

Ontological, epistemological and terminological aspects of phenotypes



Olivier Bodenreider

Lister Hill National Center
for Biomedical Communications
Bethesda, Maryland - USA



U.S. National Library of Medicine



Disclaimer

The views and opinions expressed do not necessarily state or reflect those of the U.S. Government, and they may not be used for advertising or product endorsement purposes.



Introduction

- ◆ Phenotype: observable characteristics of an organism (anatomy, physiology, behavior)
- ◆ Phenotyping is crucial to understanding how genetic variation relates to clinical manifestations
 - Precise phenotyping is required for the study of rare syndromes
 - Poor interoperability of phenotypic data
 - Across clinical data repositories
 - Between research and clinical data repositories



Issues with phenotypes in standard terminologies

- ◆ Limited coverage
 - Post-coordination supports expansion
- ◆ Limited granularity
 - Coarse phenotyping is sufficient for some purposes

terminological

- ◆ Limited interoperability
 - Xrefs, mappings
 - Different definitions / representations

ontological

- ◆ Implicit context
 - e.g., congenitality, normality

epistemological



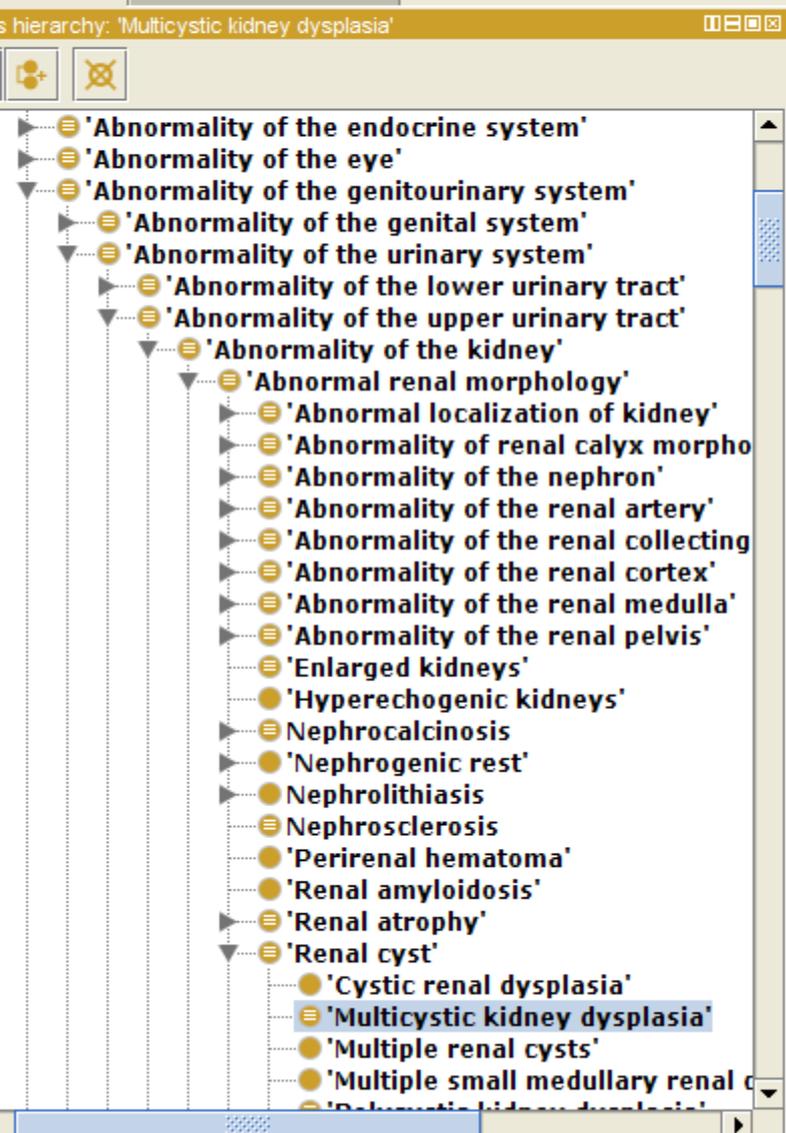
Terminological/ontological resources for phenotypes

Human Phenotype Ontology

- ◆ Developed collaboratively
 - Coordination: Peter Robinson
- ◆ Nightly builds
- ◆ Distributed as an OWL file
- ◆ 10,589 classes (as of Jan. 21, 2015)
- ◆ 16,608 names for phenotype
 - One preferred term for each class
 - 6019 exact synonyms
- ◆ Cross-references to standard terminologies
- ◆ Textual and logical definitions (PATO)
- ◆ Being integrated into the UMLS



Class hierarchy: 'Multicystic kidney dysplasia'



Annotations Usage

Annotations: 'Multicystic kidney dysplasia'

Annotations +

- label [type: string]

Multicystic kidney dysplasia
- comment [type: string]

Multicystic kidney dysplasia is the result of abnormal fetal renal development in which the affected kidney is replaced by multiple cysts and has little or no residual function. The vast majority of multicystic kidneys are unilateral. Multicystic kidney can be diagnosed on prenatal ultrasound.
- database_cross_reference [type: string]

MeSH:D021782
- database_cross_reference [type: string]

UMLS:C0345335
- definition [type: string]

Multicystic dysplasia of the kidney is characterized by multiple cysts of varying size in the kidney and the absence of a normal pelvocaliceal system. The condition is associated with ureteral or ureteropelvic atresia, and the affected kidney is nonfunctional.

Description: 'Multicystic kidney dysplasia'

Equivalent To +

- 'has part' some (polycystic and ('inherits in' some kidney) and ('has modifier' some abnormal))

SubClass Of +

- 'Renal cyst'

SubClass Of (Anonymous Ancestor)

Annotation of phenotypes in OMIM and OrphaNet

#608836 CARNITINE PALMITOYLTRANSFERASE II DEFICIENCY, LETHAL NEONATAL;;CARNITINE PALMITOYLTRANSFERAS...	CPT2
#612513 CHROMOSOME 2P16.1-P15 DELETION SYNDROME	-
#614209 MECKEL SYNDROME, TYPE 9; MKS9	B9D1
#614527 CHROMOSOME 17Q12 DELETION SYNDROME	-
MOSAIC VARIEGATED ANEUPLOIDY SYNDROME	CEP57 ; BUB1B ; BUB1 ; BUB3
BOR SYNDROME	SIX5 ; SIX1 ; EYA1
<u>BARDET-BIEDL SYNDROME</u>	MKKS ; SDCCAG8 ; WDPCP ; BBS5 ; BBS1 ; TRIM32 ; BBS2 ; IFT27 ; ARL6 ; BBS4 ; CEP290 ; BBS12 ; LZTFL1 ; MKS1 ; BBS10 ; BBIP1 ; NPHP1 ; BBS7 ; IFT172 ; BBS9 ; TTC8
SHORT RIB-POLYDACTYLY SYNDROME	-
CARNITINE PALMITOYLTRANSFERASE II DEFICIENCY	-
NON-RHIZOMELIC CHONDRODYSPLASIA PUNCTATA	-
RENAL DYSPLASIA - MEGALOCYSTIS - SIRENOMELIA	-
INDOMETHACIN EMBRYOFETOPATHY	-

Annotation of phenotypes in OrphaNet



ORPHA110		ICD-10:	Q87.8
Synonym(s):	BBS	OMIM:	209900 [↗] 600151 [↗] 605231 [↗]
Prevalence:	1-9 / 1 000 000		615981 [↗] 615982 [↗] 615983 [↗]
Inheritance:	Oligogenic or Autosomal recessive		615984 [↗] 615985 [↗] 615986 [↗]
Age of onset:	Infancy Neonatal Antenatal		615987 [↗] 615988 [↗] 615989 [↗]
		UMLS:	C0752166
		MeSH:	D020788
		MedDRA:	10056715

SUMMARY

Bardet-Biedl syndrome (BBS) is a ciliopathy with multisystem involvement. Its prevalence in Europe is estimated at between 1/125,000 and 1/175,000. This disorder is characterized by a combination of clinical signs: obesity, pigmentary retinopathy, post-axial polydactyly, polycystic kidneys, hypogenitalism, and learning disabilities, many of which appear several years after disease onset. Clinical expression is variable but most patients manifest the majority of clinical signs during the disease course. Pigmentary retinopathy is the only constant clinical sign after childhood. BBS may also be associated with several other manifestations including diabetes, hypertension, congenital cardiopathy and Hirschsprung disease (see this term). The wide clinical spectrum observed in BBS is associated with significant genetic heterogeneity. The disorder is transmitted mainly in an autosomal recessive manner but oligogenic inheritance has been reported in some cases. To date, mutations in 12 different genes (*BBS1* to *BBS12*) have been identified as being responsible for this phenotype. These genes code for proteins involved in the development and function of primary cilia. Absence or dysfunction of BBS proteins results in ciliary anomalies in organs such as the kidney or eye. However, the relationship between symptoms and ciliary dysfunction remains obscure for some of the clinical manifestations of BBS. Recognition of the clinical picture is important as the diagnosis can be confirmed by molecular analysis, allowing appropriate genetic counseling for family members and possible prenatal diagnosis. The differential diagnosis should include the Alström, McKusick-Kaufmann and Meckel-Gruber syndromes (see these terms). Patients with BBS will need multidisciplinary medical care. The renal abnormalities are the main life-threatening manifestations because they can lead to end-stage renal failure and require renal transplantation. Progressive vision loss due to retinal dystrophy, together with moderate intellectual deficit (when present), behavioral anomalies, hypomimia and obesity will affect the social life of these patients.

