

CELL ONTOLOGY HISTORY AND STRUCTURE

Cell Ontology Training
September 24, 2020
Alexander D. Diehl

 Jacobs School of Medicine and Biomedical Sciences
University at Buffalo



The Cell Ontology (CL)

- Provides a formal ontological representation for cell types across biology
- Currently has ~**2235** cell type classes (terms)
- Acts as the designated high level ontology for cell types in the OBO Foundry
- The majority of the content of the CL is focused on vertebrate cell types, with a strong emphasis on mammals and humans
- By agreement, other communities work on different cell types in a compatible manner with the CL within the context of the OBO Foundry
 - Plant Ontology
 - Drosophila Gross Anatomy Ontology
 - Cell Line Ontology

The Cell Ontology (CL)

Direct download: <http://purl.obolibrary.org/obo/cl.owl>

Information: <http://www.obofoundry.org/ontology/cl.html>

Browsers: <http://www.ontobee.org>, <https://www.ebi.ac.uk/ols/index>, <https://bioportal.bioontology.org>

Github Repository: <https://github.com/obophenotype/cell-ontology>

Protégé editor: <https://protege.stanford.edu>

Cell Ontology History

1998	Gene Ontology start
2004	Cell Ontology start
2005 – 6	First major revision of CL (immune cell types)
2007	One of the founding members of the OBO Foundry
2009 – 11	Widespread introduction of logical definitions in CL
2012 – present	Use of CL in data analysis by major scientific initiatives (FANTOM 5, ENCODE, ImmPort, HCA, KPMP, HuBMAP ...)
2020 ongoing	Improvements in representation of species specific cell types

An ontology for cell types

[Jonathan Bard](#), [Seung Y Rhee](#) & [Michael Ashburner](#) 

[Genome Biology](#) 6, Article number: R21 (2005) | [Cite this article](#)

24k Accesses | 215 Citations | 3 Altmetric | [Metrics](#)

Logical Development of the Cell Ontology

[Terrence F Meehan](#) , [Anna Maria Masci](#), [Amina Abdulla](#), [Lindsay G Cowell](#), [Judith A Blake](#), [Christopher J Mungall](#) & [Alexander D Diehl](#) 

[BMC Bioinformatics](#) 12, Article number: 6 (2011) | [Cite this article](#)

7068 Accesses | 84 Citations | 0 Altmetric | [Metrics](#)

The Cell Ontology 2016: enhanced content, modularization, and ontology interoperability

[Alexander D. Diehl](#) , [Terrence F. Meehan](#), [Yvonne M. Bradford](#), [Matthew H. Brush](#), [Wasila M. Dahdul](#), [David S. Dougall](#), [Yongqun He](#), [David Osumi-Sutherland](#), [Alan Ruttenberg](#), [Sirarat Saengvijai](#), [Ceri E. Van Slyke](#), [Nicole A. Vasilevsky](#), [Melissa A. Haendel](#), [Judith A. Blake](#) & [Christopher J. Mungall](#)

[Journal of Biomedical Semantics](#) 7, Article number: 44 (2016) | [Cite this article](#)

2383 Accesses | 44 Citations | 7 Altmetric | [Metrics](#)

The OBO Foundry

- A collective of ontology developers who are committed to collaboration and adherence to a set of shared principles
 - open use, collaborative development, non-overlapping and strictly-scoped content, and common syntax and relations
- OBO Foundry principles ensure interoperability and non-redundancy
- Most OBO Foundry ontologies are compatible with the Basic Formal Ontology, and employ formal relations defined by the Relation Ontology
- OBO Foundry ontologies are therefore tightly integrated with each other
GO/CL/Uberon



The OBO Foundry

The Open Biological and Biomedical Ontology (OBO) Foundry is a collective of ontology developers that are committed to collaboration and adherence to shared principles. The mission of the OBO Foundry is to develop a family of interoperable ontologies that are both logically well-formed and scientifically accurate. To achieve this, OBO Foundry participants voluntarily adhere to and contribute to the development of an evolving set of principles including open use, collaborative development, non-overlapping and strictly-scoped content, and common syntax and relations, based on ontology models that work well, such as the Gene Ontology (GO).

The OBO Foundry is overseen by an Operations Committee with Editorial, Technical and Outreach working groups. The processes of the Editorial working group are modelled on the journal refereeing process. A complete treatment of the OBO Foundry is given in "The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration".

On this site you will find a table of ontologies, available in several formats, with details for each, and documentation on OBO Principles.

You can contribute to this site using GitHub [OBOFoundry/OBOFoundry.github.io](https://github.com/OBOFoundry/OBOFoundry.github.io) or get in touch with us by joining our mail list <https://groups.google.com/forum/#!forum/obo-discuss>.

Download table as: [[YAML](#) | [JSON-LD](#) | [RDF/Turtle](#)]

bfo	Basic Formal Ontology 	The upper level ontology upon which OBO Foundry ontologies are built. Detail	
chebi	Chemical Entities of Biological Interest 	A structured classification of molecular entities of biological interest focusing on 'small' chemical compounds. Detail	
doid	Human Disease Ontology 	An ontology for describing the classification of human diseases organized by etiology. Detail	
go	Gene Ontology 	An ontology for describing the function of genes and gene products Detail	
obi	Ontology for Biomedical Investigations 	An integrated ontology for the description of life-science and clinical investigations Detail	
pato	Phenotype And Trait Ontology 	An ontology of phenotypic qualities (properties, attributes or characteristics) Detail	

Over 100 ontologies covering basic biological and medical domains

<http://www.obofoundry.org>

1846

What is an ontology?

