

SOP: Construction of Anatomical Structure and Cell Types and Biomarker (ASCT+B) Tables

Process for construction of ASCT+B tables

Authors: Ellen M. Quardokus, Lisel Record, Bruce W. Herr II, Hrishikesh Paul, Katy Börner

Approved by: Katy Börner (3/3/2021)

HIVE MC-IU Team, Indiana University

PI: Katy Börner

NIH Award No: 1OT2OD026671

Updated: March 3, 2021

If there are any questions, contact MC-IU via HuBMAP CCF Info infoccf@iu.edu.

Introduction

Anatomical Structures, Cell Types, plus Biomarkers (ASCT+B) Master tables aim to capture the nested *part_of* structure of anatomical human body organ systems from gross anatomical structure scale to subcellular biomarker scale. Functional tissue units (FTUs) for an organ system should be identified as well as the typology of cells and biomarkers used to uniquely identify cell types within that organ system and FTUs (e.g., gene, protein, proteoforms, lipid or metabolic markers).

Ontology terms and unique identifiers are matched to AS, CT, and B wherever possible for semantic search capability within MC-IU products: Registration User Interface (RUI), Exploration User Interface (EUI), ASCT+B Reporter. The tables are authored and reviewed by an international team of anatomists, pathologists, physicians, and other experts.

This document is intended for subject matter experts (pathologists, surgeons, dissectionists, molecular biologists, bioinformaticians, human organ or cell atlas consortia members, , ontologists, individuals from HuBMAP TMCs, HuBMAP PIs) who are involved in ASCT+B table construction or review.

Motivation

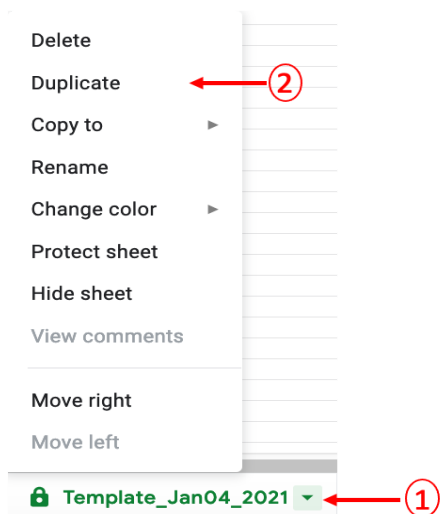
Existing ontologies (Uberon, FMA, CL, HGNC, etc.) have thousands of terms and form a complex knowledge graph. They were designed for different purposes and frequently cover developmental and disease terms. For developing a reference atlas of the healthy human adult body within

HuBMAP, we need a unified view of human anatomical structures (AS), cell types (CT), and biomarkers (B). In close collaboration with other consortia, we agreed to focus on *part_of* relationships so that the network graph can be simplified to a hierarchy. This makes it easier to develop user interfaces that enable investigators to quickly drill down, in an intuitive way, from whole-body, to organ, to organ parts, to cell types, and eventually to specific biomarkers assays associated with those cells. Plus, the ASCT+B tables capture the *located_in* relationships of CTs and AS and the *characterizes* relationships between Bs and CTs. The ASCT+B tables are not creating new ontologies. Rather, they help construct a uniform and **simplified view** of AS, CT, and Bs and their interlinkages relevant for the design of a healthy Human Reference Atlas with cross-walks to multiple existing ontologies that are revised and extended in the process.

ASCT+B flattened Master table construction

Table Set up:

1. To create a new ASCT+B table, make a copy of the standardized Google sheet flattened data template from here: https://docs.google.com/spreadsheets/d/1tK916JyG5ZSXW_cXfsyZnzXfjyoN-8B2GXLbYD6_vF0/edit?usp=sharing by clicking on the down arrow on the right side of the Google sheet tab (1) and selecting duplicate from the pop-up dialog box (2) (see picture).