Additional information

Further information on this disease

- > Classification(s) (9)
- > Gene(s) (21)
- > Publications in PubMed [↗]
- > Other website(s) (9)

Health care resources for this disease

- > Expert centres (369)
- > Diagnostic tests (54)
- > Patient organisations (81)
- > Orphan drug(s) (1)

Research activities on this disease

- > Research projects (58)
- > Clinical trials (0)
- > Registries/biobanks (28)
- > Networks (20)

Orphanet Reports series

- > Prevalence



SNOMED CT

- ◆ Developed by the International Health Terminology Standard Development Organization
- ◆ Description logics formalism
 - Supports post-coordination
- ◆ Broad coverage of clinical medicine
 - ~300,000 concepts
- ◆ Clinical findings
 - ~100,000 concepts
 - 169,000 names
- ◆ Logical definitions
- ◆ Integrated in the UMLS



SNOMED CT browser

IHTSDO SNOMED CT Browser Release: United States Edition 20150301 Perspective: Full Feedback About USA  Delivering SNOMED CT

Concept Details

Summary **Details** Diagram Refsets Members References

Parents Stated Inferred

-   Kidney disease (disorder)

 **Multicystic kidney (disorder)** 

SCTID: 204962002

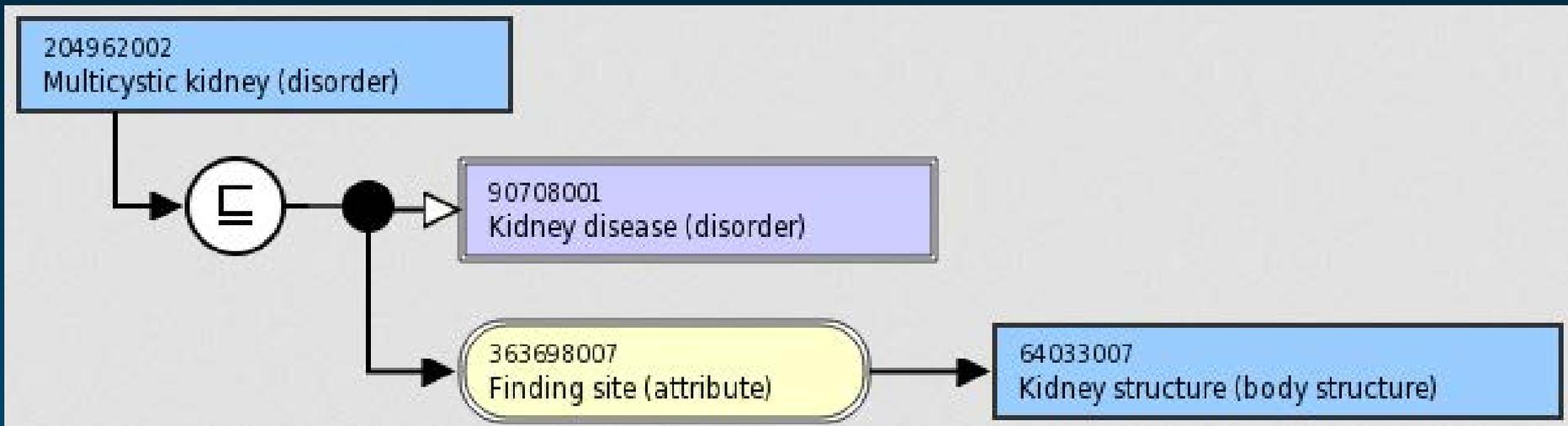
- Multicystic kidney (disorder)
- MCKD - Multicystic kidney disease
- Multicystic kidney
- Multicystic renal dysplasia

Finding site → Kidney structure

Children (0)
No children



Logical definition



Unified Medical Language System (UMLS)

- ◆ Terminology integration system
- ◆ Developed by NLM
- ◆ Integrates many (140) standard biomedical terminologies
 - SNOMED CT
 - MeSH
 - International Classification of Diseases
 - MedDRA
 - [HPO]
- ◆ 3M concepts
- ◆ 8M normalized terms



UMLS Terminology Server

A service of the U.S. National Library of Medicine | National Institutes of Health

[My Profile](#) | [Sign Out](#) | [Contact](#)



Unified Medical Language System®

UMLS Terminology Services

Metathesaurus Browser

Welcome back,
bodenreider

[UTS Home](#) [Applications](#) [SNOMED CT](#) [Resources](#) [Downloads](#) [Documentation](#) [UMLS Home](#)

Search **Tree** **Recent Searches**

Term CUI Code

multicystic kidney dysplasia

Go

Release: 2015AA

Search Type: Word

Source: All Sources

AIR

ALT

AOD

AOT

Search Results (1)

[C3714581](#) Multicystic Dysplastic Kidney

Basic View **Report View** **Raw View**

Concept: [C3714581] Multicystic Dysplastic Kidney

Semantic Types

[Congenital Abnormality](#) [T019]

[Disease or Syndrome](#) [T047]

Definitions

MSH/MH | A nongenetic defect due to malformation of the KIDNEY which appears as a bunch of grapes with multiple renal cysts but lacking the normal renal bean shape, and the collection drainage system. This condition can be detected in-utero with ULTRASONOGRAPHY.

Atoms (78) string [AUI / RSAB / TTY / Code]

[multicystic dysplastic kidney](#) [A18674881/CHV/SY/0000031000]

[multicystic dysplastic kidneys](#) [A18637802/CHV/SY/0000031000]

[multicystic kidney](#) [A18693263/CHV/PT/0000031000]

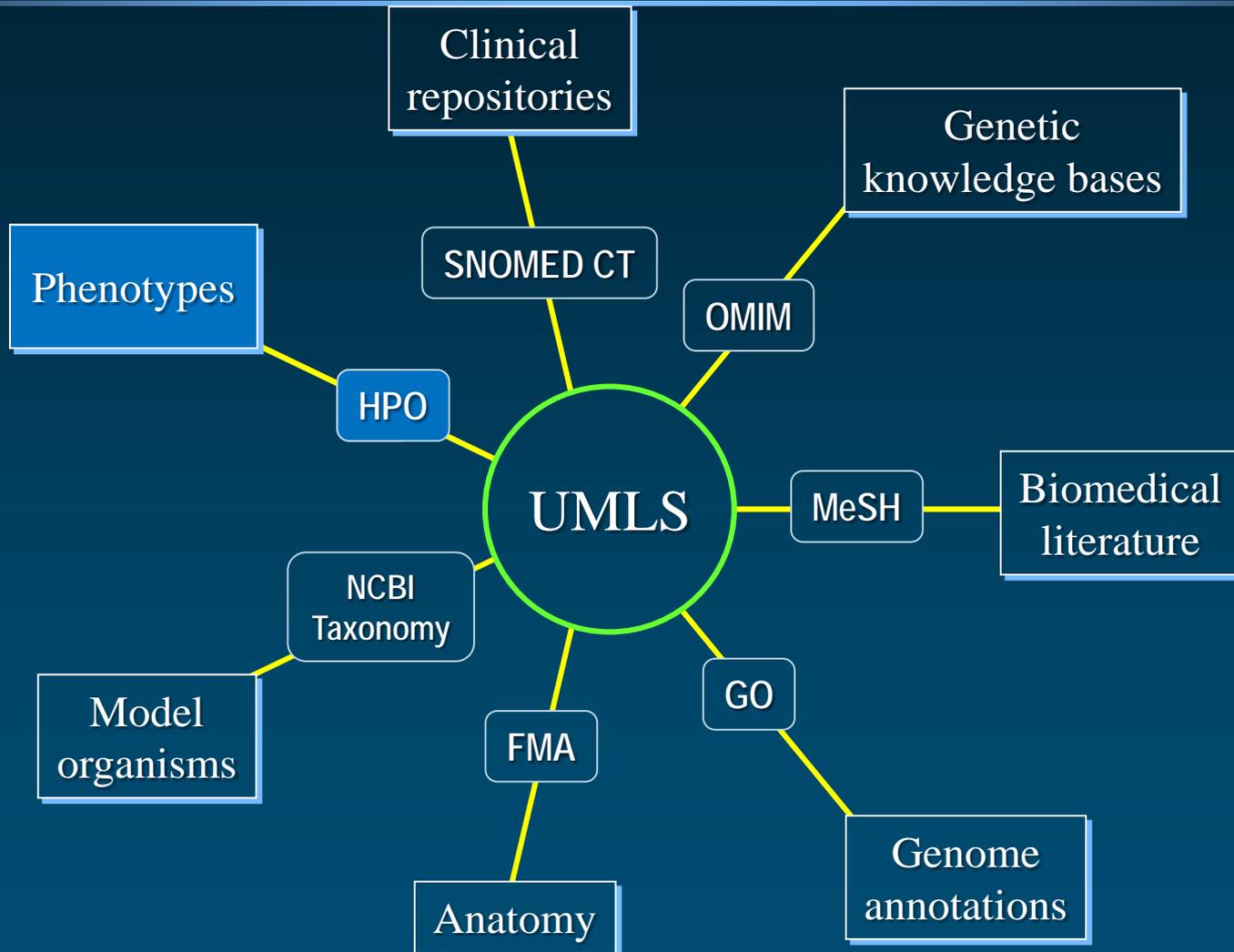
[multicystic kidney dysplasia](#) [A18563499/CHV/SY/0000031000]

