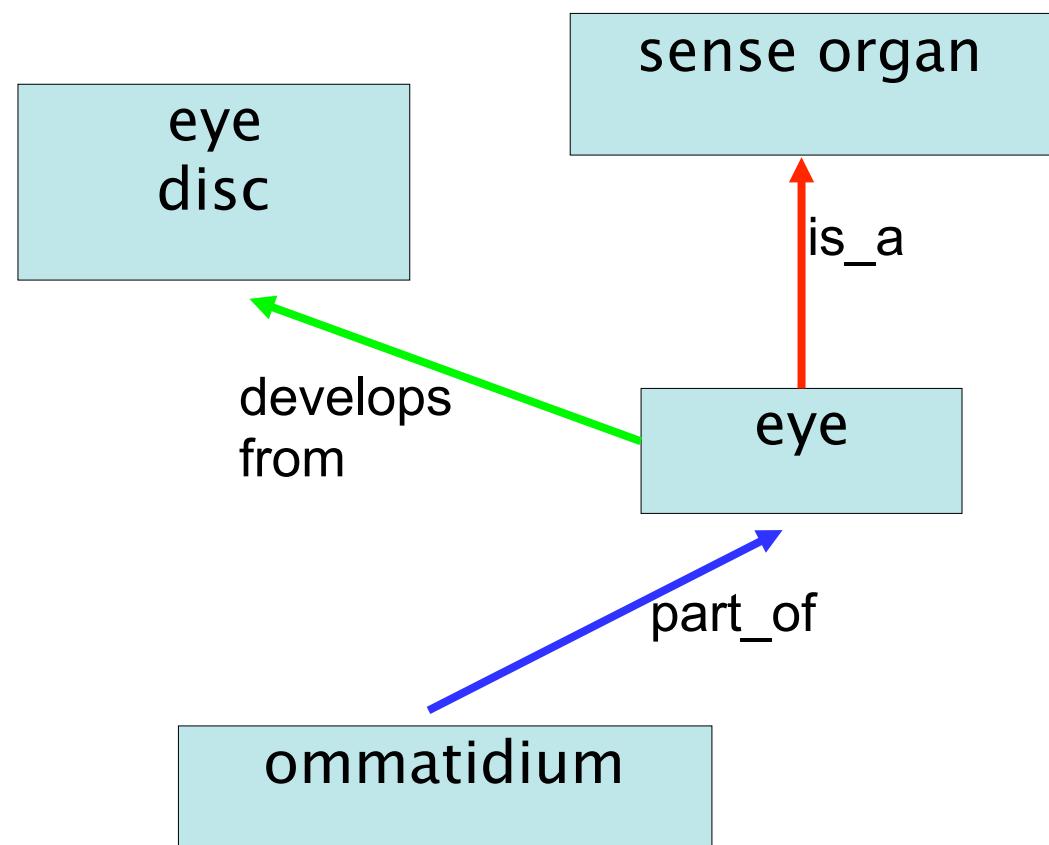
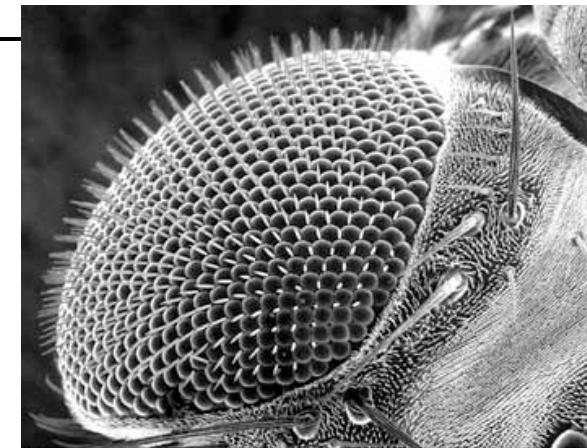


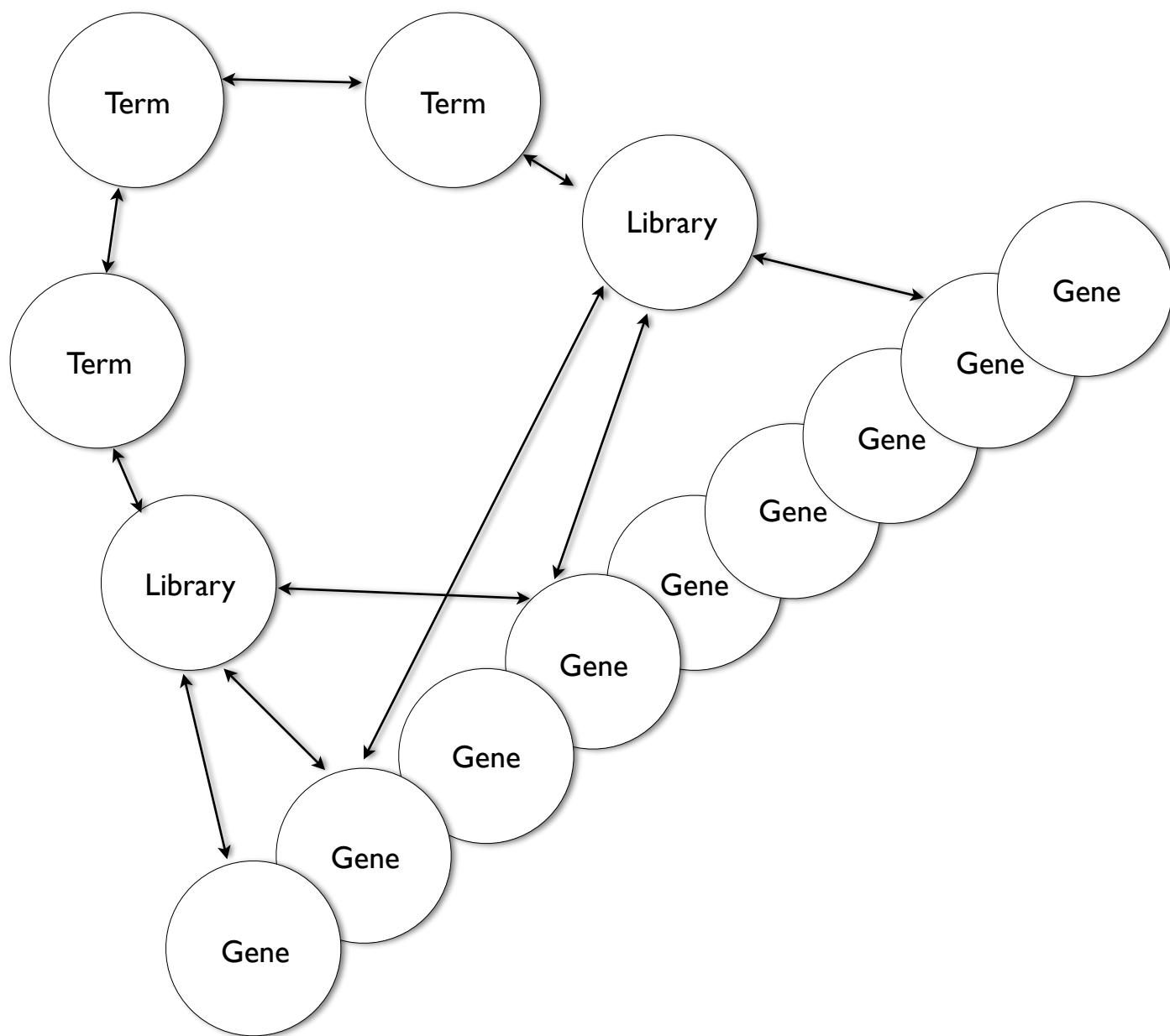
A biological ontology is:

- A precise representation of some aspect of biological reality

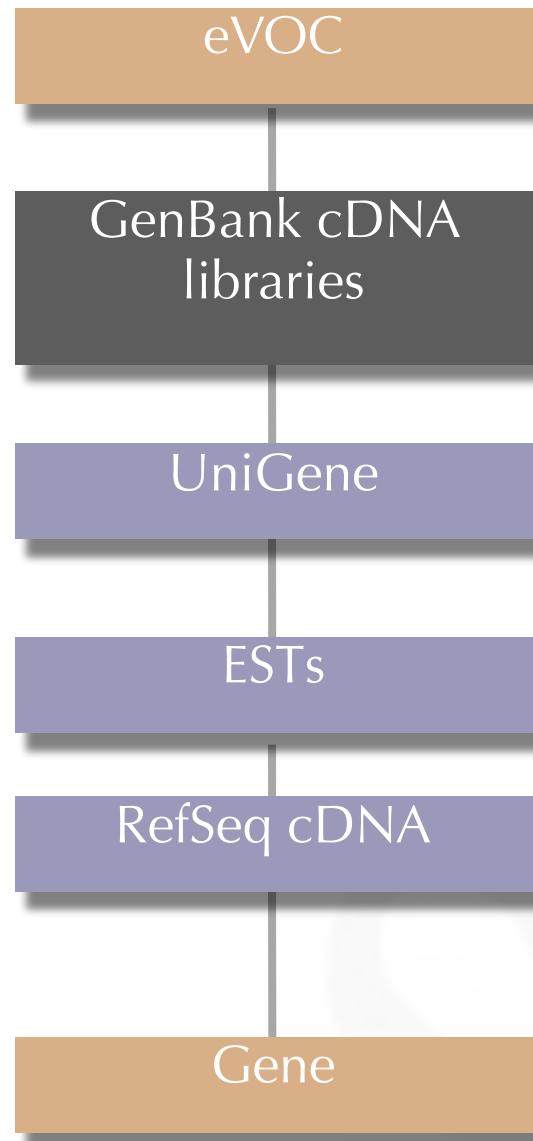
– what *kinds* of things exist?







eVOC mappings



SANBI

Prioritising disease gene candidates using gene expression

Within a region identified by classical genetic techniques:

- Examine the expression of genes in the region as a pointer to potential candidates.

Tested on the region implicated in Retinitis pigmentosa before the identification of the causative gene....

Total known genes	21 787
Genes between 8q11.23 and 8q12.1	38
Genes with expression in retina	7



GeneID	Gene Name	Description
ENSG00000168276.3	TCEA1	TRANSCRIPTION ELONGATION FACTOR A
ENSG00000120992.3	LYPLA1	LYSOPHOSPHOLIPASE I
ENSG00000104237.1	RP1	OXYGEN-REGULATED PROTEIN 1
ENSG00000147507.1	LYN	TYROSINE-PROTEIN KINASE LYN
ENSG0000008988.1	RPS20	40S RIBOSOMAL PROTEIN S20
ENSG00000169122.2	NM_147189	
ENSG00000104388.1	RAB2	RAS-RELATED PROTEIN RAB-2A

5-fold reduction in the number of potential candidates

Lori S. Sullivan, John R. Heckenlively, Sara J. Bowne, Jian Zuo, Winston A. Hide, Andreas Gal, Michael Denton, Chris F. Inglehearn, Susan H. Blanton, Stephen P. Daiger. Mutations in a novel retina-specific gene cause autosomal dominant retinitis pigmentosa Nature Genetics, 1999 Jul;22(3):255-9



SANBI

Biomedical ontologies

- SNOMed
- ICD
- Disease Ontology
- **Human Phenotype Ontology**
- OBO foundry
- NCBI
- Monarch Initiative

Monarch Initiative

Enabling navigation across a rich landscape of phenotypes, diseases, models, and genes for translational research.

Wednesday, May 27, 2015

Why the Human Phenotype Ontology?

We've often been asked, why should we use the Human Phenotype Ontology to describe patient phenotypes, rather than a more widely-used clinical vocabulary such as ICD or SNOMED? Here are the answers to some of these frequently asked questions:

1. We should use what other big NIH projects, like ClinVar, are using.

ClinVar is using HPO terms to describe phenotypes. This is done in collaboration with MedGen, which has imported HPO terms. Here is an example:

<http://www.ncbi.nlm.nih.gov/medgen/504827>

There are now many bioinformatics tools that use the HPO to empower exome diagnostics. The Monarch team has published two of these recently

- 1) Exomiser ([Robinson et al., 2014 Genome Res.](#)) => For discovering new disease genes via model organism data, several successful use cases at UDP and elsewhere
- 2) PhenIX ([Zemojtel et al., 2014 Science Translational Medicine](#)) => For clinical diagnostics of "difficult" cases. This paper was on [Russ Altman's year in review](#) at AMIA this year.

Also, a number of other groups are converging on use of the HPO, since in contrast to SNOMED, ICD, or other terminologies, it is a formal ontology that allows powerful computerized algorithms to be used. Amongst these are:

- 1) Stephen Kingsmore (Kansas City) now is using the Phenomizer to prioritize genes (currently unpublished, but see pmid: 23035047 for Stephen's previous work)
- 2) The Yandell group in Utah: Singleton MV et al; Phevar combines multiple biomedical ontologies for accurate identification of disease-causing alleles in single individuals and small nuclear families. Am J Hum Genet. 2014 Apr 3;94(4):599-610.
- 3) The Moreau group in Leuven: Sifrim A, et al; eXtasy: variant prioritization by genomic data fusion. Nat Methods. 2013 Nov;10(11):1083-4.
- 4) The Children's hospital of Philadelphia group around Jeff Pennington (BMC Genomics, accepted)
- 5) Lévesque group at McGill University Health Centre. Trakadis et al. BMC Med Genomics. 2014 May 12;7:22. doi: 10.1186/1755-8794-7-22. PhenoVar: a phenotype-driven approach in clinical genomics for the diagnosis of polymalformative syndromes.