2. Double click on the tab to change the name to organ, e.g., kidney, lung, spleen etc.

“About Table” Header Information

1. Revise the table authoring information in the first 10 rows which describe the contents:
 - a. **Provide organ name:**
“Anatomical Structures, Cell Types and Biomarkers Table for ***Organ Name*** “
 - b. **Provide Author information name and ORCID (lead author first, co-authors following) to row 3 of ASCT+B table:**
Author: Joe Doe (add up to 3 authors with ORCID IDs)
 - c. **Provide Reviewer name(s) to row 4 of ASCT+B table:**

- Jane Doe (no limit, will be listed in Acknowledgements)
- d. **Provide date started:**
Date Started: MM/DD/YYYY
 - e. **Provide date last modified:**
Last Modified: MM/DD/YYYY
 - f. **Provide version number:**
Version Number: v 0.0.1
 - g. **Provide a list of major publications:** Please provide links to major papers that provide scientific evidence for the nodes (AS, CT, B) and links (relationships between them) in the *part_of* AS tree and the AS-CT and CT-B bimodal networks.
 - h. **Provide an estimate of coverage of table for organ:** “It was estimated that roughly, this table covers about XX% of known AS, XX% of known CT, and XX% of known B.” Qualify how you arrived at this metric. In cases where published studies provide this evidence, please provide DOI.
 - i. Provide one (1) type of Functional Tissue Unit (FTU) for organ table and any accompanying references to support the following:
 - i. Provide FTU Name
 - ii. Provide FTU Function
 - iii. Provide FTU CTs
 - iv. Provide FTU Size & Shape
 - v. Provide FTU Typical total number of this FTU per one human body
 - vi. Provide FTU Key references (DOI)
3. The column headers are parsed and used by the [ASCT+B Reporter](#) to compute a visualization of the data therefore they must all conform to the format given below or they will not show in the ASCT+B reporter visualizations.

If you need to ADD several levels of depth of AS to the table, add 3 columns per AS and just change the number.

Example: AS/1 AS/1/LABEL AS/1/ID , AS/2 AS/2/LABEL AS/2/ID etc

(**n**: represents number in examples below)

AS/n: name given the anatomical structure by the subject matter expert (SME)

Example: heart

AS/n/LABEL: the name given the anatomical structure of in the ontology (this may not be exactly the same as that given by the SME.)

Example: heart

AS/n/ID: the unique ID given the anatomical structure in the ontology

Example: UBERON:0000948

The next AS columns would be “AS/2 AS/2/LABEL AS/2/ID” etc.

Adding content to Flattened ASCT+B master tables:

1. Add anatomical structures (AS) starting with the organ (i.e., heart, lung, kidney ...)

Please do not use abbreviations for the primary name of the AS, CT or B. Instead, write it out in long form.

If abbreviations are common for these terms, please add a column titled AS/1/ABBR substituting the number to match the AS number it is associated with:

Example: AS/1/ABBR

Example from Heart ASCT+B table for Anatomical structure abbreviation:

AS/2: left ventricle AS/2/ABBR: LV

Example from Kidney ASCT+B table for cell type abbreviation:

CT/1: Parietal epithelial cell, CT/1/ABBR: PEC

Example: (the arrows indicate the structure is part_of the higher order AS) nephron is *part_of* kidney. This is the language used to describe in ontological terms as well as anatomically.

kidney ← nephron ← renal corpuscle ← Bowman's (glomerular) capsule ← parietal epithelium

2. Add cell types (CT) (there will only be a single cell type column.)
To do this, because this is a flattened data table for machine readability, if an AS has four (4) cell types, then you will repeat all of the higher level structures up to the cell type column 4 times. The origin of cell types should be based on datasets with supporting literature.

Example: Last AS column set and CT columns for four rows

AS/4	CT/1
thick ascending limb of loop of Henle	Thick Ascending Limb Cell (general)
thick ascending limb of loop of Henle	Cortex-TAL cell
thick ascending limb of loop of Henle	Medulla-TAL cell
thick ascending limb of loop of Henle	Macula Densa cell

3. Add unique biomarkers (B) that define the CT for that organ system. This does NOT mean unique in relation to the whole body, just unique for the organ system that the ASCT+B master table covers. The origin of biomarkers that uniquely define cell types within a particular organ system should be based on datasets or supporting literature.