- A formal representation of the entities that exist in reality that are of interest to a field of knowledge, and the logical relationships that exist among those entities as well as between those entities and entities in related domains
- The meaning of ontology classes comes not from their names, but rather their definitions
Careful naming is important as an aid to both human users and for searching and applications like natural language processing
- Ideally ontology classes will have both textual and logical definitions
The axioms of logical definitions allow for reasoning across the ontology to infer implicit relationships among entities

What an ontology is not

- Not a terminology or nomenclature, although ontology names and ontology structure may reflect the accepted terminology or nomenclature of a field.
- Not a random collection of anything with a connection to a field of knowledge
- Not a database or knowledgebase
Ontology classes are typically defined with necessary and sufficient criteria and do not capture all facts or knowledge that is known about an entity

Domain Ontologies

Domain ontologies focus on a single area of knowledge

Representation of entities important in anatomy is split among multiple domain ontologies according to both levels of granularity and types of entity, such as whether something is a material entity, such as an organ, or a biological process that occurs in that entity

Protein Ontology – proteins, including isoforms and modified forms

GO Cellular Component – parts of the cell, including protein complexes

Cell Ontology – cell types

Uberon – tissues and organs

GO Biological Process – biological process occurring at any level of granularity

Application Ontologies

Application ontologies bring together relevant classes from multiple domain ontologies to provide an ontological representation of a complex area of knowledge.

For instance, the CCF Semantic Ontology developed by the HuBMAP project is an application ontology that includes relevant ontology classes from UBERON and the Cell Ontology that are needed to describe the anatomical parts and cell types under study by HuBMAP

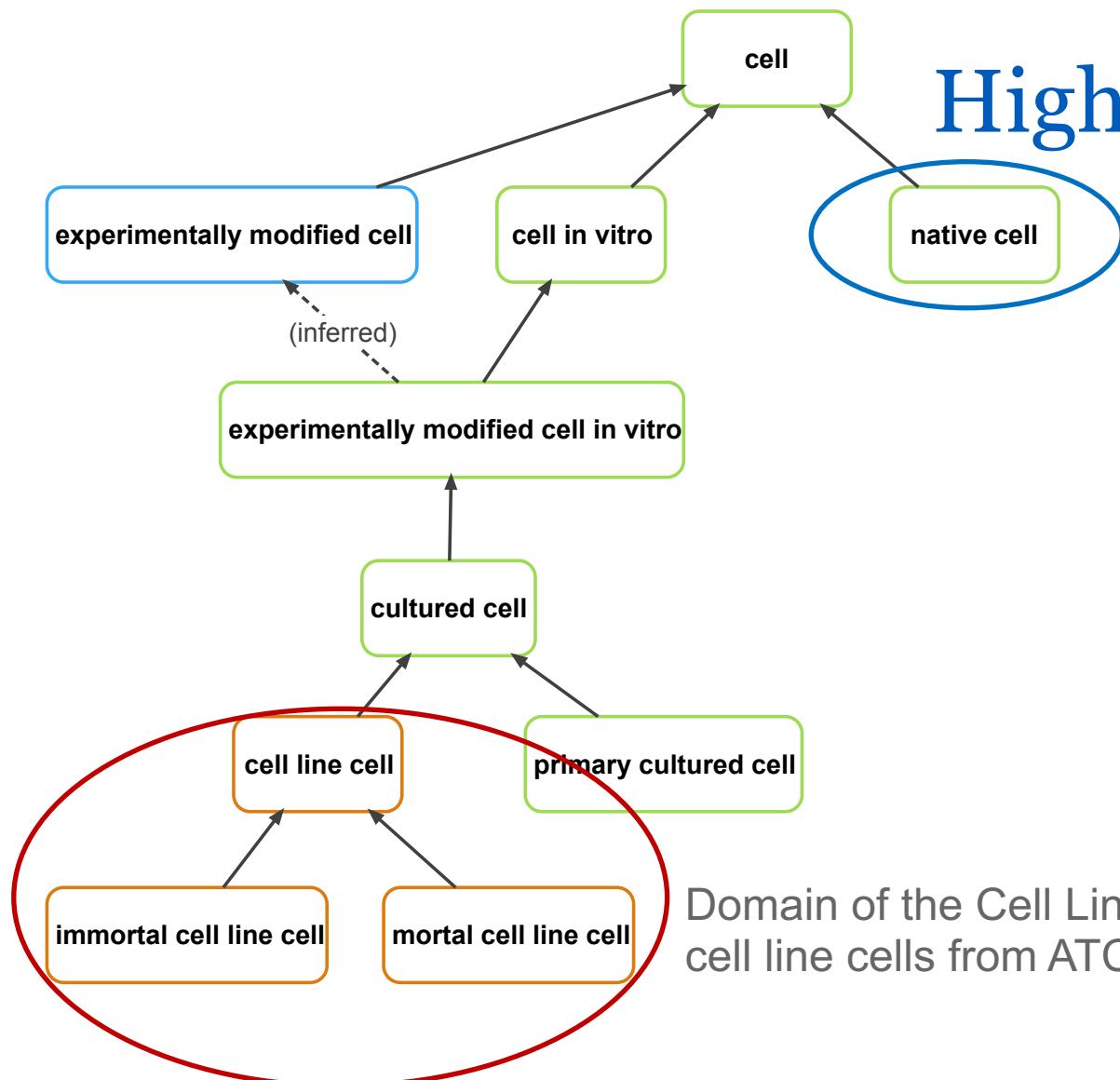
Ideally, developers of application ontologies will request new classes from core ontologies for relevant entities not currently represented in those ontologies