The above 5 citations are merely the use of HPO for prioritization of genes in NGS studies, there are numerous other bioinformatics and genetics databases using the HPO, see the HPO [homepage](#) for a partial list.

Contributors

-  [Nicole](#)
-  [Nicole Vasilevsky](#)
-  [Chris Mungall](#)
-  [Monarch Initiative](#)
-  [Harry Hochheiser](#)
-  [Peter Robinson](#)
-  [Tudor Groza](#)
-  [melissa haendel](#)
-  [Julie McMurry](#)

Search Blog and Links

Blog Archive

- ▼ 2015 (4)
 - June (1)
 - ▼ May (1)
 - [Why the Human Phenotype Ontology?](#)
 - April (1)
 - January (1)
- 2014 (8)

THE APPLICATION OF THE HUMAN PHENOTYPE ONTOLOGY

Melissa
Haendel
Oregon University

OUTLINE

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



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



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

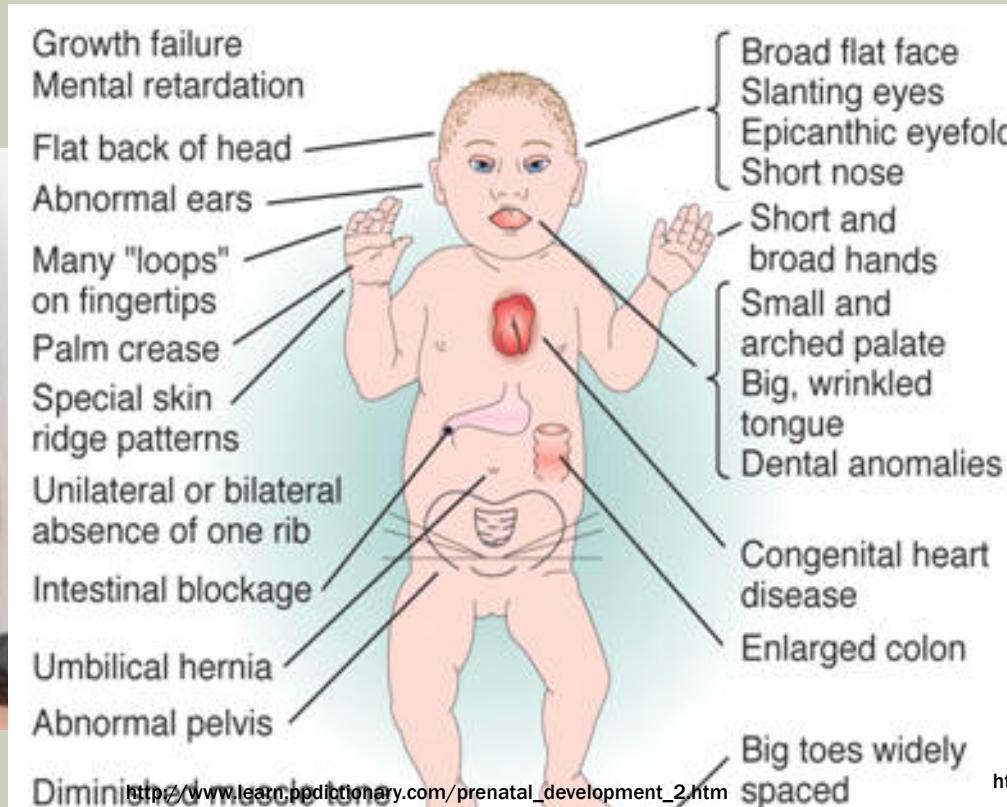


<http://www.theguardian.com/commentisfree/2009/oct/27/downs-syndrome-increase-terminations>

THE CONSTELLATION OF PHENOTYPES SIGNIFIES THE DISEASE – A ‘PROFILE’



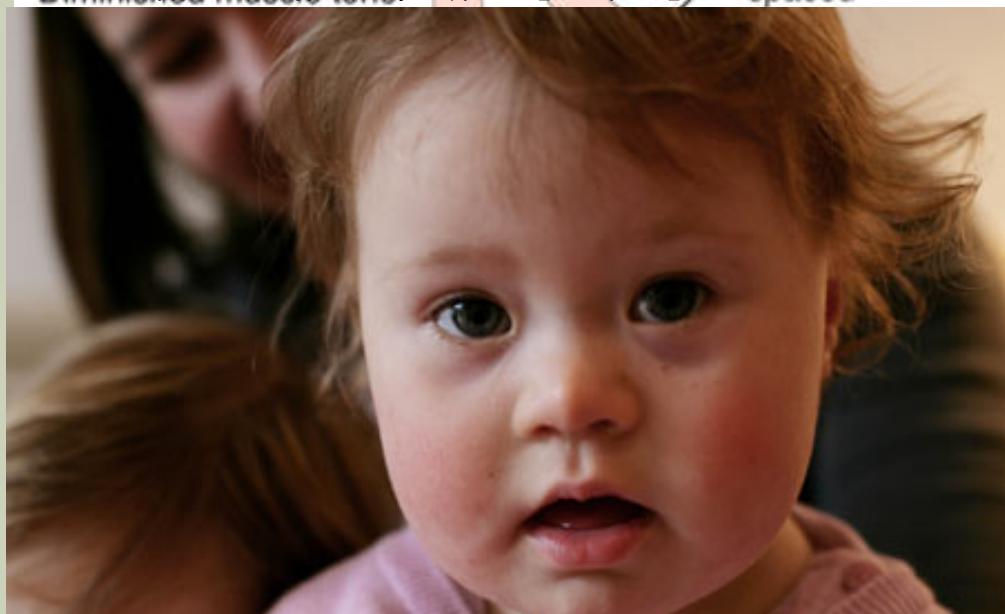
<http://www.pyroenergen.com/articles07/downs-syndrome.htm>



http://www.learn.podictionary.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
- congenital anomaly
- congenital anomaly
- cong. m.
- cong. Mal
- cong. malfor
- congenital malform
- congenital m.
- multiple congenital anomalies
- multiple congenital abnormalities
- multiple congenital abnormalities

Examples of lists:

- * dd. cong. malfor. behav. pro.
- * dd. mental retardation
- * df< delayed puberty
- * df<
- * 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: _____

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: _____

Neurological

- Hypotonia
- Seizures
- Ataxia
- Dystonia
- Chorea

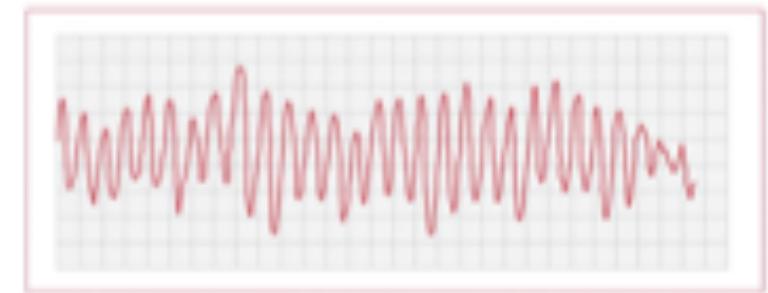
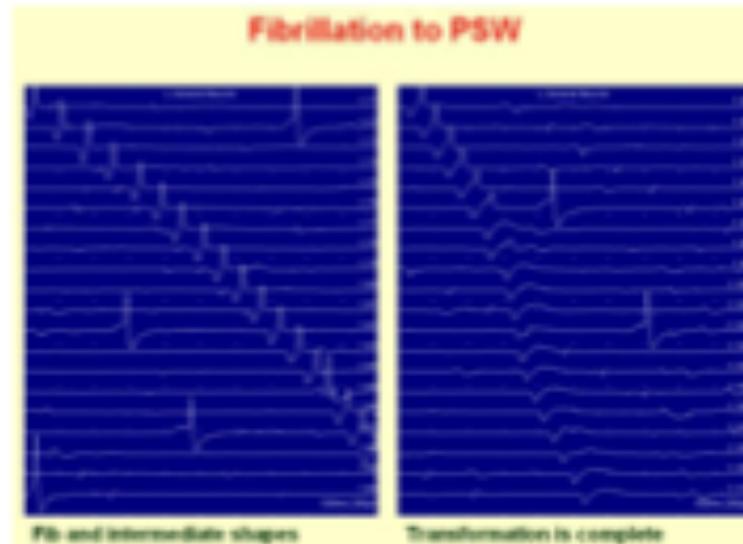
Eye Defects

- Blindness
- Coloboma
- Epicanthus
- Eyelid abnormality (Specify below)
- 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

TERMS SHOULD BE WELL DEFINED SO THEY GET USED PROPERLY



fibrillation . . .

fibrillation . . .

muscle fibrillation = fibrillation ≠ 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.

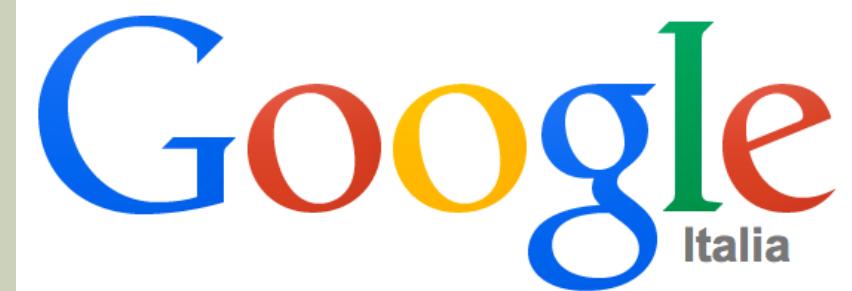
A controlled vocabulary of logically defined, inter-related terms used to annotate data

- 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

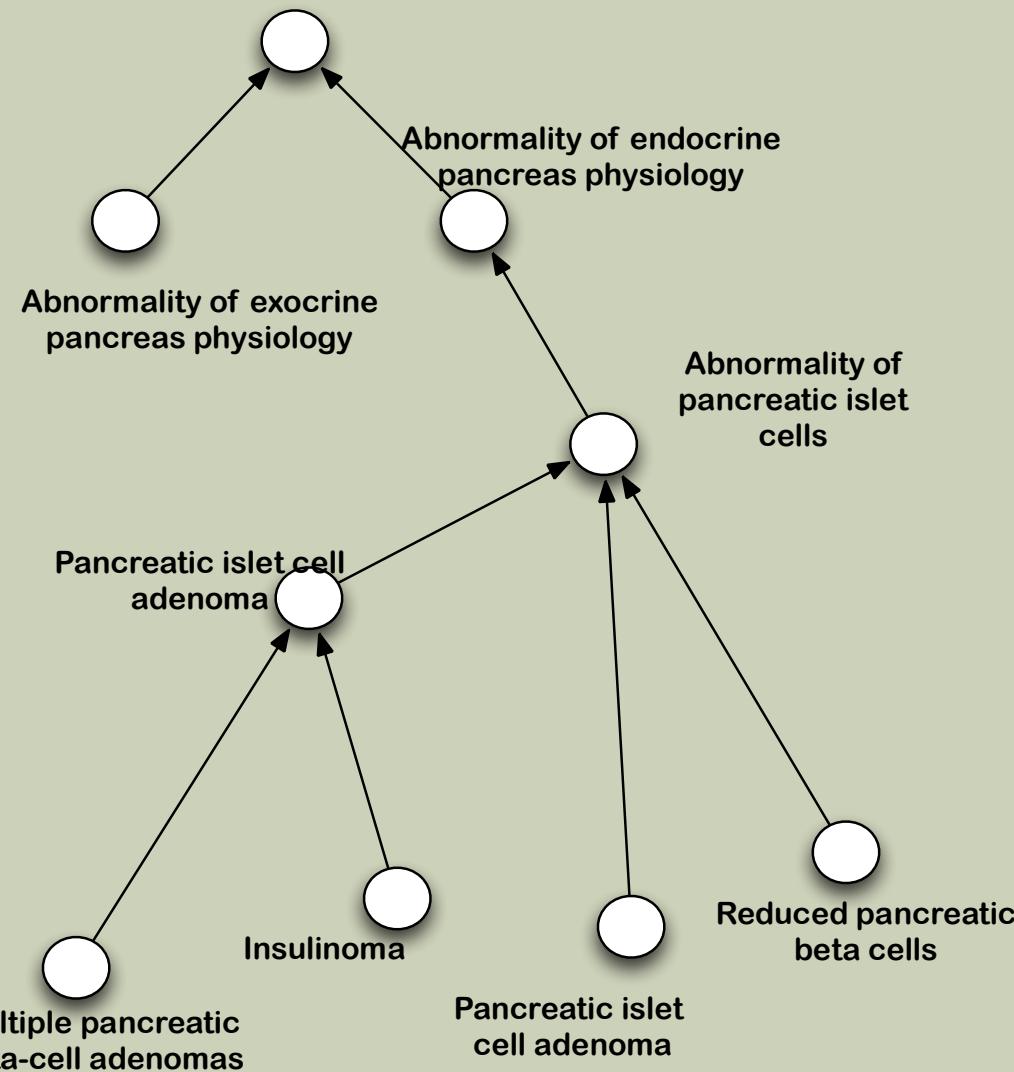
OTHER COMMON USES OF ONTOLOGIES



Siri.



