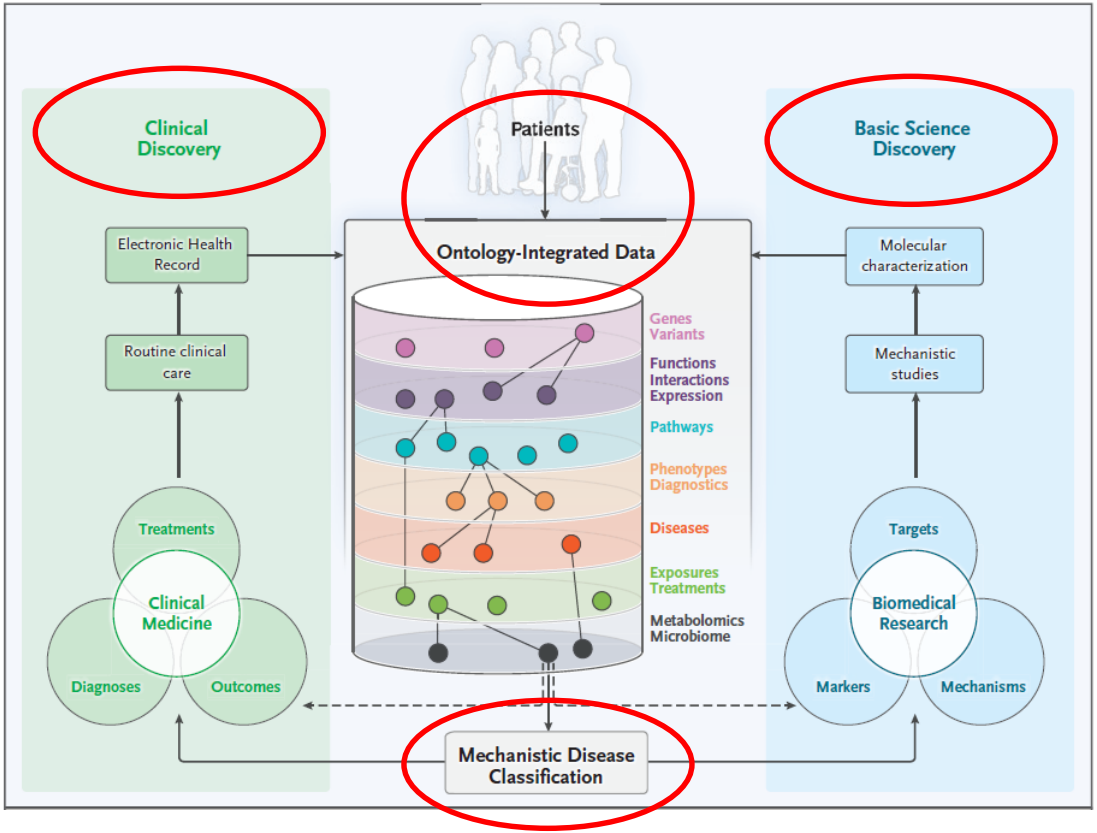
Integrative Ontology Development to Support Precision Medicine and Molecular Atlas Research

Yongqun "Oliver" He

University of Michigan Medical School

Ann Arbor, MI 48109 USA

Ontology-based mechanistic classification of disease

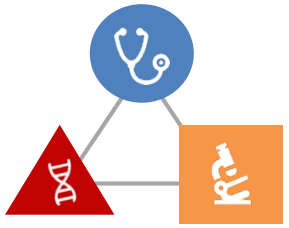


Integrating the two streams of data (**clinical** and **basic** science observations) enables more refined and dynamic classification of disease across many data types

Citation: Haendel MA, Chute CG, Robinson PN. **Classification, Ontology, and Precision Medicine**. *N Engl J Med*. 2018 Oct 11; 379(15): 1452-1462.