How We Represent Cells in the Cell Ontology

- Morphology
- Surface marker expression, singly or in combination
- Transcription factor expression or expression of other internal protein or gene, singly or in combination
- By lineage
- By function or capability

Types of Evidence Behind the Representation of Cells in CL

- Microscopy, with or without staining (histology)
- Immunofluorescence *in situ* or *in vitro*
- Flow cytometry or mass cytometry (CyTOF)
- Colony formation assays
- In vivo/in vitro lineage tracking
- Direct assays of cellular function, typically *in vitro*
- Indirect assays of cellular function *in vivo*
- Assays of gene expression, including scRNAseq



High Level Structure of CL

The Cell Ontology primarily represents native cells, normal cells in their *in vivo* locations and states

Additions to CL should be subclasses of 'native cell' in nearly all cases

Domain of the Cell Line Ontology (CLO), including cell line cells from ATCC, Coriell, and others

induced T-regulatory cell

Definition

CD4-positive alpha-beta T cell with the phenotype CD25-positive, CTLA-4-positive, and FoxP3-positive with regulatory function.

Exact Synonyms

induced regulatory T cell, induced regulatory T-cell, induced regulatory T lymphocyte, induced regulatory T-lymphocyte, induced Treg, iTreg

Related Synonyms

adaptive Treg, aTreg

Definition source

PMID:19464985

Immunity

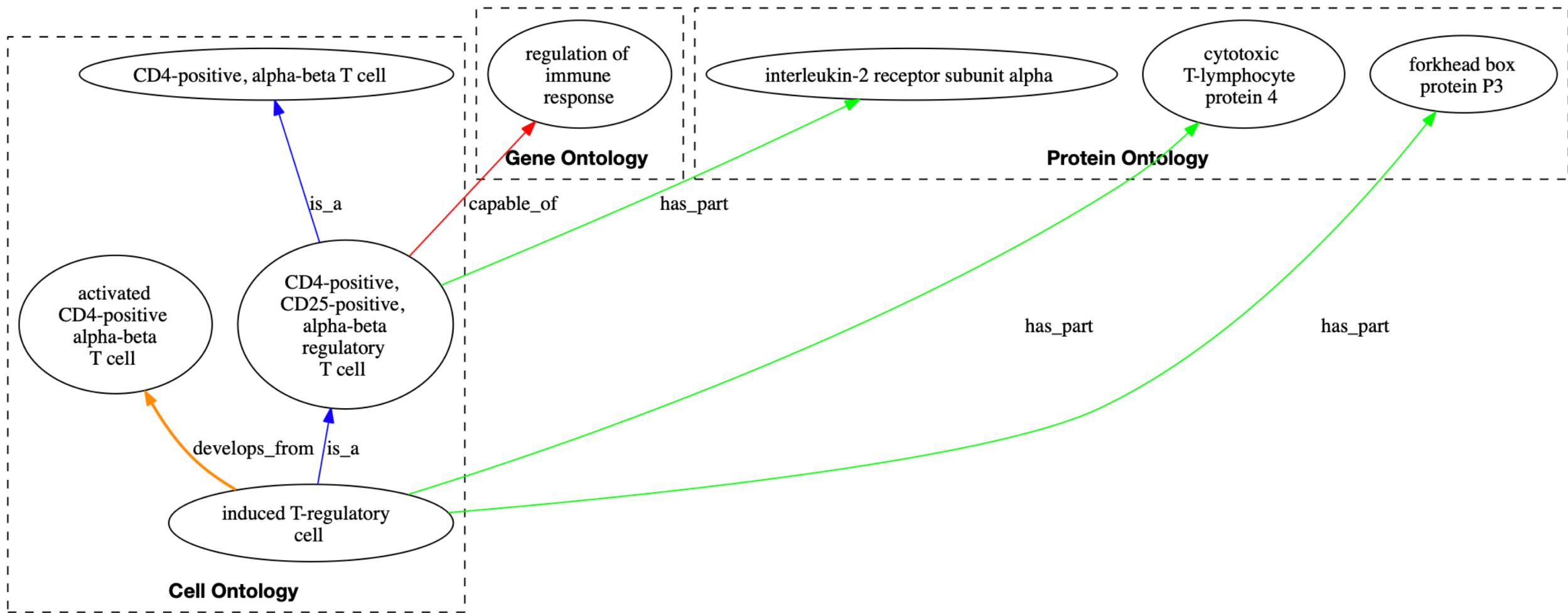
REVIEW | VOLUME 30, ISSUE 5, P626-635, MAY 22, 2009

Natural and Adaptive Foxp3⁺ Regulatory T Cells: More of the Same or a Division of Labor?