HUMAN PHENOTYPE ONTOLOGY

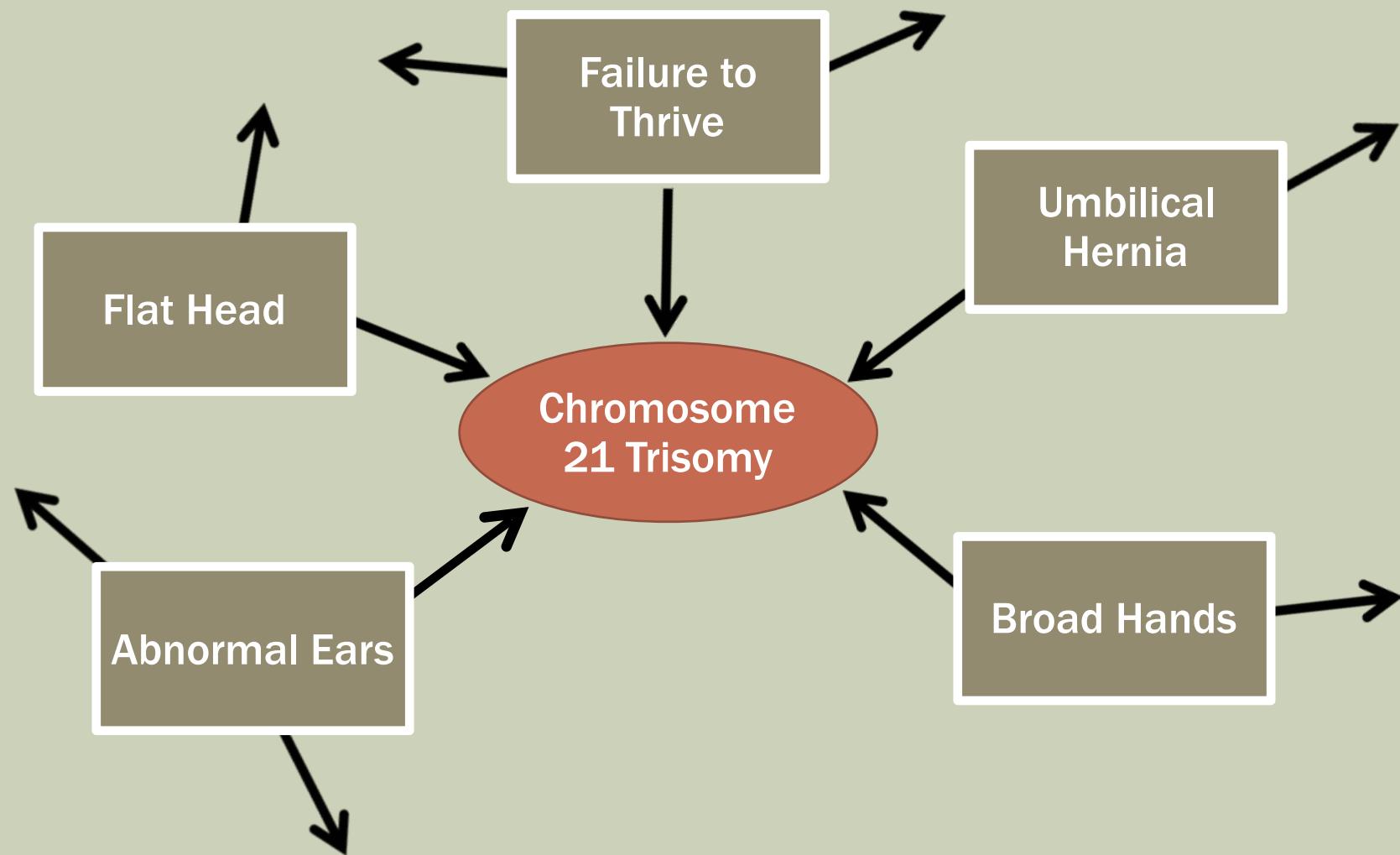


Used to annotate:

- Patients
- Disorders
- Genotypes
- Genes
- Sequence variants

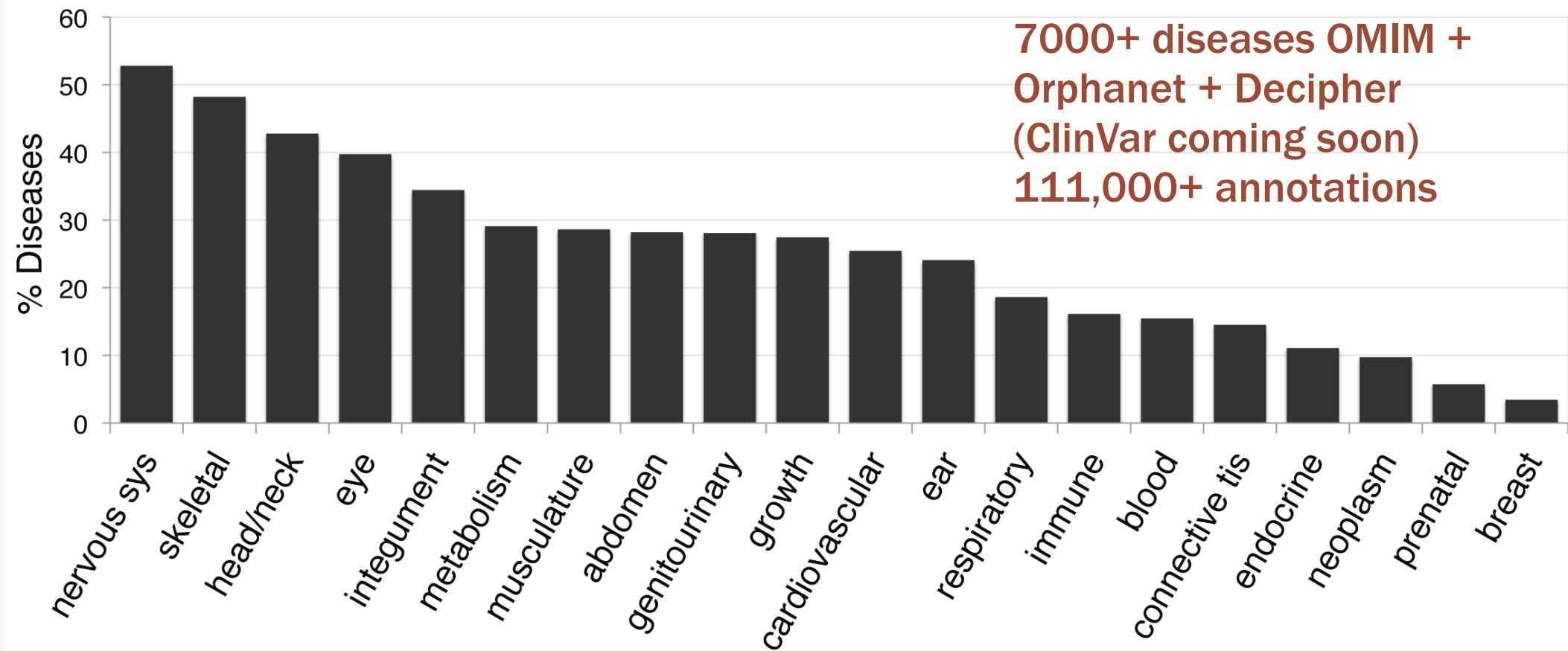
Mappings to SNOMED-CT,
UMLS, MeSH, ICD, etc.

USING A CONTROLLED VOCABULARY TO LINK PHENOTYPES TO DISEASES



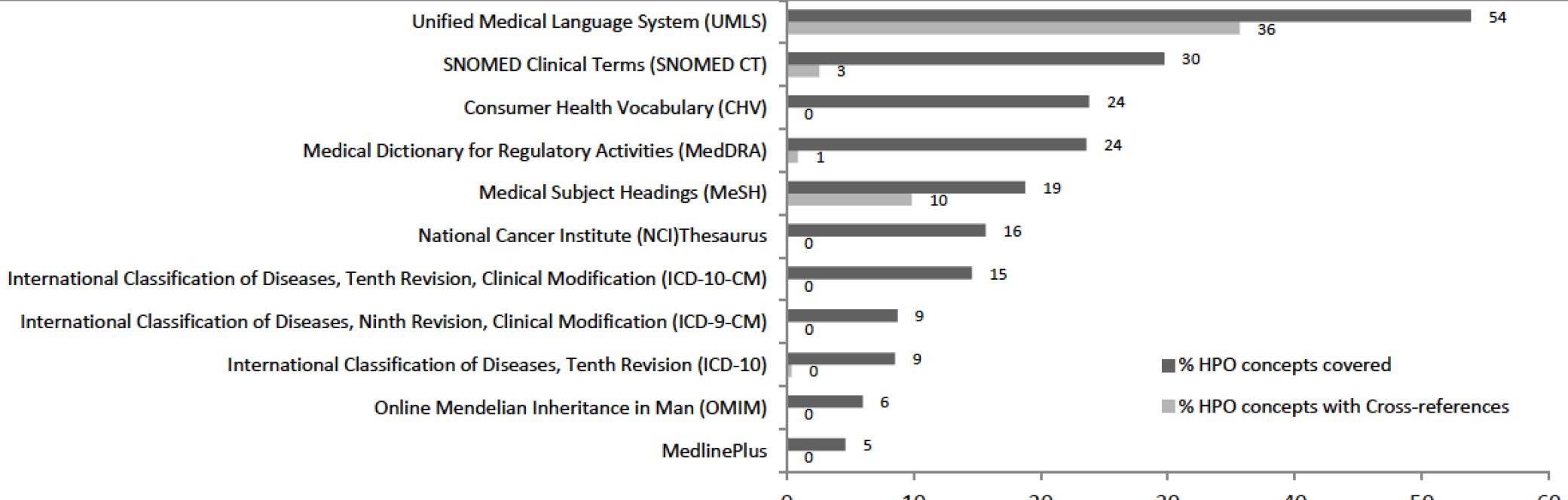
SURVEY OF ANNOTATIONS IN DISEASE CORPUS

Coverage by Phenotype Category



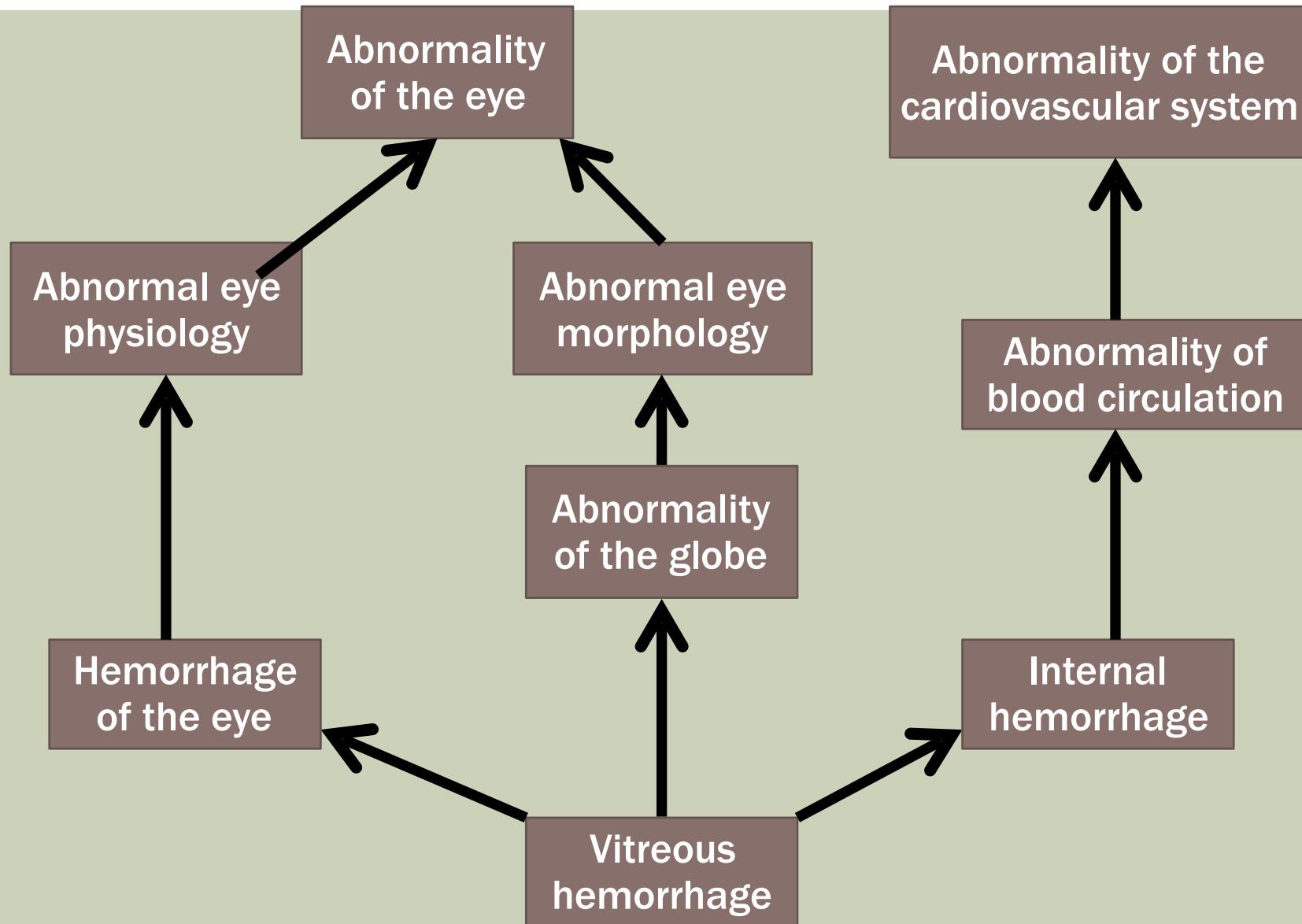
Phenotype annotations are unevenly distributed across different anatomical systems

HOW DOES HPO RELATE TO OTHER CLINICAL VOCABULARIES?

