



Modelling kidney disease using ontology: insights from the Kidney Precision Medicine Project

Edison Ong^{1,31}, Lucy L. Wang^{2,31}, Jennifer Schaub³, John F. O'Toole^{4,5}, Becky Steck³, Avi Z. Rosenberg⁶, Frederick Dowd⁷, Jens Hansen^{8,9}, Laura Barisoni¹⁰, Sanjay Jain¹¹, Ian H. de Boer¹², M. Todd Valerius¹³, Sushrut S. Waikar¹⁴, Christopher Park¹⁵, Dana C. Crawford^{16,17,18}, Theodore Alexandrov^{19,20}, Christopher R. Anderton²¹, Christian Stoeckert²², Chunhua Weng²³, Alexander D. Diehl²⁴, Christopher J. Mungall²⁵, Melissa Haendel²⁶, Peter N. Robinson²⁷, Jonathan Himmelfarb^{12,15}, Ravi Iyengar^{8,9}, Matthias Kretzler^{1,3}, Sean Mooney²⁸✉, Yongqun He^{1,29,30}✉ and Kidney Precision Medicine Project*

Abstract | An important need exists to better understand and stratify kidney disease according to its underlying pathophysiology in order to develop more precise and effective therapeutic agents. National collaborative efforts such as the Kidney Precision Medicine Project are working towards this goal through the collection and integration of large, disparate clinical, biological and imaging data from patients with kidney disease. Ontologies are powerful tools that facilitate these efforts by enabling researchers to organize and make sense of different data elements and the relationships between them. Ontologies are critical to support the types of big data analysis necessary for kidney precision medicine, where heterogeneous clinical, imaging and biopsy data from diverse sources must be combined to define a patient's phenotype. The development of two new ontologies — the Kidney Tissue Atlas Ontology and the Ontology of Precision Medicine and Investigation — will support the creation of the Kidney Tissue Atlas, which aims to provide a comprehensive molecular, cellular and anatomical map of the kidney. These ontologies will improve the annotation of kidney-relevant data, and eventually lead to new definitions of kidney disease in support of precision medicine.

Precision medicine is broadly defined as the delivery of tailored interventions or treatments to individual patients, or as described previously “the right drug for the right patient at the right time”¹. The practice of precision medicine depends acutely on emergent high-throughput technologies that are capable of generating detailed molecular phenotypes in human biosamples. Such molecular phenotypes provide an opportunity to derive more nuanced descriptions of disease, and methods to systematically incorporate this information in drug discovery and clinical care are needed.

Kidney disease is most commonly classified as either acute kidney injury (AKI)² or chronic kidney disease (CKD)³. These terms provide information about the duration of decreased kidney function or kidney damage but do not provide diagnostic specificity. Instead, these terms are better characterized as descriptions of syndromes, which may have differing underlying

causes. Existing classification criteria for AKI and CKD are designed to provide a standardized way to stage the severity of disease and cover a broad range of cases based on changes in serum creatinine level, proteinuria and urine output. However, these criteria do not help clinicians identify causal factors that can be targeted using a precision medicine approach. For example, a clinician could conclude that a patient has KDIGO CKD classification G3bA3 (REF.³) on the basis of the patient's serum creatinine level and urinary protein excretion, but this description does not provide insights into the cause of the abnormal kidney function, which could be anything from diabetes mellitus to multiple myeloma to medication-associated nephrotoxicity.

Reassessment of definitions of kidney disease and a precision medicine approach to the treatment of kidney disease requires that molecular phenotypes derived from high-throughput omics technologies and detailed

✉e-mail: yongqunh@med.umich.edu; sdmooney@uw.edu
<https://doi.org/10.1038/s41581-020-00335-w>

Key points

- Ontologies are powerful tools for organizing, integrating and linking heterogeneous data types, especially in the biomedical sciences.
- Significant additions to biomedical ontologies are necessary to better define kidney molecular and histopathological phenotypes, which is critical for kidney precision medicine.
- The Kidney Precision Medicine Project is creating a community-based Kidney Tissue Atlas to integrate molecular, cellular and anatomical knowledge of the kidney.
- The development of the Kidney Tissue Atlas Ontology and the Ontology of Precision Medicine and Investigation will facilitate data collection, harmonization and analysis in support of kidney precision medicine.
- The Kidney Precision Medicine Project has extensively adopted, reused and extended community-based reference ontologies to support the annotation of kidney data.

Natural language

A language that has evolved naturally in humans and is used in speech or writing, as opposed to a constructed or formal language.

Logical expressions

A programmatic construct that expresses logical operations over mathematical terms or entities, which allows a computer to reason over the entities in the expression.

histopathological assessments be combined with traditional clinical measurements. Harmonization and integration of these data require the development of common languages or ontologies. Ontologies adopted by the biomedical sciences provide computer-readable representations of entities of interest, such as anatomical structures, cells, molecules, genes, phenotypes and diseases. These representations can be leveraged by scientists and engineers to build computational models and systems for knowledge integration and discovery. In other words, ontologies help to bridge the language barrier between humans and computers by encoding knowledge in a form that is accessible to both. By providing a controlled vocabulary, standardized definitions and defined relationships between terms, ontologies enable validation and identification of new structured terms and relationships, which can then be leveraged in the development of predictive models. Ontology

definitions in the form of both natural language and logical expressions are created and agreed on by members of the community, and represent the state of shared knowledge within a field.

The **Kidney Precision Medicine Project (KPMP)** is an NIH-funded collaboration that aims to accelerate understanding of the most common forms of kidney disease by generating molecular and 3D imaging maps of reference kidneys and of kidneys from patients with AKI and CKD. Ontologies will be of pivotal importance to the success of the KPMP by enabling integration and analysis of different data types. This Review offers an introduction to ontologies for clinicians and researchers and provides a broad overview of ontological resources in the nephrology domain, which have thus far not been used extensively by the nephrology community. We review how reference ontologies and in particular two ontologies developed by the KPMP — the Kidney Tissue Atlas Ontology (KTAO) and the Ontology of Precision Medicine and Investigation (OPMI) — will be used to annotate kidney-relevant data and support the creation of the Kidney Tissue Atlas. These data resources will then be used to revise existing definitions of kidney disease to support a precision medicine approach to treatment. As these ontologies are shared resources, we discuss how the broader community can contribute to their development and use. We encourage others to adopt these open biomedical ontologies to annotate their data, making these data more interoperable with other community resources, with the goal of increasing shared knowledge and producing rapid advancements in the diagnosis and treatment of common forms of kidney disease.

Author addresses

¹Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI, USA.

²Allen Institute for Artificial Intelligence, Seattle, WA, USA.

³Division of Nephrology, Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA.

⁴Department of Nephrology and Hypertension, Glickman Urological and Kidney Institute, Cleveland Clinic, Cleveland, OH, USA.

⁵Department of Inflammation and Immunity, Lerner Research Institute, Cleveland Clinic, Cleveland, OH, USA.

⁶Department of Pathology, Johns Hopkins University, Baltimore, MD, USA.

⁷UW Medicine Research IT, University of Washington, Seattle, WA, USA.

⁸Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

⁹Institute for Systems Biomedicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

¹⁰Division of AI/Computational Pathology, Department of Pathology, and Division of Nephrology, Department of Medicine, Duke University, Durham, NC, USA.

¹¹Division of Nephrology, School of Medicine, Washington University in St. Louis, St Louis, MO, USA.

¹²Division of Nephrology, Department of Medicine, University of Washington, Seattle, WA, USA.

¹³Division of Renal Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA.

¹⁴Section of Nephrology, Boston University Medical Center, Boston, MA, USA.

¹⁵Kidney Research Institute, University of Washington, Seattle, WA, USA.

¹⁶Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, USA.

¹⁷Department of Genetics and Genome Sciences, School of Medicine, Case Western Reserve University, Cleveland, OH, USA.

