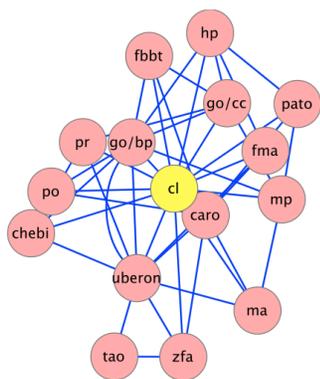


# Modularization for the Cell Ontology

Chris Mungall, Melissa Haendel, Amelia  
Ireland, Shahid Manzoor, Terry Meehan,  
David Osumi-Sutherland, Carlo Torniai,  
Alexander Diehl

# Outline

- The Cell Ontology and its neighbors
- Handling multiple species
- Extracting modules for taxonomic contexts
- The Oort – OBO Ontology Release Tool



# Inter-ontology axioms

Source	Ref	Axiom
cl	go	' <b>mature eosinophil</b> ' SubClassOf capable of some 'respiratory burst'
go	cl	'eosinophil differentiation' EquivalentTo differentiation and has_target some <b>eosinophil</b> *
cl	pro	' <b>Gr1-high classical monocyte</b> ' EquivalentTo 'classical monocyte' and has_plasma_membrane_part some 'integrin-alpha-M' and ....
cl	chebi	<b>melanophage</b> EquivalentTo 'tissue-resident macrophage' and has_part some melanin and ...
cl	uberon	' <b>epithelial cell of lung</b> ' EquivalentTo 'epithelial cell' and part_of some lung
uberon	cl	muscle tissue EquivalentTo tissue and has_part some <b>muscle cell</b>
clo	cl	HeLa cell SubClassOf derives_from some (' <b>epithelial cell</b> ' and part_of some cervix)
fbbt	cl	fbbt:neuron EquivalentTo cl:neuron and part_of some <i>Drosophila</i> **

\* experimental extension

\*\* taxon-context bridge axiom

# The cell ontology covers multiple taxonomic contexts

Ontology	Context	# of links to cl
MA	adult mouse	1
FMA	adult human	658
XAO	frog	63
AAO	amphibian	
ZFA	zebrafish	391
TAO	bony fish	385
FBbt	fruitfly	53
PO	plants	*
CL	all cells	-

# Xref macros: an easy way to formally connect to broader taxa

```
treat-xrefs-as-genus-differentia: CL part_of NCBITaxon:7955
...
[Term]
id: ZFA:0009255
name: amacrine cell
xref: CL:0000561
is_a: ZFA:0009051 ! interneuron
relationship: part_of ZFA:0000119 ! retinal inner nuclear layer
```