[multicystic kidneys](#) [A18600533/CHV/SY/0000031000]

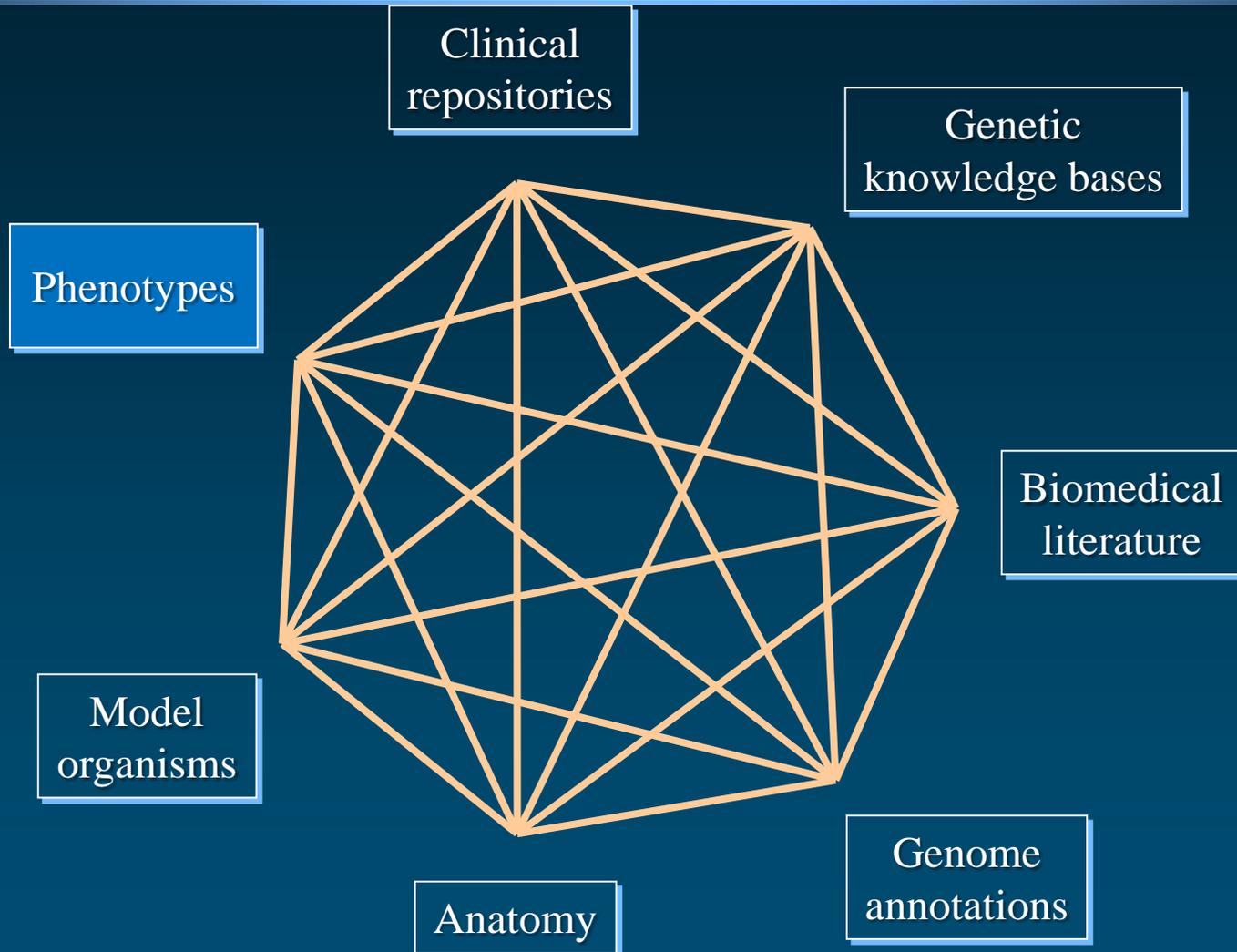
[multicystic renal dysplasia](#) [A18656310/CHV/SY/0000031000]

[Multicystic dysplastic kidney](#) [A17841572/ICD10CM/ET/O61.41]