¹⁸Cleveland Institute for Computational Biology, Cleveland, OH, USA.

¹⁹Structural and Computational Biology, European Molecular Biology Laboratory, Heidelberg, Germany.

²⁰Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA, USA.

²¹Pacific Northwest National Laboratory, Richland, WA, USA.

²²Department of Genetics, Perelman School of Medicine, University of Pennsylvania Philadelphia, Philadelphia, PA, USA.

²³Department of Biomedical Informatics, Columbia University, New York, NY, USA.

²⁴Department of Biomedical Informatics, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, Buffalo, NY, USA.

²⁵Environmental Genomics and Systems Biology, Lawrence Berkeley National Laboratory, Berkeley, CA, USA.

²⁶Department of Environmental and Molecular Toxicology, Oregon State University, Corvallis, OR, USA.

²⁷The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA.

²⁸Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA, USA.

²⁹Unit for Laboratory Animal Medicine, University of Michigan Medical School, Ann Arbor, MI, USA.

³⁰Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI, USA.

³¹These authors contributed equally: Edison Ong, Lucy L. Wang.

*A full list of members and their affiliations appears in the Supplementary information.

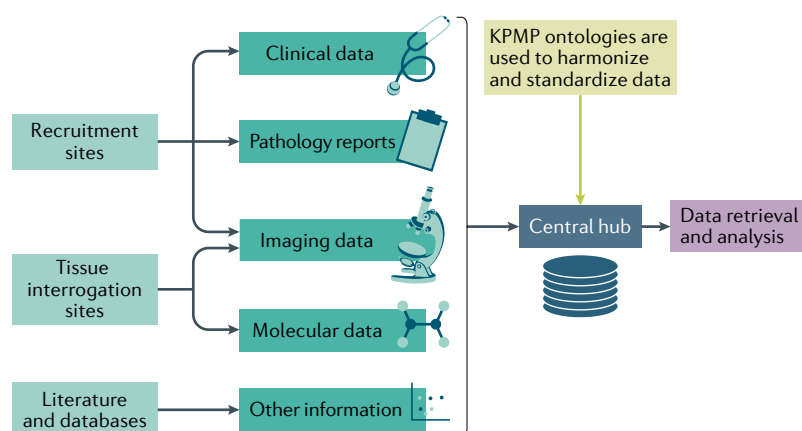


Fig. 1 | Overview of KPMP centres and the flow of KPMP data from different provenances. Clinical data and pathology reports from recruitment centres and molecular data and imaging data from tissue interrogation sites are integrated with data from the scientific literature and molecular (omics) data at the Kidney Precision Medicine Project (KPMP) central hub. The use of KPMP ontologies is pivotal to this integration. Six recruitment sites enrol participants with common forms of acute and chronic kidney disease, and collect biosamples, including a research kidney biopsy sample. Five tissue interrogation sites process and perform molecular analyses of participant biosamples. The central hub manages and aggregates data from all sites for systematic analysis and develops visualization tools and application programming interfaces to facilitate community access to data. Data harmonization and standardization with KPMP ontologies facilitate flexible data retrieval and analysis.

Development of the Kidney Tissue Atlas

One major goal of the KPMP consortium is to create a kidney tissue atlas to provide a comprehensive molecular, cellular and anatomical map of the kidney. This goal will be achieved by combining state-of-the-art molecular and cellular analyses of kidney tissue with demographic, clinical and histopathology data elements collected from participants who have generously consented to provide biopsy tissue solely for research purposes, along with detailed demographic, clinical, pathology, social history and follow-up data. The Kidney Tissue Atlas will complement concurrent atlas projects, such as the [Allen Brain Atlas](#)⁴, the [Cancer Genome Atlas](#)⁵, the [Human BioMolecular Atlas Program](#)⁶ and the [Human Cell Atlas](#)⁷. Mining of the Kidney Tissue Atlas will likely lead to novel definitions of kidney disease categories and the discovery of mechanistic drivers of these diseases. Ultimately, the Kidney Tissue Atlas is expected to provide the foundational knowledge necessary to develop new diagnostic tools and targeted therapies for the most common forms of kidney disease and injury.

The creation of such an atlas requires integration and interrogation of data — processes that are heavily reliant on ontologies (FIG. 1). Following ontological standardization, these data will be made publicly available at the [Kidney Tissue Atlas data portal](#). Annotation of data with ontology terms will facilitate user access and analysis of the data, and allow comprehensive and flexible data analysis.

Role of ontology in biomedical science

Ontology is the study of the nature of entities and their relationships in the real world⁸. With the advent of 'big data', computer scientists and informaticists have

adopted ontologies as a means to create computationally tractable models of entities and their relationships within a domain. Ontologies therefore represent a formal, structured, domain-specific, human and computer-interpretable representation of these entities and relationships⁹. They are also a foundational component of knowledge representation and reasoning — a major field of artificial intelligence that enables the modelling of scientific findings as logical expressions, which can then be interpreted by machine learning models and computer systems. Ontologies can be used for a number of purposes. For example, they can be used to represent established knowledge within a domain; to maintain standardized vocabulary within a specific field of study, across multiple locations and datasets, as well as between different consortia; to allow automated computation and decision support over structured data; and to facilitate the integration of data from distinct knowledge domains.

Ontologies share similarities with, but differ from, controlled vocabularies and taxonomies in that they not only include a controlled vocabulary and a taxonomic hierarchy but also incorporate information about other semantic relationships that provide additional information about the nature of a relationship between entities, such as the relationship between each part and its whole, or the relationship that describes spatial location. Each term (or entity) in an ontology is described by its name, synonyms, attributes and relationships to other entities.

Although most nephrology clinicians and researchers do not currently interface with ontologies, they are incorporated seamlessly into several aspects of biomedical research. For example, the Gene Ontology (GO)¹⁰ systematically classifies about 45,000 entities related to biological processes, cellular components and molecular functions of gene products for various organisms. The GO was originally developed in the late 1990s by a consortium of researchers studying the genomes of three model organisms: fruit fly, mouse and yeast. It was later used to annotate genes from other organisms, including humans, plants, animals and microorganisms. Without the GO, it would be impossible to generate consistent representations and annotations of gene products from different organisms.

In addition to annotating gene products, the GO can be used in a variety of applications, including the integration of annotated genomic data curated from the literature, the development of novel genomic analytic approaches such as gene expression functional enrichment analysis¹¹ or gene set enrichment analysis¹², and for literature mining¹³. Enrichment analyses such as gene set enrichment analysis facilitate the interpretation of otherwise uninterpretable or difficult-to-interpret big data. For example, the biological functions of hundreds or sometimes thousands of genes identified as being differentially expressed in high-throughput gene expression analyses can be summarized through enrichment analysis using GO terms¹¹. Gene-level annotations defined by the GO can be further elaborated into a network of biological pathway annotations using GO Causal Activity Modelling to determine¹⁴ the integrative effects of differentially expressed genes on biological pathways. The GO demonstrates the value of ontologies in establishing

Controlled vocabularies

A way to organize knowledge for retrieval; comprises a set of selected terms used for document indexing and information retrieval.

Taxonomies

Controlled vocabularies that have a hierarchical structure indicating subclass relationships between entities.

consistent annotation schemes for a class of biomedical entities. These annotations are interoperable and can be used to derive benefit in downstream analyses. The successes of the GO have spurred the development of many hundreds of ontologies¹⁵ in other domains, such as anatomy^{16,17}, proteins¹⁸ and disease^{19,20}.