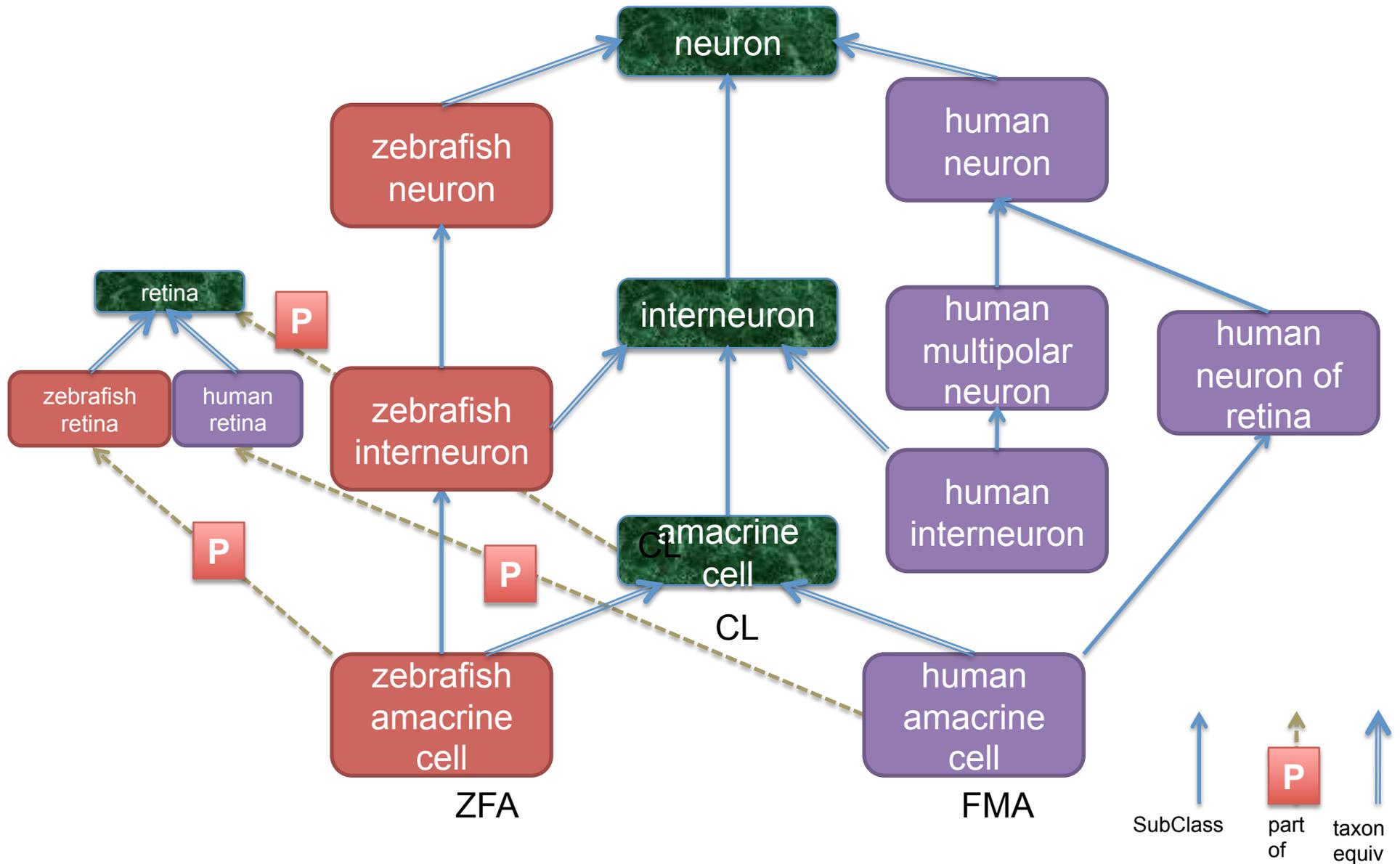
```
Class: ZFA_0009051 (amacrine cell)
EquivalentTo: CL_0000561 (amacrine cell) and BFO_0000050 (part of) some
NCBITaxon_7955 (Danio rerio)
Annotations: 'obo foundry unique label' 'zebrafish amacrine cell'
...
```

ZFA:amacrine cell is *the taxonomic equivalent of* CL:amacrine cell in zebrafish

In the context of ZFA, it is called simple 'amacrine cell';

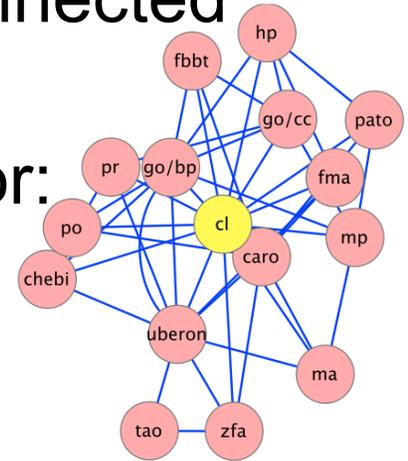
In the context of the entire OBO Foundry, it is called 'zebrafish amacrine cell'

# Bridging species through owl:imports



# Modularization for CL

- The OBO Library is becoming more connected
  - can be imported into a single ontology
- Subsets or sub-modules are required for:
  1. **Efficient Reasoning**
  2. **Improving Statistical Analyses**
  3. **User Views**
- There is a particular need for *taxonomic* subsets.
  - CL is a *multi-species* ontology
  - E.g subset of phenotype+GO+anatomy+CL for *Aves*



# 1. Subsets for Efficient Reasoning

- Goal:
  - Get subset that retains important entailments
- Multiple approaches
  - OWL API modularization tool
  - OntoFox
  - FlyBase Perl tools
  - OboEdit isa closure
- Is this even required given new fast reasoners?
  - CB
  - Elk

