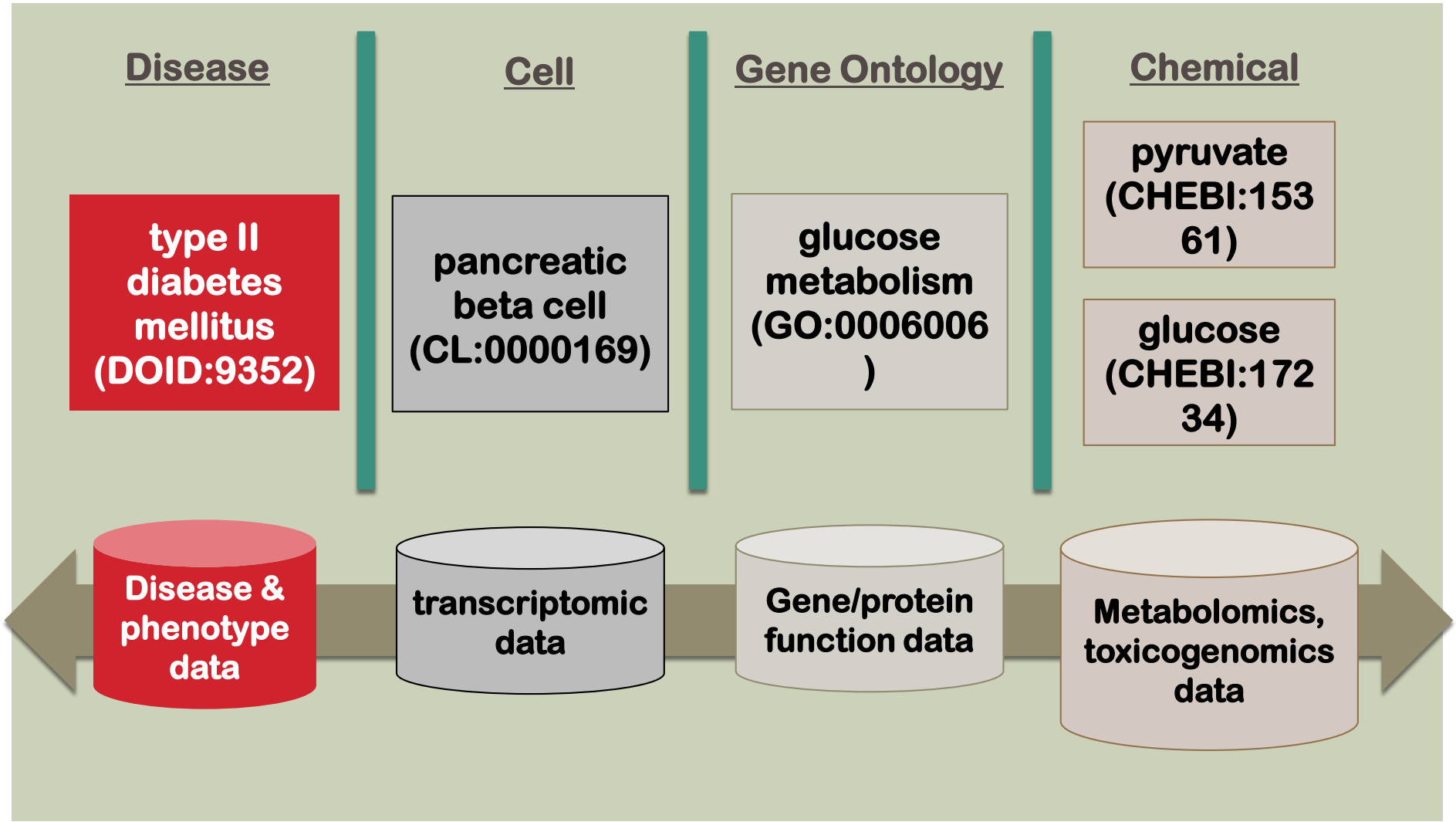


SOLUTION: BRIDGING SEMANTICS

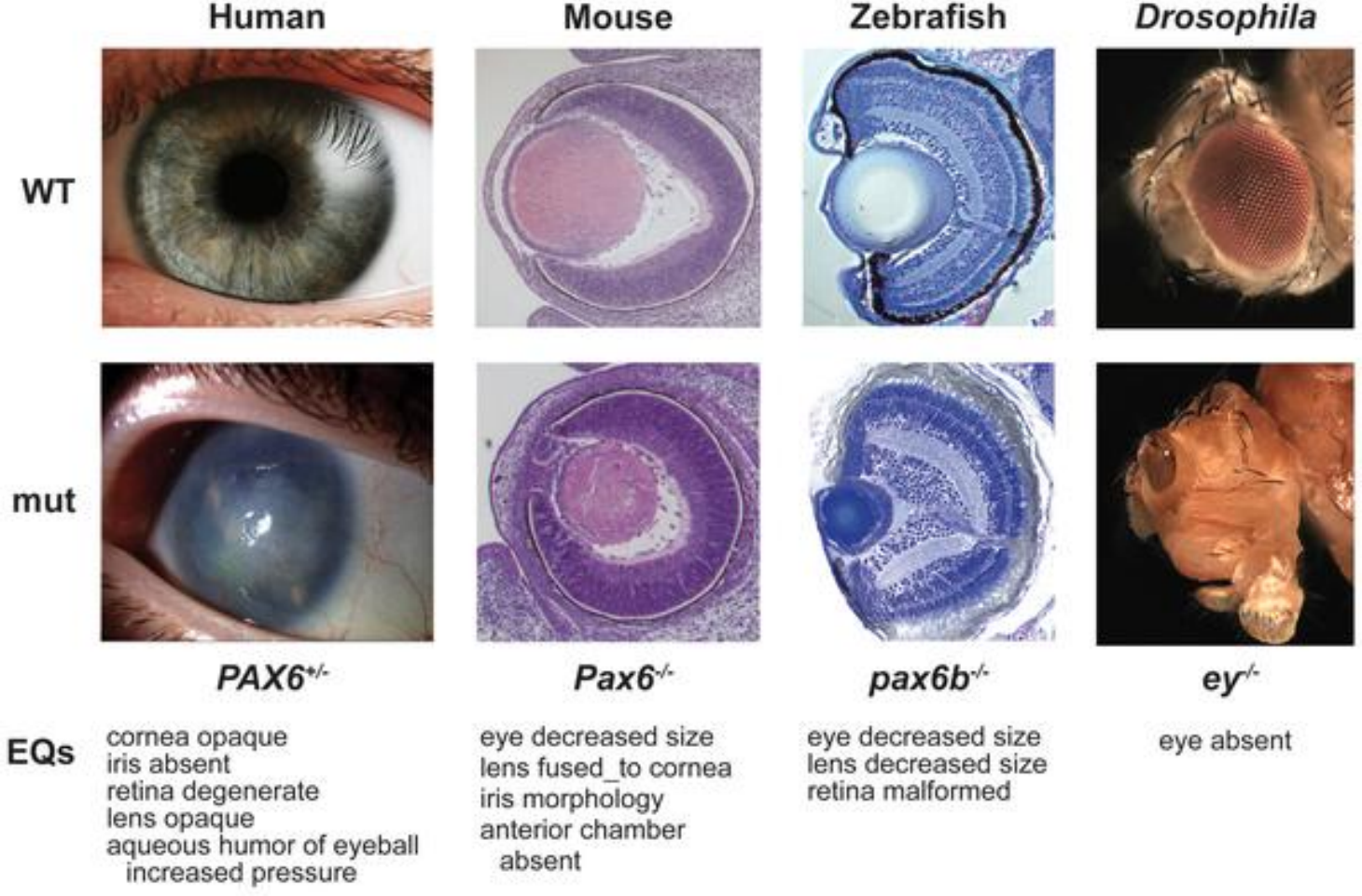


Mungall, C. J., Torniai, C., Gkoutos, G. V., Lewis, S. E., & Haendel, M. A. (2012). **Uberon, an integrative multi-species anatomy ontology. Genome Biology**, *13(1), R5. doi:10.1186/gb-2012-13-1-r5*