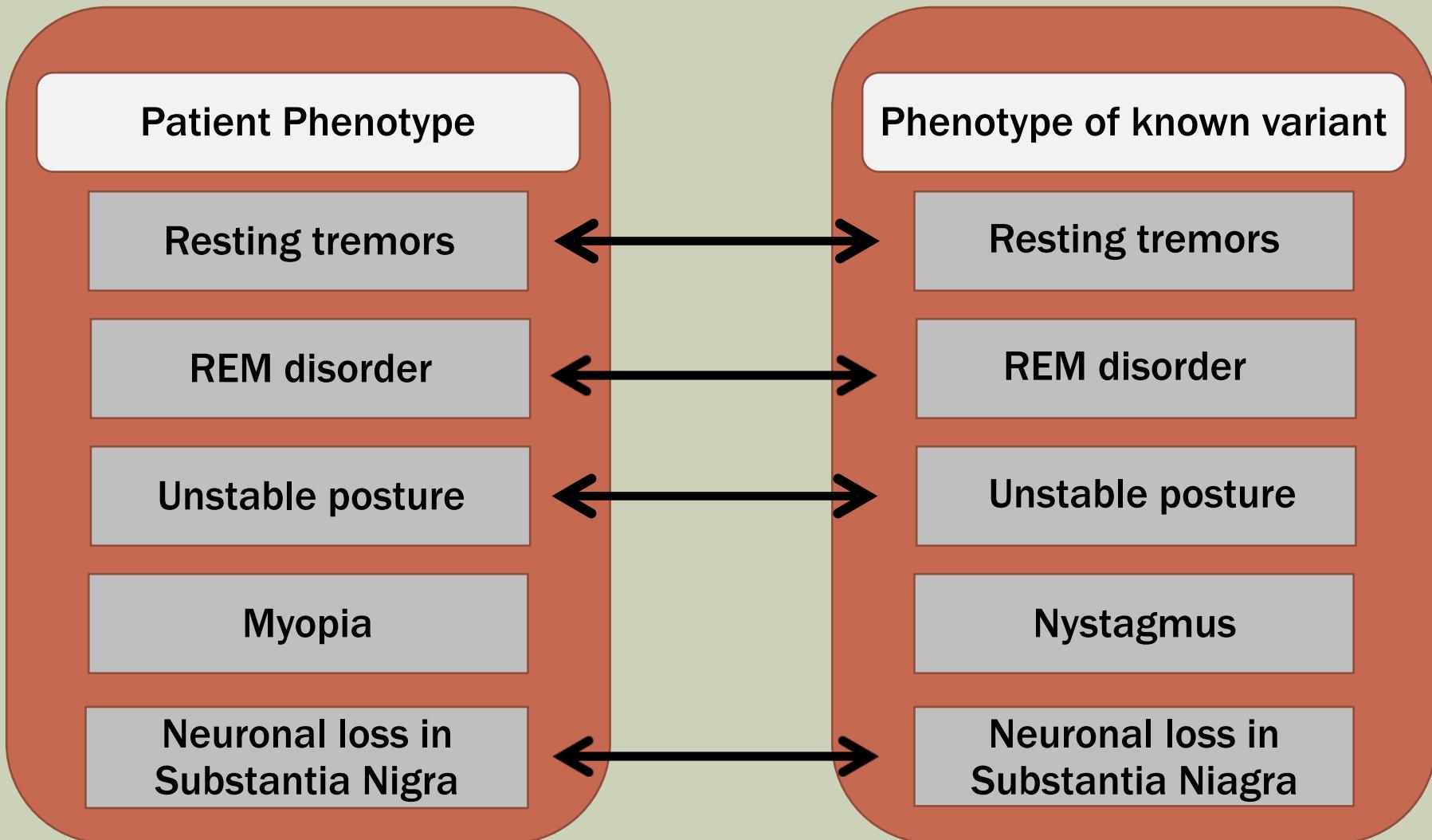


Winnenburg and Bodenreider, ISMB PhenoDay, 2014

PHENOTYPES CAN BE CLASSIFIED IN MULTIPLE WAYS



PHENOTYPE MATCHING



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?

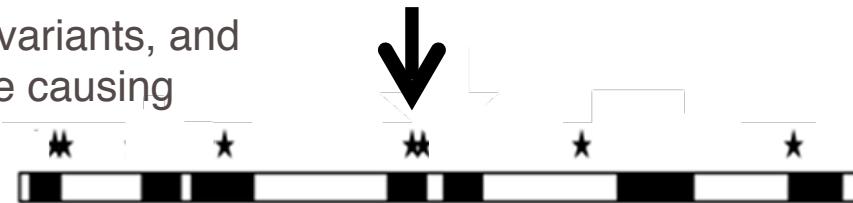
PhenIX

EXOME ANALYSIS



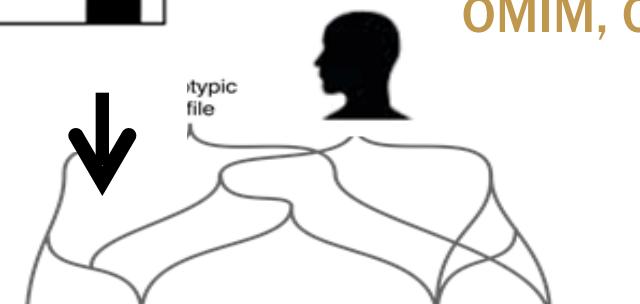
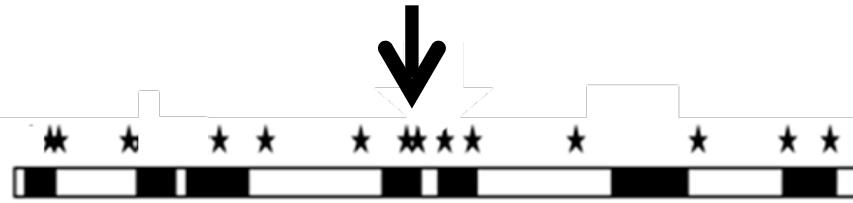
Target panel of 2741 known
Mendelian disease genes

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



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

Recessive, de novo filters

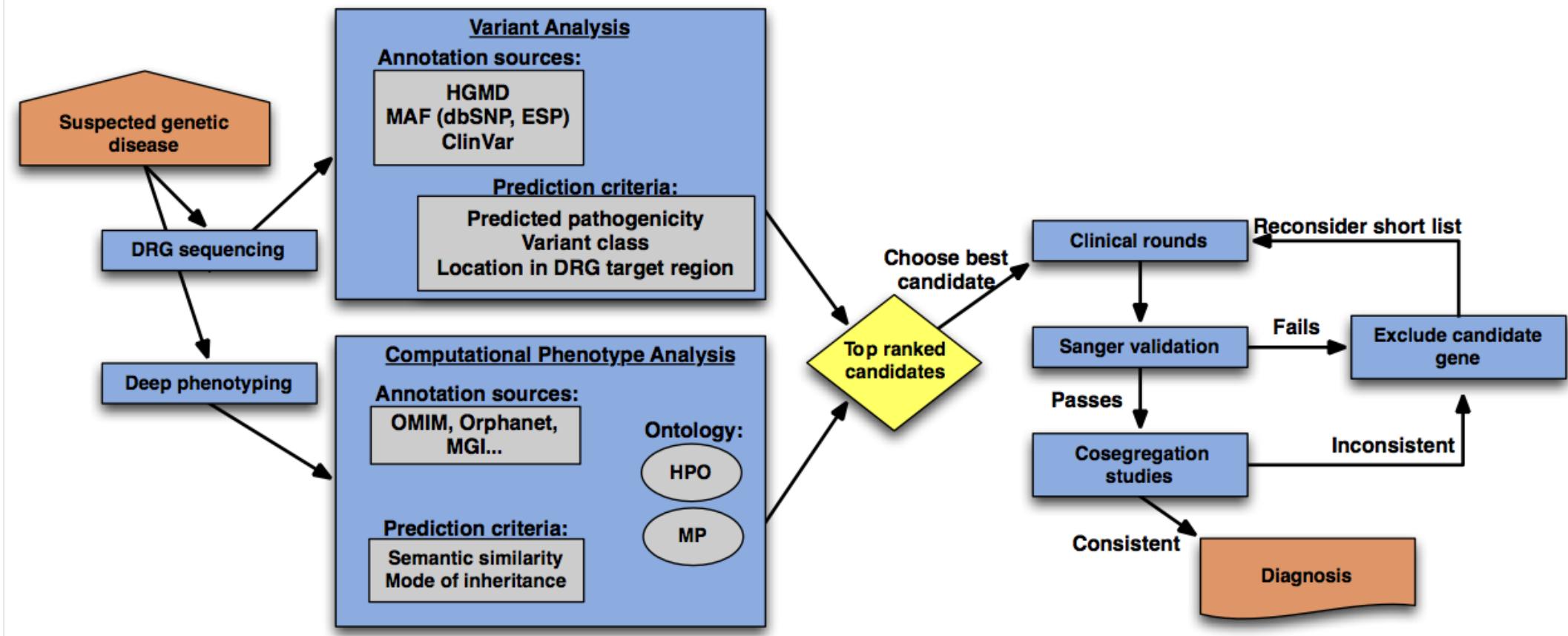


Zemojtel et al. Sci Transl Med 3 September 2014: Vol. 6,
Issue 252, p.252ra123

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	<i>MLL</i>	2	Wiedemann-Steiner syndrome (39)
P2	5y (f)	global developmental delay (HP:0001263)			
P3	6y (f)	delayed speech and language development (HP:0000750) motor delay (HP:0001270)			
P4	Death (f)	proportionate short stature (HP:0003508) microcephaly (HP:0000252)			
P5	6m (f)	feeding difficulties (HP:0011968)			
P6	Fetus (m)	congenital megaloureter (HP:0008676) cone-shaped epiphysis of the phalanges of the hand (HP:0010230)			
	Death gestat	sacral dimple (HP:0000960) hyperpigmented/hypopigmented macules (HP:0007441)			
P7	7y (m)	hypertelorism (HP:0000316)			
P8	14y (r)	abnormality of the midface (HP:0000309) flat nose (HP:0000457)			
P9	6y (f)	thick lower lip vermillion (HP:0000179)			
P10	4 between and 7y	thick upper lip vermillion (HP:0000215) full cheeks (HP:0000293)			
P11	3y (m)	short neck (HP:0000470) multiple congenital anomalies			

The Skeletome Knowledge Base

A community-driven knowledge curation platform for skeletal dysplasias.

[Take The Tour >](#)

Feedback

Comprehensive

Everything You Ever Wanted to Know About Bone Dysplasias

The Skeletome knowledge base provides information on all bone dysplasias recognised by the International Skeletal Dysplasia Society.

Community driven

Continuously Updated by the Global Bone Dysplasia Community

All entries are continuously reviewed and updated by the global community of clinicians and researchers working on bone dysplasias.

Ontology Based

Readable by Humans and Computers - the Best of Both Worlds

The Skeletome knowledge base makes extensive use of ontologies to standardise the entered information and make it accessible to computational analysis.

SKELETOME PATIENT ARCHIVE

ARCHIVE Home Patients Groups Case Finder User 1 Logout

Patients / #8 - John Doe

Summary Patient Details Sharing X-Rays Genetic Reports Clinical Summary Diagnoses Discussion

Edit

Find a clinical summary



Jul 11, 2014 (2 months ago)

saw patient again in clinic today. he now complains of hearing loss and poor vision.

Impaired vision X Hearing impairment X +

Dec 2, 2013 (9 months ago)

this is a 3 year old with cleft palate, bowed legs, and short fingers. He has complained of difficulties walking since the age of 3 years.