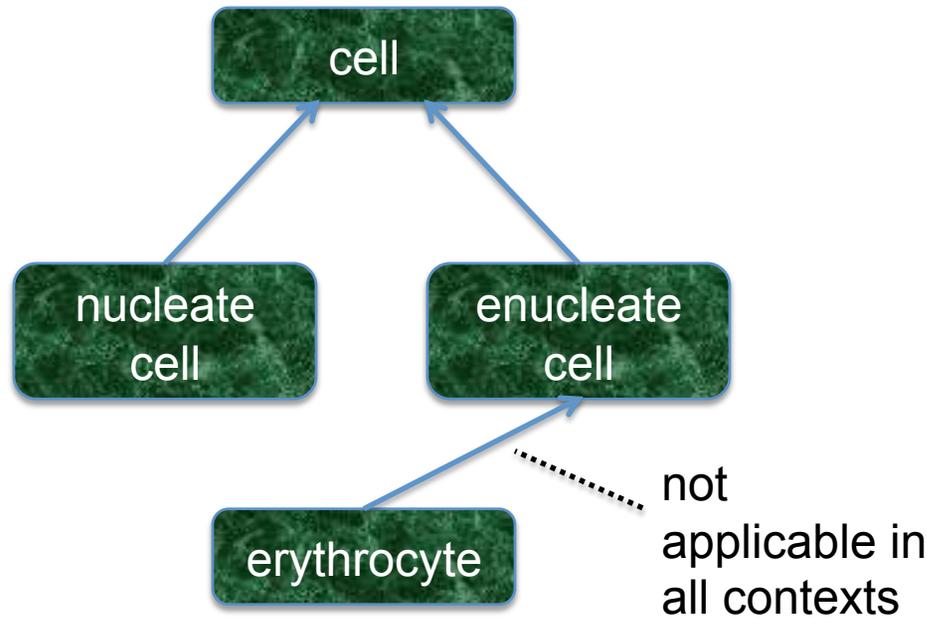
## 2. Subsets for Statistical Analysis and Probabilistic Modeling

- Goal:
  - subset that covers desired hypotheses
    - Proliferation of classes increases hypothesis space
      - Results in increased  $p$ -values
    - Some probabilistic methods scale poorly with number of classes
- Various solutions applied for GO
  - Manually generated subsets (aka slims)
    - custom tools for slim generation and extraction
  - Information-theoretic approaches
  - ...