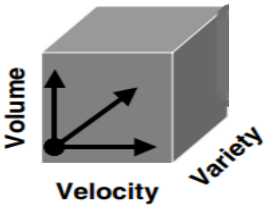
KPMP, Kidney Research, and Ontology

- Initiated 2017, Kidney Precision Medicine Project (**KPMP**), funded by NIH-NIDDK, involves >20 institutes
- **Goals**:
  - Build a kidney tissue atlas that links *clinical* phenotypes, cells, *molecules*, pathways, and *pathology* together.
  - Understand and treat human kidney diseases – Acute Kidney injury (**AKI**) and Chronic Kidney Disease (**CKD**)



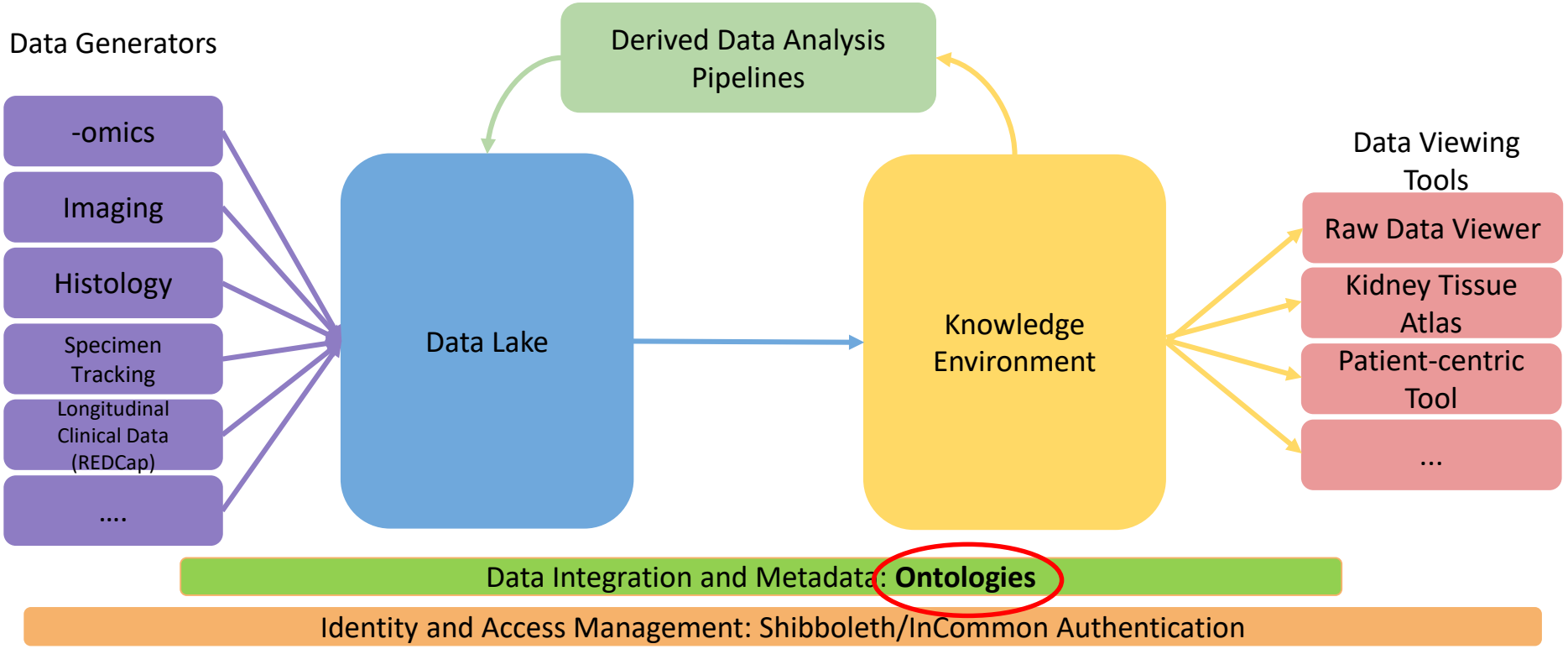
- Clinical
- Molecular
- Pathology

- **Big data challenges**:
  - high volume, velocity, variety
  - Standardization, integration, sharing, and analysis
- Ontology is critical for solving the challenges



Ontology plays a critical role in KPMP

KPMP Data Flow:



Two ontologies for KPMP

- Two community-based KPMP ontologies:
  - **KTAO: Kidney Tissue Atlas Ontology** – It's more about kidney knowledge
  - **OPMI: Ontology of Precision Medicine and Investigations** – Standardizes data and metadata types in and beyond KPMP.
  - Kidney-related info in OPMI is imported back to KTAO.