Bowing of the legs X Hypoplastic/small fingers X Cleft palate X Difficulty walking X +

Sufficiency Score

★★★★★

Bowing of the legs

1 Record

Cleft palate

1 Record

Difficulty walking

1 Record

Hearing impairment

1 Record

Hypoplastic/small fingers

1 Record

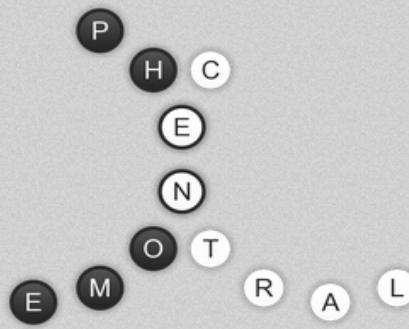
Impaired vision

1 Record

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

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

QUICK PHENOTYPE SEARCH: 🔍

CURRENT SELECTION

BEHAVIOR, COGNITION AND DEVELOPMENT

- Delayed gross motor development Delete · Add details
- Intellectual disability, moderate Delete · Add details
- NO Attention deficit hyperactivity disorder** Delete · Add details

NEUROLOGICAL

- Spasticity Delete · Add details
- NO Spinal dysraphism** Delete · Add details

CARDIAC

Defect in the atrial septum Delete · Clear details

Age of onset:

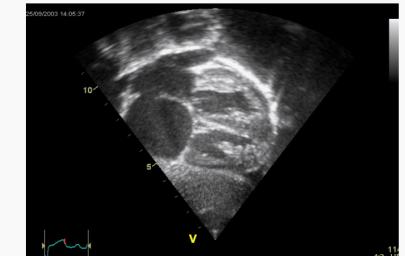
- Unknown
- Childhood onset
- Juvenile onset
- Adult onset
- Young adult onset
- Middle age onset
- Late onset

Pace of progression:

- Unknown
- Progressive disorder
- Nonprogressive disorder
- Slow progression

Comments:
No complications

Image / photo (optional): + UPLOAD AND MANAGE



Medical report (optional): None available + UPLOAD AND MANAGE

CRANIOFACIAL

RESPIRATORY

Other (enter free text and choose among suggested ontology terms)

GROWTH PARAMETERS

Weight for age

- <3rd
- >97th

Stature for age

- <3rd
- >97th

Head circumference for age

- <3rd
- >97th

Hemihypertrophy

CARDIAC

- Defect in the atrial septum
- Ventricular septal defect
- Complete atrioventricular canal defect
- Coarctation of aorta
- Tetralogy of Fallot
- Cardiomyopathy
- Arrhythmia

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 vermillion
- 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	
CURRENT PATIENT'S FEATURES	Expressive language delay
OTHER PATIENT'S FEATURES	Delayed speech and language development
THIN VERNILION BORDER	
CURRENT PATIENT'S FEATURES	Thin vermillion border
OTHER PATIENT'S FEATURES	Thin upper lip vermillion
POSTERIORLY ROTATED EARS	
CURRENT PATIENT'S FEATURES	Posteriorly rotated ears
OTHER PATIENT'S FEATURES	Low-set, posteriorly rotated ears
SHORT STATURE	
CURRENT PATIENT'S FEATURES	Short stature
OTHER PATIENT'S FEATURES	Severe short stature
ABNORMALITY OF THE EYELID	
CURRENT PATIENT'S FEATURES	Long eyelashes
OTHER PATIENT'S FEATURES	Blepharophimosis
GROWTH ABNORMALITY	
CURRENT PATIENT'S FEATURES	Growth delay
OTHER PATIENT'S FEATURES	Decreased body weight
ABNORMALITY OF THE FACE	
CURRENT PATIENT'S FEATURES	Triangular face Prominent nose
OTHER PATIENT'S FEATURES	Dental malocclusion Wide mouth Microdontia

GENE MATCHING BREAKDOWN

SRCAP

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	VARIANT	ESTIMATED HARMFULLNESS
chr16:30748691-30748691 C → T (STOPGAIN)	■■■■■ 95%	chr16:30751917-30751917 G → A (FS_INSERTION)	■■■■■ 95%
chr16:30697203-30697203 G → C (NONSYNONYMOUS)	■■■□□ 57%		

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

The screenshot shows a web browser window for the PhenomeCentral platform. The URL in the address bar is phenomecentral.org/data/F0000024. The main sidebar on the left lists various medical categories: EAR DEFECTS (Deafness, Sensorineural), BEHAVIOR, COGNITION AND DEVELOPMENT (Delayed fine motor development, Delayed gross motor development, Intellectual disability, Mild, Attention deficit hyperactivity disorder), NEUROLOGICAL (Generalized hypotonia, Absent Achilles reflex, Reduced tendon reflexes, Sensory neuropathy, Autonomic dysregulation), and OTHER (Neonatal hypotonia). Below this, a section titled "Similar cases available" shows 10 similar cases, each with a lock icon and the identifier "Undisclosed identifier".

The main content area is a modal dialog titled "Contact a non-public case owner". It is divided into two sections:

- ① Configure your message**:
 - SUBJECT**: Interested in one of your non-public cases
 - Information about you:**
 - DISCLOSE YOUR NAME
 - DISCLOSE YOUR EMAIL
 - DISCLOSE YOUR MEMBERSHIP TO PHENOMECENTRAL GROUPS
 - Information about your case (F0000024):**
 - INCLUDE DIAGNOSIS INFORMATION
 - INCLUDE A PHENOTYPE SUMMARY
 - Your requests:**
 - REQUEST MUTUAL VIEW ACCESS TO THE TWO SIMILAR CASES
 - REQUEST CONTACT INFORMATION
 - OTHER INFORMATION TO INCLUDE IN YOUR MESSAGE**: A large text input field.
- ② Preview your message**:
 - SUBJECT**: [PhenomeCentral] Interested in one of your non-public cases
 - MESSAGE**:

Hello <undisclosed recipient name>,
A PhenomeCentral user is interested in one of your non-public cases: <undisclosed case identifier>. Please see their message below.

PhenomeCentral has identified significant similarities between one of your cases and one of mine.

My patient is undiagnosed presents the following phenotypic features:

 - Absent Achilles reflex
 - Anosmia
 - Attention deficit hyperactivity disorder
 - Autonomic dysregulation
 - Decreased corneal reflex
 - Delayed fine motor development
 - Delayed gross motor development
 - Generalized hypotonia
 - Intellectual disability, mild
 - Neonatal hypotonia
 - Reduced tendon reflexes
 - Sensorineural hearing impairment
 - Sensory neuropathy

I would like to grant you the rights to view my case and to obtain view access to your case, and to learn your contact information in order to further discuss these abnormalities with you.

Regards,
Marta Girdea
marta@phenotips.org

To accept view privileges from this user and to grant them view access to <undisclosed case identifier>, follow this link:
<undisclosed URL>.

Best wishes,
The PhenomeCentral team

OUTLINE

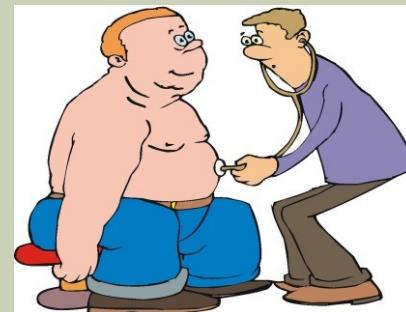
- Why phenotyping is hard
- About Ontologies
- Diagnosing known diseases
- Getting the phenotype data
- How much phenotyping is enough?
- Model organism data for undiagnosed diseases

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



MODELS RECAPITULATE VARIOUS PHENOTYPIC ASPECTS OF DISEASE

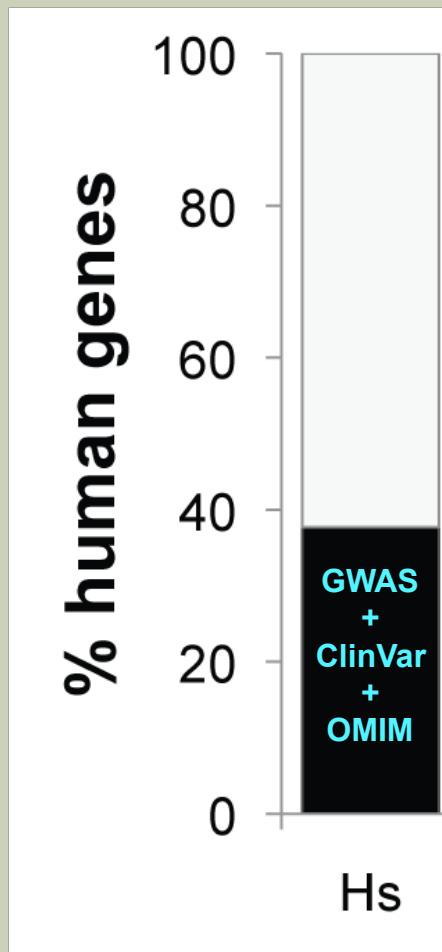
GENOTYPE

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

PHENOTYPE

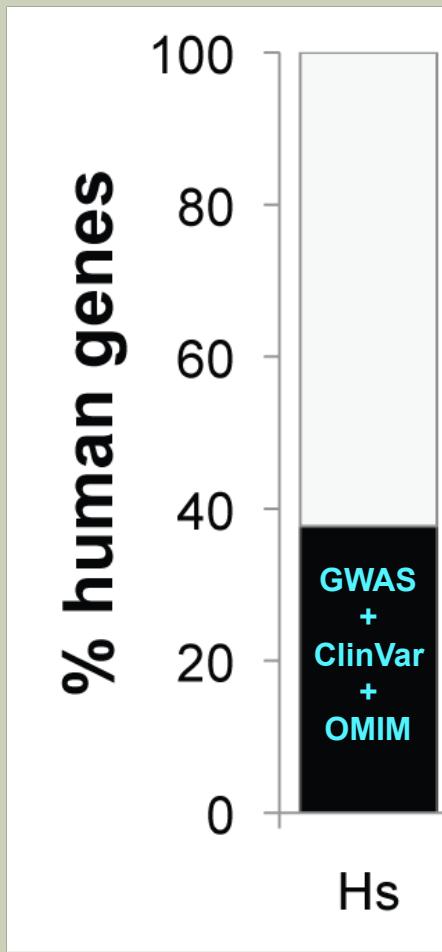
HOW MUCH PHENOTYPE DATA?

- ◆ Human genes have poor phenotype coverage



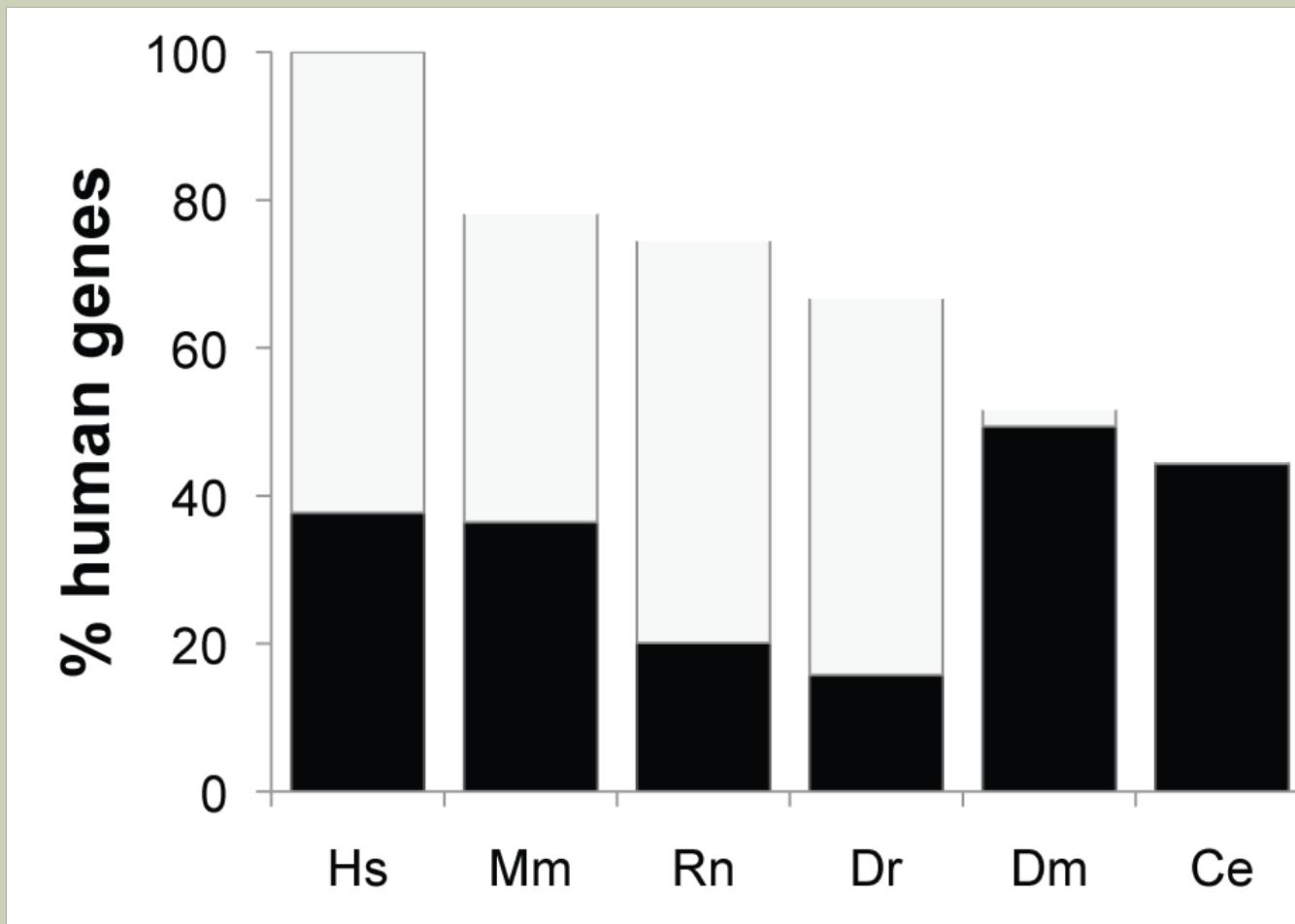
HOW MUCH PHENOTYPE DATA?