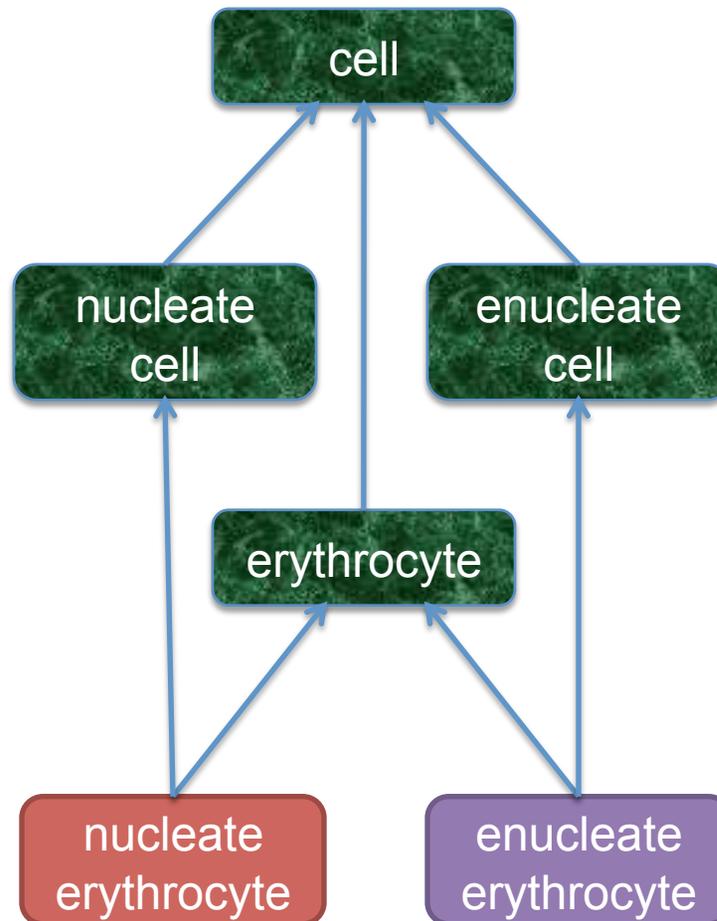
# 3. Subsets for User Views

- Users require views covering their domain
  - E.g: cells applicable for mammalian annotation
- Hide inherent complexity of multi-species ontologies
  - Reducing possible contexts reduces complexity
- How do we make these modules?

# Example of complexity arising from multiple species-contexts

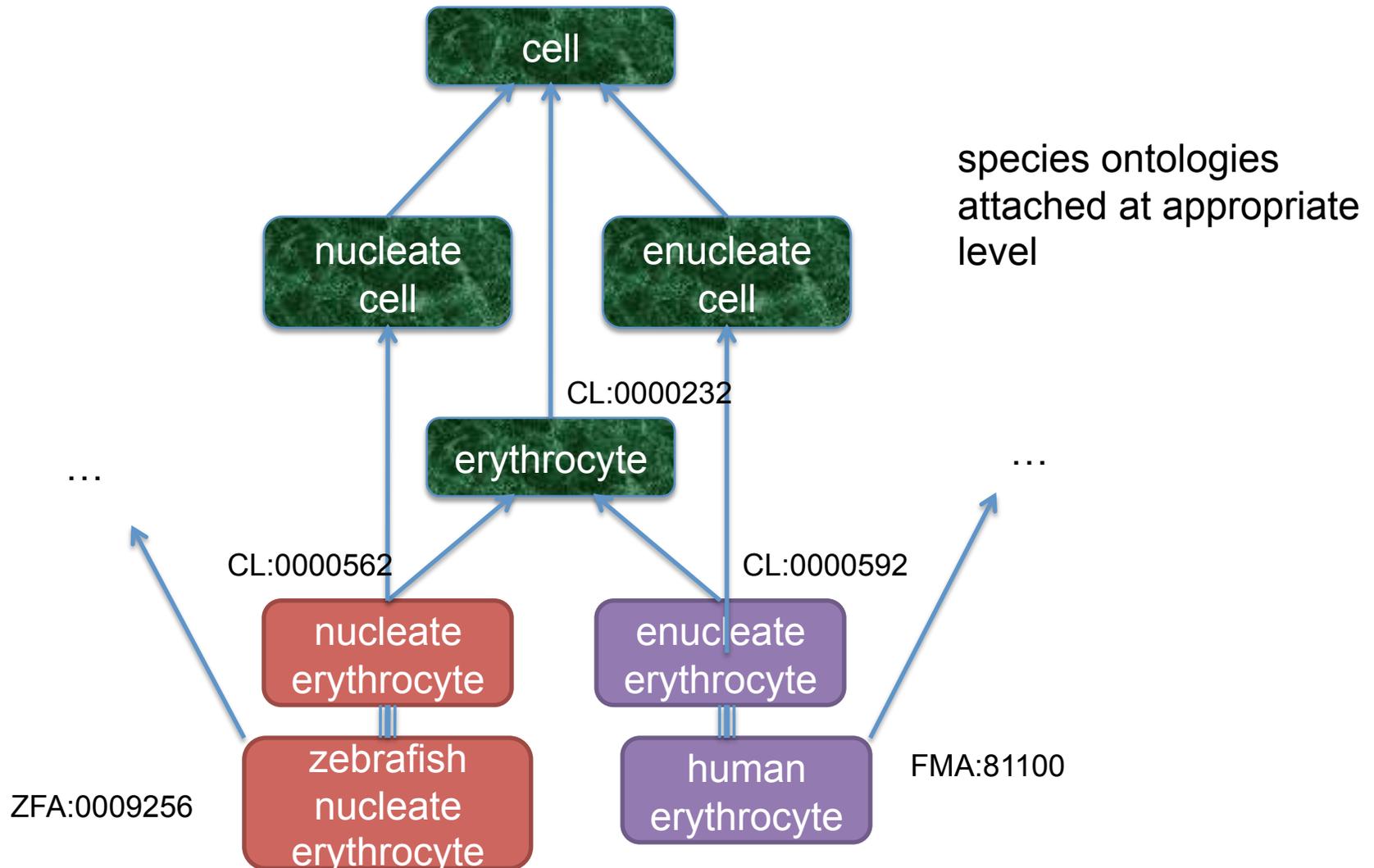


# Example of complexity arising from multiple species-contexts



“exceptional subclasses”