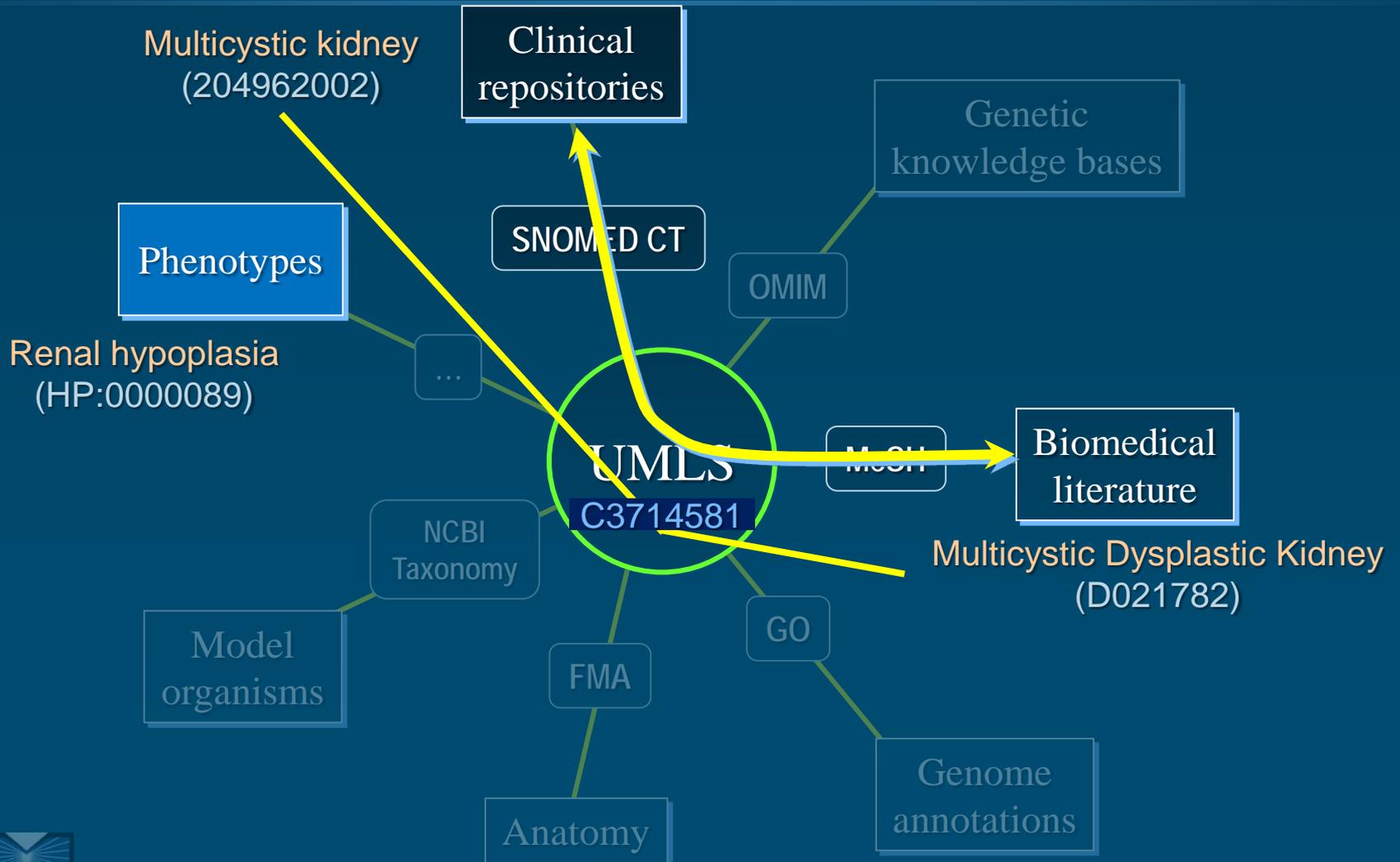
Integrating subdomains



Integrating subdomains

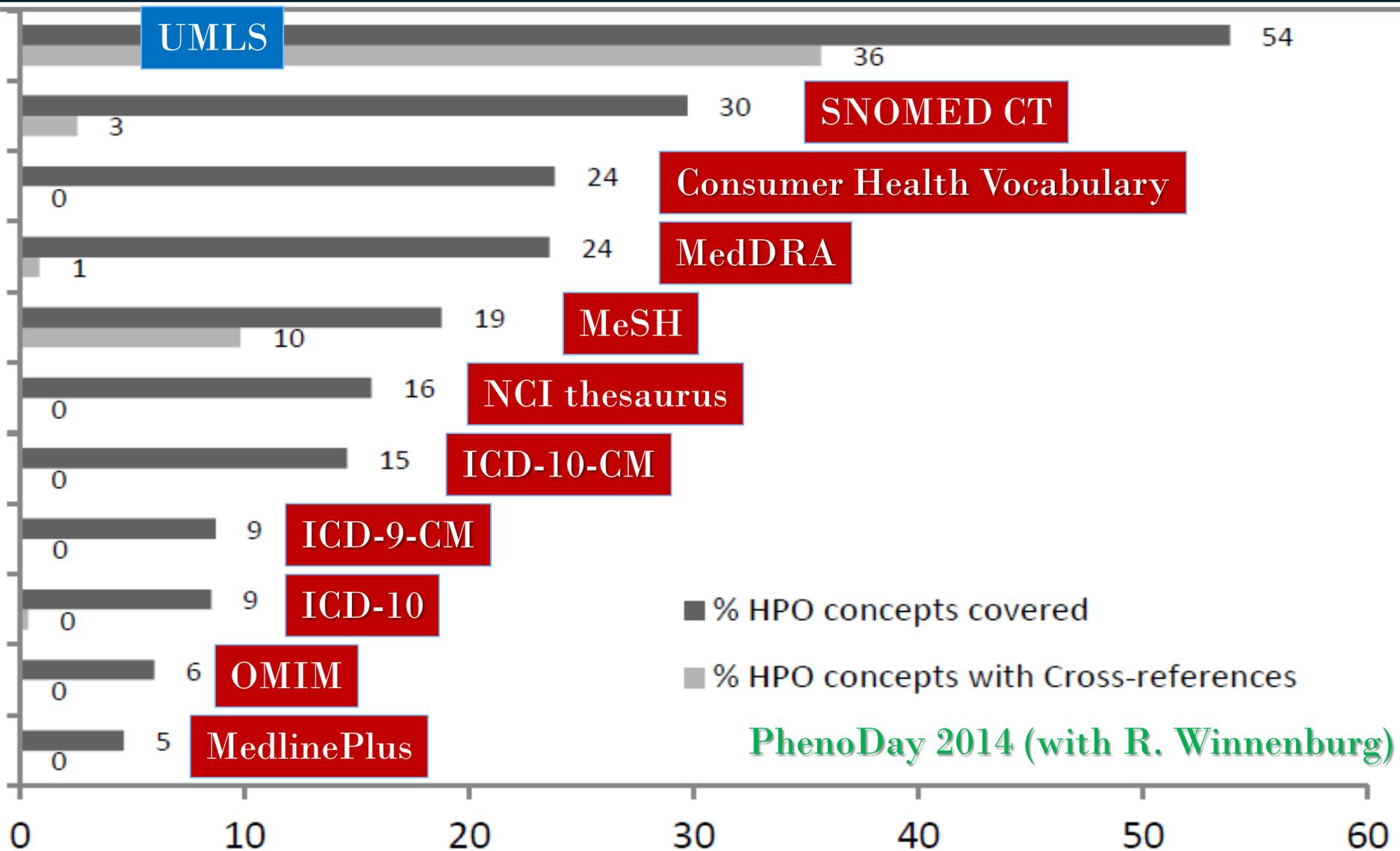


Terminology integration



Coverage
Granularity
Mapping
Representation
Context

Coverage of HPO term in standard terminologies (lexical mapping, as of 2014)



HPO terms and SNOMED CT

- ◆ Atrial fibrillation (HP_0005110)
 - Mapping to: [Atrial fibrillation](#) (49436004)
- ◆ Inlet ventricular septal defect (HP_0011622)
 - Mapping to: [Common atrioventricular canal](#) (360481003)
- ◆ Palmoplantar keratoderma (HP_0000982)
 - **No mapping**
- ◆ Hypoplastic nasal septum (HP_0005104)
 - **No mapping**
- ◆ Oval transradiancy (humeral) (HP_0003877)
 - **No mapping (not even in UMLS)**
- ◆ Lower limb peromelia (HP_0009820)
 - **No mapping (not even in UMLS)**

Mapping through pre-coordination

HPO

SNOMED CT

“Renal hypoplasia”
[HPO:HP_0000089]



“Congenital hypoplasia of kidney”
[SCTID:32659003]

synonym “renal hypoplasia”

MAPPING
THROUGH
PRE-COORDINATION



Mapping through pre-coordination

HPO

SNOMED CT

“Renal hypoplasia”

[HPO:HP_0000089]



“Congenital hypoplasia of kidney”

[SCTID:32659003]

synonym “renal hypoplasia”

MAPPING
THROUGH
PRE-COORDINATION

“Macular hypoplasia”

[HPO:HP_00001104]



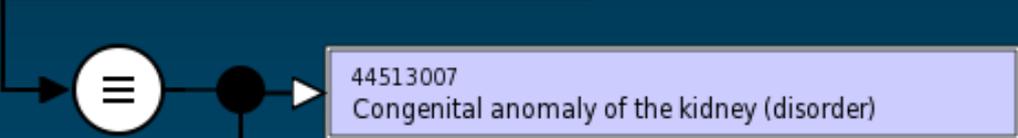
Logical definition

SNOMED CT

“Congenital hypoplasia of kidney”
[SCTID:32659003]

synonym “renal hypoplasia”

32659003
Congenital hypoplasia of kidney (disorder)



44513007
Congenital anomaly of the kidney (disorder)

246454002
Occurrence (attribute)

CONGENITAL

116676008
Associated morphology (attribute)

HYPOPLASIA

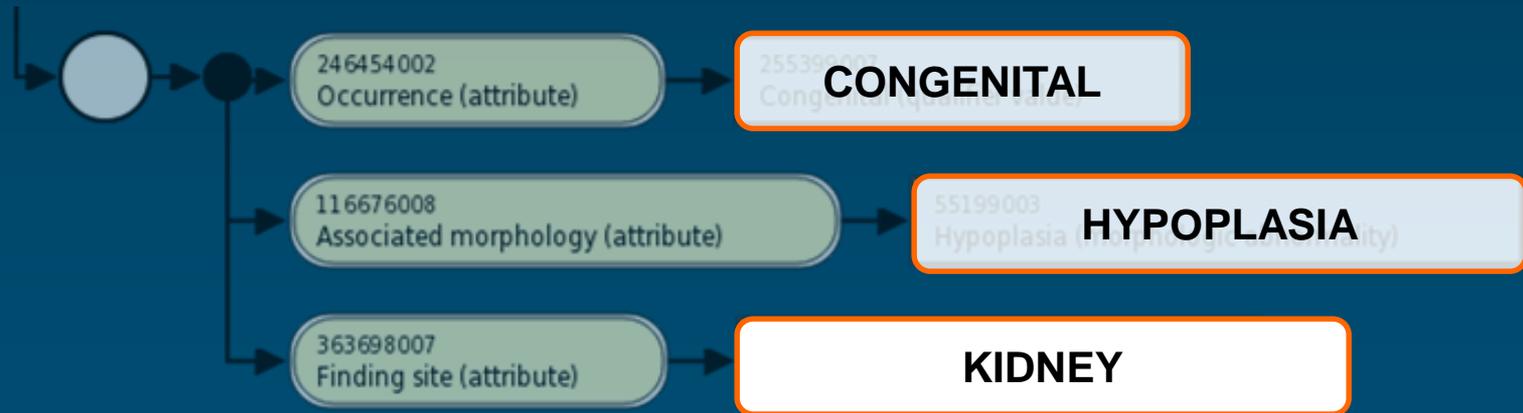
363698007
Finding site (attribute)