Open biomedical ontologies

The proliferation of biomedical ontologies has led to frequent issues of redundancy and poor interoperability²¹. Although ontology-matching algorithms²² have been developed to map terms between different ontologies, match quality is imperfect and is insufficient for addressing underlying issues of redundancy. The [Open Biological and Biomedical Ontology \(OBO\) Foundry](#)²³ was established to achieve better ontology interoperability and resolve problems arising as a consequence of overlapping representations across different biomedical ontologies. OBO Foundry ontologies are created and formatted following a set of shared principles, designed to ensure that OBO Foundry ontologies remain open, orthogonal, interoperable and logically well formed with a well-specified syntax. Only ontologies that have been developed and maintained following these principles are accepted into the consortium.

OBO currently includes more than 170 biomedical ontologies in domains such as phenotype²⁴, disease^{19,20}, anatomy¹⁶, genetics¹⁰ and proteomics¹⁸, and these ontologies have been used to successfully address a number of research questions in the biomedical sciences²⁵. For example, the Human Phenotype Ontology (HPO) supports a deep phenotyping approach to defining human diseases²⁴. One example of the application of the HPO is provided by the description of two patients with phenotypic profiles that only partially matched the standard diagnostic profile of Wiedemann–Steiner syndrome²⁵. Despite the different presentations of these two patients, they could both be matched to the syndrome profile through HPO-based inference. This ability of the HPO to assess relationships through use of fuzzy matching and causal reasoning has potential to improve diagnostic insights.

Ontologies are often used in big biomedical projects. For example, the [Library of Integrated Network-Based Cellular Signatures \(LINCS\)](#) programme aims to create a network-based understanding of biological processes by profiling changes in gene expression and various cellular processes that are induced by exposing human cells to chemical, genetic and disease perturbations²⁶. To enable systematic study of the perturbed cell responses, the LINCS programme relies heavily on ontologies to support standard representation and analysis^{27,28}. Another example is the [Encyclopedia of DNA Elements \(ENCODE\) project](#)²⁹, which is an ongoing collaborative effort that aims to identify and annotate all of the functional elements in the human genome. Ontologies have played a major part in ENCODE by facilitating the organization and standardization of experimental data, metadata and associated computational analyses. In addition, the ENCODE portal uses OBO Foundry ontologies, such as the HPO and the Ontology for Biomedical Investigations (OBI), to support ontology-driven search and data integration^{29,30}.

Interoperable, reliable and community-driven OBO Foundry ontologies are also critical to support the seamless assembly and integration of kidney data from heterogeneous sources and domains. Given its status as a sizable multicentre project, the KPMP faces challenges in the coordination of data across multiple sites and groups of personnel. The creation of a shared vocabulary for annotating patient information and tissue specimens collected at recruitment, and for summarizing molecular features and analytic results, is therefore vital to ensure the quality, interpretability and reusability of KPMP data. To achieve these goals, the KPMP focuses on adapting and extending OBO Foundry ontologies, and creating new KPMP ontologies only to address application needs. By linking KPMP ontologies to other ontology resources (through observing OBO Foundry principles and reusing terms in existing ontologies), the work of the KPMP will benefit not only consortium members but also the broader biomedical community.

Ontologies for modelling kidney disease

A number of ontological resources relevant to nephrology have already been developed. Here we describe the function of these existing ontologies and discuss how the development of new ontologies aims to fill remaining gaps.

Existing kidney ontologies. Over the past two decades, a number of ontologies and classification systems have been developed to support kidney research. These include the Genitourinary Development Molecular Anatomy Project (GUDMAP) ontology³¹, the Chronic Kidney Disease Ontology (CKDO)³² and the classification system introduced by the Renal Pathology Society (RPS)^{33–35}.

The GUDMAP consortium was formed in 2004 with the goal of creating a molecular anatomical atlas of the developing mouse kidney and urogenital tract³⁶. One component of this project was the creation of an ontology of genitourinary developmental cell types anchored to mouse anatomy³¹. The initial ontology was released in 2007 (REF.³⁷) and was created as an expansion of the ontology developed for the Edinburgh Mouse Atlas Project³⁸. The GUDMAP ontology was primarily developed to facilitate the annotation of mouse cell types but has evolved to include data from human fetal kidney and urinary tract. Molecular cell types described using KPMP ontologies can be mapped to GUDMAP terms to enable the bridging of data collected across the lifespans of human and mouse specimens. This mapping is part of future work to be done in collaboration with the curators of the GUDMAP ontology.

The CKDO is a clinically oriented ontology designed to assist in the characterization and staging of CKD³². The ontology primarily describes clinical features associated with CKD, which enables CKD to be defined on the basis of, for example, clinical diagnostic codes, or abnormal laboratory findings such as changes in estimated glomerular filtration rate and proteinuria. The CKDO is useful for identifying and classifying patients in a clinical setting using defined stages of CKD. However, it lacks the ability to connect clinical descriptions to molecular

Fuzzy matching

A technique that identifies the correspondence among phenotypic profiles that may be less than 100% perfect.

Causal reasoning

The process used to identify the causality (cause and effect) between two entities.

phenotypes or anatomy and therefore does not enable a precision approach to patient treatment.

Ontologies are not typically used by renal pathologists in clinical practice; however, the RPS has undertaken several initiatives to standardize language and reporting, and to organize, categorize and stage findings from kidney biopsy samples^{33–35}. Although these initiatives are not formal ontologies, they provide a helpful road map for the use of ontologized pathological features to drive novel classifications while enabling comparisons with existing disease definitions. For example, an international collaborative project involving an RPS working group and the KPMP pathology working group aims to harmonize language, definitions and metrics (when relevant) for histological and ultrastructural parameters across all currently used classification and scoring systems³⁹. This project involves expansion and improvement of RPS terminology and definitions to provide a framework for anthologizing histological and ultrastructural features of the kidney.

These three initiatives have each developed standardized terms in highly specific subareas of kidney physiology and disease modelling. However, none of them currently provides a framework with which to integrate the data types needed for precision medicine diagnostics and treatment, representing a gap in existing ontological resources.

Reference ontologies used by the KPMP. In addition to the aforementioned ontologies and classification systems, a number of reference ontologies relevant to kidney anatomy, function and disease are also available (TABLE 1). Most of these reference ontologies are part of the OBO Foundry and are designed to be reused by multiple groups and stakeholders. Each of these reference ontologies focuses on a group of entities relevant to a particular subdomain. For example, human phenotypes are ontologized in the HPO²⁴, Uberon (uber-anatomy ontology)¹⁶ focuses on anatomical structure, the Cell Ontology describes cell types⁴⁰ and biological processes

Table 1 | **Reference ontologies used by the KPMP for kidney modelling**

Domain	Ontology	Purpose and application to the KPMP	Number of entities	Number of relationships ^a	Ref.
Phenotype	Human Phenotype Ontology ^b	Describes clinical and pathological phenotypes	26,578	61,665	²⁴
Disease	Mondo Disease Ontology ^b	Describes relationships between patient clinical characteristics and disease terminology	111,478	136,833	¹⁹
Anatomy	Uberon (uber-anatomy ontology) ^b	Describes aspects of kidney anatomy	15,183	43,082	¹⁶
Cells and cell types	Cell Ontology ^b	Describes cell types and cellular components relevant to modelling	10,630	35,916	⁴⁰
Proteins	Protein Ontology ^b	Describes protein-related entities and the relationships between these entities	317,974	919,192	¹⁸
Biological processes	Gene Ontology ^b	Describes the association of molecular features and biological processes	50,255	106,149	¹⁰
Subcellular processes	Molecular Biology of the Cell Ontology	Describes subcellular processes (pathways) and their interactions leading to cell-level functions	6,136	19,932	⁴¹
Laboratory measurements	Ontology for Biomedical Investigations ^b	Describes laboratory values related to patient diagnostics	3,584	7,228	⁵³
Clinical measurements	Clinical Measurement Ontology ^b	Describes clinical measurements related to patient diagnostics	3,054	3,718	⁵⁴
Adverse events	Ontology of Adverse Events ^b	Describes clinical features and co-morbidities associated with patients	5,700	11,572	⁵⁵
Chemical compounds	Chemical Entities of Biological Interest ^b	Describes metabolites and other chemical entities	137,894	266,753	⁵⁶
Drugs	Drug Ontology ^b	Describes patient medications	554,934	1,112,074	⁵⁷
Kidney tissue atlas	Kidney Tissue Atlas Ontology ^{b,c}	Facilitates data collection, integration and analysis for comprehensive kidney precision medicine studies	5,593	10,838	NA
Precision medicine ontology	Ontology of Precision Medicine Investigation ^{b,c}	Describes general precision medicine projects	2,896	4,413	NA

Entity and relation counts are reported as of 8 May 2020. KPMP, Kidney Precision Medicine Project; NA, not applicable. ^aNumber of relationships or links among the entities within an ontology. ^bOpen Biological and Biomedical Ontology Foundry ontologies.