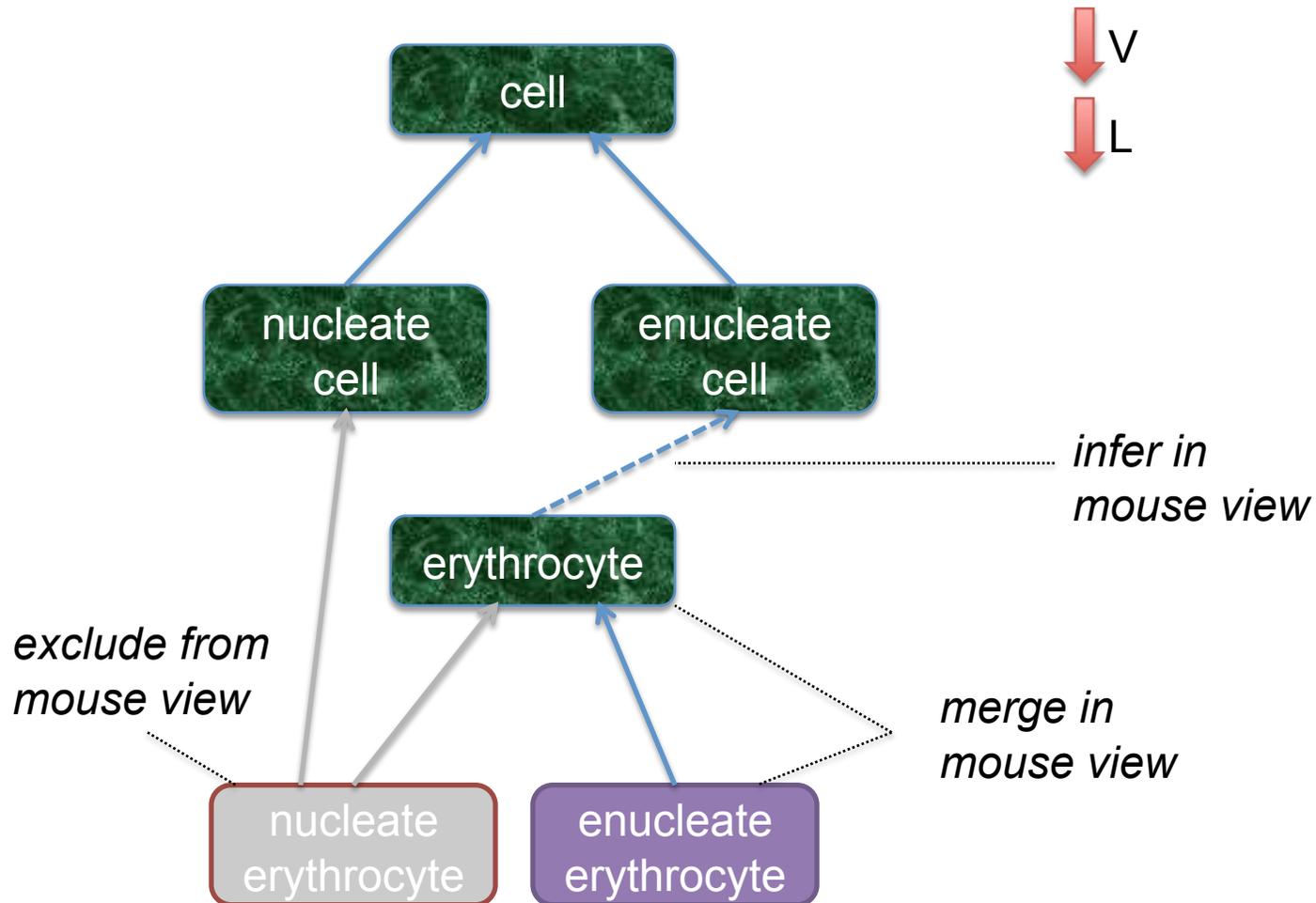
# Example of complexity arising from multiple species-contexts