PHENOTYPE REPRESENTATION REQUIRES MORE THAN "PHENOTYPE ONTOLOGIES

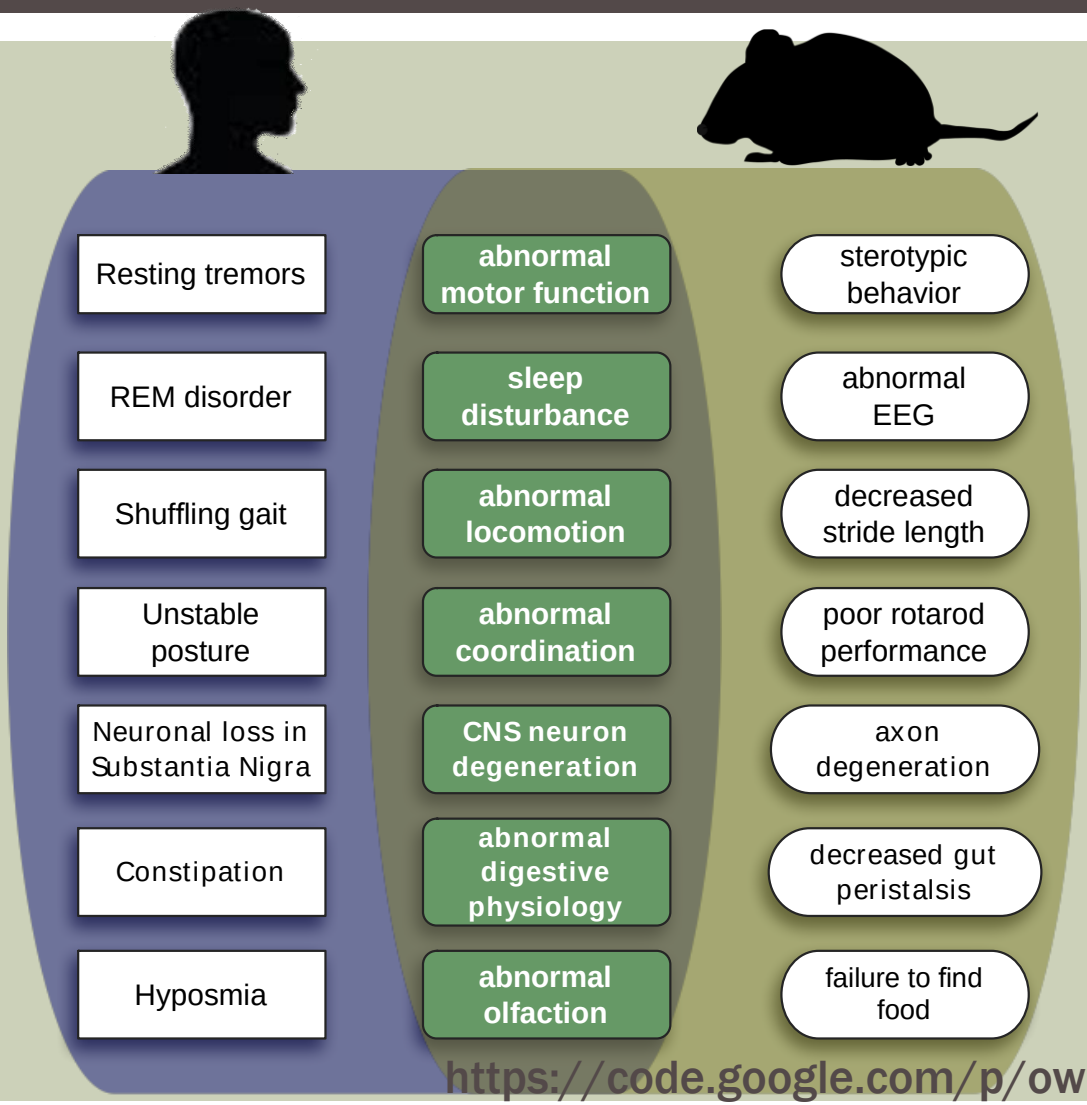


MODELS BASED ON PHENOTYPIC SIMILARITY



Washington, Haendel, et al. (2009). **Linking Human Diseases to Animal Models Using Ontology-Based Phenotype Annotation**. *PLoS Biol, 7(11). doi:10.1371/journal.pbio.1000247*

