Ontological, epistemological and terminological aspects of phenotypes

Olivier Bodenreider

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

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

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

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.
#508836 CARNITINE PALMITOYLTRANSFERASE II DEFICIENCY, LETHAL NEONATAL; CARNITINE PALMITOYLTRANSFERASE...
CPT2

#512513 CHROMOSOME 2P16.1-P15 DELETION SYNDROME
-

#514209 MECKEL SYNDROME, TYPE 9; MKS9
B9D1

#514527 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
-
Annotation of phenotypes in OrphaNet

**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, hypogonadism, 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 Alstrom, 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.
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
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
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/Q61.4]
Integrating subdomains

- Clinical repositories
- Genetic knowledge bases
- Biomedical literature
- Genome annotations
- GO
- MeSH
- OMIM
- SNOMED CT
- HPO
- NCBI Taxonomy
- FMA
- Anatomy
- Model organisms
- Phenotypes
Integrating subdomains

- Clinical repositories
- Genetic knowledge bases
- Biomedical literature
- Genome annotations
- Anatomy
- Model organisms
- Phenotypes
Terminology integration

Multicystic kidney (204962002)

Clinical repositories

Genetic knowledge bases

SNOMED CT

OMIM

Biomedical literature

Multicystic Dysplastic Kidney (D021782)

Renal hypoplasia (HP:0000089)

Phenotypes

Model organisms

NCBI Taxonomy

FMA

Anatomy

Genome annotations

MeSH

UMLS C3714581

MeSH
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

“Renal hypoplasia”
[HPO:HP_0000089]

SNOMED CT

“Congenital hypoplasia of kidney”
[SCTID:32659003]
synonym “renal hypoplasia”

MAPPING THROUGH PRE-COORDINATION

UMLS
Mapping through pre-coordination

HPO

“Renal hypoplasia”
[HPO:HP_0000089]

“Macular hypoplasia”
[HPO:HP_00001104]

SNOMED CT

“Congenital hypoplasia of kidney”
[SCTID:32659003]
synonym “renal hypoplasia”

MAPPING THROUGH PRE-COORDINATION

UMLS
Logical definition

**CONGENITAL HYPOPLASIA OF KIDNEY**

- **SNOMED CT**
  - synonym “renal hypoplasia”

- **SCTID:** 32659003

- **32659003**
  - Congenital hypoplasia of kidney (disorder)

- **44513007**
  - Congenital anomaly of the kidney (disorder)

- **246454002**
  - Occurrence (attribute)

- **116676008**
  - Associated morphology (attribute)

- **363698007**
  - Finding site (attribute)

- **CONGENITAL**

- **HYPOPLASIA**

- **KIDNEY**
Logical definition

CONGENITAL HYPOPLASIA KIDNEY

SNOMED CT

246454002 Occurrence (attribute)
116676008 Associated morphology (attribute)
363698007 Finding site (attribute)
25539007 Congenital (qualifier value)
55199003 Hypoplasia (qualifier value)
55199003 Hypoplasia (qualifier value)
"Congenital hypoplasia of macula"  
[SCTID:xxxx]

This is a post-coordinated expression…

Logical definition (modified)

SNOMED CT

CONGENITAL

HYPOPLASIA

MACULA

246454002  
Occurrence (attribute)

116676008  
Associated morphology (attribute)

363698007  
Finding site (attribute)
Logical definition (modified)

“Congenital hypoplasia of macula” [SCTID:xxxx]

... for a specific anatomical entity

This is a post-coordinated expression...
Logical definition (generalized)

Generalization

SNOMED CT

CONGENITAL

HYPOPLASIA

<ANATOMICAL STRUCTURE>
This is a template for HPO terms...
Methods

This is a template for HPO terms...

SNOMED CT

... for any anatomical entity

<TEMPLATE>
<ANATOMICAL STRUCTURE>{hypoplasia}

CONGENITAL

HYPOPLASIA

246454002 Occurrence (attribute)

116676008 Associated morphology (attribute)

363698007 Finding site (attribute)
Mapping through post-coordination

“Renal hypoplasia”  
[HPO:HP_0000089]

“Congenital hypoplasia of kidney”  
[SCTID:32659003]  
synonym “renal hypoplasia”

“Macular hypoplasia”  
[HPO:HP_00001104]

“Congenital hypoplasia of macula”  
[SCTID:xxxx]

MAPPING THROUGH PRE-COORDINATION

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
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**


<table>
<thead>
<tr>
<th>Disease</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>#614527 CHROMOSOME 17Q12 DELETION SYNDROME</td>
<td>-</td>
</tr>
<tr>
<td>MOSAIC VARIEGATED ANEUPLOIDY SYNDROME</td>
<td>CEP57 ; BUB1B ; BUB1 ; BUB3</td>
</tr>
<tr>
<td>BOR SYNDROME</td>
<td>SIX5 ; SIX1 ; EYA1</td>
</tr>
<tr>
<td>BARDET-BIEDEL SYNDROME</td>
<td>MKKS ; SDCCAG8 ; WDPCP ; BBS5 ; BBS1 ; TRIM32 ; BBS2 ; IFT27 ; ARL6 ; BBS4 ; CEP290 ; BBS12 ; LZTFL1 ; MKS1 ; BBS10 ; BBIP1 ; NPHP1 ; BBS7 ; IFT172 ; BBS9 ; TTC8</td>
</tr>
<tr>
<td>SHORT RIB-POLYDACTYLY SYNDROME</td>
<td>-</td>
</tr>
<tr>
<td>CARNITINE PALMITOYLTRANSFERASE II DEFICIENCY</td>
<td>-</td>
</tr>
<tr>
<td>NON-RHIZOMELIC CHONDRODYSPLASIA PUNCTATA</td>
<td>-</td>
</tr>
<tr>
<td>RENAL DYSPLASIA - MEGALOCYSTIS - SIRENOMELIA</td>
<td>-</td>
</tr>
<tr>
<td>INDOMETHACIN EMBRYOFETOPATHY</td>
<td>-</td>
</tr>
</tbody>
</table>
Coarse phenotyping

◆ 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”

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

“Renal hypoplasia” [HPO:HP_0000089]
Logical definition in 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)

32659003
Congenital hypoplasia of kidney (disorder)

44513007
Congenital anomaly of the kidney (disorder)

24645002
Occurrence (attribute)

116676008
Associated morphology (attribute)

363698007
Finding site (attribute)

CONGENITAL

HYPOPLASIA

KIDNEY
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 …]
Logically defined as:

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

Concepts:
- 32659003: Congenital hypoplasia of kidney (disorder)
- 44513007: Congenital anomaly of the kidney (disorder)
- 246454002: Occurrence (attribute)
- 116676008: Associated morphology (attribute)
- 363698007: Finding site (attribute)

Keywords:
- CONGENITAL
- HYPOPLASIA
- KIDNEY
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
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 +
SubClass Of +
'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


Medical Ontology Research

Contact:  olivier@nlm.nih.gov

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