ASCT+B Generator

Stephen A. Fisher[1§], Kathleen O'Neill[2], Junhyong Kim[1]

[1]Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA; [2]Department of Obstetrics and Gynecology, Hospital of the University of Pennsylvania, Philadelphia, PA 19104, USA.
[§]Correspondence to: safisher@sas.upenn.edu

Introduction

One of our year 1 goals is to generate ASCT+B tables for ovaries, uterus and the Fallopian tube. As these tables contain nested anatomical structures with any number of cells per structure, the number of records required to document an organ can be extensive. The Fallopian tube appears relatively simple with only 5 layers of nested anatomical structures captured in just over 100 records, with each record containing about 48 fields. By comparison the ovary contains 10 layers of nested anatomical structures requiring over 860 records with 70 fields per record, to document. These requirements will also continue to grow as we further detail the organs.

Our initial, manual attempt to document these organs in spreadsheets, while heroic, ultimately succumbed to numerous, hard to find copy-paste errors, typos and was simply a dizzying management task. In turn we created the ASCT+B Generator which constructs ASCT+B tables from much simpler, structured files, greatly reducing the complexity of the data needing to be maintained, while also significantly reducing the risk of human errors when creating the large, more complicated ASCT+B table. Here we describe the ASCT+B Generator input and limitations. The ASCT+B Generator code and a very basic example can be found on GitHub.

https://github.com/kimpenn/asct-b-generator

Input

Input

While the input to the ASCT+B Generator is still a data table, for the organs we've tested, it's a much smaller and less complicated table. Specifically, the table for the ovary is about 2% of the size of the actual ovary ASCT+B table. For the Fallopian tube the input table is about 25% the size of the Fallopian tube ASCT+B table.

The input table contains a set of all features used to describe an organ. Each table row describes a single feature. Features can be anatomical structures, cell types, biomarkers (e.g., genes, proteins, lipids, or metabolites), or references. Hence the number of rows is the number of features. For each feature we collect specific information, such as the label and ID as well as whether or not the feature has any "children" (i.e., sub structures or cell type descriptors). Specifically each row should contain twelve columns (see "Input Fields").

The Input File Excerpt below exemplifies some of the field values and usage.

Input Fields

- ***Name*** (Reference DOI); ***Label*** (Reference details); ***ID*** (Reference notes)

  These are the same values you'd expect to find in an ASCT+B table. For references, these fields should contain the DOI, details and notes, respectively.

- ***Type***

  The type needs to be "AS" for anatomical structures and "CT" for cell types. It doesn't matter what type values are used for the other items, so long as it's not either AS or CT.

- ***Children***

  Children is a comma separated list of structures and cell types that exist within this structure. For example, in the excerpt below the "hilum of ovary" only has a single child which is a cell type, where as the "medial ovary" has two child structures. Cell types and anatomical structures can be intermixed in the children field.

- ***Genes***; ***Proteins***; ***Proteoforms***; ***Lipids***; ***Metabolites***; ***FTUs***; ***References***

  Each of these fields is a comma separated list of features.

Usage

Command arguments

- ***name_of_top_level_entity*** - typically the name of the organ
- ***number_of_anatomical_structure_levels*** - the total number of AS levels. This can be an overestimate.
- ***input_TSV_file*** - the managed tab delimited input file containing the data to be transformed into the ASCT+B table.
- ***output_TSV_file*** - the resulting ASCT+B table output as a tab delimited file.

Example command usage

process.py organ 10 ovary-v1.txt ovary-ASCTB.xls

Issues & Limitations

- Anatomical structures can't be assigned features (e.g., genes, proteins, etc.) if they include child structures or cell types. Features such as genes and proteins are only applied to the lowest level of anatomical structures and to cell types.
- The user needs to know the number of levels of anatomical structures for the organ or at least over estimate the number.
- The output file doesn't include a header line.

Input File Excerpt

NAME (REF DOI)	LABEL (REF DETAILS)	ID (REF NOTES)	TYPE	CHILDREN	GENES	PROTEINS	PROTEOFORMS	LIPIDS	METABOLITES	FTU	REFERENCES (NAME/DOI)
ovary		UBERON:0000992	AS	central ovary, lateral ovary, medial ovary, mesovarium, ovarian ligament, hilum of ovary							
central ovary			AS	central inferior ovary, central superior ovary							
lateral ovary			AS	lateral inferior ovary, lateral superior ovary							
medial ovary			AS	medial inferior ovary, medial superior ovary							
ovarian ligament		UBERON:0008847	AS								
hilum of ovary			AS	hilar cell							
corona radiata		CL:0000713	CT								doi:10.1093/oxfordjournals.humrep.a136365
hilar cell		CL:0002095	CT			alkaline phosphatase, acid phosphatase, non-specific esterase, inhibin, calretinin, melan-A, cholesterol esters					McKay et al 1961, Boss et al 1965, Mills et al 2020, Jungbluth et al 1998, Pelkey et al 1998
primary oocyte		CL:0000654	CT								doi:10.1093/oxfordjournals.humrep.a136365
secondary oocyte		CL:0000655	CT								doi:10.1093/oxfordjournals.humrep.a136365
columnar ovarian surface epithelial columnar cell			CT			calretinin, mesothelin					Mills et al 2020, Reeves et al 1971, Hummitzsch et al 2013, Blaustein et al 1979, McKay et al 1961
flattened cuboidal ovarian surface epithelial cell			CT			oviduct-specific glycoprotein-1, E-cadherin					Mills et al 2020, Reeves et al 1971, Hummitzsch et al 2013, Blaustein et al 1979, McKay et al 1961
oviduct-specific glycoprotein-1			Protein								
mesothelin			Protein								
E-cadherin			Protein								
doi:10.1093/oxfordjournals.humrep.a136365	PMID: 3558758		Reference								
McKay et al 1961	McKay, D., Pinkerton, J., Hertig, A. & Danziger, S. (1961). The Adult Human Ovary: A Histochemical Study. Obstetrics & Gynecology, 18(1), 13-39.		Reference								