KIDNEY



Logical definition

SNOMED CT



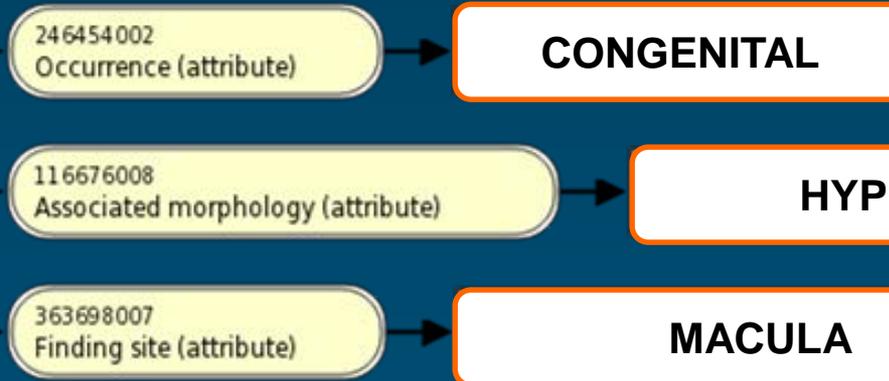
Logical definition (modified)

SNOMED CT

“Congenital hypoplasia of macula”
[SCTID:xxxx]



This is a post-coordinated expression...

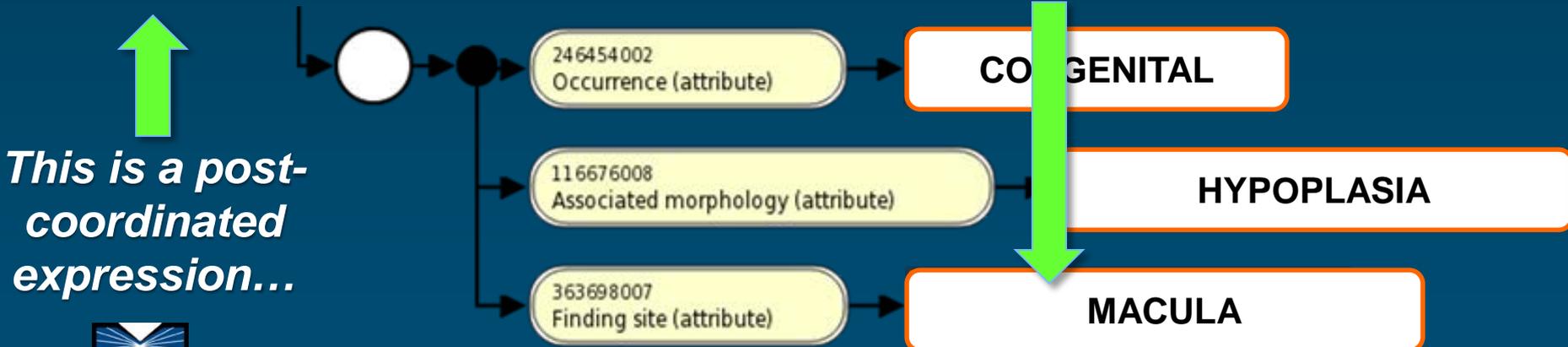


Logical definition (modified)

SNOMED CT

“Congenital hypoplasia of macula”
[SCTID:xxxx]

... for a specific
anatomical entity



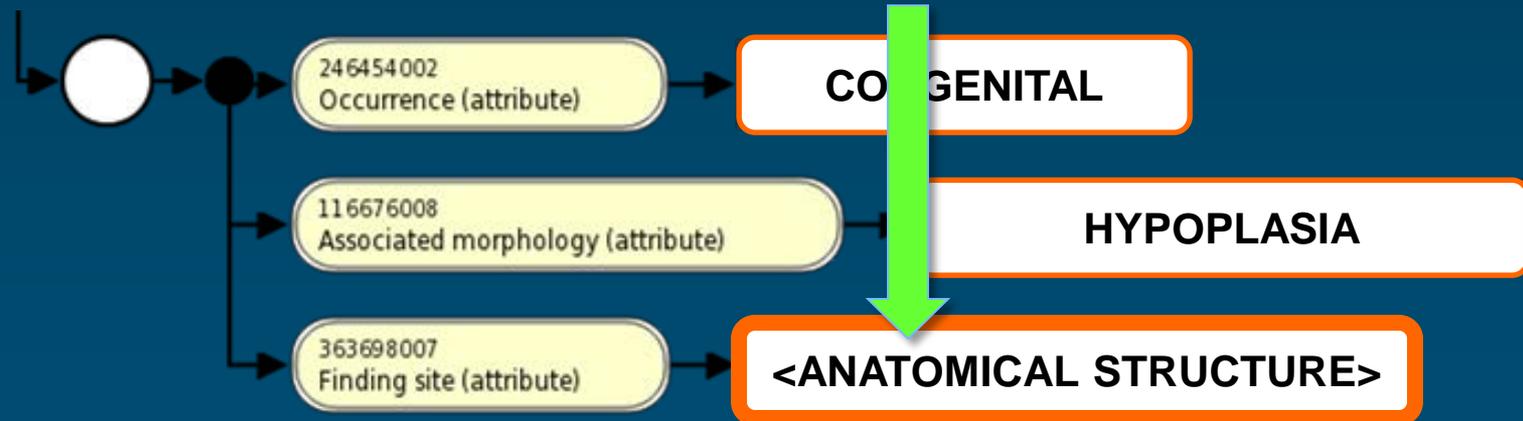
This is a post-coordinated expression...



Logical definition (generalized)