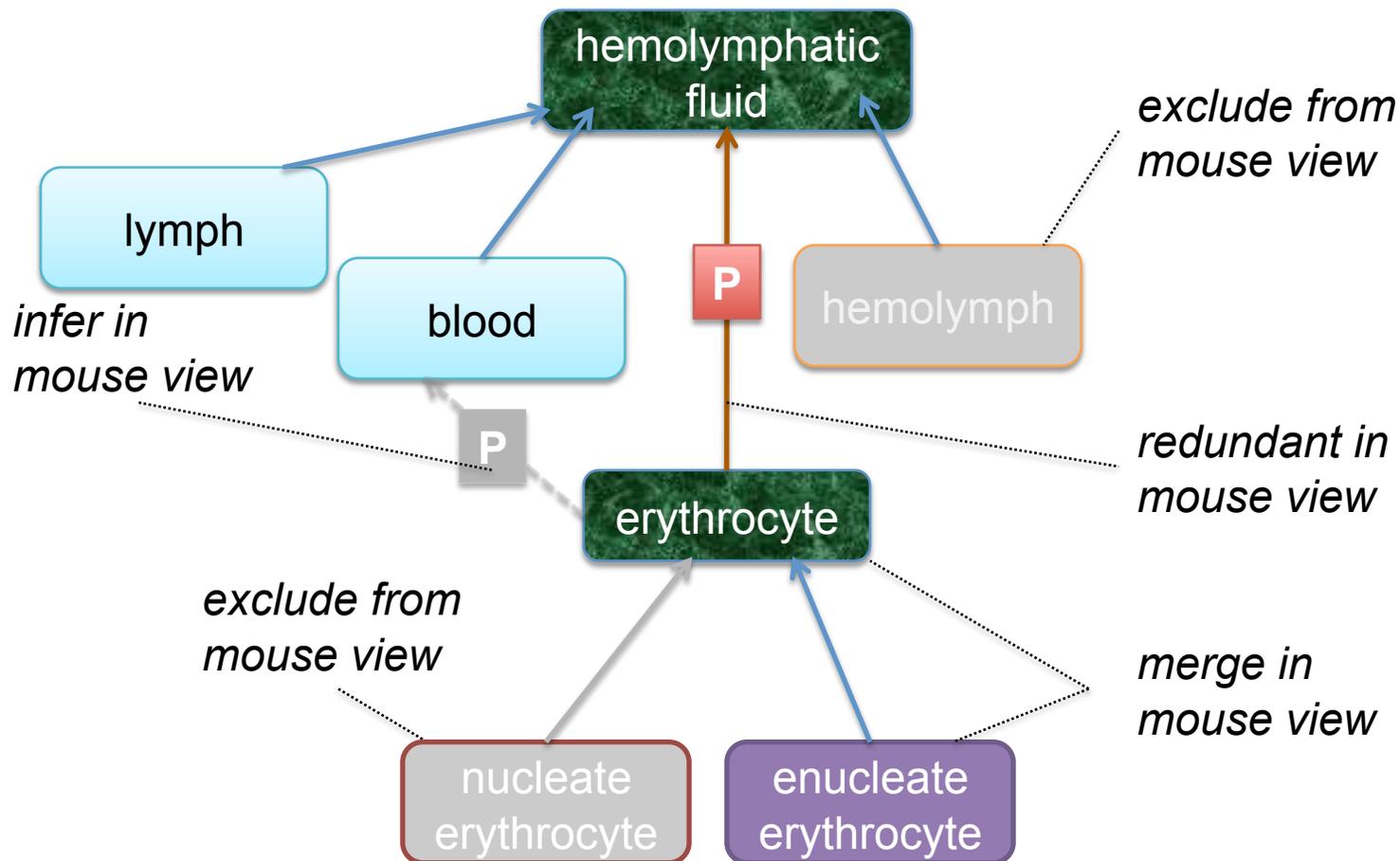
# The VPL hypothesis

- Variables:
  - **V** : Variability of entities in domain
  - **P** : Logical Precision
    - $TP/(TP+FP*c)$
  - **L** : Latticeyness of subsumption hierarchy
    - 'exception hierarchy'
- Hypothesis:
  - **Increasing variability or precision results in increased complexity of ontology**
    - (assuming DL formalism)