- Ontology development strategies
  - Follow Open Biomedical Ontology (**OBO**) **principles**: Openness, collaboration, etc. → >150 OBO library ontologies: **non-redundant**, **interoperable**
  - **Reuse/align/integrate** existing ontologies: UBERON anatomical entity, HPO (Human Phenotypes), GO, CL (Cells), OBI (Biomedical Investigations), …
  - ***Top-down***: align with top level ontologies
  - ***Bottom-up***: address use cases. – Community collaboration is important.

Ref: He Y, Xiang Z, Zheng J, Lin Y, Overton JA, Ong E. The **eXtensible ontology development (XOD)** principles and tool implementation to support ontology interoperability. *J Biomed Semantics*. 2018 Jan 12;9(1):3.

KTAO: Kidney Tissue Atlas Ontology

- KTAO GitHub website: https://github.com/KPMP/KTAO
- Deposited at:
  - BioPortal: https://bioportal.bioontology.org/ontologies/KTAO
  - Ontobee: http://www.ontobee.org/ontology/KTAO
- Statistics: includes >5000 terms.
- KTAO includes **>250** kidney disease markers and their linkages to cells/diseases

Question: how KTAO organizes these entities and link them?

Ref: He Y, Steck B, Ong E, Mariani L, Lienczewski C, Balis U, Kretzler M, Himmelfarb J, Bertram JF, Azeloglu E, Iyengar R, Hoshizaki D, Mooney SD, for the KPMP Consortium. **KTAO:** A kidney tissue atlas ontology to support community-based kidney knowledge base development and data integration (http://ceur-ws.org/Vol-2285/ICBO_2018_paper_28.pdf). *International Conference on Biomedical Ontology* 2018 (ICBO-2018), August 7-10, 2018, Corvallis, Oregon, USA. Pages 1-6.

Start with a simple case:
Kidney panel gene marker example: WT1

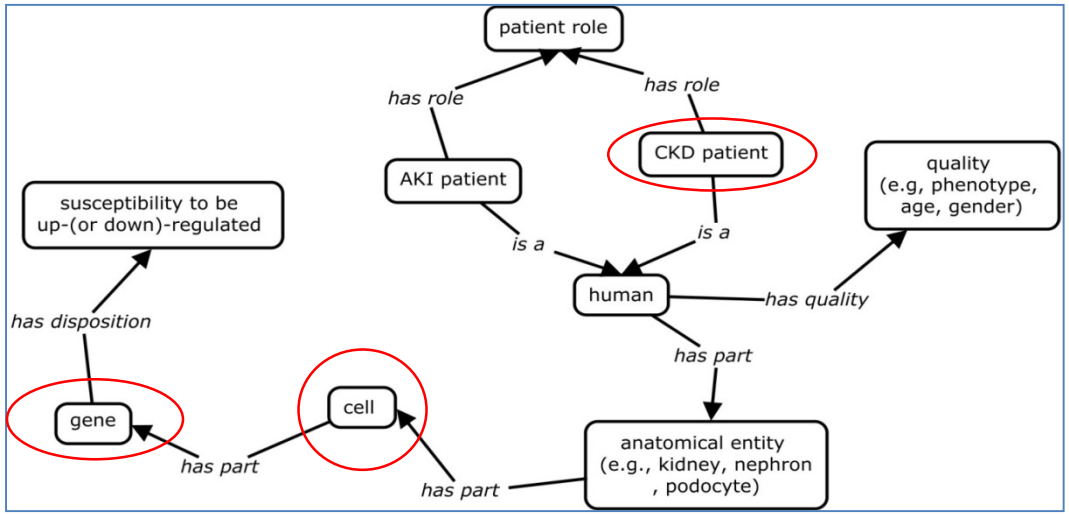
- WT1: Wilm's tumour protein
- WT1 is a transcriptional factor required for podocyte development and homeostasis

- Our gene panel data indicates:
  WT1 gene is differentially regulated in podocytes of CKD patients
  → WT1 is a CKD gene marker.
  → Up- or down-regulation may vary given conditions

- How to represent this and other knowledge in KTAO?