OWLSIM: PHENOTYPE SIMILARITY ACROSS PATIENTS OR ORGANISMS



https://code.google.com/p/owltools/wiki/OwlSim

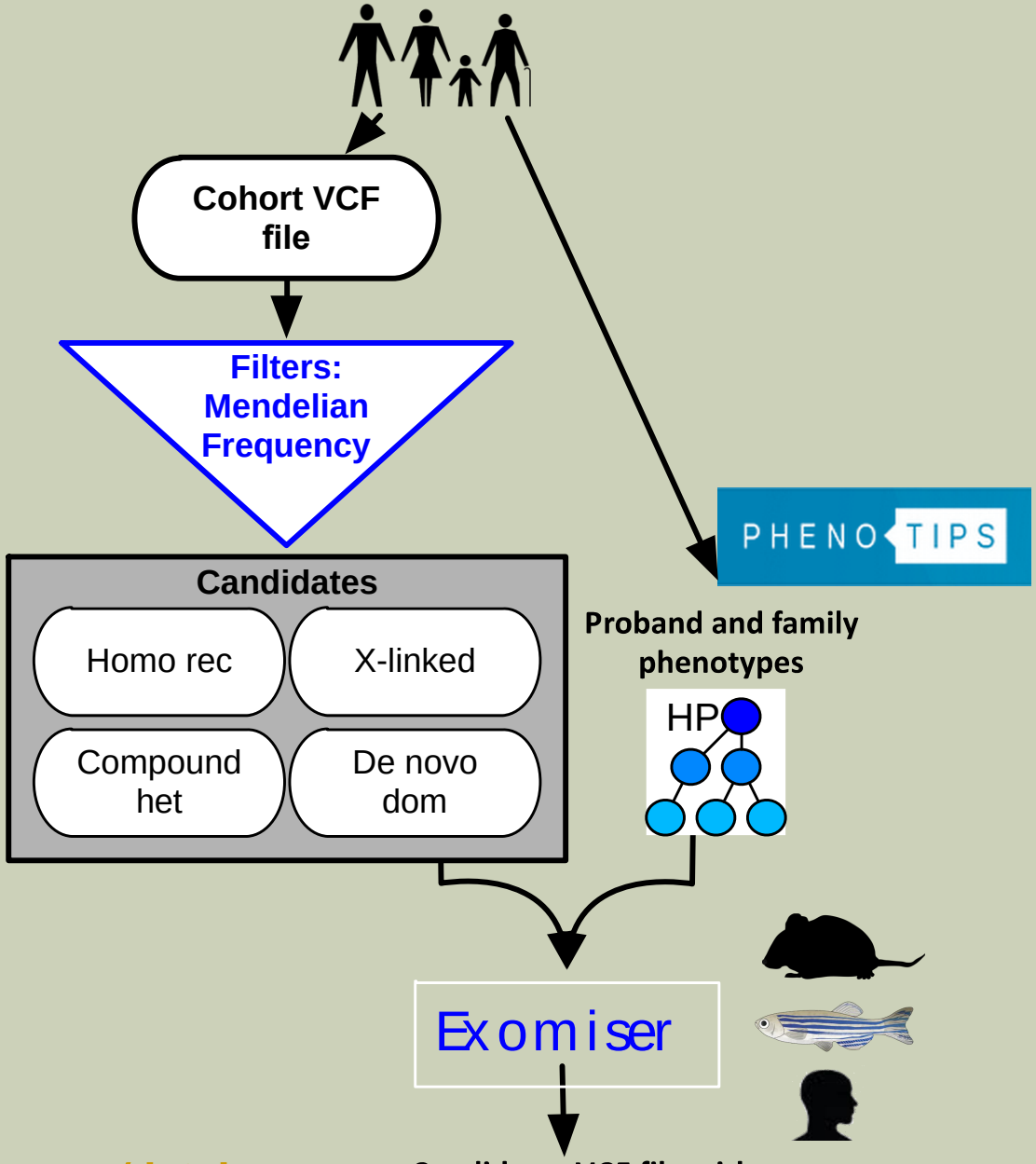
MONARCH PHENOTYPE DATA

Species	Data source	Genes	Genotypes	Variants	Phenotype annotations	Diseases
mouse	MGI	13,433	59,087	34,895	271,621	
fish	ZFIN	7,612	25,588	17,244	81,406	
fly	Flybase	27,951	91,096	108,348	267,900	
worm	Wormbase	23,379	15,796	10,944	543,874	
human	HPOA				112,602	7,401
human	OMIM	2,970			4,437	3,651
human	ClinVar	3,215		100,523	445,241	4,056
human	KEGG	2,509			3,927	1,159
human	ORPHANET	3,113			5,690	3,064
human	CTD	7,414			23,320	4,912

Also in the system: Rat; IMPC; GO annotations; Coriell cell lines; OMIA; MPD; Yeast; CTD; GWAS; Panther, Homologene orthologs; BioGrid interactions; Drugbank; AutDB; Allen Brain …157 sources to date