# Contextual views reduce complexity



# Contextual views reduce complexity



# Extracting Taxonomic Modules

## Goal

- Generate an ontology that includes a subset of the axioms in CL and surrounding ontologies

## Approach:

- Add taxonomic knowledge to multi-species ontologies (make contexts explicit)
- Use reasoners to generate subsets

# Adding taxonomic knowledge to multi-species ontologies

- Kusnierczyk *JBI 2008*
  - Taxonomy-based partitioning of the Gene Ontology
- Deegan *BMC Bioinformatics 2010*
  - Added the following links to GO:
    - `only_in_taxon`
    - `never_in_taxon`
  - Used primarily for annotation checking
- Phenoscope
  - Annotating variability across taxa

# Semantics of shortcut taxon relations

- Semantics for anatomy ontologies:
  - `only_in_taxon ?Y` → `part_of only ?Y`
  - `never_in_taxon ?Y` → `part_of only not ?Y`
- Example:
  - ‘mammary gland’ `only_in_taxon Mammalia`  
→ (expands to)
  - ‘mammary gland’ `part_of only Mammalia`

# Generating subsets

- Use standard reasoners to generate subsets
- Test each class in multi-species ontology
  - Is the corresponding taxon-specific class *satisfiable*?
  - If so, add it to the subset

Generating a subset  $S^T$  for context T from ontology O:

Initialize  $S^T = \{\}$ ,  $U^T = \{\}$

For every C in O

test if (D and T) is satisfiable

if satisfiable : add C to  $S^T$

# Example: generating chicken subset of CL

- Exclude all cell types specific to mammalian organs
  - E.g. myoepithelial cells of mammary glands
- Reasoning involves multiple ontologies, incl uberon

Does C belong in  $S^{\text{Gallus}}$  ?:

```
'mammary gland' SubClassOf part_of only mammal [uberon]
mammal disjoint from aves [ncbitaxon, enhanced]
Gallus SubClassOf Aves [ncbitaxon]
'myoepithelial cell of mammary gland' EquivalentTo 'myoepithelial cell'
and part_of some 'mammary gland' [cl]
```

→

```
('myoepithelial cell of mammary gland' and part_of some Gallus)
SubClassOf Nothing [reasoner]
```

# Pre-generated subsets with CL

- These will be part of CL releases in the future
- Ontologies:
  - **cl** : main ontology, plus imports or merged axioms
  - **cl-*<TAXON>*** : taxon subset of cl, plus import of taxon-ontologies, plus import of taxon bridge axioms
  - E.g.
    - cl-Mammalia.{obo,owl}

