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.

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)
- 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:

Data Generators

- -omics
- Imaging
- Histology
- Specimen Tracking
- Longitudinal Clinical Data (REDCap)

Data Lake

Derived Data Analysis Pipelines

Knowledge Environment

Data Viewing Tools

- Raw Data Viewer
- Kidney Tissue Atlas
- Patient-centric Tool
- ...

Ontologies

Data Integration and Metadata

Identity and Access Management: Shibboleth/InCommon Authentication
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.

KTAO: Kidney Tissue Atlas Ontology

• KTAO GitHub website: https://github.com/KPMP/KTAO
• Deposited at:
  o BioPortal: https://bioportal.bioontology.org/ontologies/KTAO
  o 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?

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?

KTAO design pattern that links kidney-related entities

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

- Reuse/align ontologies for entities
- Generate and use **new relations** to link entities. Such relations often are not in existing ontologies.
KTAO: Reuses and links UBERON (for anatomy), CL (for cell types), HPO (for phenotype), DP (for diseases), and OGG (for genes)

CKD/AKI marker genes (OGG)

Phenotypes (HPO)/Disease (MONDO)

Kidney cells in Cell Ontology (CL)

KTAO modeling of knowledge in kidney gene panel:
WT1 in podocyte regulated in CKD

Kidney anatomy terms defined in UBERON
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

<table>
<thead>
<tr>
<th>Gene</th>
<th>Exp. Cond. 1</th>
<th>Exp. Cond. 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene 1</td>
<td>Value 11</td>
<td>Value 12</td>
</tr>
<tr>
<td>Gene 2</td>
<td>Value 21</td>
<td>Value 22</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene</th>
<th>Exp. Cond. 1</th>
<th>Exp. Cond. 2</th>
<th>Clinical factor X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene 1</td>
<td>Value 11</td>
<td>Value 12</td>
<td>Value 1x</td>
</tr>
<tr>
<td>Gene 2</td>
<td>Value 21</td>
<td>Value 22</td>
<td>Value 2x</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Clinical factors support gene expression data analysis
38 KPMP Case Report Forms (CRFs)

Screening and patient tracking
- New Patient
- Eligibility assessment
- Consent
- Contact info
- Participant study status
- Medications
- Adverse Event

Enrollment
- Clinic reception
- Demographic info
- Medical history
- Personal history
- Physical measure
- Biosample collection
- Hospitalization
- PROMIS questionnaire
- Lab results
- Pre-Biopsy clinician questionnaire
- Biopsy safety checklist

Pre-Biopsy

Biopsy
- Kidney biopsy procedure details

Post-Biopsy
- Post biopsy hospitalization
- Tissue tracking
- Tissue interrogation image/data upload
- Dx Core image scanning and upload
- Patient follow-up

Pathology
- Dx image & tissue QC
- Dx Core disease category assignment
- Dx Core visual assessment
- Interrogation Core tissue QC

~3000 questions in these forms
<table>
<thead>
<tr>
<th>Metadata types</th>
<th>Metadata Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality and measurements</td>
<td>Measurement protocol details (e.g., arm and stand/sit/lay position in blood pressure measurement)</td>
</tr>
<tr>
<td>Health conditions</td>
<td>Comorbidities, pregnancy, adverse events</td>
</tr>
<tr>
<td>Medical interventions</td>
<td>drug medication, prior surgeries transplantation, dialysis, biopsy, transplantation</td>
</tr>
<tr>
<td>Substances exposed to</td>
<td>Additional prescription drugs, recreation drugs, cigarettes, alcohols</td>
</tr>
<tr>
<td>Socioeconomic factors</td>
<td>employment status, race, ethnicity, education, income, Insurance</td>
</tr>
<tr>
<td>Environmental</td>
<td>county, state, country, hospital, primary care location</td>
</tr>
<tr>
<td>Biosample</td>
<td>collection time, processing time, transportatoin tracking, biopsy location, storage location, storage time</td>
</tr>
<tr>
<td>Patient reported outcomes</td>
<td>patient experience, life quality, pain, anxiety, complication, likert scale</td>
</tr>
<tr>
<td>Patient study status tracking</td>
<td>pass or fail screening, whether informed consent signed, is active in study? is live?</td>
</tr>
<tr>
<td>Electronic health record (EHR)</td>
<td>source of EHR, record availability, processing/harmonization method</td>
</tr>
</tbody>
</table>
Clinical Common Data Models (CDMs) and OPMI

• Many clinical CDMs exist to support data standardization

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

• Weak semantics, and interoperability among CDMs
  o 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)

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

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%
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