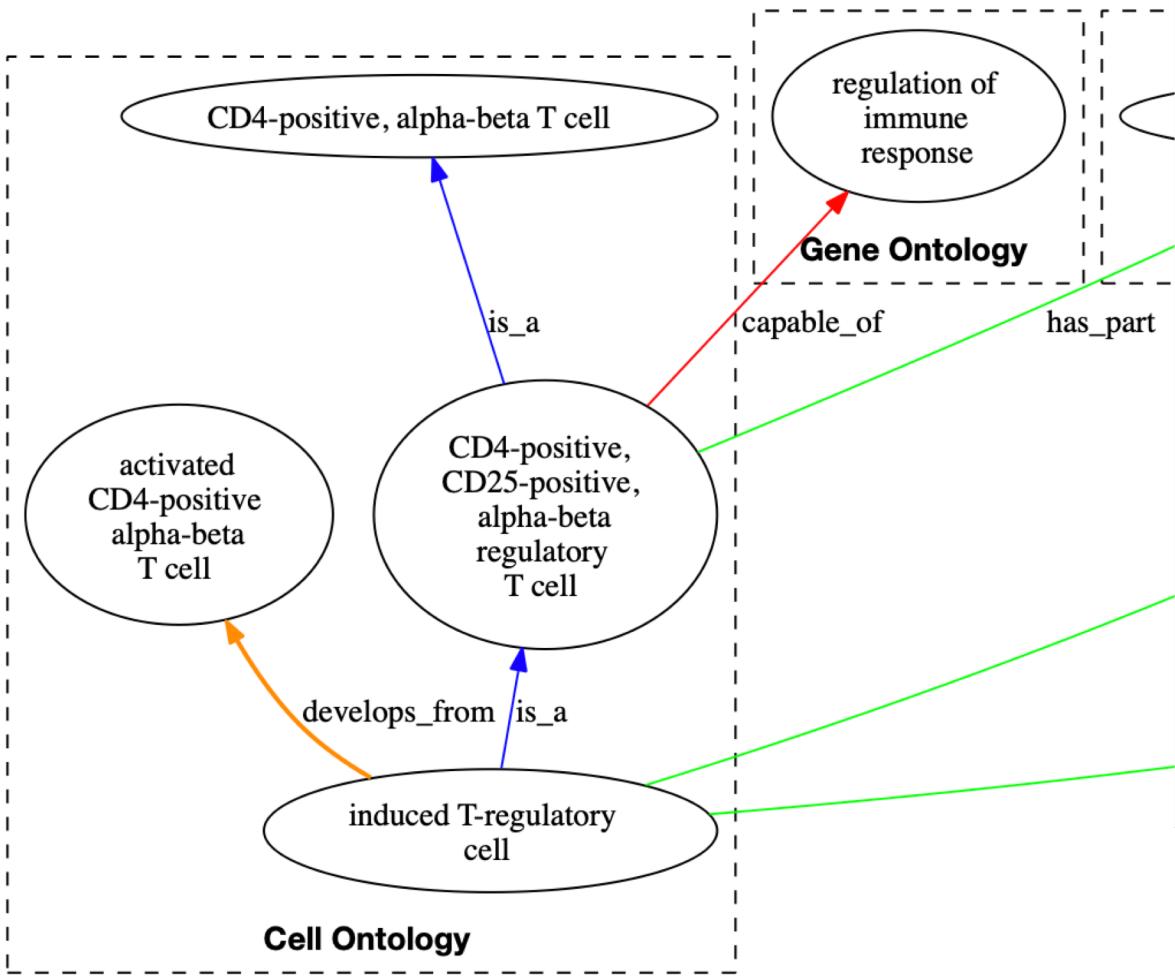
Maria A. Curotto de Lafaille • Juan J. Lafaille

Open Archive • DOI: <https://doi.org/10.1016/j.immuni.2009.05.002>

induced T-regulatory cell



induced T-regulatory cell



Description: 'induced T-regulatory cell'

Equivalent To +

- 'CD4-positive, CD25-positive, alpha-beta regulatory T cell'
and ('has part' some 'forkhead box protein P3')
and ('has plasma membrane part' some 'cytotoxic T-lymphocyte protein 4')
and ('develops from' some 'activated CD4-positive, alpha-beta T cell')