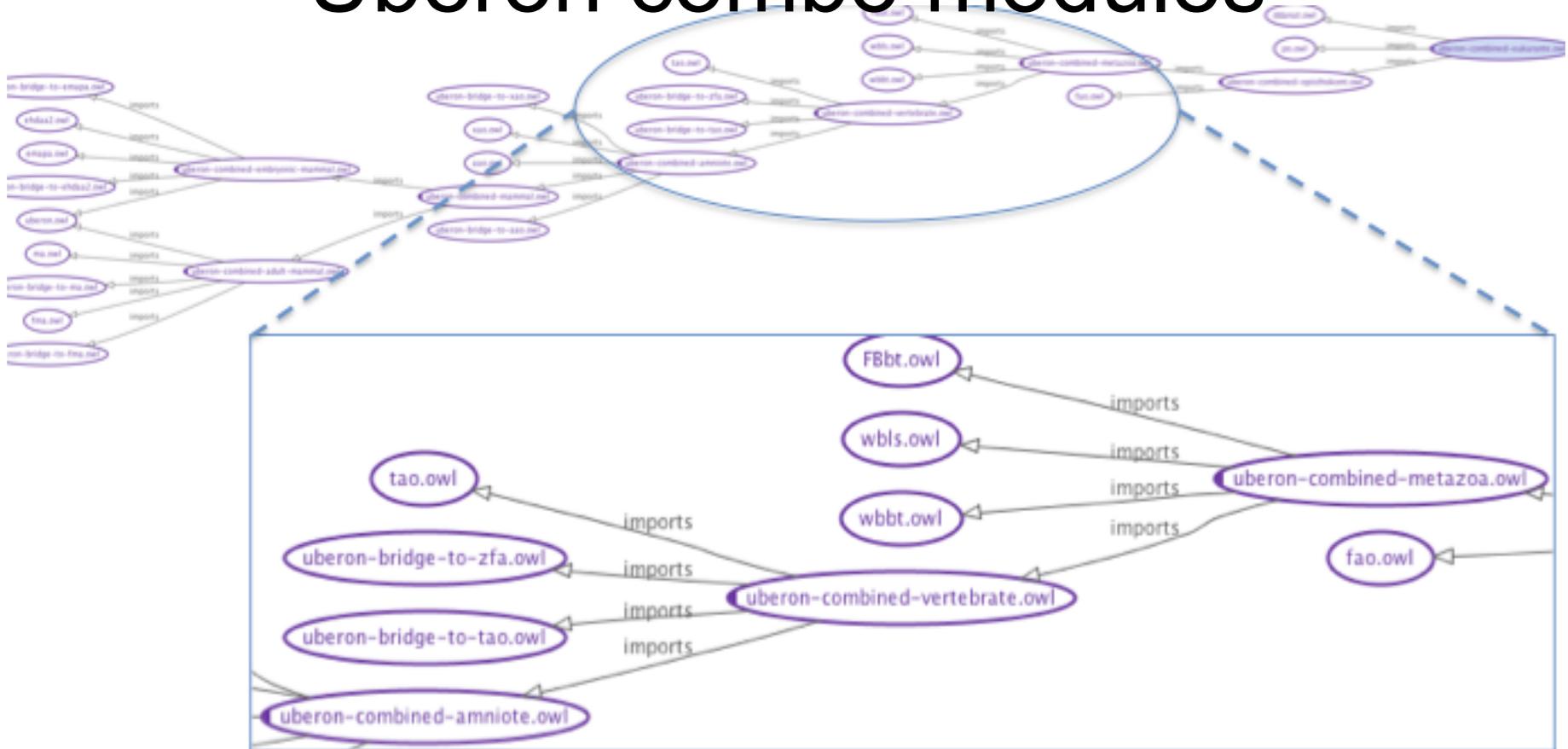
# Generation of multiple modules using The Oort

- OBO Ontology Release Tool
  - Generates obo and owl for public release
- Current capabilities
  - performs classification and consistency checking
    - Choice of OWL reasoner
  - Generates different versions:
    - main
    - simple
    - non-classified
  - Makes datastamped versions
    - <http://obofoundry.org/id-policy.html>

# Oort Extensions in progress

- Generating subset ontologies
  - E.g. go-slims
- Generation of subsets through taxon reasoning
  - E.g. cl\_Mammalia
- Expanding xref macros into bridge ontologies
  - Writes SubClassOf or EquivalentTo axioms
  - Generates obo-foundry-unique-label
  - E.g.
    - cl-bridge-to-ZFA
- Checking consistency of bridge ontology

# Example: OWL imports closure for Uberon combo modules



<http://purl.obolibrary.org/obo/uberun/mod/uberun-combined-metazoa.owl>

# Acknowledgments

- Phenoscape/RCN
  - Wasila Dahdul
  - Jim Balhoff
  - Peter Midford
  - Hilmar Lapp
  - Paula Mabee
  - Suzanna Lewis
- Plant Ontology
  - Laurel Cooper
  - Ramona Walls
  - Pankaj Jaiswal
- ZFIN
  - Yvonne Bradford
  - Ceri Van Slyke
  - Monte Westerfield
- CL
  - Richard Scheuermann
- OBO
  - Alan Ruttenberg
  - Melanie Courtot
- GO
  - GO editors