Custom Module: Get Sample Metadata

Description: Get a downloadable csv(s) of sample metadata.

Setup:

|  |  |  |  |
| --- | --- | --- | --- |
| R version | CPU cores | RAM (GB) | Max run (h) |
| Any | 16 | 128+ | 4+ |

\* RAM and Max run time listed above are suggested minimums. Values may be increased (note that increasing RAM or Max run time may increase costs). If the module exceeds either value, it will immediately fail.

Variables:

Name must be spelled exactly as shown for script to work.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Type | Name | Display name | Range | Default | Required |
| BOOLEAN | metaSummary | Metadata Column Summary |  | 0 | Yes |
| BOOLEAN | cellLevel | Cell-Level Metadata |  | 1 | Yes |
| BOOLEAN | fovLevel | FOV-Level Metadata |  | 0 | Yes |
| BOOLEAN | flowcellLevel | Flow Cell-Level Metadata |  | 0 | Yes |
| BOOLEAN | indexOnly | Index Columns Only (for the selected metadata) |  | 1 | Yes |
| BOOLEAN | qcFlagsOnly | QC Flags & Index Columns (for the selected metadata) |  | 0 | Yes |
| BOOLEAN | fullDataset | Full Metadata Array (for the selected metadata) |  | 0 | Yes |

Packages: None

Outputs: Downloadable and editable .csv file(s) of the sample metadata. Metadata can be filtered down to FOV or slide level annotations. Downloadable summary of metadata columns: column names, class, example values.

Notes: The module will give a warning if the produced file is above 600 Mb. The full cell metadata file is deleted and this warning is printed:

“Zip folder is too large for local download, please download cell metadata using the Flat File Export custom module or button"

If the resulting folder after deleting the full cell metadata is less than 600 mb, the other files will be available for download. But if this warning is produced even when not downloading the full cell metadata, split the download into more manageable pieces.

When editing the file to upload using the UpdateSampleMetadata custom module:

* ONLY ADD columns to the metadata file
* DO NOT EDIT rows from metadata file, adding or removing
* DO NOT EDIT column names in the metadata file
* DO NOT EDIT the existing data. If you need to, just add new annotations because there will be NO record of the previous data
* New column names duplicated with the existing metadata will NOT be added. Confirm the column names are unique using the metaSummary results.

Metadata Summary:

The metadata summary .tsv file provides a summary of each column in the cell metadata.

1. Column Name
2. Column Class
3. Number of unique values if a character, factor, or boolean. Empty if numeric.
4. Summary of column values
   1. Character, factor, boolean = first 40 values
      1. valueName1:valueCounts1,valueName2:valueCounts2,etc
   2. Fully unique values = first 20 values
   3. Numeric = boxplot values: min, 1stQuantile, median, mean, 3rdQuantile, max

A screenshot of a computer

Description automatically generated

Metadata column descriptions:

Many column names out of AtoMx will have Globally Unique Identifier (GUID) attached to the column name. This allows multiple pipelines to be run on the same dataset without collisions. Below is a list of the general column names, your column names might not match these exactly and it’s ok if not all columns shown here are present in your dataset.

Column names can contain their data object slot names. Any mention of RNA in a column name is indicating the slot and not the analyte. The RNA slot contains the non-control target counts regardless of analyte type.

|  |  |
| --- | --- |
| All Datasets | |
| Column Name/Prefix | Description |
| fov | Field Of View (FOV) the cell is in |
| Area | Number of pixels assigned to a given cell |
| AspectRatio | Width divided by height |
| x\_FOV\_px | x position of the cell center within the FOV, measured in pixels |
| y\_FOV\_px | y position of the cell center within the FOV, measured in pixels |
| Width | Cell’s maximum length in x dimension (pixels) |
| Height | Cell’s maximum length in y dimension (pixels) |
| Mean.PanCK | Mean fluorescence intensity within a given cell |
| Max.PanCK | Max fluorescence intensity within a given cell |
| cell\_id | Unique identifier for a single cell. Created using c\_[slide]\_[fov]\_[cell] |
| assay\_type | "RNA" or "protein" cell |
| version | Target decoding version used |
| Run\_Tissue\_name | Flowcell name |
| Panel | Panel used |
| slide\_ID\_numeric | SlideID, number is determined from |
| x\_slide\_mm | x position of the cell center within the FOV, measured in pixels |
| y\_slide\_mm | y position of the cell center within the FOV, measured in pixels |
| cell\_ID | Unique identifier for a single cell. Created using c\_[slide]\_[fov]\_[cell] |
| median\_RNA | Median number of RNA counts per FOV |
| RNA\_quantile\_0.75 | 75th quantile RNA counts per FOV |
| RNA\_quantile\_0.8 | 80th quantile RNA counts per FOV |
| RNA\_quantile\_0.85 | 85th quantile RNA counts per FOV |
| RNA\_quantile\_0.9 | 90th quantile RNA counts per FOV |
| RNA\_quantile\_0.95 | 95th quantile RNA counts per FOV |
| RNA\_quantile\_0.99 | 99th quantile RNA counts per FOV |
| nCount\_RNA | Number of RNA counts |
| nFeature\_RNA | Number of unique RNA targets |
| median\_negprobes | Median number of Negative counts per FOV |
| negprobes\_quantile\_0.75 | 75th quantile Negative counts per FOV |
| negprobes\_quantile\_0.8 | 80th quantile Negative counts per FOV |
| negprobes\_quantile\_0.85 | 85th quantile Negative counts per FOV |
| negprobes\_quantile\_0.9 | 90th quantile Negative counts per FOV |
| negprobes\_quantile\_0.95 | 95th quantile Negative counts per FOV |
| negprobes\_quantile\_0.99 | 99th quantile Negative counts per FOV |
| nCount\_negprobes | Number of Negative counts |
| nFeature\_negprobes | Number of unique Negative targets |
| Area.um2 | Area of cell in um |
| cellSegmentationSetId | GUID indicating which segmentation was run |
| cellSegmentationSetName | Name of segmentation run |
| SplitRatioToLocal | Cell is split across multiple FOVs. 0 = not split, 0-1 = small split in current FOV, >1 = larger split in current FOV |
| NucArea | Number of pixels assigned to the nucleus |
| NucAspectRatio | Nucleus width divided by height |
| Circularity | Cell area to perimeter ratio |
| Eccentricity | Ratio of the major axis to the minor axis of a cell |
| Perimeter | Number of pixels in a cell’s boundary |
| Solidity | Density of a cell, detects holes or irregular boundaries. 1 indicates perfect cell with defects decreasing the value |