SNOMED CT

Generalization



Template

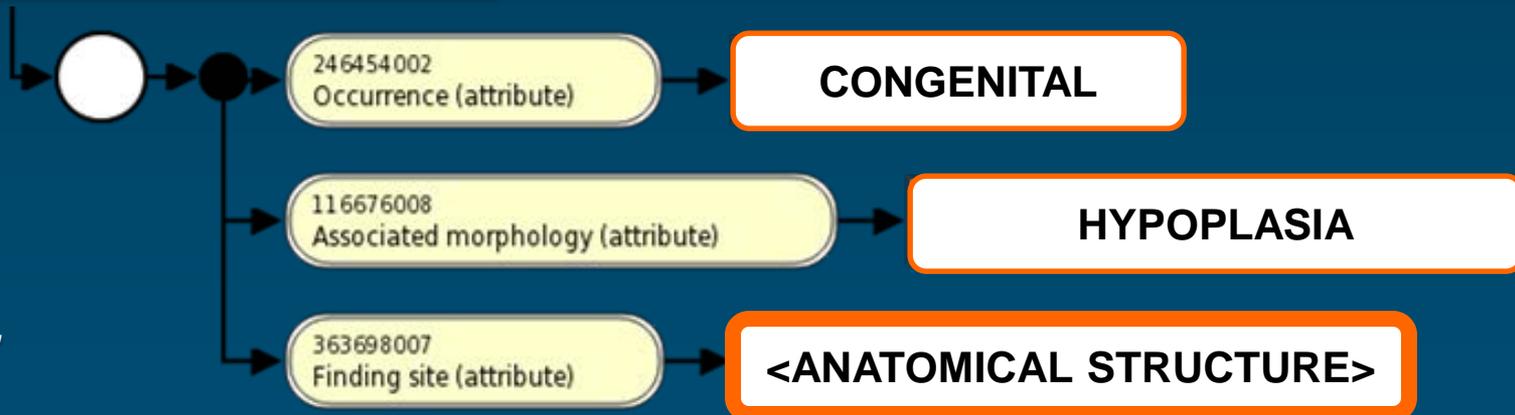
SNOMED CT

TEMPLATE

<ANATOMICAL STRUCTURE>{*hypoplasia*}



*This is a
template for
HPO terms...*



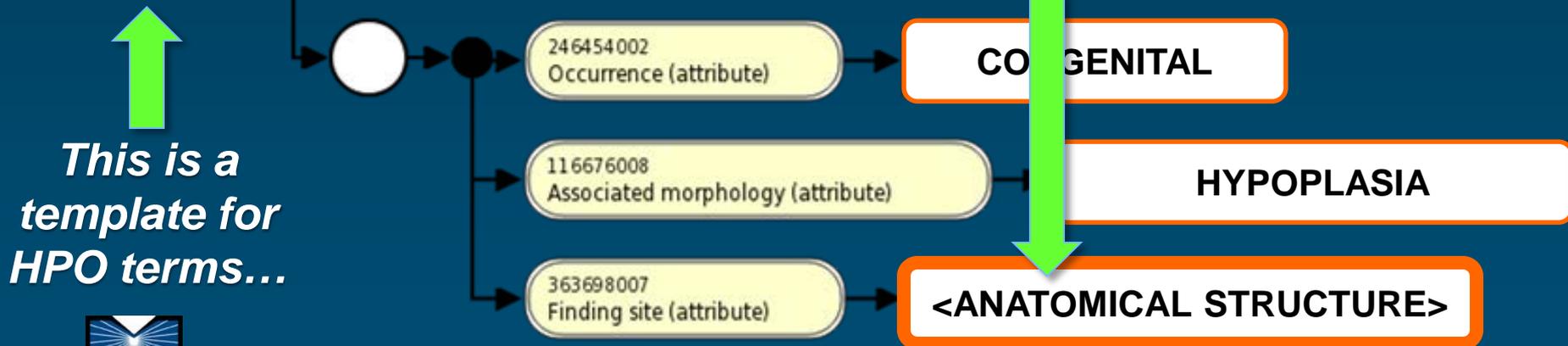
Methods

SNOMED CT

TEMPLATE

<ANATOMICAL STRUCTURE>{*hypoplasia*}

*... for any
anatomical
entity*



*This is a
template for
HPO terms...*



Mapping through post-coordination

HPO

SNOMED CT

“Renal hypoplasia”
[HPO:HP_0000089]



“Congenital hypoplasia of kidney”
[SCTID:32659003]

synonym “renal hypoplasia”

MAPPING
THROUGH
PRE-COORDINATION

“Macular hypoplasia”
[HPO:HP_00001104]

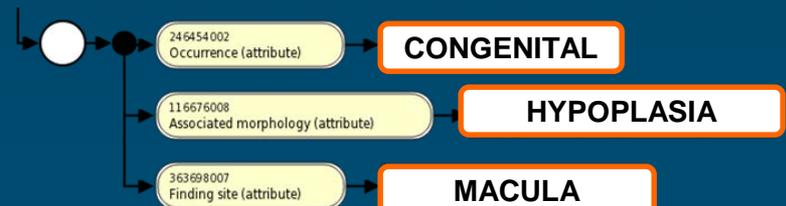


“Congenital hypoplasia of macula”
[SCTID:xxxx]

TEMPLATE

<ANATOMICAL STRUCTURE>{qualifier}

MAPPING
THROUGH
POST-COORDINATION



Post-coordination in action

- ◆ With 12 post-coordination templates, we generated post-coordinated mappings to SNOMED CT for 1617 HPO concepts
- ◆ This is in complement to the 3081 HPO concepts for which there is a pre-coordinated mapping to SNOMED CT
- ◆ Template-based mappings are usually of high quality

Medinfo 2015 (with F. Dhombres)



Issues

◆ With post-coordination

- Not end user-friendly
- Impractical in regular clinical data entry systems
- “excessive pre-coordination” – perspective of terminologists vs. clinicians

◆ With the mappings

- Context of HPO terms assumed in some cases
 - E.g., congenitality
 - HPO: Macular hypoplasia
 - SNOMEDCT: Congenital hypoplasia of the macula



Coverage
Granularity
Mapping
Representation
Context

Deep vs. coarse phenotyping

- ◆ “Next-generation sequencing demands next-generation phenotyping”
 - Hennekam, R.C. and Biesecker, L.G. (2012), *Hum Mutat*, **33**, 884-886
- ◆ Yet...

EMR-based PheWAS using ICD9 codes

These current studies using PheWAS have been performed using a custom, hierarchical grouping of International Classification of Disease, 9th edition (ICD9) codes applied to EMR data from. There are a total of 1645 PheWAS case groups (typically diseases), each with a corresponding control group. These groupings loosely follow the 3-digit (category) and section groupings defined with the ICD9 code system itself, and have been revised based on statistical co-occurrence, code frequency, and human review. For more information, see the references below.



Deep phenotyping

◆ OMIM diseases annotated with the HPO term Multicystic kidney dysplasia

<http://www.human-phenotype-ontology.org/>

#614527 CHROMOSOME 17Q12 DELETION SYNDROME	-
MOSAIC VARIEGATED ANEUPLOIDY SYNDROME	CEP57 ; BUB1B ; BUB1 ; BUB3
BOR SYNDROME	SIX5 ; SIX1 ; EYA1
BARDET-BIEDL SYNDROME	MKKS ; SDCCAG8 ; WDPCP ; BBS5 ; BBS1 ; TRIM32 ; BBS2 ; IFT27 ; ARL6 ; BBS4 ; CEP290 ; BBS12 ; LZTFL1 ; MKS1 ; BBS10 ; BBIP1 ; NPHP1 ; BBS7 ; IFT172 ; BBS9 ; TTC8
SHORT RIB-POLYDACTYLY SYNDROME	-
CARNITINE PALMITOYLTRANSFERASE II DEFICIENCY	-
NON-RHIZOMELIC CHONDRODYSPLASIA PUNCTATA	-
RENAL DYSPLASIA - MEGALOCYSTIS - SIRENOMELIA	-
INDOMETHACIN EMBRYOFETOPATHY	-

Coarse phenotyping

<https://emerge.mc.vanderbilt.edu/>

◆ eMERGE

- “National network [...] that combines DNA biorepositories with electronic medical record (EMR) systems for large scale, high-throughput genetic research in support of implementing genomic medicine”



Coarse phenotyping eMERGE

◆ GWAS

- Aim 2: “conduct genome-wide association studies (GWAS) using the phenotypes derived [from EMR data]”

◆ PheWAS

- Phenome-wide association study

◆ Phenotype definition

- Based on ICD9-CM codes, drugs and lab tests codes, and mentions in clinical narratives
- Phenotype KnowledgeBase (PheKB)



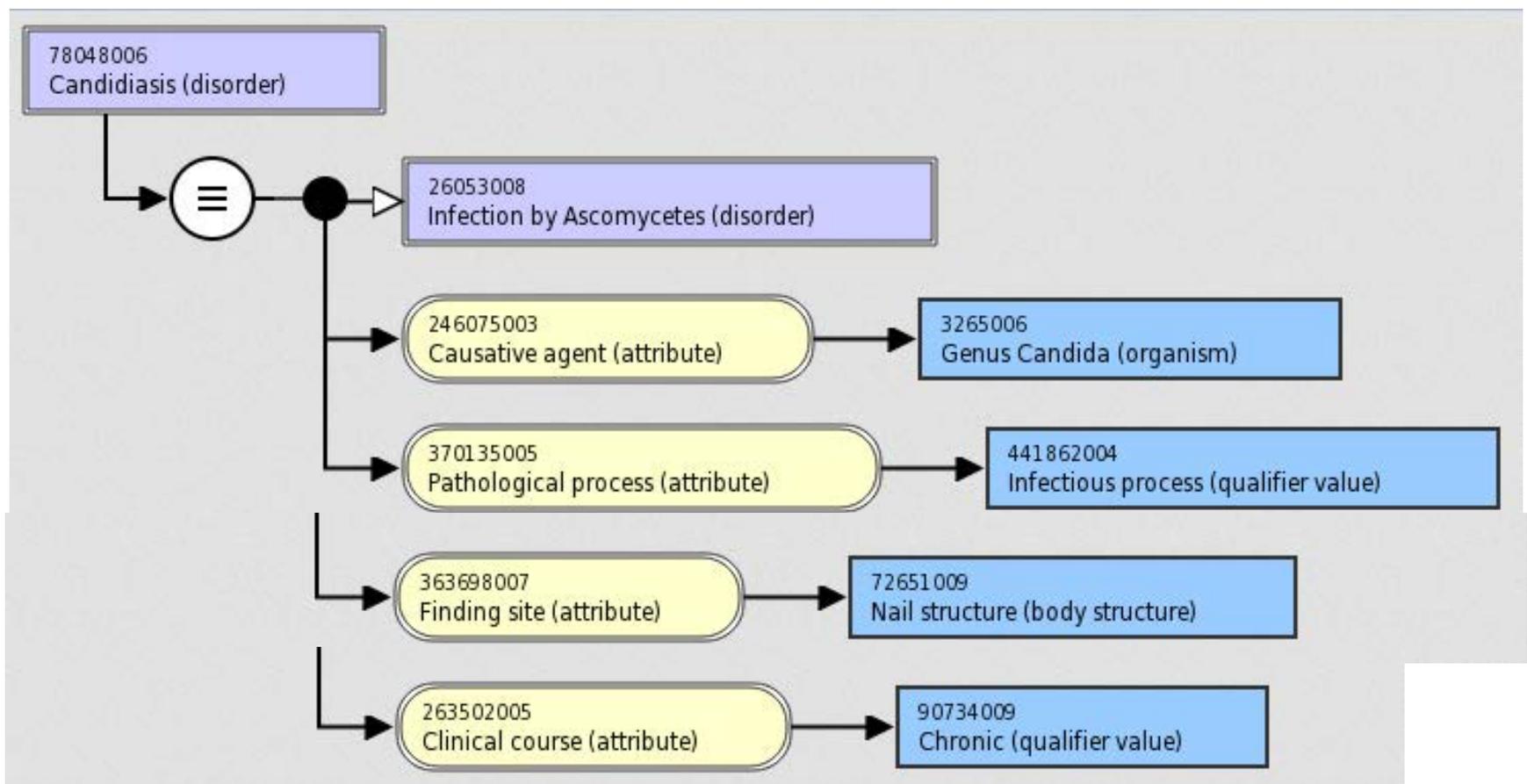
Refining ontologies

- ◆ Post-coordination can help mitigate granularity issues
- ◆ Logical definitions in SNOMED CT can be refined by
 - Laterality
 - Severity
 - Onset
 - ...