Coming soon: Animal QTLs for pig, cattle, chicken, sheep, trout, dog, horse

EXOMISER METHOD



https://www.sanger.ac.uk/resources/databases/exomiser/query/exomiser2

EXOMISER RESULTS ON NIH UNDIAGNOSED DISEASE PROGRAM PATIENTS

9 previously diagnosed families

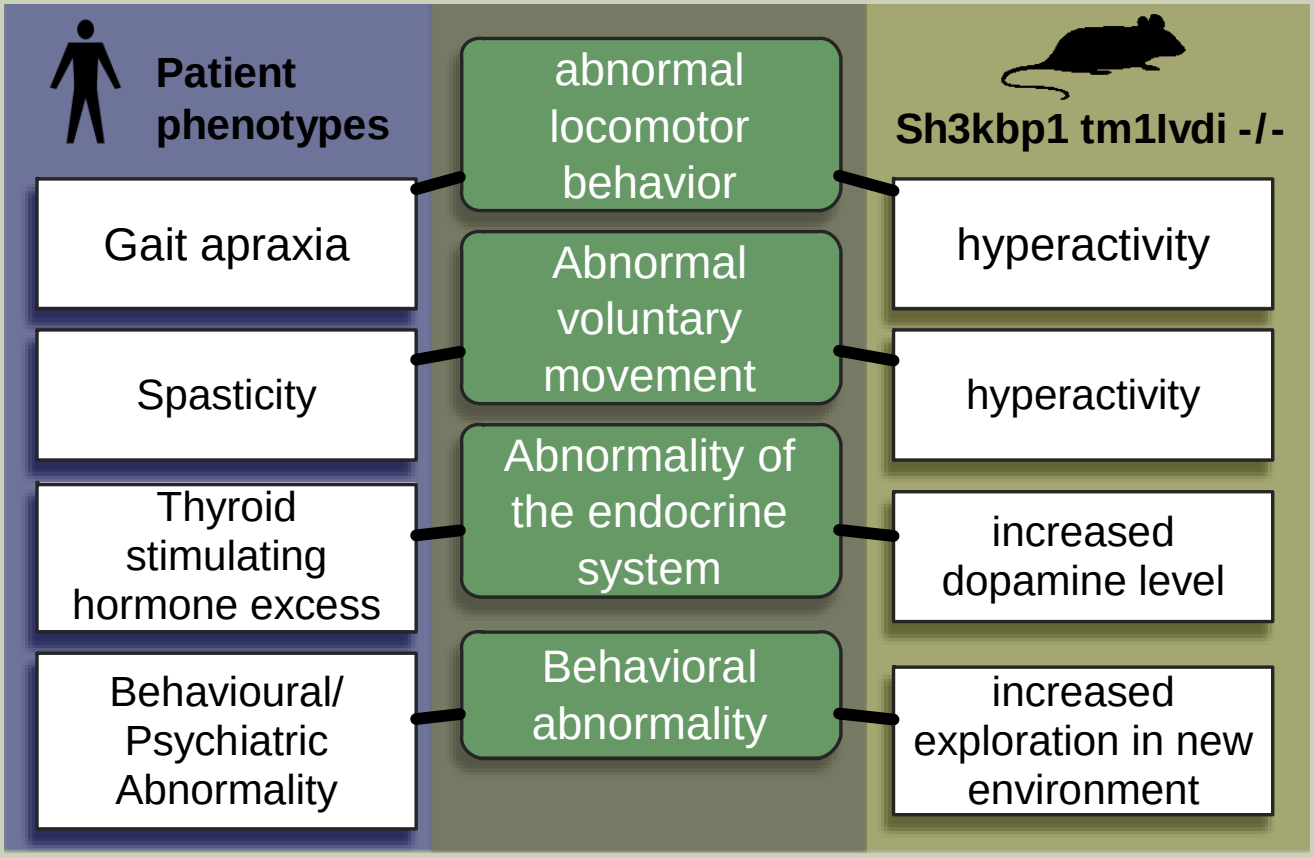
Identified causative variants with a rank of at least 7/408 potential variants

21 families without identified disorders

We have now prioritized variants in STIM1, ATP13A2, PANK2, and CSF1R in 5 different families (2 STIM1 families)

