

To process COVID-19 knowledge, we need to improve reporting and modelling of cell types

Current OBO ontologies are not sufficient to annotate COVID-19-related cell types



Tiago Lubiana (<https://lubianat.github.io/>)
Computational Systems Biology Laboratory - INOVA USP,
University of São Paulo, Brazil

Introduction:

- Animal cells exist in many different flavours
- The [Cell Ontology](#) (CL) has catalogued many of the cell types mentioned in biomedical research
- CL classes can be concatenated with descriptors from other ontologies for precise matching
- Here I describe an attempt to annotate cell types in articles that employed single-cell RNA sequencing to study COVID-19.

Approach:

- For 5 different articles, I attempted to manually annotate cell types based on the Cell Ontology and the following ontologies:
 - [NCBITaxon](#) (for species)
 - [UBERON](#) (for tissues)
 - [MMUSDV](#), [HSAPDV](#) (for stages of life)
 - [PATO](#) (for biological sex)
- I've considered all mentions of classes of cells as putative cell types.
- I decided from start to use only the information in main figures and their captions.
- Annotations were recorded in an [openly available Google Sheet](#)

Results:

OBS: All cell type mentions were alongside a description of the species and none of biological sex. Overall, failures stem either from absence in paper or failure to reconcile article terms to an ontology (see [table](#) for details)

Number of cell type mentions and matches to ontology per article

Article	Cell Type Mentions	CL	UBERON	UBERON and CL	Life Stage (MMUSDV or HSAPDV)	Life Stage, UBERON and CL
Muus et al	170	106	69	46	35	12
Liao et al	20	12	0	0	0	0
Wen et al	36	27	15	11	0	0
Qi et al	100	73	100	73	0	0
Wilk et al	34	20	0	0	0	0
All five	360	238 (66%)	184(51%)	130 (36%)	35 (9.7%)	12 (3.3%)

Examples of cell types that were not found in the Cell Ontology:

- Terms with lineage relations
 - “fetal airway progenitor”
 - “t cell lineage”
 - “multiciliated lineage”
 - “fibroblast lineage”
- Terms related to the cell-cycle
 - “cycling cells”
 - “proliferating cells”
 - “proliferative cells”
 - “dividing cells”
- Terms with specific markers:
 - “BEST4+ cells”
 - “lung KRT5-/ KRT17+ cells”
- Terms related to adaptive response:
 - “clonal B cell”
 - “non-clonal B cell”
 - “expanded B cell”
 - “non-expanded B cell”

Conclusion:

To achieve precise and comprehensive annotation of cell typed, the biomedical research community will need (1) to improve standards of reporting and (2) keep the effort to further develop the relevant ontologies.