^cKPMP-initiated ontologies.

are represented by the GO¹⁰ (which connects gene expression to cellular and tissue processes) and the Molecular Biology of the Cell Ontology (MBCO; which describes interactions between gene expression and subcellular processes)⁴¹. The Mondo Disease Ontology (MONDO)¹⁹ aims to harmonize definitions of disease and can be used to integrate the content of clinical controlled vocabularies such as those used by the *Systematized Nomenclature of Medicine (SNOMED)* and the *International Classification of Diseases (ICD)* system. Data that are annotated with reference ontology terms can be easily integrated into the ecosystem of other datasets that are annotated with terms from the same ontologies.

Each of these ontologies contains terms that are relevant to the nephrology community. For example, Uberon contains references to kidney anatomy and the HPO has terms representing abnormalities in urine microscopy and electrolyte abnormalities. Although these reference ontologies are extensive, the definitions and terms relevant to the nephrology field have not necessarily been reviewed by nephrologists or researchers in the nephrology community. To improve the value of these ontologies for the nephrology community, the KPMP has identified teams of subject matter experts who have reviewed these terms and carefully curated their definitions. In circumstances where the terms are determined to be inaccurate or incomplete, the KPMP collaborates with curators of the existing ontology to either modify or add terms as appropriate as discussed next.

Gaps in existing ontologies. Despite the abundance of ontology resources that are available for reuse, some necessary entities are not sufficiently represented. The novel nature and depth of the data collected by the KPMP and analyses required to develop the Kidney Tissue Atlas will require the introduction of new ontology terms to accurately describe and model the relationships between them — these terms must be either defined in a KPMP-specific ontology or added to an existing reference ontology. For instance, kidney-specific terms are sometimes inaccurately represented in existing reference ontologies, synonyms may be missing or taxonomic classification may need to be reorganized. As an example, existing reference ontologies lack the detailed catalogue of descriptive cell types and pathology terms that are needed by the KPMP. Kidney disease phenotypes described in the HPO are currently incomplete and lack sufficient details to annotate the full breadth of kidney disease. For example, the HPO defines focal segmental glomerulosclerosis⁴² but does not include an entry for global glomerulosclerosis. Similarly, while the Cell Ontology⁴⁰ classifies some kidney-specific cell types such as a glomerular visceral epithelial cell⁴³, it lacks the granularity to describe kidney cell phenotypes based on gene markers and molecular expression. Moreover, there are also gaps in representation of clinical data in classification systems such as the SNOMED and the ICD; these systems do not contain sufficient terminology and relationships to connect clinical terms to molecular phenotypes.

The KPMP aims to address these gaps in the molecular, pathological and clinical annotation of kidney

cells, structures and function by creating new ontological resources with which to annotate the kidney pathological and molecular features that are currently not described or are underdescribed by existing ontologies. The project will also collaborate with curators of existing ontologies to improve ontology representation for the nephrology community. Suggested changes to reference ontologies are documented and shared with the curators of each reference ontology for review and incorporation into that ontology. Similarly, if terms are missing, they are created through collaboration with the curators of the reference ontologies to develop a definition, synonyms and hierarchical classification. We anticipate that this collaborative approach will be an ongoing process as new technologies are developed and novel data become available.

In addition, entities from different ontologies are not always semantically linked, and one task of KPMP ontology development is to provide links between existing terms where appropriate. For example, a gene marker in a specific kidney cell type may not be semantically linked to its related phenotypes in another ontology. When the KPMP discovers such missing or novel associations — for example, a novel gene variant that is associated with CKD progression — the relationship is added to the KPMP ontologies. The KTAO (described in greater detail later) provides an integrative ontology framework with which to import and link these terms.

KPMP ontologies

To bridge the gaps in existing ontologies for annotating kidney-specific data, the KPMP has developed two KPMP-initiated ontologies — the KTAO and the OPMI. The KTAO is an application ontology designed to describe and integrate data relating to kidney anatomy, phenotypes, diseases, molecular features and other kidney-related concepts collected by the KPMP. Application ontologies are usually derived from reference ontologies, with the addition of highly specific terms and relationships that are applicable to a single project or end use. The purpose of the KTAO is to support the granularity needed for KPMP studies and support the needs of participating institutions within the KPMP consortium. By contrast, the OPMI is a reference ontology of concepts used to describe data for precision medicine, and is designed to support data harmonization and integration for precision medicine projects beyond the KPMP.

These two new ontologies support the creation of the Kidney Tissue Atlas (FIG. 2), and are used to annotate and standardize KPMP data at various stages of data management, including collection, analysis and long-term storage and retrieval. For example, KPMP ontologies (KTAO and OPMI) are used to standardize case report forms and the data elements collected with these forms, and to unify these clinical data with molecular data, such as kidney disease biomarkers and cell types, and anatomical entities. These ontologies are integrated with OBO Foundry ontologies and shared with the community to promote broad adoption and reuse of standardized structured knowledge. The integrated data in the KPMP Kidney Tissue Atlas can then be queried to answer questions

Case report forms

A document (paper or electronic) containing a questionnaire used for clinical research or other purposes.