SubClass Of +

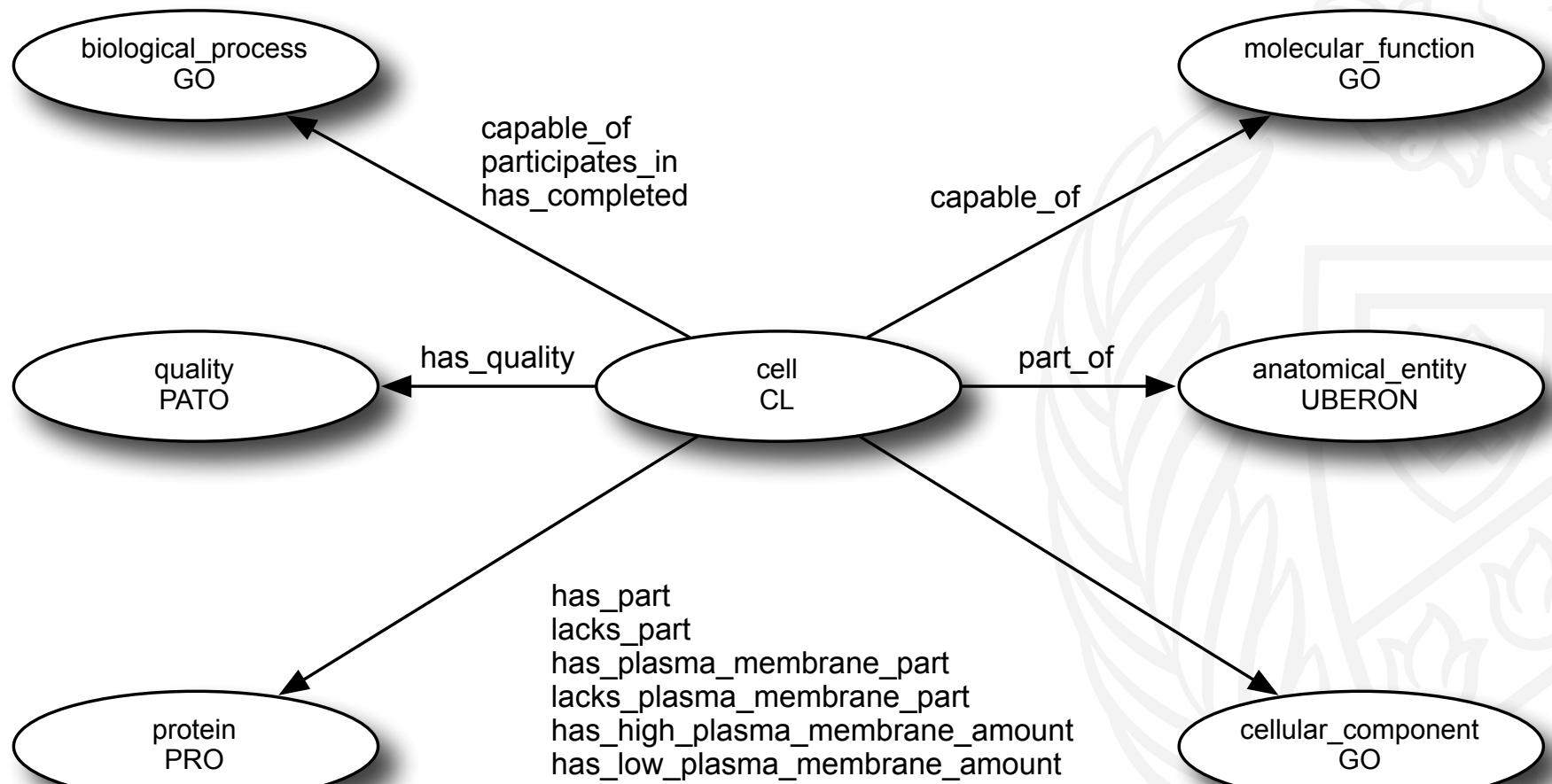
- 'CD4-positive, CD25-positive, alpha-beta regulatory T cell'

General class axioms +

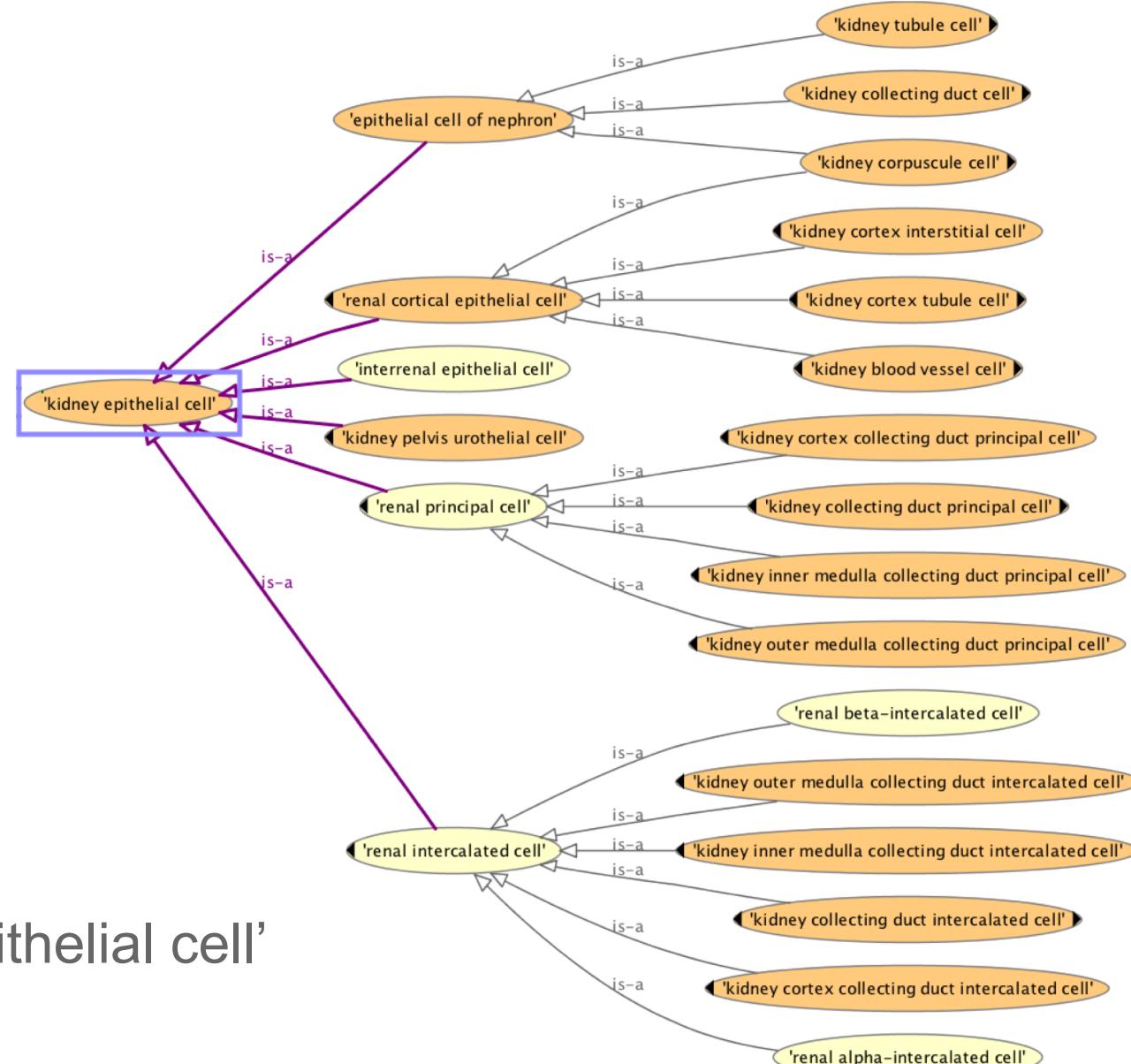
SubClass Of (Anonymous Ancestor)