Biomarkers come in many different forms. The most common are single cell transcriptomics (gene), proteins (CODEX, FACS/Flow cytometry, proteomics (mass spectrometry)). The ASCT+B reporter v2.0 will support only gene and protein biomarkers.

Newer biomarkers are emerging such as lipid profiling, metabolic profiling, and proteoform detection. A first proteoform table exists for blood and bone marrow and will be supported in v2.1.

For gene markers, use column headers: BG/1, BG/1/LABEL, BG/1/ID

For protein markers, use column headers: BP/1 BP/1/LABEL BP/1/ID

c

4. Identify at least one major Functional Tissue Unit (FTU) per organ. Functional tissue units (FTUs) such as glomerulus in kidney or crypt in colon are subregions of tissue that serve a specific function; they are important for bridging the organ-level to single cell level.

Add 3 columns as before FTU/1 , FTU/1/LABEL , FTU/1/ID.

To the “FTU” column, indicate the presence of the FTU within each row of the ASCT+B hierarchy by adding a “1” if present or leaving the spreadsheet cell empty if it is absent

5. **Provide a list of major publications (REFs) to support the AS-AS, AS-CT, CT-B relationships captured in the ASCT+B master table.** Column headings: REF/1 REF/1/DOI REF/1/NOTES. Add additional columns as required following the the REF/#, REF/#/DOI, REF/#/NOTES column header format.

REF/1: provide the bibliographic reference Authors (Year) Title. Venue.

REF/1/DOI: unique Digital Object Identifier (DOI) for the journal article.

REF/1/NOTES: Indicate what evidence within the ASCT+B table the reference provides scientific evidence to support (i.e., the part_of AS tree and linkages to CTs but also the CT-B bimodal networks). Example citation for lung: Tables S1 and S2 in

<https://www.nature.com/articles/s41586-020-2922-4> for CT-B and AS-CT associations.

Adding Ontology terms and unique Identifiers (IDs)

1. [Uberon](#), [Foundational Model of Anatomy \(FMA\)](#), [Cell Ontology \(CL\)](#) and [Human Gene Ontology Nomenclature Committee \(HGNC\)](#) are the main ontologies used to match terms from ASCT+B tables. For Lipids, LipidMAPS unique identifier (LM-ID) are described here: https://www.lipidmaps.org/data/classification/lipid_cns.html
2. You may browse these ontologies via the links provided or download the .owl files and located at the links provided for each and search, browse, explore using an ontology editor like [Protégé](#).

3. The Elk reasoner and DL Query were used in the Protégé ontology editor to compile the *part_of* and *located_in* relationships found in Uberon and the Cell Ontology between anatomical structures, cell types, and organs in a Google sheet [here](#).
4. An example of the nested anatomical structures for *part_of* and *located_in* relationships found in Uberon and the Cell Ontology between anatomical structures, cell types and organs using the Elk reasoner and DL Query output for kidney can be found [here](#).
5. A spreadsheet containing a custom download from the [HGNC](#), which provides: HGNC ID, Approved symbol, Approved name, Status, Previous symbols, Alias symbols, Chromosome, Accession numbers, RefSeq IDs, UniProt ID(supplied by UniProt), Ensembl ID(supplied by Ensembl), RefSeq(supplied by NCBI), NCBI Gene ID(supplied by NCBI), and Pubmed IDs for supporting reference literature can be found [here](#).
6. For those entities without any Uberon, FMA, or CL ontology match, fill only the columns for AS/#, CT/#, where # is the appropriate number in the hierarchy of AS or CT.
 - a. Leave blank columns with headers: AS/#/LABEL or CT/#/LABEL (these are names found in ontologies, specifically the “rdfs:label”).
 - b. Leave blank columns with headers: AS/#/ID or CT/#/ID (these are the ontology IDs usually in format Uberon:nnnn, CL:nnnn or FMA:nnnn where n=number).