- ◆ Human genes have poor phenotype coverage
What else can we leverage?

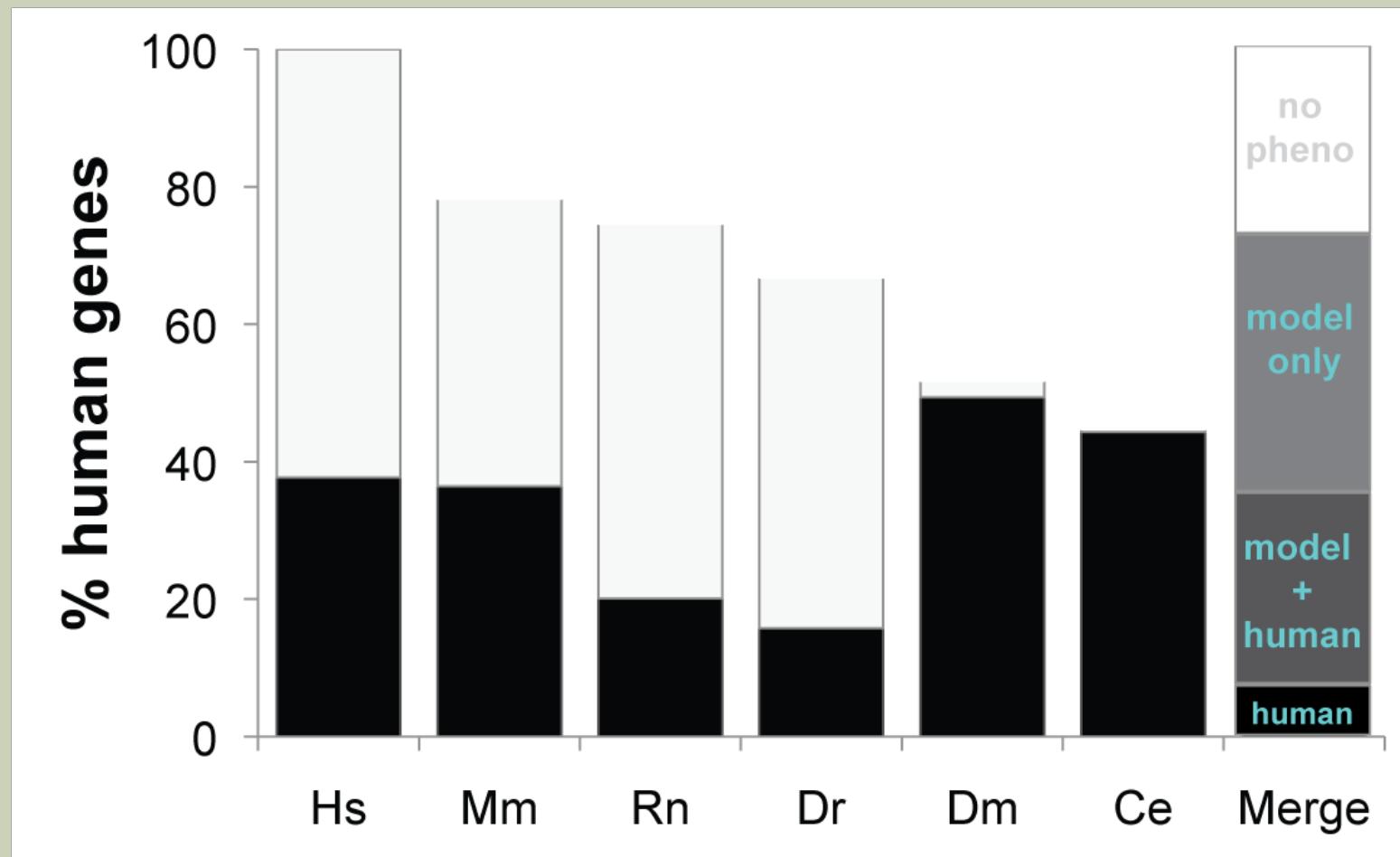


HOW MUCH PHENOTYPE DATA?

- ◆ Human genes have poor phenotype coverage
What else can we leverage? ...*animal models*

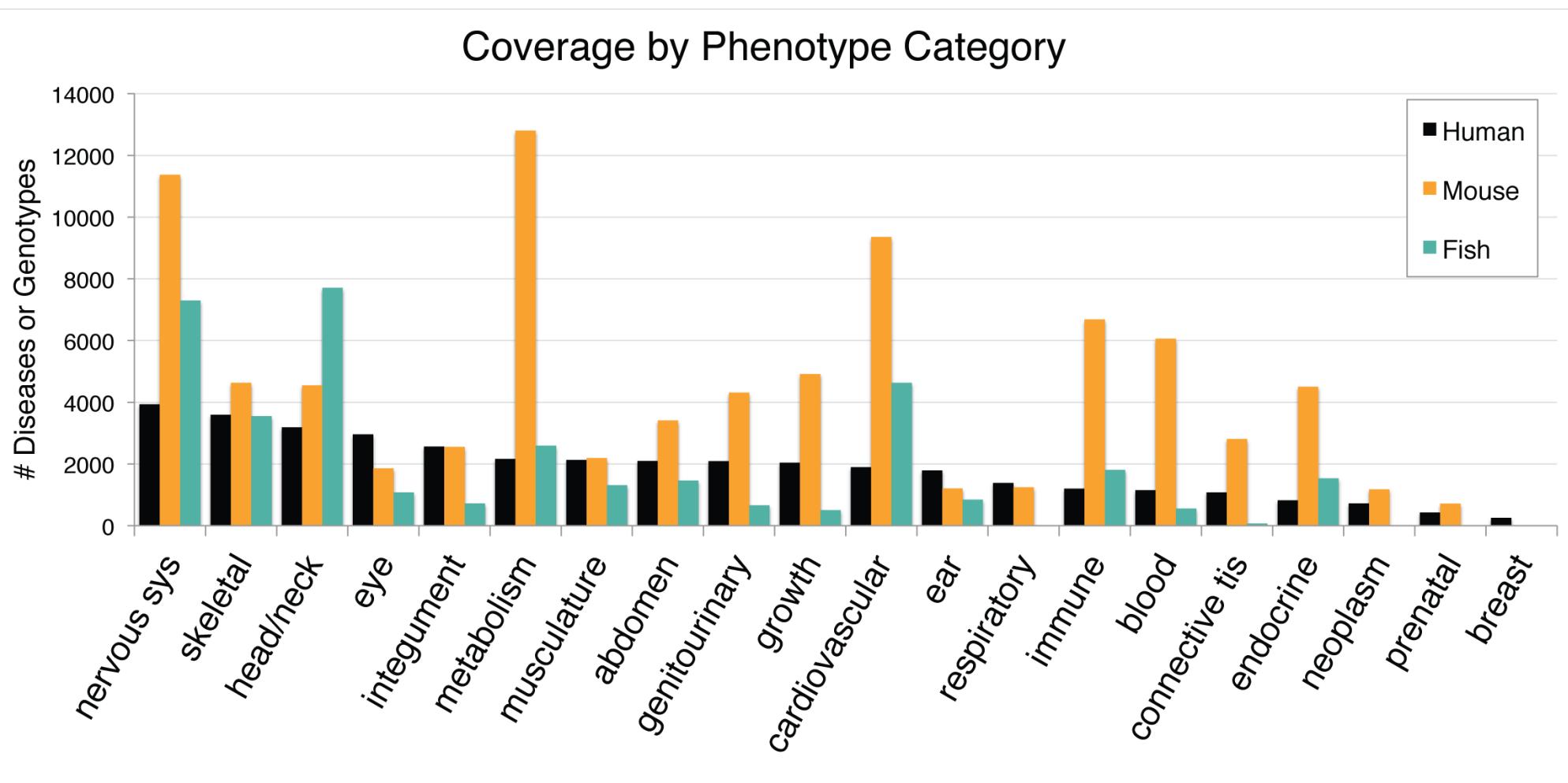


COMBINED, HUMAN AND MODEL PHENOTYPES CAN BE LINKED TO >75% HUMAN GENES



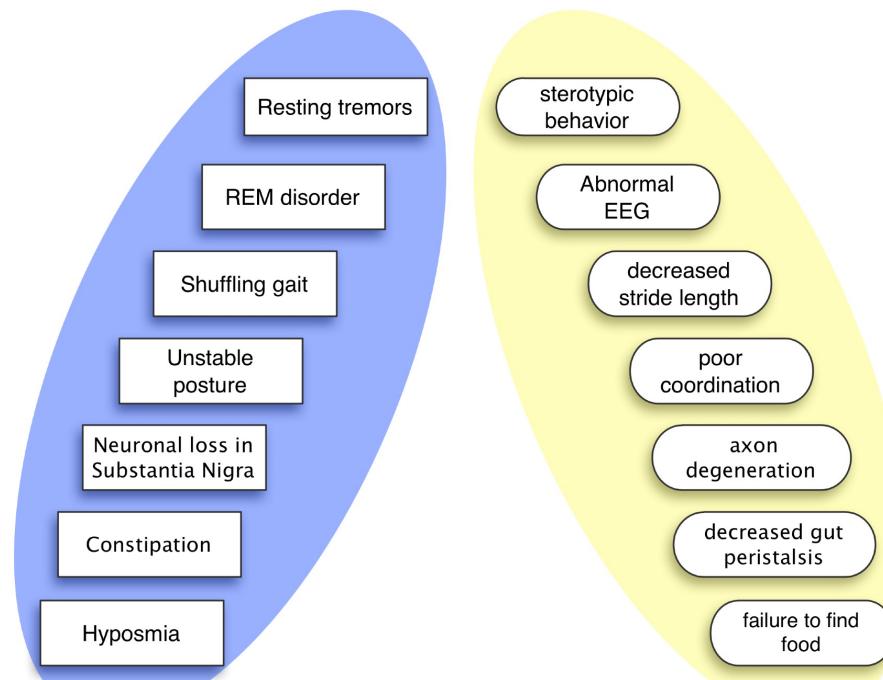
Orthology via PANTHER v9

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

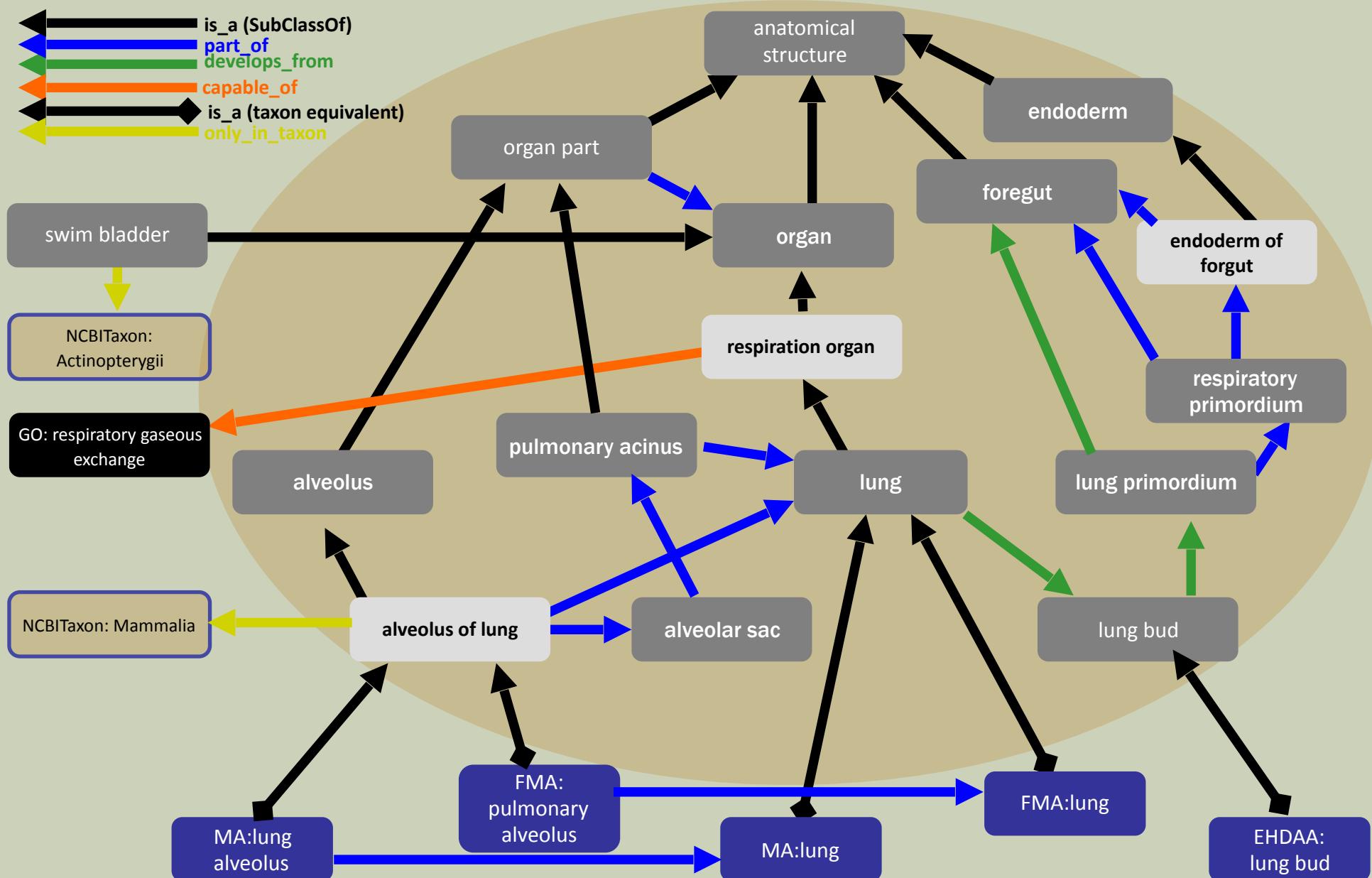


Clinician:
**Parkinson's
disease, late onset**



Researcher:
**Candidate
model system?**

SOLUTION: BRIDGING SEMANTICS



PHENOTYPE REPRESENTATION REQUIRES MORE THAN “PHENOTYPE ONTOLOGIES”