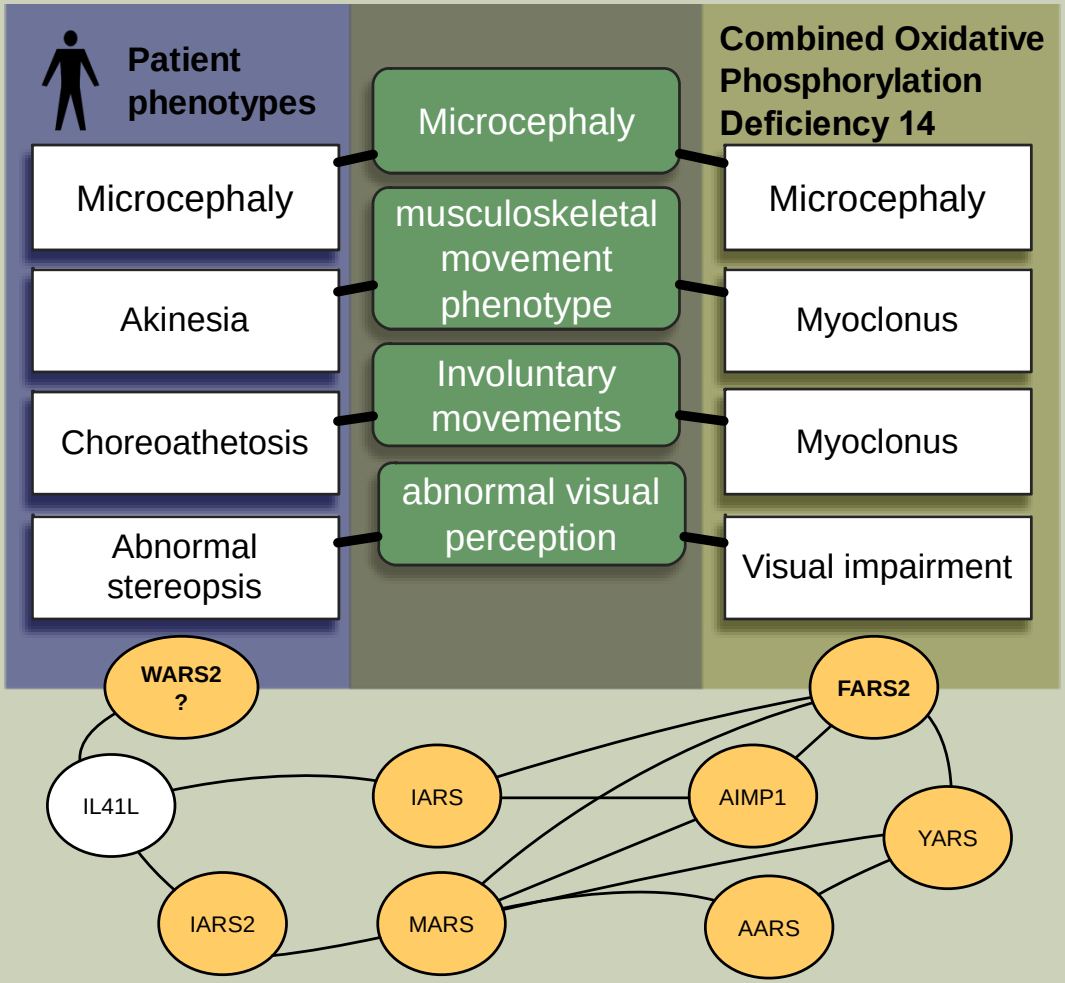
Bone et al., submitted ☺

UDP_2731

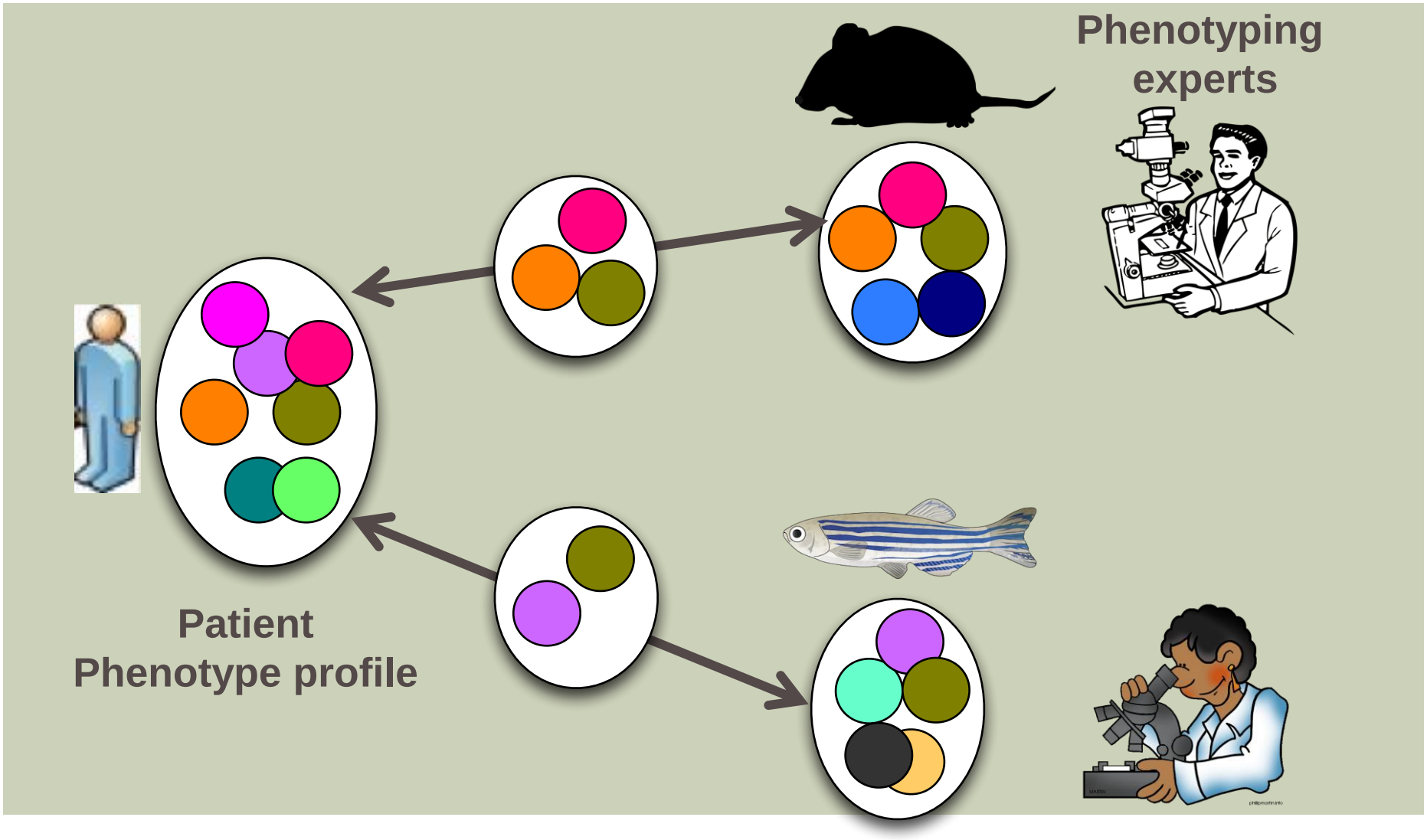

Patient phenotypes
Gait apraxia
Spasticity
Thyroid stimulating hormone excess
Behavioural/ Psychiatric Abnormality
abnormal locomotor behavior
Abnormal voluntary movement
Abnormality of the endocrine system
Behavioral abnormality
Sh3kbp1 tm1lvdi -/-
hyperactivity
hyperactivity
increased dopamine level
increased exploration in new environment

