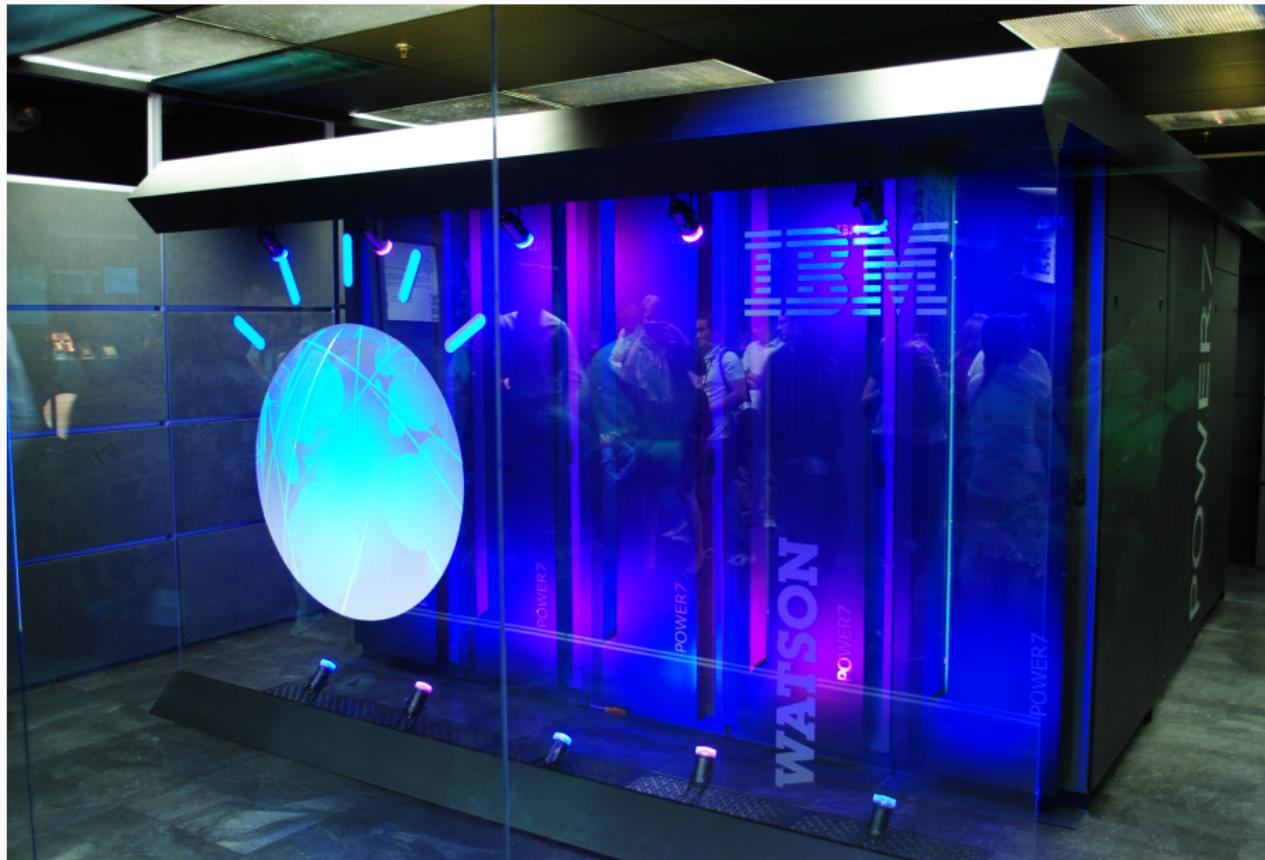


# PRINCIPLES OF BIOMEDICAL ONTOLOGIES & IMPORTANCE OF SEMANTIC INTEROPERABILITY FOR DATA INTEGRATION

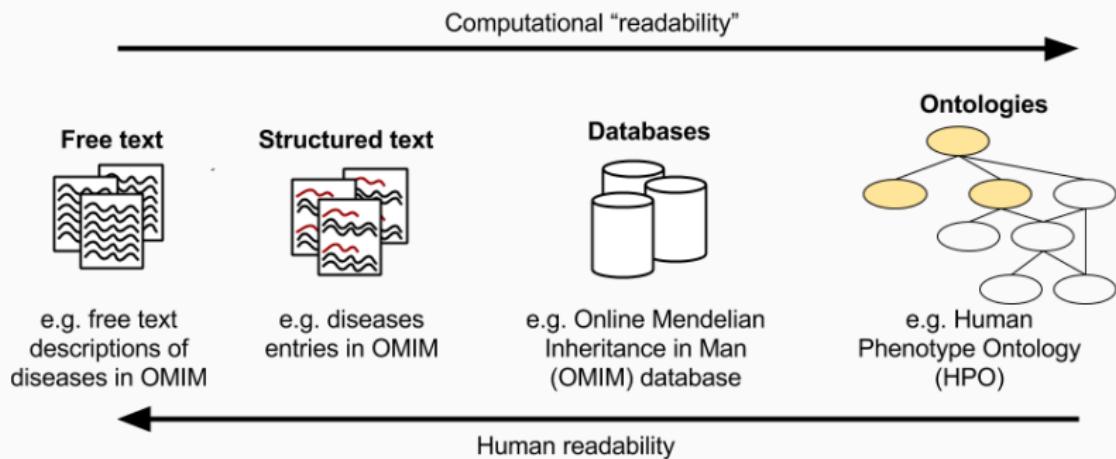
---

Anika Oellrich



Source: [https://commons.wikimedia.org/wiki/File:IBM\\_Watson.PNG](https://commons.wikimedia.org/wiki/File:IBM_Watson.PNG), author: Clockready

# COMPUTATIONAL “READABILITY”



# OUTLINE

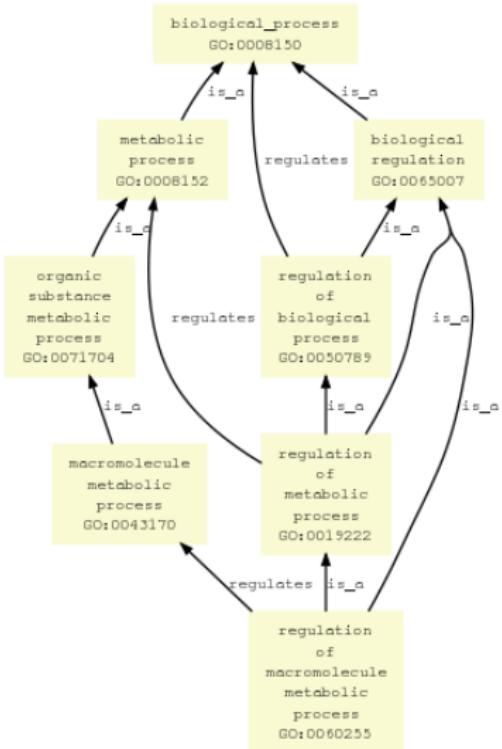
1. Ontologies in detail
2. Examples of medical resources
3. Ontologies and data integration

## A LITTLE BIT ABOUT MYSELF

- Senior PostDoc at KCL
- used ontologies for text and data mining, data integration
- worked with model organism data in the past (International Mouse Phenotyping Consortium project)
- now utilising data from electronic health records (EHRs)

## Ontologies in detail

# BEST WAY TO IMAGINE ONTOLOGY: GRAPH



Source: Generated with old Ontology Lookup Service (OLS)

Definition of e.g. *regulation of biological process*

## Class relations

### Equivalent to:

- [biological regulation](#) and [regulates some biological\\_process](#)

### Subclass of:

- [biological regulation](#)
- [regulates some biological\\_process](#)

Source: Generated with new Ontology Lookup Service (OLS)

## OPEN BIOMEDICAL ONTOLOGIES (OBO) FORMAT

```
[Term]
id: HP:0000003
name: Multicystic kidney dysplasia
alt_id: HP:0004715
def: "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."
[HPO:curators] comment: 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.
synonym: "Multicystic dysplastic kidney" EXACT []
synonym: "Multicystic kidneys" EXACT []
synonym: "Multicystic renal dysplasia" EXACT []
xref: MeSH:D021782 "Multicystic Kidney Dysplasia"
xref: UMLS:C0345335 "Multicystic Kidney Dysplasia"
is_a: HP:0000107 ! Renal cyst
```