Refinement through post-coordination

HPO: Chronic monilial nail infection

SNOMED CT: Candida infection



Coverage
Granularity
Mapping
Representation
Context

Mapping

◆ Types of mappings

- Based on strings (lexical) vs. logical definitions
- Complete (equivalence mapping) vs. partial (subsumption mapping)
- Contributed by the developers of a resource (Xrefs) vs. through a terminology integration system (UMLS, BioPortal)

◆ Usage of mappings

- Directionality may matter
- Integration vs. annotation/coding/indexing

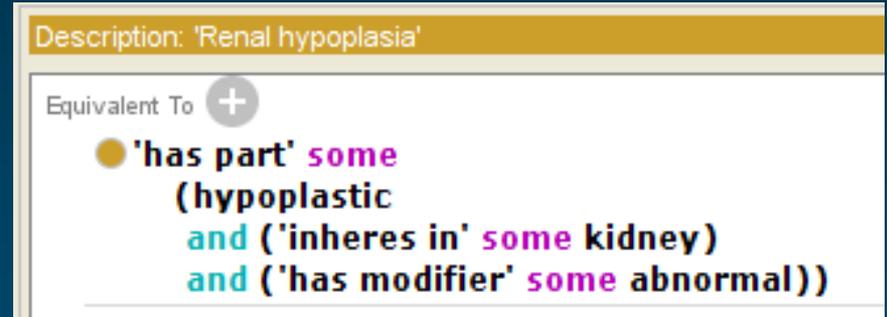


Coverage
Granularity
Mapping
Representation
Context

Language vs. representation

- ◆ Anatomical structures as phenotypes?
 - Small kidneys vs. Renal hypoplasia (synonyms)
 - Small kidneys isa Kidney (anatomical structure)
 - Hypoplasia of kidney isa Hypoplasia (clinical finding)

- ◆ No but...
 - Frequent shortcuts
 - Absent Achilles reflex
 - Enlarged cerebellum
 - [...]
 - Likely to confuse NLP systems



Description: 'Renal hypoplasia'

Equivalent To 

- 'has part' some (hypoplastic and ('inheres in' some kidney) and ('has modifier' some abnormal))

Different representations

- ◆ Both HPO and SNOMED CT provide logical definitions
 - HPO
 - OWL2 DL
 - Based on PATO
 - SNOMED CT
 - EL++
 - Based on the SNOMED CT concept model

Logical definition in HPO

HPO

“Renal hypoplasia”

[HPO:HP_0000089]

Description: 'Renal hypoplasia'

Equivalent To



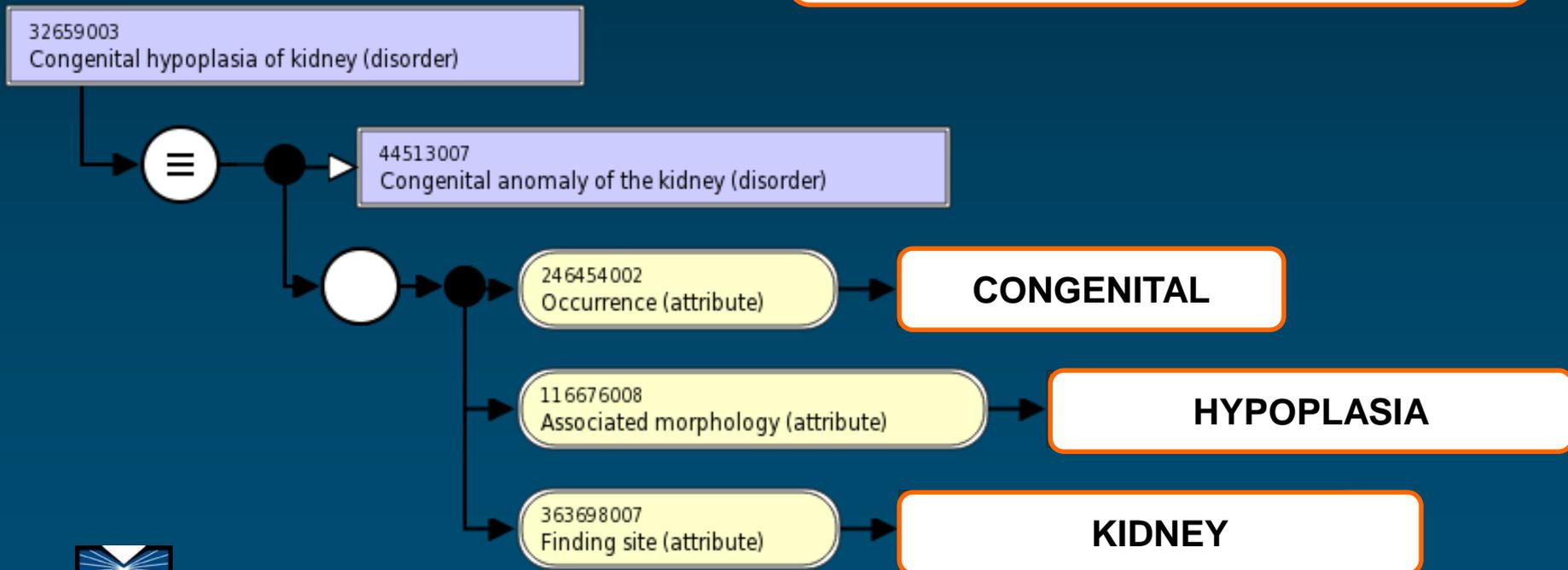
- **'has part' some**
(hypoplastic
and ('inheres in' some kidney)
and ('has modifier' some abnormal))



Logical definition in SNOMED CT

SNOMED CT

“Congenital hypoplasia of kidney”
[SCTID:32659003]

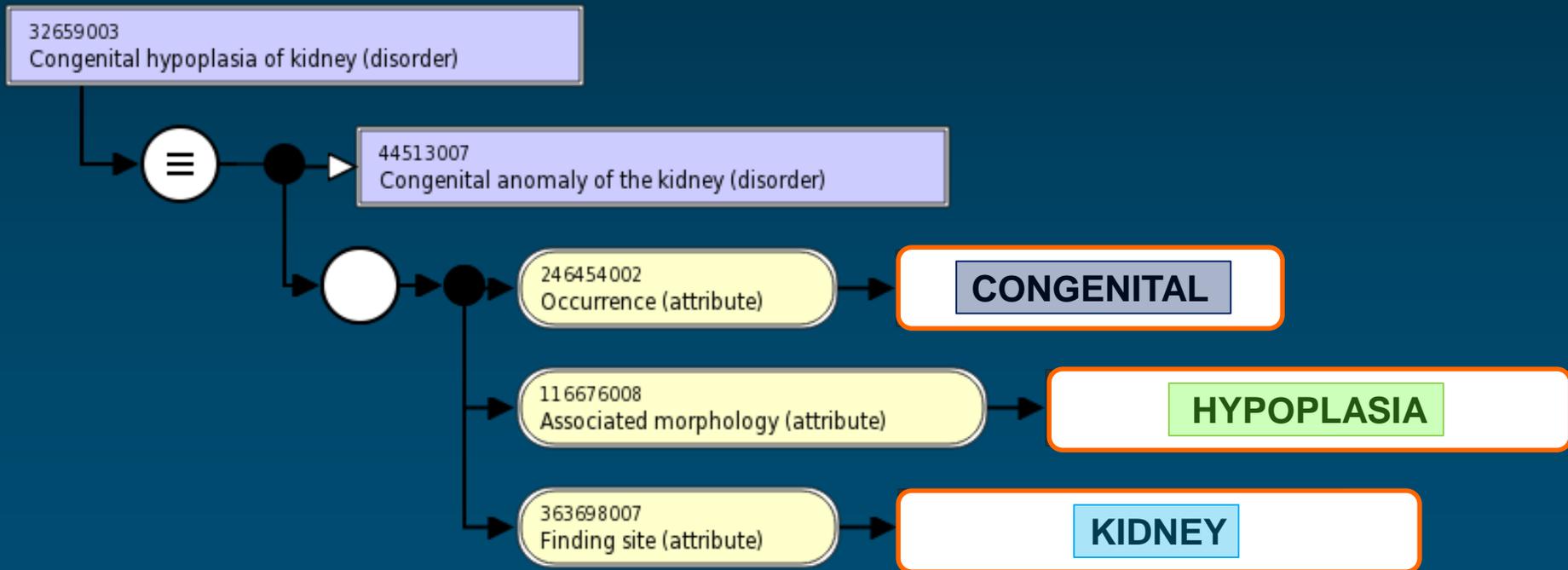


Description: 'Renal hypoplasia'