- 'develops from' some 'hematopoietic stem cell'
- 'part of' some 'immune system'
- 'hematopoietic cell'
and ('capable of' some 'ameboidal-type cell migration')
- 'mature alpha-beta T cell'
and ('has plasma membrane part' some 'CD4 molecule')
and (lacks_plasma_membrane_part some 'T cell receptor co-receptor CD8')
- 'T cell'
and ('has plasma membrane part' some 'alpha-beta T cell receptor complex')
- 'develops from' some 'common lymphoid progenitor'

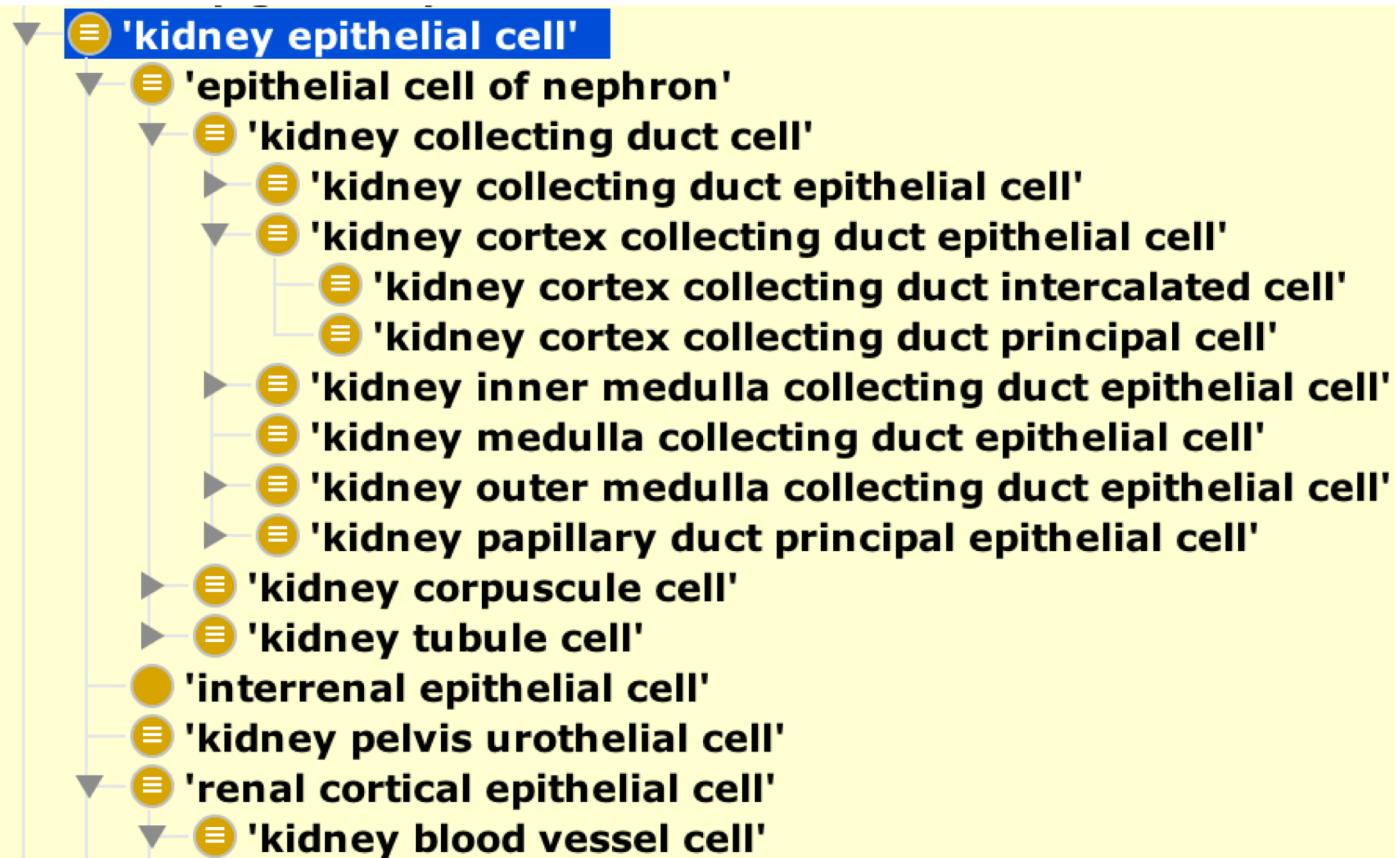
Connections to Other Ontologies



kidney epithelial cell



kidney epithelial cell



Annotations: 'kidney epithelial cell'

Annotations +

rdfs:label [type: xsd:string]

kidney epithelial cell

has_obo_namespace [type: xsd:string]

cell

definition [type: xsd:string]

An epithelial cell of the kidney.

database_cross_reference [type: xsd:string]

GOC:tfm

database_cross_reference [type: xsd:string]

KUPO:SJ

Description: 'kidney epithelial cell'

Equivalent To +

'epithelial cell'