[ftp://ftp.geneontology.org/pub/go/www/GO.format.obo-1\\_4.shtml](ftp://ftp.geneontology.org/pub/go/www/GO.format.obo-1_4.shtml)

# WEB ONTOLOGY LANGUAGE (OWL)

The screenshot shows the Protege OWL editor interface with the following details:

- Top Bar:** Shows the ontology name "hp" and its URL (<http://purl.obolibrary.org/obo/hp/releases/2016-01-13/hp.owl>).
- Left Sidebar:** Contains tabs for Active Ontology, Entities, and Individuals by class. The Individuals by class tab is selected.
- Class Hierarchy (Inferred):** A tree view showing the inheritance path of the class "Multicystic kidney dysplasia". The path includes:
  - 'Abnormality of the renal medull'
  - 'Abnormality of the renal pelvis'
  - 'Enlarged kidneys'
  - 'Hyperechogenic kidneys'
  - 'Nephrocalcinosis'
  - 'Nephrogenic rest'
  - 'Nephrolithiasis'
  - 'Nephrosclerosis'
  - 'Perirenal hematoma'
  - 'Renal amyloidosis'
  - 'Renal atrophy'
  - 'Renal cyst'
    - 'Cystic renal dysplasia'
    - 'Multicystic kidney dysplasia'** (highlighted in blue)
    - 'Multiple renal cysts'
    - 'Multiple small medullary ren'
- Annotations:** Tab for "Annotations: Multicystic kidney dysplasia".
  - Annotations:** A list of annotations:
    - label** [type: string]: Multicystic kidney dysplasia
    - id** [type: string]: HP:0000003
    - has\_alternative\_id** [type: string]: HP:0004715
  - Description:** Description of the class: "Multicystic kidney dysplasia".
    - Equivalent To:** A list of equivalent classes:
      - 'has part' some (polycystic and ('inheres in' some kidney) and ('has modifier' some abnormal))
    - SubClass Of:** A list of superclasses:
      - 'Renal cyst'
    - General class axioms:** A list of general axioms:
      - 'abnormal phenotype'
      - 'has part' some (quality and ('inheres in part of' some 'renal system') and ('has modifier' some abnormal))**
      - 'has part' some**

Example of medical resources

## HUMAN PHENOTYPE ONTOLOGY

- ~11k concepts
- ~115k annotations for hereditary & ~4k for common diseases
- reports phenotype abnormalities
- several groups working on its extension
- initial version generated from OMIM
- pre-dominantly used in research projects, e.g. in Deciphering Developmental Disorders (DDD) project

<http://human-phenotype-ontology.github.io/>, <http://www.ddduk.org/>

# SYSTEMATIZED NOMENCLATURE OF MEDICINE—CLINICAL TERMS

- SNOMED CT
- ~316k concepts
- aims to cover all aspects of healthcare (e.g. drug events, environmental factors, ...)
- it is multilingual
- pre-dominantly used in clinical environments, e.g. EHRs

[http://ihtsdo.org/fileadmin/user\\_upload/doc/download/doc\\_StarterGuide\\_Current-en-US\\_INT\\_20141202.pdf?ok](http://ihtsdo.org/fileadmin/user_upload/doc/download/doc_StarterGuide_Current-en-US_INT_20141202.pdf?ok)