ASCT+B Review (both amongst SMEs and external to SMEs)

1. Authors of ASCT+B master tables should invite external experts to review and verify AS-AS, AS-CT, CT-B relationships and ontology term choices and definitions.
2. Authors may reach out to MC-IU for assistance via HuBMAP CCF Info infoccf@iu.edu.

ASCT+B Validation (external to SMEs)

1. David Osumi-Sutherland, with input from CNS Research Analyst Ellen Quardokus has written a Python script for using new ASCT+B table format. The script takes an ASCT+B table and compares relationships captured within it to Uberon triplestore database to validate existing *part_of*, *located_in*, *has_part* relationships that appear within current build of ontologies and generates a report showing conflicts.
2. The CNS Research Analyst reviews this report and identifies missing relationships and mistakes in ASCT+B tables.
3. The CNS Research Analyst shares the output from the validation script with the SME author(s) for consultation and revision as needed.

ASCT+B Revision (both amongst SMEs and external to SMEs)

1. Authors edit as needed after review and validation against ontologies.
2. Authors edit their organ system Google sheet as needed, making certain to update the header to reflect updates. To facilitate visualizing the relationships in the ASCT+B master tables, the [ASCT+B reporter web application](#) may be used.
3. Authors inform MC-IU via HuBMAP CCF Info infoccf@iu.edu of updates and problems.

ASCT+B Approval (criteria from SMEs and internal to MC-IU)

Approval of ASCT+B tables

1. Authoring SMEs identify other independent experts that can review and approve the table for publication.
2. MC-IU approves that the table meets ASCT+B Reporter requirements--including listing of ORCID IDs for all authors, reviewers, and other experts that contributed to table construction.

ASCT+B Publishing (GitHub vs major scholarly journal)

Publishing will take three forms.

- 1) ASCT+B master tables that have been validated for *part_of* relationships, reviewed and approved are uploaded to the [CCF GitHub repository](#) for versioning. In Y4 of HuBMAP, MASTER tables will be published as part of CCF.OWL on BioPortal.
- 2) The original eleven master tables for kidney, spleen, thymus, lymph node, lung, large intestine, skin, brain, blood & bone marrow, heart and vasculature will be submitted for publication in *Nature Cell Biology*.
- 3) ASCT+B table information will be used to incorporate new terms and relationships into existing ontologies (e.g., Uberon and Cell Ontology) as appropriate.

References and Definitions

References

The below references and definitions were used in writing this Standard Operating Procedure. When available, definitions were taken from the [HuBMAP Dictionary](#), and aligned with standard terminologies used in relevant fields.

Refer to HuBMAP CCF Portal <https://hubmapconsortium.github.io/ccf/index.html>

Refer to HuBMAP 3D organ reference models:
<https://github.com/hubmapconsortium/ccf-3d-reference-object-library>

Refer to HuBMAP CCF ASCT+B Reporter
<https://github.com/hubmapconsortium/ccf-asct-reporter/blob/main/README.md>
<https://hubmapconsortium.github.io/ccf-asct-reporter/>

Bono, Bernard de, Pierre Grenon, Richard Baldock, and Peter Hunter. 2013. "Functional Tissue Units and Their Primary Tissue Motifs in Multi-Scale Physiology." *Journal of Biomedical Semantics* 4 (1): 22–22. <https://doi.org/10.1186/2041-1480-4-22>.

Mungall, C. J.; Torniai, C.; Gkoutos, G. V.; Lewis, S. E.; Haendel, M. A. Uberon, an Integrative Multi-Species Anatomy Ontology. *Genome Biol* 2012, 13 (1), R5. <https://doi.org/10.1186/gb-2012-13-1-r5>.