Disease

type II
diabetes
mellitus
(DOID:9352)

Cell

pancreatic
beta cell
(CL:0000169)

Gene Ontology

glucose
metabolism
(GO:0006006)

Chemical

pyruvate
(CHEBI:
15361)

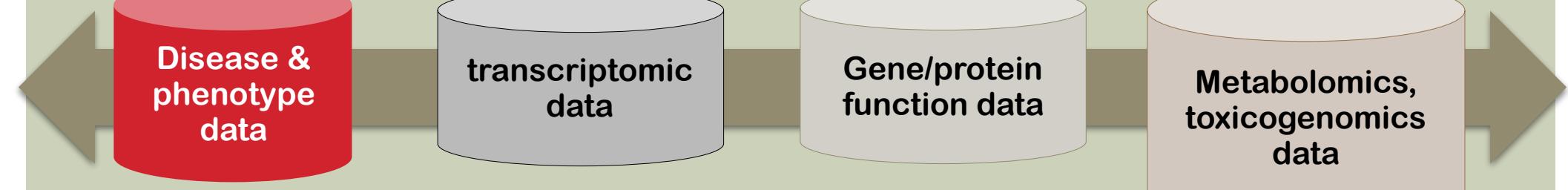
glucose
(CHEBI:
17234)

Disease &
phenotype
data

transcriptomic
data

Gene/protein
function data

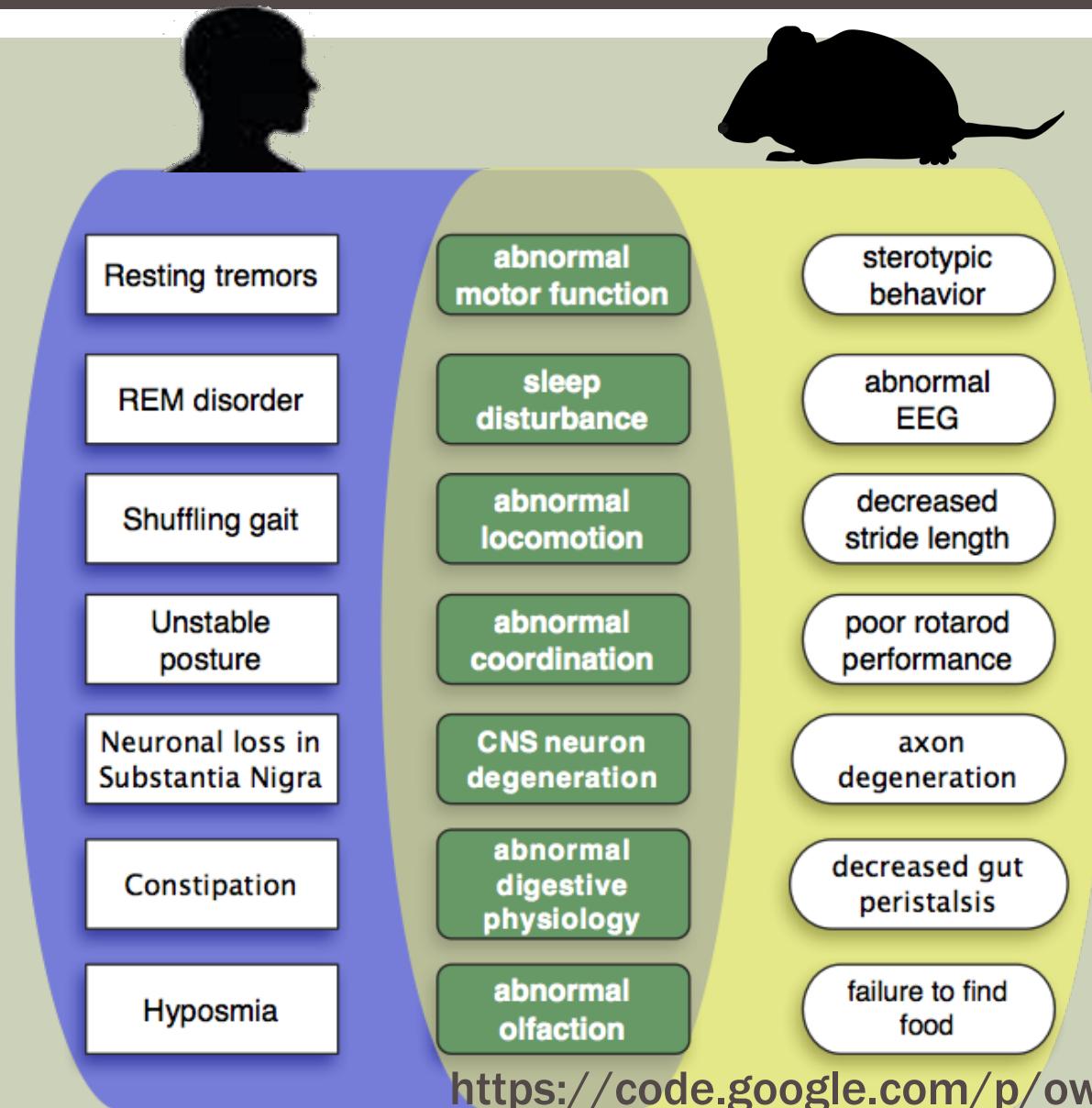
Metabolomics,
toxicogenomics
data



MODELS BASED ON PHENOTYPIC SIMILARITY

	Human	Mouse	Zebrafish	Drosophila
WT				
mut				
<i>PAX6^{+/−}</i>				
EQs	cornea opaque iris absent retina degenerate lens opaque aqueous humor of eyeball increased pressure	eye decreased size lens fused_to cornea iris morphology anterior chamber absent	eye decreased size lens decreased size retina malformed	eye absent

OWLSIM: PHENOTYPE SIMILARITY ACROSS PATIENTS OR ORGANISMS



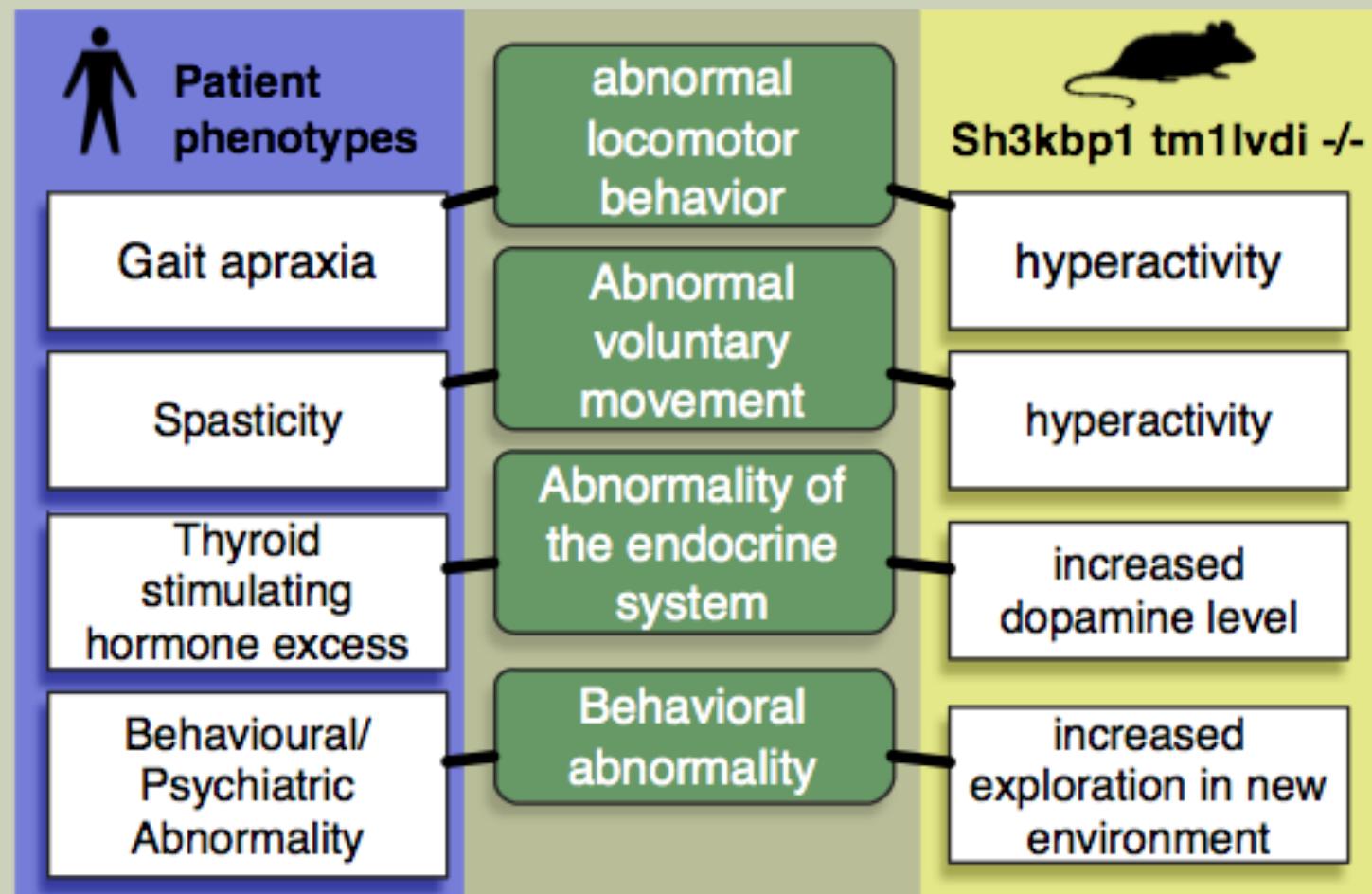
MONARCH PHENOTYPE DATA

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	
human	OMIM	2,970			4,437	
human	ClinVar	3,215		100,523	445,241	
human	KEGG	2,509			3,927	
human	ORPHANET	3,113			5,690	
human	CTD	7,414			23,320	

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

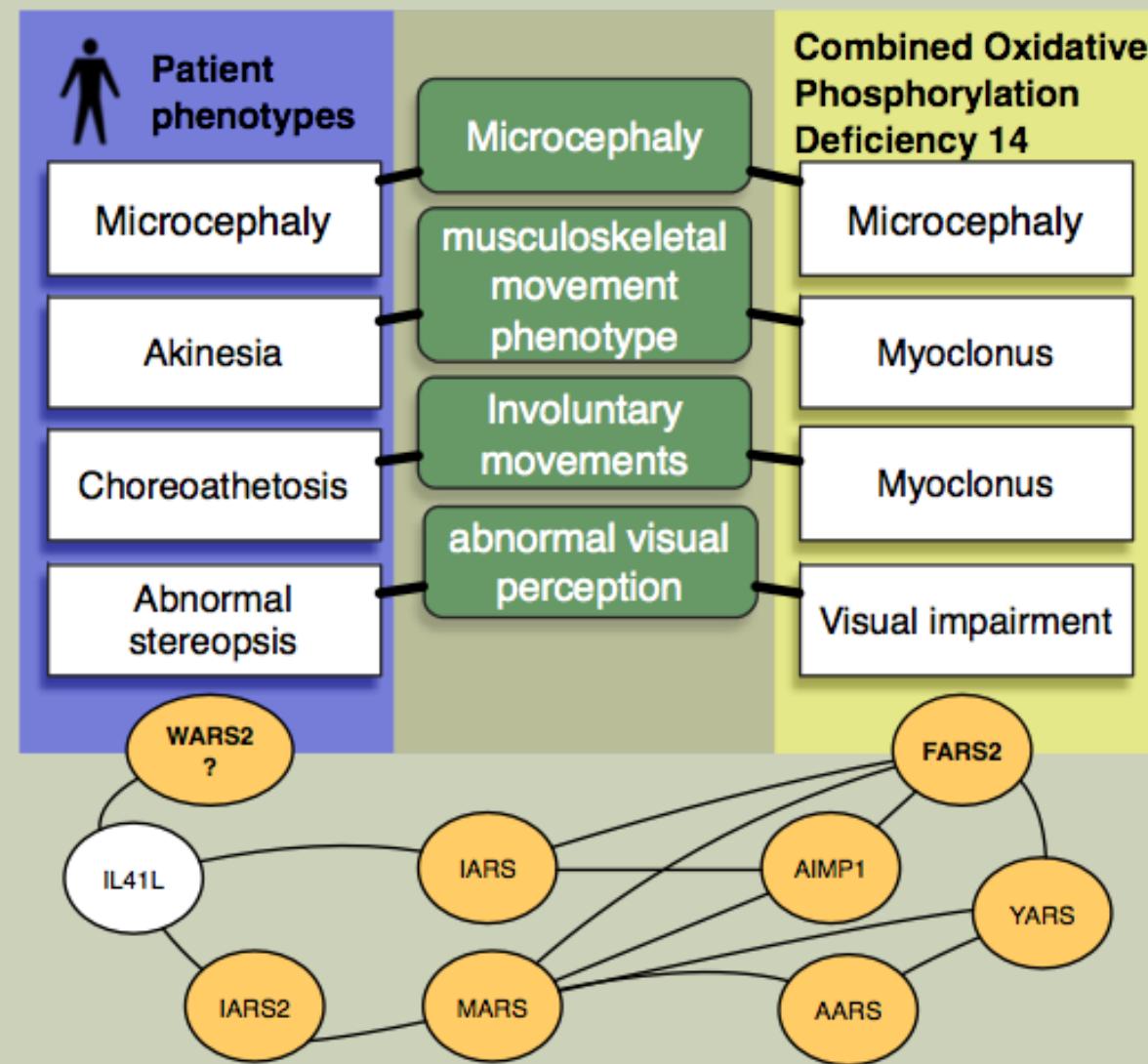
Coming soon: Animal QTLs for pig, cattle, chicken, sheep, trout, dog, horse