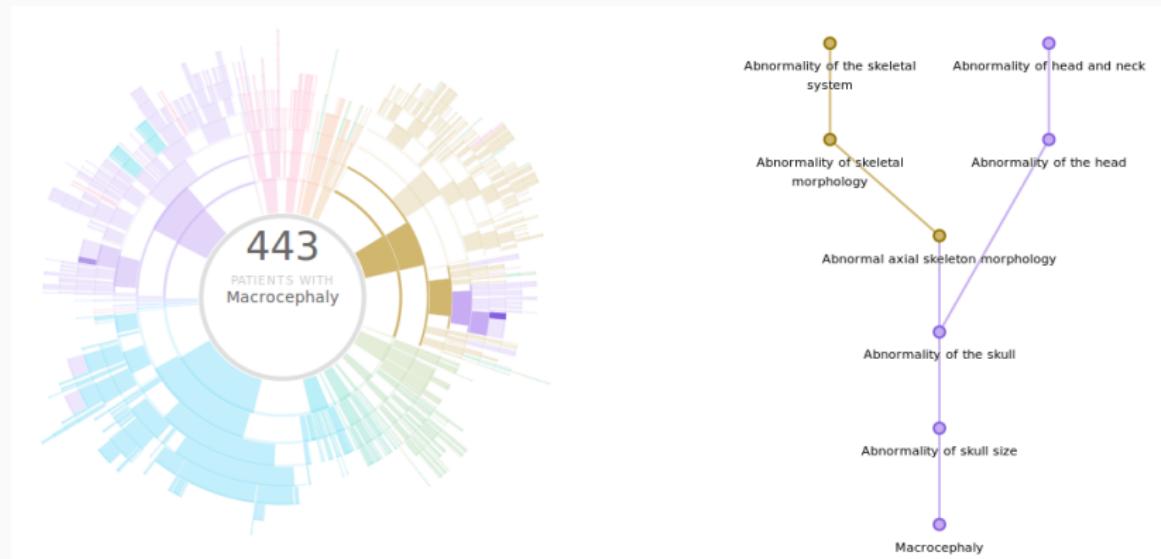
## OTHER RELEVANT RESOURCES

- Logical Observation Identifiers Names and Codes (LOINC)
- Unified Medical Language System (UMLS)
- International Classification of Diseases (ICD 9/10)
- Medical Subject Headings (MeSH)

<https://loinc.org/>, <https://www.nlm.nih.gov/research/umls/>, <http://www.who.int/classifications/icd/en/>,  
<https://www.nlm.nih.gov/mesh/>

## Ontologies and data integration

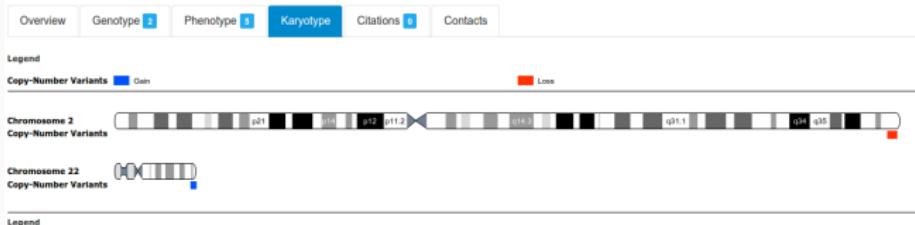
## DatabasE of genomiC varlation and Phenotype in Humans using Ensembl Resources



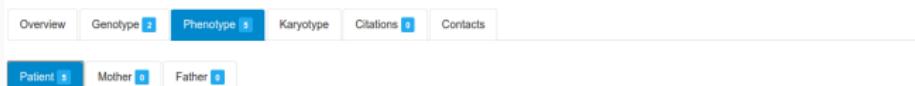
<https://decipher.sanger.ac.uk/>

# DECIPHER - PATIENT VIEW

Patient: 865



Patient: 865



Patient phenotypes

| Phenotype                    | Observation | Code       |
|------------------------------|-------------|------------|
| Abnormally low-pitched voice | present     | HP:0010300 |
| Autism                       | present     | HP:0000717 |
| Intellectual disability      | present     | HP:0001249 |
| Macrocephaly                 | present     | HP:0000256 |
| Obesity                      | present     | HP:0001513 |