Haendel, M. A.; Balhoff, J. P.; Bastian, F. B.; Blackburn, D. C.; Blake, J. A.; Bradford, Y.; Comte, A.; Dahdul, W. M.; Dececchi, T. A.; Druzinsky, R. E.; Hayamizu, T. F.; Ibrahim, N.; Lewis, S. E.; Mabee, P. M.; Niknejad, A.; Robinson-Rechavi, M.; Sereno, P. C.; Mungall, C. J. Unification of Multi-Species Vertebrate Anatomy Ontologies for Comparative Biology in Uberon. *J Biomed Semantics* 2014, 5, 21. <https://doi.org/10.1186/2041-1480-5-21>.

Elhanan, G.; Ochs, C.; Mejino, J. L. V.; Liu, H.; Mungall, C. J.; Perl, Y. From SNOMED CT to Uberon: Transferability of Evaluation Methodology between Similarly Structured Ontologies. *Artif Intell Med* 2017, 79, 9–14. <https://doi.org/10.1016/j.artmed.2017.05.002>.

Meehan, T. F.; Masci, A. M.; Abdulla, A.; Cowell, L. G.; Blake, J. A.; Mungall, C. J.; Diehl, A. D. Logical Development of the Cell Ontology. *BMC Bioinformatics* 2011, 12 (1), 6. <https://doi.org/10.1186/1471-2105-12-6>.

Diehl, A. D.; Augustine, A. D.; Blake, J. A.; Cowell, L. G.; Gold, E. S.; Gondré-Lewis, T. A.; Masci, A. M.; Meehan, T. F.; Morel, P. A.; Nijnik, A.; Peters, B.; Pulendran, B.; Scheuermann, R. H.; Yao, Q. A.; Zand, M. S.; Mungall, C. J. Hematopoietic Cell Types: Prototype for a Revised Cell Ontology. *Journal of Biomedical Informatics* 2011, 44 (1), 75–79. <https://doi.org/10.1016/j.jbi.2010.01.006>.

Courtot, M.; Meskas, J.; Diehl, A. D.; Droumeva, R.; Gottardo, R.; Jalali, A.; Taghiyar, M. J.; Maecker, H. T.; McCoy, J. P.; Ruttenberg, A.; Scheuermann, R. H.; Brinkman, R. R. FlowCL: Ontology-Based Cell Population Labelling in Flow Cytometry. *Bioinformatics* 2014, 31 (8), 1337–1339. <https://doi.org/10.1093/bioinformatics/btu807>.

Bernstein, M. N.; Ma, Z.; Gleicher, M.; Dewey, C. N. CellO: Comprehensive and Hierarchical Cell Type Classification of Human Cells with the Cell Ontology. *iScience* 2021, 24 (1), 101913. <https://doi.org/10.1016/j.isci.2020.101913>.

Travaglini, K. J.; Nabhan, A. N.; Penland, L.; Sinha, R.; Gillich, A.; Sit, R. V.; Chang, S.; Conley, S. D.; Mori, Y.; Seita, J.; Berry, G. J.; Shrager, J. B.; Metzger, R. J.; Kuo, C. S.; Neff, N.; Weissman, I. L.; Quake, S. R.; Krasnow, M. A. A Molecular Cell Atlas of the Human Lung from Single-Cell RNA Sequencing. *Nature* 2020. <https://doi.org/10.1038/s41586-020-2922-4>.

Golbreich, C.; Grosjean, J.; Darmoni, S. J. The Foundational Model of Anatomy in OWL 2 and Its Use. *Artificial Intelligence in Medicine* 2013, 57 (2), 119–132.

<https://doi.org/10.1016/j.artmed.2012.11.002>.

Rosse, C.; Mejino, J. L. V. A Reference Ontology for Biomedical Informatics: The Foundational Model of Anatomy. *Journal of Biomedical Informatics* 2003, 36 (6), 478–500.

<https://doi.org/10.1016/j.jbi.2003.11.007>.

Musen, M. A.; Protégé Team. The Protégé Project: A Look Back and a Look Forward. *AI matters* 2015, 1 (4), 4–12. <https://doi.org/10.1145/2757001.2757003>.