Equivalent To



- 'has part' some (hypoplastic and ('inheres in' some kidney) and ('has modifier' some abnormal))



Issues with representation

- ◆ Representations are not interoperable
 - Different sets of genus/differentiai
 - Entity/quality (HPO [PATO])
 - Anatomy/Morphology/Occurrence (SNOMEDCT)
 - But rules based on the DL definitions could form the basis for a new mapping approach
 - Entity → Anatomy [or Physiology or Behavior]
 - Quality → Morphology [or ...]

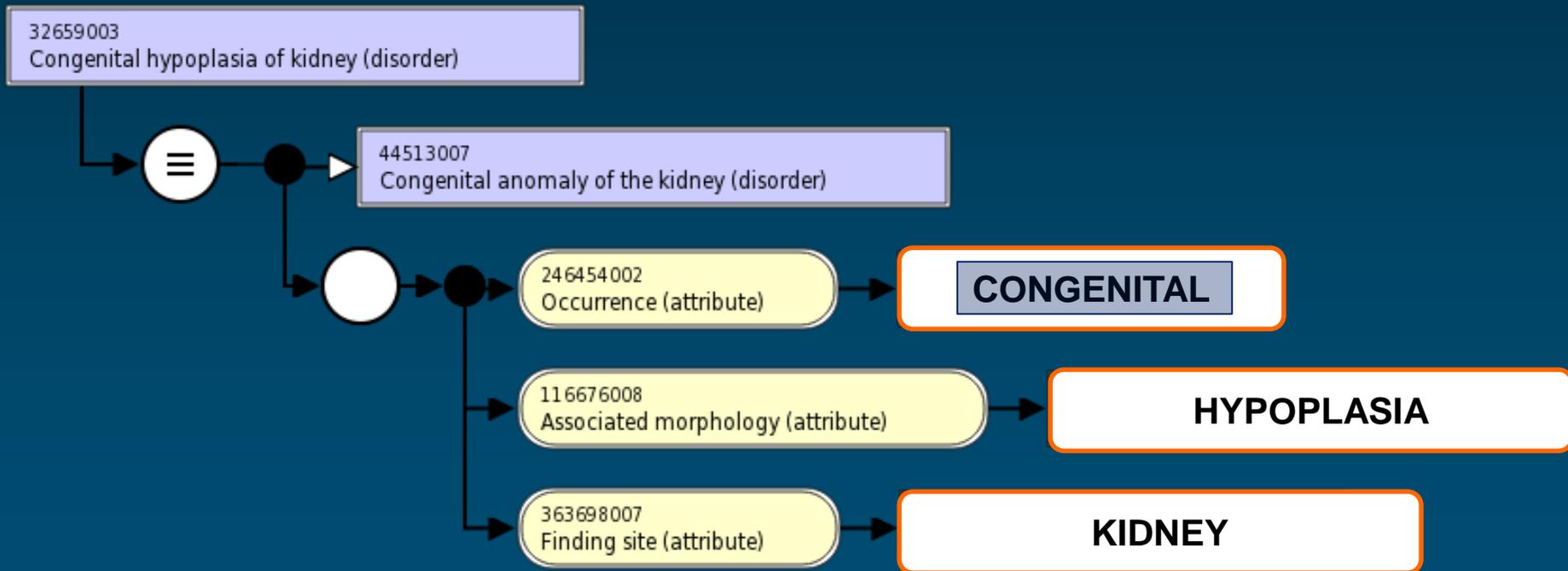
Coverage
Granularity
Mapping
Representation
Context

Description: 'Renal hypoplasia'

Equivalent To



- 'has part' some (hypoplastic and ('inheres in' some kidney) and ('has modifier' some abnormal))



Context for Renal hypoplasia

- ◆ Additional context in each representation
 - Abnormal
 - HPO: inherited from the definition of hypoplastic
 - SNOMED CT: implied from being under disorder
 - Congenital
 - SNOMED CT: part of the definition of renal hypoplasia (synonym for congenital hypoplasia of kidney)
 - HPO: implied from usage (?)

Other context issues

- ◆ Ductus arteriosus (anatomical structure)
 - Syn. for Patent ductus arteriosus (condition)
 - Ductus arteriosus is a normal anatomical structure in the fetus
 - Its persistence after birth is abnormal

Description: 'Patent ductus arteriosus'

Equivalent To



- **'has part' some**
(**'closure incomplete'**
and ('inheres in' some 'ductus arteriosus')
and ('has modifier' some abnormal))

Generalization issues

Phenotypes across diseases

- ◆ Across diseases (common/general)
 - Renal hypoplasia – always congenital
 - Absent Achilles reflex – Congenital? Abnormal?
 - Manifestation of peripheral neuropathy
 - Acquired (e.g., diabetic neuropathy)
 - Congenital (e.g., Autosomal recessive spastic ataxia of Charlevoix-Saguenay)
 - May be normal after 80

Description: 'Absent Achilles reflex'

Equivalent To +

Sub Class Of +

● 'Decreased/absent ankle reflexes'

The screenshot shows a hierarchical relationship in a medical ontology. At the top, a yellow header bar contains the text 'Description: 'Absent Achilles reflex''. Below this, there are two rows of text, each followed by a circular button with a plus sign. The first row is 'Equivalent To +' and the second row is 'Sub Class Of +'. Underneath the 'Sub Class Of +' row, there is a yellow circular icon followed by the text ''Decreased/absent ankle reflexes''.

Generalization issues

Phenotypes across species

- ◆ Across species
 - Enlarged cerebellum vs. large cerebellum
 - Enlarged = larger than normal
 - In reference to a given population
 - Species-specific
 - Large = large

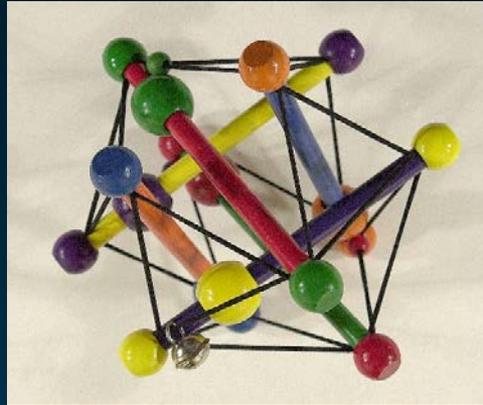


Summary

- ◆ Coverage
 - Leverage DL to refine existing concepts as needed
- ◆ Granularity
 - Not always an issue
- ◆ Mapping
 - Various kinds for various purposes
- ◆ Representation
 - Different models between HPO and SNOMED CT
- ◆ Context
 - Implicit context may impede generalization

Loosely based on 3 papers

- ◆ Winnenburg R, Bodenreider O.
Coverage of phenotypes in standard terminologies.
Proceedings of the Joint Bio-Ontologies and BioLINK at ISMB'2014
SIG session "Phenotype Day" 2014:41-44.
- ◆ Dhombres F, Winnenburg R, Case JT, Bodenreider O.
Extending the coverage of phenotypes in SNOMED CT through post-
coordination.
Stud Health Technol Inform (Proc Medinfo) 2015:(in press).
- ◆ Dhombres F, Bodenreider O.
Investigating the lexico-syntactic properties of phenotype terms –
Application to interoperability between HPO and SNOMED CT.
Proceedings of the Joint Bio-Ontologies at ISMB'2015 SIG session
"Phenotype Day" 2015:8-11.



Medical Ontology Research

Contact: olivier@nlm.nih.gov

Web: <http://mor.nlm.nih.gov>



Olivier Bodenreider

Lister Hill National Center
for Biomedical Communications
Bethesda, Maryland - USA