For more info on cell morphology metrics like Circularity, Eccentricity, Perimeter, Solidity please see the [BIDCELL paper](https://www.nature.com/articles/s41467-023-44560-w).

|  |  |
| --- | --- |
| RNA Specific Columns | |
| Column Name/Prefix | Description |
| median\_falsecode | Median number of False Code counts per FOV |
| falsecode\_quantile\_0.75 | 75th quantile False Code counts per FOV |
| falsecode\_quantile\_0.8 | 80th quantile False Code counts per FOV |
| falsecode\_quantile\_0.85 | 85th quantile False Code counts per FOV |
| falsecode\_quantile\_0.9 | 90th quantile False Code counts per FOV |
| falsecode\_quantile\_0.95 | 95th quantile False Code counts per FOV |
| falsecode\_quantile\_0.99 | 99th quantile False Code counts per FOV |
| nCount\_falsecode | Number of False Code counts |
| nFeature\_falsecode | Number of unique False Code targets |
| unassignedTranscripts | % if unassigned transcripts within an FOV |
| propNegative | Proportion of counts from negative probes |
| complexity | Proportion of counts from different targets nCount/nFeature |
| errorCtEstimate | Estimation of # of counts from genes that arise from background |
| percOfDataFromError | Percent of counts from genes that arise from background |
| qcFlagsCellCounts | Flag for cells that pass minimum number of counts threshold |
| qcFlagsCellPropNeg | Flag for cells that pass maximum proportion of negative counts |
| qcFlagsCellComplex | Flag for cells that pass minimum complexity threshold |
| qcFlagsCellArea | Flag for cells that pass minimum area threshold |
| qcCellsFlagged | Flag for cells that pass all cell QC checks |
| nCell | # cells per FOV |
| nCount | # RNA counts per FOV |
| nCountPerCell | Average RNA counts per FOV |
| nFeaturePerCell | Average RNA targets expressed per FOV |
| propNegativeCellAvg | Average propNegative per FOV |
| complexityCellAvg | Average complexity per FOV |
| errorCtPerCellEstimate | Average background counts per FOV |
| percOfDataFromErrorPerCell | Average percent background counts per FOV |
| qcFlagsFOV | Flag for FOVs that pass all FOV QC checks |
| RNA\_nbclust\_[GUID]\_clusters | InsituType cell type |
| RNA\_nbclust\_[GUID]\_posterior\_probability | InsituType probability of cell type call |
| nn\_[GUID]\_cluster\_cluster\_[GUID] | Leiden cell cluster |
| RNA\_spatialclust\_[GUID]\_neighbours\_# | # of neighbors each cell has of a specific cell type |
| spatialclust\_[GUID]\_assignments | Cell Niche |
|  |  |

|  |  |
| --- | --- |
| Protein Specific Columns | |
| Column Name/Prefix | Description |
| area.qc | Flag for cells that pass Area QC |
| mean.neg | Average negative probe count |
| negprobe.qc | Flag for cells that pass maximum number of negative counts threshold |
| n\_high\_quant | Number of targets considered high expressors |
| high.express.qc | Flag for cells that pass maximum number of high expressor threshold |
| n\_low\_quant | Number of targets considered low expressors |
| low.express.qc | Flag for cells that pass maximum number of low expressor threshold |
| remove\_flagged\_cells | Flag for FOVs that pass all protein QC checks |
| final\_cell\_type\_assignment | Protein cell type |