Jackson, R. C.; Balhoff, J. P.; Douglass, E.; Harris, N. L.; Mungall, C. J.; Overton, J. A. ROBOT: A Tool for Automating Ontology Workflows. *BMC Bioinformatics* 2019, 20 (1), 407.

<https://doi.org/10.1186/s12859-019-3002-3>.

Smith, L. M.; Kelleher, N. L. Proteoforms as the next Proteomics Currency. *Science* 2018, 359 (6380), 1106. <https://doi.org/10.1126/science.aat1884>.

Fahy, E.; Subramaniam, S.; Murphy, R. C.; Nishijima, M.; Raetz, C. R. H.; Shimizu, T.; Spener, F.; van Meer, G.; Wakelam, M. J. O.; Dennis, E. A. Update of the LIPID MAPS Comprehensive Classification System for Lipids. *J Lipid Res* 2009, 50 Suppl (Suppl), S9-14.

<https://doi.org/10.1194/jlr.R800095-JLR200>.

Fahy, E.; Subramaniam, S.; Brown, H. A.; Glass, C. K.; Merrill, A. H. J.; Murphy, R. C.; Raetz, C. R. H.; Russell, D. W.; Seyama, Y.; Shaw, W.; Shimizu, T.; Spener, F.; van Meer, G.; VanNieuwenhze, M. S.; White, S. H.; Witztum, J. L.; Dennis, E. A. A Comprehensive Classification System for Lipids. *J Lipid Res* 2005, 46 (5), 839–861. <https://doi.org/10.1194/jlr.E400004-JLR200>.

Glossary

ASCT+B Tables: refer to [HuBMAP Dictionary](#)

ASCT+B MASTER tables: The first ASCT+B table for an organ system to which other ASCT+B-like information may be compared with using the ASCT+B Reporter compare feature.

ASCT+B Reporter: refer to [HuBMAP Dictionary](#)

CO-Detection by indEXing (CODEX): refer to [HuBMAP Portal Assay Documentation](#)

Digital Object Identifier (DOI): refer to [HuBMAP Dictionary](#)

EUI (Exploration User Interface): refer to [HuBMAP Dictionary](#)

FACS: Fluorescence-activated cell sorting uses fluorescently tagged antibodies to cell surface markers most commonly to sort populations of cells into different containers based on markers present.

Flow cytometry: similar to FACS without sorting cells into containers, just visually represents cell populations based on fluorescent marker presence or absence.

Flattened data table: database that stores data in a plain text file where each row of the spreadsheet file holds one record, with fields separated in columns.

Registration User Interface (RUI): refer to [HuBMAP Dictionary](#)

Functional Tissue Unit (FTU): unit consists of a three-dimensional block of cells centered around a capillary, such that each cell in this block is within diffusion distance from any other cell in the same block

Open Researcher and Contributor ID (ORCID): <https://orcid.org/> unique identifier number researchers own and control that connects the iD with professional information – affiliations, grants, publications, etc and can be used during the publication process.

Ontology: Examples: Uberon, Foundational Model of Anatomy, Cell Ontology

Principal Investigator (PI): Person who is responsible for the overall conduct of the research at a given site.

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

Rdfs:label: [Resource Description Framework Schema \(RDFS\)](#) label is a human-readable version of a resource's name and is a standard for data interchange, developed and agreed upon by World Wide Web Consortium (W3C) and used for the preferred name in an ontology.

ROBOT: an ontology tool for automating ontology workflows <http://robot.obolibrary.org/>

Standard Operating Procedures (SOPs): SOPs are issued to specifically instruct team members in areas of responsibility, procedural steps, appropriate specifications and required records. SOPs outline procedures, which must be followed to support the reproducibility of scientific research. Procedures can take the form of a narrative, a flow chart, a process map, computer screen

HuBMAP

printouts or combination of all or any other suitable form, however must be written in appropriate, effective grammatical style (e.g. plain English).

Web Ontology Language (.OWL) file format: refer to <https://www.w3.org/OWL/> W3C Web Ontology Language (OWL) is a Semantic Web language.