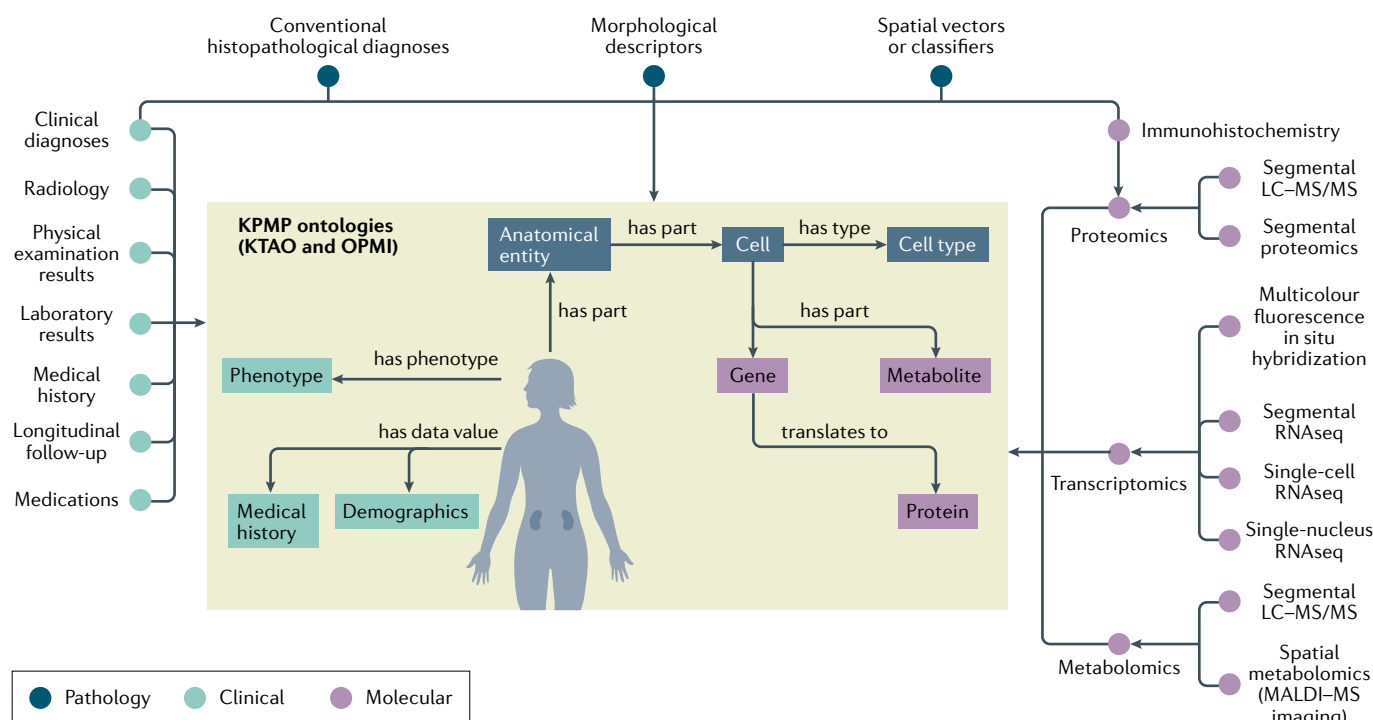


Fig. 2 | The KPMP ontology framework for supporting data representation, integration and analysis. Clinical, pathology and molecular data collected from Kidney Precision Medicine Project (KPMP) recruitment sites and tissue interrogation sites will be deposited in the KPMP Kidney Tissue Atlas. Different types of data (clinical, pathology and molecular) feed into the KPMP ontology environment. Two KPMP ontologies, the Kidney Tissue Atlas Ontology (KTAO) and the Ontology of Precision Medicine Investigation (OPMI), provide a semantic framework for modelling relationships between the heterogeneous data in the atlas. LC–MS/MS, liquid chromatography–tandem mass spectrometry; MALDI–MS, matrix-assisted laser desorption/ionization–mass spectrometry; RNAseq, RNA sequencing.

about kidney disease. For example, a researcher may want to determine the unique genes expressed in the proximal tubule of the kidney of patients with diabetic kidney disease in an effort to identify novel gene markers or targets for treatment. This question can be answered only by combining clinical features with pathological images and findings from transcriptomic, proteomic and metabolomic studies. Data from these studies must be annotated using a shared ontological framework so that they can be combined and analysed. It is anticipated that the shared Kidney Tissue Atlas data platform, supported by the KPMP ontologies, will facilitate future nephrology research by the wider community.

Kidney Tissue Atlas Ontology. As mentioned already, the KTAO is designed to logically represent the relationships between gene markers, phenotypes, diseases, cell types and anatomical entities to support the modelling of common forms of kidney disease⁴⁴. The KTAO was developed using both a top-down approach and a bottom-up approach. The top-down approach is led by ontologists and allows them to define the basic structure of the ontology and populate it with initial terms and relationships. The bottom-up approach allows the incorporation of term recommendations and editing suggestions from the end users of the ontology. To avoid repeating work done by others, ontologists involved in developing the KTAO reused appropriate terms from existing OBO Foundry ontologies, including the GO¹⁰, HPO²⁴,

MONDO¹⁹, OBI⁴⁵, UBERON¹⁶, Cell Ontology⁴⁰ and OPMI and other reference ontologies such as MBOCO⁴¹. The KTAO is strongly linked to the open biomedical ontology ecosystem, and follows the OBO Foundry principles of reuse and repurposing.

As the KPMP collaborators assess reference and diseased kidney biopsy tissue, new knowledge will be added and linked within the KTAO to create a set of well-defined kidney disease-related entities or phenomena. This ontology will enable integration of distinct data types and support user-defined searches or clustering of participants and/or samples based on a panel of clinically relevant features. Developing capability for user-defined searches and user-directed clustering is an important component of the KPMP mission and is anticipated to be an important driver of new knowledge discovery. It is expected that new entities and relationships will be identified and added to the KTAO, and existing entities and relationships will also be modified through the course of the study. Examples include the molecular definition of kidney cell types or cell states, the refinement of existing anatomical entities and/or the creation of new kidney disease classifications based on new understanding of disease pathways and mechanisms. New entities and relationships that are defined during the course of KPMP studies will initially be added to the KTAO and, when suitable, will be submitted to the corresponding reference ontologies to benefit the broader scientific community.

Ontology of Precision Medicine Investigation. The KPMP faces challenges of big data standardization and integration, which requires the synthesis of high-throughput multiscale (clinical, pathology and molecular) data into knowledge. The OPMI has been developed as a community-based open source biomedical ontology to address this challenge. The formal representation and integration of findings from basic research can be affected by various factors, including technical factors, such as those arising from the instruments used to generate data or the methods used to collect biosamples, as well as clinical and pathological factors that are unique to individual participants. Data for precision medicine purposes must be accurately captured and modelled to facilitate robust analysis, and the OPMI has been developed to achieve these aims. For example, as data are collected by the KPMP, the descriptors or measurements of the clinical data and the relationships among them as determined by the OPMI can be used to validate values during data entry, and errors can be flagged before values are stored, thereby improving the quality and reliability of the data collected by the KPMP.

The OPMI was developed following OBO Foundry principles, including openness and collaboration, and as such has been accepted as an OBO Foundry ontology. The ontology is designed as a data integration platform for general precision medicine projects, including the KPMP. The OPMI reuses many terms and relationships from existing ontologies, including the [Ontology of General Medical Science](#), OBI, HPO, Uberon, the Ontology of Adverse Events⁴⁶ and the Informed Consent Ontology⁴⁷. In addition, the OPMI represents many precision medicine-specific terms that can be imported to the KTAO and other clinical ontologies. It has been used to standardize the major metadata types and clinical factors derived from the 30 or so case report forms developed by the KPMP that together include more than 2,500 clinical questions. The standardization of data elements from these case report forms markedly improves ontology-based data integration across different institutes⁴⁸. In addition to supporting the KPMP, the OPMI has also been used by other biomedical projects. For example, the OPMI has been used as an ontology platform to model the metadata shown in the ClinicalTrials.gov database and other clinical trial repositories⁴⁹.

Applications of KPMP ontologies

As mentioned already, the aim of these new ontologies is to support kidney disease research. As illustrated by the examples below, ontologies have the potential to enhance our understanding of kidney disease by enabling deep phenotyping of kidney disease tissue, allowing us to identify new classifications and subclassifications of common kidney diseases and previously unrecognized relationships between clinical, anatomical, pathological and molecular phenotypes.

For instance, the current clinical approach to diagnose kidney disease is based on patient demographics, medical history of past and present illnesses, physical examination and laboratory tests. One of the first goals of clinical evaluation is to establish a cause of kidney disease. While nephrologists often use their clinical judgement to infer

the cause of kidney disease from the patient's medical history, laboratory values and other clinical features, a kidney biopsy sample is sometimes necessary to establish the underlying cause of kidney disease. Biopsy samples are routinely evaluated with standard histopathological approaches, including light microscopy with specialized staining, immunofluorescence microscopy and electron microscopy. The incorporation of molecular features captured by high-throughput evaluation of kidney biopsy samples is not currently the standard of care. Combining these molecular features with the standard clinical, laboratory and pathology data through the use of ontologies may reveal previously unrecognized subtypes of kidney diseases. Transcriptomic, proteomic and metabolomic data can also be integrated to redefine the classification or categorization of kidney disease and identify driver cell types and potential therapeutic targets⁵⁰.

