SCB - dataset metrics notes

Amanda Infeld
August 23, 2021

Overall metrics

Collections
- Only includes collections currently available on the cellxgene platform

Datasets
- Only includes datasets currently available on the cellxgene platform

Cells
- In the API, all datasets added in 2020 do not report cell counts
  - The cell counts for these datasets were added manually

Datasets by characteristic

Overall cleaning
- Combined the same capitalized and non-capitalized versions (e.g. 'Blood' and 'blood' were combined)

Assay

Original to grouped translation table

Original assay name	Grouped assay name
10x	10x RNA-seq
10x 3' v2	10x RNA-seq
10x 3' v3	10x RNA-seq
10x 5' v1	10x RNA-seq
10x 5' v3	10x RNA-seq
10x v2	10x RNA-seq
10x v3	10x RNA-seq
ATAC 10x v1	ATAC-seq
CITE-seq	CITE-seq

Drop-seq	Other RNA-seq
MERFISH	Spatial gene expression
Smart-seq	SS2
Smart-seq2	SS2
Smart-seq2 protocol	SS2
Visium Spatial Gene Expression	Spatial gene expression
microwell-seq	Other RNA-seq
scATAC-seq	ATAC-seq
scRNA-seq	Other RNA-seq
sci-RNA-seq	Other RNA-seq
sci-plex	Other RNA-seq
seq-Well	Other RNA-seq

Ethnicity
- One dataset labeled ethnicity as 'male' only, this label was removed and replaced with 'unknown'

Original to grouped translation table

Original ethnicity	Grouped ethnicity
unknown	unknown
na	non-human
European	European
African American	African American
Asian	Asian
Hispanic or Latin American	Hispanic or Latin American
East asian	Asian
Chinese	Asian

Finnish	European
Han Chinese	Asian

Developmental stage
- It is unclear what age ranges are included in the developmental stage categories that do not include a specific age

Adjustments

Original developmental stage	Updated developmental stage
developmental stage	unknown
human adult stage	adult
contains 'post-fertilization'	fetal stage

Tissue
- Grouped together tissues that had the tissue name and the tissue name + '(cell culture)' (e.g. 'bone marrow' and 'bone marrow (cell culture)' were combined)