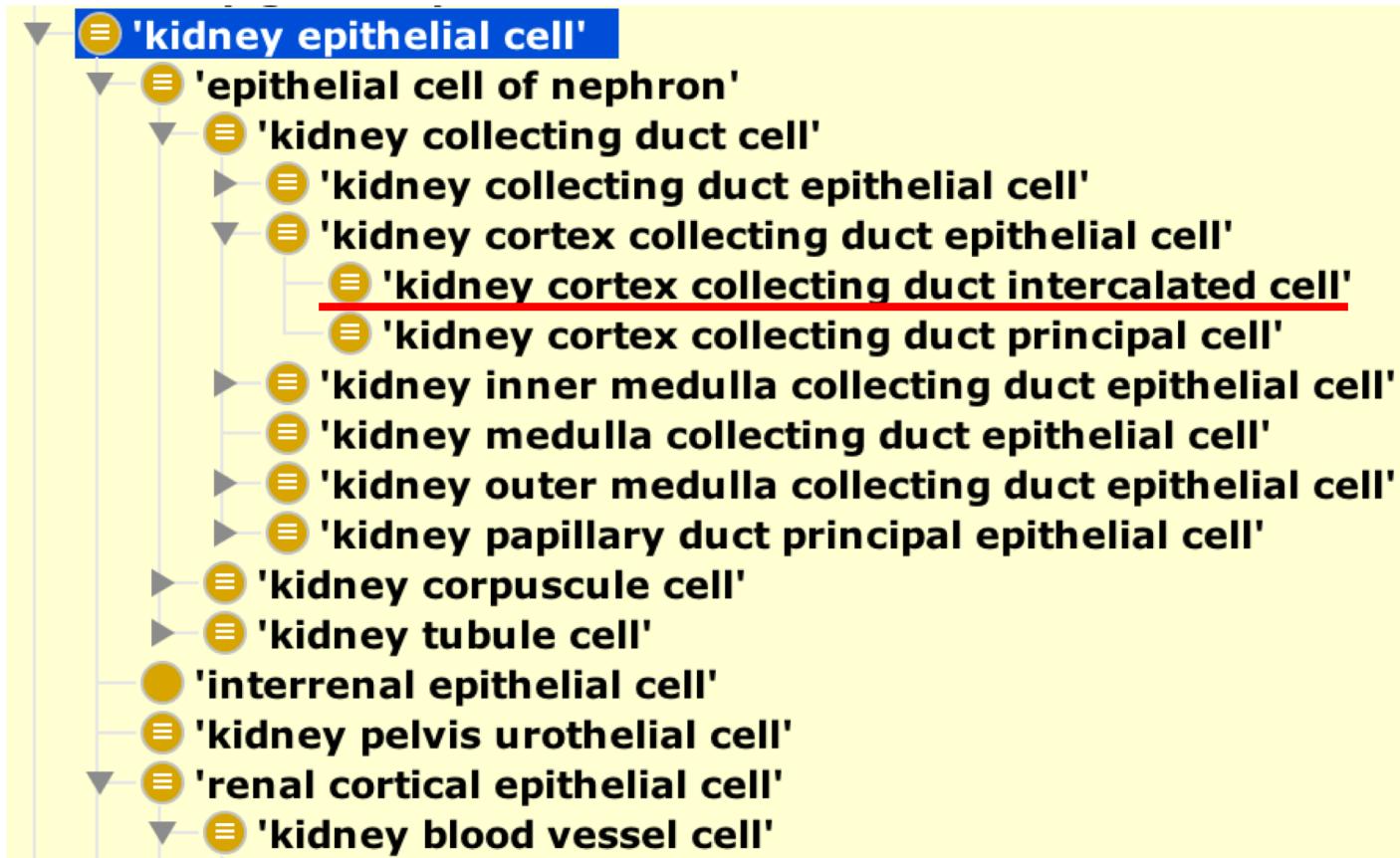
and ('part of' some **kidney**)

SubClass Of +

'kidney cell'

'epithelial cell'

kidney cortex collecting duct intercalated cell



Annotations: 'kidney cortex collecting duct intercalated cell'

Annotations 

rdfs:label [type: xsd:string]

kidney cortex collecting duct intercalated cell

has_obo_namespace [type: xsd:string]

cell

database_cross_reference [type: xsd:string]

KUPO:0001131

Description: 'kidney cortex collecting duct intercalated cell'

Equivalent To 

'renal intercalated cell'