Cell type-specific gene, protein and metabolite expression profiles translate into cell type-specific functions that regulate the functions of tissues, organs and, finally, whole organisms. Although cell ontologies allow the characterization of pathways that underlie cellular physiology from molecular profiles, the integration of cell physiology with kidney physiology and pathophysiology, as well as whole body function, requires an integrated ontology that spans multiple levels. For example, the *COL4A3* gene has an integral role in the organization of the glomerular basement membrane, which is critical for the proper filtration barrier function of the kidney. Mutations in *COL4A3* lead to disorganization of the glomerular basement membrane and Alport syndrome but can also phenocopy focal segmental glomerulosclerosis⁵¹. However, other coding sequence variations in *COL4A3* have a renoprotective role in the setting of diabetes⁵². Ontologies have the potential to link genes in specific cell types (obtained from single-nucleus or single-cell transcriptomic data) to cellular pathways and to link cell physiological function with whole body physiology, and thereby have potential to identify further examples of connectivity between changes in pathway activity and cellular dysfunction caused by disease.

Ontologies are already used to support clinical and translational examination of kidney diseases. For example, a variety of data types are collected by clinicians in standard clinical practice, including demographic data, clinical history, physical examination findings and diagnostic test results (FIG. 3a). The clinician uses this information to arrive at a diagnosis and treatment plan for the patient. For example, a clinician may evaluate a 63-year-old man with a 40-year history of poorly controlled type 2 diabetes mellitus, a slowly increasing serum creatinine level over the course of several years and proteinuria (2 g of protein per day), with normal findings following a comprehensive serological evaluation for non-diabetic kidney diseases. On the basis of these observations, the clinician may determine that the patient most likely has diabetic kidney disease and most clinicians would opt not for biopsy.

However, the availability of a biopsy sample would enable the pathology and molecular data to be collected, which may facilitate deeper understanding of the disease. The KPMP ontological framework can capture the

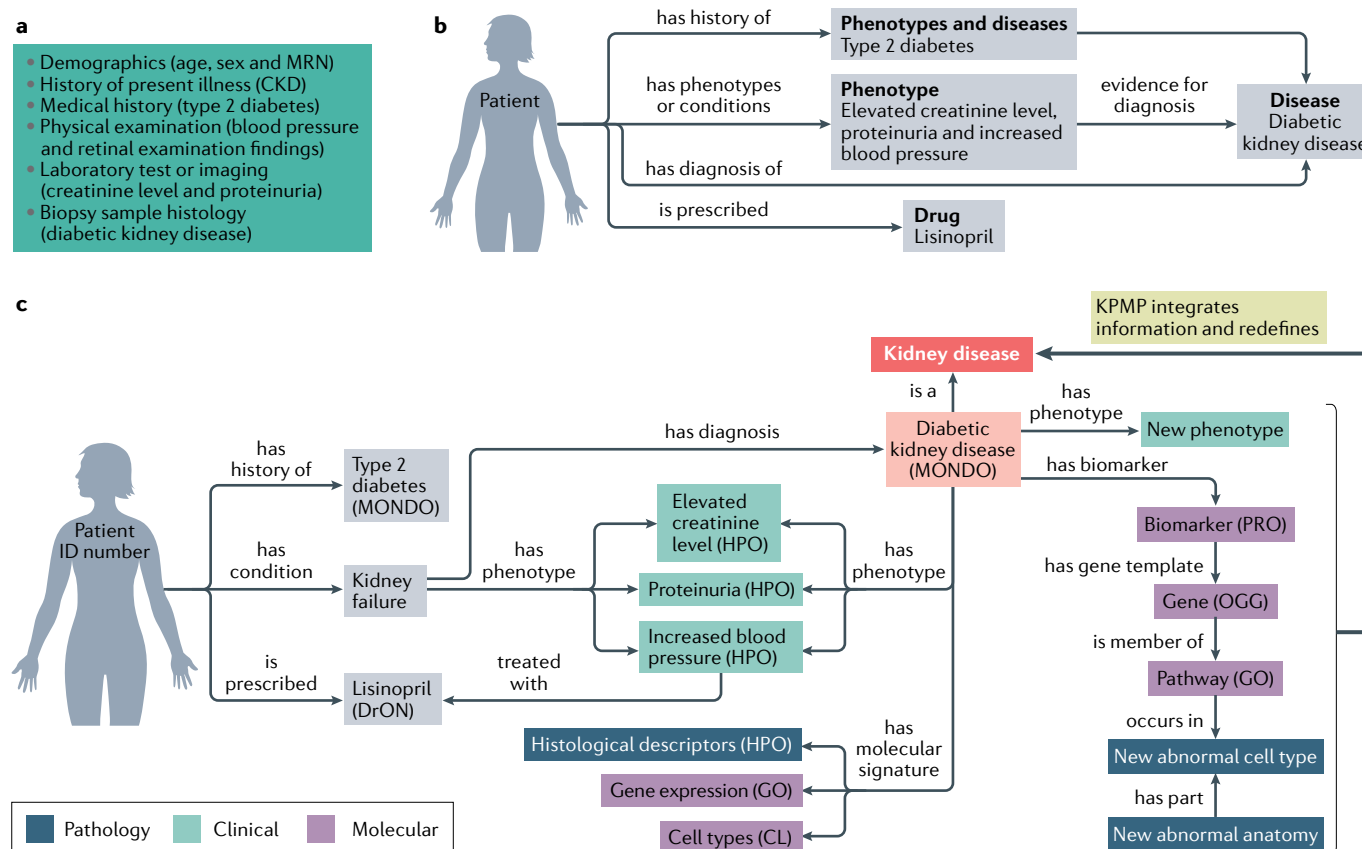


Fig. 3 | Using the Kidney Tissue Atlas Ontology to support molecular and histopathological extensions to kidney disease diagnosis.

a | Current clinical practice involves assessment of limited data types, as illustrated by the example data fields for a hypothetical patient with diabetic kidney disease. **b** | These same data elements can be used to model or support disease diagnosis and treatment using current ontologies. **c** | An integrative ontology-based approach can incorporate molecular and pathology data in addition to clinical measures, demonstrating the data harmonization goals of the Kidney Precision Medicine Project (KPMP). The KPMP Kidney Tissue Atlas, supported by ontologies such as the Kidney Tissue Atlas Ontology and the Ontology of

Precision Medicine and Investigation, will enable kidney diseases to be redefined on the basis of their molecular mechanisms and knowledge of clinical and histopathological features. This approach will aid the identification of critical cells, pathways and targets for the development of novel diagnostic methods and therapies. The different clinical and molecular phenotypes of a patient can also be modelled using existing ontologies, such as the Human Phenotype Ontology (HPO), the Cell Ontology (CL), the *Ontology of Genes and Genomes* (OGG), the Drug Ontology (DrON) and the Protein Ontology (PRO). CKD, chronic kidney disease; GO, Gene Ontology; MONDO, Mondo Disease Ontology; MRN, medical record number.

aforementioned clinical data and link them with molecular and imaging data from kidney biopsy samples, as well as other sources of knowledge, to enable a more nuanced assessment of the individual's disease presentation in the context of other reference and disease tissues (FIG. 3b). The ontology framework can be easily adapted to enable computational phenotyping of patients and the development of decision support systems to assist clinicians in diagnosis and treatment.

This approach is in contrast to the current clinical model, which does not integrate molecular and pathology data — two key components of precision medicine. A central goal of the KPMP is to develop an integrative framework using the KTAO to standardize and harmonize data obtained in standard clinical practice with novel molecular and histopathology data that will be generated through biopsy sample analysis. For example, diabetic kidney disease in a particular patient might be associated with specific biomarkers, encoded by genes and linked to particular biological pathways and functions (FIG. 3c). Single-cell and single-nucleus sequencing

technologies may enable identification of kidney cell types that drive disease and lead to the identification of new kidney disease subtypes based on molecular and cellular phenotyping. The hierarchical structure and semantic relationships provided by KPMP ontologies can be used to link diverse data types and make such discoveries possible. Integrated representation of clinical and molecular features will enable the redefinition of our understanding of kidney disease, provide clinicians with novel diagnostic and treatment options for their patients, and facilitate novel discoveries in the field of nephrology research.