UDP_2731



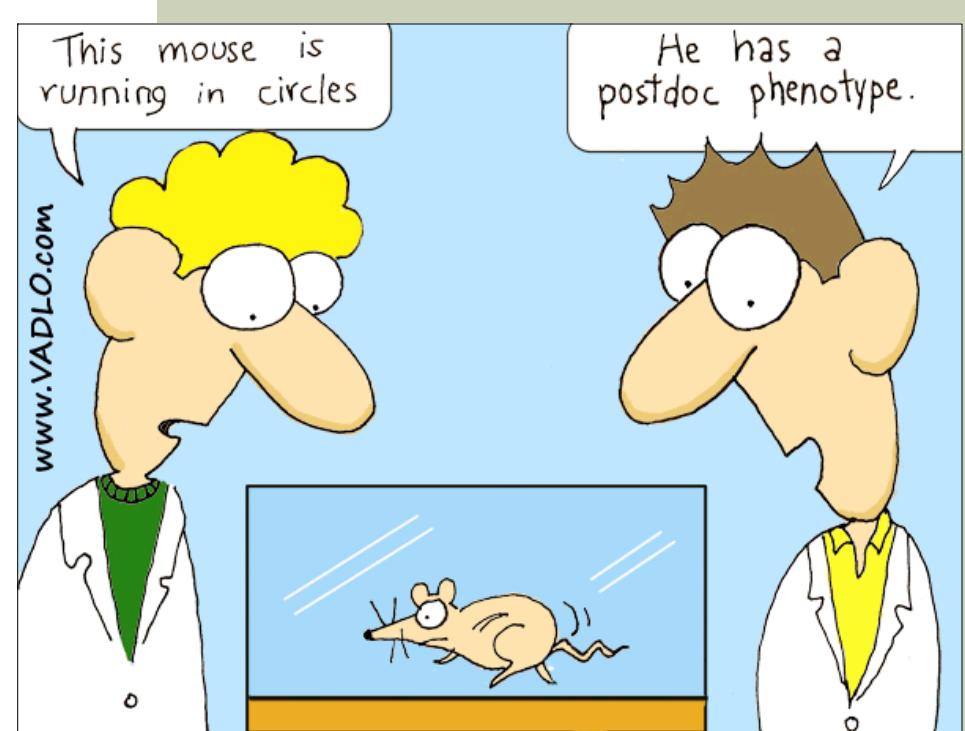
WALKING THE INTERACTOME

UDP_1166



A LOOK AT THE HPO

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



- [NA Y N] [HP:0100025] Overfriendliness
- [NA Y N] [HP:0002193] Pseudobulbar behavioral symptoms
- ▶ [NA Y N] [HP:0000711] Restlessness
- [NA Y N] [HP:0000723] Restrictive behavior
- [NA Y N] [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

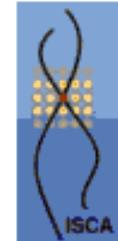
UK 100,000 Genomes Program

RIKEN

...

Close integration with other important efforts

Major credits go to OMIM and Orphanet



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

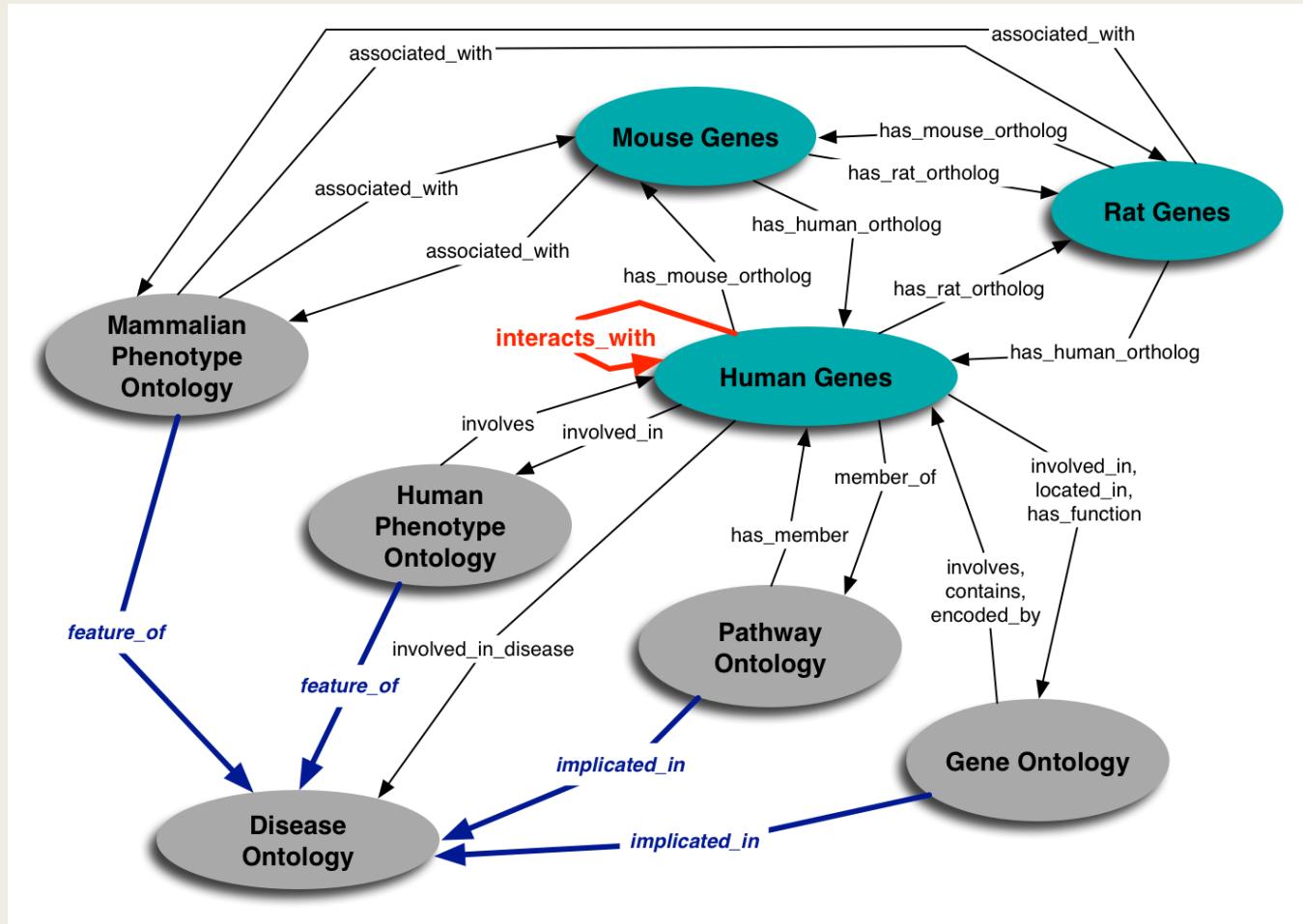
- Quantitative vs. qualitative – Much of clinical data is quantitative lab data with reference standards. It is possible to convert based on ± 3 SD, but no way to record the reference measure/population yet.
- Temporal presentation – ontologies can support temporal ordering, but data capture tools don't 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.

Predicament of modern genetics

- Vast amount of data being churned out
 - *How do we perform variant/gene prioritisation*

Idea

- Already identified *LRP2* for a rare familial disorder initially thought to be Myoshi myopathy



Predicament of modern genetics

- Vast amount of data being churned out
 - *How do we perform variant/gene prioritisation*
 - *BORG*
 - Semantic modelling to create a mind map “think like a researcher”
 - Utilises multiple integrated data sources
 - Graph database (neo4j)

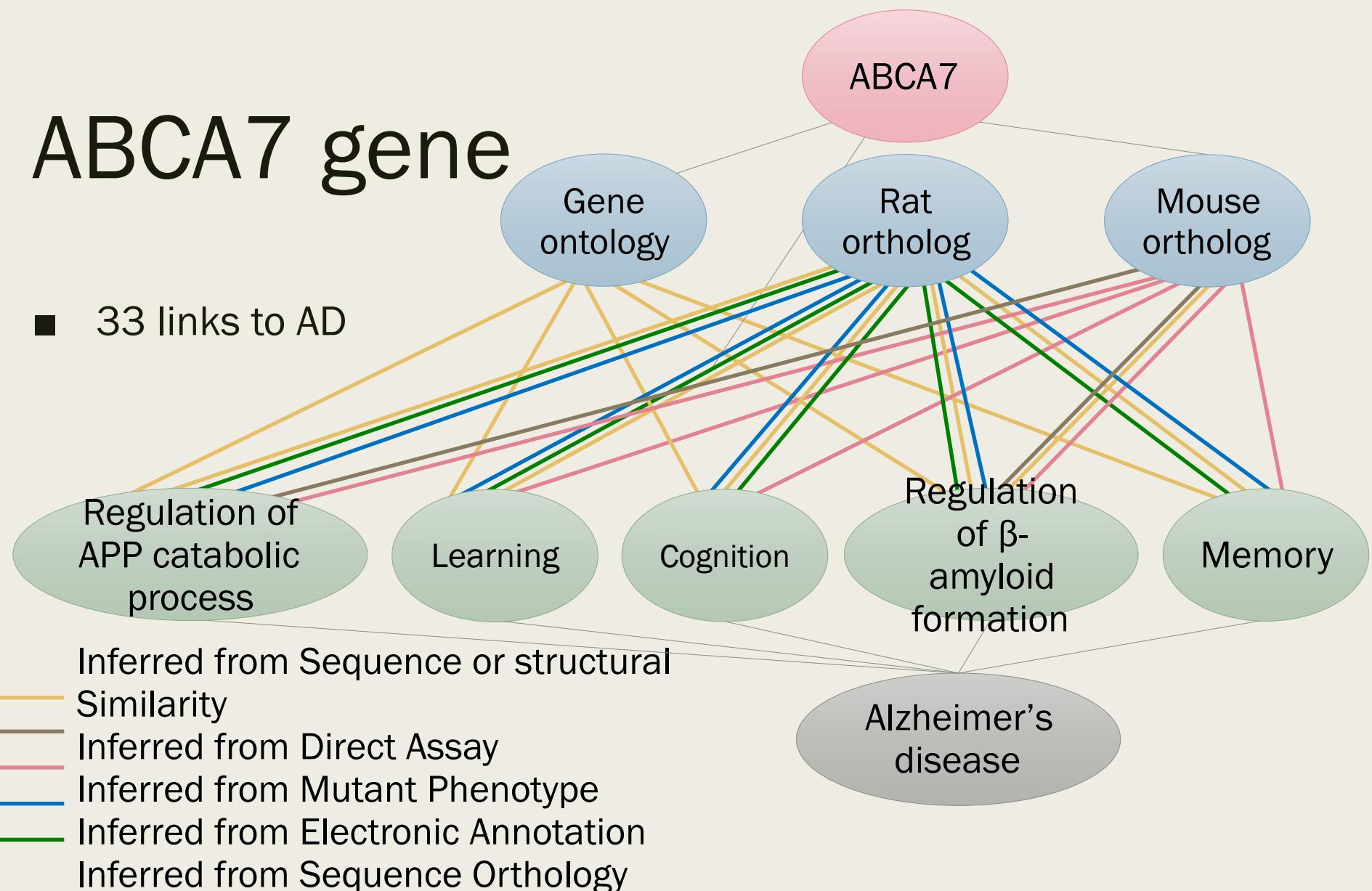
Utilising BORG for Alzheimer's disease

- Input gene list:

- *ABCA7*
 - *APOE*
 - *CD33*
 - *CLU*
 - *MEF2C*
 - *PICALM*
 - *SORL1*
- Will use these two genes as an example

ABCA7 gene

- 33 links to AD



APOE

■ 80 links to AD