Reference: M. Kann, S. Ettou, Y. L. Jung, M. O. Lenz, M. E. Taglienti, P. J. Park*, et al.*, "Genome-Wide Analysis of Wilms' Tumor 1-Controlled Gene Expression in Podocytes Reveals Key Regulatory Mechanisms," *J Am Soc Nephrol,* vol. 26, pp. 2097-104, Sep 2015.

KTAO design pattern that links kidney-related entities



- Reuse/align ontologies for entities
- Generate and use ***new relations*** to link entities. Such relations often are not in existing ontologies.

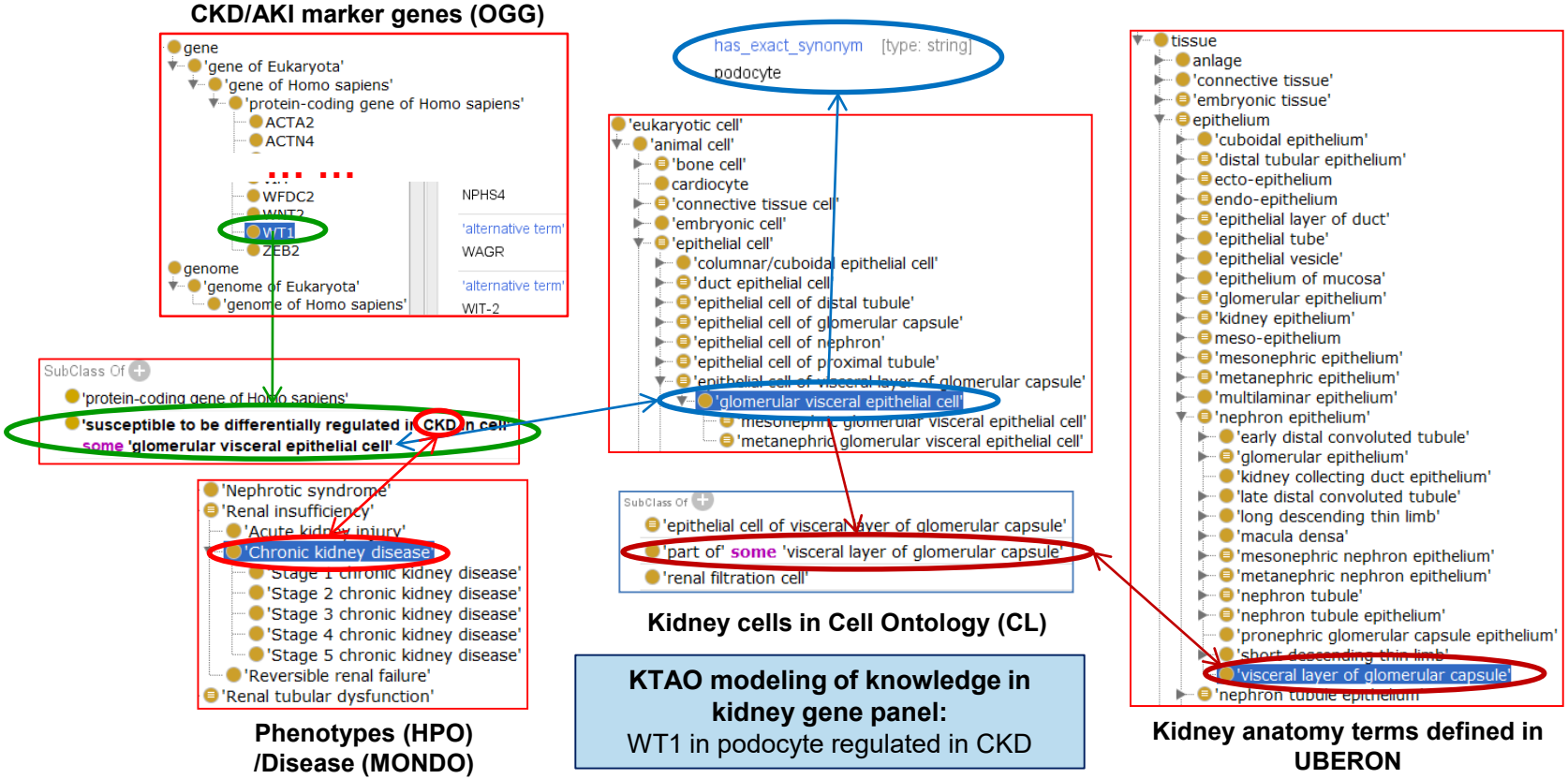
Meanwhile, generate a new KTAO relation:

'***susceptible to be differentially regulated in CKD in cell***'

→ *link a gene vs a cell*

where the gene (e.g., WT1) is susceptible to be differentially regulated in the cell (e.g., podocyte) of CKD patients

KTAO: Reuses and links UBERON (for anatomy), CL (for cell types), HPO (for phenotype), DP (for diseases), and OGG (for genes)



Usage demo: SPARQL query of KTAO knowledge

```
#Goal: To find gene markers regulated in podocytes of kidney patients

PREFIX podocyte: <http://purl.obolibrary.org/obo/CL_0000653>
SELECT distinct ?gene STR(?gene_label) AS ?gene_label ?r_label "podocyte"
FROM <http://purl.obolibrary.org/obo/merged/KTAO>
WHERE
{
   ?gene rdfs:label ?gene_label .
   ?r rdfs:label ?r_label .
   ?gene rdfs:subClassOf ?gene_restriction .
   ?gene_restriction owl:onProperty ?r; owl:someValuesFrom podocyte: .
}
```

Output format Table ▼ **Max Rows** 10 ▼

Run Query | Reset

Result | Raw Request/Permalinks | Raw Response

gene	gene_label	r_label	callret-3
http://purl.obolibrary.org/obo/OGG_3000000301	ANXA1	"susceptible to be up-regulated in CKD in cell"@en	podocyte
http://purl.obolibrary.org/obo/OGG_3000001285	COL4A3	"susceptible to be up-regulated in CKD in cell"@en	podocyte
http://purl.obolibrary.org/obo/OGG_3000053405	CLIC5	"susceptible to be up-regulated in CKD in cell"@en	podocyte
http://purl.obolibrary.org/obo/OGG_3000004868	NPHS1	"susceptible to be up-regulated in CKD in cell"@en	podocyte
http://purl.obolibrary.org/obo/OGG_3000007490	WT1	"susceptible to be differentially regulated in CKD in cell"@en	podocyte

A few lines of SPARQL query script identified 5 human genes in podocytes regulated in CKD patients

OPMI: Ontology of Precision Medicine and Investigation

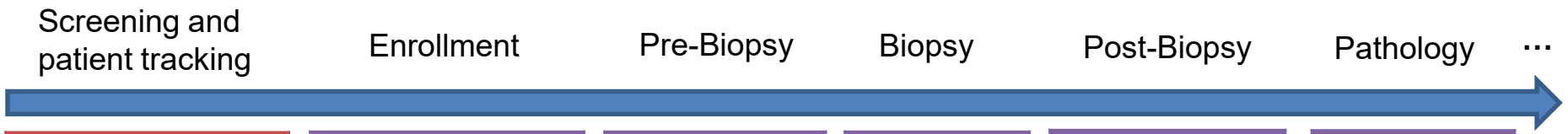
- OPMI is an OBO library ontology
- Currently focuses on clinical data/metadata:
  - Clinical Case Report Forms (CRFs).
  - Clinical common data models (CDMs)
- Clinical factors affect Omics and imaging results:
  - ***Mouse or cell model***: not many
  - ***Human***: likely hundreds → difficult to handle
  - KPMP data are all ***human*** data
  - OPMI critical to capture clinical factors

Gene	Exp. Cond. 1	Exp. Cond. 2
Gene 1	Value 11	Value 12
Gene 2	Value 21	Value 22
… …	… …	… …

Gene	Exp. Cond. 1	Exp. Cond. 2	Clinical factor X
Gene 1	Value 11	Value 12	Value 1x
Gene 2	Value 21	Value 22	Value 2x
… …	… …	… …	… …