Conclusions and future directions

Successful applications of precision medicine require that large numbers of phenotypic traits, including molecular, genomic, clinical and other traits, be documented for each individual. The relationships between various traits and treatment outcomes provide a framework with which to predict the needs of each individual and select the best treatment plan for an individual patient.

To develop these frameworks, patterns must be identified through the collection of data from a large, diverse group of individuals and these data must be standardized to allow for proper comparison. Ontologies provide both a means for the terminological standardization of data and support for the incorporation of structured terms and relationships into predictive models for clinical deployment.

By creating a resource such as the Kidney Tissue Atlas, the KPMP aims to create a repository of clinical and biospecimen data that can be used to support kidney precision medicine. A key goal of the KPMP is to use these data, particularly molecular data, to define novel subtypes of current (and currently insufficient) disease classifications. With these new disease subtypes, clinicians and researchers can discover more targeted and effective therapies. The molecular phenotypes needed for these novel disease classifications are especially challenging to describe, as the molecular features used to define these phenotypes are continuous, whereas traditional phenotypes are discrete. How best to define novel molecular and cellular phenotypes is an open question that the KPMP hopes to answer as more data and insight into this issue are acquired. This challenge mirrors the global challenge faced by precision medicine: the disconnect between the recognition of each individual as a unique case deserving specialized treatment, and the need to classify individuals into groups in order to assess the statistical efficacy of those treatments.

Ontologies have a critical infrastructural role in the aforementioned tasks. They provide a mechanism for harmonizing and integrating data collected from disparate centres and organizations across different categories and domains. When large combined datasets are analysed, computational methods are necessary to discover correlations and relationships between input features. Manual harmonization of large datasets is impractical and expensive, and thus built-in annotation of shared ontology terms is critical and makes these sorts of analyses feasible.

The novel data generated by the KPMP will require additions to be made to existing open biomedical ontologies to support data annotation in the future. Members of the KPMP are therefore working with other ontology groups and developers to incorporate kidney-specific

terminology and relationships into reference ontologies such as the HPO. A standard operating procedure has been established to support the collaboration between the HPO (and other ontologies) and KTAO development teams. As the KPMP builds up tissue collection and analysis, ontology terms will be used to annotate patient data, specimens and analysis results. Development of KPMP ontologies and suggested additions and changes to references ontologies are ongoing as annotation needs are continuously re-evaluated.

In addition to assisting in data annotation and analysis, the KTAO framework will become a living representation of our understanding and knowledge of kidney diseases. As described, KPMP studies are expected to generate new kidney disease subtypes, biomarkers and disease-specific pathways, which will be integrated into the KTAO and other reference ontologies. These updated ontologies can then be used to further improve kidney-specific data annotation and analysis. We also expect that the KPMP ontology framework could be used to support the development of new tools. For example, the Kidney Tissue Atlas visualization tool can use the KTAO entity hierarchy to provide better browsing and querying of tissue samples. Molecular data and pathways annotated using KTAO terms can also be used for advanced biomarker and pathway analysis.

Thus, ontologies are essential for kidney precision medicine and provide practical benefits for data organization and knowledge discovery. The strength of a shared KPMP data resource and ontologies depends on the contributions and efforts of the surrounding research community. Ontological improvements made by the KPMP aim to help enable standardized data sharing for the nephrology clinical research community. By making data more interoperable through annotation with the same shared ontologies, the pool of data that can be harnessed for research grows substantially. It is our hope that members of the nephrology community will support this shared ecosystem and use these ontologies as a fundamental organizational layer in data analyses. Only through consistent investment in data interoperability can greater gains be derived from resources that are so laboriously built and shared.

Published online: 16 September 2020

- Abrahams, E. Right drug-right patient-right time: personalized medicine coalition. *Clin. Transl. Sci.* **1**, 11–12 (2008).
- Khwaja, A. KDIGO clinical practice guidelines for acute kidney injury. *Nephron Clin. Pract.* **120**, c179–c184 (2012).
- Stevens, P. E. & Levin, A. Kidney Disease: Improving Global Outcomes Chronic Kidney Disease Guideline Development Work Group Members. Evaluation and management of chronic kidney disease: synopsis of the kidney disease: improving global outcomes 2012 clinical practice guideline. *Ann. Intern. Med.* **158**, 825–830 (2013).
- Hawrylycz, M. et al. in *Springer Handbook of Bio-/Neuroinformatics*. (ed. Kasabov, N.) 1111–1126 (Springer, 2014).
- Cancer Genome Atlas Research Network. et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).
- HuBMAP Consortium. The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. *Nature* **574**, 187–192 (2019).
- Regev, A. et al. The Human Cell Atlas. *Elife* **6**, e27041 (2017).
- Cimiano, P. in *Ontology Learning and Population from Text: Algorithms, Evaluation and Applications* (ed. Cimiano, P.) 9–17 (Springer, 2006).
- Gruber, T. R. Toward principles for the design of ontologies used for knowledge sharing? *Int. J. Human-Computer Stud.* **43**, 907–928 (1995).
- Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
- Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).
- Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
- Doms, A. & Schroeder, M. GoPubMed: exploring PubMed with the Gene Ontology. *Nucleic Acids Res.* **33**, W783–W786 (2005).
- Thomas, P. D. et al. Gene Ontology Causal Activity Modeling (GO-CAM) moves beyond GO annotations to structured descriptions of biological functions and systems. *Nat. Genet.* **51**, 1429–1433 (2019).
- Whetzel, P. L. et al. BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. *Nucleic Acids Res.* **39**, W541–W545 (2011).
- Mungall, C. J., Tornai, C., Gkoutos, G. V., Lewis, S. E. & Haendel, M. A. Uberon, an integrative multi-species anatomy ontology. *Genome Biol.* **13**, R5 (2012).
- Rosse, C. & Mejino, J. L. V. Jr. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *J. Biomed. Inform.* **36**, 478–500 (2003).
- Natale, D. A. et al. Protein Ontology (PRO): enhancing and scaling up the representation of protein entities. *Nucleic Acids Res.* **45**, D339–D346 (2017).
- Kibbe, W. A. et al. Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. *Nucleic Acids Res.* **43**, D1071–D1078 (2015).