WALKING THE INTERACTOME

UDP_1166

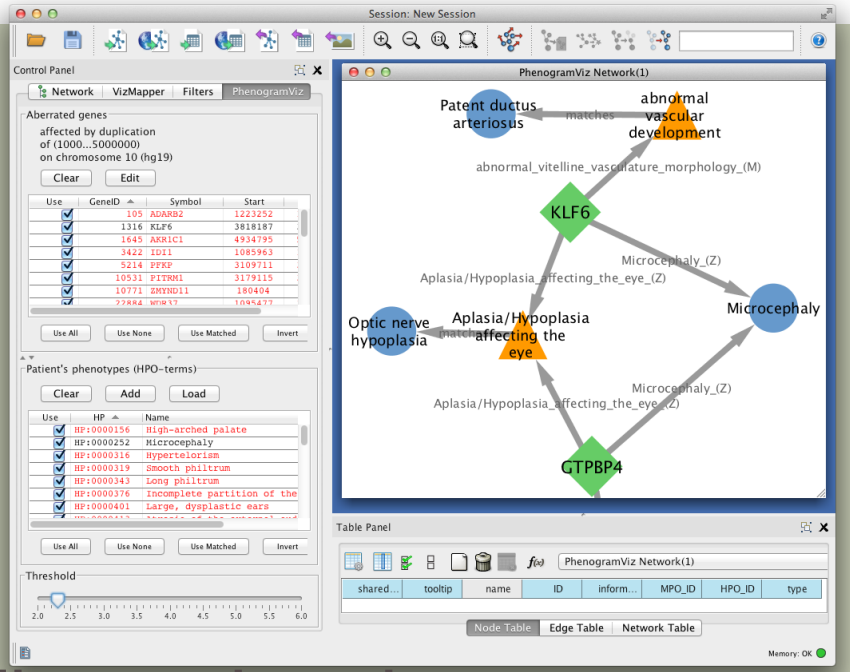


FINDING COLLABORATORS FOR FUNCTIONAL VALIDATION



PHENOVIZ: INTEGRATE ALL HUMAN, MOUSE, AND FISH DATA TO UNDERSTAND CNVS

Desktop application for differential diagnostics in CNVs



- Explain manifestations of CNV diseases based on genes contained in CNV

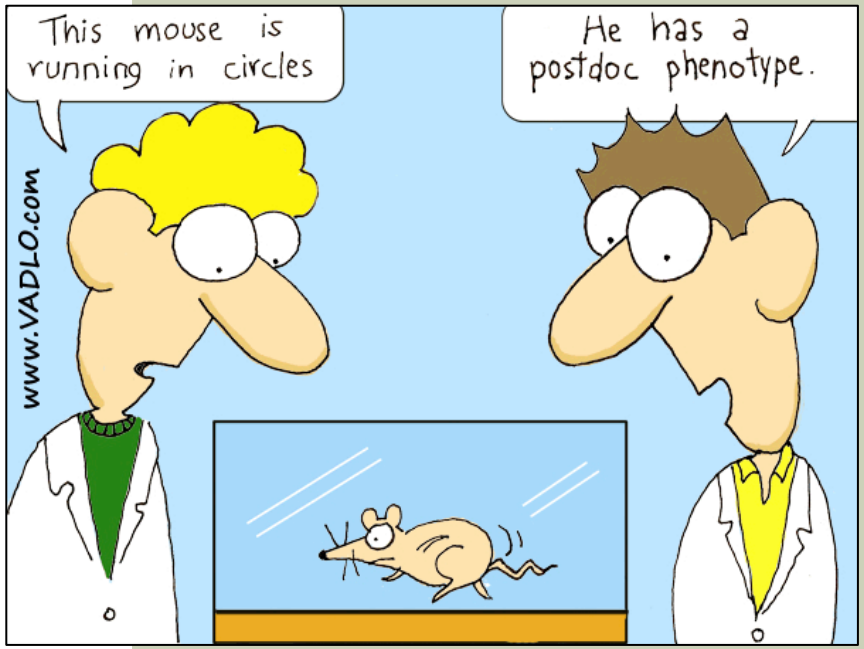
E.g., Supravalcular aortic stenosis in Williams syndrome can be explained by haploinsufficiency for elastin