Clinical factors support gene expression data analysis

38 KPMP Case Report Forms (CRFs)



Screening and patient tracking	Enrollment	Pre-Biopsy	Biopsy	Post-Biopsy	Pathology
New Patient	Clinic reception	Pre-Biopsy clinician questionnaire	Kidney biopsy procedure details	Post biopsy hospitalization	Dx image & tissue QC
Eligibility assessment	Demographic info	Biopsy safety checklist		Tissue tracking	Dx Core disease category assignment
Consent	Medical history			Tissue interrogation image/data upload	Dx Core visual assessment
Contact info	Personal history			Dx Core image scanning and upload	Interrogation Core tissue QC
Participant study status	Physical measure			Patient follow-up	
Medications	Biosample collection				
Adverse Event	Hospitalization				
	PROMIS questionnaire				
	Lab results				

~3000 questions in these forms

KPMP Clinical Metadata Identified from CRFs

Metadata types	Metadata Examples
Quality and measurements	Measurement protocol details (e.g., arm and stand/sit/lay position in blood pressure measurement)
Health conditions	Comorbities, pregnancy, adverse events
Medical interventions	drug medication, prior surgeries transplantation, dialysis, biopsy, transplantation
Substances exposed to	Additional prescription drugs, recreation drugs, cigarettes, alcohols
Socioeconomic factors	employment status, race, ethnicity, education, income, Insurance
Environmental	county, state, country, hospital, primary care location
Biosample	collection time, processing time, transportatoin tracking, biopsy location, storage location, storage time
Patient reported outcomes	patient experience, life quality, pain, anxiety, complication, likert scale
Patient study status tracking	pass or fail screening, whether informed consent signed, is active in study? is live?
Electronic health record (EHR)	source of EHR, record availability, processing/harmonization method

Clinical Common Data Models (CDMs) and OPMI

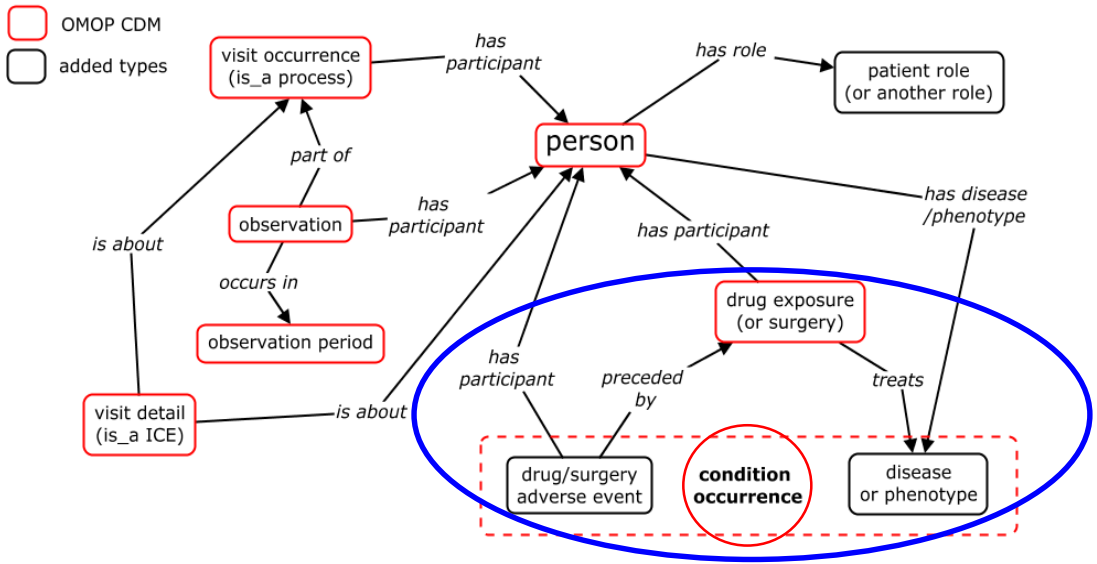
- Many clinical CDMs exist to support data standardization

- Observational Medical Outcomes Partnership (OMOP) CDM
  - Developed by OHDSI
  - >1 billion patient records

- Weak semantics, and interoperability among CDMs
  - Can OPMI help solve these issues?