20. Shefchek, K. A. et al. The Monarch Initiative in 2019: an integrative data and analytic platform connecting phenotypes to genotypes across species. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkz997> (2019).
21. Kamdar, M. R., Tudorache, T. & Musen, M. A. A systematic analysis of term reuse and term overlap across biomedical ontologies. *Semant. Web* **8**, 853–871 (2017).
22. Euzenat, J. & Shvaiko, P. *Ontology Matching*. <https://doi.org/10.1007/978-3-642-38721-0> (Springer, 2013).
23. Smith, B. et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat. Biotechnol.* **25**, 1251–1255 (2007).
24. Robinson, P. N. & Mundlos, S. The Human Phenotype Ontology. *Clin. Genet.* **77**, 525–534 (2010).
25. Haendel, M. A., Chute, C. G. & Robinson, P. N. Classification, ontology, and precision medicine. *N. Engl. J. Med.* **379**, 1452–1462 (2018).
26. Keenan, A. B. et al. The Library of Integrated Network-Based Cellular Signatures NIH program: system-level cataloging of human cells response to perturbations. *Cell Syst.* **6**, 13–24 (2018).
27. Vempati, U. D. et al. Metadata standard and data exchange specifications to describe, model, and integrate complex and diverse high-throughput screening data from the Library of Integrated Network-based Cellular Signatures (LINCS). *J. Biomol. Screen.* **19**, 803–816 (2014).
28. Ong, E. et al. Ontological representation, integration, and analysis of LINCS cell line cells and their cellular responses. *BMC Bioinformatics* **18**, 556 (2017).
29. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
30. Malladi, V. S. et al. Ontology application and use at the ENCODE DCC. *Database* 2015 (2015).
31. McMahon, A. P. et al. GUDMAP: the genitourinary developmental molecular anatomy project. *J. Am. Soc. Nephrol.* **19**, 667–671 (2008).
32. Cole, N. I. et al. An ontological approach to identifying cases of chronic kidney disease from routine primary care data: a cross-sectional study. *BMC Nephrol.* **19**, 85 (2018).
33. Bajema, I. M. et al. Revision of the International Society of Nephrology/Renal Pathology Society classification for lupus nephritis: clarification of definitions, and modified National Institutes of Health activity and chronicity indices. *Kidney Int.* **93**, 789–796 (2018).
34. Leung, N. et al. The evaluation of monoclonal gammopathy of renal significance: a consensus report of the International Kidney and Monoclonal Gammopathy Research Group. *Nat. Rev. Nephrol.* **15**, 45–59 (2019).
35. Sethi, S. et al. Mayo Clinic/Renal Pathology Society consensus report on pathologic classification, diagnosis, and reporting of GN. *J. Am. Soc. Nephrol.* **27**, 1278–1287 (2016).
36. Harding, S. D. et al. The GUDMAP database—an online resource for genitourinary research. *Development* **138**, 2845–2853 (2011).
37. Little, M. H. et al. A high-resolution anatomical ontology of the developing murine genitourinary tract. *Gene Expr. Patterns* **7**, 680–699 (2007).
38. Hayamizu, T. F. et al. EMAP/EMAPA ontology of mouse developmental anatomy: 2013 update. *J. Biomed. Semant.* **4**, 15 (2013).
39. Haas, M. et al. Consensus definitions for glomerular lesions by light and electron microscopy: recommendations from a working group of the Renal Pathology Society. *Kidney Int.* <https://doi.org/10.1016/j.kint.2020.08.006> (2020).
40. Diehl, A. D. et al. The Cell Ontology 2016: enhanced content, modularization, and ontology interoperability. *J. Biomed. Semant.* **7**, 44 (2016).
41. Hansen, J., Meretzky, D., Woldesenbet, S., Stolovitzky, G. & Iyengar, R. A flexible ontology for inference of emergent whole cell function from relationships between subcellular processes. *Sci. Rep.* **7**, 17689 (2017).
42. No authors listed. Focal segmental glomerulosclerosis. *Human Phenotype Biology* http://purl.obolibrary.org/obo/HP_0000097 (2020).
43. No authors listed. Cell Ontology. glomerular visceral epithelial cell. *Ontobee*. http://purl.obolibrary.org/obo/CL_0000653 (2020).
44. He, Y. et al. KTAO: a Kidney Tissue Atlas Ontology to support community-based kidney knowledge base development and data integration. in *Proc. 9th International Conference on Biological Ontology* (RWTH, 2018).
45. Peters, B., & The OBI Consortium. Ontology for biomedical investigations. *Nat. Preced.* <https://doi.org/10.1038/npre.2009.3623> (2009).
46. Kang, Y., Fink, J. C., Doerfler, R. & Zhou, L. Disease specific ontology of adverse events: ontology extension and adaptation for chronic kidney disease. *Comput. Biol. Med.* **101**, 210–217 (2018).
47. Y. Lin, et al. Development of a BFO-based Informed Consent Ontology (ICO). in *Proc. 5th International Conference on Biomedical Ontologies* (ICBO) (2014).
48. He, Y. et al. OPM: the Ontology of Precision Medicine and Investigation and its support for clinical data and metadata representation and analysis. in *Proc. 11th International Conference on Biomedical Ontology* (2019).
49. Hripcsak, G. et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Stud. Health Technol. Inform.* **216**, 574–578 (2015).
50. Eddy, S., Mariani, L. H. & Kretzler, M. Integrated multi-omics approaches to improve classification of chronic kidney disease. *Nat. Rev. Nephrol.* <https://doi.org/10.1038/s41581-020-0286-5> (2020).
51. Stokman, M. F. et al. The expanding phenotypic spectra of kidney diseases: insights from genetic studies. *Nat. Rev. Nephrol.* **12**, 472–483 (2016).
52. Salem, R. M. et al. Genome-wide association study of diabetic kidney disease highlights biology involved in glomerular basement membrane collagen. *J. Am. Soc. Nephrol.* **30**, 2000–2016 (2019).
53. Bandrowski, A. et al. The Ontology for Biomedical Investigations. *PLoS ONE* **11**, e0154556 (2016).
54. Smith, J. R. et al. The clinical measurement, measurement method and experimental condition ontologies: expansion, improvements and new applications. *J. Biomed. Semant.* **4**, 26 (2013).
55. He, Y. et al. OAE: the Ontology of Adverse Events. *J. Biomed. Semant.* **5**, 29 (2014).
56. Hastings, J. et al. ChEBI in 2016: improved services and an expanding collection of metabolites. *Nucleic Acids Res.* **44**, D1214–D1219 (2016).
57. Hanna, J., Joseph, E., Brochhausen, M. & Hogan, W. R. Building a drug ontology based on RxNorm and other sources. *J. Biomed. Semant.* **4**, 44 (2013).

Acknowledgements

The KPMP project is supported by the NIH National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) U2C project 1U2CDK114886. The authors appreciate the discussion with, editing by and support from Deborah Hoshizaki from the NIDDK and the KPMP consortium.

Author contributions

This review was coordinated by the KPMP, with S.M. and Y.H. providing conceptual guidance. E.O., J.S., J.F.O., B.S., F.D., L.B., S.J., C.P., S.M. and Y.H. contributed to the design, development, coordination and application of the KPMP ontologies. E.O., L.L.W., J.S., J.F.O., S.M. and Y.H. wrote the manuscript with support from B.S., S.J., I.H.D., S.S.W., D.C.C., C.S., C.W., A.D.D., C.J.M. and M.K. J.S., J.F.O., S.J., I.H.D., M.T.V., S.S.W., J. Himmelfarb and M.K. provided guidance on clinical nephrology, A.Z.R. and L.B. provided guidance on pathology, B.S. and F.D. provided guidance on data systems, D.C.C., T.A., C.R.A., C.S. and R.I. provided guidance on omics analyses and J. Hansen, C.W., A.D.D., C.M., M.H. and P.N.R. provided guidance on reference ontologies. All authors reviewed and approved the manuscript.

Competing interests

The authors declare no conflicts of interest.

Peer review information

Nature Reviews Nephrology thanks B. Smith and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41581-020-00335-w>.

RELATED LINKS

Allen Brain Atlas: <https://portal.brain-map.org/>
Encyclopedia of DNA Elements (ENCODE) project: <https://www.encodeproject.org/>
Human BioMolecular Atlas Program: <https://commonfund.nih.gov/hubmap>
Human Cell Atlas: <https://www.humancellatlas.org/>
International Classification of Diseases (ICD): <https://www.who.int/classifications/icd/en/>
Kidney Precision Medicine Project: <https://kpmp.org/>
Kidney Tissue Atlas data portal: <https://atlas.kpmp.org/>
Library of Integrated Network-Based Cellular Signatures: <http://www.lincsproject.org/>
Ontology of General Medical Science: <https://github.com/OGMS/ogms>
Ontology of Genes and Genomes: <http://www.obofoundry.org/ontology/ogg.html>
Open Biological and Biomedical Ontology (OBO) Foundry: <http://www.obofoundry.org/>
Systematized Nomenclature of Medicine (SNOMED): <https://www.snomed.org/snomed-ct/>
The Cancer Genome Atlas: <https://www.cancer.gov/tcga>

© Springer Nature Limited 2020