# PHENOGRAMVIZ TO SEE GENES THAT COULD IMPLICATE PHENOTYPE

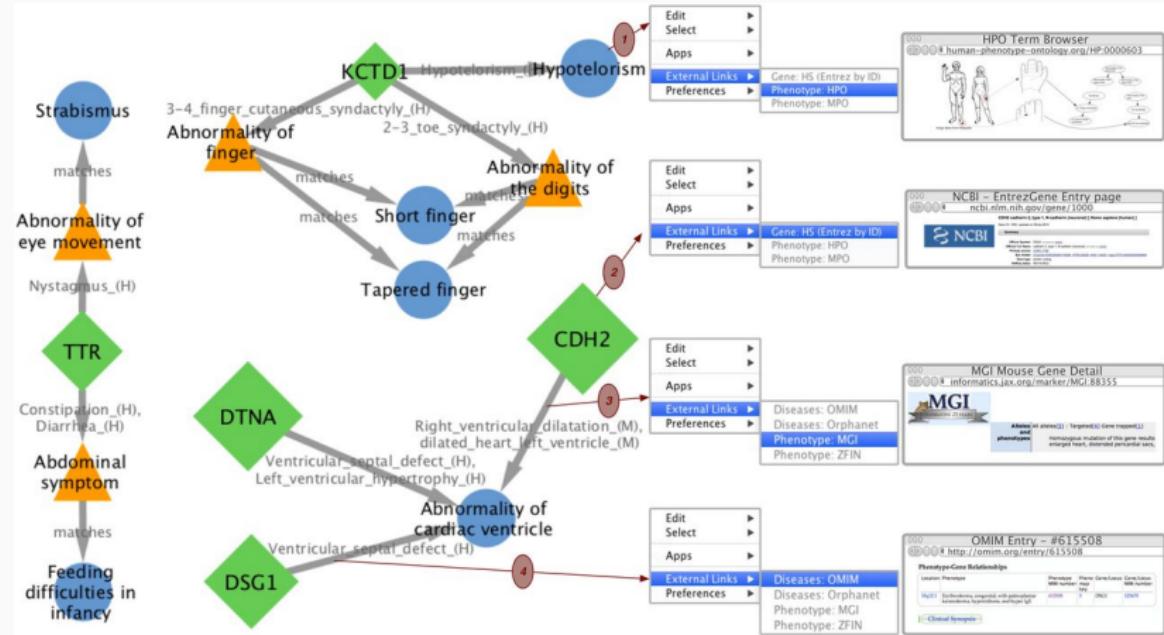


Figure taken from: Köhler, Sebastian, et al. "Clinical interpretation of CNVs with cross-species phenotype data." Journal of medical genetics 51.11 (2014): 766-772.



## Disease: CLN4B disease

Your feedback welcome!

Source: OMIM:162350

Overview

Compare

Phenotypes (13)

Genes (4)

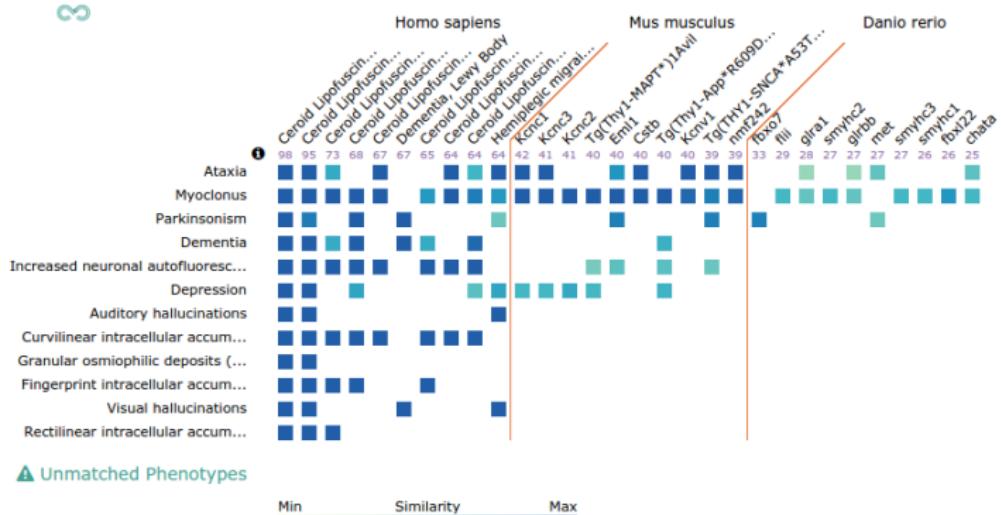
Genotypes (0)

Models (0)

Variants (2)

Pathways (1)

## Phenotype Similarity Comparison



## SUMMARY

Ontologies ...

- define meaning of concepts
- have a hierarchy on which concepts are sorted
- are defined with formal languages
- can help data visualisation
- applied for semantic similarity measures