OPMI to ontologize OMOP CDM

For example, OPMI modeling of *adverse event* (AE):



OMOP CDM:

- Do not differentiate different types of "condition occurrence", like AE and natural disease / phenotype.

OPMI modeling:

- AE always occurs after medical intervention, e.g., drug use, **surgery**.
- Support data analysis.

Citation: He Y, Ong E, Zheng J, Wan L, Schaub J, Kretzler M. Ontological representation of OMOP CDM using the OBO framework. 2018 OHDSI Symposium, Oct 12, 2018, Bethesda, MD, USA.

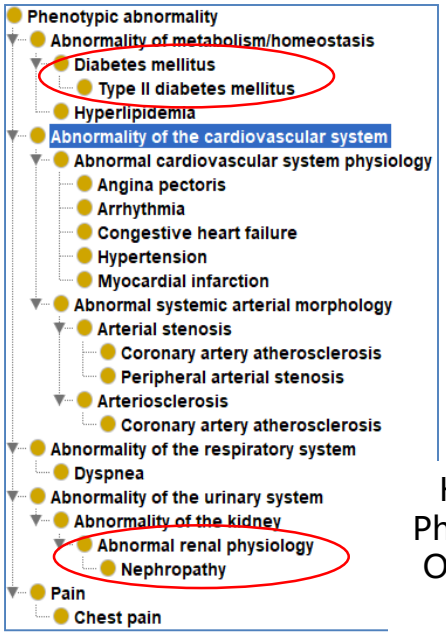
Heart surgery-associated Acute Kidney Injury (AKI) AE using IQVIA OHDSI/OMOP data

(*Prior knowledge*: incidence of AKI after heart surgery is up to 30-50%)

Algorithm:

30 days before surgery (no AKI) → **Heart surgery** → **AKI AE within 14 days**

Conditions 30 days before heart surgery:



Human Phenotype Ontology (**HPO**)

→ **Operation on heart (SNOMED: 4275564)** and its subclasses →

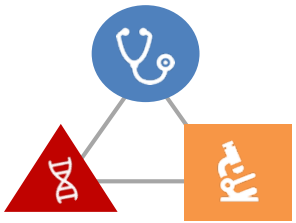
- *15,548* patients in the AKI AE cohort.
- Sex effect:
  - *Male*: 72%;
  - *Female*: 28%
- Age effect:
  - Age > 55: 78.5%

Finding: Phenotypes including **Type II diabetes** & **Nephropathy** are often observed before heart surgery-associated AKI adverse event.

Summary

- KTAO standardizes all components of Kidney Atlas and their relations.
- OPMI standardizes data/metadata, and common data models (CDMs).

Discussion



- **Ongoing**: Ontology representation of clinical tissue interrogation, assays, pathology, molecular pathway, for advanced data integration.
- How the KPMP ontology research supports HubMAP (Human BioMolecular Atlas Program)?

Acknowledgements

University of Michigan
- Edison Ong
- Jennifer Schaub
- Ling Wan
- Becky Steck
- Ulysses Balis
- Matthias Kretzler

Univ. of Washington
- Sean Mooney
- Fred Dowd
- Christopher Park
- Jonathan Himmelfarb

Mount Sinai
- Ravi Ulyengar

Cleveland Clinic
- John O'Toole

Duke University
- Laura Barisoni

Washington Univ.
- Sanjay Jain

IQVIA, UK
- Anastasios Siapos
- Christian Reich
- Sarah Seager

NIH NIDDK
- Deborah K. Hoshizaki
- Robert A. Star

Ontology/terminology community
- Chris Mungall (Lawrence Berkeley NL)
- Melissa Haendel (OHSU)
- Chris Stockert (UPenn)
- Jie Zheng (UPenn)
- Chunhua Weng (Columbia U.)
- Harold Solbrig (Mayo Clinic)
- Alexander Diehl (Univ. at Buffalo)
- Barry Smith (Univ. at Buffalo)
- Hong Yu (Guizhou Medical School)
- Peter Robinson (Jackson Lab)

KPMP Funding: NIH NIDDK
U2C Project #: 1U2CDK114886-01