and ('part of' **some **cortical collecting duct**)**

SubClass Of 

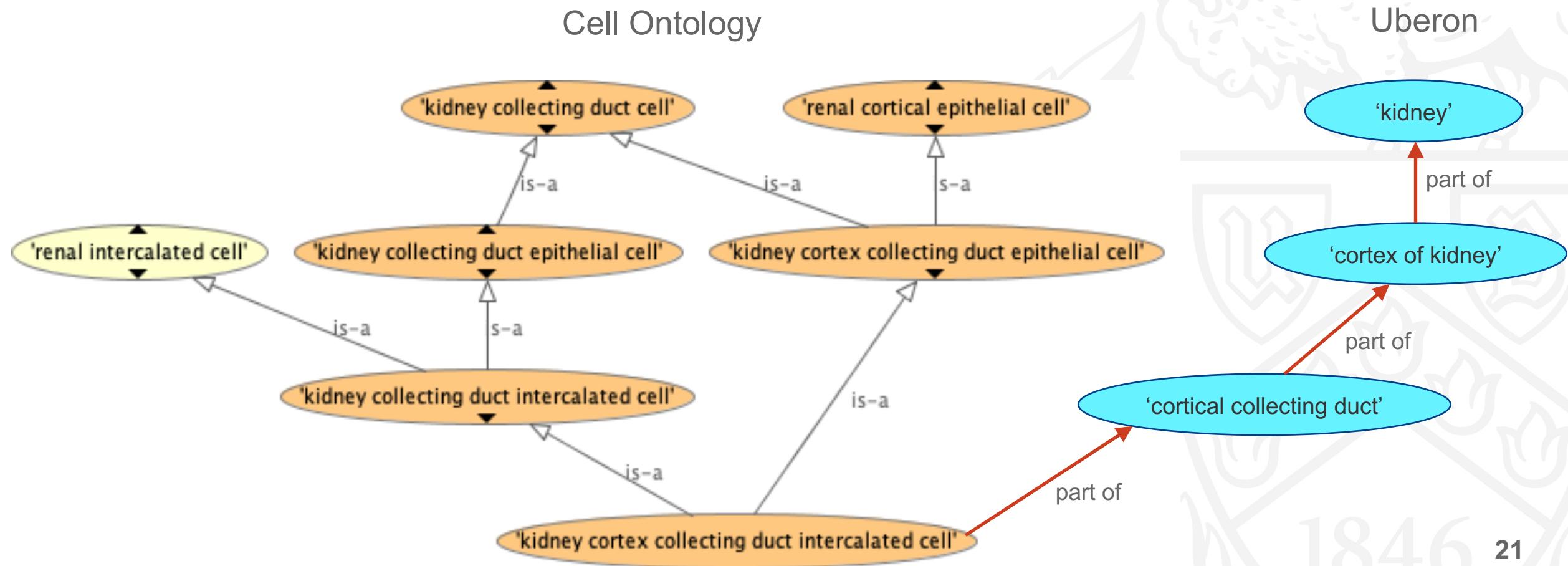
'kidney cortex collecting duct epithelial cell'

'part of' **some 'cortical collecting duct'**

'renal intercalated cell'

'kidney collecting duct intercalated cell'

kidney cortex collecting duct intercalated cell



A Deluge of Cell Type Data is on the Way...

- High Dimensional Flow Cytometry
- Mass Cytometry (CyTOF)
- Single Cell RNA Sequencing and CITE-seq
- Enhanced *in situ* methods for studying gene/protein expression

- Major Projects including
 - Human Cell Atlas
 - Kidney Precision Medicine Project
 - HuBMAP

Novel cell types and the Cell Ontology

Ideally, cell types identified via mass cytometry or scRNA seq can be identified as subtypes of known cell types in the Cell Ontology

New cell types in the Cell Ontology ideally should be substantiated by multiple lines of evidence

The Cell Ontology serves multiple communities, all of whom may contribute new cell types or request revisions or improvements to particular cell type classes or the hierarchy

CITE-Seq results show that cell populations defined by protein expression may correspond to multiple clusters based on scRNAseq, due to stochastic variations in gene expression. Thus we need to be careful not to create excessively granular cell types based solely on scRNAseq.

Challenges in Cell Naming and Definition

Standards for naming of novel cell types need to be established. We need names to be meaningful and hopefully agreeable across interested communities. Synonyms can be used to capture alternative names for the same cell type

Good definitions are more important than good names. Good textual definitions and reference are important

Methods to elucidate minimal necessary and sufficient criteria to molecularly define novel cell types must ensure that such definitions are unambiguous with regard to similar cell types across the body

We need to resolve conflicts when separate groups submit differing ways of defining the same cell type (choice between combined definition or create new sibling or subclasses)

Considerations in Use of CL in Data Annotation

Cell type classes in the Cell Ontology used to annotate particular cell type data should be determined based on both name and definition

The data about an observed cell or cell population should not contradict the definition of the CL class

The most granular class in CL that can be used should generally be picked

Not all criteria of the CL need to be fulfilled if information is lacking, but in some cases a more general cell type may be more appropriate

CELLS 2020 Workshop (tomorrow)

11 am – 1:30 pm ET Friday, September 25

Four talks about aspects of ontologies related to describing cell types

- Shrikant Pawar et al., Depositing single cell sequencing data to shared data portals
- David Osumi-Sutherland, Building semantics driven resources for neurobiology
- Yongqun He et al., The Cell Line Ontology 2020: expanded content and ontology interoperability
- Alexander D. Diehl, et al., Innate lymphoid cells and a new paradigm for the implementation of taxon restrictions in the Cell Ontology

Website/agenda: <https://sites.google.com/view/cells-2020-workshop/home>

Zoom: <https://buffalo.zoom.us/j/98084911233?pwd=bGRrOTVrVW1Ydm1KQ2V2NVZ0aUJKdz09>

Cell Ontology Team and Contributors

Christopher Mungall
Terrence Meehan
Nicole Vasilevsky
David Osumi-Sutherland
Melissa Haendel
Bjoern Peters
James Overton
Randi Vita
Steve Kleinstein
Patrick Dunn
Nico Matentzoglu
Richard Scheuermann
Brian Aevermann

David Dougall
Ceri Van Slyke
Yvonne Bradford
Anna Maria Masci
Ryan Brinkmann
Jonathan Bona
Oliver He
Sirarat Sarntivijai
Lindsay Cowell
Darren Natale
Judith Blake
Alan Ruttenberg
Barry Smith

Gene Ontology Consortium
Protein Ontology Consortium
Human Immunology Project
Consortium

NHGRI, NIAID, NIGMS
And many more...