- Double the number of explanations using model data

Doelken, Köhler, et al. (2013) *Dis Model Mech* **6**:358-72

A LOOK AT THE HPO

- [HP:0000708] Behavioural/Psychiatric Abnormality
  - [HP:0100851] **Abnormal emotion/affect behavior**
    - [HP:0006919] Abnormal aggressive, impulsive or violent behavior
    - [HP:0100852] Abnormal fear/anxiety-related behavior
      - [HP:0000756] Agoraphobia
      - [HP:0000739] Anxiety
      - [HP:0000712] Emotional lability
      - [HP:0001575] Mood changes
      - [HP:0000720] Mood swings
    - [HP:0012154] Anhedonia
    - [HP:0000741] Apathy
    - [HP:0000729] Autism spectrum disorder
    - [HP:0100024] Conspicuously happy disposition
    - [HP:0000716] Depression
    - [HP:0010529] Echolalia
    - [HP:0000719] Inappropriate behavior
      - [HP:0000734] Disinhibition
      - [HP:0000748] Inappropriate laughter
      - [HP:0008768] Inappropriate sexual behavior
    - [HP:0000732] Inflexible adherence to routines or rituals
    - [HP:0000737] Irritability
    - [HP:0000757] Lack of insight
    - [HP:0000745] Lack of motivation
    - [HP:0000721] Lack of spontaneous play
    - [HP:0000744] Low frustration tolerance
    - [HP:0002300] Mutism
    - [HP:0010865] Oppositional defiant disorder
    - [HP:0100025] Overfriendliness



- [HP:0100025] Overfriendliness
- [HP:0002193] Pseudobulbar behavioral symptoms
- [HP:0000711] Restlessness
- [HP:0000723] Restrictive behavior
- [HP:0100962] Shyness

WHO USES THE HPO?

Databases & Bioinformatics Resources Using HPO

DECIPHER (Sanger Institute)
DDD (Sanger Institute)
ECARUCA
FORGE (Genome Canada)
GWAS Central
IRDiRC
ISCA
NCBI Genetic Testing Registry
NIH Undiagnosed diseases program
UK 100,000 Genomes Program
RIKEN
...

Close integration with other important efforts
Major credits go to OMIM and Orphanet

monarch INITIATIVE

ISCA

FORGE CANADA CONSORTIUM

OMIM

orphanet

- Bayés, Àlex, et al. Nature neuroscience 2011
- Castellano, Sergi, et al. PNAS 2014
- Corpas, Manuel, et al. " Current Protocols in Human Genetics 2012
- Sifrim, Alejandro, et al. Nature methods 2013
- Lappalainen, Ilkka, et al. Nucleic acids research 2013
- Firth, Helen V., and Caroline F. Wright. Developmental Medicine & Child Neurology 2011
- Many more...

ADVANTAGES OF HPO

- Widely used, flexible, freely available, and community supported resource
- Prioritization of candidate variants through tools such as PhenIX and Exomizer, and others
- Extensive links to model organism ontologies, allowing selection of optimal models for wet-lab validation and research, and collaborators
- Intuitive clinical interfaces built into tools such as PhenoTips, Certagenia, and others
- Ability to easily share data with key international projects (Decipher/DDD, RD-Connect, PhenomeCentral, Matchmaker Exchange, etc.)

LIMITATIONS

- Quantitative vs. qualitative – Much of clinical data is quantitative lab data with reference standards. It is possible to convert based on ±3 SD, but no way to record the reference measure/population yet.
- Temporal presentation – ontologies can support temporal ordering, but data capture tools don't yet capture this and the comparison algorithms don't yet take it into account
- Severity – semantic encoding is available, but simple in comparison to phenotype-specific measures
- Emerging ontology – some areas have poor coverage, such as nervous system, behavior, and imaging results. Need to represent the assays in these contexts.

ACKNOWLEDGMENTS

NIH-UDP
- William Bone
- Murat Sincan
- David Adams
- Amanda Links
- David Draper
- Joie Davis
- Neal Boerkoel
- Cyndi Tifft
- Bill Gahl

OHSU
- Nicole Vasilesky
- Matt Brush
- Bryan Laraway
- Shahim Essaid

Lawrence Berkeley
- Nicole Washington
- Suzanna Lewis
- Chris Mungall

UCSD
- Amarnath Gupta
- Jeff Grethe
- Anita Bandrowski
- Maryann Martone

U of Pitt
- Chuck Boromeo
- Jeremy Espino
- Becky Boes
- Harry Hochheiser

Sanger
- Anika Oehlrich
- Jules Jacobson
- Damian Smedley

Toronto
- Marta Girdea
- Sergiu Dumitriu
- Heather Trang
- Mike Brudno

JAX
- Cynthia Smith

Charité
- Sebastian Kohler
- Sandra Doelken
- Sebastian Bauer
- Peter Robinson

Funding:

NIH Office of Director: 1R24OD011883

NIH-UDP: HHSN268201300036C, HHSN268201400093P

WHERE TO GET HPO, AND HOW TO REQUEST NEW CONTENT

We need you!

Browse in the following places:

http://www.human-phenotype-ontology.org/
http://purl.bioontology.org/ontology/HP

Get the file:

http://purl.obolibrary.org/obo/hp.owl

Request content:

https://sourceforge.net/p/obo/human-phenotype-requests/new/

More Documentation:

https://code.google.com/p/phenotype-ontologies/